Méthodes de démélange non-linéaires pour l'imagerie hyperspectrale

Nguyen Nguyen Hoang

UNIVERSITÉ DE NICE SOPHIA-ANTIPOLIS

ÉCOLE DOCTORALE

SCIENCES FONDAMENTALES ET APPLIQUEES

THÈSE

pour obtenir le grade de

Docteur de l'Université de Nice Sophia-Antipolis

Mention : Sciences de l'univers

Présentée et soutenue par

Nguyen Hoang Nguyen

Méthodes de démélange non-linéaires pour l'imagerie hyperspectrale

soutenue le 3 décembre 2013

Jury :

Rapporteurs :	Emmanuel Duflos	Ecole Centrale de Lille
	Jean-Yves Tourneret	Institut National Polytechnique de Toulouse
Examinateurs :	Jocelyn Chanussot	Grenoble INP
	Pierre Borgnat	Ecole Normale Supérieure de Lyon
	Paul Honeine	Université de Technologie de Troyes
Directeurs :	Cédric Richard	Université de Nice Sophia-Antipolis
	Céline Theys	Université de Nice Sophia-Antipolis

Table des matières

Table des figures

Liste des tableaux

Liste des Abréviations

$\boldsymbol{A}$	Matrice des abondances de taille $L \times N$
$\boldsymbol{\alpha}$	Vecteur d'abondance du pixel-vecteur observé
$\boldsymbol{I}_N$	Matrice identité de taille $(N \times N)$
$\boldsymbol{m}_{\lambda_\ell}^\top$	ℓ-ème ligne de $\boldsymbol{M}^T$, soit le vecteur de R signatures spectrales des endmembers dans la ℓ-ème bande spectrale
$\boldsymbol{m}_i$	Vecteur spectral d'un matériau
$\boldsymbol{R}$	Matrice des pixels de taille $L \times N$
$\boldsymbol{V}_{\mathcal{S}}$	Volume d'un simplexe $\mathcal{S}$
$\boldsymbol{x}$	Vecteur des données entrées
$\boldsymbol{y}$	Vecteur des donnée sortie
$\boldsymbol{n}$	Vecteur de bruit
$\boldsymbol{r}$	Pixel-vecteur observé
$\mathcal{H}$	Espace de Hilbert à noyau reproduisant
$\mathcal{X}$	Espace de données entrées (Cas général)
$\kappa(\cdot,\cdot)$	Noyau définissant l'espace $\mathcal{H}$
$\mathbf{K}$	Matrice de Gram correspondant au noyau $\kappa(\cdot,\cdot)$
$\mathbb{R}$	Ensemble des réels
$\mathbb{R}^+$	Ensemble des réels non-négatifs
$\mathbb{R}^d$	Espace réel de dimension d
$\boldsymbol{M}$	Matrice dont les colonnes sont les spectres des matériaux
L	Nombre de bandes spectrales
N	Nombre de pixel-vecteurs observés
R	Nombre des composés purs ou endmembers
y_i	Sortie correspondant à l'entrée $\boldsymbol{x}_i$
FBR	Fonction de base radiale
FCLS	Fully constrained least-squares
ISOMAP	isometric feature mapping
KPCA	Kernel principal component Analysis

LLE Local Linear Embedding

MDS Multi dimensional scaling

MKL Multi-kernel Learning - Apprentissage à noyau multiple

MNF Minimum Noise Fraction

MSE Erreur quadratique moyenne

NMF Factorisation en matrices non-négatives

PBSI Partially-linear based system identification

PCA Principal component Analysis

PNMM Post non-linear mixing model - modèle de mélange post non-linéaire

RKHS Espace de Hilbert à noyau reproduisant

RMSE Root-mean-square error

SGA Simplexe Growing Algorithm

SVD Singular values decomposition

SVM Séparateur à vaste marge

VCA Vertex component analysis

VD Dimension virtuelle

Introduction

La présente thèse de doctorat s'inscrit dans le cadre de l'exploitation des grandes masses de données issues de l'imagerie hyperspectrale. Ce domaine connaît actuellement un développement rapide dans de nombreux domaines des sciences de l'observation, de la résolution microscopique aux échelles astronomiques, dans des cadres aussi bien civils que militaires. Parmi les questions ouvertes, celle de la segmentation des cubes de données, qui vise à isoler les objets d'intérêt qui y sont présents, est tout à fait centrale et constitue un véritable verrou technologique. La superposition spatiale des zones d'intérêt conduit à s'intéresser au problème de démélange où, dans un contexte non supervisé, il s'agit d'estimer le nombre de composés de référence représentés dans un ensemble de vecteurs hyperspectraux, ainsi que la signature spectrale et la fraction d'abondance de chacun dans le mélange [Keshava & Mustard 2002]. Ce travail porte sur le problème de démélange des vecteurs hyperspectraux dans un cadre supervisé, où l'on suppose que les composés de référence ont été déterminés *a priori* en utilisant des approches d'extraction appropriées. Le lecteur intéressé est invité à consulter [Winter 1999, Nascimento & Bioucas-Dias 2005, Boardman 1993]. Dans ces conditions, le problème de démélange se résume à l'estimation des fractions d'abondance de chaque composé de référence en chaque pixel de l'image.

L'estimation des abondances est généralement envisagée sur la base d'un modèle de mélange linéaire. Différentes méthodes sont envisagées dans [Heinz & Chang 2001, Dobigeon *et al.* 2009, Theys *et al.* 2009, Honeine & Richard 2011a]. Par exemple, la méthode FCLS présentée dans [Heinz & Chang 2001] estime les abondances en minimisant un coût quadratique sous contrainte de positivité et de somme unité. La stratégie géométrique décrite dans [Honeine & Richard 2011a] vise à calculer les ratios de volumes de polyèdres dans l'espace engendré par les pixel-vecteurs hyperspectraux. Le principal avantage de la première est la convexité du problème d'optimisation. Un très faible coût de calcul caractérise la seconde.

Les interactions entre les matériaux peuvent générer des phénomènes non-linéaires qui devraient être pris en compte dans le calcul des abondances, au risque sinon de fausser celui-ci. Des modèles non-linéaires ont été récemment introduits dans la littérature afin de tenir compte de ces effets. Citons par exemple le modèle intime [Hapke 1981], et des approximations telles que le modèle bilinéaire généralisé [Halimi *et al.* 2011a]. Les méthodes de démélange non-linéaire tentent d'inverser ces modèles et d'estimer les abondances. Dans [Halimi *et al.* 2011a], un algorithme de déconvolution non-linéaire pour le modèle de mélange bilinéaire généralisé est proposé, et constitue la base de nombreux travaux. Basée sur l'inférence bayésienne, cette méthode présente toutefois une complexité calculatoire élevée et est consacrée exclusivement au modèle bilinéaire généralisé. Dans [Raksuntorn & Du 2010, Nascimento & Bioucas-Dias 2009], les auteurs envisagent de considérer des modèles de mélange plus généraux en générant des spectres de

référence artificiels par croisement des signatures de composés purs afin de prendre en compte les éventuels effets de diffusion de la lumière sur les différents matériaux. Cependant, il n'est pas aisé d'identifier les termes croisés qu'il conviendrait de générer. Par ailleurs, un croisement de l'ensemble des composés purs identifiés conduirait sans nul doute à une explosion de la taille du problème. Une autre stratégie possible consiste à utiliser des méthodes d'apprentissage de variétés, telles que Isomap [de Silva & Tenenbaum 2003] et LLE [Roweis & Saul 2000], afin d'y résoudre un problème de démélange linéaire. Enfin, dans [Chen *et al.* 2013], les auteurs formulent un nouveau paradigme basé sur des noyaux reproduisants. Le modèle repose sur un processus de mélange linéaire, complété d'interactions non-linéaires exprimées dans un espace de Hilbert à noyau reproduisant. Cette famille de modèles présente une interprétation physique claire et permet de prendre en compte des interactions complexes entre les composés de référence.

Cette thèse vise à explorer de nouvelles voies pour le démélange non-linéaire des données hyperspectrales. Tout d'abord, en considérant le cas où les spectres des composés de référence sont connus, nous exploitons les méthodes classiques de démélange linéaire en y intégrant de nouvelles idées par apprentissage de variétés. Plusieurs méthodes sont présentées et étudiées pour l'intégration directe ou indirecte d'informations dans les étapes de démélange. Ensuite, nous envisageons le cas où des données d'apprentissage de vérité terrain sont disponibles, et résolvons le problème de démélange dans ce contexte. L'approche proposée consiste en l'identification de l'application de l'espace des observations, celui des pixel-vecteurs, vers l'espace des vecteurs d'abondance. Un tel problème est dit de la pré-image [Bkir *et al.* 2004], selon la terminologie en vogue en machine learning.

Le document est structuré ainsi :

- **Chapitre 1** : État de l'art
  Ce chapitre présente les terminologies de l'imagerie hyperspectrale. Plusieurs problèmes de traitement des données hyperspectrales sont introduits, notamment la réduction de dimension, la classification, la détection et enfin le démélange.
- **Chapitre 2** : Modèle de mélange et méthodes de démélange linéaire
  Ce chapitre présente des modèles de mélange usité en imagerie hyperspectrale. La complexité et l'adaptation de chaque modèle est discutée. Puis, quelques méthodes pour le démélange linéaire sont détaillées. Ces algorithmes constituent un socle pour la présentation de nos propres méthodes de démélange non-linéaires.
- **Chapitre 3** : Ingénierie des noyaux
  Ce chapitre introduit des méthodes d'ingénierie pour élaborer des noyaux reproduisants adaptés aux applications envisagées. Les terminologies, les définitions, les caractéristiques et quelques méthodes à noyaux sont présentées. Les méthodes de noyaux sont au centre des préoccupations développées dans les chapitres à suivre.
- **Chapitre 4** : Démélange non-linéaire par apprentissage de variétés. Dans ce chapitre, nous analysons les méthodes de démélange classiques opérant sur

des variétés préalablement estimées. Nous présentons également des tests pour valider les algorithmes proposés.

– **Chapitre 5** : Démélange non-linéaire par méthode de pré-image.Nous y proposons une méthode originale de démélange basée sur le problème de la pré-image. Une évolution de cet algorithme, incorporant des informations spatiales, est introduite pour un gain en robustesse. Des simulations sont également réalisées pour valider les algorithmes.

Chapitre 1

État de l'art

Sommaire

Dans ce chapitre, nous introduisons les principes de l'acquisition d'images hyperspectrales ainsi que les principales applications. Finalement, nous présentons brièvement nos contributions dans ce domaine.

1.1 Acquisition d'images hyperspectrales et applications

L'imagerie hyperspectrale, comme les autres modes d'imagerie spectrale, acquiert des informations sur une partie du spectre électromagnétique. Si l'œil humain est sensible à la lumière dans le domaine dit du visible selon trois bandes spectrales principales (rouge, vert et bleu), l'imagerie spectrale segmente le spectre en un nombre plus ou moins important de bandes. Cette technologie vise notamment à acquérir des informations au-delà des limites du visible.

Les capteurs et systèmes de traitement correspondants sont élaborés afin de pouvoir être déployés dans nombre d'applications tels que l'agriculture, la minéralogie, la physique et la surveillance de la terre. Les capteurs acquièrent des signaux provenant des objets imagés sur une large plage du spectre électromagnétique. Il existe certains matériaux ayant des signatures spectrales uniques qui permettent de les identifier sans ambiguïté.

1.1.1 Acquisition des images hyperspectrales

Les capteurs hyperspectraux recueillent des informations selon une collection d'images, chacune d'elle correspondant à une plage du spectre électromagnétique. Ces images sont ensuite combinées pour former un cube de données hyperspectrales destiné à être traité et analysé. La figure 1.1 présente un exemple de cube hyperspectral généré par des capteurs aéroportés du « Airborne Visible/ Infrared Imaging Spectrometer » de la NASA (AVIRIS) [NASA 1992]. De tels cubes de données peuvent aussi être générés par des satellites comme Hyperion. Par ailleurs, des capteurs portatifs sont aussi utilisés pour certaines expérimentations et études [Huertas 1999]. Chaque pixel d'une image hyperspectrale consiste en un vecteur dont chacune des composantes renvoie à une propriété de réflectance du matériau dans une bande de fréquence donnée. La figure 1.2 illustre ce propos. La qualité d'une image hyperspectrale dépend de sa résolution spectrale aussi bien que de sa résolution spatiale. La résolution spectrale est définie par la largeur de chaque bande de fréquence. Plus les bandes sont étroites, plus les informations recueillies sont discriminantes, facilitant par là même l'identification des objets imagés. En revanche, il convient de noter que des cellules petites ont des capacités de détection plus faibles et sont sujettes à des rapports signal-sur-bruit (SNR) défavorables, ce qui limite l'efficacité des méthodes de traitement employées. Les cubes hyperspectraux fournissent une information de réflectance spectrale, qui est le rapport de l'énergie réfléchie sur l'énergie incidente pour chaque bande spectrale. La réflectance est une grandeur sans unité qui se situe dans l'intervalle [0,1]. La figure 1.3 illustre la procédure d'acquisition des données de réflectance. Ces réflectances varient en fonction de la bande spectrale pour la plupart des matériaux, car l'énergie dans ces bandes spectrales est dispersée ou absorbée à différents degrés. Ces spectres de réflectance varient selon les matériaux imagés. La figure 1.4 illustre ce fait. Les valeurs en creux de ces courbes indiquent les bandes pour lesquelles les matériaux absorbent significativement l'énergie incidente. Ces bandes sont appelées bandes d'absorption. La

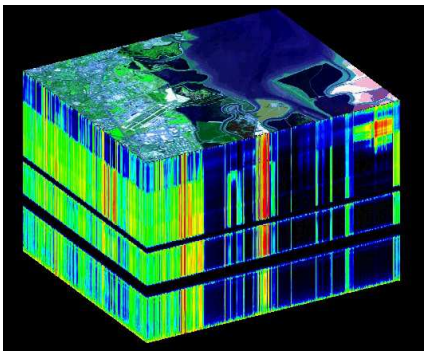


FIGURE 1.1 – Cube de données acquis par AVIRIS (NASA) [NASA 1992]

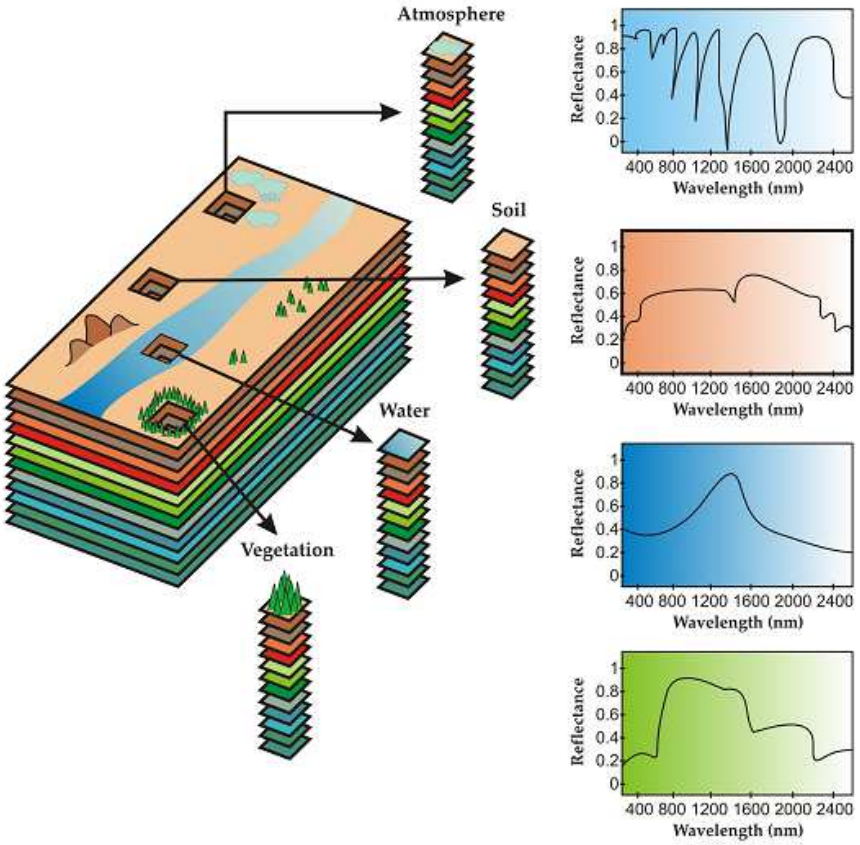


FIGURE 1.2 – Cube hyperspectral [Bioucas-Dias *et al.* 2012]

forme générale d'une courbe spectrale et la position des bandes d'absorption peuvent être utilisées pour identifier et distinguer les différents matériaux. Par exemple, la végétation a une plus grande réflectance que les sols dans le proche infrarouge et une réflexion moindre de la lumière rouge.

Dans un but de comparaison, il convient de situer l'imagerie hyperspectrale par rapport à l'imagerie multispectrale. La distinction repose sur le nombre de bandes spectrales ou le type de mesures. Le choix d'une technique dépend de l'objectif de l'application. L'imagerie multispectrale propose plusieurs images en bandes discrètes et étroites. Pour cette raison, l'imagerie multispectrale ne produit pas le spectre continu d'un objet. L'imagerie hyperspectrale renvoie à un continuum de bandes étroites. Ainsi un capteur constitué de seulement 20 bandes peut également être hyperspectral s'il couvre l'intervalle de 500 à 700 nm avec 20 bandes, chacune ayant 10 nm de largeur. Un autre terme aussi mentionné est l'imagerie ultraspectrale. Il est réservé aux capteurs de type interféromètre ayant une résolution spectrale très fine. La limitation de cette technologie est le volume très élevé de données.

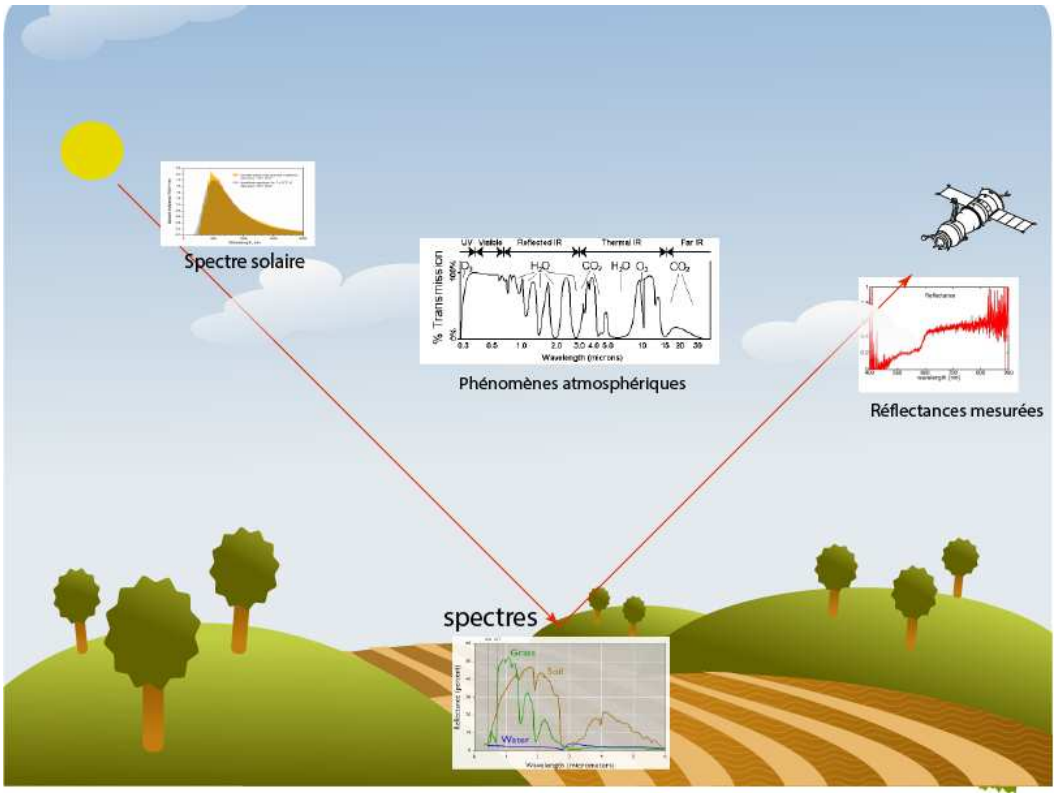


FIGURE 1.3 – Procédure d'acquisition des réflectances [Manolakis *et al.* 2003]

1.1.2 Applications

A l'origine, cette technologie a été développée pour l'exploitation minière et la géologie grâce à la possibilité d'identifier des minéraux différents, de rechercher du

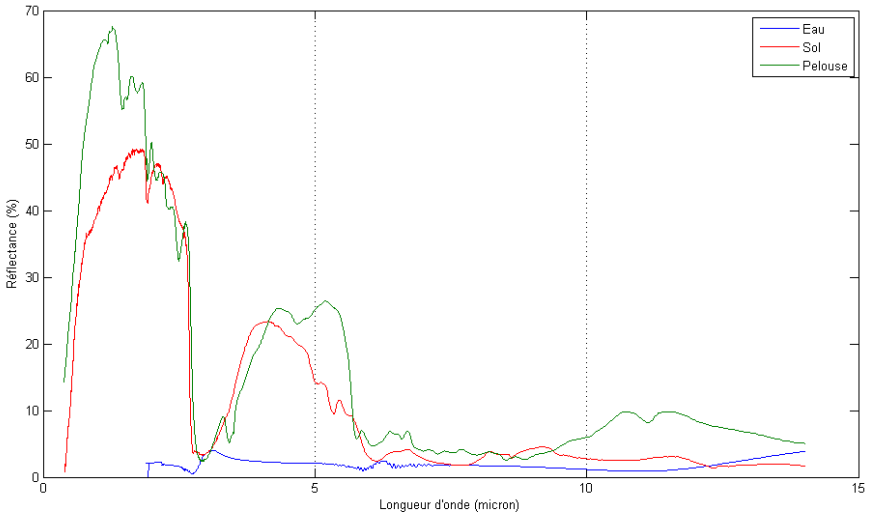


FIGURE 1.4 – Réflectances de matériaux différents en fonction des bandes de fréquence [Microimage 2013]

pétrole, etc ... À l'heure actuelle, de nombreuses applications dans des domaines variés coexistent. Parmi elles, on peut citer l'écologie, la surveillance de la terre, la chimie, l'astronomie. On relève également des applications dans le traitement des manuscrits historiques. Les applications de l'imagerie hyperspectrale ne cessent actuellement de croître.

La cartographie géologique et l'exploration minière sont parmi les applications principales qui bénéficient de la technologie hyperspectrale. De nombreux minéraux et roches présentent des motifs caractéristiques dans leur spectre électromagnétique, ce qui permet d'identifier leur composition chimique et les abondances relatives des différents matériaux présents. Les capteurs à bande étroite peuvent être embarqués sur des plates-formes aéroportées et spatiales, et disposer d'une résolution suffisante pour identifier les caractéristiques du milieu imagé et distinguer à distance les matériaux naturels le constituant. La plupart des études utilisant des données hyperspectrales, dans le cadre d'applications géologiques, concernent les zones de terres arides ou semi-arides car sous ces climats, la végétation et les sols sont rares. Par conséquent, les caractéristiques géologiques sont mieux préservées et exposées que dans les régions tropicales, [Noomen 2007, Werff 2006].

L'astronomie utilise depuis peu des imageurs hyperspectraux pour étudier la formation et l'évolution des objets astronomiques. Comme cette technologie donne accès à un cube de données où l'axe des longueurs d'ondes est échantillonné très finement, il est possible d'étudier avec précision le voisinage d'une raie d'émission ou d'absorption en utilisant une large plage de bandes spectrales. Ceci permet d'observer l'univers en volume et en profondeur. Néanmoins, la taille des cubes de données demeure très élevée. A titre d'exemple, notons que les cubes de données acquis par MUSE (Multi-Unit Spectroscopic Explorer) [Bacon & et Al 2012] atteignent jus-

qu'à $2400 \times 2400 \times 3000$ pixels. Ceci pose la question de la complexité et des coûts calculatoires pour les algorithmes appliqués à ce type de données.

L'imagerie hyperspectrale contribue aussi à la protection de l'environnement. La plupart des pays exigent une surveillance en continu des émissions produites par les centres de production d'énergie au charbon et au mazout, par les incinérateurs de déchets municipaux qui peuvent s'avérer dangereux, par les cimenteries, ainsi que par de nombreux autres types de sources industrielles. Ces activités de contrôle sont généralement pratiquées à l'aide de systèmes d'échantillonnage extractifs couplés à des techniques de spectroscopie infrarouge. Certains types de mesures permettent également l'évaluation de la qualité de l'air, mais ils ne sont pas très populaires en raison de l'incertitude sur les mesures. Le Telops Hyper-Cam, un imageur infrarouge hyperspectral, offre maintenant la possibilité d'obtenir une image complète des émissions résultant des cheminées industrielles à partir d'un emplacement distant, sans avoir besoin de systèmes d'échantillonnage extractifs, [Savary & et al. 2010].

La technologie hyperspectrale joue également des rôles importants dans la surveillance militaire. L'imagerie hyperspectrale y est particulièrement utile en raison de contre-mesures que les entités militaires adverses prennent maintenant pour contrer la surveillance aérienne.

L'avantage de l'imagerie hyperspectrale est que les méthodes n'ont besoin d'aucune connaissance fine de l'échantillonnage puisque le spectre entier est acquis en chaque pixel, à condition que les post-traitements puissent exploiter les données dans leur ensemble. L'imagerie hyperspectrale peut aussi profiter des relations spatiales entre les différents spectres dans une zone, ceci permettant d'utiliser des modèles spectro-spatiaux plus complexes pour une segmentation ou une classification de l'image plus précise. Cependant, les inconvénients principaux sont le coût et la complexité. Il est nécessaire de recourir à des calculateurs rapides, et à des capacités de stockage importantes car le volume d'un cube hyperspectral peut excéder des centaines de méga-octets. Tous ces facteurs augmentent considérablement le coût d'acquisition et de traitement des données hyperspectrales.

1.2 Aspects d'analyse d'image hyperspectrale

Comme mentionné précédemment, l'imagerie hyperspectrale offre de nombreuses opportunités dans des domaines différents. Cependant, cette technologie pose aussi plusieurs défis dans le traitement et l'analyse. On peut citer :

La variabilité spectrale : la variabilité de la signature spectrale d'un matériau peut être très importante dans les applications de télédétection en raison des variations des conditions atmosphériques, du bruit du capteur, de la composition des matériaux, des matériaux alentour, etc ... Dans certain cas, il s'avère difficile d'identifier un matériau donné à partir d'une bibliothèque de spectres.

Pixel de mélange la zone couverte au sol par un seul pixel est généralement assez grande. Dans ces conditions, la réflectance d'un pixel est le mélange de tous les

spectres des matériaux qui résident dans cette zone. On a alors à traiter des pixels de mélange au lieu de pixels purs où un seul matériau serait représenté.

Volume de données important et grande dimension de données : de nombreuses techniques de traitement du signal, usuellement réservées à des données de faible dimension, peuvent ne pas supporter un tel passage à l'échelle.

Suppression des interférences : de nombreux signaux invisibles pour un capteur d'imagerie ordinaire peuvent devenir visibles pour un capteur hyperspectral. Dans ces conditions, il faut prévoir de supprimer ces interférences afin d'offrir un meilleur contraste aux signaux d'intérêt.

Pour maîtriser ces problèmes et s'adapter aux demandes de plusieurs applications, l'imagerie hyperspectrale peut renvoyer à plusieurs tâches telles que

Réduction de dimension : réduire la dimension spectrale afin de faciliter l'analyse des données tout en préservant le résultat de l'analyse.

Classification : attribuer une étiquette de classe (ou la probabilité de classe) à chaque pixel (en mode supervisé ou non-supervisé).

Détection : rechercher les pixels correspondant à des signatures spectrales spécifiques connues ou non.

Détection de changements : trouver des changements importants entre deux données hyperspectrales d'une même zone géographique.

Démélange : estimer la fraction d'abondance de chaque matériau présent dans la zone couverte par le capteur.

La figure 1.5 illustre un procédé complet de l'exploitation d'images hyperspectrales. Dans les sections suivantes, nous introduisons en détail chacune de ces tâches.

1.3 Réduction de dimension

La taille importante des données et la grande dimension des images hyperspectrales entraînent des difficultés dans la transmission des données, le stockage et l'analyse. Heureusement, la dimension spectrale peut être réduite en exploitant la corrélation spectrale. Il est possible de réduire cette dimension en se basant, soit sur des transformations appropriées, grâce à une PCA par exemple, soit par sélection et regroupement de bandes. Les algorithmes utilisant des transformations ont pour but de trouver un sous-espace optimal au sens de certains objectifs : SNR maximal, variance maximale, etc ... Les données sont ensuite projetées sur ce sous-espace. Par ailleurs, les méthodes basées sur la sélection de bandes visent à choisir un sous-ensemble de bandes tout en préservant la qualité des résultats. Comme pour d'autres tâches, la réduction de dimension peut être supervisée ou non-supervisée, linéaire ou non-linéaire. Dans cette partie, nous allons nous concentrer sur quelques méthodes de réduction de dimension connues utilisant des transformations et projections différentes selon l'objectif recherché.

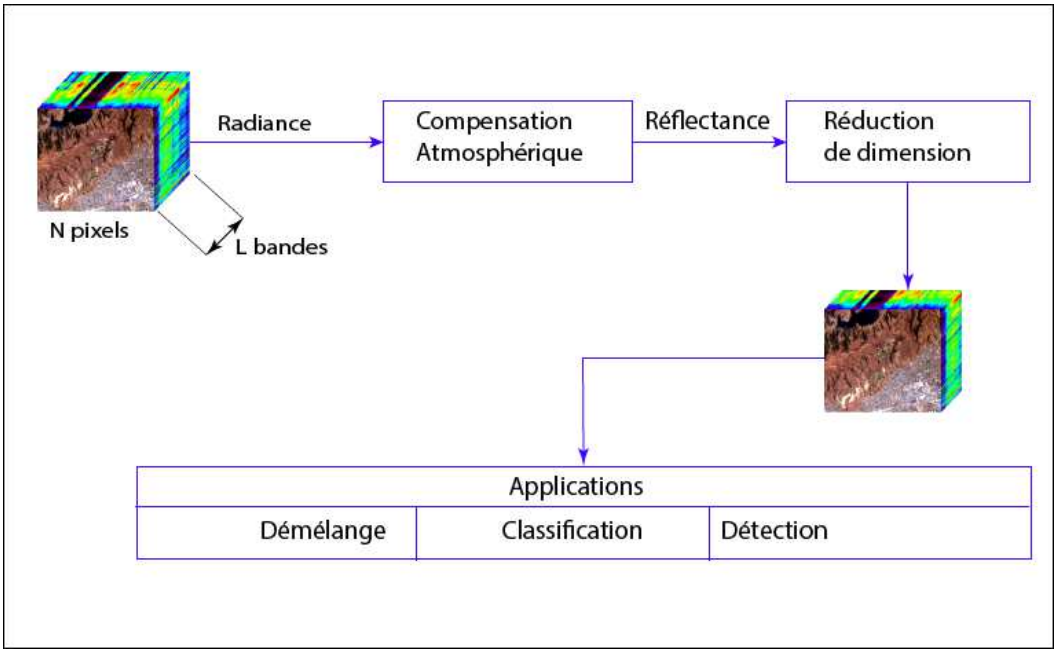


FIGURE 1.5 – Procédure de traitement et d'analyse [Shaw & Manolakis 2002]

1.3.1 Réduction basée sur les transformations

Une méthode classique non-supervisée est l'analyse en composantes principales (PCA) [Jolliffe 1986], qui consiste à transformer des variables corrélées en nouvelles variables décorrélées. Il s'agit d'une approche à la fois géométrique (les variables sont représentées dans un nouvel espace, selon des directions d'inertie maximale) et statistique (la recherche porte sur des axes indépendants expliquant au mieux la variabilité - la variance - des données). Lorsque l'on veut compresser un ensemble de variables, les premiers axes de l'analyse en composantes principales sont les meilleurs choix, du point de vue de l'inertie ou de la variance. Ces nouvelles variables sont nommées composantes principales, et s'expriment selon des axes principaux. Cette approche permet de réduire le nombre de variables et de rendre l'information moins redondante. On commence par calculer la matrice covariance $\boldsymbol{C}$ d'un tableau de données $\boldsymbol{X}$ de taille $m \times n$, avec m la dimension et n le nombre d'observations

$$\boldsymbol{C} = \frac{1}{m} \sum_{i=1}^{m} \boldsymbol{x}_i \boldsymbol{x}_i^T$$

Les données sont ensuite projetées sur les k premiers vecteurs propres de cette matrice $\boldsymbol{C}$, permettant de réduire la taille des données de m à k. La figure 1.6 illustre les deux composantes principales d'un ensemble de données de dimension deux. Il est possible d'appliquer la PCA afin de réduire la taille d'un cube hyperspectral et d'extraire ses caractéristiques, dites composantes principales, et d'alléger ainsi la charge des autres traitements. Si un tel traitement favorise une reconstruction fidèle des données malgré la réduction de leur dimension, il n'est pas nécessaire-

ment approprié lorsqu'il précède une opération de classification puisqu'il ne vise pas à préserver la séparabilité de classes. En raison de ces faiblesses, d'autres mé-

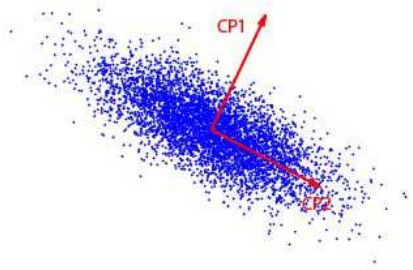


FIGURE 1.6 – Composantes principales d'un ensemble de données

thodes ont été introduites. On peut citer parmi elles Noise-Adjusted PCA (NAPCA) [Chang & Du 1999] ou Minimum Noise Fraction (MNF) [Green *et al.* 1988]. Ces méthodes affinent la PCA conventionnelle en classant les composantes principales en fonction du SNR. Ces méthodes cependant demandent une estimation précise du bruit, qui peut être difficile dans certain cas. De plus, le critère du SNR n'est pas approprié dans les cas où les interférences sont importantes. Pour estimer la variance du bruit en chaque bande spectrale (pour une matrice de covariance de bruit diagonale), il peut être supposé que le signal en chaque pixel a une forte corrélation avec ses voisins (spatiale et spectrale). Ainsi, le résidu d'une prédiction linéaire peut être associé à la composante de bruit. La variance du bruit peut être estimée comme la variance des composantes de bruit en tous les pixels. L'estimation des paramètres du bruit peut être plus précise si la prédiction linéaire est menée dans des zones homogènes.

Une autre méthode est l'analyse discriminante linéaire de Fisher (LDA) [Duda *et al.* 2000]. Il s'agit d'une technique supervisée standard de réduction de dimension dans le domaine de la reconnaissance des formes. Elle consiste à projeter les données de grandes dimensions sur un espace de faible dimension, où toutes les classes sont bien séparées. Ceci se traduit par la maximisation d'un quotient de Rayleigh. Néanmoins, cette méthode peut donner des résultats inattendus si les échantillons d'une classe sont distribués selon une loi multimodale. L'analyse discriminante de Fisher locale (LFDA) [Sugiyama 2006] est conçue pour surmonter cette difficulté. Pour l'imagerie hyperspectrale, qui peut présenter des caractéristiques multimodales dans chaque classe, LFDA peut donner de meilleures performances que le FDA original.

Dans de nombreux cas, les données hyperspectrales peuvent ne pas vivre dans

un hyperplan. Cela mène à des résultats biaisés qui influent sur les procédures de traitement à suivre. Les récents développements dans le domaine de l'apprentissage des variétés permettent de pallier cette difficulté. Si la variété est de faible dimension, les données peuvent être visualisées dans cet espace de faible dimension.

Parmi ces méthodes, on peut citer les plus connues : KPCA [Mika *et al.* 1999], ISOMAP [de Silva & Tenenbaum 2003], LLE [Roweis & Saul 2000]. KPCA est une méthode développée à partir de PCA en utilisant l'astuce du noyau [Aronszajn 1950]. KPCA commence par calculer la matrice de covariance des données représentées dans un espace de Hilbert à noyau reproduisant, de grande dimension.

$$\boldsymbol{C} = \frac{1}{m}\sum_{i=1}^{m}\Phi(\boldsymbol{x}_i)\Phi(\boldsymbol{x}_i)^T$$

avec $\Phi(\cdot)$ une application implicite donnée conséquente au choix d'un noyau. Les données transformées sont ensuite projetées sur les k premiers vecteurs propres de cette matrice, à l'image de PCA. Cette approche utilise l'astuce du noyau pour alléger la charge calculatoire. Malheureusement, il n'est pas trivial de trouver un noyau performant pour un problème donné, de sorte que KPCA ne donne pas toujours de bons résultats pour certains problèmes. Par exemple, il est connu qu'il donne de piètres performances pour des points situés sur une variété de type Swiss Roll, voir ci-après.

Isomap [de Silva & Tenenbaum 2003] est l'une des méthodes de réduction de dimensionnalité les plus largement utilisées. Pour cette méthode, les distances géodésiques sont traduites selon une métrique euclidienne grâce à la méthode MDS [Cox & Cox 1994]. L'algorithme fournit une méthode simple pour estimer la géométrie intrinsèque d'une variété de données grâce à une estimation approximative de la distance de chaque point à ses voisins. Isomap est très efficace et généralement applicable à un large éventail de type de données. La figure 1.7 donne un résultat obtenu avec Isomap. Cette méthode sera discutée dans la suite. Notons qu'il existe une évolution de cette méthode, le Landmark-Isomap [Bengio *et al.* 2003]. Cet algorithme vise à augmenter l'efficacité calculatoire au prix d'une certaine perte de précision des résultats.

L'algorithme LLE [Roweis & Saul 2000] a été proposé en même temps qu'Isomap par une équipe concurrente. Il a plusieurs avantages par rapport à Isomap, notamment une optimisation plus rapide reposant sur des matrices éparses, et une meilleure qualité des résultats sur de nombreux problèmes. LLE commence par déterminer un ensemble de plus proches voisins pour chaque chaque point. Il calcule ensuite un ensemble de poids pour chaque point, qui décrit au mieux celui-ci en tant que combinaison linéaire de ses voisins. A partir de ces informations, il utilise enfin une technique d'optimisation à base de vecteurs propres pour définir l'espace de représentation des données le plus adéquat.

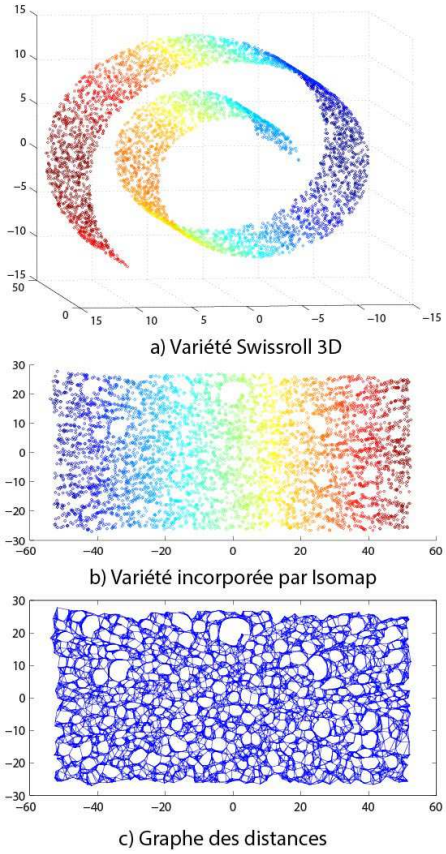


FIGURE 1.7 – Traitement des données Swiss Roll avec Isomap [Lei *et al.* 2012]

1.3.2 Notions sur l'utilisation des méthodes de réduction de dimensions

Dans certains cas, l'utilisation d'une méthode de réduction de dimension peut améliorer les performances dans le contexte de l'analyse des images hyperspectrales. Cependant, il est possible que l'utilisation de telles techniques puissent dégrader les performances, par rapport à celle utilisant la dimension originale. Enfin, la question du choix de la dimension de l'espace de représentation final demeure ouverte. Dans le cadre d'un problème de classification supervisée, celui-ci est lié au nombre de classes.

1.4 Classification

Étant donné un ensemble d'observations (pixel-vecteurs), le problème de classification consiste à attribuer une étiquette unique à chaque pixel. Pour résoudre ce problème, on peut recourir préalablement à un algorithme de démélange afin d'extraire les coefficients d'abondance en guise de vecteurs de paramètres. La taille du problème s'en trouve ainsi réduite. La figure 1.8 illustre un procédé de classification

standard.

Pour la classification de données hyperspectrales toujours, il existe plusieurs types de problèmes de classification. Ainsi est-il possible d'envisager un cadre supervisé ou non-supervisé selon l'existence d'une vérité terrain ou non. L'approche non-supervisée peut s'avérer délicate étant donné l'absence d'information sur le nombre de classes à discriminer. En ce qui concerne l'approche supervisée, la malédiction de la dimensionnalité constitue un écueil certain étant donnée la dimension des pixel-vecteurs au regard du nombre de données d'apprentissage généralement disponibles. Afin d'améliorer les performances des algorithmes, il est recommandé d'exploiter conjointement les informations spectrales et spatiales. En conclusion, il est en pratique important de recourir à des algorithmes de pré-traitement efficaces et rapides.

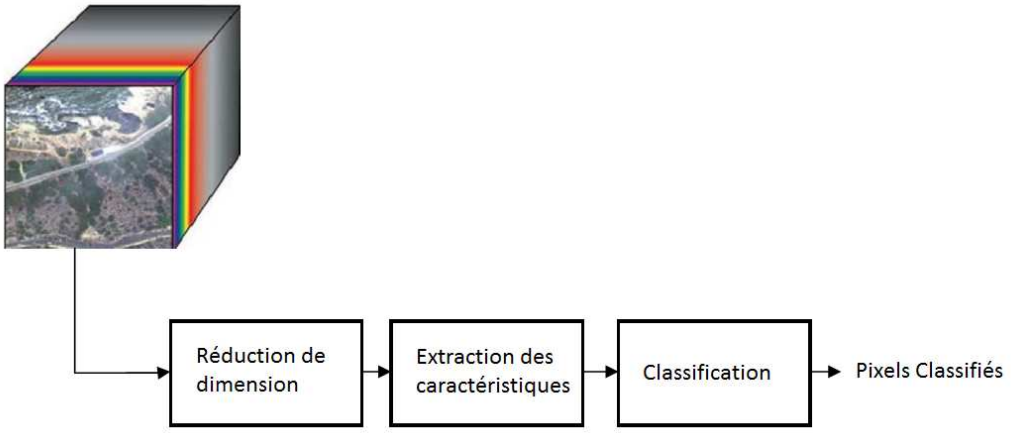


FIGURE 1.8 – Procédure de classification

1.4.1 Classification non-supervisée

La classification non-supervisée doit faire face à l'absence de données étiquetées. Plusieurs approches pour surmonter ce problème ont été proposées en analyse de données hyperspectrales : le clustering, la segmentation hiérarchique, l'approche objet. Le clustering [Elavarasi *et al.* 2011] vise à établir automatiquement des regroupements dans l'espace des caractéristiques en estimant les centres de chaque classe. La question centrale reste le choix du nombre de classes.

La segmentation hiérarchique a pour but de produire un ensemble de segmentations de l'image à différents niveaux de détails. Les segmentations grossières peuvent être produites en fusionnant des régions de segmentations plus fines. Un algorithme typique est la segmentation hiérarchique récursive (RHSEG)[Tarabalka *et al.* 2012]. Il s'agit d'un hybride de la technique de croissance de région et du clustering spectral. RHSEG effectue une segmentation multi-résolution, dans laquelle les régions aux niveaux les plus fins peuvent être associées à des régions à des niveaux plus

grossiers. Chaque région peut être suivie dans la hiérarchie de segmentations afin de désigner le niveau de segmentation optimal pour une région donnée.

La classification d'images basée sur les objets est aussi dite analyse d'images basée sur les objets (OBIA) ou GEOBIA [Chena *et al.* 2012] dans le domaine de la télédétection. La technique fonctionne à des échelles multiples et utilise les informations spectrales, morphologiques, tailles, textures, structure et contextuelles pour regrouper de façon non-supervisée des pixels dans des objets. Le logiciel le plus connu pour cette technologie est « Definiens e-cognition »(http ://www.ecognition.com). Il peut segmenter l'image originale en objets connectés, éventuellement à différents niveaux, et choisir automatiquement ou semi-automatiquement un niveau optimal pour chaque objet.

1.4.2 Classification supervisée

La classification supervisée utilise des pixels déjà étiquetés pour élaborer un classifieur. Il existe plusieurs stratégies pour la classification supervisée. D'abord, nous commençons par les approches classiques.

- Le classifieur utilisant la distance minimale qui assigne un pixel à la classe la plus proche.
- Le classifieur du maximum de vraisemblance évalue la probabilité d'attribution d'un pixel à une classe à l'aide à la fois de la variance et de la covariance des échantillons d'apprentissage disponibles.
- Les classifieurs discriminants qui visent à classer un ensemble de pixel-vecteurs dans des classes différentes en utilisant une fonction discriminante qui minimise l'erreur de classification. Plusieurs types de fonctions discriminantes peuvent être appliquées : plus-proches-voisins, arbres de décision, fonctions linéaires, fonctions non-linéaires, etc ...

Ces types de classifieurs classiques sont très sensibles à l'augmentation de la dimension des données. Pour les grandes dimensions, il faut disposer de suffisamment de données d'apprentissage. L'utilisation des noyaux est une bonne solution pour ce problème. Ce principe est aujourd'hui l'un des plus populaires pour la classification supervisée hyperspectrale. La classification utilisant les méthodes à noyaux a commencé avec l'utilisation des séparateurs à vaste marge (SVM) [Suykens *et al.* 2002, Camps-Valls & Bruzzone 2005]. L'astuce du noyau permet aux algorithmes d'opérer dans un espace de Hilbert de grande dimension, voire infinie, dans lequel les données d'apprentissage peuvent devenir linéairement séparables. Les SVM visent donc à trouver un hyperplan à marge maximale pour séparer les données d'apprentissage dans cet espace. La figure 1.9 illustre le principe de cette approche. Des difficultés subsistent cependant pour la mise en œuvre de ce type d'approche, en particulier le choix du noyau. D'autres algorithmes de classification populaires reposent sur les réseaux de neurones [Benediktsson *et al.* 1990]. Le perceptron multicouche avec apprentissage par rétro-propagation du gradient a été largement utilisé. Comparés aux méthodes à noyau, les réseaux de neurones sont davantage sensibles à la malédiction de la dimensionnalité. Toutefois, si un post-traitement spatial est ap-

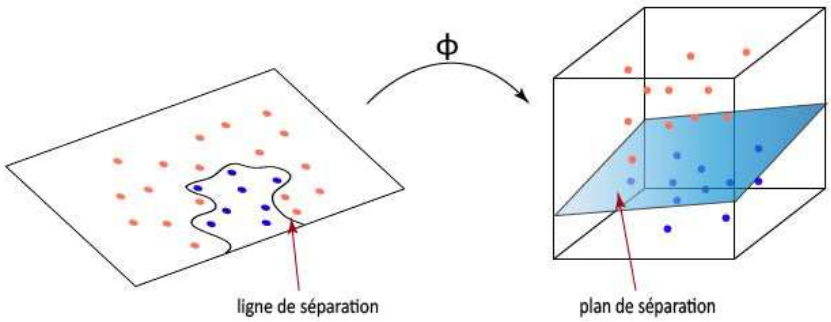


FIGURE 1.9 – Injection des données dans un espace de grande dimension

pliqué, les réseaux de neurones peuvent également s'avérer efficaces. Parmi les autres méthodes de classification supervisées, on peut citer la classification par sous-espaces [Li *et al.* 2012] et l'apprentissage de variétés [Ma *et al.* 2010].

1.4.3 Classification combinant informations spatiales et spectrales

Plusieurs algorithmes tendent à combiner les informations spatiales et spectrales pour améliorer les résultats de classification.

- Classification basée sur les profils morphologiques : ils utilisent des opérations d'ouverture et de fermeture afin d'inclure des informations spatiales [Benediktsson *et al.* 2003]. Lorsqu'elle est couplée à un classifieur SVM, cette méthode peut donner de très bons résultats. D'autres algorithmes ont aussi été développés selon le même principe : les profils morphologiques étendus [Plaza *et al.* 2005], les profils d'attributs morphologiques [Mura *et al.* 2010], les approches morphologiques des bassins hydrographiques [Tarabalka *et al.* 2010b]
- Classification intégrant une segmentation [Tarabalka *et al.* 2010a] : cette méthode combine le RHSEG (un classifieur non-supervisé) et un SVM pour classer chaque pixel en intégrant des informations spatiales.
- Classification associant champs de Markov et SVM [Tarabalka *et al.* 2010c] : un champ de Markov est utilisé dans le post-traitement pour exploiter des informations spatiales.

1.4.4 Progrès récents en classification

Dans cette section, on énumère quelques tendances nouvelles en classification des données hyperspectrales.

- Classification semi-supervisée : ce procédé utilise à la fois des échantillons étiquetés et non étiquetés. Pour les applications réelles, les données d'en-

traînement sont généralement difficiles à obtenir en grand nombre. On utilise donc également des échantillons non étiquetés qui devraient être relativement faciles à générer. Un algorithme de ce type est le SVM transductif (TSVMs) [Bruzzone *et al.* 2006]. Il est basé sur l'utilisation conjointe des échantillons étiquetés et non-étiquetés dans le cadre d'un processus itératif à apprentissage transductif.

- Apprentissage actif : il sélectionne les échantillons les plus informatifs dans un groupe de candidats. Jusqu'à présent, de nombreuses stratégies d'apprentissage actives homme-machine ont été proposées. Quand le cadre de machine à machine peut être utilisé, on parle d'auto-apprentissage semi-supervisé.
- Démélange des images avant classification [Dopido *et al.* 2011] : le démélange spectral peut être utilisé pour l'extraction de caractéristiques avant classification supervisée. Ce démélange peut être réalisé de façon supervisée (données étiquetées) ou non-supervisée (image originale).
- Intégration de la classification et d'un démélange spectral avec apprentissage actif [Dopido *et al.* 2012].

1.5 Détection

La détection d'une cible dans une image hyperspectrale se rapporte à l'utilisation de la haute résolution spectrale des images d'hyperspectrales pour cartographier les emplacements d'un objectif, ou une caractéristique avec une signature spectrale ou spatiale particulière. La détection de la cible en imagerie hyperspectrale renvoie à plusieurs techniques de traitement, et modes d'acquisition et fusion de l'information : reconnaissance par la forme, la signature hyperspectrale, la texture, etc... dans une ou plusieurs bandes. Les cibles d'intérêt sont souvent plus petites que la taille des pixels de l'image (détection de cibles sub-pixelliques), ou sont mélangées à d'autres types de couverture non ciblés au sein d'un pixel. Il faut donc recourir parfois à des pré-traitements tels que le démélange spectral pour détecter les cibles. Les images hyperspectrales sont utiles à la détection de cibles car elles renvoient à un continuum de bandes spectrales, et fournissent de grandes quantités de données à haute résolution spectrale. Avec une image hyperspectrale, les propriétés spectrales telles que le contraste, la variabilité, la similitude et la discriminabilité, peuvent être utilisées pour détecter des cibles au niveau sub-pixellique.

Bien que la détection puisse être envisagée comme un problème de classification à deux classes, il y a quelques différences majeures. Tout d'abord, le nombre de pixels couverts par la cible peut être petit et il est difficile d'avoir assez d'échantillons d'entraînement pour estimer les statistiques de la classe cible. De plus, en raison de la dominance des pixels de fond (non-cible), la minimisation de la probabilité d'erreur n'est plus un bon choix. Au lieu de cela, il est préférable de maximiser la probabilité de détection en gardant la probabilité de fausse alarme bornée.

Étant donné un spectre observé $\boldsymbol{r}$, un test classique consiste à comparer le rap-

port de vraisemblance défini ci-dessous à un seuil.

$$\Lambda(\boldsymbol{r}) = \frac{p(\boldsymbol{r}|\text{cible présente})}{p(\boldsymbol{r}|\text{cible absente})} \tag{1.1}$$

Le seuil doit être fixé de sorte à satisfaire un compromis entre la probabilité de bonne détection et la probabilité de fausse alarme, décrit par la courbe opérationnelle du récepteur (COR). En pratique, le choix du seuil devrait être effectué sans recourir à une intervention extérieure. Dans ce contexte, un détecteur de type CFAR veille à maintenir le taux de fausse alarme constant, indépendamment des paramètres de nuisance du problème.

1.6 Démélange des images hyperspectrales

En imagerie hyperspectrale, les capteurs acquièrent souvent des scènes dans lesquelles de nombreuses substances matérielles hétérogènes contribuent au spectre mesuré. Ce dernier est alors qualifié de pixel de mélange. Les pixels de mélange sont fréquents dans les images hyperspectrales en raison d'une résolution spatiale insuffisante du capteur, ou en raison des effets de mélanges intimes. La figure 1.10 illustre des pixels de mélange dans une image hyperspectrale. En plus des effets de mélanges spectraux, il y a beaucoup d'autres sources de perturbation qui peuvent affecter considérablement les images hyperspectrales. Par exemple, les perturbateurs atmosphériques sont une source potentielle de mélange. Par ailleurs, des effets de diffusion/réflexion multiples peuvent également entraîner des inexactitudes du modèle. Enfin, les ombres et les conditions d'éclairage variées peuvent intervenir. Les

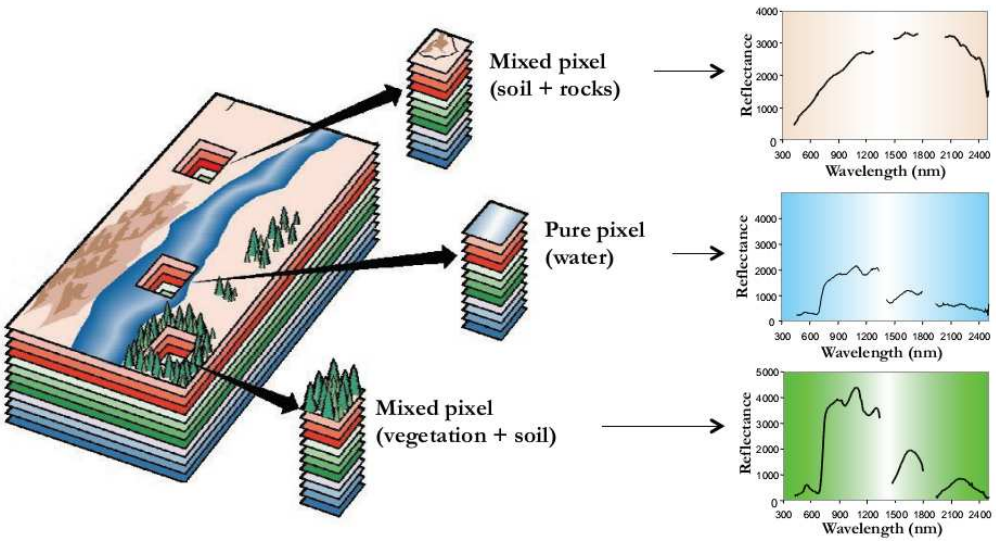


FIGURE 1.10 – Pixels de mélange dans une image hyperspectrale [Ciznicki *et al.* 2012]

pixels de mélange peuvent s'avérer très gênants dans les procédés d'analyse et de

traitement des images hyperspectrales. Par exemple, une procédure de classification pixel par pixel peut être aisément trompée du fait de la présence de mélange. Un remède consisterait à augmenter la résolution spatiale de l'image. Néanmoins, les effets de mélange intime dans des matériaux homogènes tels que les sols et les sables interviennent indépendamment d'une résolution même accrue. Notons que de tels mélanges entre matériaux sont de nature non-linéaire vis-à-vis des spectres mesurés. Les méthodes de démélange offrent des capacités d'interprétation impor-

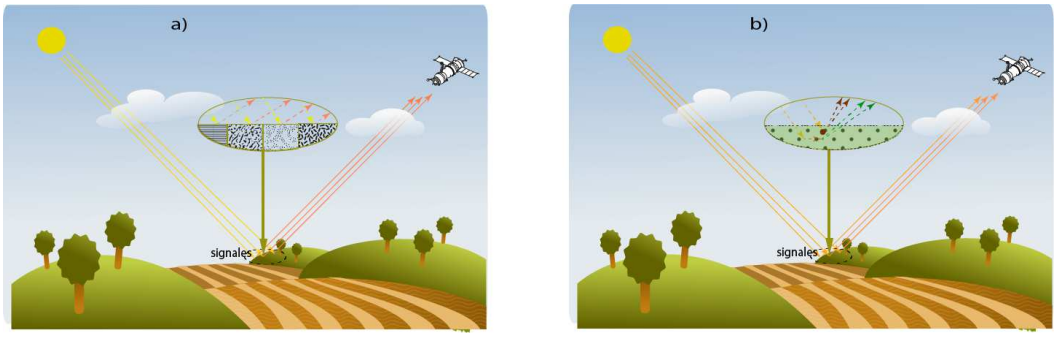


FIGURE 1.11 – Deux types de pixels dans une image hyperspectrale : a) mélange linéaire b) mélange intime non-linéaire [Keshava & Mustard 2002]

tante dans de nombreux scénarios tactiques pour lesquels les détails sub-pixelliques sont essentiels [Keshava & Mustard 2002]. Le démélange spectral est une procédure selon laquelle un spectre mesuré est décomposé en un ensemble de spectres élémentaires, ou endmembers, et un ensemble de fractions correspondantes, ou abondances. Plus clairement, ces derniers indiquent la proportion de chacun des endmembers représentés dans le pixel de mélange. Les endmembers correspondent aux éléments constitutifs identifiés dans la scène, comme l'eau, les sols, les métaux, etc. Un pixel-vecteur est dit pur s'il ne se réfère qu'à un seul élément. Une image hyperspectrale peut ne contenir aucun pixel pur.

Deux familles d'algorithmes de démélange peuvent être identifiées selon la nature des mélanges considérés : linéaire et non-linéaire. Dans un modèle de mélange linéaire, les composants macroscopiquement purs sont supposés être répartis de façon homogène dans des patchs distincts de la scène hyperspectrale. Dans un mélange non-linéaire, les composants microscopiquement purs sont intimement mélangés. Les algorithmes dédiés à ce type de mélange visent à identifier une fonction non-linéaire reliant les spectres des matériaux et leur abondance au pixel-vecteur acquis. Une connaissance a priori approfondie des matériaux est importante.

D'une manière générale, le problème du démélange spectral est un cas particulier du problème inverse généralisé qui vise à estimer les paramètres du système en utilisant une ou plusieurs observations d'un signal ayant interagi avec le système avant d'arriver au capteur. Un procédé de démélange complet commence par un cube de données sur lequel une correction atmosphérique a déjà été appliquée. Une réduction de dimension peut être appliquée selon l'algorithme de démélange choisi. Après cette étape, une extraction des endmembers est pratiquée pour trouver les

spectres constitutifs de la scène. Les algorithmes existants estiment les abondances après ou conjointement à l'extraction des endmembers. La procédure complète est schématiquement représentée sur la figure 1.12. Chacune des étapes est détaillée dans les sections suivantes.

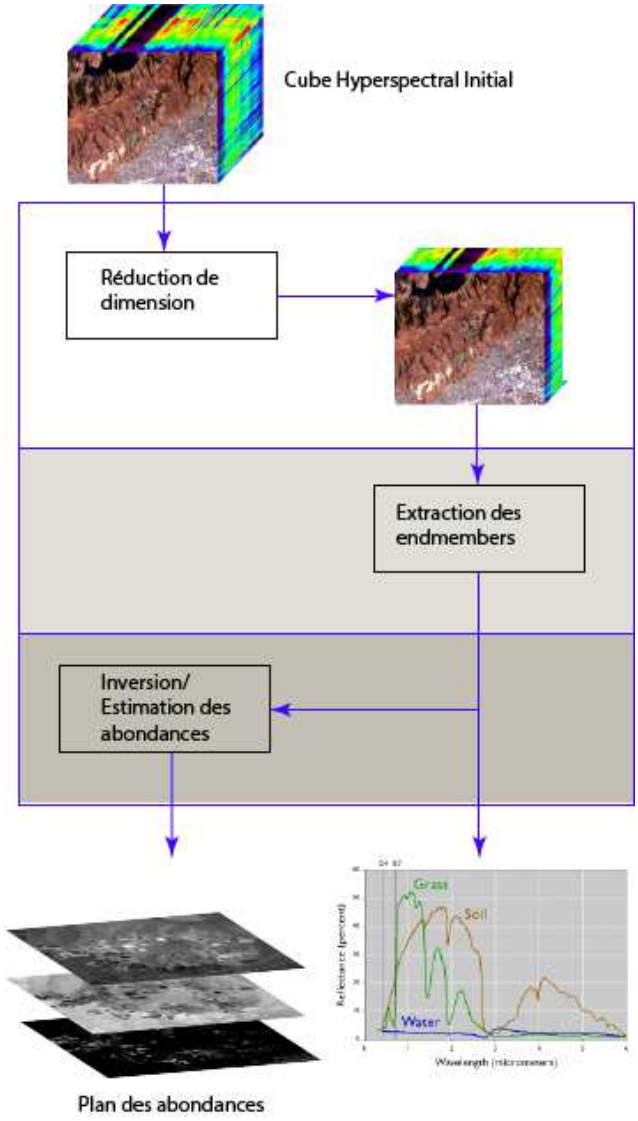


FIGURE 1.12 – Procédure de démélange [Parente & Plaza 2010]

1.6.1 Modèles de mélange

La définition d'un modèle physique de mélange spectral est une étape essentielle pour le démélange. Ces modèles visent à décrire comment les composantes constitutives d'un pixel se combinent pour produire les spectres mesurés. Ils tentent de reproduire la physique fondamentale dictant la phénoménologie hyperspectrale. Les algorithmes de démélange visent à utiliser ces modèles pour effectuer l'opération

d'inversion, qui a pour but d'estimer les endmembers et leurs abondances en chaque pixel-vecteur de l'image.

A partir des deux principaux types de mélanges existants, on a déduit deux familles générales de modèles de mélange :

Pour le modèle linéaire [Keshava & Mustard 2002], la surface imagée correspondant à chaque pixel est constituée de patchs homogènes. Le rayonnement réfléchi combine les spectres des éléments constitutifs dans les mêmes proportions que leurs abondances respectives. Dans ce cas, il existe une relation linéaire entre les abondances des substances et le spectre mesuré. Soit $\boldsymbol{r} = \{r_1, r_2, \ldots, r_L\}$ le spectre de réflectance acquis par le capteur avec L le nombre de bandes, $\boldsymbol{m}_i$ le spectre du i^{e} matériau constitutif, $\boldsymbol{\alpha} = \{\alpha_1, \alpha_2, \ldots, \alpha_R\}$ le vecteur des abondances avec R le nombre de matériaux, et $\boldsymbol{n}$ un bruit. On a donc

$$\boldsymbol{r} = \alpha_1 \boldsymbol{m}_1 + \alpha_2 \boldsymbol{m}_2 + \cdots + \alpha_R \boldsymbol{m}_R + \boldsymbol{n} \tag{1.2}$$

Dans ce modèle, les abondances doivent répondre aux contraintes de positivité et de somme unité :

$$\begin{aligned} &\alpha_i \geq 0, \qquad \forall i \in 1, \ldots, R \\ &\sum_{i=1}^{R} \alpha_i = 1. \end{aligned} \tag{1.3}$$

Pour le modèle non-linéaire, le modèle de mélange non-linéaire est décidément plus complexe en raison des interactions entachant le pixel-vecteur. On peut décrire ce modèle avec l'équation :

$$\boldsymbol{r} = \boldsymbol{\Psi}(\boldsymbol{\alpha}, \boldsymbol{M}) + \boldsymbol{n} \tag{1.4}$$

où Ψ est une fonction non-linéaire. Quelques modèles de cette famille ont déjà été introduits, par exemple, le modèle bilinéaire généralisé [Halimi *et al.* 2011a], le modèle intime [Nascimento & Bioucas-Dias 2010], ou le modèle post non-linéaire [Altmann *et al.* 2011b].

1.6.2 Réduction de dimension

L'étape de réduction de dimension a été présentée dans la section 1.3. Dans cette partie, nous allons davantage nous intéresser à la recherche du nombre de dimensions nécessaires, c'est-à-dire du nombre de sources primitives qui constituent l'image hyperspectrale, ou encore le nombre de endmembers. On peut citer quelques méthodes utiles : VD [Chang *et al.* 2010], HySime [Bioucas-Dias & Nascimento 2008], ou encore EML [Bin *et al.* 2012]

- Virtual Dimensionality (VD) : Cette méthode part du principe que chaque composé élémentaire est présent majoritairement dans une bande spectrale, où l'influence des autres composés est négligeable. Selon ce point de vue, l'estimation de la VD repose sur la comparaison des valeurs propres des matrices de covariance et de corrélation des pixels vecteurs observés.

– Hyperspectral Subspace identification minimum error (HySime) : Le principe de cette approche est proche d'une analyse en composantes principales, tout en prenant en compte les caractéristiques du bruit entachant les observations.
– Eigenvalue Likelihood Maximization (ELM) : Cette méthode apporte une modification à VD en prenant en compte le fait que les valeurs propres correspondant au bruit sont identiques sur les matrices de covariance et de corrélation. En outre, les valeurs propres correspondant au signal (endmembers) sont plus grandes dans la matrice de corrélation que dans la matrice de covariance. La technique exploite ce fait et donne une méthode entièrement automatique qui ne demande aucun paramètre en entrée comme VD, ni d'estimation du bruit comme Hysime.

1.6.3 Extraction des endmembers

Afin d'extraire des endmembers, plusieurs méthodes ont été développées. Certaines de ces méthodes supposent l'existence de pixels purs dans la scène et les considèrent comme des endmembers. Parmi celles-ci, on peut citer des méthodes classiques comme NFINDR [Winter 1999], SGA [Chang *et al.* 2006], ou encore VCA [Nascimento & Bioucas-Dias 2005].

L'algorithme N-FINDR a pour but de trouver les endmembers en maximisant le volume du simplexe défini par ces derniers. Les sommets qui définissent ce simplexe sont les endmembers recherchés, à condition qu'il existe un pixel pur pour chaque endmember dans l'image traitée. La figure 1.13 illustre ce propos. Une étape de réduction de dimension est appliquée avant d'utiliser cet algorithme car les calculs de volume demandent d'évaluer le déterminant d'une matrice carrée. L'algorithme

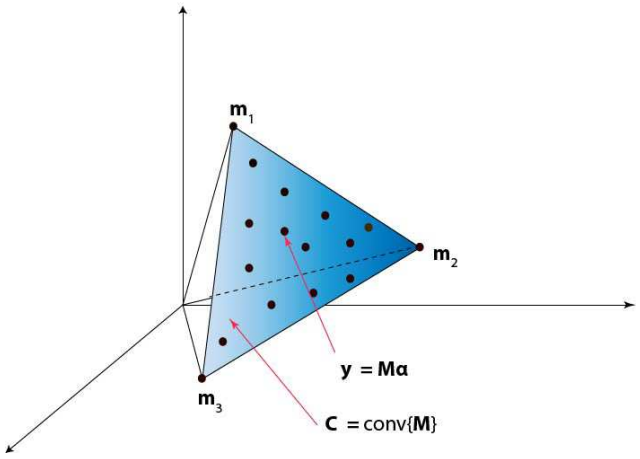


FIGURE 1.13 – Ensemble de données hyperspectrales comprenant 3 endmembers

SGA détermine les endmembers de manière successive. Tout d'abord, l'algorithme commence par un point initialisé de façon aléatoire $\boldsymbol{t}$, puis recherche un autre point

qui maximise le volume du simplexe $\mathcal{S}_{1i} = \{\boldsymbol{t}, \boldsymbol{r}_i\}$. Ce point est choisi comme le premier endmember et étiqueté $\boldsymbol{m}_1$. SGA poursuit sa recherche avec les simplexes $\mathcal{S}_{2i} = (\boldsymbol{m}_1, \boldsymbol{r}_i)$. Le point $\boldsymbol{r}_i$ maximisant le volume de ce simplexe est sélectionné comme étant le deuxième endmember et étiqueté $\boldsymbol{m}_2$. Le processus continue jusqu'à ce que les R endmembers requis aient été déterminés.

L'algorithme VCA repose sur le fait que la transformation affine d'un simplexe est également un simplexe. On projette les données selon une direction qui est orthogonale au sous-espace engendré par les endmembers qui ont été trouvés jusque là. Le point le plus éloigné fournit un endmember supplémentaire. L'algorithme enchaine ces itérations jusqu'à ce qu'il atteigne le nombre souhaité de endmembers. VCA traite la distance entre un point et le sous-espace engendré par les endmembers. Cet algorithme est similaire à SGA à la différence que le VGA utilise un processus de caractérisation du bruit afin de réduire la sensibilité au bruit. Ceci est réalisé en utilisant la décomposition en valeurs singulières (SVD) pour obtenir la projection qui représente au mieux les données au sens de la puissance maximale.

Ces méthodes peuvent être raffinées en ajoutant des information spatiales. Le pré-traitement spatial ne modifie pas l'algorithme spectral. Pour ces modifications, on suppose que les signatures pures demeurent dans des zones spatialement homogènes. Un indice d'homogénéité spatiale est évalué. Il est utilisé afin de définir des régions homogènes grâce à un algorithme de classification, qui servent à guider l'extraction des endmembers [Martin & Plaza 2011]. Il existe aussi des algorithmes intégrant les informations spatio-spectrales sous forme de transformations morphologiques comme AMEE [Plaza *et al.* 2002] ou SSEE [Rogge *et al.* 2007].

Très souvent, on ne trouve pas de pixels purs dans une scène hyperspectrale. Ces cas requièrent l'utilisation d'approches spécifiques. Parmi elles, on peut citer MVSA [Li & Bioucas-Dias 2008], SISAL [Hendrix *et al.* 2012], MVC-NMF [Miao & Qi 2007] et ICE [Berman *et al.* 2004]. Ces méthodes recherchent un simplexe de volume minimum qui enferme tous les pixels de la scène. Elles autorisent la violation de la contrainte de positivité par les fractions d'abondance car, en présence du bruit ou des perturbations, les vecteurs spectraux peuvent se situer en dehors du simplexe. Par ailleurs, MVC-NMF résout un problème d'optimisation appliqué aux données d'origine qui consiste à minimiser conjointement l'erreur de reconstruction et le volume du simplexe.

Le choix d'une méthode d'extraction dépend beaucoup du type de données, des applications, et du choix de l'algorithme de démélange utilisé.

1.6.4 Méthodes linéaires d'estimation des abondances

Une fois les endmembers extraits, il faut estimer leurs abondances respectives. Ce problème a le plus souvent été traité en utilisant le modèle de mélange linéaire. Une idée primitive est de trouver les abondances qui minimisent l'erreur de reconstruction $\|\boldsymbol{r} - \boldsymbol{M}\boldsymbol{\alpha}\|^2$, où $\boldsymbol{M}$ désigne la matrice des endmembers rangés en colonne. Cependant, ce résultat ne tient pas compte des contraintes de positivité et de somme unité. La méthode FCLS vise à résoudre ce problème [Heinz & Chang 2001] en minimisant

l'erreur quadratique sous ces contraintes. Le principal avantage de cette approche est la convexité du problème d'optimisation.

Il existe d'autres approches reposant sur le modèle linéaire. Des exemples sont décrits dans [Dobigeon *et al.* 2009, Theys *et al.* 2009, Honeine & Richard 2011a]. La stratégie géométrique décrite dans [Honeine & Richard 2011a] vise à estimer les abondances en calculant les ratios de volumes de polyèdres dans l'espace décrit par les pixel-vecteurs. Cette méthode a un faible coût calculatoire.

Certains algorithmes d'extraction peuvent fournir une estimation des abondances simultanément, notamment les méthodes ne reposant pas sur l'hypothèse de pixel pur. Il existe aussi des techniques basées sur les principes de séparation aveugle de sources, qui ne nécessitent pas que les endmembers soient connus. On peut citer l'analyse en composantes indépendantes (ICA) [Hyvarinen & Oja 2000] et la factorisation en matrices non-négatives (NMF).

Ces méthodes reposent sur une approximation linéaire de la loi de mélange des images hyperspectrales. Dans les cas de mélanges intimes par exemple, les résultats peuvent s'avérer sévèrement biaisés. Il est alors nécessaire de recourir à des méthodes de démélange non-linéaire.

1.6.5 Méthodes non-linéaires d'estimation des abondances

Plusieurs algorithmes et modèles de mélange ont été développés pour faire face aux non-linéarités dans les scènes hyperspectrales. Comme mentionné dans les sections précédentes, des modèles non-linéaires peuvent être introduits pour tenir compte de ces effets, par exemple, le modèle bilinéaire généralisé [Halimi *et al.* 2011a], le modèle de mélange post non-linéaire [Jutten & Karhunen 2003a] et le modèle intime [Hapke 1981]. Les méthodes de démélange non linéaire visent à inverser ces modèles et à estimer les abondances. Dans [Halimi *et al.* 2011a], un algorithme de démélange non-linéaire pour le modèle de mélange bilinéaire a été proposé. Basé sur l'inférence bayésienne, cette méthode présente toutefois une complexité calculatoire élevée et n'est consacrée qu'au modèle bilinéaire. Dans [Raksuntorn & Du 2010, Nascimento & Bioucas-Dias 2009], les auteurs élargissent l'ensemble des endmembers en ajoutant des termes croisés de signatures artificielles pour modéliser les effets de diffusion de la lumière sur les différents matériaux. Cependant, il n'est pas facile d'identifier les termes croisés qui devraient être sélectionnés et ajoutés au dictionnaire des endmembers. Si tous les termes croisés étaient envisagés, la taille de l'ensemble des endmembers pourrait croître de façon très significative. Une autre stratégie possible consiste à utiliser des méthodes d'apprentissage de variétés tels que Isomap [de Silva & Tenenbaum 2003] et LLE [Roweis & Saul 2000], comme nous le proposons dans ce document, ces méthodes permettant l'utilisation de méthodes linéaires dans la variété ainsi définie. Enfin, dans [Chen *et al.* 2012], les auteurs formulent un nouveau paradigme basé sur les noyaux reproduisants. Il repose sur l'hypothèse que le mécanisme de mélange peut être décrit par un mélange spectral linéaire des endmembers, complété par un terme de fluctuations non-linéaires défini dans un espace de Hilbert à noyau reproduisant.

Cette famille de modèles partiellement linéaires a une interprétation physique claire et permet de prendre en compte des interactions complexes entre les endmembers.

L'étape d'estimation des abondances peut être accomplie dans le contexte où les abondances sont connues pour certains pixels, appelés données d'entraînement. Un processus d'apprentissage est ensuite appliqué afin d'estimer les abondances des pixels restants, par exemple [Tourneret *et al.* 2008, Themelis *et al.* 2010, Altmann *et al.* 2011b]. Une méthode [Plaza & Plaza 2010] utilise un réseau de neurones pour apprendre les propriétés des mélanges non-linéaires à partir d'un ensemble d'apprentissage, en mode supervisé ou semi-supervisé. Le problème le plus important est la disponibilité limitée des données d'entraînement. Dans [Altmann *et al.* 2011b], le modèle considéré est une combinaison linéaire de fonctions à base radiale. Les pondérations sont estimées sur la base des échantillons d'apprentissage.

1.7 Orientation des travaux

Dans ce travail, nous étudions des méthodes d'apprentissage des variétés et les intégrons dans des méthodes de démélange, pour mieux extraire les endmembers et estimer leurs abondances.

D'abord, nous considérons des méthodes d'apprentissage de variétés afin de pouvoir y appliquer des techniques de démélange linéaire. Nous montrons que, dans un contexte d'apprentissage supervisé, l'estimation des abondances à partir de données d'apprentissage peut être envisagée comme un problème de pré-image [Honeine & Richard 2011a]. Bien que l'application de l'espace des observations vers l'espace des caractéristiques est de première importance pour les méthodes à noyau reproduisant, l'application inverse peut être également utile. Pour résoudre le problème pré-image, dans le cadre de notre application, nous nous concentrons sur l'élaboration d'une application de l'espace (transformé) des pixel-vecteurs, défini au préalable grâce à un noyau reproduisant, vers l'espace de faible dimension des vecteurs d'abondance. Nous considérons également le problème de sélection du noyau reproduisant. Nous montrons que les noyaux partiellement linéaires constituent une solution appropriée, la composante non-linéaire du noyau pouvant être avantageusement définie à partir de techniques d'apprentissage de variétés. L'introduction d'une régularisation spatiale de type TV (total variation) est également discutée et expérimentée [Iordache *et al.* 2011] pour tirer partie de l'information spatiale. En intégrant ce type d'information, nous améliorons la robustesse des algorithmes.

Chapitre 2

Modèles de mélange et méthodes de démélange linéaire

Sommaire

Dans ce chapitre, nous décrivons quelques modèles de mélange hyperspectraux ainsi que leur sens physique. Puis nous introduisons les principaux algorithmes de démélange issus de la littérature, qui seront utilisés par la suite à titre de comparaison. Finalement, nous évoquerons des pistes en vue d'améliorer ces techniques.

2.1 Modèle linéaire

Déjà présenté au Chapitre 1, le modèle linéaire vise à représenter un scénario simple de mélange. Il correspond au cas où la surface imagée en un pixel donné est constituée de patchs homogènes. Le rayonnement réfléchi combine alors les spectres des éléments constitutifs dans les mêmes proportions que leurs abondances respectives. La figure 2.1 illustre ce principe, en montrant que le modèle de mélange considéré pour une scène dépend également de l'échelle considérée. Aussi simple soit-il, ce modèle n'est pas à négliger car il favorise l'implémentation de méthodes de démélange efficaces d'un point de vue calculatoire et aisés à développer. Ce modèle s'écrit

$$\boldsymbol{r} = \boldsymbol{M}\boldsymbol{\alpha} + \boldsymbol{n} \tag{2.1}$$

où $\boldsymbol{r}$ est la signature spectrale observée en un pixel-vecteur, $\boldsymbol{M}$ est une matrice de taille $L \times R$ des signatures spectrales des éléments constitutifs, avec L le nombre de bandes spectrales, R le nombre de composés purs, et $\boldsymbol{\alpha}$ est le vecteur des abondances correspondant aux éléments purs regroupés en colonne dans la matrice $\boldsymbol{M}$. Cette équation peut s'écrire de manière plus compacte à l'échelle de la scène à traiter sous la forme

$$\boldsymbol{R} = \boldsymbol{M}\boldsymbol{A} + \boldsymbol{N} \tag{2.2}$$

où $\boldsymbol{R}$ est la matrice des pixel-vecteurs de taille $L \times N$, avec N le nombre d'observations dans la scène hyperspectrale, $\boldsymbol{N}$ la matrice du bruit correspondant à chaque pixel. Ce modèle linéaire doit satisfaire les contraintes de positivité et de somme unité décrites dans (1.3). Ces contraintes imposent que chaque pixel-vecteur se situe dans un simplexe convexe dont les sommets sont les composés constitutifs de la scène. La figure 1.13 illustre une telle configuration.

L'équation (2.2) suggère de résoudre le problème de démélange linéaire correspondant par moindres carrés contraints si $\boldsymbol{M}$ est connue. Dans le cas contraire, on peut envisager une méthode de type NMF pour estimer simultanément les spectres des composés constitutifs et leurs abondances dans la scène.

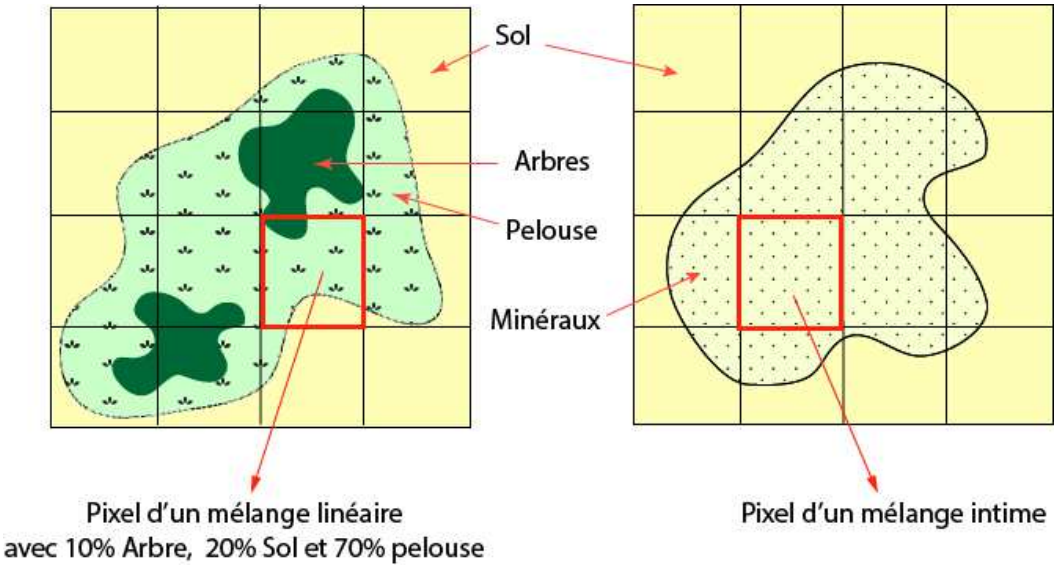


FIGURE 2.1 – a) Mélange linéaire, b) Mélange intime [Du & Plaza 2012]

2.2 Modèles non-linéaires

Dans certaines scènes hyperspectrales, les effets non-linéaires sont inévitables et peuvent biaiser les résultats si l'on utilise des méthodes linéaires. Ces effets peuvent être macroscopiques et générés par des réflexions multiples, ou encore par des interactions intimes entre les matériaux dans la zone imagée. Dans ce contexte, il apparait

incontournable de développer des modèles de démélange non-linéaires. Aussi, plusieurs modèles de ce type ont été développés dans la littérature. Ils tendent à mimer les phénomènes physiques sous-jacents, tout en veillant à respecter un compromis entre la précision du modèle et sa complexité vis-à-vis d'une procédure d'inversion. Nous décrivons ci-dessous les modèles les plus usités jusqu'ici.

2.2.1 Modèle bilinéaire généralisé

Ce modèle vise à généraliser le modèle linéaire, en considérant des termes croisés pour approcher les effets de réflexions multiples macroscopiques. Un tel scénario peut être rencontré pour des scènes acquises sur des zones forestières, où les interactions entre le sol et les feuillages de natures différentes sont multiples. Le modèle bilinéaire néglige les interactions impliquant plus de deux matériaux, ce qui justifie son appellation. L'observation $\boldsymbol{r}$ est modélisée par

$$\boldsymbol{r} = \sum_{i=1}^{R} \alpha_i \boldsymbol{m}_i + \sum_{j=1}^{R-1} \sum_{k=j+1}^{R} \beta_{j,k} \boldsymbol{m}_j \odot \boldsymbol{m}_k + \boldsymbol{n} \tag{2.3}$$

dans lequel $\odot$ représente le produit de Hadamard défini par

$$\boldsymbol{m}_j \odot \boldsymbol{m}_k = \begin{pmatrix} m_{1,j} \\ \vdots \\ m_{L,j} \end{pmatrix} \odot \begin{pmatrix} m_{1,k} \\ \vdots \\ m_{L,k} \end{pmatrix} = \begin{pmatrix} m_{1,j}\, m_{1,k} \\ \vdots \\ m_{L,j}\, m_{L,k} \end{pmatrix} \tag{2.4}$$

Les coefficients β_{jk} représentent la proportion des réflectances multiples causées par la réflexion entre les deux matériaux $\boldsymbol{m}_j$ et $\boldsymbol{m}_k$. Le trajet des photons de réflexions multiples étant plus long que le trajet normal, la réflectance s'en trouve réduite. Ces coefficients peuvent donc s'écrire sous la forme $\beta_{jk} = \gamma_{jk}\alpha_j\alpha_k$ avec $0 \leqslant \gamma_{jk} \leqslant 1$. En outre, bien que les termes d'interaction d'ordre supérieur soient également reçus par le capteur hyperspectral, les expériences menées dans [Somers *et al.* 2009] ont montré que ces termes peuvent être négligés. Différentes variantes du modèle bilinéaire ont été proposées dans la littérature. Ces modèles diffèrent par les contraintes d'additivité imposées sur les abondances. Le modèle proposé dans [Nascimento & Bioucas-Dias 2009], nommé « modèle de Nascimento », repose sur les contraintes suivantes :

$$\sum_{i=1}^{R} \alpha_i + \sum_{j=1}^{R-1} \sum_{k=j+1}^{R} \beta_{jk} = 1 \tag{2.5}$$

Ce modèle considère les termes croisés comme de nouveaux éléments constitutifs de la scène avec les coefficients β_{jk} pour abondances correspondantes. Le modèle décrit dans [Fan *et al.* 2009] diffère de ce dernier par

$$\sum_{i=1}^{R} \alpha_i = 1 \qquad \beta_{jk} = \alpha_j\alpha_k \tag{2.6}$$

Ce modèle attribue une relation d'égalité entre les coefficients β et le produit des abondances. Cette hypothèse n'a pas de justification physique solide. Finalement, le modèle bilinéaire généralisé défini dans [Halimi *et al.* 2011a] est donné par

$$\boldsymbol{r} = \boldsymbol{M}\boldsymbol{\alpha} + \sum_{j=1}^{R-1} \sum_{k=j+1}^{R} \gamma_{jk}\alpha_j\alpha_k \boldsymbol{m}_j \odot \boldsymbol{m}_k + \boldsymbol{n} \tag{2.7}$$

sous les contraintes

$$\sum_{i=1}^{R} \alpha_i = 1 \qquad \alpha_i > 0 \tag{2.8}$$

$$0 \leqslant \gamma_{jk} \leqslant 1 \qquad \gamma_{jk} \geqslant 0 \tag{2.9}$$

où γ_{jk} désigne l'interaction entre les éléments constitutifs j et k. Ce modèle est une généralisation intéressante du modèle linéaire prenant en compte des phénomènes de diffusion multiples macroscopiques.

2.2.2 Modèle intime

En sus des effets non-linéaires macroscopiques, les images hyperspectrales sont aussi corrompues par des phénomènes à l'échelle microscopique. A titre d'exemple, dans le cadre d'applications géologiques, les minéraux sont mélangés à une échelle sub-pixellique. Ces échelles spatiales sont généralement plus petites que la longueur du trajet suivi par les photons. La figure 2.1b illustre ce phénomène.

En se basant sur la théorie du transfert radiatif, on a développé plusieurs modèles théoriques ayant pour but de décrire avec précision les interactions subies par la lumière lors de la rencontre avec ces mélanges. Une approche est donnée par Hapke dans [Hapke 1981], où il décrit ces interactions intimes physiques en utilisant plusieurs quantités significatives. Il s'agit d'un modèle post non-linéaire. Suivant cette idée, plusieurs modèles de mélange non-linéaires simplifiés ont été proposés pour caractériser la relation non-linéaire entre des éléments constitutifs dans un mélange. Dans [Jia & Qian 2007], les auteurs définissent un modèle analytique pour exprimer les réflectances mesurées en fonction des paramètres intrinsèques aux mélanges, par exemple, la fraction de masse, les caractéristiques des particules individuelles (densité, taille) et les albédos de mono-dispersion. Toutefois, ce modèle dépend aussi fortement de paramètres empiriques car il nécessite la connaissance parfaite de la position et de l'orientation du capteur par rapport à la zone observée. Pour cette raison, l'estimation des paramètres de ce type de modèles s'avère difficile en pratique.

2.2.3 Modèles post non-linéaires

Les deux modèles précédents décrivent indépendamment deux types d'effets non-linéaires à deux échelles différentes. Afin de décrire conjointement les effets microscopiques et macroscopiques, un modèle dual à deux termes a été proposé dans

[Close *et al.* 2012, Jutten & Karhunen 2003a]

$$\boldsymbol{r} = \sum_{i=1}^{R} \alpha_i \boldsymbol{m}_i + \alpha_{R+1}\, g\left(\sum_{j=1}^{R} f_j \boldsymbol{w}_j\right) + \boldsymbol{n}. \tag{2.10}$$

Le premier terme décrit le mélange linéaire d'échelle macroscopique tandis que le second rend compte d'un mélange intime via un élément constitutif additionnel pondéré par un coefficient d'abondance.

D'autre part, dans [Altmann *et al.* 2012], les auteurs ont proposé un modèle à même de décrire une grande variété de classes de non-linéarités. Il est obtenu en considérant un développement au second ordre d'un modèle de mélange post non-linéaire (PPNMM). Plus précisément, le pixel-vecteur observé est modélisé par

$$\boldsymbol{r} = g\left(\sum_{i=1}^{R} \alpha_i \boldsymbol{m}_i\right) + \boldsymbol{n} \tag{2.11}$$

dans lequel g est une fonction non-linéaire définie par un polynôme de degré deux.

$$\begin{array}{ccc} g : [0,1]^L & \longrightarrow & R^L \\ \boldsymbol{x} & \longmapsto & \boldsymbol{x} + b\,(\boldsymbol{x} \odot \boldsymbol{x}) \end{array} \tag{2.12}$$

Cette fonction est paramétrée par un coefficient réel b afin d'atténuer ou magnifier la composante non-linéaire. Le modèle peut se réécrire

$$\boldsymbol{r} = \boldsymbol{M}\boldsymbol{\alpha} + b(\boldsymbol{M}\boldsymbol{\alpha}) \odot (\boldsymbol{M}\boldsymbol{\alpha}) + \boldsymbol{n} \tag{2.13}$$

On note qu'il se réduit au modèle linéaire lorsque $b = 0$. En outre, les auteurs ont montré qu'il est suffisamment flexible pour décrire la plupart des modèles bilinéaires. Enfin, en modifiant l'expression de g, il est possible de décrire de nombreux types de non-linéarités si tant est que le modèle de démélange demeure suffisamment simple pour être inversé.

2.3 Méthodes linéaires d'estimation des composés purs

Après avoir discuté de différents modèles de mélange, on va s'intéresser à la mise en œuvre d'une procédure complète de démélange. Nous allons présenter des méthodes basées sur ces modèles. Comme mentionné au chapitre 1, une chaîne classique commence avec la réduction de dimension du problème, puis l'extraction des composés constitutifs. La réduction de dimension est une option afin de réduire la complexité calculatoire. L'extraction des composés purs peut être réalisée avec ou sans réduction de dimension. Dans cette section, on détaille quelques méthodes connues d'extraction des composés purs.

Un composé pur peut être défini comme une signature spectrale idéalement pure pour une classe. Cette notion doit être distinguée de celle de pixel pur traditionnellement utilisée dans l'exploitation des données hyperspectrales, où le terme « pixel » désigne un pixel-vecteur de dimension L. Un pixel est dit pur si sa signature spectrale correspond à celle d'un composé pur, aussi appelé composé constitutif dans la suite.

2.3.1 Méthode N-FINDR

La méthode N-FINDR a été initialement introduite dans [Winter 1999]. Il s'agit de l'un des algorithmes les plus utilisés pour l'extraction des composés purs. Il repose sur la croissance d'un simplexe dans un espace de dimension L dont les R sommets sont sélectionnés parmi les observations. Etant donnée une configuration du simplexe à une itération donnée, N-FINDR essaye de remplacer une observation faisant office de sommet par une autre observation afin de faire croître le volume du simplexe. Ce procédé est illustré par la figure 2.2 et est décrit par l'algorithme suivant :

- Chaque sommet est successivement remplacé par le pixel-vecteur en cours de test ;
- Le volume du simplexe formé après chaque substitution est calculé ;
- Si le volume du simplexe croît grâce à l'une de ces substitutions, le pixel-vecteur en cours de test remplace le sommet concerné ;
- Le processus est répété tant que tous les pixel-vecteur constituant la scène n'ont pas été testés.

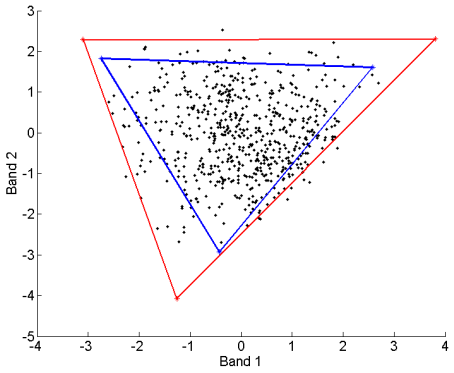


FIGURE 2.2 – N-FINDR

La mise en œuvre de cette méthode nécessite de surmonter plusieurs difficultés. La première concerne le nombre désiré R de composés purs. Pour cela, on peut recourir à des procédures telles que VD [Chang *et al.* 2010] ou encore Hysime [Bioucas-Dias & Nascimento 2008]. La deuxième question concerne le choix de l'ensemble initial des pixel-vecteurs constituant les sommets du simplexe. Une initialisation judicieuse est essentielle pour obtenir un résultat final correct et accélérer la convergence de l'algorithme. Pour cela, un algorithme efficace a été introduit dans [Neville *et al.* 1999]. Cet algorithme IEA (Iterative Error Analysis) repose sur des procédures de démélange partiel utilisant, par exemple, FCLS [Heinz & Chang 2001], afin de produire une séquence de pixel-vecteurs initiaux.

Un dernier problème repose sur le fait qu'un seul pixel-vecteur candidat au remplacement est considéré à la fois. En conséquence, cet algorithme glouton ne conduit pas de recherche exhaustive de la solution, rendue impossible compte tenu de la taille du problème. Plusieurs stratégies ont été proposées afin d'améliorer cette

situation, chacune poursuivant son propre objectif et adoptant un critère de convergence spécifique. Parmi eux, on peut citer : [Xiong *et al.* 2011, Sanchez *et al.* 2010, Plaza & Chang 2005].

Décrivons l'algorithme [Plaza & Chang 2005] constitué des étapes précédemment décrites :

1. Estimation du nombre R de composés purs : Utiliser VD ou Hysime ;
2. Réduction de dimension : Utiliser MNF pour réduire la dimension des pixel-vecteurs observés de L à $R-1$, et faciliter ainsi le calcul du volume des simplexes ;
3. Initialisation : Soit $\{\boldsymbol{m}_1^{(0)}, \boldsymbol{m}_2^{(0)}, \ldots, \boldsymbol{m}_R^{(0)}\}$ l'ensemble des composés purs initialisés par l'algorithme IEA ;
4. Calcul de volume : A l'itération k, calculer le volume du simplexe $\mathcal{S}_k$ dont les sommets sont les composés purs $\{\boldsymbol{m}_1^{(k)}, \boldsymbol{m}_2^{(k)}, \ldots, \boldsymbol{m}_R^{(k)}\}$ en utilisant
$$\boldsymbol{V}_{\mathcal{S}} = \frac{1}{(R-1)!} \left| \det \begin{pmatrix} 1 & 1 & \ldots & 1 \\ \boldsymbol{m}_1^{(k)} & \boldsymbol{m}_2^{(k)} & \ldots & \boldsymbol{m}_R^{(k)} \end{pmatrix} \right| \tag{2.14}$$
5. Calcul répété du volume : Pour chaque pixel-vecteur $\boldsymbol{r}$, calculer les volumes des simplexes $\mathcal{S}(\boldsymbol{r}, \boldsymbol{m}_2^{(k)}, \ldots, \boldsymbol{m}_R^{(k)})$, $\mathcal{S}(\boldsymbol{m}_1^{(k)}, \boldsymbol{r}, \ldots, \boldsymbol{m}_R^{(k)})$, $\mathcal{S}(\boldsymbol{m}_1^{(k)}, \boldsymbol{m}_2^{(k)}, \ldots, \boldsymbol{r})$. Si aucune de ces substitutions ne permet d'accroitre $\boldsymbol{V}(\mathcal{S}_k)$, ne pas modifier l'ensemble des composés purs avec $\boldsymbol{r}$. Sinon, procéder au remplacement par $\boldsymbol{r}$ du sommet à l'ensemble ayant conduit au plus grand accroissement de $\boldsymbol{V}(\mathcal{S})$.

2.3.2 SGA ou OSP

L'algorithme SGA (Simplex Growing Algorithm) [Chang *et al.* 2006] vise à extraire des endmembers désirés en utilisant la croissance d'une séquence de simplexes. Il commence avec deux sommets, et ajoute successivement des sommets au simplexe, extraits des pixel-vecteurs observés. L'algorithme s'achève lorsque le nombre de sommets atteint vaut R, préalablement estimé par VD ou Hysime. La figure 2.3 décrit le processus de SGA. Afin de sélectionner un pixel approprié comme sommet initial, un processus de sélection tel que celui ci-après est utilisé.

1. Choisir de façon aléatoire un pixel-vecteur $\boldsymbol{t}$ dans l'ensemble de données ;
2. Trouver un pixel-vecteur $\boldsymbol{m}_1$ parmi les pixel-vecteurs $\boldsymbol{r}$ de la scène qui maximise le déterminant
$$\left| \det \begin{pmatrix} 1 & 1 \\ \boldsymbol{t} & \boldsymbol{r} \end{pmatrix} \right|$$
Pour calculer ce dernier, il est nécessaire de recourir à une réduction de L à 1 en utilisant simplement une PCA ou MNF.

Il convient de noter que la production du premier sommet $\boldsymbol{m}_1$ est déterminée par le pixel généré aléatoirement $\boldsymbol{t}$. Le choix d'un autre pixel $\boldsymbol{t}$ peut générer un autre sommet $\boldsymbol{m}_1$. Toutefois, les expériences réalisées dans [Chang *et al.* 2006] montrent que le pixel $\boldsymbol{t}$ n'a que peu d'effet sur le résultat final. L'algorithme SGA est donc décrit comme ci-dessous :

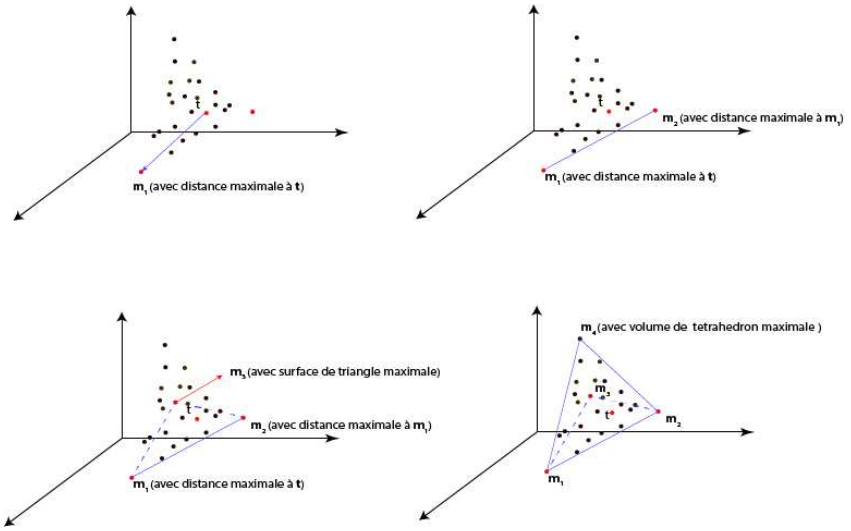


FIGURE 2.3 – Procédure SGA [Chang *et al.* 2006]

1. Initialisation : Utiliser VD ou Hysime pour estimer R. Initialiser $\boldsymbol{m}_1$ en utilisant le processus décrit ci-dessus. On note $n = 1$;
2. Pour $n \geq 1$ et pour chaque pixel $\boldsymbol{r}$ testé, calcul du volume $\boldsymbol{V}(\boldsymbol{m}_1, \boldsymbol{m}_2, ...\boldsymbol{m}_n, \boldsymbol{r})$ du simplexe à partir de

$$\boldsymbol{V}(\boldsymbol{m}_1, \boldsymbol{m}_2, ...\boldsymbol{m}_n, \boldsymbol{r}) = \frac{1}{n!} \left| \det \begin{pmatrix} 1 & 1 & \dots & 1 & 1 \\ \boldsymbol{m}_1 & \boldsymbol{m}_2 & \dots & \boldsymbol{m}_n & \boldsymbol{r} \end{pmatrix} \right| \tag{2.15}$$

   Une technique de réduction de dimension est nécessaire pour que la matrice considérée dans l'expression ci-dessus soit carrée ;
3. Trouver $\boldsymbol{m}_{n+1}$ tel que :

$$\boldsymbol{m}_{n+1} = \arg\max_{\boldsymbol{r}} \boldsymbol{V}(\boldsymbol{m}_1, \boldsymbol{m}_2, ...\boldsymbol{m}_n, \boldsymbol{r}) \tag{2.16}$$

   L'algorithme se poursuit jusqu'à ce que tous les sommets aient été déterminés.

Cet algorithme a un faible coût calculatoire. Par ailleurs, l'ensemble des sommets obtenus est consistent, ce qui est un point fort par rapport aux autres algorithmes. En revanche, il nécessite une réduction de dimension à chaque itération, ce qui constitue une faiblesse certaine de la méthode.

2.3.3 VCA

VCA (Vertex Component Analysis) [Nascimento & Bioucas-Dias 2005] est une méthode non-supervisée reposant sur le fait que toute transformation affine d'un simplexe est aussi un simplexe, dont les sommets restent inchangés. Cet algorithme, comme les précédents, suppose l'existence de pixels purs dans la scène hyperspectrale. Il projette de manière itérative des données dans une direction orthogonale au sous-espace défini par les sommets déjà déterminés. Le nouveau sommet correspond

au point dont la projection est la plus éloignée du simplexe courant. L'exécution de l'algorithme se poursuit jusqu'à ce que tous les sommets aient été déterminés. Les performances de VCA sont au moins aussi bonnes que celles de N-FINDR, pour une complexité moindre. Du fait de la projection, notons qu'il est insensible à la présence d'un facteur d'illumination γ tel que $\boldsymbol{r} = \boldsymbol{M}\gamma\boldsymbol{\alpha} + \boldsymbol{n}$. La figure 2.4 illustre les projections effectuées.

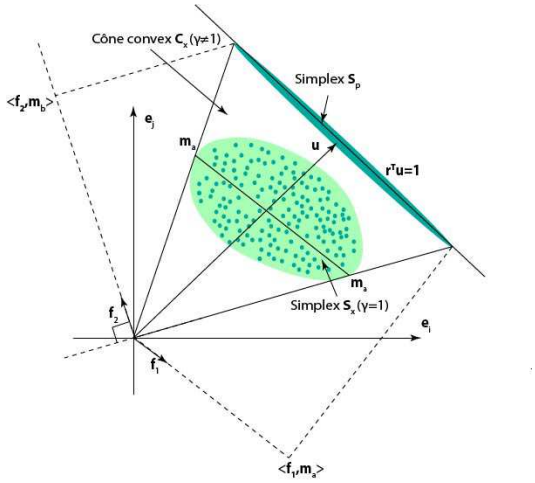


FIGURE 2.4 – Procédure VCA [Nascimento & Bioucas-Dias 2005]

2.4 Méthodes linéaires d'estimation des abondances

Cette étape fait suite à l'extraction des composés purs. Les algorithmes ci-après supposent donc que les composés ont été préalablement déterminés. Ils sont souvent qualifiés d'algorithmes de démélange supervisés. En outre, ils reposent sur des modèles de mélange déjà présentés au cours de la première section. Dans cette partie, on se consacre aux algorithmes de démélange linéaire. Les méthodes non-linéaires reposent sur des principes plus complexes, et seront donc évoqués dans la suite.

2.4.1 FCLS

Pour un modèle de mélange linéaire $\boldsymbol{r} = \boldsymbol{M}\boldsymbol{\alpha} + \boldsymbol{n}$, sans contraintes, on peut utiliser la méthode des moindres carrés pour estimer le vecteur des abondances, soit

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{r} - \boldsymbol{M}\boldsymbol{\alpha}\|^2 \tag{2.17}$$

La solution de ce problème est la solution très connue :

$$\hat{\boldsymbol{\alpha}} = (\boldsymbol{M}^\top \boldsymbol{M})^{-1}\boldsymbol{M}^\top \boldsymbol{r} \tag{2.18}$$

Une solution analytique des moindres carrés sous contrainte de type égalité existe incluant la contrainte de somme unité. Le problème réside dans l'introduction

de la contrainte de non-négativité. Plusieurs algorithmes ont été proposés afin de satisfaire à ces contraintes, dont FCLS (Fully Constrained Least Squares) [Heinz & Chang 2001]. Le problème

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{r} - \boldsymbol{M}\boldsymbol{\alpha}\|^2 \tag{2.19}$$

sous les contraintes $\mathbf{0} \preceq \boldsymbol{\alpha} \preceq \mathbf{1}$ peut être aisément résolu par programmation quadratique. Pour intégrer la contrainte de somme unité, FCLS ajoute un vecteur ligne à la matrice $\boldsymbol{M}$ dont tous les éléments sont égaux à 1. On fait de même avec le pixel-vecteur observé $\boldsymbol{r}$ et on introduit un paramètre δ ($\delta \geq 0$) tels que :

$$\tilde{\boldsymbol{r}} = \begin{pmatrix} \delta \boldsymbol{r} \\ 1 \end{pmatrix} \qquad \text{et} \qquad \tilde{\boldsymbol{M}} = \begin{pmatrix} \delta \boldsymbol{M} \\ \mathbf{1} \end{pmatrix} \tag{2.20}$$

Le problème (2.19) se résout alors avec $\tilde{\boldsymbol{r}}$ et $\tilde{\boldsymbol{M}}$. Notons que ceci revient à ajouter un terme de pénalité au terme d'attache aux données avec un paramètre de régularisation $1/(2\delta^2)$, la contrainte de somme unité n'est donc pas strictement vérifiée en procédant ainsi.

2.4.2 Démélange basé sur la factorisation de matrice non-négative

Des algorithmes ont été développés afin d'estimer simultanément la matrice des composés purs et les abondances des pixel-vecteurs de la scène [Miao & Qi 2007, Berman *et al.* 2004, Hendrix *et al.* 2012, Li & Bioucas-Dias 2008]. Généralement, ces algorithmes n'exploitent pas l'hypothèse d'existence de pixels purs. La base théorique pour ces méthodes est la factorisation en matrices non-négatives. Dans le cadre d'une application à l'imagerie hyperspectrale, il s'agit de factoriser la matrice $\boldsymbol{R}$ non-négative en deux matrices non-négatives : matrice des endmembers $\boldsymbol{M}$ de taille $L \times R$, et la matrice des abondances $\boldsymbol{A}$ de taille $R \times N$, avec N le nombre d'observations, (2.2). Le problème peut s'écrire ainsi :

$$\begin{aligned} &\min_{\boldsymbol{M},\boldsymbol{A}} \frac{1}{2}\|\boldsymbol{R} - \boldsymbol{M}\boldsymbol{A}\|_F^2 \\ &\qquad \text{sujet à} \quad \boldsymbol{A} \succeq 0 \quad \text{et} \quad \boldsymbol{M} \succeq 0 \end{aligned} \tag{2.21}$$

La solution de ce problème n'est pas unique, celui-ci n'étant par ailleurs pas conjointement convexe par rapport à $\boldsymbol{M}$ et $\boldsymbol{A}$. Il convient donc d'ajouter des contraintes afin de restreindre l'espace des solutions. L'algorithme MVC-NMF [Miao & Qi 2007] propose d'ajouter la contrainte de volume minimum. Le problème peut donc se formuler ainsi

$$\begin{aligned} &\min_{\boldsymbol{M},\boldsymbol{A}} \frac{1}{2}\|\boldsymbol{R} - \boldsymbol{M}\boldsymbol{A}\|_F^2 + \lambda \mathcal{J}(\boldsymbol{M}) \\ &\qquad \text{sujet à} \quad \boldsymbol{M} \succeq 0 \quad \text{et} \quad \boldsymbol{A} \succeq 0 \quad \text{et} \quad \mathbf{1}_R^\top \boldsymbol{A} = \mathbf{1}_N^\top \end{aligned} \tag{2.22}$$

dans lequel $\mathbf{1}_N^\top$ est un vecteur ligne de taille N dont les éléments sont égaux à 1. Le premier terme est l'erreur de reconstruction qui veille à trouver un simplexe circonscrivant toutes les données. Le deuxième est la restriction de volume du simplexe qui

tend à le rendre aussi compact que possible. Le terme $\mathcal{J}(\boldsymbol{M})$ est donné par

$$\mathcal{J}(\boldsymbol{M}) = \frac{\left|\det \begin{pmatrix} 1 & 1 & \dots & 1 \\ \boldsymbol{m}_1 & \boldsymbol{m}_2 & \dots & \boldsymbol{m}_R \end{pmatrix}\right|^2}{2(R-1)!} \tag{2.23}$$

Une autre méthode qui permet aussi de réaliser le démélange sans hypothèse d'existence de pixel-vecteurs purs est MVSA/SISAL. La différence majeure entre MVC-NMF et MVSA/SISAL est que celui-ci permet la violation des contraintes de positivité.

L'avantage de ces méthodes se situe dans la possibilité d'estimer simultanément l'ensemble des pixels purs et des abondances. En outre, elles permettent d'exploiter des informations spatiales. L'algorithme présenté dans [Iordache *et al.* 2011] utilise la factorisation en matrices non-négatives en intégrant des informations spatiales pour améliorer les performances. Toutefois, les principales faiblesses résident toujours dans la complexité calculatoire et la non-convexité du problème.

2.4.3 Démélange basé sur des propriétés géométriques

Le problème d'estimation des abondances peut se résoudre par une formulation géométrique du problème [Honeine & Richard 2011a]. On peut considérer les abondances comme des coordonnées barycentriques. Celles-ci peuvent être obtenues, soit comme un rapport de volumes de simplexes, soit comme un rapport des distances. Il apparait que ces volumes et distances sont inévitablement calculés lors de la phase d'extraction des composés purs, avec N-FINDR par exemple, et qu'il est donc possible d'exploiter ces résultats déjà acquis pour évaluer les abondances sans aucun calcul supplémentaire. En relaxant la contrainte de positivité et en conservant la contrainte de somme unité, le problème de démélange s'écrit comme un système linéaire décrit ci-dessous

$$\begin{pmatrix} 1 & 1 & \dots & 1 \\ \boldsymbol{m}_1 & \boldsymbol{m}_2 & \dots & \boldsymbol{m}_R \end{pmatrix} \boldsymbol{\alpha} = \begin{pmatrix} 1 \\ \boldsymbol{r} \end{pmatrix} \tag{2.24}$$

Ce système est supposé carré et de taille $R \times R$, signifiant par la même qu'une opération de réduction de dimension des $\boldsymbol{m}_i$ devra avoir été pratiquée au préalable. En utilisant la règle de Cramer, la solution de ce système linéaire peut s'écrire ainsi

$$\alpha_i = \frac{\det \begin{pmatrix} 1 & \dots & 1 & \dots & 1 \\ \boldsymbol{m}_1 & \dots & \boldsymbol{r} & \dots & \boldsymbol{m}_R \end{pmatrix}}{\det \begin{pmatrix} 1 & \dots & 1 & \dots & 1 \\ \boldsymbol{m}_1 & \dots & \boldsymbol{m}_i & \dots & \boldsymbol{m}_R \end{pmatrix}} \tag{2.25}$$

Etant donné le simplexe $\mathcal{S} = \{\boldsymbol{m}_1, \boldsymbol{m}_2, \dots, \boldsymbol{m}_R\}$, on remarque que le dénominateur de l'expression ci-dessus est proportionnel au volume de $\mathcal{S}$ donné par

$$\boldsymbol{V}_{\mathcal{S}} = \frac{1}{(R-1)!} \det \begin{pmatrix} 1 & 1 & \dots & 1 \\ \boldsymbol{m}_1 & \boldsymbol{m}_2 & \dots & \boldsymbol{m}_R \end{pmatrix} \tag{2.26}$$

De même, le numérateur est proportionnel au volume du simplexe $\mathcal{S}\backslash\{\boldsymbol{m}_i\} \cup \{\boldsymbol{r}\}$. Il en résulte que

$$\alpha_i = \frac{V_{\boldsymbol{S}\backslash\ \{\boldsymbol{m}_i\}\cup\{\boldsymbol{r}\}}}{\boldsymbol{V_S}} \tag{2.27}$$

avec $i = 1, 2, \ldots, R$. Ces volumes sont préalablement calculés lors de la mise en œuvre de N-FINDR. On peut aussi écrire ces équations sous la forme de distances, qui nous permet d'intégrer ces estimations dans les algorithmes basés sur les distances comme VCA ou ICE.

$$\boldsymbol{V}_{\mathcal{S}} = \frac{1}{(R-1)!}\delta(\boldsymbol{m}_i)\boldsymbol{V}_{\boldsymbol{S}\backslash\{\boldsymbol{m}_i\}} \tag{2.28}$$

où $\delta(\boldsymbol{m}_i)$ est la distance entre le sommet $\boldsymbol{m}_i$ et le sous-espace engendré par les autres sommets de $\mathcal{S}$. L'estimation des abondances peut donc s'écrire :

$$\alpha_i = \frac{\delta(\boldsymbol{r})}{\delta(\boldsymbol{m}_i)} \tag{2.29}$$

pour $i = 1, 2, \ldots, R$. Dans les calculs ci-dessus, la contrainte de non-négativité est violée si il existe des pixels en dehors du simplexe $\mathcal{S}$ formé par les composés purs. En outre, toutes les termes de volume et de distance sont orientés. Leur signe peut être positif ou négatif en fonction de l'ordre de la séquence définie par les sommets des simplexes dans le cas des volumes, ou du côté de l'hyperplan où se trouve le pixel $\boldsymbol{r}$ pour la formule de distance.

2.5 Conclusion

Dans ce chapitre, nous avons introduit les modèles de mélange les plus utilisés. Ces modèles permettent d'élaborer des algorithmes d'extraction de composés purs et d'estimation des abondances. Les modèles et méthodes linéaires présentés dans ce chapitre demeurent basiques lorsqu'ils sont confrontés à des mélanges complexes. Des algorithmes de démélange non-linéaires sont développés dans la suite à cet effet.

Chapitre 3

Noyaux reproduisants : ingénierie et méthodes

Sommaire

Dans ce chapitre, nous présentons les concepts fondamentaux de la théorie des noyaux reproduisants [Aronszajn 1950]. Cette théorie est essentielle dans notre étude pour développer des techniques non-linéaires pour le démélange non-linéaire. Ensuite, nous présentons des méthodes basées sur cette théorie et utilisées dans cette thèse. Voir [Pothin 2007] pour une présentation détaillée.

3.1 Concepts de base

L 'idée générale des méthodes à noyau consiste à injecter les données de l'espace des observations $\mathcal{X}$ dans un espace fonctionnel $\mathcal{H}$ grâce au concours d'une application non linéaire :

$$\begin{array}{rcl} \phi : \mathcal{X} & \rightarrow & \mathcal{H} \\ \boldsymbol{x} & \mapsto & \phi(\boldsymbol{x}) \end{array} \tag{3.1}$$

Cette étape est préalable à la mise en œuvre d'un algorithme d'apprentissage dans ce nouvel espace $\mathcal{H}$. Un exemple illustrant l'intérêt d'une telle transformation des données est proposé dans la figure 3.1. Le but y est de séparer le jeu de données

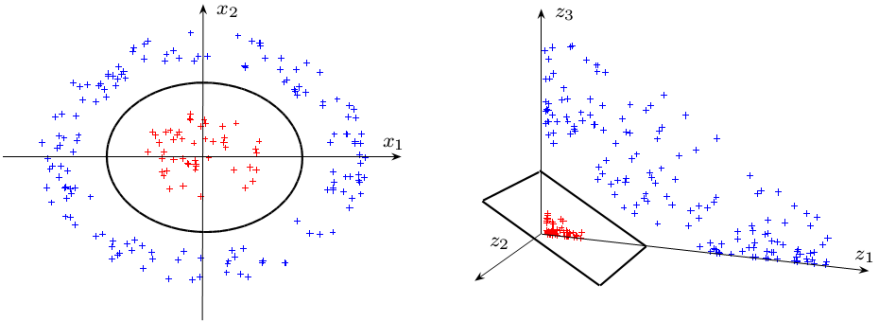


FIGURE 3.1 – Exemples de classification binaire dans $\mathbb{R}^2$. La courbe de séparation est une ellipse dans l'espace initial. Après transformation par $\phi(\boldsymbol{x}) = (x_1^2, \sqrt{2}x_1x_2, x_2^2)$, les données deviennent linéairement séparables.

en deux classes. Alors que dans l'espace d'entrée, ce problème est non-séparable au moyen d'un discriminant linéaire, il le devient en considérant la transformation $\phi(\cdot)$ qui, à tout vecteur du plan, associe les trois polynômes de degré 2 constitués à partir de ses coordonnées

$$\begin{array}{rcl} \phi : \mathbb{R}^2 & \to & \mathbb{R}^3 \\ (x_1, x_2) & \mapsto & (x_1^2, \sqrt{2}x_1x_2, x_2^2) \end{array} \tag{3.2}$$

En utilisant cette transformation, la surface discriminante dans $\mathcal{H}$ est un hyperplan séparateur. Pour $\boldsymbol{x} \in \mathbb{R}^p$ et ϕ une transformation fournissant les produits des composantes de $\boldsymbol{x}$ de degré d, la dimension de $\mathcal{H}$ est de $\frac{(d+p-1)!}{d!(p-1)!}$. Si cette approche est valable pour des exemples simples, dans des espaces $\mathcal{X}$ de faible dimension, elle ne peut clairement pas être appliquée à des problèmes plus complexes. Par exemple, pour $d = 5$ et $\boldsymbol{x}$ un vecteur composé des pixels d'une image de dimension 16×16, l'espace induit est de dimension 10^{10}. Cependant, pour certaines transformations ϕ et leur espace image $\mathcal{H}$ correspondant, il est possible de calculer le produit scalaire dans $\mathcal{H}$ sans avoir à expliciter ϕ. Ceci est réalisé par le biais de fonctions appelées noyaux, définies comme suit :

Définition 1 *Un noyau est une fonction $\kappa : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ qui, pour tout $\boldsymbol{x}$, $\boldsymbol{x}'$ de $\mathcal{X}$,*

$$\kappa(\boldsymbol{x}, \boldsymbol{x}') = \langle \phi(\boldsymbol{x}), \phi(\boldsymbol{x})' \rangle_{\mathcal{H}} \tag{3.3}$$

où ϕ est une transformation de $\mathcal{X}$ vers un espace, généralement fonctionnel, $\mathcal{H}$ muni du produit scalaire $\langle \cdot, \cdot \rangle_{\mathcal{H}}$.

Reprenons l'exemple 3.1. Le produit scalaire entre deux transformées $\phi(x)$ et $\phi(z)$ peut se calculer ainsi

$$\begin{aligned} \langle \boldsymbol{x}, \boldsymbol{z} \rangle_{\mathcal{H}} &= \left(x_1^2, \sqrt{2}x_1x_2, x_2^2\right) \left(z_1^2, \sqrt{2}z_1z_2, z_2^2\right)^\top \\ &= \left((x_1, x_2)\,(z_1, z_2)^\top\right)^2 \\ &= \langle \boldsymbol{x}, \boldsymbol{z} \rangle^2 \\ &= \kappa(\boldsymbol{x}, \boldsymbol{z}) \end{aligned} \tag{3.4}$$

De manière plus générale, pour $\boldsymbol{x}$, $\boldsymbol{x}' \in \mathbb{R}^p$ et $d \in \mathbb{N}$, on peut montrer que les fonctions de la forme

$$\kappa(\boldsymbol{x}, \boldsymbol{x}') = \langle \boldsymbol{x}, \boldsymbol{x}' \rangle^d \tag{3.5}$$

correspondent au produit scalaire dans l'espace composé par tous les produits de degré d des éléments de $\boldsymbol{x}$ et $\boldsymbol{x}'$ [Poggio 1975]. Évidemment, n'importe quelle fonction κ ne peut représenter un produit scalaire. Nous explorons à la section suivante les conditions que κ doit remplir pour pouvoir avoir une telle fonction.

3.2 Caractérisation des noyaux reproduisants

Dans cette section, nous présentons formellement les propriétés des espaces associés à une fonction noyau, ainsi que les conditions qu'une fonction doit remplir pour être un noyau reproduisant. Commençons par introduire quelques définitions nécessaires pour la suite.

Définition 2 *Un espace $\mathcal{H}$ est muni d'un produit scalaire s'il existe une forme bilinéaire $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ symétrique et à valeurs réelles telle que pour tout $\boldsymbol{x} \in \mathcal{H}$, on aie $\langle \boldsymbol{x}, \boldsymbol{x}' \rangle \geq 0$, avec égalité uniquement pour $\boldsymbol{x} = 0$. On dit alors que $\mathcal{H}$ est un espace pré-hilbertien.*

Pour rappel, le produit scalaire vérifie les propriétés suivantes pour $f, g, h \in \mathcal{H}$ et $\alpha \in R$:

1. $\langle f, g \rangle_{\mathcal{H}} = \langle g, f \rangle_{\mathcal{H}}$
2. $\langle f + g, h \rangle_{\mathcal{H}} = \langle f, h \rangle_{\mathcal{H}} + \langle g, h \rangle$
3. $\langle \alpha\, f, g \rangle_{\mathcal{H}} = \alpha \langle f, g \rangle_{\mathcal{H}}$
4. $\langle f, f \rangle_{\mathcal{H}} = 0 \Leftrightarrow f = 0$

Exemple 1 *On propose ci-après plusieurs exemples de produits scalaires usuels.*

1. *Le produit scalaire usuel dont dérive la norme euclidienne dans $\mathbb{R}^p$ est donné par :*

$$\langle \boldsymbol{x}, \boldsymbol{x}' \rangle = \boldsymbol{x}^\top \boldsymbol{x}' = \sum_{i=1}^{p} x_i x_i', \qquad \forall \boldsymbol{x}, \boldsymbol{x}' \in \mathbb{R}^p$$

2. *Un produit scalaire pour l'espace des fonctions de carré intégrables $\mathcal{L}^2(\mathcal{X})$,*

$$\mathcal{L}^2(\mathcal{X}) = \{ f : \int_{\mathcal{X}} |f(\boldsymbol{x})|^2 d\boldsymbol{x} < \infty \}$$

avec $\mathcal{X}$ un ensemble compact, est donné

$$\langle f, g \rangle_{\mathcal{L}^2(\mathcal{X})} = \int_{\mathcal{X}} f(\boldsymbol{x})\, \overline{g(\boldsymbol{x})}\, d\boldsymbol{x}, \qquad \forall f, g \in \mathcal{L}^2(\mathcal{X}),$$

où $\overline{g(x)}$ désigne le complexe conjugué de $g(\boldsymbol{x})$

3. *L'espace des polynômes homogènes de degré d des composantes de $\boldsymbol{x} \in \mathbb{R}^p$, défini par*

$$\mathcal{H}_d = \{f : f = \sum_{|\alpha|=d} w_\alpha \, x^\alpha\},$$

est muni du produit scalaire dit de Weyl [Cucker & Smale 2002] :

$$\langle f, g \rangle_{\mathcal{H}_d} = \sum_{|\alpha|=d} \frac{w_\alpha v_\alpha}{C_\alpha^d},$$

pour tout f et $g \in \mathcal{H}_d$ $f = \sum w_\alpha x^\alpha$. Ici α est un ensemble d'indices tel que $\alpha = \{\alpha_1, \alpha_2, \ldots, \alpha_p\} \in (\mathbb{N} \cup \{0\})^p, |\alpha| = \sum_{j=1}^p \alpha_j, x^\alpha = x_1^{\alpha_1} \ldots x_p^{\alpha_p}$, avec

$$C_\alpha^d = \frac{d!}{\alpha_1! \ldots \alpha_p!},$$

le coefficient mutinomial associé à la paire (α, d).

Définition 3 *Un espace $\mathcal{H}$ muni d'un produit scalaire $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ est un espace de Hilbert s'il est complet pour la norme $\|f\|_{\mathcal{H}}^2 = \langle f, f \rangle_{\mathcal{H}}$, c'est-à-dire si toutes les séries de Cauchy convergent dans $\mathcal{H}$.*

Définition 4 *On appelle espace de Hilbert à noyau reproduisant (RKHS), un espace de Hilbert $\mathcal{H}$ de fonctions définies sur $\mathcal{X}$ à valeurs réelles, et telles que pour tout $\boldsymbol{x} \in \mathcal{X}$, les fonctions d'évaluation $\delta_{\boldsymbol{x}}(f)$ définies par $\delta_{\boldsymbol{x}}(f) = f(\boldsymbol{x})$ sont bornées.*

Par définition, $\mathcal{H}$ est un RKHS s'il existe un réel M tel que

$$|f(\boldsymbol{x})| < M \|f\|_{\mathcal{H}}, \qquad \forall \boldsymbol{x} \in \mathcal{X}, \forall f \in \mathcal{H} \tag{3.6}$$

Cette propriété conduit au résultat suivant

Théorème 1 *Soit $\mathcal{H}$ un espace de Hilbert de fonctions définies sur $\mathcal{X}$ et muni de la base orthonormée $\{\psi_k\}_{k=0}^\infty$. L'espace $\mathcal{H}$ est un RKHS si*

$$\sum_{k=0}^{\infty} |\psi_k(\boldsymbol{x})|^2 < \infty, \tag{3.7}$$

pour tout $\boldsymbol{x} \in \mathcal{X}$. Le produit scalaire associé à $\mathcal{H}$ est alors défini par :

$$\kappa(\boldsymbol{x}, \boldsymbol{x}') = \sum_{k=0}^{\infty} \psi_k(\boldsymbol{x}) \, \psi_k(\boldsymbol{x}') \tag{3.8}$$

Propriété 1 *Soit $\mathcal{H}$ un RKHS de noyau κ. La propriété reproduisante de Aronszajin [Aronszajn 1950] est définie pour $f \in \mathcal{H}$ par :*

$$\langle \kappa(\boldsymbol{x}, \cdot), f \rangle_{\mathcal{H}} = f(\boldsymbol{x}) \tag{3.9}$$

En particulier,

$$\langle \kappa(\boldsymbol{x}, \cdot), \kappa(\boldsymbol{x}', \cdot) \rangle_{\mathcal{H}} = \kappa(\boldsymbol{x}, \boldsymbol{x}') \tag{3.10}$$

Définition 5 *Une matrice réelle* $\mathbf{K}$ *de taille* $n \times n$ *et de terme général* $\mathbf{K}_{ij}$ *est dite semi-définie positive si elle est symétrique et vérifie*

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j \mathbf{K}_{ij} \geq 0 \tag{3.11}$$

pour toute séquence de nombres réels $\{\alpha_i\}_{i=1,2,\ldots,n}$. *Si, en outre, l'inégalité est stricte, alors* $\mathbf{K}$ *est dite définie positive.*

De façon équivalente, les valeurs propres d'une matrice semi-définie positive sont toutes positives ou nulles, et celles d'une matrice définie positive sont strictement positives.

Définition 6 *On appelle matrice de Gram la matrice semi-définie positive dont le terme général peut s'écrire* $\mathbf{K}_{ij} = \kappa(\boldsymbol{x}_i, \boldsymbol{x}_j)$

Définition 7 *Une fonction* $\kappa : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$ *qui, pour toute séquence* $\{\boldsymbol{x}_i\}_{i=1,2,\ldots,n}$, *donne lieu à une matrice de Gram, est dite semi-définie positive.*

On note ici que n'importe quel produit scalaire est semi-défini positif.

3.2.1 Théorème de Moore-Aronszajn

D'après les propriétés d'un produit scalaire, un noyau κ doit de toute évidence être symétrique, et vérifier l'inégalité de Cauchy-Schwartz :

$$\kappa(\boldsymbol{x}, \boldsymbol{x}')^2 = \langle \phi(\boldsymbol{x}), \phi(\boldsymbol{x}') \rangle_{\mathcal{H}}^2 \leq \|\phi(\boldsymbol{x})\|_{\mathcal{H}}^2 \, \|\phi(\boldsymbol{x}')\|_{\mathcal{H}}^2 = \kappa(\boldsymbol{x}, \boldsymbol{x}) \, \kappa(\boldsymbol{x}', \boldsymbol{x}') \tag{3.12}$$

Toutefois, ces deux conditions ne garantissent pas à elles seules l'existence d'un espace $\mathcal{H}$ associé à κ. Le théorème suivant fournit une condition nécessaire et suffisante d'admissibilité d'une fonction à noyau :

Théorème 2 *A toute fonction* κ *semi-définie positive sur* $\mathcal{X} \times \mathcal{X}$, *il correspond un RKHS unique de fonctions définies sur* $\mathcal{X}$, *à valeurs réelles, et réciproquement.*

Preuve Un moyen de construire l'espace de Hilbert à noyau reproduisant associé à κ est de considérer la transformation ϕ suivante :

$$\begin{array}{rcl} \phi : \mathcal{X} & \to & \mathcal{H} \\ \boldsymbol{x} & \mapsto & \kappa(\boldsymbol{x}, \cdot) \end{array}$$

où $\phi(\boldsymbol{x}) = \kappa(\boldsymbol{x}, \cdot)$ désigne une fonction semi-définie positive sur $\mathcal{X}$, obtenue en fixant le premier argument de ϕ à $\boldsymbol{x}$, et $\mathcal{H}$ l'espace engendré par les fonctions $\kappa(\boldsymbol{x}, \cdot)$ avec $\boldsymbol{x} \in \mathcal{X}$. Étant donné deux fonctions de $\mathcal{H}$

$$f(\cdot) = \sum_{i=1}^{n} \alpha_i \kappa(\boldsymbol{x}_i, \cdot) \qquad g(\cdot) = \sum_{j=1}^{n'} \beta_j \kappa(\boldsymbol{x}'_j, \cdot) \tag{3.13}$$

avec $n, n' \in \mathbb{N}$, $\alpha_i, \beta_j \in \mathbb{R}$ et $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$, on considère la forme bilinéaire suivante :

$$\langle f, g \rangle_{\mathcal{H}} = \sum_{i=1}^{n} \sum_{j=1}^{n'} \alpha_i \beta_j \kappa(\boldsymbol{x}_i, \boldsymbol{x}'_j) \tag{3.14}$$

dont on cherche à démontrer qu'il s'agit d'un produit scalaire dans $\mathcal{H}$. On note en premier lieu qu'en utilisant les définitions de f, g et la symétrie de κ, on peut mettre cette expression sous la forme :

$$\langle f, g \rangle_{\mathcal{H}} = \sum_{j=1}^{n'} \beta_j f(\boldsymbol{x}_j) \sum_{i=1}^{n} \alpha_i g(\boldsymbol{x}_i) \tag{3.15}$$

De (3.15), il est clair que $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ est à valeur réelle, symétrique et bilinéaire. Par ailleurs, comme κ est semi-défini positif, on a :

$$\langle f, f \rangle_{\mathcal{H}} = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j \kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) \qquad \forall f \in \mathcal{H} \tag{3.16}$$

Par conséquent, pour toute famille des fonctions h_i et de coefficients $\gamma_i \in \mathbb{R}$, on a

$$\sum_{i=1}^{N} \sum_{j=1}^{N} \gamma_i \gamma_j \langle h_i, h_j \rangle_{\mathcal{H}} = \langle \sum_{i=1}^{N} \gamma_i h_i, \sum_{j=1}^{N} \gamma_j h_j \rangle_{\mathcal{H}} \geq 0 \tag{3.17}$$

ce qui montre que $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ est semi-défini positif. Il reste enfin à prouver la propriété

$$\langle f, f \rangle_{\mathcal{H}} = 0 \Leftrightarrow f = 0 \qquad \forall f \in \boldsymbol{H} \tag{3.18}$$

pour démontrer que (3.14) est un produit scalaire. Or, d'après l'équation (3.15), toute fonction $f \in \mathcal{H}$ vérifie :

$$\langle f, \kappa(\boldsymbol{x}, \cdot) \rangle_{\mathcal{H}} = \sum_{i=1}^{n} \alpha_i \kappa(\boldsymbol{x}, \boldsymbol{x}_i) = f(\boldsymbol{x}) \tag{3.19}$$

En particulier

$$\langle \kappa(\boldsymbol{x}, \cdot), \kappa(\boldsymbol{x}', \cdot) \rangle_{\mathcal{H}} = \kappa(\boldsymbol{x}, \boldsymbol{x}') \tag{3.20}$$

A partir des équations (3.12), (3.19) et (3.20), nous en déduisons que

$$f(\boldsymbol{x})^2 = \langle f, \kappa(\boldsymbol{x}, \cdot) \rangle_{\mathcal{H}}^2 \leq \langle f, f \rangle_{\mathcal{H}} \, \kappa(\boldsymbol{x}, \boldsymbol{x}) \tag{3.21}$$

qui prouve la propriété (3.18). Ainsi $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ est un produit scalaire sur $\mathcal{H}$. Afin de transformer $\mathcal{H}$ en un espace de Hilbert, il suffit de le compléter conformément à [Aronszajn 1950] de manière à ce que toute suite de Cauchy y converge. Grâce à (3.19), $\mathcal{H}$ est un espace de Hilbert à noyau reproduisant. D'après (3.20), la fonction κ représente un produit scalaire dans $\mathcal{H}$. Elle est appelée noyau reproduisant de $\mathcal{H}$.

La réciproque du théorème peut être démontrée en s'appuyant sur le théorème de représentation de Riesz [Reed & Simon 1980]. D'après ce théorème, il existe pour

tout $\boldsymbol{x} \in \mathcal{X}$ une fonction $\kappa(\boldsymbol{x}, \cdot) \in \mathcal{H}$ qui vérifie (3.19). Or, il résulte de (3.20) que la fonction $\kappa(\boldsymbol{x}, \boldsymbol{x}')$ définie de $\mathcal{X} \times \mathcal{X}$ dans $\mathbb{R}$ est symétrique. Elle vérifie également

$$\begin{aligned}\sum_{i=1}^{M}\sum_{j=1}^{M}\alpha_i\alpha_j\kappa(\boldsymbol{x}_i,\boldsymbol{x}_j)_{\mathcal{H}} &= \langle\sum_{i=1}^{M}\alpha_i\kappa(\boldsymbol{x}_i,\cdot),\sum_{j=1}^{M}\alpha_j\kappa(\boldsymbol{x}_j,\cdot)\rangle_{\mathcal{H}} \\ &= \|\sum_{i=1}^{M}\alpha_i\kappa(\boldsymbol{x}_i,\cdot)\|^2_{\mathcal{H}} \geq 0\end{aligned} \tag{3.22}$$

pour tout $\alpha_1, \ldots, \alpha_M \in \mathbb{R}$. Par conséquent, κ est semi-définie positive.

Le RKHS associé à κ est généralement appelé espace canonique des caractéristiques, et $\kappa(\boldsymbol{x}, \cdot)$ son application canonique associée. Notons que bien que ces deux derniers soient uniques d'après le théorème de Moore-Aronszajn, il existe une infinité de fonctions ϕ et d'espaces $\mathcal{H}$ tels que l'on ait (3.3).

3.2.2 Théorème de Mercer

On vient de montrer que toute fonction semi-définie positive correspond à un produit scalaire dans un espace de Hilbert. Dans cette section, nous présentons la condition de Mercer qui confirme ce résultat dans le cas continu, en fournissant les outils nécessaires pour expliciter l'espace dans lequel les données sont injectées.

D'après la théorie de Hilbert-Schmidt [Courant & Hilbert 1962], toute fonction continue et symétrique $\kappa(\boldsymbol{x}, \boldsymbol{x}')$ admet une décomposition de la forme

$$\kappa(\boldsymbol{x}, \boldsymbol{x}') = \sum_{i=0}^{\infty} \lambda_i \, \psi_i(\boldsymbol{x}) \, \psi_i(\boldsymbol{x}') \tag{3.23}$$

où λ_i et ψ_i sont les valeurs propres et les vecteurs propres de l'opérateur intégral

$$(T_\kappa f)(\cdot) = \int_{\mathcal{X}} \kappa(\boldsymbol{x}, \cdot) \, f(\boldsymbol{x}) \, d\boldsymbol{x} \tag{3.24}$$

Une condition suffisante pour que (3.23) soit un produit scalaire est que les coefficients λ_i soient positifs. On peut en effet dans ce cas construire aisément une transformation ϕ satisfaisant (3.3). Par exemple :

$$\begin{array}{rcl}\phi : \mathcal{X}^2 & \to & \ell^2 \\ \boldsymbol{x} & \mapsto & (\sqrt{\lambda_0}\,\psi_0(\boldsymbol{x}), \sqrt{\lambda_1}\,\psi_1(\boldsymbol{x}), \ldots)^\top\end{array} \tag{3.25}$$

On montre alors :

$$\begin{aligned}\langle\phi(\boldsymbol{x}), \phi(\boldsymbol{x})'\rangle_{\ell_2} &= \langle\left[\sqrt{\lambda_i}\psi_i(\boldsymbol{x})\right]_i, \left[\sqrt{\lambda_i}\psi_i(\boldsymbol{x}')\right]_i\rangle_{\ell_2} \\ &= \sum_i \sqrt{\lambda_i}\,\psi_i(\boldsymbol{x})\,\sqrt{\lambda_i}\,\psi_i(\boldsymbol{x}') \\ &= \sum_i \lambda_i\,\psi_i(\boldsymbol{x})\,\psi_i(\boldsymbol{x}') \\ &= \kappa(\boldsymbol{x}, \boldsymbol{x}')\end{aligned} \tag{3.26}$$

Le théorème suivant établit une condition nécessaire et suffisante pour que les paramètres λ_i soient positifs.

Théorème 3 *Pour garantir qu'une fonction κ continue et symétrique puisse s'écrire sous la forme (3.23), avec des coefficients λ_i positifs, il est nécessaire et suffisant que la condition*

$$\int_{\mathcal{X}} \int_{\mathcal{X}} \kappa(\boldsymbol{x}, \boldsymbol{x}')\, f(\boldsymbol{x})\, \overline{f(\boldsymbol{x}')}\, d\boldsymbol{x}\, d\boldsymbol{x}' \geq 0 \tag{3.27}$$

soit vérifiée pour tout $f \in \mathcal{L}^2(\mathcal{X})$

Preuve. Par commodité, notons $\mathcal{H} = \mathcal{L}^2(X)$. D'après la théorie de Hilbert-Schmidt [Courant & Hilbert 1962], les valeurs $\{\lambda_i\}_i$ et fonctions propres $\{\psi_i\}_i$ apparaissant dans (3.23) vérifient

$$\int_{\mathcal{X}} \kappa(\boldsymbol{x}, \cdot)\, \psi_i(\boldsymbol{x})\, d\boldsymbol{x} = \lambda_i\, \psi(\cdot) \tag{3.28}$$

avec $\lambda_i \in \mathbb{R}, \psi_i \in \mathcal{H}$ et $\langle \psi_i, \psi_j \rangle_{\mathcal{H}} = \delta_{ij}$. En pré-multipliant (3.28) par $\int_{\mathcal{X}} \overline{\psi_j(\boldsymbol{x})} d\boldsymbol{x}$ et en utilisant la propriété de symétrie de κ, on obtient

$$\int_{\mathcal{X}} \int_{\mathcal{X}} \kappa(\boldsymbol{x}, \boldsymbol{x}')\, \psi_i(\boldsymbol{x})\, \overline{\psi_j(\boldsymbol{x}')}\, d\boldsymbol{x}\, d\boldsymbol{x}' = \lambda_i\, \delta_{ij} \tag{3.29}$$

où la dernière égalité découle de l'orthogonalité des fonctions propres. D'après (3.29), garantir (3.27) pour toute fonction f de l'espace engendré par $\{\psi_i\}_i$ est nécessaire et suffisant pour que les valeurs propres λ_i soient positives. Pour prouver le théorème, il reste à montrer que ceci reste vrai pour tout $f \in \mathcal{H}$. Or, $\{\psi_i\}_i$ forme une base orthonormée de $\mathcal{H}$. Toute fonction $f \in \mathcal{H}$ peut donc se décomposer sous la forme :

$$f = \sum_i \alpha_i \psi_i \tag{3.30}$$

où les α_i sont des coefficients réels. En combinant (3.30) et (3.29), on trouve

$$\begin{aligned} \int_{\mathcal{X}} \int_{\mathcal{X}} \kappa(\boldsymbol{x}, \boldsymbol{x}')\, f(\boldsymbol{x})\, \overline{f(\boldsymbol{x}')}\, d\boldsymbol{x}\, d\boldsymbol{x}' &= \sum_i \sum_j \alpha_i\, \alpha_j \int_{\mathcal{X}} \int_{\mathcal{X}} \kappa(\boldsymbol{x}, \boldsymbol{x}')\, \psi_i(\boldsymbol{x})\, \overline{\psi_j(\boldsymbol{x}')}\, d\boldsymbol{x}\, d\boldsymbol{x}' \\ &= \sum_i \lambda_i\, \alpha_i^2 \end{aligned} \tag{3.31}$$

Supposons que l'opérateur intégral dans cette expression soit négatif. Les coefficients α_i^2 apparaissant dans le terme de droite, cela implique qu'il existe au moins une valeur propre $\lambda_i < 0$. En imposant la positivité de l'opérateur intégral pour tout $f \in \mathcal{H}$, on impose ainsi la positivité des coefficients λ_i comme annoncé par le théorème.

La condition (3.27) stipule que toute matrice de Gram générée par κ doit être semi-définie positive. En d'autres termes, κ doit être une fonction semi-définie positive, en accord avec le résultat de Moore-Aronszajn. Les fonctions κ qui vérifient la propriété (3.27) sont généralement appelés noyaux de Mercer.

Remarque 1 *Le théorème de Moore-Aronjazn s'applique aux fonctions κ continues ou définies point par point. Le théorème de Mercer ne s'applique qu'aux fonctions κ continues.*

3.2.3 Théorème du représentant

En formulant un problème d'optimisation dans un RKHS $\mathcal{H}$, on montre, sous certaines conditions, que la solution optimale peut s'exprimer sous la forme d'une combinaison de fonctions de base, indépendamment de la dimension de $\mathcal{H}$. Ceci est formalisé par le théorème suivant.

Théorème 4 *Soit $\mathcal{H}$ un RKHS de noyau κ, $\Omega : [0, \infty) \to \mathbb{R}$ une fonction monotone strictement décroissante, $\mathcal{X}$ un ensemble et $c : (\mathcal{X}, \mathbb{R}^2)^n \to \mathbb{R} \cup \{\infty\}$ une fonction coût. La fonction $f \in \mathcal{H}$ minimisant le risque régularisé*

$$c((\boldsymbol{x}_1, y_1, f(\boldsymbol{x}1)), \ldots, (\boldsymbol{x}_n, y_n, f(\boldsymbol{x}_n))) + \Omega(\|f\|_{\mathcal{H}}) \tag{3.32}$$

admet une représentation de la forme $f(\boldsymbol{x}) = \sum_{i=1}^{n} \alpha_i \kappa(\boldsymbol{x}_i, \boldsymbol{x})$.

3.3 Construction d'un noyau

Nous avons vu au cours de la section précédente que la condition de (semi)-définie positivité est nécessaire et suffisante pour décider si une fonction candidate est un noyau valide. Une conséquence importante de ce résultat est qu'il justifie une série de règles permettant de combiner des noyaux en un noyau mieux adapté aux données. Un exemple rudimentaire de noyau construit à partir d'un noyau de référence est le noyau normalisé :

$$\tilde{\kappa}(\boldsymbol{x}, \boldsymbol{x}') = \frac{\kappa(\boldsymbol{x}, \boldsymbol{x}')}{\sqrt{\kappa(\boldsymbol{x}, \boldsymbol{x}) \kappa(\boldsymbol{x}', \boldsymbol{x}')}} \tag{3.33}$$

L'interprétation géométrique d'une telle transformation est donnée plus loin. D'autres transformations sont possibles. Commençons par énoncer les plus élémentaires.

3.3.1 Règles simples

Lemme 1 *Soit κ_1 et κ_2 deux noyaux définis sur $\mathcal{X} \times \mathcal{X}$, $a \in \mathbb{R}_+$, f une fonction à valeurs réelles, $\boldsymbol{B}$ une matrice semi-définie positive, κ_3 un noyau défini sur $\mathcal{Z} \times \mathcal{Z}$ et $g : \mathcal{X} \mapsto \mathcal{Z}$. Dans ce cas, toutes les fonctions suivantes sont des noyaux :*

1. $\kappa(\boldsymbol{x}, \boldsymbol{x}') = \kappa_1(\boldsymbol{x}, \boldsymbol{x}') + \kappa_2(\boldsymbol{x}, \boldsymbol{x}')$
2. $\kappa(\boldsymbol{x}, \boldsymbol{x}') = a\kappa_1(\boldsymbol{x}, \boldsymbol{x}')$
3. $\kappa(\boldsymbol{x}, \boldsymbol{x}') = \kappa_1(\boldsymbol{x}, \boldsymbol{x}')\kappa_2(\boldsymbol{x}, \boldsymbol{x}')$
4. $\kappa(\boldsymbol{x}, \boldsymbol{x}') = f(\boldsymbol{x})f(\boldsymbol{x}')$
5. $\kappa(\boldsymbol{x}, \boldsymbol{x}') = \boldsymbol{x}^\top \boldsymbol{B} \boldsymbol{x}'$
6. $\kappa(\boldsymbol{x}, \boldsymbol{x}') = \kappa(g(\boldsymbol{x}), g(\boldsymbol{x}'))$

3.3.2 Règles avancées

Il a été montré dans [Hofmann *et al.* 2005] que la classe des fonctions noyaux était fermée aux opérations énoncées dans le lemme précédent, ce qui signifie qu'il n'existe pas d'autres opérations sur les fonctions conservant la propriété de semi-définie positivité. Nous répertorions à présent quelques règles « avancées » qui combinent ces opérations élémentaires.

Corollaire 1 *Etant donné κ_1 un noyau et f une fonction à valeurs réelles, la fonction :*

$$\kappa(\boldsymbol{x}, \boldsymbol{x}') = f(\boldsymbol{x})\, f(\boldsymbol{x}')\, \kappa_1(\boldsymbol{x}, \boldsymbol{x}') \tag{3.34}$$

est un noyau.

Par application des règles 3) et 4), on prouve que (3.34) est un noyau pour tout f à valeurs réelles. A fortiori, pour f à valeurs positives, (3.34) est un noyau. On parle dans ce cas de noyau (quasi)-isogone. Un noyau isogone conserve les angles entre les vecteurs transformés.

Remarque 2 *Le noyau normalisé (3.33) est un noyau isogone où $f(\boldsymbol{x}) = \frac{1}{\sqrt{\kappa(\boldsymbol{x},\boldsymbol{x})}}$*

Corollaire 2 *Soit m noyaux $\{\kappa_1, \kappa_2, \ldots, \kappa_m\}$ définis sur $\mathcal{X} \times \mathcal{X}$, et $\boldsymbol{\alpha} \in \mathbb{R}_+^m$. Les fonctions suivantes :*

$$\begin{aligned} \kappa(\boldsymbol{x}, \boldsymbol{x}') &= \sum_{i=1}^{m} \alpha_i\, \kappa(\boldsymbol{x}, \boldsymbol{x}') \\ \kappa(\boldsymbol{x}, \boldsymbol{x}') &= \prod_{i=1}^{m} \kappa_i(\boldsymbol{x}, \boldsymbol{x}') \end{aligned} \tag{3.35}$$

sont des noyaux.

Ces propriétés impliquent que la fonction

$$\kappa(\boldsymbol{x}, \boldsymbol{x}') = p^+(\kappa_1(\boldsymbol{x}, \boldsymbol{x}')) \tag{3.36}$$

où $p^+(\cdot)$ est un polynôme à coefficients positifs et κ_1 un noyau sur $\mathcal{X} \times \mathcal{X}$ est un noyau. En particulier, en prenant la limite du développement en série de la fonction exponentielle, on prouve que :

$$\kappa(\boldsymbol{x}, \boldsymbol{x}') = \exp(\kappa_1(\boldsymbol{x}, \boldsymbol{x}')) \tag{3.37}$$

est un noyau. De (3.37), on en déduit que $\exp(\langle \boldsymbol{x}, \boldsymbol{x}' \rangle / \sigma^2)$ est un noyau pour $\sigma \in \mathbb{R}_+$. Après normalisation (voir (3.33)), on obtient le noyau gaussien [Aizeman *et al.* 1964, Boser *et al.* 1992] :

$$\frac{\exp(\langle \boldsymbol{x}, \boldsymbol{x}' \rangle / \sigma^2)}{\sqrt{\exp(\|\boldsymbol{x}\|^2/\sigma^2) \exp(\|\boldsymbol{x}'\|^2/\sigma^2)}} = \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{x}'\|^2}{2\sigma^2}\right) \tag{3.38}$$

Corollaire 3 *Soit* $\boldsymbol{x} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N]^\top$ *et* $\boldsymbol{x}' = [\boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_N]^\top \in \mathcal{X}$. *Si* $\kappa_1, \ldots, \kappa_N$ *sont des noyaux définis sur* $\mathcal{X}_1 \times \mathcal{X}_1, \ldots, \mathcal{X}_N \times \mathcal{X}_N$ *respectivement, avec* $\boldsymbol{x}_i$, $\boldsymbol{x}'_i \in \mathcal{X}$, *alors leur somme directe* $\kappa_1 \oplus \cdots \oplus \kappa_N$ *définie sur* $\mathcal{X} \times \mathcal{X}$ *par :*

$$\kappa_1 \oplus \cdots \oplus \kappa_N(\boldsymbol{x}, \boldsymbol{x}') = \kappa_1(\boldsymbol{x}_1, \boldsymbol{x}'_1) + \cdots + \kappa_N(\boldsymbol{x}_N, \boldsymbol{x}'_N) \tag{3.39}$$

est un noyau. De manière similaire, leur produit tensoriel $\kappa_1 \otimes \kappa_N$ *défini par*

$$\kappa_1 \otimes \cdots \otimes \kappa_N(\boldsymbol{x}, \boldsymbol{x}') = \kappa_1(\boldsymbol{x}_1, \boldsymbol{x}'_1) \times \ldots \times \kappa_N(\boldsymbol{x}_N, \boldsymbol{x}'_N) \tag{3.40}$$

est un noyau.

Remarque 3 *La somme directe et le produit tensoriel de noyaux combinent des noyaux définis sur des espaces d'entrées pouvant être différents. Cela permet de traiter des données de nature différente par exemple.*

Corollaire 4 *Etant donné* f *une fonction à valeurs réelles et* $q(f)$ *une distribution de probabilité dans la classe de fonctions* $\mathcal{F}$, *la fonction suivante est un noyau :*

$$\kappa(\boldsymbol{x}, \boldsymbol{x}') = \int_{\mathcal{F}} f(\boldsymbol{x})\, f(\boldsymbol{x}')\, q(f)\, df \tag{3.41}$$

Ce résultat peut être démontré en notant que, puisque $f(\boldsymbol{x})f(\boldsymbol{x}')$ est un noyau pour tout $f \in \mathcal{F}$ et que $q(f) \geq 0$, la limite de l'opérateur intégral est une combinaison linéaire à coefficients positifs de noyaux. Les noyaux construits selon la règle (3.45) sont généralement appelés noyaux covariants. Pour $\mathcal{F}$ l'ensemble des fonctions à valeurs $\{\pm 1\}$, il est facile de montrer que le noyau obtenu est normalisé :

$$\kappa(\boldsymbol{x}, \boldsymbol{x}') = \int_{\mathcal{F}} f(\boldsymbol{x})\, f(\boldsymbol{x}')\, q(f)\, df = \int_{\mathcal{F}} q(f)\, df = 1 \tag{3.42}$$

3.3.3 Effet sur le feature space associé

Nous venons de voir comment créer de nouveaux noyaux à partir de noyaux existants. Il est souvent utile de comprendre l'effet qu'a une certaine opération sur l'espace image associé. Par exemple, si l'on considère la somme de deux noyaux, on peut écrire :

$$\begin{aligned} \kappa_1(\boldsymbol{x}, \boldsymbol{x}') + \kappa_2(\boldsymbol{x}, \boldsymbol{x}') &= \langle \phi_1(\boldsymbol{x}), \phi_1(\boldsymbol{x}') \rangle_{\mathcal{H}_1} + \langle \phi_2(\boldsymbol{x}), \phi_2(\boldsymbol{x}') \rangle_{\mathcal{H}_2} \\ &= \left\langle \begin{pmatrix} \phi_1(\boldsymbol{x}) \\ \phi_2(\boldsymbol{x}) \end{pmatrix}, \begin{pmatrix} \phi_1(\boldsymbol{x}') \\ \phi_2(\boldsymbol{x}') \end{pmatrix} \right\rangle_{\mathcal{H}} \\ &= \langle \phi(\boldsymbol{x}), \phi(\boldsymbol{x}') \rangle_{\mathcal{H}} \end{aligned} \tag{3.43}$$

qui signifie que $\phi(\cdot) \in \mathcal{H}$ est la concaténation des vecteurs $\phi_1(\cdot) \in \mathcal{H}_1$ et $\phi_2(\cdot) \in \mathcal{H}_2$. Il est important de souligner que les interprétations pour des noyaux ne sont pas uniques. De l'équation (3.43) par exemple, on peut proposer $\phi(\boldsymbol{x}) = (\phi_2(\boldsymbol{x}), \phi_1(\boldsymbol{x}))$

montrant ainsi que ϕ n'est pas unique. Un autre exemple qui montre que l'espace image lui-même n'est pas unique est donné par $\kappa_1 + \kappa_1 = 2\kappa_1$.

$$2\kappa_1(\boldsymbol{x}, \boldsymbol{x}') = 2\langle \phi_1(\boldsymbol{x}), \phi(\boldsymbol{x}') \rangle_{\mathcal{H}} = \langle \sqrt{2}\phi_1(\boldsymbol{x}), \sqrt{2}\phi(\boldsymbol{x}') \rangle_{\mathcal{H}} \tag{3.44}$$

En effet, puisque qu'il existe au moins deux interprétations différentes possibles pour $\phi(\boldsymbol{x})$. L'une en terme de vecteurs concaténés, l'autre en terme d'une mise à l'échelle. Une autre approche utilisable pour concevoir des noyaux consiste à transformer directement l'espace image, et chercher à exprimer le résultat en terme de produits scalaires uniquement.

Exemple 2 *Le noyau normalisé* (3.33) *est obtenu par normalisation des vecteurs* $\phi(\cdot)$. *En effet :*

$$\left\langle \frac{\phi(\boldsymbol{x})}{\|\phi(\boldsymbol{x})\|_{\mathcal{H}}}, \frac{\phi(\boldsymbol{x}')}{\|\phi(\boldsymbol{x}')\|_{\mathcal{H}}} \right\rangle_{\mathcal{H}} = \frac{\langle \phi(\boldsymbol{x}), \phi(\boldsymbol{x}') \rangle_{\mathcal{H}}}{\|\phi(\boldsymbol{x})\|_{\mathcal{H}} \|\phi(\boldsymbol{x}')\|_{\mathcal{H}}} = \frac{\kappa(\boldsymbol{x}, \boldsymbol{x}')}{\sqrt{\kappa(\boldsymbol{x}, \boldsymbol{x})\kappa(\boldsymbol{x}', \boldsymbol{x}')}} = \tilde{\kappa}(\boldsymbol{x}, \boldsymbol{x}') \tag{3.45}$$

On mesure ainsi le cosinus de l'angle θ *entre les deux vecteurs* $\phi(\boldsymbol{x})$ *et* $\phi(\boldsymbol{x}')$.

Exemple 3 *Considérons l'opération de centrage, qui transforme* $\phi(\boldsymbol{x})$ *en* $\bar{\phi}(\boldsymbol{x})$ *par translation d'un vecteur obtenu par combinaison linéaire d'individus de l'ensemble d'apprentissage. Formellement :*

$$\bar{\phi}(\boldsymbol{x}) \triangleq \phi(\boldsymbol{x}) + \sum_{i=1}^{n} \gamma_i \, \phi(\boldsymbol{x}_i) \tag{3.46}$$

où $\gamma_i \in \mathbb{R}, i = 1, \ldots, n$. *On obtient pour produit scalaire :*

$$\begin{aligned}
\langle \bar{\phi}(\boldsymbol{x}), \bar{\phi}(\boldsymbol{x}') \rangle &= \left\langle \phi(\boldsymbol{x}) + \sum_{i=1}^{n} \gamma_i \phi(\boldsymbol{x}_i), \phi(\boldsymbol{x}') + \sum_{i=1}^{n} \gamma_i \phi(\boldsymbol{x}'_i) \right\rangle \\
&= \langle \phi(\boldsymbol{x}), \phi(\boldsymbol{x}') \rangle - \sum_{i=1}^{n} \gamma_i \langle \phi(\boldsymbol{x}), \phi(\boldsymbol{x}_i) \rangle - \sum_{i=1}^{n} \gamma_i \langle \phi(\boldsymbol{x}_i), \phi(\boldsymbol{x}') \rangle \\
&\quad + \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_i \gamma_j \langle \phi(\boldsymbol{x}_i), \phi(\boldsymbol{x}_j)) \rangle \\
&= \kappa(\boldsymbol{x}, \boldsymbol{x}') - \sum_{i=1}^{n} \gamma_i \kappa(\boldsymbol{x}, \boldsymbol{x}_i) \\
&\quad - \sum_{i=1}^{n} \gamma_i \kappa(\boldsymbol{x}', \boldsymbol{x}_i) + \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_i \gamma_j \kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) \\
&\triangleq \bar{\kappa}(\boldsymbol{x}, \boldsymbol{x}')
\end{aligned} \tag{3.47}$$

Par construction, $\bar{\kappa}$ *est toujours valide. Dans le cas particulier* $\gamma_i = 1/n$, *on retrouve la méthode de centrage de données [Cristianini 2001]. Pour un problème de classification bi-classe, choisir*

$$\gamma_i = 1/n^+ \tag{3.48}$$

où n^+(resp. n^-) est le nombre d'individus d'étiquettes $y_i = +1$(resp. -1), permet de positionner l'origine des données à mi-distance des centres d'inerties des deux classes.

Remarque 4 *La matrice de Gram $\bar{\mathbf{K}}$ associée au noyau centré $\bar{\kappa}$ peut se noter :*

$$\bar{\mathbf{K}} = \mathbf{K} - \mathbf{\Gamma}^\top \mathbf{K} - \mathbf{K}\mathbf{\Gamma} + \mathbf{\Gamma}^\top \mathbf{K}\mathbf{\Gamma} \tag{3.49}$$

où $\mathbf{\Gamma} = [\boldsymbol{\gamma}, \ldots, \boldsymbol{\gamma}] \in \mathbb{R}^{n\times n}$, $\boldsymbol{\gamma}$ le vecteur colonne tel que $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_n) \in \mathbb{R}^n$. Dans le cas particulier $\gamma_i = 1/n$, en notant $\mathbf{1}$ la matrice symétrique de dimension $n \times n$ de terme général 1, on obtient :

$$\bar{\mathbf{K}} = \mathbf{K} - \frac{1}{n}\mathbf{1}\mathbf{K} - \frac{1}{n}\mathbf{K}\mathbf{1} + \frac{1}{n^2}\mathbf{1}\mathbf{K}\mathbf{1} \tag{3.50}$$

3.4 Exemples des noyaux

Dans cette section, nous citons quelques exemples de noyaux susceptibles d'être utilisés en pratique : les noyaux projectifs et les noyaux stationnaires. Ces noyaux sont principalement utilisés sur des données vectorielles.

3.4.1 Noyaux projectifs

Les noyaux projectifs sont fonction du produit scalaire des observations. Une condition nécessaire pour qu'une fonction de type $\kappa(\boldsymbol{x}, \boldsymbol{x}') = \kappa(\langle \boldsymbol{x}, \boldsymbol{x}' \rangle)$ soit un noyau est qu'elle vérifie pour tout $\xi \geq 0$ [Schölkopf *et al.* 1999]

$$\begin{aligned} \kappa(\xi) &\geq 0 \\ \delta_\xi \kappa(\xi) &\geq 0 \\ \delta_\xi \kappa(\xi) + \xi \delta_\xi^2 \kappa(\xi) &\geq 0 \end{aligned} \tag{3.51}$$

où δ_ξ désigne l'opérateur de dérivée partielle par rapport à ξ. Cette condition est nécessaire, mais pas suffisante. Pour des noyaux définis sur la sphère unité, une condition nécessaire et suffisante est donnée par le théorème suivant :

Théorème 5 *[Schoenberg 1942, Ovari 2000] Soit $\kappa(\langle \boldsymbol{x}, \boldsymbol{x}' \rangle)$ défini sur la sphère unité dans l'espace de Hilbert $\mathcal{H}$.*

1. *Si $\mathcal{H}$ est de dimension infinie, alors κ est un noyau si son développement en série de Taylor est à coefficients de Taylor positifs uniquement,*

$$\kappa(\xi) = \sum_{i=0}^{\infty} \alpha_i \xi^i, \qquad \alpha_i \geq 0 \tag{3.52}$$

2. *Si $\mathcal{H}$ est de dimension finie, alors κ est un noyau si son développement en polynôme de Legendre P_i^N est à coefficients positifs uniquement,*

$$\kappa(\xi) = \sum_{i=0}^{\infty} \alpha_i P_i^N, \qquad \alpha_i \geq 0 \tag{3.53}$$

Les noyaux projectifs les plus répandus sont les noyaux polynomiaux homogènes et non-homogènes [Boser *et al.* 1992, Vapnik 1995], que nous définissons formellement dans les exemples 4 et 5 qui suivent.

Exemple 4 *Le noyau polynomial homogène*

$$\kappa(\boldsymbol{x}, \boldsymbol{x}') = \langle \boldsymbol{x}, \boldsymbol{x}' \rangle^d \tag{3.54}$$

défini sur $\mathcal{X} \times \mathcal{X}$ *avec* $\mathcal{X} \subset \mathbb{R}^p$, $\mathcal{X} \neq \emptyset$ *et* $d \in \mathbb{N}_+$.

La validité de ces noyaux est facilement démontrée en appliquant la règle (3) du Lemme 1 autant de fois que nécessaire. Pour mieux comprendre le rôle de l'exposant d, explicitons une transformation ϕ possible associée à ce noyau. Par application du théorème binomial, on détermine :

$$\langle \boldsymbol{x}, \boldsymbol{x}' \rangle^d = \left(\sum_{j=1}^{p} x_j x'_j \right)^d = \sum_{j_1, \ldots, j_d = 1}^{p} x_{j_1} \ldots x_{j_d} x'_{j_1} \ldots x'_{j_d} = \sum_{|\alpha| = d} C_d^\alpha x^\alpha x'^\alpha \tag{3.55}$$

d'où l'on reconnaît une transformation dans l'espace des polynômes de degré d. Puisque $C_d^\alpha > 0$, poser :

$$\begin{aligned} \phi : \mathbb{R}^p &\to \mathcal{H}_d \\ \boldsymbol{x} &\mapsto \{\sqrt{C_d^\alpha} x^\alpha\}_{|\alpha| = d} \end{aligned} \tag{3.56}$$

nous permet de trouver immédiatement $\langle \boldsymbol{x}, \boldsymbol{x}' \rangle^d = \langle \phi(\boldsymbol{x}), \phi(\boldsymbol{x}) \rangle_{\mathcal{H}_d}$. Par exemple, pour $\boldsymbol{x} \in \mathbb{R}^2$:

$$\begin{aligned} (d = 1) &: \phi(\boldsymbol{x}) = (x_1, x_2) \\ (d = 2) &: \phi(\boldsymbol{x}) = (x_1^2, \sqrt{2} x_1 x_2, x_2^2) \\ (d = 3) &: \phi(\boldsymbol{x}) = (x_1^3, \sqrt{3} x_1^2 x_2, \sqrt{3} x_1 x_2^2, x_2^3) \\ (d = 4) &: \phi(\boldsymbol{x}) = (x_1^4, \sqrt{4} x_1^3 x_2, \sqrt{6} x_1^2 x_2^2, \sqrt{4} x_1 x_2^3, x_2^4) \\ &\vdots \end{aligned}$$

Par dénombrement, on montre [Poggio 1975, Taylor & Cristianini 2004] :

$$\dim(\mathcal{H}_{d+}) = \binom{p - 1 + d}{p - 1}_{\mathbb{N}} \tag{3.57}$$

où $(\cdot)_{\mathbb{N}}$ est le coefficient binomial des entiers naturels.

Exemple 5 *Le noyau polynomial non-homogène*

$$\kappa(\boldsymbol{x}, \boldsymbol{x}') = (\langle \boldsymbol{x}, \boldsymbol{x}' \rangle + \theta)^d \tag{3.58}$$

défini sur $\mathcal{X} \times \mathcal{X}$ *avec* $\mathcal{X} \subset \mathbb{R}^p$, $\mathcal{X} \neq \emptyset$ *et* $d \in \mathbb{N}_+$ *et* $\theta > 0$

On a pour expansion multinomiale :

$$(\langle \boldsymbol{x}, \boldsymbol{x}' \rangle + \theta)^d = \sum_{s=0}^{d} \binom{d}{s}_{\mathbb{N}} \theta^{d-s} \langle \boldsymbol{x}, \boldsymbol{x}' \rangle^s = \sum_{|\alpha| \leq d} \binom{d}{|\alpha|}_{\mathbb{N}} \theta^{d-|\alpha|} C_d^\alpha x^\alpha x'^\alpha$$

d'où l'on reconnaît une combinaison linéaire, à coefficients positifs, de noyaux homogènes. En appliquant (3.35), on trouve que (3.58) est bien un noyau. De manière similaire à l'exemple précédent, on montre qu'un espace image associé est :

$$\begin{aligned} \phi : \mathbb{R}^p &\to \mathcal{H}_{d+} \\ \boldsymbol{x} &\mapsto \{\sqrt{\binom{d}{|\alpha|}_{\mathbb{N}} \theta^{d-|\alpha|} C_d^\alpha} x^\alpha\}_{|\alpha| \leq d} \end{aligned} \tag{3.59}$$

où $\mathcal{H}_{d+}$ est l'espace des produits des éléments de $\boldsymbol{x}$ d'ordre inférieur ou égal à d. Par dénombrement [Taylor & Cristianini 2004] :

$$\dim(\mathcal{H}_{d+}) = \binom{p+d}{d}_{\mathbb{N}} \tag{3.60}$$

Remarque 5 *Le poids des composantes d'un noyau polynomial est distribué selon* $\sqrt{C_\alpha^d}$ *dans le cas homogène. Dans le cas non-homogène, augmenter le paramètre* θ *permet de diminuer la contribution relative des polynômes de plus haut degré.*

Les noyaux polynômiaux homogènes et non-homogènes ne sont pas continus en d. Une question qui vient naturellement à l'esprit est de savoir si l'on peut lever cette discontinuité en considérant le cas d'un exposant d réel. Le résultat suivant montre que dans le cas polynomial non-homogène, un exposant réel conduit (sous certaines hypothèses) à une fonction qui n'est pas un noyau.

3.4.2 Noyaux stationnaires

Les noyaux stationnaires, appelés également noyaux invariants par translation, dépendent de $\boldsymbol{x} - \boldsymbol{x}'$. Le théorème suivant permet de vérifier si une fonction de type $\kappa(\boldsymbol{x}, \boldsymbol{x}') = \kappa(\boldsymbol{x} - \boldsymbol{x}')$ est un noyau [Bochner 1955]

Théorème 6 *Une fonction continue* $\kappa(\boldsymbol{x}, \boldsymbol{x}') = \kappa(\boldsymbol{x} - \boldsymbol{x}')$ *est semi-définie positive sur* $\mathcal{X}$ *si elle est de la forme*

$$\kappa(\boldsymbol{x}, \boldsymbol{x}') = \int_{\mathbb{R}^p} cos(\boldsymbol{r}^\top (\boldsymbol{x} - \boldsymbol{x}')) f(d\boldsymbol{r}) \tag{3.61}$$

où f *est une mesure positive finie sur* $\mathbb{R}^p$.

Un noyau stationnaire qui dépend de $\|\boldsymbol{x} - \boldsymbol{x}'\|$ est appelé noyau radial. Une condition suffisante pour qu'une fonction de type $\kappa(\|\boldsymbol{x} - \boldsymbol{x}'\|_{\mathcal{X}})$ soit un noyau radial est qu'elle vérifie pour tout $\xi > 0$ et tout entier $k \geq 0$ [Cucker & Smale 2002]

$$(-1)^k \delta_\xi \kappa(\xi) \geq 0 \tag{3.62}$$

Le noyau radial le plus fréquemment utilisé est le noyau gaussien déjà vu en (3.38). Celui-ci est caractérisé par un continuum de valeurs propres, ce qui signifie que les composantes de ϕ ne sont pas en nombre fini. Comme il vérifie $\kappa(\boldsymbol{x}, \boldsymbol{x}') = 1$ pour tout $\boldsymbol{x}, \boldsymbol{x}' \in \boldsymbol{X}$, les données transformées sont de norme unitaire. De plus, étant donné que $\|\phi(x)\| = 1$, on a :

$$\cos(\angle(\phi(\boldsymbol{x}), \phi(\boldsymbol{x}'))) = \langle \phi(\boldsymbol{x}), \phi(\boldsymbol{x}') \rangle = \kappa(\boldsymbol{x}, \boldsymbol{x}') > 0 \tag{3.63}$$

Par conséquent, tous les échantillons occupent le même orthant dans l'espace transformé. Une base orthonormée pour l'espace RKHS associé à ce noyau a déjà été trouvée, voir [Steinwart *et al.* 2004]. Un résultat plus modeste qui se contente d'exhiber un espace image est présenté dans [Pothin 2007].

Remarque 6 *Etant donné* $\kappa(\boldsymbol{x}, \boldsymbol{x}') = e^{-\frac{\|\boldsymbol{x}-\boldsymbol{x}'\|^2}{2\sigma^2}}$ *défini pour* $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$, $\mathcal{X}$ *un sous-ensemble non vide de* $\mathbb{R}^p$, *et* $0 < \sigma < \infty$, *on a :*

$$\kappa(\boldsymbol{x}, \boldsymbol{x}') = \langle \phi(\boldsymbol{x}), \phi(\boldsymbol{x}') \rangle_{l_2} \tag{3.64}$$

avec $\phi : \mathcal{X} \to l_2$ *la transformation définie par :*

$$\begin{aligned} \phi : \mathbb{R}^2 &\to l_2 \\ \boldsymbol{x} &\mapsto e^{-\frac{\|\boldsymbol{x}\|^2}{2\sigma^2}} \left[\sqrt{\frac{C_\alpha^p}{\sigma^{2k} k!}} x^\alpha \right]_{\|\alpha\|=k, k=0} \end{aligned} \tag{3.65}$$

où C_α^k *sont les coefficients multinomiaux et* x^α *un produit des composantes de* $\boldsymbol{x}$ *d'ordre* $|\alpha|$.

Remarque 7 *D'après* (3.65), *le paramètre* σ *influe dans la transformation* ϕ *implicite au noyau gaussien sous forme d'une pondération à la fois globale et locale des composantes* $\boldsymbol{x}$. *Ce résultat montre également que la contribution des polynômes décroit plus vite que* $k!$ *avec l'ordre* k *pour tout* $\sigma \geq 1$

3.5 Exemple d'application à la régression

En injectant les données dans un espace image de grande dimension, les noyaux permettent de développer des versions non-linéaires de méthodes linéaires. Sur le plan théorique, la fonction noyau définit un espace hilbertien, dit auto-reproduisant et isométrique par la transformation non-linéaire de l'espace initial et dans lequel est résolu le problème linéaire. L'introduction de noyaux, spécifiquement adaptés à une problématique donnée, nous permettent d'avoir une grande flexibilité pour s'adapter à des situations très diverses (reconnaissance de formes, de séquences génomiques, de caractères, détection de spams, diagnostics...). À noter que, sur le plan algorithmique, ces algorithmes sont plus pénalisés par le nombre d'observations que par le nombre de variables. Néanmoins, des versions performantes des algorithmes

permettent de prendre en compte des bases de données volumineuses dans des temps de calcul acceptables.

Les méthodes à vecteurs supports peuvent être mises en oeuvre en situation de régression, c'est-à-dire pour l'approximation de fonctions quand $\boldsymbol{Y}$ (les sorties) est quantitative. Dans le cas non linéaire, le principe consiste à rechercher une estimation de la fonction par sa décomposition sur une base fonctionnelle. Si l'on considère un terme d'attache aux données de type « moindres carrés » pénalisés, la formulation avec contraintes dans le RKHS $\mathcal{H}$ s'écrit :

$$\begin{cases} \min_{f\in\mathcal{H}} & \frac{1}{2}\|f\|^2_{\mathcal{H}} + \frac{1}{2\lambda}\sum_{i=1}^n \xi_i^2 \\ \text{avec} & f(\boldsymbol{x}_i) = y_i + \xi_i \quad i = 1,\ldots,n \end{cases} \tag{3.66}$$

La solution prend la forme d'une fonction qui est une combinaison linéaire de noyaux. Les coefficients sont les variables du problème dual données par la résolution du système linéaire suivant :

$$(\mathbf{K} + \lambda \boldsymbol{I})\boldsymbol{\alpha} = \boldsymbol{y} \tag{3.67}$$

Dans cette formulation (qui peut être reprise pour la classification et l'estimation de densités) toutes les observations sont vecteur support ($\mathcal{A} = \{1,\ldots,n\}$). Afin d'obtenir de la parcimonie, il convient de modifier le terme d'attache aux données. Parmi toutes les solutions envisageables, celle retenue dans le cadre de la support vector regression [Smola & Schölkopf 2003], est une fonction de déviation en valeur absolue tronquée (appelé aussi le cout t-insensible). Le problème avec contraintes s'écrit :

$$\begin{cases} \min_{f\in\mathcal{H},b\in\mathbb{R}} & \frac{1}{2}\|f\|^2_{\mathcal{H}} + C\sum_{i=1}^n \xi_i \\ \text{avec} & |f(\boldsymbol{x}_i) + b - y_i| \leq t + \xi_i \\ & 0 \leq \xi_i \quad i = 1,\ldots,n \end{cases} \tag{3.68}$$

et la formulation équivalente en terme de cout pénalisé est :

$$\min_{f\in\mathcal{H},b\in\mathbb{R}} \frac{1}{2}\|f\|^2_{\mathcal{H}} + C\sum_{i=1}^n \max(0, |f(\boldsymbol{x}_i) + b - y_i| - t) \tag{3.69}$$

Le dual est un programme quadratique bi-paramétrique.

3.6 Conclusion

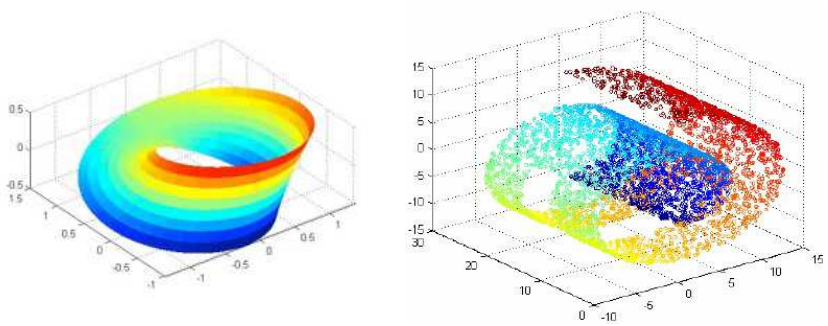
Dans ce chapitre, nous avons présenté quelques principes fondamentaux sur les méthodes à noyaux. Nous avons aussi introduit des règles pour concevoir des noyaux et donné des exemples simples de noyaux. Enfin, nous avons illustré leur mise en œuvre dans le cadre d'un problème de régression. Une telle approche sera reprise dans la suite pour le démélange des données hyperspectrales.

Chapitre 4

Démélange non-linéaire par apprentissage de variétés

Sommaire

Dans ce chapitre, nous étudions les méthodes d'apprentissage de variétés. Nous exhibons également la relation entre ces méthodes et une méthode à noyau, le KACP. Nous montrons aussi qu'une méthode d'apprentissage de variétés peut être utilisée comme une étape de réduction de dimension, en préalable à la mise en œuvre d'un algorithme de traitement linéaire, par exemple de démélange, obtenant au final ainsi une chaine de traitement non-linéaire. Une telle méthode peut également être utilisée directement pour l'extraction des composés hyperspectraux purs et l'estimation de leurs abondances grâce à des distances géodésiques. Finalement, nous présentons des résultats de simulation afin d'illustrer l'efficacité de ces principes dans le cas de mélanges fortement non-linéaires.



(a) Ruban de moëbius (b) Variété de Swiss-roll

FIGURE 4.1 – Exemples de variété.

4.1 Variétés et apprentissage de variétés

4.1.1 Notions de variétés et d'apprentissage de variétés

Une variété de dimension n, avec $n \in \mathbb{R}$, est un espace topologique localement euclidien. Par exemple, une courbe est une *variété de dimension un*, et une surface est une *variété de dimension deux*. Une variété désigne donc un ensemble de points qui appartiennent à une région qui s'apparente localement à un tel espace. Des autres exemples sont tels que nous pouvons obtenir un cercle en repliant un segment sur lui-même, un cylindre ou un cône en repliant une bande plane sur elle-même. Un autre exemple classique de variété à bord est le ruban de Moëbius 4.1a. Dans cette thèse, nous utiliserons fréquemment une variété de test « Swissroll », qui est représentée sur la figure 4.1b.

L'apprentissage de variétés est une approche récente devenue populaire pour la réduction de dimension d'un ensemble de données distribué sur une variété. Parfois, ces deux notions sont confondues. L'apprentissage de variété a pour objectif de représenter une telle distribution de données dans un espace euclidien de dimension plus faible, en conservant les propriétés caractéristiques. Les méthodes utilisées reposent sur l'idée principale que la dimension de nombreuses variétés est faible, et que celles-ci peuvent être décrites par une fonction définie à partir de quelques paramètres sous-jacents. En d'autres termes, les échantillons concernés sont intrinsèquement distribués selon un espace de dimension faible, mais plongés dans un espace de dimension élevée. Pour illustrer plus en détails la notion de variété, prenons l'exemple d'une courbe représentée dans la figure 4.2. On note que la courbe est définie dans $\mathbb{R}^3$, mais a un volume et une surface nulles. La dimension extrinsèque de « trois » est obsolète puisque la courbe peut être paramétrée par une seule variable. Une façon de formaliser cette intuition se fait par l'idée d'une variété : la courbe est une variété unidimensionnelle car elle « s'apparente localement à » $\mathbb{R}^1$.

Définition 8 *Un homéomorphisme est une fonction continue dont l'inverse est également une fonction continue*

Définition 9 *Une variété $\mathcal{M}$ de dimension d est un ensemble localement homéo-*

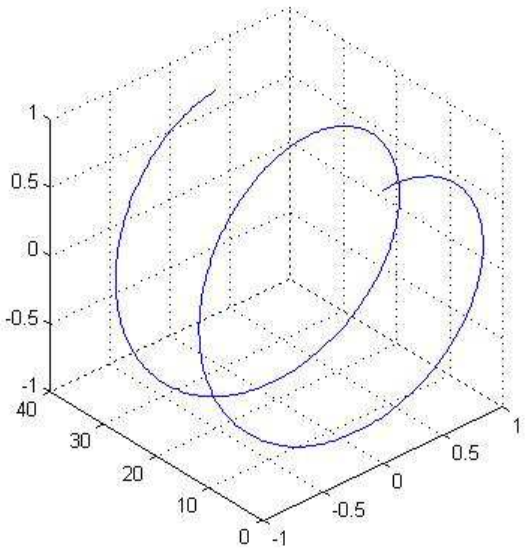


FIGURE 4.2 – Variété unidimensionnelle incorporée dans $\mathbb{R}^3$.

morphe à $\mathbb{R}^d$. *Autrement dit, pour tout* $\boldsymbol{x} \in \mathcal{M}$, *il existe un voisinage autour de* $\boldsymbol{x}$, *noté* $\mathcal{N}_{\boldsymbol{x}}$, *et un homéomorphisme* $f : \mathbb{R}_d \to \mathcal{N}_{\boldsymbol{x}}$. *Ces voisinages sont des cartes locales, et l'ensemble des cartes est appelé atlas.*

4.1.2 Apprentissage de variétés

Malgré le caractère populaire des méthodes de réduction de dimension mentionnées au Chapitre 1, elles ont de nombreuses limites étant donné leur nécessité d'opérer dans un espace euclidien, qui ne s'accommode pas d'une distribution des données dans une variété. Par exemple, pour les données Swissroll 4.1b, bien que les données soient intuitivement dans un espace de dimension 2, une ACP n'est pas en mesure d'extraire correctement cette structure bidimensionnelle. Les méthodes d'apprentissage de variétés sont le pendant de l'ACP opérant sur des variétés.

Soient des données $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n \in \mathbb{R}^D$ supposées distribuées dans une variété de dimension intrinsèque d. Pour simplifier, on suppose qu'une unique carte est suffisante pour représenter celles-ci. Le problème s'écrit :

Problème 1 *Soient des données* $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n \in \mathbb{R}^D$ *distribuées sur une variété* $\mathcal{M}$ *de dimension* d, *qui peut être décrite par une unique carte définie par* $f : \mathcal{M} \to \mathbb{R}^d$. *On recherche* $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n \in \mathbb{R}^d$ *tels que* $\boldsymbol{y}_1 \overset{def}{=} f(\boldsymbol{x}_i)$.

La figure 4.3 illustre un swissroll et son injection dans un espace à deux dimensions grâce à la méthode Isomap. On note que la relation de voisinage entre les échantillons est préservée. Ceci résulte du fait qu'il existe un homéomorphisme f de $\mathcal{M}$ dans $\mathbb{R}^2$.

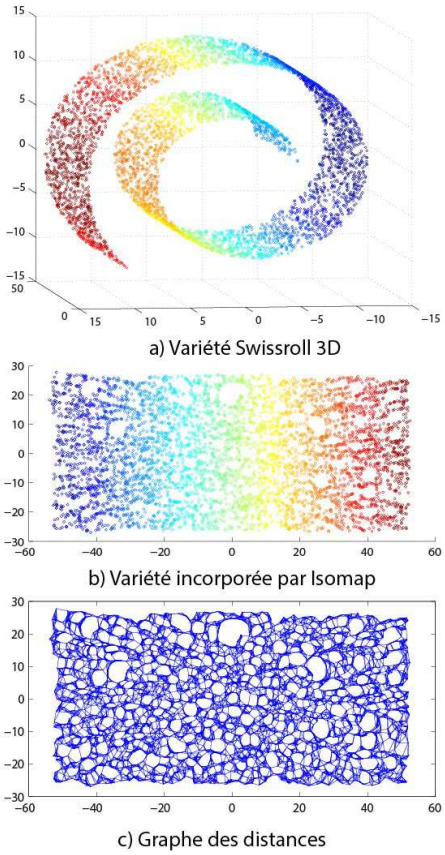


FIGURE 4.3 – Variété swissroll et son injection dans $\mathbb{R}^2$ avec Isomap [Lei *et al.* 2012]

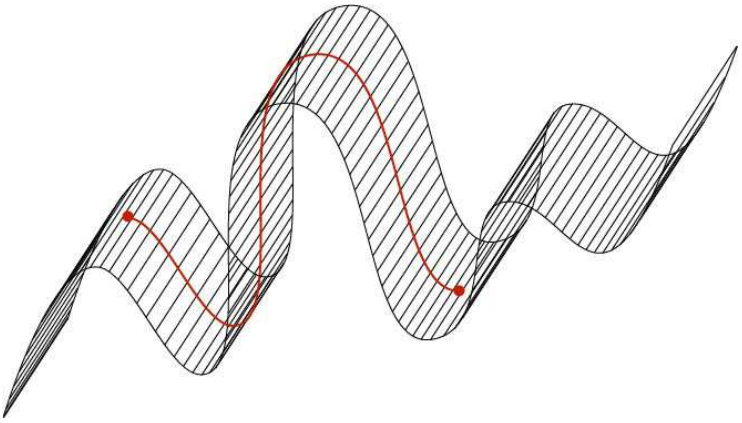


FIGURE 4.4 – Distance géodésique sur une variété surfacique.

4.2 Algorithmes d'apprentissage de variétés

Au cours de cette section, nous allons passer en revue les deux principaux algorithmes d'apprentissage de variétés existants : Isomap [de Silva & Tenenbaum 2003] et LLE [Roweis & Saul 2000]. Ceux-ci seront par la suite appliqués au traitement des données hyperspectrales. Notons ici déjà que chaque algorithme discuté nécessite la définition d'un voisinage de taille k. Au sein de chaque voisinage, la variété peut être assimilée à un espace euclidien.

4.2.1 Méthode ISOMAP

Isomap [de Silva & Tenenbaum 2003] - abréviation de « isometric feature mapping » - est l'un des premiers algorithmes qui ait été développé pour l'apprentissage de variétés. Il peut être considéré comme une extension de la méthode « Multidimensional Scaling (MDS) » [Cox & Cox 1994] classique pour intégrer des informations de dissemblance dans un espace euclidien. Isomap se compose de deux étapes principales :

1. Estimation des distances géodésiques (distances sur le long d'une variété) entre les points d'entrée à l'aide d'une méthode d'estimation du plus court chemin sur un graphe de « k plus proches voisins ». La figure 4.4 illustre la distance géodésique sur une variété surfacique.
2. Utilisation de la méthode MDS pour trouver des points dans l'espace euclidien de faible dimension, dont les distances correspondent au mieux aux distances géodésiques trouvées à l'étape 1.

4.2.1.1 Estimation des distances géodésiques

On suppose que les données sont issues d'une variété $\mathcal{M}$ de dimension d incluse dans $\mathbb{R}^D$, et nous souhaitons trouver un homéomorphisme f de $\mathcal{M}$ dans $\mathbb{R}^2$. Isomap recherche une application isométrique, i.e., qui préserve les distances entre les points.

En d'autres termes, si $\boldsymbol{x}_i$, $\boldsymbol{x}_j$ sont des points sur une variété et $G(\boldsymbol{x}_i, \boldsymbol{x}_j)$ la distance géodésique entre eux. On recherche alors une application $f : \mathcal{M} \to \mathbb{R}^d$ telle que

$$\|f(\boldsymbol{x}_i) - f(\boldsymbol{x}_j)\| = G(\boldsymbol{x}_i, \boldsymbol{x}_j) \tag{4.1}$$

En outre, on suppose que la variété est suffisamment régulière afin que la distance géodésique entre des points voisins soit (approximativement) euclidienne. Ainsi la distance euclidienne de $\mathbb{R}^D$ constitue-t-elle une approximation raisonnable de la distance géodésique entre deux points voisins de la variété. Pour des points éloignés, il convient de procéder autrement. Pour ce faire, l'algorithme Isomap construit un graphe de « k plus proches voisins » dont les arrêtes sont pondérées par les distances euclidiennes. Puis il approxime la distance géodésique entre 2 points quelconques en sommant les poids des arrêtes du plus court chemin entre eux, obtenu par la méthode deDijkstra [Dijkstra 1959] ou de Floyd [Floyd 1962].

4.2.1.2 Multidimensional Scaling (MDS)

Une fois les distances géodésiques calculées entre toutes paires de points, l'algorithme Isomap estime les images des points dans $\mathbb{R}^D$ de sorte que les distances euclidiennes entre eux soient (approximativement) égales aux distances géodésiques. MDS est une technique classique qui peut être utilisée pour cela.

Le problème plus généralement traité par MDS est le suivant. Etant donnée une matrice $\boldsymbol{D} \in \mathbb{R}^{n \times n}$ de dissemblances, l'algorithme construit un ensemble de points dont les distances euclidiennes entre eux correspondent (approximativement) aux dissemblances fournies par $\boldsymbol{D}$. Il existe de nombreuses fonctions de coût pour cette tâche, et une grande variété d'algorithmes pour minimiser ces fonctions de coût. L'algorithme MDS repose sur le théorème suivant, qui établit un lien entre l'espace des matrices de distances euclidiennes et l'espace des matrices de Gram.

Théorème 7 *Une matrice non-négative symétrique $\boldsymbol{D} \in \mathbb{R}^{n \times n}$, de termes diagonaux nuls, est une matrice de distance euclidienne si et seulement si*

$$\boldsymbol{B} \stackrel{def}{=} -\frac{1}{2}\boldsymbol{H}\boldsymbol{D}\boldsymbol{H} \qquad \textit{avec} \qquad \boldsymbol{H} \stackrel{def}{=} \boldsymbol{I} - \frac{1}{n}\boldsymbol{1}\boldsymbol{1}^\top$$

est semi-définie positive. Ici, $\boldsymbol{B}$ est la matrice de Gram pour une configuration de moyenne centrée avec les composantes de $\boldsymbol{D}$ pour distances inter-points.

Ce théorème inspire une technique pour convertir une matrice de distance euclidienne en une matrice de Gram. Soit $\boldsymbol{X}$ l'ensemble des points images dans $\mathbb{R}^d$, et $\boldsymbol{B}$ la matrice de Gram correspondante $\boldsymbol{B} = \boldsymbol{X}\boldsymbol{X}^\top$. Pour obtenir $\boldsymbol{X}$ de $\boldsymbol{B}$, on procède à une décomposition spectrale de $\boldsymbol{B} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$. On pose alors $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Lambda}^{1/2}$. Lorsque $\boldsymbol{D}$ n'est pas une matrice euclidienne, par exemple suite à l'application de Isomap, la matrice $\boldsymbol{B}$ définie ci-dessus ne peut être une matrice semi-définie positive et donc une matrice de Gram valide. Pour surmonter celà, MDS projette $\boldsymbol{B}$ sur le cône des matrices semi-définies positives en annulant ses valeurs propres négatives.

En général, $\boldsymbol{X} := \boldsymbol{U}\boldsymbol{\Lambda}_+^{1/2}$ est une configuration à n dimensions. Afin de réaliser une réduction de dimension de D à d avec $D > d$, on projette généralement $\boldsymbol{X}$ sur ses d composantes principales.

Soit $\boldsymbol{X} \stackrel{\text{def}}{=} \boldsymbol{U}\boldsymbol{\Lambda}_+^{1/2}$ la configuration à n dimensions. En conséquence, $\boldsymbol{U}\boldsymbol{\Lambda}_+^{1/2}\boldsymbol{U}^\top$ est la décomposition spectrale de $\boldsymbol{X}\boldsymbol{X}^\top$. Supposons que l'on utilise une ACP pour projeter les données $\boldsymbol{X}$ sur un espace à d dimensions. On rappelle que cet algorithme renvoie $\boldsymbol{X}\boldsymbol{V}$, où $\boldsymbol{V} \in \mathbb{R}^{n\times d}$ dont les colonnes $\boldsymbol{v}$ sont données par les vecteurs propres de $\boldsymbol{X}^\top\boldsymbol{X} : \boldsymbol{v}_1, \ldots, \boldsymbol{v}_d$. Ces vecteurs propres sont calculés par

$$\begin{aligned}
\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{v}_i &= \xi_i\boldsymbol{v}_i \\
(\boldsymbol{U}\boldsymbol{\Lambda}_+^{1/2})^\top(\boldsymbol{U}\boldsymbol{\Lambda}_+^{1/2})\boldsymbol{v}_i &= \xi_i\boldsymbol{v}_i \\
(\boldsymbol{\Lambda}_+^{1/2}\boldsymbol{U}^\top\boldsymbol{U}\boldsymbol{\Lambda}_+^{1/2})\boldsymbol{v}_i &= \xi_i\boldsymbol{v}_i \\
\boldsymbol{\Lambda}_+\boldsymbol{v}_i &= \xi_i\boldsymbol{v}_i
\end{aligned}$$

Cette dernière expression montre que la troncature effectuée suite à MDS peut être conjointement pratiquée par l'ACP, lors de la phase de réduction à d dimensions. Par ailleurs, les valeurs propres ξ_i révèlent l'importance relative de chaque dimension.

4.2.1.3 Algorithme

L'algorithme Isomap consiste à estimer les distances géodésiques en utilisant le plus court chemin sur un graphe, puis trouver une injection de ces distances dans un espace euclidien à l'aide de la méthode MDS. La figure 4.5 résume l'algorithme. Une caractéristique particulièrement utile pour Isomap est qu'il fournit automatiquement une estimation du nombre de dimensions de la variété sous-jacente. En particulier, le nombre de valeurs propres non nulles trouvées par MDS donne cette dimension sous-jacente. Cependant, en présence de bruit et d'incertitudes dans le processus d'estimation, il est à noter que l'écart entre les valeurs propres dominantes et les valeurs propres secondaires, ou supposément nulles, peut être altéré. Enfin, Isomap dispose d'une garantie d'optimalité [Bernstein *et al.* 2000]. En effet, Isomap est assuré de récupérer le paramétrage d'une variété sous les hypothèses suivantes.

1. La variété est isométrique intégrée dans $\mathbb{R}^D$.
2. L'espace paramétrique sous-jacent est convexe, c'est-à-dire la distance entre deux points quelconques dans l'espace paramétrique est donnée par une distance géodésique qui se trouve entièrement à l'intérieur de l'espace paramétrique.
3. La variété est bien échantillonnée partout.
4. La variété est compacte.

La preuve de cette optimalité commence par montrer que, avec suffisamment d'échantillons, les distances sur le graphe se rapprochent des distances géodésiques sous-jacentes. A la limite, ces distances sont euclidiennes et MDS retourne précisément les points injectés dans l'espace de dimension réduite. En outre, MDS est continue : de petites variations en entrée se traduisent par de petites variations en

---

Entrées : $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots \boldsymbol{x}_n \in \mathbb{R}^D, k$

1. Concevoir un graphe de voisinage à partir des k plus proches voisins avec les poids $w_{ij} := \|\boldsymbol{x}_i - \boldsymbol{x}_j\|$ pour les arêtes $(\boldsymbol{x}_i, \boldsymbol{x}_j)$.
2. Calculer les distances de plus court chemin entre toutes les paires de points à l'aide de l'algorithme Dijkstra ou Floyd. Ranger les carrés de ces distances dans une matrice $\boldsymbol{D}$ avec $\boldsymbol{D}_{ii} = 0$ et $\boldsymbol{D}_{ij} \geq 0$.
3. Poser $\boldsymbol{B} \overset{\text{def}}{=} -\frac{1}{2}\boldsymbol{H}\boldsymbol{D}\boldsymbol{H}$, où $\boldsymbol{H} \overset{\text{def}}{=} \boldsymbol{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^\top$ est une matrice de centrage.
4. Calculer la décomposition spectrale de $\boldsymbol{B} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$.
5. Construire $\boldsymbol{\Lambda}_+$ en posant $[\boldsymbol{\Lambda}_+]_{ij} = \max(0, \boldsymbol{\Lambda}_{ij})$.
6. Fixer la dimension intrinsèque d, correspondant au nombre de composantes non nulles de $\boldsymbol{\Lambda}_+$.
7. Calculer $\boldsymbol{X}^* = \boldsymbol{U}\boldsymbol{\Lambda}_+^{1/2}$.
8. Réduire la dimension en gardant les d premiers lignes de $\boldsymbol{X} : \boldsymbol{Y} = [\boldsymbol{x}]_{n\times d}$.
9. Retourner $\boldsymbol{Y}$.

FIGURE 4.5 – ISOMAP

sortie de l'algorithme. En raison de cette robustesse aux perturbations, la solution de MDS converge asymptotiquement vers la carte recherchée. Isomap a été développé selon deux directions principales : soit gérer les transformations conformes, soit gérer de très grands ensembles de données [Jordan *et al.* 2001].

La première approche est appelée C-Isomap et aborde la question qui est que beaucoup de types d'injections ne conservent pas les distances, mais les angles. C-Isomap a une garantie d'optimalité théorique semblable à celle de Isomap, mais exige que la variété soit échantillonnée uniformément.

La deuxième extension, Landmark Isomap, est basée sur des points de repère pour accélérer le calcul des distances entre les points pour MDS. Dans Isomap, ces deux étapes nécessitent $O(n^3)$ opérations, ce qui est bloquant pour les grands ensembles de données. L'idée de base consiste à sélectionner un petit ensemble de l points de repère, où $D < \ell < n$. Ensuite, les distances géodésiques approximatives de chaque point aux points de repère sont calculées. Ensuite, MDS est exécuté sur la matrice $\ell \times \ell$ de distances entre les points de repère. Enfin, les points restants sont incorporés en leurs distances aux points de repère. Cette technique réduit la complexité d'Isomap de $O(n^3)$ à $O(dn\ell \log n + \ell^3)$.

4.2.2 Méthode LLE

L'algorithme LLE (Locally Linear Embedding) [Roweis & Saul 2000] repose sur une intuition très différente. L'idée vient de la visualisation d'une variété comme une collection de « patches » se chevauchant. Si la taille de voisinage est petite et la variété suffisamment lisse, ces patches sont localement (hyper)-plans. En outre, la projection de la variété sur ces patches n'altère pas la configuration des données. L'idée est donc d'identifier ces patches et de caractériser la géométrie entre eux, et de trouver une injection dans $\mathbb{R}^d$ qui conserve cette géométrie locale. On suppose que ces patches locaux se chevauchent de sorte que les reconstructions locales se combinent en un système global. Comme dans toutes les méthodes d'apprentissage de variétés, le nombre de voisins choisi est un paramètre déterminant dans la mise en œuvre de l'algorithme. Soit $\mathcal{N}(i)$ l'ensemble des k plus proches voisins du point $\boldsymbol{x}_i$. La première étape de l'algorithme modélise la variété comme une collection de patches, et vise à caractériser la géométrie locale. Pour cela, la méthode vise à représenter $\boldsymbol{x}_i$ comme une combinaison convexe de ses k plus proches voisins. Les coefficients de pondération sont choisis afin de minimiser l'erreur quadratique de reconstruction de chaque point :

$$J(\boldsymbol{W}) = \sum_i \left\| \boldsymbol{x}_i - \sum_{j \in \mathcal{N}_i} w_{ij}\, \boldsymbol{x}_j \right\|^2 \tag{4.2}$$

La matrice de poids $\boldsymbol{W}$ est utilisée comme un substitut à la géométrie locale des patches. Intuitivement, le vecteur $[\boldsymbol{W}]_i$ révèle la disposition des points autour de $\boldsymbol{x}_i$. En plus, il existe des contraintes sur les poids. 1) Chaque ligne doit se sommer à un, et 2) $[\boldsymbol{W}]_{ij} = w_{ij} = 0$ si $j \notin \mathcal{N}(i)$. La contrainte 2) implique que LLE est une méthode locale. La contrainte 1) rend les poids invariants aux translations globales, i.e., si chaque $\boldsymbol{x}_i$ subit une translation de $\boldsymbol{\alpha}$, alors

$$\left\| \boldsymbol{x}_i + \boldsymbol{\alpha} - \sum_{j \in \mathcal{N}_i} w_{ij}(\boldsymbol{x}_j + \boldsymbol{\alpha}) \right\|^2 = \left\| \boldsymbol{x}_i - \sum_{j \in \mathcal{N}_i} w_{ij} \boldsymbol{x}_j \right\|^2 \tag{4.3}$$

et $\boldsymbol{W}$ ne s'en trouve pas affecté. En outre, $\boldsymbol{W}$ est invariant aux rotations et mises à l'échelle globale.

Une solution analytique $\boldsymbol{W}$ à ce problème existe, obtenue à partir de la méthode des facteurs de Lagrange. Ainsi, les poids de reconstruction pour chaque point $\boldsymbol{x}_i$ sont donnés par

$$w_{ij} = \frac{\sum_{k \in \mathcal{N}(i)} c_{jk}^{-1}}{\sum_{\ell,m \in \mathcal{N}(i)} c_{\ell m}^{-1}} \tag{4.4}$$

où $c_{\ell m}^{-1}$ est la composante (ℓ, m) de l'inverse de la matrice de covariance locale $\boldsymbol{C}$, dont les entrées sont définies par

$$c_{\ell m} \overset{\text{def}}{=} (\boldsymbol{x}_i - \boldsymbol{x}_\ell)^\top (\boldsymbol{x}_i - \boldsymbol{x}_m) \qquad \text{avec} \qquad \ell, m \in \mathcal{N}(i). \tag{4.5}$$

La seconde étape de l'algorithme vise à trouver les images $\boldsymbol{y}_i$ des points $\boldsymbol{x}_i$ dans $\mathbb{R}^d$ respectant la configuration estimée au moyen de la matrice $\boldsymbol{W}$. Pour ce faire, on

considère la fonction coût suivante :

$$J(\boldsymbol{Y}) = \sum_i \left\| \boldsymbol{y}_i - \sum_j w_{ij} \boldsymbol{y}_j \right\|^2 \tag{4.6}$$

à minimiser par rapport à $\boldsymbol{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n]$. On peut écrire $J(\boldsymbol{Y}) = \boldsymbol{Y}^\top \boldsymbol{M} \boldsymbol{Y}$ avec

$$\boldsymbol{M} \overset{\text{def}}{=} (\boldsymbol{I} - \boldsymbol{W})^\top (\boldsymbol{I} - \boldsymbol{W}). \tag{4.7}$$

La minimisation du critère ci-dessus nécessite d'imposer des contraintes sur $\boldsymbol{Y}$. Sans perte de généralité, puisque le problème (4.6) est invariant par homothétie, on impose $\boldsymbol{Y}\boldsymbol{Y}^\top = \boldsymbol{I}$ ce qui contraint par ailleurs la transformation à être de rang d. Enfin, on pose $\boldsymbol{Y}\mathbf{1} = \mathbf{0}$ afin de centrer les données $\boldsymbol{y}_i$. On montre que la résolution de ce problème contraint nécessite la maximisation d'un quotient de Rayleigh, dont la solution est obtenue en retenant les $d+1$ vecteurs propres de $\boldsymbol{M}$ associés aux plus faibles valeurs propres. Le dernier, valant $\mathbf{1}$ et associé à la valeur propre 0, ne doit pas être retenu car il correspond à la translation des données.

Conformal Eigenmap [Sha & Saul 2005] est une technique qui peut être utilisée en conjonction avec LLE pour fournir une estimation automatique de d, la dimension de la variété. LLE est résumé dans la figure 4.7, et illustré par la figure 4.6.

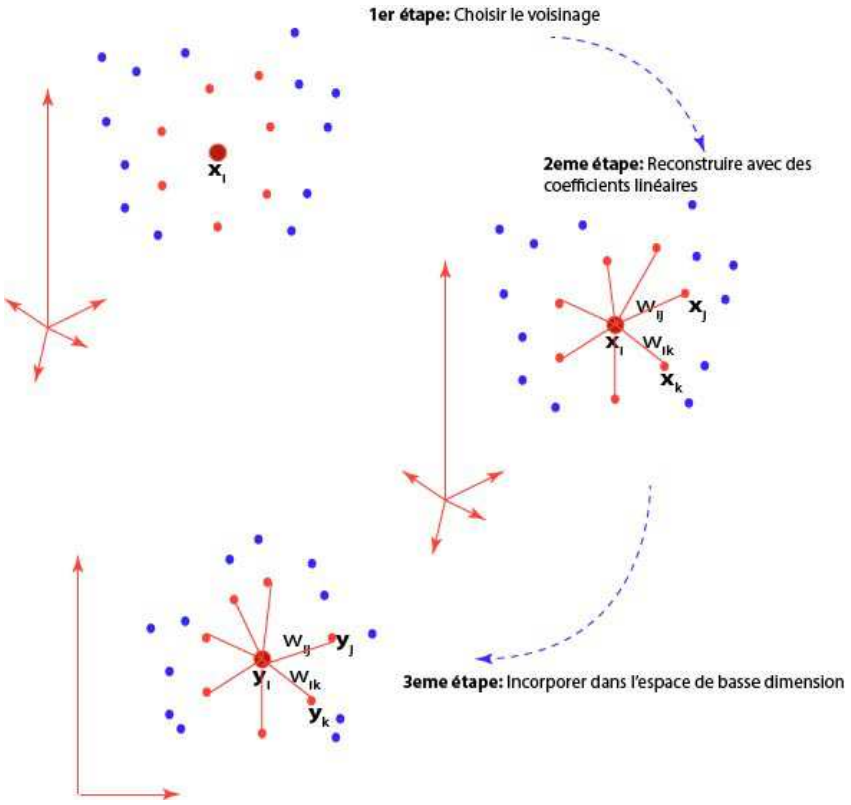


FIGURE 4.6 – Procédure d'apprentissage de variété avec LLE [Roweis & Saul 2000]

Entrées : $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots \boldsymbol{x}_n \in \mathbb{R}^D, k, d$

1. Calculer les poids de reconstruction. Pour chaque point x_i, calculer $\boldsymbol{C}$ et $\boldsymbol{W}$
2. Prendre $\boldsymbol{M} = (\boldsymbol{I}-\boldsymbol{W})^\top(\boldsymbol{I}-\boldsymbol{W})$ et calculer les vecteurs propres $\boldsymbol{M}_{n\times d}$ associés aux d plus petites valeurs propres non nulles
3. Renvoyer le résultat : $\boldsymbol{Y} = \boldsymbol{M}_{n\times d}$

FIGURE 4.7 – LLE

4.3 Apprentissage de variétés et KACP

Les méthodes présentées ci-dessus peuvent s'exprimer à partir d'une formulation à noyau. Isomap, par exemple, utilise MDS pour l'injection des données. MDS est équivalent à ACP, ce qui permet d'entrevoir ici la mise en œuvre de KACP et de l'astuce du noyau utilisé. Dans [Ham *et al.* 2003], les auteurs ont montré la relation entre les méthodes d'apprentissage de variétés et les méthodes KACP.

Tous les algorithmes de réduction de dimension partagent une caractéristique commune en cela qu'ils induisent d'abord une structure locale de voisinage sur les données, puis exploite cette structure locale pour reconstruire la variété dans un espace de dimension inférieure. Cette relation de voisinage local est définie en utilisant les k voisins les plus proches dans l'espace euclidien, et peut être décrite par un graphe. Cependant, la façon dont ces algorithmes utilisent cette structure de voisinage est tout à fait différente. Envisagée dans ce contexte, l'intégration des noyaux dans ces algorithmes partagent donc une stratégie similaire. On construit d'abord une matrice de Gram des données d'apprentissage. La diagonalisation de cette matrice à noyau définit une injection gardant la structure de faible dimension de la variété. Nous allons d'abord présenter la méthode KACP, puis nous nous attacherons à la réinterprétation des algorithmes selon KACP.

4.3.1 Kernel ACP

Dans les méthodes de noyau, les noyaux peuvent être considérés comme une mesure de similarité non-linéaire. Soit un ensemble non-vide $\mathcal{X}$ et un noyau défini positif κ, des données d'entrée $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\} \in \mathcal{X}$ définissant un sous-espace de $\mathbb{R}^D$. KACP calcule les composantes principales de ces données $\phi(\boldsymbol{x}_1), \phi(\boldsymbol{x}_2), \ldots, \phi(\boldsymbol{x}_n)$ dans l'espace de Hilbert $\mathcal{H}$. Car $\mathcal{H}$ peut être de dimension infinie, le problème ACP doit être transformé en un problème qui peut être résolu en termes de noyau. Dans ce but, nous considérons la matrice de covariance dans $\mathcal{H}$:

$$\boldsymbol{C} := \frac{1}{n}\sum_{i=1}^{n} \phi(\boldsymbol{x}_i)\phi(\boldsymbol{x}_i)^\top \tag{4.8}$$

Pour diagonaliser la matrice $\boldsymbol{C}$ alors même que $\mathcal{H}$ peut être de dimension infinie, on recherche les valeurs propres $\lambda \geq 0$ et les vecteurs propres correspondant $\boldsymbol{v}$ satisfaisant l'équation :

$$\boldsymbol{C}\boldsymbol{v} = \lambda \boldsymbol{v} \tag{4.9}$$

Notons que toute solution de (4.9) s'inscrit dans un sous-espace engendré par les images $\phi(\boldsymbol{x}_i)$. Les solutions $\boldsymbol{v}$, vecteurs propres de $\boldsymbol{C}$, peuvent être représentés par :

$$\boldsymbol{v} = \sum_{i=1}^{n} \alpha_i \phi(\boldsymbol{x}_i) \tag{4.10}$$

Le problème (4.9) est équivalent à :

$$\phi(\boldsymbol{x}_\ell)\boldsymbol{C}\boldsymbol{v} = \lambda \phi(\boldsymbol{x}_\ell)\boldsymbol{v} \qquad \forall \ell = 1, \ldots, n \tag{4.11}$$

En substituant (4.8), (4.10) dans (4.11), on obtient :

$$n\lambda \mathbf{K}\boldsymbol{\alpha} = \mathbf{K}^2\boldsymbol{\alpha} \tag{4.12}$$

où $\boldsymbol{\alpha}$ est un vecteur colonne de composantes α_i, $i = 1, \ldots, n$. Le problème se réduit à trouver les coefficients α_i en résolvant cette décomposition en éléments propres :

$$n\lambda \boldsymbol{\alpha} = \mathbf{K}\boldsymbol{\alpha} \tag{4.13}$$

pour les valeurs propres λ non nulles.

On peut démontrer que l'extraction de la ℓ^e caractéristique de $\phi(\boldsymbol{x})$, au moyen du ℓ^e vecteur $\boldsymbol{\alpha}_j^{(\ell)}$, a pour forme :

$$\langle \phi(\boldsymbol{x}), \boldsymbol{v}^{(\ell)} \rangle_{\mathcal{H}} = \frac{1}{\sqrt{\lambda^{(\ell)}}} \sum_{i=1}^{n} \alpha_i^{(\ell)} \kappa(\boldsymbol{x}_i, \boldsymbol{x}) = \frac{1}{\sqrt{\lambda^{(\ell)}}} \mathbf{K}\boldsymbol{\alpha}^{(\ell)} \tag{4.14}$$

Le facteur $\frac{1}{\sqrt{\lambda^{(\ell)}}}$ demeure pour assurer $\boldsymbol{v}^{(\ell)\top}\boldsymbol{v}^{(\ell)} = 1$. Comme pour ACP, KACP nécessite que les données dans $\mathcal{H}$ soient à moyenne nulle. En général, ceci nécessite un pré-traitement même si les données sont à moyenne nulle dans l'espace $\mathcal{X}$. Le centrage des données peut être effectué en substituant la matrice $\boldsymbol{K}$ par

$$\mathbf{K}' = (\boldsymbol{I} - \boldsymbol{e}\boldsymbol{e}^\top)\mathbf{K}(\boldsymbol{I} - \boldsymbol{e}\boldsymbol{e}^\top) \tag{4.15}$$

avec $\boldsymbol{e} = \frac{1}{\sqrt{n}}[1, \ldots, 1]^\top$.

4.3.2 Isomap sous forme KACP

Comme pour MDS, Isomap construit d'abord une matrice de distances entre les paires de données. Cependant, au lieu d'utiliser la distance euclidienne, Isomap construit un graphe symétrique et approxime la distance géodésique grâce à un algorithme de plus court chemin. MDS est appliquée à cette matrice de distance géodésique et l'injection recherchée est donnée par les vecteurs propres correspondant aux valeurs propres les plus petites. Comme souligné dans [Williams 2002],

on peut interpréter MDS grâce à KACP. De la même façon, on peut prendre les distances utilisées dans Isomap et considérer la matrice de Gram suivante :

$$\mathbf{K}_{\text{Isomap}} = -\frac{1}{2}(\boldsymbol{I} - \boldsymbol{e}\boldsymbol{e}^\top)\boldsymbol{S}(\boldsymbol{I} - \boldsymbol{e}\boldsymbol{e}^\top) \tag{4.16}$$

où $\boldsymbol{S}$ est la matrice des distances géodésiques au carré, et $\boldsymbol{e} = \frac{1}{\sqrt{n}}[1, \ldots, 1]^\top$ est le vecteur destiné à centrer $\mathbf{K}_{\text{Isomap}}$. Néanmoins, nous ne disposons d'aucune garantie sur le caractère défini positif de cette matrice. Toutefois, dans la limite continue pour une variété lisse, la distance géodésique entre les points est proportionnelle à la distance euclidienne dans l'espace des paramètres de faible dimension de la variété [Grimes & Donoho 2002]. Il est connu que $\kappa(\boldsymbol{x}, \boldsymbol{x}') = -\|\boldsymbol{x} - \boldsymbol{x}'\|^\beta$ est définie positive pour $0 < \beta \leq 2$. Dans la limite continue, $-\boldsymbol{S}$ est donc définie positive tout comme $\mathbf{K}_{\text{Isomap}}$. Voir pages 49–51 dans [Schölkopf & Smola 2001]). Dans l'équation (4.13) appliquée à la matrice $\mathbf{K}_{\text{Isomap}}$, parce que l'application fournie par KACP est donnée par les vecteurs propres associés aux plus grandes valeurs propres de $\mathbf{K}_{\text{Isomap}}$, on constate en pratique que l'utilisation des projections par les vecteurs propres dominants de KACP donne un résultat comparable à Isomap.

4.3.3 LLE sous forme KACP

L'algorithme LLE construit une matrice de poids fournissant la meilleure approximation d'une donnée par ses voisins, au sens des moindres carrés. Cette matrice est définie par $\boldsymbol{M} = (\boldsymbol{I} - \boldsymbol{W})^\top(\boldsymbol{I} - \boldsymbol{W})$ qui a une valeur propre maximale $\lambda_{\max}$, et une plus petite valeur propre nulle associée au vecteur propre uniforme $\boldsymbol{e}$. Comme les autres vecteurs propres sont orthogonaux à $\boldsymbol{e}$, la somme de leurs composantes vaut 0. Pour LLE, l'application recherchée est définie à partir des vecteurs propres associés aux valeurs propres des rangs $n - d$ à $n - 1$. Nous définissons :

$$\mathbf{K} := (\lambda_{\max}\boldsymbol{I} - \boldsymbol{M}) \tag{4.17}$$

Par la construction, $\mathbf{K}$ est une matrice définie positive, son vecteur propre principal est $\boldsymbol{e}$, et les coordonnées des vecteurs propres des rangs 2 à $d + 1$ fournissent l'application recherchée. De manière équivalente, on peut considérer

$$\mathbf{K}_{\text{LLE}} = (\boldsymbol{I} - \boldsymbol{e}\boldsymbol{e}^\top)\mathbf{K}(\boldsymbol{I} - \boldsymbol{e}\boldsymbol{e}^\top) \tag{4.18}$$

et utiliser ses vecteurs propres associés aux valeurs propres dominantes des rangs 2 à $d+1$. Ceci établit une connexion directe avec KACP, qui diagonaliserait également la matrice $\mathbf{K}_{\text{LLE}}$.

4.4 Démélange des images hyperspectrales par apprentissage de variétés

Nous avons présenté le modèle de mélange linéaire et des modèles non-linéaires pour les images hyperspectrales au chapitre 2. Les méthodes d'extraction des

composés purs et d'estimation des abondances sont souvent deux étapes indépendantes dans la procédure de démélange. Nous nous concentrons ici sur les méthodes d'extraction qui supposent l'existence de pixels purs telles que N-FINDR, SGA et VCA, mais adoptons le formalisme géométrique développé dans [Honeine & Richard 2011a]. En effet, avec ces mêmes algorithmes, il permet conjointement d'extraire les composés purs et d'estimer conjointement les abondances des pixel-vecteurs de l'image traité. Les abondances y sont définies comme des coordonnées barycentriques dans un espace de dimension réduite, et peuvent être évaluées par des rapports de volumes de simplexes ou de distances.

Les approches rappelées ci-dessus reposent sur un modèle de démélange linéaire. Ils conduisent à des résultats sérieusement biaisés si le modèle de mélange sous-jacent est non-linéaire. Une stratégie possible est de recourir à des méthodes d'apprentissage de variété pour prendre en compte la géométrie des données. Plus précisément, il s'agit là d'exploiter les distances géodésiques estimées par Isomap, et la représentation des données dans un espace euclidien de dimension réduite fourni par LLE. Dans la suite, une version non-linéaire des algorithmes N-FINDR, SGA et VCA est développée. Il s'agit d'une contribution originale de cette thèse.

4.4.1 Démélange non-linéaire avec N-FINDR

L'algorithme N-FINDR a pour but de trouver les composés purs en maximisant le volume du simplexe formé par des pixel-vecteurs de l'image traitée. Les sommets définissant ce simplexe sont les composés purs recherchés, qui doivent être présents dans l'ensemble de données.

Pour développer une version non-linéaire de N-FINDR, une première approche consiste à adopter LLE comme pré-traitement des données plutôt que toute autre approche de réduction de dimension linéaire telle que l'ACP recommandée par les auteurs. LLE réalise une injection des données dans un système de coordonnées global unique de dimension inférieure. La mise en œuvre de N-FINDR sur cette nouvelle représentation aboutit au final à un schéma d'inversion non-linéaire.

Avec Isomap, l'utilisateur aboutit à une matrice de distances géodésiques entre les pixel-vecteurs de l'image traitée. L'exploitation de ces informations nécessite une reformulation de l'algorithme N-FINDR. Le calcul de volume est généralement effectué en considérant le déterminant ci-après

$$\boldsymbol{V}_{\mathcal{S}} = \frac{1}{(R-1)!} \left| \det \begin{pmatrix} 1 & 1 & \dots & 1 \\ \boldsymbol{m}_1 & \boldsymbol{m}_2 & \dots & \boldsymbol{m}_R \end{pmatrix} \right| \tag{4.19}$$

où les $\boldsymbol{m}_i$ désignent les sommets considérés. Comme rappelé ci-dessus, le calcul de ce déterminant nécessite de recourir à une réduction de dimension par projection des données dans un espace de dimension $k-1$, ce qui pourrait être fait avec LLE si on adoptait cette méthode d'apprentissage de variétés. Sinon, il est possible de recourir au déterminant de Cayley-Menger comme cela a été montré de façon indépendante dans [Honeine & Richard 2011a, Heylen *et al.* 2011]. Cette formulation

repose sur les distances euclidiennes inter-sommets et ne nécessite pas de réduction de dimension.

Soit d_{ij} une distance entre les sommets $\boldsymbol{m}_i$ et $\boldsymbol{m}_j$. Le volume au carré du simplexe défini par les $\{\boldsymbol{m}_i\}_{i=1,\dots,R}$ dans $\mathbb{R}^L$ est donné par

$$\boldsymbol{V}_{\mathcal{S}}^2 = \frac{(-1)^R}{2^{R-1}((R-1)!)^2} \det(\boldsymbol{C}_{1,\dots,R}) \tag{4.20}$$

dans laquelle

$$\boldsymbol{C}_{1,\dots,R} = \begin{pmatrix} \boldsymbol{D}^2_{1,\dots,R} & \mathbf{1} \\ \mathbf{1}^\top & 0 \end{pmatrix} \tag{4.21}$$

où $\mathbf{1}$ est un vecteur de 1, et $\boldsymbol{D}^2_{1,\dots,R}$ est la matrice telle que $\boldsymbol{D}^2_{1,\dots,R}(i,j) = d^2_{ij}$. On note que l'on peut également réécrire (4.20) en remplaçant

$$\det(\boldsymbol{C}_{1,\dots,R}) = -(\boldsymbol{d}_1^\top \boldsymbol{C}^{-1}_{2,\dots,R} \boldsymbol{d}_1) \det(\boldsymbol{C}_{2,\dots,R})$$

avec $\boldsymbol{d}_1 = (d^2_{12}, \dots, d^2_{1R}, 1)^\top$. Enfin, on peut également calculer la distance entre le sommet $\boldsymbol{m}_1$ et le sous-espace engendré par les autres sommets $(\boldsymbol{m}_2, \dots, \boldsymbol{m}_R)$ ainsi

$$\delta(\boldsymbol{m}_1) = \left(\frac{\boldsymbol{d}_1^\top \boldsymbol{C}^{-1}_{2,\dots,R} \boldsymbol{d}_1}{2} \right)^{\frac{1}{2}} \tag{4.22}$$

Avec ces équations, si les distances choisies sont euclidiennes, les résultats obtenus sont les mêmes qu'avec N-FINDR [Honeine & Richard 2011a]. On préconise l'utilisation des distances géodésiques, le calcul des volumes étant effectué le long de la variété. Les coordonnées dans ce cas sont les coordonnées sur la variété. Il s'agit là d'une méthode de démélange non-linéaire non-supervisée originale, permettant d'extraire conjointement les composés purs et leurs abondances au sein de l'image traitée. Aucune hypothèse sur la nature du modèle non-linéaire n'est faite.

4.4.2 Démélange non-linéaire avec SGA

A la différence de N-FINDR, l'algorithme SGA détermine successivement les sommets du simplexe. Tout d'abord, l'algorithme débute par un sommet choisi de façon aléatoire $\boldsymbol{t}$, puis recherche un autre point au sein de l'image qui maximise le volume du simplexe $\mathcal{S}_{1i} = \{\boldsymbol{t}, \boldsymbol{r}_i\}$. Ce point est choisi comme le premier composé pur et étiqueté $\boldsymbol{m}_1$. SGA poursuit sa recherche avec les simplexes $\mathcal{S}_{2i} = (\boldsymbol{m}_1, \boldsymbol{r}_i)$. Le nouveau point $\boldsymbol{r}_i$ qui maximise le volume du simplexe est sélectionné comme deuxième composé pur et étiqueté $\boldsymbol{m}_2$. Le processus est itéré jusqu'à ce que les R composés purs requis aient été déterminés.

SGA réclame des calculs de volume à chaque recherche d'un nouveau sommet. Cependant, dans sa formulation classique, il demande aussi une étape de réduction de dimension à chaque itération pour le calcul du volume par (4.19). Itérer cette procédure, avec des distances géodésiques pour Isomap, ou un espace euclidien de dimension réduite pour LLE, estimés à chaque pas de l'algorithme s'avèrerait trop

couteux. Comme pour N-FINDR, on préconise d'utiliser LLE comme pré-traitement à l'algorithme SGA conventionnel, éventuellement avec (2.27) pour estimer conjointement les abondances, ou exploiter la formulation de Cayley-Menger (4.20) avec les distances géodésiques fournies par Isomap.

4.4.3 Démélange non-linéaire avec VCA

L'algorithme VCA repose sur le fait que la transformation affine d'un simplexe est également un simplexe. En conséquence, il projette l'ensemble des données sur un sous-espace orthogonal au sous-espace engendré par les composés purs déjà sélectionnés. Le point le plus éloigné constitue le nouveau composé pur, et la procédure est itérée jusqu'à ce que le nombre souhaité d'éléments ait été obtenu. Cette procédure est similaire à SGA, mais basée sur le calcul de la distance d'un sommet de simplexe à sa base. Ceci peut être effectué grâce à la formulation de Cayley-Menger (4.20) avec les distances géodésiques fournies par Isomap. Ou alors, la méthode LLE peut être mis en œuvre comme un traitement non-linéaire préalable à l'algorithme VCA originel.

4.5 Expérimentation

Cette section vise à tester les algorithmes décrits ci-dessus, sur des données simulées et sur une scène réelle de petite taille.

4.5.1 Données artificielles

Les données artificielles considérées ici appartiennent à $\mathbb{R}^3$, et sont distribuées sur une variété bidimensionnelle de type "Swissroll". Celle-ci est paramétrée par une variable σ, permettant d'en augmenter la courbure à mesure que cette variable croît. Les coordonnées d'un point $\boldsymbol{r}_i$ en fonction de l'abondance locale $\boldsymbol{\alpha}_i$ (coordonnées sur le simplexe original) s'exprime par

$$\begin{cases} r_{i1} = \alpha_{i1} \sin(\sigma\alpha_{i1}) + 1 \\ r_{i2} = \alpha_{i1} \cos(\sigma\alpha_{i1}) + 1 \\ r_{i3} = \alpha_{i2} + 1 \end{cases} \tag{4.23}$$

Afin de respecter la contrainte de somme à un, l'abondance du 3$^{\text{e}}$ composé pur est défini par $\alpha_{i3} = 1 - (\alpha_{i1} + \alpha_{i2})$. En fixant l'un des coefficients d'abondance à 1, les autres étant à 0, on retrouve les composés purs dont les coordonnées dans $\mathbb{R}^3$ sont

$$\begin{cases} \boldsymbol{m}_1 = (\sin(\sigma) + 1, \cos(\sigma) + 1, 1)^\top \\ \boldsymbol{m}_2 = (1, 1, 2)^\top \\ \boldsymbol{m}_3 = (1, 1, 1)^\top \end{cases} \tag{4.24}$$

La figure 4.8 représente la variété pour deux valeurs de σ différentes. Plus σ augmente, plus la courbure de la variété est importante. Le démélange des données de

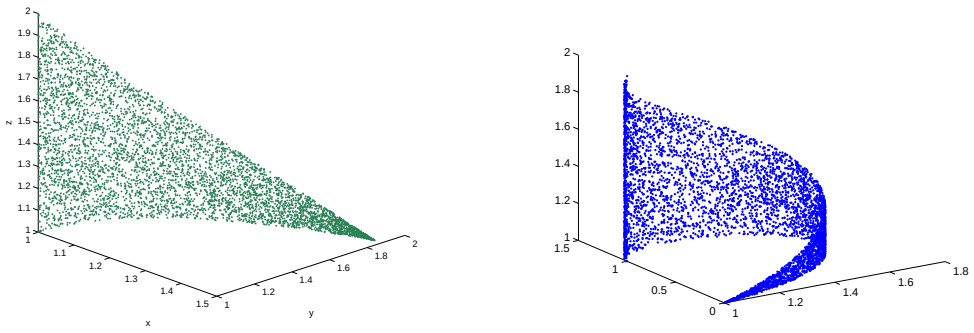


FIGURE 4.8 – Variété "Swissroll" pour $\sigma = 0.5$ (gauche) et $\sigma = \pi$ (droite)

la variété Swissroll a été effectué sur un ensemble de 1000 échantillons en utilisant les algorithmes basés sur l'apprentissage de variété. Le nombre de composés purs a été fixé à $R = 3$. L'extraction des composés purs a été réalisée avec N-FINDR, SGA et VCA, dans leur version linéaire et non-linéaire.

Le protocole linéaire a consisté à appliquer préalablement une ACP pour réduire la dimension des données à 2. Puis les trois algorithmes ont été appliqués pour estimer conjointement les composés purs et leurs abondances dans les données. Le protocole non-linéaire a consisté, d'une part à appliquer LLE comme un pré-traitement non-linéaire puis à mettre en œuvre les 3 méthodes linéaires, d'autre part à calculer les distances géodésiques par Isomap et appliquer les 3 méthodes linéaires avec celles-ci.

Dans les simulations, pour Isomap et LLE, la taille k du voisinage a été fixée à 10. On a observé que de bonnes performances étaient obtenues pour $5 \leq k \leq 20$. Pour $k < 5$, de médiocres performances ont pu être observées. Les algorithmes ont été comparés en fonction de la valeur du paramètre σ dans l'intervalle $[0, 5]$. Le graphe supérieur dans les figures 4.9, 4.10 et 4.11 fournit la valeur moyenne de l'angle spectral minimum entre les composés purs estimés $\widehat{\boldsymbol{m}}_i$ et la vérité terrain $\boldsymbol{m}_i$, soit

$$\theta = \frac{1}{R} \sum_{i=1}^{R} \min_{j} \arccos \left(\frac{\widehat{\boldsymbol{m}}_i \cdot \boldsymbol{m}_j}{\|\widehat{\boldsymbol{m}}_i\| \|\boldsymbol{m}_j\|} \right) \tag{4.25}$$

On peut constater que les algorithmes non-linéaires donnent des valeurs acceptables sur toute la gamme de σ, et obtiennent de meilleurs résultats que les algorithmes linéaires originaux. Cependant, lorsque la courbure de la non-linéarité augmente, ces algorithmes ne parviennent pas à fonctionner de manière satisfaisante. Le graphe inférieur dans les figures 4.9, 4.10 et 4.11 fournit l'erreur entre les abondances estimées et leurs valeurs de référence, soit

$$E = \frac{1}{R} \sum_{k=1}^{R} \min_{k'} \left(\frac{1}{N} \sum_{i=1}^{N} |\widehat{\alpha}_{ik} - \alpha_{ik'}| \right) \tag{4.26}$$

où $N = 1000$ désigne la taille de l'ensemble de données. Une fois encore, les algorithmes non-linéaires présentent de meilleures performances que leurs alter ego linéaires. Les figures 4.12 et 4.13 présentent les mêmes résultats d'une manière différente, ce qui permet de comparer les performances de N-FINDR, SGA et VCA suivant qu'ils sont associés à Isomap ou LLE respectivement. Il apparait que Isomap est plus performant que LLE, mais nécessite davantage de calculs.

4.5.2 Données réelles

Nous avons testé un algorithme de référence, VCA, qui donne des résultats fiables dans les cas bruité ou non bruité, avec les trois méthodes de réduction de dimension : linéaire avec ACP, et non-linéaire avec Isomap et LLE. La scène choisie est une sous-image de taille 50×50 du site de Moffet Field (CA, USA). Dans ces images 4.14, trois matériaux principaux sont représentés : le sol, l'eau et les végétaux. Cette image présente de fortes non-linéarités aux interfaces entre ces matériaux. Il apparait que les méthodes non-linéaires améliorent le contraste sur ces zones, avec un net avantage à LLE, ce qui confirme ce qui avait été observé sur des données artificielles.

4.6 Conclusion

Dans ce chapitre, nous avons étudié des méthodes d'apprentissage de variétés, et fait le lien avec la méthode non-linéaire KACP de réduction de dimension. Nous avons proposé un protocole original pour l'extraction non-linéaire des composés purs d'une image hyperspectrale, et l'estimation conjointe de leurs abondances. Les perspectives de travail concernent ici une étude approfondie du formalisme géométrique [Honeine & Richard 2011a] adopté, pour une meilleure prise en compte des contraintes sur les abondances estimées. En l'état, cette approche ne permet que de constater a posteriori que le mélange estimé pour certains pixel-vecteurs n'y satisfait pas, mettant de fait en cause la validité locale du modèle de démélange choisi sans offrir pour autant de solution.

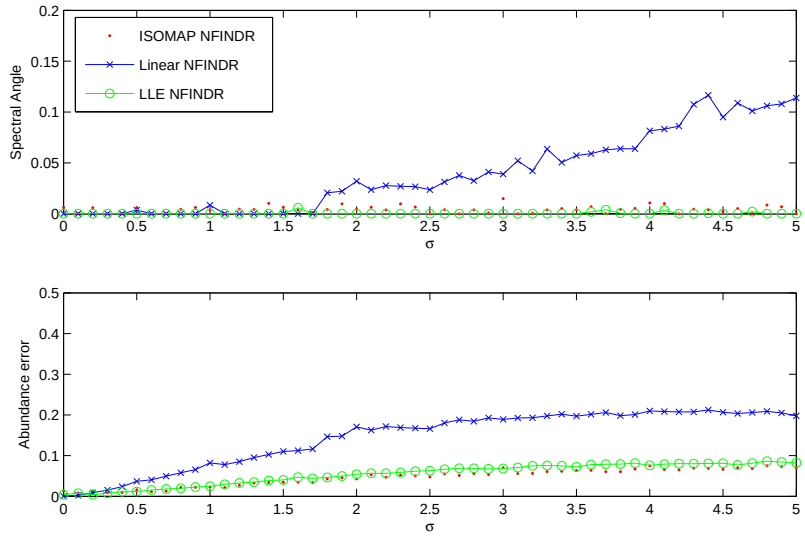


FIGURE 4.9 – Algorithme N-FINDR avec ACP (linéaire), ISOMAP et LLE (non-linéaire).

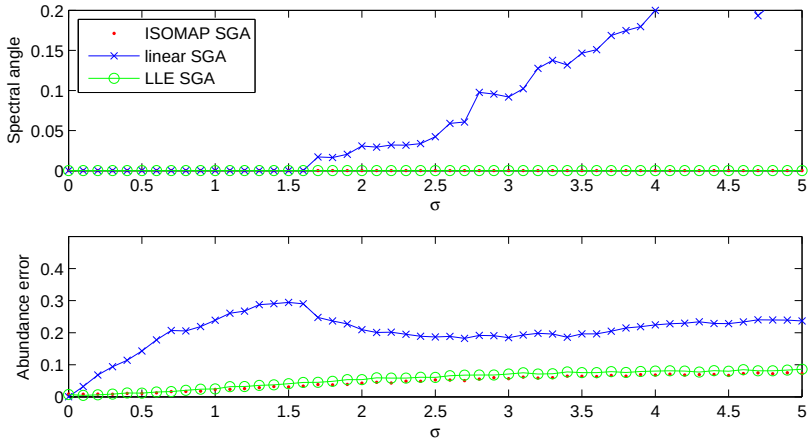


FIGURE 4.10 – Algorithme SGA avec ACP (linéaire), ISOMAP et LLE (non-linéaire).

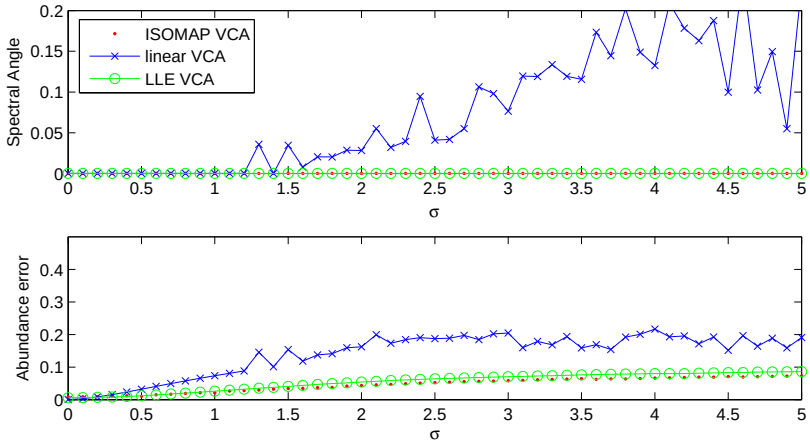


FIGURE 4.11 – Algorithme VCA avec ACP (linéaire), ISOMAP et LLE (non-linéaire).

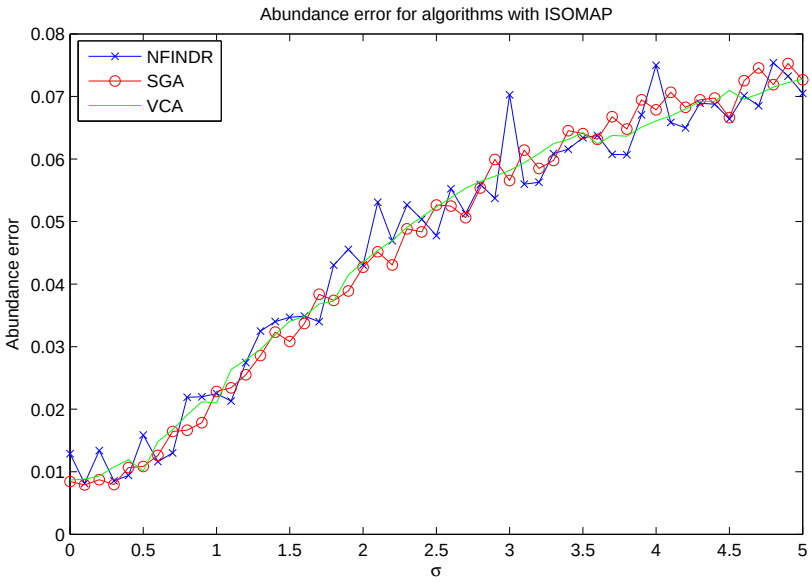


FIGURE 4.12 – Erreur sur les abondances par N-FINDR, SGA et VCA en utilisant ISOMAP

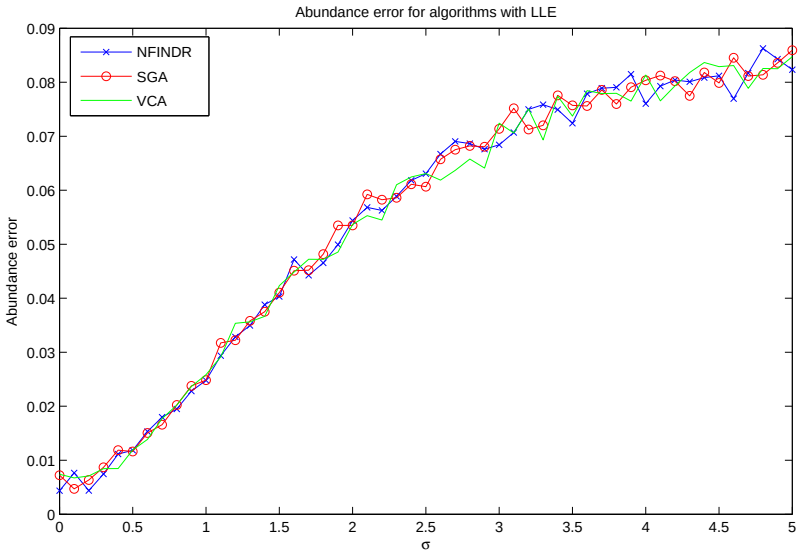
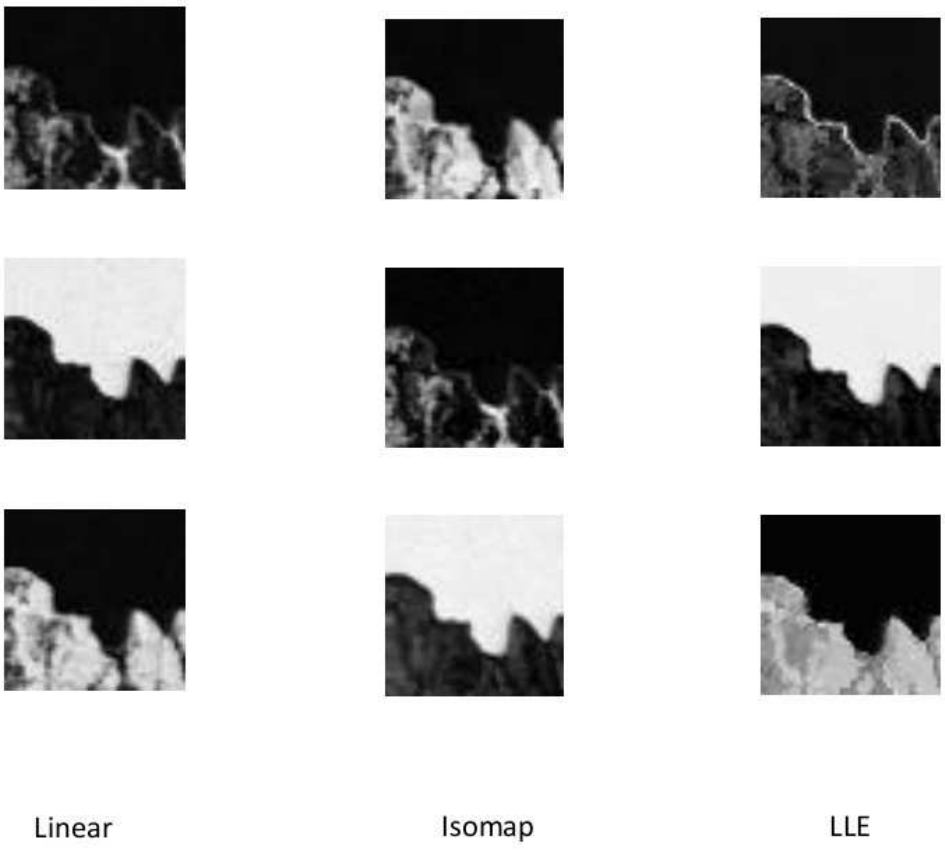


FIGURE 4.13 – Erreur sur les abondances par N-FINDR, SGA et VCA en utilisant LLE

FIGURE 4.14 – VCA avec trois méthodes de réduction de dimension sur Moffet Field.



Chapitre 5

Démélange non-linéaire par une méthode de pré-image

Sommaire

Dans ce chapitre, nous présentons une méthode de démélange non-linéaire basée sur le principe de pré-image, ainsi nommé par la communauté du machine learning. Voir [Honeine & Richard 2011b] pour une présentation détaillée de ce problème et des solutions existantes. Après avoir détaillé cette approche, nous nous intéressons au problème de sélection de modèle. Nous mettons l'accent sur les modèles partiellement linéaires, en montrant que la composante non-linéaire de ce type de modèle peut être avantageusement déterminée par apprentissage de variété. Finalement, nous considérons le problème de démélange ainsi défini en considérant un terme de régularisation spatial complémentaire.

5.1 Méthodes de pré-image

5.1.1 Introduction

Les méthodes de noyaux, comme présentées dans le Chapitre 3, visent à développer des méthodes d'identification non-linéaires efficaces. L'un des principes directeur de ces approches est connu sous le nom d'« astuce de noyau », et exploite le fait qu'un grand nombre de techniques de traitement de données ne dépendent pas explicitement des données elles-mêmes, mais d'une mesure de similarité entre elles telle qu'un produit scalaire. Pour concevoir une extension non-linéaire de techniques originellement linéaires, on peut appliquer une transformation non-linéaire aux données, et mettre en œuvre ces méthodes linéaires dans l'espace des caractéristiques ainsi défini. Selon l'astuce du noyau, cette transformation peut être aisément réalisée en remplaçant les produits scalaires par un noyau reproduisant. Ceci est équivalent au calcul de produits scalaires dans l'espace des caractéristiques. De tels algorithmes présentent des performances accrues par rapport à leurs homologues linéaires, pour une complexité calculatoire identique.

Tandis que la transformation non-linéaire de l'espace des entrées $\mathcal{X}$ vers l'espace des caractéristiques $\mathcal{H}$ est centrale dans les méthodes à noyau, la transformation inverse de $\mathcal{H}$ à $\mathcal{X}$ peut également être d'un intérêt primordial. Ceci est par exemple le cas pour l'algorithme KACP qui fournit un moyen de débruiter des signaux ou images dans l'espace $\mathcal{H}$, mais n'offre pas les clés pour retrouver ces données dans leur espace de représentation initial $\mathcal{X}$. Dans le cadre d'une application au démélange non-linéaire des données hyperspectrales, il s'agit de faire correspondre à un spectre de mélange observé, un antécédent sous la forme d'un vecteur d'abondance. Malheureusement, il s'avère que la transformation inverse n'existe pas en général, seuls quelques éléments de $\mathcal{H}$ ayant une pré-image dans $\mathcal{X}$. Le problème de la pré-image consiste ici à trouver un antécédent approximatif dans $\mathcal{X}$ à tout élément de $\mathcal{H}$. Il s'agit là essentiellement d'un problème de réduction de dimension, ces deux questions étant intimement liées dans leur évolution historique.

Le problème de la pré-image a attiré un intérêt considérable au cours des quinze dernières années. Il trouve son origine dans [Mika *et al.* 1999], où une méthode de point fixe potentiellement instable est présentée. Dans [Kwok & Tsang 2003], les auteurs suggèrent une relation entre les distances dans $\mathcal{X}$ et dans $\mathcal{H}$. L'application d'une technique de type MDS donne une estimation de la transformation inverse recherchée. Cette approche a ouvert la voie à une série d'autres techniques qui utilisent des données d'apprentissage dans les deux espaces comme information préalable, comme l'apprentissage de variété [Roweis & Saul 2000, de Silva & Tenenbaum 2003] et les méthodes « out-of-sample »[Bengio *et al.* 2003, Arias *et al.* 2007].

5.1.2 Formulation du problème de pré-image

Un problème est mal posé si au moins l'une des trois conditions suivantes, qui caractérisent les problèmes bien posés au sens de Hadamard, est violée :

1. une solution existe

2. la solution est unique
3. elle dépend continûment des données (condition de stabilité)

Malheureusement, l'identification de l'antécédent $\boldsymbol{x}^*$ dans $\mathcal{X}$ d'un élément quelconque ψ dans $\mathcal{H}$ est un problème mal posé du simple fait que $\dim(\mathcal{H}) \gg \dim(\mathcal{X})$. Aussi existerait-il $\boldsymbol{x}^*$ tel que $\phi(\boldsymbol{x}^*) = \psi$, celui-ci ne serait pas nécessairement unique.

Soit $\boldsymbol{\psi}$ dans l'espace des caractéristiques $\mathcal{H}$. En raison du théorème de représentant (Théorème 4, Chapitre 3), on peut écrire $\psi = \sum_{i=1}^{n} \alpha_i \kappa(\boldsymbol{x}_i, \cdot)$ avec $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ des données apprentissage dans $\mathcal{X}$. On cherche à résoudre le problème d'optimisation

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x} \in \mathcal{X}} \| \sum_{i=1}^{n} \alpha_i \kappa(\boldsymbol{x}_i, \cdot) - \kappa(\boldsymbol{x}, \cdot) \|_{\mathcal{H}}^2 \tag{5.1}$$

Du fait du caractère reproduisant du noyau, on peut estimer une pré-image $\boldsymbol{x}^*$ en minimisant la fonction coût suivante

$$\mathcal{J}(\boldsymbol{x}) = \kappa(\boldsymbol{x}, \boldsymbol{x}) - 2 \sum_{i=1}^{n} \alpha_i \kappa(\boldsymbol{x}, \boldsymbol{x}_i) \tag{5.2}$$

Afin de résoudre ce problème, on présente ci-après deux approches populaires.

5.1.2.1 Algorithmes de descente de gradient

Les méthodes de descente de gradient sont des techniques d'optimisation parmi les plus simples. Elles nécessite le calcul du gradient de la fonction objectif (5.2), notée $\bigtriangledown_{\boldsymbol{x}} \mathcal{J}(\boldsymbol{x})$. Dans sa forme la plus simple, la valeur estimée $\boldsymbol{x}(n)$ est mise à jour en $\boldsymbol{x}(n+1)$ en suivant une direction opposée à celle du gradient de la fonction coût qu'il s'agit de minimiser. Ainsi

$$\boldsymbol{x}(n+1) = \boldsymbol{x}(n) - \eta \bigtriangledown_{\boldsymbol{x}} \mathcal{J}(\boldsymbol{x}(n)) \tag{5.3}$$

où η est un pas, souvent optimisé en utilisant une procédure de recherche linéaire. Des alternatives plus efficaces existent, telles que la méthode de Newton. Malheureusement, la fonction objectif est intrinsèquement non-convexe. Ainsi un algorithme de gradient de descente a-t-il peu de chance de succès étant donné la présence de minimas locaux, nécessitant des exécutions répétées afin de retenir finalement une solution acceptable.

5.1.2.2 Méthodes de point fixe

L'expression du noyau reproduisant engendrant $\mathcal{H}$ fournit des informations utiles pour élaborer des techniques d'optimisation plus appropriées. Plus précisément, le gradient de l'expression (5.2) dispose d'une expression analytique pour la plupart des noyaux. En annulant celui-ci, on simplifie la procédure d'optimisation en aboutissant directement à une technique de point fixe. En prenant par exemple le noyau Gaussien, la minimisation de la fonction objectif (5.2) aboutit à

$$\bigtriangledown_{\boldsymbol{x}} \mathcal{J}(\boldsymbol{x}) = -\frac{2}{\sigma^2} \sum_{i=1}^{n} \alpha_i \exp\left(\frac{\|\boldsymbol{x} - \boldsymbol{x}_i\|^2}{2\sigma^2} \right) (\boldsymbol{x} - \boldsymbol{x}_i) \tag{5.4}$$

En annulant cette expression, on est amené au schéma suivant :

$$\boldsymbol{x}(n+1) = \frac{\sum_{i=1}^{n} \alpha_i \kappa(\boldsymbol{x}(n), \boldsymbol{x}_i) \boldsymbol{x}_i}{\sum_{i=1}^{n} \alpha_i \kappa(\boldsymbol{x}(n), \boldsymbol{x}_i)} \tag{5.5}$$

avec $\kappa(\boldsymbol{x}(n), \boldsymbol{x}_i) = \exp(\frac{\|\boldsymbol{x}(n)-\boldsymbol{x}_i\|^2}{2\sigma^2})$. Ce schéma peut être reproduit pour d'autres noyaux, par exemple le noyau polynomial de degré p.

Malheureusement, cette méthode de point fixe n'échappe pas aux minima locaux du fait de la non-convexité de la fonction coût et a tendance à être instable. Pour éviter cette situation, des problèmes régularisés peuvent être facilement formulés, comme dans [Abrahamsen & Hansen 2009]. Une propriété intéressante de la méthode du point fixe est que la solution du problème s'écrit sous la forme $\boldsymbol{x}^* = \sum_{i=1}^{n} \beta_i \boldsymbol{x}_i$ pour certains coefficients $\beta_1, \beta_2, \ldots, \beta_n$ à déterminer. Ainsi l'espace de recherche de ces paramètres est contrôlé, par opposition aux techniques de descente du gradient qui explorent l'espace entier.

5.1.2.3 MDS et pré-image

Le problème de la pré-image s'apparente à une réduction de dimension. Dans les deux cas, on cherche à incorporer les données dans un espace de dimension beaucoup plus faible. On rappelle que MDS injecte les données dans un espace de faible dimension en tachant de préserver les distances. Ce principe a été appliqué avec succès au problème de pré-image dans [Kwok & Tsang 2003]. Considérons chaque distance dans l'espace des caractéristiques $\delta_i^2 = \|\psi - \kappa(\boldsymbol{x}_i, \cdot)\|_{\mathcal{H}}^2$ et son homologue dans l'espace d'entrée $\|\boldsymbol{x}^* - \boldsymbol{x}_i\|^2$. De façon idéale, ces distances sont conservées par la transformation, ce qui signifie :

$$\|\boldsymbol{x}^* - \boldsymbol{x}_i\|^2 \approx \|\psi - \kappa(\boldsymbol{x}_i, \cdot)\|_{\mathcal{H}}^2 \tag{5.6}$$

pour tout $i = 1, 2, \ldots, n$. Il est facile de vérifier que s'il existe un i tel que $\boldsymbol{\psi} = \phi(\boldsymbol{x}_i)$, nous obtenons la pré-image $\boldsymbol{x}^* = \boldsymbol{x}_i$. Une manière pour résoudre ce problème consiste à minimiser l'erreur quadratique moyenne entre ces distances, soit

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}} \sum_{i=1}^{n} \left| \|\boldsymbol{x}^* - \boldsymbol{x}_i\|^2 - \|\psi - \kappa(\boldsymbol{x}_i, \cdot)\|_{\mathcal{H}}^2 \right|^2 \tag{5.7}$$

Pour résoudre ce problème d'optimisation, une première stratégie consiste à recourir à une méthode de point fixe à laquelle on aboutit naturellement en annulant le gradient de cette fonction coût. Ceci conduit à l'expression suivante :

$$\boldsymbol{x}^* = \frac{\sum_{i=1}^{n} (\|\boldsymbol{x}^* - \boldsymbol{x}_i\| - \delta_i^2)\, \boldsymbol{x}_i}{\sum_{i=1}^{n} (\|\boldsymbol{x}^* - \boldsymbol{x}_i\| - \delta_i^2)} \tag{5.8}$$

Une autre approche pour résoudre ce problème consiste à examiner séparément les égalités (5.6), pour chaque i, ce qui mène aux n équations :

$$2\langle \boldsymbol{x}^*, \boldsymbol{x}_i \rangle = \langle \boldsymbol{x}^*, \boldsymbol{x}^* \rangle + \langle \boldsymbol{x}_i, \boldsymbol{x}_i \rangle - \delta_i \tag{5.9}$$

pour $i = 1, 2, \ldots, n$. Dans ces expressions, l'inconnu apparaît également du côté droit, avec $\langle \boldsymbol{x}^*, \boldsymbol{x}^* \rangle$. Une façon de s'en départir est de considérer des données centrées $\boldsymbol{x}_i$. Dans ces conditions, en prenant la moyenne empirique de chaque membre de l'équation ci-dessus, on aboutit à

$$\langle \boldsymbol{x}^*, \boldsymbol{x}^* \rangle = \frac{1}{n} \sum_{i=1}^{n} (\delta_i^2 - \langle \boldsymbol{x}_i, \boldsymbol{x}_i \rangle) \tag{5.10}$$

Soit $\boldsymbol{\varepsilon}$ le vecteur ayant pour entrées $\frac{1}{n} \sum_{i=1}^{n} (\delta_i^2 - \langle \boldsymbol{x}_i, \boldsymbol{x}_i \rangle)$. Sous forme de matrice, le problème s'écrit

$$2\boldsymbol{X}^\top \boldsymbol{x}^* = \mathrm{diag}(\boldsymbol{X}^\top \boldsymbol{X}) - [\delta_1^2, \delta_2^2, \ldots, \delta_n^2]^\top + \boldsymbol{\varepsilon} \tag{5.11}$$

où $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n]$ et $\mathrm{diag}(\cdot)$ est l'opérateur diagonal, avec $\mathrm{diag}(\boldsymbol{X}^\top \boldsymbol{X})$ le vecteur colonne d'entrées $\langle \boldsymbol{x}_i, \boldsymbol{x}_i \rangle$. La pré-image est donc obtenue par la solution analytique suivante

$$\boldsymbol{x}^* = \frac{1}{2} (\boldsymbol{X}\boldsymbol{X}^\top)^{-1} \boldsymbol{X} (\mathrm{diag}(\boldsymbol{X}^\top \boldsymbol{X}) - [\delta_1^2, \delta_2^2, \ldots, \delta_n^2]^\top) \tag{5.12}$$

Pour que cette technique puisse être pratiquement mise en œuvre, à l'image de l'algorithme LLE, seul un certain voisinage est considéré pour cette méthode de pré-image. Cette approche ouvre des horizons sur une famille de techniques, déduites des méthodes de réduction de dimension et d'apprentissage de variété [Etyngier *et al.* 2007].

Au-delà de la préservation des distances tel qu'avec MDS, il est possible de développer des méthodes de pré-image préservant les produits scalaires. En utilisant une telle stratégie, la mesure angulaire est également conservée. Pour cette raison, cette méthode est dite par transformation conforme. Une technique proposée récemment dans [Honeine & Richard 2009] pour résoudre le problème de la pré-image repose sur une telle transformation. Cette approche est adoptée dans la suite, où elle sera détaillée, et appliquée au problème de démélange des données hyperspectrales.

5.2 Démélange supervisé par calcul de pré-image

Le principe de recherche d'une pré-image a été utilisé dans plusieurs applications, dont certaines dans [Honeine & Richard 2011b]. Parmi elles, on peut citer l'utilisation d'une pré-image conjointement avec KPCA pour l'extraction des caractéristiques, le débruitage d'images, l'auto-localisation dans les réseaux de capteurs, etc... Nous utilisons ici ce principe pour le démélange des images hyperspectrales.

On suppose disposer de pixel-vecteurs spectraux et des abondances correspondantes, ce qui constitue l'ensemble d'apprentissage, sur lequel on envisage d'appliquer une méthode d'apprentissage sous la forme d'une recherche d'une fonction de pré-image. Celle-ci est destinée à estimer les abondances de pixel-vecteurs non-étiquetés. Une telle approche est dite supervisée par rapport au cadre non-supervisé où, ni les spectres des composés purs, ni les abondances, ne sont connus. Voir quelques exemples d'approches supervisées dans [Tourneret *et al.* 2008,

Themelis *et al.* 2010, Altmann *et al.* 2011b]. Dans [Altmann *et al.* 2011b], la transformation qui estime les abondances pour tout pixel-vecteur est une combinaison linéaire de fonctions à base radiale. Les poids de ces fonctions sont estimés à partir des échantillons d'apprentissage. Un algorithme de moindres carrés orthogonaux est ensuite appliqué afin de réduire le nombre de fonctions de base intervenant dans le modèle. Dans ce chapitre, nous montrons que le processus d'apprentissage pour l'estimation des abondances à partir des données d'apprentissage peut être considéré comme un problème de pré-image. La résolution de ce problème vise à estimer une transformation inverse, de l'espace de grande dimension des pixel-vecteurs, vers l'espace de faible dimension des vecteurs d'abondance. On considère également le problème sélection de noyau, et montrons que les noyaux partiellement linéaires sont des modèles appropriés. La composante non-linéaire du noyau peut être conçue par des techniques de l'apprentissage de variétés telles que celles présentées au chapitre précédent. Une technique de régularisation spatiale est également discutée et expérimentée dans ce chapitre. La « régularisation TV »a été utilisée avec succès dans la littérature [Iordache *et al.* 2011]. En intégrant ce type d'information, nous montrons que l'on améliore les performances de nos algorithmes.

5.2.1 Modèle de mixage

Soit $\boldsymbol{r} = [r_1, r_2, \ldots, r_L]^\top$ un pixel-vecteur observé, avec L le nombre de bandes spectrales. Nous supposons que le pixel-vecteur $\boldsymbol{r}$ est un mélange de R spectres de composés purs $\boldsymbol{m}_i$. Notons $\boldsymbol{M} = [\boldsymbol{m}_1, \boldsymbol{m}_2, \ldots, \boldsymbol{m}_R]$ la matrice de taille $(L \times R)$ des signatures spectrales des composés purs, et $\boldsymbol{\alpha}$ le vecteur de dimension R des abondances associées à $\boldsymbol{r}$. On considère tout d'abord le modèle de mélange linéaire déjà présenté au chapitre 2, où chaque pixel-vecteur observé est une combinaison linéaire des spectres des composés purs, pondérés par les abondances, c'est-à-dire :

$$\boldsymbol{r} = \boldsymbol{M}\boldsymbol{\alpha} + \boldsymbol{n} \tag{5.13}$$

où $\boldsymbol{n}$ est un vecteur de bruit. Le vecteur d'abondance $\boldsymbol{\alpha}$ est généralement déterminé en minimisant une fonction coût, par exemple, l'erreur de reconstruction quadratique moyenne, sous contrainte de non-négativité et de somme unité

$$\begin{aligned} &\alpha_i \geq 0, \qquad \forall i \in 1, \ldots, R \\ &\sum_{i=1}^{R} \alpha_i = 1. \end{aligned} \tag{5.14}$$

Le modèle ci-dessus suppose que les vecteurs d'abondance $\boldsymbol{\alpha}$ se situent dans un simplexe de R sommets. Une conséquence directe est que les pixel-vecteurs $\boldsymbol{r}$ se trouvent également dans un simplexe dont les sommets sont les R spectres des composés purs. Il existe de nombreuses situations où l'on rencontre des phénomènes de diffusion multiple. Le modèle (5.13) peut donc être inapproprié et être avantageusement remplacé par un modèle non-linéaire. Soit le mécanisme de mélange général

$$\boldsymbol{r} = \boldsymbol{\Psi}(\boldsymbol{\alpha}, \boldsymbol{M}) + \boldsymbol{n} \tag{5.15}$$

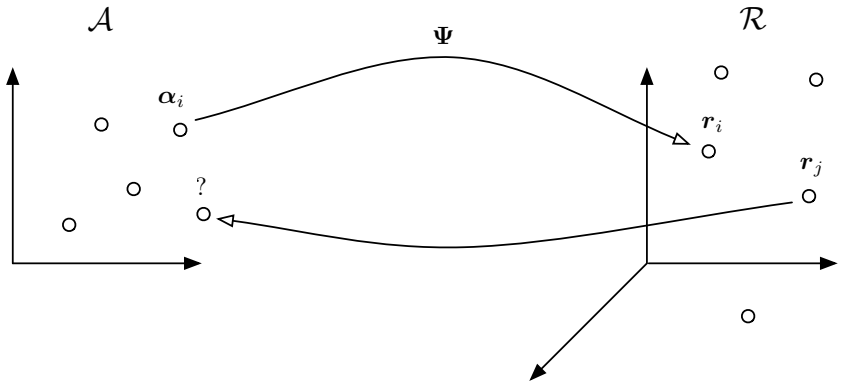


FIGURE 5.1 – Problème élémentaire de pré-image.

avec $\boldsymbol{\Psi}$ une fonction inconnue qui définit les interactions entre les composantes spectrales de la matrice $\boldsymbol{M}$, sous contraintes (5.14).

Comme illustrés dans la figure 5.1, les modèles (5.13) et (5.15) reposent tous deux sur une transformation de l'espace d'entrée de faible dimension $\mathcal{A}$ des vecteurs d'abondance $\boldsymbol{\alpha}$, en espace de sortie de grande dimension $\mathcal{R}$ des données hyperspectrales $\boldsymbol{r}$. Dans ce contexte, nous considérons le problème de l'estimation des abondances comme un problème pré-image [Honeine & Richard 2011a]. La résolution du problème de pré-image vise à estimer la transformation inverse qui permet d'obtenir $\boldsymbol{\alpha}$ à partir de $\boldsymbol{r}$, à partir des données d'apprentissage.

5.2.2 Méthode d'estimation de pré-image utilisée

Cette section présente un cadre original, basé sur le problème pré-image, pour le démélange supervisé des données hyperspectrales. Voir la figure 5.2. Afin de permettre au modèle de mieux représenter certains phénomènes de mélange complexes, on inscrit la méthodologie dans le contexte des espaces de Hilbert à noyau reproduisant (RKHS) en lieu et place de $\mathcal{R}$.

Soit $\mathcal{H}$ un espace de Hilbert de fonctions à valeurs réelles ψ définies sur $\mathcal{R}$, et soit $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ le produit scalaire associé à $\mathcal{H}$. On suppose l'existence d'une fonctionnelle d'évaluation δ_r, définie par $\delta_r[\psi] = \psi(\boldsymbol{r})$, linéaire par rapport à ψ et bornée pour tout $\boldsymbol{r}$ de $\mathcal{R}$. En vertu du théorème de représentation de Riesz, il existe une fonction définie positive unique $\boldsymbol{r} \mapsto \kappa(\boldsymbol{r}, \cdot)$ dans $\mathcal{H}$, qui satisfait [Aronszajn 1950]

$$\psi(\boldsymbol{r}') = \langle \psi, \kappa(\cdot, \boldsymbol{r}') \rangle_{\mathcal{H}}, \quad \forall \psi \in \mathcal{H} \tag{5.16}$$

pour tout $\boldsymbol{r}' \in \mathcal{R}$. Voir [Aronszajn 1950] pour une introduction détaillée. En remplaçant ψ par $\kappa(\cdot, \boldsymbol{r})$ dans (5.16), on obtient

$$\kappa(\boldsymbol{r}, \boldsymbol{r}') = \langle \kappa(\cdot, \boldsymbol{r}), \kappa(\cdot, \boldsymbol{r}') \rangle_{\mathcal{H}} \tag{5.17}$$

pour tout $\boldsymbol{r}$, $\boldsymbol{r}' \in \mathcal{R}$. L'équation (5.17) est à l'origine du terme générique *noyau reproduisant* pour qualifier κ. En notant par $\boldsymbol{\Phi}$ la transformation qui associe la

fonction noyau $\kappa(\cdot, \boldsymbol{r})$ à chaque donnée d'entrée $\boldsymbol{r}$, l'équation (5.17) implique que

$$\kappa(\boldsymbol{r}, \boldsymbol{r}') = \langle \boldsymbol{\Phi}(\boldsymbol{r}), \boldsymbol{\Phi}(\boldsymbol{r}') \rangle_{\mathcal{H}}. \tag{5.18}$$

Le noyau évalue donc le produit scalaire de toute paire d'éléments de $\mathcal{R}$ injectés dans l'espace $\mathcal{H}$ sans aucune connaissance explicite de $\boldsymbol{\Phi}$ et $\mathcal{H}$. Dans le domaine du machine learning, cet idée-clé est connue sous l'appellation d'*astuce du noyau*. Comme illustrée dans la figure 5.2, la transformation inverse vers l'espace $\mathcal{A}$ afin d'identifier $\boldsymbol{\alpha}$ étant donné $\kappa(\cdot, \boldsymbol{r})$ de $\mathcal{H}$ est une tâche essentielle. La méthode d'estimation de pré-image mise en œuvre ici a été récemment proposée dans [Honeine & Richard 2011b]. Elle vise à définir une transformation qui préserve les produits scalaires, dans l'espace d'entrée $\mathcal{A}$ et dans l'espace de fonctions $\mathcal{H}$. Etant donné $\boldsymbol{r}$, il permet donc d'estimer $\boldsymbol{\alpha}$ à partir de $\kappa(\cdot, \boldsymbol{r})$. La section suivante est consacrée à cette approche, et à son application au démélange supervisé.

5.3 Démélange supervisé

Étant donné un ensemble de données d'apprentissage $\{(\boldsymbol{\alpha}_1, \boldsymbol{r}_1), \ldots, (\boldsymbol{\alpha}_n, \boldsymbol{r}_n)\}$, on souhaite estimer la pré-image $\boldsymbol{\alpha}$ en $\mathcal{A}$ de tout élément $\kappa(\cdot, \boldsymbol{r})$ de $\mathcal{H}$. L'approche proposée comporte deux étapes : d'abord, l'apprentissage de la transformation inverse, puis l'estimation de la pré-image.

5.3.1 Étape 1 : Apprentissage de la transformation inverse

Par le théorème de représentation [Schölkopf *et al.* 2000], on peut limiter les investigations à l'espace engendré par les n fonctions noyau $\{\kappa(\cdot, \boldsymbol{r}_1), \ldots, \kappa(\cdot, \boldsymbol{r}_n)\}$. Concentrons-nous seulement sur un sous-espace de celui-ci, engendré par ℓ fonctions à déterminer notées $\{\psi_1, \ldots, \psi_\ell\}$ avec $\ell \leq n$, de la forme

$$\psi_k = \sum_{i=1}^{n} \lambda_{ki}\, \kappa(\cdot, \boldsymbol{r}_i), \quad k = 1, \ldots, \ell. \tag{5.19}$$

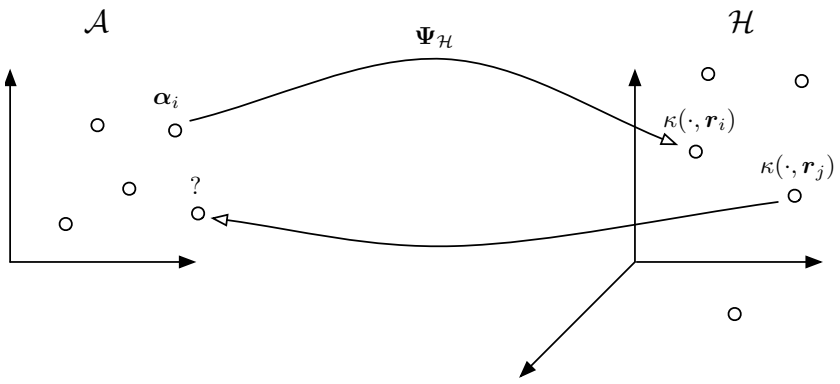


FIGURE 5.2 – Problème de préimage

Nous considérons également l'opérateur d'analyse $C : \mathcal{H} \to \mathbb{R}^\ell$ défini par

$$C\varphi = [\langle \varphi, \psi_1 \rangle_{\mathcal{H}} \dots \langle \varphi, \psi_\ell \rangle_{\mathcal{H}}]^\top. \tag{5.20}$$

A ce stade, il est important de noter que la k^{e} composante de la représentation de tout élément $\kappa(\cdot, \boldsymbol{r})$ de $\mathcal{H}$, définie par la famille $\{\psi_1, \dots, \psi_\ell\}$, est donnée par la projection de celui-ci sur ψ_k, soit

$$\langle \kappa(\cdot, \boldsymbol{r}), \psi_k \rangle_{\mathcal{H}} = \psi_k(\boldsymbol{r}) = \sum_{i=1}^{n} \lambda_{ki}\, \kappa(\boldsymbol{r}, \boldsymbol{r}_i). \tag{5.21}$$

La fonction noyau $\kappa(\cdot, \boldsymbol{r})$, qui est l'image de $\boldsymbol{r}$ dans $\mathcal{H}$, est donc représentée par le vecteur de longueur ℓ suivant

$$\boldsymbol{\psi}_r = [\psi_1(\boldsymbol{r})\, \psi_2(\boldsymbol{r})\, \dots\, \psi_\ell(\boldsymbol{r})]^\top. \tag{5.22}$$

Afin de définir complètement l'opérateur d'analyse C, en estimant les paramètres inconnus λ_{ki} à partir des données d'apprentissage $(\boldsymbol{\alpha}_i, \boldsymbol{r}_i)$, nous proposons de considérer la relation suivante entre les produits scalaires dans l'espace des abondances $\mathcal{A}$ et, avec un abus de notation, dans l'espace des fonctions de $\mathcal{H}$ paramétrées par les pixel-vecteurs

$$\boldsymbol{\alpha}_i^\top \boldsymbol{\alpha}_j = \boldsymbol{\psi}_{r_i}^\top \boldsymbol{\psi}_{r_j} + \varepsilon_{ij}, \quad \forall\, i, j = 1, \dots, n \tag{5.23}$$

où ε_{ij} désigne le manque d'ajustement du modèle ci-dessus. Notons qu'il n'y a aucune contrainte sur les fonctions d'analyse ψ_k, mise à part leur forme (5.19) et la contrainte d'ajustement (5.23). Évaluons maintenant les paramètres λ_{ki} dans (5.21) de sorte que la variance empirique de ε_{ij} soit minimum, c'est-à-dire :

$$\min_{\psi_1, \dots, \psi_\ell} \frac{1}{2} \sum_{i,j=1}^{n} (\boldsymbol{\alpha}_i^\top \boldsymbol{\alpha}_j - \boldsymbol{\psi}_{r_i}^\top \boldsymbol{\psi}_{r_j})^2 + \eta\, P(\psi_k, \dots, \psi_\ell) \tag{5.24}$$

où P est une fonction de régularisation, et η un paramètre réglable utilisé pour contrôler le compromis entre l'ajustement des données et la régularisation de la solution. Nous allons utiliser la pénalisation ℓ_2 ici, définie par

$$P(\psi_k, \dots, \psi_\ell) = \sum_{k=1}^{\ell} \|\psi_k\|_{\mathcal{H}}^2 \tag{5.25}$$

Le problème d'optimisation peut être réécrit sous forme matricielle :

$$\min_{\boldsymbol{L}} \frac{1}{2} \|\boldsymbol{A} - \boldsymbol{K}\boldsymbol{L}^\top \boldsymbol{L}\boldsymbol{K}\|_F^2 + \eta\, \mathrm{tr}(\boldsymbol{L}^\top \boldsymbol{L}\boldsymbol{K}) \tag{5.26}$$

où $\boldsymbol{A}$ et $\boldsymbol{K}$ sont les matrices de Gram, dont les (i, j)-èmes entrées respectivement définies par $\boldsymbol{\alpha}_i^\top \boldsymbol{\alpha}_j$ et $\kappa(\boldsymbol{r}_i, \boldsymbol{r}_j)$, et $\boldsymbol{L}$ la matrice d'entrée (i, j) donnée par λ_{ij}. En prenant la dérivée de ce coût par rapport à $\boldsymbol{L}^\top \boldsymbol{L}$ plutôt que par rapport à $\boldsymbol{L}$, nous obtenons

$$\hat{\boldsymbol{L}}^\top \hat{\boldsymbol{L}} = \boldsymbol{K}^{-1}(\boldsymbol{A} - \eta\, \boldsymbol{K}^{-1})\boldsymbol{K}^{-1} \tag{5.27}$$

Dans ce qui suit, nous montrons que seul $\hat{\boldsymbol{L}}^\top \hat{\boldsymbol{L}}$ est nécessaire pour calculer la pré-image.

5.3.2 Étape 2 : Estimation de la pré-image

Considérons d'abord le cas d'une fonction quelconque φ de $\mathcal{H}$ donnée, qui peut s'écrire comme suit

$$\varphi = \sum_{i=1}^{\ell} \phi_i \, \kappa(\cdot, \boldsymbol{r}_i) + \varphi^{\perp}, \tag{5.28}$$

les coefficients ϕ_i étant une donnée du problème, et avec $\varphi^{\perp}$ un complément orthogonal à l'espace engendré par les fonctions $\kappa(\cdot, \boldsymbol{r}_i)$. La k^{e} entrée de la représentation de φ est donc donnée par

$$\langle \varphi, \psi_k \rangle_{\mathcal{H}} = \sum_{i,j=1}^{n} \phi_i \, \lambda_{kj} \, \kappa(\boldsymbol{r}_i, \boldsymbol{r}_j). \tag{5.29}$$

Soit $\boldsymbol{\varphi}$ la représentation de φ par l'opérateur d'analyse C. La minimisation de la variance de l'erreur d'ajustement définie dans (5.23), par rapport à la pré-image $\boldsymbol{\alpha}$ étant donné $\boldsymbol{\varphi}$, entre $\boldsymbol{\alpha}^{\top}\boldsymbol{\alpha}_i$ et $\boldsymbol{\varphi}^{\top}\boldsymbol{\psi}_{r_i}$ pour $i = 1, \ldots, n$, conduit au problème d'optimisation

$$\begin{aligned} \hat{\boldsymbol{\alpha}} &= \arg\min_{\boldsymbol{\alpha}} \frac{1}{2} \|\boldsymbol{\Lambda}^{\top}\boldsymbol{\alpha} - \boldsymbol{K}\hat{\boldsymbol{L}}^{\top}\hat{\boldsymbol{L}}\boldsymbol{K}\boldsymbol{\phi}\|^2 \\ &= \arg\min_{\boldsymbol{\alpha}} \frac{1}{2} \|\boldsymbol{\Lambda}^{\top}\boldsymbol{\alpha} - (\boldsymbol{A} - \eta \, \boldsymbol{K}^{-1})\boldsymbol{\phi}\|^2 \end{aligned} \tag{5.30}$$

sous contraintes de non-négativité et de somme unité (1.3). $\boldsymbol{\Lambda}$ est la matrice dont la i^{e} colonne est le vecteur $\boldsymbol{\alpha}_i$, et $\boldsymbol{\phi}$ est le vecteur dont la i^{e} entrée est ϕ_i, $i = 1, \ldots, n$.

On considère maintenant le cas particulier où l'on cherche la pré-image $\boldsymbol{\alpha}$ de fonctions de la forme $\kappa(\cdot, \boldsymbol{r})$. Le problème d'apprentissage ci-dessus, par rapport à la pré-image $\boldsymbol{\alpha}$ de $\kappa(\cdot, \boldsymbol{r})$, se ramène à

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \frac{1}{2} \|\boldsymbol{\Lambda}^{\top}\boldsymbol{\alpha} - (\boldsymbol{A} - \eta \, \boldsymbol{K}^{-1})\boldsymbol{K}^{-1}\boldsymbol{\psi}_r\|^2 \tag{5.31}$$

sous les contraintes de non-négativité et de somme unité toujours, avec $\boldsymbol{\psi}_r$ le vecteur défini dans (5.22). Ce problème d'optimisation convexe peut être résolu par la méthode FCLS [Heinz & Chang 2001], ou par [Chen *et al.* 2011].

5.4 Sélection de noyaux

La fonction noyau $\kappa(\cdot, \boldsymbol{r})$ injecte les observations $\boldsymbol{r}$ dans un espace à dimension très élevé, voir infinie, $\mathcal{H}$. Elle caractérise l'espace des solutions pour la relation non-linéaire entre les données d'entrée et les données sorties $\boldsymbol{\alpha}$ et $\boldsymbol{r}$. Les exemples classiques sont le noyau Gaussien $\kappa(\boldsymbol{r}_i, \boldsymbol{r}_j) = \exp\bigl(-\|\boldsymbol{r}_i - \boldsymbol{r}_i\|^2/2\sigma^2\bigr)$, avec σ la largeur de bande, et les noyaux polynomiaux non-homogènes $\kappa(\boldsymbol{r}_i, \boldsymbol{r}_j) = (1 + \boldsymbol{r}_i^{\top}\boldsymbol{r}_j)^q$, avec $q \in \mathbb{N}^*$. Nous allons à présent proposer quelques noyaux spécifiques, et les tester dans la section suivante. D'une part, nous proposons de concevoir directement des noyaux à partir des informations extraites des données. D'autre part, nous allons présenter un noyau partiellement linéaire qui a prouvé son efficacité dans les problèmes de démélange non-linéaire des données hyperspectrales [Chen *et al.* 2012].

5.4.1 Noyaux basés sur l'apprentissage de variété

Dans [Ham *et al.* 2003] et comme représenté dans la chapitre 2, le problème de l'apprentissage de variété est traité avec KACP. Le principe de mise en évidence de la structure sous-jacente des données peut être envisagé comme une méthode de réduction de dimension non linéaire, basé sur des informations locales avec l'algorithme LLE [Roweis & Saul 2000], ou une distance géodésique avec l'algorithme Isomap [de Silva & Tenenbaum 2003]. Ces techniques peuvent être utilisées pour concevoir des noyaux qui préservent certaines caractéristiques de la structure de variété de l'espace $\mathcal{R}$ auquel les $\boldsymbol{r}_i$ appartiennent, dans l'espace $\mathcal{H}$ des $\kappa(\cdot, \boldsymbol{r}_i)$.

À titre d'exemple, on considère les noyaux à base radiale $\kappa(\boldsymbol{r}_i, \boldsymbol{r}_j) = f(\|\boldsymbol{r}_i - \boldsymbol{r}_i\|)$ avec $f \in \mathcal{C}_\infty$. Une condition suffisante pour que cette classe de noyaux soit définie positive est la monotonie complète de la fonction f, qui est définie comme suit,

$$(-1)^k f^{(k)}(r) \geq 0, \quad \forall r \geq 0 \tag{5.32}$$

où $f^{(k)}$ désigne la k^{eme} dérivée de f [Cucker & Smale 2002]. Au lieu d'utiliser la distance euclidienne $d_{ij} = \|\boldsymbol{r}_i - \boldsymbol{r}_j\|$ avec $f(\cdot)$, on peut utiliser les distances fournies par Isomap. Cette approche consiste à construire un graphe d'adjacence symétrique à l'aide d'un critère de k plus proches voisins, et d'appliquer l'algorithme de Dijkstra pour calculer le plus court chemin sur ce graphe entre chaque donnée. Malheureusement, la matrice de Gram $\boldsymbol{K}_{\text{iso}}$ construite de telle manière n'a aucune garantie d'être définie positive. Cette difficulté peut être surmontée en utilisant MDS, qui envoie les données dans un sous-espace euclidien de faible dimension, où les longueurs d'arêtes sont au mieux préservées. Une alternative est de forcer la matrice $\boldsymbol{K}_{\text{iso}}$ à être définie positive en utilisant l'une des approches décrites dans [Munoz & Diego 2006].

5.4.2 Noyaux partiellement linéaires

Le modèle (5.13) suppose une relation linéaire entre les coefficients d'abondance et les pixel-vecteurs. Il existe toutefois de nombreuses situations décrites dans un précédent chapitre où ce modèle est inapproprié, et peut être remplacé par un modèle non-linéaire. Dans [Chen *et al.* 2012], les auteurs justifient et étudient un noyau partiellement linéaire, constitué d'une composante linéaire paramétrée par la matrice $\boldsymbol{M}$ de spectres des composés purs, et d'une composante non-linéaire supposée combler les manques du modèle pour mimer les phénomènes de réflexions multiples. De nombreuses expérimentations, sur des données simulées et réelles, ont permis d'illustrer la flexibilité et l'efficacité de cette classe de modèles. Dans le même esprit que cette étude, pour le problème de démélange par une méthode de pré-image, nous suggérons d'utiliser des noyaux de la forme

$$\kappa'(\boldsymbol{r}_i, \boldsymbol{r}_j) = (1-\gamma)\boldsymbol{r}_i^\top \boldsymbol{\Sigma} \boldsymbol{r}_j + \gamma \kappa'(\boldsymbol{r}_i, \boldsymbol{r}_j) \tag{5.33}$$

avec $\kappa'(\boldsymbol{r}_i, \boldsymbol{r}_j)$ un noyau reproduisant, $\boldsymbol{\Sigma}$ une matrice semi-définie positive, et γ un paramètre de l'intervalle $[0, 1]$ utilisé pour ajuster un compromis entre les composantes linéaires et non-linéaires du mélange. Dans les expérimentations à suivre,

nous avons utilisé ce noyau avec $\boldsymbol{\Sigma} = (\boldsymbol{M}\boldsymbol{M}^\top)^\dagger$, soit

$$\kappa'(\boldsymbol{r}_i, \boldsymbol{r}_j) = (1-\gamma)\boldsymbol{r}_i^\top (\boldsymbol{M}\boldsymbol{M}^\top)^\dagger \boldsymbol{r}_j + \gamma\kappa'(\boldsymbol{r}_i, \boldsymbol{r}_j) \tag{5.34}$$

où $(\cdot)^\dagger$ désigne l'opérateur pseudo-inverse. On peut montrer qu'ainsi, pour $\gamma = 0$, le noyau ci-dessus conduit à l'estimateur du maximum de vraisemblance en présence d'un scénario de mélange linéaire.

5.5 Régularisation spatiale

5.5.1 Formulation

La section précédente a été consacrée à l'estimation des abondances par apprentissage d'une transformation inverse. A présent, nous visons à améliorer les résultats en intégrant la corrélation spatiale entre les pixel-vecteurs dans un voisinage. Le problème a été abordé dans [Iordache *et al.* 2011], dans le cas linéaire, en ajoutant une régularisation de type « variation totale »(TV). Cette solution a été étendue au problème de démélange non-linéaire par méthodes à noyau dans [Chen *et al.* 2012], notamment avec le noyau partiellement linéaire évoqué précédemment. Ici, nous adaptons ce principe à la méthode de pré-image développée dans ce chapitre.

Soit $\boldsymbol{A}$ la matrice des vecteurs d'abondance, c'est-à-dire $\boldsymbol{A} = [\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n]$. Afin de prendre en considération les relations entre pixel-vecteurs voisins, le problème de démélange peut alors être résolu en minimisant la fonction de coût générale suivante par rapport à la matrice $\boldsymbol{A}$

$$J(\boldsymbol{A}) = J_{\text{err}}(\boldsymbol{A}) + \nu\, J_{\text{sp}}(\boldsymbol{A}) \tag{5.35}$$

sous les contraintes de non-négativité imposées à chaque entrée de $\boldsymbol{A}$, et de somme unité imposée sur chaque colonne de $\boldsymbol{A}$. Pour faciliter les notations, ces deux contraintes seront exprimées par

$$\begin{aligned} &\boldsymbol{A} \succeq \boldsymbol{0} \\ &\boldsymbol{A}^\top \boldsymbol{1}_R = \boldsymbol{1}_N \end{aligned} \tag{5.36}$$

La fonction $J_{\text{err}}(\boldsymbol{A})$ représente l'erreur de modélisation, tandis que $J_{\text{sp}}(\boldsymbol{A})$ est un terme de régularisation pour favoriser la similitude des abondances de pixel-vecteurs voisins. Le paramètre non-négatif ν contrôle le compromis entre la fidélité aux données et les similitudes inter pixel-vecteurs.

Afin de prendre les relations spatiales entre pixel-vecteurs en considération, la fonction de régularisation suivante est utilisée

$$J_{\text{sp}}(\boldsymbol{A}) = \sum_{n=1}^{N} \sum_{m \in \mathcal{N}(n)} \|\boldsymbol{\alpha}_n - \boldsymbol{\alpha}_m\|_1 \tag{5.37}$$

où $\|\cdot\|_1$ désigne la norme ℓ_1 d'un vecteur, et $\mathcal{N}(n)$ l'ensemble des voisins du pixel n. Ce terme de régularisation favorise l'homogénéité spatiale des pixel-vecteurs voisins,

qui devrait être caractérisée par des vecteurs d'abondance proches au sens d'une certaine métrique. Sans perte de généralité, dans ce chapitre, nous limitons le voisinage du pixel n en prenant les 4 pixels les plus proches, soit indexés par $n-1$ et $n+1$ (contiguïté de ligne), $n-w$ et $n+w$ (contiguïté de colonne). Nous définissons les matrices, de taille $(N \times N)$, $\boldsymbol{H}_{\leftarrow}$ et $\boldsymbol{H}_{\rightarrow}$ qui calculent la différence entre chaque vecteur d'abondance et son voisin de gauche, et de droite, respectivement. De même, nous notons $\boldsymbol{H}_{\uparrow}$ et $\boldsymbol{H}_{\downarrow}$ les matrices calculant la différence entre les vecteurs d'abondance et leur voisin du haut, et du bas, respectivement. Avec ces notations, la fonction de régularisation (5.37) peut être réécrite sous forme matricielle ainsi

$$J_{\text{sp}}(\boldsymbol{A}) = \|\boldsymbol{A}\boldsymbol{H}\|_{1,1} \tag{5.38}$$

avec $\boldsymbol{H}$ la matrice $\left(\boldsymbol{H}_{\leftarrow}\, \boldsymbol{H}_{\rightarrow}\, \boldsymbol{H}_{\uparrow}\, \boldsymbol{H}_{\downarrow}\right)$ de taille $(N \times 4N)$ et $\|\,\|_{1,1}$ la somme des normes ℓ_1 des colonnes d'une matrice. On relève que cette fonction de régularisation est convexe mais non lisse, ce qui nécessite des algorithmes d'optimisation adaptés. En considérant à la fois l'erreur de modélisation et le terme de régularisation, le problème d'optimisation devient

$$\begin{aligned} \min_{\boldsymbol{A}} \sum_{n=1}^{N} \frac{1}{2} \|\boldsymbol{\Lambda}^\top \boldsymbol{\alpha}_n - (\boldsymbol{A} - \eta\, \boldsymbol{K}^{-1}) \boldsymbol{K}^{-1} \boldsymbol{\psi}_r\|^2 + \nu\, \|\boldsymbol{A}\, \boldsymbol{H}\|_{1,1} \\ \text{sous contrainte de} \quad \boldsymbol{A} \succeq 0 \quad \text{et} \quad \boldsymbol{A}^\top \mathbf{1}_R = \mathbf{1}_N \end{aligned} \tag{5.39}$$

où ν contrôle le compromis entre la fidélité aux données et les similitudes inter pixel-vecteurs. Plus simplement, nous écrirons $\boldsymbol{A} \in \mathcal{S}_{+1}$ pour désigner les contraintes de positivité et de somme unité sur $\boldsymbol{A}$.

5.5.2 Solution

Même si le problème (5.39) est convexe, il ne peut être résolu facilement en raison du terme de régularisation non lisse. Afin de remédier à cet inconvénient, nous le réécrivons sous la forme équivalente suivante

$$\begin{aligned} \min_{\boldsymbol{A} \in \mathcal{S}_{+1}} \sum_{n=1}^{N} \frac{1}{2} \|\boldsymbol{\Lambda}^\top \boldsymbol{\alpha}_n - (\boldsymbol{A} - \eta\, \boldsymbol{K}^{-1}) \boldsymbol{K}^{-1} \boldsymbol{\psi}_r\|^2 + \nu\, \|\boldsymbol{U}\|_{1,1} \\ \text{sous contrainte de} \quad \boldsymbol{V} = \boldsymbol{A} \quad \text{et} \quad \boldsymbol{U} = \boldsymbol{V}\boldsymbol{H} \end{aligned} \tag{5.40}$$

où deux matrices $\boldsymbol{U}$ et $\boldsymbol{V}$, et deux contraintes, ont été introduites. Cette approche à variables séparées a été initialement proposée dans [Goldstein & Osher 2009]. La matrice $\boldsymbol{U}$ permet de découpler la fonction de régularisation non-lisse du problème quadratique. La matrice $\boldsymbol{V}$ permet de relaxer les liens entre les pixel-vecteurs. Comme étudié dans [Goldstein & Osher 2009], cette méthode dite de Bregman est particulièrement efficace pour faire face à une large classe de problèmes régularisés

par la norme ℓ_1. En appliquant ce cadre à (5.39), la formulation suivante est obtenue

$$\begin{aligned}\boldsymbol{A}^{(k+1)},\boldsymbol{V}^{(k+1)},\boldsymbol{U}^{(k+1)} = \\ \underset{\boldsymbol{A}\in\mathcal{S}_{+1},\boldsymbol{V},\boldsymbol{U}}{\arg\min} \sum_{n=1}^{N} \frac{1}{2}\|\boldsymbol{\Lambda}^\top\boldsymbol{\alpha}_n - (\boldsymbol{A}-\eta\,\boldsymbol{K}^{-1})\boldsymbol{K}^{-1}\boldsymbol{\psi}_r\|^2 \\ + \nu\,\|\boldsymbol{U}\|_{1,1} + \frac{\zeta}{2}\|\boldsymbol{A}-\boldsymbol{V}-\boldsymbol{D}_1^{(k)}\|_F^2 + \frac{\zeta}{2}\|\boldsymbol{U}-\boldsymbol{V}\boldsymbol{H}-\boldsymbol{D}_2^{(k)}\|_F^2\end{aligned} \tag{5.41}$$

avec

$$\begin{aligned}\boldsymbol{D}_1^{(k+1)} &= \boldsymbol{D}_1^{(k)} + \left(\boldsymbol{V}^{(k+1)} - \boldsymbol{A}^{(k+1)}\right) \\ \boldsymbol{D}_2^{(k+1)} &= \boldsymbol{D}_2^{(k)} + \left(\boldsymbol{V}^{(k+1)}\boldsymbol{H} - \boldsymbol{U}^{(k+1)}\right)\end{aligned} \tag{5.42}$$

où $\|\cdot\|_F^2$ désigne la norme de Frobenius, et ζ un paramètre positif. Parce que l'on est parvenu à séparer certaines composantes de la fonction de coût, on est à présent en mesure de minimiser celle-ci efficacement en minimisant itérativement par rapport à $\boldsymbol{A}$, $\boldsymbol{V}$ et $\boldsymbol{U}$. Les trois étapes correspondantes sont les suivantes :

5.5.2.1 Étape 1 : Optimisation par rapport à $\boldsymbol{A}$

Le problème d'optimisation (5.41) se limite à

$$\boldsymbol{A}^{(k+1)} = \underset{\boldsymbol{A}\in\mathcal{S}_{+1}}{\arg\min} \sum_{n=1}^{N} \frac{1}{2}\left(\|\boldsymbol{\Lambda}^\top\boldsymbol{\alpha}_n - (\boldsymbol{A}-\eta\,\boldsymbol{K}^{-1})\boldsymbol{K}^{-1}\boldsymbol{\psi}_r\|^2 + \zeta\|\boldsymbol{\alpha}_n - \boldsymbol{\xi}_n^{(k)}\|^2\right) \tag{5.43}$$

où $\boldsymbol{\xi}_n^{(k)} = \boldsymbol{V}_n^{(k)} - \boldsymbol{D}_{1,n}^{(k)}$. Ici, $\boldsymbol{V}_n$ et $\boldsymbol{D}_{1,n}$ désignent respectivement la n^{e} colonne de $\boldsymbol{V}$ et $\boldsymbol{D}_1$. On peut observer que ce problème peut être décomposé en sous-problèmes, chacun impliquant un vecteur d'abondance $\boldsymbol{\alpha}_n$. Ceci résulte de l'utilisation de la matrice $\boldsymbol{V}$ dans la formulation (5.41). Considérons le problème d'optimisation local

$$\begin{aligned}\boldsymbol{\alpha}_n^{(k+1)} = \underset{\boldsymbol{\alpha}_n}{\arg\min}\frac{1}{2}\|\boldsymbol{\Lambda}^\top\boldsymbol{\alpha}_n - (\boldsymbol{A}-\eta\,\boldsymbol{K}^{-1})\boldsymbol{K}^{-1}\boldsymbol{\psi}_r\|^2 + \zeta\|\boldsymbol{\alpha}_n - \boldsymbol{\xi}_n^{(k)}\|^2 \\ \text{sous contrainte de} \quad \boldsymbol{\alpha}_n \succeq 0 \\ \boldsymbol{\alpha}_n^\top\mathbf{1}_R = 1\end{aligned} \tag{5.44}$$

L'estimation de $\boldsymbol{\alpha}_n$ est un problème quadratique avec contraintes qui peut être résolu aisément. Ce processus doit être répété pour $n = 1,\ldots,N$ afin d'obtenir $\boldsymbol{A}^{(k+1)}$. Les matrices $\boldsymbol{A}^{(k+1)}$ et $\boldsymbol{V}^{(k+1)}$, dont le calcul est présenté ci-après, permettent d'évaluer la matrice $\boldsymbol{D}_1^{(k+1)}$ en utilisant l'équation (5.42).

5.5.2.2 Étape 2 : Optimisation par rapport à $\boldsymbol{V}$

Le problème d'optimisation (5.41) se trouve réduit à

$$\boldsymbol{V}^{(k+1)} = \underset{\boldsymbol{V}}{\arg\min}\,\|\boldsymbol{A}^{(k)} - \boldsymbol{V} - \boldsymbol{D}_1^{(k)}\|_F^2 + \|\boldsymbol{U}^{(k)} - \boldsymbol{V}\boldsymbol{H} - \boldsymbol{D}_2^{(k)}\|_F^2 \tag{5.45}$$

L'annulation de la dérivée de (5.45) par rapport à $\boldsymbol{V}$ conduit à

$$\left(\boldsymbol{A}^{(k)} - \boldsymbol{V} - \boldsymbol{D}_1^{(k)}\right) - \left(\boldsymbol{U}^{(k)} - \boldsymbol{V}\boldsymbol{H} - \boldsymbol{D}_2^{(k)}\right)\boldsymbol{H}^\top = 0 \tag{5.46}$$

dont la solution est alors donnée par

$$\boldsymbol{V}^{(k+1)} = \left(\boldsymbol{A}^{(k)} - \boldsymbol{D}_1^{(k)} + (\boldsymbol{U}^{(k)} - \boldsymbol{D}_2^{(k)})\boldsymbol{H}^\top\right)(\boldsymbol{I} + \boldsymbol{H}\boldsymbol{H}^\top)^{-1} \tag{5.47}$$

5.5.2.3 Étape 3 : Optimisation par rapport à $\boldsymbol{U}$

Enfin, le problème d'optimisation qu'il reste à considérer est le suivant

$$\boldsymbol{U}^{(k+1)} = \arg\min_{\boldsymbol{U}} \nu\|\boldsymbol{U}\|_{1,1} + \frac{\zeta}{2}\|\boldsymbol{U} - \boldsymbol{V}^{(k)}\boldsymbol{H} - \boldsymbol{D}_2^{(k)}\|_F^2 \tag{5.48}$$

Sa solution peut être exprimée par la fonction de seuil souple (soft marin), soit

$$\boldsymbol{U}^{(k+1)} = \text{Thresh}\left(\boldsymbol{V}^{(k)}\boldsymbol{H} + \boldsymbol{D}_2^{(k)}, \frac{\sigma}{\zeta}\right) \tag{5.49}$$

où $\text{Thresh}(\cdot, \tau)$ désigne l'application composante par composante de la fonction de seuil souple définie ainsi

$$\text{Thresh}(x, \tau) = \text{sign}(x)\ \max(|x| - \tau, 0) \tag{5.50}$$

En conclusion, le problème (5.40) est résolu en appliquant de manière itérative (5.41) et (5.42), où l'optimisation de (5.41) peut être effectuée par les étapes 1 à 3. Ces itérations se poursuivent jusqu'à ce qu'un critère d'arrêt soit satisfait. On peut montrer que, si le problème (5.41) a une solution $\boldsymbol{A}^*$ quel que soit $\zeta > 0$, alors la séquence générée $\boldsymbol{A}^{(k)}$ converge vers l'optimum $\boldsymbol{A}^*$ [Eckstein & Bertsekas 1992].

5.6 Expérimentations

5.6.1 Sans régularisation spatiale

Dans cette section, on présente plusieurs tests afin de valider les algorithmes, avant de prendre en compte la régularité spatiale. Les scènes de test considérées sont, soit synthétiques, soit réelles.

5.6.1.1 Abondances uniformes sur le simplexe

Nous avons généré arbitrairement des vecteurs d'abondance selon une distribution uniforme sur le simplexe défini par les contraintes (5.14), puis les avons utilisés pour synthétiser une image de 50×50 de pixel-vecteurs à partir de spectres extraits de la bibliothèque ENVI. Ces spectres se composent de 420 bandes contiguës, couvrant la plage des longueurs d'onde de $0,3951$ à $2,56$ micromètres. Les trois modèles

considérés étaient : le modèle linéaire, le modèle bilinéaire généralisé avec $\gamma_{ij} = 1$, et le modèle post non-linéaire défini par [Jutten & Karhunen 2003b]

$$\boldsymbol{r} = (\boldsymbol{M}\boldsymbol{\alpha})^{\xi} + \boldsymbol{n} \tag{5.51}$$

où $(\cdot)^{\xi}$ désigne la valeur exponentielle ξ appliquée à chaque entrée du vecteur en argument. Le paramètre ξ a été fixé à $0,7$. Les images ont été corrompues avec un bruit blanc Gaussien additif, selon deux niveaux SNR = 30dB et SNR = 15dB.

Pour une meilleure comparaison des performances, nous avons considéré les algorithmes suivants qui n'opèrent pas à partir d'une base d'apprentissage : FCLS, KFCLS avec un noyau gaussien de bande passante $\sigma = 4$, et l'algorithme bayésien dédié au modèle bilinéaire généralisé (BilBay) [Halimi *et al.* 2011b].

Ces algorithmes ont été comparés à des méthodes nécessitant la connaissance d'une base d'apprentissage, dont celui présenté précédemment

- **Méthode à fonctions radiales de base et OLS**[Altmann *et al.* 2011a] : Cette méthode représente la fonction inverse non-linéaire en utilisant une combinaison linéaire de fonctions radiales de base (FBRs). Les valeurs moyennes des fonctions de base coïncident avec les données d'apprentissage. Les coefficients sont estimés à partir des données d'apprentissage. Le nombre de fonctions utilisées est limité en utilisant une méthode de type OLS. Ceci permet d'établir un compromis entre les performances obtenues et le coût calculatoire.
- **Méthode proposée, avec plusieurs choix de noyaux** : Trois types de noyaux ont été considérés : le noyau polynomial homogène de degré $d = 2$, le noyau gaussien de largeur de bande $\sigma = 4$, et enfin le noyau partiellement linéaire avec $\gamma = 0.1$ et une composante non-linéaire gaussienne de paramètre $\sigma = 4$. Le terme de régularisation η a été fixé à $\eta = 10^{-3}$. Les noyaux basés sur l'apprentissage de variétés (Isomap) étant approprié pour les données structurées selon une variété, nous avons choisi d'en réserver l'usage pour des données Swissroll présentées dans le chapitre précédent.

La taille de la base d'apprentissage a été fixée à $T = 200$. L'erreur quadratique de reconstruction

$$\text{RMSE} = \sqrt{\sum_{i=1}^{N} \|\hat{\boldsymbol{a}}_i - \boldsymbol{a}_i\|^2}$$

entre les abondances estimées et les abondances réelles a été calculée pour quantifier l'efficacité des algorithmes. Deux scènes constituées de $R = 3$ et $R = 5$ ont été considérées. Les résultats sont présentés dans le tableau 5.1 et le tableau 5.2.

Nous pouvons constater que FCLS s'est avéré efficace pour le démélange des données linéairement combinées, puisque le modèle de mélange correspond au modèle de démélange. L'algorithme bayésien issu du modèle bilinéaire généralisé s'est avéré performant lorsqu'il a été confronté aux modèles de mélange linéaire et bilinéaire, mais inefficace face à un modèle de mélange post non-linéaire sauf à très faible SNR. Avec les algorithmes basés sur les données d'apprentissage, de très bonnes performances et une robustesse vis-à-vis du bruit ont été observées. La méthode FBRs a mené à de bons résultats, mais est peu flexible vis-à-vis des non-linéarités que

TABLE 5.1 – Scène 1 (trois matériaux) : comparaison des RMSE

	SNR = 30 dB			SNR = 15 dB		
	linéaire	bilinéaire	PNMM	linéaire	bilinéaire	PNMM
FCLS	0.0037	0.0758	0.0604	0.0212	0.0960	0.0886
KFCLS	0.0054	0.2711	0.2371	0.0296	0.2694	0.2372
BilBay	0.0384	0.0285	0.1158	0.1135	0.1059	0.1191
RBFN-OLS	0.0144	0.0181	0.0170	0.0561	0.0695	0.0730
Pré-image (polynomial)	0.0139	0.0221	0.0129	0.0592	0.0601	0.0764
Pré-image (Gaussien)	0.0086	0.0104	0.0103	0.0422	0.0561	0.0597
Pré-image (partiel. linéaire)	0.0072	0.0096	0.0098	0.0372	0.0395	0.0514

TABLE 5.2 – Scène 2 (cinq matériaux) : comparaison des RMSE

	SNR = 30 dB			SNR = 15 dB		
	linéaire	bilinéaire	PNMM	linéaire	bilinéaire	PNMM
FCLS	0.0134	0.1137	0.1428	0.0657	0.1444	0.1611
KFCLS	0.0200	0.2051	0.1955	0.0890	0.1884	0.1572
BilBay	0.0585	0.0441	0.1741	0.1465	0.1007	0.1609
RBFN-OLS	0.0200	0.0236	0.0259	0.0777	0.0805	0.0839
Pré-image (polynomial)	0.025	0.0267	0.0348	0.0905	0.0903	0.1
Pré-image (Gaussien)	0.0186	0.0233	0.0245	0.0775	0.0778	0.0875
Pré-image(partiel. linéaire)	0.0148	0.0184	0.0203	0.0636	0.0616	0.0763

l'expérimentateur voudrait pouvoir choisir. L'approche proposée dispose de cette flexibilité totale, et s'avère nettement couteuse en temps de calcul.

5.6.1.2 Données de Swissroll

L'algorithme proposé a été mis en œuvre sur des données artificiellement générées selon la variété Swissroll, en considérant un noyau partiellement linéaire. Successivement, pour la composante non-linéaire, un noyau Gaussien de largeur $\sigma = 4$, et un noyau construit à partir des distances géodésiques, ont été considérés.

Nous pouvons faire une comparaison rapide des algorithmes utilisant ou non le noyau basé sur l'apprentissage avec un ensemble de données artificielles de variété Swissroll comme décrit dans le chapitre précédent. La figure (5.3) illustre que ce dernier a conduit à de meilleures performances, même pour un paramètre de non-linéarité important.

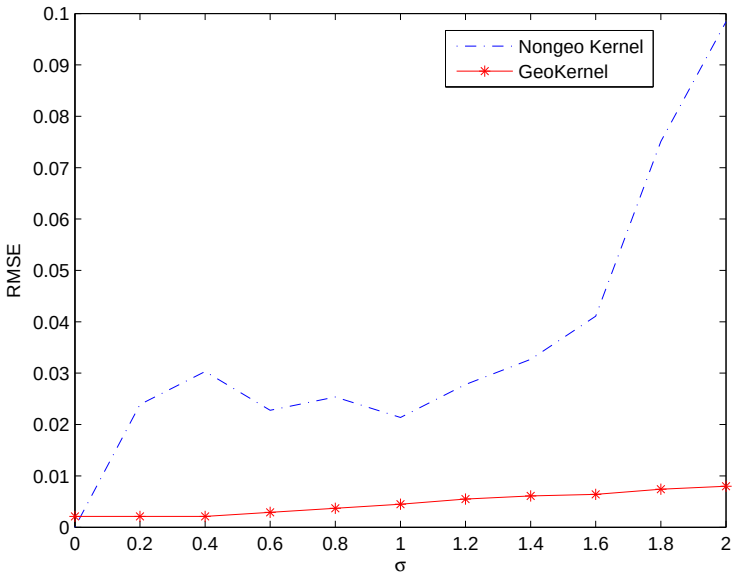


FIGURE 5.3 – Démélange des données Swissroll par calcul de pré-image

5.6.1.3 Données réelles

La scène considérée pour cette expérience est celle acquise par AVIRIS sur Moffet Field. Elle est composée de $L = 189$ bandes spectrales. Une sous-image de taille 50×50 pixels a été choisie. Les composés purs, au nombre de trois (végétation, sol, eau) ont été extraits à l'aide de l'algorithme VCA. La figure 5.4 montre le résultat obtenu avec l'algorithme de pré-image à noyau partiellement linéaire, en utilisant les mêmes paramètres que dans la section précédente. Le graphe des abondances obtenues distingue clairement les trois composantes présentes dans l'image.

5.6.2 Application de la régularisation spatiale

Deux images constituées de pixel-vecteurs spatialement corrélés ont été générées pour les expériences suivantes. Les spectres des composés purs utilisés ont été tirés au hasard dans la bibliothèque ENVI.

Le premier cube de données IM1, composées de 50×50 pixel-vecteurs, est généré à l'aide de cinq signatures. Les pixel-vecteurs de fond correspondent à un mélange des composés purs selon les proportions $[0.1149, 0.0741, 0.2003, 0.2055, 0.4051]^\top$. La première ligne de la figure 5.5 fournit les cartes d'abondance en vérité terrain des 5 composés purs considérés. Les données de réflectance ont été générées à partir d'un modèle bilinéaire de la forme

$$\boldsymbol{r}_i = \boldsymbol{M}\boldsymbol{\alpha}_i + \sum_{p=1}^{R} \sum_{q=p+1}^{R} \alpha_{i,p}\, \alpha_{i,q}\, \boldsymbol{m}_p \otimes \boldsymbol{m}_q + \boldsymbol{v}_i \tag{5.52}$$

où $\otimes$ désigne le produit d'Hadamard, dans les proportions fournies par les cartes

FIGURE 5.4 – Démélange des données Moffet Field par calcul de pré-image

d'abondance évoquées ci-dessus. Les observations ont été corrompues par un bruit blanc Gaussien additif de SNR égal à 20 dB.

Le deuxième cube de données, noté IM2 et contenant 75×75 pixels mixtes, a été généré en utilisant 5 signatures spectrales de référence. Les cartes d'abondance des composés purs ont été générées de la même manière que pour l'image DC2 dans [Iordache *et al.* 2011]. La première ligne de cette figure constitue la vérité terrain des abondances de ces 5 matériaux. Des zones spatialement homogènes avec des transitions nettes peuvent être clairement observées. Sur la base de ces cartes d'abondance, un cube de données hyperspectrales a été généré avec le modèle bilinéaire rappelé ci-dessus, appliquée aux 5 signatures spectrales de référence. La scène a été corrompue par un bruit blanc gaussien additif avec un SNR de 20 dB.

Les algorithmes, avec et sans régularisation spatiale, ont été testés pour comparer leurs performances et démontrer l'intérêt d'ajouter ce type d'information. En guise de référence, la méthode FCLS a été considérée dans ces expériences. Les paramètres de réglage des algorithmes ont été fixés à partir d'expériences préliminaires sur des données indépendantes, via une simple recherche sur les grilles définies ci-après.

1. Méthode linéaire FCLS [Heinz & Chang 2001] : les paramètres de régularisation λ de l'ensemble $\{0.0001, 0.001, 0.01, 0.1\}$ ont été testés.
2. L'algorithme de pré-image sans la régularisation spatiale ℓ_1 : Le noyau partiellement linéaire avec $\gamma = 0,1$ a été retenu, avec un noyau Gaussien de largeur de bande $\sigma = 4$. Le nombre de données d'apprentissage a été arbitrairement fixé à $T = 200$ et le paramètre de régularisation η à 10^{-3}.
3. L'algorithme de pré-image avec la régularisation spatiale ℓ_1 : les mêmes para-

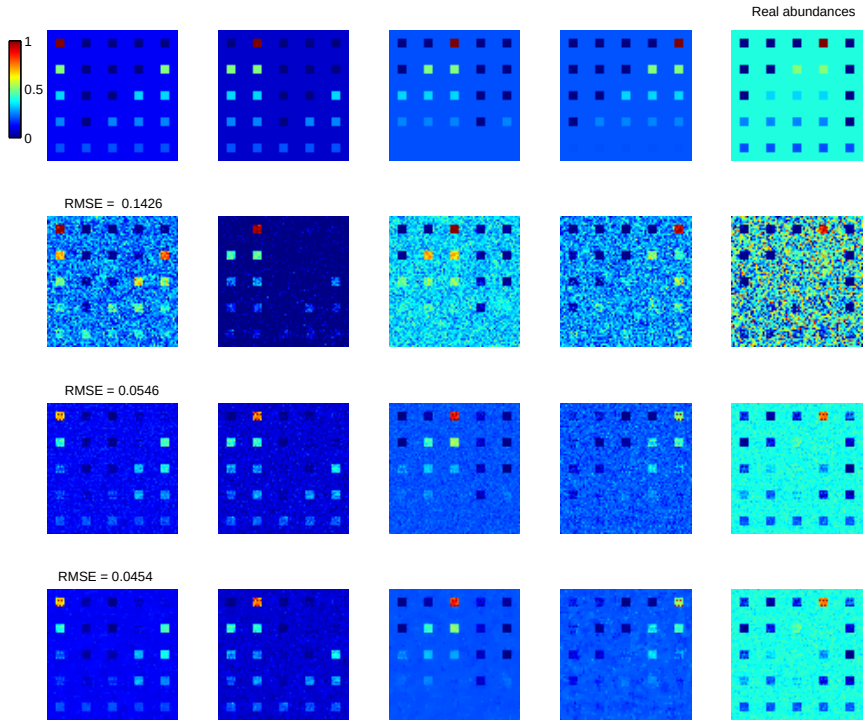


FIGURE 5.5 – Démélange non linéaire avec régularisation spatiale pour IM1. De haut en bas : cartes d'abondance originales, FCLS, méthode de pré-image sans régularisation spatiale, méthode de pré-image avec régularisation spatiale.

mètres que ci-dessus ont été choisis. Concernant les paramètres de régularisation spatiale ζ et ν, ils ont été fixés comme indiqué ci-après.

Avec l'image IM1, les tests ont conduit à $\lambda = 0,01$ pour FCLS, et $\zeta = 1$, $\nu = 0,1$ pour l'algorithme proposé. Avec l'image IM2, ces tests préliminaires ont par ailleurs mené à $\lambda = 0,01$ pour FCLS, $\zeta = 20$, $\nu = 0,5$ pour l'algorithme proposé.

Les abondances estimées sont présentées dans les figures 5.5-5.6. L'erreur de reconstruction est présenté dans le tableau 5.3. Pour les deux images, on peut constater que l'algorithme FCLS n'a pas mené à des performances satisfaisantes. On note une amélioration des performances pour la méthode de pré-image, même si les erreurs d'estimation masquent partiellement les structures spatiales. Enfin, la méthode spatialement régularisée présente une erreur de reconstruction plus faible et de carte d'abondance plus claire. L'utilisation de l'information spatiale apporte évidemment des avantages pour le processus de démélange non linéaire.

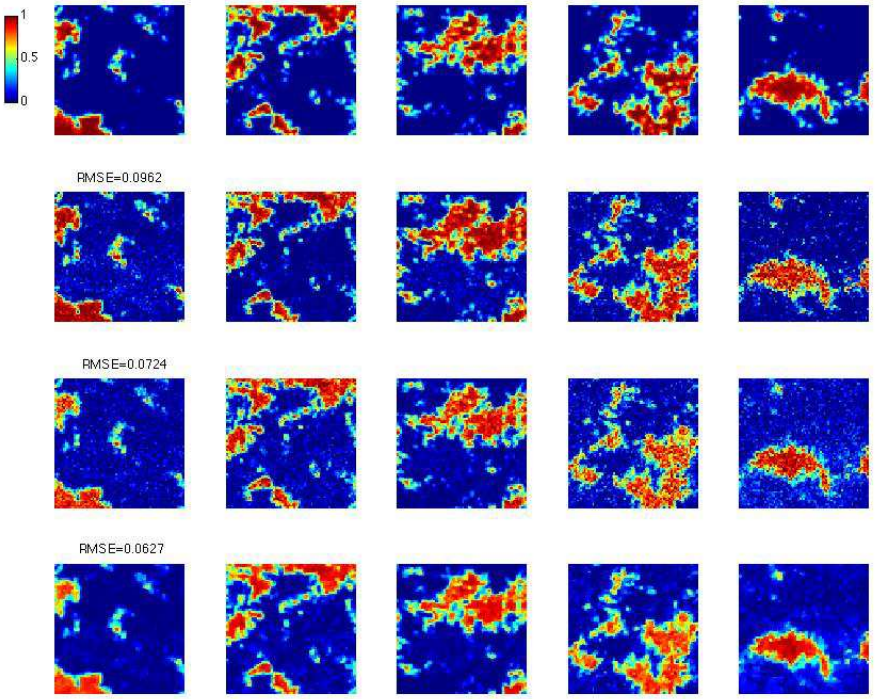


FIGURE 5.6 – Démélange non linéaire avec régularisation spatiale pour IM2. De haut en bas : cartes d'abondance originales, FCLS, méthode de pré-image sans régularisation spatiale, méthode de pré-image avec régularisation spatiale.

TABLE 5.3 – Comparaison des RMSE des algorithmes pour IM1 et IM2

Algorithmes	IM1	IM2
FCLS	0.1426	0.0962
Pré-image sans rég. spatiale	0.0546	0.0724
Pré-image avec rég. spatiale	0.0454	0.0627

5.7 Conclusion

Dans ce chapitre, nous avons présenté un algorithme de démélange hyperspectral à l'aide du principe de pré-image. Avec un nombre de données d'apprentissage suffisant, cet algorithme a montré qu'il est apte à estimer la fonction de mélange non-linéaire des matériaux. Nous avons également introduit dans cette méthode une régularisation efficace basée sur l'information spatiale. Cette modification nous a permis d'améliorer considérablement la qualité des images reconstruites.

Conclusion

Dans cette thèse, nous avons présenté les aspects de la technologie d'imagerie hyperspectrale en concentrant sur le défi de démixage non-linéaire. Pour cette tâche, nous avons proposé trois solutions. La première consiste à intégrer les avantages de l'apprentissage de variétés dans les méthodes de démixage classique pour concevoir leurs versions non-linéaires. Les résultats avec les données générées sur une variété bien connue - le « Swissroll »- donne des résultats prometteurs. Les méthodes fonctionnent beaucoup mieux avec l'augmentation de la non-linéarité. Cependant, l'absence de contrainte de non-négativité dans ces méthodes reste une question ouverte pour des améliorations à trouver. La deuxième proposition vise à utiliser la méthode de pré-image avec la régression de la matrice de noyau pour estimer une transformation inverse de l'espace de données entrées des pixels vers l'espace des abondances. L'ajout des informations spatiales sous forme « variation totale » est également introduit pour rendre l'algorithme plus robuste au bruit. Néanmoins, le problème d'obtention des données de réalité terrain nécessaires pour l'étape d'apprentissage limite l'application de ce type d'algorithmes.

Bibliographie

[Abrahamsen & Hansen 2009] T.J. Abrahamsen et L.K. Hansen. *Input space regularization stabilizes pre-images for kernel PCA de-noising.* Machine Learning for Signal Processing, 2009. MLSP 2009. IEEE International Workshop on, vol. 1, pages 1–4, 2009. (Cité en page 84.)

[Aizeman *et al.* 1964] M. Aizeman, E. Braverman et L. Rozonoer. *Theoretical foundations of the potential function method in pattern recognition learning.* In Automation and Remote Control, 1964. (Cité en page 50.)

[Altmann *et al.* 2011a] Y. Altmann, N. Dobigeon, S. McLaughlin et J.-Y. Tourneret. *Nonlinear unmixing of hyperspectral images using radial basis functions and orthogonal least squares.* In Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS), Vancouver, Canada, July 2011. (Cité en page 96.)

[Altmann *et al.* 2011b] Y. Altmann, A. Halimi, N. Dobigeon et J.-Y. Tourneret. *A polynomial post nonlinear model for hyperspectral image unmixing.* In Proc. IEEE IGARSS, Vancouver, Canada, July 2011. (Cité en pages 23, 26 et 86.)

[Altmann *et al.* 2012] Y. Altmann, A. Halimi, N. Dobigeon et J.-Y. Tourneret. *Supervised Nonlinear Spectral Unmixing Using a Postnonlinear Mixing Model for Hyperspectral Imagery.* Image Processing, IEEE Transactions on, vol. 21, no. 6, pages 3017–3025, 2012. (Cité en page 33.)

[Arias *et al.* 2007] P. Arias, G. Randall et G. Sapiro. *Connecting the out-of-sample and pre-image problems in kernel methods.* In Proc. IEEE CVPR, 2007. (Cité en page 82.)

[Aronszajn 1950] N. Aronszajn. *Theory of reproducing kernels.* Transactions of the American Mathematical Society, vol. 68, 1950. (Cité en pages 14, 41, 44, 46 et 87.)

[Bacon & et Al 2012] R. Bacon et M. Accardo et Al. *News of the MUSE.* Rapport technique, Universitté de Lyon 1, 2012. (Cité en page 9.)

[Benediktsson *et al.* 1990] J. A. Benediktsson, P. H. Swain et O.K. Ersoy. *Neural network approaches versus statistical methods in classification of multisource remote sensing data.* IEEE Transactions on Geoscience and Remote Sensing, vol. 28, pages 540–552, 1990. (Cité en page 17.)

[Benediktsson *et al.* 2003] J. A. Benediktsson, M. Pesaresi et K. Amason. *Classification and feature extraction for remote sensing images from urban areas based on morphological transformations.* IEEE Transactions on Geoscience and Remote Sensing, vol. 41, pages 1940–1949, 2003. (Cité en page 18.)

[Bengio *et al.* 2003] Y. Bengio, J.-F. Paiement, P. Vincent, O. Delalleau, N. Le Roux et M. Ouimet. *Out-of-sample extensions for LLE, Isomap, MDS, Eigenmaps, and Spectral Clustering.* In Proc. NIPS, 2003. (Cité en pages 14 et 82.)

[Berman *et al.* 2004] M. Berman, H. Kiiveri, R. Lagerstrom, A. Ernst, R. Dunne et J. F. Huntington. *ICE : a statistical approach to identifying endmembers in hyperspectral images.* IEEE Trans. Geosci. Remote Sensing, vol. 42, pages 2085–2095, 2004. (Cité en pages 25 et 38.)

[Bernstein *et al.* 2000] M. Bernstein, V. De Silva, J. C. Langford et J. B. Tenenbaum. *Graph Approximations to Geodesics on Embedded Manifolds*, 2000. (Cité en page 65.)

[Bin *et al.* 2012] L. Bin, J. Chanussot, S. Doute et Z. Liangpei. *Empirical automatic estimation of the number of endmembers in hyperspectral images.* IEEE GRSL, 2012. (Cité en page 23.)

[Bioucas-Dias & Nascimento 2008] J.M. Bioucas-Dias et J.M.P. Nascimento. *Hyperspectral subspace identification.* IEEE Transactions Geoscience and Remote Sensing, vol. 46, pages 2435–2445, 2008. (Cité en pages 23 et 34.)

[Bioucas-Dias *et al.* 2012] J. M. Bioucas-Dias, M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader et J. Chanussot. *Hyperspectral Unmixing Overview : Geometrical, Statistical, and Sparse Regression-Based Approaches*, 2012. (Cité en page 7.)

[Bkir *et al.* 2004] G. Bkir, J. Weston et B. Schölkopf. Learning to find pre-images, volume 16, pages 449–456. MIT Press, Cambridge, MA, 2004. (Cité en page 2.)

[Boardman 1993] J.W. Boardman. *Atomatic spectral unmixing of AVIRIS data using convex geometry concepts.* In Proc. AVIRIS workshop, volume 1, pages 11–14, 1993. (Cité en page 1.)

[Bochner 1955] S. Bochner. Harmonic analysis and the theory of probability. University of California Press, 1955. (Cité en page 55.)

[Boser *et al.* 1992] B. Boser, I. Guyon et V. Vapnik. *A training algorithm for optimal margin classifers.* In Proc. of the 5th Annual Workshop on Computational Learning Theory, pages 144–152, 1992. (Cité en pages 50 et 54.)

[Bruzzone *et al.* 2006] L. Bruzzone, M. Chi et M. Marconcini. *A novel transductive SVM for semisupervised classification of remote-sensing images.* IEEE Transactions on geoscience and remote sensing, vol. 44, pages 3363–3373, 2006. (Cité en page 18.)

[Camps-Valls & Bruzzone 2005] G. Camps-Valls et L. Bruzzone. *Kernel-based methods for hyperspectral image classification.* IEEE Transactions on Geoscience and Remote Sensing, vol. 43, no. 6, pages 1351–1362, 2005. (Cité en page 17.)

[Chang & Du 1999] C.-I. Chang et Q. Du. *Interference and Noise-Adjusted Principal Components Analysis.* IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING Transactions on geoscience and remote sensing, vol. 37, pages 2387–2396, 1999. (Cité en page 12.)

[Chang *et al.* 2006] C.-I. Chang, C.C. Wu, W. Liu et Y.C. Ouyang. *A new growing method for simplex-based endmember extraction algorithm.* IEEE Transac-

tions on Geoscience and Remote Sensing, vol. 44, no. 10, pages 2804–2819, 2006. (Cité en pages 23, 35 et 36.)

[Chang *et al.* 2010] C.-I. Chang, W. Xiong, W. Liu, C.C. Wu et C.C.C. Chen. *Linear spectral mixture analysis-based approaches to estimation of virtual dimensionality in hyperspectral imagery , vol. 48, no. 11, pp. 3960-3979, Nov. 2010.* IEEE Transactions on Geoscience and Remote Sensing, vol. 48, pages 3960–3979, 2010. (Cité en pages 23 et 34.)

[Chen *et al.* 2011] J. Chen, C. Richard, J.-C. M. Bermudez et P. Honeine. *Nonnegative Least-Mean-Square Algorithm.* IEEE Transactions on Signal Processing, vol. 59, no. 11, pages 5225 –5235, nov. 2011. (Cité en page 90.)

[Chen *et al.* 2012] J. Chen, C. Richard et P. Honeine. *Nonlinear unmixing of hyperspectral data based on a linear-mixture/nonlinear-fluctuation model.* IEEE Transactions on signal processing, 2012. (Cité en pages 26, 90, 91 et 92.)

[Chen *et al.* 2013] J. Chen, C. Richard et P. Honeine. *Nonlinear Unmixing of Hyperspectral Data Based on a Linear-Mixture/Nonlinear-Fluctuation Model.* IEEE Transactions on Signal Processing, vol. 61, no. 2, pages 480–492, Jan 2013. (Cité en page 2.)

[Chena *et al.* 2012] G. Chena, G. J. Haya et B. St-Ongeb. *A GEOBIA framework to estimate forest parameters from lidar transects, Quickbird imagery and machine learning : A case study in Quebec, Canada.* International Journal of Applied Earth Observation and Geoinformation, vol. 15, pages 28–37, 2012. (Cité en page 16.)

[Ciznicki *et al.* 2012] M. Ciznicki, K. Kurowski et A. Plaza. *Graphics processing unit implementation of JPEG2000 for hyperspectral image compression.* Journal of Applied Remote Sensing, vol. 6, no. 1, pages 061507–1–061507–14, 2012. (Cité en page 20.)

[Close *et al.* 2012] R. Close, P. Gader, J. Wilson et A. Zare. *Using physics-based macroscopic and microscopic mixture models for hyperspectral pixel unmixing.* Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVIII, vol. 8390, pages 83901L–83901L–13, 2012. (Cité en page 33.)

[Courant & Hilbert 1962] R. Courant et D. Hilbert. Methods of mathematical physics. Interscience, 1962. (Cité en pages 47 et 48.)

[Cox & Cox 1994] T.F. Cox et M.A.A. Cox. Multidimensional Scaling. Chapman and Hall, London, 1994. (Cité en pages 14 et 63.)

[Cristianini 2001] N. Cristianini. *Support vector and kernel machines.* Rapport technique, Tutorial at ICML, 2001. (Cité en page 52.)

[Cucker & Smale 2002] F. Cucker et S. Smale. *On the mathematical foundations of learning.* Bulletin of the American Mathematical Society, vol. 39, pages 1–49, 2002. (Cité en pages 44, 55 et 91.)

[de Silva & Tenenbaum 2003] V. de Silva et J. B. Tenenbaum. A global geometric framework for nonlinear dimensionality reduction, pages 705–712. MIT Press, 2003. (Cité en pages 2, 13, 14, 26, 63, 82 et 91.)

[Dijkstra 1959] E.W. Dijkstra. *A note on two problems in connexion with graphs.* Numerische Mathematik, pages 269–271, 1959. (Cité en page 64.)

[Dobigeon *et al.* 2009] N. Dobigeon, S. Moussaoui, M. Coulon, J.-Y. Tourneret et A.O. Hero. *Joint Bayesian endmember extraction and linear unmixing for hyperspectral imagery.* IEEE Transactions on Signal Processing, vol. 57, no. 11, pages 4355–4368, 2009. (Cité en pages 1 et 25.)

[Dopido *et al.* 2011] I. Dopido, M. Zortea, A. Villa, A. Plaza et P. Gamba. *Unmixing prior to supervised classification of hyperspectral images.* IEEE Geoscience and Remote Sensing Letters, vol. 8, pages 760–764, 2011. (Cité en page 18.)

[Dopido *et al.* 2012] I. Dopido, J. Li, A. Plaza et P. Gamba. *Integration of classification and spectral unmixing.* IEEE Tyhrrenian Workshop on Remote Sensing, 2012. (Cité en page 18.)

[Du & Plaza 2012] Q. Du et A. Plaza. Recent advances in hyperspectral data analysis. Tutorial on hyperspectral imaging IGARSS2012, Munich, 2012. (Cité en page 30.)

[Duda *et al.* 2000] R.O. Duda, P.E. Hart et D.H. Stork. Pattern classification (2nd ed.). Wiley Interscience, 2000. (Cité en page 13.)

[Eckstein & Bertsekas 1992] J. Eckstein et D.P. Bertsekas. *On the Douglas—Rachford splitting method and the proximal point algorithm for maximal monotone operators.* Mathematical Programming, vol. 55, pages 293–318, 1992. (Cité en page 95.)

[Elavarasi *et al.* 2011] S. Anitha Elavarasi, J. Akilandeswari et B. Sathiyabhama. *A survey on partition clustering algorithms.* International Journal of Enterprise Computing and Business Systems, vol. 1, 2011. (Cité en page 16.)

[Etyngier *et al.* 2007] P. Etyngier, F. Ségonne et R. Keriven. *Shape priors using Manifold Learning Techniques.* In in &quot ;11th IEEE International Conference on Computer Vision, Rio de Janeiro, 2007. (Cité en page 85.)

[Fan *et al.* 2009] W. Fan, B. Hu, J. Miller et M. Li. *Comparative study between a new nonlinear model and common linear model for analysing laboratory simulated-forest hyperspectral data.* International Journal of Remote Sensing, vol. 30, no. 11, pages 2951–2962, 2009. (Cité en page 31.)

[Floyd 1962] R.W. Floyd. *Algorithm 97 : Shortest path.* Communications of the ACM, vol. 5, page 345, 1962. (Cité en page 64.)

[Goldstein & Osher 2009] T. Goldstein et S. Osher. *The Split Bregman Method for L1-Regularized Problems.* SIAM J. Imaging Sci., 2(2), vol. 2, pages 323–343, 2009. (Cité en page 93.)

[Green *et al.* 1988] A.A. Green, M.D. Craig et S. Cheng. *The Application Of The Minimum Noise Fraction Transform To The Compression And Cleaning Of*

Hyper-spectral Remote Sensing Data. Geoscience and Remote Sensing Symposium, IGARSS, vol. 3, page 1807, 1988. (Cité en page 13.)

[Grimes & Donoho 2002] C. Grimes et D. L. Donoho. *When does isomap recover the natural parametrization of families of articulated images*. Rapport technique, Stanford University, 2002. (Cité en page 71.)

[Halimi *et al.* 2011a] A. Halimi, Y. Altman, N. Dobigeon et J.-Y. Tourneret. *Nonlinear unmixing of hyperspectral images using a generalized bilinear model*. IEEE Transactions on Geoscience and Remote Sensing, vol. 49, no. 11, pages 4153–4162, 2011. (Cité en pages 1, 23, 25 et 32.)

[Halimi *et al.* 2011b] A. Halimi, Y. Altman, N. Dobigeon et J.-Y. Tourneret. *Nonlinear unmixing of hyperspectral images using a generalized bilinear model*. IEEE Transactions Geoscience and Remote Sensing, vol. 49, no. 11, pages 4153–4162, November 2011. (Cité en page 96.)

[Ham *et al.* 2003] J. Ham, D.D. Lee, S. Mika et B. Schölkopf. *A kernel view of the dimensionality reduction of manifolds*. Rapport technique TR-110, Max Planck Institut für biologische Kybernetik, 2003. (Cité en pages 69 et 91.)

[Hapke 1981] B. Hapke. *Bidirectional reflectance spectroscopy, 1, Theory*. Journal of Geophysical Research, vol. 86, no. B4, pages 3039–3054, 1981. (Cité en pages 1, 25 et 32.)

[Heinz & Chang 2001] D.C. Heinz et C.-I. Chang. *Fully constrained least squares linear mixture analysis for material quantification in hyperspectral imagery*. IEEE Transactions on Geoscience and Remote Sensing, vol. 39, no. 3, pages 529–545, 2001. (Cité en pages 1, 25, 34, 37, 90 et 99.)

[Hendrix *et al.* 2012] E.M.T. Hendrix, I. Garcia, J. Plaza, G. Martin et A. Plaza. *A new minimum volume enclosing algorithm for endmember identification and abundance estimation in hyperspectral data , vo. 50, no. 7, pp. 2744-2757, 2012*. IEEE Transactions on Geoscience and Remote Sensing, vol. 50, pages 2744–2757, 2012. (Cité en pages 25 et 38.)

[Heylen *et al.* 2011] R. Heylen, D. Burazenoviv et P. Scheunders. *Non-linear spectral unmixing by geodesic simplex volume maximization*. IEEE Journal of Signal Processing, vol. 5, no. 3, pages 534–542, June 2011. (Cité en page 72.)

[Hofmann *et al.* 2005] T . Hofmann, B. Schölkopf et A.J. Smola. *A tutorial review of RKHS methods in Machine Learning*, 2005. (Cité en page 50.)

[Honeine & Richard 2009] P. Honeine et C. Richard. *Solving the pre-image problem in kernel machines : A direct method*. In Machine Learning for Signal Processing, 2009. MLSP 2009. IEEE International Workshop on, 2009. (Cité en page 85.)

[Honeine & Richard 2011a] P. Honeine et C. Richard. *Geometric Unmixing of Large Hyperspectral Images : A Barycentric Coordinate Approach*. IEEE Transactions on Geoscience and Remote Sensing, 2011. (Cité en pages 1, 25, 26, 39, 72, 73, 76 et 87.)

[Honeine & Richard 2011b] P. Honeine et C. Richard. *Preimage problem in kernel-based machine learning.* IEEE Signal Processing Magazine, vol. 28, no. 2, pages 77–88, March 2011. (Cité en pages 81, 85 et 88.)

[Huertas 1999] A. Huertas. *Use of Hyperspectral Data with Intensity Images for Automatic Building Modeling*, 1999. (Cité en page 6.)

[Hyvarinen & Oja 2000] A. Hyvarinen et E. Oja. *Independent component analysis : algorithms and applications.* Neural Networks, vol. 13, pages 411–430, 2000. (Cité en page 25.)

[Iordache *et al.* 2011] M.D. Iordache, J.M. Bioucas-Dias et A. Plaza. *Total variation regulatization in sparse hyperspectral unmixing.* In Hyperspectral Image and Signal Processing : Evolution in Remote Sensing (WHISPERS), 2011 3rd Workshop on, pages 1– 4. Hyperspectral Image and Signal Processing : Evolution in Remote Sensing (WHISPERS), 2011 3rd Workshop on, 6-9 June 2011. (Cité en pages 27, 39, 86, 92 et 99.)

[Jia & Qian 2007] S. Jia et Y. Qian. *Spectral and Spatial Complexity-Based Hyperspectral Unmixing.* Geoscience and Remote Sensing, IEEE Transactions on, vol. 45, no. 12, pages 3867–3879, 2007. (Cité en page 32.)

[Jolliffe 1986] I.T. Jolliffe. Principal component analysis, page 487. Springer-Verlag, 1986. (Cité en page 12.)

[Jordan *et al.* 2001] M.I. Jordan, Y. LeCun et S.A. Solla. Advances in neural information processing systems. Proceedings of the First 12 Conferences, 2001. (Cité en page 66.)

[Jutten & Karhunen 2003a] C. Jutten et J. Karhunen. *Advances in nonlinear blind source separation.* In Proc. International Symposium on Independent Component Analysis and Blind Signal Separation (ICA), pages 245–256, 2003. (Cité en pages 25 et 33.)

[Jutten & Karhunen 2003b] C. Jutten et J. Karhunen. *Advances in nonlinear blind source seperation.* Proc. International Symposium on Independent Component Analysis and Blind Signal Seperation (ICA), pages 245–256, 2003. (Cité en page 96.)

[Keshava & Mustard 2002] N. Keshava et J.F. Mustard. *Spectral unmixing.* IEEE Signal Processing Magazine, vol. 19, no. 1, pages 44–57, 2002. (Cité en pages 1, 20 et 21.)

[Kwok & Tsang 2003] J.T. Kwok et I.W. Tsang. *The pre-image problem in kernel methods.* In Proc. ICML, 2003. (Cité en pages 82 et 84.)

[Lei *et al.* 2012] Y.-K. Lei, Z.-H. You, Z. Ji, L. Zhu et D.-S Huang. *Assessing and predicting protein interactions by combining manifold embedding with multiple information integration.* BMC Bioinformatics, vol. 13, no. Suppl 7, page S3, 2012. (Cité en pages 15 et 62.)

[Li & Bioucas-Dias 2008] J. Li et J. Bioucas-Dias. *Minimum volume simplex analysis : A fast algorithm to unmix hyperspectral data.* in Proceedings of the

IEEE International Geoscience and Remote Sensing Symposim (IGARSS), vol. 3, pages 250–253, 2008. (Cité en pages 25 et 38.)

[Li *et al.* 2012] J. Li, J. Bioucas-Dias et A. Plaza. *Spectral-spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields.* IEEE Transactions on Geoscience and Remote Sensing, vol. 50, pages 809–823, 2012. (Cité en page 18.)

[Ma *et al.* 2010] L. Ma, M. Crawford et J. Tian. *Local manifold learning-based k-nearest-neighbor for hyperspectral image classification.* IEEE Transactions on Geoscience and Remote Sensing, vol. 48, pages 4099–4109, 2010. (Cité en page 18.)

[Manolakis *et al.* 2003] D. Manolakis, D. Marden et G. A. Shaw. *Hyperspectral Image Processing for Automatic Target Detection Applications*, 2003. (Cité en page 8.)

[Martin & Plaza 2011] G. Martin et A. Plaza. *Region-based spatial preprocessing for end-members extraction and hyperspectral unmixing.* IEEE Transactions Geoscience and Remote Sensing letters, vol. 8, pages 745–749, 2011. (Cité en page 24.)

[Miao & Qi 2007] L. Miao et H. Qi. *Endmember extraction from highly mixed data using minimum volume constrained nonnegativematrix factorization.* IEEE Transactions Geoscience and Remote Sensing, vol. 45, pages 765–777, 2007. (Cité en pages 25 et 38.)

[Microimage 2013] Inc. Microimage. Introduction to hyperspectral imaging, 2013. (Cité en page 9.)

[Mika *et al.* 1999] S. Mika, B. Schölkopf, A. Smola, K.-R. Müller, M. Scholz et G. Rätsch. *Kernel PCA and de-noising in feature spaces.* In Proc. NIPS, 1999. (Cité en pages 13 et 82.)

[Munoz & Diego 2006] A. Munoz et I.de Diego. *From indefinite to positive semi-definite matrices.* In Dit-Yan Yeung, James Kwok, Ana Fred, Fabio Roli et Dick de Ridder, editeurs, Structural, Syntactic, and Statistical Pattern Recognition, volume 4109 of *Lecture Notes in Computer Science*, pages 764–772. Springer Berlin / Heidelberg, 2006. (Cité en page 91.)

[Mura *et al.* 2010] M. Dalla Mura, J. A. Benediktsson, B. Waske et L. Bruzzone. *Morphological attribute profiles for the analysis of very high resolution images.* IEEE Transactions on Geoscience and Remote Sensing, vol. 48, pages 3747–3762, 2010. (Cité en page 18.)

[NASA 1992] NASA. *http ://aviris.jpl.nasa.gov/*, 1992. (Cité en pages 6 et 7.)

[Nascimento & Bioucas-Dias 2005] J.M.P. Nascimento et J.M. Bioucas-Dias. *Vertex Component Analysis : A fast algorithm to unmix hyperspectral data.* IEEE Transactions on Geoscience and Remote Sensing, vol. 43, no. 4, pages 898–910, April 2005. (Cité en pages 1, 23, 36 et 37.)

[Nascimento & Bioucas-Dias 2009] J.M.P. Nascimento et J.M. Bioucas-Dias. *Nonlinear mixture model for hyperspectral unmixing.* In Proc. SPIE, volume 7477, 2009. (Cité en pages 1, 26 et 31.)

[Nascimento & Bioucas-Dias 2010] J.M.P. Nascimento et J.M. Bioucas-Dias. *Unmixing hyperspectral intimate mixtures.* In Proc. SPIE, volume 7830, 2010. (Cité en page 23.)

[Neville *et al.* 1999] R. A. Neville, K. Staenz, T. Szeredi, J. Lefebvre et P. Hauff. *Automatic endmember extraction from hyperspectral data for mineral exploration.* In Proc. 4th Int. Airborne Remote Sens. Conf. and Exhib./21st Can. Symp. Remote Sens., Ottawa, ON, Canada, pages 21 –24, 1999. (Cité en page 34.)

[Noomen 2007] M.F. Noomen. *Hyperspectral reflectance of vegetation affected by underground hydrocarbon gas.* PhD thesis, Enschede, ITC 151p, 2007. (Cité en page 9.)

[Ovari 2000] Z. Ovari. *Kernels, eigenvalues and Support Vector Machines.* PhD thesis, Australian National University , Canberra, 2000. (Cité en page 53.)

[Parente & Plaza 2010] M. Parente et A. Plaza. *Survey of geometric and statistical unmixing algorithms for hyperspectral images.* In Hyperspectral Image and Signal Processing : Evolution in Remote Sensing (WHISPERS), 2010 2nd Workshop on, pages 1–4, 2010. (Cité en page 22.)

[Plaza & Chang 2005] A. Plaza et C.-I Chang. *An improved N-FINDR algorithm in implementation.* Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XI, Proceedings of the SPIE,, vol. 5806, pages 298–306, 2005. (Cité en page 35.)

[Plaza & Plaza 2010] J. Plaza et A. Plaza. *Spectral mixture analysis of hyperspectral scenes using intelligently selected training samples.* IEEE Geoscience and Remote Sensing Letters, vol. 7, no. 2, pages 371–375, Arilp 2010. (Cité en page 26.)

[Plaza *et al.* 2002] A. Plaza, P. Martinez, R. Perez et J. Plaza. *Spatial/spectral endmember extraction by multidimension and morphological operations.* IEEE Transactions Geoscience and Remote Sensing, vol. 40, pages 2025–2041, 2002. (Cité en page 24.)

[Plaza *et al.* 2005] A. Plaza, P. Martinez, J. Plaza et R.M. Perez. *Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations.* IEEE Transactions on Geoscience and Remote Sensing, vol. 43, pages 466–479, 2005. (Cité en page 18.)

[Poggio 1975] T . Poggio. *On optimal nonlinear associative recall.* In Biological Cybernetics, vol. 19, pages 201–209, 1975. (Cité en pages 43 et 54.)

[Pothin 2007] J.-B Pothin. *Décision par méthodes a noyaux en traitement du signal : Techniques de selection et d' élaboration de noyaux adaptés.* PhD thesis, Université de Technologie de Troyes, 2007. (Cité en pages 41 et 56.)

[Raksuntorn & Du 2010] N. Raksuntorn et Q. Du. *Nonlinear spectral mixture analysis for hyperspectral imagery in an unknown environment.* IEEE Geoscience and Remote Sensing Letters, vol. 7, no. 4, pages 836–840, 2010. (Cité en pages 1 et 26.)

[Reed & Simon 1980] M. Reed et B. Simon. Method of modern mathematical physics. vol. 1 : Functional analysis. Academic Press, San Diego, 1980. (Cité en page 46.)

[Rogge *et al.* 2007] D.M. Rogge, B. Rivard, J. Zhang, A. Sanchez, J. Harris et J. Feng. *Integration of spatial–spectral information for the improved extraction of endmembers.* Remote Sensing of Environment, vol. 110, pages 287–303, 2007. (Cité en page 24.)

[Roweis & Saul 2000] S. Roweis et L. Saul. *Nonlinear dimensionality reduction by locally linear embedding.* Science, pages 2323–2326, 2000. (Cité en pages 2, 14, 26, 63, 67, 68, 82 et 91.)

[Sanchez *et al.* 2010] S. Sanchez, G. Martin et A. Plaza. *Parallel implementation of the N-FINDR endmember extraction algorithm on commodity graphics processing units.* Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International, pages 955–958, 2010. (Cité en page 35.)

[Savary & et al. 2010] P. Tremblayand S. Savary et M. Rolland et al. *Standoff gas identification and quantification from turbulent stack plumes with an imaging Fourier-transform spectrometer.* In Proceedings of SPIE Vol. 7673, 76730H, 2010. (Cité en page 10.)

[Schoenberg 1942] I. J. Schoenberg. Positive definite functions on spheres. Duke Math, 1942. (Cité en page 53.)

[Schölkopf & Smola 2001] B. Schölkopf et A.J. Smola. Learning with Kernels. MIT Press, Cambridge, MA, December 2001. (Cité en page 71.)

[Schölkopf *et al.* 1999] B. Schölkopf, J. Platt, J. Shawe Taylor, A. J. Smola et R. C. Williamson. *Estimating the support of a high-dimensional distribution.* Rapport technique, Microsoft Research, 1999. (Cité en page 53.)

[Schölkopf *et al.* 2000] B. Schölkopf, R. Herbrich et R. Williamson. *A generalized representer theorem.* Rapport technique NC2-TR-2000-81, NeuroCOLT, Royal Holloway College, University of London, UK, 2000. (Cité en page 88.)

[Sha & Saul 2005] F. Sha et L.K. Saul. *Analysis and extension of spectral methods for nonlinear dimensionality reduction.* Proceeding ICML '05 Proceedings of the 22nd international conference on Machine learning, pages 784–791, 2005. (Cité en page 68.)

[Shaw & Manolakis 2002] G. Shaw et D. Manolakis. *Signal processing for hyperspectral image exploitation.* Signal Processing Magazine, IEEE, vol. 19, no. 1, pages 12–16, 2002. (Cité en page 11.)

[Smola & Schölkopf 2003] Alex J. Smola et Bernhard Schölkopf. *A Tutorial on Support Vector Regression.* Rapport technique, NeuroCOLT, 2003. (Cité en page 57.)

[Somers *et al.* 2009] B. Somers, C. Kenneth, D. Stephanie, J. Stuckens, D. Van der Zande, W.W. Verstraeten et P. Coppin. *Nonlinear Hyperspectral Mixture Analysis for tree cover estimates in orchards.* Remote Sensing of Environment, vol. 113, no. 15, pages 1183–1193, June 2009. (Cité en page 31.)

[Steinwart *et al.* 2004] I. Steinwart, D. Hush et C. Scovel. *An explicit description of the reproducing kernel Hilbert spaces of gaussian RBF kernels.* Rapport technique, Los Alamos National Laboratory, 2004. (Cité en page 56.)

[Sugiyama 2006] M. Sugiyama. *Local Fisher Discriminant Analysis for Supervised Dimensionality Reduction.* ICML '06 Proceedings of the 23rd international conference on Machine learning, pages 905–912, 2006. (Cité en page 13.)

[Suykens *et al.* 2002] J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor et J. Vandewalle. Least squares support vector machines. World Scientific, Singapore, 2002. (Cité en page 17.)

[Tarabalka *et al.* 2010a] Y. Tarabalka, J. A. Benediktsson, J. Chanussot et J. C. Tilton. *Multiple spectral-spatial classification approach for hyperspectral data.* IEEE Transactions on Geoscience and Remote Sensing, vol. 48, pages 4122–4132, 2010. (Cité en page 18.)

[Tarabalka *et al.* 2010b] Y. Tarabalka, J. Chanussot et J. A. Benediktsson. *Segmentation and classification of hyperspectral images using watershed transformation.* Pattern Recognition, vol. 43, pages 2367–2379, 2010. (Cité en page 18.)

[Tarabalka *et al.* 2010c] Y. Tarabalka, M. Fauvel, J. Chanussot et J. A. Benediktsson. *SVM and MRF-based method for accurate classification of hyperspectral images.* IEEE Geoscience and Remote Sensing Letters, vol. 7, pages 736–740, 2010. (Cité en page 18.)

[Tarabalka *et al.* 2012] Y. Tarabalka, J. C. Tilton, J. A. Benediktsson et J. Chanussot. *A marker-based approach for the automated selection of a single segmentation from a hierarchical set of image segmentations.* IEEE JSTARS, vol. 5, pages 262–272, 2012. (Cité en page 16.)

[Taylor & Cristianini 2004] J. S. Taylor et N. Cristianini. Kernel methods for pattern analysis. Cambridge University Press, 2004. (Cité en pages 54 et 55.)

[Themelis *et al.* 2010] K. Themelis, A.A. Rontogiannis et K. Khoutroumbas. *Semi-supervised Hyperspectral Unmixing via the weighted Lasso.* In in Proc. ICASSP, pages 1194–1197, 2010. (Cité en pages 26 et 86.)

[Theys *et al.* 2009] C. Theys, N. Dobigeon, J.-Y. Tourneret et H. Lanteri. *Linear unmixing of hyperspectral images using a scaled gradient method.* Statistical Signal Processing 2009. SSP '09. IEEE/SP 15th Workshop, pages 729–713, Sept. 2009. (Cité en pages 1 et 25.)

[Tourneret *et al.* 2008] J.-Y. Tourneret, N. Dobigeon et C.-I Chang. *Semi-supervised linear spectral unmixing using a hierarchical Bayesian model for hyperspectral imagery.* IEEE Transactions on Signal Processing, vol. 5, no. 7, pages 2684–2695, 2008. (Cité en pages 26 et 86.)

[Vapnik 1995] V.N. Vapnik. The nature of statistical learning theory. Springer, New York, NY, 1995. (Cité en page 54.)

[Werff 2006] H. Werff. *Knowledge based remote sensing of complex objects : recognition of spectral and spatial patterns resulting from natural hydrocarbon.* PhD thesis, Utrecht University, ITC, 2006. (Cité en page 9.)

[Williams 2002] C.K.I. Williams. *On a Connection between Kernel PCA and Metric Multidimensional Scaling.* Machine Learning, pages 11–19, 2002. (Cité en page 71.)

[Winter 1999] M.E. Winter. *N-FINDR : an algorithm for fast autonomous spectral end-member determination in hyperspectral data.* In Proc. SPIE Spectrometry V, volume 3753, pages 266–277, 1999. (Cité en pages 1, 23 et 34.)

[Xiong *et al.* 2011] W. Xiong, C.-I Chang, C.-C Wu, K. Kalpakis et H.-M Chen. *Fast Algorithms to Implement N-FINDR for Hyperspectral Endmember Extraction.* Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of, vol. 4, pages 545–564, 2011. (Cité en page 35.)