Approche inverse pour la reconstruction des environnements circumstellaires en polarimétrie avec l'instrument d'imagerie directe ESO/VLT SPHERE IRDIS.

Laurence Denneulin

N° d'ordre NNT :xxx

THÈSE de DOCTORAT DE L'UNIVERSITE DE LYON
opérée au sein de
l'Université Claude Bernard Lyon 1

École Doctorale ED52
Physique et Astrophysique (PHAST)

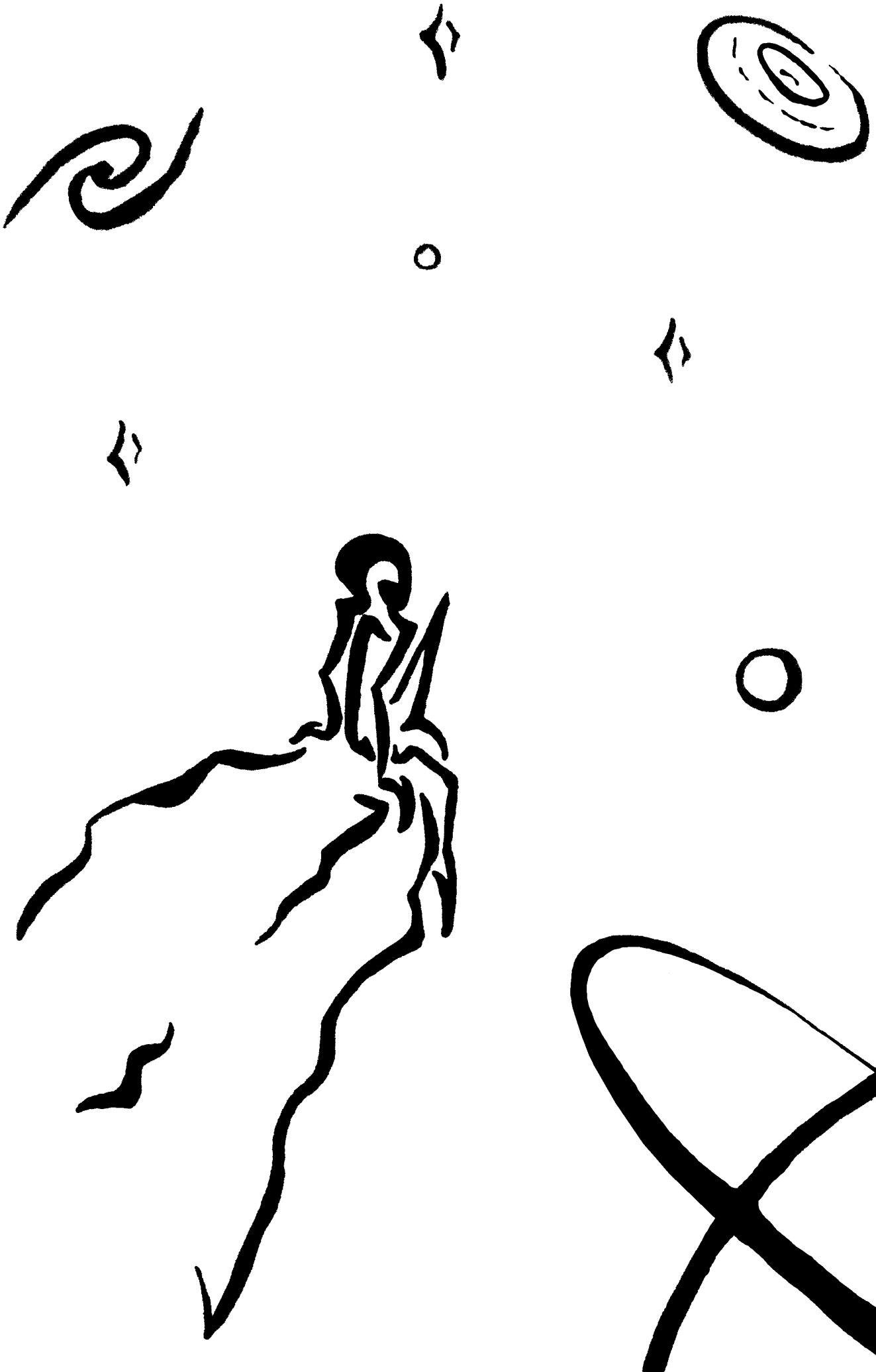
Spécialité de doctorat : Astrophysique

Soutenue publiquement le 15/10/2020, par :
Laurence Denneulin

Approche inverse pour la reconstruction des environnements circumstellaires en polarimétrie avec l'instrument d'imagerie directe ESO/VLT SPHERE IRDIS.

Devant le jury composé de :

M. Masnou Simon :	Président
M. Schmid Hans Martin :	Rapporteur
M. Talbot Hugues :	Rapporteur
Mme Benisty Myriam :	Examinatrice
Mme Repetti Audrey :	Examinatrice
Mme Maud Langlois :	Directrice de thèse
Mme Nelly Pustelnik :	Co-encadrante de thèse
M. Éric Thiébaut :	Co-encadrant de thèse



Résumé

L'étude des environnements circumstellaires nous permet d'en apprendre plus sur la formation des exoplanètes. Malgré les avancées instrumentales, permettant une plus grande résolution des environnements, leur observation reste difficile du fait du grand contraste entre les environnements et leurs étoiles hôtes. En effet, celles-ci sont 1000 à 10 000 fois plus brillantes, voir 10 000 000 fois plus brillantes dans le cas des exoplanètes.

Lors de l'acquisition directe d'image de ces environnements, le signal de l'environnement est mélangé au résidu de lumière stellaire. Or, la lumière de l'environnement est partiellement linéairement polarisée, tandis que le résidu de lumière stellaire ne l'est pas. Le sous-instrument *Infrared Dual-band Imaging and Spectroscopy* (IRDIS) de l'instrument de l'*European Southern Observatory (ESO) appelé Spectro-Polarimeter High-contrast Expolanet REsearch* (SPHERE), situé sur l'un des quatre *Very Large Telescopes* (VLT) dans le désert d'Atacama au Chili, acquiert des jeux de données où la polarisation linéaire est modulée par rotation, selon plusieurs cycles d'angles connus. Ainsi, par combinaison de ces données multivariées, il est possible de démélanger la lumière diffusée par l'environnement circumstellaire et la lumière de l'étoile.

Lors du démélange par les méthodes de l'état-de-l'art, les données sont combinées sans prendre en compte la précision des mesures, c'est-à-dire la statistique du bruit de photon, qui domine le signal d'intérêt, et du bruit de lecture du détecteur, ainsi que les données manquantes. De plus, si les images d'un cycle de rotation d'angle contiennent des images non-exploitables, à cause de la turbulence atmosphérique par exemple, les images du cycle sont toutes supprimées. Aussi, tout recentrage, toute rotation ou toute déconvolution des données est fait indépendamment du démélange. De ce fait, la propagation des erreurs n'est pas contrôlée.

Les méthodes de types « problèmes inverses » permettent, à partir d'un modèle direct des données, de procéder au démélange tout en ayant le contrôle sur la propagation des erreurs. Une telle approche peut être retrouvée pour d'autres types d'observations en astrophysique, mais n'a jamais été développée pour l'imagerie directe en polarimétrie.

Le but de ma thèse, est de reconstruire, sous un certain critère d'optimalité, des cartes de la lumière polarisée des environnements circumstellaires, une carte des angles de polarisation associés et une carte des résidus lumineux de l'étoile et de la lumière non-polarisée de l'environnement. Pour y parvenir, dans cette thèse, je propose un modèle physique des données non-linéaire, séparable en chaque pixel, basé sur le formalisme de Jones et paramétré en l'intensité non-polarisée, l'intensité polarisée linéairement et l'angle de polarisation linéaire. Je dérive ensuite une formulation alternative linéaire de ce modèle, paramétrée en les paramètres de Stokes, permettant d'expliciter le lien entre un tel modèle et les méthodes de l'état-de-l'art. J'étends par la suite ce modèle au cas non-séparable, sans et avec convolution par la réponse impulsionnelle spatiale (PSF : *Point Spread Function*), et dérive, dans le cas sans convolution, un nouveau modèle non-linéaire, paramétré en l'intensité non-polarisée et les paramètres de

Stokes d'intensité polarisée horizontale et verticale.

Pour chacun de ces modèles, je propose différentes méthodes d'estimation des paramètres, basées sur la minimisation de critères sous contraintes et, dans le cas non-séparable, régularisés. Parmi les régularisations utilisées, je compare notamment les pénalisations différentiables et non-différentiables sur le gradient des pixels ou sur les valeurs propres du hessien des pixels. Dans le cas linéaire, une contrainte épigraphique reliant les paramètres de Stokes est proposée. Dans le cas non-linéaire, on impose une contrainte de positivité sur les intensités. Afin de régler le poids des régularisations j'utilise l'estimateur non-biaisé du risque de Stein (SURE : *Stein Unbiased Risk Estimator*).

L'ensemble de ces méthodes d'estimation sont mises en œuvre sur des données synthétiques qui reproduisent les données types rencontrées en imagerie directe en polarimétrie et sur des données réelles. Selon les propriétés des fonctions considérées dans le critère, j'utilise, pour la minimisation, différents algorithmes. Dans le cas séparable, je procède à une inversion directe par annulation du gradient. Dans le cas non-linéaire, les paramètres sont estimés de manière hiérarchique. Dans le cas différentiable j'utilise une méthode de descente de gradient avec préconditionnement à mémoire limitée de Broyden-Fletcher-Goldfarb-Shanno (BFGS), qui peut être également adaptée dans le cas de la contrainte de positivité sur les intensités. Dans le cas linéaire, avec régularisation différentiable et contrainte épigraphique, j'utilise l'algorithme Forward-Backward avec *backtracking*, afin d'éviter le calcul de la constante de Lipschitz du gradient de la partie différentiable du critère qui peut être difficile. Dans le cas des régularisations non-différentiables, j'utilise l'algorithme primal-dual de Condat-Vũ intégrant une étape de *backtracking*.

Je montre alors que dans le cas séparable, la prise en compte des données manquantes permet d'utiliser les cartes des cycles incomplets, réduisant ainsi l'erreur sur l'estimation des cartes. Je montre ensuite, dans le cas non-séparable, que la prise en compte dans le modèle des transformations du détecteur (rotations, translations) et des mauvais pixels réduit à coup sûr l'erreur faite sur les intensités non-polarisées et polarisées estimées, dans le cas d'une régularisation indépendante sur les intensités. Ce n'est cependant pas le cas pour l'angle dont l'erreur est mieux réduite par une régularisation sur les paramètres de Stokes d'intensité polarisée. Je montre également que sans déconvolution, dans le cas de disques de faible intensité, le réglage automatique des poids de régularisation par SURE a tendance à sur-régulariser l'estimation et qu'un choix manuel est alors plus pertinent.

Dans le cas de la déconvolution, je montre qu'une régularisation différentiable par *a priori* de lissage avec préservation des bords telle que l'approximation hyperbolique de la Variation Totale (TV-h) donne, en le temps imparti, de meilleurs résultats que la Variation Totale (TV) et qu'une pénalisation par la norme de Shatten du hessien encore trop coûteuse en temps de calculs sur de gros volumes de données. Je montre également qu'inclure la convolution par la PSF dans le modèle réduit fortement l'erreur d'estimation par rapport à une déconvolution *a posteriori*.

Enfin je montre sur différents jeux de données astrophysiques l'apport des meilleures méthodes présentées dans cette thèse au niveau de la reconstruction des structures et de l'intensité retrouvée. Je conclue que la prise en compte du modèle complet incluant la convolution, pour la reconstruction des environnements circumstellaires à différents niveaux de régularisation, permettait la résolution plus fine des structures polarisées brillantes et une meilleure détection des structures de faible intensité polarisée.

Abstract

Survey of circumstellar environments allows a better understanding of exoplanets formation. Despite instrumentation enhancement, allowing for a bigger resolution of these environments, their observation remains difficult due to high contrast between the environments and their host stars. In fact the host stars are 1000 to 10 000 times brighter than the environment, even 10 000 000 times brighter for exoplanets.

When images of these circumstellars environnements are acquired in direct imaging, the signal of the environnement is mixed to star light residuals. Yet, the light of the environment is partially linearly polarized while the light of the star is unpolarized. The instrument Infrared Dual-band Imaging and Spectroscopy (IRDIS) of the European Southern Observatory's (ESO) Spectro-Polarimeter High-contrast Expolanet REsearch (SPHERE) instrument, installed at one of the four Very Large Telescopes (VLT) in Atacama in Chile, acquires datasets where the polarization is modulated according to a known angle cycle. Then, combining these multivariate data, it is possible to extract the light scattered by the environment from the light of the stars.

When data are combined with the state-of-the-art methods, neither the photon noise statistics of the data, witch dominate the signal of interest, nor the read out noise of the detector, nor the missing data. Moreover, if any image in an angle rotation cycle is missing, the rest of the cycle is note used. Finally, any centering, rotation, or deconvolution by the Point Spread Function are made independently of the data reduction. The bad pixels and dead pixels are interpolated before the processing. The issue of such a processing is that the propagation of the errors in the data is not handled.

The « inverse problem » methods aim to estimate the light of the environment using a direct model of the data, while controlling the error propagation in the reconstruction. This approach is already used in several field in astronomy, but they have not been developed yet for high contrast direct imaging in polarimetry. The aim of my PhD thesis is to reconstruct, under a given quality criterion, a map of the polarized light of the circumstellar environments, a map of the corresponding polarization angles and a map of the residual star light and of the unpolarized light of the environment.

To achieve this goal, in this thesis I develop a non-linear physical model of the data, pixelwise independent, based on Jones formalism and parameterized in the unpolarized intensity, the linearly polarized intensity and the linear polarization angle. Then, I derive an alternative formulation, parameterized in the Stokes parameters, providing the link between such a physical model and the state-of-the-art methods. Throughout this thesis, I extend this model to a pixelwise dependent formulation, without and with the convolution by the Point Spread Function (PSF). In the case without the convolution, I derive a new non-linear model, parameterized in the unpolarized intensity, and the Stokes parameters of the horizontal and vertical polarized intensities. For each of this model, I develop several methods to estimate the parameters, based on the minimization of a constrained criterion and, in the piwelwise dependent case, with

regularizations. Among these regularizations, I compare differentiable and non-differentiable penalizations applied on horizontal and vertical image gradients coefficients and on the singular values of the pixels Hessian matrice. In the linear case, I impose an epigraphical constraint between the Stokes parameters. It corresponds in the linear case to a non-negative constraint on the polarized and unpolarized intensities. To auto-calibrate the regularization weights, I use the Stein Unbiased Risk Estimator (SURE).

The whole methods are applied on simulated dataset, created to reproduce typical astrophysical datasets encountered in circumstellar environment polarimetrical direct imaging. Depending of the properties of the functions considered in the objective function, the research of its minimum is done with different algorithms. In the pixelwise independent case, I proceed with a direct inversion. In the non-linear case, the parameters of interest are estimated hierarchically. In the smooth pixelwise dependent case, I use a preconditionned gradient descent with with the limited memory preconditionnement of Broyden-Fletcher-Goldfarb-Shanno (BFGS). This algorithm is suitable to estimate the parameters under a non-negative constraint. In the linear case with an epigraphical constraint and a smooth regularization, I use the Forward-Backward with backtracking. The backtracking avoids us the calculus of the gradient Lipschitz constant which can be difficult. In the case of non-smooth regularization, I use the preconditionned primal-dual Condat-Vũ algorithm with a backtracking step. Then, I show that in the pixelwise independent case, taking into account the missing data allows us for the use of incomplete data cycle, reducing the maps estimation error. I also point out that in the pixelwise dependent case, taking the detector transformations (rotations, translations) and the dead pixels into account in the model reduce the error on the polarized and non-polarized intensities in the non-linear case of an independent regularization on both quantities. The error on the angle benefit from a linear model with a regularization on the Stokes parameters. I also show that in the cas of low disk polarization the auto-calibration of the regularization weights with SURE tend to over-regularize the estimation, giving a higher bound for a better manual choice. In the case of the deconvolution, I show that a smooth regularization with smoothing prior and edge preserving as the Total Variation hyperbolic approximation (TV-h) gives better result, in a given time, than than the non-smooth penalization of the Total Variation (TV) and the Shatten norm of the Hessian. I also show that the reconstructions estimated from the global model including the convolution have a smaller estimation error than reconstructions with a posterior deconvolution.

Finally, I show on several astrophysical datasets the benefits of the best methods that I develop in this thesis, for the polarized intensity and the angle of polarization. I conclude that taking in account a global model including the convolution for the reconstruction of circumstellar environment, with several level of regularization, allows for a thinner resolution of bright polarized structures and a best detection of low polarized intensity structures.

Remerciements

Je souhaiterais tout d'abord remercier Maud, Nelly et Éric pour avoir accepté de m'encadrer pour ces trois années de thèse et d'avoir eu foi en mon potentiel de mener cette thèse à bien, surtout quand ce n'était pas mon cas. Ces trois années auprès de vous ont été très enrichissantes, aussi bien scientifiquement qu'humainement. Merci d'avoir réussi à supporter mon tempérament pas toujours facile, et mes fautes d'orthographes (certaines perles valaient le coup) ! Merci pour les sachets de thés, les dosettes de café, les repas, les demis de bière... j'espère que nous pourrons toujours nous réunir autour d'un verre ou d'un café pour discuter amicalement ! Je souhaiterais également remercier mes rapporteurs Hugues Talbot et Hans Martin Schmid pour avoir accepté de faire le rapport de ma thèse. Je remercie également Myriam Benisty,
Simon Masnou et Audrey Repetti pour avoir accepté de participer au jury de ma soutenance. Je remercie également les membres de l'équipes Harissa : Renaud, Isabelle, Michel, Ferreol, Jean-François, Gil, pour leur aide et pour toutes ces discussions enrichissantes aussi bien dans le cadre scientifique que général. Je remercie d'autre part les membres de l'équipe SiSyPhe du laboratoire de physique de l'ENS de Lyon pour m'avoir acceptée dans leurs bureaux. Je remercie tous les doctorants, post-doctorants, ingénieurs et étudiant qui ont pu croiser ma route pendant ces trois années et aider à les rendre plus agréable :

- Tout d'abord merci à Samuel, mon «co-bureau »de l'autre coté du couloir pour cette deuxième moitié de thèse, pour ces échanges scientifiques ou autre fort intéressants, pour ton aide et ton soutiens, surtout en cette dernière année de thèse compliquée. Merci à Clément, stagiaire furtif et véritable co-bureau dont une magnifique citation orne encore mon mur, merci pour les vidéos de chat mignonnes qui étaient plus que nécessaires. Merci à vous deux pour ces petites sorties jeux de société qui m'auront permis de me changer les idées !
- Merci à Arnaud, thésard contemporain, surtout pour ces derniers mois de confinement où les discussions avec toi ont permis de me rassurer un peu sur l'avancement de mon manuscrit.
- Merci aux membres de l'AstroGang, aussi bien musical que du quizz (même si je n'y suis finalement allée que très peu) ! Un merci tout particulier à Mathieu, Karen et Diane, mais aussi Anthony, David L. et Peter, dans l'espoir de pouvoir un jour rejouer de la musique (et à Dark Souls) avec vous ! Merci aussi à David C. et Benjamin pour toutes ces discussions intéressantes. Merci à Joanne également, d'être venue vers moi dès mon arrivée à l'observatoire et m'avoir permis, avec Anthony, de m'intégrer dans le groupe. J'espère vous revoir bientôt !
- Merci à tous les autres doctorants du sommet de la colline : Kieran, Marion, Maxime, Adé, Mathieu et Valentin, au plaisir de se refaire une raclette !
- Je voudrais également remercier Barbara, Hadi, Marion et Jean-Michel de l'ENS de Lyon,

pour m'avoir acceptée et intégrée parmi vous, c'était toujours un plaisir de discuter avec vous.
— Enfin je souhaiterai remercier les doctorants de l'OSU que j'ai pu côtoyer lors du conseil pédagogique ou de composante ou de l'organisation du congrès doctorants, en particulier Lucia pour le premier et Anaïs.
— Mes amis de longue date, qui ont également choisi le parcours de doctorants et avec qui j'ai pu continuer à échanger sur les joies d'un tel projet, en particulier Robin, Armand et Cécile.
— Enfin je souhaite remercier ceux qui ne rentrent dans aucune de ces catégories mais que je n'oublie pas pour autant : Wassila, Jenny, Antony B., Olivier, Lola, j'ai passé de très bons moments à vos côtés !

Enfin je souhaiterais remercier ma famille et mes amis pour m'avoir soutenue tout au long de la thèse et surtout sur les derniers mètres. Un remerciement tout particulier pour Martin, mon compagnon et «co-bureau »pendant ces mois de confinements. Merci de m'avoir soutenue et épaulée du mieux que tu le pouvais, surtout sur les dernières semaines de rédaction, qui auront été les plus dures.

Enfin, une pensée pour ceux qui nous ont quittés pendant ces trois années en laissant un vide derrière eux.

Notations et acronymes

Notation

Notations pour les ensembles

$\mathbb{N}$	L'ensemble des entiers naturels.
$\mathbb{Z}$	L'ensemble des entiers relatifs.
$\mathbb{R}$	L'ensemble des réels.
$\mathbb{C}$	L'ensemble des complexes
$\mathcal{H}$	Un espace d'Hilbert
$\mathcal{C}$	Un sous-espace d'Hilbert
$\mathcal{M}_N(\mathcal{C})$	L'ensemble des matrices carré de taille N dont les éléments sont dans $\mathcal{C}$
$\mathcal{S}_N(\mathcal{C})$	L'ensemble des matrices symétriques de taille N dont les éléments sont dans $\mathcal{C}$
$\mathcal{P}_\rho$	L'ensemble des matrices symétriques strictement ρ-positive

Symbols mathématiques

$\in, \ni$	Appartenance à un ensemble
$\subset$	Inclusion dans un ensemble
$\times$	Multiplication terme à terme
$\otimes$	Produit de Kronecker
i	Identité complexe
e	Exponentielle
$*$	Produit de convolution
$\succeq, \preceq$	Ordre de Loewner
∇	Opérateur gradient continue
∇^2	Opérateur hessien continue
$\mathbf{D}$	Opérateur gradient discret ou différences finies
$\mathbf{D}^2$	Opérateur hessien discret
$\cdot_{\mid\mathcal{C}}$	Restriction à l'ensemble $\mathcal{C}$
$\cdot^\top$	Opérateur transposé
$\cdot^*$	Opérateur adjoint
$\bar{\cdot}$	Opérateur conjugué

Problèmes inverses

$\widehat{\cdot}$	Estimation d'un estimateur
$\overline{\cdot}$	Vérité terrain
$\mathbb{E}[\cdot]$	L'espérance d'une variable aléatoire.
$\mathcal{L}(\cdot\|\boldsymbol{x})$	Densité de probabilité d'une réalisation de variable aléatoire sachant $\boldsymbol{x}$
$\mathrm{Var}(\cdot)$	Variance d'une variable aléatoire.
$\mathrm{Cov}(\cdot)$	Covariance d'une variable aléatoire.
min, max	Minimum et maximum d'une fonction.
argmin , argmax	Argument minimal et argument maximal d'une fonction.
Argmin , Argmax	Ensembles respectifs des argmin et argmax d'une fonction.
prox	Opérateur proximal d'une fonction
∂	Sous-différentielle d'une fonction
$\partial_{\boldsymbol{x}}$	Dérivée directionnelle dans la direction de $\boldsymbol{x}$
$\Psi(\cdot)$	Fonction objectif
$\Phi(\cdot)$	Attache aux données
$\mathcal{R}(\cdot)$	fonction de régularisation
$\iota_{\mathcal{C}}$	fonction indicatrice de l'ensemble $\mathcal{C}$
$\mathbb{P}_{\mathcal{C}}$	Projection orthogonale sur l'ensemble $\mathcal{C}$.

Notations générales

x	un scalaire
X	un scalaire correspondant à la dimension d'un ensemble.
$\boldsymbol{x}$	un vecteur
$\mathbf{M}$	une matrice où un opérateur linéaire
t	le temps
$[t]$	une itération
$((x)^{[t]})_{t\in\mathbb{N}}$	une suite d'itérés
K	Le nombre d'images d'un jeu de données
M	Le nombre de pixels par image du jeu de données
L	Le nombre de composantes d'un estimateurs
N	Le nombre de pixels par composantes d'un estimateur
$\boldsymbol{d}$	Données
$\boldsymbol{d}^{\mathrm{S}}$	Données *pré-traitées*
$\boldsymbol{d}^{\mathrm{nS}}$	Données *calibrées*
$f_{\dots}$	Un modèle des données
$f_{\dots}^{\mathrm{S}}$	Un modèle séparable des données en les pixels
$f_{\dots}^{\mathrm{nS}}$	Un modèle non-séparable des données en les pixels
$f_{\dots}^{\mathrm{nS\text{-}C}}$	Un modèle non-séparable des données incluant la convolution

I^{p}	Carte d'intensité polarisée
I^{u}	Carte d'intensité non-polarisée
Θ	Carte d'angle de polarisation
I	Paramètre de Stokes de la polarisation totale
Q	Paramètre de Stokes de l'intensité polarisée horizontalement
U	Paramètre de Stokes de l'intensité polarisée verticale
$\mathcal{F}(\cdot)$	Fonction différentiable
$\mathcal{G}(\cdot)$	Fonction non-différentiable
$g(\cdot)$	Fonction de pénalisation
λ	Hyperparamètre de la contribution des régularisations

Acronymes

En astrophysique

ESO	European Southern Observatory
VLT	Very Large Telescope
SPHERE	Spectro-Polarimeter High-contrast Exoplanet REsearch
IRDIS	InfraRed Dual-band Imaging and Spectroscopy
IFS	Integral Field Spectrograph
ZIMPOL	Zürich IMaging Polarimeter
DPI	Dual Polarimetry Imaging
ALMA	Atakama Large Milimeter Array
GPI	Gemini Planet Imager

Méthodes et algorithmes

MnLS	Méthode non-Linéaire Séparable
MLS	Méthodes Linéaire Séparable
MLnS	Méthode Linéaire non-Séparable
MnLnS	Méthode non-Linéaire non-Séparable
FB	Forward-Backward
VMFB	Variable-Metric Forward Backward
PD	Primal-Dual
VMPD	Variable Metric Primal-Dual
FDCR	Fréchet-Darmois-Cramér-Rao
BFGS	Broyden-Fletcher-Goldfarb-Shanno
VMLM-B	Variable Metric Limited Memory and Bound
SURE	Stein Unbiased Risk estimator
TV	Total Variation
TGV	Total Generalized Variation

Table des matières

Résumé **I**

Abstract **III**

Remerciements **V**

Table des notations **VII**

Introduction **1**

1 État-de-l'art et formalisme **7**
1.1 L'étude des environnements circumstellaires en polarimétrie 8
1.1.1 Les environnements circumstellaires 8
1.1.2 Méthodes d'observation des environnements circumstellaires 10
1.1.3 La polarimétrie en imagerie directe 12
1.1.4 L'instrument ESO/VLT-SPHERE 15
1.1.5 Méthodes de reconstruction en imagerie directe polarimétrique 18
1.2 Excursion au cœur des *«problèmes inverses »* 27
1.2.1 Problème direct 27
1.2.2 Attaches aux données 28
1.2.3 Régularisations 29
1.2.4 Contraintes strictes 33
1.2.5 Estimation de l'erreur 34
1.3 Les méthodes de résolution 35
1.3.1 Comment trouver le minimum d'une fonction 36
1.3.2 Méthodes différentiables : Quasi-Newton et ℓ-BFGS 39
1.3.3 L'algorithme Forward-Backward à métrique variable 41
1.3.4 Les algorithmes primaux-duaux et l'algorithme Condat-Vũ à métrique variable 43
1.3.5 Leurs utilisations en astrophysique 45

2 Modèle direct séparable des données de l'instrument ESO/VLT-SPHERE IRDIS **47**
2.1 MnLS : Modèle non-Linéaire Séparable avec le formalisme de Jones 48
2.1.1 Le modèle direct non-linéaire séparable 48
2.1.2 Résolution séparable du modèle non-linéaire 52
2.1.3 Simulation des données pré-traitées 54
2.1.4 Application sur données simulées et données astrophysiques 56

2.2 MLS : Modèle Linéaire Séparable sur les paramètres de Stokes 64
2.2.1 Le modèle direct séparable linéaire 64
2.2.2 Résolution du problème séparable 65
2.2.3 Application sur données simulées et données astrophysiques 66

3 Modèle direct non-séparable des données de l'instrument ESO/VLT-SPHERE IRDIS 73
3.1 Modèle direct linéaire non-séparable 74
3.1.1 Le modèle direct non-séparable linéaire 74
3.1.2 Résolution par approche inverse différentiable 75
3.1.3 Simulation des données calibrées 77
3.1.4 Application sur données simulées et données astrophysiques 77
3.2 Modèle direct non-linéaire non-séparable 87
3.2.1 Le modèle direct non-séparable non-linéaire 87
3.2.2 Résolution par approche inverse différentiable 88
3.2.3 Application sur données simulées et données astrophysiques 88

4 Modèle direct non-séparable incluant la convolution 97
4.1 Du modèle direct à la fonction objectif 98
4.1.1 Modèle linéaire non-séparable prenant en compte la convolution . . . 98
4.1.2 Fonction objectif 101
4.2 MLnS-D : Méthodes Linéaires non-Séparables avec Déconvolution 105
4.2.1 Résolution par approche différentiable 106
4.2.2 Résolution par approche différentiable avec projection épigraphique . 106
4.2.3 Résolution par approche non-différentiable 109
4.2.4 Comparaison des performances sur données simulées 113
4.3 Applications sur données simulées et données astrophysiques 116
4.3.1 Résultats sur données simulées 116
4.3.2 Résultats sur données astrophysiques 127
4.3.3 Synthèse des résultats 132

5 Érreur d'estimation 135
5.1 Erreur d'estimation dans le cas d'un bruit gaussien centré 136
5.1.1 Information de Fisher d'un modèle gaussien centré 137
5.1.2 Étude théorique de la borne de FDCR dans le cas non-biaisé 138
5.1.3 Estimation de l'erreur dans le cas d'un estimateur artificiellement biaisé 140
5.2 Érreur des modèles séparables 144
5.2.1 Borne de FDCR de la MnLS 145
5.2.2 De la borne de FDCR de la MLS à une approximation de la variance pour la MnLS 145
5.2.3 Application pour un pixel sur données simulées 146
5.3 Érreur des modèles non-séparables 149
5.3.1 Critère SURE pour le cas multidimensionnel 149
5.3.2 Application aux méthodes non-séparables 150

6 Étude astrophysique **153**
6.1 Calibration . . . 153
6.1.1 Calibration du détecteur . . . 153
6.1.2 Calibration instrumentale . . . 159
6.1.3 Pré-traitement des données *calibrées* . . . 161
6.2 Reconstruction et étude de différentes cibles astrophysiques . . . 162
6.2.1 RXJ 1615 . . . 162
6.2.2 T Tauri . . . 166
6.2.3 Étude d'un nouvel objet . . . 172

7 Conclusion **179**
7.1 Contributions . . . 179
7.2 Perspectives . . . 181

Bibliographie **192**

Annexes **192**

A Convergences des algorithmes : définitions et théorèmes **193**
A.1 Définitions et propositions . . . 193
A.2 On the convergence of Krasnosel'skii-Mann . . . 194
A.3 Forward-Backward . . . 196
A.3.1 Preuve du théorème 1.3.3 . . . 198
A.4 Convergence du primal-dual . . . 202

B Papiers de conférences et de colloque acceptés et papier A&A soumis **207**

Introduction

Les environnements circumstellaires sont divers et variés. De l'étoile simple au système d'étoile binaire, du disque protoplanétaire au disque de débris, observés de face, de profil, à différentes longueurs d'ondes, différentes étendues, d'un objet à l'autre, les données peuvent-être très différentes. Tout système d'exoplanètes commence, après effondrement d'un nuage moléculaire pour former une étoile jeune, par un disque protoplanétaire. Celui-ci est alors constitué de gaz et de poussières fines, dont la taille est de l'ordre du micromètre. Il est alors le berceau de futurs exoplanètes.

Au début du processus de formation des exoplanètes, le gaz et la poussière tournent autour de l'étoile. Commence ensuite une phase où la poussière s'accrète pour former des grains, dont la taille est de l'ordre du kilomètre, qui nettoient petit à petit le gaz, on parle de disque de transition. Ces grains sont alors assez massifs pour attirer à eux plus de matière jusqu'à former des planètes. Finalement, quand il n'y a quasiment plus de gaz, qu'il ne reste que de très gros grains et qu'il peut y avoir une ou plusieurs planètes dans l'environnement, on parle de disque de débris. Dans notre système solaire, un tel disque correspondrait à la ceinture de Kuiper, dans laquelle se situe la planète naine Pluton.

L'étude de ces environnements circumstellaires est la clef pour comprendre comment se forment les systèmes stellaires et les exoplanètes. La taille de ces environnements circumstellaires peut varier entre quelques dizaines à plusieurs centaines de fois la distance Terre-Soleil, distance appelée Unité Astronomique (AU), selon la masse totale de l'environnement. Cependant, ces environnements se situent à plusieurs centaines d'années lumières par rapport à nous, soit plusieurs millions de milliards de kilomètres. De ce fait leurs tailles apparentes, ou tailles angulaires, calculées comme l'angle entre le bord de l'objet, notre œil et le bord opposé, varie seulement entre $0,01$ et 2 secondes d'arc.

L'œil nu ne permet d'observer distinctement, ou de résoudre, que des objets célestes de taille angulaire supérieure à une minute d'arc. Pour pouvoir observer ces environnements avec une meilleure résolution, il faut des télescopes de très grande taille, comme les Très Grands Télescopes au Chili, VLT pour Very Large Telescope en anglais, dont les miroirs principaux font plus de huit mètres ou le radiotélescope ALMA, pour Atacama Large Millimeter Array en anglais, dont l'écartement des antennes peut aller jusqu'à seize kilomètres.

Pour acquérir des informations sur ces environnements circumstellaires, différentes techniques d'imagerie existent. D'une part l'interférométrie, comme ALMA qui fonctionne dans le domaine des longueurs d'ondes millimétriques et qui permet d'étudier la structure interne du disque. D'autre part l'imagerie directe, qui permet d'observer la surface de ces environnements dans les longueurs d'onde infra-rouge.

L'observation des environnements circumstellaires reste cependant très difficile. En effet, ces environnements circumstellaires sont très proches d'une étoile hôte émettant beaucoup plus de lumière, d'un contraste bien supérieur à 1000. Or, lors de l'observation, la lumière

de l'étoile ne correspond pas à un point mais, du fait de différents effets de diffraction, à une tâche étendue dans laquelle est noyée la lumière de l'environnement lui-même. L'analogie de l'observation des environnements circumstellaires ou des exoplanètes est souvent faite avec l'observation d'une luciole qui vole à côté d'un phare à des centaines de kilomètres de l'observateur, car les rapports d'intensité et de séparation angulaire sont similaires.

L'utilisation d'un masque, appelé coronographe, permet d'éteindre une partie de la lumière de l'étoile et d'ainsi réduire la brillance de la tâche stellaire. Différents types de coronographe existent, mais ils ne sont jamais parfaits et sur les images acquises par les instruments à haut contraste, un résidu de lumière stellaire est toujours présent. Afin de pouvoir étudier les environnements circumstellaires à partir des images acquises par ces instruments, il faut donc appliquer des techniques d'observations et de traitement permettant de distinguer la lumière de l'environnement des résidus de lumière de l'étoile. C'est dans ce contexte que se place ma thèse.

Pour procéder au démélange du signal de l'environnement et du signal stellaire, il existe différentes techniques. La première est la différentiation angulaire. Cette méthode se fonde sur l'hypothèse que lorsqu'on fait tourner artificiellement le champ observé, les résidus de lumière stellaire restent fixes, tandis que l'environnement observé tourne. En prenant différentes images à différentes orientations connues, la diversité produite aide à séparer les composantes environnement et résidu. Avec cette méthode, il est possible d'observer principalement des exoplanètes et des disques dont la forme n'est pas invariante par rotation apparente, en ajoutant cependant des artefacts importants sur leurs morphologies.

La seconde méthode, plus fine pour étudier leur morphologie, utilise la différence de propriétés entre la lumière de l'étoile et la lumière diffusée par la poussière du disque. En effet, la lumière diffusée par la poussière du disque est issue de la réflexion de la lumière de l'étoile sur les grains qui composent le disque. Les ondes qui composent la lumière de l'étoile oscillent de manière aléatoire lors de leurs propagation. En se reflétant sur les grains de poussières, elles se mettent chacune à osciller dans des plans uniques, perpendiculaires aux angles d'incidence des ondes sur la poussière. En faisant tourner artificiellement ces plans d'oscillation d'un certain nombre d'angles connus, les ondes composant la lumière stellaire ne sont pas affectées. Il est alors possible d'identifier et d'extraire la lumière émise par la poussière du résidu de lumière stellaire. La façon dont oscille l'onde est la polarisation de la lumière. La lumière émise par l'étoile est dite non-polarisée. La lumière réfléchie par la poussière du disque est polarisée linéairement et l'angle que fait son plan d'oscillation est appelé angle de polarisation linéaire. C'est sur cette méthode d'observation, qui permet ce démélange en imagerie polarimétrique, qu'est basée ma thèse.

L'instrument avec lequel sont prises les images sur lesquelles j'ai travaillé est l'instrument d'imagerie directe SPHERE-IRDIS de l'European Southern Observatory (ESO), installé sur l'un des Très Grands Télescopes au Chili. Il permet avec son mode d'imagerie polarimétrique d'acquérir un ensemble d'images en faisant tourner artificiellement l'angle de polarisation des objets célestes étudiés.

À partir de ces observations, il est possible de reconstruire trois images, ou cartes de pixels, qui constituent notre signal d'intérêt. Une première image contenant la lumière non-polarisée de l'étoile et de son environnement, une seconde image, contenant la lumière polarisée linéairement de l'environnement circumstellaire, et une carte contenant les angles de polarisation

linéaire de cet environnement. La figure 1 représente la lumière polarisée linéairement de différents environnements circumstellaires.

Les méthodes de la Double Différence et du Double Ratio [Tinbergen, 2005] permettent traditionnellement de reconstruire le signal d'intérêt par combinaison des données. Ces méthodes sont directes, dans le sens où les données sont transformées puis combinées entre elles pour reconstruire les trois composantes. L'inconvénient de telles méthodes, est qu'elles ne tiennent pas compte du caractère aléatoire du signal enregistré dans les images, qui se traduit par un bruit statistique, et qu'elles propagent plus facilement les erreurs statiques et statistiques.

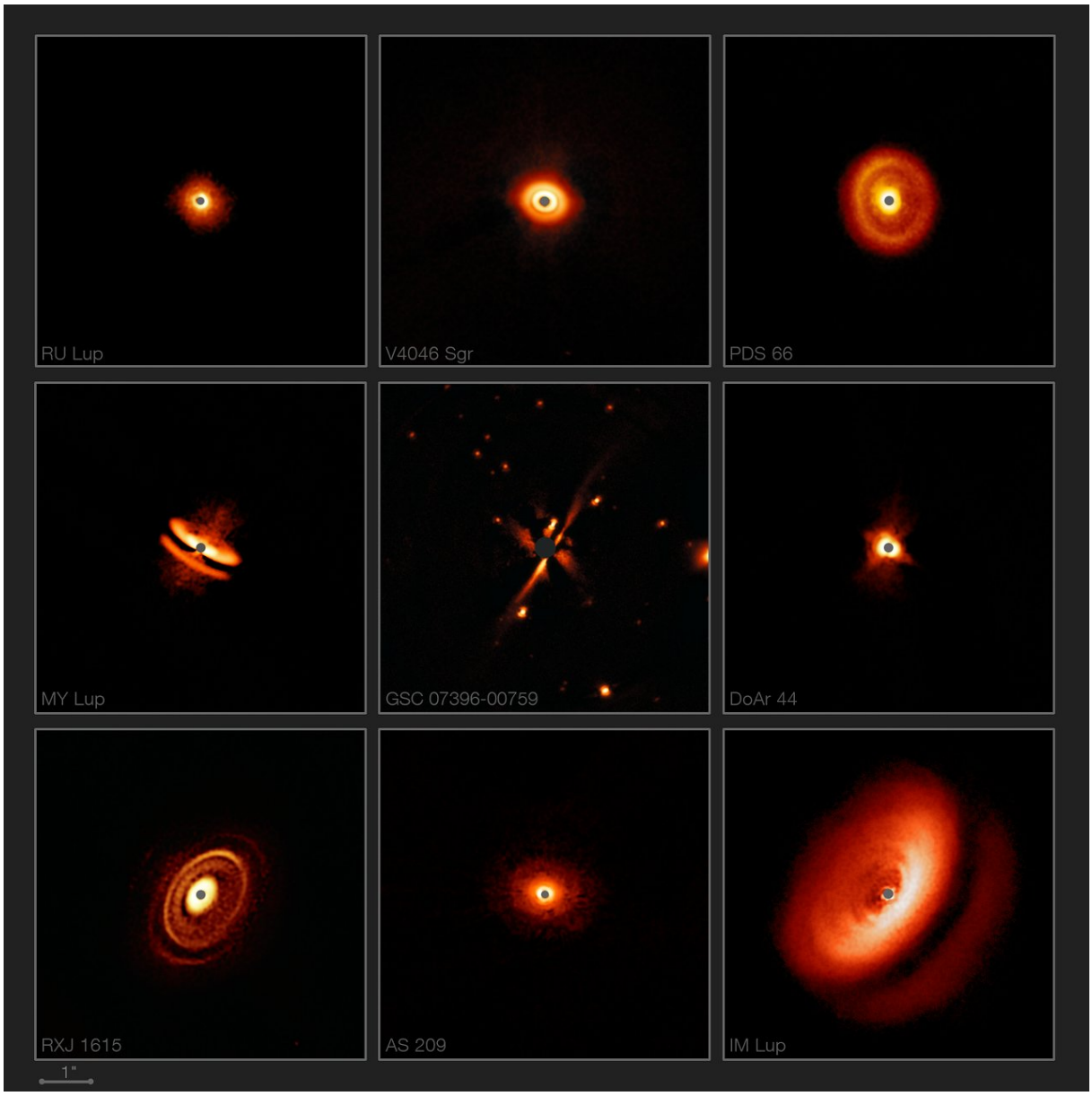


Figure 1 – Disques circumstellaires observés avec SPHERE. On peut observer la diversité de forme et d'orientation de ces différents disques, dont l'étendue se situe entre 0, 5 et 3 secondes d'arc. [Avenhaus et al., 2018, Sissa et al., 2018].

Dans cette thèse, je m'intéresse à reconstruire des images des environnements circumstellaires à partir des acquisitions en polarimétrie de l'instrument ESO-VLT SPHERE IRDIS. À l'issu de mes travaux de thèse, j'aboutis à un ensemble de méthodes de complexités différentes, de type *«problèmes inverses »*, prenant en compte dans un modèle physique des données, toutes les transformations de la lumière ayant lieu dans l'instrument, de l'entrée dans le télescope jusqu'à l'acquisition par la caméra, permettant ainsi, par ajustement du modèle sur les données, d'extraire de manière optimale les cartes d'intensité non-polarisée, d'intensité-polarisée et d'angle de polarisation, ainsi que leurs cartes d'erreurs d'estimation respectives.

Dans le chapitre 1, je présente l'état-de-l'art observationnel pour les environnements circumstellaires et les exoplanètes. J'explique alors ce qu'est la polarisation de la lumière et présente plus précisément l'instrument ESO-VLT SPHERE/IRDIS ainsi que le formalisme et les méthodes existantes pour le traitement des données. Je présente ensuite ce que sont les *«problèmes inverses »*, c'est-à-dire comment à partir d'un modèle direct des données, arriver à un problème d'optimisation numérique sous forme de critère à minimiser, composé d'une attache aux données et éventuellement de régularisations et de contraintes strictes. Dans une dernière section, je présente quelques algorithmes permettant la résolution de ces *« problèmes inverses »* dans le cas de régularisations différentiables et non-différentiables.

Les méthodes de type *«problèmes inverses »* sont utilisées depuis plusieurs décennies en astrophysique, notamment en radio-interférométrie. Cependant, en imagerie directe à haut contraste, leur utilisation n'est pas encore répandue. Pourtant, étant donné la difficulté liée aux niveaux de contraste, il semble pertinent d'utiliser des méthodes de type *« problèmes inverses »* pour l'extraction des images qui nous intéressent, car ils permettent la prise en compte des erreurs dans les données, et donc par extension, des données manquantes et des effets instrumentaux. Dans le chapitre 2, je développe donc un modèle direct des acquisitions, prenant en compte les transformations de la polarisation de la lumière induites par l'instrument au cours de l'observation, paramétré en chaque pixel de manière non-linéaire, en l'intensité non-polarisée, l'intensité polarisée linéairement et l'angle de polarisation. Pour ce modèle j'utilise le formalisme de Jones. Je présente une méthode d'ajustement du modèle aux données, prenant an compte les données manquantes, et je compare les images extraites à celles obtenues avec les méthodes de l'état-de-l'art. Dans un second temps, je présente un modèle linéaire paramétré par les paramètres de Stokes ainsi qu'une méthode d'ajustement du modèle aux données prenant en compte les données manquantes. Je présente le lien entre une telle méthode et les méthodes de l'état-de-l'art et compare les images extraites pour de telles méthodes avec la méthode utilisant le formalisme de Jones. Dans ce chapitre, je montre l'apport de la prise en compte des données manquantes dans le modèle par rapport au fait d'enlever une partie du jeu de données.

Dans le chapitre 3, je complète le modèle linéaire par les paramètres de Stokes, en incluant les transformations au niveau de la caméra, afin de pouvoir prendre en compte les pixels défectueux. La transformation se représente comme un opérateur mathématique linéaire, qui lors de l'ajustement du modèle aux données, du fait de son mauvais conditionnement, va amplifier le bruit dans les images extraites. Pour l'ajustement du modèle aux données, je propose alors un critère régularisé favorisant des images lisses. Je compare les images extraites pour des régularisations de Tikhonov [Tikhonov, 1963] et Variation Totale Hyperbolique [Charbonnier et al., 1997], après minimisation du critère par une méthode quasi-Newton, à celles obtenues avec les méthodes de l'état-de-l'art. L'intensité polarisée se trouvant dans tous les paramètres de Stokes, je propose alors un changement de variables permettant de ré-

gulariser sur l'intensité polarisée et l'intensité non-polarisée de manière indépendante. Après minimisation du critère régularisé associé par une méthode quasi-Newton, je compare à la méthode linéaire et aux méthodes de l'état-de-l'art. Dans ce chapitre, je montre aussi l'apport d'une approche différentiable régularisée pour la prise en compte des transformations de la caméra et des pixels défectueux dans le modèle, par rapport aux méthodes appliquées sur des données déjà transformées. Je montre également qu'une régularisation par la Variation Totale Hyperbolique donne des résultats avec une erreur moindre par rapport à une régularisation quadratique de Tikhonov.

Dans le chapitre 4, je complète une dernière fois le modèle en incluant le flou résultant de l'instrument. Ce flou se représente par un opérateur mathématique linéaire par rapport aux paramètres, mais il est également mal conditionné. Pour limiter l'augmentation du bruit dans les images estimées à partir de ce modèle, j'utilise d'une part une régularisation différentiable, telle que la Variation Totale Hyperbolique [Charbonnier et al., 1997], et non-différentiables, comme la Variation Totale [Rudin et al., 1992] ou encore la norme de Shatten sur la matrice hessienne [Lefkimmiatis et al., 2013]. Je présente alors différentes méthodes de minimisation des critères régularisés selon la différentiabilité des régularisations telles que Forward-Backward [Combettes and Vũ, 2013, Combettes et al., 2014] ainsi que l'algorithme primal-dual Condat-Vũ [Condat, 2013, Vũ, 2015] dans le cas non différentiable, dans lesquels je propose d'ajouter une étape de backtracking. Je compare alors les images extraites à celles obtenues avec les méthodes de l'état-de-l'art. Je montre l'apport de la prise en compte du flou dans le modèle par rapport aux méthodes où le flou est traité après reconstruction.

Dans le chapitre 5, je présente les méthodes d'estimation de l'erreur faite sur les images reconstruites pour chaque pixel à l'aide de la borne inférieure de Fréchet-Darmois-Cramér-Rao (FDCR). La contribution des régularisations pour la reconstruction des images est gérée par un hyperparamètre nécessitant un réglage minutieux, car la qualité des reconstructions en dépend. Je présente une méthode d'estimation des hyperparamètres à partir de l'estimateur non-biaisé du risque de Stein (SURE) [Ramani et al., 2008, Deledalle et al., 2014]. Ces méthodes sont utilisées au long des chapitres précédents. Je présente à chaque fois le cas simple et le cas appliqué à mon problème de reconstruction en polarimétrie.

Enfin, dans le chapitre 6, après avoir expliqué comment calibrer les données de l'instrument ESO/VLT-SPHERE IRDIS, je montre l'apport des meilleures méthodes présentées dans cette thèse sur différents environnements circumstellaires, par rapport aux méthodes de l'état-de-l'art. Pour finir, j'applique alors la meilleure méthode pour l'étude de l'environnement circumstellaire autour du système d'étoile EM SR 13, qui est un disque autour d'une étoile binaire, associée à une troisième étoile.

Chapitre 1

État-de-l'art et formalisme

Ce chapitre a pour vocation d'établir les bases en astrophysique et en traitement du signal sur lesquelles s'appuient ma thèse. L'étude des environnements circumstellaires présente deux intérêts majeurs. Le premier est d'étudier l'évolution des disques circumstellaires afin de comprendre la formation des systèmes exoplanétaires. Cette étude se fait par la confrontation de résultats de simulations hydrodynamiques, faites sous certaines hypothèses, aux environnements observés. Le second intérêt est la détection d'exoplanètes, soit par acquisition directe et traitement des images pouvant contenir des exoplanètes, soit par déduction de leur présence grâce à l'étude morphologique des environnements. Une fois les planètes détectées par imagerie directe, il est possible d'étudier leurs orbites et leurs compositions atmosphériques. Dans les deux cas, il est nécessaire pour pouvoir étudier ces environnements, d'avoir les outils, aussi bien instrumentaux que numériques, permettant d'avoir une résolution angulaire et un contraste optimaux.

Dans la section 1.1, je présente le contexte astrophysique de l'étude des environnements circumstellaires ainsi que l'évolution des méthodes d'observation de ces environnements à haut contraste. Afin de détailler le fonctionnement de l'observation en polarimétrie, j'introduis la polarisation de la lumière et son utilisation en astrophysique. Je présente ensuite le fonctionnement de l'instrument d'imagerie directe ESO/VLT-SPHERE IRDIS, qui a permis d'obtenir les données d'environnements circumstellaires en polarimétrie analysées dans mon travail de thèse. Je conclus cette section en présentant les méthodes de l'état-de-l'art permettant de démélanger le signal polarisé des environnements circumstellaires du signal non-polarisé.

De manière générale, les méthodes de l'état-de-l'art utilisées en imagerie à haut contraste en polarimétrie, procèdent au démélange étape par étape. En effet, après calibration des données pour détecter les pixels morts, enlever le fond et réajuster la réponse des pixels, les données sont découpées, recentrées et tournées, avant de procéder au démélange par combinaison directe des données. Si une déconvolution a lieu pour enlever le flou instrumental, elle est généralement faite *a posteriori*. Une telle approche ne permet pas de connaître comment se propagent les erreurs d'estimation et de les contrôler. Les méthodes de type *«problèmes inverses »*, introduites dans la section 1.2, permettent ce contrôle. Elles reposent premièrement sur une étape de modélisation du processus d'acquisition, pouvant intégrer modulation de la polarisation, convolution et rotations, reliant les paramètres inconnus que l'on souhaite estimer et les données acquises par l'instrument. En second, elles reposent sur une étape d'inversion fréquemment associée à la minimisation d'une fonction de vraisemblance pénalisée. La résolution sous forme de *«problème inverse »*, nous permet ici de démélanger le signal polarisé de l'envi-

ronnement circumstellaire et le signal non-polarisé du résidu stellaire et de l'environnement, comme présenté sur la figure 1.2b.

De nombreux algorithmes permettent d'obtenir une solution à la minimisation d'une fonction de vraisemblance pénalisée. Le choix de l'algorithme va à la fois dépendre de l'architecture du problème et de sa capacité à résoudre le problème en un temps réduit, c'est-à-dire de quelques heures à quelques jours selon la taille des données et de la taille des paramètres, en nombre de pixels. Dans la section 1.3, tout en introduisant des notions essentielles à la résolution de *«problèmes inverses »*, j'introduis les algorithmes principaux de l'état-de-l'art en traitement du signal et les conditions nécessaires à leur convergence. Je présente enfin leurs utilisations en astrophysique.

1.1 L'étude des environnements circumstellaires en polarimétrie

Au cours des derniers siècles, l'architecture et la composition de notre système solaire ont été de mieux en mieux comprises grâce au développement et au perfectionnement des instruments permettant d'observer l'univers. Par l'agrandissement du diamètre des télescopes, l'amélioration des systèmes d'optique adaptative et la résolution pixélique des caméras, ces instruments ont permis la découverte de nouveaux mondes : des systèmes exoplanétaires plus ou moins semblables au nôtre. La recherche d'exoplanètes et la caractérisation de celles-ci est un des piliers de l'astronomie moderne. L'étude de leurs propriétés et la recherche d'une planète aux propriétés proches de la nôtre est par ailleurs un sujet qui passionne le grand public.

1.1.1 Les environnements circumstellaires

La catégorie des environnements circumstellaires regorge d'une grande variété d'objets d'études, en lien avec les interactions stellaires et la formation des exoplanètes.

Les disques circumstellaires, dont un florilège est présenté sur la figure 1, sont au centre de l'étude de ces environnements. Parmi ces disques on distingue en particulier les disques protoplanétaires, les disques de transition et les disques de débris. La formation des disques circumstellaires débute lors de l'effondrement de nuages moléculaires, formant ainsi une ou plusieurs étoiles. Lors de cette formation s'accrète un disque de poussière et de gaz, appelé disque protoplanétaire [Lynden-Bell and Pringle, 1974]. Ces disques d'accrétion ressemblent globalement à des tores, où la poussière en surface est composée de grains dont la taille est de l'ordre du micromètre, tandis que les grains qui se trouvent dans le plan médian du disque sont plus gros, c'est-à-dire de l'ordre du millimètre.

Dans le cas de l'observation des disques protoplanétaires, donc des disques au stade le plus jeune, l'étude de leur composition et de leurs morphologies, en lien avec des simulations hydrodynamiques, nous permet d'étudier les *scenarii* de formation de ces disques comme dans l'étude du disque HD142527 [Price et al., 2018]. Sur la figure 1, les disques IM Lup, RU Lup, les disques autour de l'étoile AS 209 et du système binaire V4046 Sgr [Avenhaus et al., 2018], sont des disques protoplanétaires.

À l'intérieur de ces disques protoplanétaires, la poussière s'accrète ensuite jusqu'à former des grains dont la taille est de l'ordre du kilomètre. On parle alors de planétésimaux. Leur masse

devient alors suffisamment grande pour attirer à eux plus de matière et continuer à croître jusqu'à devenir des planètes. Lorsque la majorité poussière interne du disque est accrétée mais qu'il reste encore du gaz, on parle alors de disque de transition.

Les disques de transition sont particulièrement intéressants à observer, car leurs morphologies peuvent donner des signes de la formation active d'une ou plusieurs planètes. En effet, lors de leurs formations, les planètes « nettoient » leurs orbites de la poussière, ce qui peut alors créer des *gaps*, c'est-à-dire des anneaux sans poussières et sans gaz, comme dans le cas de RXJ 1615, MY Lup et PDS 66 sur la figure 1 [Avenhaus et al., 2018], ou encore comme le cas de PDS 70 [Keppler et al., 2018, Keppler et al., 2019, Haffert et al., 2019] dont l'observation récente a permis de découvrir une seconde exoplanète. Dans le cas où les exoplanètes sont trop petites pour être visibles, le scénario de formation de ces *gaps* peut être expliqué par des simulations hydrodynamiques, comme dans le cas de HL Tau [Dipierro et al., 2015], ou encore HD 163296 [Pinte et al., 2019]. En attirant à elles la matière du fait de la gravité, les planètes peuvent également former des «bras spiraux »comme en particulier dans le cas de RY Lup [Langlois et al., 2018], visible sur la figure 1.1, ou encore MWC 758 [Benisty et al., 2015], où la présence d'exoplanètes est soutenue par le couplage des observations à des simulations hydrodynamiques.

Enfin, lorsque les planètes ont nettoyé tout le gaz et la poussière, qu'il ne reste quasiment plus de gaz, et que les disques sont composés de grains assez gros, on parle alors de disques de débris. Ces disques sont composés d'anneaux de poussières qui n'ont jamais pu être accrétées, comme dans le cas de HR 4796A observé avec le Gemini Planet Imager et l'instrument SPHERE/ZIMPOL [Perrin et al., 2015, Milli et al.,]. À la fin du processus de formation planétaire, l'observation par imagerie directe permet à la fois la détection de nouvelles exoplanètes mais également l'étude de leurs masses et de leurs compositions. L'observation par imagerie directe des exoplanètes de faibles masses, c'est-à-dire inférieur à 10% de la masse de Jupiter, est actuellement impossible car le niveau de contraste de ces objets est trop élevé. Dans [Hunziker et al., 2020], les auteurs étudient le contraste maximal qu'il est possible d'atteindre pour l'observation d'exoplanètes en lumière réfléchie et polarisée dans le visible. Il est cependant possible d'observer les planètes géantes, suffisament jeunes et chaudes qui émettent de

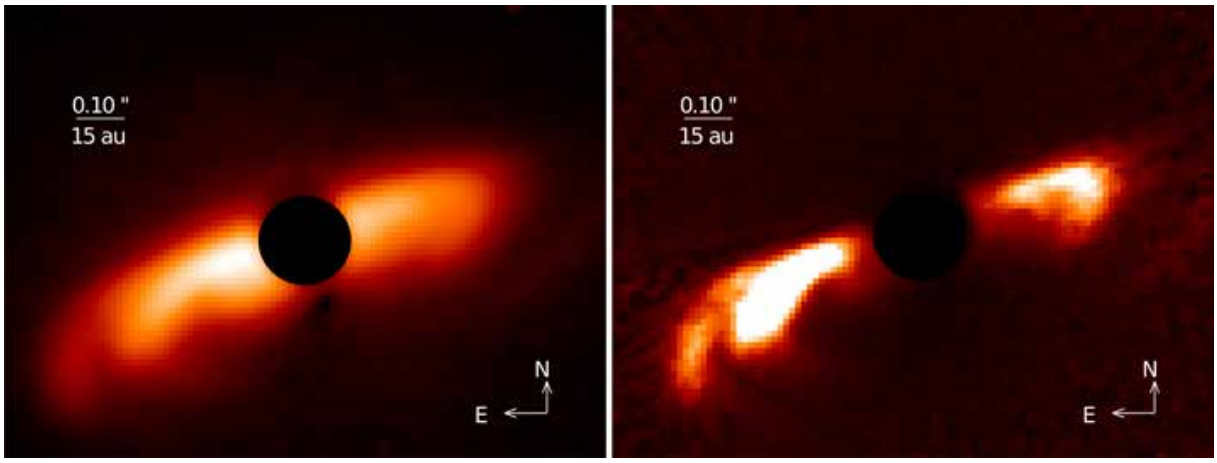


FIGURE 1.1 – Inensité polarisée du disque de transition Ry Lup extraite du papier [Langlois et al., 2018]. L'image de droite correspond à l'image de gauche où le flou instrumental a été enlevé.

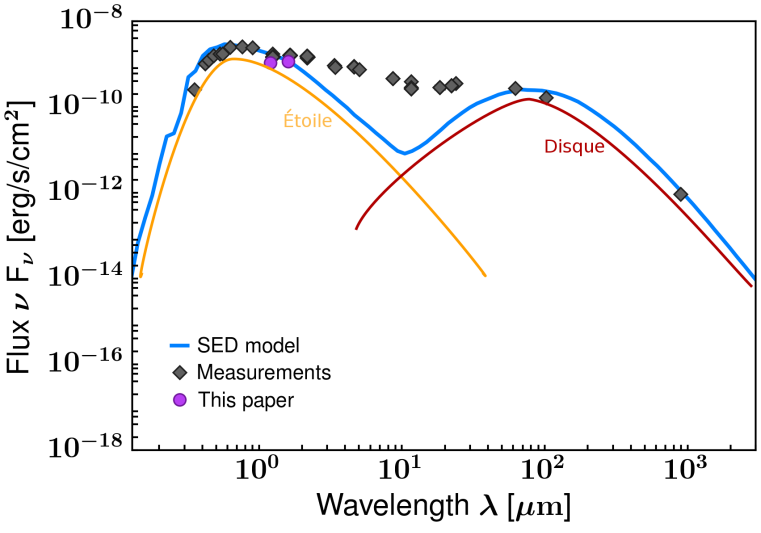
la lumière non-polarisée dans le proche infrarouge, telles que celle du système HR 8799 observées avec SPHERE [van Holstein et al., 2017] ou encore du PDS70 [Keppler et al., 2018]. Il est est également possible d'utiliser l'imagerie directe pour étudier et caractériser les naines brunes [Mesa et al., 2016].

1.1.2 Méthodes d'observation des environnements circumstellaires

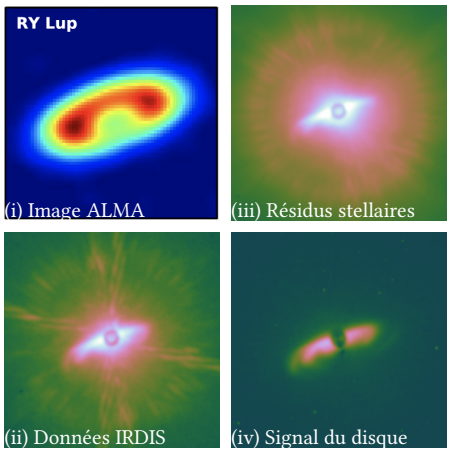
Pour pouvoir observer distinctement les structures de ces environnements, ou détecter des exoplanètes proches de l'étoile, c'est-à-dire ayant une faible séparation angulaire, il est nécessaire d'avoir une résolution instrumentale suffisante. Selon le critère de Rayleigh, la taille angulaire minimale d'un objet que l'on peut observer avec un télescope doit être supérieure à la résolution angulaire Θ, qui dépend à la fois du diamètre D du télescope et de la longueur d'onde λ à laquelle on observe, selon la relation suivante :

$$\Theta \simeq 1,22\lambda/D \qquad \text{(en radian)}. \tag{1.1}$$

Ainsi, plus le diamètre D est grand, plus Θ est petit, ce qui permet d'observer des objets de plus petite taille angulaire ou de faible séparation angulaire. De même, plus les longueurs d'ondes de la lumière émise par les objets sont grandes, plus il est nécessaire d'avoir un grand diamètre de télescope pour résoudre les objets. Par exemple, pour une longueur d'onde micrométrique, avec un diamètre de $8,2$m tel que celui du VLT, il est possible de résoudre des objets dont la séparation angulaire est de plusieurs millisecondes d'arc (mas). Cela correspondrait à distinguer deux objets séparés d'environ quelques centimètres à 200km de nous. Dans le cas des ondes millimétriques, un tel diamètre de télescope ne permet de résoudre des objets séparés que de plusieurs secondes d'arc. Pour résoudre des objets de plus faible séparation angulaire dans les longueurs d'ondes millimétriques, il est nécessaire d'avoir des diamètres de télescopes



(a) Distribution d'énergie spectrale (SED : *Spectral Energy Distribution*) de la cible Ry Lup (source : [Langlois et al., 2018]). La courbe orange correspond à b.(ii)-(iv) et la courbe rouge à b.(i).



(b) Image du disque ALMA (i) et extrait des données de l'instrument ESO/VLT-SPHERE IRDIS (ii) dans laquelle le signal du disque (iv) est mélangé aux résidus stellaires (iii).

FIGURE 1.2 – Distribution spectrale d'énergie et illustration du problème de démélange de Ry Lup.

de plusieurs centaines voir milliers de mètres. La taille du miroir est également responsable du nombre de photons captés. Comme il n'est pas nécessaire, dans le domaine millimétrique, d'avoir un très grand miroir pour acquérir les photons, la solution pour avoir une grande résolution angulaire est de combiner les signaux de plusieurs antennes ou petits télescopes, dispersés sur le diamètre. C'est ce qu'on appelle l'interférométrie.

La poussière et le gaz composant les disques circumstellaires émettent de la lumière dans le visible et l'infrarouge proche (NIR : *Near-Infrared*). Selon la taille des grains de poussière, ces longueurs d'ondes sont de l'ordre du micromètre et du millimètre (10^{-6} et 10^{-3} mètres) . La figure 1.2a montre la répartition du flux du disque Ry Lup [Langlois et al., 2018] en fonction de la longueur d'onde. On voit que la contribution de l'étoile est surtout présente dans les longueurs d'ondes micrométriques, et la contribution du disque est plus forte dans les longueurs d'ondes millimétriques.

En interférométrie, l'arrivée du réseau d'antenne ALMA qui fonctionne aux longueurs d'ondes millimétriques a alors pu permettre de sonder en profondeur les environnements circumstellaires. En effet, ALMA à la particularité de pouvoir sonder des poussières de taille millimétrique se situant dans le plan médian des disques circumstellaires, car les plus petits grains ne sont pas opaques à de telles longueurs d'ondes, et ainsi d'avoir accès à la répartition de la masse des disques.

En imagerie directe, l'arrivée du télescope spatial Hubble (HST : *Hubble Space Telescope*), a révolutionné l'astronomie moderne. Sont arrivés ensuite les très grands télescopes terrestres, tels que les Very Large Telescopes (VLT), le télescope Subaru ou le télescope Gemini. L'imagerie directe dans le proche infrarouge permet d'étudier la lumière réfléchie par les grains micrométriques situés à la surface du disque [Mulders et al., 2013], car le disque est opaque pour de telles longueurs d'ondes. Le problème des télescopes terrestres est que lors du passage dans l'atmosphère, la lumière des étoiles est entachée d'aberrations. Lors d'une observation cela se traduit en une perte de résolution. Le but de l'optique adaptative est de compenser les effets de l'atmosphère afin de restituer la tâche de diffraction de la lumière de l'étoile, produite par le télescope, exempte d'aberration. Grâce à leurs instruments d'optiques adaptatives extrêmes, ces très grands télescopes ont rendu possible la résolution des environnements à haut contraste.

Cependant, comme évoqué précédemment, la lumière des environnements circumstellaires est bien plus faible que la lumière de l'étoile hôte. La lumière de l'environnement est alors noyée dans la tâche de diffraction de la lumière stellaire, dont la brillance est mille à dix mille fois supérieure.

L'utilisation d'un coronographe permet, en reproduisant un effet semblable à celui d'une éclipse, d'atténuer la brillance de la tâche de diffraction d'un facteur 10 à 100. Les coronographes ne sont cependant pas parfait et il reste un résidu de lumière stellaire toujours plus brillant que le disque lui-même. Il est cependant possible, par différentes techniques d'observation dédiées, de démélanger les deux sources (disques, ou exoplanètes, et résidus stellaires). La figure 1.2b montre un extrait de données de l'instrument ESO/VLT-SPHERE IRDIS dans laquelle sont mélangées la partie disque et la partie résidus stellaires.

D'une part, pour la détection d'exoplanètes ou de disque, la technique de la différence angulaire d'image (ADI : *Angular Differential Imaging*) peut-être utilisée. Cette méthode utilise le fait que le résidu de lumière stellaire est fixe par rapport à l'instrument, tandis que l'environnement est fixe par rapport au ciel. Ainsi, en faisant tourner artificiellement le ciel à l'aide d'un ensemble de miroirs, appelé dérotateur, il est possible d'extraire l'environnement, qui a

tourné, du résidu de lumière stellaire, qui n'est pas affecté par la rotation. Il est alors possible d'avoir accès à la lumière émise par la surface du disque, mais qui est alors entachée d'artefacts liés à l'auto soustraction de la lumière lisse du disque. Cette technique d'observation est aussi peu efficace lorsque l'environnement est relativement invariant par rotation comme dans le cas d'un disque vu de face.

D'autre part, il y a la différence polarimétrique d'image (DPI : Differential Polarimetric Imaging) [van Holstein et al., 2020], qui est la technique utilisée dans cette thèse et qui permet d'avoir accès à la morphologie de disques sans artefacts. Nous détaillons dans la section suivante ce qu'est la polarimétrie et les méthodes d'observation en polarimétrie.

1.1.3 La polarimétrie en imagerie directe

1.1.3.1 Qu'est-ce que la polarimétrie

La lumière peut être vue comme une onde plane. Une onde plane est composée d'un champ électrique et d'un champ magnétique qui se propagent dans la même direction de propagation de manière proportionnelle. On appelle plan d'onde le plan $(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^2 \times \mathbb{R}^2$ orthogonal à la direction de propagation. La figure 1.3 est une illustration d'onde plane.

Le champ électrique se propage en oscillant à une certaine longueur d'onde. On appelle polarisation de la lumière, la façon dont ce champ électrique se comporte dans le plan d'onde en oscillant. La figure 1.4 montre ces différents états de polarisation.

On représente le champ électrique dans son plan d'onde par un vecteur complexe

$$\boldsymbol{e} = \boldsymbol{p} \times \mathrm{e}^{\mathrm{i}(\omega t - \kappa z)} \in \mathbb{C}^2 \tag{1.2}$$

où $\times$ représente le produit terme à terme des vecteurs, $z \in \mathbb{R}$ la position le long de la direction de propagation, $\kappa = 2\pi/\lambda$ avec $\lambda \in \mathbb{R}$ la longueur d'onde et $\omega = 2\pi c/\lambda$ où c est la vitesse de propagation de l'onde. Ce champ électrique oscille avec une certaine amplitude complexe $\boldsymbol{p} \in \mathbb{C}^2$, c'est son vecteur de polarisation. Ce vecteur de polarisation est celui qui va « tracer »la forme de la polarisation dans le plan d'onde.

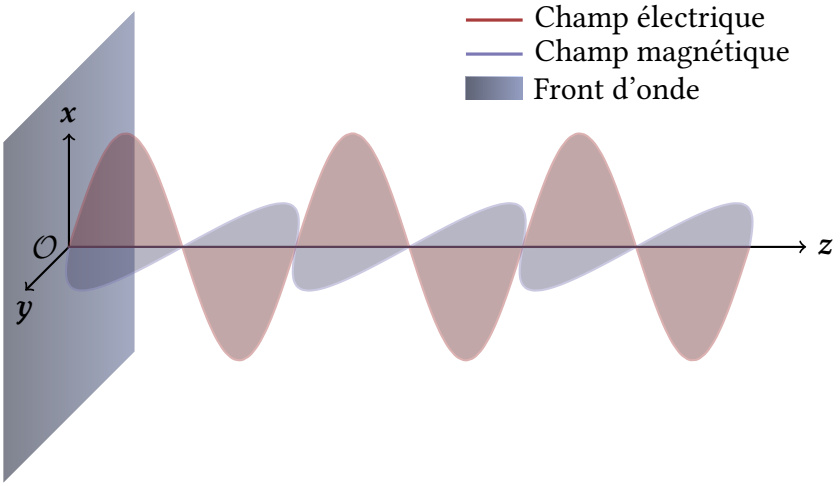


Figure 1.3 – Schéma d'une onde électromagnétique plane, se propageant dans la direction du vecteur $\boldsymbol{z}$.

Ce vecteur complexe est composé d'une phase $\vartheta \in]-\pi, \pi]^2$ et d'une amplitude $\boldsymbol{a} \in \mathbb{R}_+^2$ tel que :

$$\boldsymbol{p} = \boldsymbol{a} \times \mathrm{e}^{\mathrm{i}\vartheta} = \begin{pmatrix} \boldsymbol{a}^{(x)} \mathrm{e}^{\mathrm{i}\vartheta^{(x)}} \\ \boldsymbol{a}^{(y)} \mathrm{e}^{\mathrm{i}\vartheta^{(y)}} \end{pmatrix}, \tag{1.3}$$

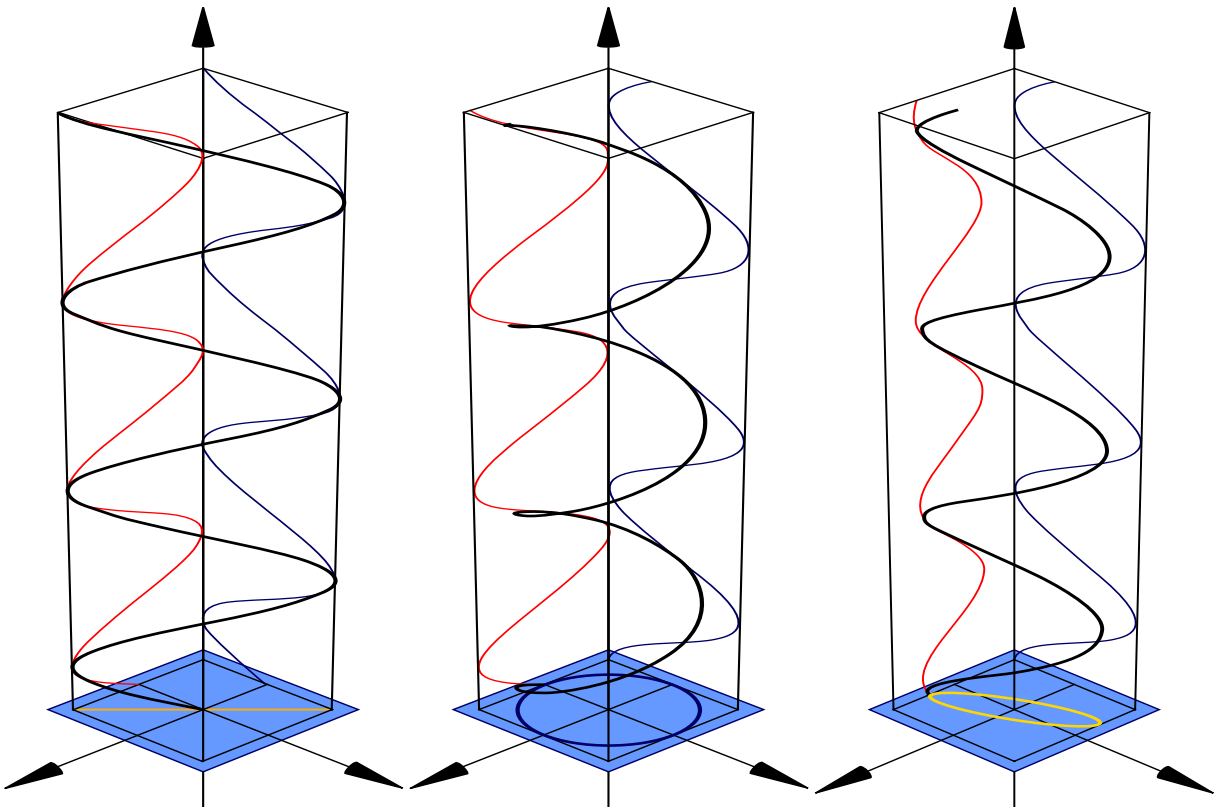
où $\times$ représente le produit terme à terme des vecteurs.

Lorsque les composantes de $\boldsymbol{a}$ et de ϑ sont quelconques, lors de la propagation de l'onde, le vecteur de polarisation trace une ellipse dans le plan d'onde, comme décrit sur la figure 1.4c. On parle alors de polarisation elliptique. Il existe deux cas particuliers à cette polarisation.

Le premier cas est quand la différence entre les deux composantes de la phase est de $\frac{\pi}{2} \pm \pi$, c'est-à-dire $\vartheta^{(x)} = \theta = \vartheta^{(y)} \pm \frac{\pi}{2}$, et que les deux composantes de l'amplitude sont identiques, c'est-à-dire $\boldsymbol{a}^{(x)} = \boldsymbol{a}^{(y)} = a$. Dans un tel cas, le vecteur de polarisation trace un cercle dans le plan d'onde, comme sur la figure 1.4b, on parle alors de polarisation circulaire. En effet, l'équation du vecteur de polarisation dans un tel cas est :

$$\boldsymbol{p} = a\mathrm{e}^{\mathrm{i}\vartheta} \begin{pmatrix} 1 \\ \pm\mathrm{i} \end{pmatrix}. \tag{1.4}$$

En calculant la position de ce vecteur en fonction de ϑ, on obtient bien l'équation d'un cercle de rayon a.



(a) Polarisation linéaire (b) Polarisation circulaire (c) Polarisation elliptique

FIGURE 1.4 – Les différents états de polarisation de la lumière (source : *Wikipedia*, figures libres de droit).

Le second cas est quand la différence entre les deux composantes de la phase est un multiple de π, c'est-à-dire $\vartheta^{(x)} = \theta = \vartheta^{(y)} \pm \pi$, et que les deux composantes de l'amplitude sont identiques, c'est-à-dire $\boldsymbol{a}^{(x)} = \boldsymbol{a}^{(y)} = a$. Le champ électrique oscille alors de manière rectiligne, comme sur la figure 1.4a et on parle alors de polarisation linéaire. Dans le cas de la polarisation linéaire, l'équation du vecteur de polarisation linéaire est :

$$\boldsymbol{p} = a\mathrm{e}^{\mathrm{i}\vartheta}\begin{pmatrix}1\\-1\end{pmatrix} = a\begin{pmatrix}\cos(\theta)\\\sin(\theta)\end{pmatrix} \tag{1.5}$$

Ce vecteur en coordonnées polaires défini donc une droite paramétrée par l'orientation θ. On appelle θ angle de polarisation linéaire. C'est l'angle que fait le vecteur par rapport à l'axe des abscisses dans le plan $(\boldsymbol{x}, \boldsymbol{y})$.

De manière géométrique, la polarisation linéaire est un cas où le petit rayon de l'ellipse est nul, tandis que la polarisation circulaire est le cas où le grand rayon et le petit rayon sont égaux.

Différents instruments permettent de polariser la lumière ou de transformer la polarisation de la lumière polarisée. Les polariseurs par absorption, sont généralement basés sur l'utilisation de cristaux présentant un dichroïsme. Ils servent à transformer toute lumière en une lumière polarisée linéairement, dont l'angle de polarisation linéaire est perpendiculaire à l'orientation du polariseur (orientation spécifique sous forme de micro fentes). L'utilisation de matériaux biréfringents permet de transformer ou moduler artificiellement les phases des vecteurs de polarisation. C'est le cas de la lame demi-onde, qui orientée d'un certain angle α, fait tourner le vecteur de polarisation de -2α. Une lame quart-d'onde va de plus retarder l'une des composantes de la phase de $\pi/2$, transformant ainsi la lumière polarisée linéairement en lumière polarisée circulaire et inversement. Une description mathématique de l'effet des outils optiques utilisés dans l'instrument ESO/VLT-SPHERE IRDIS sur le champ électrique sera faite dans la suite de ce chapitre.

1.1.3.2 La polarimétrie en astrophysique

Différentes sources de lumière polarisée ou partiellement polarisée peuvent être trouvées dans l'univers. En général, la lumière polarisée est le résultat de la diffusion de lumière non-polarisée sur certains matériaux. Son observation permet de nombreuses études en astrophysique, comme l'étude du milieu interstellaire et des régions de formation d'étoile dans les nébuleuses, l'étude du Fond Diffus Cosmologique (CMB : *Cosmic Microwave Background*) [Adam et al., 2016], ou encore dans notre cas, l'observation des environnements circumstellaires.

Les environnements circumstellaires sont proches d'une étoile hôte. La lumière d'une étoile est *non-polarisée*, c'est-à-dire qu'à un temps t donné, l'amplitude $\boldsymbol{a} \in \mathbb{R}^2$ et la phase $\vartheta \in]-\pi, \pi]^2$ de son vecteur de polarisation $\boldsymbol{p} \in \mathbb{C}^2$ sont des vecteurs aléatoires, de composantes respectivement indépendantes et identiquement distribuées. Cette lumière se propage dans toutes les directions et se diffuse sur la poussière.

Lors de sa diffusion par des grains de poussières, la lumière d'une étoile se polarise partiellement, c'est la diffusion de Rayleigh ou de Mie selon la taille des grains. Dans le cas des particules sphériques, la lumière diffusée perpendiculairement à l'angle d'incidence de la lumière est presque entièrement linéairement polarisée. De plus, l'angle de polarisation linéaire est alors perpendiculaire à l'angle d'incidence.

Les grains de poussière composant les environnements circumstellaires sont assimilés à des agrégats vaguement sphériques. L'étude de la lumière des environnements circumstellaires permet l'observation de particules dont la taille dépend de la longueur d'onde à laquelle on les observe [Mulders et al., 2013]. Comme évoqué plus tôt, il est alors possible avec l'interféromètre ALMA d'observer la lumière émise par les grains millimétriques présents en particulier dans le plan médian du disque. Aux longueurs d'ondes de l'infrarouge proche (NIR : *Near InfraRed*), c'est-à-dire entre $0,9\mu$m (micromètres) et $2,5\mu$m, les grains vont être du même ordre de grandeur. Il est possible en imagerie directe dans le proche infrarouge d'étudier la lumière réfléchie par ces grains micrométriques présents à la surface du disque. Les deux instruments permettant aujourd'hui ce type d'observations sont le Gemini Planer Imager (GPI) [Macintosh et al., 2014] et l'instrument Spectro-Polarimeter High-contrast Exoplanet REsearch (SPHERE) [Beuzit et al., 2019]. C'est sur les observations en polarimétrie de ce dernier qu'est basé le travail de ma thèse.

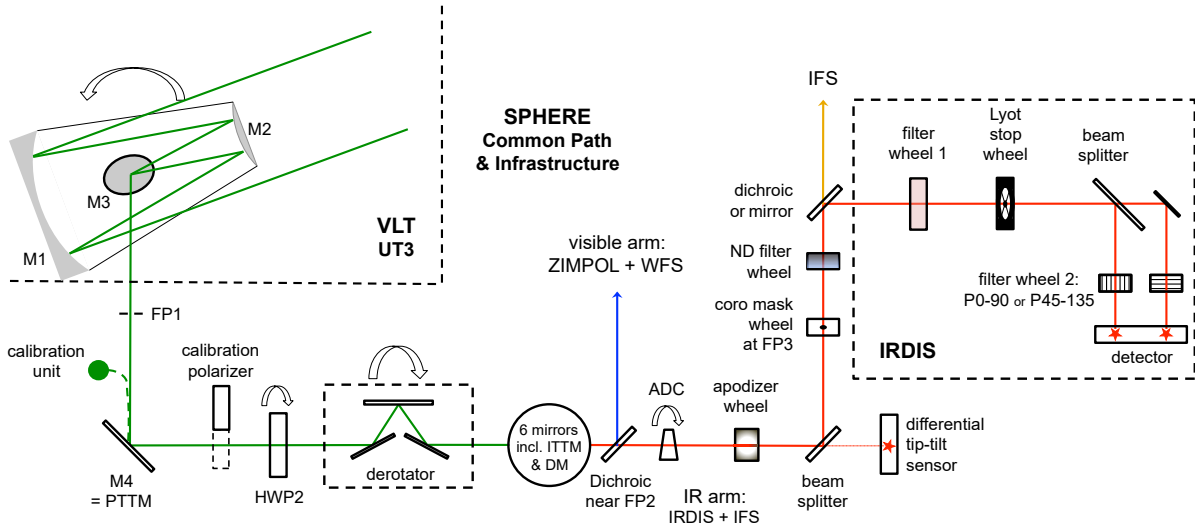
1.1.4 L'instrument ESO/VLT-SPHERE

L'instrument Spectro-Polarimeter High-contrast Exoplanet REsearch (SPHERE) [Beuzit et al., 2019] est un instrument de l'observatoire européen, European Southern Observatory (ESO). Il est placé sur l'une des deux plateformes Nasmyth du troisième des quatre très grands télescopes (VLT-UT3) situés dans le désert d'Atakama au Chili. Il est composé de trois sous systèmes instrumentaux permettant l'acquisition de données dans des modes différents. Ceux-ci sont visibles sur la photo 1.5a.

- Le spectrographe intégrale de champs (IFS : *Integral Field Spectrograph*)[Claudi et al., 2008], permet d'acquérir des images de cibles astrophysiques à différentes longueurs d'ondes, dans le domaine de l'infra-rouge proche. Il est utilisé pour la détection d'exoplanètes [Flasseur et al., 2018], pour l'étude d'objets distants du système solaire comme les astéroïdes [Berdeu et al., 2020] mais aussi pour l'étude des environnements circumstellaires [Pairet et al., 2018a, Pairet et al., 2019].
- Le polarimètre ZIMPOL, pour Zürich IMaging Polarimeter) [Schmid et al., 2018] permet une observation en polarimétrie et en imagerie à des longueurs d'ondes du visible entre $0,5\mu$m et $0,9\mu$m. Il est utilisé pour l'étude des environnements circumstellaires et l'étude d'objets distants de notre système solaire.
- Le spectrographe à image duale IRDIS (*InfraRed Dual-Band Imaging and Spectroscopy*), permet à la fois une imagerie à différentes longueurs d'ondes, et une imagerie en polarimétrie. Il permet, grâce à son mode double-bande, d'obtenir simultanément deux images à deux longueurs d'ondes proches ou deux polarisations différentes. Ce sont respectivement les modes Dual Band Imaging (DBI) [Vigan et al., 2014] et Dual Polarimetry Imaging (DPI) [Langlois et al., 2014, de Boer et al., 2020]. Ces modes sont utilisés pour l'étude des environnements circumstellaires, la détection d'exoplanètes, mais aussi l'étude d'objets de notre système solaire. Il possède également un mode longue fente [Vigan et al., 2010]. Ce mode est particulièrement utilisé pour la caractérisation d'exoplanètes en moyenne résolution spectrale.

C'est sur les données DPI de ce dernier instrument qu'est basée ma thèse. Mon but étant de modéliser l'action de l'instrument sur le champ électrique de la lumière, il est nécessaire de comprendre comment fonctionne l'instrument SPHERE/IRDIS. La figure 1.5b représente le cheminement de la lumière à travers l'instrument.

(a) Photo de l'instrument ESO/VLT-SPHERE sur la plateforme Nasmyth du VLT-UT3 (crédits : Maud Langlois).



(b) Schéma de l'instrument ESO/VLT-SPHERE/IRDIS (source : [de Boer et al., 2020]).

Figure 1.5 – Le spectro-polarimètre SPHERE.

Lors de son arrivée sur le télescope, la lumière est réfléchie par le miroir primaire M1, puis par le miroir secondaire M2 et enfin le miroir M3 qui envoie la lumière vers les instruments d'une des deux plateformes Nasmyth, dont SPHERE, ou, si le miroir M3 n'est pas utilisé, directement au foyer Cassegrain du télescope.

Une fois dans l'instrument SPHERE, la lumière est réfléchie par un quatrième miroir M4. Elle passe alors par un polariseur, qui peut être activé ou désactivé, dont l'utilité lors de son activation est de calibrer la polarisation instrumentale [van Holstein et al., 2020]. En mode polarimétrique, la lumière passe alors par une lame demi-onde qui, orientée d'un angle α, fait tourner la polarisation de la lumière de 2α. Cette lame demi-onde est tournée au long des acquisitions selon le cycle d'angle $\alpha \in \{0°, 45°, 22,5°, 67,5°\}$, correspondant à une rotation de la polarisation respective d'angle $2\alpha \in \{0°, 90°, 45°, 135°\}$.

La lumière résultante passe alors par le dérotateur : un ensemble de trois miroirs qui tourne en continu lors de l'acquisition de la lumière. Le but de cet ensemble de miroirs est, soit de compenser la rotation du ciel au cours de la séquence d'observation, on parle alors de mode champs stabilisé, soit de compenser la rotation de la pupille lors du suivi de la cible par le télescope, on parle alors de mode pupille stabilisée. Le premier cas correspond à ce qui est utilisé pour le mode DPI et le second cas pour le mode ADI. La perspective actuelle étant de combiner DPI et ADI et donc d'extraire des mesures polarimétriques en mode pupille stabilisée. La lumière passe ensuite à travers un système d'optique adaptative extrême [Fusco et al., 2006, Petit et al., 2012], qui permet de compenser la turbulence atmosphérique à partir d'étoiles de références. La lumière est alors séparée en lumière visible, qui est envoyée vers ZIMPOL, et proche infrarouge, qui est envoyée vers l'IFS et l'instrument IRDIS. Avant d'entrer dans les instruments IFS et IRDIS, la lumière passe par un apodiseur, qui n'est utilisée qu'avec le coronographe de Lyot dans IRDIS pour apodiser le motif de diffraction des objets brillants, puis à travers un masque coronographique [Carbillet et al., 2011] situé dans le plan focal. Les instruments IFS et IRDIS peuvent être utilisés ensemble ou séparément. Dans le premier cas, la lumière est alors séparée par une lame dichroïque. Dans le second cas, la lumière est envoyée vers l'une ou l'autre des unités à l'aide d'un miroir.

Une fois arrivée dans l'instrument IRDIS, la lumière passe à travers une roue à filtre permettant l'observation dans les bandes Y (de $0,97\mu$m à $1,11\mu$m), J (de $1,12\mu$m à $1,37\mu$m), H (de $1,48\mu$m à $1,77\mu$m) et K_S (de $2,03\mu$m à $2,33\mu$m). La lumière passe ensuite à travers le stop de Lyot qui, allié à l'apodiseur et au masque, permet de réduire l'effet de diffraction des objets brillants d'un facteur 10 à 100 selon la longueur d'onde.

Ensuite, la lumière est séparée en deux par une lame séparatrice. Les deux faisceaux lumineux résultants sont envoyés simultanément à travers une seconde roue à filtre qui peut être activée pour le mode DBI ou bien utilisée avec deux analyseurs croisés pour le mode DPI. Ces analyseurs sont constitués de deux polariseurs orientés respectivement d'un angle de 0° et de 90°. Le rôle de ces polariseurs est de filtrer la lumière dont la polarisation n'est pas alignée avec eux.

Enfin les deux faisceaux de lumière sont imagés sur un même détecteur infrarouge, sur une zone de 1024×2048 pixels, chaque faisceau étant projeté sur une moitié de 1024×1024 pixels. On parle de partie gauche et de partie droite du détecteur pour les faisceaux arrivant respectivement des analyseurs orientés de 0° et de 90°. En éclairant les différentes parties du détecteur, les photons composant la lumière excitent des couches d'électrons. Le nombre d'électrons excités en ce pixel est alors enregistré. Il est proportionnel au nombre de photons ayant atteint ce pixel. Le facteur de proportionnalité entre le nombre de photons incidents et

le nombre d'électrons enregistrés est appelé rendement quantique. Ce nombre d'électrons est ensuite multiplié par un autre facteur, le gain du détecteur, permettant de convertir un nombre d'électrons en valeur numérique.

Finalement, après une séquence d'observation avec l'instrument ESO/VLT-SPHERE IRDIS en mode polarimétrie, on acquière un jeu de données dans lequel la lame demi-onde a fait plusieurs cycles. Les images sont composées de deux sous-images gauche et droite, avec une polarisation différente. Lors des modulations de la polarisation dans l'instrument, les résidus de lumière stellaire étant non-polarisés reste inchangés, tandis que la lumière polarisée linéairement du disque tourne. C'est grâce à cette variété qu'il est possible de démélanger les deux sources.

1.1.4.1 Calibration et prétraitement des données

Les données brutes acquises par le détecteur de l'instrument IRDIS ne sont pas exploitables telles quelles. Tout d'abord, elles sont polluées par le fond thermique de l'instrument et du ciel, qui rayonne plus ou moins fort dans le proche infrarouge selon la longueur d'onde. De plus, le détecteur n'est pas éclairé de manière uniforme, c'est-à-dire que les parties gauche et droite du détecteur sont éclairées différemment. Enfin, un certain nombre de pixels sont défectueux et polluent les données.

Dans l'état-de-l'art, les données sont *calibrées* et *pré-traitées* à partir du *pipeline* de l'instrument [Pavlov et al., 2008]. D'une part, un fond moyen est estimé à partir de données de ciel sans objet d'intérêt. D'autre part, on calcule le champ plat, c'est-à-dire l'éclairement du détecteur en chaque pixel, à partir de données où le détecteur est éclairé par une lampe uniforme spatialement, interne à l'instrument. Enfin, une carte de pixels morts ou défectueux est calculée. Les données *calibrées* sont alors obtenues par soustraction du fond et division par le champ plat.

Les acquisitions sont triées afin d'enlever celles qui ne peuvent être exploités, à cause par exemple de la turbulence atmosphérique ou du bougé de l'instrument. Les cycles de lame demi-onde contenant ces données sont alors mis de côté et ne sont pas inclus dans les cubes de données *pré-traitées*.

Ces cubes sont alors obtenues par interpolation des mauvais pixels, découpage des images gauche et droite et recentrage des images, afin que les centres des étoiles hôtes, calculés à partir de données spécifiques, soient à la même position sur les images gauche et droite. Cela est nécessaire pour pouvoir utiliser les méthodes de l'état-de-l'art.

Dans le cas où le champ d'observation est tourné d'un angle artificiel, il est nécessaire de tourner les données avant ou après la reconstruction.

Il faut noter que d'autres éléments entrent en compte dans la calibration des données, tels que le bruit de lecture, le gain et le rendement quantique en chaque pixel du détecteur. Cependant, ces calibrations ne sont pas prises en compte par le pipeline, ou alors de manière uniforme pour tout le détecteur. Dans le chapitre 6, je présente une façon d'estimer conjointement les paramètres de calibration des données basée sur l'approche inverse.

1.1.5 Méthodes de reconstruction en imagerie directe polarimétrique

Le but des méthodes de reconstruction en imagerie directe polarimétrique est, à partir d'un jeu de données, de démélanger la lumière non-polarisée du résidu stellaire et du disque et la

lumière polarisée du disque, sous forme d'intensité. On notera tout au long de cette thèse respectivement I^{u} et I^{p} pour les intensités, ainsi que θ l'angle de polarisation linéaire du disque. Avant d'étudier les méthodes de démélange de la lumière polarisée I^{p} de la lumière non-polarisée I^{u} à partir des données de l'instrument ESO-VLT SPHERE, il est nécessaire d'énoncer les différents formalismes permettant de représenter l'action des différents éléments optiques de l'instrument sur la polarisation de la lumière qui le traverse.

1.1.5.1 Les différents formalismes

Il existe deux formalismes pour décrire les signaux polarisés et leurs transformations, soit sous forme de champ électrique, soit sous forme d'intensité. Le premier est le formalisme de Jones. Il représente les transformations de la polarisation par les éléments optiques au niveau du champ électrique dans $\mathbb{C}^2$. Le second est le formalisme de Mueller et représente les transformations de la polarisation par les éléments optiques sur les paramètres de Stokes dans $\mathbb{R}^4$, qui correspondent à des intensités. L'intensité d'un champ électrique quelconque $\boldsymbol{e} \in \mathbb{R}^2$ est donné par l'espérance temporelle de son module carré, c'est-à-dire :

$$\mathcal{I} = \mathbb{E}_t\left[|\boldsymbol{e}|^2\right]. \tag{1.6}$$

Il est donc possible de passer d'un formalisme à l'autre en passant du champ électrique à l'intensité en utilisant la relation (1.6) ci-dessus.

Dans le cas du formalisme de Jones, qui est plus intuitif, le changement de polarisation du champ électrique induit par un élément optique est représenté par une matrice de $\mathcal{M}_2(\mathbb{C})$, c'est-à-dire une matrice complexe de taille 2×2. Cette matrice est encadrée de part et d'autre par des matrices de rotation réelles permettant de se placer dans le repère de l'optique pour effectuer la transformation, puis de revenir dans le repère initial.

Chaque élément optique dans l'instrument atténue l'amplitude du vecteur de polarisation d'un facteur $\varepsilon \geq 0$ et retarde la phase d'un angle $\delta \in]-\pi, \pi]$. Pour tous éléments optiques ne transformant pas la polarisation dans l'instrument, comme les miroirs, les lames séparatrice et dichroïques, l'idéal serait que l'atténuation et le retard soient nuls, pour ne pas induire de polarisation instrumentale, mais ce n'est jamais le cas. On appelle alors polarisation instrumentale toutes ces transformations non souhaitées de la polarisation. Il est possible de la calibrer à l'aide de données de calibration [van Holstein et al., 2020].

La matrice de Jones d'un élément optique appliquant une atténuation ε de l'amplitude et un retard δ de la phase est donnée par :

$$\mathbf{J}_{\varepsilon,\delta} = \begin{pmatrix} 1 & 0 \\ 0 & \varepsilon \mathrm{e}^{-\mathrm{i}\delta} \end{pmatrix}. \tag{1.7}$$

La matrice de rotation associée, pour un instrument orienté d'un angle φ par rapport au repère initial est donnée par :

$$\mathbf{R}_{\varphi} = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix}, \tag{1.8}$$

où l'axe principal du repère initial est choisi comme l'axe Sud-Nord. Dans le cas de la lame demi-onde, on a alors $\varepsilon = 1$ et $\delta = \pi$ tandis que l'angle φ correspond aux positions $\alpha \in \{0°, 45°, 22,5°, 67,5°\}$ que prend la lame demi-onde, ce qui donne

$$\mathbf{J}_{\alpha} = \mathbf{R}_{\alpha}^{\top} \mathbf{J}_{1,\pi} \mathbf{R}_{\alpha} = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix} = \begin{pmatrix} \cos 2\alpha & \sin 2\alpha \\ \sin 2\alpha & -\cos 2\alpha \end{pmatrix}, \tag{1.9}$$

où $^\top$ représente la transposée de la matrice. Pour les analyseurs croisés, cela correspond au cas où $\varepsilon = 0$ et où φ vaut respectivement 0° et 90°.

Par ailleurs, on remarque que le rang de la matrice de Jones dans le cas des analyseurs est réduit à 1. Il est donc possible, plutôt que de représenter le passage par les analyseurs par un produit matrice-vecteur, de le représenter par le produit scalaire Hermitien entre le vecteur représentant l'analyseur et le vecteur du champ électrique avant passage dans les analyseurs. Les deux méthodes sont équivalentes et cela évite le passage par la norme du champ électrique lors du passage à l'intensité. On exprime alors l'analyseur orienté d'un angle ψ, par le vecteur de Jones suivant :

$$\boldsymbol{a}_\psi = \begin{pmatrix} \cos(\psi) \\ -\sin(\psi) \end{pmatrix}. \tag{1.10}$$

En effet, le vecteur de Jones d'un analyseur aligné sur l'axe principal est donné par le vecteur :

$$\boldsymbol{a} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

De ce fait, si l'analyseur est orienté d'un angle ψ par rapport à l'axe principal, cela revient à appliquer la matrice de rotation d'angle ψ à ce vecteur, soit :

$$\boldsymbol{a}_\psi = \begin{pmatrix} \cos(\psi) & \sin(\psi) \\ -\sin(\psi) & \cos(\psi) \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \cos(\psi) \\ -\sin(\psi) \end{pmatrix}.$$

Dans le cas du formalisme de Mueller, on ne travaille plus sur le champ électrique mais sur les paramètres de Stokes I, Q, U et V. Ces paramètres sont liés au champ électrique $\boldsymbol{e} \in \mathbb{C}^2$ par la relation suivante :

$$\begin{cases} \mathrm{I} = |\boldsymbol{e}^{(x)}|^2 + |\boldsymbol{e}^{(y)}|^2 \\ \mathrm{Q} = |\boldsymbol{e}^{(x)}|^2 - |\boldsymbol{e}^{(y)}|^2 \\ \mathrm{U} = 2\mathfrak{Re}\left(\boldsymbol{e}^{(x)}\overline{\boldsymbol{e}^{(y)}}\right) \\ \mathrm{V} = -2\mathfrak{Im}\left(\boldsymbol{e}^{(x)}\overline{\boldsymbol{e}^{(y)}}\right), \end{cases} \tag{1.11}$$

où $\bar{\cdot}$ représente le conjugué d'un nombre complexe et $\mathfrak{Re}, \mathfrak{Im}$ représentent respectivement la partie réelle et la partie imaginaire d'un nombre complexe. Le paramètre I représente l'intensité totale, polarisée et non-polarisée de la lumière. Les paramètres Q et U représentent respectivement les intensités polarisées linéairement horizontale et verticale, c'est-à-dire projetée dans le repère principal et projetée dans un repère tourné de 45°. Enfin le paramètre V représente la polarisation circulaire.

L'influence du changement de polarisation sur les paramètres de Stokes est modélisée par une matrice de $\mathcal{M}_4(\mathbb{R})$, c'est-à-dire une matrice réelle de taille 4×4.

La matrice de Mueller d'un élément optique induisant une atténuation ε de l'amplitude et un retard δ de la phase est donnée par :

$$\boldsymbol{M}_{\varepsilon,\delta} = \frac{1}{2} \begin{pmatrix} 1+\varepsilon^2 & 1-\varepsilon^2 & 0 & 0 \\ 1-\varepsilon^2 & 1+\varepsilon^2 & 0 & 0 \\ 0 & 0 & 2\varepsilon\cos\delta & 2\varepsilon\sin\delta \\ 0 & 0 & -2\varepsilon\sin\delta & 2\varepsilon\cos\delta \end{pmatrix}. \tag{1.12}$$

La matrice de rotation associée selon l'orientation φ de l'élément optique est donnée par :

$$\mathcal{R}_{\varphi} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2\varphi & \sin 2\varphi & 0 \\ 0 & -\sin 2\varphi & \cos 2\varphi & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{1.13}$$

Le passage d'un formalisme à l'autre consiste seulement, à partir du formalisme de Jones en calculer :

$$\boldsymbol{M}_{\varepsilon,\delta} = \boldsymbol{A}\left(\boldsymbol{J}_{\varepsilon,\delta} \otimes \boldsymbol{J}^{*}_{\varepsilon,\delta}\right)\boldsymbol{A}^{-1}, \tag{1.14}$$

où $\boldsymbol{J}_{\varepsilon,\delta}$ est la matrice de Jones donnée par l'équation (1.7), $\otimes$ représente le produit de Kronecker, * représente l'adjoint de la matrice et

$$\boldsymbol{A} = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 0 \\ 0 & -\mathrm{i} & \mathrm{i} & 0 \end{pmatrix}. \tag{1.15}$$

Dans la littérature, le formalisme de Mueller sur les paramètres de Stokes est généralement favorisé par rapport au formalisme de Jones dans le cas d'étude d'environnements partiellement polarisés. C'est pourquoi pour l'observation à haut-contraste des environnements circumstellaires, c'est ce formalisme qui est utilisé.

Dans le cas de l'instrument ESO/VLT-SPHERE IRDIS, la polarisation circulaire V ne peut pas être mesurée, car l'instrument ne dispose pas d'une lame quart-d'onde. Dans la suite on ne parle alors plus que des paramètres I, Q et U.

1.1.5.2 Les méthodes de l'état-de-l'art

Rappelons que notre but est, à partir des données en polarimétrie de l'instrument ESO-VLT SPHERE IRDIS, de démélanger la lumière polarisée linéairement de l'environnement de la lumière non-polarisée de l'étoile et de l'environnement. Ce que l'on souhaite obtenir à la fin, sont des cartes d'intensité non-polarisée de l'étoile et de l'environnement I^{u}, d'intensité polarisée linéairement de l'environnement I^{p} et d'angles de polarisation linéaire θ. Ces cartes sont composées de N pixels et on note $n \in \{1, \dots N\}$ l'indice du pixel de la carte. Les paramètres de Stokes I, Q et U sont liés à ces quantités par les relations, pour tout pixel $n \in \{1, \dots N\}$:

$$\begin{cases} \mathrm{I}_n = I^{\mathrm{u}}_n + I^{\mathrm{p}}_n \\ \mathrm{Q}_n = I^{\mathrm{p}}_n \cos 2\theta_n \\ \mathrm{U}_n = I^{\mathrm{p}}_n \sin 2\theta_n \end{cases} \tag{1.16}$$

Le but des méthodes de l'état-de-l'art est de retrouver les cartes de paramètres I, Q et U par combinaisons des intensités acquises par le détecteur. Pour comprendre leur fonctionnement physique, il faut comprendre comment l'intensité et l'angle de polarisation sont modifiés par les éléments optiques principaux de l'instrument, comme la lame demi-onde et les analyseurs.

Tout d'abord, regardons ce qu'il se passe lors du passage de la lumière polarisée linéairement par la lame demi-onde. Soit un champ électrique $\boldsymbol{e}_n$ polarisé linéairement, qui va atteindre

le détecteur en le pixel $n \in \{1,\dots N\}$, de vecteur de polarisation $\boldsymbol{p}_n$ d'amplitude p_n et d'angle de polarisation θ_n. Son intensité est donnée par

$$I_n^{\mathrm{p}} = \mathbb{E}_t\left[|\boldsymbol{e}_n|^2\right] = \mathbb{E}_t[(p_n\cos\theta_n)^2 + (p_n\sin\theta_n)^2] = \mathbb{E}_t[p_n^2] = p_n^2, \tag{1.17}$$

car son amplitude est constante par rapport au temps. On voit par ailleurs que pour de la lumière polarisée linéairement, l'intensité est le carré de l'amplitude. Lors du passage par la lame demi-onde orientée d'un angle α, alors l'angle de polarisation se retrouve retardé de 2α. En effet, le vecteur de polarisation du champ électrique $\boldsymbol{e}_n^{\mathrm{out}}$ sortant est donné par :

$$\boldsymbol{p}_n^{\mathrm{out}} = \begin{pmatrix}\cos(2\alpha) & \sin(2\alpha)\\ \sin(2\alpha) & -\cos(2\alpha)\end{pmatrix}\begin{pmatrix}p_n\cos\theta_n\\ p_n\sin\theta_n\end{pmatrix} = \begin{pmatrix}p_n\cos(\theta_n-2\alpha)\\ p_n\sin(\theta_n-2\alpha)\end{pmatrix}.$$

Le champ électrique résultant est donc polarisé linéairement avec un angle de polarisation linéaire de $\theta_n - 2\alpha$. De ce fait, l'intensité associée est le carré de l'amplitude soit I_n^{p} également. Lors du passage de la lumière linéairement polarisée par la lame demi-onde, son angle de polarisation tourne d'un angle de 2α tandis que son intensité reste inchangée.

Regardons maintenant comment se comporte la lumière non-polarisée lors du passage par la lame demi-onde. La lumière est non-polarisée signifie qu'elle possède à chaque temps $t > 0$ une polarisation aléatoire, de vecteur d'amplitude et de vecteur de phase aux composantes aléatoires, centrées, indépendantes et identiquement distribuées. Cet aléa peut se transcrire de manière simplifiée par le fait qu'elle possède toutes les orientations de polarisation linéaire possibles, c'est-à-dire qu'à un temps $t > 0$ donné, la polarisation non-linéaire peut s'écrire comme un vecteur de polarisation linéaire aléatoire $\boldsymbol{u}_{t,n}$, qui atteint le détecteur en un pixel $1 \in \{n,\dots N\}$ d'amplitude aléatoire $u_{t,n} > 0$, centrée et de variance $\sigma_{u_n}^2$, et d'angle de polarisation aléatoire $\vartheta_{t,n} \in]-\pi,\pi]$. On a alors d'après 1.17, que l'intensité de la lumière non-polarisée est donnée par $I_n^{\mathrm{u}} = \mathbb{E}_t[u_{t,}^2] = \sigma_{u_n}^2$. On a après passage dans la lame demi-onde au temps $t > 0$:

$$\boldsymbol{u}_{t,n}^{\mathrm{out}} = \begin{pmatrix}u_{t,n}\cos(\vartheta_{t,n}-2\alpha)\\ u_{t,n}\sin(\vartheta_{t,n}-2\alpha)\end{pmatrix},$$

et donc d'après 1.17,$I_n^{\mathrm{uout}} = \sigma_{u,n}^2$. La lame demi-onde n'affecte donc pas l'intensité non-polarisée. De plus comme $\vartheta_{t,n}$ prend toutes les valeurs comprises entre $]-\pi,\pi]$ de manière identiquement distribuée, on comprend bien que le retard de 2α induit par la lame demi-onde ne va avoir aucune influence sur la moyenne temporelle des angles.

Concentrons-nous maintenant sur l'influence des analyseurs sur la polarisation de la lumière. L'action d'un polariseur sur l'intensité de la lumière polarisée linéairement est donné par la loi de Malus suivante :

Définition 1.1.1. *Loi de Malus : Soit une onde polarisée linéairement, d'amplitude p et d'angle de polarisation θ, passant par un polariseur parfait orienté d'un angle ψ. Si on note I^p l'intensité de cette onde, l'intensité transmise par le polariseur est donnée par :*

$$I^{pout} = I^p\cos^2(\theta-\psi). \tag{1.18}$$

Démonstration : *(Fin page 23)* Soit $\boldsymbol{e}$ le champ électrique d'une onde polarisée linéairement défini comme précédemment. En passant à travers un polariseur parfait orienté d'un angle ψ, le vecteur de polarisation $\boldsymbol{p}^{\mathrm{out}}$ du champ électrique $\boldsymbol{e}^{\mathrm{out}}$ résultant est donné par :

$$\boldsymbol{p}^{\text{out}} = \begin{pmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} \cos\psi & \sin\psi \\ -\sin\psi & \cos\psi \end{pmatrix}\begin{pmatrix} p\cos\theta \\ p\sin\theta \end{pmatrix} = \begin{pmatrix} p\cos(\theta-\psi)\cos(\psi) \\ p\cos(\theta-\psi)\sin(\psi) \end{pmatrix}.$$

En prenant l'espérance temporelle du module carré du champ électrique $\boldsymbol{e}^{\text{out}}$ pour avoir l'intensité résultante, on obtient que :

$$I^{\text{pout}} = \mathbb{E}_t\Big[|\boldsymbol{e}^{\text{out}}|^2\Big] = \mathbb{E}_t\Big[\,(p\cos(\theta-\psi)\cos(\psi))^2 + (p\cos(\theta-\psi)\sin(\psi))^2\,\Big] = p^2\cos^2(\theta-\psi).$$

car p et θ sont fixes dans le temps. Or d'après 1.17 $p^2 = I^{\text{p}}$ d'où

$$I^{\text{pout}} = I^{\text{p}}\cos^2(\theta-\psi).$$

□

La loi de Malus donne directement l'expression de la valeur de l'intensité polarisée à la sortie de l'analyseur. Étudions maintenant le comportement de l'intensité non-polarisée au travers de l'analyseur. En appliquant la loi de Malus dans le cas de la lumière non-polarisée, nous avons alors la proposition suivante :

Proposition 1.1.2. *Soit une onde de lumière non-polarisée d'intensité I^u entrante dans un polariseur. Alors, l'intensité sortante de l'analyseur $I^{uout} = I^u/2$.*

Démonstration : *(Fin page 23)* En effet, soit $\boldsymbol{u}_t$ le vecteur de polarisation du champ électrique de l'onde de lumière non-polarisée, d'intensité I^{u}, d'amplitude aléatoire $u_t > 0$, centrée et de variance $\sigma_u^2 = I^{\text{u}}$, et d'angle de polarisation aléatoire $\vartheta_t \in]-\pi,\pi]$ de loi uniforme sur $]-\pi,\pi]$.

On a alors d'après la loin de Malus :

$$I^{\text{uout}} = \mathbb{E}_t\Big[u_t^2\cos^2(\vartheta_t-\psi)\Big] \tag{1.19}$$

Par indépendance de u_t et ϑ_t on a :

$$\begin{aligned}
I^{\text{uout}} &= \mathbb{E}_t\Big[u_t^2\Big]\mathbb{E}_t\Big[\cos^2(\vartheta_t-\psi)\Big] \\
&= I^{\text{u}}\mathbb{E}_t\Big[\cos^2(\vartheta_t-\psi)\Big] \\
&= \frac{I^{\text{u}}}{2\pi}\int_{-\pi}^{\pi}\cos^2(\vartheta_t-\psi)\,\mathrm{d}\vartheta_t \\
&= \frac{I^{\text{u}}}{2\pi}\int_{-\pi}^{\pi}\cos^2(\vartheta_t)\,\mathrm{d}\vartheta_t \quad \text{par invariance de la loi uniforme par rotation sur }]-\pi,\pi] \\
&= \frac{2I^{\text{u}}}{\pi}\int_{0}^{\pi/2}\cos^2(\vartheta_t)\,\mathrm{d}\vartheta_t \\
&= \frac{I^{\text{u}}}{\pi}\int_{0}^{\pi/2}\cos(2\vartheta_t)+1\,\mathrm{d}\vartheta_t \\
&= \frac{I^{\text{u}}}{\pi}\left[\frac{\sin(2\vartheta_t)}{2}+\vartheta_t\right]_0^{\pi/2} = \frac{I^{\text{u}}\pi}{2\pi} = \frac{I^{\text{u}}}{2}.
\end{aligned}$$

□

Finalement on a la propriété suivante :

Proposition 1.1.3. *Soit une onde de lumière partiellement polarisée, d'intensité $\mathcal{I} = I^u + I^p$, passant à travers une lame demi-onde orientée d'un angle α puis à travers un analyseur orienté d'un angle ψ, alors l'intensité sortante est donnée par :*

$$\mathcal{I}_{\alpha,\psi} = \frac{I^u}{2} + I^p \cos^2(\theta - (2\alpha + \psi)). \tag{1.20}$$

Démonstration : *(Fin page 24)* En effet, après passage dans la lame demi onde, tous les angles de la lame demi-onde sont retardés de 2α. En remplaçant alors dans la loi de Malus θ par $\theta - 2\alpha$ et ϑ_t par $\vartheta_t - 2\alpha$, on obtient que $\mathcal{I}^{\text{out}}_{\alpha,\psi} = I^{\text{uout}}_{\alpha,\psi} + I^{\text{pout}}_{\alpha,\psi} = I^{\text{u}}/2 + I^{\text{p}} \cos^2(\theta - (2\alpha + \psi))$. □

Comme évoqué plus tôt, lors de l'acquisition des données, la lame demi-onde est positionnée selon le cycle d'angles $\alpha \in \{0°, 45°, 22,5°, 67,5°\}$. La lumière passe ensuite par deux analyseurs croisés, d'angles $\psi \in \{0°, 90°\}$.

Pour un pixel $n \in \{1, \dots N\}$, les intensités atteignant le détecteur pour chaque combinaison de position de lame demi-onde, d'après l'équation 1.20, sont les suivantes : Pour $\alpha = 0°$, on a :

$$\begin{cases} \mathcal{I}_{n,0°,0°} = \frac{I^{\text{u}}_n}{2} + I^{\text{p}}_n \cos^2(\theta_n) = \frac{1}{2}(I^{\text{u}}_n + I^{\text{p}}_n + I^{\text{p}}_n \cos 2\theta_n) = \frac{1}{2}(\text{I}_n + \text{Q}_n), \\ \mathcal{I}_{n,0°,90°} = \frac{I^{\text{u}}_n}{2} + I^{\text{p}}_n \sin^2(\theta_n) = \frac{1}{2}(\text{I}_n - \text{Q}_n) \end{cases} \tag{1.21}$$

De même, pour les acquisitions où $\alpha = 45°$, on a :

$$\begin{cases} \mathcal{I}_{n,45°,0°} = \frac{1}{2}(\text{I}_n - \text{Q}_n), \\ \mathcal{I}_{n,45°,90°} = \frac{1}{2}(\text{I}_n + \text{Q}_n). \end{cases} \tag{1.22}$$

Pour les acquisitions où $\alpha = 22,5°$, on a :

$$\begin{cases} \mathcal{I}_{n,22,5°,0°} = \frac{I^{\text{u}}_n}{2} + I^{\text{p}}_n \cos^2(\theta_n + 45°) = \frac{1}{2}(I^{\text{u}}_n + I^{\text{p}}_n - I^{\text{p}}_n \sin 2\theta_n) = \frac{1}{2}(\text{I}_n - \text{U}_n), \\ \mathcal{I}_{n,22,5°,90°} = \frac{1}{2}(\text{I}_n + \text{U}_n). \end{cases} \tag{1.23}$$

Enfin, pour les acquisitions où $\alpha = 67,5°$, on a :

$$\begin{cases} \mathcal{I}_{n,67,5°,0°} = \frac{1}{2}(\text{I}_n + \text{U}_n), \\ \mathcal{I}_{n,67,5°,90°} = \frac{1}{2}(\text{I}_n - \text{U}_n). \end{cases} \tag{1.24}$$

Il est important de noter que l'expression des intensités $\mathcal{I}^{\text{out}}_{\alpha,\psi}$ ne correspond pas à l'intensité enregistrée par le détecteur, mais à l'intensité atteignant le détecteur. En effet, comme évoqué précédemment, la quantité enregistrée par le détecteur n'est qu'une proportion de l'espérance temporelle du nombre photo-électrons enregistrés, lié au rendement du détecteur. Cette espérance temporelle est aléatoire, ce qui n'est pas pris en compte ici. Ne sont également pas pris en compte le bruit de lecture du détecteur ainsi que la polarisation instrumentale que pourraient induire les éléments optiques de l'instrument, autres que la lame demi-onde et que les analyseurs.

La méthode la plus utilisée dans l'état-de-l'art, qui est la méthode de la double différence [Tinbergen, 2005], ne tient pas compte de ces paramètres et suppose que l'intensité acquise par le détecteur après calibration pour un pixel $n \in \{1, \dots N\}$ est $\mathcal{I}^{\text{out}}_{n,\alpha,\psi}$. Elle consiste en additionner ou soustraire ces intensités afin de retrouver les paramètres de Stokes. Pour un cycle de lame demi-onde, on a :

$$\begin{cases} \mathrm{I}_n = \frac{\mathcal{I}_{n,0°,0°} + \mathcal{I}_{n,0°,90°} + \mathcal{I}_{n,45°,0°} + \mathcal{I}_{n,45°,90°} + \mathcal{I}_{n,22,5°,0°} + \mathcal{I}_{n,22,5°,90°} + \mathcal{I}_{n,67,5°,0°} + \mathcal{I}_{n,67,5°,90°}}{4}, \\ \mathrm{Q}_n = \frac{\mathcal{I}_{n,0°,0°} - \mathcal{I}_{n,0°,90°} - \mathcal{I}_{n,45°,0°} + \mathcal{I}_{n,45°,90°}}{2}, \\ \mathrm{U}_n = \frac{-\mathcal{I}_{n,22,5°,0°} + \mathcal{I}_{n,22,5°,90°} + \mathcal{I}_{n,67,5°,0°} - \mathcal{I}_{n,67,5°,90°}}{2} \end{cases} \tag{1.25}$$

Le problème de cette méthode est qu'elle n'est pas très robuste aux effets instrumentaux. Il en existe donc une seconde moins biaisée qui est celle du double ratio [Tinbergen, 2005, Avenhaus et al., 2014]. On commence par calculer les valeurs :

$$R_{\mathrm{Q}_n} = \sqrt{\frac{\mathcal{I}_{n,0°,0°}/\mathcal{I}_{n,0°,90°}}{\mathcal{I}_{n,45°,0°}/\mathcal{I}_{n,45°,90°}}} \text{ et } R_{\mathrm{U}_n} = \sqrt{\frac{\mathcal{I}_{n,22,5°,0°}/\mathcal{I}_{n,22,5°,90°}}{\mathcal{I}_{n,67,5°,0°}/\mathcal{I}_{n,67,5°,90°}}},$$

que l'on combine ensuite comme :

$$p_{\mathrm{Q}_n} = \frac{R_{\mathrm{Q}_n} - 1}{R_{\mathrm{Q}_n} + 1} \text{ et } p_{\mathrm{U}_n} = \frac{R_{\mathrm{U}_n} - 1}{R_{\mathrm{U}_n} + 1}.$$

On calcule enfin :

$$\begin{cases} I_{\mathrm{Q}_n} = \frac{\mathcal{I}_{n,0°,0°} + \mathcal{I}_{n,0°,90°} + \mathcal{I}_{n,45°,0°} + \mathcal{I}_{n,45°,90°}}{2} \\ I_{\mathrm{U}_n} = \frac{\mathcal{I}_{n,22,5°,0°} + \mathcal{I}_{n,22,5°,90°} + \mathcal{I}_{n,67,5°,0°} + \mathcal{I}_{n,67,5°,90°}}{2}. \end{cases}$$

et donc finalement on a

$$\begin{cases} \mathrm{I}_n = \frac{I_{\mathrm{Q}_n} + I_{\mathrm{U}_n}}{2} \\ \mathrm{Q}_n = p_{\mathrm{Q}_n} I_{\mathrm{Q}_n} \\ \mathrm{U}_n = p_{\mathrm{U}_n} I_{\mathrm{U}_n}. \end{cases} \tag{1.26}$$

Une fois ces trois paramètres de Stokes retrouvés, il est possible de calculer l'angle de polarisation θ_n, l'intensité polarisée I^{p}_n et l'intensité non-polarisée I^{u}_n par les relations suivantes :

$$\begin{cases} I^{\mathrm{p}}_n = \sqrt{\mathrm{Q}^2_n + \mathrm{U}^2_n} \\ \theta_n = (1/2) \arctan(\mathrm{U}_n/\mathrm{Q}_n) \\ I^{\mathrm{u}}_n = \mathrm{I}_n - I^{\mathrm{p}}_n. \end{cases} \tag{1.27}$$

Remarque : Pour minimiser les effets de polarisation instrumentale, lors de l'estimation I^{p}_n, il est utile de projeter chaque pixel de l'image obtenue, orthogonalement et tangentiellement, sur le cercle centré en l'étoile et dont le rayon correspond à la distance de ce pixel au centre. Comme évoqué précédemment, la lumière polarisée est issue de la diffusion de la lumière de l'étoile par la poussière, d'un angle de polarisation orthogonal à l'angle d'incidence de la lumière venant de l'étoile. L'hypothèse est que la polarisation du disque en chaque pixel est colinéaire au cercle en ce pixel. En projetant orthogonalement, il ne reste alors que l'intensité polarisée sans polarisation instrumentale. De même, en projetant tangentielle, il ne reste alors

que la polarisation instrumentale. On note ces deux projections dans ce référentiel radial $P_{n,\perp}$ et $P_{n,\parallel}$. Elles sont données par :

$$\begin{cases} P_{n,\perp} = \mathrm{Q}_n \cos(2\varphi_n) + \mathrm{U}_n \sin(2\varphi_n) \\ P_{n,\parallel} = \mathrm{U}_n \cos(2\varphi_n) - \mathrm{Q}_n \sin(2\varphi_n), \end{cases} \tag{1.28}$$

où $\varphi_n = \arctan\left(\frac{x_n - x_0}{y_n - y_0}\right)$ avec (x_n, y_n) correspond aux coordonnées cartésiennes du pixel n dans le plan de l'image et (x_0, y_0) celles du centre de l'étoile.

Une façon de minimiser la polarisation instrumentale, est alors de minimiser le signal sur $P_{n,\parallel}$ par rapport à l'intensité non-polarisée comme proposé dans [Avenhaus et al., 2018]. Soit $C = (c_1, c_2, c_3, c_4)^{\mathrm{T}} \in \mathbb{R}^4$ et $\varepsilon > 0$, on pose pour tout $n \in \{1, \dots N\}$:

$$\begin{cases} P_{n,\perp} = (\mathrm{Q}_n - c_2 I^{\mathrm{u}}_n - c_4)\cos(2(\varphi_n + \varepsilon)) + (\mathrm{U}_n - c_1 I^{\mathrm{u}}_n - c_3)\sin(2(\varphi_n + \varepsilon)) \\ P_{n,\parallel} = (\mathrm{U}_n - c_1 I^{\mathrm{u}}_n - c_3)\cos(2(\varphi_n + \varepsilon)) - (\mathrm{Q}_n - c_2 I^{\mathrm{u}}_n - c_4)\sin(2(\varphi_n + \varepsilon)). \end{cases} \tag{1.29}$$

On résout alors le critère

$$\min_{C,\varepsilon} \sum_{n=1}^{N} \left(P_{n,\parallel}\right)^2 = \min_{C,\varepsilon} f(C, \varepsilon). \tag{1.30}$$

On remarque que le critère est linéaire en C mais pas en ε, on va donc chercher le minimum de la fonction

$$f_{\hat{C}}(\varepsilon) = \min_{C} f(C, \varepsilon). \tag{1.31}$$

Comme cette fonction n'est pas convexe, on discrétise le petit ensemble auquel appartient ε, que l'on suppose être $]-1, 1[$, et on minimise en C pour chacun des points de discrétisation. La minimisation en C revient à résoudre le problème $AC = B$ où A est une matrice symétrique de taille 4×4 mais de rang 2 uniquement. Il est donc nécessaire d'utiliser la décomposition en valeurs singulières pour trouver le vecteur C minimal. Finalement, on choisit la valeur de ε et le vecteur C associé donnant la plus petite valeur du critère.

Un algorithme permettant de trouver ce minimum sans avoir à discrétiser tout l'intervalle est l'algorithme de type Brent [Brent, 1973].

Le problème des méthodes de l'état-de-l'art est que la propagation des erreurs statiques et statistiques n'est pas maitrisée, aussi bien lors des interpolations des pixels, qu'au moment du pré-traitement des données et lors de la combinaison de celles-ci. C'est pourquoi dans cette thèse nous proposons une méthode de reconstruction des paramètres I^{u}, I^{p} et θ par une approche de type *«problèmes inverses »*, qui permet cette maitrise des erreurs et permet d'estimer ces paramètres de manière optimale.

Une autre façon de gérer la polarisation instrumentale est de la calibrer à l'aide de données de calibration spécifiques, comme proposé par [van Holstein et al., 2020]. L'auteur représente l'ensemble des transformations de la polarisation dans l'instrument par une matrice de Mueller $\mathbf{M}$ dont les entrées sont inconnues. À partir de données où la polarisation est connue, il vient alors estimer par un moindre carré les éléments de la matrice $\mathbf{M}$.

1.2 Excursion au cœur des *«problèmes inverses »*

Nous sommes dans une époque où la taille des jeux de données, que ce soit en astrophysique, en imagerie médicale, ou bien d'autre domaines, ne cesse d'augmenter. Il est donc nécessaire de développer en continu de nouveaux outils numériques permettant à la fois le stockage et le traitement de ces données, dans un moindre coût aussi bien en temps de calcul qu'en mémoire.

On appelle méthode d'apprentissage, ou *Machine Learning*, toute méthode permettant d'apprendre une information à partir de données. De la simple régression linéaire aux réseaux de neurones, en passant par les arbres de décision et les décompositions multi-échelles, la faune des méthodes disponibles est gigantesque et ne cesse de s'agrandir.

L'approche de type *«problèmes inverses »* peut être considérée comme un sous-ensemble de ces méthodes. Développée au cours des dernières décennies, cette approche lie à la fois méthodes probabilistes et méthodes d'optimisation continue. Ces deux types de méthodes pouvant être reliées par formulation du maximum *a posteriori* [Tarantola, 2005].

Le but des méthodes de type *«problèmes inverses »* est, à partir d'un certain ensemble de données $\boldsymbol{d} \in \mathbb{R}^K$ suivant une certaine loi de probabilité dont le modèle est connu, où K est le nombre de réalisations de cette loi de probabilité, de trouver un estimateur $\boldsymbol{x} \in \mathbb{R}^L$ des paramètres du modèle des données, où L est la taille de l'estimateur. Cet estimateur s'obtient en général par la résolution d'un problème de minimisation de la forme :

$$\widehat{\boldsymbol{x}} \in \underset{\boldsymbol{x}\in\mathbb{R}^L}{\operatorname{Argmin}} \; [\Phi(\boldsymbol{x}) + \mathcal{R}(\boldsymbol{x}) + \iota_{\mathcal{C}}(\boldsymbol{x})] \tag{1.32}$$

où $\Phi(\boldsymbol{x}) : \mathbb{R}^L \to \mathbb{R}$ est un terme d'attache aux données, $\mathcal{R}(\boldsymbol{x}) : \mathbb{R}^L \to \mathbb{R}$ est un terme de régularisation facultatif, qui peut être décomposé comme $\mathcal{R}(\boldsymbol{x}) = \lambda g(\mathbf{G}\boldsymbol{x})$ où $\mathbf{G} : \mathbb{R}^L \to \mathbb{R}^G$ est un opérateur linéaire, $g : \mathbb{R}^G \to]-\infty, +\infty]$ une fonction de pénalisation dont la contribution doit être ajustée par le paramètre $\lambda \geq 0$, et $\iota_{\mathcal{C}}(\boldsymbol{x}) : \mathbb{R}^L \to \{0, +\infty\}$ est l'indicatrice d'une contrainte stricte $\mathcal{C}$ facultative sur le domaine de définition de $\boldsymbol{x}$. Nous présentons ces trois termes en détail dans la suite.

1.2.1 Problème direct

La modélisation directe d'un ensemble de données $\boldsymbol{d} = (\boldsymbol{d}_k)_{k\in\{1,\dots K\}}$ en fonction du signal d'intérêt $\overline{\overline{\boldsymbol{x}}} = (\overline{\overline{\boldsymbol{x}}}_\ell)_{\ell\in\{1,\dots L\}}$ s'écrit de manière générale sous la forme :

$$\boldsymbol{d} = \mathcal{B}(f(\overline{\overline{\boldsymbol{x}}})), \tag{1.33}$$

où $f : (\mathbb{R})^L \to (\mathbb{R})^K$ est un opérateur de déformation du signal fixé, qui n'est pas nécessairement linéaire, et $\mathcal{B} : (\mathbb{R})^K \to (\mathbb{R})^K$ un opérateur de dégradation aléatoire. Dans le cas où f est linéaire, on peut alors l'écrire sous la forme $f(\boldsymbol{x}) = \mathbf{A}\boldsymbol{x}$ où $\mathbf{A} : \mathcal{M}_{K\times L}(\mathbb{R})$ est une matrice.

Dans cette thèse, je travaille sous l'hypothèse que les données $\boldsymbol{d} \in \mathbb{R}^K$ suivent une distribution normale, de moyenne $f(\overline{\overline{\boldsymbol{x}}}) : \mathbb{R}^L \to \mathbb{R}^K$ où le paramètre $\overline{\overline{\boldsymbol{x}}}$ est l'inconnue que l'on cherche à estimer, et de covariance Σ, on note $\boldsymbol{d} \sim \mathcal{N}(f(\overline{\overline{\boldsymbol{x}}}), \Sigma)$. La densité de probabilité d'une telle loi est donnée par :

$$\forall \boldsymbol{x} \in \mathbb{R}^L, \quad \mathcal{L}(\boldsymbol{d}|\boldsymbol{x}) = \frac{\mathrm{e}^{-\frac{1}{2}\left(\boldsymbol{d}-f(\boldsymbol{x})\right)^\top \Sigma^{-1}\left(\boldsymbol{d}-f(\boldsymbol{x})\right)}}{\sqrt{(2\pi)^{\mathring{K}} \det \Sigma}} \tag{1.34}$$

où $^\top$ représente l'opérateur transpose et $\mathring{K}$ le nombre de données valides.

Je travaille sous l'hypothèse que les réalisations de $\boldsymbol{d}$ sont indépendantes pixel à pixel et dans le temps, c'est-à-dire que le nombre d'électrons enregistrés par un pixel ne dépend ni du nombre enregistré par ses voisins, ni du nombre enregistré à l'acquisition précédente. Cette hypothèse implique que la matrice de covariance Σ est diagonale avec les entrées $\Sigma_{k,k} = \sigma_k^2$.

1.2.2 Attaches aux données

La méthode la plus couramment utilisée pour estimer les paramètres du modèle de données est de maximiser la vraisemblance des données, c'est-à-dire :

$$\widehat{\boldsymbol{x}} \in \underset{\boldsymbol{x}\in\mathbb{R}^L}{\operatorname{Argmax}}\ \mathcal{L}(\boldsymbol{d}|\boldsymbol{x}). \tag{1.35}$$

Selon l'expression de la densité, il peut être plus simple de minimiser la co-logvraisemblance des données :

$$-\log \mathcal{L}(\boldsymbol{d}|\boldsymbol{x}). \tag{1.36}$$

Dans le cas Gaussien tel que celui présenté à l'équation (1.34), où Σ est connu, trouver un estimateur de $\overline{\overline{x}}$ correspond alors à minimiser la co-logvraisemblance :

$$-\log \mathcal{L}(\boldsymbol{d}|\boldsymbol{x}) = \frac{1}{2}\Big(\boldsymbol{d} - f(\boldsymbol{x})\Big)^\top \Sigma^{-1}\Big(\boldsymbol{d} - f(\boldsymbol{x})\Big) = \frac{1}{2}\|\boldsymbol{d} - f(\boldsymbol{x})\|^2_{\Sigma^{-1}}. \tag{1.37}$$

La norme au carrée pondérée par l'inverse de la matrice de covariance $\|\cdot\|^2_{\Sigma^{-1}}$ correspond au carré de la distance de Mahalanobis [Mahalanobis, 1936].

Dans le cas de données séparables, la co-logvraisemblance des données peut se réécrire sous la forme :

$$-\log \mathcal{L}(\boldsymbol{d}|\boldsymbol{x}) = \sum_{k=1}^{K} \frac{1}{2\sigma_k^2}\Big(\boldsymbol{d}_k - f_k(\boldsymbol{x})\Big)^2. \tag{1.38}$$

Dans les deux cas, estimer le paramètre $\boldsymbol{x}$ revient à résoudre le problème de minimisation :

$$\widehat{\boldsymbol{x}} \in \underset{\boldsymbol{x}\in\mathbb{R}^L}{\operatorname{Argmin}}\ -\log \mathcal{L}(\boldsymbol{d}|\boldsymbol{x}). \tag{1.39}$$

On pose alors $\Phi(\boldsymbol{x}) = -\log \mathcal{L}(\boldsymbol{d}|\boldsymbol{x})$ comme attache aux données. D'après le théorème de Fermat sur les points stationnaires, résoudre (1.39) consiste à trouver $\widehat{\boldsymbol{x}} \in \mathbb{R}^L$ tel que $\nabla\Phi(\widehat{\boldsymbol{x}}) = 0$. L'unicité de $\widehat{\boldsymbol{x}}$ n'est cependant pas respectée si Φ n'est pas strictement convexe. L'ajout de contraintes peut permettre de réduire l'espace des solutions.

D'un point de vue statistique, l'estimation $\widehat{\boldsymbol{x}}$ de $\overline{\overline{\boldsymbol{x}}}$ est une variable aléatoire. Or, la variance de $\widehat{\boldsymbol{x}}$ dépend du conditionnement de Φ, ce qui peut être problématique dans le cas où $\nabla\Phi$ est mal conditionné, car cela va augmenter la variance de $\widehat{\boldsymbol{x}}$.

En effet, plaçons-nous pour commencer dans le cas où f est linéaire, c'est-à-dire que l'on peut écrire $f(\boldsymbol{x}) = \mathbf{A}\boldsymbol{x}$ avec $\mathbf{A} : \mathcal{M}_{K\times L}(\mathbb{R})$. Supposons que les données $\boldsymbol{d} \in \mathbb{R}^K$ peuvent s'écrire sous la forme $\boldsymbol{d} = \mathbf{A}\overline{\overline{\boldsymbol{x}}} + \boldsymbol{\beta}$ où $\boldsymbol{\beta} \sim \mathcal{N}(0, \Sigma)$. Supposons également que $\mathbf{A}^*\Sigma^{-1}\mathbf{A}$ est inversible. On a :

$$\nabla\Phi(\widehat{\boldsymbol{x}}) = 0 \Leftrightarrow \mathbf{A}^\top\Sigma^{-1}\Big(\boldsymbol{d} - \mathbf{A}\widehat{\boldsymbol{x}}\Big) = 0 \tag{1.40}$$

$$\Leftrightarrow \mathbf{A}^*\Sigma^{-1}\mathbf{A}\boldsymbol{x} = \mathbf{A}^*\Sigma^{-1}\Big(\mathbf{A}\overline{\overline{\boldsymbol{x}}} + \beta\Big) \tag{1.41}$$

$$\Leftrightarrow \widehat{\boldsymbol{x}} = \overline{\overline{\boldsymbol{x}}} + \Big(\mathbf{A}^*\Sigma^{-1}\mathbf{A}\Big)^{-1}\mathbf{A}^*\Sigma^{-1}\beta. \tag{1.42}$$

Or si $\mathbf{A}^*\Sigma^{-1}\mathbf{A}$ est mal conditionnée, on a alors :

$$\|\widehat{\boldsymbol{x}} - \overline{\boldsymbol{x}}\|^2 = \|\left(\mathbf{A}^*\Sigma^{-1}\mathbf{A}\right)^{-1}\mathbf{A}^*\Sigma^{-1}\beta\|^2 \tag{1.43}$$

et donc toute petite perturbation dans β va se retrouver amplifiée et donc augmenter la variance de l'estimation $\widehat{\boldsymbol{x}}$.

Pour pallier ce problème, il est alors usuel d'ajouter à l'attache aux données un terme permettant d'améliorer le conditionnement et de baisser ainsi la variance de $\widehat{\boldsymbol{x}}$. D'un point de vue statistique, cela consiste en un *a priori* sur la loi de probabilité de $\widehat{\boldsymbol{x}}$. En terme d'optimisation, cela consiste en l'ajout d'un terme de régularisation.

1.2.3 Régularisations

Le choix du type de régularisation utilisée pour résoudre le *«problème inverse »* va dépendre des structures attendues dans les images des environnements reconstruits. C'est-à-dire si les environnements sont des objets étendus et réguliers, avec des bords francs ou alors seulement composés d'éléments ponctuels. Dans les valeurs des paramètres $\boldsymbol{x}$, les structures lisses vont se traduire par des faibles variations d'intensité pixel à pixel, les bords francs par de grands sauts d'intensité le long des bords, et enfin pour les éléments ponctuels, par peu de pixels avec de l'intensité parmi de nombreux pixels nuls.

Ces termes de régularisation peuvent tous s'écrire sous la forme :

$$\mathcal{R}(\boldsymbol{x}) = \lambda g(\mathbf{G}\boldsymbol{x}) \tag{1.44}$$

où $\lambda \geq 0$ est l'hyperparamètres qui va réguler la contribution de la régularisation, $g : \mathbb{R}^G \to]-\infty, +\infty]$ la fonction de coût et $\mathbf{G} : \mathbb{R}^N \to \mathbb{R}^G$ un opérateur linéaire. Cela peut être vu par analogie, comme un Lagrangien ou $\lambda \geq 0$ serait le multiplicateur de Lagrange.

Dans le cas des environnements circumstellaires, on s'attend à avoir des images d'objets étendus dans le champ, dont les intensités en chaque pixel sont relativement homogènes. Les bords peuvent être assez francs et il est également possible, dans le cas où il y a des étoiles dans le champ, d'avoir une composante parcimonieuse.

Dans le cas de recherche d'un paramètre $\boldsymbol{x}$ parcimonieux, il s'agit tout simplement d'utiliser la norme ℓ_1 sur les valeurs de $\boldsymbol{x}$, ce qui donne pour tout $\lambda \geq 0$ par :

$$\mathcal{R}(\boldsymbol{x}) = \lambda \|\boldsymbol{x}\|_{\ell_1}. \tag{1.45}$$

En augmentant la valeur du paramètre λ, on favorise une solution *«creuse »*, c'est-à-dire avec peu d'élément non-nuls.

Pour une pénalisation sur la régularité des valeurs, on va en général pénaliser le gradient des images. Pour un pixel $n \in \{1, \dots N\}$ d'une image, on note son gradient par $(\boldsymbol{D}\boldsymbol{x})_n \in \mathbb{R}^2$. Il est composé d'une différence horizontale et verticale des pixels, c'est-à-dire si l'on note (x_n, y_n) les coordonnées cartésiennes du pixel $n \in \{1, \dots N\}$ dans le plan de l'image :

$$(\boldsymbol{D}\boldsymbol{x})_n = \begin{pmatrix} (\boldsymbol{D}^{\mathrm{h}}\boldsymbol{x})_n \\ (\boldsymbol{D}^{\mathrm{v}}\boldsymbol{x})_n \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_{x_n+1, y_n} - \boldsymbol{x}_{x_n, y_n} \\ \boldsymbol{x}_{x_n, y_n+1} - \boldsymbol{x}_{x_n, y_n} \end{pmatrix}. \tag{1.46}$$

Pour le calcul du gradient au bord de l'image, on suppose soit que l'image est périodique, soit continue et donc de gradient nul.

La pénalisation sur le gradient la plus simple consiste en une simple pénalisation quadratique du gradient, c'est-à-dire pour tout $\lambda \geq 0$:

$$\mathcal{R}(\boldsymbol{x}) = \frac{\lambda}{2}\|\mathbf{D}\boldsymbol{x}\|_2^2, \tag{1.47}$$

C'est la régularisation de Tikhonov, qui fut introduite en tout premier lieu dans [Tikhonov, 1963]. Cette pénalisation permet lors de l'estimation d'un paramètre d'obtenir une solution relativement lisse. L'avantage d'une telle pénalisation est qu'elle est quadratique et donc différentiable. Cela permet soit d'obtenir une expression explicite de la solution soit l'utilisation d'algorithmes très rapides pour la recherche du minimum. L'inconvénient d'une telle régularisation est qu'elle ne gère que très difficilement les sauts d'intensités dans une image et va avoir tendance à lisser les bords, ce qui peut être gênant lorsque l'objet observé a des bords francs.

La régularisation la plus classique permettant de préserver les bords est la Variation Totale [Rudin et al., 1992] (TV : *Total Variation*). Cette pénalisation, proposée par Rudin et al. en 1992, consiste en une pénalisation par la norme ℓ_1 des valeurs du gradient, afin que celui-ci soit parcimonieux, c'est-à-dire pour tout $\lambda \geq 0$:

$$\mathcal{R}(\boldsymbol{x}) = \lambda\|\mathbf{D}\boldsymbol{x}\|_{\ell_1}. \tag{1.48}$$

La norme ℓ_1 étant moins sensible aux valeurs extrêmes que la norme ℓ_2, une pénalisation par la norme ℓ_1 sur le gradient va donc privilégier un gradient parcimonieux et donc une solution constante par morceaux, ce qui correspond à des bords francs. Il existe également une version isotrope de cette régularisation donnée par :

$$\mathcal{R}(\boldsymbol{x}) = \lambda\|\mathbf{D}\boldsymbol{x}\|_{\ell_{1,2}} = \lambda \sum_{n=1}^{N} \sqrt{(\mathbf{D}^{\mathrm{h}}\boldsymbol{x})_n^2 + (\mathbf{D}^{\mathrm{v}}\boldsymbol{x})_n^2}. \tag{1.49}$$

Il est possible d'utiliser une approximation hyperbolique différentiable de la variation totale, ou TV hyperbolique (TV-h) [Charbonnier et al., 1997], qui pour tout $\lambda, \mu \geq 0$ est donné par :

$$\mathcal{R}(\boldsymbol{x}) = \lambda\left(\sqrt{\|\mathbf{D}\boldsymbol{x}\|_2^2 + \mu^2} - \mu\right). \tag{1.50}$$

Le paramètre μ permet de tendre soit vers une régularisation quadratique, avec des bords lissés comme Tikhonov, soit lorsque μ tend vers zéro, tendre vers TV avec des bords plus francs. La figure 1.6 montre le comportement des normes ℓ_2, ℓ_1 et l'approximation hyperbolique de ℓ_1, pour différentes valeurs de λ et μ.

Une autre fonction connue permettant d'approximer la norme ℓ_1 est la fonction de Huber [Huber, 1992], qui consiste à prendre la norme ℓ_1 pour les valeurs supérieures à une certaine valeur de $\mu \geq 0$ et la norme ℓ_2 pour les valeurs en dessous de μ. L'expression de la Variation Totale approximée avec la norme de Huber est donnée par :

$$\mathcal{R}(\boldsymbol{x}) = \lambda\|\mathbf{D}\boldsymbol{x}\|_{\mathrm{Huber}}. \tag{1.51}$$

où

$$\forall \boldsymbol{x} \in \mathbb{R}^2, \quad \|\boldsymbol{x}\|_{\mathrm{Huber}} = \begin{cases} \|\boldsymbol{x}\|_{\ell_1} - \mu, & \text{si } \|\boldsymbol{x}\|^2 \geq \mu \\ \frac{1}{2\mu}\|\boldsymbol{x}\|^2 & \text{sinon.} \end{cases} \tag{1.52}$$

Autour de 0 la norme est alors quadratique. De plus, si $\mu \to 0$ alors on retrouve également la norme ℓ_1. Le comportement de la norme de Huber est également présenté sur la figure 1.6.

On peut observer sur la figure 1.6 que les régularisations en utilisant une approximation hyperbolique de la norme ℓ_1 et Huber sont très similaires, le passage à quelque chose de quadratique étant plus doux avec l'approximation hyperbolique.

L'inconvénient de TV et de son approximation hyperbolique lorsque μ est choisi petit, est qu'elle crée, dans les reconstructions, comme des aplats de couleurs, appelé communément effet *cartoon*.

Pour palier aux effets *cartoon* tout en gardant des bords francs, il existe d'autres régularisations. Parmi les plus usuelles, on peut citer la pénalisation par la norme de Shatten ℓ_1 sur l'opérateur hessien [Lefkimmiatis et al., 2013, Chierchia et al., 2014]. On définit le hessien d'une image en chaque pixel $n \in \{1, \dots N\}$ de $\boldsymbol{x} \in \mathbb{R}^N$ par :

$$(\mathbf{D}^2\boldsymbol{x})_n = (\mathbf{D}(\mathbf{D}\boldsymbol{x}))_n \begin{pmatrix} (\mathbf{D}^{\mathrm{h}}(\mathbf{D}^{\mathrm{h}}\boldsymbol{x}))_n & (\mathbf{D}^{\mathrm{h}}(\mathbf{D}^{\mathrm{v}}\boldsymbol{x}))_n \\ (\mathbf{D}^{\mathrm{v}}(\mathbf{D}^{\mathrm{h}}\boldsymbol{x}))_n & (\mathbf{D}^{\mathrm{h}}(\mathbf{D}^{\mathrm{v}}\boldsymbol{x}))_n \end{pmatrix} \tag{1.53}$$

où $\mathbf{D}^{\mathrm{h}}\mathbf{D}^{\mathrm{h}}$, $\mathbf{D}^{\mathrm{h}}\mathbf{D}^{\mathrm{v}}$ et $\mathbf{D}^{\mathrm{v}}\mathbf{D}^{\mathrm{v}}$ représentent les dérivées secondes selon les directions respectivement horizontales, diagonales et verticales.

Pour un pixel, la norme de Shatten consiste à appliquer la norme ℓ_p avec $p \in [1, +\infty[$ aux valeurs propres des matrices hessiennes.

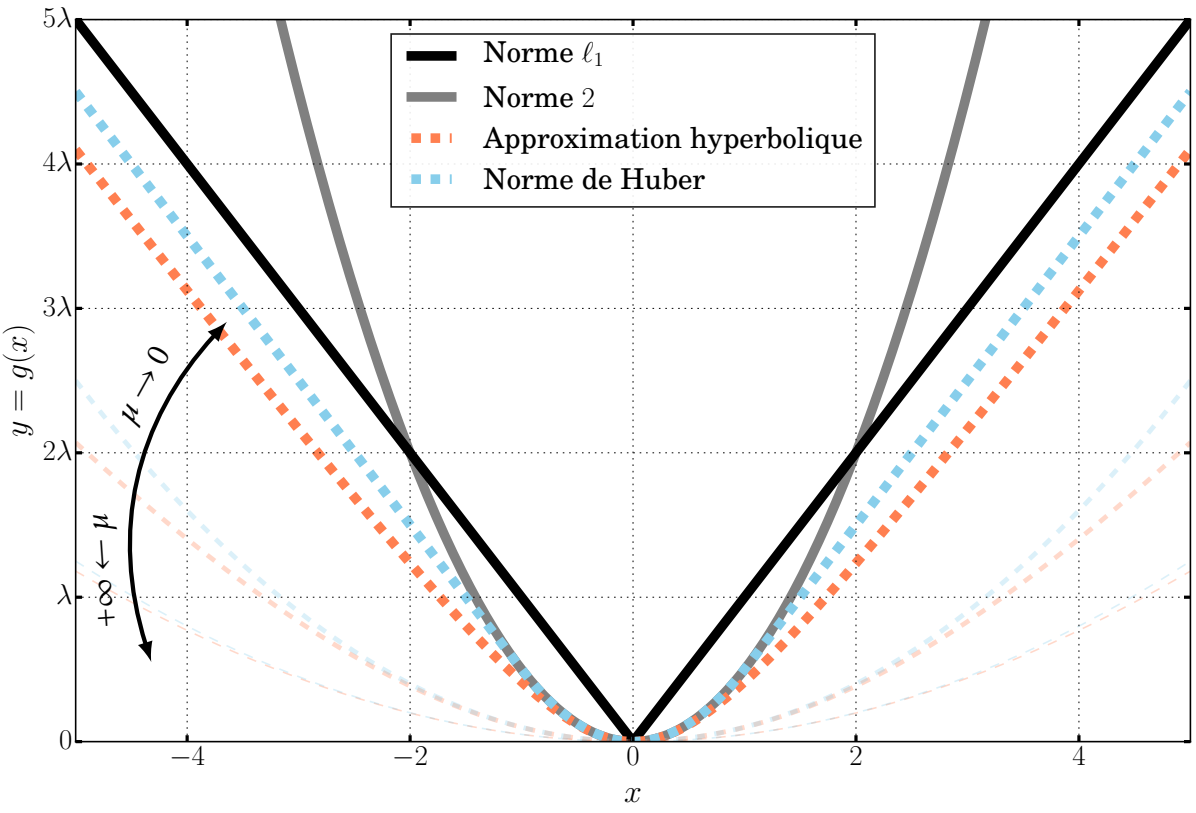


FIGURE 1.6 – Comparaison de la norme ℓ_2, de la norme ℓ_1, de son approximation hyperbolique et de la fonction de Huber pour différentes valeurs de μ.

La régularisation par la norme S_1 est alors donnée par :

$$\mathcal{R}(\boldsymbol{x}) = \lambda \sum_{n=1}^{N} S_1\left(\mathbf{D}^2 \boldsymbol{x}_n\right). \tag{1.54}$$

où la norme de Shatten est donnée par :

$$\mathcal{S}_1(\mathbf{D}^2 \boldsymbol{x}_n) := \left|s_1((\mathbf{D}^2 \boldsymbol{x})_n)\right| + \left|s_2((\mathbf{D}^2 \boldsymbol{x})_n)\right|, \tag{1.55}$$

où $s_1(\mathbf{D}^2 \boldsymbol{x}), s_2(\mathbf{D}^2 \boldsymbol{x}) \geq 0$ sont les valeurs singulières de l'opérateur hessien.

Cette pénalisation permet d'imposer une contrainte plus douce sur les petites variations entre pixels et donc d'éviter les aplats de couleurs.

Une autre régularisation permettant d'éviter l'effet *cartoon* est la Variation Totale Généralisée [Bredies et al., 2010], (TGV : *Total Generalized Variation*). Cette régularisation est un compromis entre une pénalisation sur le gradient et une pénalisation sur l'opérateur hessien d'une image. Elle consiste à pénaliser conjointement le gradient de l'image et son hessien. Elle s'écrit alors de la manière suivante :

$$\mathcal{R}(\boldsymbol{x}) = \lambda \left\{ \min_{\boldsymbol{y} \in (\mathbb{R}^2)^N} \left[\gamma_1 \|\mathbf{D}\boldsymbol{x} - \boldsymbol{y}\|_{\ell_1} + \gamma_0 \|\mathbf{D}\boldsymbol{y}\|_{\ell_1} \right] \right\}, \tag{1.56}$$

où $\gamma_0 \geq 0$ et $\gamma_1 \geq 0$ sont deux hyperparamètres à régler. Cette régularisation n'est pas non-plus différentiable.

Enfin, une dernière régularisation, permettant d'éviter l'impression d'aplats de couleurs tout en ayant des bords francs, est l'utilisation d'une pénalisation multi-échelle, c'est-à-dire sur une décomposition en ondelettes ou en trames de $\boldsymbol{x}$. Soit $\boldsymbol{\psi} = (\boldsymbol{\psi}_o)_{o \in \{1,\dots O\}}$ une base d'ondelettes ou un ensemble de trame, où $\boldsymbol{\psi}_o \in \mathbb{R}^N$ pour tout $o \in \{1, \dots O\}$ où O est la taille de la base ou la redondance des trames, alors la régularisation sur la décomposition en ondelette est donnée par :

$$\mathcal{R}(\boldsymbol{x}) = \lambda \|\mathbf{P}_{\boldsymbol{\psi}} \boldsymbol{x}\|_{\ell_1}, \tag{1.57}$$

où $\mathbf{P}_{\boldsymbol{\psi}}$ représente la transformée en ondelettes ou sur une trame.

Il existe de nombreuses bases d'ondelettes, la plus simple étant celle des ondelettes de Haar. Une autre base bien connue est celle des ondelettes de Daubechies, qui sont utilisées pour la compression au format JPEG 2000. Parmi les trames, en astrophysique, les curvelet sont utilisées dès le début des années 2000 pour le débruitage d'images [Starck et al., 2003]. Des trames d'ondelettes particulièrement adaptées pour les environnements circumstellaires sont les Shearlets, utilisées pour la reconstruction de disque en ADI dans [Pairet et al., 2018b]. Cet ensemble de trame trouve son intérêt dans la reconstruction d'objets au bords francs dont l'orientation est autre qu'horizontale et verticale. En effet dans le cas des régularisations sur le gradient des images, les régularisations sont faites sur des différences horizontales et verticales des pixels. Les éléments de la base de Shearlets prennent différentes orientations sur tout le cercle trigonométrique, à différentes échelles, permettant ainsi une décomposition des contours optimale. Pour plus de contenu sur les ondelettes, se référer à [Jacques et al., 2011, Pustelnik et al., 2016].

Toutes ces régularisations ont un intérêt pour la reconstruction des environnements circumstellaires et les avantages et inconvénients dans le contexte de l'analyse des environnements circumstellaires seront discutés tout au long de ce manuscrit.

Ces régularisations sont des contraintes douces, dans le sens où on cherche l'image régularisée qui minimise une fonction de coût assez douce. Le poids de cette régularisation est par ailleurs géré par l'hyperparamètre λ : plus λ sera élevé, plus le poids de la contrainte va être fort et donc la solution régularisée, et inversement.

Cependant, dans certain cas on peut vouloir s'assurer que la contrainte soit absolument respectée. On parle alors de contraintes strictes.

1.2.4 Contraintes strictes

Lors de la minimisation d'un critère de la forme

$$\widehat{\boldsymbol{x}} \in \underset{\boldsymbol{x}\in\mathbb{R}^L}{\operatorname{Argmin}} \; [\Phi(\boldsymbol{x}) + \mathcal{R}(\boldsymbol{x})], \tag{1.58}$$

toute contrainte stricte sur la solution $\widehat{\boldsymbol{x}} \in \mathbb{R}^L$ peut-être vue comme une contrainte d'appartenance à un domaine $\mathcal{C}$, dans lequel la contrainte est vérifiée. C'est-à-dire que la minimisation est faite uniquement sur le domaine $\mathcal{C}$:

$$\widehat{\boldsymbol{x}} \in \underset{\boldsymbol{x}\in\mathcal{C}}{\operatorname{Argmin}} \; [\Phi(\boldsymbol{x}) + \mathcal{R}(\boldsymbol{x})]. \tag{1.59}$$

En posant alors la fonction indicatrice :

$$\iota_{\mathcal{C}}(\boldsymbol{x}) = \begin{cases} 0 & \text{si } \boldsymbol{x} \in \mathcal{C} \\ +\infty & \text{sinon,} \end{cases} \tag{1.60}$$

minimiser (1.59) revient à minimiser

Rappel éq. (1.32)

$$\widehat{\boldsymbol{x}} \in \underset{\boldsymbol{x}\in\mathbb{R}^L}{\operatorname{Argmin}} \; [\Phi(\boldsymbol{x}) + \mathcal{R}(\boldsymbol{x}) + \iota_{\mathcal{C}}(\boldsymbol{x})].$$

On peut faire ressortir trois types de contraintes qui ont un intérêt dans notre cas.

1. Les contraintes de dynamique : ces contraintes consistent en une restriction directe des valeurs des paramètres à un intervalle donné dont les bornes sont fixes. En imagerie classique, il est assez courant de contraindre les valeurs Rouges Vertes Bleues (RVB) dans l'intervalle $[0, 255]$. Dans notre cas, l'intensité enregistrée se situant entre $[0, +\infty[$, $\mathcal{C}$ est alors l'ensemble

   $$\mathcal{C} = \{\boldsymbol{x} \in \mathbb{R}^L \quad \text{tels que} \quad \forall \ell \in \{1, \dots L\}, \quad \boldsymbol{x}_\ell \geq 0\}. \tag{1.61}$$

   J'utilise une telle contrainte dans les chapitres 2 et 3.

2. Les contraintes linéaires : ces contraintes consistent en une restriction de la combinaison linéaire des paramètres à un intervalle donné dont les bornes sont fixes. Cela correspond à la contrainte sur l'ensemble suivant [Theodoridis et al., 2011] :

   $$\mathcal{C} = \{\boldsymbol{x} \in \mathbb{R}^L \quad \text{tels que} \quad \forall \ell \in \{1, \dots L\}, \; \mathbf{A}\boldsymbol{x} \geq \boldsymbol{y}\}, \tag{1.62}$$

   où $\boldsymbol{y} \in \mathbb{R}^K$ est un vecteur et $\mathbf{A} \in \mathcal{M}_{K\times L}(\mathbb{R})$ un opérateur linéaire fixe. Je présente une contrainte linéaire dans le chapitre 4, pour contraindre les angles de polarisation à une certaine orientation.

3. Les contraintes épigraphiques : ces contraintes consistent à borner d'une fonction des paramètres à un intervalle, dont les bornes ne sont pas fixes mais dépendent aussi des paramètres à estimer. Appliquer une telle contrainte consiste à projeter les paramètres sur l'épigraphe de la fonction [Chierchia et al., 2013, Theodoridis et al., 2011]. Cela correspond à la contrainte sur l'ensemble :

$$\mathcal{C} = \{\boldsymbol{x}, \xi \in \mathbb{R}^L \times \mathbb{R} \quad \text{tels que} \quad \forall \ell \in \{1, \dots L\},\ f(\boldsymbol{x}) \leq \xi\}, \tag{1.63}$$

où $\boldsymbol{x}, \xi \in \mathbb{R}^L \times \mathbb{R}$ sont les paramètres à estimer, reliés par la fonction $f : \mathbb{R}^L \to \mathbb{R}$. Elles sont utilisées en polarimétrie, notamment dans [Birdi et al., 2019], car les paramètres de Stokes sont tous liés entre eux par la contrainte épigraphique :

$$\mathcal{C} = \left\{(\mathrm{I}, \mathrm{Q}, \mathrm{U}, \mathrm{V}) \in (\mathbb{R}^N)^L \quad \text{tels que} \quad \forall n \in \{1, \dots N\} \sqrt{\mathrm{Q}_n^2 + \mathrm{U}_n^2 + \mathrm{V}_n^2} \leq \mathrm{I}\right\}.$$

J'utilise une telle contrainte dans le chapitre 4.

1.2.5 Estimation de l'erreur

Un estimateur du maximum de vraisemblance est asymptotiquement non-biaisé, ce qui signifie qu'avec un nombre de réalisations suffisant, le biais de l'estimation est négligeable. Cependant, comme expliqué plus tôt, dans le cas où le modèle des données est mal conditionné, l'estimation du maximum de vraisemblance sans *a priori* donne une estimation de paramètres avec une grande variance. On a alors un sur-ajustement des données, ce qui n'est pas idéal.

Rajouter de l'information *a priori* sous forme de régularisation et de contraintes permet de diminuer cette variance, plus il y a d'information *a priori* plus la solution est biaisée. C'est ce qu'on appelle alors sous-ajustement, ce qui n'est pas idéal non plus [1]. Il faut donc avoir un bon compromis entre attache aux données et informations *a priori* afin d'assurer un bon compromis biais/variance, comme illustré sur la figure 1.7. Ce compromis est représenté par l'erreur totale, et sera atteint en son minimum.

Afin de s'assurer un bon compromis biais/variance des reconstructions, il est donc nécessaire de pouvoir estimer l'erreur faite sur les paramètres estimés. De plus, pouvoir estimer l'erreur faite par une méthode sur un paramètre permet de comparer efficacement les méthodes de reconstruction.

La littérature foisonne de bornes inférieures de cette erreur, aussi bien dans le cas non-biaisé que dans le cas biaisé [Todros and Tabrikian, 2010a, Todros and Tabrikian, 2010b]. La borne inférieure de Fréchet-Darmois-Cramér-Rao (FDCR) est la plus utilisée dans la littérature. Cependant cette borne ne permet pas de calculer une borne à l'erreur totale mais seulement à la covariance des paramètres dans le cas non-biaisé. Plusieurs dérivées de cette borne existent pour le cas biaisé, comme la borne de FDCR uniforme [Fessler and Hero, 1993]. J'utilise la borne de FDCR dans le chapitre 2 et présente une étude un peu plus approfondie de cette borne dans le chapitre 5.

Une autre possibilité, plutôt que de calculer une borne inférieure à l'erreur, est d'estimer directement l'erreur à l'aide de l'estimateur non-biaisé du risque de Stein (SURE : *Stein Unbiased Risk Estimator*) [Stein, 1981, Ramani et al., 2008, Eldar, 2008, Deledalle et al., 2014]. Une

1. Comparaison grossière : c'est comme faire deviner un animal à quelqu'un, il y a beaucoup de choix la variance de la solution est grande. En lui disant que c'est un mammifère, à quatre pattes, poilu ou encore tacheté, la solution se précise, mais la personne pourrait très bien partir dans la mauvaise direction.

telle borne est utilisée dans les chapitre 3 et 4. Une étude plus approfondie est présentée dans le chapitre 5.

Dans cette thèse nous utilisons essentiellement la borne de FDCR et SURE dans le domaine des données.

1.3 Les méthodes de résolution

Pour résoudre un problème de minimisation de la forme

Rappel éq. (1.32)

$$\widehat{\boldsymbol{x}} \in \underset{\boldsymbol{x}\in\mathbb{R}^L}{\operatorname{Argmin}} \; [\Phi(\boldsymbol{x}) + \mathcal{R}(\boldsymbol{x}) + \iota_{\mathcal{C}}(\boldsymbol{x})],$$

la littérature regorge d'algorithmes dont l'utilisation va dépendre des différentes propriétés de Φ et de $\mathcal{R}$ et de l'ensemble $\mathcal{C}$.

Les performances d'un algorithme se jugent à la fois sur son taux de convergence, c'est-à-dire le nombre d'itérations nécessaires pour aboutir à une certaine précision sur la solution, et le nombre d'allocations en mémoire requises à chaque itération. Ces deux critères ne sont pas dissociables, en effet un algorithme où chaque itération requiert une grande allocation mémoire prend du temps pour faire une seule itération, par exemple de l'ordre de la seconde. À l'inverse, un algorithme où les itérations nécessitent peu d'allocation mémoire, et où l'itération est de l'ordre du centième de seconde peut faire cent fois plus d'itérations en un même temps. Cela peut être critique pour un algorithme qui nécessite plus de 10^4 itérations pour converger.

Dans cette section je présente les trois principaux algorithmes sur lesquels s'appuient mes travaux de thèse. Ces algorithmes ont des complexités différentes et peuvent être ap-

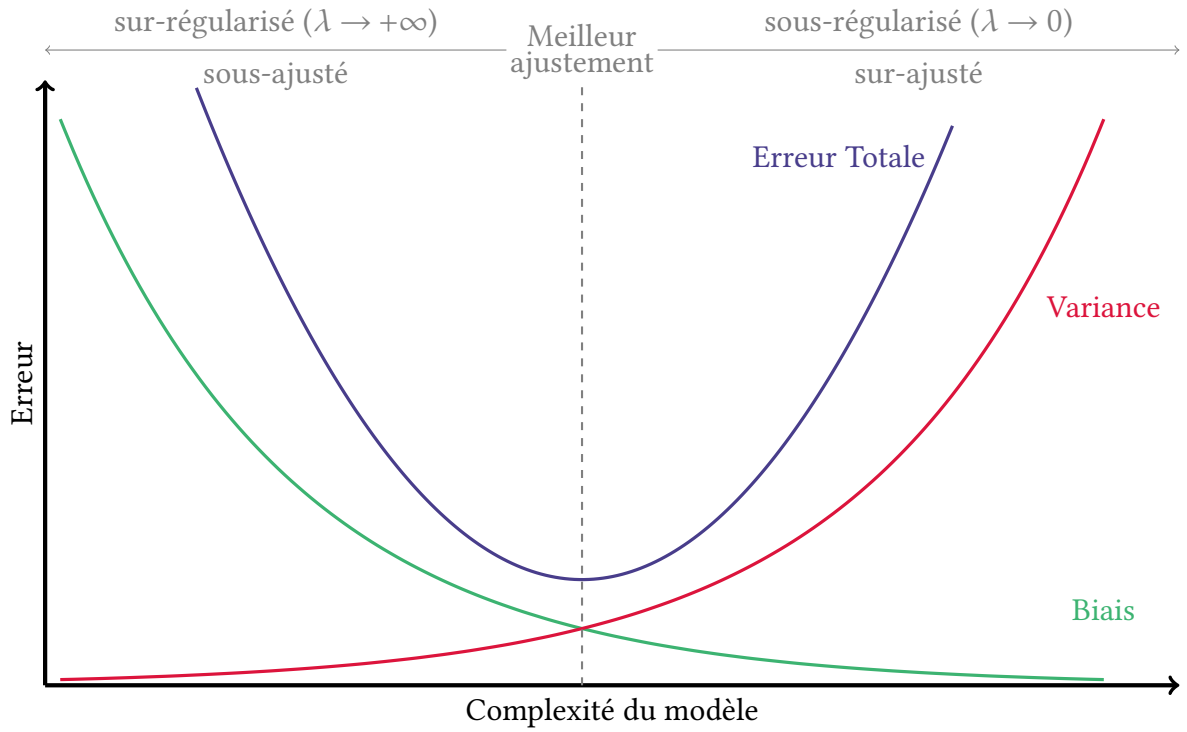


FIGURE 1.7 – Figure sur le compromis biais variance.

pliqués dans le cas différentiable et non-différentiable. Tous ces algorithmes peuvent prendre en compte un préconditionneur. Au préalable j'introduis les outils nécessaires à la recherche de minimum.

1.3.1 Comment trouver le minimum d'une fonction

Soit $\mathcal{H}$ un espace de Hilbert. Comme évoqué précédemment, dans le cas différentiable, la règle de Fermat indique que trouver le minimum $\widehat{\boldsymbol{x}} \in \mathcal{H}$ d'une fonctionnelle $\mathcal{G}$ consiste à trouver $\widehat{\boldsymbol{x}} \in \mathcal{H}$ tel que $\nabla\mathcal{G}(\widehat{\boldsymbol{x}}) = 0$. De plus, dans le cas strictement convexe, $\widehat{\boldsymbol{x}}$ est unique.

Dans la suite, $\mathcal{G}$ n'est pas toujours différentiable (cf. les régularisations présentées dans la section 1.2.3), c'est pourquoi nous avons besoin de définir des outils permettant la recherche de minimum dans le cas non-différentiable : les notions de sous-différentielle et d'opérateur proximal, permettant d'étendre la notion de gradient et de descente de gradient au cas général.

Définition 1.3.1. *Soit $\mathcal{G} : \mathcal{H} \to \mathbb{R}$ une fonction propre, c'est-à-dire que son domaine est différent de l'ensemble vide. Alors sa sous-différentielle $\partial\mathcal{G} : \mathcal{H} \to 2^{\mathcal{H}}$ [2] est donnée par :*

$$\forall \boldsymbol{x} \in \mathcal{H} \quad \partial\mathcal{G}(\boldsymbol{x}) = \Big\{ \boldsymbol{u} \in \mathcal{H} \quad \textit{tels que} \quad \forall \boldsymbol{y} \in \mathcal{H}, \langle \boldsymbol{y} - \boldsymbol{x}, \boldsymbol{u} \rangle + \mathcal{G}(\boldsymbol{x}) \leq \mathcal{G}(\boldsymbol{y}) \Big\}. \tag{1.64}$$

Dans le cas où $\mathcal{G}$ est différentiable, on a alors $\partial\mathcal{G}(x) = \{\nabla\mathcal{G}(x)\}$. Avec cette définition de sous-différentielle, la règle de Fermat s'écrit [Bauschke and Combettes, 2011, Théorème 16.2] :

$$\widehat{\boldsymbol{x}} \in \underset{\boldsymbol{x} \in \mathcal{H}}{\operatorname{Argmin}}\, \mathcal{G}(\boldsymbol{x}) \Leftrightarrow 0 \in \partial\mathcal{G}(\widehat{\boldsymbol{x}}). \tag{1.65}$$

Que ce soit dans le cas différentiable ou non-différentiable, le problème de recherche de 0 du gradient ou de la sous-différentielle de $\mathcal{G}$ peut être vu comme une recherche de point fixe d'une certaine fonction. En effet, soit $\tau \geq 0$:

$$0 \in \partial_{\mathcal{G}}(\widehat{\boldsymbol{x}}) \Leftrightarrow \widehat{\boldsymbol{x}} - \widehat{\boldsymbol{x}} \in \tau\partial_{\mathcal{G}}(\widehat{\boldsymbol{x}}). \tag{1.66}$$

On a alors deux problèmes de point fixes possibles :

$$\widehat{\boldsymbol{x}} \in \Big(\mathbf{Id} - \tau\partial_{\mathcal{G}}\Big)(\widehat{\boldsymbol{x}}). \tag{1.67}$$

et

$$\widehat{\boldsymbol{x}} \in \Big(\mathbf{Id} + \tau\partial_{\mathcal{G}}\Big)(\widehat{\boldsymbol{x}}) \tag{1.68}$$

Dans le cas différentiable la recherche de minimum se fait en général à partir du problème de point fixe 1.67, en construisant des itérations de la forme

$$\boldsymbol{x}^{[t+1]} = \boldsymbol{x}^{[t]} - \tau\nabla\mathcal{G}(\boldsymbol{x}^{[t]}). \tag{1.69}$$

Cela correspond à la descente de gradient explicite. Dans le cas non-différentiable, il a été largement documenté que la méthode de descente de sous-gradient, donnée à partir du schéma explicite $\boldsymbol{x}^{[t+1]} \in \Big(\mathbf{Id} - \tau\partial_{\mathcal{G}}\Big)(\boldsymbol{x}^{[t]})$ n'offrait pas la même efficacité de descente. En effet, celui-ci s'appuie sur un pas de descente τ décroissant afin de garantir la convergence des itérés, conduisant à des instabilités numériques [Polyak, 1987].

2. $2^{\mathcal{H}}$ correspond à l'ensemble des sous-ensembles de $\mathcal{H}$.

Dans le cas non-différentiable, il a donc été proposé d'effectuer des itérations implicites à partir du problème de point fixe 1.68, ce qui conduit à un schéma numérique de la forme :

$$\boldsymbol{x}^{[t]} \in \Big(\mathbf{Id} - \tau \partial_{\mathcal{G}}\Big)(\boldsymbol{x}^{[t+1]}) \tag{1.70}$$

Un tel schéma correspond alors à trouver les points fixes de l'opérateur proximal de la fonction $\mathcal{G}$, dont la définition est donnée par :

Définition 1.3.2. *Soit $\mathcal{G} : \mathcal{H} \to \mathbb{R}$ une fonction propre, convexe et semi-continue inférieurement*[3] *et soit $\tau \geq 0$. Alors, son opérateur proximal est donné par :*

$$\forall \boldsymbol{x} \in \mathcal{H} \quad \operatorname{prox}_{\tau\mathcal{G}}(\boldsymbol{x}) = \underset{\boldsymbol{y}\in\mathcal{H}}{\operatorname{argmin}} \frac{1}{2\tau}\|\boldsymbol{x}-\boldsymbol{y}\|^2 + \mathcal{G}(\boldsymbol{y}). \tag{1.71}$$

On a alors la relation :

$$\forall \boldsymbol{x} \in \mathcal{H} \quad \widehat{\boldsymbol{y}} = \operatorname{prox}_{\tau\mathcal{G}}(\boldsymbol{x}) \Leftrightarrow \boldsymbol{x} - \widehat{\boldsymbol{y}} \in \tau\partial_{\mathcal{G}}(\widehat{\boldsymbol{y}}). \tag{1.72}$$

Ainsi, si $\widehat{\boldsymbol{x}}$ est un point fixe de $\operatorname{prox}_{\tau\mathcal{G}}$, dont l'ensemble est noté **Fix**, on a alors :

$$\begin{aligned} \widehat{\boldsymbol{x}} \in \mathbf{Fix}\ \operatorname{prox}_{\mathcal{G}} &\Leftrightarrow \widehat{\boldsymbol{x}} = \operatorname{prox}_{\mathcal{G}}(\widehat{\boldsymbol{x}}) \\ &\Leftrightarrow \widehat{\boldsymbol{x}} - \widehat{\boldsymbol{x}} \in \partial_{\mathcal{G}}(\widehat{\boldsymbol{x}}) \\ &\Leftrightarrow 0 \in \partial_{\mathcal{G}}(\widehat{\boldsymbol{x}}) \\ &\Leftrightarrow \widehat{\boldsymbol{x}} \in \underset{\boldsymbol{x}\in\mathcal{H}}{\operatorname{Argmin}}\ \mathcal{G}(\boldsymbol{x}). \end{aligned} \tag{1.73}$$

Lorsque l'attache aux données est différentiable, comme dans le cas de la distance de Mahalanobis, minimiser Φ revient à trouver $\nabla\Phi(\widehat{\boldsymbol{x}}) = 0$.

Dans le cas de la variation totale, la norme ℓ_1 n'est pas différentiable pour tout $\boldsymbol{x} \in \mathbb{R}^G$, on ne peut donc pas définir son gradient pour la recherche de minimum du critère (1.32). Il est alors nécessaire de définir sa sous-différentielle. La norme ℓ_1 d'un vecteur étant la somme des valeurs absolues de chaques composantes, la sous différentielle consiste en le produit des sous-différentielles de la valeur absolue en chaque composante. La sous-différentielle de la norme ℓ_1 multipliée par un facteur $\lambda \geq 0$ est donc donnée par :

$$\partial\lambda\|\boldsymbol{x}\|_{\ell_1} = \underset{n\in\{1,\dots N\}}{\times} \begin{cases} -\lambda & \text{if } \boldsymbol{x}_n < 0, \\ [-\lambda, \lambda] & \text{if } \boldsymbol{x}_n = 0, \\ \lambda & \text{if } \boldsymbol{x}_n > 0, \end{cases} \tag{1.74}$$

et son opérateur proximal est donné par le seuillage doux :

$$\forall \boldsymbol{x} \in \mathbb{R}^N, \quad \operatorname{prox}_{\|\cdot\|_{\ell_1}}(\boldsymbol{x}) = \left(\max\left\{0, 1 - \frac{\lambda}{|\boldsymbol{x}_n|}\right\}\boldsymbol{x}_n\right)_{n\in\{1,\dots N\}}. \tag{1.75}$$

Dans le cas de la norme $\ell_{1,2}$ défini dans l'équation (1.49), son opérateur proximal est :

$$\forall \boldsymbol{x} \in (\mathbb{R}^2)^N, \quad \operatorname{prox}_{\|\cdot\|_{\ell_{1,2}}}(\boldsymbol{x}) = \left(\max\left\{0, 1 - \frac{\lambda}{|\boldsymbol{x}_n|_2}\right\}\boldsymbol{x}_n\right)_{n\in\{1,\dots N\}}. \tag{1.76}$$

3. Voir définition A.1.1

Dans le cas de l'approximation hyperbolique de ℓ_1, qui est différentiable, la sous-différentielle est réduite à son gradient donné pour tout $\mu \geq 0$ par :

$$\forall \boldsymbol{x} \in (\mathbb{R}^2)^N, \quad \nabla\left(\sqrt{\|\boldsymbol{x}\|_2^2 + \mu^2} - \mu\right) = \left(\frac{\lambda \boldsymbol{x}_n}{\sqrt{\|\boldsymbol{x}_n\|_2^2 + \mu^2}}\right)_{n\in\{1,\dots N\}}. \tag{1.77}$$

Cela permet d'utiliser son gradient pour la recherche de minimum.

Dans le cas de la norme de Huber, sa sous-différentielle est donnée par son gradient, donné par :

$$\forall \boldsymbol{x} \in (\mathbb{R}^2)^N, \forall n \in \{1,\dots N\}, \quad \left(\nabla\|\boldsymbol{x}\|_{\text{Huber}}\right)_n = \begin{cases} \text{Signe}(\boldsymbol{x}_n), & \text{si } \|\boldsymbol{x}_n\|^2 \geq \mu \\ \frac{\boldsymbol{x}_n}{\mu} & \text{sinon.} \end{cases} \tag{1.78}$$

La figure 1.8 représente ces sous-différentielles pour différentes valeurs de μ et de λ. On voit que dans la transition du passage quadratique au passage ℓ_1 est plus douce avec l'approximation hyperbolique qu'avec la fonction de Huber.

Dans le cas de la norme de Shatten sur l'opérateur hessien, la régularisation n'est pas différentiable, il est donc également nécessaire de se rapporter à sa sous-différentielle lors de la recherche de minimum à partir d'itérations basées sur l'opérateur proximal. L'opérateur proxi-

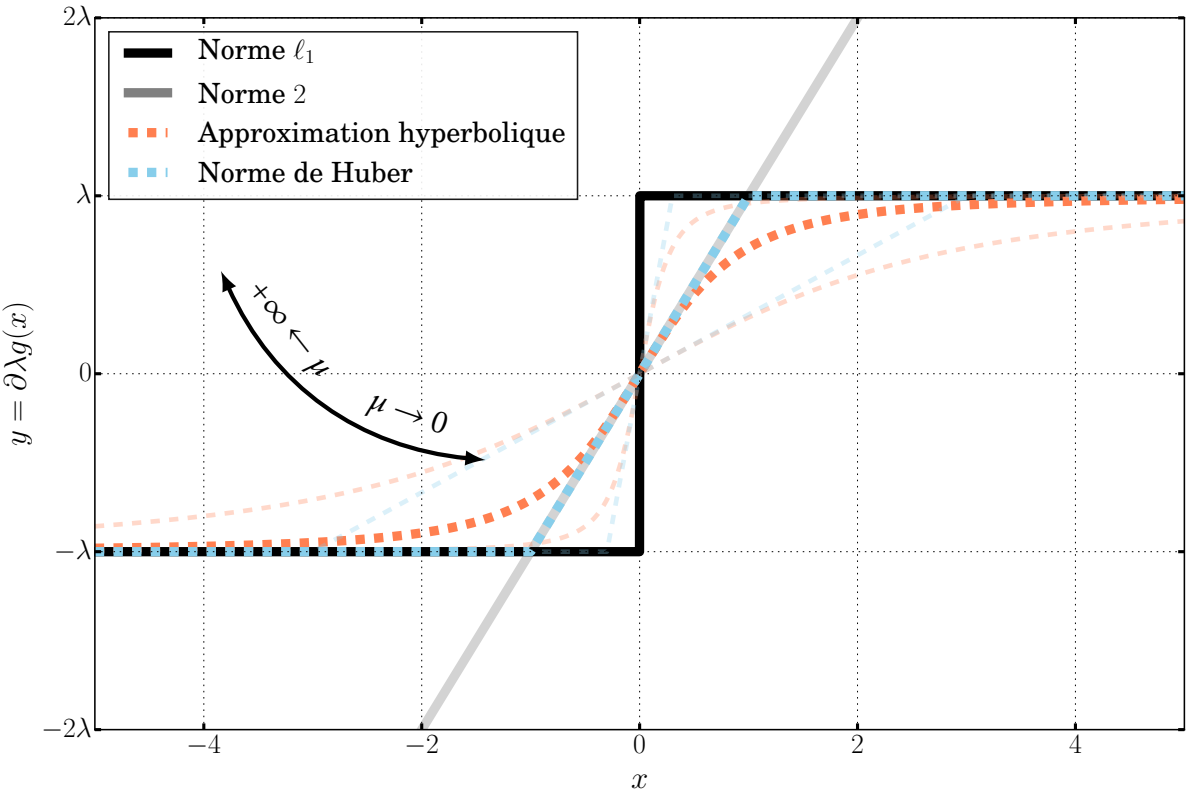


FIGURE 1.8 – Comparaison des sous-différentielles de la norme ℓ_2, de la norme ℓ_1, de son approximation hyperbolique et de la fonction de Huber pour différentes valeurs de μ.

mal de la norme de Shatten ℓ_1 est donnée pour toute matrice $\mathbf{D} \in \mathcal{M}_2(\mathbb{R})$, dont la décomposition en valeurs singulières est donnée par $\mathbf{D} = \mathbf{U}\,\mathrm{Diag}(s_1, s_2)\mathbf{V}^\top$, par [Lefkimmiatis et al., 2013, Chierchia et al., 2013] :

$$\mathrm{prox}_{\mathcal{S}_1}(\mathbf{D}) = \mathbf{U}\,\mathrm{prox}_{\|\cdot\|_{\ell_1}}(\mathrm{Diag}(s_1, s_2))\,\mathbf{V}^\top, \tag{1.79}$$

où $\mathrm{prox}_{\|\cdot\|_{\ell_1}}$ est donné par l'équation (1.76). Cela découle de [Bauschke and Combettes, 2011, Corolaire 23.25].

Dans le cas des contraintes strictes, la fonction $\iota_{\mathcal{C}}$ n'est pas différentiable. Pour résoudre le problème (1.32), il est donc nécessaire de définir les opérateurs proximaux associés, qui dans ce cas se réduit à la projection orthogonale sur l'ensemble $\mathcal{C}$ [Theodoridis et al., 2011], définie comme

$$\mathbb{P}_{\mathcal{C}}(\boldsymbol{x}) = \underset{\boldsymbol{y}\in\mathcal{C}}{\mathrm{argmin}}\ \|\boldsymbol{x} - \boldsymbol{y}\|_2^2. \tag{1.80}$$

1.3.2 Méthodes différentiables : Quasi-Newton et ℓ-BFGS

Dans le cas où Φ et $\mathcal{R}$ sont différentiables et $\mathcal{C} = \mathbb{R}^L$, alors en posant $\Psi(\boldsymbol{x}) = \Phi(\boldsymbol{x}) + \mathcal{R}(\boldsymbol{x})$ pour tout $\boldsymbol{x} \in \mathcal{C}$, trouver le minimum de (1.32) revient à trouver $\widehat{\boldsymbol{x}} \in \mathbb{R}^L$ tel que $\nabla\mathcal{F}(\widehat{\boldsymbol{x}}) = 0$. La méthode la plus classique pour trouver $\widehat{\boldsymbol{x}}$ est la méthode du gradient explicite, dont l'itération principale peut être écrite comme :

$$\forall t \geq 0 \quad \boldsymbol{x}^{[t+1]} = \boldsymbol{x}^{[t]} + \alpha^{[t]}\boldsymbol{r}^{[t]}; \tag{1.81}$$

où $\alpha^{[t]}$ est le pas de descente et $\boldsymbol{r}^{[t]}$ la direction de descente.

Le choix classique de $\alpha^{[t]} \geq 0$ est de le prendre tel qu'il minimise le critère, en l'itérée suivante, dans la direction $\boldsymbol{r}^{[t]}$.

Le choix de la direction de descente consiste à prendre $\boldsymbol{r}^{[t]} = -\boldsymbol{M}^{[t]}\nabla\mathcal{F}(\boldsymbol{x}^{[t]})$ où $\boldsymbol{M}^{[t]}$ est une matrice symétrique définie positive. C'est un choix judicieux de la matrice $\boldsymbol{M}^{[t]}$ combiné à un choix judicieux de pas $\alpha^{[t]}$ qui va accélérer la descente.

Le choix le plus simple consiste à prendre la matrice $\boldsymbol{M}^{[t]} = \mathbf{Id}$, c'est la méthode de la plus grande pente. On a alors $\boldsymbol{r}^{[t]} = -\nabla\mathcal{F}(\boldsymbol{x}^{[t]})$.

Dans le cas où $\nabla^2\mathcal{F}$ est définie positive, la méthode quasi-Newton qui consiste à choisir $\mathbf{M}^{[t]} = \left(\nabla^2\mathcal{F}\right)^{-1}$ améliore le taux de convergence.

Pour éviter de calculer $\nabla^2\mathcal{F}$ ou dans le cas où la hessienne n'est pas inversible, il est possible d'approximer l'inverse de la hessienne. C'est ce que fait la méthode de Broyden-Fletcher-Goldfarb-Shanno (BFGS) qui calcule cet inverse à chaque itération, à partir de l'itération précédente [Nocedal and Wright, 1999].

De nombreuses autres méthodes existent pour faire cette approximation, mais BFGS reste la plus efficace. Elle est de plus invariante aux changements de variables linéaires, ce qui peut être très pratique lorsque $\mathcal{F}$ peut s'exprimer de différentes façons selon le choix de paramètres à estimer, ce qui est le cas dans cette thèse.

L'idée de BFGS vient du fait que pour deux itérations successives, d'après le développement de Taylor du gradient on a :

$$\nabla\mathcal{F}(\boldsymbol{x}^{[t+1]}) \approx \nabla\mathcal{F}(\boldsymbol{x}^{[t]}) + \left(\nabla^2\mathcal{F}(\boldsymbol{x}^{[t]})\right)\left(\boldsymbol{x}^{[t+1]} - \boldsymbol{x}^{[t]}\right) \tag{1.82}$$

On note $\boldsymbol{y}^{[t]} = \nabla\mathcal{F}(\boldsymbol{x}^{[t+1]}) - \nabla\mathcal{F}(\boldsymbol{x}^{[t]})$ et $\boldsymbol{s}^{[t]} = \boldsymbol{x}^{[t+1]} - \boldsymbol{x}^{[t]}$, on a donc la condition $\boldsymbol{y}^{[t]} = \left(\nabla^2\mathcal{F}(\boldsymbol{x}^{[t]})\right)\boldsymbol{s}^{[t]}$ et donc chaque approximation $\nabla^2\mathcal{F}^{[t]}$ doit vérifier cette relation. De même à chaque itération, en posant $\mathbf{H}^{[t]}$ est approximativement égale à l'inverse de la matrice hessienne $\nabla^2\mathcal{F}^{[t]}$, ce qui amène à imposer $\mathbf{H}^{[t+1]}\boldsymbol{y}^{[t]} = \boldsymbol{s}^{[t]}$.

De plus, l'itération $\mathbf{H}^{[t+1]}$ doit être symétrique définie positive et suffisamment proche de $\mathbf{H}^{[t]}$. Pour cela on choisit $\mathbf{H}^{[t+1]}$ comme solution au problème de minimisation suivant :

$$\min_{\mathbf{H}} \|\mathbf{H} - \mathbf{H}^{[t]}\|_{F_W} \quad \text{sous contraintes} \quad \begin{cases} \mathbf{H} = \mathbf{H}^\top, \\ \mathbf{H}\boldsymbol{y}^{[t]} = \boldsymbol{s}^{[t]}, \end{cases} \tag{1.83}$$

où $\|\cdot\|_{F_W} = \|\mathbf{W}^{1/2}\cdot\mathbf{W}^{1/2}\|_F$ avec $\|.\|_F$ la norme de Frobenius, et $\mathbf{W}$ vérifiant

$$\mathbf{W}\boldsymbol{s}^{[t]} = \boldsymbol{y}^{[t]}. \tag{1.84}$$

L'unique solution de ce problème donne l'estimation de $\mathbf{H}^{[t+1]}$ suivante :

$$\mathbf{H}^{[t+1]} = \left(\mathbf{Id} - \frac{\boldsymbol{s}^{[t]}\boldsymbol{y}^{[t]\top}}{\langle \boldsymbol{y}^{[t]}, \boldsymbol{s}^{[t]}\rangle}\right)\mathbf{H}^{[t]}\left(\mathbf{Id} - \frac{\boldsymbol{y}^{[t]}\boldsymbol{s}^{[t]\top}}{\langle \boldsymbol{y}^{[t]}, \boldsymbol{s}^{[t]}\rangle}\right) + \frac{\boldsymbol{s}^{[t]}\boldsymbol{s}^{[t]\top}}{\langle \boldsymbol{y}^{[t]}, \boldsymbol{s}^{[t]}\rangle} \tag{1.85}$$

Le problème du calcul d'une telle matrice est que c'est une matrice de taille L^2 où L est la dimension des paramètres à estimer. Dans le cas où L est grand, c'est-à-dire supérieur à 10^3, le stockage d'une telle matrice est trop couteux. En 1980, Nocedal propose alors une version à mémoire limitée de BFGS, appelée ℓ-BFGS, permettant de calculer $\boldsymbol{r}^{[t+1]} = \mathbf{H}^{[t]}\nabla\mathcal{F}(\boldsymbol{x}^{[t]}$ en ne stockant que les m dernières itérations [Nocedal, 1980]. L'avantage d'une telle méthode est le stockage de seulement $m \times L$ éléments au lieu de L^2.

L'algorithme ℓ-BFGS consiste en l'étape de récursion pour calculer $\mathbf{r}^{[t]}$ dans l'algorithme 1 présenté ci-dessous. La convergence globale d'un tel algorithme fut démontrée dans le cas de problèmes strictement convexes dans [Liu and Nocedal, 1989].

On choisit le paramètre $\gamma^{[t]}$ comme

$$\gamma^{[t]} = \frac{s^{[t-1]\top}y^{[t-1]}}{y^{[t-1]\top}y^{[t-1]}},$$

qui minimise la norme Euclidienne entre $\gamma y^{[t-1]}$ et $s^{[t-1]}$ [Nocedal and Wright, 1999]. Le choix, de $\gamma^{[0]}$ et la possibilité de changer $\mathbf{H}^{[0]}$ en fonction de la courbure fait toute la force de cet algorithme.

Il est également possible d'intégrer des contraintes de bornes séparables, pouvant s'exprimer comme dans l'équation 1.61. Il s'agit alors de l'algorithme VMLM-B présenté dans [Thiébaut, 2002]. L'algorithme ℓ-BFGS [Nocedal and Wright, 1999] est plus connu mais plus coûteux en calculs et moins à même de faire face à des problèmes fortement non-linéaires.

Le respect de la contrainte est vérifié au moment de la recherche de la direction de descente. Celle-ci est alors choisie de manière à respecter les bornes.

L'avantage d'une méthode différentiable est que la convergence locale est assurée dans un certain voisinage de la solution, même lorsque le problème n'est pas convexe. Il est donc possible d'utiliser cette méthode pour estimer des paramètres de modèles non-linéaires, à partir de modèles non-convexes, dont les reconstructions auraient une erreur moindre par rapport à des modèles linéaires convexes.

Algorithme 1 : Algorithme à métrique variable et mémoire limitée (VMLM : Variable Metric Limited Memory)

Initialiser $\boldsymbol{x}^{[0]}(\mathbb{R}^N)^L$, $\mathbf{H}^{[0]} = \gamma^{[0]}\mathbf{Id}$, $m > 2$:
pour $t = 0, 1, \ldots$ **faire**
  $\boldsymbol{q} = \nabla\mathcal{F}(\boldsymbol{x}^{[t]})$
  pour $i = t-1, t-2, \ldots, t-m$ **faire**
    $\rho^{[i]} = 1/\langle \boldsymbol{y}^{[i]}, \boldsymbol{s}^{[i]} \rangle$
    $a^{[i]} = \rho^{[i]} \boldsymbol{s}^{[i]\top} \boldsymbol{q}$
    $\boldsymbol{q} = \boldsymbol{q} - a^{[i]} \boldsymbol{y}^{[i]}$
  $\boldsymbol{r} = \mathbf{H}^{[0]} \boldsymbol{q}$
  pour $i = t-m, t-m+1, \ldots, t-1$ **faire**
    $\rho^{[i]} = 1/\langle \boldsymbol{y}^{[i]}, \boldsymbol{s}^{[i]} \rangle$
    $b = \rho^{[i]} (\boldsymbol{y}^{[i]})^\top \boldsymbol{r}$
    $\boldsymbol{r} = \boldsymbol{r} + \left(a^{[i]} - b\right) \boldsymbol{s}^{[i]}$
  Trouver $\alpha^{[t]}$ optimal vérifiant les conditions de Wolfe
  $\boldsymbol{x}^{[t+1]} = \boldsymbol{x}^{[t]} - \alpha^{[t]} \boldsymbol{r}$

Cependant, dans le cas où la contrainte n'est plus séparable, typiquement le cas d'une contrainte épigraphique, trouver la direction de descente respectant une telle contrainte est plus difficile, la solution est alors de se tourner vers la méthode du gradient projeté, qui peut être vu comme un cas particulier de l'algorithme Forward-Backward.

1.3.3 L'algorithme Forward-Backward à métrique variable

Dans le cas où Φ et de $\mathcal{R}$ sont différentiables et $\mathcal{C}$ est une contrainte non-séparable, seul le cas spécifique de l'algorithme Forward-Backward correspondant au gradient projeté nous intéresse. L'algorithme est alors présenté dans ce contexte. Pour traiter des régularisations $\mathcal{R}$ non-différentiables, un algorithme primal-dual sera considéré dans la sous-section suivante.

En posant $\mathcal{F}(\boldsymbol{x}) = \Phi(\boldsymbol{x}) + \mathcal{R}(\boldsymbol{x})$ pour tout $\boldsymbol{x} \in C$, résoudre (1.32) revient à trouver un point fixe de la projection sur $\mathcal{C}$ de l'étape de descente de gradient (1.81). En effet, posons $\mathcal{G} = \iota_{\mathcal{C}}$, trouver le minimum $\widehat{\boldsymbol{x}} \in \mathbb{R}^L$ de (1.32) revient à résoudre :

$$
\begin{aligned}
\widehat{\boldsymbol{x}} \in \underset{\boldsymbol{x}\in\mathbb{R}^L}{\operatorname{Argmin}}\ \mathcal{F} + \mathcal{G} &\Leftrightarrow 0 \in \partial\Big(\nabla\mathcal{F}(\widehat{\boldsymbol{x}}) + \partial\mathcal{G}(\widehat{\boldsymbol{x}}\Big) \\
&\Leftrightarrow 0 \in \Big\{\nabla\mathcal{F}(\widehat{\boldsymbol{x}}) + \partial\mathcal{G}(\widehat{\boldsymbol{x}}\Big\} \\
&\Leftrightarrow (\widehat{\boldsymbol{x}} - \widehat{\boldsymbol{x}}) \in \Big\{\nabla\mathcal{F}(\widehat{\boldsymbol{x}}) + \partial\mathcal{G}(\widehat{\boldsymbol{x}})\Big\} \\
&\Leftrightarrow \Big(\widehat{\boldsymbol{x}} - \nabla\mathcal{F}(\widehat{\boldsymbol{x}}) - \widehat{\boldsymbol{x}}\Big) \in \partial\mathcal{G}(\widehat{\boldsymbol{x}}) \\
&\Leftrightarrow \widehat{\boldsymbol{x}} = \operatorname{prox}_{\mathcal{G}}(\widehat{\boldsymbol{x}} - \nabla\mathcal{F}(\widehat{\boldsymbol{x}})) \qquad \text{d'après (1.72)} \\
&\Leftrightarrow \widehat{\boldsymbol{x}} \in \mathbf{Fix}\ \operatorname{prox}_{\mathcal{G}}(\boldsymbol{x} - \nabla\mathcal{F}(\boldsymbol{x})). \qquad (1.86)
\end{aligned}
$$

Soit β la constante de Lipschitz de $\nabla\mathcal{F}$, c'est-à-dire telle que :

$$
\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{H} \times \mathcal{H} \|\nabla\mathcal{F}(\boldsymbol{x}) - \nabla\mathcal{F}(\boldsymbol{y})\| \leq \beta \|\boldsymbol{x} - \boldsymbol{y}\|, \qquad (1.87)
$$

et $(\mathbf{P}^{[t]})_{t\geq 0} \in \mathcal{P}_\rho(\mathbb{R}^L)$ une suite de matrices symétriques ρ-positives de préconditionnement, où l'ensemble $\mathcal{P}_\rho(\mathbb{R}^L)$ est défini par l'équation (A.3). Soit $\mu = \sup_{t\geq 0} \|\mathbf{P}^{[t]}\|_{t\geq 0}\|$. Soit $\mathbf{P} \in \mathcal{P}_0(\mathbb{R}^L)$, et $\mathcal{GH} \to \mathbb{R}$ propre, convexe et semi-continue inférieurement :

$$\forall \boldsymbol{x} \in \mathcal{H}, \quad \operatorname{prox}_{\mathcal{G}}^{\mathbf{P}}(\boldsymbol{x}) = \underset{\boldsymbol{y}\in\mathcal{H}}{\operatorname{argmin}} \ \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{y}\|_{\mathbf{P}}^2 + \mathcal{G}(\boldsymbol{y}), \tag{1.88}$$

où $\|\cdot\|_{\mathbf{P}}^2 = \langle \cdot, \mathbf{P}\cdot\rangle$. Alors, le problème de point fixe (1.86) peut être résolu par l'algorithme Forward-Backward à métrique variable dont les itérations sont présentées dans l'algorithme 2.

Algorithme 2 : Variable Metric Forward-Backward (VMFB)

Soit $\boldsymbol{x}^{[0]} \in \mathcal{H}$ et $\gamma \in]0, 2/\beta\mu[$, alors :
pour $t = 0, 1, \ldots$ **faire**
$\quad \boldsymbol{x}^{[t+1]} = \operatorname{prox}_{\gamma\mathcal{G}}^{\mathbf{P}^{[t]}}\left(\boldsymbol{x}^{[t]} - \gamma\mathbf{P}^{[t]}\nabla\mathcal{F}(\boldsymbol{x}^{[t]})\right)$.

Les garanties de convergences pour un tel algorithme sont données par le théorème suivant :

Théoreme 1.3.3 ([Combettes and Vũ, 2013]). *Soit $\mathcal{F}, \mathcal{G} \in \Gamma_0(\mathcal{H}, \mathcal{K})$, où $\nabla\mathcal{F}$ β-Lipschitz. Soit $\rho \in]0, +\infty[$ et $(\mathbf{P}^{[t]})_{t\in\mathbb{N}}$ une suite de matrices dans $\mathcal{P}_\rho(\mathcal{H})$ telles que*

$$(1 + \eta^{[t]})\mathbf{P}^{[t+1]} \geq \mathbf{P}^{[t]}, \tag{1.89}$$

avec la suite $(\eta^{[t]})_{t\in\mathbb{N}} \in \mathbb{R}_+$ tel que $\sum_t |\eta^{[t]}| < \infty$. Soit $\mu = \sup_{t\in\mathbb{N}} \|\mathbf{P}^{[t]}\|$ et $\gamma \in]0, 2/\beta\mu[$. Alors, la suite d'itérations $(\boldsymbol{x}^{[t]})_{t\in\mathbb{N}}$ de l'algorithme 2 converge vers une solution $\widehat{\boldsymbol{x}} \in \operatorname{Argmin} \mathcal{F} + \mathcal{G}$.

La preuve de ce théorème peut-être trouvée en annexe page 198.

En prenant $\mathbf{P}^{[t]} = \mathbf{Id}$ pour tout $t \geq 0$, on retrouve les itérations classiques de Forward-Backward.

Les conditions de convergences de l'algorithme 2 reposent sur la constante de Lipschitz β du gradient de $\mathcal{F}$. Dans le cas linéaire, c'est-à-dire où $\nabla^2\mathcal{F} = \mathbf{A}$ avec $\mathbf{A}$ une matrice symétrique définie positive, le calcul de β peut se faire à l'aide de la méthode de la puissance dont les itérations sont rappelées par l'algorithme 3. En effet, dans un tel cas on a

$$\beta = \|\mathbf{A}\| = \sqrt{\|\mathbf{A}\|^2} = \sqrt{\|\mathbf{A}^*\mathbf{A}\|} = \sqrt{\lambda_{\max}(\mathbf{A}^*\mathbf{A})},$$

où * représente l'adjoint de l'opérateur. La méthode de la puissance permettant de calculer la valeur propre maximale d'un opérateur $\mathbf{G} \in \mathbb{R}^G$, il est donc possible d'utiliser ces itérations en posant $\mathbf{G} = \mathbf{A}^*\mathbf{A}$.

Pour un opérateur linéaire $G \in \mathbb{R}^G$, les itérations de la méthode de la puissance sont don-

nées par l'algorithme suivant :

Algorithme 3 : Méthode de la puissance

Soit $\boldsymbol{x}^{[0]} \in \mathcal{N}(0, Id_L)$ et $\varepsilon \geq 0$.
pour $t = 0, 1, \ldots$ **faire**
  Initialiser $\rho^{[0]} = 0$;
  tant que $|\rho^{[t]} - \rho^{[t-1]}|/\rho^{[t]} \geq \varepsilon$ **faire**
    $\boldsymbol{x}^{[t+1]} = \mathbf{G}^*\mathbf{G}\boldsymbol{x}_k$;
    $\rho^{[t+1]} = \|\boldsymbol{x}^{[t+1]}\|_2^2/\|\boldsymbol{x}^{[t]}\|_2^2$;
  Retourner $\rho^{[t]}$;

Dans le cas de l'algorithme 2 on a alors $\beta = \sqrt{\rho^{[t]}}$.

1.3.4 Les algorithmes primaux-duaux et l'algorithme Condat-Vũ à métrique variable

Dans le cas où $\mathcal{R}(\boldsymbol{x}) = \lambda g(\mathbf{G}\boldsymbol{x})$ n'est plus différentiable, l'utilisation du schéma Forward-Backward est plus compliquée. En effet, si une forme explicite de l'opérateur proximal de g peut souvent exister, ce n'est pas toujours le cas pour sa composition avec un opérateur linéaire $\mathbf{G}$. Dans une tel situation, la pratique usuelle est de résoudre le problème dual associé au problème 1.32, donné par :

$$\widehat{\boldsymbol{y}} \in \underset{\boldsymbol{y}}{\operatorname{Argmin}} \left\{ [(\iota_{\mathcal{C}} + \Phi)^*(-\mathbf{G}^*\boldsymbol{y}) + g^*(\boldsymbol{y}) \right\}, \tag{1.90}$$

où * représente le dual des fonctions et l'adjoint de l'opérateur linéaire. Cependant cette formulation peut poser problème lorsque Φ est composée d'un opérateur linéaire non-inversible, car Φ^* fait intervenir l'inverse de cet opérateur.

Dans ce cas, il est d'usage de résoudre le problème de point selle :

$$\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}} \in \underset{\boldsymbol{x} \in \mathbb{R}^L}{\operatorname{Argmin}} \max_{\boldsymbol{y} \in \mathbb{R}^G} \left\{ (\iota_{\mathcal{C}} + \Phi)(\boldsymbol{x}) - g^*(\boldsymbol{y}) + \langle \mathbf{G}\boldsymbol{x}, \boldsymbol{y} \rangle \right\}. \tag{1.91}$$

Un tel problème peut alors être résolu par l'algorithme primal-dual à métrique variable de Condat-Vũ [Condat, 2013, Vũ, 2015] dont les itérations sont les suivantes :

Algorithme 4 : Algorithme à métrique variable de Condat-Vu (VMCV)

$\boldsymbol{x}^{[0]} \in \mathcal{H}$ et $\boldsymbol{y}^{[0]} \in \mathcal{K}$.
Soit $\mathbf{U} > 0$ et $\sigma^{[t]} > 0$.
pour $t = 0, 1, \ldots$ **faire**
  $\boldsymbol{x}^{[t+1]} = \operatorname{prox}_{\iota_{\mathcal{C}}}^{\mathbf{U}^{[t]}} \left(\boldsymbol{x}^{[t]} - \mathbf{U}^{[t]} \left(\nabla\Phi(\boldsymbol{x}^{[t]}) + \mathbf{G}^*\boldsymbol{y}^{[t]} \right) \right)$;
  $\boldsymbol{y}^{[t+1]} = \operatorname{prox}_{\sigma^{[t]} g^*} \left(\boldsymbol{y}^{[t]} + \sigma^{[t]}\mathbf{G} \left(2\boldsymbol{x}^{[t+1]} - \boldsymbol{x}^{[t]} \right) \right)$;

Nous donnons maintenant quelques conditions nécessaires à la convergence.

Proposition 1.3.4 ([Vũ, 2015]). *Soit Φ, g, $\iota_{\mathcal{C}}$ des fonctions convexes, semi-continues inférieurement et propres telles que Φ est β-Lipschitz et différentiable avec $\beta > 0$. Soit $\mathbf{U}^{[t]}$ une matrice symetrique définie positive et $\sigma_t > 0$, tel que $\mathbf{U}^{[t+1]} \geq \mathbf{U}^{[t]}$ et $\sigma^{[t+1]} \leq \sigma^{[t]}$. Alors, si*

$$\frac{1 - \|\sqrt{\sigma^{[t]}}\mathbf{G}\sqrt{\mathbf{U}^{[t]}}\|}{\max\{\|\mathbf{U}^{[t]}\|, \sigma^{[t]}\}} \geq \beta/2, \tag{1.92}$$

la suite $(\boldsymbol{x}^{[t]})_{t>0}$ *de l'algorithme 4 converge vers* $\widehat{\boldsymbol{x}}$*, une solution de* (1.32).

La preuve de cette proposition peut être trouvée en annexe à la page 202.

Remarque : L'algorithme 4 peut être vu comme une itération de l'algorithme 4 :

$$\boldsymbol{z}^{[t+1]} = \operatorname{prox}^{\mathbf{P}^{[t]}}_{\partial g^*}\left(\boldsymbol{z}^{[t]} - \mathbf{P}^{[t]}\boldsymbol{\nabla\Phi}(\boldsymbol{z}^{[t]})\right) \tag{1.93}$$

avec :

$$\boldsymbol{z}^{[t]} = \begin{pmatrix}\boldsymbol{x}^{[t]}\\ \boldsymbol{y}^{[t]}\end{pmatrix}, \quad \partial \boldsymbol{g}^* = \begin{pmatrix}\partial \iota_{\mathcal{C}} & \mathbf{G}^*\\ -\mathbf{G} & \partial g^*\end{pmatrix}, \quad \boldsymbol{\nabla\Phi} = \begin{pmatrix}\nabla\Phi & 0\\ 0 & 0\end{pmatrix} \quad \text{and} \quad \mathbf{P}^{[t]} = \begin{pmatrix}\mathbf{U}^{[t]^{-1}} & -\mathbf{G}^*\\ -\mathbf{G} & \sigma^{[t]^{-1}}\end{pmatrix}^{-1}, \tag{1.94}$$

En effet, on a :

$$\begin{cases}\boldsymbol{x}^{[t+1]} = \left(\mathbf{Id} + \mathbf{U}^{[t]}\partial\iota_{\mathcal{C}}\right)^{-1}\left(\boldsymbol{x}^{[t]} - \mathbf{U}^{[t]}\left(\nabla\Phi(\boldsymbol{x}^{[t]}) - \mathbf{G}^*\boldsymbol{y}^{[t]}\right)\right),\\ \boldsymbol{y}^{[t+1]} = \left(\mathbf{Id} + \sigma^{[t]}\partial g^*\right)^{-1}\left(\boldsymbol{y}^{[t]} + \sigma^{[t]}G\left(2\boldsymbol{x}^{[t+1]} - \boldsymbol{x}^{[t]}\right)\right)\end{cases} \tag{1.95}$$

$$\Leftrightarrow \begin{cases}\left(\mathbf{Id} + \mathbf{U}^{[t]}\partial\iota_{\mathcal{C}}\right)\boldsymbol{x}^{[t+1]} \ni \boldsymbol{x}^{[t]} - \mathbf{U}^{[t]}\left(\nabla\Phi(\boldsymbol{x}^{[t]}) - \mathbf{G}^*\boldsymbol{y}^{[t]}\right),\\ \left(\mathbf{Id} + \sigma^{[t]}\partial g^*\right)\boldsymbol{y}^{[t+1]} \ni \boldsymbol{y}^{[t]} + \sigma^{[t]}\mathbf{G}\left(2\boldsymbol{x}^{[t+1]} - \boldsymbol{x}^{[t]}\right)\end{cases} \tag{1.96}$$

$$\Leftrightarrow \begin{cases}\boldsymbol{x}^{[t+1]} - \mathbf{U}^{[t]}\partial\iota_{\mathcal{C}}(\boldsymbol{x}^{[t+1]}) \ni \boldsymbol{x}^{[t]} - \mathbf{U}^{[t]}\left(\nabla\Phi(\boldsymbol{x}^{[t]}) - \mathbf{G}^*\boldsymbol{y}^{[t]}\right)\\ \boldsymbol{y}^{[t+1]} + \sigma^{[t]}\partial g^*(\boldsymbol{y}^{[t+1]}) \ni \boldsymbol{y}^{[t]} + \sigma^{[t]}\mathbf{G}\left(2\boldsymbol{x}^{[t+1]} - \boldsymbol{x}^{[t]}\right)\end{cases} \tag{1.97}$$

$$\Leftrightarrow \begin{cases}\mathbf{U}^{[t]^{-1}}\left(\boldsymbol{x}^{[t+1]} - \boldsymbol{x}^{[t]}\right) = -\nabla\Phi(\boldsymbol{x}^{[t]}) + \mathbf{G}^*\boldsymbol{y}^{[t]},\\ \boldsymbol{y}^{[t+1]} + \sigma^{[t]}\partial g^*(\boldsymbol{y}^{[t+1]}) \ni \boldsymbol{y}^{[t]} + \sigma^{[t]}\mathbf{G}\left(2\boldsymbol{x}^{[t+1]} - \boldsymbol{x}^{[t]}\right)\end{cases} \tag{1.98}$$

$$\Leftrightarrow \begin{cases}-\nabla\Phi(\boldsymbol{x}^{[t]}) = \mathbf{U}^{[t]^{-1}}\left(\boldsymbol{x}^{[t+1]} - \boldsymbol{x}^{[t]}\right) - \mathbf{G}^*\boldsymbol{y}^{[t]},\\ 0 \in \boldsymbol{y}^{[t+1]} + \sigma^{[t]}\partial g^*(\boldsymbol{y}^{[t+1]}) - \boldsymbol{y}^{[t]} - \sigma^{[t]}\mathbf{G}\left(2\boldsymbol{x}^{[t+1]} - \boldsymbol{x}^{[t]}\right)\end{cases} \tag{1.99}$$

En divisant la seconde ligne par $\sigma^{[t]}$ et en faisant apparaître $G^*\boldsymbol{y}^{[t+1]}$ dans la première ligne on a :

$$\Leftrightarrow \begin{cases}-\nabla\Phi(\boldsymbol{x}^{[t]}) = \mathbf{U}^{[t]^{-1}}\left(\boldsymbol{x}^{[t+1]} - \boldsymbol{x}^{[t]}\right) - \mathbf{G}^*\boldsymbol{y}^{[t]} + \mathbf{G}^*\boldsymbol{y}^{[t+1]} - \mathbf{G}^*\boldsymbol{y}^{[t+1]},\\ 0 \in \sigma^{[t]^{-1}}\left(\boldsymbol{y}^{[t+1]} - \boldsymbol{y}^{[t]}\right) + \partial g^*(\boldsymbol{y}^{[t+1]}) - \mathbf{G}\boldsymbol{x}^{[t+1]} - \mathbf{G}\left(\boldsymbol{x}^{[t+1]} - \boldsymbol{x}^{[t]}\right)\end{cases} \tag{1.100}$$

$$\Leftrightarrow \boldsymbol{z}^{[t]} - \mathbf{P}^{[t]}\boldsymbol{\nabla\Phi}(\boldsymbol{z}^{[t]}) \in \left(\mathbf{Id} + \mathbf{P}^{[t]}\partial\boldsymbol{g}\right)(\boldsymbol{z}^{[t+1]}), \tag{1.101}$$

$$\Leftrightarrow \boldsymbol{z}^{[t+1]} = \operatorname{prox}^{\mathbf{P}^{[t]}}_{\partial \boldsymbol{g}^*}\left(\boldsymbol{z}^{[t]} - \mathbf{P}^{[t]}\boldsymbol{\nabla\Phi}(\boldsymbol{z}^{[t]})\right). \tag{1.102}$$

Le choix des matrices de préconditionement $(\mathbf{U}^{[t]})_{t\geq 0}$ est un sujet de recherche toujours d'actualité. Dans le cas où pour tout $t \geq 0$, $\mathbf{U}^{[t]} = \tau^{[t]}\mathbf{Id}$ on retrouve alors la forme de l'algorithme présenté par Condat [Condat, 2013]. Un choix de matrice $(\mathbf{U}^{[t]})_{t\geq 0}$ diagonales tel que celui proposé par Lorenz et Pock dans [Lorenz and Pock, 2015] permet d'avoir des expressions explicites des opérateurs proximaux, ce que n'est pas le cas lorsque les matrices $(\mathbf{U}^{[t]})_{t\geq 0}$ pleines, tel que le précondtionneur proposé dans [Li and Zhang, 2016]. Dans un tel cas, il est alors nécessaire de faire des sous-itérations pour calculer l'opérateur proximal à partir de la formulation (1.88).

1.3.5 Leurs utilisations en astrophysique

L'utilisation de méthodes de type *«problèmes inverses »* en astrophysique date de plusieurs décennies. En effet, dès 1985 Titterington propose des approches régularisées pour l'estimation de paramètres en astrophysique [Titterington, 1985]. On retrouve également l'utilisation du maximum de vraisemblance pour la déconvolution aveugle d'images convoluées par une PSF entachée d'aberrations [Thiébaut and Conan, 1995].

Ces méthodes sont cependant principalement utilisées en interférométrie optique, pour la soustraction de tavelures [Borde and Traub, 2006] ou la reconstruction de front d'ondes, et surtout en radio-interférométrie avec le célèbre algorithme CLEAN [Högbom, 1974], qui va engendrer de nombreuses variantes.

Pour ce qui est des méthodes non-lisses, on retrouve l'utilisation de méthodes parcimonieuses en 2003 dans le débruitage à l'aide de curvelets [Starck et al., 2003], lors de l'utilisation d'un algorithme de *«matching pursuit »*. Une autre utilisation des méthodes parcimonieuses fut faite en 2012 en radio-interférométrie avec l'algorithme SARA [Carrillo et al., 2012]. Cette première méthode fut ensuite déclinée en plusieurs variantes selon les applications, comme en particulier Polca-SARA [Birdi et al., 2019] pour l'application en polarimétrie qui utilise une contrainte épigraphique. Cette application est particulièrement intéressante, car l'algorithme utilisé est proche de l'algorithme primal-dual avec contrainte épigraphique présenté dans le chapitre 4. On retrouve de plus des méthodes d'optimisation non-convexe dans le cas de reconstruction jointe à la calibration d'antenne [Repetti et al., 2017]. Dans le cas de l'utilisation de ces algorithmes, se pose de plus en plus la question du réglage automatique des hyperparamètres, dont une application utilisant SURE peut être trouvée pour l'application en déconvolution multispectrale [Ammanouil et al., 2019].

Ces dernières années, l'utilisation de méthodes d'apprentissage et de méthodes inverses a explosé en astrophysique. Que ce soit pour l'estimation du fond diffus cosmologique (CMB : *Cosmic Microwave Background*)[Adam et al., 2016] mais aussi plus récemment l'observation du trou noir supermassif MG87 [Akiyama et al., 2019] pour laquelle de nombreuses méthodes ont été développées.

L'utilisation de méthodes inverses en imagerie directe à haut contraste est cependant beaucoup plus récente et timide. Dans le cas des données SPHERE on cite notement pour la reconstruction des disques en ADI [Pairet et al., 2018b] avec l'IFS, et [Berdeu et al., 2020] pour l'autocalibration des données de l'IFS.

Dans les autres applications des méthodes d'apprentissages, on peut également citer les approches par patch de détection des exoplanètes [Flasseur et al., 2018] ou la reconstruction d'objets étendus [Flasseur et al., 2019], ou encore les travaux de reconstructions à l'aide de réseaux de neurones [Flamary, 2017].

L'utilisation de tels algorithmes pour la résolution de *«problèmes inverses »* n'a jamais encore été faite dans le cas de l'imagerie directe en polarimétrie à haut contraste. Dans cette thèse, je m'appuie à la fois sur des approches différentiable, telles que celles présentées dans les chapitres 2 et 3, mais également sur les derniers résultats en optimisation non-différentiable présentés dans le chapitre 4. Mon but est alors d'illustrer l'apport de chaque méthode, pour la reconstruction des environnements circumstellaires, ainsi que leurs limitations.

Chapitre 2

Modèle direct séparable des données de l'instrument ESO/VLT-SPHERE IRDIS

Dans ce chapitre, je m'intéresse à l'écriture d'un modèle direct séparable des données *pré-traitées*[1] de l'instrument ESO-VLT-SPHERE IRDIS. L'idée est de modéliser par des opérations mathématiques, toute la chaîne de modulation de la polarisation de la lumière, depuis son entrée dans le télescope jusqu'à la caméra lors de son acquisition. Ainsi, il est possible d'avoir une expression des données en fonction du signal d'intérêt.

Dans la suite, je note $L \in \mathbb{N}$ le nombre de composantes du signal d'intérêt et $N \in \mathbb{N}$ le nombre de pixels par composante. Les signaux d'intérêt sont notés $\boldsymbol{x} \in (\mathbb{R}^N)^L$, où les L vecteurs $\boldsymbol{x}_\ell = (\boldsymbol{x}_{\ell,n})_{n\in\{1,\dots,N\}} \in \mathbb{R}^N$ correspondent aux différentes composantes, ou paramètres d'intérêt. L'entrée $\boldsymbol{x}_{\ell,n} \in \mathbb{R}$ représente le pixel $n \in \{1,\dots N\}$ de la composante $\ell \in \{1,\dots L\}$. Les paramètres d'intérêt inconnus, que nous cherchons à estimer, sont notés $\overline{\overline{\boldsymbol{x}}} \in (\mathbb{R}^N)^L$. Leurs estimations sont notées par un chapeau, *i.e.* $\widehat{\boldsymbol{x}} \in (\mathbb{R}^N)^L$.

D'autre part, je note $K \in \mathbb{N}$ le nombre d'acquisitions composant un jeu de données et $M \in \mathbb{N}$ le nombre de pixels de chaque acquisition $k \in \{1,\dots,K\}$. Nous notons par $\boldsymbol{d} \in (\mathbb{R}^M)^K$ les jeux de données, où les vecteurs $\boldsymbol{d}_k = (\boldsymbol{d}_{k,m})_{m\in\{1,\dots,M\}} \in \mathbb{R}^M$ représentent les différentes observations et $\boldsymbol{d}_{k,m} \in \mathbb{R}$ représente le pixel $m \in \{1,\dots M\}$ de l'acquisition $k \in \{1,\dots K\}$.

Le but de la modélisation directe est d'aboutir à une expression des données de l'instrument $\boldsymbol{d} = (\boldsymbol{d}_{k,m})_{k\in\{1,\dots K\},m\in\{1,\dots,M\}}$ en fonction du signal d'intêret $\overline{\overline{\boldsymbol{x}}} = (\overline{\overline{\boldsymbol{x}}}_{\ell,n})_{\ell\in\{1,\dots L\},n\in\{1,\dots,N\}}$ sous la forme :

$$\boldsymbol{d} = \mathcal{B}(f(\overline{\overline{\boldsymbol{x}}})), \tag{2.1}$$

où $f : (\mathbb{R}^N)^L \to (\mathbb{R}^M)^K$ est un opérateur de déformation du signal fixé et $\mathcal{B} : (\mathbb{R}^M)^K \to (\mathbb{R}^M)^K$ un opérateur de dégradation aléatoire.

Dans le cadre des données *pré-traitées*, un pixel des données $m \in \{1,\dots M\}$ est une combinaison des L composantes du signal d'intérêt en un unique pixel $n \in \{1,\dots N\}$. Par abus de notation dans la suite de ce chapitre on notera j,n le pixel m de la partie j du détecteur, où $j = 1$ correspond à la partie gauche et $j = 2$ à la partie droite (*cf.* 1.1.4 pour le fonctionnement de l'instrument IRDIS). On notera le modèle séparable des données :

$$\boldsymbol{d}^{\mathrm{S}}_{j,k,n} = \mathcal{B}_{j,k,n}(f^{\mathrm{S}}_{j,k}(\overline{\overline{\boldsymbol{x}}}_n)), \tag{2.2}$$

1. Pour avoir des données *pré-traitées*, les pixels défectueux identifiés des données *calibrées* ont été interpolés puis les images ont été découpées et recentrées afin que les centres des étoiles coïncident (cf. sections 1.1.4 et 6.1).

où $f^{\mathrm{S}}_{j,k} : \mathbb{R}^L \to \mathbb{R}$ modélise l'instrument supposé connu, et $\mathcal{B}_{j,k,n} : \mathbb{R} \to \mathbb{R}$ représente le bruit de mesure. Finalement, la modélisation séparable consiste à identifier et établir l'ensemble des fonctions $(f^{\mathrm{S}}_{j,k})_{j\in\{1,2\},k\in\{1,\dots K\}}$, ainsi que la statistique du bruit de mesure $(\mathcal{B}_{j,k,n})_{j\in\{j\in\{1,2\},k\in\{1,\dots K\},n\in\{1,\dots N\}}$ de la formulation (2.2). Le choix du formalisme, ici Jones ou Stokes, aura une influence sur la forme des $(f^{\mathrm{S}}_{j,k})_{j\in\{1,2\},k\in\{1,\dots K\}}$ et sur les paramètres composants le signal d'intérêt $\boldsymbol{x}$.

Ce chapitre est organisé comme suit. Dans la section 2.1, je présente un premier modèle direct séparable des données *pré-traitées* de l'instrument, basé sur le formalisme de Jones, paramétré en I^{u}, I^{p} et θ, conduisant au signal d'intérêt $\boldsymbol{x} = (I^{\mathrm{u}}, I^{\mathrm{p}}, \theta)^\top$. En considérant ce formalisme, une méthode d'estimation hiérarchique des paramètres de ce modèle non-linéaire séparable, appelée Méthode non-Linéaire Séparable (MnLS), est présentée. Enfin, j'étudie et compare les performances de la MnLS avec les méthodes de l'état-de-l'art présentées dans la sous-section 1.1.5.

Dans la section 2.2, je propose une reformulation linéaire du modèle direct basée sur le formalisme de Stokes. Cela constitue un second modèle direct paramétré en I, Q et U, conduisant au signal d'intérêt $\boldsymbol{x} = (\mathrm{I}, \mathrm{Q}, \mathrm{U})^\top$. Je propose ensuite une méthode d'estimation des paramètres de ce modèle linéaire séparable, appelée Méthode Linéaire Séparable (MLS). J'étudie et compare les performances de la MLS avec la MnLS et les méthodes de l'état-de-l'art.

2.1 MnLS : Modèle non-Linéaire Séparable avec le formalisme de Jones

2.1.1 Le modèle direct non-linéaire séparable

Dans cette section, je développe le calcul permettant d'obtenir un modèle direct des données de l'instrument ESO/VLT-SPHERE IRDIS, à partir du formalisme de Jones. Ce choix de formalisme, plutôt que l'utilisation du formalisme de Mueller pour un modèle paramétré par les paramètres de Stokes, vient d'un désir initial de comprendre la physique de l'instrument et son action sur la lumière, représentée par son champ électrique (cf. 1.1.5).

Soit un champ électrique complexe $\boldsymbol{e}_n \in \mathbb{C}^2$, où $n \in \{1, \dots, N\}$ est la discrétisation du champ à l'entrée du télescope. Pour modéliser le comportement de $\boldsymbol{e}_n$ à travers le télescope, de l'instrument SPHERE et de l'instrument IRDIS jusqu'au détecteur, la chaîne instrumentale est décomposée en plusieurs étapes, illustrées sur la figure 2.1 et présentées ci-dessous. Pour tout $n \in \{1, \dots, N\}$:

1. Le champ électrique passe par un premier ensemble de miroirs qui induit une première

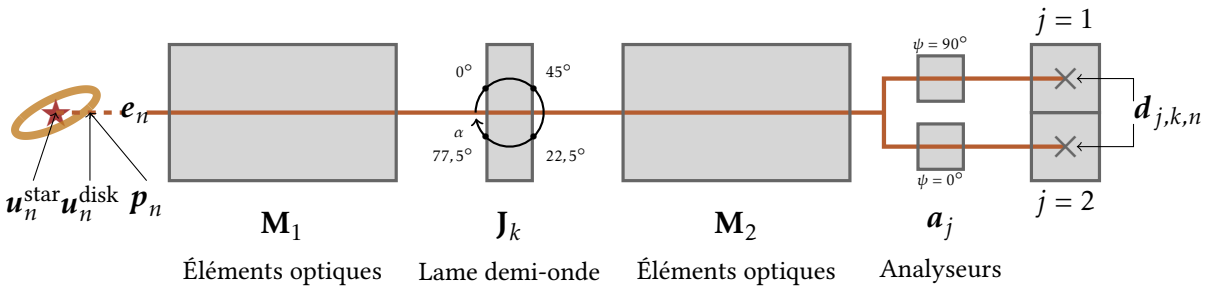


Figure 2.1 – Shéma simplifié de l'instrument ESO/VLT-SPHERE IRDIS.

transformation. Celle-ci est modélisée par une matrice $\mathbf{M}_1 \in \mathcal{M}(\mathbb{C}^2)$, définie par l'équation (1.7), et représente l'atténuation et le déphasage des miroirs. Le champ électrique après passage par l'ensemble de ces éléments optiques est donné par $\mathbf{M}_1 \boldsymbol{e}_n$.

2. Le champ résultant passe ensuite par la lame demi-onde qui, au fil des acquisitions, tourne selon le cycle d'angle $\alpha \in \{0° ; 45° ; 22,5° ; 77,5°\}$. Pour l'acquisition $k \in \{1, \dots K\}$, la lame demi-onde est tournée d'un angle α par rapport à l'axe principal, ce qui induit un déphasage connu dépendant de α. Celui-ci est modélisé par une matrice $\mathbf{J}_k \in \mathcal{S}(\mathbb{R}^2)$ définie par l'équation (1.9), paramétrée par l'orientation α de la lame-demi-onde à l'acquisition k. On note dans la suite $k_\alpha \in \{1, \dots K_\alpha\}$, l'ensemble des acquisitions pour lesquelles la lame demi-onde est orientée d'un même angle α. Par ailleurs, comme la lame demi-onde prend quatre positions, on a $K_\alpha = K/4$. De plus, ce nombre K_α est lui-même composé du nombre d'acquisitions par position $K_{\alpha_{\text{acq}}}$ et du nombre de fois où l'on fait tourner la lame-demi onde $K_{\alpha_{\text{rot}}}$, c'est-à-dire que $K = 4K_{\alpha_{\text{acq}}}K_{\alpha_{\text{rot}}}$. Finalement, le champ électrique après passage par la lame demi-onde à une acquisition $k \in \{1, \dots K\}$ est donné par $\mathbf{J}_k\mathbf{M}_1\boldsymbol{e}_n$.
3. Le champ électrique passe ensuite par un second ensemble d'éléments optiques qui induisent une seconde transformation $\mathbf{M}_2 \in \mathcal{M}(\mathbb{C}^2)$, définie par l'équation (1.7). Le champ électrique résultant avant séparation et passage dans les analyseurs est $\mathbf{M}_2\mathbf{J}_k\mathbf{M}_1\boldsymbol{e}_n$.
4. Le champ électrique est ensuite séparé par une lame séparatrice. Les champs électriques résultants passent alors à travers deux analyseurs, orientés chacun d'un angle $\psi_j \in \{0°, 90°\}$, avec $j = 1$ ou $j = 2$. Le passage par l'un ou l'autre des analyseurs est modélisé par la projection du champ électrique incident dans la base du polariseur $\boldsymbol{a}_j \in \mathbb{R}^2$ orienté de l' angle ψ_j. Cette projection est mathématiquement équivalente au produit scalaire hermitien $\langle \boldsymbol{a}_j, \mathbf{M}_2\mathbf{J}_k\mathbf{M}_1\boldsymbol{e}_n \rangle$.
5. Chaque champ scalaire résultant arrive en un pixel $n \in \{1, \dots, N\}$ de la partie $j \in \{1, 2\}$ du détecteur, qui enregistre un nombre de photo-électrons. Ce nombre est la mesure des électrons excités par les photons incidents en ce pixel. Il est un pourcentage du nombre de photons incidents et le facteur entre ces deux quantités est appelé le rendement quantique. Le nombre de photons incidents est proportionnel à la moyenne temporelle du module carré du champ électrique atteignant le pixel, c'est-à-dire pour un pixel n de la partie j du détecteur à l'acquisition k, à l'intensité : $\mathbb{E}_t\left[\left|\langle \boldsymbol{a}_j, \mathbf{M}_2\mathbf{J}_k\mathbf{M}_1\boldsymbol{e}_n \rangle_{\mathbb{C}^2}\right|^2\right]_m$. Afin de s'affranchir de ces facteurs, pour simplifier les expressions, on suppose que le champ électrique $\boldsymbol{e}_n$ est exprimé dans une unité qui tient compte des deux facteurs de proportionnalité. Cette unité est communément appelée Unités Digitales Arbitraires (ADU).
6. Enfin, ce nombre de photons fluctue. L'intensité acquise est donc une variable aléatoire suivant une loi de Poisson paramétrée par sa moyenne, qui est égale à sa variance. Étant donné les niveaux d'intensité enregistré en présence de signal, ce bruit de Poisson peut-être approximé par un bruit gaussien de même moyenne et de même variance. Á ce bruit s'ajoute également le bruit de lecture du détecteur qui est gaussien, centré et de variance σ^2_{ro}, supposé uniforme. Finalement, les données s'expriment sous la forme :

$$j \in \{1,2\}, \forall k \in \{1, \dots K\}, \forall n \in \{1, \dots N\}, \quad \boldsymbol{d}^{\text{S}}_{j,k,n} = \mathcal{B}_{j,k,n}\left(\mathbb{E}_t\left[\left|\langle \boldsymbol{a}_j, \mathbf{M}_2\mathbf{J}_k\mathbf{M}_1\boldsymbol{e}_n \rangle_{\mathbb{C}^2}\right|^2\right]\right), \tag{2.3}$$

où $\mathcal{B}_{j,k,n}(y)$ donne une réalisation de la variable gaussienne $\mathcal{N}(y, y + \sigma^2_{\text{ro}})$.

On suppose que pour tout $n \in \{1,\dots,N\}$, le champ électrique $\boldsymbol{e}_n = \boldsymbol{u}_n + \boldsymbol{p}_n$, c'est-à-dire qu'il est composé de la somme de deux champs électriques $\boldsymbol{u}_n \in \mathbb{C}^2$ et $\boldsymbol{p}_n \in \mathbb{R}^2$, respectivement non-polarisé et polarisé. Le champ électrique non-polarisé $\boldsymbol{u}_n \in \mathbb{C}^2$, d'amplitude moyenne $u_n = \sqrt{I_n^{\mathrm{u}}/2}$, est lui-même la somme de deux champs électriques non-polarisés $\boldsymbol{u}_n^{\mathrm{star}}$ et $\boldsymbol{u}_n^{\mathrm{disk}}$. Ces champs électriques représentent respectivement la lumière résiduelle de l'étoile masquée par le coronographe, ainsi que la lumière des étoiles pouvant se trouver dans le champ, et la partie de la lumière émise par le disque qui n'est pas polarisée. Le champ électrique polarisé linéairement $\boldsymbol{p}_n \in \mathbb{R}^2$, d'amplitude $p_n = \sqrt{I_n^{\mathrm{p}}}$ et d'angle de polarisation θ_n, représente la lumière du disque diffusée par la poussière.

Proposition 2.1.1. *Soit $\boldsymbol{v}_{j,k} = (\mathbf{M}_2\mathbf{J}_k\mathbf{M}_1)^* \boldsymbol{a}_j$ pour $j \in \{1,2\}$, où * représente l'adjoint pour le produit scalaire hermitien. Soit $\overline{\boldsymbol{x}}_n = (\overline{I_n^u}, \overline{I_n^p}, \overline{\theta}_n)^\top$, alors le modèle non-linéaire séparable des données s'écrit $\forall k \in \{1,\dots K\}$, $\forall n \in \{1,\dots N\}$ et $\forall j \in \{1,2\}$ de la forme $\boldsymbol{d}_{j,k,n}^S = \mathcal{B}_{j,k,n}(f_{j,k}(\overline{\boldsymbol{x}}_n))$ avec*

$$\forall \boldsymbol{x}_n = (I_n^u, I_n^p, \theta_n)^\top, \quad f_{j,k}^S(\boldsymbol{x}_n) = \frac{I_n^u \|\boldsymbol{v}_{j,k}\|_{\mathbb{C}^2}^2}{2} + I_n^p \left|\langle \boldsymbol{v}_{j,k}, \boldsymbol{c}(\theta_n)\rangle_{\mathbb{C}^2}\right|^2, \tag{2.4}$$

où $\boldsymbol{c}(y) = (\cos y, \sin y)^\top$.

Démonstration : *(Fin page 52)* En effet, pour tout $n \in \{1,\dots N\}$, pour la partie $j \in \{1,2\}$ du détecteur et pour l'acquisition $k \in \{1,\dots K\}$, par définition de l'adjoint et de $\boldsymbol{v}_{j,k}$ on a :

$$\mathbb{E}_t\left[\left|\langle \boldsymbol{a}_j, \mathbf{M}_2\mathbf{J}_k\mathbf{M}_1\boldsymbol{e}_n\rangle_{\mathbb{C}^2}\right|^2\right] = \mathbb{E}_t\left[\left|\langle \boldsymbol{v}_{j,k}, \boldsymbol{e}_n\rangle_{\mathbb{C}^2}\right|^2\right].$$

En développant l'expression, on a par linéarité de l'espérance temporelle :

$$\mathbb{E}_t\left[\left|\langle \boldsymbol{v}_{j,k}, \boldsymbol{e}_n\rangle_{\mathbb{C}^2}\right|^2\right] = |\boldsymbol{v}_{j,k}^{(x)}|^2\mathbb{E}\left[|\boldsymbol{e}_n^{(x)}|^2\right] + |\boldsymbol{v}_{j,k}^{(y)}|^2\mathbb{E}\left[|\boldsymbol{e}_n^{(y)}|^2\right] + \boldsymbol{v}_{j,k}^{(x)}\overline{\boldsymbol{v}}_{j,k}^{(y)}\mathbb{E}\left[\boldsymbol{e}_n^{(x)}\overline{\boldsymbol{e}}_n^{(y)}\right] + \boldsymbol{v}_{j,k}^{(y)}\overline{\boldsymbol{v}}_{j,k}^{(x)}\mathbb{E}\left[\boldsymbol{e}_n^{(y)}\overline{\boldsymbol{e}}_n^{(x)}\right],$$

où $\overline{\cdot}$ représente le complexe conjugué d'un champ électrique et $\boldsymbol{e}_n^{(x)}$ et $\boldsymbol{e}_n^{(y)}$ (resp. $\boldsymbol{v}_{j,k}^{(x)}$ et $\boldsymbol{v}_{j,k}^{(y)}$) représentent respectivement les première et deuxième composantes du champ électrique $\boldsymbol{e}_n$ (resp. $\boldsymbol{v}_{j,k}$).

Or, par définition du champ électrique non-polarisé (cf. section 1.1.3), les composantes de $\boldsymbol{u}_n$ sont centrées et indépendantes.

On a donc d'une part :

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{e}_n^{(x)}\overline{\boldsymbol{e}}_n^{(y)}\right] &= \mathbb{E}\left[\boldsymbol{e}_n^{(x)}\right]\mathbb{E}\left[\overline{\boldsymbol{e}}_n^{(y)}\right] \\ &= \mathbb{E}\left[\boldsymbol{u}_n^{(x)} + \boldsymbol{p}_n^{(x)}\right]\mathbb{E}\left[\overline{\boldsymbol{u}}_n^{(y)} + \overline{\boldsymbol{p}}_n^{(y)}\right] \\ &= \boldsymbol{p}_n^{(x)}\overline{\boldsymbol{p}}_n^{(y)},\end{aligned}$$

Par symétrie, on trouve de même que $\mathbb{E}_t\left[\overline{\boldsymbol{e}}_n^{(x)}\boldsymbol{e}_n^{(y)}\right] = \overline{\boldsymbol{p}}_n^{(x)}\boldsymbol{p}_n^{(y)}$. Par ailleurs, comme $\boldsymbol{p}_n$ représente un champ électrique polarisé linéairement, par définition, on a $\boldsymbol{p}_n \in \mathbb{R}^2$ donc $\overline{\boldsymbol{p}}_n = \boldsymbol{p}_n$.

D'autre part, comme les composantes $\boldsymbol{u}_n^{\text{star}}$ et $\boldsymbol{u}_n^{\text{disk}}$ de $\boldsymbol{u}_n = \boldsymbol{u}_n^{\text{star}} + \boldsymbol{u}_n^{\text{disk}}$ sont indépendantes, on a :

$$\begin{aligned}
\mathbb{E}_t\left[|\boldsymbol{e}_n^{(x)}|^2\right] &= \mathbb{E}_t\left[|\boldsymbol{u}_n^{\text{star}(x)} + \boldsymbol{u}_n^{\text{disk}(x)} + \boldsymbol{p}_n^{(x)}|^2\right] \\
&= \mathbb{E}_t\left[|\boldsymbol{u}_n^{\text{star}(x)} + \boldsymbol{u}_n^{\text{disk}(x)}|^2\right] + \mathbb{E}_t\left[\boldsymbol{u}_n^{\text{star}(x)} + \boldsymbol{u}_n^{\text{star}(x)}\right]\boldsymbol{p}_n^{(x)} \\
&\quad + \mathbb{E}_t\left[\overline{\boldsymbol{u}}_n^{\text{star}(x)} + \overline{\boldsymbol{u}}_n^{\text{disk}(x)}\right]\boldsymbol{p}_n^{(x)} + {\boldsymbol{p}_n^{(x)}}^2 \\
&= \mathbb{E}_t\left[|\boldsymbol{u}_n^{\text{star}(x)}|^2\right] + \mathbb{E}_t\left[|\boldsymbol{u}_n^{\text{disk}(x)}|^2\right] + {\boldsymbol{p}_n^{(x)}}^2.
\end{aligned}$$

Or, par définition de la covariance, on a :

$$\mathbb{E}\left[|\boldsymbol{e}_n^{(x)}|^2\right] = \text{Cov}\left(\boldsymbol{u}_n^{\text{star}(x)}, \overline{\boldsymbol{u}}_n^{\text{star}(x)}\right) + \text{Cov}\left(\boldsymbol{u}_n^{\text{disk}(x)}, \overline{\boldsymbol{u}}_n^{\text{disk}(x)}\right) + {\boldsymbol{p}_n^{(x)}}^2.$$

On remarque que si $u \in \mathbb{C}$ est un nombre complexe aléatoire de moyenne $\mathbb{E}_t[u]$, alors :

$$\text{Cov}(u, \overline{u}) = \text{Var}(\mathfrak{Re}(u)) + \text{Var}(\mathfrak{Im}(u)). \tag{2.5}$$

En effet, posons $u - \mathbb{E}[u] = a + \mathrm{i}b$ avec $(a, b) \in \mathbb{R}^2$ aléatoires et centrés. On a alors :

$$\begin{aligned}
\text{Cov}(u, \overline{u}) &= \mathbb{E}\left[(u - \mathbb{E}[u])\left(\overline{u - \mathbb{E}[u]}\right)\right], \\
&= \mathbb{E}\left[(a + \mathrm{i}b)(a - \mathrm{i}b)\right], \\
&= \mathbb{E}\left[a^2 + b^2\right], \\
&= \mathbb{E}\left[a^2\right] + \mathbb{E}\left[b^2\right], \\
&= \text{Var}(a) + \text{Var}(b), \\
&= \text{Var}(\mathfrak{Re}(u)) + \text{Var}(\mathfrak{Im}(u)).
\end{aligned}$$

D'où le fait que :

$$\begin{aligned}
\mathbb{E}_t\left[|\boldsymbol{e}_n^{(x)}|^2\right] =& \text{Var}_t\left(\mathfrak{Re}(\boldsymbol{u}_n^{\text{star}(x)})\right) + \text{Var}_t\left(\mathfrak{Im}(\boldsymbol{u}_n^{\text{star}(x)})\right) \\
&+ \text{Var}_t\left(\mathfrak{Re}(\boldsymbol{u}_n^{\text{disk}(x)})\right) + \text{Var}_t\left(\mathfrak{Im}(\boldsymbol{u}_n^{\text{disk}(x)})\right) + {\boldsymbol{p}_n^{(x)}}^2.
\end{aligned} \tag{2.6}$$

De même :

$$\begin{aligned}
\mathbb{E}_t\left[|\boldsymbol{e}_n^{(y)}|^2\right] =& \text{Var}_t\left(\mathfrak{Re}(\boldsymbol{u}_n^{\text{star}(y)})\right) + \text{Var}_t\left(\mathfrak{Im}(\boldsymbol{u}_n^{\text{star}(y)})\right) + \\
&\text{Var}_t\left(\mathfrak{Re}(\boldsymbol{u}_n^{\text{disk}(y)})\right) + \text{Var}_t\left(\mathfrak{Im}(\boldsymbol{u}_n^{\text{disk}(y)})\right) + {\boldsymbol{p}_n^{(y)}}^2.
\end{aligned} \tag{2.7}$$

Finalement, comme $\boldsymbol{u}_n^{\text{star}}$ et $\boldsymbol{u}_n^{\text{disk}}$ ont des composantes identiquement distribuées on a :

$$\text{Var}_t\left(\mathfrak{Re}(\boldsymbol{u}_n^{\text{star}(x)})\right) = \text{Var}_t\left(\mathfrak{Re}(\boldsymbol{u}_n^{\text{star}(y)})\right) = {u_{\mathfrak{Re}}}_n^{\text{star}}, \tag{2.8}$$

$$\text{Var}_t\left(\mathfrak{Im}(\boldsymbol{u}_n^{\text{star}(x)})\right) = \text{Var}_t\left(\mathfrak{Im}(\boldsymbol{u}_n^{\text{star}(y)})\right) = {u_{\mathfrak{Im}}}_n^{\text{star}}, \tag{2.9}$$

$$\text{Var}_t\left(\mathfrak{Re}(\boldsymbol{u}_n^{\text{disk}(x)})\right) = \text{Var}_t\left(\mathfrak{Re}(\boldsymbol{u}_n^{\text{disk}(y)})\right) = {u_{\mathfrak{Re}}}_n^{\text{disk}}, \tag{2.10}$$

$$\text{Var}_t\left(\mathfrak{Im}(\boldsymbol{u}_n^{\text{disk}(x)})\right) = \text{Var}_t\left(\mathfrak{Im}(\boldsymbol{u}_n^{\text{disk}(y)})\right) = {u_{\mathfrak{Im}}}_n^{\text{disk}}. \tag{2.11}$$

Posons alors $u_n^2 = u_{\Re\mathfrak{e}\,n}^{\text{star}} + u_{\mathrm{Im}\,n}^{\text{star}} + u_{\Re\mathfrak{e}\,n}^{\text{disk}} + u_{\mathrm{Im}\,n}^{\text{disk}}$, on a alors :

$$\begin{aligned}
&\mathbb{E}_t\left[|\langle \boldsymbol{v}_{j,k}, \boldsymbol{e}_n\rangle_{\mathbb{C}^2}|^2\right] \\
&= |\boldsymbol{v}_{k,j}^{(x)}|^2\left(u_n^2 + \boldsymbol{p}_n^{(x)^2}\right) + |\boldsymbol{v}_{j,k}^{(y)}|^2\left(u_n^2 + p_n^{(y)^2}\right) + \boldsymbol{v}_n^{j,x}\overline{\boldsymbol{v}}_{j,k}^{(y)}\boldsymbol{p}_n^{(x)}\boldsymbol{p}_n^{(y)} + \boldsymbol{v}_{j,k}^{(y)}\overline{\boldsymbol{v}}_{j,k}^{(x)}\boldsymbol{p}_n^{(x)}\boldsymbol{p}_n^{(y)}, \\
&= (\boldsymbol{v}_{j,k}^{(x)}\overline{\boldsymbol{v}}_{j,k}^{(x)} + \boldsymbol{v}_{j,k}^{(y)}\overline{\boldsymbol{v}}_{j,k}^{(y)})u_n^2 + \boldsymbol{v}_{j,k}^{(x)}\overline{\boldsymbol{v}}_{j,k}^{(x)}\boldsymbol{p}_n^{(x)^2} + \boldsymbol{v}_{j,k}^{(y)}\overline{\boldsymbol{v}}_{j,k}^{(y)}\boldsymbol{p}_n^{(y)^2} + \left(\boldsymbol{v}_{j,k}^{(x)}\overline{\boldsymbol{v}}_{j,k}^{(y)} + \boldsymbol{v}_{j,k}^{(y)}\overline{\boldsymbol{v}}_{j,k}^{(x)}\right)\boldsymbol{p}_n^{(x)}\boldsymbol{p}_n^{(y)}, \\
&= \langle \boldsymbol{v}_{j,k}, \boldsymbol{v}_{j,k}\rangle_{\mathbb{C}^2} u_n^2 + |\langle \boldsymbol{v}_{j,k}, \boldsymbol{p}_j\rangle_{\mathbb{C}^2}|^2.
\end{aligned}$$

Or comme $\boldsymbol{p}_n = p_n\begin{pmatrix}\cos\theta_n \\ \sin\theta_n\end{pmatrix}$, on obtient que :

$$\mathbb{E}_t\left[|\langle \boldsymbol{v}_{j,k}, \boldsymbol{e}_n\rangle_{\mathbb{C}^2}|^2\right] = \|\boldsymbol{v}_{j,k}\|_{\mathbb{C}^2}^2 u_n^2 + p_n^2|\langle v_k, \boldsymbol{c}_{\theta_n}\rangle_{\mathbb{C}^2}|^2, \tag{2.12}$$

où $\boldsymbol{c}_{\theta_n} = \begin{pmatrix}\cos\theta_n \\ \sin\theta_n\end{pmatrix}$. En remarquant que $p_n^2 = I_n^{\mathrm{p}}$ et que $u_n^2 = I_n^{\mathrm{u}}/2$ du fait du théorème de la moyenne, on obtient $\forall j = 1, 2$, $\forall k \in \{1, \dots K\}$, $\forall n \in \{1, \dots N\}$:

$$d_{j,k,n}^{\mathrm{S}} = \mathcal{B}_{j,k,n}\left(I_n^{\mathrm{u}}\frac{\|\boldsymbol{v}_{j,k}\|_{\mathbb{C}^2}^2}{2} + I_n^{\mathrm{p}}\left|\langle \boldsymbol{v}_{j,k}, \boldsymbol{c}_{\theta_n}\rangle_{\mathbb{C}^2}\right|^2\right). \tag{2.13}$$

où $\boldsymbol{c}_{\theta_n} = \begin{pmatrix}\cos\theta_n \\ \sin\theta_n\end{pmatrix}$, $\frac{I_n^{\mathrm{u}}}{2} = u_n^2$ et $I_n^{\mathrm{p}} = p_n^2$.

□

2.1.2 Résolution séparable du modèle non-linéaire

Pour estimer les paramètres I^{u}, I^{p} et θ à partir du modèle des données 2.1.1, le plus évident est de chercher les paramètres maximisant la vraisemblance des données. Pour un pixel donné $n \in \{1, \dots N\}$, cela revient à résoudre le critère des moindres carrés pondérés sous contrainte de positivité suivant :

$$\left(\widehat{I^{\mathrm{u}}}_n, \widehat{I^{\mathrm{p}}}_n, \widehat{\theta}_n\right)^\top \in \underset{I_n^{\mathrm{u}}\geq 0, I_n^{\mathrm{p}}\geq 0, \theta_n}{\mathrm{Argmin}} \quad \sum_{j,k}\frac{\mathbf{W}_{j,k,n}}{2}\left(\boldsymbol{d}_{j,k,n}^{\mathrm{S}} - f_{j,k}^{\mathrm{S}}\left((I_n^{\mathrm{u}}, I_n^{\mathrm{p}}, \theta_n)^\top\right)\right)^2, \tag{2.14}$$

où

$$\mathbf{W}_{j,k,n} = \begin{cases}\mathrm{Var}(\boldsymbol{d}_{j,k,n}^{\mathrm{S}})^{-1}, \\ 0 \quad \text{si donnée manquante (cf. sec.1.1.4.1).}\end{cases} \tag{2.15}$$

Posons le critère

$$\forall n \in \{1, \dots, N\}, \quad \Phi_n(\boldsymbol{x}) = \sum_{j,k}\frac{\mathbf{W}_{j,k,n}}{2}\left(\boldsymbol{d}_{j,k,n}^{\mathrm{S}} - f_{j,k}^{\mathrm{S}}(\boldsymbol{x})\right)^2. \tag{2.16}$$

En s'appuyant sur la section 1.3.1, d'après le théorème de Fermat sur les points stationnaires, résoudre le problème (2.14) revient à résoudre

$$\forall n \in \{1, \dots, N\} \qquad \nabla\Phi_n((\widehat{I_n^{\mathrm{u}}}, \widehat{I_n^{\mathrm{p}}}, \widehat{\theta_n})^\top) = 0 \quad \text{sous la contrainte} \quad \begin{cases}\widehat{I_n^{\mathrm{u}}} \geq 0, \\ \widehat{I_n^{\mathrm{p}}} \geq 0.\end{cases} \tag{2.17}$$

où $\nabla\Phi_n$ représente le gradient de la fonction ϕ_n. Du fait de la non-linéarité de $f^{\mathrm{S}}_{j,k}$ en θ_n, les paramètres $(I^{\mathrm{u}}_n, I^{\mathrm{p}}_n, \theta_n)^\top$ ne peuvent pas être estimés par inversion directe de $\nabla\Phi_n$. On peut cependant séparer le problème en θ_n afin de résoudre le problème (2.14) de manière hiérarchique ou alternée. C'est-à-dire résoudre :

$$\forall n \in \{1,\dots,N\} \qquad \widehat{\theta_n} \in \underset{\theta_n}{\mathrm{Argmin}} \left[\min_{I^{\mathrm{u}}_n \geq 0, I^{\mathrm{p}}_n \geq 0} \quad \sum_{j,k} \frac{\mathbf{W}_{j,k,n}}{2} \left(\boldsymbol{d}^{\mathrm{S}}_{j,k,n} - f^{\mathrm{S}}_{j,k}\left((I^{\mathrm{u}}_n, I^{\mathrm{p}}_n, \theta_n)^\top \right) \right)^2 \right], \tag{2.18}$$

où les paramètres I^{u}_n et I^{p}_n sont obtenus par inversion du système $(I^{\mathrm{u}}_n, I^{\mathrm{p}}_n)^\top = \mathbf{V}^{-1}\boldsymbol{b}$ avec :

$$\mathbf{V} = \begin{pmatrix} \sum_{j,k} \mathbf{W}_{j,k,n} \|\boldsymbol{v}_{j,k}\|^4_{\mathbb{C}^2}/4 & \sum_{j,k} \boldsymbol{W}_{j,k,n} \|\boldsymbol{v}_{j,k}\|^2_{\mathbb{C}^2} |\langle \boldsymbol{v}_{j,k}, \boldsymbol{c}(\theta)\rangle_{\mathbb{C}^2}|^2/2 \\ \sum_{j,k} \mathbf{W}_{j,k,n} \|\boldsymbol{v}_{j,k}\|^2_{\mathbb{C}^2} |\langle \boldsymbol{v}_{j,k}, \boldsymbol{c}(\theta)\rangle_{\mathbb{C}^2}|^2/2 & \sum_{j,k} \boldsymbol{W}_{j,k,n} |\langle \boldsymbol{v}_{j,k}, \boldsymbol{c}_{\theta_n}\rangle_{\mathbb{C}^2}|^4 \end{pmatrix} \tag{2.19}$$

et

$$\boldsymbol{b} = \begin{pmatrix} \sum_{j,k} \mathbf{W}_{j,k,n} \|\boldsymbol{v}_{j,k}\|^2_{\mathbb{C}^2} \boldsymbol{d}^{\mathrm{S}}_{j,k,n}/2 \\ \sum_{j,k} \mathbf{W}_{j,k,n} |\langle \boldsymbol{v}_{j,k}, \boldsymbol{c}(\theta)\rangle_{\mathbb{C}^2}|^2 \boldsymbol{d}^{\mathrm{S}}_{j,k,n} \end{pmatrix} \tag{2.20}$$

Dans le cas où la contrainte de positivité n'est pas respectée, c'est-à-dire que $I^{\mathrm{u}}_n < 0$ ou $I^{\mathrm{p}}_n < 0$, il est nécessaire d'explorer toutes les possibilités de projections et de choisir celle qui minimise le critère (2.16). Dans cette thèse, je procède à la recherche d'un minimum du critère (2.16) de manière hiérarchique, par l'algorithme 5.

Algorithme 5 : Méthode non-Linéaire Séparable (MnLS)

pour $n = 1,\dots,N$ **faire**
  pour $\theta \in]-\pi/2, \pi/2]$ **faire**
    $(\widehat{I^{\mathrm{u}}_n}, \widehat{I^{\mathrm{p}}_n})^\top = \mathbf{V}^{-1}\boldsymbol{b}$ où $\mathbf{V}$ et $\boldsymbol{b}$ sont obtenus respectivement par (2.19) et (2.20).
    si $\widehat{I^{u}_n} \geq 0$ **et** $\widehat{I^{p}_n} \geq 0$ **alors**
      retourner $(\widehat{I^{\mathrm{u}}_n}, \widehat{I^{\mathrm{p}}_n})^\top$.
    sinon
      $\widehat{I^{\mathrm{u}}_n}^{(1)} = 0$
      $\widehat{I^{\mathrm{p}}_n}^{(1)} = \max\left\{ \frac{\left(\sum_{j,k} \mathbf{W}_{j,k,n} |\langle \boldsymbol{v}_{j,k}, \boldsymbol{c}(\theta)\rangle_{\mathbb{C}^2}|^2 \boldsymbol{d}_{j,k,n}\right)}{\left(\sum_{j,k} \boldsymbol{W}_{j,k,n} |\langle \boldsymbol{v}_{j,k}, \boldsymbol{c}(\theta)\rangle_{\mathbb{C}^2}|^4\right)}, 0 \right\}$
      $\widehat{I^{\mathrm{u}}_n}^{(2)} = \max\left\{ \frac{\left(2\sum_{j,k} \mathbf{W}_{j,k,n} \|\boldsymbol{v}_{j,k}\|^2_{\mathbb{C}^2} \boldsymbol{d}_{j,k,n}\right)}{\left(\sum_{j,k} \mathbf{W}_{j,k,n} \|\boldsymbol{v}_{j,k}\|^4_{\mathbb{C}^2}\right)}, 0 \right\}$
      $\widehat{I^{\mathrm{p}}_n}^{(2)} = 0$
      si $\Phi_n\left((\widehat{I^{u}_n}^{(1)}, \widehat{I^{p}_n}^{(1)}, \theta)^\top \right) \leq \Phi_n\left((\widehat{I^{u}_n}^{(2)}, \widehat{I^{p}_n}^{(2)}, \theta)^\top \right)$ **alors**
        retourner $(\widehat{I^{\mathrm{u}}_n}, \widehat{I^{\mathrm{p}}_n})^\top = (\widehat{I^{\mathrm{u}}_n}^{(1)}, \widehat{I^{\mathrm{p}}_n}^{(1)})^\top$.
      sinon
        retourner $(\widehat{I^{\mathrm{u}}_n}, \widehat{I^{\mathrm{p}}_n})^\top = (\widehat{I^{\mathrm{u}}_n}^{(2)}, \widehat{I^{\mathrm{p}}_n}^{(2)})^\top$.
  retourner $\widehat{\theta_n}$ qui minimise $\Phi_n\left((\widehat{I^{\mathrm{u}}_n}, \widehat{I^{\mathrm{p}}_n}, \theta)^\top \right)$.

La recherche du θ_n minimal peut se faire soit en discrétisant son intervalle de définition et choisissant le θ_n donnant le meilleur critère parmi toutes les valeurs discrétisées, soit via un algorithme de descente sur le critère n'impliquant pas de gradient, tel que la méthode de Brent [Brent, 1973]. Cette méthode consiste à trouver un minimum à partir d'un intervalle donné, par interpolations quadratiques successives d'une discrétisation de trois points, raffinés autour du minimum de l'interpolation précédente.

2.1.3 Simulation des données pré-traitées

Afin de pouvoir tester la MnLS, dont les étapes sont données par l'algorithme 5, il est nécessaire de créer des données synthétiques qui soient proches des données réelles. Pour créer de tels jeux de données, j'ai simulé les cartes I^{u}, I^{p} et θ, de taille $N = 128 \times 128$ (ce qui correspond à la taille moyenne des régions d'intérêts sur données réelles), présentées dans la figure 2.2. Les structures les plus difficiles à reconstruire étant les structures peu polarisées fines et de grand contraste vis-à-vis de l'intensité non-polarisée, j'ai donc reproduit ces difficultés dans ma simulation de données.

Pour cela, j'ai synthétisé un disque constitué de trois anneaux de même brillance. Ce disque

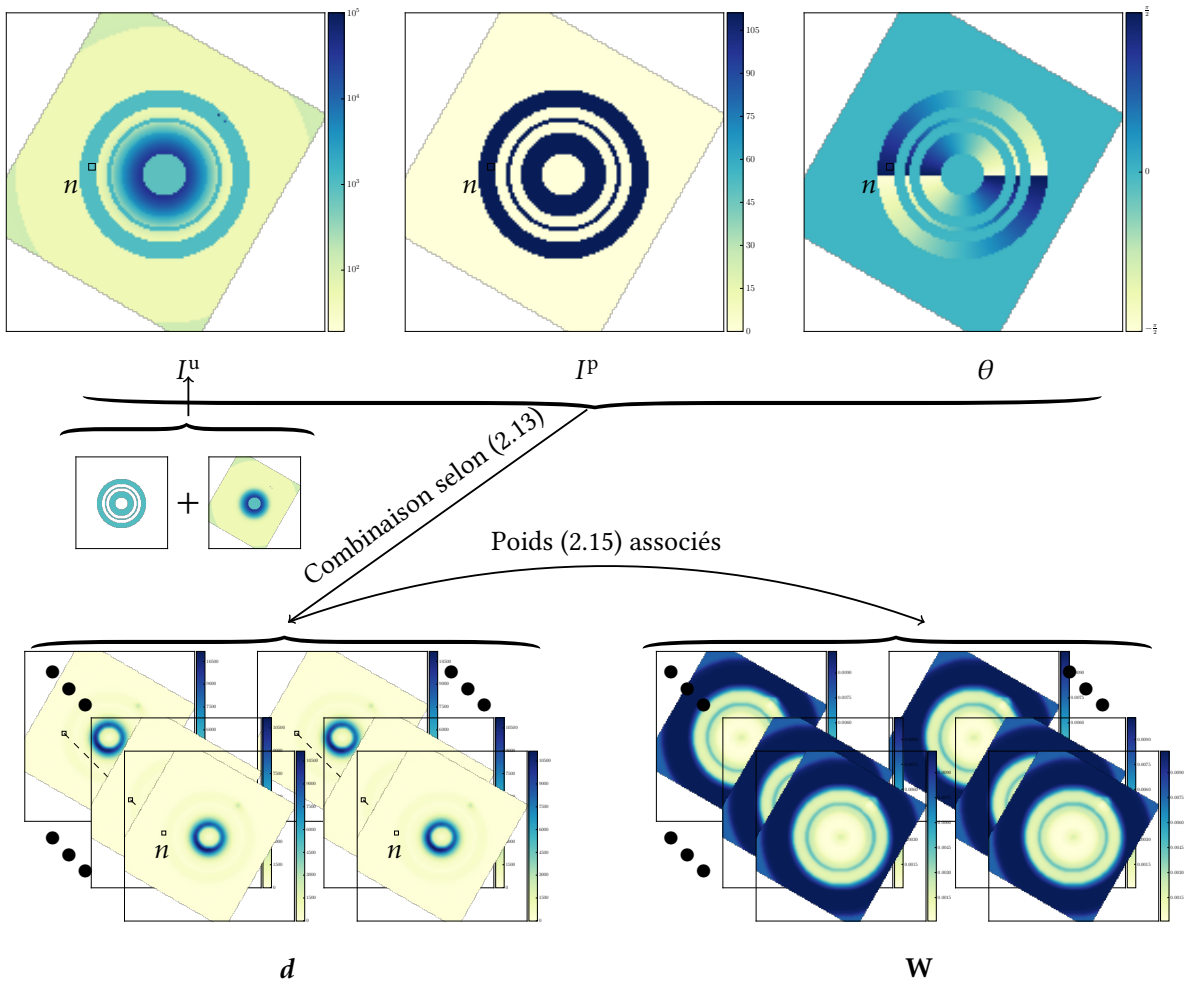


FIGURE 2.2 – Simulation des données calibrées et des données pré-traitées, pour $\tau^{\mathrm{disk}} = 10\%$.

est partiellement polarisé, c'est-à-dire qu'il a une composante polarisée linéairement et une composante non-polarisée. La composante polarisée linéairement est ici représentée par les cartes d'intensité I^{p} et d'angle $\theta \in]-\pi, \pi]$, tournant autour du centre des différents anneaux. La composante non-polarisée du disque est notée I^{udisk} et je représente le rapport entre les deux composantes par le taux de polarisation :

$$\tau^{\mathrm{disk}} = \frac{I^{\mathrm{p}}}{I^{\mathrm{udisk}} + I^{\mathrm{p}}}. \tag{2.21}$$

Dans la suite, cette quantité est exprimée en pourcentage. Les situations où la reconstruction de l'intensité polarisée sera la plus complexe correspondent à des cas où $\tau^{\mathrm{disk}} < 10\%$. En astrophysiques, les cas les plus complexes ont un taux de polarisation du disque de 3%, les cas les plus brillant sont de l'ordre de 30%.

La composante I^{udisk} est mélangée à la composante non-polarisée de l'étoile I^{ustar} et d'éventuels compagnons (étoiles proches de l'étoile mère, de plus ou moins fortes intensités), les deux cartes sont présentées sur la figure 2.2. Sur la composante non-polarisée de l'étoile apparaissent les résidus au bord du coronographe et la limite de l'optique adaptative. L'intensité non-polarisée totale est ici représentée par la carte $I^{\mathrm{u}} = I^{\mathrm{ustar}} + I^{\mathrm{udisk}}$. Il est important de noter que le démélange de ces deux composantes non-polarisées n'est pas possible à partir des données DPI de l'instrument SPHERE IRDIS, sans ADI. Le taux de polarisation du disque τ^{disk} n'est donc pas accessible en pratique à partir de ces données polarimétriques.

Les cartes simulées sont ensuite combinées selon le modèle des données (2.13) pour les différentes positions de lame demi-onde et orientation d'analyseurs. Les jeux seront composés de $K_{\alpha_{\mathrm{rot}}} = 8$ cycles complets de positions de lame demi-onde, avec $K_{\alpha_{\mathrm{acc}}} = 2$ acquisitions par positions, soit $K_\alpha = 16$ images par position de lame demi-onde, pour un total de $K = 4K_\alpha = 64$ images par jeu.

Avant de générer les différentes réalisations du bruit associées, les cartes combinées sont convoluées par une PSF instrumentale. Sont alors effectuées les $K \times 2N$ réalisations de bruit, où $2N$ représente l'ensemble des pixels des parties gauche et droite du détecteur (*i.e* $j = 1$ et $j = 2$). Ces réalisations sont faites à partir d'une graine de générateur aléatoire fixe, afin de constituer le cube de données *pré-traitées*, c'est-à-dire que les données gauches et droites du détecteur sont centrées géométriquement, tournées et découpées de manière subpixelique, de telle sorte que chaque pixel $n \in \{1, \dots, N\}$ du paramètre d'intérêt corresponde au même pixel $n \in \{1, \dots N\}$ des données gauche et droite du détecteur, telles que présentées sur la figure 2.2. Les cartes de poids associées à chaque acquisition sont simulées en parallèle d'après (2.15). Dans le cas des données simulées, ces cartes de poids sont redondantes, car toutes les acquisitions qui correspondent à une même position de lame demi-onde auront la même variance.

Afin de pouvoir comparer les performances des méthodes développées ici avec celles de l'état-de-l'art, j'ai simulé des jeux de données ayant des niveaux de difficultés différents pour la reconstruction. Pour cela, j'ai créé six cubes pour six taux de polarisation du disque $\tau^{\mathrm{disk}} \in \{3\%, 7\%, 10\%, 15\%, 25\%, 50\%\}$. Un taux de polarisation faible du disque correspond à une partie polarisée peu brillante et inversement. Comme ce taux n'est pas disponible en pratique, pour juger la difficulté des cas, j'introduis le taux de polarisation total, donné en chaque pixel $n \in \{1, \dots, N\}$ par la formule

$$\tau^{\mathrm{total}}(\boldsymbol{x}_n) = \frac{I^{\mathrm{p}}_n}{I^{\mathrm{u}}_n + I^{\mathrm{p}}_n}, \tag{2.22}$$

et le rapport signal-à-bruit (ou SNR pour Signal-to-Noise Ratio en anglais), défini en chaque pixel $n \in \{1,\dots,N\}$ par

$$\mathrm{SNR}(\boldsymbol{x}_n) = \frac{\sqrt{K_\alpha} I_n^{\mathrm{p}}}{\sqrt{(I_n^{\mathrm{u}} + I_n^{\mathrm{p}})/2 + \sigma_{\mathrm{ro}}^2}}, \tag{2.23}$$

où K_α correspond au nombre d'acquisitions par position de lame demi-onde et σ_{ro}^2 la variance du bruit de lecture du détecteur. Le déroulé des calculs permettant d'obtenir une telle formule du SNR est présenté dans le chapitre 5. La différence entre τ^{total} et τ^{disk} est que ce dernier ne prend pas en compte les résidus de lumière de l'étoile dans l'intensité polarisée. La figure 2.3 montre les cartes de SNR et de taux de polarisation totale des paramètres simulés en fonction des taux de polarisation du disque τ^{disk}. Au centre, là où les résidus de l'étoile sont les plus forts, le SNR et le τ^{total} sont les plus faibles, surtout dans le cas où la polarisation du disque est très faible. En revanche, on gagne très vite en SNR dès que l'on s'éloigne du centre, ou que la partie polarisée du disque est plus brillante (*i.e.* $\geq 10\%$).

Avant de produire les $K \times 2N$ réalisations de bruit de chaque jeu de données, la graine d'aléa est réinitialisée à une même valeur, fixée au préalable. Cela permet, d'une part, la reproductibilité des résultats. D'autre part, pour un pixel donné, les réalisations du bruit sont alors proportionnelles d'un jeu de donnée à l'autre. En effet, pour faire une réalisation d'une variable gaussienne centrée, de variance σ^2, on simule une réalisation centrée de variance 1 que l'on multiplie par l'écart-type σ. En réinitialisant la graine, on s'assure que toutes les $K \times 2N$ réalisations de variables centrées et réduites faites, sont identiques aux précédentes. Ainsi, seul l'écart-type, qui va dépendre du taux de polarisation, varie. Finalement, on s'attend donc à avoir une erreur proportionnelle à la variance des données.

2.1.4 Application sur données simulées et données astrophysiques

2.1.4.1 Applications sur données simulées

Pour cette application, nous allons étudier deux configurations. La première correspond au cas sans données manquantes, c'est-à-dire que $\mathbf{W}_{j,k,n} = \mathrm{Cov}(\boldsymbol{d}_{j,k,n})$ pour tout $j \in \{1,2\}$,, tout $k \in \{1,\dots K\}$ et tout $n \in \{1,\dots N\}$. Le second correspond au cas où certaines images des

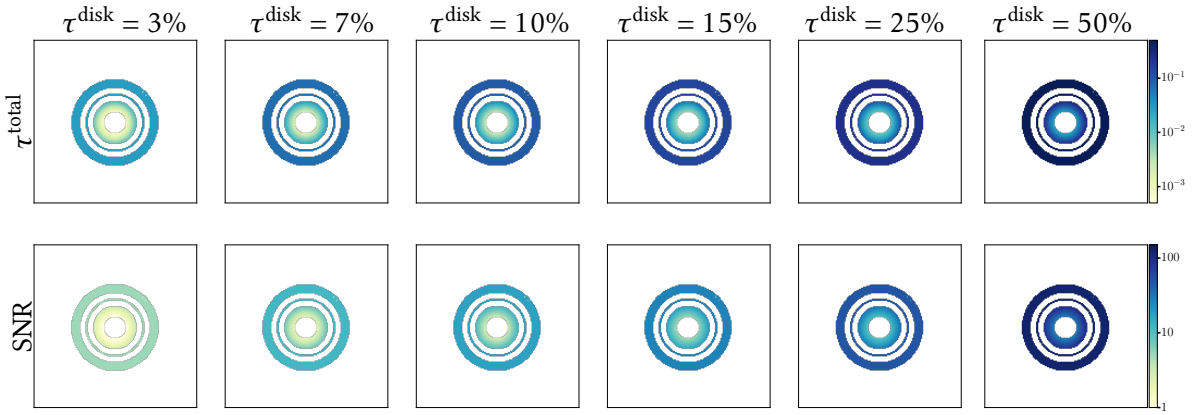


FIGURE 2.3 – Cartes des valeurs τ^{total} et de SNR des cartes simulées pour les différents τ^{disk}.

données sont non exploitables, c'est-à-dire que pour ces données k, pour tout $n \in \{1, \dots N\}$ et tout $j \in \{1, 2\}$, $\mathbf{W}_{j,k,n} = 0$. Ces données non exploitables sont représentées en rouge dans le tableau 2.1.

On applique donc tout d'abord la MnLS (Algorithme 5) aux données simulées dans le cas où il n'y a pas de données manquantes. La figure 2.4 présente les cartes des paramètres reconstruits avec la MnLS et les méthodes de la Double Différence, du Double Ratio, pour les taux de polarisation du disque $\tau^{\text{disk}} \in \{3\%, 10\%50\%\}$.

On voit que pour $\tau^{\text{disk}} = 3\%$, la reconstruction des paramètres I^{p} et θ est plus bruitée pour toutes les méthodes. Lorsque le taux de polarisation augmente et donc que le SNR augmente, la reconstruction des paramètres I^{p} et θ est plus précise. En revanche, il est difficile de dire, visuellement, quelle méthode parmi les trois donne les meilleurs résultats.

Pour juger la qualité des reconstructions et pouvoir comparer les méthodes, j'ai choisi comme critère de qualité l'Erreur Quadratique Moyenne normalisée (EQMn) ou (MSEn : *normalized Mean Square Error*), calculée sur l'ensemble des pixels valides du paramètre d'intérêt. L'EQMn d'une composante $\ell \in \{1, \dots, L\}$ d'une estimation $\widehat{\boldsymbol{x}} \in (\mathbb{R}^N)^L$ du paramètre vrai $\overline{\boldsymbol{x}} \in (\mathbb{R}^N)^L$ est définie pour tout $\ell \in \{1, \dots L\}$ par :

$$\text{EQMn}(\widehat{\boldsymbol{x}}_\ell, \overline{\boldsymbol{x}}_\ell) = \frac{\sum_{n \in \widetilde{N}_\ell} (\widehat{\boldsymbol{x}}_{\ell,n} - \overline{\boldsymbol{x}}_{\ell,n})^2}{\sum_{n=1}^{N} \overline{\boldsymbol{x}}_{\ell,n}^2}, \tag{2.24}$$

où $\widetilde{N}_\ell$ est l'ensemble des pixels valides ou représentatifs du paramètre estimé. Pour le paramètre I^{u}, tous les pixels sont considérés valides. Pour les paramètres I^{p} et θ, on réduit l'ensemble des pixels représentatifs aux endroits où il y a du signal, c'est-à-dire aux trois anneaux qui composent le disque non convolué. Une bonne valeur de l'EQMn sera proche de 0. Dans le cas de l'EQMn angulaire, on *«déplie »* l'angle en sommant

$$\text{EQMn}(\widehat{\theta}, \overline{\theta}) = \sum_{n \in \widetilde{N}_3} \left(\frac{\arg(\mathrm{e}^{2\mathrm{i}(\widehat{\theta}_n - \overline{\theta}_n)})}{2\widetilde{N}_3} \right)^2. \tag{2.25}$$

Dans le cas ou l'on n'a pas de données manquantes, la figure 2.5a représente la racine des valeurs de l'EQMn des estimateurs $\widehat{I^{\text{u}}}$, $\widehat{I^{\text{p}}}$ et $\widehat{\theta}$ par rapport aux paramètres vrais $\overline{I^{\text{u}}}$, $\overline{I^{\text{p}}}$ et $\overline{\theta}$, pour les différentes méthodes. La figure 2.5b représente le rapport entre les racines des EQMn par rapport à la meilleure EQMn obtenu pour chaque τ^{disk}.

On observe que pour I^{u}, il y a très peu d'erreur et que la méthode du double ratio est toujours celle qui a l'EQMn la plus élevée. À faible taux de polarisation, la double différence et la MnLS ont la même EQMn, tandis qu'à haut taux de polarisation, la MnLS a une EQMn plus basse que les méthodes de l'état-de-l'art.

Pour I^{p}, les trois méthodes semblent avoir une EQMn identique. On voit cependant d'après la figure 2.5b que le Double Ratio a toujours les valeurs d'EQMn les plus élevées. De plus à faible taux de polarisation, la Double Différence et la MnLS ont une EQMn similaire. À haut taux de polarisation, la méthode de la Double Différence a une EQMn plus élevée que la MnLS.

Enfin pour θ, les trois méthodes semblent également donner des résultats similaires. On voit cependant, sur la figure 2.5b, qu'à faible taux de polarisation, l'erreur sur l'angle faite par la méthode du double Ratio est plus petite que celle des autres méthodes. En revanche, à haut taux de polarisation celle-ci est plus élevée. La méthode de la Double Différence a une EQMn sensiblement similaire à celle de la MnLS.

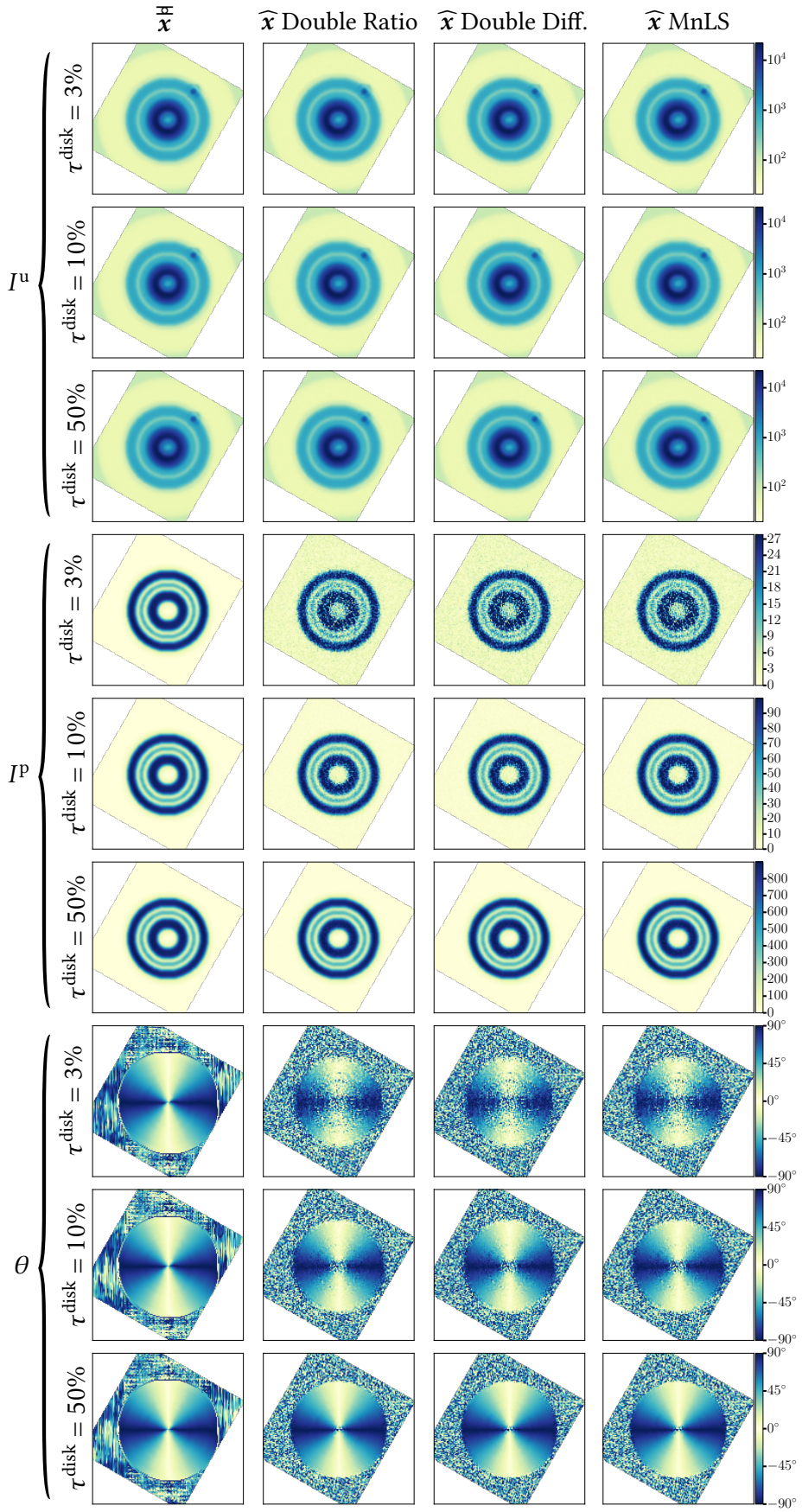


FIGURE 2.4 – Cartes reconstruites avec les différentes méthodes pour $\tau^{\mathrm{disk}} \in \{3\%, 10\%, 50\%\}$.

TABLE 2.1 – Table des correspondances de chaque position et cycle de lame demi-onde pour les $K = 64$ acquisitions. Les données en rouge sont non-exploitables dans le cas des données manquantes. Les données grisées correspondent aux données supprimées pour faire fonctionner les méthodes de l'état-de-l'art.

		Numéro du cycle de lame demi-onde $K_{\alpha_{\text{rot}}}$							
		1	2	3	4	5	6	7	8
α	0°	1 2	9 10	17 18	25 26	33 34	41 42	49 50	57 58
	45°	3 4	11 12	19 20	27 28	35 36	43 44	51 52	59 60
	22,5°	5 6	13 14	21 22	29 30	37 38	45 46	53 54	61 62
	77,5°	7 8	15 16	23 24	31 32	39 40	47 48	55 56	63 64

On remarque cependant les rapports d'erreurs sont de l'ordre du millième d'intensité ou d'angle, un rapport de 1 signifiant que les méthodes sont identiques.

Dans le cas où certaines des acquisitions ne sont pas exploitables, la figure 2.5c représente les valeurs de l'EQMn des estimateurs $\widehat{I^{\text{u}}}$, $\widehat{I^{\text{p}}}$ et $\widehat{\theta}$ des paramètres vrais $\overline{I^{\text{u}}}$, $\overline{I^{\text{p}}}$ et $\overline{\theta}$, pour les différentes méthodes. La figure 2.5b représente le rapport entre les racines des EQMn par rapport à la meilleure EQMn obtenu pour chaque τ^{disk}. Dans le cas de la MnLS, ces données sont prises en compte en mettant tous les poids $\mathbf{W}_{k,n}$ avec $k \in \{4, 6, 39, 43, 62\}$ et $n \in \{1, \dots, N\}$ à 0. Dans le cas des méthodes de l'état-de-l'art, celles-ci nécessitent toutes les acquisitions des cycles de lame demi-onde afin de fonctionner. Il est donc nécessaire de supprimer du cube les tranches de cycles contenant les poses non utilisables.

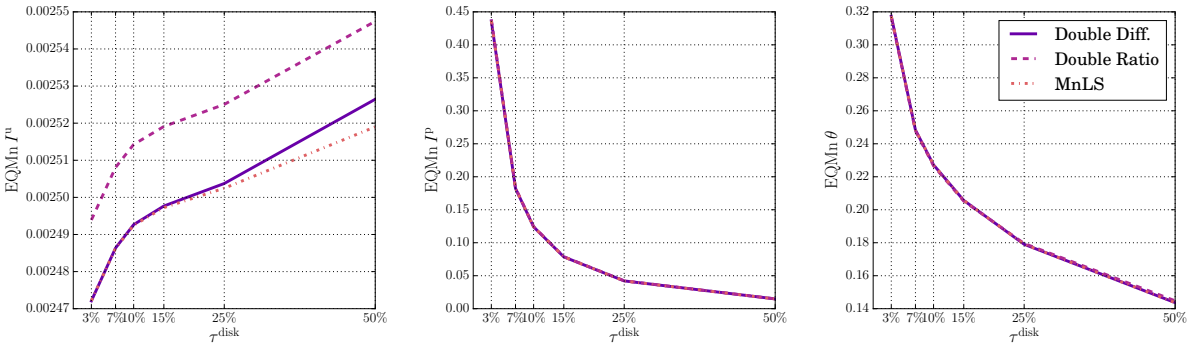
Le tableau 2.1 montre la composition du cube par rapport aux positions de lame demi-onde et au différents cycles. Ne sont supprimées uniquement les tranches de cycles contenant les données inutilisables, soit un total de 4 tranches et donc 16 images. Cela baisse donc le SNR pour ces méthodes, il n'est donc pas aberrant d'observer sur la figure 2.5c que la MnLS donne une meilleure EQMn que ces méthodes de l'état-de-l'art, pour l'ensemble des paramètres.

On voit en effet que pour les trois paramètres d'intérêt, la MnLS est celle qui a l'EQMn la plus basse pour tout taux de polarisation du disque τ^{disk}. Le rapport d'erreur est alors de l'ordre du dixième pour les intensité et d'un demi dixième pour l'angle.

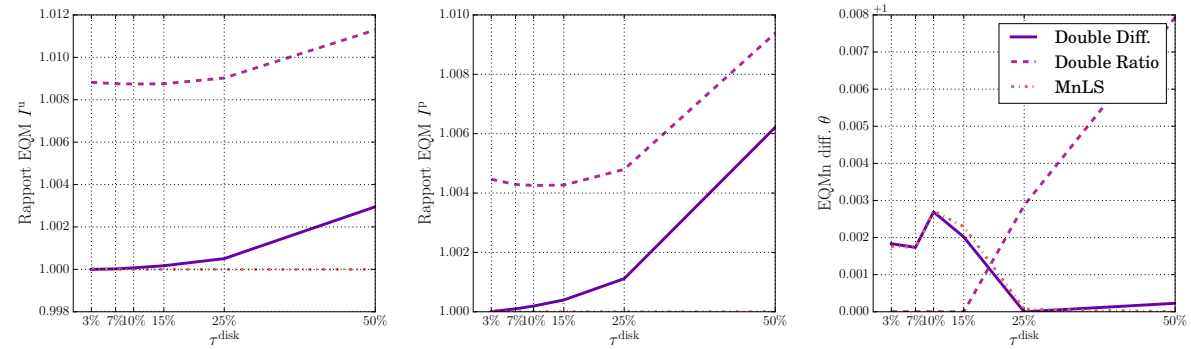
En conclusion, sur données simulées, sans données manquantes les trois méthodes sont équivalentes au millième prêt. La MnLS est tout de même la méthode qui permet d'avoir la plus petite erreur, sauf sur l'angle à faible taux de polarisation, où le Double Ratio est la méthode avec l'erreur la plus faible. Avec données manquantes, la MnLS est la méthode qui a l'erreur la plus faible, grâce à la prise en compte des poids, cela permet de ne pas baisser considérablement le SNR en rejetant des cycles de lame demi-onde. On en déduit que sur données simulées, dans le cas de données manquantes la MnLS est meilleure que la méthode du Double Ratio et de la Double Différence.

2.1.4.2 Application sur données astrophysiques

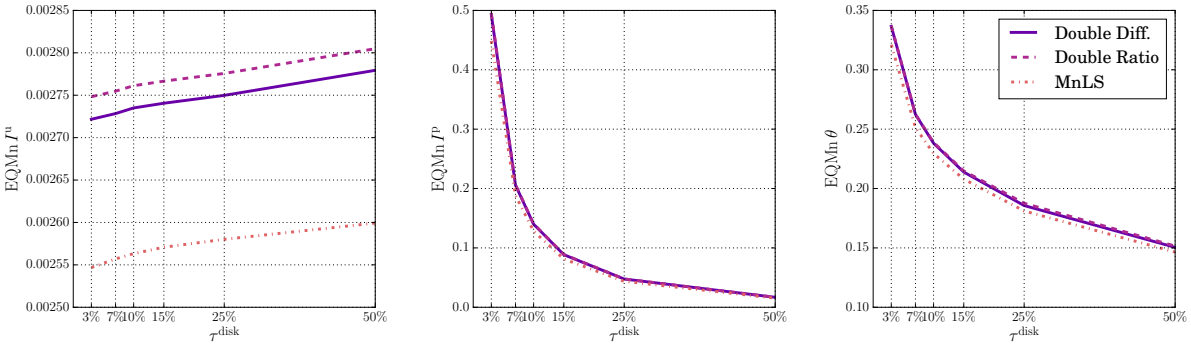
On se propose maintenant d'appliquer la MnLS sur un jeu de données astrophysiques *pré-traitées*, et de comparer les reconstructions avec les résultats obtenus par les méthodes de la Double Différence et du Double Ratio.



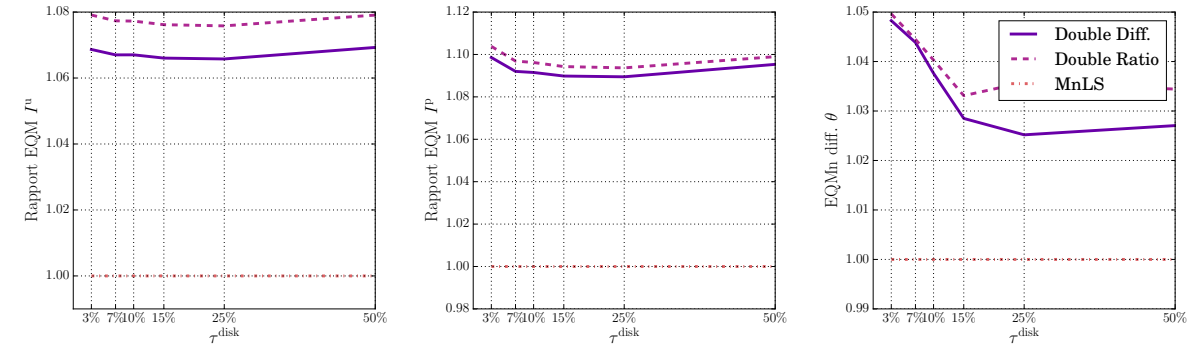
(a) EQMn entre les $\widehat{\boldsymbol{x}}$ estimés avec les différentes méthodes et $\overline{\boldsymbol{x}}^{\mathsf{H}}$ sans données manquantes.



(b) Rapport entre les EQMn estimés avec la meilleure EQMn donnée pour chaque τ^{disk}.



(c) EQMn entre les $\widehat{\boldsymbol{x}}$ estimés avec les différentes méthodes pour chaque τ^{disk} et $\overline{\boldsymbol{x}}^{\mathsf{H}}$ avec les acquisitions $k \in \{4, 6, 39, 43, 62\}$ inutilisables.



(d) Rapport entre les EQMn estimés avec la meilleure EQMn donnée pour chaque τ^{disk} avec les acquisitions $k \in \{4, 6, 39, 43, 62\}$ inutilisables.

FIGURE 2.5 – Erreurs quadratiques moyennes normalisées des différents paramètres pour les différentes méthodes. Les figures a) et c) représentent l'EQMn EQMn($\widehat{\boldsymbol{x}}, \overline{\boldsymbol{x}}^{\mathsf{H}}$) pour les différentes méthodes. Les figures b) et d) représentent le rapport entre ces EQMn et la meilleure valeur de toutes les EQMn pour chaque valeur de τ^{disk}.

Pour tester les méthodes, j'ai choisi la cible RXJ 1615 [Avenhaus et al., 2018] observée en bande H. Une étude plus précise des résultats astrophysiques est présentée dans la section 6.2.1. Ce jeu de données comporte cinq images qui ne sont pas exploitables et un cycle de lame demi-onde incomplet. Dans le cas des méthodes de l'état-de-l'art, les cycles correspondants sont supprimés, comme pour le cas des données simulées. Dans le cas de la MnLS, les poids des images sont simplement mis à 0. Le cycle incomplet est également complété par deux images de poids nul.

La figure 2.6 présente les cartes des paramètres reconstruits avec la MnLS et les méthodes de la Double Différence et du Double Ratio. On voit que dans le cas de la Double Différence, les résidus d'interpolation des pixels morts sont amplifiés. La carte d'intensité polarisée $\widehat{I^{\mathrm{p}}}$ reconstruite avec la MnLS semble moins bruitée par ces résidus d'interpolation que les cartes reconstruites avec les méthodes de l'état-de-l'art. En revanche les reconstructions $\widehat{I^{\mathrm{u}}}$ et $\widehat{\theta}$ sont relativement similaire.

Comme la vérité terrain est inconnu, il n'est pas possible de calculer l'EQMn. Cependant, les paramètres estimés par la MnLS sont asymptotiquement sans biais, car ce sont des esti-

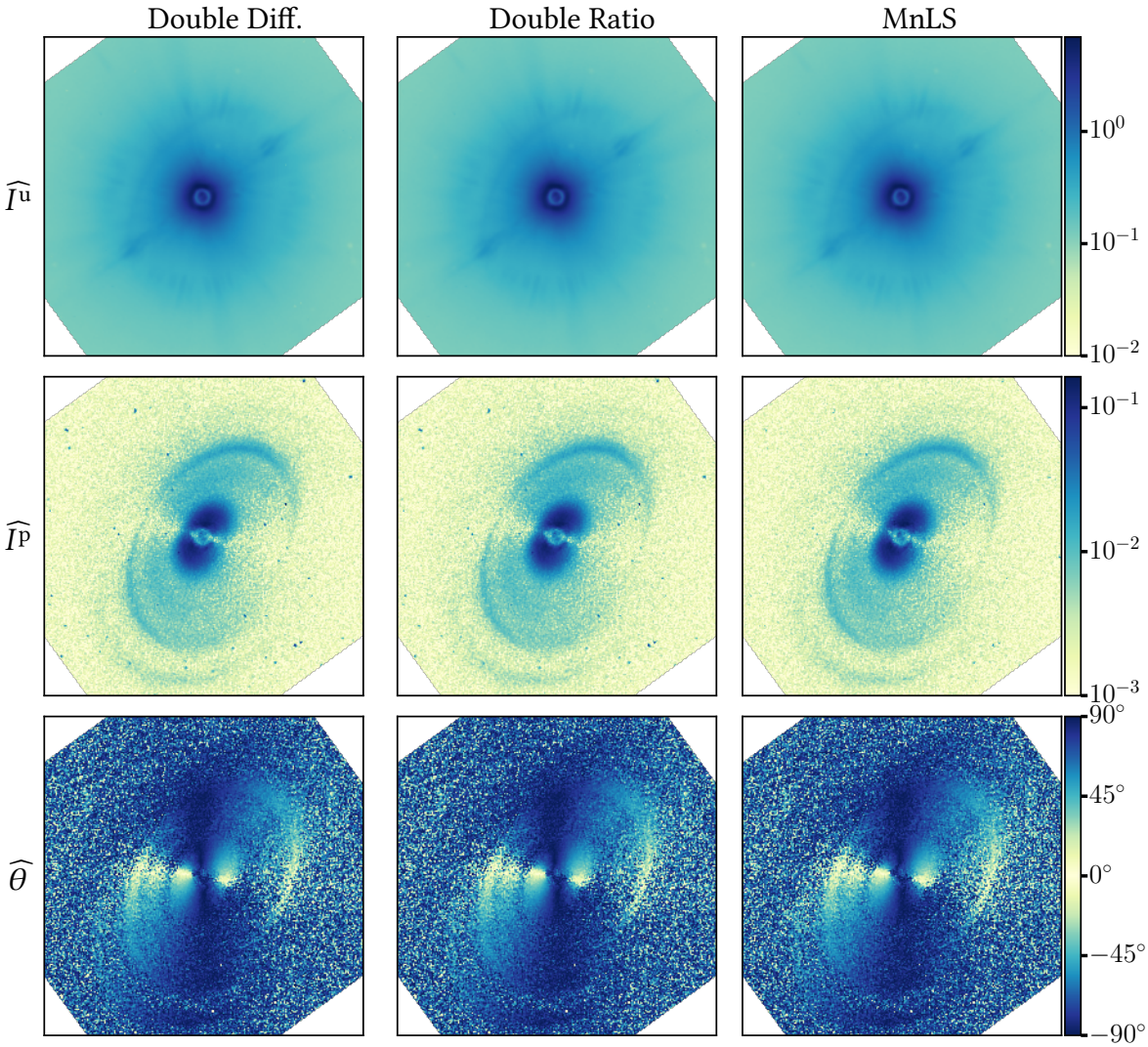


Figure 2.6 – Cartes reconstruites avec les différentes méthodes de la cible RXJ 1615 observée en bande H.

mateurs de moindres carrés. En émettant l'hypothèse que les paramètres reconstruits par le Double Ratio et la Double Différence sont également sans biais, on peut alors utiliser la variance des paramètres comme représentation de l'erreur.

La figure 2.8 présente les cartes d'approximation de l'erreur de chaque paramètre pour chaque méthode en utilisant la formule 5.28. Les calculs associés à chaque méthode sont développés dans la section 5.2 où une étude plus précise de l'estimation de l'erreur y est présentée. La figure 2.8a représente l'écart-type minimal des intensités estimées pour chaque méthode et de l'angle en degrés. La figure 2.8b représente l'erreur relative faite sur les intensités estimées, c'est-à-dire que pour chaque pixel $n \in \{1, \dots N\}$ de chaque paramètre, l'erreur absolue a été divisée par l'intensité estimée.

On voit sur la figure 2.8a que l'erreur angulaire est très élevée pour toutes les reconstructions. Cependant, lorsqu'on regarde les cartes de SNR et de taux de polarisation total estimés, tracées sur la figure 2.7, de tels résultats semblent cohérents avec ceux obtenus sur données simulées à la section 5.2, sur la figure 5.3a, car le SNR est très faible. On voit d'ailleurs dans les cartes d'erreurs de la figure 2.8b, que bien qu'en présence de données manquantes, l'erreur relative sur I^{u} faite avec la méthode de la Double Différence est la plus faible comme dans le cas sur données simulées à faible taux de polarisation du disque sur données simulées. Cela est dû au fait que l'estimateur de la Double Différence est biaisé [Tinbergen, 2005]. Dans l'erreur relative du Double Ratio on voit par ailleurs apparaître les araignées.

2.1.4.3 Synthèse des résultats

Dans cette section nous présentons une méthode non-linéaire permettant de prendre en compte la statistique des données ainsi que les données manquantes. Si l'avantage d'une telle méthode semble évident sur données simulées avec données manquantes, sur données réelles, le gain semble moins évident à quantifier. Mais, cette méthode a l'avantage de disposer de barre d'erreurs fiables. L'avantage est également visible au niveau des résidus d'interpolation des pixels morts. Enfin, le bénéfice de l'utilisation d'une méthode inverse pourra être plus évident lorsque le modèle sera plus complexe, c'est-à-dire lorsqu'on y aura inclue le pré-traitement et la déconvolution.

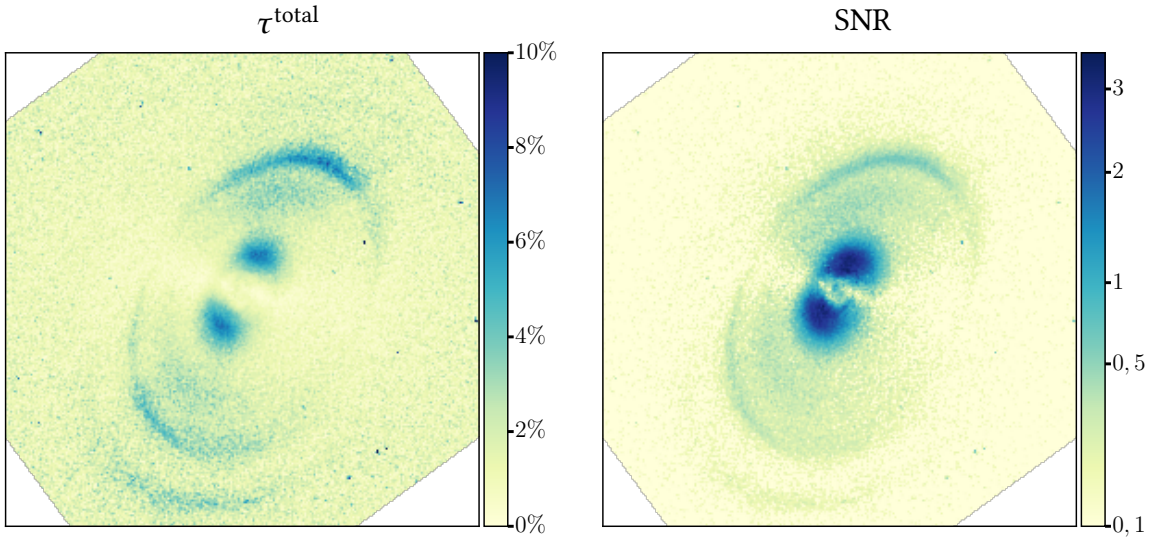
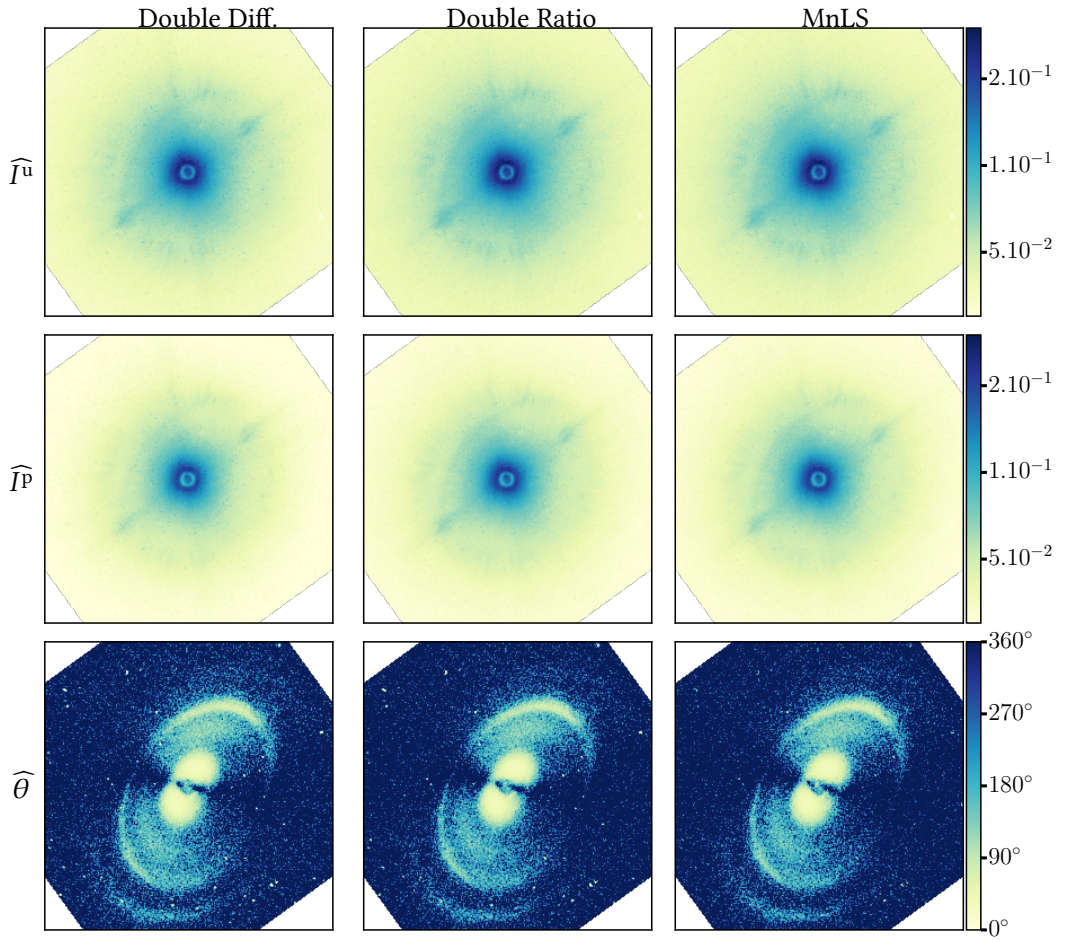
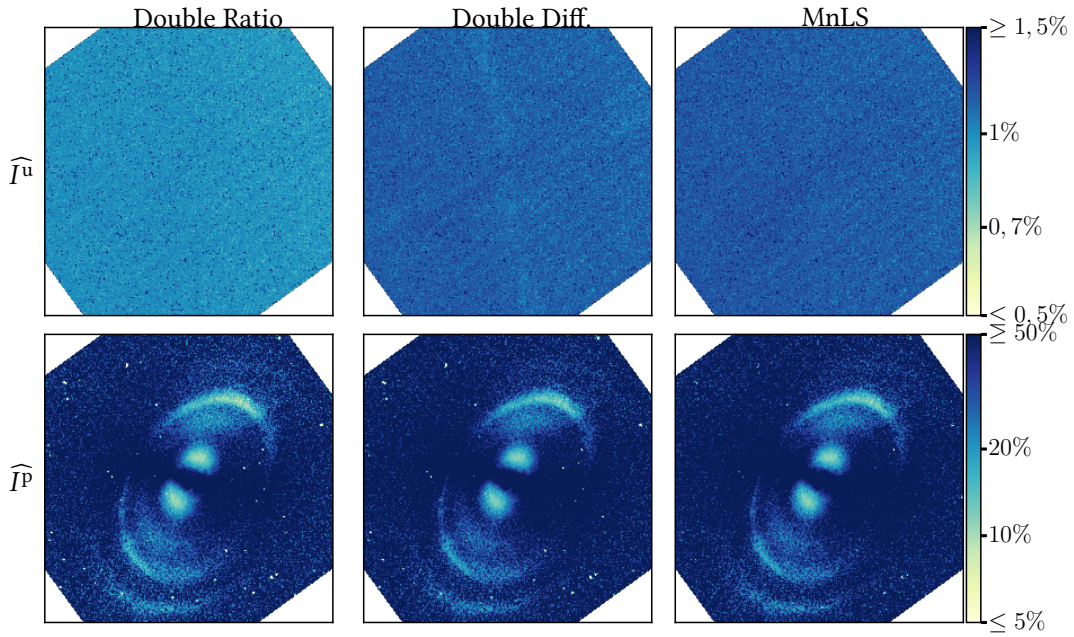


FIGURE 2.7 – Taux de polarisation total τ^{total} et rapport signal sur bruit (SNR) estimés à partir des cartes reconstruites par la MnLS et des formules (2.22) et (2.23).



(a) Erreur absolue des différents paramètres



(b) Erreur relative faites sur les intensités.

FIGURE 2.8 – Cartes d'erreur des différentes méthodes de la cible RXJ 1615 en bande H.

2.2 MLS : Modèle Linéaire Séparable sur les paramètres de Stokes

2.2.1 Le modèle direct séparable linéaire

Dans la section précédente, nous avons vu que la prise en compte des poids, dans la reconstruction des paramètres I^{u}, I^{p} et θ par la MnLS, permettait d'obtenir de meilleurs résultats qu'avec les méthodes de l'état-de-l'art. Cela est d'autant plus visible dans le cas où des observations ne sont pas utilisables. L'étape suivante serait de pouvoir prendre en compte, dans le modèle, les pixels morts du détecteur. Cependant, ces pixels morts sont dans le domaine du détecteur, c'est-à-dire lorsque le modèle des données se fait sur pour chaque image calibrée du détecteur, de taille M, et non des données *pré-traitées* qui ont été transformées pour qu'un pixel des données corresponde au même pixel des cartes reconstruites. Pour pouvoir prendre en compte ces pixels morts, il est donc essentiel de modéliser mathématiquement ces transformations et de les prendre en compte dans le modèle. Ce sera l'objet du chapitre 3.

Dans le cadre de l'ajout de ces transformations dans le modèle, celui-ci deviendrait non-séparable et donc plus complexe. Le modèle précédent étant non-linéaire et non-convexe en θ, la recherche non-séparable du minimum global de (2.16) serait fastidieuse. Une alternative au formalisme du Jones, qui est celle principalement utilisée dans la littérature dans le cas de reconstruction de sources partiellement polarisées [Avenhaus et al., 2014], est de paramétrer les données, pour tout $n \in \{1, \dots N\}$, par les paramètres de Stokes I_n, Q_n et U_n. L'avantage d'un tel modèle est qu'il est linéaire, ce qui permet dans le cas séparable d'avoir une solution explicite.

Soit $\boldsymbol{x} = (\mathrm{I}, \mathrm{Q}, \mathrm{U})^\top \in (\mathbb{R}^N)^L$ le vecteur composé des paramètres de Stokes, c'est-à-dire d'après la section 1.1.5.2, avec :

Rappel éq. (1.16)

$$\begin{cases} \mathrm{I}_n = I^{\mathrm{u}}_n + I^{\mathrm{p}}_n, \\ \mathrm{Q}_n = I^{\mathrm{p}}_n \cos 2\theta_n, \\ \mathrm{U}_n = I^{\mathrm{p}}_n \sin 2\theta_n. \end{cases}$$

On a alors la proposition suivante :

Proposition 2.2.1. *Soit* $\overline{\overline{\boldsymbol{x}}}_n = (\overline{\overline{\mathrm{I}}}_n, \overline{\overline{\mathrm{Q}}}_n, \overline{\overline{\mathrm{U}}}_n)^\top$, *alors le modèle linéaire séparable des données s'écrit* $k \in \{1, \dots K\}$, $n \in \{1, \dots N\}$ *et* $\forall j \in \{1, 2\}$ *de la forme* $\boldsymbol{d}^S_{j,k,n} = \mathcal{B}_{j,k,n}(f^S_{j,k}(\overline{\overline{\boldsymbol{x}}}_n))$ *avec*

$$f^S_{j,k}(\boldsymbol{x}) = \sum_{\ell=1}^{3} v_{j,k,\ell}\, \overline{\overline{\boldsymbol{x}}}_{\ell,n}, \tag{2.26}$$

où

$$v_{j,k,1} = \frac{\|\boldsymbol{v}_{j,k}\|^2_{\mathbb{C}^2}}{2}, \quad v_{j,k,2} = \frac{|\boldsymbol{v}^{(x)}_{j,k}|^2 - |\boldsymbol{v}^{(y)}_{j,k}|^2}{2} \quad \textit{et} \quad v_{j,k,3} = \Re\mathfrak{e}\left(\boldsymbol{v}^{(x)}_{j,k} \boldsymbol{v}^{(y)}_{j,k}\right), \tag{2.27}$$

avec les $(\boldsymbol{v}_{j,k})_{j\in\{1,2\},k\in\{1,\dots K\}}$ *tels que définis dans la proposition 2.2.1.*

Démonstration : *(Fin page 65)* En repartant de l'expression $f_{j,k}$ dans (2.4) on a :

$$
\begin{aligned}
&\frac{\|\boldsymbol{v}_{j,k}\|^2_{\mathbb{C}^2}}{2} I^{\mathrm{u}}_n + I^{\mathrm{p}}_n |\langle \boldsymbol{v}_{j,k}, c_{\theta_n}\rangle_{\mathbb{C}^2}| \\
=&\frac{\|\boldsymbol{v}_{j,k}\|^2_{\mathbb{C}^2}}{2} I^{\mathrm{u}}_n + I^{\mathrm{p}}_n \left(|\boldsymbol{v}^{(x)}_{j,k}|^2 \cos^2\theta_n + |\boldsymbol{v}^{(y)}_{j,k}|^2 \sin^2\theta_n + 2\mathfrak{Re}\left(\boldsymbol{v}^{(x)}_{j,k}\boldsymbol{v}^{(y)}_{j,k}\right)\cos\theta_n \sin\theta_n \right) \\
=&\frac{\|\boldsymbol{v}_{j,k}\|^2_{\mathbb{C}^2}}{2} I^{\mathrm{u}}_n + I^{\mathrm{p}}_n \left(\frac{\|\boldsymbol{v}_{j,k}\|^2_{\mathbb{C}^2}}{2} - \frac{\|\boldsymbol{v}_{j,k}\|^2_{\mathbb{C}^2}}{2} + |\boldsymbol{v}^{(x)}_{j,k}|^2 \cos^2\theta_n + |\boldsymbol{v}^{(y)}_{j,k}|^2 \sin^2\theta_n + \mathfrak{Re}\left(\boldsymbol{v}^{(x)}_{j,k} v^{(y)}_{j,k}\right)\sin 2\theta_n \right) \\
=&\frac{\|\boldsymbol{v}_{j,k}\|^2_{\mathbb{C}^2}}{2} \left(I^{\mathrm{u}}_n + I^{\mathrm{p}}_n \right) + I^{\mathrm{p}}_n \left(- \frac{|\boldsymbol{v}^{(x)}_{j,k}|^2 + |\boldsymbol{v}^{(y)}_{j,k}|^2}{2} + |\boldsymbol{v}^{(x)}_{j,k}|^2 \cos^2\theta_n + |\boldsymbol{v}^{(y)}_{j,k}|^2 \sin^2\theta_n \right) + \mathfrak{Re}\left(\boldsymbol{v}^{(x)}_{j,k} v^{(y)}_{j,k}\right) \mathrm{U}_n \\
=&\frac{\|\boldsymbol{v}_{j,k}\|^2_{\mathbb{C}^2}}{2} \mathrm{I}_n + I^{\mathrm{p}}_n \left(\frac{|\boldsymbol{v}^{(x)}_{j,k}|^2}{2} \left(\cos^2\theta_n - \sin^2\theta_n\right) - \frac{|\boldsymbol{v}^{(y)}_{j,k}|^2}{2} \left(cos^2\theta_n - \sin^2\theta_n\right) \right) + \mathfrak{Re}\left(\boldsymbol{v}^{(x)}_{j,k} v^{(y)}_{j,k}\right) \mathrm{U}_n \\
=&\frac{\|\boldsymbol{v}_{j,k}\|^2_{\mathbb{C}^2}}{2} \mathrm{I}_n + I^{\mathrm{p}}_n \left(\frac{|\boldsymbol{v}^{(x)}_{j,k}|^2 - |\boldsymbol{v}^{(y)}_{j,k}|^2}{2} \cos 2\theta_n \right) + \mathfrak{Re}\left(\boldsymbol{v}^{(x)}_{j,k} v^{(y)}_{j,k}\right) \mathrm{U}_n \\
=&\frac{\|\boldsymbol{v}_{j,k}\|^2_{\mathbb{C}^2}}{2} \mathrm{I}_n + \frac{|\boldsymbol{v}^{(x)}_{j,k}|^2 - |\boldsymbol{v}^{(y)}_{j,k}|^2}{2} \mathrm{Q}_n + \mathfrak{Re}\left(\boldsymbol{v}^{(x)}_{j,k} v^{(y)}_{j,k}\right) \mathrm{U}_n \\
=&v_{j,k,1}\mathrm{I}_n + v_{j,k,2}\mathrm{Q}_n + v_{j,k,3}\mathrm{U}_n
\end{aligned}
$$

□

2.2.2 Résolution du problème séparable

Dans la suite on note $\boldsymbol{x}_n = (\mathrm{I}_n, \mathrm{Q}_n, \mathrm{U}_n)$ le vecteur de $L = 3$ paramètres à estimer au pixel $n \in \{1, \ldots, N\}$. L'estimation des paramètres I_n, Q_n et U_n est obtenue pour $n \in \{1, \ldots, N\}$ par la résolution du problème minimisation suivant :

$$
\widehat{\boldsymbol{x}}_n = \underset{\boldsymbol{x}_n \in \mathbb{R}^L}{\operatorname{argmin}} \sum_{j,k} \mathbf{W}_{j,k,n} \left(\boldsymbol{d}^{\mathrm{S}}_{j,k,n} - \sum_{\ell} v_{j,k,\ell} \boldsymbol{x}_{\ell,n} \right)^2, \tag{2.28}
$$

où les poids $\mathbf{W} \in (\mathbb{R}^M)^K$ sont donnés par (2.15). Posons alors

$$
\Phi_n(\boldsymbol{x}_n) = \sum_{j,k} \mathbf{W}_{j,k,n} \left(\boldsymbol{d}^{\mathrm{S}}_{j,k,n} - \sum_{\ell} v_{j,k,\ell} \boldsymbol{x}_{\ell,n} \right)^2. \tag{2.29}
$$

Résoudre le problème (2.28) revient à résoudre $\forall n \in \{1, \ldots, N\}$, l'équation $\nabla\Phi_n(\widehat{\boldsymbol{x}}_n) = 0$ où $\nabla\Phi_n$ représente le gradient de la fonction ϕ_n, ce qui correspond à résoudre, pour tout $n \in \{1, \ldots, N\}$, le système $\mathbf{V}_n \boldsymbol{x}_n = \boldsymbol{b}_n$ avec

$$
\mathbf{V}_n = \sum_{j,k} \mathbf{W}_{j,k,n} \begin{pmatrix} \boldsymbol{v}^2_{j,k,1} & \boldsymbol{v}_{j,k,1}\boldsymbol{v}_{j,k,2} & \boldsymbol{v}_{j,k,1}\boldsymbol{v}_{j,k,3} \\ \boldsymbol{v}_{j,k,1}\boldsymbol{v}_{j,k,2} & \boldsymbol{v}^2_{j,k,2} & \boldsymbol{v}_{j,k,2}\boldsymbol{v}_{j,k,3} \\ \boldsymbol{v}_{j,k,1}\boldsymbol{v}_{j,k,3} & \boldsymbol{v}_{j,k,2}\boldsymbol{v}_{j,k,3} & \boldsymbol{v}^2_{j,k,3} \end{pmatrix} \quad \text{and} \quad \boldsymbol{b}_n = \sum_{j,k} \mathbf{W}_{j,k,n} \boldsymbol{d}^{\mathrm{S}}_{j,k,n} \begin{pmatrix} \boldsymbol{v}_{j,k,1} \\ \boldsymbol{v}_{j,k,2} \\ \boldsymbol{v}_{j,k,3} \end{pmatrix} \tag{2.30}
$$

.

Enfin, pour retrouver les paramètres d'intérêt, il suffit de calculer :

Rappel éq. (1.27)

$$\begin{cases} I_n^{\mathrm{p}} = \sqrt{\mathrm{Q}_n^2 + \mathrm{U}_n^2} \\ \theta_n = (1/2)\arctan\left(\mathrm{U}_n/\mathrm{Q}_n\right) \\ I_n^{\mathrm{u}} = \mathrm{I}_n - I_n^{\mathrm{p}}, \end{cases}$$

ce qui correspond à la méthode suivante :

Algorithme 6 : Méthode Linéaire Séparable (MLS)

pour $n = 1, \ldots, N$ **faire**
- $\widehat{x} = \mathbf{V} \setminus \boldsymbol{b}$;
- $\widehat{I_n^{\mathrm{p}}} = \sqrt{\widehat{\boldsymbol{x}}_2^2 + \widehat{\boldsymbol{x}}_3^2}$
- $\widehat{\theta_n} = \arctan(\widehat{\boldsymbol{x}}_3/\widehat{\boldsymbol{x}}_2)/2$
- $\widehat{I_n^{\mathrm{u}}} = \widehat{\boldsymbol{x}}_1 - I_n^{\mathrm{p}}$
- **Retourner** $\left(I_n^{\mathrm{u}}, I_n^{\mathrm{p}}, \theta\right)$.

Remarque : La contrainte de positivité sur I_n^{p} est vérifiée par sa définition en les paramètres de Stokes. En revanche ce n'est pas vrai pour I_n^{u}. En effet, pour être certains que $I_n^{\mathrm{u}} \geq 0$, il faudrait vérifier que $I_n^{\mathrm{u}} + I_n^{\mathrm{p}} \geq I^{\mathrm{p}}$, ce qui mène à vérifier la contrainte épigraphique $\mathrm{I}_n \geq \sqrt{\mathrm{Q}_n^2 + \mathrm{U}_n^2}$. L'introduction d'une telle contrainte est traitée dans le chapitre 4.

2.2.3 Application sur données simulées et données astrophysiques

2.2.3.1 Applications sur données simulées

On se propose d'appliquer la MLS aux mêmes données simulées que précédemment. La figure 2.9 représente les cartes des paramètres I^{p}, I^{u} et θ, reconstruites par les méthodes de la double différence, du double ratio, par la MnLS et par la MLS, pour $\tau^{\mathrm{disk}} \in \{3\%, 10\%, 50\%\}$. De la même manière que précédemment, les résultats sont plus bruités pour les reconstructions à faible de taux de polarisation tandis que les zones plus brillantes sont mieux reconstruites.

La figure 2.10 représente les EQMn des quatre méthodes ainsi que les rapports entre les EQMn et la meilleure valeur d'EQMn pour chaque τ^{disk}. On voit que la MnLS et la MLS ont exactement la même EQMn. Les deux méthodes sont donc complètement équivalentes.

On remarque cependant, d'après la figure 2.9, que la contrainte de positivité est toujours vérifiée pour la MLS. On pourrait donc s'attendre que dans le cas où l'on a des problèmes dans les données, telles que de la saturation, qui font que le modèle n'est plus vrai, la méthode linéaire donne de meilleurs résultats.

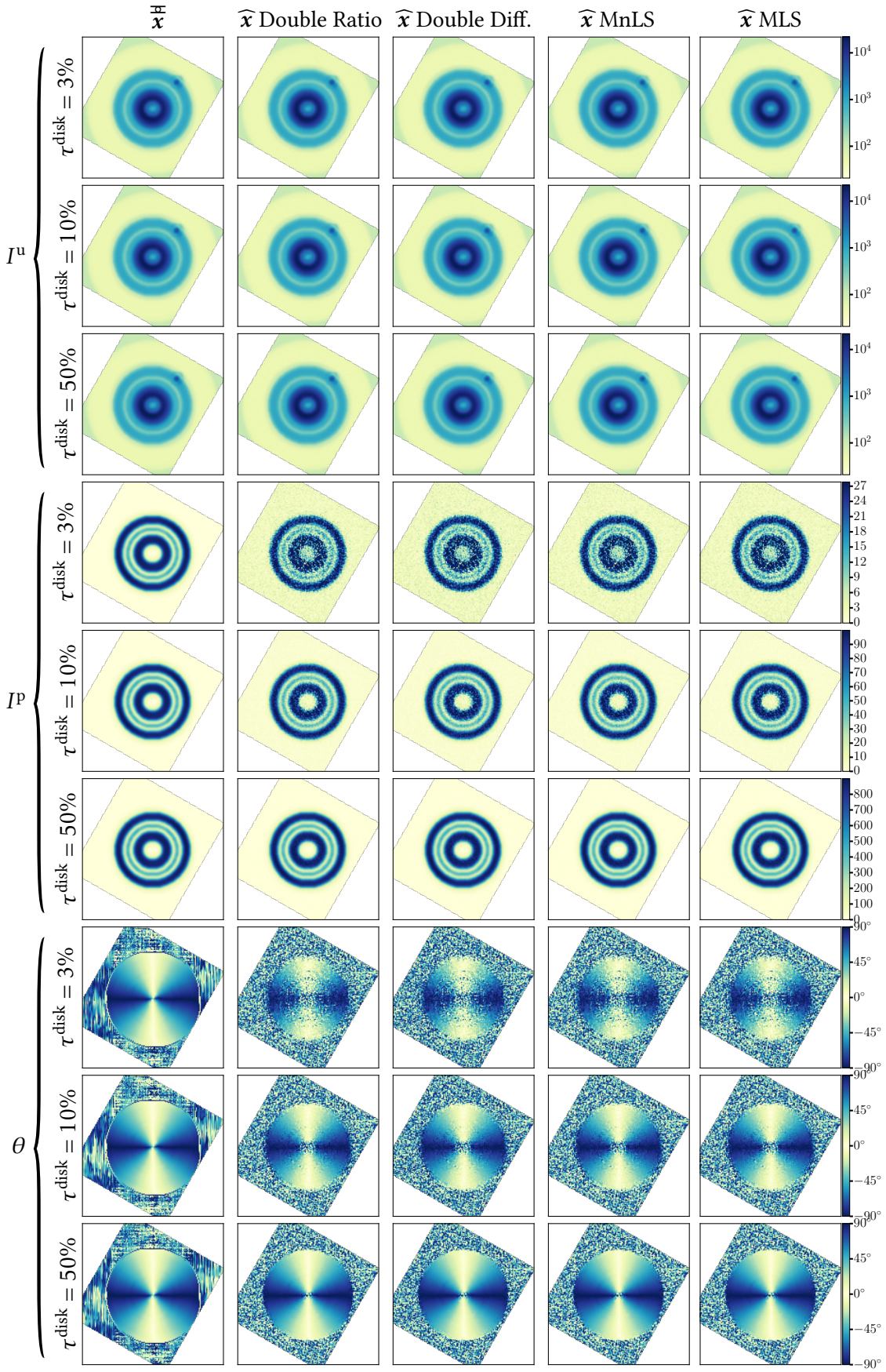
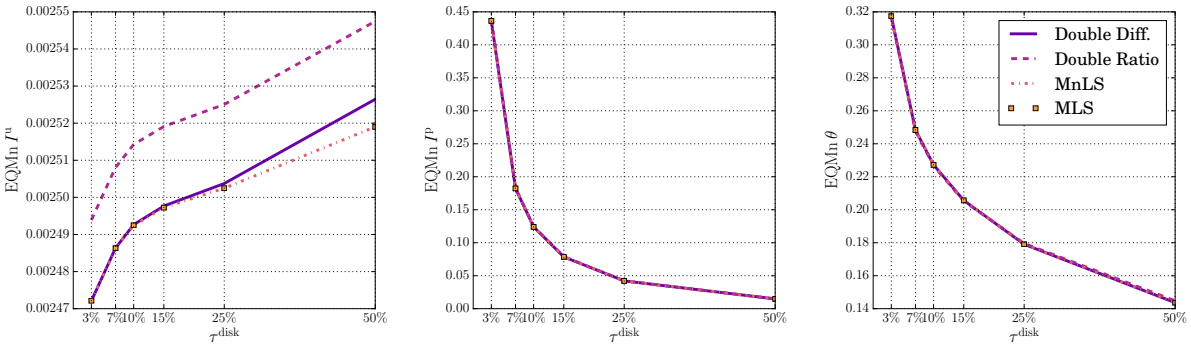
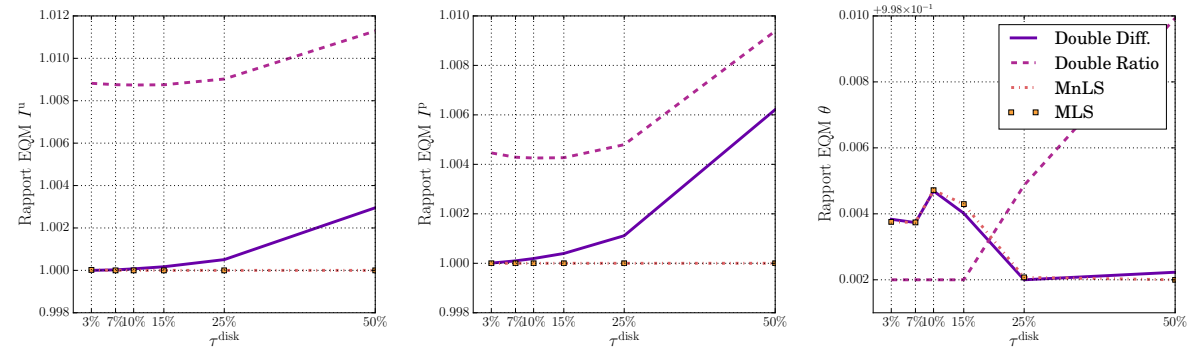


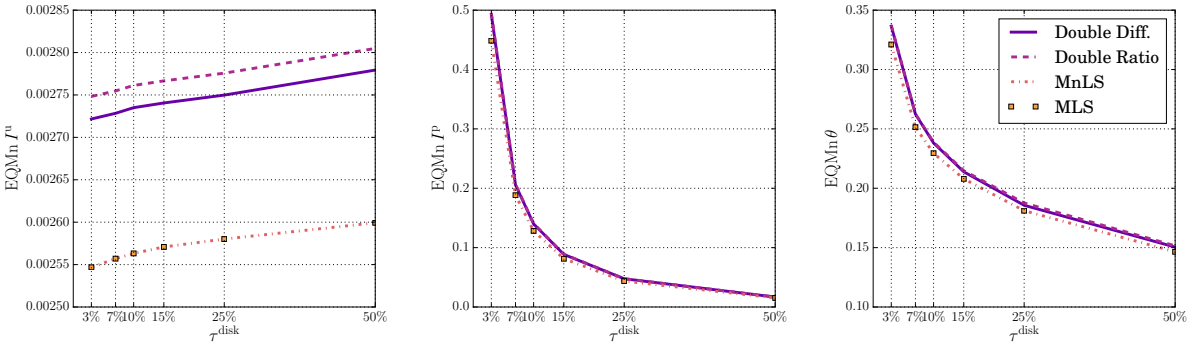
FIGURE 2.9 – Cartes reconstruites avec les différentes méthodes pour $\tau^{\mathrm{disk}} \in \{3\%, 10\%, 50\%\}$.



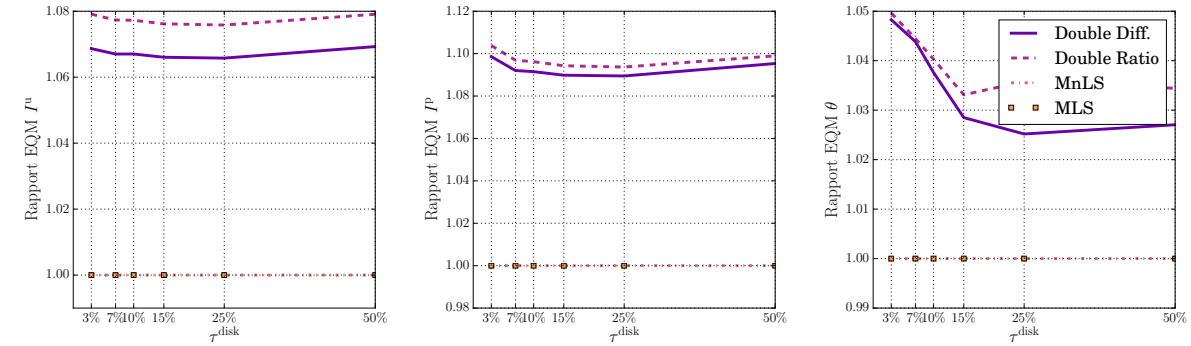
(a) EQMn entre les $\widehat{\boldsymbol{x}}$ estimés avec les différentes méthodes et $\overline{\overline{\boldsymbol{x}}}$, sans données manquantes.



(b) Différence entre les EQMn estimés avec les méthodes de l'état-de-l'art et estimés avec la MnLS, sans données manquantes.



(c) EQMn entre les $\widehat{\boldsymbol{x}}$ estimés avec les différentes méthodes et $\overline{\overline{\boldsymbol{x}}}$, avec données manquantes.



(d) Différence entre les EQMn estimés avec les méthodes de l'état-de-l'art et estimés avec la MnLS, avec données manquantes.

FIGURE 2.10 – Erreurs quadratiques moyennes normalisées des différents paramètres pour les différentes méthodes. Les figures a) et c) représentent l'EQMn EQMn($\widehat{\boldsymbol{x}}, \overline{\overline{\boldsymbol{x}}}$) pour les différentes méthodes. Les figures b) et d) représentent le rapport entre ces EQMn et la meilleure valeur de toutes les EQMn pour chaque valeur de τ^{disk}.

Remarque : Dans le cas où il n'y a pas de polarisation instrumentale, c'est-à-dire que les matrices $\mathbf{M}_1$ et $\mathbf{M}_2$ dans (2.3) sont des matrices identités, alors le modèle des données de la proposition 2.2.1est semblable au modèle des données supposé pour les méthodes de la double différence et du double ratio (*cf.* section 1.1.5).

En effet pour les acquisitions où $\alpha = 0°$, on a (*cf.* tableau 2.1) :

$$k \in \{1, \dots K_{0°}\} \quad \begin{cases} \boldsymbol{d}^{\mathrm{S}}_{1,k,n} = \mathcal{B}_{1,k,n}(0.5\mathrm{I}_n + 0.5\mathrm{Q}_n), \\ \boldsymbol{d}^{\mathrm{S}}_{2,k,n} = \mathcal{B}_{2,k,n}(0.5\mathrm{I}_n - 0.5\mathrm{Q}_n). \end{cases} \tag{2.31}$$

Pour les acquisitions où $\alpha = 45°$, on a :

$$k \in \{1, \dots K_{45°}\} \quad \begin{cases} \boldsymbol{d}^{\mathrm{S}}_{1,k,n} = \mathcal{B}_{1,k,n}(0.5\mathrm{I}_n - 0.5\mathrm{Q}_n), \\ \boldsymbol{d}^{\mathrm{S}}_{2,k,n} = \mathcal{B}_{2,k,n}(0.5\mathrm{I}_n + 0.5\mathrm{Q}_n). \end{cases} \tag{2.32}$$

Pour les acquisitions où $\alpha = 22,5°$, on a :

$$k \in \{1, \dots K_{22,5°}\} \quad \begin{cases} \boldsymbol{d}^{\mathrm{S}}_{1,k,n} = \mathcal{B}_{1,k,n}(0.5\mathrm{I}_n + 0.5\mathrm{U}_n), \\ \boldsymbol{d}^{\mathrm{S}}_{2,k,n} = \mathcal{B}_{2,k,n}(0.5\mathrm{I}_n - 0.5\mathrm{U}_n). \end{cases} \tag{2.33}$$

Pour les acquisitions où $\alpha = 77,5°$, on a :

$$k \in \{1, \dots K_{77,5°}\} \quad \begin{cases} \boldsymbol{d}^{\mathrm{S}}_{1,k,n} = \mathcal{B}_{1,k,n}(0.5\mathrm{I}_n - 0.5\mathrm{U}_n), \\ \boldsymbol{d}^{\mathrm{S}}_{2,k,n} = \mathcal{B}_{2,k,n}(0.5\mathrm{I}_n + 0.5\mathrm{U}_n). \end{cases} \tag{2.34}$$

En conclusion, sur données simulées et dans le cadre d'un modèle séparable, la MnLS et la MLS donnent des résultats identiques, ce qui est rassurant vis-à-vis de l'utilisation du modèle linéaire dans un cadre plus complexe tel que le cas non-séparable des chapitres suivants.

2.2.3.2 Applications sur données astrophysiques

On se propose maintenant d'appliquer la MLS sur le même jeu de données astrophysiques *pré-traitées* que dans la section précédente, et de comparer les reconstructions avec les résultats obtenus par la MnLS et les méthodes de la Double Différence et du Double Ratio. La figure 2.6 présente les cartes des paramètres reconstruits avec la MnLS et les méthodes de la Double Différence et du Double Ratio. La figure 2.8 présente les cartes d'approximation de l'erreur de chaque paramètre. La figure 2.8a représente l'erreur absolue faite sur les intensités estimées pour chaque méthode ainsi que l'erreur angulaire en degré. La figure 2.8b représente l'erreur relative faite sur les intensités estimées, c'est-à-dire que pour chaque pixel $n \in \{1, \dots N\}$ de chaque paramètre, l'erreur absolue a été divisée par l'intensité estimée.

On voit que les cartes de reconstruction et d'erreur de la MLS sont identiques aux cartes de la MnLS.

2.2.3.3 Synthèse des résultats

Dans cette section nous avons présenté un modèle linéaire permettant de faire le lien avec les méthodes de l'état-de-l'art. Nous avons vu que la MLS et la MnLS donnent des résultats

identiques. La pertinence de la contrainte de positivité peut alors être questionnée, cependant il est nécessaire de vérifier également sont utilité dans un modèle plus complexe avant de conclure.

Étant donnés les résultats obtenus avec la MLS et la MnLS, les conclusions vis-à-vis de l'état-de-l'art restent les mêmes que dans la section précédente.

Conclusion du chapitre

Dans ce chapitre nous avons présenté deux modèles séparables des données. Le premier est un modèle non-linéaire basé sur le formalisme de Jones, paramétré par les intensités I^{u}, I^{p} et l'angle de polarisation θ. Le second est linéaire et est paramétré par les paramètres de Stokes I, Q et U. Nous avons présenté deux méthodes d'estimation associées respectivement à ces deux modèles et avons montré qu'elles étaient identiques sauf dans le cas où le modèle des données faisait que la contrainte de positivité n'était pas respectée. Nous avons également montré que ces méthodes donnent des résultats au moins aussi bons que les méthodes de l'état-de-l'art et qu'elles avaient un avantage dans la prise en compte de la précision des données dans des cartes de poids, surtout dans le cas de données manquantes.

En conclusion, l'utilisation d'une méthode permettant de gérer les erreurs et les données manquantes, telles la MnLS et de la MLS, pour la reconstruction à partir de données *pré-traitées*, apporte un avantage qualitatif vis-à-vis de l'utilisation des méthodes de l'état-de-l'art.

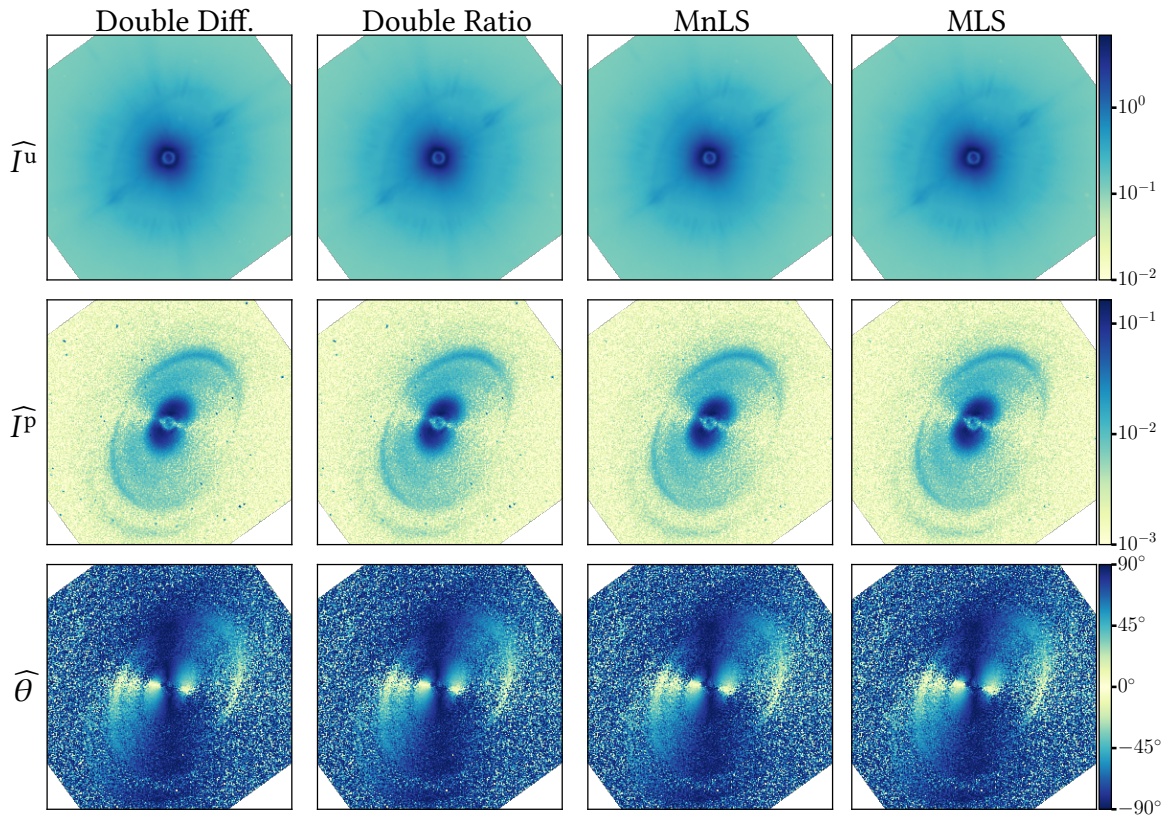
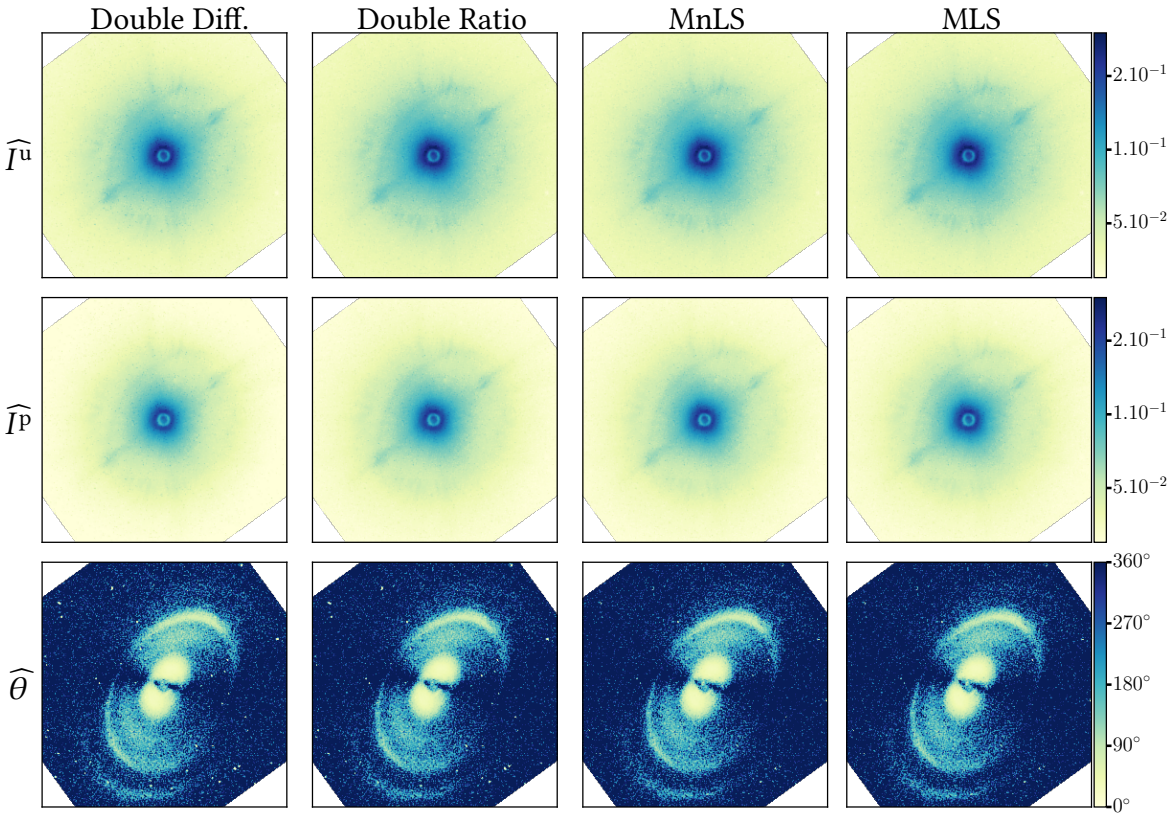
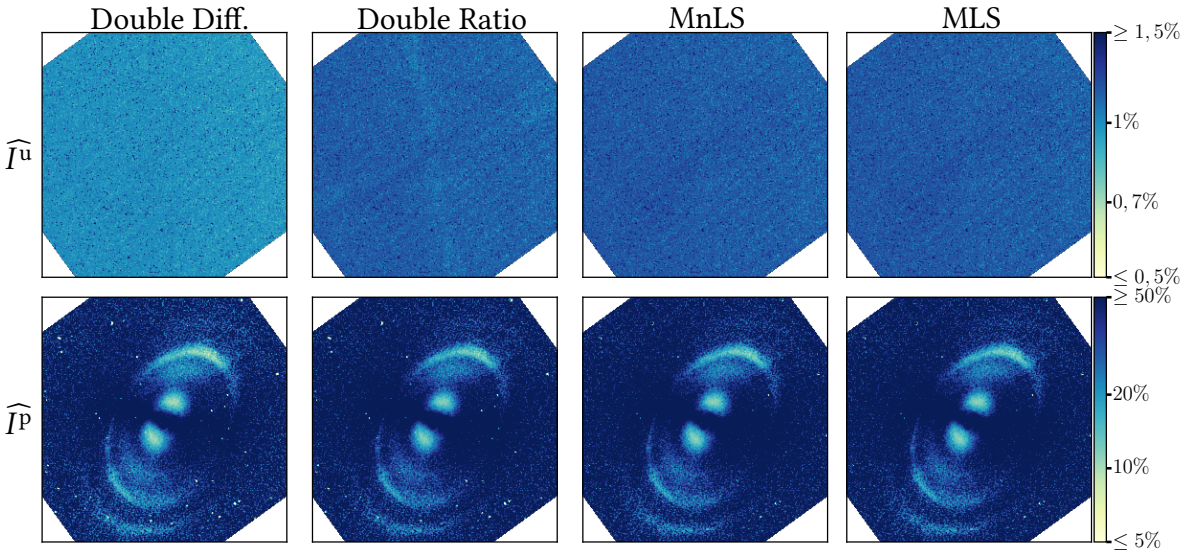


FIGURE 2.11 – Cartes reconstruites avec les différentes méthodes de la cible RXJ 1615 observée en bande H.



(a) Erreur absolue des différents paramètres



(b) Erreur relative faites sur les intensités.

FIGURE 2.12 – Cartes d'erreurs de reconstruction de la cible RXJ 1615 pour les différentes méthodes.

Chapitre 3

Modèle direct non-séparable des données de l'instrument ESO/VLT-SPHERE IRDIS

Dans le chapitre précédent, nous avons établi deux modèles directs séparables des données, dans le cas de données *pré-traitées*. Le problème d'un tel modèle est qu'il ne nous permet pas de prendre en compte les pixels morts ou défectueux du détecteur. En effet pour obtenir les données *pré-traitées*, notées $\boldsymbol{d}^{\mathrm{S}}$, les données *calibrées* contenant les mauvais pixels, notées $\boldsymbol{d}^{\mathrm{nS}}$, ont été coupées, translatées et tournées. Ces transformations des données impliquent une interpolation des pixels qui risque de propager les erreurs si celles-ci ne sont pas traitées *a priori*. De ce fait, avant de faire les transformations, il est nécessaire de traiter les mauvais pixels. Dans le cas séparable, ceux-ci sont donc traités, par interpolation des plus proches voisins. Or, d'après les résultats obtenus dans le chapitre précédent, la prise en compte des données manquantes dans des cartes de poids apporte une amélioration non négligeable sur l'erreur de la solution. Il semble donc pertinent de se demander si la prise en compte dune carte des pixels défectueux dans le modèle aura un effet aussi significatif.

Pour prendre en compte les pixels morts dans le modèle, il est alors nécessaire de prendre également en compte l'ensemble des transformations (découpage, translations et rotations) permettant de passer du domaine des cartes reconstruites au domaine des données calibrées, que nous précisons dans cette section.

Rappelons que le but de la modélisation directe est d'avoir une expression des données $\boldsymbol{d} = (\boldsymbol{d}_{k,m})_{k\in\{1,\dots K\},m\in\{1,\dots,M\}}$ en fonction du signal d'intérêt $\overline{\overline{\boldsymbol{x}}} = (\boldsymbol{x}_{\ell,n})_{\ell\in\{1,\dots L\},n\in\{1,\dots,N\}}$ sous la forme :

Rappel éq. (2.1)

$$\boldsymbol{d} = \mathcal{B}(f(\overline{\overline{\boldsymbol{x}}})),$$

où $f : (\mathbb{R}^N)^L \to (\mathbb{R}^M)^K$ est un opérateur de déformation fixe, et $\mathcal{B} : (\mathbb{R}^M)^K \to (\mathbb{R}^M)^K$ un opérateur de bruit aléatoire.

Dans ce chapitre, nous nous intéressons à un modèle des données *calibrées*, non-séparables en les pixels $n \in \{1,\dots,N\}$. Cela signifie qu'un pixel $m \in \{1,\dots M\}$ sur le détecteur correspond à une combinaison de tous les pixels $n \in \{1,\dots,N\}$ des paramètres d'intêret. On notera le modèle des données :

$$\boldsymbol{d}_k^{\mathrm{nS}} = \mathcal{B}_k(f_k^{\mathrm{nS}}(\overline{\overline{\boldsymbol{x}}})), \tag{3.1}$$

où $f_k^{\mathrm{nS}} : (\mathbb{R}^N)^L \to \mathbb{R}^M$ modélise l'instrument et les transformations du détecteur pour passer du domaine des cartes reconstruites au domaine des données, et $\mathcal{B}_k : \mathbb{R}^M \to \mathbb{R}^M$ représente le bruit des données.

Le problème d'une telle modélisation, est qu'un opérateur d'interpolation est souvent mal conditionné. De ce fait, son utilisation lors de la résolution par approche inverse risque de propager les erreurs. Il est donc nécessaire, lors de la résolution, d'ajouter un terme de régularisation des cartes reconstruites.

Le but de ce chapitre est double. D'une part, modéliser l'ensemble des fonctions $(f_k)_{k\in\{1,\dots,K\}}$ et la distribution de l'ensemble des opérateurs de bruit $(\mathcal{B}_k)_{k\in\{1,\dots,L\}}$ de la formulation (3.1). D'autre part, montrer l'apport de l'utilisation d'une méthode non-séparable régularisée prenant en compte les mauvais pixels, par rapport à une méthode séparable où ils auraient été interpolés *a priori* (cf. Chapitre 2).

Ce chapitre est organisé comme suit. Dans la section 3.1, je présente un premier modèle direct, linéaire, des données calibrées de l'instrument, basé sur le formalisme de Stokes, paramétré par le signal d'intérêt $\boldsymbol{x}^{\text{stokes}} = (\mathrm{I}, \mathrm{Q}, \mathrm{U})^\top$. Je propose ensuite une méthode d'estimation des paramètres de ce modèle prenant en compte un terme de régularisation différentiable. Nous l'appellerons Méthode Linéaire non-Séparable (MLnS). Je compare les performances de la MLnS avec les performances de la Méthode non-Linéaire Séparable (MnLS), et de la méthode de l'état-de-l'art dite du Double Différence pour la reconstruction à partir de données *pré-traitées*, sur données simulées et données astrophysiques.

Dans la section 3.2, je propose une reformulation non-linéaire du modèle direct linéaire, paramétré par le signal d'intérêt $\boldsymbol{x}^{\text{non-lin}} = (I^{\mathrm{u}}, \mathrm{Q}, \mathrm{U})^\top$. Je présente ensuite une méthode d'estimation des paramètres, prenant en compte un terme de régularisation non-linéaire des paramètres ainsi qu'une contrainte de positivité sur I^{u}. Cette méthode est appelée Méthode non-Linéaire non-Séparable (MnLnS). Enfin j'étudie et compare les performances de la MnLnS avec celles de la MLnS pour la reconstruction à partir données calibrées, ainsi qu'avec les performances de la MnLS et de la Double Différence pour la reconstruction à partir de données *pré-traitées*.

3.1 Modèle direct linéaire non-séparable

3.1.1 Le modèle direct non-séparable linéaire

Nous représentons les paramètres de Stokes par les vecteurs $(\boldsymbol{x}_\ell)_{\ell\in\{1,\dots L\}} \in \mathbb{R}^N$. Rappelons que le modèle séparable des données en les paramètres de Stokes s'écrit pour toute acquisition $k \in \{1,\dots K\}$ et pour tout pixel $n \in \{1,\dots N\}$, pour la partie $j \in \{1,2\}$ du détecteur, de la forme $\boldsymbol{d}_{j,k,n}^{\mathrm{nS}} = \mathcal{B}_{j,k,n}(f_{j,k}^{\mathrm{nS}}(\boldsymbol{x}))$ avec :

Rappel éq. (2.26)
$$f_{j,k}^{\mathrm{nS}}(\boldsymbol{x}) = \sum_{\ell=1}^{3} v_{j,k,\ell}\, \overline{\boldsymbol{x}}_{\ell,n},$$

où les $v_{j,k,\ell}$ représentent l'action de l'instrument, à l'acquisition k, pour la partie j du détecteur, sur la composante ℓ du signal d'intérêt.

On définit par les opérateurs d'interpolation $\mathbf{T}_{j,k} : \mathbb{R}^N \to \mathbb{R}^M/2$, l'ensemble des translations et rotations des images gauches et droites du détecteur permettant de passer du domaine des cartes reconstruites au domaine des données. Ils consistent en une transformée des coor-

données des pixels, suivit d'une interpolation des valeurs des pixels voisins suivant un noyau d'interpolation donné. On a alors la proposition suivante :

Définition 3.1.1. *Soit* $\overline{\overline{\boldsymbol{x}}} = (\overline{\overline{\mathrm{I}}}, \overline{\overline{\mathrm{Q}}}, \overline{\overline{\mathrm{U}}})^\top$ *et* $(\mathbf{T}_{j,k})_{j\in\{1,2\},k\in\{1,\dots K\}} : \mathbb{R}^N \to \mathbb{R}^M/2$ *un ensemble d'opérateurs linéaires, alors le modèle linéaire non-séparable des données* $\boldsymbol{d} \in (\mathbb{R}^M)^K$ *s'écrit* $\forall k \in \{1,\dots K,\}$, *de la forme* $\boldsymbol{d}_k = \mathcal{B}_k(f_k(\overline{\overline{\boldsymbol{x}}}))$ *avec*

$$f_k^{nS}(\boldsymbol{x}) = \sum_{\ell=1}^{3} \begin{bmatrix} \mathbf{T}_{1,k}\boldsymbol{v}_{1,k,\ell} \\ \mathbf{T}_{2,k}\boldsymbol{v}_{2,k,\ell} \end{bmatrix} \overline{\overline{\boldsymbol{x}}}_\ell \tag{3.2}$$

où

Rappel éq. (2.27) $v_{j,k,1} = \frac{\|\boldsymbol{v}_{j,k}\|^2_{\mathbb{C}^2}}{2}, \quad v_{j,k,2} = \frac{|\boldsymbol{v}_{j,k}^{(x)}|^2 - |\boldsymbol{v}_{j,k}^{(y)}|^2}{2} \quad et \quad v_{j,k,3} = \mathfrak{Re}\left(\boldsymbol{v}_{j,k}^{(x)}\boldsymbol{v}_{j,k}^{(y)}\right).$

Remarque : Les opérateurs $(\mathbf{T}_{j,k})_{k\in\{1,\dots,K\},j=1,2}$ étant linéaires par rapports aux cartes reconstruites, les $(\mathbf{T}_{j,k})_{k\in\{1,\dots,K\},j=1,2}$ commutent avec $(\sum_{\ell=1}^{L} \boldsymbol{v}_{j,k,\ell}\overline{\overline{\boldsymbol{x}}}_\ell)_{k\in\{1,\dots,K\},j=1,2}$. La formulation (3.2) est donc équivalente à

$$f_k^{\mathrm{nS}}(\boldsymbol{x}) = \begin{bmatrix} \mathbf{T}_{1,k} f_{1,k}^{\mathrm{S}} \\ \mathbf{T}_{1,k} f_{2,k}^{\mathrm{S}} \end{bmatrix}(\boldsymbol{x}), \tag{3.3}$$

où les $(f_{j,k}^{\mathrm{S}})_{j\in\{1,2\},k\in\{1,\dots K\}}$ sont donnés par (2.26), avec $(f_{j,k}^{\mathrm{S}}(\boldsymbol{x}))_n = f_{j,k}(\boldsymbol{x}_n)$ pour tout $n \in \{1,\dots N\}$.
Si on souhaite changer de paramétrisation, il s'agit donc simplement de remplacer les $(f_{j,k}^{\mathrm{S}})_{j\in\{1,2\},k\in\{1,\dots K\}}$, par exemple par le modèle non-linéaire (2.4) dans le cas où $\boldsymbol{x} = (I^{\mathrm{u}}, I^{\mathrm{p}}, \theta)^\top$.

L'avantage de l'expression (3.3) est qu'elle ne nécessite d'appliquer les opérateurs $(\mathbf{T}_{j,k})_{k\in\{1,\dots,K\},j=1,2}$ qu'une seule fois chacun au lieu de L fois, ce qui fait un gain considérable d'opérations lors de l'utilisation de cette fonction dans un algorithme nécessitant beaucoup d'itérations pour converger.

3.1.2 Résolution par approche inverse différentiable

Pour estimer les paramètres I, Q et U du modèle des données exprimé dans la proposition 3.1.1, comme dans le chapitre précédent, on cherche les paramètres maximisant la vraisemblance des données. Cela correspond à trouver les paramètres minimisant le carré de la distance de Mahalanobis associée au modèle, c'est-à-dire :

$$\widehat{\boldsymbol{x}} \in \underset{\boldsymbol{x}\in\mathbb{R}^L}{\mathrm{Argmin}} \sum_{k=1}^{K} \frac{1}{2} \left\| \boldsymbol{d}_k^{\mathrm{nS}} - f_k^{\mathrm{nS}}(\boldsymbol{x}) \right\|^2_{\mathbf{W}_k}. \tag{3.4}$$

où

$$\forall k \in \{1,\dots K\},\ \forall m \in \{1,\dots M\}, \quad \mathbf{W}_{k,m} = \begin{cases} \mathrm{Cov}(\boldsymbol{d}_{k,m}^{\mathrm{nS}})^{-1}, \\ 0 \quad \text{si pixel ou acquisition invalides.} \end{cases} \tag{3.5}$$

La différence entre cette formulation et la formulation (2.29) est que ce critère ne peut pas être résolu indépendamment en chaque pixel $n \in \{1,\dots N\}$, car un pixel $m \in \{1,\dots M\}$ ne correspond pas à un unique pixel $n \in \{1,\dots N\}$.

Les données sont supposées indépendantes dans le temps et pixel-à-pixel, d'où le fait que $(\mathrm{Cov}(\boldsymbol{d})^{-1})_{k,m} = \mathrm{Cov}(\boldsymbol{d}_{k,m})^{-1}$. Dans la suite on notera l'attache aux données

$$\Phi(\boldsymbol{x}) = \sum_{k=1}^{K} \frac{1}{2} \|\boldsymbol{d}_k - f_k(\boldsymbol{x})\|^2_{\mathbf{W}_k}. \tag{3.6}$$

La fonction Φ est convexe et différentiable de gradient β-Lipschitz, avec $\beta = \sum_{k=1}^{K} \|f_k\|$. Le problème d'une telle attache aux données est que les opérateurs de translations, ont leurs plus petites valeurs propres nulles ou proches de zéro. Ils sont donc mal conditionnés. De ce fait, comme cela a été montré dans la partie 1.2.3, une résolution par approche inverse d'une telle attache aux données amplifierait le bruit dans la reconstruction. C'est pourquoi il est nécessaire d'ajouter un terme de régularisation.

Dans ce chapitre, nous comparons deux régularisations. D'une part, la régularisation de Tikhonov, dont la formule est donnée par l'équation (1.47), qui consiste en une norme ℓ_2 sur le gradient des paramètres reconstruits. Comme nous appliquons cette régularisation indépendamment sur les cartes I, Q et U, la fonction de régularisation s'écrit alors pour $\lambda_\ell \geq 0$ comme :

$$\mathcal{R}(\boldsymbol{x}) = \frac{1}{2} \sum_{\ell=1}^{L} \lambda_\ell \|\mathbf{D}\boldsymbol{x}_\ell\|_2^2, \tag{3.7}$$

où $\mathbf{D}$ correspond ici à l'opérateur de différences finies normalisées. La fonction $\mathcal{R}$ est convexe, différentiable de gradient λ-Lipschitz.

D'autre part, l'approximation hyperbolique de la Variation Totale (TV-h), dont la formule est donnée par l'équation (1.50). Une telle régularisation devrait nous permettre des bords plus francs dans les cartes d'intensités. Le terme de régularisation est donné $\forall \lambda_\ell, \mu_\ell \geq 0$ par

$$\mathcal{R}(\boldsymbol{x}) = \sum_{\ell=1}^{L} \lambda_\ell \sum_{n} \left(\sqrt{\|\mathbf{D}_n \boldsymbol{x}_\ell\|_2^2 + \mu_\ell^2} - \mu_\ell \right), \tag{3.8}$$

qui est différentiable de gradient $\max_\ell\{\lambda_\ell/\mu_\ell\}$-Lipschitz.

Finalement, reconstruire les paramètres I, Q et U à partir du modèle des données de la définition 3.1.1 revient à résoudre le problème :

$$\widehat{\boldsymbol{x}} \in \underset{\boldsymbol{x} \in (\mathbb{R}^N)^L}{\mathrm{Argmin}}\ \Phi(\boldsymbol{x}) + \mathcal{R}(\boldsymbol{x}), \tag{3.9}$$

avec $\Phi, \mathcal{R} : (\mathbb{R}^N)^L \to \mathbb{R}$ convexes et différentiables. Posons maintenant $\mathcal{F}(\boldsymbol{x}) = \Phi(\boldsymbol{x}) + \mathcal{R}(\boldsymbol{x})$. Comme $\mathcal{F}$ est convexe car somme de deux fonctions convexes alors on a :

$$\widehat{\boldsymbol{x}} \in \underset{\boldsymbol{x}}{\mathrm{Argmin}}\ \mathcal{F}(\boldsymbol{x}) \Leftrightarrow \nabla \mathcal{F}(\widehat{\boldsymbol{x}}) \ni 0, \tag{3.10}$$

où ∇ représente ici le gradient de la fonctionnelle.

Le système n'étant pas inversible, il est nécessaire de passer par des méthodes itératives telles que les descentes de gradient ou les méthodes de Newton. C'est pourquoi nous avons

choisi, pour résoudre le problème (3.9), une méthode quasi-Newton. Nous utilisons une méthode de gradient préconditionné avec un préconditionnement par Broyden-Fletcher-Goldfarb-Shanno à mémoire limitée (ℓ-BFGS), introduite dans la partie 1.3.2.

Les paramètres $\widehat{I^{\mathrm{u}}}$, $\widehat{I^{\mathrm{p}}}$ et $\widehat{\theta}$ sont ensuite retrouvés par combinaison des paramètres de Stokes :

Rappel éq. (1.27)

$$\begin{cases} \widehat{I^{\mathrm{p}}}_n = \sqrt{\widehat{\mathrm{Q}}_n^2 + \widehat{\mathrm{U}}_n^2} \\ \widehat{\theta}_n = (1/2)\arctan\left(\widehat{\mathrm{U}}_n / \widehat{\mathrm{Q}}_n\right) \\ \widehat{I^{\mathrm{u}}_n} = \widehat{\mathrm{I}}_n - \widehat{I^{\mathrm{p}}_n}. \end{cases}$$

Cependant, tout comme dans le cas séparable, la positivité en chaque pixel $n \in \{1, \dots N\}$ du paramètre I^{u} ne peut pas être assurée.

3.1.3 Simulation des données calibrées

Pour pouvoir tester les méthodes non-séparables, il est nécessaire d'avoir des données synthétiques calibrées. Celles-ci sont créées à partir des mêmes cartes $\overline{\overline{I^{\mathrm{u}}}}$, $\overline{\overline{I^{\mathrm{p}}}}$ et $\overline{\overline{\theta}}$ que celles simulées dans la partie 2.1.3. En utilisant le fait que (2.4) et (2.26) sont équivalentes, je combine d'abord les cartes en utilisant la fonction (2.4), puis je convolue le résultat par une PSF instrumentale afin de ressembler aux données astrophysiques, où les objets ont été convoluées par la PSF instrumentale. J'applique ensuite les opérateurs linéaires selon (2.4) pour passer dans le domaine des données calibrées. J'effectue alors $K \times M$ réalisations de bruit à partir d'une graine de générateur aléatoire fixe, afin de constituer le cube de données calibrées. Sont également introduits 10% de pixels morts choisis aléatoirement. Les cartes de poids associées à chaque acquisition sont simulées en parallèle d'après (3.5).

Comme dans la section précédente, j'ai simulés des jeux de données pour six taux de polarisation du disque $\tau^{\mathrm{disk}} \in \{3\%, 7\%, 10\%15\%, 25\%, 50\%\}$.

Pour pouvoir comparer les méthodes non-séparables avec les méthodes séparables, il est nécessaire de pré-traiter les données calibrées simulées. Les pixels morts sont donc interpolés en faisant une interpolation linéaire des plus proches voisins. Les données aux pixels morts interpolés sont ensuite tournées et translatées en utilisant la transformées de coordonnées inverse et en interpolant selon le même noyau de convolution que pour la simulation des données calibrées. Les mêmes étapes sont également appliquées sur les cartes de poids.

3.1.4 Application sur données simulées et données astrophysiques

Dans cette section, nous étudions les performances et la qualité des résultats obtenus par la Méthode Linéaire non-Séparable en comparant les résultats obtenus, sur données simulées et sur données astrophysiques *calibrées*, aux résultats obtenus par la Méthode Linéaire Séparable (MLS) et le Double Ratio sur données *pré-traitées*.

La MLnS est une méthode régularisée par un ensemble de paramètres $(\lambda_\ell)_{\ell \in \{1,\dots L\}}$ qui nécessitent d'être réglés de manière optimale, c'est-à-dire de manière à optimiser un critère de qualité donné. Dans le cas de l'application sur données simulées, il est possible de choisir les hyperparamètres comme ceux minimisant la somme des Erreurs Quadratiques Moyennes normalisées des composantes du paramètre d'intérêt $\boldsymbol{x} \in (\mathbb{R}^N)^L$.

Nous rappelons que pour la composante $\ell \in \{1,\dots,L\}$ d'une estimation $\widehat{\boldsymbol{x}} \in (\mathbb{R}^N)^L$ du

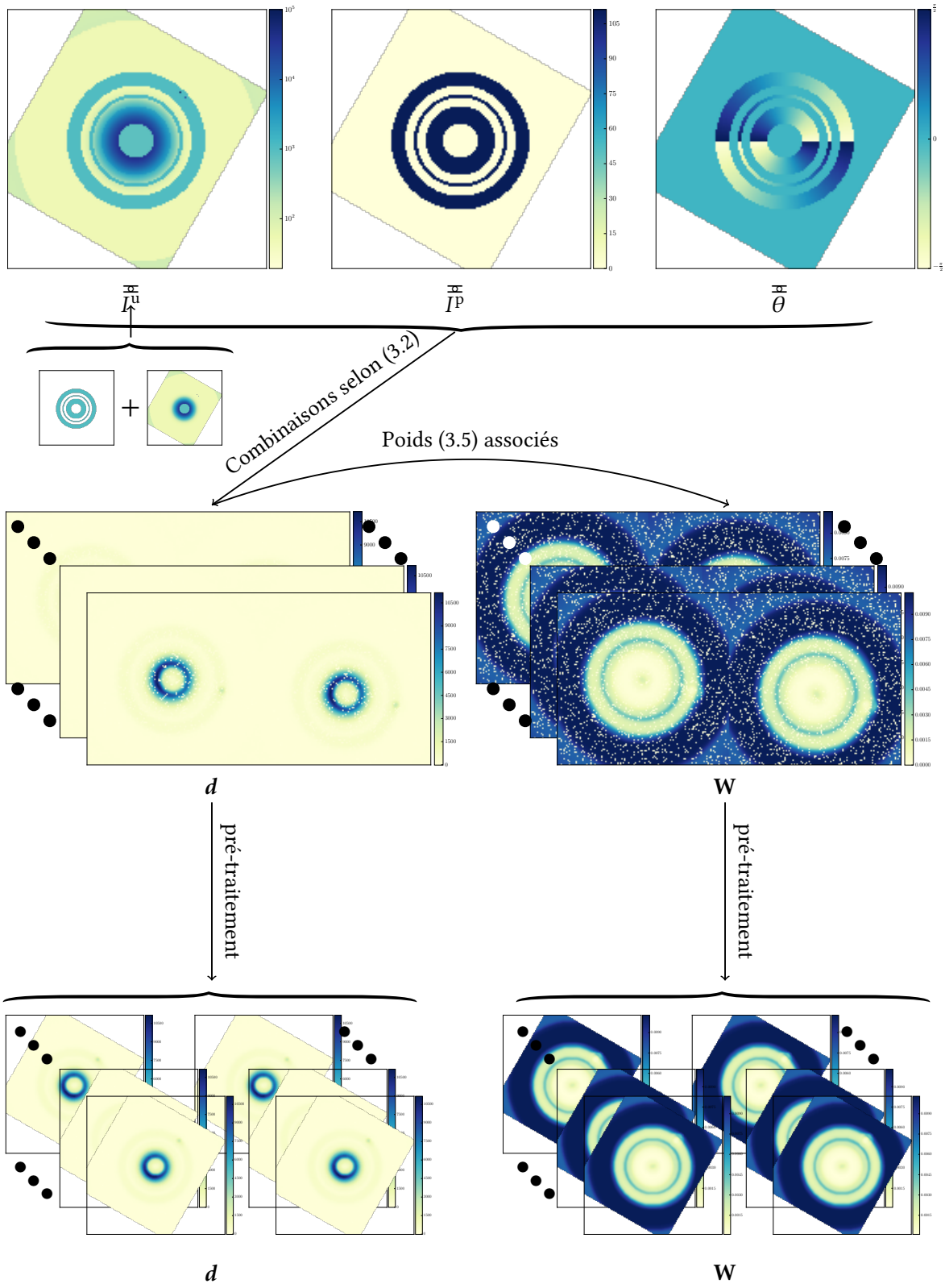


FIGURE 3.1 – Simulation des données calibrées et des données pré-traitées, pour $\tau^{\mathrm{disk}} = 10\%$.

paramètre vrai $\overline{\overline{\boldsymbol{x}}} \in (\mathbb{R}^N)^L$, l'Erreur Quadratique Moyenne normalisée est définie pour tout $\ell \in \{1, \dots L\}$ par :

Rappel éq. (2.24)
$$\mathrm{EQMn}(\widehat{\boldsymbol{x}}_\ell, \overline{\overline{\boldsymbol{x}}}_\ell) = \frac{\sum_{n \in \widetilde{N}_\ell} (\widehat{\boldsymbol{x}}_{\ell,n} - \overline{\overline{\boldsymbol{x}}}_{\ell,n})^2}{\sum_{n=1}^{N} \overline{\overline{\boldsymbol{x}}}_{\ell,n}^2},$$

où $\widetilde{N}_\ell$ est l'ensemble des pixels valides ou représentatifs de la composante $\ell \in \{1, \dots L\}$.

Dans la suite on note alors $(\lambda_\ell)^{\text{EQMn-P}}_{\ell \in \{1,\dots L\}}$ l'ensemble des hyperparamètres minimisant la somme des EQMn dans le domaine des paramètres. On note $\widehat{\boldsymbol{x}}^{\text{EQMn-P}} = (\widehat{\mathrm{I}}^{\text{EQMn-P}}, \widehat{\mathrm{Q}}^{\text{EQMn-P}}, \widehat{\mathrm{U}}^{\text{EQMn-P}})^\top$ les paramètres reconstruits par la méthode linéaire pour un tel ensemble d'hyperparamètres et $\widehat{I^{\mathrm{u}}}^{\text{EQMn-P}}$, $\widehat{I^{\mathrm{p}}}^{\text{EQMn-P}}$ et $\widehat{\theta}^{\text{EQMn-P}}$ les paramètres associés.

Dans le cas de la reconstruction sur données astrophysiques, rappelons que comme la vérité terrain n'est pas connue, il n'est pas possible d'utiliser l'EQMn pour juger la qualité de la reconstruction. De plus, comme dans cette section nous régularisons la solution, nous biaisons notre reconstruction. De ce fait il n'est pas non-plus possible d'utiliser la borne de Fréchet-Darmois-Cramer-Rao (FDCR) pour avoir une borne inférieure à l'EQM des paramètres. En effet, la borne de FDCR n'est qu'une borne inférieure de la covariance. Cela ne nous donne donc aucune information sur le biais et donc sur l'EQM.

C'est pourquoi nous utilisons alors l'estimateur non-biaisé du risque de Stein (SURE : *Stein Unbiased Risk Estimator*), qui est un estimateur de l'EQM dans le domaine des données, c'est-à-dire :

Rappel éq. (5.66)
$$\mathrm{EQM}^{\text{données}}(\widehat{\boldsymbol{x}}) = \sum_{k=1}^{K} \mathbb{E}\left[(f_k^{\mathrm{nS}}(\widehat{\boldsymbol{x}}) - f_k^{\mathrm{nS}}(\overline{\overline{\boldsymbol{x}}}))^2 \right].$$

Pour une étude plus approfondie, se référer aux sections 5.1.3 pour une application simple et 5.3 pour l'application à ces méthodes. Dans la suite on note alors $(\lambda_\ell)^{\text{SURE}}_{\ell \in \{1,\dots L\}}$ l'ensemble des hyperparamètres minimisant le critère SURE. On note $\widehat{\boldsymbol{x}}^{\text{SURE}} = (\widehat{\mathrm{I}}^{\text{SURE}}, \widehat{\mathrm{Q}}^{\text{SURE}}, \widehat{\mathrm{U}}^{\text{SURE}})^\top$ les paramètres reconstruits par la méthode linéaire pour un tel ensemble d'hyperparamètres et $\widehat{I^{\mathrm{u}}}^{\text{SURE}}$, $\widehat{I^{\mathrm{p}}}^{\text{SURE}}$ et $\widehat{\theta}^{\text{SURE}}$ les paramètres intensités et angle associés.

Dans la section 5.3, il est montré que pour le cas linéaire, l'ensemble de hyperparamètres $(\lambda_\ell)^{\text{EQMn-P}}_{\ell \in \{1,\dots L\}}$ est très différent de $(\lambda_\ell)^{\text{SURE}}_{\ell \in \{1,\dots L\}}$. C'est pourquoi dans le cas de données simulées, nous comparons les résultats obtenus pour chacun des deux ensembles d'hyperparamètres. Dans le cas des données astrophysique, nous comparons les résultats pour un ensemble d'hyperparamètres donnant une valeur du critère SURE minimale.

Par ailleurs, les paramètres Q et U contenant les mêmes structures, on émet l'hypothèse qu'ils ont la même régularité, c'est-à-dire $\lambda_2 = \lambda_3$. De ce fait, dans le cas de la MLnS, nous avons seulement deux hyperparamètres à estimer.

3.1.4.1 Résultats sur données simulées

Comme dans le chapitre précédent, nous avons appliqué la méthode sur données simulées pour différent taux de polarisation du disque $\tau^{\mathrm{disk}} \in \{3\%; 7\%; 10\%; 15\%; 25\%; 50\%\}$. Pour la correspondance en termes de SNR et de taux total de polarisation τ^{tot}, se référer à la figure 2.3.

Sur la figure 3.2 sont affichées les reconstructions obtenues pour $\tau^{\mathrm{disk}} \in \{3\%; 10\%; 50\%\}$, par la méthode de la Double Différence et de la Méthode non-Linéaire Séparable (MnLS) sur données *pré-traitées* et par la Méthode Linéaire non-Séparable (MLnS) sur données *calibrées*,

avec régularisation de Tikhonov et a préservation de bord TV-h, pour des valeurs de $\mu = 0,1$ et $\mu = 1$. On fixe les hyperparamètres $(\lambda_\ell)_{\ell\in\{1,\dots L\}}$ comme ceux minimisant d'une part l'EQMn dans le domaine des paramètres, notée EQMn-P, et d'autre par le critère SURE. Visuellement, il est difficile de voir une différence vis-à-vis du paramètre I^{u} reconstruit. En revanche, pour le paramètre I^{p}, on voit que pour $\tau^{\mathrm{disk}} \in \{3\%; 10\%\}$ il y a clairement une différence entre les reconstructions avec $(\lambda_\ell)^{\mathrm{SURE}}_{\ell\in\{1,\dots L\}}$ et les reconstructions avec $(\lambda_\ell)^{\mathrm{EQMn\text{-}P}}_{\ell\in\{1,\dots L\}}$.

En effet les valeurs reconstruites dans les anneaux proches du centre pour les hyperparamètres $(\lambda_\ell)^{\mathrm{SURE}}_{\ell\in\{1,\dots L\}}$ ont une intensité plus faible et semblent plus régularisés. De plus, dans le cas des hyperparamètres réglés avec SURE, la régularisation semble plus forte, lors de la régularisation, par TV-h pour $\mu = 0,1$ qu'avec Tikhonov. Notons que les anneaux au centre correspondent à une zone où le contraste est très élevé et donc le SNR très faible. Par ailleurs on voit que lorsque μ augmente, *i.e.* $\mu = 1$, la reconstruction tend à ressembler à celle obtenue avec une régularisation de Tikhonov. Lorsqu'on regarde les cartes d'angles θ reconstruites, pour $\tau^{\mathrm{disk}} = 3\%$, la reconstruction avec $(\lambda_\ell)^{\mathrm{SURE}}_{\ell\in\{1,\dots L\}}$ semble bien meilleure.

Pour juger la qualité des reconstruction de manière plus rigoureuse, la figure 3.3 représente l'EQMn des paramètres $\widehat{I^{\mathrm{u}}}$, $\widehat{I^{\mathrm{p}}}$ et $\widehat{\theta}$ pour les différentes méthodes en fonction du taux de polarisation du disque. Dans la figure 3.3a sont tracés les EQMn dans le cas où les hyperparamètres minimisent le critère SURE. Dans la figure 3.3b sont tracés les EQMn dans le cas où les hyperparamètres minimisent l'erreur quadratique moyenne des paramètres.

Pour l'intensité non-polarisée I^{u}, on voit que l'EQMn des paramètres reconstruits avec la MLnS, pour les hyperparamètres $(\lambda_\ell)^{\mathrm{SURE}}_{\ell\in\{1,\dots L\}}$ est toujours plus faible que celle des reconstructions séparables. Pour le choix d'hyperparamètres $(\lambda_\ell)^{\mathrm{EQMn\text{-}P}}_{\ell\in\{1,\dots L\}}$, l'erreur des méthodes non-séparables est rarement plus faible que l'erreur faite sur les méthodes séparables.

Pour l'intensité polarisée I^{p}, on voit que pour les hyperparamètres $(\lambda_\ell)^{\mathrm{SURE}}_{\ell\in\{1,\dots L\}}$, l'EQMn est plus élevée que l'EQMn des méthodes séparables, sauf à très faible taux de polarisation. On voit par ailleurs que l'erreur est plus faible pour une reconstruction avec TV-h, mais que le choix de μ ne semble pas grandement l'influencer. Pour le choix des hyperparamètres $(\lambda_\ell)^{\mathrm{EQM\text{-}P}}_{\ell\in\{1,\dots L\}}$, on voit que pour $\tau^{\mathrm{disk}} < 25\%$, pour Tikhonov et TV-h avec $\mu = 1$, les reconstructions sont meilleures que celles de l'état-de-l'art. Dans le cas de $\mu = 0,1$ c'est le cas pour $\tau^{\mathrm{disk}} < 15\%$.

Enfin pour θ, on voit que l'EQMn de la reconstruction avec TV-h est toujours plus faible que celle des méthodes de l'état-de-l'art, aussi bien avec $(\lambda_\ell)^{\mathrm{SURE}}_{\ell\in\{1,\dots L\}}$ qu'avec $(\lambda_\ell)^{\mathrm{EQM\text{-}P}}_{\ell\in\{1,\dots L\}}$. Pour $\mu = 0,1$, quand $\tau^{\mathrm{disk}} \in [7\%, 15\%]$ l'erreur faite sur $\widehat{\theta}$ pour $(\lambda_\ell)^{\mathrm{EQM\text{-}P}}_{\ell\in\{1,\dots L\}}$ est de $0,10°$ plus faible que l'erreur angulaire avec $(\lambda_\ell)^{\mathrm{SURE}}_{\ell\in\{1,\dots L\}}$. Elle reste plus faible pour les autre cas également mais pour une erreur inférieure à $0,10°$. Pour Tikhonov, si pour $\tau^{\mathrm{disk}} < 25\%$ la MLnS avec les hyperparamètres $(\lambda_\ell)^{\mathrm{SURE}}_{\ell\in\{1,\dots L\}}$ est meilleure que les méthodes séparables ce n'est pas le cas pour des taux de polarisation du disque plus élevés. Dans le cas de $(\lambda_\ell)^{\mathrm{EQM\text{-}P}}_{\ell\in\{1,\dots L\}}$, la méthode est moins bonne que les méthodes séparables pour $\tau^{\mathrm{disk}} \geq 10\%$. La méthode avec TV-h pour $\mu = 1$ donne une erreur angulaire plus élevée qu'avec $\mu = 0,1$ mais moins élevée qu'avec Tikhonov. Par ailleurs dans le cas d'une régularisation par Tikhonov ou pour TV-h avec $\mu = 1$ l'EQMn est plus faible pour $(\lambda_\ell)^{\mathrm{SURE}}_{\ell\in\{1,\dots L\}}$ que pour $(\lambda_\ell)^{\mathrm{EQMN\text{-}P}}_{\ell\in\{1,\dots L\}}$, sauf pour TV-h lorsque $\tau^{\mathrm{disk}} > 25\%$.

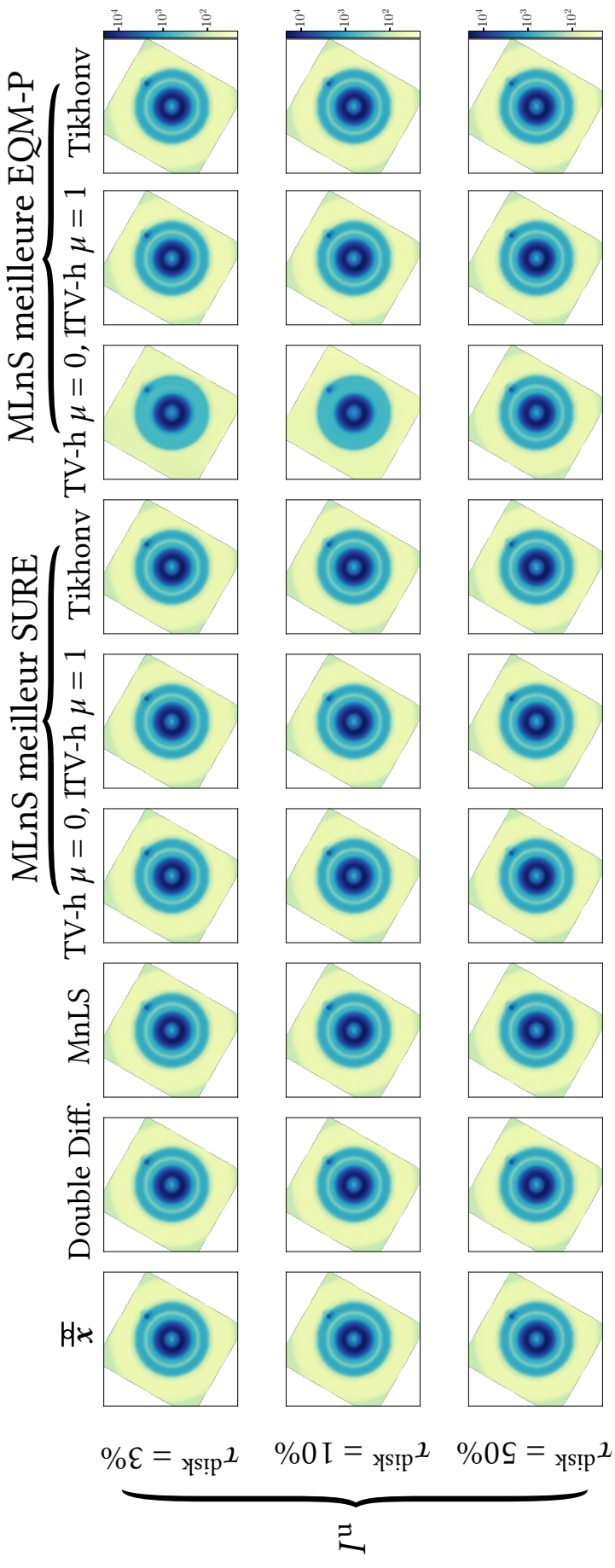


FIGURE 3.2 – Cartes reconstruites avec les différentes méthodes pour $\tau^{\text{disk}} \in \{3\%, 10\%, 50\%\}$.

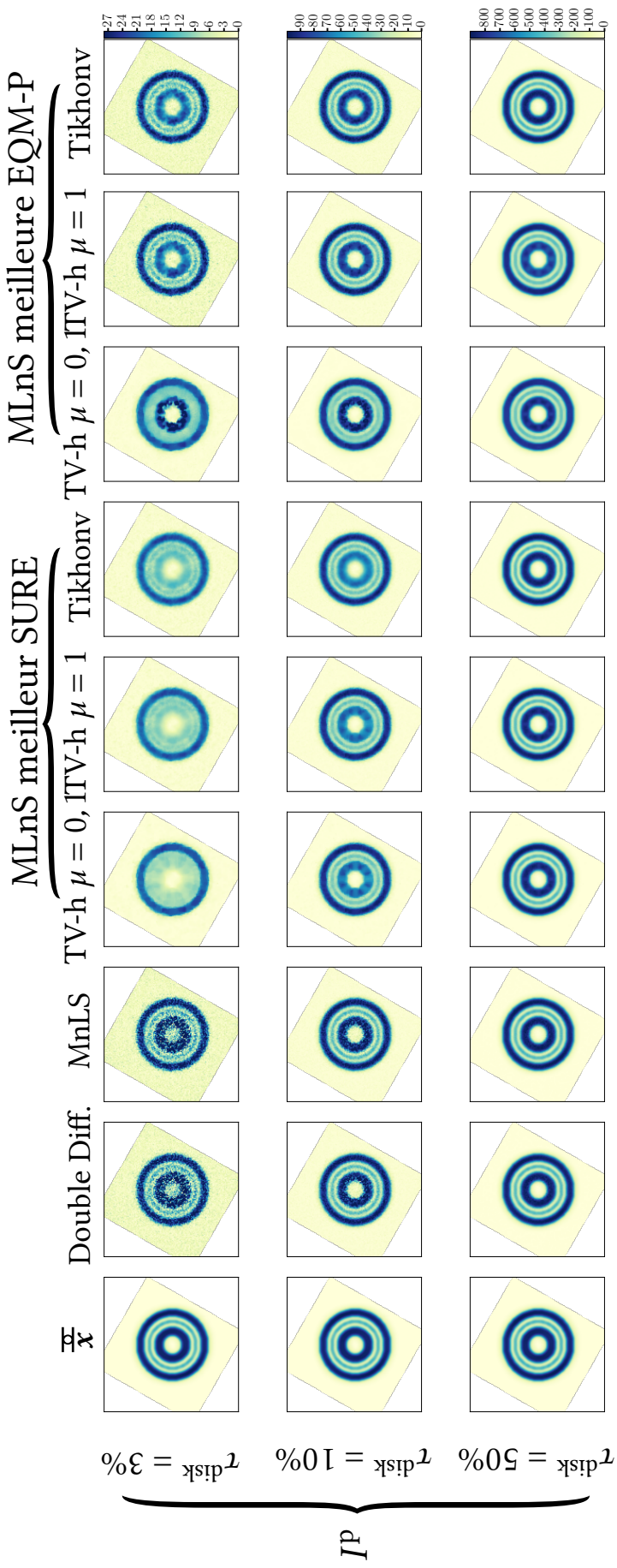


FIGURE 3.2 – Cartes reconstruites avec les différentes méthodes pour $\tau^{\text{disk}} \in \{3\%, 10\%, 50\%\}$.

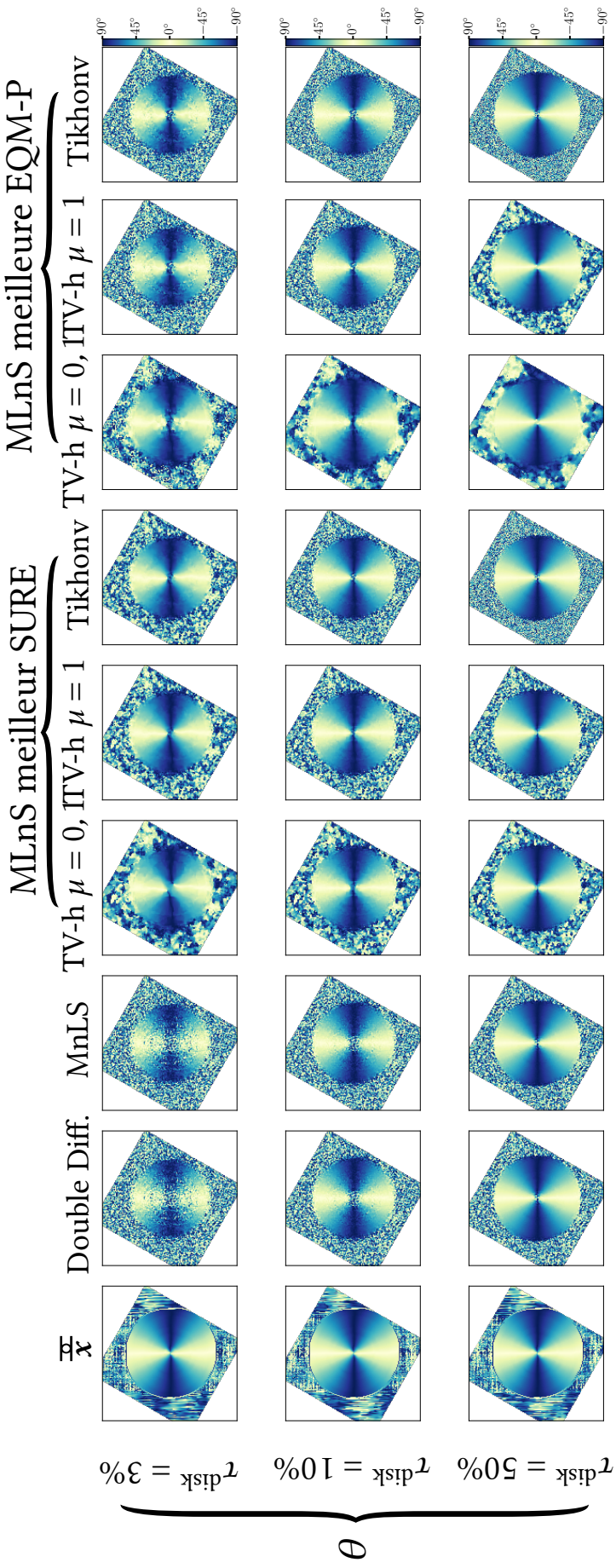
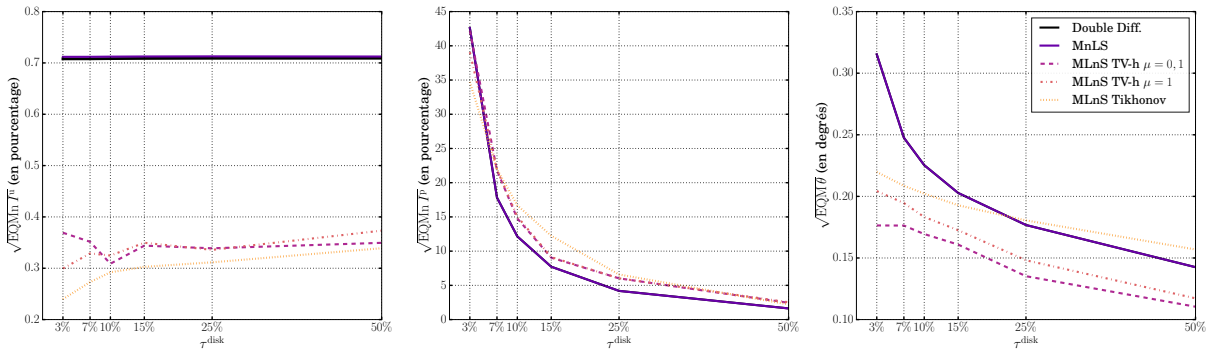


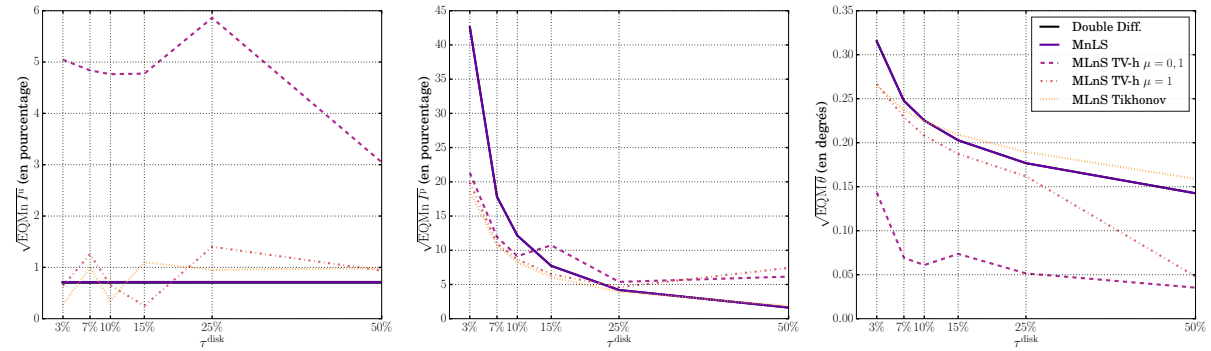
FIGURE 3.2 – Cartes reconstruites avec les différentes méthodes pour $\tau^{\text{disk}} \in \{3\%, 10\%, 50\%\}$.

D'après ces résultats, on voit donc que le choix des hyperparamètres $(\lambda_\ell)^{\text{SURE}}_{\ell\in\{1,\dots L\}}$ ou des hyperparamètres $(\lambda_\ell)^{\text{EQMn-P}}_{\ell\in\{1,\dots L\}}$ et des régularisations ont des avantages et inconvénients selon les paramètres estimés. Le problème est que lors de l'application sur données astrophysiques, la vérité terrain n'est pas connue et donc l'accès à l'EQMn-P n'est pas possible. On peut donc supposer que dans le cas de faible taux de polarisation du disque et donc de faible SNR, le réglage par SURE n'est pas pertinent. Il faut cependant remarquer que l'estimation est faite sur les paramètres I, Q et U et non sur les paramètres I^{u}, I^{p} et θ, et que le paramètre I^{p} est mélangé à I^{u} dans I. Le paramètre I^{u} est donc à la fois régularisé dans I et à la fois dans Q et U. Par ailleurs, on voit que pour un choix de paramètres minimisant l'EQMn des paramètres, lorsque le taux de polarisation du disque est faible, et donc que I^{p} est faible par rapport à I^{u}, l'erreur d'estimation est faible pour I^{u} et élevée pour I^{p}. Ensuite, lorsque le taux de polarisation augmente l'erreur sur I^{u} augmente et l'erreur sur I^{p} décroit. Pour un taux de polarisation du disque de 50%, on voit que l'erreur sur I^{u} augmente. On peut en déduire que selon le taux de polarisation et donc de la brillance de I^{u} par rapport à I^{p}, la régularisation sur I va favoriser soit I^{u} soit I^{p}.

À l'inverse, lorsque le choix des hyperparamètres est celui minimisant le critère SURE, lorsque le taux de polarisation augmente, l'erreur sur I^{u} n'augmente que très peu, tandis que l'erreur sur I^{p} décroit.



(a) EQMn entre les $\widehat{\boldsymbol{x}}$ estimés avec les différentes méthodes et $\overline{\overline{\boldsymbol{x}}}$ avec $(\lambda_\ell)^{\text{SURE}}_{\ell\in\{1,\dots L\}}$.



(b) EQMn entre les $\widehat{\boldsymbol{x}}$ estimés avec les différentes méthodes et $\overline{\overline{\boldsymbol{x}}}$ avec $(\lambda_\ell)^{\text{EQMn-P}}_{\ell\in\{1,\dots L\}}$.

FIGURE 3.3 – Erreurs Quadratiques Moyennes normalisées des différents paramètres pour les différentes méthodes. La figure a) représente l'EQMn $\text{EQMn}(\widehat{\boldsymbol{x}}, \overline{\overline{\boldsymbol{x}}})$ pour les différentes méthodes. La figure b) représente le rapport entre ces EQMn et la meilleure valeur de toutes les EQMn pour chaque valeur de τ^{disk}.

3.1.4.2 Résultats sur données astrophysiques

Afin de juger les performances de la méthode sur données astrophysiques, nous appliquons maintenant la MLnS aux données *calibrées* de la cible RXJ 1615 [Avenhaus et al., 2018] observée en bande H, avec la régularisation de Tikhonov et la régularisation à préservation de bords TV-h. Nous comparons les résultats pour les hyperparamètres réglés avec SURE d'une part et d'autre part à la main. Pour la régularisation TV-h avec SURE, nous réglons cette fois-ci le paramètre μ conjointement aux paramètres λ. Sur la figure 3.4, nous comparons visuellement les résultats avec les résultats obtenus par la méthode de la Double Différence et la MLnS sur les données *pré-traitées*.

Pour l'intensité non-polarisée I^{u}, il n'est pas évident de comparer la qualité des résultats de manière visuelle. En revanche pour I^{p} et θ, on voit qu'un réglage des hyperparamètres par SURE sur-régularise complètement les estimations. Cela reste cohérent avec les résultats obtenus sur données simulées à faible rapport signal à bruit. En effet, comme il a été vu dans le chapitre précédent, le disque RXJ 1615 a un SNR très faible, ce qui peut expliquer cette sur-régularisation par le critère SURE.

Un réglage manuel des hyperparamètres permet une solution moins régularisée et plus proches des résultats séparables. Ici j'ai choisi des hyperparamètres suffisamment élevés pour améliorer le conditionnement du problème, et donc éviter l'augmentation du bruit dû aux opérateur $(\mathbf{T}_{j,k})_{\{j\in\{1,2\},k\in\{1,\dots K\}\}}$, mais pas trop haut, afin d'éviter un lissage trop intense des structures.

À noter que la recherche des hyperparamètres par recherche du minimum du critère SURE à l'aide d'une méthode de type Powell [Powell, 2006], pour un jeu de donnée de taille $300 \times 600 \times 88$, prend environ 10 heures, là où une recherche manuelle prend une vingtaine de minutes.

On remarque qu'avec les méthodes séparables, il reste quelques résidus aux endroits où les pixels morts ont été interpolés. Dans le cas des méthodes séparables, les pixels morts ont été lissés. Le disque autour du coronographe et la reconstruction de l'angle sont également plus lisses. Cependant au niveau de l'intensité polarisée I^{u}, les anneaux externes ne semblent pas plus régularisés que dans le cas des méthodes séparables. Dans les deux cas, les résultats obtenus avec Tikhonov et TV-h semblent équivalents.

Il serait souhaitable de pouvoir juger de l'erreur faite sur les paramètres de manière quantitative. Cependant, comme nous l'avons dit précédemment, il n'est pas possible dans le cas régularisé d'utiliser la borne de FDCR pour avoir une borne inférieure à l'EQM des paramètres, comme cela avait été fait dans le chapitre précédent.

Plusieurs pistes pourraient cependant être explorées pour pouvoir afficher des cartes d'erreurs sur les pixels, de la même façon que dans le chapitre précédent. La première piste serait d'utiliser les méthodes permettant d'estimer les bornes de FDCR en présence de biais [Fessler and Hero, 1993]. Une autre piste serait de se tourner la version de SURE généralisée dans le domaine des paramètres [Eldar, 2008], cependant cette version de SURE nécessite d'avoir un modèle $(f_k)_{k\in\{1,\dots K\}}$ qui soit inversible, ce qui n'est pas le cas ici. Ces pistes n'ont cependant pas encore eu le temps d'être explorées et font partie des perspectives pour la suite de cette thèse.

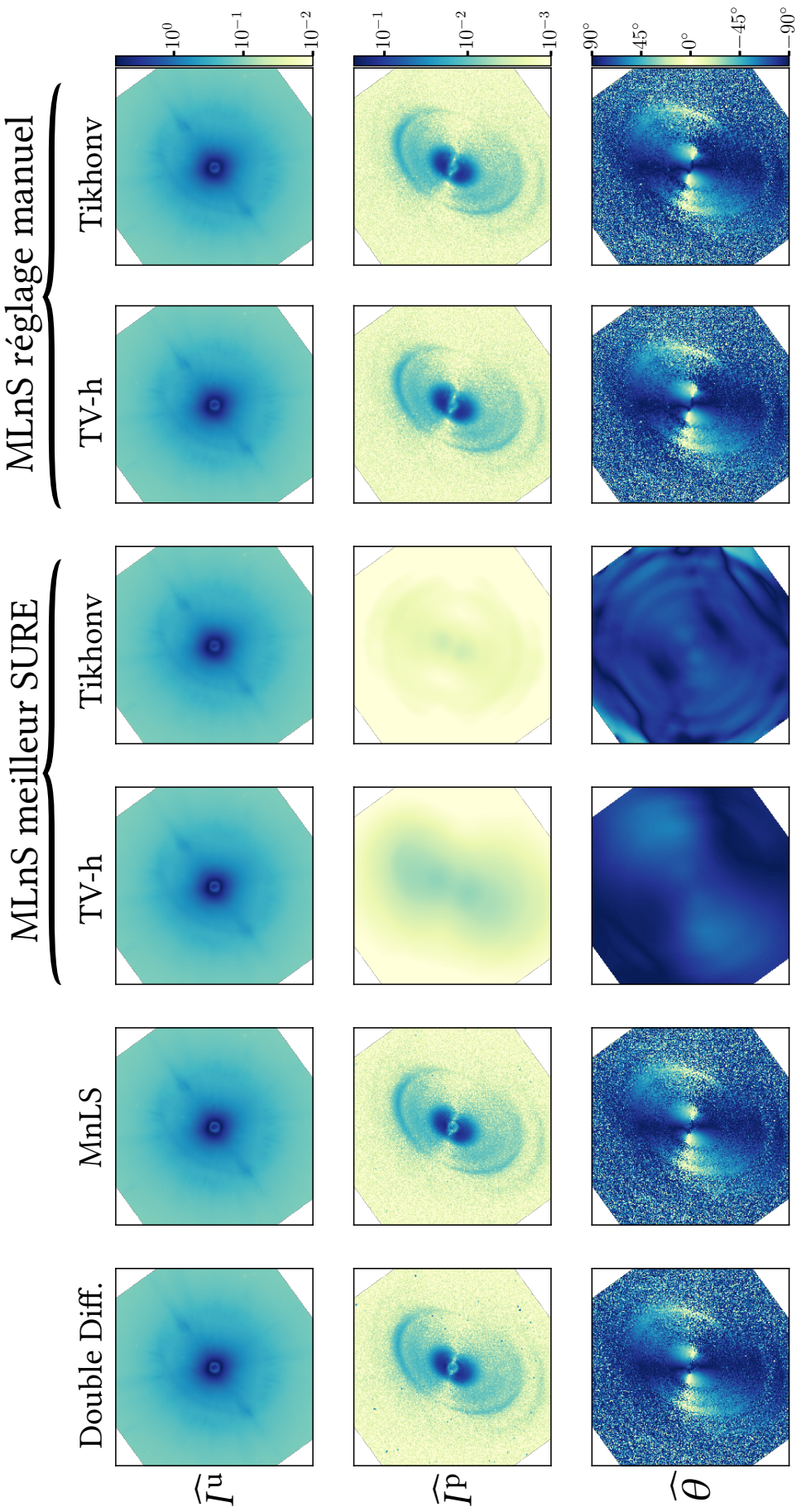


FIGURE 3.4 – Cartes reconstruites avec les différentes méthodes de la cible RXJ 1615 en H.

3.1.4.3 Synthèse des résultats

Nous avons pu voir dans cette section que la prise en compte des transformations dans le modèle et l'estimation des paramètres du modèle par un critère régularisé peut avoir un apport positif, surtout dans les cas de faible rapport signal à bruit. Cependant, il est nécessaire pour cela d'avoir un bon réglage des hyperparamètres. Or dans le cas de faible SNR, le réglage par minimisation du critère SURE sur-régularise la reconstruction du paramètre I^{p}. Une hypothèse est que I^{p} étant à la fois présent dans I d'une part et Q et U d'autre part, une régularisation indépendante sur ces paramètres n'est peut-être pas pertinent. C'est pourquoi dans la section suivante nous proposons un changement de variable permettant de régulariser I^{u} et I^{p} de manière indépendante.

3.2 Modèle direct non-linéaire non-séparable

3.2.1 Le modèle direct non-séparable non-linéaire

Dans la section précédente nous avons vu une méthode non-séparable régularisée permettant de reconstruire le signal d'intérêt à partir du modèle linéaire des données.

Le choix du modèle linéaire venait d'un besoin de plus de simplicité pour la résolution dans le cas de contraintes non différentiable, et du besoin d'éviter les problèmes de multiplicité des solutions pour θ. Cependant, la contrainte de positivité sur I^{u} n'es pas assurée et le paramètre I^{p} est à la fois régularisé dans I avec I^{u}, pour un hyperparamètre donné, et dans Q et U, pour une autre valeur d'hyperparamètre. Séparer I^{u} des autres paramètres semble donc plus judicieux.

C'est pourquoi, nous proposons le modèle des données en les paramètres $\boldsymbol{x} = (I^{\mathrm{u}}, \mathrm{Q}, \mathrm{U})^\top$:

$$\forall k \in \{1, \dots, K\} \quad \boldsymbol{d}_k^{\mathrm{nS}} = \mathcal{B}_k\left(f_k^{\mathrm{nS}}\left(q(\overline{\overline{\boldsymbol{x}}})\right)\right) \tag{3.11}$$

où les $(f_k^{\mathrm{nS}})_{k\in\{1,\dots,K\}}$ sont définis par l'équation (3.2) et $q : (\mathbb{R}^N)^L \to (\mathbb{R}^N)^L$ correspond au changement de variable séparable :

$$q(\boldsymbol{x}) = \begin{pmatrix} \boldsymbol{x}_1 + \sqrt{\boldsymbol{x}_2^2 + \boldsymbol{x}_3^2} \\ \boldsymbol{x}_2 \\ \boldsymbol{x}_3 \end{pmatrix}, \tag{3.12}$$

tel que $q(\boldsymbol{x})_n = q(Vx_n)$, qui implique que $q((I^{\mathrm{u}}, \mathrm{Q}, \mathrm{U})^\top) = (\mathrm{I}, \mathrm{Q}, \mathrm{U})^\top$. En effet, on a $\forall k \in \{1, \dots, K\}$ et $j = 1, 2$:

$$v_{j,k,1}\mathrm{I} + v_{j,k,2}\mathrm{Q} + v_{j,k,3}\mathrm{U} \tag{3.13}$$

$$= v_{j,k,1}\left(I^{\mathrm{u}} + \sqrt{\mathrm{Q}^2 + \mathrm{U}^2}\right) + v_{j,k,2}\mathrm{Q} + v_{j,k,3}\mathrm{U} \tag{3.14}$$

$$= \sum_{\ell=1}^{3} v_{j,k,\ell} q((I^{\mathrm{u}}, \mathrm{Q}, \mathrm{U})^\top)_\ell. \tag{3.15}$$

Ce changement de variable permet donc l'utilisation d'une contrainte de borne sur I^{u} et la régularisation séparée sur I^{u}, Q et U.

3.2.2 Résolution par approche inverse différentiable

On se propose maintenant de reconstruire les paramètres I^{u}, I^{p} et θ à partir du modèle des données 3.11 paramétré en $\boldsymbol{x} = (I^{\mathrm{u}}, \mathrm{Q}, \mathrm{U})^\top$. Le modèle direct n'étant plus linéaire, introduire le changement de variables dans l'attache aux données (3.6) casse la convexité. La convergence des différents algorithmes vers un minimum global n'est donc plus assurée. C'est pour cela que pour plus de simplicité, nous restons dans le cas différentiable. Le problème non linéaire et non convexe à résoudre s'écrit alors :

$$\widehat{\boldsymbol{x}} \in \underset{\boldsymbol{x}}{\operatorname{Argmin}} \left\{ \Phi\Big(q(\boldsymbol{x})\Big) + \mathcal{R}\Big(q(\boldsymbol{x})\Big) \right\}, \tag{3.16}$$

où Φ et $\mathcal{R}$ sont différentiable et données respectivement par les équations (3.6) et (3.7) ou (3.8).

On remarque cependant, qu'une régularisation unique I^{p} semblerait plus justifiée qu'une régularisation faite sur Q et U de manière indépendante. C'est pourquoi nous proposons de résoudre le problème :

$$\widehat{\boldsymbol{x}} \in \underset{\boldsymbol{x}}{\operatorname{Argmin}} \left\{ \Phi\Big(q(\boldsymbol{x})\Big) + \mathring{\mathcal{R}}\Big(r(\boldsymbol{x})\Big) \right\}, \tag{3.17}$$

où $r : (\mathbb{R}^N)^L \to (\mathbb{R}^N)^2$ correspond au changement de variable séparable :

$$r(\boldsymbol{x}) = \begin{pmatrix} \boldsymbol{x}_1 \\ \sqrt{\boldsymbol{x}_2^2 + \boldsymbol{x}_3^2} \end{pmatrix}, \tag{3.18}$$

tel que $r(\boldsymbol{x})_n = r(V x_n)$, qui implique que $r((I^{\mathrm{u}}, \mathrm{Q}, \mathrm{U})^\top) = (I^{\mathrm{u}}, I^{\mathrm{p}})^\top$ et $\mathring{g} : (\mathbb{R}^N)^{\mathring{L}} \to \mathbb{R}$ s'écrit pour $\lambda_\ell \geq 0$:

$$\mathring{\mathcal{R}}(\boldsymbol{x}) = \sum_{\ell=1}^{\mathring{L}} \lambda_\ell \|\mathbf{D}\boldsymbol{x}_\ell\|_2^2, \tag{3.19}$$

dans le cas de la régularisation quadratique de Tikhonov et

$$\mathring{\mathcal{R}}(\boldsymbol{x}) = \sum_{\ell=1}^{\mathring{L}} \lambda_\ell \left(\sqrt{\|\mathbf{D}\boldsymbol{x}_\ell\|_2^2 + \mu_\ell^2} - \mu_\ell \right), \tag{3.20}$$

dans le cas de la régularisation à préservation des bords.

Comme précédemment, pour résoudre le problème défini par l'équation (3.17), on utilise la méthode quasi-newton.

3.2.3 Application sur données simulées et données astrophysiques

Dans cette section, nous étudions les performances et la qualité des résultats obtenus par la Méthode non-Linéaire non-Séparable (MnLnS) en comparant les résultats obtenus, sur données simulées et sur données astrophysiques *calibrées*, aux résultats de la Méthode Linéaire non-Séparable (MLnS) et aux résultats obtenus par la Méthode Linéaire Séparable (MLS) et le Double Ratio sur données *pré-traitées*.

Dans le cas de la MLnS et de la MnLnS, comme dans la section précédente, les reconstructions sont faites pour un choix d'hyperparamètres minimisant d'une part la somme des

EQMn dans le domaine des paramètres, et d'autre part le critère SURE. Comme précédemment on note d'une part $\widehat{\boldsymbol{x}}^{\text{EQMn-P}}$ le paramètre d'intérêt estimé par l'une ou l'autre des deux méthodes pour un choix d'hyperparamètres minimisant l'EQMn dans le domaine des paramètres. D'autre part, on note $\widehat{\boldsymbol{x}}^{\text{SURE}}$ le paramètre d'intérêt estimé par l'une des deux méthodes, pour un choix d'hyperparamètres minimisant le critère SURE.

3.2.3.1 Résultats sur données simulées

Comme précédemment, nous avons appliqué la méthode sur données simulées pour différents taux de polarisation du disque $\tau^{\text{disk}} \in \{3\%; 7\%; 10\%; 15\%; 25\%; 50\%\}$. Dans le cas de la régularisation par TV-h, pour plus de simplicité lors de la recherche des hyperparamètres, nous fixons le paramètre μ et comparons les résultats pour $\mu = 0,1$ et $\mu = 1$.

Sur la figure 3.5 sont affichées les reconstructions obtenues pour $\tau^{\text{disk}} \in \{3\%; 10\%; 50\%\}$, par la méthode de la Double Différence et de la Méthode non-Linéaire Séparable (MnLS) sur données *pré-traitées* et par la Méthode Linéaire non-Séparable (MLnS) et la Méthode non-Linéaire non-Séparable sur données *calibrées*, minimisant d'une part l'EQMn dans le domaine des paramètres, notée EQMn-P, et d'autre par le critère SURE. Les reconstructions faites par la MLnS et MnLnS sont comparées pour des régularisations par Tikhonov et TV-h avec $\mu = 0,1$ et $\mu = 1$. Visuellement, pour $\tau^{\text{disk}} \geq 10\%$, la MnLnS semble donner des résultats similaires selon le choix d'hyperparamètres $(\lambda_\ell)^{\text{EQMn-P}}_{\ell\in\{1,\dots L\}}$ ou $(\lambda_\ell)^{\text{SURE}}_{\ell\in\{1,\dots L\}}$, ce qui est positif. En revanche, dans le cas où $\tau^{\text{disk}} = 3\%$, si pour I^{u} il est difficile de dire si la reconstruction par la MnLnS apporte quelque chose, pour I^{p} on voit que d'une par la reconstruction avec la MnLnS semble plus régularisée qu'avec la MLnS. Comme précédemment, la reconstruction avec $(\lambda_\ell)^{\text{SURE}}_{\ell\in\{1,\dots L\}}$ est également plus régularisée que la reconstruction avec $(\lambda_\ell)^{\text{EQMn-P}}_{\ell\in\{1,\dots L\}}$, sauf dans le cas où $\mu = 0,1$. Pour $\tau^{\text{disk}} = 10\%$, les intensités retrouvées par la MnLnS pour l'anneau du centre et l'anneau fin, semblent les plus proches de la vérité terrain que les méthodes séparables et que la MLnS réglée par SURE. Pour $\tau^{\text{disk}} = 50\$$, il est difficile de dire quelle méthode est la plus efficace. Pour l'estimation de θ en revanche, on voit que les cartes ne sont pas régularisées, et même plus bruitées que les cartes obtenues avec les méthodes de l'état-de-l'art. Notons que l'angle n'est pas pris en compte dans les régularisations, un tel résultat est donc logique.

Sur la figure 3.6 sont tracés les racines des EQMn des paramètres, en pourcentage pour les intensités et en degré pour l'angle. Dans la figure 3.3a sont tracées les EQMn dans le cas où les hyperparamètres minimisent le critère SURE. Dans la figure 3.3b sont tracés les EQMn dans le cas où les hyperparamètres minimisent l'erreur quadratique moyenne des paramètres.

On voit que pour les paramètres I^{p} et θ, la reconstruction par la MnLnS a une erreur très similaire pour tous les choix des hyperparamètres $(\lambda_\ell)^{\text{EQMn-P}}_{\ell\in\{1,\dots L\}}$ ou $(\lambda_\ell)^{\text{SURE}}_{\ell\in\{1,\dots L\}}$ et des régularisations. On voit de plus que pour I^{p}, pour un taux de polarisation du disque inférieur à 25%, toutes les reconstructions par la MnLnS ont une EQMn plus basse que les reconstructions par les méthodes séparables et que la MLnS où les hyperparamètres minimisent le critère SURE. Cependant, pour $\tau^{\text{disk}} < 10\%$, l'erreur obtenue avec la MLnS, pour un réglage des hyperparamètres minimisant l'EQM-P, est plus basse que la MnLnS pour tous les choix de régularisation, équivalente à celles des méthodes séparables et plus élevée que les erreurs de la MLnS, peu importe le choix des hyperparamètres.

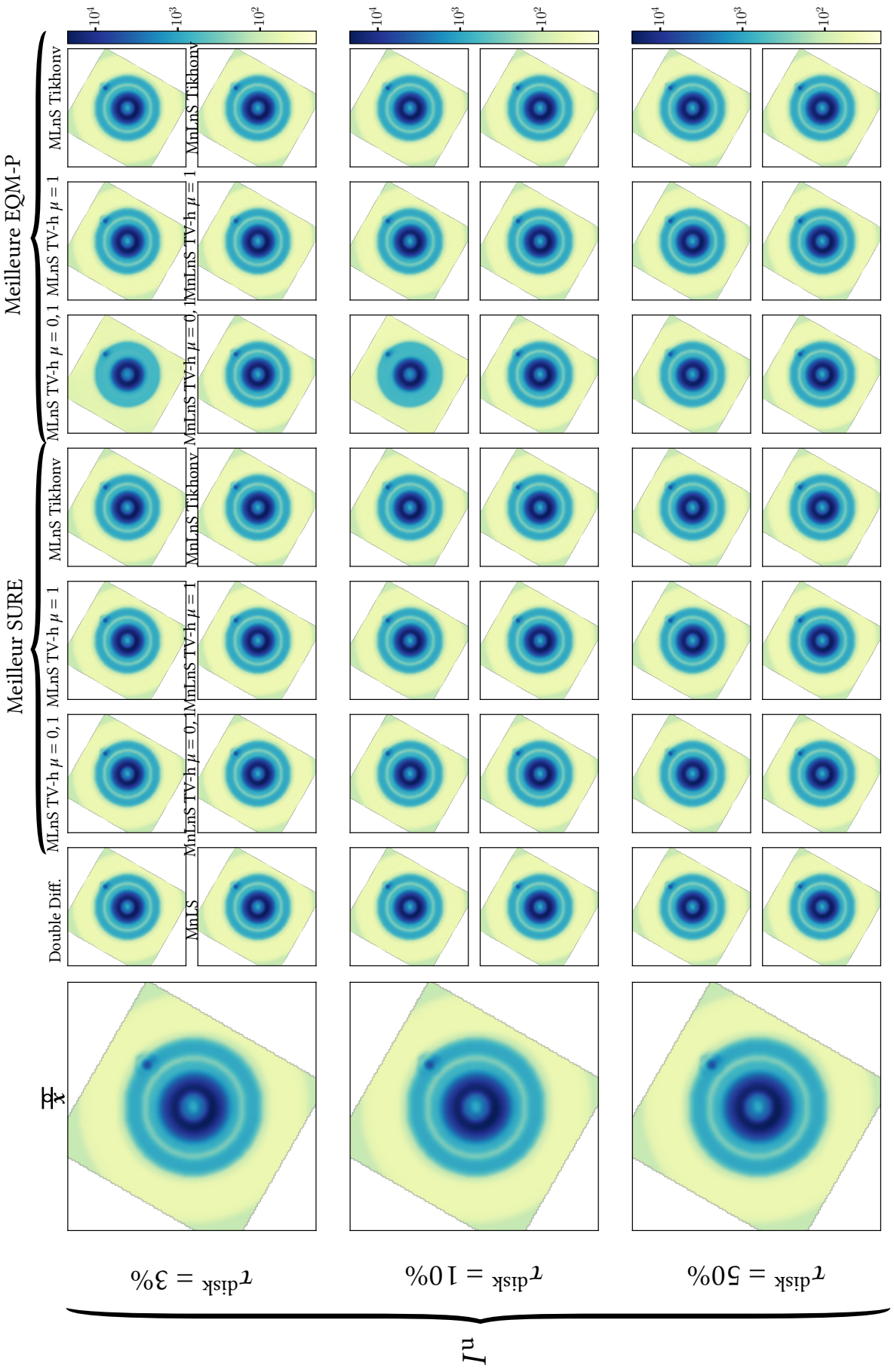


FIGURE 3.5 – Cartes reconstruites avec les différentes méthodes pour $\tau^{\text{disk}} \in \{3\%, 10\%, 50\%\}$.

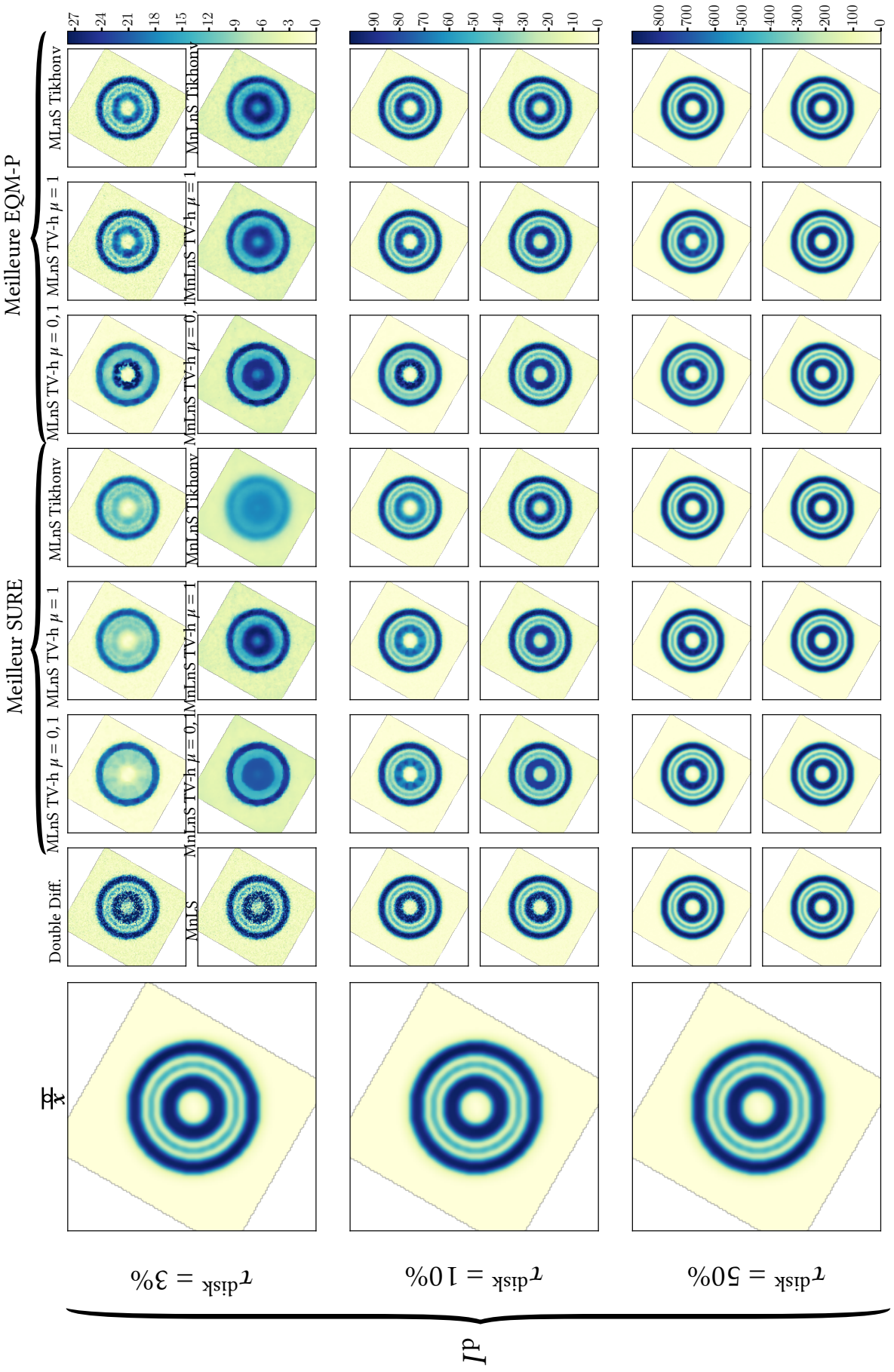


FIGURE 3.5 – Cartes reconstruites avec les différentes méthodes pour $\tau^{\text{disk}} \in \{3\%, 10\%, 50\%\}$.

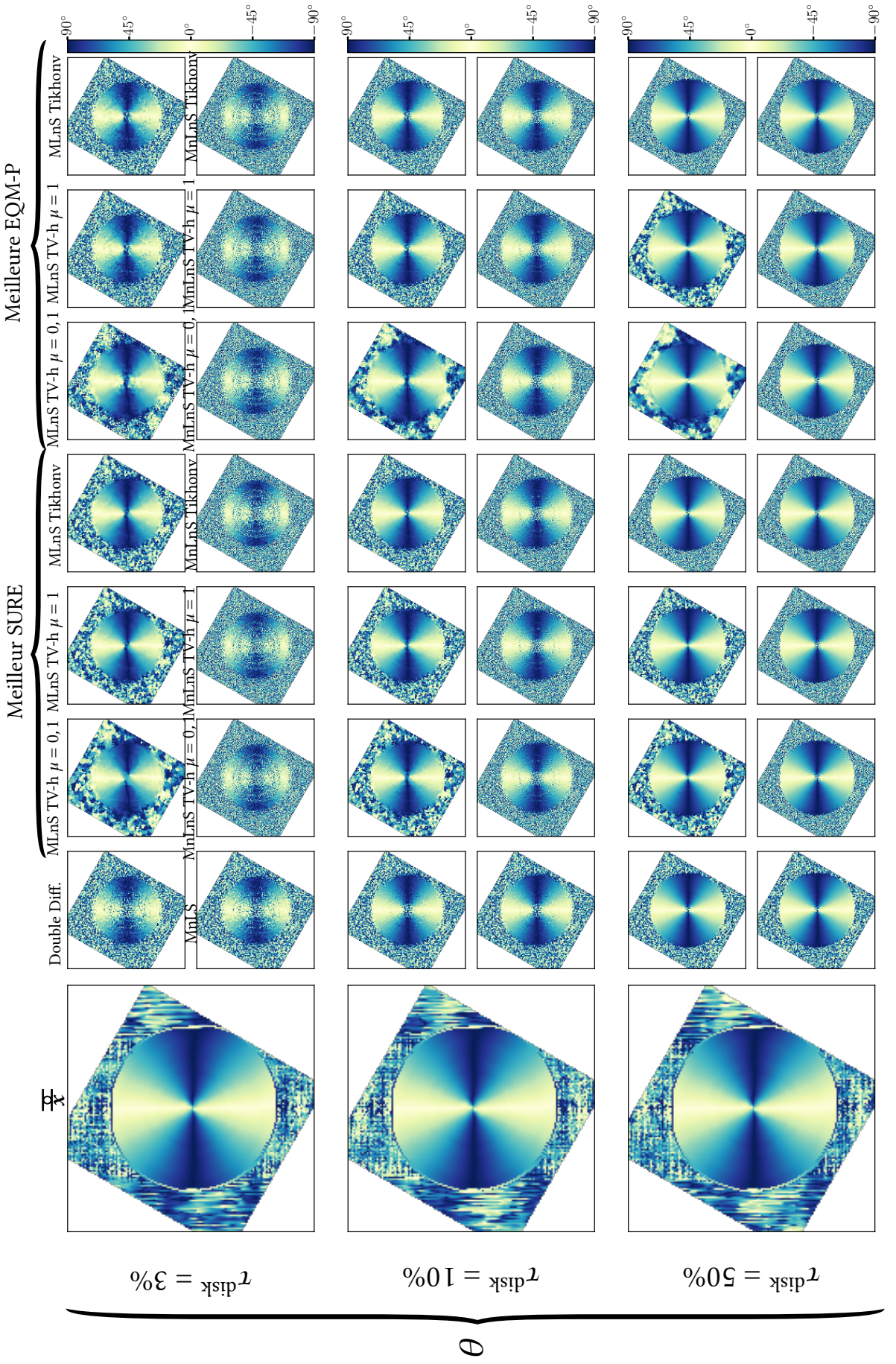
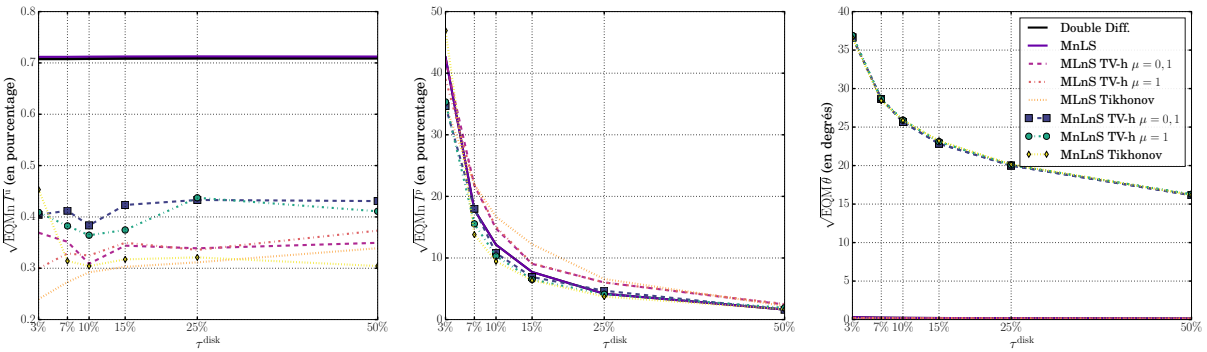


FIGURE 3.5 – Cartes reconstruites avec les différentes méthodes pour $\tau^{\text{disk}} \in \{3\%, 10\%, 50\%\}$.

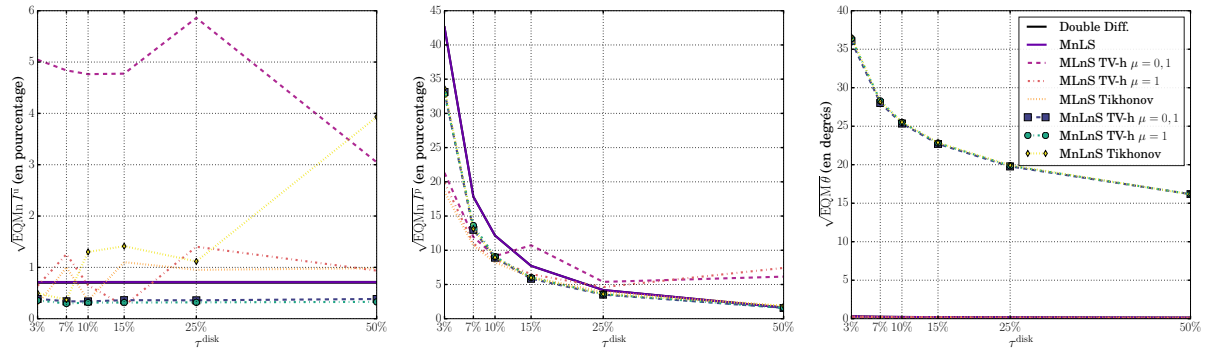
Pour θ, on voit que la MnLnS a une EQMn très grande, qui décroit avec le taux de polarisation. Ce qui est cohérent avec ce que l'on observe sur la figure 3.5. Enfin, pour le paramètre I^{u}, les résultats diffèrent selon le choix de régularisation. Lors de l'estimation par la MnLnS avec $(\lambda_\ell)^{\mathrm{SURE}}_{\ell\in\{1,\dots L\}}$, la reconstruction avec Tikhonov a une erreur inférieure aux reconstructions obtenues par TV-h pour tous les choix d'hperparamètres, sauf pour $\tau^{\mathrm{disk}} = 3\%$. À l'inverse, les reconstructions avec TV-h ont une erreur plus élevée que la MLnS mais inférieure à celle des méthodes de l'état-de-l'art. Pour $(\lambda_\ell)^{\mathrm{EQMn\text{-}P}}_{\ell\in\{1,\dots L\}}$, la reconstruction par la MnLS avec une régularisation par TV-h donne l'erreur la plus basse pour tous les taux de polarisation du disque. Dans le cas d'une régularisation par Tikhonov, pour $\tau^{\mathrm{disk}} > 7\%$, l'erreur est supérieure à celle faite par les méthodes de l'état-de-l'art.

Le choix de l'hyperparamètre μ semble avoir une plus grande influence dans la reconstruction utilisant les hyperparamètres $(\lambda_\ell)^{\mathrm{SURE}}_{\ell\in\{1,\dots L\}}$ qu'avec $(\lambda_\ell)^{\mathrm{EQM\text{-}P}}_{\ell\in\{1,\dots L\}}$. En effet, si lorsque l'EQMn pour $\mu = 1$ et pour $\mu = 0,1$ diffèrent, c'est le choix $\mu = 1$ qui donne l'EQMn la plus basse.

D'après ces résultats, la Méthode non-Linéaire non-Séparable (MnLnS) semble avoir un avantage par rapport à la Méthode Linéaire non-Séparable (MLnS) dans le fait que, pour une telle méthode, les hyperparamètres $(\lambda_\ell)^{\mathrm{EQMn\text{-}P}}_{\ell\in\{1,\dots L\}}$ et $(\lambda_\ell)^{\mathrm{SURE}}_{\ell\in\{1,\dots L\}}$ donnent des résultats relativement équivalents. Dans le cadre astrophysique où la vérité terrain ne serait pas connue et où le choix des hyperparamètres doit nécessairement passer par SURE, la MnLnS donnerait



(a) EQMn entre les $\widehat{\boldsymbol{x}}$ estimés avec les différentes méthodes et $\overline{\overline{\boldsymbol{x}}}$ avec $(\lambda_\ell)^{\mathrm{SURE}}_{\ell\in\{1,\dots L\}}$.



(b) EQMn entre les $\widehat{\boldsymbol{x}}$ estimés avec les différentes méthodes et $\overline{\overline{\boldsymbol{x}}}$ avec $(\lambda_\ell)^{\mathrm{EQMn\text{-}P}}_{\ell\in\{1,\dots L\}}$.

FIGURE 3.6 – Erreurs Quadratiques Moyennes normalisées des différents paramètres pour les différentes méthodes. La figure a) représente l'EQMn $\mathrm{EQMn}(\widehat{\boldsymbol{x}}, \overline{\overline{\boldsymbol{x}}})$ pour les différentes méthodes. La figure b) représente le rapport entre ces EQMn et la meilleure valeur de toutes les EQMn pour chaque valeur de τ^{disk}.

donc une meilleure reconstruction de l'intensité polarisée I^{p} par rapport aux méthodes séparables et à la MLnS, avec TV-h dans le cas où $\tau^{\mathrm{disk}} = 3\%$, avec Tikhonov sinon. En revanche pour la reconstruction de l'angle θ, c'est la MLnS qui donnerait la meilleure reconstruction de l'angle avec une reconstruction par TV-h pour $\mu = 0,1$ (cf. figure 3.2). Enfin, pour I^{u}, c'est la reconstruction avec Tikhonov qui donne la meilleure reconstruction.

3.2.3.2 Résultats sur données astrophysiques

Nous appliquons maintenant la MnLnS avec régularisation par TV-h et Tikhonov, aux données *calibrées* de la cible RXJ 1615 [Avenhaus et al., 2018] observée en bande H. Sur la figure 3.2, nous comparons visuellement les résultats obtenus par la MnLnS et la MLnS, régularisées avec TV-h et Tikhonov, avec la méthode de la Double Différence et la MLnS sur les données *pré-traitées*. Dans le cas de la MLnS et de la MnLnS nous réglons les hyperparamètres, d'une part par minimisation du critère SURE et d'autre part manuellement. Dans le cas de TV-h, μ est réglé conjointement aux $(\lambda_\ell)_{\ell\in\{1,\dots L\}}$ lors de la minimisation du critère SURE.

Pour I^{u}, il est difficile de comparer les différentes reconstructions car les résultats sont très semblables, mais ce paramètre n'est pas le plus important lors de l'étude astrophysique. En revanche, pour I^{p} on voit que parmi les reconstructions pour un choix des hyperparamètres minimisant le critère SURE, la reconstruction par la MnLnS avec une régularisation par Tikhonov est la meilleure, la régularisation avec TV-h étant sur-régularisée. La solution avec Tikhonov reste cependant trop régularisée car on perd les structures de faibles dynamiques. Pour le choix des hyperparamètres réglés manuellement, la reconstruction avec la MnLnS semble être la meilleure. En effet, du fait de la polarisation instrumentale, sur les reconstructions le disque semble coupé en deux, ce qui n'est en réalité pas le cas (voir figure 1 [Avenhaus et al., 2018] pour l'image du disque sans polarisation instrumentale). Or sur la reconstruction avec la MnLnS, la séparation semble moins marquée. Dans le cas de la reconstruction de l'angle θ avec la MnLnS, en dehors de la reconstruction avec Tv-h où il ne semble rester que du bruit, les reconstructions avec Tikhonov réglé par SURE et TV-h et Tikhonov réglé manuellement donnent des résultats identiques. Comme pour les données simulées, l'angle est même plus bruité que les reconstructions obtenues avec les méthodes de l'état-de-l'art.

3.2.3.3 Synthèse des résultats

Nous avons pu voir dans cette section que le changement de variable et l'estimation des paramètres du modèle par un critère régularisé indépendamment sur I^{u} et I^{p} a un impact très positif pour la reconstruction de I^{p}. Cependant, le réglage des hyperparamètres reste compliqué. Dans le cas de faible SNR, le réglage par minimisation du critère SURE sur-régularise la reconstruction du paramètre I^{p}, bien qu'en moindre mesure avec la régularisation de Tikhonov par rapport à la MLnS. Cependant, l'angle n'étant pas régularisé la reconstruction est pire qu'avec les méthodes de l'état-de-l'art.

Conclusion du chapitre

Dans ce chapitre nous avons vu deux méthodes non-séparables permettant de prendre en compte dans le modèle les transformations du détecteur (translations, rotations) ainsi que les pixels morts. La première est linéaire, paramétrée en $(\mathrm{I}, \mathrm{Q}, \mathrm{U})$ et régularisée indépendamment

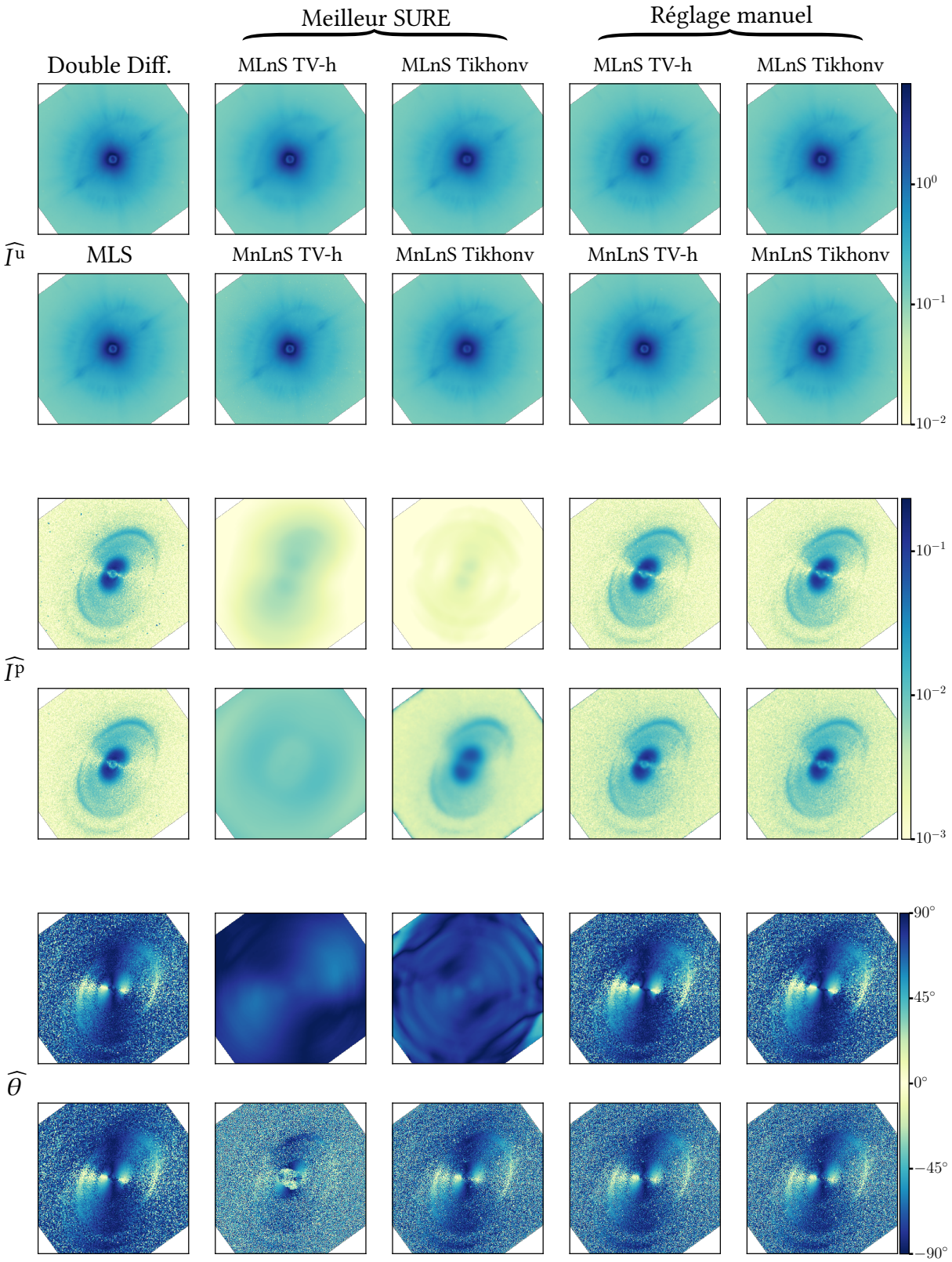


FIGURE 3.7 – Cartes reconstruites avec les différentes méthodes de la cible RXJ 1615 observée en bande H.

sur les trois composantes. La seconde est non-linéaire, paramétrée en $(I^{\mathrm{u}}, \mathrm{Q}, \mathrm{U})$ et régularisée indépendamment sur I^{u} et sur $I^{\mathrm{p}} = \sqrt{\mathrm{Q}^2 + \mathrm{U}^2}$. Nous avons estimé ces paramètres par minimisation d'un critère régularisé d'une part par Tikhonov, d'autre part par la régularisation à préservation de bords TV-h, et dans le cas non-linéaire avec une contrainte de positivité sur I^{u}.

Nous avons montré que le réglage des hyperparamètres par minimisation du critère SURE donnait, dans le cas de faible SNR, des résultats sur-régularisés. Nous avons cependant montré que dans le cas non-linéaire, les reconstructions étaient tout de même moins régularisées que les méthodes linéaires. De plus, dans un tel cas, la régularisation par Tikhonov permet d'obtenir la meilleure reconstruction de l'intensité polarisée I^{p}. Dans le cas du réglage manuel des hyperparamètres, l'intensité polarisée I^{p} semble également mieux reconstruite avec la MnLnS.

En revanche, la reconstruction de l'angle ne bénéficie pas de la MnLnS et son erreur est même empirée par une telle reconstruction. La MLnS avec régularisation par TV-h est alors plus à même d'obtenir une solution proche de la vérité terrain.

Un bon compromis entre ces méthodes serait alors d'utiliser une régularisation structurée, régularisant non pas sur I^{u}, mais conjointement sur Q et U. L'implémentation d'une telle contrainte n'a pas été fait dans cette thèse mais est une des principales perspectives pour l'amélioration de la méthode.

Chapitre 4

Modèle direct non-séparable incluant la convolution

Dans le chapitre précédent, nous avons établi un modèle direct des données *calibrées* non-séparable prenant en compte le prétraitement des données. Ce modèle n'est pas complet, car il ne prend pas en compte le flou instrumental, induit par la diffraction de la lumière dans l'instrument. Ce flou se modélise par la convolution des paramètres d'intérêt, modélisés par les vecteurs $(\boldsymbol{x}_\ell)_{\ell\in\{1,\dots L\}} \in \mathbb{R}^N$, par la réponse impulsionnelle de l'instrument (PSF : *Point Spread Function*), représentée par l'opérateur $\mathbf{A} : \mathbb{R}^N \to \mathbb{R}^N$.

Dans la littérature sur les environnements circumstellaires étudiés en polarimétrie, la déconvolution des paramètres d'intérêt est faite indépendamment de la réduction des données, comme sur la cible RY Lup dans [Langlois et al., 2018]. Or, nous avons vu dans la section précédente que procéder à l'estimation des paramètres d'intérêt sur données *calibrées*, notées $\boldsymbol{d}^{\mathrm{nS}}$, à partir d'un modèle plus précis des données, plutôt que l'estimation sur données *pré-traitées*, notées $\boldsymbol{d}^{\mathrm{S}}$, permettait de réduire l'erreur d'estimation sur les paramètres. Il semble donc pertinent de se demander si la prise en compte de la convolution dans le modèle pourrait à nouveau réduire l'erreur faite sur les paramètres reconstruits. En effet, la déconvolution des données permet de gagner en résolution spatiale, c'est-à-dire d'augmenter la précision des détails dans les images.

Rappelons que le modèle direct des données s'écrit de la forme, pour toute acquisition $k \in \{1,\dots K\}$:

Rappel éq. (3.1)
$$\boldsymbol{d}_k = \mathcal{B}_k(f_k(\overline{\overline{\boldsymbol{x}}})),$$

où $f_k : (\mathbb{R}^N)^L \to \mathbb{R}^M$ modélise l'instrument, les transformations du détecteur pour passer du domaine des cartes reconstruites au domaine des données et maintenant également la convolution par la PSF. Enfin $\mathcal{B}_k : \mathbb{R}^M \to \mathbb{R}^M$ représente le bruit additif dans les données.

Comme dans la section précédente, les opérateurs $(f_k)_{k\in\{1,\dots K\}}$ sont mal conditionnés car la PSF contient peu de valeurs propres non-nulles. La minimisation seule de la vraisemblance sous la forme

Rappel éq. (3.6)
$$\Phi(\boldsymbol{x}) = \sum_{k=1}^{K} \tfrac{1}{2} \|\boldsymbol{d}_k - f_k(\boldsymbol{x})\|^2_{\mathbf{W}_k}.$$

va amplifier le bruit. C'est pourquoi nous devons ajouter à cette fonction objectif une contrainte

de régularité $\mathcal{R}$ sur les valeurs des pixels des cartes reconstruites. Afin de contraindre la positivité des intensités, nous ajoutons également la contrainte épigraphique $\iota_{\mathcal{C}}$ définie par l'équation (4.26).

Le but de ce chapitre est dans un premier temps de motiver le choix de la fonction objectif

Rappel éq. (1.32)
$$\widehat{\boldsymbol{x}} \in \underset{\boldsymbol{x}\in(\mathbb{R}^N)^L)}{\operatorname{Argmin}} \left[\Phi(\boldsymbol{x}) + \mathcal{R}(\boldsymbol{x}) + \iota_{\mathcal{C}}(\boldsymbol{x})\right],$$

où $\Phi : (\mathbb{R}^N)^L \to \mathbb{R}$ est un terme d'attache aux données, $\mathcal{R} : (\mathbb{R}^N)^L \to \mathbb{R}$ est un terme de régularisation des paramètres et $\iota_{\mathcal{C}} : (\mathbb{R}^N)^L \to \{0, +\infty\}$ est une contrainte sur le domaine des paramètres. Nous étudions plusieurs contraintes de régularité $\mathcal{R}$, dont le choix est motivé par la structure des objets que l'on souhaite reconstruire, par exemple lisse, à bords francs ou encore parcimonieuse. Dans un second temps, nous proposons différents algorithmes adaptés aux problèmes de minimisation considérés.

Plus précisément, dans la section 4.1, je développe les étapes permettant d'aboutir à la fonction objectif $\mathcal{F} : (\mathbb{R}^N)^L \to]-\infty, +\infty]$. Tout d'abord, j'établis le modèle des données prenant en compte la convolution. Nous nommerons celui-ci, pour tout $k \in \{1, \dots K\}$ $f_k^{\text{nS-C}}$. J'en déduis alors le terme d'attache aux données Φ. J'introduis ensuite différents termes de régularisation $\mathcal{R}$ et spécifie la contrainte épigraphique $\iota_{\mathcal{C}}$ considérée dans notre problème, ainsi que d'autres contraintes de domaine qui pourraient être ajoutées au modèle.

Dans la section 4.2, j'introduis différents algorithmes permettant de minimiser la fonction objectif Φ en fonction des contraintes inclues. Je compare sur les données simulées, les performances numériques des différents algorithmes pour les différentes régularisations en termes de vitesse de convergence du critère et de l'erreur sur la reconstruction. Par cette comparaison, je cherche à trouver une méthode qui permet d'obtenir des résultats de bonne qualité en un temps de convergence raisonnable, c'est-à-dire quelques minutes pour un petit jeu de données (*e.g.* $128 \times 256 \times 88$) et pas plus de 24h pour un grand jeu de données (*e.g.* $1024 \times 2048 \times 128$).

Enfin, dans la section 4.3, je compare sur données simulées les performances qualitatives des différentes méthodes en un temps de calcul limité, en terme d'Erreur Quadratique Moyenne. Je compare alors les résultats obtenus, par les méthodes ayant les meilleures performances, aux résultats obtenus avec la méthode de la Double Différence et la Méthode non-Linéaire Séparable (MnLS), appliquées sur données *pré-traitées* puis déconvoluées. J'applique également ces méthodes sur le même jeu de données astrophysiques que précédemment et compare les résultats par rapport à une déconvolution *a posteriori* des méthodes séparables.

4.1 Du modèle direct à la fonction objectif

4.1.1 Modèle linéaire non-séparable prenant en compte la convolution

Dans cette section je développe les calculs permettant d'obtenir l'expression des données prenant en compte la convolution, à partir du modèle linéaire séparable des données présenté dans la proposition 2.2.1. Rappelons que ce modèle est paramétré par $\boldsymbol{x} = (\mathrm{I}, \mathrm{Q}, \mathrm{U})^\top$ où, I, Q et U sont les paramètres de Stokes. Le modèle des données prenant en compte la convolution est alors donné par la proposition suivante.

Proposition 4.1.1. *On considère les paramètres d'intérêt* $\overline{\overline{\boldsymbol{x}}} \in (\mathbb{R}^N)^L$*, soit* $\mathbf{A} : \mathbb{R}^N \to \mathbb{R}^N$ *un opérateur de convolution par la PSF instrumentale et* $(\mathbf{T}_{j,k})_{j\in\{1,2\},k\in\{1,\dots K\}} : \mathbb{R}^N \to \mathbb{R}^{M/2}$ *des opérateurs de transformation linéaire. Soit* $(v_{j,k,\ell})_{j\in\{1,2\},k\in\{1,\dots k\},\ell\in\{1,\dots L\}} \in \mathbb{R}$ *l'ensemble des modulations séparables de la polarisation dans l'instrument, alors le modèle linéaire non-séparable des données incluant la déconvolution s'écrit* $k \in \{1,\dots K\}$ *de la forme* $\boldsymbol{d}_k^{nS} = \mathcal{B}_k(f_k^{nS\text{-}C}(\overline{\overline{\boldsymbol{x}_n}}))$ *avec :*

$$f_k^{nS\text{-}C}(\boldsymbol{x}) = \sum_\ell \begin{bmatrix} \mathbf{T}_{1,k} v_{1,k,\ell} \\ \mathbf{T}_{2,k} v_{2,k,\ell} \end{bmatrix} (\mathbf{A}\boldsymbol{x}_\ell), \tag{4.1}$$

où les $(\mathcal{B}_k)_{k\in\{1,\dots K\}} : \mathbb{R}^M \to \mathbb{R}^M$ *produisent une réalisation de la variable gaussienne* $\mathcal{N}(\boldsymbol{x}, Diag(\boldsymbol{x}) + \sigma_{ro}^2)$ *où* σ_{ro}^2 *est la variance du bruit de lecture de l'instrument.*

Démonstration : *(Fin page 101)* Pour cela il est essentiel de se rapporter au champ électrique afin de comprendre comment se comporte la convolution par la PSF vis-và-vis de l'intensité enregistrée par le détecteur. Soit $\boldsymbol{e} \in \mathbb{C}^2$ un champ électrique entrant dans le télescope. On note Ω_m la position angulaire d'un pixel $m \in \{1,\dots M\}$ sur le détecteur et Ω_{ciel} ce à quoi il correspond sur le ciel. On a alors, à l'acquisition $k \in \{1,\dots K\}$ et pour la partie $j \in \{1,2\}$ du détecteur, la convolution en deux dimensions suivante :

$$\boldsymbol{e}_{j,k}^{\text{out}}(\Omega_m) = \int \mathrm{h}(\Omega_m - \Omega_{\text{ciel}}) \langle \boldsymbol{v}_{j,k}, \boldsymbol{e}(\Omega_{\text{ciel}}) \rangle_{\mathbb{C}^2} \mathrm{d}^2\Omega_{\text{ciel}}, \tag{4.2}$$

où h est la fonction de transfert optique de l'instrument. Cette équation découle de deux hypothèses : la linéarité de la réponse instrumentale, qui implique que

$$\boldsymbol{e}_{j,k}^{\text{out}}(\Omega_m) = \int \mathrm{h}(\Omega_m, \Omega_{\text{ciel}}) \langle \boldsymbol{v}_{j,k}, \boldsymbol{e}(\Omega_{\text{ciel}}) \rangle_{\mathbb{C}^2} \mathrm{d}^2\Omega_{\text{ciel}}, \tag{4.3}$$

et la stationnarité de la réponse linéaire qui implique que $texth(\Omega_m, \Omega_{\text{ciel}}) = \mathrm{h}(\Omega_m - \Omega_{\text{ciel}})$. La linéarité est elle-même due au théorème de superposition pour les champs électriques et magnétiques, c'est-à-dire que le champ électrique en sortie et la somme des champs électriques entrant.

On a donc l'intensité sur le détecteur :

$$\begin{aligned}
&\mathbb{E}_t\left[|\boldsymbol{e}_{j,k}^{\text{out}}(\Omega_j)|^2\right] \\
=&\mathbb{E}_t\left[\int \mathrm{h}(\Omega_m - \Omega_{\text{ciel}}) \langle \boldsymbol{v}_{j,k}, e(\Omega_{\text{ciel}}) \rangle_{\mathbb{C}^2} \mathrm{d}^2\Omega_{\text{ciel}} \int \overline{\mathrm{h}\left(\Omega_m - \Omega'_{\text{ciel}}\right) \langle \boldsymbol{v}_{j,k}, e\left(\Omega'_{\text{ciel}}\right) \rangle_{\mathbb{C}^2}} \mathrm{d}^2\Omega'_{\text{ciel}}\right] \\
=&\mathbb{E}_t\left[\iint \mathrm{h}(\Omega_m - \Omega_{\text{ciel}}) \overline{\mathrm{h}\left(\Omega_m - \Omega'_{\text{ciel}}\right)} \langle \boldsymbol{v}_{j,k}, e(\Omega_{\text{ciel}}) \rangle_{\mathbb{C}^2} \overline{\langle \boldsymbol{v}_{j,k}, e\left(\Omega'_{\text{ciel}}\right) \rangle_{\mathbb{C}^2}} \mathrm{d}^2\Omega_{\text{ciel}} \mathrm{d}^2\Omega'_{\text{ciel}}\right] \\
=&\iint \mathbb{E}_t\left[\langle \boldsymbol{v}_{j,k}, e(\Omega_{\text{ciel}}) \rangle_{\mathbb{C}^2} \langle \overline{v}_{k,\ell}, \overline{e}\left(\Omega'_{\text{ciel}}\right) \rangle_{\mathbb{C}^2}\right] \mathrm{h}(\Omega_m - \Omega_{\text{ciel}}) \overline{\mathrm{h}}\left(\Omega_m - \Omega'_{\text{ciel}}\right) \mathrm{d}^2\Omega_{\text{ciel}} \mathrm{d}^2\Omega'_{\text{ciel}},
\end{aligned} \tag{4.4}$$

où

$$\mathbb{E}_t\left[\langle \boldsymbol{v}_{j,k}, e(\Omega_{\text{ciel}})\rangle_{\mathbb{C}^2}\langle \overline{v}_{k,\ell}, \overline{e}\left(\Omega'_{\text{ciel}}\right)\rangle_{\mathbb{C}^2}\right] = |\boldsymbol{v}_{j,k}^{(x)}|^2\mathbb{E}_t\left[\overline{\boldsymbol{e}}^{(x)}(\Omega_{\text{ciel}})\,\boldsymbol{e}^{(x)}\left(\Omega'_{\text{ciel}}\right)\right] \tag{4.5}$$

$$+|\boldsymbol{v}_{j,k}^{(y)}|^2\mathbb{E}_t\left[\overline{\boldsymbol{e}}^{(y)}(\Omega_{\text{ciel}})\,\boldsymbol{e}^{(y)}\left(\Omega'_{\text{ciel}}\right)\right] \tag{4.6}$$

$$+\boldsymbol{v}_{j,k}^{(x)}\overline{\boldsymbol{v}}_{j,k}^{(y)}\mathbb{E}_t\left[\overline{\boldsymbol{e}}^{(x)}(\Omega_{\text{ciel}})\,\boldsymbol{e}^{(y)}\left(\Omega'_{\text{ciel}}\right)\right] \tag{4.7}$$

$$+\overline{\boldsymbol{v}}_{j,k}^{(x)}\boldsymbol{v}_{j,k}^{(y)}\mathbb{E}_t\left[\overline{\boldsymbol{e}}^{(y)}(\Omega_{\text{ciel}})\,\boldsymbol{e}^{(x)}\left(\Omega'_{\text{ciel}}\right)\right] \tag{4.8}$$

Or, l'objet observé dans le ciel n'est pas cohérent spatialement dans le temps, ce qui implique que les espérances (4.5), (4.6), (4.7) et (4.8) sont nulles, sauf dans le cas où $\Omega'_{\text{ciel}} = \Omega_{\text{ciel}}$. On introduit donc la distribution de Dirac δ, qui est telle que pour une fonction g quelconque :

$$\begin{cases} \int_{-\infty}^{+\infty} g(x)\delta(x-y)\,\mathrm{d}x = g(y), \\ \int_{-\infty}^{+\infty} \delta(x)\,\mathrm{d}x = 1. \end{cases} \tag{4.9}$$

De ce fait, on a pour (4.7) :

$$\begin{aligned} &\int \mathbb{E}_t\left[\overline{\boldsymbol{e}}^{(x)}(\Omega_{\text{ciel}})\,\boldsymbol{e}^{(y)}\left(\Omega'_{\text{ciel}}\right)\right]\mathrm{h}(\Omega_m-\Omega_{\text{ciel}})\,\overline{\mathrm{h}}\left(\Omega_m-\Omega'_{\text{ciel}}\right)\mathrm{d}^2\Omega_{\text{ciel}} \\ =&\int \mathbb{E}_t\left[\overline{\boldsymbol{e}}^{(x)}(\Omega_{\text{ciel}})\,\boldsymbol{e}^{(y)}\left(\Omega'_{\text{ciel}}\right)\right]\delta\left(\Omega'_{\text{ciel}}-\Omega_{\text{ciel}}\right)\mathrm{h}(\Omega_m-\Omega_{\text{ciel}})\,\overline{\mathrm{h}}\left(\Omega_m-\Omega'_{\text{ciel}}\right)\mathrm{d}^2\Omega_{\text{ciel}} \\ =&\mathbb{E}_t\left[\overline{\boldsymbol{e}}^{(x)}(\Omega_{\text{ciel}})\,\boldsymbol{e}^{(y)}(\Omega_{\text{ciel}})\right]\mathrm{h}(\Omega_m-\Omega_{\text{ciel}})\,\overline{\mathrm{h}}(\Omega_m-\Omega_{\text{ciel}}). \end{aligned}$$

De même pour les espérances (4.5), (4.6) et (4.8). En remplaçant dans l'équation (4.4), on a donc :

$$\begin{aligned} \mathbb{E}_t\left[|\boldsymbol{e}_{j,k}^{\text{out}}(\Omega_m)|^2\right] = \int \Bigg[&|\boldsymbol{v}_{j,k}^{(x)}|^2\mathbb{E}_t\left[|\overline{\boldsymbol{e}}^{(x)}(\Omega_{\text{ciel}})|^2\right] + |\boldsymbol{v}_{j,k}^{(y)}|^2\mathbb{E}_t\left[|\overline{e}^{(y)}(\Omega_{\text{ciel}})|^2\right] \\ &+\boldsymbol{v}_{j,k}^{(x)}\overline{\boldsymbol{v}}_{j,k}^{(y)}\mathbb{E}_t\left[\overline{\boldsymbol{e}}^{(x)}(\Omega_{\text{ciel}})\,\boldsymbol{e}^{(y)}(\Omega_{\text{ciel}})\right] \\ &+\overline{\boldsymbol{v}}_{j,k}^{(x)}\boldsymbol{v}_{j,k}^{(y)}\mathbb{E}_t\left[\boldsymbol{e}^{(x)}(\Omega_{\text{ciel}})\,\overline{\boldsymbol{e}}^{(y)}(\Omega_{\text{ciel}})\right]\Bigg]|\mathrm{h}(\Omega_m-\Omega_{\text{ciel}})|^2\,\mathrm{d}^2\Omega_{\text{ciel}}. \end{aligned} \tag{4.10}$$

D'où, d'après les démonstrations des propositions 2.2.1 et 2.1.1, le fait qu'on a d'une part :

$$\begin{aligned} \mathbb{E}_t\left[|\boldsymbol{e}_{j,k}^{\text{out}}(\Omega_m)|^2\right] &= \int\left(\|\boldsymbol{v}_{j,k}\|_{\mathbb{C}^2}^2\frac{I^{\text{u}}}{2} + I^{\text{p}}|\langle \boldsymbol{v}_{j,k}, \boldsymbol{c}(\theta)\rangle_{\mathbb{C}^2}|^2\right)|\mathrm{h}(\Omega_m-\Omega_{\text{ciel}})|^2\,\mathrm{d}^2\Omega_{\text{ciel}} \\ &= \mathbf{A}\left(\|\boldsymbol{v}_{j,k}\|_{\mathbb{C}^2}^2\frac{I^{\text{u}}}{2} + I^{\text{p}}|\langle \boldsymbol{v}_{j,k}, \boldsymbol{c}(\theta)\rangle_{\mathbb{C}^2}|^2\right), \end{aligned} \tag{4.11}$$

et d'autre part

$$\begin{aligned} \mathbb{E}_t\left[|\boldsymbol{e}_{j,k}^{\text{out}}(\Omega_m)|^2\right] &= \int\left(v_{j,k,1}\mathrm{I} + v_{j,k,2}\mathrm{Q} + v_{j,k,3}\mathrm{U}\right)|\mathrm{h}(\Omega_m-\Omega_{\text{ciel}})|^2\,\mathrm{d}^2\Omega_{\text{ciel}} \\ &= \sum_{\ell=1}^{3}\left(v_{j,k,\ell}\int \boldsymbol{x}_\ell\,|\mathrm{h}(\Omega_m-\Omega_{\text{ciel}})|^2\,\mathrm{d}^2\Omega_{\text{ciel}}\right) \\ &= \sum_{\ell=1}^{3} v_{j,k,\ell}\mathbf{A}\boldsymbol{x}_\ell. \end{aligned} \tag{4.12}$$

où $\boldsymbol{x} = (\mathrm{I}, \mathrm{Q}, \mathrm{U})^\top$ et

$$v_{j,k,1} = \frac{\|\boldsymbol{v}_{j,k}\|^2_{\mathbb{C}^2}}{2}, \quad v_{j,k,2} = \frac{|\boldsymbol{v}^{(x)}_{j,k}|^2 - |\boldsymbol{v}^{(y)}_{j,k}|^2}{2} \quad \text{et} \quad v_{j,k,3} = \mathfrak{Re}\left(\boldsymbol{v}^{(x)}_{j,k}\boldsymbol{v}^{(y)}_{j,k}\right). \tag{2.27}$$

□

Remarque : En notant $\mathbf{A}\boldsymbol{x} = (\mathbf{A}\boldsymbol{x}_\ell)_{\ell\in\{1,\dots L\}}$, la formulation (4.1) est donc équivalente à

$$f_k^{\text{nS-C}}(\boldsymbol{x}) = \begin{bmatrix}\mathbf{T}_{1,k} f^{\mathrm{S}}_{1,k} \\ \mathbf{T}_{1,k} f^{\mathrm{S}}_{2,k}\end{bmatrix}(\mathbf{A}\boldsymbol{x}) = f_k^{\mathrm{nS}}(\mathbf{A}\boldsymbol{x}), \tag{4.13}$$

où les $(f^{\mathrm{S}}_{j,k})_{j\in\{1,2\},k\in\{1,\dots K\}}$ sont données par le modèle linéaire séparable (2.26), avec $(f^{\mathrm{S}}_{j,k}(\boldsymbol{x}))_n = f^{\mathrm{S}}_{j,k}(\boldsymbol{x}_n)$ pour tout $n \in \{1,\dots N\}$, et les $(f_k^{\mathrm{nS}})_{k\in\{1,\dots K\}}$ par le modèle non-séparable (3.2). Dans le cas d'un changement de variable non-linéaire $\boldsymbol{x} = g(\boldsymbol{y})$ des paramètres d'intérêts, tel que celui présenté dans la section 3.2, l'opérateur de convolution ne commute pas avec g et on a alors

$$f_k^{\text{nS-C}}(g(\boldsymbol{y})) = \begin{bmatrix}\mathbf{T}_{1,k} f^{\mathrm{S}}_{1,k} \\ \mathbf{T}_{1,k} f^{\mathrm{S}}_{2,k}\end{bmatrix}(\mathbf{A}\boldsymbol{x}) = f_k^{\mathrm{nS}}(\mathbf{A}g(\boldsymbol{y})), \tag{4.14}$$

L'opérateur de convolution $\mathbf{A}$ peut également commuter avec les transformations de la polarisation $(v_{j,k,\ell})_{j\in\{1\,;\,2\},k\in\{1,\dots K\},\ell\in\{1,\dots L\}}$, et avec la transformation du détecteur $(\mathbf{T}_{j,k})_{j\in\{1\,;\,2\},k\in\{1,\dots K\}}$ mais une telle écriture du modèle $f_k^{\text{nS-C}}$ a une complexité plus élevée que la formulation (4.14) dans notre cas. En effet, la convolution serait alors appliquée deux fois pour chacune des $k \in \{1,\dots K\}$ acquisitions au lieu d'une application par composante $\ell \in \{1,\dots L\}$ et dans notre application $L = 3$ et $K \geq 4$.

Notons que les $(f^{\mathrm{S}}_{j,k})_{j\in\{1,2\},k\in\{1,\dots K\}}$ doivent nécessairement être linéaires pour que la formulation (4.14) soit vraie. Dans le cas non-linéaire, un changement de variable tel que celui fait pour le modèle (3.11) est nécessaire.

4.1.2 Fonction objectif

Pour estimer les paramètres $\overline{\boldsymbol{x}}$ du modèle des données établi dans la proposition 4.1.1, nous minimisons un problème de la forme

Rappel éq. (1.32)
$$\widehat{\boldsymbol{x}} \in \underset{\boldsymbol{x}\in(\mathbb{R}^N)^L}{\operatorname{Argmin}} \left[\Phi(\boldsymbol{x}) + \mathcal{R}(\boldsymbol{x}) + \iota_{\mathcal{C}}(\boldsymbol{x})\right],$$

où $\Phi : (\mathbb{R}^N)^L \to \mathbb{R}$ est un terme d'attache aux données, $\mathcal{R} : (\mathbb{R}^N)^L \to \mathbb{R}$ est un terme de régularisation des paramètres et $\iota_{\mathcal{C}} : (\mathbb{R}^N)^L \to \{0, +\infty\}$ permet d'imposer une contrainte sur les paramètres, c'est-à-dire la contrainte $\boldsymbol{x} \in \mathcal{C}$. Le but de cette section est de donner la forme des différentes fonctions composant la fonction objectif.

4.1.2.1 Attache aux données

Celle-ci, comme dans la section précédente, correspond au carré de la distance de Mahalanobis qui est donné pour tout $\boldsymbol{x} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_L) \in (\mathrm{R}^N)^L$ par :

Rappel éq. (3.6)
$$\Phi(\boldsymbol{x}) = \textstyle\sum_{k=1}^{K} \frac{1}{2} \left\| \boldsymbol{d}_k^{\mathrm{nS}} - f_k^{\text{nS-C}}(\boldsymbol{x}) \right\|^2_{\mathbf{W}_k}.$$

où $\forall k \in \{1,\dots K\}$, $\forall m \in \{1,\dots M\}$

Rappel éq. (3.5)
$$\mathbf{W}_{k,m} = \begin{cases} \mathrm{Cov}(\boldsymbol{d}_{k,m}^{\mathrm{nS}})^{-1}, \\ 0 \quad \text{si pixel ou acquisition invalides,} \end{cases}$$

et $f_k = f_k^{\text{nS-C}}$ défini par l'équation (4.1). On suppose que les réalisations de bruit sont indépendantes pour les M pixels des K acquisitions.

4.1.2.2 Le choix des régularisations

Comme dans la section précédente, les opérateurs $f_k^{\text{nS-C}}$ sont mal conditionnés et nécessitent l'ajout de contraintes de régularisation, permettant ainsi de diminuer la variance de la reconstruction obtenue par minimisation seule de l'attache aux données.

De manière générale, les environnements circumstellaires sont spatialement relativement homogènes sans grandes variations d'intensité, avec des contours assez précis. C'est pourquoi nous avons choisi d'utiliser des régularisations de type lissage à bord franc. Pour plus de détails sur les régularisations, se référer à la section 1.2.3. Dans cette section, je détaille les différentes formes que va prendre la fonction $\mathcal{R}$ en fonction de la régularisation utilisée, sur les paramètres $\boldsymbol{x} = (\mathrm{I}, \mathrm{Q}, \mathrm{U})^\top$.

1. Régularisation de Tikhnov [Tikhonov, 1963] (équation (1.47)) : Comme on applique cette régularisation indépendamment sur les cartes I, Q et U, la fonction de régularisation s'écrit alors pour $\lambda_\ell \geq 0$ comme :
$$\mathcal{R}(\boldsymbol{x}) = \sum_{\ell=1}^{L} \lambda_\ell \|\mathbf{D}\boldsymbol{x}_\ell\|_2^2. \tag{4.15}$$
La fonction $\mathcal{R}$ est convexe, différentiable de gradient $\lambda\|\mathbf{D}\|^2$-Lipschitz.
2. Régularisation par la Variation Totale (TV) [Rudin et al., 1992] (équation (1.48)) : Le terme de régularisation est donné pour $\lambda_\ell \geq 0$ par :
$$\mathcal{R}(\boldsymbol{x}) = \sum_{\ell=1}^{L} \lambda_\ell \|\mathbf{D}\boldsymbol{x}_\ell\|_{\ell_{1,2}}, \tag{4.16}$$
où la norme $\ell_{1,2}$ est définie dans l'équation (1.49). Une telle régularisation est convexe, mais non-différentiable. La recherche de minimum, dans le cas d'une telle contrainte, peut se faire soit à l'aide d'outils d'optimisation non-différentiable, comme décrits dans la section 1.3, ou alors on peut avoir recours à son approximation hyperbolique qui est différentiable.
3. Régularisation par *a priori* de lissage avec préservation des bords (TV-h : approximation hyperbolique de TV) [Charbonnier et al., 1997] (équation (1.50)) : Le terme de régularisation est donné $\forall \lambda_\ell, \mu_\ell \geq 0$ par
$$\mathcal{R}(\boldsymbol{x}) = \sum_{\ell=1}^{L} \lambda_\ell \left(\sqrt{\|\mathbf{D}\boldsymbol{x}_\ell\|_2^2 + \mu_\ell^2} - \mu_\ell \right), \tag{4.17}$$
qui est différentiable de gradient $\max_\ell\{\lambda_\ell/\mu_\ell\}\|\mathbf{D}\|^2$-Lipschitz.

Un problème bien connu de la régularisation par TV est son effet *« cartoon »*, qui donne une impression d'aplats de couleurs. La régularisation TV-h est aussi sujette à cet effet mais en moindre mesure du fait du paramètre μ, qui rend la contrainte plus quadratique sur les zones où le gradient spatial est proche de 0. L'utilisation d'autre contraintes peut pallier ce problème d'effet *«catoon »*.

4. Régularisation par la norme de Schatten sur la matrice hessienne (Shatten) [Lefkimmiatis et al., 2013] (équation (1.54)) : L'intérêt d'une telle régularisation est que la contrainte de bord francs est plus souple, ce qui permet d'éviter les effets *« cartoons »* et l'aplatissement de détails assez fin, c'est-à-dire de quelques pixels de largeur ou d'intensité faible. L'expression de la régularisation par la norme de Shatten à l'ordre 1 sur l'opérateur Hessien est donné par :

$$\mathcal{R}(\boldsymbol{x}) = \sum_{\ell=1}^{L} \lambda_\ell S_1\left(\mathbf{D}^2 \boldsymbol{x}_\ell\right). \tag{4.18}$$

où S_1 est la norme de Shatten à l'ordre de 1, définie par l'équation (1.55). Elle consiste à sommer la valeur absolue des valeurs propres de l'opérateur Hessien $\mathbf{D}^2$ de $\boldsymbol{x}$. Cette régularisation est également convexe mais non-différentiable.

5. Régularisation par la Variation Totale Généralisée (TGV) [Bredies et al., 2010] (équation (1.56)) : Cette régularisation est un compromis entre la Variation Totale et la pénalisation de la norme de Shatten. L'avantage d'une telle pénalisation est de laisser apparaître des bords aussi francs qu'avec TV, tout en évitant les effets *« cartoon »*. Afin de pouvoir inclure cette régularisation dans notre problème, tout en évitant d'avoir recours à des sous-itérations pour estimer le paramètre $\boldsymbol{y} \in (\mathbb{R}^{2N})^L$ de la régularisation, on augmente la dimension du problème en posant comme paramètres d'intérêt $\boldsymbol{z} = (\boldsymbol{x}, \boldsymbol{y})^\top$. Je résous alors le critère (1.32) en $\boldsymbol{z}$, avec $\Phi(\boldsymbol{z}) = \Phi(\boldsymbol{x})$, $\iota_{\mathcal{C}}(\boldsymbol{z}) = \iota_{\mathcal{C}}(\boldsymbol{x})$ et

$$\mathcal{R}(\boldsymbol{z}) = \sum_{\ell=1}^{L} \lambda_\ell \Big[\gamma_1 \|\mathbf{D}\boldsymbol{x}_\ell - \boldsymbol{y}_\ell\|_{\ell_1} + \gamma_0 \|\boldsymbol{D}\boldsymbol{y}_\ell\|_{\ell_1} \Big]. \tag{4.19}$$

6. Régularisation par TV + ℓ_1 : Dans certains jeux de données, il peut y avoir des étoiles, proches de l'étoile hôte, qui font partie de l'environnement. Leur déconvolution en utilisant une régularisation de lissage à bord franc peut être compliquée du fait des rebonds qui en découlent. En effet, un réglage des hyperparamètres optimal pour la reconstruction du disque peut être sous-optimale pour la reconstruction du point source et inversement. C'est pourquoi, sur le même exemple que la variation totale généralisée, j'ai étudié la décomposition de $\boldsymbol{x}$ en une composante constante par morceaux, qui correspondrait au disque et au résidu stellaire et une composante parcimonieuse, qui correspondrait aux étoiles présentes dans le champ. L'expression d'une telle contrainte est donnée $\forall \lambda_\ell \geq 0$ et $\forall \boldsymbol{x}_\ell \in \mathcal{H}$ avec $\boldsymbol{x}_\ell = \boldsymbol{x}_\ell^{\text{smooth}} + \boldsymbol{x}_\ell^{\text{sparse}}$ par

$$\mathcal{R}(\boldsymbol{x}_\ell) = \lambda_\ell^{\text{smooth}} \|\mathbf{D}\boldsymbol{x}_\ell^{\text{smooth}}\|_{\ell_1} + \lambda_\ell^{\text{sparse}} \|\boldsymbol{x}_\ell^{\text{sparse}}\|_{\ell_1} \tag{4.20}$$

Les régularisations par TGV et par TV+ℓ_1 ne seront pas présentées dans ce manuscrit mais les résultats sont présentés dans [Denneulin et al., 2019] (cf. Annexe B). J'ai conclu de cette contribution que la régularisation par TGV n'apportait pas d'avantages par rapport à une régularisation par TV ou Shatten dans notre contexte d'étude.

4.1.2.3 Contraintes sur le domaine de définition des solutions

On note $\mathcal{C} \subset (\mathbb{R}^N)^L$ le domaine de définition des solutions. Contraindre les solutions à ce domaine consiste en l'ajout d'une indicatrice à la fonction objectif, c'est-à-dire :

$$\iota_{\mathcal{C}}(\boldsymbol{x}) = \begin{cases} 0 & \text{si } x \in \mathcal{C} \\ +\infty & \text{sinon.} \end{cases} \tag{4.21}$$

Concrètement, si cette contrainte n'est pas vérifiée, la fonction objectif sera égale à l'infini. Elle assure donc que le minimum de la fonction objectif soit dans le domaine.

Cette contrainte n'est pas différentiable. Il est donc nécessaire de passer par son opérateur proximal qui équivaut à la projection sur l'ensemble $\mathcal{C}$ (*cf.* section 1.2.4 pour la peuve). Les contraintes proposées dans ce chapitre sont :

1. Contrainte angulaire (ne sera pas présentée dans les résultats) : Cette contrainte linéaire est inspirée des projections définies pour chaque pixel $n \in \{1, \dots N\}$ par :

   Rappel éq. (1.28)
   $$\begin{cases} P_{n,\perp} = \mathrm{Q}_n \cos(2\varphi_n) + \mathrm{U}_n \sin(2\varphi_n) \\ P_{n,\parallel} = \mathrm{U}_n \cos(2\varphi_n) - \mathrm{Q}_n \sin(2\varphi_n). \end{cases}$$

   Cette contrainte utilise le fait que l'on sait où se trouve l'étoile par rapport au disque et donc qu'on a une idée de l'orientation de la polarisation par rapport à celle-ci.
   Pour rappel, cette contrainte vient du fait que l'angle de polarisation de la lumière polarisée linéairement est perpendiculaire à l'angle d'incidence du rayon. Comme le centre de l'étoile est connu, il est possible d'avoir une information *a priori* sur l'angle de polarisation. En effet, dans le cas d'un disque vu de face, l'angle de polarisation à une position donnée correspond alors à la position sur le cercle trigonométrique centré en l'étoile. On appelle $\psi \in \mathbb{R}^N$ la carte d'angles *a priori* telle que $\forall n \in \{1, \dots N\}$, $\psi_n = \arctan(\Omega_n)$ où Ω_n est la position angulaire du pixel n. On contraint donc les paramètres

   $$\begin{cases} \mathrm{Q} = I^{\mathrm{p}} \cos 2\Theta \\ \mathrm{U} = I^{\mathrm{p}} \sin 2\Theta, \end{cases} \tag{4.22}$$

   en les projetant parallèlement et orthogonalement, pour chaque pixel $n \in \{1, \dots N\}$, dans la base $(\cos 2\psi, \sin 2\psi)$ comme présenté à l'équation (1.28). Ainsi, si $\theta = \psi$, on a $P_{\parallel} = I_p$ et $P_{\perp} = 0$. Dans la littérature, $P_{\parallel}$ et $P_{\perp}$ sont aussi respectivement appelés Q_ϕ et U_ϕ.
   Comme I^{p} est inconnue il n'est pas possible d'utiliser $P_{\parallel} = I_p$ comme contrainte. Il est cependant possible de contraindre Q et U à vérifier $P_{\perp} = 0$, ce qui correspond à contraindre linéairement Q et U au domaine :

   $$\mathcal{C} = \left\{ (\mathrm{I}, \mathrm{Q}, \mathrm{U})^\top \in (\mathbb{R}^N)^3 \text{ tels que } \forall n \in \{1, \dots, N\},\ \mathrm{U}_n \cos(2\psi_n) - \mathrm{Q}_n \sin(2\psi_n) = 0 \right\}. \tag{4.23}$$

   L'opérateur proximal de l'indicatrice de cette contrainte, qui peut être réécrite comme :

   $$\mathcal{C} = \left\{ (\mathrm{I}, \mathrm{Q}, \mathrm{U})^\top \in (\mathbb{R}^N)^L \text{ tels que } \forall n \in \{1, \dots, N\}, \quad \left\langle \begin{pmatrix} \mathrm{I}_n \\ \mathrm{Q}_n \\ \mathrm{U}_n \end{pmatrix}, \begin{pmatrix} 0 \\ -\sin 2\psi_n \\ \cos 2\psi_n \end{pmatrix} \right\rangle = 0 \right\},$$

consiste à appliquer la projection suivant :

$$\forall n \in \{1,\dots N\}, \quad \mathbb{P}_{\mathcal{C}}\left(\mathrm{I}_n, \mathrm{Q}_n, \mathrm{U}_n\right) = \begin{pmatrix} \mathrm{I}_n \\ \mathrm{Q}_n \\ \mathrm{U}_n \end{pmatrix} - \left\langle \begin{pmatrix} \mathrm{I}_n \\ \mathrm{Q}_n \\ \mathrm{U}_n \end{pmatrix}, \begin{pmatrix} 0 \\ -\sin 2\psi_n \\ \cos 2\psi_n \end{pmatrix} \right\rangle \begin{pmatrix} 0 \\ -\sin 2\psi_n \\ \cos 2\psi_n \end{pmatrix} \tag{4.24}$$

Le problème d'une telle contrainte est qu'elle n'est utile que quand le disque est vu complètement de face. Elle est donc bien trop forte en conditions réelles et risque de trop biaiser la solution. C'est pourquoi nous ne l'utiliserons pas dans les applications. Cependant, lorsque le disque est vu complètement de face et que les angles sont connus, une telle contrainte peut être imposée pour maximiser l'intensité polarisée.

2. Contrainte épigraphique : Le but d'une telle contrainte est de contraindre la positivité des pixels des paramètres I^{u} et I^{p}, dans le cas de la résolution du modèle linéaire en I, Q et U. En effet, dans le cas linéaire, la contrainte :

$$\mathcal{C} = \left\{ (I^{\mathrm{u}}, I^{\mathrm{p}}, \theta)^\top \in (\mathbb{R}^N)^L, \quad \text{tels que} \quad \forall n \in \{1,\dots N\}, \quad I^{\mathrm{u}} \geq 0 \text{ et } I^{\mathrm{p}} \geq 0 \right\} \tag{4.25}$$

peut être traduite par la contrainte épigraphique entre les composantes du paramètre d'intérêt $\boldsymbol{x} = (\mathrm{I}, \mathrm{Q}, \mathrm{U})^\top$:

$$\mathcal{C} = \left\{ (\mathrm{I}, \mathrm{Q}, \mathrm{U})^\top \in (\mathbb{R}^N)^L, \quad \text{tels que} \quad \forall n \in \{1,\dots,N\}, \sqrt{\mathrm{Q}_n^2 + \mathrm{U}_n^2} \leq \mathrm{I}_n \right\}. \tag{4.26}$$

De fait, vérifier la contrainte donnée par la formulation (4.25) implique que l'on doit vérifier pour tout $n \in \{1,\dots N\}$ la contrainte $\mathrm{I}_n \geq 0$, car $\mathrm{I}_n = I_n^{\mathrm{u}} + I_n^{\mathrm{p}}$ en s'assurant la positivité de I_n^{u} et de I_n^{p}. La contrainte $I_n^{\mathrm{p}} \geq 0$ est assurée par sa définition vis-à-vis des paramètres de Stokes Q_n et U_n : $I_n^{\mathrm{p}} = \sqrt{\mathrm{Q}_n^2 + \mathrm{U}_n^2}$. Il reste alors à vérifier $I_n^{\mathrm{u}} \geq 0$, qui est vérifiée si et seulement si $I_n^{\mathrm{u}} + I_n^{\mathrm{p}} \geq 0$, donc si $\mathrm{I}_n \geq I_n^{\mathrm{p}}$. En développant l'expression de I_n^{p} par rapport aux paramètres de Stokes Q_n et U_n, on retrouve l'expression de la contrainte épigraphique.
On retrouve l'utilisation d'une telle contrainte épigraphique en astrophysique dans l'utilisation de la polarimétrie en radio-interférométrie [Repetti et al., 2017]. Cependant elle n'a pas encore été exploitée jusqu'à aujourd'hui dans les méthodes de reconstruction en imagerie directe à haut contraste.

4.2 MLnS-D : Méthodes Linéaires non-Séparables avec Déconvolution

Dans la section précédente, nous avons développé un modèle complet des données, prenant en compte les transformations du détecteur et la convolution des paramètres par la PSF. Nous avons ensuite énoncé les différents éléments composants la fonction objectif $\Phi : (\mathbb{R}^N)^L \to \mathbb{R}$ défini par l'équation (3.6) et proposé une liste de régularisations $\mathcal{R} : (\mathbb{R}^N)^L \to \mathbb{R}$ et de contraintes $\iota_{\mathcal{C}}$ sur le domaine des paramètres.

Dans cette section, notre but est d'étudier trois algorithmes permettant de résoudre le problème de minimisation (1.32) selon le choix des régularisations et des contraintes de domaine. Ces méthodes correspondent à trois configurations de critères : un critère entièrement différentiable, un critère avec fonction objectif et régularisations différentiables, avec une contrainte épigraphique, et enfin un critère avec une régularisation non-différentiable comportant un opérateur linéaire.

4.2.1 Résolution par approche différentiable

Lorsque le choix des régularisations permet d'avoir une fonction objectif différentiable et une simple contrainte de dynamique, il est possible d'utiliser la méthode de gradient préconditionné avec un préconditionnement par Broyden-Fletcher-Goldfarb-Shanno à mémoire limitée (ℓ-BFGS), donné par l'algorithme 1, comme dans le chapitre précédent.

Nous appliquons cet algorithme dans le cas de la régularisation à préservation de bord TV hyperbolique.

Comme il a été expliqué dans la section 1.3.2, cette méthode, si elle permet d'inclure une contrainte de positivité dans le cas non-linéaire (cf. section 3.2), ne nous permet pas d'inclure la contrainte épigraphique dans le cas linéaire, car elle n'est pas séparable en les paramètres. C'est pourquoi pour inclure une telle contrainte, nous utilisons l'algorithme du gradient projeté, qui peut-être vu comme un cas particulier de forward-backward, présenté dans la section 1.3.3.

4.2.2 Résolution par approche différentiable avec projection épigraphique

Comme il a été présenté dans la section 1.3.3, introduire la contrainte épigraphique (4.26) nécessite le calcul de sa projection. L'opérateur proximal de cette indicatrice est donné par la projection sur l'épigraphe de la fonction définie pour tout $\boldsymbol{x} \in \mathbb{R}^2$ par $\sqrt{\boldsymbol{x}_1^2 + \boldsymbol{x}_2^2}$. D'après [Chierchia et al., 2012, Proposition 3.4, Corollaire 3.5], la projection sur l'épigraphe des paramètres I, Q et U est donnée pour tout $n \in \{1, \dots N\}$, par :

$$\left(\left(\mathrm{Q}_n^{\perp}, \mathrm{U}_n^{\perp}\right), \mathrm{I}_n^{\perp}\right) = \mathbb{P}_{\mathrm{epi}\|\cdot\|_2}\left((\mathrm{Q}_n, \mathrm{U}_n), \mathrm{I}_n\right) \tag{4.27}$$

$$\Rightarrow \begin{cases} (\mathrm{Q}_n^{\perp}, \mathrm{U}_n^{\perp}) = \begin{cases} (0,0) & \text{si } \|(\mathrm{Q}_n, \mathrm{U}_n)\|_2 \leq -\mathrm{I}_n, \\ (\mathrm{Q}_n, \mathrm{U}_n) & \text{si } \|(\mathrm{Q}_n, \mathrm{U}_n)\|_2 \leq \mathrm{I}_n, \\ (\mathrm{Q}_n, \mathrm{U}_n)\left(1 + \frac{\mathrm{I}}{\|(\mathrm{Q}_n, \mathrm{U}_n)\|_2}\right)/2 & \text{sinon}; \end{cases} \\ \\ \mathrm{I}_n^{\perp} = \max\left\{\left\|(\mathrm{Q}_n^{\perp}, \mathrm{U}_n^{\perp})\right\|_2, \mathrm{I}\right\}. \end{cases} \tag{4.28}$$

On utilise ici l'algorithme Forward-Backward préconditionné, ou Variable Metric Forward-Backward (VMFB) en anglais, donné par l'algorithme 2. Pour la clarté de l'algorithme, on réécrit le critère (1.32) sous la forme :

$$\widehat{\boldsymbol{x}} \in \underset{\boldsymbol{x} \in (\mathbb{R}^N)^L}{\mathrm{Argmin}}\ \mathcal{F}(\boldsymbol{x}) + \mathcal{G}(\boldsymbol{x}). \tag{4.29}$$

où $\mathcal{F}(\boldsymbol{x}) = \Phi(\boldsymbol{x}) + \mathcal{R}(\boldsymbol{x})$ est convexe et différentiable, de constante de Lipschitz $\beta > 0$ et $\mathcal{G}(\boldsymbol{x}) = \iota_{\mathcal{C}}(\boldsymbol{x})$ propre convexe et semi-continue inférieurement.

La vitesse de convergence de l'algorithme 2 repose essentiellement sur la constante de Lipschitz $\beta > 0$ de $\mathcal{F}$. Il est donc important d'avoir une bonne estimation de cette constante, car une mauvaise estimation peut mener soit à la non-convergence de l'algorithme, si l'estimation est plus petite que β, soit à une convergence plus lente, si l'estimation est plus grande que β.

Cette constante peut être calculée en utilisant la méthode de la puissance, dont les itérations sont données par l'algorithme 3 dans la section 1.3.3, cependant cette méthode peut avoir des

difficultés à converger lorsque les opérateurs ont des valeurs trop grandes ou trop petites, à cause des erreurs numériques que cela induit.

Pour plus de simplicité et de robustesse, nous utilisons donc une étape de backtracking, permettant de s'affranchir du calcul de la constante de Lipschitz, comme fait dans l'algorithme ISTA avec backtracking dans [Beck and Teboulle, 2009], ce qui mène aux itérations suivantes :

Algorithme 7 : Variable Metric Forward-Backward with Backtracking (VMFBwB)

Initialiser $\boldsymbol{x}^{[0]}(\mathbb{R}^N)^L$, $\beta^{[0]} > 0$ et $\eta > 1$.
pour $t = 0, 1, \ldots$ **faire**
  pour $i = 0, 1, \ldots$ **faire**
    $\mathring{\beta}^{[i]} = \eta^i \beta^{[t]}$ et mettre à jour $\mathbf{P}^{[i]}$ tel que (4.32) soit respectée.
    $\mathring{\boldsymbol{x}}^{[i]} = \mathrm{prox}_{\mathcal{G}}^{\mathbf{P}^{[i]}}(\boldsymbol{x}[t] - \mathbf{P}^{[i]} \nabla \mathcal{F}(\boldsymbol{x}^{[t]}))$
    si $\mathcal{F}(\mathring{\boldsymbol{x}}^{[i]}) \leq \mathcal{F}(\boldsymbol{x}^{[t]}) + \langle \mathring{\boldsymbol{x}}^{[i]} - \boldsymbol{x}^{[t]}, \nabla \mathcal{F}(\boldsymbol{x}^{[t]}) \rangle + \frac{\beta^{[i]}}{2} \|\mathring{\boldsymbol{x}}^{[i]} - \boldsymbol{x}^{[t]}\|^2$ **alors**
      $\beta^{[t+1]} = \mathring{\beta}^{[i]}$,
      $\boldsymbol{x}^{[t+1]} = \mathring{\boldsymbol{x}}^{[i]}$.
      fin pour

Nous énonçons maintenant les propriétés garantissant la convergence de l'algorithme 7. L'itération principale de l'algorithme 7 est un cas particulier de l'algorithme 2 dans le cas où $\gamma = 1$. D'après le théorème 1.3.3, l'algorithme 7 converge si :

$$\forall t \in \mathbb{N}, \quad \|\mathbf{P}^{[t]}\| \leq \frac{2}{\beta}. \tag{4.30}$$

Dans le cas du backtracking, β n'est pas connu. Les garanties de convergence de VFBwB reposent sur le théorème 4.2.2 suivant, qui s'appuie sur le théorème 4.2.1.

Théoreme 4.2.1. *Soit $\beta^{[0]} > 0$, $\eta > 1$ $i \in \mathbb{N}$, et $(\beta^{[t]})_{t\in\mathbb{N}}$ une suite croissante telle que $\beta^{[t+1]} = \eta^i \beta^{[t]}$ où i est choisi selon l'algorithme 7. Alors, $\exists \widehat{t} \geq 0$ tel que $\beta^{[\widehat{t}]} \geq \beta$. De plus, en posant $\widehat{\beta} = \beta^{[\widehat{t}]}$, alors $\forall t \geq \widehat{t}$, $\beta^{[t]} = \widehat{\beta}$.*

Démonstration : *(Fin page 107)* Comme $\mathcal{F}$ est à gradient β-Lipschitz, cela implique que :

$$\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{H}, \mathcal{K}, \quad \mathcal{F}(\boldsymbol{x}) \leq \mathcal{F}(\boldsymbol{y}) + \langle \boldsymbol{x} - \boldsymbol{y}, \nabla \mathcal{F}(\boldsymbol{y}) \rangle + \frac{\beta}{2} \|\boldsymbol{x} - \boldsymbol{y}\|^2. \tag{4.31}$$

La suite $(\beta^{[t]})_{t\in N}$ est croissante, par conséquent $\exists \widehat{t} > 0$ tel que $\beta^{[\widehat{t}]} \geq \beta$. Alors, à chaque itération $t \geq \widehat{t}$, comme pour la valeur $\widehat{\beta}$, la propriété (4.31) est respectée et par conséquence, $\forall t \geq \widehat{t}$, $\beta^{[t+1]} = \eta^0 \widehat{\beta} = \widehat{\beta}$. □

Théoreme 4.2.2. *Soit $\mathcal{F} \in \Gamma_0(\mathcal{H})$ et $\mathcal{G} \in \Gamma_0(\mathcal{K})$, où $\mathcal{F}$ est différentiable avec $\nabla\mathcal{F}$ différentiable et β-Lipschitz. Soit $\beta^{[0]} \geq 0$, $\eta > 1$ et $(\beta^{[t]})_{t\in\mathbb{N}}$ une suite croissante telle que $\beta^{[t+1]} = \eta^i \beta^{[t]}$ où i est obtenu selon l'algorithme 7. Soit $(\mathbf{P}^{[t]})_{t\in\mathbb{N}} \in \mathcal{P}_\rho(\mathcal{H})$ une suite de matrices symétriques définies positives, telle que*

$$\forall t \geq 0, \quad \|\mathbf{P}^{[t]}\| \leq 2/\beta^{[t]}. \tag{4.32}$$

Alors la suite $(\boldsymbol{x}^{[t]})_{t\in N}$ définie dans l'algorithme 7 converge vers un minimum $\overline{\overline{\boldsymbol{x}}}$ de l'équation (4.29).

Démonstration : *(Fin page 108)* D'après le théorème 4.2.1, $\exists \widehat{t}$ tel que $\forall t \geq \widehat{t}$, $\beta^{[t]} = \widehat{\beta} \geq \beta$. Ce qui implique que $\exists \widehat{t}$ tel que

$$\forall t \geq \widehat{t}, \quad \|\mathbf{P}^{[t]}\| \leq \frac{2}{\widehat{\beta}} \leq \frac{2}{\beta}.$$

De ce fait, $\forall t \geq \widehat{t}$, les conditions requises par le théorème 1.3.3 sont respectées. Par conséquent, la suite $(\boldsymbol{x}^{[t]})_{t \in N}$ converge vers une solution de (4.29).

□

Dans cette section, j'ai choisi de prendre un préconditionneur diagonal. Un tel choix est motivé par le fait que dans le cas où $\mathbf{P}^{[t]}$ n'est pas diagonal, l'expression de $\text{prox}_{\mathcal{G}}^{\mathbf{P}^{[i]}}$ n'est plus explicite. J'ai utilisé un préconditionneur diagonal similaire à celui proposé par [Lorenz and Pock, 2015, Lemme 10], dans le cas spécifique à Forward-Backward, c'est-à-dire où les termes différentiables duaux et les opérateurs linéaires primaux et duaux sont nuls. Soit $\gamma \in]0, 2[$ et soit une matrice diagonale $\mathbf{B} \in \mathcal{M}_L(\mathbb{R})$, d'éléments diagonaux $(\mathbf{B}_\ell)_{\ell \in \{1,\dots L\}}$, telle que $\nabla \mathcal{F}$ soit co-coercitive vis-à-vis de $\mathbf{B}^{-1}$. On choisit alors $\mathbf{P}^{[t]} = \gamma \mathbf{B}^{-1}$. Dans notre cas, une telle matrice correspond alors à la matrice où $\forall \ell \in \{1, \dots L\}$, $\mathbf{B}_\ell = \beta$ avec β la constante de Lipschitz de $\Phi + \mathcal{R}$. Le préconditionneur choisi correspond alors à $\mathbf{P}^{[t]} = \gamma / \beta^{[t]} \mathbf{Id}$, où $\beta^{[t]}$ est obtenu par backtracking.

Grâce à l'approximation quadratique, il est également possible pour chaque valeur $\beta^{[t]}$ donnée, de connaître le pas optimal γ. En effet il s'agit du pas minimisant l'approximation quadratique, qui est obtenue pour $\gamma = 1$.

Remarque : Pour accélérer la vitesse de convergence, une idée intéressante est de relâcher la contrainte de croissance sur la suite $(\beta^{[t]})_{t \in N}$, c'est-à-dire de permettre à chaque itération que $\beta^{[t+1]} < \beta[t]$ si l'itération courante se situe bien en dessous de la quadratique. Cependant j'ai remarqué lors des implémentations de la méthode que, lorsque la contrainte épigraphique est prise en compte, si la contrainte est relaxée, l'algorithme diverge très rapidement.

Dans notre cas, le backtracking est particulièrement intéressant pour accélérer la vitesse de convergence lorsque le paramètre μ de la régularisation TV-h est choisi très petit. En effet, on a $\mathcal{F}(\boldsymbol{x}) = \Phi(\boldsymbol{x}) + \mathcal{R}(\boldsymbol{x})$. De ce fait, soit β_Φ la constante de Lipshcitz de $\nabla\Phi$, la constante de Lipschitz β de $\nabla \mathcal{F}$ vérifie $\beta \leq \beta_\Phi + \lambda/\mu$. En effet

$$\forall \boldsymbol{y} \in \mathbb{R}^2 \nabla^2 \mathcal{R}(\boldsymbol{y}) = \lambda \frac{\mu^2 - \|\boldsymbol{y}\|^2}{(\|\boldsymbol{y}\|^2 + \mu^2)^{3/2}} \leq \frac{\lambda}{\mu}. \tag{4.33}$$

Donc si $\mu \to 0$, $\beta \to +\infty$. Or, sans backtracking, le pas de descente de l'algorithme VMFBwB est proportionnel à β^{-1}, donc si on choisit μ trop petit le pas de descente va être très petit et l'algorithme va converger plus lentement. Le fait que la vitesse de convergence dépende de μ est problématique, car on souhaiterait pouvoir régler ce paramètre aussi grand ou aussi petit qu'il doive l'être afin d'avoir une qualité de résultats optimal.

Le backtracking peut alors aider à converger plus vite car la valeur de $\beta^{[t]}$ calculée n'augmente que lorsque l'itération associé se trouve au-dessus de la quadratique. Comme la valeur de μ influe surtout sur la forme du critère autour du minimum, c'est-à-dire qu'il est plus « arrondi »lorsque μ est grand et plus «pointu »lorsque μ est petit (cf. figure 1.6), les premiers

itérés sont alors en dessous de l'approximation quadratique du critère, même pour les petites valeurs de $\beta^{[t]}$, ce qui permet de faire des pas plus grands au début de la descente.

La figure 4.1 illustre ce comportement sur données polarimétriques simulées. Dans le cas de FB sans backtracking, lorsque μ est petit, l'algorithme met 10 fois plus de temps à converger. En revanche, dans le cas avec backtracking, l'algorithme converge toujours à la même vitesse, peu importe la valeur de μ. On voit de plus que le backtracking permet bien de converger vers la même solution que sans backtracking.

4.2.3 Résolution par approche non-différentiable

Dans le cas où la contrainte $\mathcal{R}(\boldsymbol{x}) = \lambda g(\mathbf{G}\boldsymbol{x})$ n'est plus différentiable et où $\mathbf{G} \neq \mathbf{Id}$, l'utilisation du schéma Forward-Backward est plus compliquée. En effet, si une forme explicite de l'opérateur proximal de g existe, ce n'est pas le cas pour sa composition avec l'opérateur linéaire $g \circ \mathbf{G}$. Dans de telles situations, la pratique usuelle est de résoudre le problème dual associé du problème 1.32, donné par :

Rappel éq. (1.90)

$$\widehat{\boldsymbol{y}} \in \underset{\boldsymbol{y}}{\operatorname{Argmin}} \left\{ [(\iota_{\mathcal{C}} + \Phi)^*(-\mathbf{G}^*\boldsymbol{y}) + g^*(\boldsymbol{y}) \right\},$$

où * représente le dual des fonctions et l'adjoint de l'opérateur linéaire. Cependant, ce n'est pas possible ici, car la fonction duale de Φ fait intervenir les inverses des opérateurs,$(\mathbf{T}_{j,k})^{-1}_{k\in\{1,\dots,K\},j=1,2}$ et $\mathbf{A} \in \mathcal{M}_L(\mathbb{R})$, qui n'existent pas dans notre cas. C'est pour cette raison que j'ai fait le choix

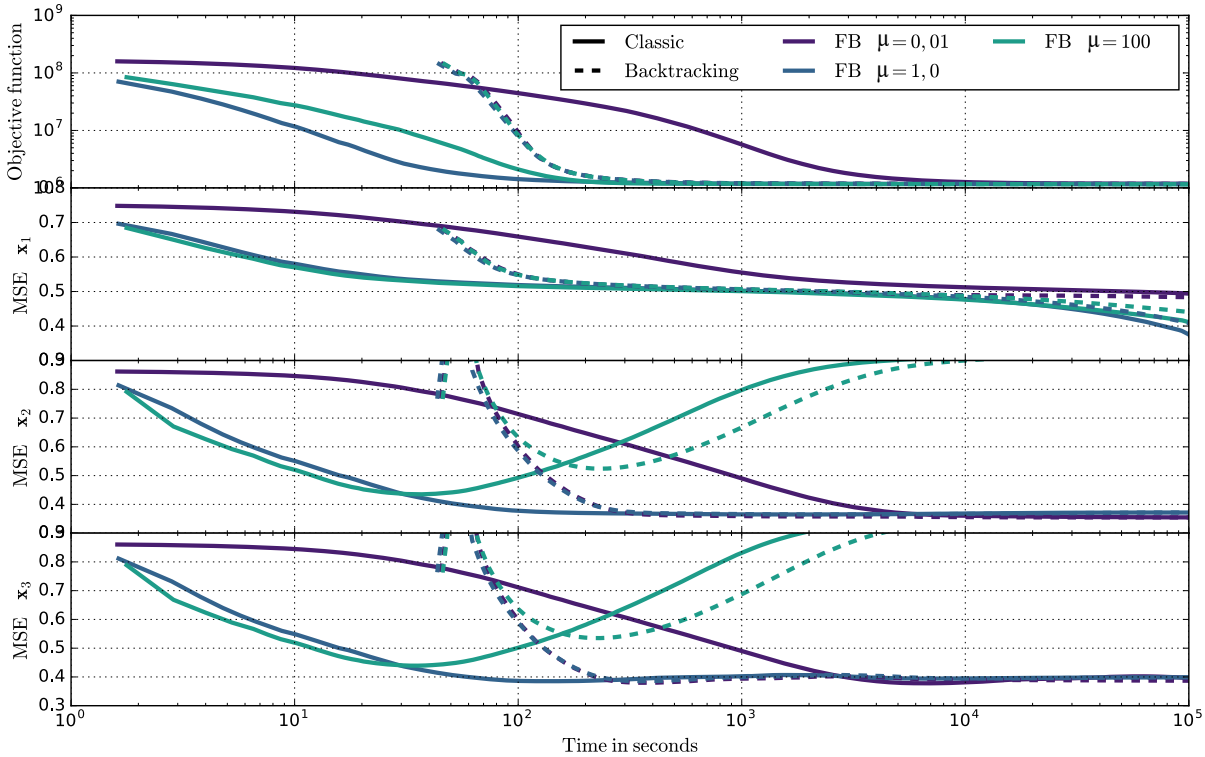


FIGURE 4.1 – Comparaison de la vitesse de convergence du critère et de l'EQMn des paramètres $\boldsymbol{x}_1$ = I, $\boldsymbol{x}_2$ = Q et $\boldsymbol{x}_3$ = U, pour l'algorithme VMFB avec $\forall\ t \geq 0$, $\mathbf{P}^{[t]} = 2/\beta$, avec et sans backtracking, pour différentes valeurs de μ.

d'une instance de l'algorithme primal-dual à métrique variable de Condat-Vũ [Condat, 2013, Vũ, 2015] (VMCV) présenté dans la section 1.3.4, dans le cas particulier ou le terme non-différentiable primal est $\iota_{\mathcal{C}}$, dont les itérations sont données par l'algorithme 4.

Afin de s'affranchir du calcul de la constante de Lipschitz, comme pour l'algorithme 7, j'introduis une étape de backtracking dans l'algorithme 4, ce qui conduit aux itérations données par l'algorithme 8. Les garanties de convergence de VFCVwB reposent sur la proposition 4.2.3

Algorithme 8 : Variable Metric Condat-Vũ algorithm with Backtracking (VMCVwB)

Soit $\boldsymbol{x}^{[0]} \in (\mathbb{R}^N)^L$, $\boldsymbol{y}^{[0]} \in \mathbb{R}^{K_1} \times \ldots \times \mathbb{R}^{K_L}$, $\beta^{[0]} \geq 0$ et $\eta > 1$:
pour $t = 0, 1, \ldots$ **faire**
  pour $i = 0, 1, \ldots$ **faire**
    $\mathring{\beta}^{[i]} = \eta^i \beta^{[t]}$ et on définit $\mathring{\mathbf{U}}^{[i]} \geq 0$ et $\mathring{\sigma}^{[i]} \geq 0$ tels que la condition (4.34) soit respectée.
    $\mathring{\boldsymbol{x}}^{[i]} = \operatorname{prox}_{\iota_{\mathcal{C}}}^{\mathring{\mathbf{U}}^{[i]}}\left(\boldsymbol{x}^{[t]} - \mathbf{U}^{[i]}\left(\nabla\Phi(\boldsymbol{x}^{[t]}) + \mathbf{G}^*\boldsymbol{y}^{[t]}\right)\right)$;
    $\mathring{\boldsymbol{y}}^{[i]} = \operatorname{prox}_{\mathring{\sigma}^{[i]} g^*}\left(\boldsymbol{y}^{[t]} + \mathring{\sigma}^{[i]}\mathbf{G}\left(2\mathring{\boldsymbol{x}}[i] - \boldsymbol{x}^{[t]}\right)\right)$;
    si $\Phi(\mathring{\boldsymbol{x}}^{[i]}) \leq \Phi(\boldsymbol{x}^{[t]}) + \langle \mathring{\boldsymbol{x}}^{[i]} - \boldsymbol{x}^{[t]}, \nabla\Phi(\boldsymbol{x}^{[t]})\rangle + \frac{\mathring{\beta}^{[i]}}{2}\|\mathring{\boldsymbol{x}}^{[i]} - \boldsymbol{x}^{[t]}\|^2$ **alors**
      $\beta^{[t+1]} = \mathring{\beta}^{[i]}$,
      $\boldsymbol{x}^{[t+1]} = \mathring{\boldsymbol{x}}^{[i]}$,
      $\boldsymbol{y}^{[t+1]} = \mathring{\boldsymbol{y}}^{[i]}$, **fin pour**

suivante, qui s'appuie sur le théorème 4.2.1.

Proposition 4.2.3. *Soit $\Phi, g, \iota_{\mathcal{C}}$ des fonctions convexes, semi-continues inférieurement et propres telles que Φ est β-Lipschitz et différentiable avec $\beta > 0$. Soit $\beta^{[0]} \geq 0$, $\eta > 1$ et $(\beta^{[t]})_{t\in\mathbb{N}}$ une suite croissante telle que $\beta^{[t+1]} = \eta^i\beta^{[t]}$ où i est obtenu selon l'algorithme 8. Soit $\mathbf{U}^{[t]}$ une matrice symétrique définie positive et $\sigma_t > 0$, tel que $\mathbf{U}^{[t+1]} \geq \mathbf{U}^{[t]}$ et $\sigma^{[t+1]} \leq \sigma^{[t]}$. Alors, si*

$$\frac{1 - \|\sqrt{\sigma^{[t]}}\mathbf{G}\sqrt{\mathbf{U}^{[t]}}\|}{\max\{\|\mathbf{U}^{[t]}\|, \sigma^{[t]}\}} \geq \frac{\beta^{[t]}}{2}, \tag{4.34}$$

la suite $(\boldsymbol{x}^{[t]})_{t>0}$ de l'algorithme 8 converge vers $\widehat{\boldsymbol{x}}$, une solution de (1.32).

Remarque : Lors de l'étape de backtracking, on vérifie à chaque itéré, sa position vis-à-vis de l'approximation quadratique du critère. Dans le cas primal-dual, il s'agit de l'approximation quadratique de l'équation :

Rappel éq. (1.91) $$\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}} \in \operatorname*{Argmin}_{\boldsymbol{x}\in\mathbb{R}^L} \max_{\boldsymbol{y}\in\mathbb{R}^G}\left\{(\iota_{\mathcal{C}} + \Phi)(\boldsymbol{x}) - g^*(\boldsymbol{y}) + \langle \mathbf{G}\boldsymbol{x}, \boldsymbol{y}\rangle\right\},$$

qui est équivalente à l'approximation quadratique de $\Phi(\boldsymbol{x})$ et de $\Phi(\boldsymbol{x}) + \langle \mathbf{G}\boldsymbol{x}, \boldsymbol{y}\rangle$.

Le choix des matrices $(\mathbf{U}^{[t]})_{t>0}$ fut longuement étudié dans cette thèse et plusieurs pistes furent explorées. En effet, j'ai implémenté cette méthode pour différentes pénalisations non-différentiables : TV, Shatten sur le Hessien, mais aussi TGV et TV+ ℓ_1. Le temps de convergence de l'algorithme étant long, c'est-à-dire de plusieurs heures à plusieurs jours selon la taille des

données, la précision recherchée et la régularisation utilisée, il fallut trouver des stratégies d'accélération afin de pouvoir comparer les régularisations avec des paramètres optimaux. Trois configurations ont été implémentées :

— Sans préconditionnement : c'est-à-dire $\mathbf{U}^{[t]} = \tau^{[t]}\mathbf{Id}$ tels que $\tau^{[t]}$ et $\sigma^{[t]}$ vérifient la condition de convergence présentée par Condat dans [Condat, 2013], qui est donnée par

$$\frac{1}{\tau^{[t]}} - \sigma^{[t]}\|\mathbf{G}\|^2 \geq \frac{\beta^{[t]}}{2}. \tag{4.35}$$

Une telle contrainte est moins restrictive sur le choix des hyperparamètres de convergence $\tau^{[t]}$ et $\sigma^{[t]}$ que la contrainte dont la formule est(1.92), et permet un choix d'hyperparamètres de convergence qui ne vérifient pas (1.92). La différence entre les deux conditions est que la contrainte de Condat, dans la preuve de la proposition 1.3.4, n'assure seulement que $\mathbf{M}^{[t]}$ est définie positive par rapport aux variables primales.

— Avec préconditionneur diagonal : inspiré par le préconditionneur diagonal proposé dans [Lorenz and Pock, 2015, Lemme 10], tout comme le préconditionneur diagonal utilisé pour l'algorithme Forward-Backward, on pose la matrice diagonale $\mathbf{B} \in \mathcal{M}_L(\mathbb{R})$, d'éléments diagonaux $(\mathbf{B}_\ell)_{\ell\in\{1,\dots L\}}$, telle que $\forall \ell \in \{1,\dots L\}$, $\mathbf{B}_\ell = \beta$. Le préconditionneur diagonal est alors donné pour $\mu > 0$, $\gamma \in]0,2[$ et $s \in [0,2]$ par $\mathbf{U}^{[t]} = \tau^{[t]}\mathbf{Id}$:

$$\tau^{[t]} = \frac{1}{\frac{\beta^{[t]}}{\gamma} + \mu\|\mathbf{G}\|^{2-s}} \quad \text{et} \quad \sigma^{[t]} = \frac{\mu}{\|\mathbf{G}\|^s}. \tag{4.36}$$

Un tel choix de préconditionneur vérifie la condition (1.92) qui est suffisante à garantir la convergence de l'algorithme 8.

— Avec préconditionnement par l'inverse de l'approximation du Hessien de Φ : c'est-à-dire un préconditioneur non diagonal de la forme $\mathbf{U}^{[t]} = (\nabla^2\Phi + \varepsilon^{[t]}\mathbf{Id})^{-1}$, noté dans la suite $\mathbf{U}(\varepsilon^{[t]})$ et inspiré par le précondtionneur proposé dans [Li and Zhang, 2016]. Cependant pour un tel choix de préconditionneur non diagonal, il n'existe pas de formulation explicite de l'opérateur proximal de $\iota_{\mathcal{C}}$, car celui-ci n'est plus séparable en chaque pixel. C'est pourquoi j'ai choisi de résoudre le problème avec $\iota_{\mathcal{C}}$ dans le dual. Nous résolvons donc le problème (1.32) avec l'algorithme 9.
Les conditions nécessaires et suffisantes à la convergence de l'algorithme 4 mènent à la condition suivante sur la convergence de l'algorithme 9 sans étape de backtracking :

Proposition 4.2.4. *On suppose que* $\lambda_{\min}(\nabla^2\Phi) = 0$ *et* $\mathbf{U}(\varepsilon^{[t]}) = \left(\nabla^2\Phi + \varepsilon^{[t]}\mathbf{Id}\right)^{-1}$. *Soit* $r > 0$ *et* $\sigma^{[t]} = r(\varepsilon^{[t]})^{-1}$, *et*

$$\varepsilon^{[t]} \geq \max\{1, r\}\frac{\beta}{2} + \sqrt{r}\|\mathbf{G}\|, \tag{4.37}$$

alors, la suite $(\boldsymbol{x}^{[t]})_{t\in\mathbb{N}}$ *générée par l'Algorithme 8 converge vers* $\overline{\overline{\boldsymbol{x}}}$, *une solution de* (1.32).

Démonstration : *(Fin page 113)* Comme vue dans la preuve de convergence de l'algorithme 4, l'algorithme 9 peut être vu comme une instance particulière de l'algoritme Forward-Backward :

$$\boldsymbol{u}^{[t+1]} = \operatorname{prox}_{\partial g}^{\mathbf{M}^{[t]^{-1}}}\left(\boldsymbol{u}^{[t]} - \mathbf{M}^{[t]^{-1}}\nabla\mathcal{F}(\boldsymbol{u}^{[t]})\right) \tag{4.38}$$

Algorithme 9 : VMPDwB

Soit $\boldsymbol{x}^{[0]} \in (\mathbb{R}^N)^L$, $\boldsymbol{y}^{[0]} \in \mathbb{R}^{K_1} \times \ldots \times \mathbb{R}^{K_L}$, $\boldsymbol{y}^{[0]} \in (\mathbb{R}^N)^L$, $\beta^{[0]} \geq 0$ et $\eta > 1$:
pour $t = 0, 1, \ldots$ **faire**
  pour $i = 0, 1, \ldots$ **faire**
    $\mathring{\beta}^{[i]} = \eta^i \beta^{[t]}$ et soit $\mathring{\varepsilon}^{[i]} \geq 0$ et $\mathring{\sigma}^{[i]} \geq 0$ tel que la condition (4.37) soit respectée.
    $\mathring{\boldsymbol{x}}^{[i]} = \boldsymbol{x}^{[t]} - \mathbf{U}(\mathring{\varepsilon}^{[i]})\left(\nabla\Phi(\boldsymbol{x}^{[t]}) + \mathbf{G}^*\boldsymbol{y}^{[t]} + \boldsymbol{z}^{[t]}\right)$;
    $\mathring{\boldsymbol{y}}^{[i]} = \mathrm{prox}_{\mathring{\sigma}^{[i]} g^*}\left(\boldsymbol{y}^{[t]} + \mathring{\sigma}^{[i]}\mathbf{G}\left(2\mathring{\boldsymbol{x}}[i] - \boldsymbol{x}^{[t]}\right)\right)$;
    $\mathring{\boldsymbol{z}}^{[i]} = \mathrm{prox}_{\mathring{\sigma}^{[i]} \iota_C^*}\left(\boldsymbol{z}^{[t]} + \mathring{\sigma}^{[i]}\left(2\mathring{\boldsymbol{x}}^{[i]} - \boldsymbol{x}^{[t]}\right)\right)$;
    si $\Phi(\mathring{\boldsymbol{x}}^{[i]}) \geq \Phi(\boldsymbol{x}^{[t]}) + \langle \mathring{\boldsymbol{x}}^{[i]} - \boldsymbol{x}^{[t]}, \nabla\Phi(\boldsymbol{x}^{[t]})\rangle + \frac{\mathring{\beta}^{[i]}}{2}\|\mathring{\boldsymbol{x}}^{[i]} - \boldsymbol{x}^{[t]}\|^2$ **alors**
      $\beta^{[t+1]} = \mathring{\beta}^{[i]}$,
      $\boldsymbol{x}^{[t+1]} = \mathring{\boldsymbol{x}}^{[i]}$,
      $\boldsymbol{y}^{[t+1]} = \mathring{\boldsymbol{y}}^{[i]}$,
      $\boldsymbol{z}^{[t+1]} = \mathring{\boldsymbol{z}}^{[i]}$.
      fin pour

où :

$$\boldsymbol{u}^{[t]} = \begin{pmatrix} \boldsymbol{x}^{[t]} \\ \boldsymbol{y}^{[t]} \\ \boldsymbol{z}^{[t]} \end{pmatrix}, \qquad \partial\boldsymbol{g} = \begin{pmatrix} 0 & \mathbf{G}^* & \mathbf{Id} \\ -\mathbf{G} & \partial g & 0 \\ -\mathbf{Id} & 0 & \partial\iota_C^* \end{pmatrix},$$

$$\nabla\boldsymbol{\mathcal{F}} = \begin{pmatrix} \nabla\Phi & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{et} \quad \mathbf{M}^{[t]} = \begin{pmatrix} \mathbf{U}(\varepsilon^{[t]})^{-1} & -\mathbf{G}^* & -\mathbf{Id} \\ -\mathbf{G} & \sigma^{[t]^{-1}} & 0 \\ -\mathbf{Id} & 0 & \sigma^{[t]^{-1}} \end{pmatrix}, \tag{4.39}$$

Alors, si $\mathbf{M}^{[t]}$ est choisie telle que le Théoreme 1.3.3 soit respecté, alors la suite d'itérés $(\boldsymbol{x}^{[t]})_{t\in\mathbb{N}}$ converge vers une solution $\overline{\overline{\boldsymbol{x}}}$ de (1.32). Les conditions sur $\varepsilon^{[t]}$ sont déduites de la condition (1.92).
Comme on suppose $\lambda_{\min}(\nabla^2\Phi) = 0$, on a lors $\|\mathbf{U}(\varepsilon^{[t]})\| = \varepsilon^{-1}$. De ce fait la condition (1.92) pour un tel choix de matrice de préconditionnement implique que :

$$\min\left\{\varepsilon^{[t]}\left(1 - \varepsilon^{[t]^{-\frac{1}{2}}}\|\mathbf{G}\|\sigma^{[t]^{\frac{1}{2}}}\right), \sigma^{[t]^{-1}}\left(1 - \varepsilon^{[t]^{-\frac{1}{2}}}\|\mathbf{G}\|\sigma^{[t]^{\frac{1}{2}}}\right)\right\} \geq \frac{\beta}{2}, \tag{4.40}$$

$$\Leftrightarrow \min\left\{\varepsilon^{[t]}, \sigma^{[t]^{-1}}\right\}\left(1 - \varepsilon^{[t]^{-\frac{1}{2}}}\|\mathbf{G}\|\sigma^{[t]^{\frac{1}{2}}}\right) \geq \frac{\beta}{2}, \tag{4.41}$$

$$\tag{4.42}$$

Posons $\sigma^{[t]} = r\varepsilon^{[t]^{-1}}$, on a alors :

$$\varepsilon^{[t]}\min\left\{1, r^{-1}\right\}\left(1 - \varepsilon^{[t]^{-1}}\|\mathbf{G}\|r^{\frac{1}{2}}\right) \geq \frac{\beta}{2}, \tag{4.43}$$

$$\Leftrightarrow \min\left\{1, r^{-1}\right\}\left(\varepsilon^{[t]} - \|\mathbf{G}\|r^{\frac{1}{2}}\right) \geq \frac{\beta}{2}, \tag{4.44}$$

$$\Leftrightarrow \varepsilon^{[t]} \geq \frac{\beta}{2}\max\{1, r\} + \sqrt{r}\|\mathbf{G}\|. \tag{4.45}$$

Ceci conclue la preuve de la proposition 4.2.4. □

La proposition 4.2.4 implique que

$$\forall r, t > 0, \quad \varepsilon^{[t]} > \beta/2,$$

où β est la constante de Lipschitz de $\nabla\Phi$. De plus, comme $\mathbf{U}(\varepsilon^{[t]})$ est symétrique définie positive, on a :

$$\|\mathbf{U}(\varepsilon^{[t]})\| = \frac{1}{\lambda_{\min}(\nabla^2 h + \varepsilon \mathbf{Id})} \leq \varepsilon^{-1} \leq \frac{2}{\beta}.$$

Cela implique que le pas de descente est limité par $2/\beta$, ce qui résulte en un plus grand nombre d'itérations lorsque β augmente.

Dans la suite de ce manuscrit je n'utilise que le préconditionement diagonal $\mathbf{U}(\tau^{[t]}) = \tau^{[t]}\mathbf{Id}$, car nous avons observé qu'en pratique, le préconditionneur $\mathbf{U}(\varepsilon^{[t]})$ n'accélère finalement pas la méthode. En effet, sur un exemple «jouet »nous avons fait converger l'algorithme pour les deux configurations de préconditionnement en fixant les mêmes paramètres $\sigma^{[t]} = \sigma$ et $\|\mathbf{U}(\tau^{[t]})\| = \|\mathbf{U}(\varepsilon^{[t]})\| = \tau = \varepsilon^{-1}$, de telle sorte qu'au moins la condition (4.35) soit respectée. La figure 4.2, présente les résultats obtenus. On voit que le comportement des courbes de convergences est exactement le même pour les deux choix de matrice de préconditionnement.

4.2.4 Comparaison des performances sur données simulées

Nous comparons maintenant les performances des algorithmes VMLMB pour la régularisation par TV-h, VMFBwB pour la régularisation TV-h avec Projection Épigraphique (PE), et VMCVwB pour les régularisations TV et Shatten avec PE, sur données simulées de la manière présentée dans la section 3.1.3. Nous comparons les performances en terme de nombre d'itérations ainsi que de temps de calcul nécessaires pour arriver à la convergence du critère et de l'Erreur Quadratique Moyenne normalisée des paramètres I^{u}, I^{p} et Θ, en nombre d'itérations, pour le taux de convergence, et en temps de calcul en secondes pour la vitesse. Nous nous plaçons dans le cas fixe de polarisation du disque de $\tau^{\mathrm{disk}} = 10\%$.

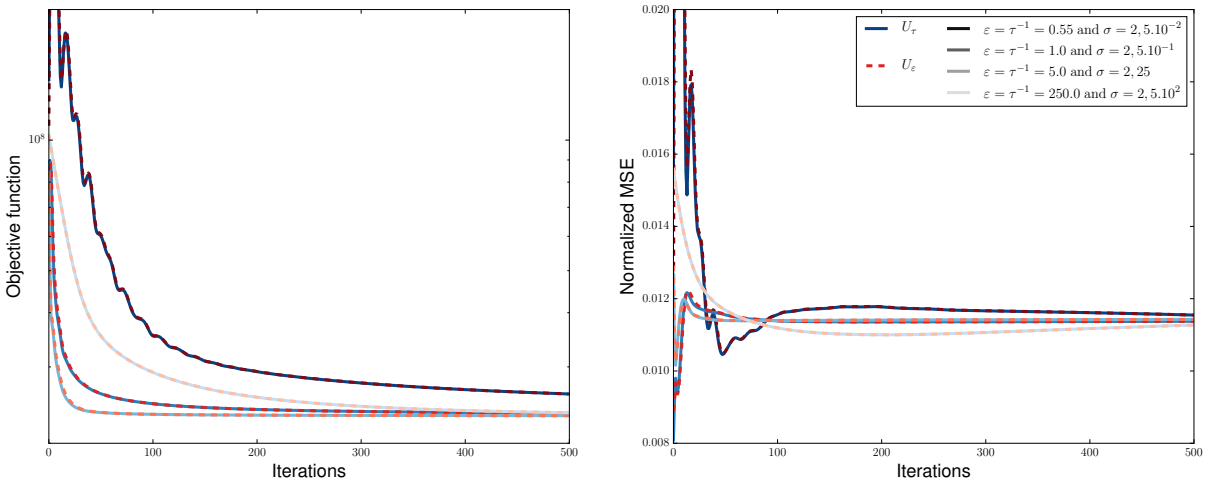


FIGURE 4.2 – Convergence de la fonction objectif et de l'Erreur Quadratique Moyenne en fonction des itérations pour le préconditionneur diagonal et par l'inverse de l'approximation de la hessienne.

Nous comparons les performances de VMLM et VMFBwB avec TV-h pour $\mu = 0,1$, avec l'algorithme VMCVwB pour les régularisations TV et Shatten sur l'opérateur Hessien.

La figure 4.3 montre l'évolution du critère, d'une part en nombre d'itérations et d'autre part en temps (mesuré en seconde). Nous étudions les configurations où les hyperparamètres $(\lambda_\ell)_{\ell\in\{1,\dots L\}}$ donnent la meilleure valeur du critère SURE et d'autre part donnent la meilleure valeur de l'EQMn. Le temps en secondes nécessaire à la convergence est relatif à la machine utilisée. Le temps indiqué ici permet donc uniquement d'avoir un ordre de grandeur du temps de calcul des méthodes et du rapport de temps entre les méthodes, dans le cas de leur résolution sur la même machine. On voit que les méthodes proximales ont une allure de convergence équivalente. Cependant, le temps de convergence en secondes pour Shatten est quasiment doublé par rapport aux autre méthodes proximales. La méthode différentiable VMLM permet de gagner un facteur 10 au niveau du taux de convergence. Au niveau du temps de calcul, selon le choix des hyperparamètres, l'algorithme VMLM converge en entre 5 et 30 minutes, alors que plus d'une heure de calculs est nécessaire aux algorithmes proximaux.

La figure 4.4 montre l'évolution de l'EQM au cours des itérations, d'une part en nombre d'itérations et d'autre part en temps en seconde. Nous étudions les configurations où les hyperparamètres $(\lambda_\ell)_{\ell\in\{1,\dots L\}}$ donnent la meilleure valeur du critère SURE et d'autre part donnent la meilleure valeur de l'EQMn. On voit alors que si le critère semble avoir convergé pour VMLMB, ce n'est pas le cas pour l'Erreur Quadratique Moyenne. On voit que dans le cas du meilleur critère SURE, l'erreur a du mal à se stabiliser, tandis que dans le cas de la meilleure EQM des paramètres, elle converge en 5 minutes pour I^{p} et θ.

Étant donné le temps de convergence des méthodes proximales, cela peut être gênant pour la recherche des hyperparamètres optimaux, car la lenteur de l'algorithme oblige à couper la

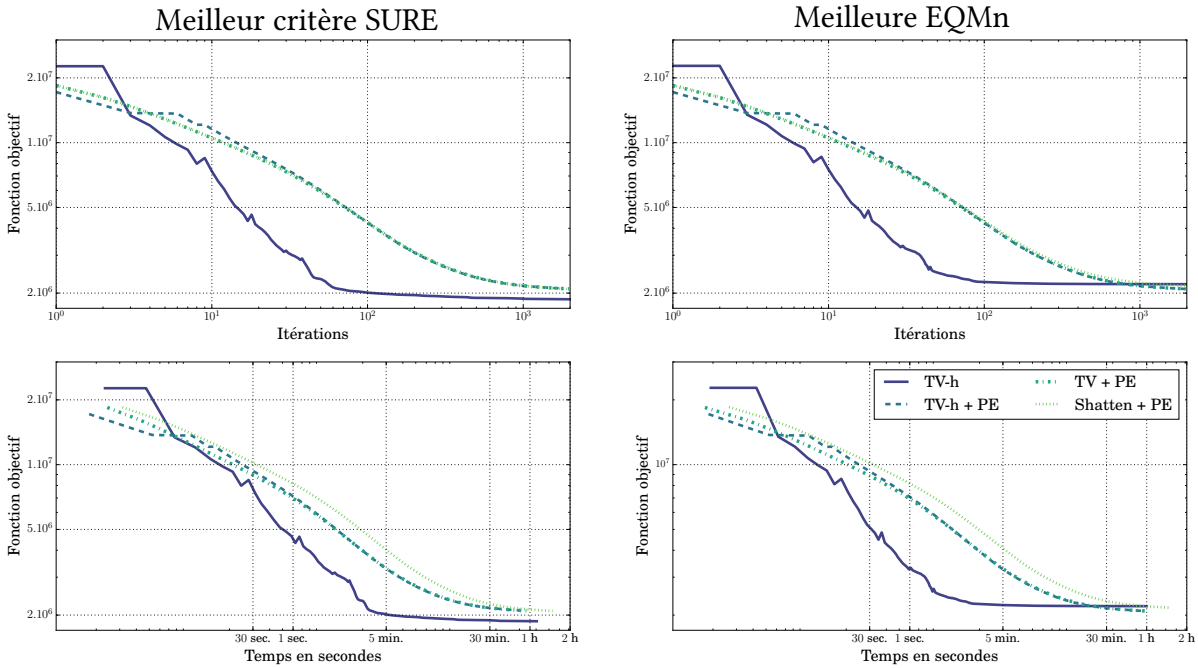


Figure 4.3 – Comparaison du taux de convergence du critère en itérations et de la vitesse de convergence du critère en secondes, des algorithmes VMLM, pour la reconstruction avec TV-h, VMFBwB, pour la reconstruction avec TV-h + PE, et VMCVwB pour TV +PE et Shatten + PE.

convergence à un certain nombre d'itérations pour s'assurer de trouver les hyperparamètres en un temps réaliste. De ce fait, il est possible que le couple d'hyperparamètres associés soit le meilleur pour cet itéré, mais qu'avec une convergence un peu plus longue, un autre couple donne un critère SURE ou une EQMn plus faible.

C'est pourquoi nous avons fixé dans la suite le temps de résolution comme critère décisif dans le choix des méthodes. Par exemple, pour un jeu de données de taille similaire aux don-

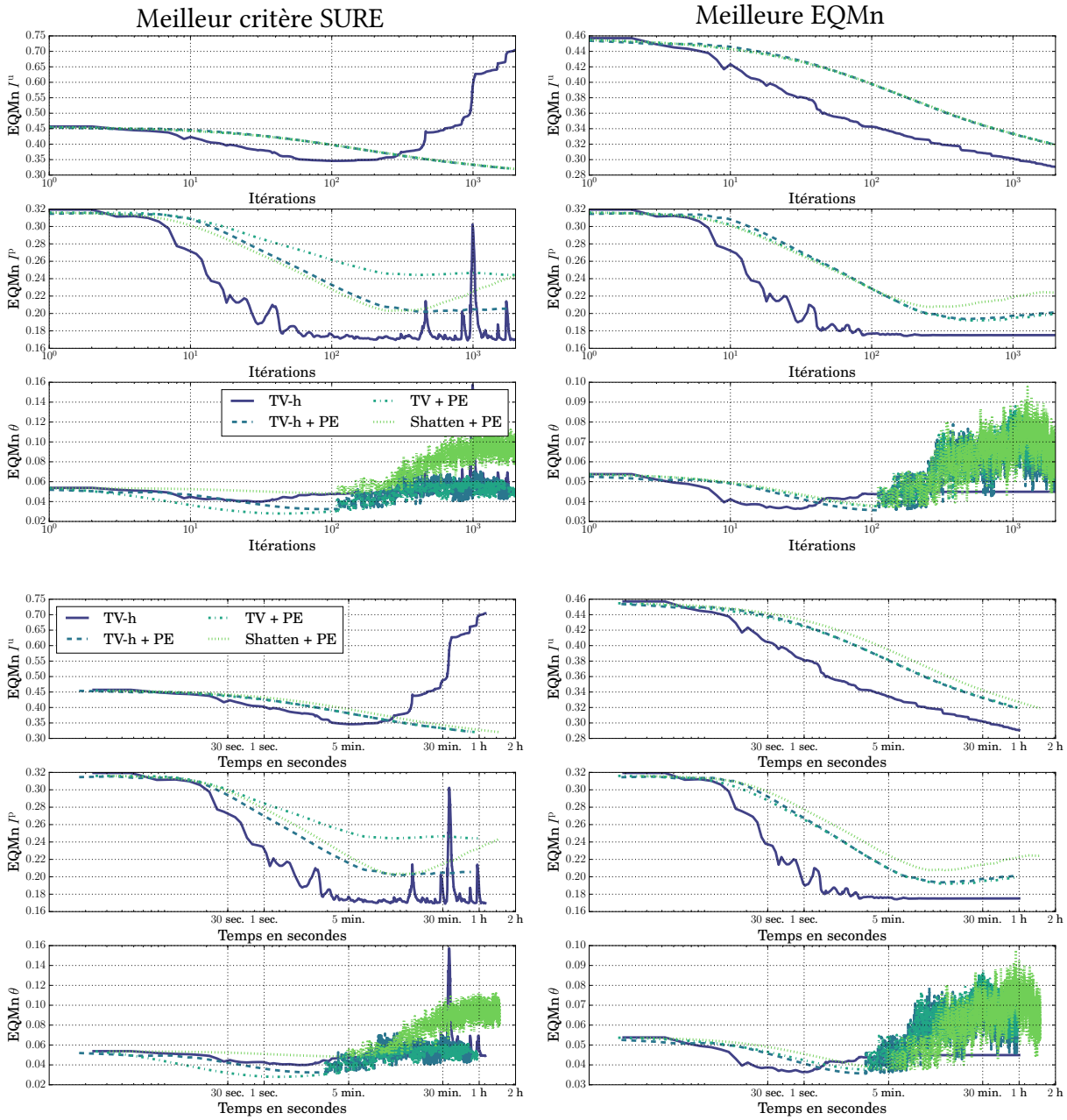


FIGURE 4.4 – Comparaison du taux de convergence de l'EQMn en itérations et de la vitesse de convergence de l'EQMn en secondes, des algorithmes VMLM, pour la reconstruction avec TV-h, VMFBwB, pour la reconstruction avec TV-h + PE, et VMCVwB pour TV +PE et Shatten + PE.

nées simulées ($128 \times 128 \times 64$), nous fixons le temps maximal de résolution pour toutes les méthodes à 1 heure.

4.3 Applications sur données simulées et données astrophysiques

On se propose maintenant de comparer de manière qualitative les résultats obtenus, sous une contrainte de temps, avec les différentes méthodes sur données simulées et sur données astrophysiques. On compare dans un premier temps les résultats obtenus par les Méthodes Linéaires non-Séparables avec Déconvolution (MLnS-D) présentées dans ce chapitre à partir de données *calibrées*, pour les différentes régularisations, avec et sans contrainte épigraphique dans le cas de TV-h. On compare dans un second temps les résultats obtenus par la MLnS-D aux résultats obtenus par la Méthode non-Linéaire Séparable (MnLS), présentée dans la section 2.2 et le Double Ratio sur données *pré-traitées*, déconvolués *a posteriori*.

Comme dans la section précédente, les reconstructions sont faites pour un choix d'hyperparamètres minimisant d'une part la somme des EQMn dans le domaine des paramètres, et d'autre part le critère SURE. Comme précédemment, on note d'une part $\widehat{\boldsymbol{x}}^{\text{EQMn-P}}$ le paramètre d'intérêt estimé par l'une ou l'autre des deux méthodes pour un choix d'hyperparamètres minimisant l'EQMn dans le domaine des paramètres. D'autre part, on note $\widehat{\boldsymbol{x}}^{\text{SURE}}$ le paramètre d'intérêt estimé par l'une des deux méthodes, pour un choix d'hyperparamètres minimisant le critère SURE. D'après les cartes de critère SURE et de valeur d'EQMn présentées sur la figure 5.5, pour chacun des cas, les hyperparamètres optimaux sont du même ordre de grandeur pour toutes les régularisations.

Afin d'éviter le calcul des hyperparamètres optimaux pour les méthodes proximales, du fait des temps de calcul important de ces méthodes, on ne calculera alors que les paramètres optimaux minimisant le critère SURE et l'EQMn dans le cas différentiable de TVh sans contrainte épigraphique (PE : Projection Épigraphique). Á titre indicatif, pour un nombre d'itérations réduit à 400 et un jeu de données de taille $128 \times 256 \times 88$, il faut environ une demi-journée pour trouver l'hyperparamètre optimal avec la régularisation Shatten.

4.3.1 Résultats sur données simulées

On applique les MLnS-D sur données simulées pour différents taux de polarisation du disque $\tau^{\text{disk}} \in \{3\%; 7\%; 10\%; 15\%; 25\%; 50\%\}$. Le temps maximal de convergence est fixé à 1h. Dans le cas de l'utilisation de la régularisation TV-h, nous fixons $\mu = 0,1$. Le critère de qualité choisi est, comme dans les chapitres précédents, l'Erreur Quadratique Moyenne normalisée dont on rappelle les expressions pour l'intensité et l'angle :

Rappel éq. (2.24)
$$\text{EQMn}(\widehat{\boldsymbol{x}}_\ell, \overline{\boldsymbol{x}}_\ell) = \frac{\sum_{n \in \widetilde{N}_\ell} (\widehat{x}_{\ell,n} - \overline{x}_{\ell,n})^2}{\sum_{n=1}^{N} \overline{x}_{\ell,n}^2},$$

et

Rappel éq. (2.25)
$$\text{EQMn}(\widehat{\theta}, \overline{\theta}) = \sum_{n \in \widetilde{N}_3} \left(\frac{\arg(e^{2i(\widehat{\theta}_n - \overline{\theta}_n)})}{2\widetilde{N}_3} \right)^2.$$

où $\widetilde{N}_\ell$ est l'ensemble des pixels valides ou représentatifs du paramètre estimé. Pour le paramètre I^{u}, tous les pixels sont considérés valides. Pour les paramètres I^{p} et θ, on réduit l'ensemble des pixels représentatifs aux anneaux composant le disque. Une bonne valeur de l'EQMn sera proche de 0.

La figure 4.5 montre les résultats visuels de la déconvolution dans les cas de polarisation du disque $\tau^{\mathrm{disk}} \in \{3\%; 10\%; 50\%\}$. Tout d'abord, on voit que l'intensité non-polarisée I^{u} est mieux reconstruite dans le cas d'hyperparamètres minimisant l'EQMn. Dans le cas de l'hyperparamètre minimisant le critère SURE, les résultats semblent sous-régularisés. On voit d'autre part que pour chacune des méthodes, les résultats sont assez similaires pour tous les taux de polarisation. On voit enfin que sans contrainte épigraphique, avec l'hyperparamètre minimisant le critère SURE, la reconstruction de I^{u} présente bien plus d'artefacts de reconstruction, dus à la déconvolution de la binaire simulée. En présence de la contrainte épigraphique, il n'est cependant pas évident de comparer les méthodes TV-h, TV et Shatten.

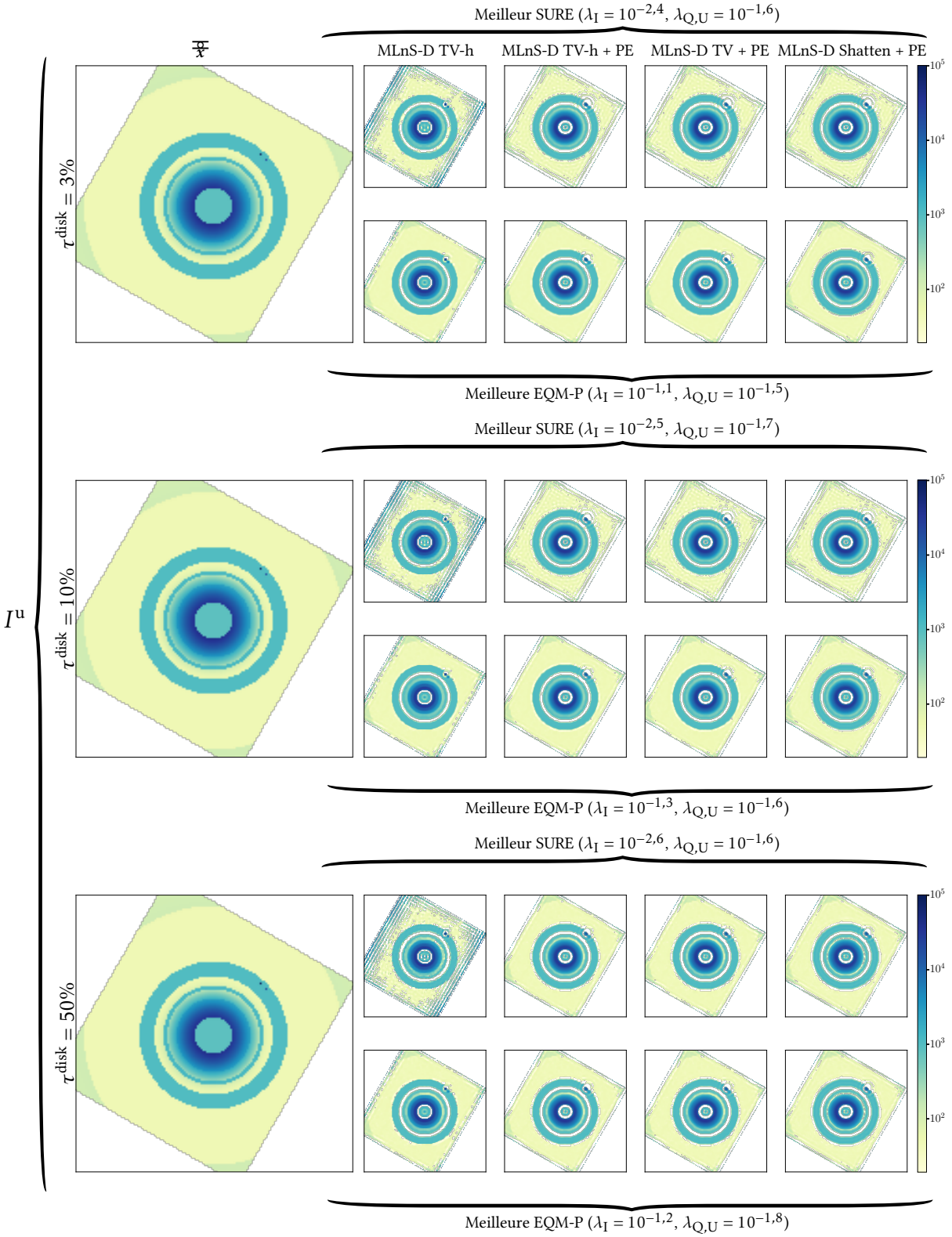
Pour ce qui est de l'intensité polarisée I^{p}, qui est le paramètre qui nous intéresse le plus pour son exploitation astrophysique (avec θ), on observe que, dans le cas de faible taux de polarisation du disque du disque ($\tau^{\mathrm{disk}} \in \{3\%\,;\,10\%\}$), l'utilisation de la contrainte épigraphique induit une erreur là où la binaire est présente dans I^{u}. Dans un tel cas, la reconstruction avec TV-h semble donc la plus performante des méthodes. On remarque aussi que l'intensité de l'anneau fin semble mieux restituée avec Shatten, mais que des artefacts apparaissent entre les deux anneaux. Les bords des anneaux reconstruits avec Shatten sont également plus doux. Pour $\tau^{\mathrm{disk}} = 50\%$, il n'est pas évident de distinguer les méthodes.

Les reconstructions de l'angle de polarisation θ, en présence de la contrainte épigraphique et à faible polarisation du disque, présentent les mêmes artefacts que I^{p}, dus à la déconvolution de la binaire présente dans I^{u}. Pour $\tau^{\mathrm{disk}} = 3\%$, la meilleure reconstruction semble être obtenue avec Shatten, où l'anneau fin semble plus visible. De plus l'effet *«cartoon »* est moins présent, de même pour $\tau^{\mathrm{disk}} = 10\%$. Pour $\tau^{\mathrm{disk}} = 50\%$, il n'est pas évident de comparer l'efficacité des méthodes.

La figure 4.6 représente l'EQMn des différents paramètres estimés dans les différentes configurations de régularisations et contraintes. La figure 4.6a représente l'EQMn des paramètres dans le cas des hyperparamètres minimisant le critère SURE lors de la régularisation avec TV-h. La figure 4.6b représente l'EQMn des paramètres dans le cas des hyperparamètres minimisant le critère SURE lors de la régularisation avec TV-h. On observe que pour chacun des paramètres $\widehat{I^{\mathrm{p}}}^{\text{EQMn-P}}$, $\widehat{I^{\mathrm{p}}}^{\text{SURE}}$, $\widehat{\theta}^{\text{EQMn-P}}$ et $\widehat{\theta}^{\text{SURE}}$, les EQMn sont relativement équivalente. Il en est de même pour $\widehat{I^{\mathrm{u}}}^{\text{EQMn-P}}$, $\widehat{I^{\mathrm{u}}}^{\text{SURE}}$, lors de la reconstruction avec contrainte épigraphique.

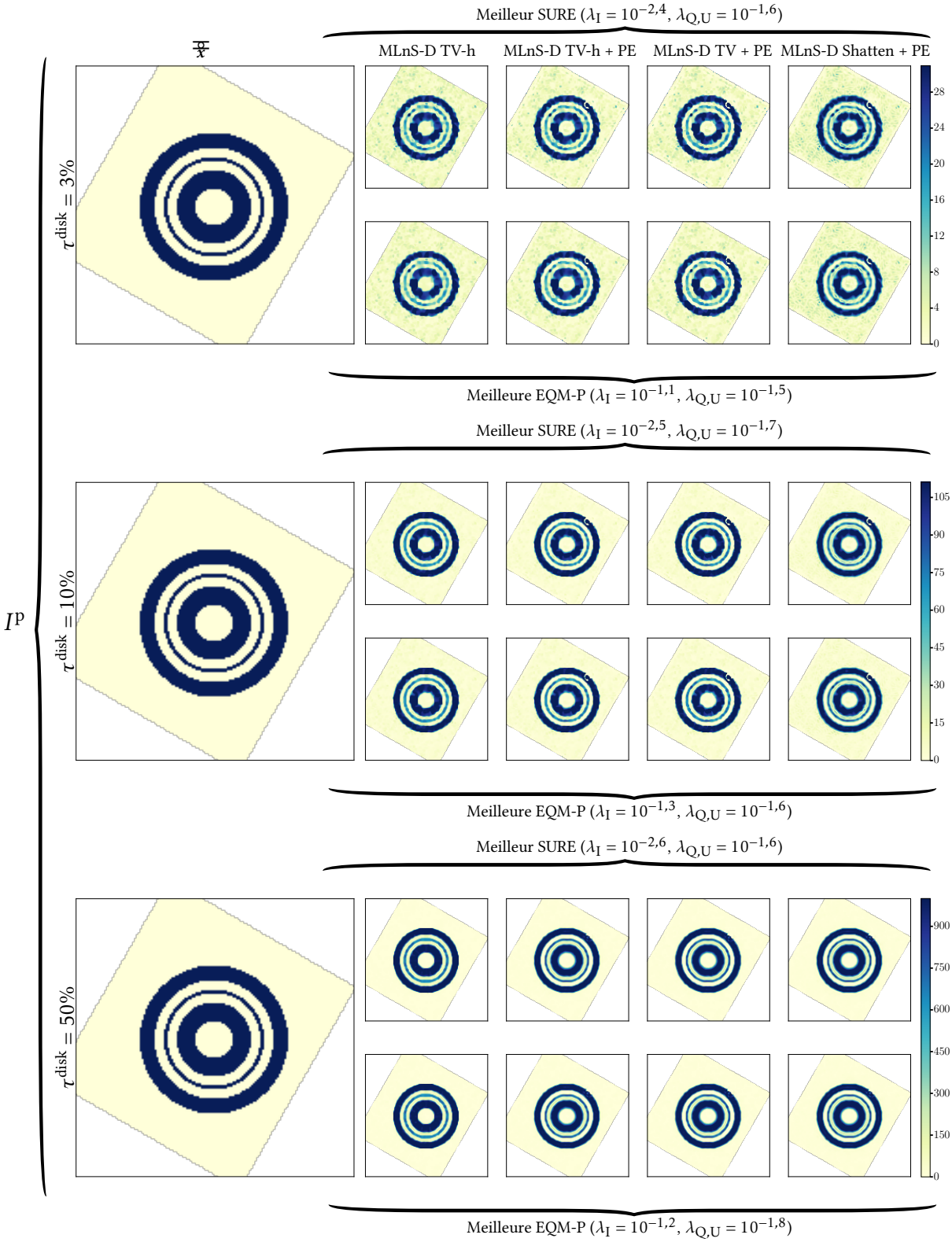
Pour I^{u} on voit, que l'EQMn est la plus basse pour la reconstruction régularisée par TV-h, dans le cas où l'hyperparamètre λ minimise l'EQMn. La régularisation donnant la meilleure reconstruction ensuite est celle par la régularisation de Shatten, quand λ minimise le critère SURE et l'EQMn.

Pour I^{p}, c'est une reconstruction régularisée par TV-h sans contrainte épigraphique qui donne la meilleure EQMn. On voit par ailleurs que l'EQMn pour TV-h avec contrainte épigraphique et TV est identique. La reconstruction régularisée par Shatten est celle qui a l'EQMn la plus élevée. Une telle différence entre les EQMn avec et sans contrainte épigraphique vient des artefacts de reconstruction autour de la binaire. Pour Shatten, cela vient sûrement aussi des bords qui sont plus doux et donc des valeurs moins élevées au bords par rapport à la vérité terrain où les bords sont francs.



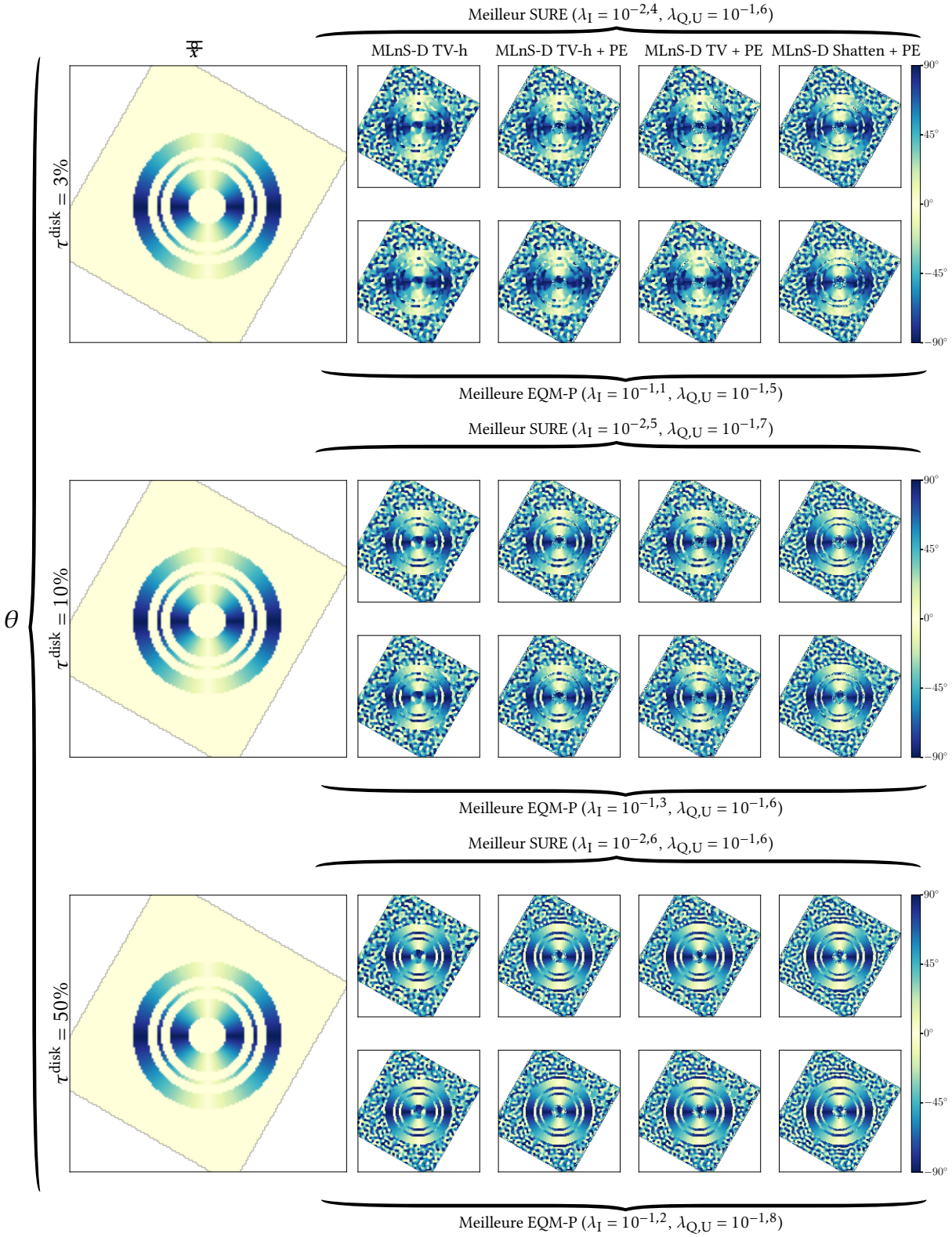
(a) Reconstruction des cartes de I^u.

FIGURE 4.5 – Cartes reconstruites avec les différentes méthodes pour $\tau^{\text{disk}} \in \{3\%, 10\%, 50\%\}$.



(b) Reconstruction des cartes de I^P.

FIGURE 4.5 – Cartes reconstruites avec les différentes méthodes pour $\tau^{disk} \in \{3\%, 10\%, 50\%\}$.



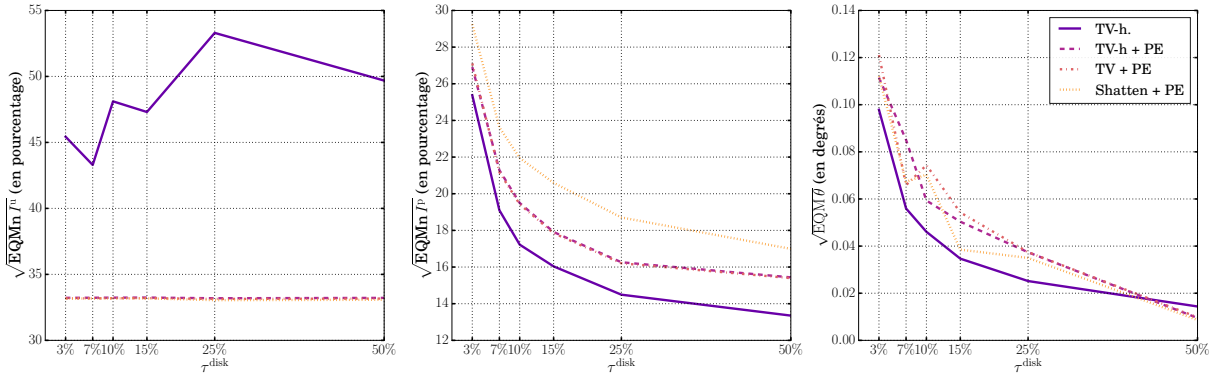
(c) Reconstructions des cartes de θ.

FIGURE 4.5 – Cartes reconstruites avec les différentes méthodes pour $\tau^{disk} \in \{3\%, 10\%, 50\%\}$.

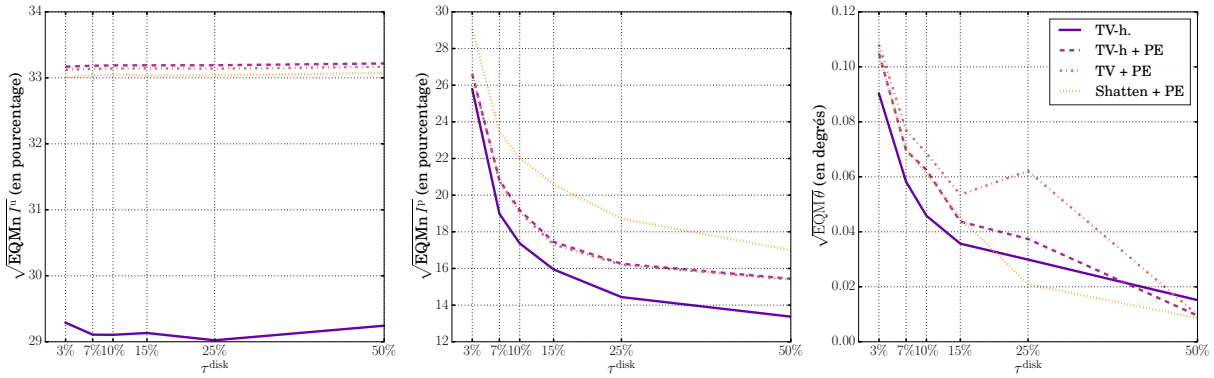
Enfin pour θ, dans le cas où les hyperparamètres $(\lambda_\ell)_{l\in\{1,\dots L\}}$ minimisant le critère SURE, la reconstruction par TV-h donne les valeurs d'EQMn les plus basses, sauf dans le cas où $\tau^{disk} = 50\%$, où c'est une régularisation par Shatten qui a l'EQMn la plus basse. Dans le cas où les hyperparamètres $(\lambda_\ell)_{l\in\{1,\dots L\}}$ minimisant l'EQMn, l'EQMn la plus basse est obtenue par TV-h, pour $\tau^{\text{disk}} \leq 15\%$, et par avec Shatten pour $\tau^{\text{disk}} \geq 25\%$. La reconstruction avec TV est celle qui donne l'EQMn la plus élevée.

On voit donc que dans la majeure partie des cas c'est la reconstruction par TV-h sans projection épigraphique qui donne les meilleurs résultats. Cependant il faut prendre en compte que d'une part, en pratique, il n'est pas possible de calculer l'EQMn car la vérité terrain n'est pas connue. D'autre part, les reconstructions avec projection épigraphique ont été arrêtées avant convergence. Il serait donc intéressant d'explorer les méthodes d'accélération numériques, autres que le préconditionnement, afin d'accélérer la convergence des méthodes proximales et comparer les résultats à convergence.

Dans le cas présent, c'est-à-dire dans le cas de la reconstruction par les MLnS-D sous contrainte temporelle, on peut conclure qu'une résolution régularisée par la TV-h sans contrainte épigraphique est la meilleure solution. En effet, c'est cette méthode qui donne la meilleure reconstruction de I^{p} pour la minimisation du critère SURE, ainsi que de l'angle lorsque $\tau^{\text{disk}} \leq 50\%$ dans le temps imparti. Cela est un critère important dans le cas de la



(a) EQMn entre les $\widehat{\boldsymbol{x}}$ estimés avec les différentes méthodes et $\overline{\overline{\boldsymbol{x}}}$ avec $(\lambda_\ell)^{\text{SURE}}_{\ell\in\{1,\dots L\}}$.



(b) EQMn entre les $\widehat{\boldsymbol{x}}$ estimés avec les différentes méthodes et $\overline{\overline{\boldsymbol{x}}}$ avec $(\lambda_\ell)^{\text{EQMn-P}}_{\ell\in\{1,\dots L\}}$.

FIGURE 4.6 – Erreurs quadratiques moyennes normalisées des différents paramètres pour les différentes méthodes. La figure a) représente l'EQMn $\text{EQMn}(\widehat{\boldsymbol{x}}, \overline{\overline{\boldsymbol{x}}})$ pour les différentes méthodes. La figure b) représente le rapport entre ces EQMn et la meilleure valeur de toutes les EQMn pour chaque valeur de τ^{disk}.

résolution de cartes plus grandes à partir d'un jeu de données plus grand (par exemple :$300 \times 300 \times 88$ pour RXJ 1615), surtout pour la recherche des hyperparamètres optimaux.

Afin de montrer l'apport d'un modèle complet incluant la convolution par rapport à une déconvolution *a posteriori*, nous comparons maintenant les estimations de I^{u} présentées sur la figure 4.5, dans le cas des hyperparamètres minimisant l'EQMn, avec la déconvolution des résultats séparables obtenus par les méthodes de la Double Différence et de la Méthode non-Linéaire Séparable. Pour faire la déconvolution, ont résout le problème :

$$\widehat{\boldsymbol{x}} \in \underset{\boldsymbol{x} \geq 0_N \in \mathbb{R}^N}{\operatorname{Argmin}} \left\{ \frac{1}{2} \|\widehat{\boldsymbol{x}}^{\mathrm{sep}} - \mathbf{A}\boldsymbol{x}\|^2 + \mathcal{R}(\boldsymbol{x}) \right\}, \tag{4.46}$$

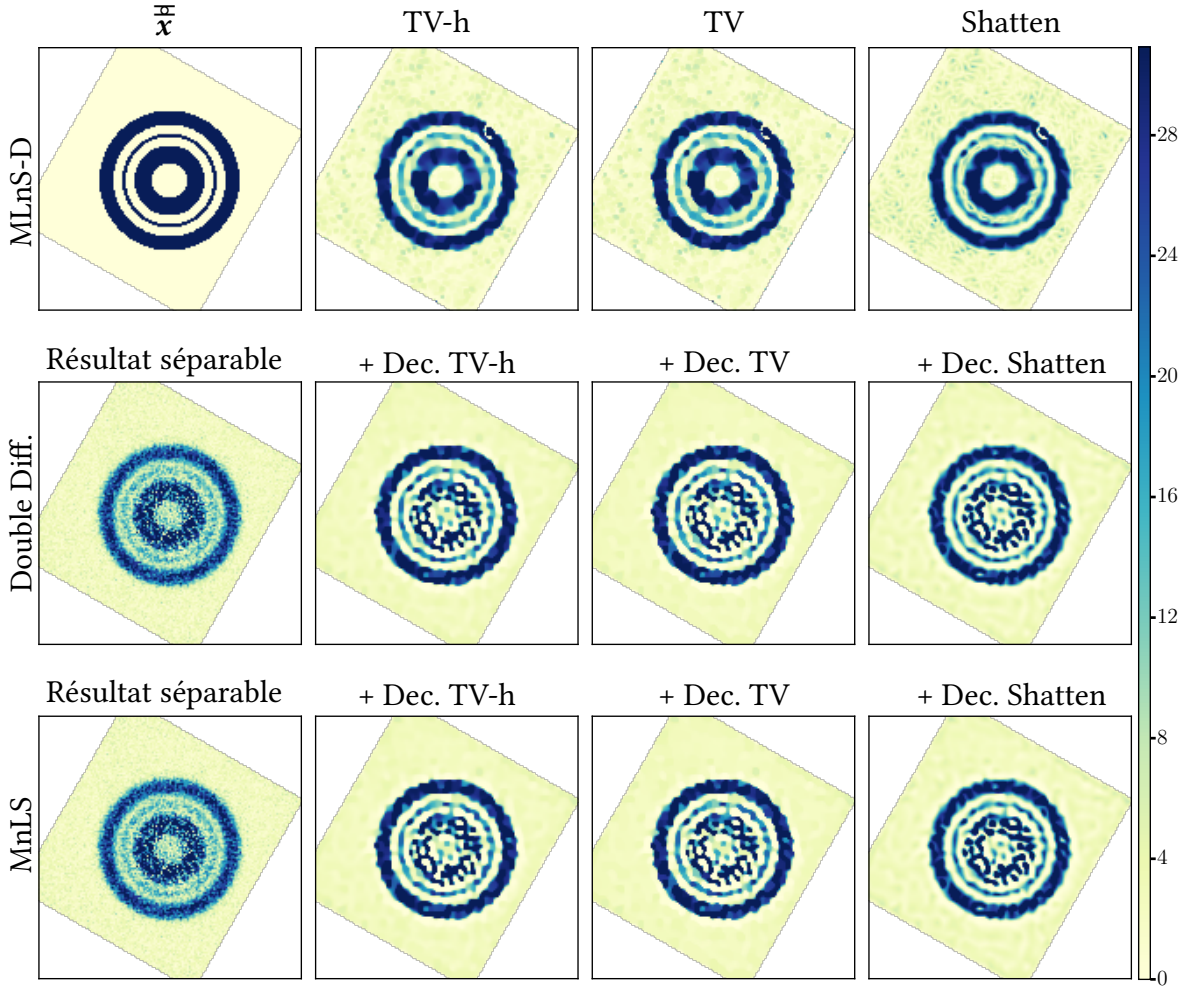
où $\widehat{\boldsymbol{x}}^{\mathrm{sep}}$ représente le paramètre I^{u} estimé avec l'une ou l'autre des deux méthodes séparables, où $\boldsymbol{A} : \mathcal{M}_N(\mathbb{R})$ représente la même convolution par la PSF que dans le cas non-séparable, où $\mathcal{R}$ correspond à la régularisation, et $\boldsymbol{x} \geq 0_N$ signifie que pour tout $n \in \{1, \dots N\}$, $\boldsymbol{x}_n \geq 0$. Les hyperparamètres sont réglés manuellement car un réglage minimisant le critère SURE est sous-régularisé et le réglage minimisant l'EQMn semble sur-régularisé. La sur-régularisation du critère lors de la minimisation de l'EQmn vient du fait que celui-ci est calculé uniquement sur les zones contenant du signal, c'est-à-dire les anneaux. Or dans les cas de faible polarisation du disque, les choix des hyperparamètres maximisant l'intensité polarisée dans les anneaux induisent un lissage entre les anneaux.

La figure 4.7 les résultats visuels obtenus pour I^{p} pour les taux de polarisation du disque $\tau^{\mathrm{disk}} \in \{3\%, 10\%, 50\%\}$, pour les régularisations TV-h, TV et Shatten, dans le cas d'une part de la reconstruction par la Méthode Linéaire non-Séparable avec Déconvolution (MLnS-D) avec contrainte épigraphique, et dans le cas de la déconvolution *a posteriori* des résultats séparables de la double différence et de la Méthode non-Linéaire Séparable, avec contrainte de positivité. On voit tout d'abord que, pour chaque taux de polarisation du disque, les reconstructions par déconvolution *a posteriori* sont très similaires pour toutes les régularisations.

Pour $\tau^{\mathrm{disk}} = 3\%$, on voit que la reconstruction avec un modèle global est bien meilleure, surtout pour l'anneau central. Cependant, la reconstruction séparable ne présentant pas l'artefact de déconvolution autour de la binaire, il n'est pas non plus présent sur les résultats déconvolués.

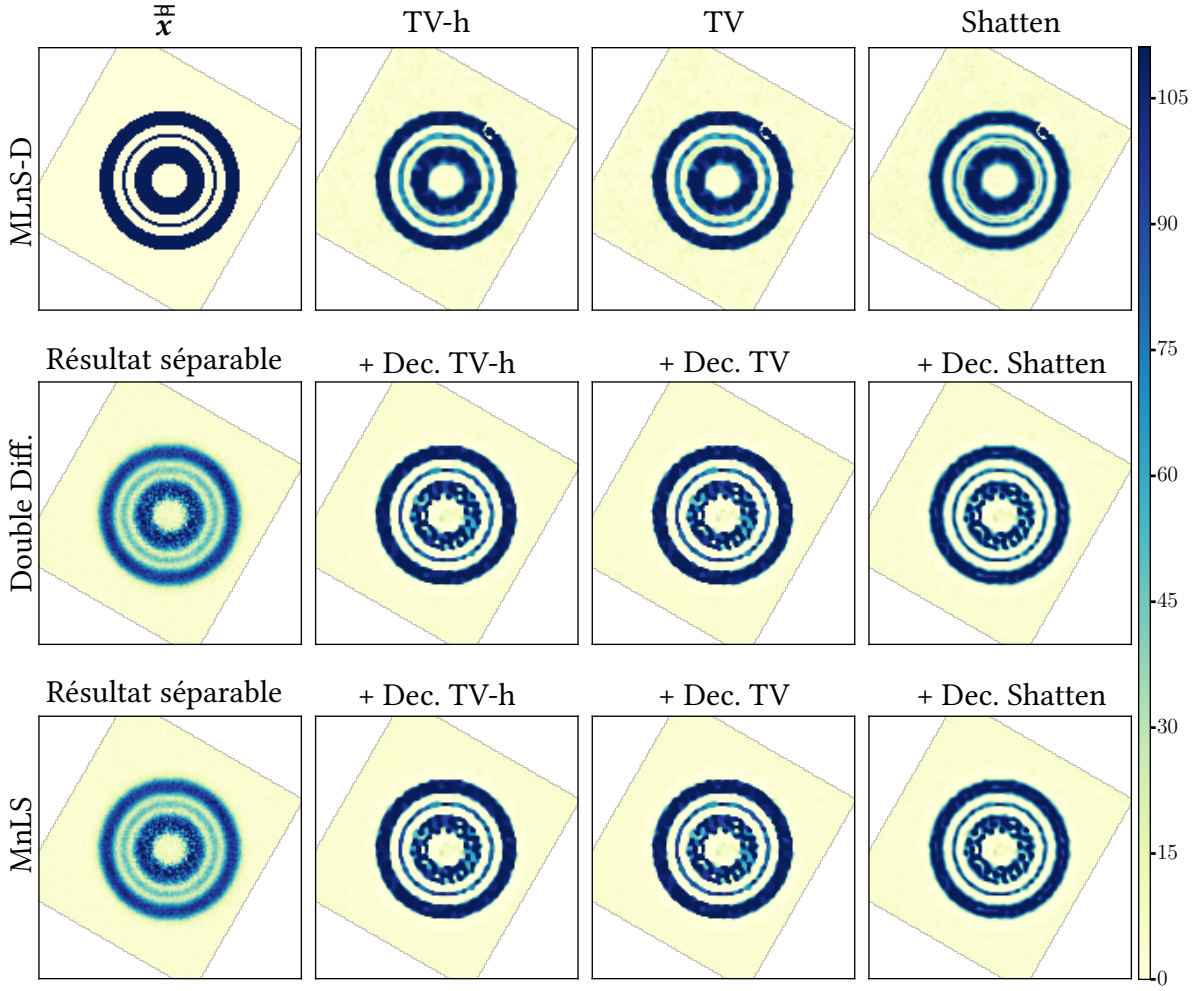
Pour $\tau^{\mathrm{disk}} = 10\%$, on voit que dans le cas de TV-h et de TV, les deux anneaux externes sont mieux restitués avec les méthodes séparables déconvoluées *a posteriori*, tandis que l'anneau du centre est mieux reconstruit avec les MLnS-D. Dans le cas de la régularisation par Shatten, la reconstruction de l'anneau externe comporte quelques artefacts dans le cas de la déconvolution des résultats séparables. L'anneau fin est cependant particulièrement bien reconstruit par l'ensemble des méthodes de déconvolution *a posteriori*.

Pour $\tau^{\mathrm{disk}} = 50\%$, les reconstructions par déconvolution a posteriori semblent meilleures que les reconstructions par les MLnS-D. En effet, la finesse et l'intensité de l'anneau fin est bien mieux retrouvée avec la déconvolution directe de I^{p} obtenu par les méthodes séparables. Les bords sont également plus francs, sauf dans le cas de la déconvolution des méthodes séparables avec une régularisation par Shatten, où l'intensité retrouvée semble être légèrement plus basse que l'intensité vraie et les bords sont plus doux.



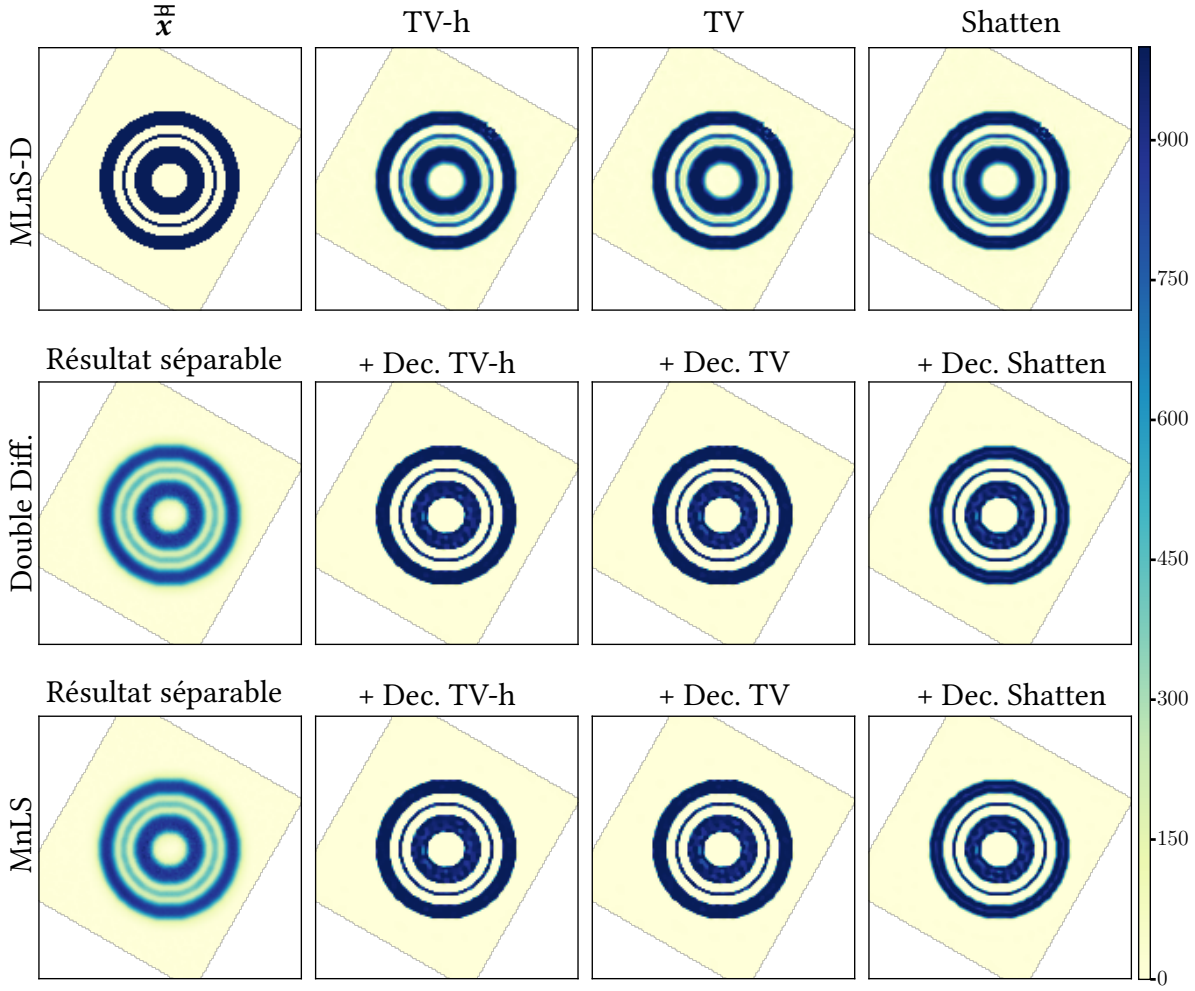
(a) Comparaison pour $\tau^{\mathrm{disk}} = 3\%$.

FIGURE 4.7 – Comparaison visuelle des reconstructions du paramètre I^{p} pour $\tau^{\mathrm{disk}} \in \{3\%, 10\%, 50\%\}$, obtenues avec TV-h, TV et Shatten, avec contrainte épigraphique dans le cas du modèle complet et contrainte de positivité dans le cas de la déconvolution des méthodes séparables. La première colonne correspond sur la première ligne à la vérité terrain puis aux résultats séparables sans déconvolution de la Double Différence et de la MnLS. Pour la troisième à la cinquième colonne, la première ligne correspond aux reconstructions obtenues à partir du modèle complet présenté dans ce chapitre, pour les différentes régularisations. La deuxième et la troisième ligne correspondent respectivement aux reconstructions par déconvolution *a posteriori* des résultats séparables de la Double Différence et de la MnLS.



(b) Comparaison pour $\tau^{\mathrm{disk}} = 10\%$.

FIGURE 4.7 – Comparaison visuelle des reconstructions du paramètre I^{p} pour $\tau^{\mathrm{disk}} \in \{3\%, 10\%, 50\%\}$, obtenues avec TV-h, TV et Shatten, avec contrainte épigraphique dans le cas du modèle complet et contrainte de positivité dans le cas de la déconvolution des méthodes séparables. La première colonne correspond sur la première ligne à la vérité terrain puis aux résultats séparables sans déconvolution de la Double Différence et de la MnLS. Pour la troisième à la cinquième colonne, la première ligne correspond aux reconstructions obtenues à partir du modèle complet présenté dans ce chapitre, pour les différentes régularisations. La deuxième et la troisième ligne correspondent respectivement aux reconstructions par déconvolution *a posteriori* des résultats séparables de la Double Différence et de la MnLS.



(c) Comparaison pour $\tau^{\mathrm{disk}} = 50\%$.

FIGURE 4.7 – Comparaison visuelle des reconstructions du paramètre I^{p} pour $\tau^{\mathrm{disk}} \in \{3\%, 10\%, 50\%\}$, obtenues avec TV-h, TV et Shatten, avec contrainte épigraphique dans le cas du modèle complet et contrainte de positivité dans le cas de la déconvolution des méthodes séparables. La première colonne correspond sur la première ligne à la vérité terrain puis aux résultats séparables sans déconvolution de la Double Différence et de la MnLS. Pour la troisième à la cinquième colonne, la première ligne correspond aux reconstructions obtenues à partir du modèle complet présenté dans ce chapitre, pour les différentes régularisations. La deuxième et la troisième ligne correspondent respectivement aux reconstructions par déconvolution *a posteriori* des résultats séparables de la Double Différence et de la MnLS.

La figure 4.8 compare les EQMn des reconstructions de I^{p} pour les différentes régularisations, dans le cas d'une part de la reconstruction par la Méthode Linéaire non-Séparable avec Déconvolution (MLnS-D), et dans le cas de la déconvolution *a posteriori* des résultats séparables de la double différence et de la Méthode non-Linéaire Séparable. On voit tout d'abord que pour les deux méthodes séparables, l'EQMn est identique. De plus, on constate d'une part que pour la régularisation par TV-h et TV, la reconstruction par la MLnS-D, les reconstructions ont une EQMn plus basse que par la MnLS, pour $\tau \leq 25\%$, même deux fois plus basse pour $\tau^{\mathrm{disk}} = 3\%$. Cependant, pour $\tau \geq 25\%$, la déconvolution a posteriori des méthodes séparables a une EQMn plus basse. D'autre part, dans le cas de la reconstruction par Shatten, ce sont toujours les reconstructions par la MLnS-D qui ont les valeurs d'EQMn les plus basses. Ces résultats confirment bien les observations faites sur les résultats visuels.

Il ressort de tels résultats que dans le cas de la reconstruction d'objets peu brillant, l'utilisation d'un modèle linéaire incluant la convolution apporte de meilleurs résultats qu'une déconvolution *a posteriori* des paramètres d'intérêt. Dans le cas d'objet très brillant, la baisse de performance de la MLnS-D vis-à-vis des méthodes séparables avec déconvolution *a posteriori* peut tout aussi bien venir d'un problème de réglage des hyperparamètres que de convergence des algorithmes ou d'un mauvais choix des paramètres du modèle pour la régularisation.

En effet, un tel comportement ressemble fortement au comportement de la MLnS sans déconvolution, vis-à-vis des résultats séparables dans la section 3.1.4, où à τ^{disk} grand, la MLnS est moins bonne que les méthodes séparables.

Rappelons que lors de l'utilisation du modèle linéaire, la régularisation est faite sur les paramètres de Stokes I, Q et U reliées aux intensités I^{u} et I^{p} par les relations

Rappel éq. (1.16)

$$\begin{cases} \mathrm{I}_n = I^{\mathrm{u}}_n + I^{\mathrm{p}}_n, \\ \mathrm{Q}_n = I^{\mathrm{p}}_n \cos 2\theta_n, \\ \mathrm{U}_n = I^{\mathrm{p}}_n \sin 2\theta_n. \end{cases}$$

Dans le cas linéaire, le paramètre I^{p} est donc régularisé simultanément dans I et Q. Or nous avons vu dans la section 3.2.3 que l'utilisation d'un changement de variable permettant de

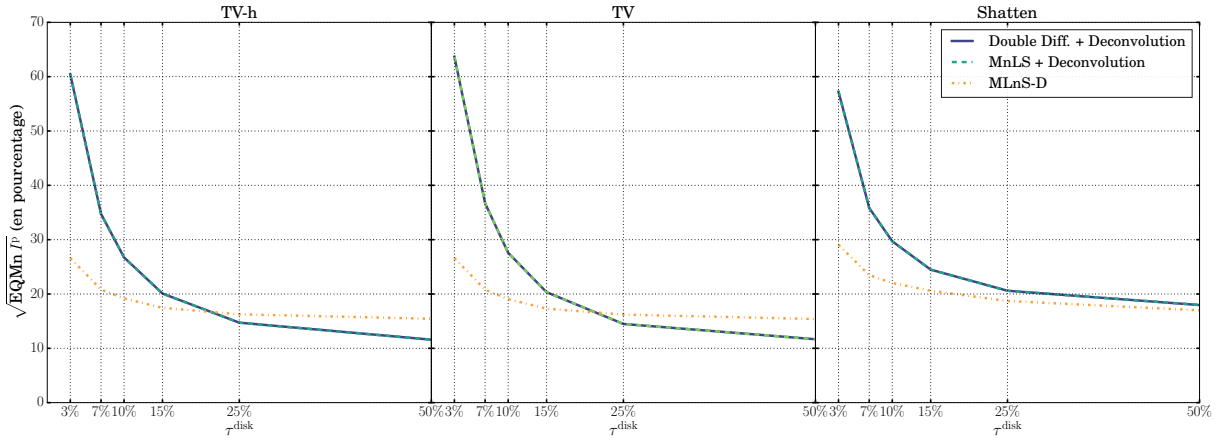


FIGURE 4.8 – Comparaison des EQMn des reconstructions du paramètre I^{p} en fonction du taux de polarisation du disque, obtenues avec TV-h, TV et Shatten, avec contrainte épigraphique dans le cas du modèle complet et contrainte de positivité dans le cas de la déconvolution des méthodes séparables.

régulariser sur I^{u} et I^{p} séparément permettait d'avoir une erreur d'estimation plus faible que celle des méthodes séparables, pour tous les taux de polarisation du disque.

On peut donc supposer que la qualité de reconstruction, des paramètres du modèle incluant la déconvolution, bénéficierait du changement de variable non-linéaire en les paramètres $(I^{\mathrm{u}}, \mathrm{Q}, \mathrm{U})^\top$, avec une régularisation jointe et structurée sur les paramètres Q et U, aussi bien a faible taux de polarisation du disque, qui sont les cas les plus intéressants en astrophysiques, les cas $\tau^{\mathrm{disk}} \geq 25\%$ n'étant pas représentatifs des données réelles. De tels travaux sont en cours et seront présentés dans un papier en cours de préparation.

En conclusion, sur données simulées l'utilisation de MLns-D avec TV-h sans contrainte épigraphique permet à faible taux de polarisation du disque, et donc faible SNR, d'obtenir des résultats satisfaisant en un temps restreint. L'utilisation de la contrainte épigraphique permet un meilleur réglage des hyperparamètres λ_I avec SURE, cependant elle nécessite l'utilisation de méthodes proximales, qui nécessitent plus d'itérations et de temps en seconde pour arriver à convergence. C'est pourquoi l'utilisation de TV-h sans contrainte épigraphique me semble la plus pertinente sur données astrophysiques. De plus, l'inclusion de la convolution dans le modèle linéaire, permet à faible SNR de fortement diminuer l'erreur de reconstruction vis-à-vis des résultats séparables déconvolués *a posteriori*. Un changement de variable tel que celui fait dans la section 3.2 pourrait permettre d'une part une meilleure estimation de I^{p} a haut SNR, mais aussi de traiter la contrainte épigraphique sous forme de contrainte de positivité sur I^{u} et I^{p} lors de la résolution avec VMLM-B.

4.3.2 Résultats sur données astrophysiques

J'étudie dans cette partie les résultats obtenus dans le temps impartit avec la MLnS-D, régularisée par TV-h sans et avec contrainte épigraphique et avec TV, aux données *calibrées* de la cible RXJ 1615 [Avenhaus et al., 2018] observée en bande H. Pour des raisons de temps de calcul, les résultats de déconvolution avec régularisation par Shatten ne seront pas présentés dans ce manuscrit sur cette cible.

Avant de faire la déconvolution, j'estime la PSF en ajustant une fonction de Moffat sur une reconstruction de la PSF, obtenue par minimisation de la vraisemblance régularisée des données de calibrations de PSF. De plus amples détails sont présentés dans la section 6.1. Je masque également les pixels où le modèle de déconvolution est faux, comme les pixels correspondant au coronographe ou a des saturations du détecteur, en mettant les poids des pixels correspondants à 0 dans les cartes de poids $(\mathbf{W}_k)_{k\in\{1,\dots K\}} \in \mathbb{R}^M$.

Sur la figure 4.9 sont présentés les résultats obtenus pour les différentes configurations de régularisations et de contraintes, pour un ensemble d'hyperparamètres minimisant le critère SURE, dans le cas d'une régularisation par TV-h sans contrainte épigraphique. On voit que dans le cas de TV-h, les cartes reconstruites $\widehat{I^{\mathrm{p}}}$ et $\widehat{\theta}$ sont sur-régularisées, tandis que la carte reconstruite avec TV semble être régularisée de manière satisfaisante. La précision relative atteinte sur le critère pour TV au bout de 2000 itérations est de l'ordre de 10^{-6}. Il est surprenant que les résultats obtenus avec TV-h ne soient pas similaires à ceux obtenus par TV pour un si petit hyperparamètre μ. La différence entre une régularisation par TV-h sans et avec projection épigraphique est légère. Elle se voit plus particulièrement sur l'anneau central, qui semble mieux reconstruit avec la contrainte, et sur le paramètre I^{u}, qui semble alors moins régularisé avec contrainte également.

En comparaison avec les résultats séparables et non-séparables sans déconvolution, pour les cartes reconstruites $\widehat{I^{\mathrm{P}}}$ et $\widehat{\theta}$, tels que ceux obtenus sur la figure 3.7, on remarque deux améliorations. D'une part, la «coupure »du disque due à la polarisation instrumentale est moins visible avec la déconvolution. D'autre part, dans le cas de la reconstruction par TV, la séparation entre les anneaux est plus marquée. On peut également deviner les contours de l'anneau le plus externe qui est très faible, ainsi qu'une petite structure indiquée en haut à gauche, que

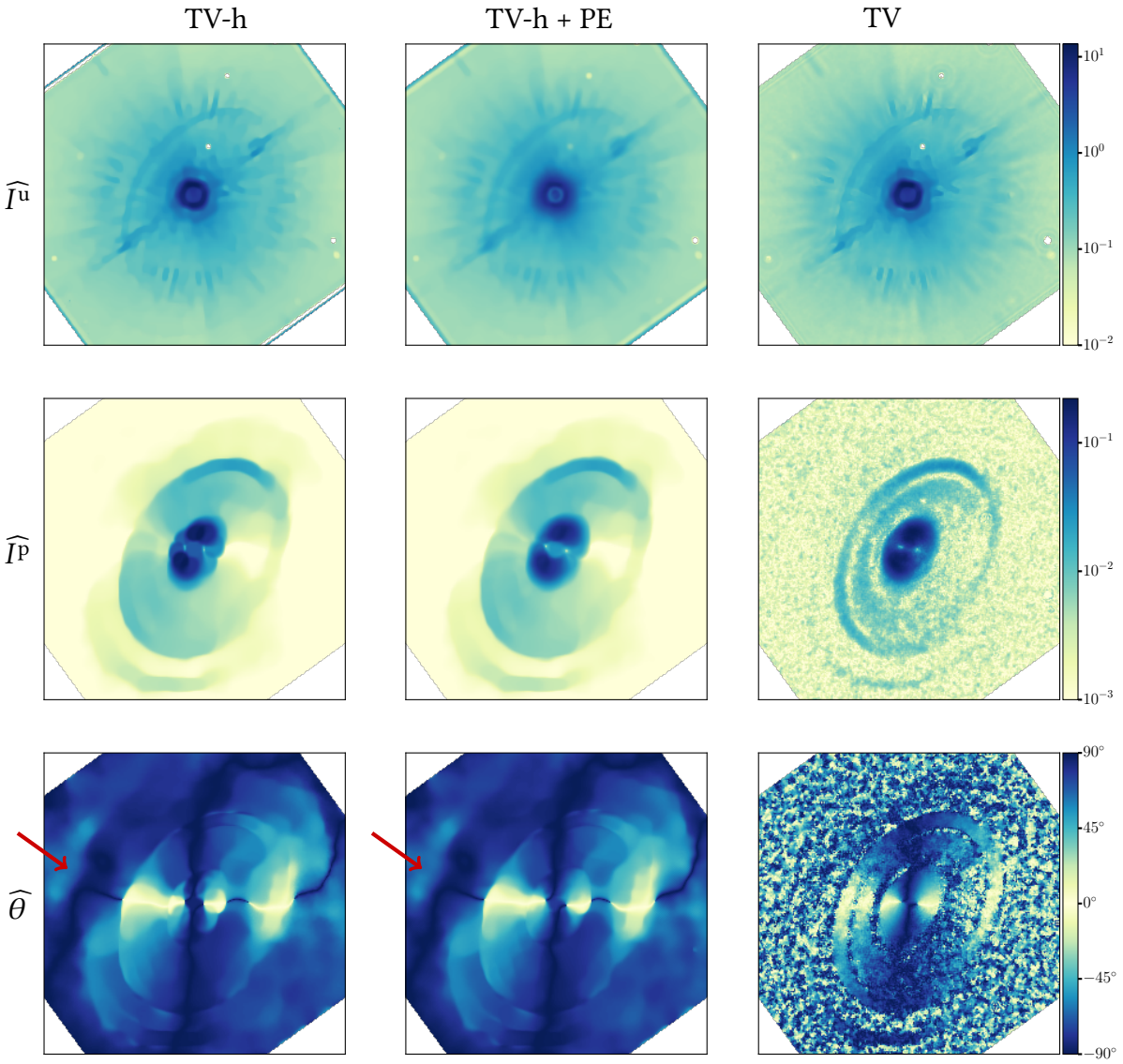


FIGURE 4.9 – Cartes reconstruites avec les différentes méthodes pour les régularisations TV-h, sans et avec contrainte épigraphique et TV avec contrainte épigraphique. Les hyperparamètres sont réglés de sorte à minimiser le critère SURE dans le cas d'une régularisation par TV-h sans contrainte épigraphique. Cela résulte en $\mu = 10^{-5}$, $\lambda_{\mathrm{I}} \approx 1$ et $\lambda_{\mathrm{Q}} \approx 1$. La convergence pour TV et TV-h avec contrainte épigraphique est arrêtée au bout 2000 itérations. Si pour TV-h avec contrainte épigraphique l'algorithme semble avoir pratiquement convergé, car les résultats sont semblables à ceux sans contrainte épigraphique, où l'algorithme a convergé, ce n'est pas le cas de TV, où les résultats sont moins régularisés.

TABLE 4.1 – Jeux d'hyperparamètres pour la reconstruction régularisée par TV-h. Les résultats associés sont présentés dans la figure 4.10

	μ	λ_I	λ_Q, λ_U		μ	λ_I	λ_Q, λ_U		μ	λ_I	λ_Q, λ_U
(a)	10^{-3}	1	1	(b)	10^{-3}	1	$10^{-0.4}$	(c)	10^{-3}	1	10^{-1}
(d)	10^{-4}	1	1	(e)	10^{-4}	1	$10^{-0.5}$	(f)	10^{-4}	1	10^{-1}

l'on peut retrouver dans l'image sans polarisation instrumentale sur la figure 1 où la polarisation instrumentale a été enlevée. Cette structure est par ailleurs bien plus visible dans les cartes d'angles θ sur-régularisées, ce qui s'explique par le petit SNR sur cette structure liée au flux très faible.

Les résultats obtenus pour les hyperparamètres minimisant le critère SURE ne sont pas satisfaisant dans le cas de la régularisation par TV-h. Je propose donc d'étudier différents ensembles d'hyperparamètres et de comparer les résultats au bout de 2000 itérations maximum. Cela correspond à un peu moins de 2 heures de calcul par ensemble d'hyperparamètres.

La figure 4.10 présente des résultats pour 6 jeux d'hyperparamètres, dont les valeurs sont présentés dans la table 4.1, le cas (c) correspondant à un cas sous-régularisé et le cas (d) à un cas sur-régularisé. Sur la carte (e), on numérote les différents anneaux ou structures de 1 à 5. Les cas où l'anneau externe n°1 est le plus visible sont les cas (a), (c) et (f). On voit par ailleurs que dans le cas (e), la structure n°5 du disque externe est visible. On voit de plus sur les cas (a) et (b) que l'anneau n°3 semble se dédoubler au-dessus de l'anneau n°4.

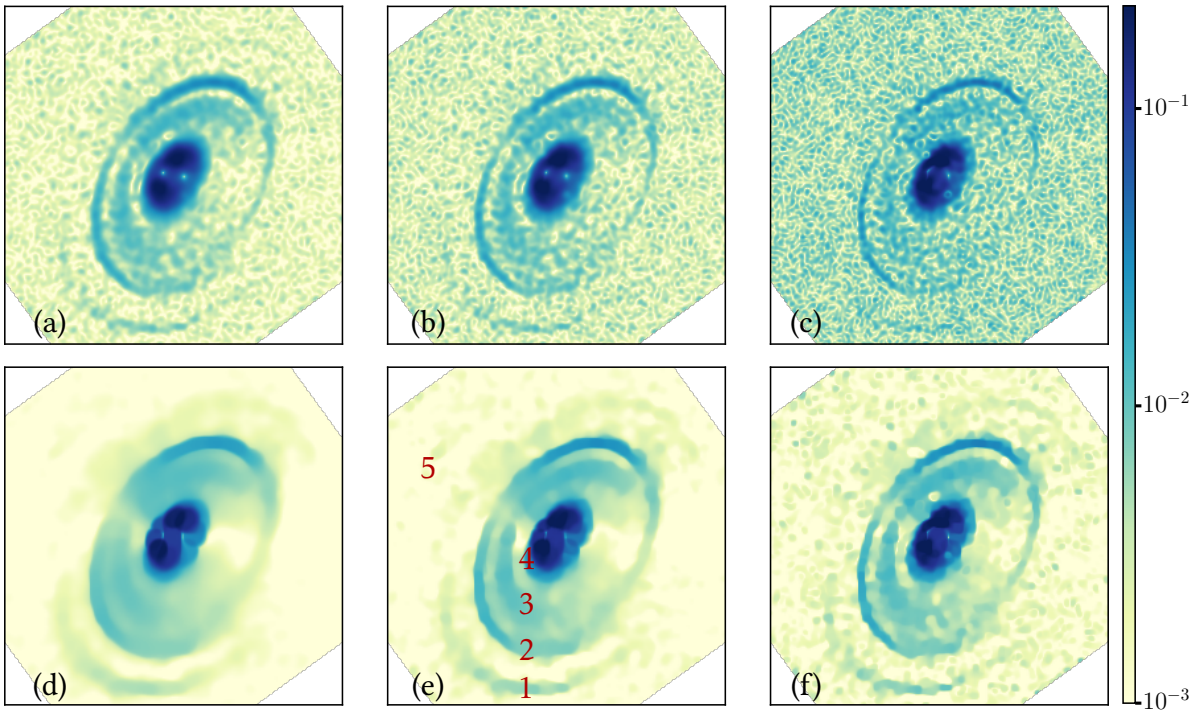


FIGURE 4.10 – Comparaison des différentes reconstructions de I^p avec une régularisation TV-h sans contrainte épigraphique pour différentes valeurs d'hyperparamètres choisis manuellement. En cas de non-convergence de l'algorithme, celui-ci est arrêté au bout de 2000 itérations.

La figure 4.11 est associée à deux autres choix de jeux d'hyperparamètres, pour lesquels les résultats ne semblent ni sur-régularisés, ni sous-régularisés. La différence entre les résultats de chaque colonne est très minime, mais on remarque pour $\widehat{I^{\mathrm{P}}}$ que si la reconstruction de droite semble légèrement plus bruitée, le dédoublement de l'anneau n°3 est plus visible. Cependant la structure n°5 est plus visible dans le cas de la colonne de gauche, mais pas autant que sur la carte θ dans les cas sur-régularisé. On remarque par ailleurs que le dédoublement de l'anneau est également visible dans $\widehat{\theta}$, on peut donc supposer que ce dédoublement est bien réel et non un artefact de déconvolution.

Il ressort de tels résultats que la déconvolution permet d'une part de mieux détecter les objets de faible intensité et mieux préciser les structures brillantes. Cependant, le réglage des hyperparamètres de régularisation est délicat et il ne semble pas y avoir de réglage « parfait ». En effet, les solutions sur-régularisées et sous-régularisées apporte des informations différentes et complémentaires. Si le critère SURE semble sur-régularisé, il permet une meilleure détection de la structure n°5. À l'inverse, un réglage manuel sous-régularisant la solution augmente le nombre d'artefacts mais permet d'observer le dédoublement de l'anneau n°3. Il y a donc une plage d'hyperparamètres donnant des résultats tout autant intéressants et le choix d'un unique hyperparamètre demande de faire un compromis entre une résolution fine des structures brillantes, mais la perte des structures de faible intensité, et la meilleure détection des structures de faible intensité mais une perte de résolution des structures fines.

Pour montrer l'efficacité de l'inclusion de la convolution dans le modèle direct des données astrophysiques, par rapport à une déconvolution *a posteriori* des résultats séparables, nous comparons les résultats précédents à une déconvolution *a posteriori* des résultats séparables de la Double Différence et de la MnLS.

La figure 4.12 montre les résultats obtenus par déconvolution des résultats séparables de la Double Différence et de la MnLS, pour les régularisations avec TV-h, TV et Shatten avec contrainte de positivité. Les hyperparamètres sont fixés manuellement de sorte que la reconstruction ne soit ni sur-régularisé, ni sous-régularisée.

On remarque tout d'abord que les reconstructions obtenues à partir des résultats séparables de la Double Différence contiennent un grand nombre artefacts autour des résidus d'interpolation des pixels morts. Ceux-ci ayant été pondérés dans le cas de la MnLS, les résultats de déconvolution de ma méthode ne présentent pas ces artefacts. Ces erreurs mises à part, les déconvolutions des résultats de la Double Différence et de la MnLS sont assez similaires pour chacune des régularisations. On remarque par ailleurs que les résultats obtenus avec TV sont plus lisses que les résultats obtenus avec TV-h et Shatten. En comparaison avec la MLnS-D, l'anneau n°1 semble plus bruité lors de la déconvolution à posteriori. De plus, le dédoublement de l'anneau n°3 et la structure n°5 ne sont pas visibles. Enfin, la «coupure »due à la polarisation instrumentale semble exacerbée par la déconvolution.

On peut donc en conclure que la prise en compte de la convolution dans le modèle complet que j'ai développé donne des résultats bien meilleurs que ceux de la déconvolution *a posteriori* des méthodes séparables. En effet, les résultats de la MLnS-D sont plus précis, en termes de morphologie et de détection des structures de faible intensité, qu'une déconvolution *a posteriori*.

Il est important de faire remarquer que ce disque a un SNR par pixel très faible (cf. figure 2.7). Une telle conclusion est donc en accord avec les comparaisons faites sur données simulées.

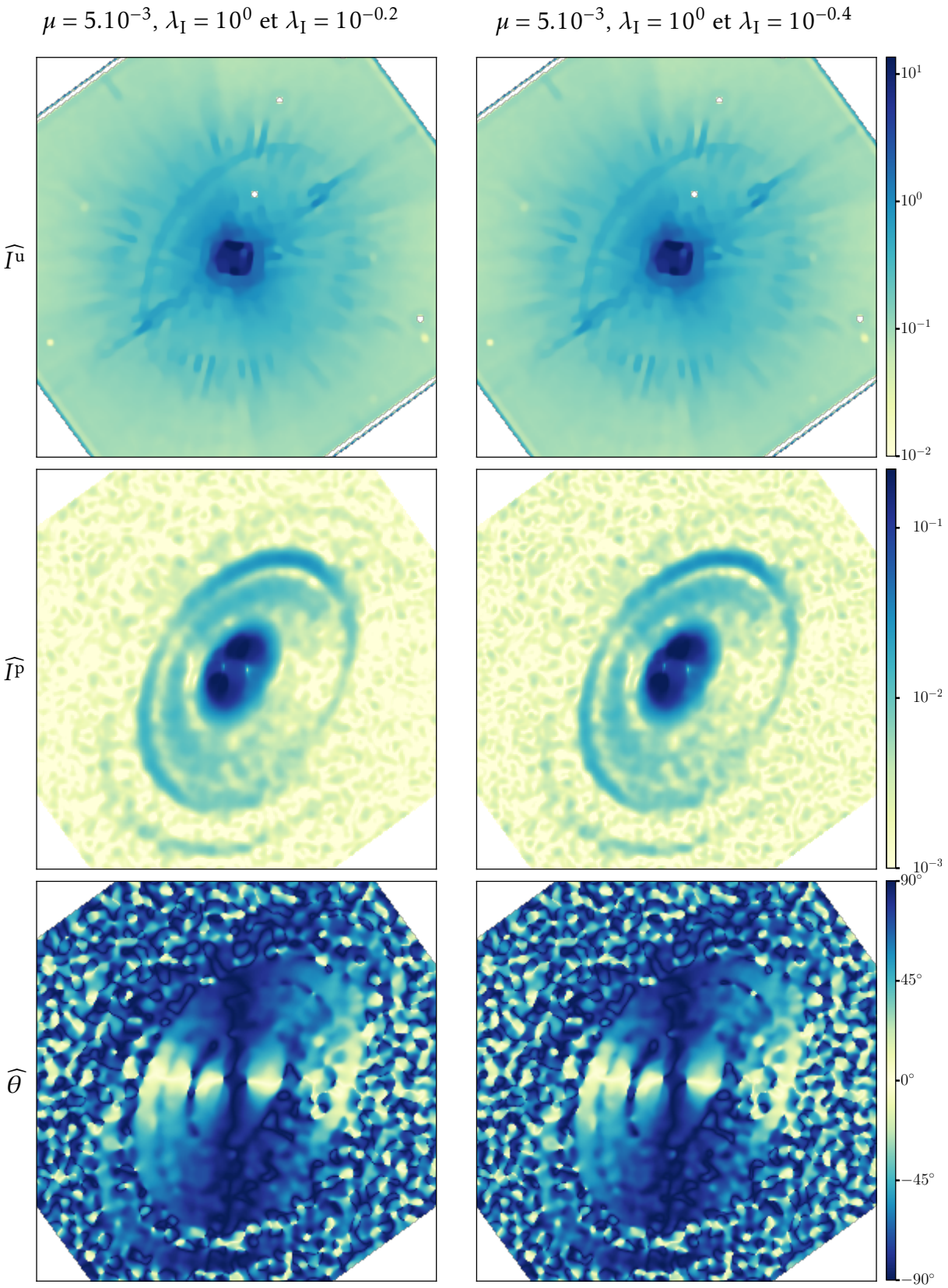


FIGURE 4.11 – Cartes reconstruites avec la régularisation TV-h, sans contrainte épigraphique pour un réglage d'hyperparamètres manuel visuellement satisfaisant.

4.3.3 Synthèse des résultats

Nous avons pu voir dans cette section que la déconvolution avec régularisation par TV-h sans contrainte épigraphique permet d'obtenir des résultats de bonne qualité, avec une erreur faible en un temps de calcul moindre par rapport aux méthodes proximales, peu importe la difficulté du cas. La contrainte épigraphique permet par ailleurs d'éviter les artefacts de déconvolution sur $\widehat{I^{\mathrm{u}}}$ sans sur-régulariser le carte reconstruite.

La comparaison des résultats de la MLnS-D aux résultats de déconvolution *a posteriori* des résultats séparables de la double différence et de la MnLS a montré, dans le cas sur données simulées, les limites d'un modèle direct linéaire et pousse à envisager un modèle non linéaire permettant de régulariser I^{u} et I^{p} de manière indépendante. Sur données astrophysiques, l'inclusion de la convolution dans le modèle améliore grandement la qualité de reconstruction.

Nous avons vu dans le cas des données simulées que le choix des hyperparamètres minimisant le critère SURE permettait d'obtenir des résultats équivalents par rapport à un choix d'hyperparamètres minimisant l'Erreur Quadratique Moyenne normalisée, sauf dans le cas de TV-h sans projection épigraphique, pour la reconstruction de la carte I^{u}. Cependant cette méthode étant celle qui donne les meilleurs résultats pour I^{p} et θ, il semble encore plus pertinent de se tourner vers un modèle non linéaire permettant d'inclure la contrainte épigraphique sous forme de contrainte de positivité sur I^{u} et I^{p}. Sur données réelles, les hyperparamètres minimisant le critère SURE sur-régularise la solution, ce qui met en avant des structures de très faibles intensités, moins visible dans des cas de choix d'hyperparamètres semblant plus « raisonnable ».

Enfin, nous avons vu lors de l'application sur données réelles que la régularisation des environnements se fait à plusieurs échelles, impliquant un réglage des hyperparamètres délicat,

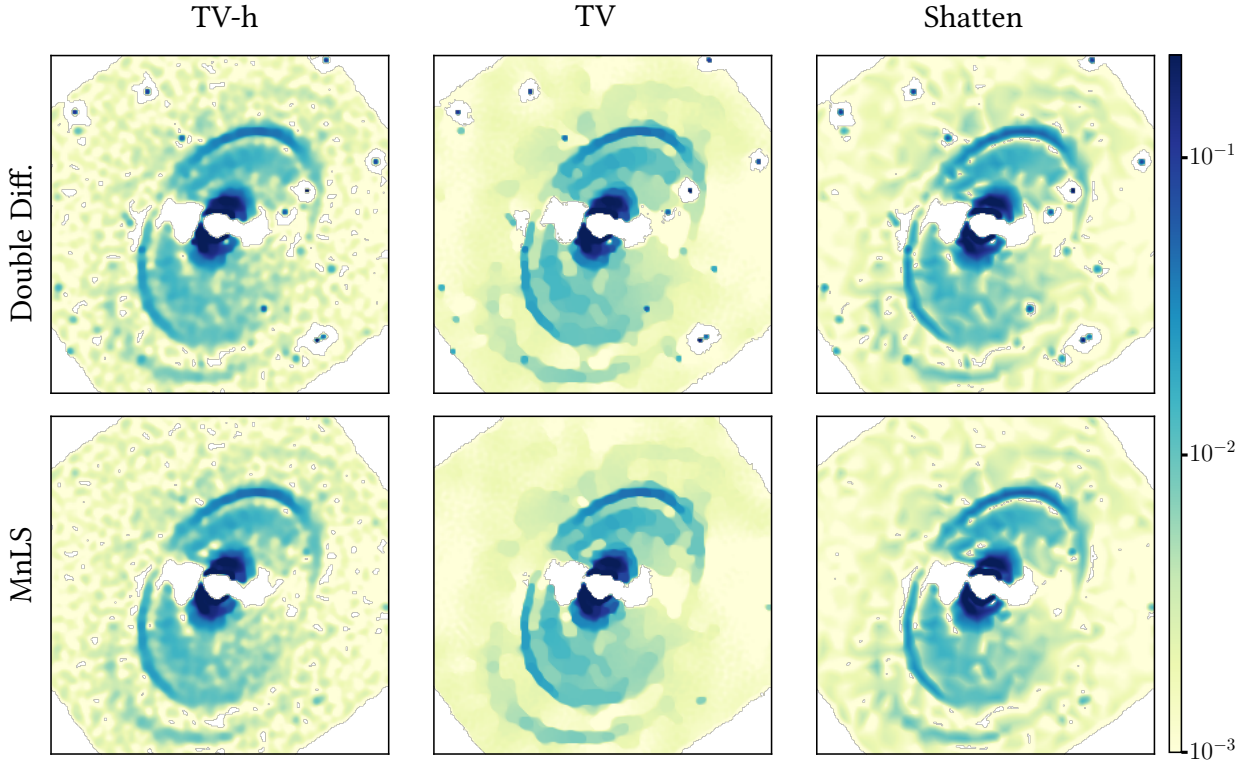


FIGURE 4.12 – Comparaison des cartes de résultats séparables de I^{p} déconvolués *a posteriori* à partir de la méthode du de la Double Différence et de la MnLS.

et ouvre le chemin vers l'étude de régularisations plus adaptées à ce problème.

Conclusion du chapitre

Dans ce chapitre, j'ai présenté le modèle complet incluant la convolution. J'ai alors développé un critère régularisé permettant d'estimer les paramètres de ce modèle, sans et avec contrainte épigraphique. J'ai présenté trois algorithmes permettant de minimiser le critère selon les choix de régularisation. Dans le cas d'une régularisation différentiable sans contrainte épigraphique j'ai utilisé l'algorithme VMLM, dans le cas de l'utilisation de la contrainte épigraphique j'ai utilisé l'algorithme VMFBwB et enfin dans le cas de l'utilisation de régularisation non-différentiable avec contrainte épigraphique, j'ai utilisé l'algorithme VMCVwB.

J'ai observé que le backtracking permettait d'améliorer le temps de convergence, dans le cas de la régularisation TV-h avec projection épigraphique. J'ai également observé que la méthode différentiable VMLM gagnait un facteur 30 en temps de calcul en secondes vis-à-vis des méthodes proximales VMFBwB et VMCVwB. Les temps de calcul des méthodes proximales est donc contraignant vis-à-vis du réglage optimal des hyperparamètres.

J'ai également observé que l'inclusion de la convolution dans le modèle permettait de diminuer fortement l'erreur d'estimation par rapport à une déconvolution disjointe de la reconstruction des cartes, d'un facteur 2 dans le cas de plus faible SNR. Une baisse de performance du modèle complet a fort SNR semble indiquer qu'il existe un choix de paramétrisation du modèle et de régularisation plus pertinent.

J'ai également observé, que pour les résultats obtenus dans le temps imparti, la régularisation par TV-h sans contrainte épigraphique donne de manière générale les résultats ayant l'erreur la plus petite. Dans le cas sur données simulées, nous avons vu qu'il y avait plusieurs choix d'hyperparamètres possibles permettant l'observation de structures différentes. C'est pourquoi il me semble nécessaire lors de la reconstruction de comparer les résultats pour au moins deux ensembles d'hyperparamètres, entrainant une légère sur-régularisation et une sous-régularisation, afin de s'assurer de ne manquer aucune information dans l'image.

Ce chapitre ouvre donc de nombreuses perspectives pour raffiner les résultats de déconvolution. D'une part, la qualité des images reconstruites avec déconvolution reste contrainte par la calibration des données, c'est-à-dire à une bonne soustraction du fond ainsi qu'une bonne estimation de la covariance des données, des pixels morts, de la PSF et du centre de l'étoile. Une amélioration de la calibration pourrait permettre de réduire les artefacts de déconvolution, cependant nous sommes limités par le nombre de données de calibration disponible. L'utilisation d'un critère repondéré, afin d'affiner la calibration de manière jointe à la reconstruction, est donc une perspective intéressante.

D'autre part, il serait intéressant de comparer des régularisations structurées où multi-échelle pour la reconstruction des disques. L'accélération numérique des algorithmes proximaux pourrait permettre une telle comparaison, cependant étant donné les performances de la méthode VMLM-B, il peut être intéressant d'étudier l'approximation hyperbolique des contraintes non-différentiables, telle que celle de TV pour TV-h. De plus, l'utilisation d'un modèle non-linéaire permettrait d'introduire la contrainte de positivité sur les intensités et une régularisation des paramètres plus pertinente tout en restant différentiable.

Enfin, le réglage automatique des hyperparamètres par SURE dans le cas de données astrophysiques ne donne qu'un aperçu de réglage optimal et l'utilisation d'autres critère tel que

SURE généralisé, pourrait permettre un meilleur réglage automatique. Cependant il semble intéressant de pouvoir garder la main sur ces paramètres afin de pouvoir choisir le réglage qui semble le plus approprié selon la résolution souhaitée.

Chapitre 5

Érreur d'estimation

Lors des précédents chapitres, nous nous sommes intéressés à l'estimation des paramètres $\overline{\overline{\boldsymbol{x}}} \in (\mathbb{R}^N)^L$ du modèle des données $\boldsymbol{d} \in (\mathbb{R}^M)^K$, s'écrivant sous la forme :

Rappel éq. (2.1)
$$\boldsymbol{d} = \mathcal{B}(f(\overline{\overline{\boldsymbol{x}}})),$$

où $f : (\mathbb{R}^N)^L \to (\mathbb{R}^M)^K$ est un opérateur de déformation linéaire fixé, prenant différentes formes selon que l'on soit séparable ou non, linéaire ou non, et $\mathcal{B} : (\mathbb{R}^M)^K \to (\mathbb{R}^M)^K$ un opérateur de dégradation aléatoire, correspondant dans notre cas à un bruit additif gaussien centré de matrice de covariance Σ.

Pour estimer ces paramètres, nous avons développé différentes méthodes basées sur l'approche inverse. Pour juger la qualité et la fiabilité des reconstructions obtenues à partir des différentes méthodes, le critère que nous avons choisi est l'Érreur Quadratique Moyenne (EQM) entre les paramètres estimés $\widehat{\boldsymbol{x}}$ et les paramètres vrais $\overline{\overline{\boldsymbol{x}}}$.

Dans le cas où la vérité terrain n'est pas connue, nous avons eu recours, dans le cas séparable, à la borne de Fréchet-Darmois-Cramer-Rao (FDCR), qui permet d'avoir une borne inférieure à l'EQM dans le cas non biaisé. Dans le cas non-séparable, nous avons eu recours au risque de Stein, qui est un estimateur de l'EQM.

Le but de ce chapitre est d'apporter plus d'explications théorique et empirique vis-à-vis de l'utilisation de ces différentes méthodes de l'estimation de l'erreur.

Ce chapitre est découpé comme suit. Dans la section 5.1, je développe le calcul de la borne de FDCR dans le cas de l'estimation de paramètres d'un modèle de données de le forme :

$$\forall k \in \{1, \dots K\}, \quad \boldsymbol{d}_k = f_k(\overline{\overline{\boldsymbol{x}}}) + \beta_k, \tag{5.1}$$

où $\overline{\overline{\boldsymbol{x}}} \in \mathbb{R}^L$ où L est le nombre de composantes $(\overline{\overline{\boldsymbol{x}}}_\ell)_{\ell \in \{1,\dots L\}} \in \mathbb{R}$, $f_k : \mathbb{R}^L \to \mathbb{R}^K$ et $\beta_k \sim \mathcal{N}(0, \sigma_k^2)$. J'étudie ensuite le comportement de la borne de FDCR selon que les opérateurs $(f_k)_{k \in \{1,\dots K\}}$ soient linéaires ou non. Je développe le calcul de l'approximation de la covariance de l'estimateur d'un modèle de données, à partir de la FDCR d'une autre paramétrisation de ce modèle. J'étudie ensuite la question plus complexe des modèles régularisés, nous introduisons alors l'estimateur non-biaisé du risque de Stein (SURE : *Stein Unbiased Risk Estimator*) dans un cas simple et montrons son efficacité pour le réglage automatique des hyperparamètres de régularisation.

Dans une seconde section nous développons tout d'abord le calcul de la borne de FDCR dans le cadre de la Méthode non-Linéaire Séparable (Algo. 5). Nous développons ensuite le

calcul de la borne de FDCR dans le cadre de la Méthode non-Linéaire Séparable (Algo. 6), à partir de laquelle nous calculons une approximation de la covariance des paramètres estimés par la MnLS. Nous étudions ensuite le comportement des EQM des paramètres estimés par la MnLS en fonction du SNR, ainsi que les différentes bornes.

Dans une troisième section, nous dérivons le critère SURE pour estimer l'Erreur Quadratique Moyenne du modèle dans le cas d'un modèle non-séparable et présentons des cartes de critère SURE pour les différentes méthodes non-séparables développées dans ce manuscrit, sans déconvolution présentées au chapitre 3 et avec déconvolution présentées au chapitre 4.

5.1 Erreur d'estimation dans le cas d'un bruit gaussien centré

Lors de l'estimation des paramètres d'un modèle, il est important d'avoir une information sur la qualité et la fiabilité des résultats. Les deux valeurs contenant cette information, sont la variance et le biais des estimateurs. Un estimateur de qualité est un estimateur dont la variance est faible. Comme présenté dans la section 1.2, l'ajout de contraintes dans le modèle peut aider à la minimisation de la variance. Cependant, il ne faut pas négliger le biais, qui indique à quel point on s'éloigne de notre solution. Le problème de l'ajout de contraintes est que cela augmente le biais. Il faut donc faire attention à garder un bon compromis entre la variance et le biais des estimateurs c'est-à-dire minimiser l'Erreur Quadratique Moyenne (EQM), qui pour un estimateur $\widehat{\boldsymbol{x}} \in \mathbb{R}^L$ du paramètre vrai $\overline{\boldsymbol{x}} \in \mathbb{R}^L$ est donné par :

$$\mathrm{EQM}(\widehat{\boldsymbol{x}}_\ell, \overline{\boldsymbol{x}}_\ell) = \mathbb{E}\Big[(\widehat{\boldsymbol{x}}_\ell - \overline{\boldsymbol{x}}_\ell)^2\Big]. \tag{5.2}$$

En effet celle-ci vérifie pour tout $\ell \in \{1, \dots L\}$:

$$\mathrm{EQM}(\widehat{\boldsymbol{x}}_\ell, \overline{\boldsymbol{x}}_\ell) = \mathrm{Var}(\widehat{\boldsymbol{x}}_\ell) + \mathrm{Biais}(\widehat{\boldsymbol{x}}_\ell, \overline{\boldsymbol{x}}_\ell)^2, \tag{5.3}$$

où

$$\mathrm{Var}(\widehat{\boldsymbol{x}}_\ell) = \mathbb{E}\Big[\Big(\widehat{\boldsymbol{x}}_\ell - \mathbb{E}[\widehat{\boldsymbol{x}}_\ell]\Big)^2\Big] \tag{5.4}$$

et

$$\mathrm{Biais}(\widehat{\boldsymbol{x}}_\ell, \overline{\boldsymbol{x}}_\ell) = \mathbb{E}[\widehat{\boldsymbol{x}}_\ell - \overline{\boldsymbol{x}}_\ell]. \tag{5.5}$$

La figure 1.7 illustre le fait que le minimum de l'EQM est atteint lorsqu'on a un compromis entre minimisation de la variance et minimisation du biais.

Dans le cas des données simulées, la vérité terrain est connue, il est donc possible de calculer directement l'Erreur Quadratique Moyenne et de juger la qualité des estimateurs via l'EQM normalisée. Cependant, dans le cas de l'application sur données astrophysiques, la vérité terrain n'est pas connue, il n'est pas possible de calculer la EQMn. Il est cependant possible d'avoir une borne minimale à l'EQM à partir de la borne de FDCR.

En effet, pour un estimateur sans biais, on a

$$\mathrm{Cov}(\widehat{\boldsymbol{x}}) \succeq \mathrm{FDCR}(\boldsymbol{x}), \tag{5.6}$$

où $\succeq$ représente l'ordre de Loewner défini par l'équation (A.2) c'est-à-dire :

$$\forall \boldsymbol{x} \in \mathbb{R}^L, \langle \boldsymbol{x}, \mathrm{Cov}(\widehat{\boldsymbol{x}})\boldsymbol{x}\rangle \geq \langle \boldsymbol{x}, \mathrm{FDCR}(\boldsymbol{x})\boldsymbol{x}\rangle. \tag{5.7}$$

et

$$\mathrm{FDCR}(\boldsymbol{x}) = \mathbf{I}(\boldsymbol{x})^{-1}, \tag{5.8}$$

où $\mathbf{I}(\boldsymbol{x})$ représente la matrice d'information de Fisher du paramètre $\boldsymbol{x}$. Or, un estimateur des moindres carrés est asymptotiquement sans biais. C'est-à-dire que plus il y a de réalisations du bruit, et donc que le SNR est grand, plus le biais s'approche de 0. Dans ce cas, on a pour tout $\ell \in \{1, \dots L\}$ que $\mathrm{EQMn}(\widehat{\boldsymbol{x}}_\ell, \overline{\widehat{\boldsymbol{x}}}_\ell) \approx \mathrm{Var}(\widehat{\boldsymbol{x}}_\ell) \geq \mathrm{FDCR}(\boldsymbol{x})_{\ell,\ell}$.

5.1.1 Information de Fisher d'un modèle gaussien centré

La borne de FDCR s'obtient en prenant la pseudo-inverse de l'information de Fisher $\mathbf{I}(\boldsymbol{x})$. Dans le cas où l'information de Fisher est inversible, alors la borne de FDCR est l'inverse de $\mathcal{I}(\boldsymbol{x})$. De manière générale, l'information de Fisher est obtenue par :

$$\mathbf{I}(\boldsymbol{x}) = \mathbb{E}\Big[\nabla\mathcal{L}(\boldsymbol{d}|\boldsymbol{x})\nabla\mathcal{L}(\boldsymbol{d}|\boldsymbol{x})^\top\Big] = -\mathbb{E}\Big[\nabla^2\mathcal{L}(\boldsymbol{d}|\boldsymbol{x})\Big], \tag{5.9}$$

où $\mathcal{L}(\boldsymbol{d}|\cdot) : \mathbb{R}^L \to \mathbb{R}$ représente la co-logvraisemblance des données $\boldsymbol{d} \in \mathbb{R}^K$ paramétrées en $\boldsymbol{x} \in \mathbb{R}^L$, où ∇ représente le gradient et ∇^2 la matrice hessienne de $\mathcal{L}$.

Dans d'un modèle gaussien centré tel que celui exprimé à l'équation (5.1), la co-logvraissemblance des données est donnée, à une constante près, par :

$$\mathcal{L}(\boldsymbol{d}|\boldsymbol{x}) = -\sum_{k=1}^{K} \frac{\big(\boldsymbol{d}_k - f_k(\boldsymbol{x})\big)^2}{2\sigma_k^2}. \tag{5.10}$$

Proposition 5.1.1. *Soit $\boldsymbol{d} \in \mathbb{R}^K$ un ensemble de données suivant une loi gaussienne centrée telle que $\boldsymbol{d}_k \sim \mathcal{N}(f_k(\boldsymbol{x}), \sigma_k^2)$, où $\boldsymbol{x} \in \mathbb{R}^L$ est le paramètre à estimer. Alors, en posant $\mathbf{W}_k = (\sigma_k^2)^{-1}$, l'information de Fisher est donnée pour tout $\boldsymbol{x} \in \mathbb{R}^L$ par :*

$$\mathbf{I}(\boldsymbol{x}) = \sum_k \nabla f_k(\boldsymbol{x})\mathbf{W}_k\nabla f_k(\boldsymbol{x})^\top. \tag{5.11}$$

Démonstration : *(Fin page 137)* On a :

$$\nabla^2\mathcal{L}(\boldsymbol{d}|\boldsymbol{x}) = -\sum_k \Big[\nabla f_k(\boldsymbol{x})\mathbf{W}_k\nabla f_k(\boldsymbol{x})^\top - \nabla^2 f_k(\boldsymbol{x})\mathbf{W}_k\big(\boldsymbol{d}_k - f_k(\boldsymbol{x})\big)\Big]. \tag{5.12}$$

Comme $\mathbb{E}[\boldsymbol{d}_k] = f_k(\boldsymbol{x})$, on a par linéarité de l'espérance :

$$-\mathbb{E}\Big[\nabla^2\mathcal{L}(\boldsymbol{d}|\boldsymbol{x})\Big] = \sum_k \mathbb{E}\Big[\nabla f_k(\boldsymbol{x})\mathbf{W}_k\nabla f_k(\boldsymbol{x})^\top - \nabla^2 f_k(\boldsymbol{x})\mathbf{W}_k\big(\boldsymbol{d}_k - f_k(\boldsymbol{x})\big)\Big] \tag{5.13}$$

$$= \sum_k \Big[\nabla f_k(\boldsymbol{x})\mathbf{W}_k\nabla f_k(\boldsymbol{x})^\top - \nabla^2 f_k(\boldsymbol{x})\mathbf{W}_k\big(\mathbb{E}[\boldsymbol{d}_k] - f_k(\boldsymbol{x})\big)\Big] \tag{5.14}$$

$$= \sum_k \nabla f_k(\boldsymbol{x})\mathbf{W}_k\nabla f_k(\boldsymbol{x})^\top. \tag{5.15}$$

□

5.1.2 Étude théorique de la borne de FDCR dans le cas non-biaisé

Dans le cadre général, l'expression de la borne de FDCR est :

$$\mathrm{FDCR}(\boldsymbol{x}) = \mathbf{B}\mathbf{I}(\boldsymbol{x})^{-1}\mathbf{B}^{\top} \tag{5.16}$$

où $\mathbf{B}$ est la matrice jacobienne donnée par

$$\forall \ell_1, \ell_2 \in \{1, \dots L\}, \quad (\mathbf{B})_{\ell_1,\ell_2} = \frac{\partial \mathbb{E}[\widehat{\boldsymbol{x}}]_{\ell_1}}{\partial \boldsymbol{x}_{\ell_2}}. \tag{5.17}$$

Sachant qu'un estimateur sans biais signifie que $\mathbb{E}[\widehat{\boldsymbol{x}}] = \overline{\boldsymbol{x}}$, dans un tel cas $\mathbf{B} = \mathbf{Id}$, où $\mathbf{Id}$ représente la matrice identité de $\mathcal{M}_L(\mathbb{R})$, et donc $\mathrm{FDCR}(\boldsymbol{x}) = \mathbf{I}(\boldsymbol{x})^{-1}$.

Dans le cas où $f_k : \mathbb{R}^L \to \mathbb{R}$ est une fonction linéaire, c'est-à-dire que f_k peut s'écrire pour tout $k \in \{1, \dots k\}$ de la forme $f_k(\boldsymbol{x}) = \mathbf{A}_k\boldsymbol{x}$ avec $\mathbf{A}_k \in \mathcal{M}_L(\mathbb{R})$, alors la borne de FDCR est exacte, c'est-à-dire qu'elle ne dépend pas de $\boldsymbol{x}$. En effet, l'information de Fisher dans un tel cas vaut :

$$\mathbf{I}(\boldsymbol{x}) = \sum_k \mathbf{A}_k\mathbf{A}_k^*\mathbf{W}_k\mathbf{A}_k^*\mathbf{A}_k. \tag{5.18}$$

où * représente l'adjoint au sens matriciel.

Dans le cas non-linéaire en revanche, ∇f_k dépend de $\boldsymbol{x}$. La borne de FDCR exacte se calcule alors en le paramètre vrai $\overline{\boldsymbol{x}}$. Cependant, dans le cas où $\overline{\boldsymbol{x}}$ n'est pas connue, il n'est pas possible de calculer la borne de FDCR exacte. Elle peut cependant se calculer en l'estimation $\widehat{\boldsymbol{x}}$ de $\boldsymbol{x}$, car on s'attend à ce que la borne de FDCR ne varie que très peu à l'échelle des erreurs sur les paramètres. En revanche la non-linéarité peut poser problème dans les cas où il existe $\ell \in \{1, \dots L\}$ tel que $\left(\nabla f_k(\boldsymbol{x})\right)_\ell = 0$. En effet, dans un tel cas, la matrice d'information de Fisher n'est plus de rang plein. Elle ne peut alors donner aucune information vis-à-vis de l'erreur sur le paramètre $\boldsymbol{x}_\ell$.

Dans le cas où $\left(\nabla f_k(\boldsymbol{x})\right)_\ell \approx 0$, c'est-à-dire d'un ordre inférieur à 10^{-3}, le problème reste le même. La matrice est très mal conditionnée ce qui peut introduire des erreurs numériques. De plus, la borne de FDCR a alors des valeurs très grandes, c'est-à-dire d'un ordre supérieur à 10^{-3}, ce qui implique une borne minimale très grande qui peut être bien plus grande que l'erreur estimée empiriquement.

Cependant, dans le cas où l'on calcule la borne de FDCR à partir de $\widehat{\boldsymbol{x}}$, il est possible d'avoir des valeurs de $\left(\nabla f_k(\widehat{\boldsymbol{x}})\right)_\ell \geq \left(\nabla f_k(\overline{\boldsymbol{x}})\right)_\ell$ de plusieurs ordres de grandeurs, la borne de FDCR calculée en $\widehat{\boldsymbol{x}}$ est alors plus petite que la borne de FDCR exacte calculée à partir de $\overline{\boldsymbol{x}}$. Ce qui peut mener à sous-estimer l'erreur et donc à considérer comme juste des erreurs de reconstruction.

Pour palier ce problème, l'idée est de faire apparaître un changement de variable permettant d'utiliser une borne de FDCR exacte ne dépendant pas de $\boldsymbol{x}$.

Supposons que les données $\boldsymbol{d}_k \in \mathbb{R}$ peuvent être décrite par un modèle linéaire, paramétré en l'estimateur $\boldsymbol{x} \in \mathbb{R}^L$, de moyenne $\mathbb{E}[\boldsymbol{x}]$ et de covariance $\mathrm{Cov}(\boldsymbol{x})$, mais aussi par un modèle non-linéaire paramétré en $q(\boldsymbol{x}) \in \mathbb{R}^L$, de moyenne $\mathbb{E}[q(\boldsymbol{x})]$ et de covariance $\mathrm{Cov}(q(\boldsymbol{x}))$, où $q : \mathbb{R}^L \to \mathbb{R}^L$ est une fonction bijective. Cette fonction permet de passer des paramètres du modèle linéaire au modèle non-linéaire. Comme on a $\mathrm{Cov}(\boldsymbol{x}) \geq \mathrm{FDCR}(\boldsymbol{x})$ avec $\mathrm{FDCR}(\boldsymbol{x})$ qui ne dépend pas de $\boldsymbol{x}$, on voudrait pouvoir exprimer $\mathrm{Cov}(q(\boldsymbol{x}))$ en fonction de $\mathrm{Cov}(\boldsymbol{x})$ et donc d'obtenir une borne inférieure à $\mathrm{Cov}(q(\boldsymbol{x}))$ dépendant de $\mathrm{FDCR}(\boldsymbol{x})$ et non de $\mathrm{FDCR}(q(\boldsymbol{x}))$. Pour cela on utilise la proposition suivante :

Proposition 5.1.2. *Soit $\boldsymbol{x} \in R^L$ une variable aléatoire de moyenne $\mathbb{E}[\boldsymbol{x}]$ et de covariance $Cov(\boldsymbol{x})$, et $q(\boldsymbol{x}) \in \mathbb{R}^L$ une variable aléatoire fonction de $\boldsymbol{x}$, de moyenne $\mathbb{E}[q(\boldsymbol{x})]$ de covariance $Cov(q(\boldsymbol{x}))$. Alors*

$$Cov(q(\boldsymbol{x})) = \mathbf{J}_q\Big(\mathbb{E}[\boldsymbol{x}]\Big) Cov(\boldsymbol{x}) \mathbf{J}_q\Big(\mathbb{E}[\boldsymbol{x}]\Big)^\top, \tag{5.19}$$

où $\mathbf{J}_q$ représente la matrice jacobienne de q et $^\top$ l'opérateur transposé.

Démonstration : *(Fin page 140)* Pour exprimer $\mathrm{Cov}(q(\boldsymbol{x}))$ en fonction de $\mathrm{Cov}(\boldsymbol{x})$, il suffit de faire un développement limité de $q(\boldsymbol{x})$ au voisinage de $\mathbb{E}[\boldsymbol{x}]$ et de calculer la moyenne $\mathbb{E}[q(\boldsymbol{x})]$ et la covariance $\mathrm{Cov}(q(\boldsymbol{x}))$ à partir de ces approximations. En posant pour tout $\ell \in \{1, \dots L\}$ $q_\ell(\boldsymbol{x})$ la ℓième composante de $q(V x)$, on a le développement limité suivant, pour tout $\ell_1 \in \{1, \dots L\}$:

$$\begin{aligned} q_{\ell_1}(\boldsymbol{x}) = q_{\ell_1}\Big(\mathbb{E}[\boldsymbol{x}]\Big) + \sum_{\ell_2=1}^{L} \Big(\boldsymbol{x}_{\ell_2} - \mathbb{E}[\boldsymbol{x}_{\ell_2}]\Big) \frac{\partial q_{\ell_1}\Big(\mathbb{E}[\boldsymbol{x}]\Big)}{\partial \boldsymbol{x}_{\ell_2}} \\ + \frac{1}{2} \sum_{\ell_2=1}^{L} \sum_{\ell_3=1}^{L} \Big(\boldsymbol{x}_{\ell_2} - \mathbb{E}[\boldsymbol{x}_{\ell_2}]\Big)\Big(\boldsymbol{x}_{\ell_3} - \mathbb{E}[\boldsymbol{x}_{\ell_3}]\Big) \frac{\partial^2 q_{\ell_1}\Big(\mathbb{E}[\boldsymbol{x}]\Big)}{\partial \boldsymbol{x}_{\ell_2} \partial \boldsymbol{x}_{\ell_3}} + o\Big(\big\|\boldsymbol{x} - \mathbb{E}[\boldsymbol{x}]\big\|^2\Big). \end{aligned} \tag{5.20}$$

On a alors à partir de cette approximation, d'une part pour tout $\ell_1 \in \{1, \dots L\}$:

$$\mathbb{E}\Big[q_{\ell_1}(\boldsymbol{x})\Big] \approx q_{\ell_1}\Big(\mathbb{E}[\boldsymbol{x}]\Big) + \frac{1}{2} \sum_{\ell_2=1}^{L} \sum_{\ell_3=1}^{L} \mathrm{Cov}(\boldsymbol{x})_{\ell_2,\ell_3} \frac{\partial^2 q_{\ell_1}\Big(\mathbb{E}[\boldsymbol{x}]\Big)}{\partial \boldsymbol{x}_{\ell_2} \partial \boldsymbol{x}_{\ell_3}}, \tag{5.21}$$

car $\mathbb{E}\Big[\Big(\boldsymbol{x}_{\ell_2} - \mathbb{E}[\boldsymbol{x}_{\ell_2}]\Big)\Big] = 0$ et $\mathbb{E}\Big[\Big(\boldsymbol{x}_{\ell_2} - \mathbb{E}[\boldsymbol{x}_{\ell_2}]\Big)\Big(\boldsymbol{x}_{\ell_3} - \mathbb{E}[\boldsymbol{x}_{\ell_3}]\Big)\Big] = \mathrm{Cov}(\boldsymbol{x})_{\ell_2,\ell_3}$. D'autre part, pour tout $\ell_1, \ell_2 \in \{1, \dots L\}$:

$$\mathrm{Cov}(q(\boldsymbol{x}))_{\ell_1,\ell_2} \approx \sum_{\ell_3=1}^{L} \sum_{\ell_4=1}^{L} \mathrm{Cov}(\boldsymbol{x})_{\ell_3,\ell_4} \frac{\partial q_{\ell_1}\Big(\mathbb{E}[\boldsymbol{x}]\Big)}{\partial \boldsymbol{x}_{\ell_3}} \frac{\partial q_{\ell_2}\Big(\mathbb{E}[\boldsymbol{x}]\Big)}{\partial \boldsymbol{x}_{\ell_4}}. \tag{5.22}$$

En effet :

$$\mathrm{Cov}(q(\boldsymbol{x}))_{\ell_1,\ell_2} = \mathbb{E}\Big[q(\boldsymbol{x}_{\ell_1}) q(\boldsymbol{x}_{\ell_2})\Big] - \mathbb{E}\Big[q(\boldsymbol{x}_{\ell_1})\Big] \mathbb{E}\Big[q(\boldsymbol{x}_{\ell_2})\Big] \tag{5.23}$$

avec d'une part d'après l'équation (5.21) :

$$\begin{aligned} \mathbb{E}\Big[q(\boldsymbol{x}_{\ell_1})\Big] \mathbb{E}\Big[q(\boldsymbol{x}_{\ell_2})\Big] = & q_{\ell_1}\Big(\mathbb{E}[\boldsymbol{x}]\Big) q_{\ell_2}\Big(\mathbb{E}[\boldsymbol{x}]\Big) + \frac{1}{2} \sum_{\ell_3=1}^{L} \sum_{\ell_4=1}^{L} \mathrm{Cov}(\boldsymbol{x})_{\ell_3,\ell_4} \Bigg(q_{\ell_1}\Big(\mathbb{E}[\boldsymbol{x}]\Big) \frac{\partial^2 q_{\ell_2}\Big(\mathbb{E}[\boldsymbol{x}]\Big)}{\partial \boldsymbol{x}_{\ell_3} \partial \boldsymbol{x}_{\ell_4}} \\ & + q_{\ell_2}\Big(\mathbb{E}[\boldsymbol{x}]\Big) \frac{\partial^2 q_{\ell_1}\Big(\mathbb{E}[\boldsymbol{x}]\Big)}{\partial \boldsymbol{x}_{\ell_3} \partial \boldsymbol{x}_{\ell_4}} \Bigg) + o\Big(\big\|\boldsymbol{x} - \mathbb{E}[\boldsymbol{x}]\big\|^2\Big) \end{aligned} \tag{5.24}$$

et d'autre part d'après l'équartion (5.20) :

$$\mathbb{E}\Big[q(\boldsymbol{x}_{\ell_1})q(\boldsymbol{x}_{\ell_2})\Big] = q_{\ell_1}\Big(\mathbb{E}[\boldsymbol{x}]\Big)q_{\ell_2}\Big(\mathbb{E}[\boldsymbol{x}]\Big) + \frac{1}{2}\sum_{\ell_3=1}^{L}\sum_{\ell_4=1}^{L}\mathrm{Cov}(\boldsymbol{x})_{\ell_3,\ell_4}\Bigg(q_{\ell_1}\Big(\mathbb{E}[\boldsymbol{x}]\Big)\frac{\partial^2 q_{\ell_2}\Big(\mathbb{E}[\boldsymbol{x}]\Big)}{\partial \boldsymbol{x}_{\ell_3}\partial \boldsymbol{x}_{\ell_4}} + q_{\ell_2}\Big(\mathbb{E}[\boldsymbol{x}]\Big)\frac{\partial^2 q_{\ell_1}\Big(\mathbb{E}[\boldsymbol{x}]\Big)}{\partial \boldsymbol{x}_{\ell_3}\partial \boldsymbol{x}_{\ell_4}}\Bigg) + \sum_{\ell_3=1}^{L}\sum_{\ell_4=1}^{L}\mathrm{Cov}(\boldsymbol{x})_{\ell_3,\ell_4}\frac{\partial q_{\ell_1}\Big(\mathbb{E}[\boldsymbol{x}]\Big)}{\partial \boldsymbol{x}_{\ell_3}}\frac{\partial q_{\ell_2}\Big(\mathbb{E}[\boldsymbol{x}]\Big)}{\partial \boldsymbol{x}_{\ell_4}} + o\Big(\big\|\boldsymbol{x}-\mathbb{E}[\boldsymbol{x}]\big\|^2\Big). \tag{5.25}$$

Les termes des équations (5.24) et (5.25) se simplifient, d'où l'équation (5.22) et donc

$$\forall \ell_1,\ell_2 \in \{1,\dots L\},\quad \mathrm{Cov}(q(\boldsymbol{x}))_{\ell_1,\ell_2} = \nabla q_{\ell_1}\Big(\mathbb{E}[\boldsymbol{x}]\Big)^\top \mathrm{Cov}(\boldsymbol{x})\nabla q_{\ell_2}\Big(\mathbb{E}[\boldsymbol{x}]\Big), \tag{5.26}$$

ce qui est équivalent, comme par définition de la matrice jacobienne

$$\mathbf{J}_q(\boldsymbol{x}) = \begin{pmatrix}\nabla q_1(\boldsymbol{x})^\top \\ \vdots \\ \nabla q_L(\boldsymbol{x})^\top\end{pmatrix} = \begin{pmatrix}\frac{\partial q_1(\boldsymbol{x})}{\partial \boldsymbol{x}_1} & \cdots & \frac{\partial q_1(\boldsymbol{x})}{\partial \boldsymbol{x}_L} \\ \vdots & \ddots & \vdots \\ \frac{\partial q_L(\boldsymbol{x})}{\partial \boldsymbol{x}_1} & \cdots & \frac{\partial q_L(\boldsymbol{x})}{\partial \boldsymbol{x}_L}\end{pmatrix}, \tag{5.27}$$

à l'équation (5.19).

□

Enfin, il est possible d'avoir une borne inférieure à la covariance de $q(\boldsymbol{x})$, et donc à l'EQM, à partir de la relation suivante :

$$\mathrm{EQM}\Big(q(\boldsymbol{x})\Big) \approx \mathrm{Cov}\Big(q(\boldsymbol{x})\Big) \geq \mathbf{J}_q\Big(\mathbb{E}[\boldsymbol{x}]\Big)\mathrm{FDCR}(\boldsymbol{x})\mathbf{J}_q^\top\Big(\mathbb{E}[\boldsymbol{x}]\Big). \tag{5.28}$$

En effet, d'après la relation (5.7) entre la covariance et la borne de FDCR, on a en particulier :

$$\begin{aligned}\forall \boldsymbol{y} \in \mathbb{R}^L, \langle \mathbf{J}_q\Big(\mathbb{E}[\boldsymbol{x}]\Big)^\top \boldsymbol{y}, \mathrm{Cov}(\widehat{\boldsymbol{x}})\mathbf{J}_q\Big(\mathbb{E}[\boldsymbol{x}]\Big)^\top \boldsymbol{y}\rangle &\geq \langle \mathbf{J}_q\Big(\mathbb{E}[\boldsymbol{x}]\Big)^\top \boldsymbol{y}, \mathrm{FDCR}(\boldsymbol{x})\mathbf{J}_q\Big(\mathbb{E}[\boldsymbol{x}]\Big)^\top \boldsymbol{y}\rangle \\ \Leftrightarrow \langle \boldsymbol{y}, \mathbf{J}_q\Big(\mathbb{E}[\boldsymbol{x}]\Big)\mathrm{Cov}(\widehat{\boldsymbol{x}})\mathbf{J}_q\Big(\mathbb{E}[\boldsymbol{x}]\Big)^\top \boldsymbol{y}\rangle &\geq \langle \boldsymbol{y}, \mathbf{J}_q\Big(\mathbb{E}[\boldsymbol{x}]\Big)\mathrm{FDCR}(\boldsymbol{x})\mathbf{J}_q\Big(\mathbb{E}[\boldsymbol{x}]\Big)^\top \boldsymbol{y}\rangle.\end{aligned} \tag{5.29}$$

5.1.3 Estimation de l'erreur dans le cas d'un estimateur artificiellement biaisé

Dans le cas d'un modèle des données $\boldsymbol{d} \in \mathbb{R}^K$ de la forme $\boldsymbol{d} = f(\overline{\boldsymbol{x}}) + \boldsymbol{\beta}$ avec $\boldsymbol{\beta} \sim \mathcal{N}(0, \Sigma^2)$, où $f : \mathbb{R}^L \to \mathbb{R}^K$ est mal conditionnée, c'est-à-dire de valeur propre minimale nulle ou proche de 0, l'estimation du paramètre $\boldsymbol{x} \in \mathbb{R}^L$ à partir du maximum de vraisemblance amplifie le bruit. Cela se traduit par une très grande variance des paramètres estimés.

Afin de réduire cette variance, il est possible d'introduire de l'information *a priori* sur l'estimateur, sous forme de régularisation et de contrainte. Cependant, en complexifiant ainsi le modèle, on augmente le biais de la solution. Avoir une information sur la covariance des données ne suffit donc plus à nous donner une borne inférieure à l'EQM. En effet, comme représenté dans la figure 1.7, la variance la plus petite ne donne pas forcément la valeur de

l'EQM la plus petite. Pour minimiser l'EQM, il faut équilibrer la complexité du modèle pour avoir le meilleur compromis entre le biais et la variance. Cette équilibre ce gère à partir d'un ou plusieurs hyperparamètres, qui vont servir à "doser" la contribution de la contrainte. Le choix de ces hyperparamètres doit donc se faire en cherchant l'ensemble des hyperparamètres minimisant l'EQM.

Dans le cas de l'application sur donnée simulées, la vérité terrain est connue. Il est donc possible de calculer l'EQM et d'équilibrer les contraintes de manière à la minimiser. En revanche, dans le cas de données réelles, il faut avoir recours à d'autres méthodes.

Un grand nombre de bornes et d'approximations sont disponibles dans la littérature, une liste non exhaustive peut être trouvée dans la section 1.2.5.

Dans, cette thèse j'ai fait le choix d'utiliser l'estimateur non biaisé du risque de Stein (SURE : *Stein Unbiased Risk Estimator*). Le but est de changer de métrique et d'au lieu de minimiser l'EQM entre $\widehat{\boldsymbol{x}}$ et $\overline{\overline{\boldsymbol{x}}}$, c'est-à-dire :

$$\mathrm{EQM}^{\text{paramètres}}(\widehat{\boldsymbol{x}}) = \mathbb{E}\Big[\|\widehat{\boldsymbol{x}} - \overline{\overline{\boldsymbol{x}}}\|^2\Big], \tag{5.30}$$

de minimiser l'estimation de l'EQM pondéré par l'inverse de la variance des données, c'est-à-dire $\mathbf{W} = \Sigma^{-2}$, entre $f(\widehat{\boldsymbol{x}})$ et $f(\overline{\overline{\boldsymbol{x}}})$, c'est-à-dire :

$$\mathrm{EQM}^{\text{données}}(\widehat{\boldsymbol{x}}) = \mathbb{E}\Big[\|f(\widehat{\boldsymbol{x}}) - f(\overline{\overline{\boldsymbol{x}}})\|_{\mathbf{W}}^2\Big]. \tag{5.31}$$

On appelle également cette erreur, erreur de prédiction, là où l'erreur sur les paramètres est appelée erreur d'estimation. On espère que les deux erreurs aient un profil similaire et que leurs minima soient situés dans une même région.

Comme $\overline{\overline{\boldsymbol{x}}}$ est inconnu, on minimise alors SURE qui est un estimateur non biaisé de (5.31). Notons $\mu \in \mathbb{R}$ l'hyperparamètre à régler et $\widehat{\boldsymbol{x}}_\mu \in R^L$ le paramètre d'intérêt estimé pour cette valeur d'hyperparamètre. On note également $\mathcal{K} = \{k \in \{1, \dots K\} \text{ tels que } \mathbf{W}_k \neq 0\}$, c'est-à-dire l'ensemble des mesures valides, ou encore, l'ensemble des mesures dont la variance est finie. Le critère SURE est alors donné par :

$$\mathrm{SURE}(\widehat{\boldsymbol{x}}_\mu) = \left\|\boldsymbol{d} - f(\widehat{\boldsymbol{x}}_\mu)\right\|_{\mathbf{W}}^2 + 2\ \mathrm{tr}\left(\frac{\partial f(\widehat{\boldsymbol{x}}_\mu)}{\partial \boldsymbol{d}}\right)\Big|_{\mathcal{K}} - \mathbf{Card}(\mathcal{K}) \tag{5.32}$$

où $\mathbf{Card}(\cdot)$ représente le cardinal de l'ensemble, c'est-à-dire le nombre d'éléments qu'il contient. On a alors d'une part :

$$\mathbb{E}\Big[\mathrm{SURE}(\widehat{\boldsymbol{x}}_\mu)\Big] = \mathrm{EQM}^{\text{données}}(\widehat{\boldsymbol{x}}_\mu)$$

En effet :

$$\begin{aligned}
\mathbb{E}\Big[\|f(\widehat{\boldsymbol{x}}_\mu) - f(\overline{\overline{\boldsymbol{x}}})\|_{\mathbf{W}}^2\Big] &= \mathbb{E}\Big[\|f(\widehat{\boldsymbol{x}}_\mu) - f(\overline{\overline{\boldsymbol{x}}}) - \boldsymbol{\beta} + \boldsymbol{\beta}\|_{\mathbf{W}}^2\Big] \\
&= \mathbb{E}\Big[\|f(\widehat{\boldsymbol{x}}_\mu) - \boldsymbol{d} + \boldsymbol{\beta}\|_{\mathbf{W}}^2\Big] \\
&= \mathbb{E}\Big[\|f(\widehat{\boldsymbol{x}}_\mu) - \boldsymbol{d}\|_{\mathbf{W}}^2 + 2(f(\widehat{\boldsymbol{x}}_\mu) - \boldsymbol{d})^\top \mathbf{W}\boldsymbol{\beta} + \|\boldsymbol{\beta}\|_{\mathbf{W}}^2\Big]. \qquad (5.33) \\
& \qquad (5.34)
\end{aligned}$$

Supposons que $(f(\widehat{\boldsymbol{x}}_\mu) - \boldsymbol{d})$ est à dérivée faible. On utilise alors le lemme de Stein et le fait que $\boldsymbol{\beta} = \boldsymbol{d} - f(\overline{\overline{\boldsymbol{x}}})$, dont les étapes dans notre cas sont les suivantes :

$$\mathbb{E}\Big[(f(\widehat{\boldsymbol{x}}_\mu) - \boldsymbol{d})^\top \mathbf{W}\boldsymbol{\beta}\Big] = \int (f(\widehat{\boldsymbol{x}}_\mu) - \boldsymbol{d})^\top \mathbf{W}\boldsymbol{\beta} \mathrm{e}^{-\beta^\top \mathbf{W}\beta} d\boldsymbol{d} \tag{5.35}$$

$$= -\int (f(\widehat{\boldsymbol{x}}_\mu) - \boldsymbol{d})^\top \frac{\partial \mathrm{e}^{-\beta^\top \mathbf{W}\beta}}{\partial \boldsymbol{d}} d\boldsymbol{d}\Big|_{\mathcal{K}} \tag{5.36}$$

car $\frac{\partial \mathrm{e}^{-\beta^\top \mathbf{W}\beta}}{\partial \boldsymbol{d}} = \boldsymbol{\beta}\mathbf{W}\mathrm{e}^{-\beta^\top \mathbf{W}\beta}$. Nous avons fait apparaître la restriction à $\mathcal{K}$ induite par les entrées nulles de la matrice $\mathbf{W}$.

En utilisant la définition de la dérivée faible, on a alors :

$$-\int (f(\widehat{\boldsymbol{x}}_\mu) - \boldsymbol{d})^\top \frac{\partial \mathrm{e}^{-\beta^\top \mathbf{W}\beta}}{\partial \boldsymbol{d}} d\boldsymbol{d}\Big|_{\mathcal{K}_k} = -\sum_{k\in\mathcal{K}} \int (f_k(\widehat{\boldsymbol{x}}_\mu) - \boldsymbol{d}_k) \frac{\partial \mathrm{e}^{-\beta^\top \mathbf{W}\beta}}{\partial \boldsymbol{d_k}} d\boldsymbol{d} \tag{5.37}$$

$$= \sum_{k\in\mathcal{K}} \int \frac{\partial (f_k(\widehat{\boldsymbol{x}}_\mu) - \boldsymbol{d_k})}{\partial \boldsymbol{d_k}} \mathrm{e}^{-\beta^\top \mathbf{W}\beta} d\boldsymbol{d} \tag{5.38}$$

$$= \sum_{k\in\mathcal{K}} \mathbb{E}\Big[\frac{\partial (f_k(\widehat{\boldsymbol{x}}_\mu) - \boldsymbol{d}_k)}{\partial \boldsymbol{d}_k}\Big] \tag{5.39}$$

$$= \mathbb{E}\Big[\mathrm{tr}\Big(\frac{\partial (f(\widehat{\boldsymbol{x}}_\mu) - \boldsymbol{d})}{\partial \boldsymbol{d}}\Big)\Big|_{\mathcal{K}}\Big] \tag{5.40}$$

par linéarité de la somme. Finalement, on obtient que

$$\begin{aligned}\mathbb{E}\Big[\langle (f(\widehat{\boldsymbol{x}}_\mu) - \boldsymbol{d}), \mathbf{W}\boldsymbol{\beta}\rangle\Big] &= \mathbb{E}\Big[\mathrm{tr}\Big(\frac{\partial f(\widehat{\boldsymbol{x}}_\mu) - \boldsymbol{d})}{\partial \boldsymbol{d}}\Big)\Big|_{\mathcal{K}}\Big] \\ &= \mathbb{E}\Big[\mathrm{tr}\Big(\frac{\partial f(\widehat{\boldsymbol{x}}_\mu)}{\partial \boldsymbol{d}} - \mathbf{Id}\Big)\Big|_{\mathcal{K}}\Big] \\ &= \mathbb{E}\Big[\mathrm{tr}\Big(\frac{\partial f(\widehat{\boldsymbol{x}}_\mu)}{\partial \boldsymbol{d}}\Big)\Big|_{\mathcal{K}}\Big] - \mathbf{Card}(\mathcal{K}) \end{aligned} \tag{5.41}$$

En remplaçant le terme croisé dans (5.33) par (5.41), on a donc :

$$\mathbb{E}\Big[\|f(\widehat{\boldsymbol{x}}_\mu) - f(\overline{\overline{\boldsymbol{x}}})\|^2_{\mathbf{W}}\Big] = \mathbb{E}\Big[\|f(\widehat{\boldsymbol{x}}_\mu) - \boldsymbol{d}\|^2_{\mathbf{W}} + 2\mathrm{tr}\Big(\frac{\partial f(\widehat{\boldsymbol{x}}_\mu)}{\partial \boldsymbol{d}}\Big)\Big|_{\mathcal{K}} - 2\mathbf{Card}(\mathcal{K}) + \|\boldsymbol{\beta}\|^2_{\mathbf{W}}\Big]. \tag{5.42}$$

D'autre part, par définition de la covariance, on a que :

$$\|\boldsymbol{\beta}\|^2_{\mathbf{W}} = \mathbf{Card}(\mathcal{K}). \tag{5.43}$$

D'où

$$\mathbb{E}\Big[\|f(\widehat{\boldsymbol{x}}_\mu) - f(\overline{\overline{\boldsymbol{x}}})\|^2_{\mathbf{W}}\Big] = \mathbb{E}\Big[\|f(\widehat{\boldsymbol{x}}_\mu) - \boldsymbol{d}\|^2_{\mathbf{W}} + 2\mathrm{tr}\Big(\frac{\partial f(\widehat{\boldsymbol{x}}_\mu)}{\partial \boldsymbol{d}}\Big)\Big|_{\mathcal{K}} - \mathbf{Card}(\mathcal{K})\Big] \tag{5.44}$$

$$= \mathbb{E}\Big[\mathrm{SURE}(\widehat{\boldsymbol{x}}_\mu)\Big]. \tag{5.45}$$

Pour calculer la trace, nous appuyons sur l'approximation de la divergence présentée dans [Ramani et al., 2008]. Il montre que

$$\mathrm{tr}\left(\frac{\partial f(\widehat{\boldsymbol{x}}_\mu)}{\partial \boldsymbol{d}}\right) = \mathrm{div} f(\widehat{\boldsymbol{x}}_\mu) \tag{5.46}$$

où div représente la divergence du modèle par rapport aux données et propose une méthode de Monte-Carlo pour estimer cette divergence. Soit $\boldsymbol{p} \sim \mathcal{N}(0_{\mathbb{R}^K}, \mathbf{Id}_{\mathbb{R}^K \times \mathbb{R}^K})$ une perturbation centrée et réduite et une petite valeur $\varepsilon > 0$. On note alors

$$\boldsymbol{d} = \boldsymbol{d} + \varepsilon \boldsymbol{p} \tag{5.47}$$

les données perturbées par une bruit gaussien centré de variance ε^2 et $\delta\widehat{\boldsymbol{x}}_\mu$ l'estimateur des données $\delta\boldsymbol{d}$ à partir du modèle f. Alors on a :

$$\begin{aligned}
\mathrm{div} f(\widehat{\boldsymbol{x}}_\mu) &\approx \frac{\langle \boldsymbol{p}, f(\delta\widehat{\boldsymbol{x}}_\mu) - f(\widehat{\boldsymbol{x}}_\mu)\rangle}{\varepsilon}\Big|_{\mathcal{K}} && (5.48)\\
&= \frac{\langle \varepsilon\boldsymbol{p}, f(\delta\widehat{\boldsymbol{x}}_\mu) - f(\widehat{\boldsymbol{x}}_\mu)\rangle}{\varepsilon^2}\Big|_{\mathcal{K}} && (5.49)\\
&= \frac{\langle \varepsilon\boldsymbol{p}, f(\delta\widehat{\boldsymbol{x}}_\mu) - f(\widehat{\boldsymbol{x}}_\mu)\rangle}{\varepsilon^2}\Big|_{\mathcal{K}} && (5.50)\\
&= \mathbf{Card}(\mathcal{K})\frac{\langle \delta\boldsymbol{d} - \boldsymbol{d}, f(\delta\widehat{\boldsymbol{x}}_\mu) - f(\widehat{\boldsymbol{x}}_\mu)\rangle}{\langle \varepsilon\boldsymbol{p}, \varepsilon\boldsymbol{p}\rangle}\Big|_{\mathcal{K}} && (5.51)\\
&= \mathbf{Card}(\mathcal{K})\frac{\langle \delta\boldsymbol{d} - \boldsymbol{d}, f(\delta\widehat{\boldsymbol{x}}_\mu) - f(\widehat{\boldsymbol{x}}_\mu)\rangle}{\langle \delta\boldsymbol{d} - \boldsymbol{d}, \delta\boldsymbol{d} - \boldsymbol{d}\rangle}\Big|_{\mathcal{K}} && (5.52)
\end{aligned}$$

5.1.3.1 Application pour un cas simple

On se place dans le cas simple où $\overline{\overline{\boldsymbol{x}}}$ correspond à une image en niveaux de gris de L pixels. On simule alors un jeu de donnée suivant la formulation (5.1) où $k \in \{1, \dots K\}$ correspond à un pixel de l'image dégradée. La dégradation est obtenue par la translation de l'image d'un demi pixel, représentée par un opérateur linéaire pour tout $k \in \{1, \dots K\}$, $f_k(\boldsymbol{x}) = (\mathbf{A}\boldsymbol{x})_k$ avec $\mathbf{A} \in \mathcal{M}_{K\times L}(\mathbb{R})$, puis par l'ajout d'un bruit gaussien centré indépendant non-identiquement distribué, tel que $\beta_k \sim \mathcal{N}(0, f_k(\overline{\overline{\boldsymbol{x}}}))$.

Pour estimer $\widehat{\boldsymbol{x}}$ à partir de $\boldsymbol{d}$, on minimise alors le critère régularisé par la régularisation de Tikhonov :

$$\widehat{\boldsymbol{x}} \in \underset{\boldsymbol{x}\in\mathbb{R}^L}{\mathrm{Argmin}} \left[\frac{1}{2}\|\boldsymbol{d} - \mathbf{A}\boldsymbol{x}\|^2_{\mathbf{W}} + \frac{\lambda}{2}\|\boldsymbol{D}\boldsymbol{x}\|^2\right], \tag{5.53}$$

où $\mathbf{W}_k = 1/f_k(\overline{\overline{\boldsymbol{x}}})$ et $\boldsymbol{D}$ représente les différences finies.

On choisit alors ε comme la médiane des $(\boldsymbol{d}_k)_{k\in\{1,\dots K\}}$ et on simule un jeu de données perturbées $\delta\boldsymbol{d}$ selon la formulation (5.47), afin de calculer le critère SURE.

On estime alors les cartes $\widehat{\boldsymbol{x}}_\mu$ et $\delta\widehat{\boldsymbol{x}}_\mu$ à partir du problème (5.53) avec l'algorithme VMLMB, présenté dans la section 1.3.2, pour différentes valeurs de $\lambda \in [10^{-2}, 10^4]$ à patir du modèle des données respectives $\boldsymbol{d}$ et $\delta\boldsymbol{d}$. On calcule alors l'EQM des paramètres, donnée par l'équation (5.30) , l'EQM du modèle, donnée par l'équation (5.31), ainsi que le critère SURE, par l'équation (5.32).

La figure 5.1 montre le tracé des EQMs des paramètres et du modèle ainsi que le critère SURE en fonction de la valeur de l'hyperparamètre λ. Sont également tracé l'attache aux données et la trace du critère SURE normalisée par le nombre de mesures valides $\mathbf{Card}(\mathcal{K})$. On voit que le critère SURE correspond bien à l'EQM du modèle dont il est l'estimateur.

De plus, on remarque que la valeur de l'hyperparamètre donnant l'EQM des paramètres minimale est la même que celle qui minimise l'EQM du modèle et le critère SURE. Enfin on voit que lorsque λ augmente, l'attache aux données du critère SURE augmente tandis que la trace décroit.

5.2 Érreur des modèles séparables

Dans de le cas de la Méthode non-Linéaire Séparable (MnLS) et de la Méthode Linéaire Séparable(MLS), les données sont exprimées pour tout pixel $n \in \{1,\dots N\}$, acquisition $k \in \{1,\dots K\}$ et partie $j \in \{1,2\}$ du détecteur par

Rappel éq. (2.2)
$$\boldsymbol{d}^{\mathrm{S}}_{j,k,n} = \mathcal{B}_{j,k,n}(f^{\mathrm{S}}_{j,k}(\overline{\overline{\boldsymbol{x}}}_n)),$$

où $f_{j,k} : \mathbb{R} \to \mathbb{R}$ pour tout $j \in \{1,2\}, k \in \{1,\dots K\}$ et où $\mathcal{B}_{j,k,n}(x)$ donne une réalisation de la loi gaussienne $\mathcal{N}(x, \sigma^2_{j,k,n})$. La fonction de co-logvraisemblance des données associées est donnée pour tout pixel $n \in \{1,\dots N\}$ par $\mathcal{L}(\boldsymbol{d}_n|\boldsymbol{x}_n) = -\Phi_n(\boldsymbol{x}_n)$ où :

Rappel éq. (2.16)
$$\forall n \in \{1,\dots,N\}, \quad \Phi_n(\boldsymbol{x}) = \sum_{j,k} \frac{\mathbf{W}_{j,k,n}}{2} \left(\boldsymbol{d}^{\mathrm{S}}_{j,k,n} - f^{\mathrm{S}}_{j,k}(\boldsymbol{x})\right)^2.$$

D'après l'équation (5.11), on a que la matrice de Fisher de $\boldsymbol{x}_n \in \mathbb{R}^L$ est donnée pour tout $n \in \{1,\dots N\}$ par :

$$\mathcal{I}_{\Phi_n}(\boldsymbol{x}_n) = \sum_{j,k} \nabla f_{j,k}(\boldsymbol{x}_n) \mathbf{W}_{j,k,n} \nabla f_{j,k}(\boldsymbol{x}_n)^\top. \tag{5.54}$$

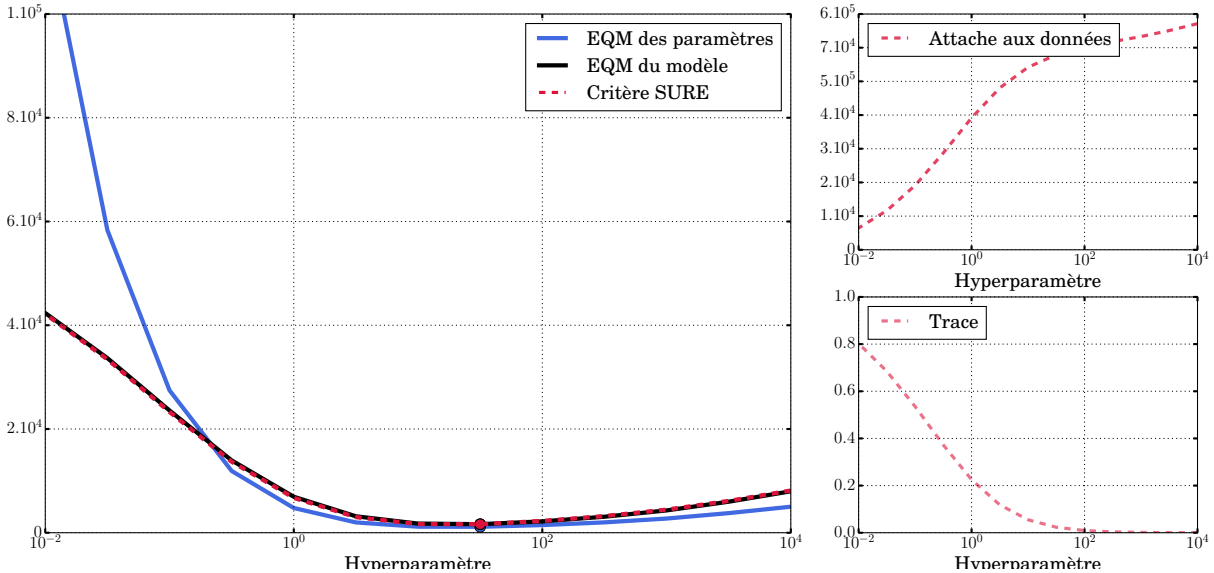


Figure 5.1 – Comparaison du critère SURE avec l'EQM calculée dans le domaine des paramètres et l'EQM calculée dans le domaine des données.

5.2.1 Borne de FDCR de la MnLS

Rappelons que dans le cas de la MnLS, la fonction $f_{j,k}$ est donnée pour $\boldsymbol{x} = (I^{\mathrm{u}}, I^{\mathrm{p}}, \theta)^\top \in (\mathbb{R}^N)^L$, par

Rappel éq. (2.4)
$$f_{j,k}(\boldsymbol{x}) = \frac{\boldsymbol{x}_1 \|\boldsymbol{v}_{j,k}\|^2_{\mathbb{C}^2}}{2} + \boldsymbol{x}_2 \left|\langle \boldsymbol{v}_{j,k}, \boldsymbol{c}(\boldsymbol{x}_3)\rangle_{\mathbb{C}^2}\right|^2,$$

où $\boldsymbol{c}(x) = (\cos x, \sin x)^\top$. On déduit de (5.54) que l'information de Fisher associé est donnée par :

$$\mathbf{I}_{\Phi_n}(\boldsymbol{x}_n) = \begin{pmatrix} \mathbf{V} & \begin{matrix} \mathbf{I}_{1,n} \\ \mathbf{I}_{2,n} \end{matrix} \\ \begin{matrix} \mathbf{I}_{1,n} & \mathbf{I}_{2,n} \end{matrix} & \mathbf{I}_{3,n} \end{pmatrix} \tag{5.55}$$

où $\mathbf{V}$ est la matrice donnée par

Rappel éq. (2.19)
$$\mathbf{V} = \begin{pmatrix} \sum_{j,k} \mathbf{W}_{j,k,n} \|\boldsymbol{v}_{j,k}\|^4_{\mathbb{C}^2}/4 & \sum_{j,k} \boldsymbol{W}_{j,k,n} \|\boldsymbol{v}_{j,k}\|^2_{\mathbb{C}^2} |\langle \boldsymbol{v}_{j,k}, \boldsymbol{c}_{\theta_n}\rangle_{\mathbb{C}^2}|^2/2 \\ \sum_{j,k} \mathbf{W}_{j,k,n} \|\boldsymbol{v}_{j,k}\|^2_{\mathbb{C}^2} |\langle \boldsymbol{v}_{j,k}, \boldsymbol{c}_{\theta_n}\rangle_{\mathbb{C}^2}|^2/2 & \sum_{j,k} \boldsymbol{W}_{j,k,n} |\langle \boldsymbol{v}_{j,k}, \boldsymbol{c}_{\theta_n}\rangle_{\mathbb{C}^2}|^4 \end{pmatrix},$$

et

$$\begin{cases} \mathbf{I}_{1,n} = I^{\mathrm{p}}_n \sum_{j,k} \mathbf{W}_{j,k,n} \|\boldsymbol{v}_{j,k}\|^2_{\mathbb{C}^2} \dfrac{\partial |\langle \boldsymbol{v}_{j,k}, \boldsymbol{c}_{\theta_n}\rangle_{\mathbb{C}^2}|^2}{\partial \theta_n}, \\ \mathbf{I}_{2,n} = I^{\mathrm{p}}_n \sum_{j,k} \mathbf{W}_{j,k,n} |\langle \boldsymbol{v}_{j,k}, \boldsymbol{c}_{\theta_n}\rangle_{\mathbb{C}^2}|^2 \dfrac{\partial |\langle \boldsymbol{v}_{j,k}, \boldsymbol{c}_{\theta_n}\rangle_{\mathbb{C}^2}|^2}{\partial \theta_n}, \\ \mathbf{I}_{3,n} = I^{\mathrm{p}}_n \sum_{j,k} \mathbf{W}_{j,k,n} \left(\dfrac{\partial |\langle \boldsymbol{v}_{j,k}, \boldsymbol{c}_{\theta_n}\rangle_{\mathbb{C}^2}|^2}{\partial \theta_n} \right)^2. \end{cases} \tag{5.56}$$

Dans le cas de la MnLS, les fonctions $(f_{j,k})_{j\in\{1,2\},k\in\{1,\dots K\}}$ sont non-linéaires et donc la matrice de Fisher $\mathbf{I}_{\Phi_n}(\boldsymbol{x}_n)_n$ dépend de $\boldsymbol{x}_n$, en particulier de I^{p}_n et de θ_n. Comme il a été vu précédemment, cela pose problème pour l'estimation de la borne de FDCR dans le cas où $I^{\mathrm{p}}_n \approx 0$. En effet dans un tel cas, on a $\mathbf{I}_{\ell,n} = 0$ pour tout $\ell \in \{1, \dots L\}$, on ne peut donc pas avoir une borne inférieure à l'erreur angulaire. C'est pourquoi nous avons introduit le changement de variable (5.28).

5.2.2 De la borne de FDCR de la MLS à une approximation de la variance pour la MnLS

Dans le cas de la MLS, la fonction $f_{j,k}$ est donnée pour $\boldsymbol{x} = (\mathrm{I}, \mathrm{Q}, \mathrm{U})^\top \in (\mathbb{R}^N)^L$, par

Rappel éq. (2.26)
$$\forall n \in \{1, \dots N\} f^{\mathrm{S}}_{j,k}(\boldsymbol{x}_n) = \textstyle\sum_{\ell=1}^3 v_{j,k,\ell} \overline{\boldsymbol{x}}^{\mathbf{H}}_{\ell,n},$$

où les $v_{j,k,\ell} \in \mathbb{R}$ pour tout $j \in \{1,2\}$, $k \in \{1, \dots K\}$ et $\ell \in \{1, \dots L\}$, avec

Rappel éq. (2.27)
$$\begin{cases} v_{j,k,1} = \dfrac{\|\boldsymbol{v}_{j,k}\|^2_{\mathbb{C}^2}}{2} \\ v_{j,k,2} = \dfrac{|\boldsymbol{v}^{(x)}_{j,k}|^2 - |\boldsymbol{v}^{(y)}_{j,k}|^2}{2} \\ v_{j,k,3} = \mathfrak{Re}\left(\boldsymbol{v}^{(x)}_{j,k} \boldsymbol{v}^{(y)}_{j,k} \right), \end{cases}.$$

On a donc :

$$\mathbf{I}_{\Phi_n}(\boldsymbol{x}_n) = \mathbf{V}_n \tag{5.57}$$

où

Rappel éq. (2.30)
$$\mathbf{V}_n = \sum_{j,k} \mathbf{W}_{j,k,n} \begin{pmatrix} \boldsymbol{v}_{j,k,1}^2 & \boldsymbol{v}_{j,k}\boldsymbol{v}_{j,k,2} & \boldsymbol{v}_{j,k,1}\boldsymbol{v}_{j,k,3} \\ \boldsymbol{v}_{j,k,1}\boldsymbol{v}_{j,k,2} & \boldsymbol{v}_{j,k,2}^2 & \boldsymbol{v}_{j,k,2}\boldsymbol{v}_{j,k,3} \\ \boldsymbol{v}_{j,k,1}\boldsymbol{v}_{j,k,3} & \boldsymbol{v}_{j,k,2}\boldsymbol{v}_{j,k,3} & \boldsymbol{v}_{j,k,3}^2 \end{pmatrix}$$

Dans ce cas là, les fonctions $(f_{j,k})_{j\in\{1,2\},k\in\{1,\dots K\}}$ sont linéaires et donc l'information de Fisher ne dépend pas de $\boldsymbol{x}$. De ce fait, la borne de FDCR est exacte et ne dépend pas des paramètres. Il est donc possible d'avoir une borne inférieure de la covariance des paramètres estimés $\widehat{\mathrm{I}}$, $\widehat{\mathrm{Q}}$ et $\widehat{\mathrm{U}}$, en utilisant le fait que

Rappel éq. (5.6)
$$\mathrm{Cov}(\widehat{\boldsymbol{x}}) \succeq \mathrm{FDCR}(\boldsymbol{x}).$$

On peut également utiliser cette borne de FDCR pour calculer une approximation de la covariance des paramètres estimés $(\widehat{I_n^{\mathrm{u}}}, \widehat{I_n^{\mathrm{p}}}, \widehat{\theta_n})$ à partir de l'équation (5.28). Dans notre cas, la fonction $q : \mathbb{R}^L \to \mathbb{R}^L$ correspond au changement de variable des paramètres de Stokes I, Q, U aux paramètres I^{u}, I^{p}, θ donné par :

$$q\begin{pmatrix} \mathrm{I}_n \\ \mathrm{Q}_n \\ \mathrm{U}_n \end{pmatrix} = \begin{pmatrix} \mathrm{I}_n - \sqrt{\mathrm{Q}_n^2 + \mathrm{U}_n^2} \\ \sqrt{\mathrm{Q}_n^2 + \mathrm{U}_n^2} \\ \frac{1}{2}\arctan(\frac{\mathrm{U}_n}{\mathrm{Q}_n}) \end{pmatrix} = \begin{pmatrix} I_n^{\mathrm{u}} \\ I_n^{\mathrm{p}} \\ \theta_n \end{pmatrix}. \tag{5.58}$$

Les paramètres en lesquels le modèle est linéaire sont les paramètres $\widehat{\boldsymbol{x}} = (\widehat{\mathrm{I}}, \widehat{\mathrm{Q}}, \widehat{\mathrm{U}})^\top$. La jacobienne de q, est donnée pour tout $n \in \{1, \dots N\}$ par :

$$\mathbf{J}_q\begin{pmatrix} \mathrm{I}_n \\ \mathrm{Q}_n \\ \mathrm{U}_n \end{pmatrix} = \begin{pmatrix} 1 & -\frac{\mathrm{Q}_n}{\sqrt{\mathrm{Q}_n^2+\mathrm{U}_n^2}} & -\frac{\mathrm{U}_n}{\sqrt{\mathrm{Q}_n^2+\mathrm{U}_n^2}} \\ 0 & \frac{\mathrm{Q}_n}{\sqrt{\mathrm{Q}_n^2+\mathrm{U}_n^2}} & \frac{\mathrm{U}_n}{\sqrt{\mathrm{Q}_n^2+\mathrm{U}_n^2}} \\ 0 & -\frac{\mathrm{U}_n}{2(\mathrm{Q}_n^2+\mathrm{U}_n^2)} & \frac{\mathrm{Q}_n}{2(\mathrm{Q}_n^2+\mathrm{U}_n^2)} \end{pmatrix}. \tag{5.59}$$

Remarque : Il est également possible d'estimer une borne inférieure à la variance des paramètres estimés avec les méthodes de l'état-de-l'art en utilisant la formule (5.19) en posant $\boldsymbol{x} = \boldsymbol{d}$ et $q(\boldsymbol{x})$ la fonction des paramètres de Stokes donnée soit par (1.25) pour la double différence ou (1.26) pour le double ratio.

5.2.3 Application pour un pixel sur données simulées

On se propose d'étudier le comportement de l'erreur des estimateurs obtenus par la MnLS pour un pixel en fonction du SNR et du taux total de polarisation.

Sur la figure 5.2a, sont tracées les courbes de niveau du SNR et du taux de polarisation τ^{total}, dont les valeurs sont données pour un pixel $n \in \{1, \dots N\}$ par :

Rappel éq. (2.22), (2.23)
$$\tau^{\mathrm{total}}(\boldsymbol{x}_n) = \frac{I_n^{\mathrm{p}}}{I_n^{\mathrm{u}}+I_n^{\mathrm{p}}} \qquad \text{et} \qquad \mathrm{SNR}(\boldsymbol{x}_n) = \frac{\sqrt{K} I_n^{\mathrm{p}}}{2\sqrt{(I_n^{\mathrm{u}}+I_n^{\mathrm{p}})/2+\sigma_{\mathrm{ro}}^2}}$$

en fonction de l'intensité polarisée I_n^{p} et de l'intensité non-polarisée I_n^{u}. Le SNR est calculé ici dans le pire des cas, c'est-à-dire où $K = 4$ soit une seule rotation de lame demi-onde. En arrière-plan est tracée la carte des Erreurs Angulaires (EA) faites sur l'angle de polarisation θ_n, calculées empiriquement, pour $S = 1000$ simulations, à partir de la formule :

$$\mathrm{EA}_\theta(\widehat{\theta}_n, \overline{\theta}_n) = \sqrt{\sum_{s=1}^{S} \mathrm{ang}\Big(\exp\big(2\mathrm{i}(\widehat{\theta}_{s,n} - \overline{\theta}_n)\big)\Big)^2 / 2S}, \tag{5.60}$$

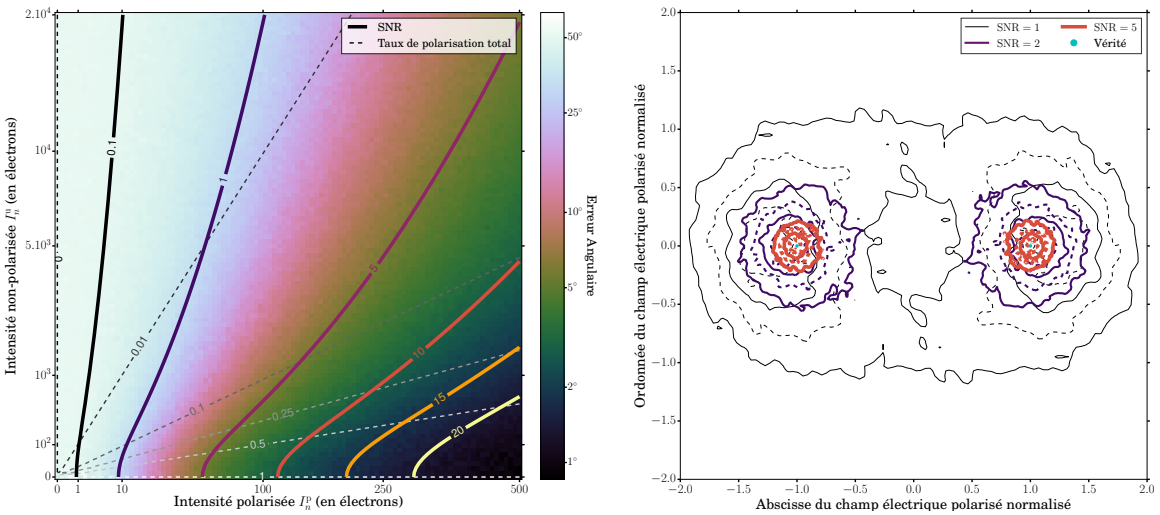
où i est la notation complexe et ang l'angle de l'exponentielle complexe. L'angle vrai $\overline{\theta}_n$ est choisi aléatoirement pour chaque couple de valeur $(I_n^{\mathrm{p}}, I_n^{\mathrm{u}})$.

Les valeurs de I^{u} correspondraient, pour les valeurs les plus élevées (*i.e.* $> 10^3$), à celles du résidu de l'étoile au bord du coronographe où de la fonction d'étalement de point d'un éventuel compagnon brillant.

On voit que les courbes de niveau du SNR et les niveaux d'erreurs angulaires correspondent. On peut donc supposer que les erreurs angulaires sont corrélées au SNR. On remarque par ailleurs que pour avoir une erreur angulaire inférieur à 5°, il faut avoir un SNR supérieur à 5. On peut donc en déduire que dans le cas de disques peu brillants (*i.e* $I^{\mathrm{p}} < 100$), avec $K = 4$ acquisitions il sera impossible d'avoir une précision angulaire de 2°. La solution est alors d'augmenter le nombre d'acquisitions afin d'avoir un plus grand nombre de réalisations du bruit et donc un meilleur SNR.

La figure 5.2b illustre le fait que lorsque le SNR diminue, la distribution des estimations $\widehat{\theta}_n$ s'élargit. Les courbes de niveaux correspondent aux différents quartiles de la distribution des points

$$\sqrt{\frac{\widehat{I^{\mathrm{u}}}}{\overline{I^{\mathrm{u}}}}} \begin{pmatrix} \cos\left(\widehat{\theta}_n\right) \\ \sin\left(\widehat{\theta}_n\right) \end{pmatrix}. \tag{5.61}$$



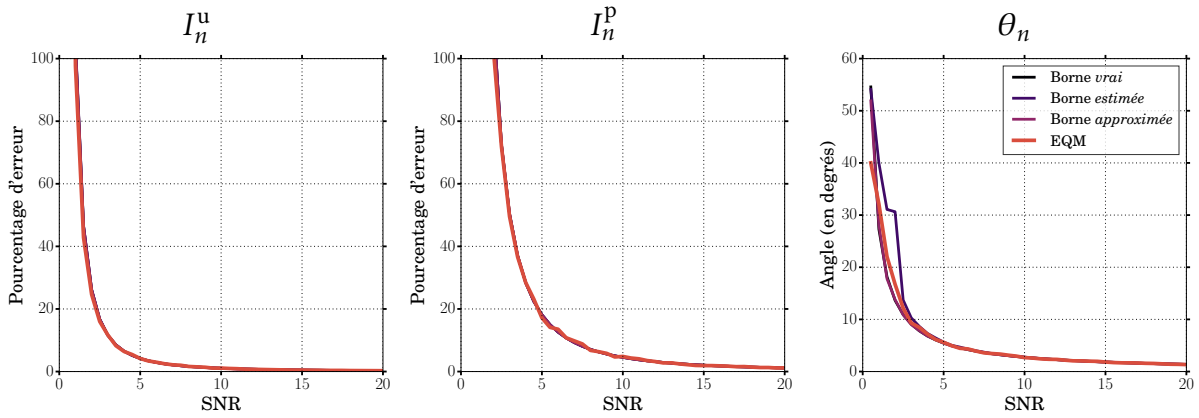
(a) (b)

FIGURE 5.2 – Estimation empirique de l'erreur angulaire en fonction du SNR et du tau de polarisation. Figure (a) : Valeurs du SNR, de τ^{total} et de l'erreur angulaire en fonction de I_n^{p} et de I_n^{u}. Figure (b) : distribution des valeurs des champs électriques estimés normalisés, donnés par (5.61), pour chaque réalisation du bruit à différent SNR.

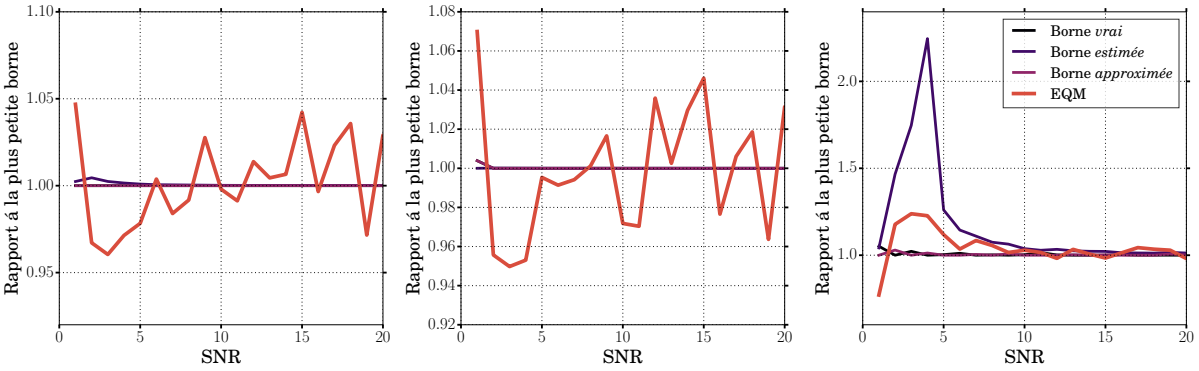
On voit que plus le SNR diminue, plus les réalisations s'étalent en « croissant de lune », jusqu'à se rejoindre pour former un cercle. Cela signifie qu'à partir d'un certain SNR, estimer l'angle où choisir un angle aléatoirement entre $[0, 2\pi[$ est équivalent. Dans le cas d'un SNR aussi faible, il n'y a donc pas de sens à présenter la mesure de l'angle.
Dans une application sur données réelles, afin de savoir si une mesure est suffisamment bonne pour être présentée, on utilise le fait que pour une estimation $\widehat{\boldsymbol{x}}$ non biaisé :

$$\mathrm{Var}(\widehat{\boldsymbol{x}}_\ell) = \mathrm{Cov}(\widehat{\boldsymbol{x}})_{\ell,\ell} \geq \mathrm{FDCR}(\boldsymbol{x})_{\ell,\ell}. \tag{5.62}$$

Il a été vu précédemment, que dans le cas d'estimateurs non-linéaires, la borne de FDCR dépend de I^{p} et de θ. Nous avons également introduit un changement de variable permettant de calculer une borne inférieure à la covariance de $(\widehat{I^{\mathrm{u}}_n}, \widehat{I^{\mathrm{p}}_n}, \widehat{\theta}_n)$ à partir de la FDCR du modèle paramétré en $(\widehat{\mathrm{I}}_n, \widehat{\mathrm{Q}}_n, \widehat{\mathrm{U}}_n)$. La question que l'on peut se poser est : comment se comporte les EQMs des paramètres $\boldsymbol{x} = (\widehat{\mathrm{I}}_n, \widehat{\mathrm{Q}}_n, \widehat{\mathrm{U}}_n)$ par rapport à la FDCR de la MnLS, calculée en $\widehat{\boldsymbol{x}}$ (borne *estimée*) et en $\overline{\overline{\boldsymbol{x}}}$ (borne *vrai*) ainsi qu'à l'approximation par changement de variable (borne *approximée*). La figure 5.3 représente pour chaque paramètre, l'EQM et les bornes *estimées*, *vraies* et *approximées* correspondantes, pour $S = 10000$ simulations. Les EQM de I^{u}_n et I^{p}_n sont



(a) Valeurs des racines des différentes bornes et de l'EQM des paramètres de la MnLS en fonction du SNR.



(b) Rapport à la borne la plus basse pour chaque valeur du SNR

Figure 5.3 – Comparaison des racines des différentes bornes et de l'EQM, calculé entre les paramètres estimés et vrai de la MnLS, pour différents SNR.

obtenus par :

$$\mathrm{EQM}(\widehat{x}, \overline{\overline{x}}) = \frac{1}{S}\sum_{1}^{S}(\widehat{x_s} - \overline{\overline{x}})^2. \tag{5.63}$$

L'espérance $\mathbb{E}[\boldsymbol{x}]$ pour le calcul de la borne *approximée*, est fait de manière empirique, c'est-à-dire que $\mathbb{E}[\boldsymbol{x}] = \frac{1}{S}\sum_{s=1}^{S}\boldsymbol{x}_s$. Enfin, pour le calcul de la borne *estimée*, est calculée par

$$\widehat{\mathrm{FDCR}}\Big((I^{\mathrm{u}}{}_n, I^{\mathrm{p}}{}_n, \theta_n)\Big). = \frac{1}{S}\sum_{s=1}^{S}\mathrm{FDCR}\Big((\widehat{I^{\mathrm{u}}}_{s,n}, \widehat{I^{\mathrm{p}}}_{s,n}, \widehat{\theta}_{s,n})\Big). \tag{5.64}$$

On voit sur la figure 5.3b que la borne *approximée* est bien une borne inférieure aux bornes *vraies* et *estimées*. De plus la borne *vraie* semble elle-même être une borne inférieure à la borne *estimée*, qui semble la plus loin de la MSE. Lorsque le SNR diminue, on voit que la borne *estimée* pour θ_n tend à être beaucoup plus grandes que la borne *vrai*. Le choix de la borne *approximée* semble alors la plus pertinente comme estimation de l'erreur.

5.3 Érreur des modèles non-séparables

Rappelons que dans le cas non-séparable, les données sont un ensemble de vecteurs $(\boldsymbol{d}_k^{\mathrm{nS}})_{k\in\{1,\dots K\}} \in \mathbb{R}^M$ où $K \in \mathbb{N}$ représente le nombre d'acquisitions et $M \in \mathbb{N}$ le nombre de pixels par acquisition. Les paramètres d'intérêts sont, comme dans le cas séparable, un ensemble de vecteurs $(\boldsymbol{x}_\ell)_{\ell\in\{1,\dots L\}} \in \mathbb{R}^N$ où $L \in \mathbb{N}$ représente le nombre de composantes et $N \in \mathbb{N}$ le nombre de pixels par composantes. Les données et les paramètres d'intérêt sont reliés par le modèle

Rappel éq. (3.1)
$$\boldsymbol{d}_k^{\mathrm{nS}} = \mathcal{B}_k(f_k^{\mathrm{nS}}(\overline{\overline{\boldsymbol{x}}})) \tag{3.1}$$

où $f_k^{\mathrm{nS}} : (\mathbb{R}^N)^L \to \mathbb{R}^M$ modélise l'instrument, les transformations du détecteur pour passer du domaine des cartes reconstruites au domaine des données, peut prendre en compte la convolution par la PSF, et $\mathcal{B}_k : \mathbb{R}^M \to \mathbb{R}^M$ représente le bruit des données de matrice de covariance $\mathbf{W}^{-1}$.

5.3.1 Critère SURE pour le cas multidimensionnel

Les méthodes utilisées pour reconstruire les paramètres d'intérêts $\boldsymbol{x} \in (\mathbb{R}^N)^L$ à partir d'un tel modèle, c'est-à-dire toutes les méthodes présentées dans les chapitres 3 et 4, contiennent différentes régularisations permettant de réduire la variance des reconstructions mais augmentant le biais. C'est pourquoi, comme dans la section 5.1.3 nous n'étudions plus la variance mais l'Erreur Quadratique Moyenne des paramètres reconstruits. Dans le cas où il y a plusieurs composantes, il s'agit de la somme sur toutes les composantes de leurs EQMs soit :

$$\mathrm{EQM}^{\text{paramètres}}(\widehat{\boldsymbol{x}}) = \sum_{\ell=1}^{L}\mathbb{E}\Big[\|\widehat{\boldsymbol{x}}_\ell - \overline{\overline{\boldsymbol{x}}}_\ell\|^2\Big], \tag{5.65}$$

Comme précédemment, comme $\overline{\overline{\boldsymbol{x}}}$ n'est pas connu, nous utilisons SURE pour estimer l'EQM dans le domaine des données au lieu de l'EQM dans le domaine des paramètres reconstruits,

soit :

$$\mathrm{EQM}^{\text{données}}(\widehat{\boldsymbol{x}}) = \sum_{k=1}^{K} \mathbb{E}\Big[(f_k^{\mathrm{nS}}(\widehat{\boldsymbol{x}}) - f_k^{\mathrm{nS}}(\overline{\overline{\boldsymbol{x}}}))^2\Big]. \tag{5.66}$$

Notons $\nu \in \mathbb{R}^h$ l'ensemble des hyperparamètres à régler, où $h \in \mathbb{N}^*$ correspond au nombre d'hyperparamètres à régler. Notons alors $\widehat{\boldsymbol{x}}_\nu \in (\mathbb{R}^N)^L$ les paramètres d'intérêt estimés pour un ensemble d'hyperparamètres ν donné, et pour tout $k \in \{1, \dots K\}$ l'ensemble des pixels $m \in \{1, \dots M\}$ valides de l'acquisition k : $\mathcal{M}_k = \{m \in \{1, \dots M\} \text{ tels que } \mathbf{W}_{k,m} \neq 0\}$. Alors le critère SURE est donné par :

$$\mathrm{SURE}(\widehat{\boldsymbol{x}}_\nu) = \sum_{k=1}^{K} \Big[\big\| \boldsymbol{d}_k^{\mathrm{nS}} - f_k^{\mathrm{nS}}(\widehat{\boldsymbol{x}}_\nu) \big\|_{\mathbf{W}_k}^2 + 2 \operatorname{tr}\Big(\frac{\partial f_k^{\mathrm{nS}}(\widehat{\boldsymbol{x}}_\nu)}{\partial \boldsymbol{d}_k^{\mathrm{nS}}} \Big)\Big|_{\mathcal{M}_k} - \mathrm{Card}(\mathcal{K}_k) \Big], \tag{5.67}$$

où $\mathrm{Card}(\cdot)$ représente le cardinal de l'ensemble, c'est-à-dire le nombre d'éléments qui le composent. Dans ce cas, le calul de la trace est le suivant. Soit $\boldsymbol{p}_k \sim \mathcal{N}(0_{\mathbb{R}^M}, \mathbf{Id}_{\mathbb{R}^M \times \mathbb{R}^M})$ une perturbation et $\varepsilon > 0$. On note $\delta \boldsymbol{d}_k^{\mathrm{nS}} = \boldsymbol{d}_{\boldsymbol{k}}^{\mathrm{nS}} + \varepsilon \boldsymbol{p}$ et $\delta \widehat{\boldsymbol{x}}_\nu$ l'estimateur des données $\delta \boldsymbol{d}_k$ à partir du modèle f_k. Alors :

$$\operatorname{tr}\Big(\frac{\partial f_k(\widehat{\boldsymbol{x}}_\nu)}{\partial \boldsymbol{d}_k^{\mathrm{nS}}} \Big)\Big|_{\mathcal{K}_k} \approx \mathrm{Card}(\mathcal{M}_k) \frac{\langle \delta \boldsymbol{d}_k^{\mathrm{nS}} - \boldsymbol{d}_k^{\mathrm{nS}}, f_k^{\mathrm{nS}}(\delta \widehat{\boldsymbol{x}}_\nu) - f_k^{\mathrm{nS}}(\widehat{\boldsymbol{x}}_\nu) \rangle}{\langle \delta \boldsymbol{d}_k^{\mathrm{nS}} - \boldsymbol{d}_k^{\mathrm{nS}}, \delta \boldsymbol{d}_k^{\mathrm{nS}} - \boldsymbol{d}_k^{\mathrm{nS}} \rangle}\Big|_{\mathcal{M}_k}. \tag{5.68}$$

5.3.2 Application aux méthodes non-séparables

Dans cette section, nous étudions les critères SURE obtenus pour les différentes méthodes développées dans les chapitres 3 et 4, c'est-à-dire $\nu = (\mu_\ell, \lambda_\ell)_{\ell \in \{1,\dots L\}}$.

Sur la figure 5.4 sont tracées les cartes de critère SURE, d'EQM du modèle et d'EQM des paramètres en fonction des hyperparamètres $(\lambda_\ell)_{\ell \in \{1,\dots L\}}$ dans le cas de la MLnS et de la MnLnS. Dans le cas de la MLnS, on appelle respectivement $\lambda_1 = \lambda_{\mathrm{I}}$, $\lambda_2 = \lambda_{\mathrm{Q}}$ et $\lambda_3 = \lambda_{\mathrm{U}}$ et on fixe $\lambda_{\mathrm{Q}} = \lambda_{\mathrm{U}}$. Dans le cas de la MnLnS, on appelle $\lambda_1 = \lambda_{I^{\mathrm{u}}}$ et $\lambda_2 = \lambda_{I^{\mathrm{p}}}$. Dans le cas de la régularisation à préservation de bord TV-h, on fixe les paramètres $(\mu_\ell)_{\ell \in \{1,\dots L\}}$ manuellement et on pose alors pour tout $\ell \in \{1, \dots L\}$ $\mu_\ell = \mu = 0,1$.

On observe tout d'abord que dans presque tous les cas, le critère SURE estime bien l'EQM du modèle. La méthode semble avoir des difficultés dans le cas non-linéaire avec régularisation à préservation de bord TV-h, où l'estimation semble compliquée. On remarque de plus que contrairement au cas simple présenté dans la section précédente, le couple d'hyperparamètres minimisant l'EQM des paramètres est différent que le couple d'hyperparamètres minimisant l'EQM du modèle et le critère SURE. Cependant, dans le cas de la MnLnS, les critères sont plus semblables, c'est-à-dire que les zones dans lesquels se trouvent les minimums semblent plus proches.

Sur la figure 5.5, sont tracées les cartes de critère SURE, d'EQM du modèle et d'QM des paramètres en fonction des hyperparamètres $(\lambda_\ell)_{\ell \in \{1,\dots L\}}$, dans le cas de la MLnS avec déconvolution. Les cartes sont tracées pour la régularisation à préservation de bord TV-h sans et avec contrainte épigraphique, puis avec régularisation TV et Shatten sur le hessien.

On observe que les critères SURE estiment très bien l'EQM du modèle. On remarque de plus que les zones contenant les couples de paramètres minimisant d'une part l'EQM des paramètres et d'autre par l'EQM du modèle sont très proches en comparaison aux méthodes sans déconvolution.

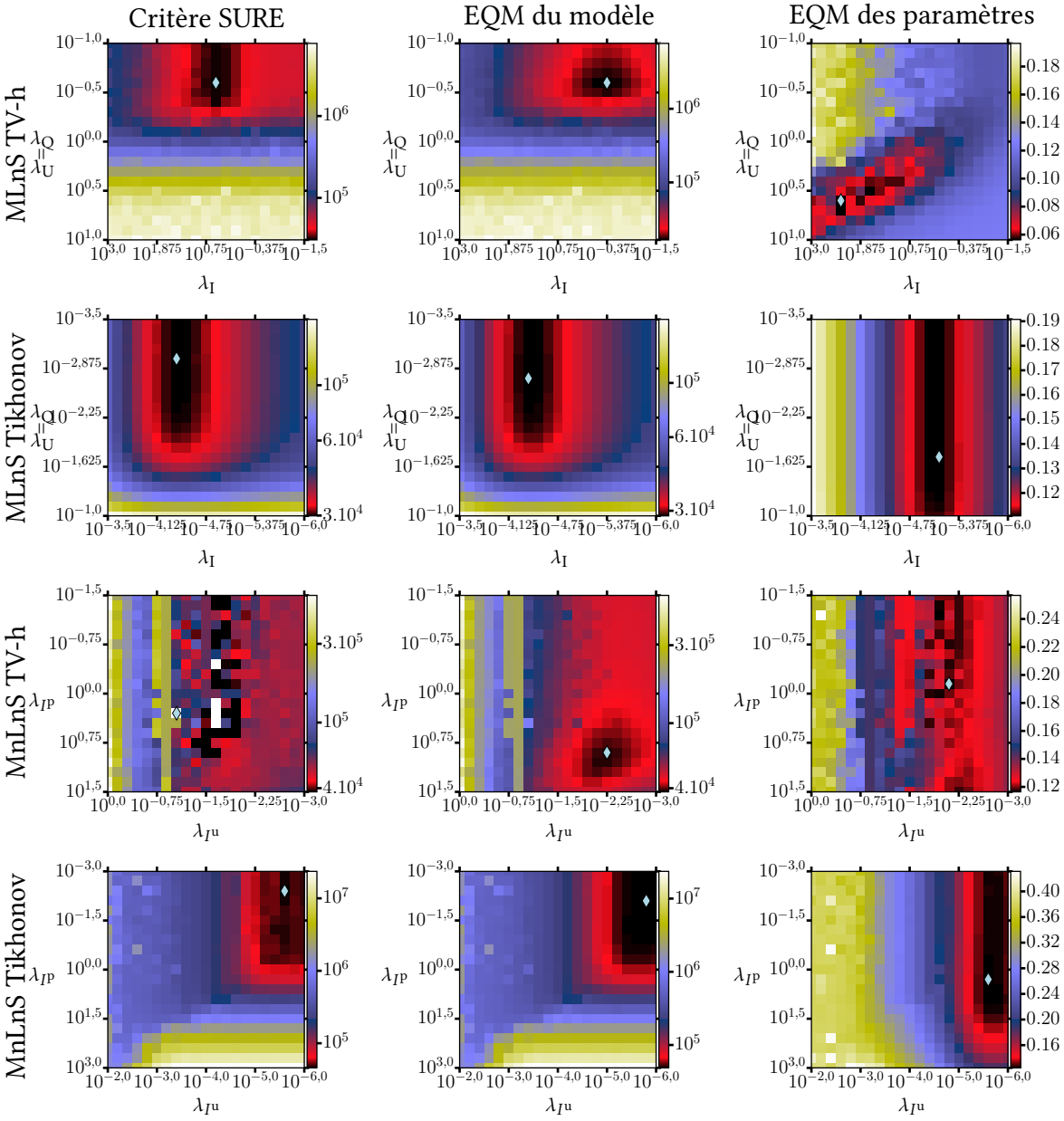


FIGURE 5.4 – Comparaison des cartes de critère SURE et d'Erreur Quadratique Moyenne dans le domaine des données et dans le domaine des paramètres.

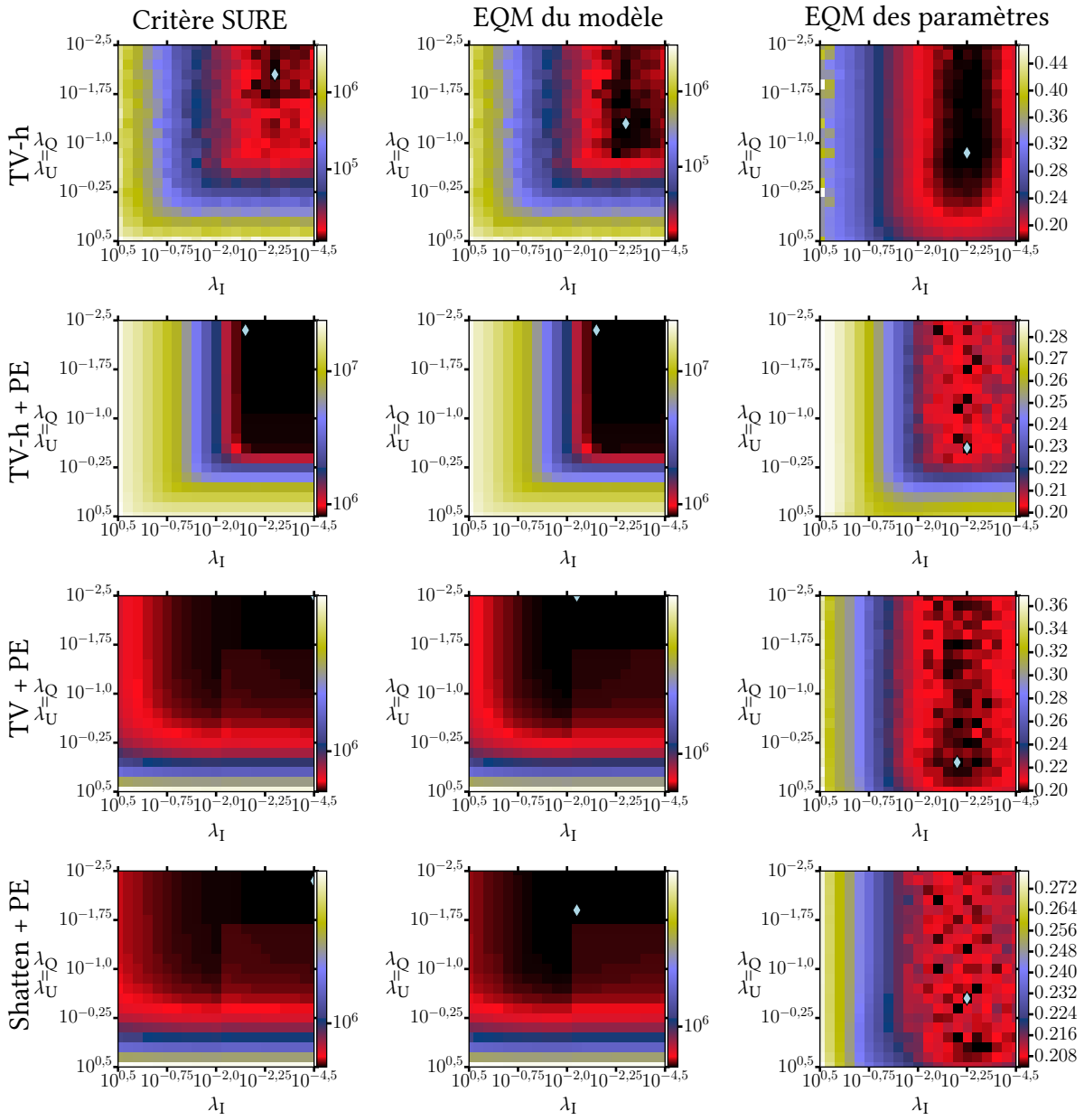


FIGURE 5.5 – Comparaison des cartes de critère SURE et d'Erreur Quadratique Moyenne dans le domaine des données et dans le domaine des paramètres.

Chapitre 6

Étude astrophysique

6.1 Calibration

En astrophysique, la calibration des données est essentielle, pour prendre en compte le bruit et les artefacts liés à la mesure, avant de procéder à une reconstruction à partir des données. Elle permet d'une part, d'étalonner la pollution lumineuse induite par le ciel et l'instrument, ainsi que le comportement du détecteur en terme d'erreur sur les valeurs des pixels, c'est ce qu'on appelle ici la calibration du détecteur. D'autre part, elle permet grâce à des données bien précise, d'estimer la PSF de l'instrument et les transformations (translations, décentrage, rotations) des images sur le détecteur, c'est ce qu'on appelle ici calibration instrumentale. Il est également possible de calibrer la polarisation instrumentale, [van Holstein et al., 2020] mais cela n'a pas pu être appliqué avant l'écriture de ce manuscrit (voir article soumis au journal A&A en annexes).

6.1.1 Calibration du détecteur

Lorsque la lumière, composée de photons, arrive sur le détecteur, celui-ci enregistre un certain nombre de photo-électrons par pixel. Pour un pixel donné $m \in \{1,\dots M\}$, à l'acquisition $k \in \{1,\dots K\}$ cette quantité dépend du temps d'intégration $t_k \geq 0$ et de l'efficacité du détecteur, c'est-à-dire de son rendement quantique $\boldsymbol{q}_m$. Ce nombre de photons est alors converti en ADU (Analog–to-Digital Unit) dépendant du gain $\boldsymbol{g}_m$ de conversion du détecteur. Le nombre de photo-électrons enregistré est aléatoire et suit une loi de Poisson de paramètre égal à sa moyenne et à sa variance. Pour toute acquisition, on modélise la donnée $\boldsymbol{d}_{k,m}$ brute obtenue en un pixel $m \in \{1,\dots M\}$ à l'acquisition $k \in \{1,\dots K\}$ par :

$$\forall m \in \{1,\dots M\} \quad \boldsymbol{d}_{k,m} = \frac{\left(\boldsymbol{q}_m \boldsymbol{f}_{k,m} + \boldsymbol{c}_m\right) t_k + \mathbf{b}_m + \beta_{k,m}}{\boldsymbol{g}_m}, \tag{6.1}$$

où :

- $\boldsymbol{q}_m$ est le rendement quantique du pixel $m \in \{1,\dots M\}$ du détecteur, supposé constant au cours des acquisitions. Pour simplifier les calculs, il peut être supposé connu et le même pour tous les pixels ($\approx 85\%$ pour IRDIS) ;
- $\boldsymbol{g}_m$ est le gain de conversion du pixel $m \in \{1,\dots M\}$ du détecteur, supposé constant au cours des acquisitions. Pour simplifier les calculs, peut être supposé connu et le même pour tous les pixels ($\approx 1.85 e^-$/pix pour IRDIS) ;

- $\boldsymbol{f}_{k,m}$ le flux atteignant le détecteur au pixel $m \in \{1,\dots M\}$ à l'acquisition $k \in \{1,\dots K\}$ (en e^-/sec) ;
- $\boldsymbol{c}_m = \boldsymbol{c}_m^{\text{sky}} + \boldsymbol{c}_m^{\text{instru}} + \boldsymbol{c}_m^{\text{obsc}}$ est le fond thermique en e^-/sec, supposé constant au cours des acquisitions, au pixel $m \in \{1,\dots M\}$, où :
  - $\boldsymbol{c}_m^{\text{sky}}$ correspond à la contribution du ciel ;
  - $\boldsymbol{c}_m^{\text{instru}}$ correspond à la contribution de l'instrument ;
  - $\boldsymbol{c}_m^{\text{obsc}}$ est le courant d'obscurité du détecteur, c'est-à-dire la contribution pour $t = 0$, dans le cas d'IRDIS il est négligeable et peut être considéré nul.

  Les contributions $\boldsymbol{c}_m^{\text{sky}}$ et $\boldsymbol{c}_m^{\text{instru}}$ sont dues au fait que l'on observe dans l'infrarouge proche, tout rayonnement thermique émet dans l'infrarouge et est donc capté par le détecteur.
- $\boldsymbol{\beta}_{k,m} = \boldsymbol{\beta}_{k,m}^{\text{photons}} + \boldsymbol{\beta}_{k,m}^{\text{ro}}$ est un terme de bruit, pour le pixel $m \in \{1,\dots M\}$ à l'acquisition $k \in \{1,\dots K\}$, contenant toutes les fluctuations où :
  - $\boldsymbol{\beta}_{k,m}^{\text{photons}}$ est le bruit de photon qui suit une loi de Poisson (en e^-) ;
  - $\boldsymbol{\beta}_{k,m}^{\text{ro}}$ est le bruit de lecture du détecteur qui suit une loi centrée (en e^-), où ro signifie *read out* ;
- $\boldsymbol{b}_m$ est le biais du pixel $m \in \{1,\dots M\}$ du détecteur (en e^-), supposé constant au cours des acquisitions, il peut être considéré nul pour l'instrument IRDIS.

L'acquisition des données lors d'une nuit d'observation se fait en plusieurs temps. D'abord sont prises les données nécessaires à la calibration de l'instrument pour le centrage et la PSF. Puis sont faites les acquisitions de l'objet d'intérêt. Enfin sont alors acquises d'autres mesures nécessaires à la calibration instrumentale (centres et PSF) ainsi que les données de calibrations du fond thermique du ciel. Dans la journée sont alors faites un ensemble de mesures ne nécessitant pas l'ouverture du télescope ni une température basse à l'intérieur du télescope.

Les mesures de calibrations sont faites sous forme d'images de taille $M = 2048 \times 1024$ représentées par des matrices où chaque entrée représente le l'intensité obtenue en ADU pour un pixel. Ces mesures sont appelées différemment selon ce qui est enregistré.

Le fond thermique de l'instrument : Il est estimé à partir des données de calibration appelées *DARK BACKGROUND*. Ces données sont acquises avec l'instrument fermé avec un temps d'intégration $t_k = t \geq 0$ fixe. Ce qui correspond dans le modèle des données (6.1), au cas où, pour tout $m \in \{1,\dots M\}$ et tout $k \in \{1,\dots K\}$, on a $\boldsymbol{f}_{k,m} = 0$ et $\boldsymbol{c}_{k,m}^{\text{sky}} = 0$.

Le fond thermique du ciel : Il est estimé à partir des données de calibration appelées *SKY BACKGROUND*. Celles-ci sont prises en pointant l'instrument vers le ciel, à un endroit où il n'y a pas de cible particulière et en déplaçant le télescope de manière continue, avec le même temps d'intégration $t_k = t \geq 0$ fixé que les données objet. Ces données correspondent au cas dans le modèle des données (6.1), où pour tout $m \in \{1,\dots M\}$ et tout $k \in \{1,\dots K\}$, on a $\boldsymbol{f}_{k,m} = 0$. La contribution du ciel n'est pas aussi importante pour toutes les longueurs d'ondes. En particulier, en bande H celle-ci est négligeable par rapport à la contribution thermique de l'instrument lui-même.

Le champ plat : Il est estimé à partir des données de calibration appelées *FLAT FIELD*. Celles-ci sont obtenues en intégrant sur différent temps d'intégration $t_k \geq 0$ la lumière interne à l'instrument, dont le flux est constant et uniforme. Elles ont vocation à renormaliser

les valeurs de chaque pixels, en fonction de la proportion d'éclairement du détecteur en ce pixel. Elles correspondent dans le modèle des données (6.1), au cas où pour tout $m \in \{1,\dots M\}$, $k \in \{1,\dots K\}$, où $\boldsymbol{c}_{k,m}^{\text{sky}} = 0$. Le coefficient $\boldsymbol{f}_{k,m} = \boldsymbol{f}_m$ correspond au flux de la lampe éclairant le détecteur, qui ne dépend pas du rendement du détecteur et est supposé constant au fil des acquisitions $k \in \{1,\dots K\}$.

Comme on suppose le flux envoyé en chaque pixel constant au cours du temps, c'est-à-dire que pour tout $m \in \{1,\dots M\}$ la quantité $(\boldsymbol{q}_m \boldsymbol{f}_m + \boldsymbol{c}_m)\, t_k$ augmente linéairement avec le temps d'intégration t_k, cela permet d'avoir une estimation précise des pixels défectueux. En effet, ceux-ci ont un comportement incohérent par rapport aux autres. Ainsi en vérifiant si un pixel a bien une augmentation linéaire de flux au cours du temps, on peut vérifier si ce pixel est défectueux ou non, car le modèle (6.1) ne prend pas en compte la saturation. Afin de faire cette estimation, on souhaiterait faire un test d'hypothèse suivant la loi des données. Pour estimer la dispersion de celles-ci, on utilise les données objet.

Les données objet : Ces données, appelées données SCIENCE, sont les données contenant la cible d'intérêt et sont acquises de la manière présentées dans la section 1.1.4. La quantité $\boldsymbol{f}_{k,m}$ correspond donc aux données *calibrées* traitées tout au long de ce manuscrit. Le but de la calibration est donc d'obtenir la quantité $\boldsymbol{f}_{k,m}$ ainsi que la précision d'une telle quantité, sous forme de cartes de poids $\boldsymbol{W}_{k,m} = \text{Cov}(\boldsymbol{f})^{-1}_{k,m}$. On suppose que les données dont indépendantes vis-à-vis des acquisitions $k \in \{1,\dots K\}$ et vis-à-vis des autres pixels $m \in \{1,\dots M\}$. On peut donc écrire $\boldsymbol{W}_{k,m} = \text{Cov}(\boldsymbol{f}_{k,m})^{-1}$. Dans le cas où le pixel est défectueux ou de données manquantes, on a alors $\text{Cov}(\boldsymbol{f})_{k,m} = +\infty$ et donc $\boldsymbol{W}_{k,m} = 0$.

On utilise ces données pour étudier la statistique du bruit des données, de la même manière que présentée dans [Foi et al., 2008]. Pour un pixel donné et un temps $t_k \geq 0$, pour tout pixel $m \in \{1,\dots M\}$, à l'acquisition $k \in \{1,\dots K\}$, on pose :

$$\boldsymbol{y}_{k,m} = \left(\boldsymbol{q}_m \boldsymbol{f}_{k,m} + \boldsymbol{c}_m\right) t_k, \tag{6.2}$$

l'espérance du nombre de photo-électrons et

$$\boldsymbol{n}(\boldsymbol{z}_{k,m}) = \sqrt{\boldsymbol{y}_{k,m} + {\sigma_{\text{ro}}}^2_{k,m}}, \tag{6.3}$$

l'écart-type du nombre de photo-électrons. On pose également $(\xi_{k,m})_{k\in\{1,\dots K\},m\in\{1,\dots M\}}$ une variable aléatoire centrée réduite, c'est-à-dire telle que :

$$\forall k \in \{1,\dots K\}, \forall m \in \{1,\dots M\}, \quad \begin{cases} \mathbb{E}\left[\xi_{k,m}\right] = 0, \\ \text{Var}\left(\xi_{k,m}\right) = 1. \end{cases} \tag{6.4}$$

On a alors :

$$\boldsymbol{d}_{k,m} = \frac{\boldsymbol{y}_{k,m} + \boldsymbol{z}_{k,m}\xi_{k,m} + b}{g}, \tag{6.5}$$

De cela on déduit que :

$$\begin{cases} \mathbb{E}\left[\boldsymbol{d}_{k,m}\right] = \frac{\boldsymbol{y}_{k,m} + \boldsymbol{b}_m}{\boldsymbol{g}_m}, \\ \text{Var}\left(\boldsymbol{d}_{k,m}\right) = \left(\frac{\boldsymbol{z}}{\boldsymbol{g}}\right)^2. \end{cases} \tag{6.6}$$

D'après le modèle direct des données de calibration (6.1), pour un pixel $m \in \{1, \dots M\}$ donné et une configuration $k \in \{1, \dots K\}$, on a alors :

$$\begin{cases} \mathbb{E}\left[\boldsymbol{d}_{k,m}\right] = \frac{(\boldsymbol{q}_m \boldsymbol{f}_{k,m} + \boldsymbol{c}_m) t_k + \boldsymbol{b}_m}{\boldsymbol{g}_m} \\ \mathrm{Var}\left(\boldsymbol{d}_{k,m}\right) = \frac{(\boldsymbol{q}_m \boldsymbol{f}_{k,m} + \boldsymbol{c}_m) t_k + {\sigma_{\mathrm{ro}}}^2_{k,m}}{\boldsymbol{g}_m^2}. \end{cases} \tag{6.7}$$

Afin de simplifier les expressions par rapport au gain $(g_m)_{m \in \{1,\dots M\}}$, on pose alors $\mathring{\boldsymbol{f}}_{k,m} = \boldsymbol{f}_{k,m}/\boldsymbol{g}_m$, $\mathring{\boldsymbol{c}}_m = \boldsymbol{c}_m/\boldsymbol{g}_m$, $\mathring{\boldsymbol{b}}_m = \boldsymbol{b}_m/\boldsymbol{g}_m$ et $\mathring{\sigma}^2_{\mathrm{ro}\,k,m} = \sigma^2_{\mathrm{ro}\,k,m}/\boldsymbol{g}_m$. On a alors :

$$\boldsymbol{d}_{k,m} = \left(\mathring{\boldsymbol{f}}_{k,m} + \mathring{\boldsymbol{c}}_m\right) t_k + \mathring{\boldsymbol{b}}_m + \mathring{\beta}_{k,m},$$

et

$$\begin{cases} \mathbb{E}\left[\boldsymbol{d}_{k,m}\right] = (\mathring{\boldsymbol{f}}_{k,m} + \mathring{\boldsymbol{c}}_m) t_k + \mathring{\boldsymbol{b}}_m, \\ \mathrm{Var}\left(\boldsymbol{d}_{k,m}\right) = \frac{(\mathring{\boldsymbol{f}}_{k,m} + \mathring{\boldsymbol{c}}_m) t_k + \mathring{\sigma}^2_{\mathrm{ro}\,k,m}}{\boldsymbol{g}_m}, \end{cases} \tag{6.8}$$

On souhaite donc estimer conjointement $\vartheta_{k,m} = (\mathring{\boldsymbol{f}}_{k,m}, \mathring{\boldsymbol{c}}_{k,m}, \mathring{\boldsymbol{b}}_m, \boldsymbol{g}_m, {\sigma_{\mathrm{ro}}}^2_{k,m})$. Le problème étant non-linéaire, l'idée est de scinder le problème en deux et d'itérer de manière alternée et séparable en les pixels :

$$\left(\widehat{\mathring{\boldsymbol{f}}}_{k,m}, \widehat{\mathring{\boldsymbol{c}}}_m, \right) = \underset{\mathring{\boldsymbol{f}}_{k,m} \geq 0, \mathring{\boldsymbol{c}}_{k,m} \geq 0}{\mathrm{argmin}} \sum_{k=1}^{K} \left[\frac{\left(\boldsymbol{d}_{k,m} - (\mathring{\boldsymbol{f}}_{k,m} + \mathring{\boldsymbol{c}}_m) t_k - \mathring{\boldsymbol{b}}_m\right)^2}{v_{k,m}} + \log\left(v_{k,m}\right) \right], \tag{6.9}$$

où $v_{k,m}$ est dans un premier temps la variance empirique puis la variance estimée après la deuxième étape :

$$\begin{aligned} \widehat{\mathring{\sigma}^2_{\mathrm{ro}}}_{k,m} = \underset{\mathring{\sigma}^2_{\mathrm{ro}\,k,m}}{\mathrm{argmin}} \quad \Bigg\{ \min_{\boldsymbol{g}_m} \sum_{k=1}^{K} \Bigg[\boldsymbol{g}_m \frac{\left(\boldsymbol{d}_{k,m} - (\mathring{\boldsymbol{f}}_{k,m} + \mathring{\boldsymbol{c}}_m) t_k - \mathring{\boldsymbol{b}}_m\right)^2}{(\mathring{\boldsymbol{f}}_{k,m} + \mathring{\boldsymbol{c}}_m) t_k + \mathring{\sigma}^2_{\mathrm{ro}\,k,m}} \\ + \log\left((\mathring{\boldsymbol{f}}_{k,m} + \mathring{\boldsymbol{c}}_m) t_k + \mathring{\sigma}^2_{\mathrm{ro}\,k,m}\right) - \log\left(\boldsymbol{g}_m\right) \Bigg] \Bigg\}. \end{aligned} \tag{6.10}$$

Un article présentant cette méthode de calibration du détecteur, de manière jointe à partir de la formulation (6.10), et de détection des pixels morts, avec une application au détecteur de l'instrument IRDIS est en cours d'écriture (Thiébaut et al. en préparation). Cette méthode de calibration est originale car toutes les mesures sont ajustées en même temps *a contrario* avec les méthodes de calibration de l'état-de-l'art présentée dans la section 1.1.4, qui consiste en une simple soustraction de la moyenne du fond thermique et d'une division par une moyenne temporelle des données *FLAT FILED*. De plus cette nouvelle méthode de calibration est optimale au sens du maximum de vraisemblance.

Sont présentées dans la suite les pistes d'estimation des pixels morts étudiées dans cette thèse.

Détection des pixels morts par régression linéaire : Dans un premier temps on se contente de résoudre le problème linéaire 6.9. On propose alors une sélection des pixels valides sur les critères : $\mathring{\boldsymbol{f}}_{k,m} \geq 0$, $\mathring{\boldsymbol{c}}_m \geq 0$ et sur la loi du χ^2 centré puis réduite

$$\frac{\left(\chi^2_m - K^{\text{free}}\right)}{\sqrt{2K^{\text{free}}}} < \upsilon, \tag{6.11}$$

où $K^{\text{free}} = K - P^{\text{param}}$ correspond au nombre de degrés de liberté, avec P^{param} le nombre de paramètres à estimés, et

$$\chi^2_m = \sum_{k=1}^{K} \frac{\left(\boldsymbol{d}_{k,m} - (\widehat{\mathring{\boldsymbol{f}}}_{k,m} + \widehat{\mathring{\boldsymbol{c}}}_m)t_k\right)^2}{v_m}.$$

J'impose alors un seuil manuel à cette valeur variant autour de $\upsilon = 3,5$.

Les données nous permettant facilement de vérifier ce critère sont les données *FLAT FIELD*. Elles contiennent en général cinq temps d'intégration différents avec 40 réalisations du bruit chacune. J'estime le flux par un moindre carré pondéré par la variance des flats puis j'impose deux conditions : une condition de positivité sur le flux et une condition de linéarité qui s'exprime sur le χ^2.

Je commence par faire cette étude sur un échantillon de 3×3 pixels. Je trace tout d'abord leurs évolutions avec l'augmentation du temps d'intégration. Afin de condenser les données multiples, je trace leurs moyennes empiriques. La figure 6.1 montre un tel tracé pour un extrait de 3×3 pixels. D'après les évolutions tracées sur la figure 6.1, on peut émettre l'hypothèse que les pixels 4, 5, 6, 8 et 9 sont valides tandis que les autres sont défectueux. On appellera

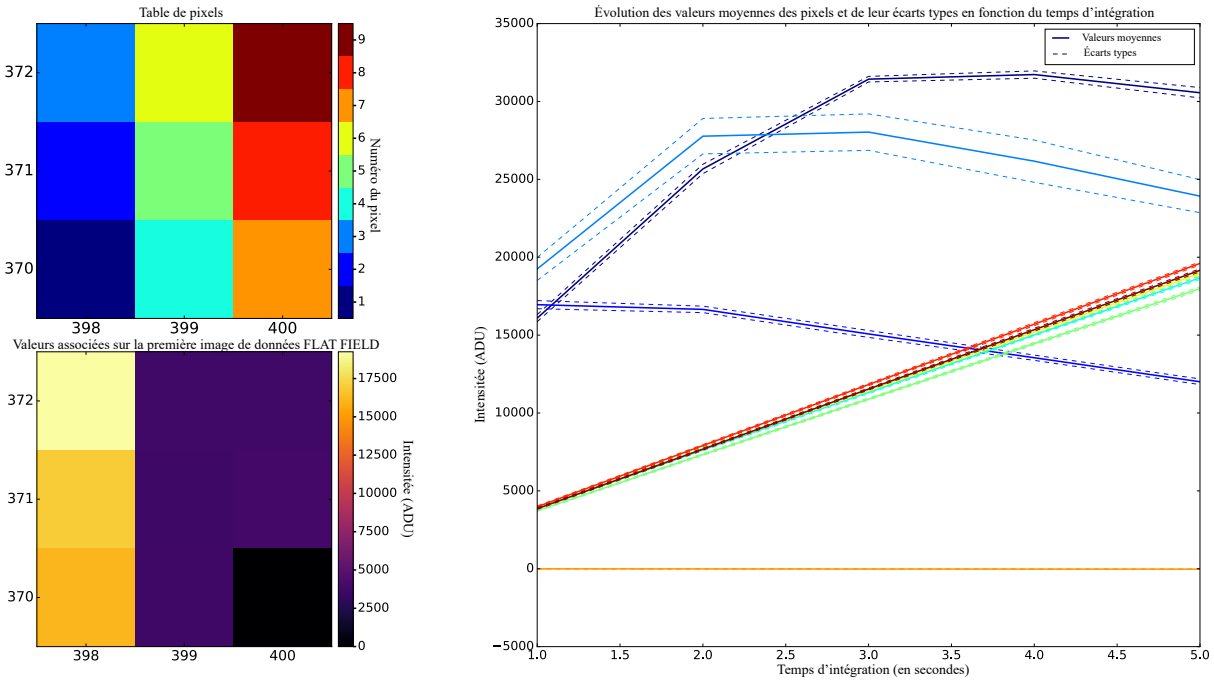


FIGURE 6.1 – Exemple du comportement de neuf pixels du détecteur de l'instrument ESO/VLT-SPHERE IRDIS en fonction du temps d'intégration à flux entrant fixe.

pixels froids ceux à valeur constante nulle ou proche de 0, comme le pixel 7 et ceux à valeur constante très élevée, sont appelés pixels chauds.

À partir des valeurs obtenues pour les paramètres $\mathring{f}_{k,m}$, $\mathring{c}_{k,m}$, je trace, sur la figure 6.2, la moyenne alors estimée par-dessus les moyennes empiriques pour l'échantillon de taille 5×5. Je trace également les cartes de booléens correspondant, d'une part au respect des conditions $\mathring{f}_{k,m} \geq 0$ et $\mathring{c}_m \geq 0$, et d'autre part une valeur du χ^2 centré réduit inférieur à $3,5$. Enfin je trace la carte résultante de l'intersection des deux cartes précédentes.

On remarque que d'après la figure 6.2, le pixel à la position $(369, 400)$ et considéré comme mort par la contrainte sur le χ^2, car sa moyenne est légèrement plus élevée que celle des autres pixels valides. De plus l'hypothèse émise plus tôt sur les pixels de la figure 6.1 est bien vérifiée.

Il est important de noter que le nombre de données de calibrations du détecteur pour IRDIS est limité, notamment au niveau des fonds instrumentaux et du ciel. Les premiers sont souvent disponibles en un seul exemplaire et les seconds ont souvent entre 5 et 10 réalisations seulement. Il faut donc prendre cela en compte afin de pouvoir avoir une méthode globale qui puisse s'appliquer à tous les filtres. De plus dans les jeux de données anciens, il n'y a qu'une acquisition seule par temps de pause, ce qui biaise fortement la détection des pixels morts.

Remarque : Du fait des possibilités offertes par les méthodes d'interpolation et les méthodes régularisées pour combler les données manquantes, il est préférable de sélectionner trop de pixel morts. De ce fait, en pratique on croise différents critères de sélection : une sé-

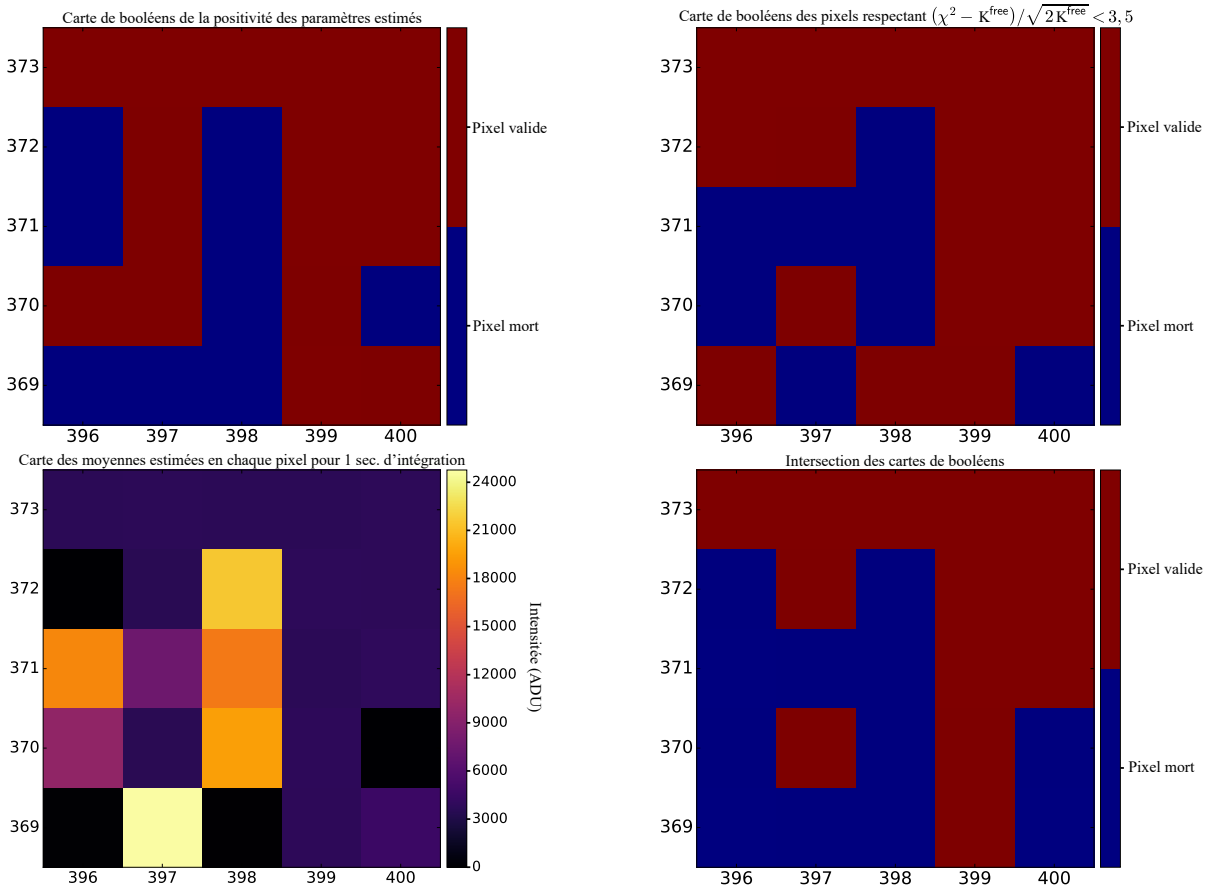


FIGURE 6.2 – Cartographie de vingt-cinq pixels morts du détecteur de l'instrument ESO/VLT-SPHERE IRDIS à partir des données *FLAT FIELD*

lection sur la covariance de tous les paramètres de calibration estimés conjointement et un critère sur la valeur de la co-logvraisemblance.

6.1.2 Calibration instrumentale

Avant de pouvoir appliquer l'algorithme à des données calibrées, il reste encore quelques paramètres à ajuster.

Calcul des centres gauche et droite : Il est nécessaire pour permettre, dans le cas séparable, le recentrage des images gauche/droite puis temporel à un même pixel $n \in \{1, \dots N\}$, à l'aide de translations respectives, suivies d'une interpolation. Dans le cas non-séparable, l'inverse de ces translations gauche et droite, suivies d'une interpolation est inclue dans le modèle des données, pour aller de l'espace des paramètres à l'espace des données.

Pour calculer les centres des images gauche et droite, on utilise les données de calibration intitulées *STAR CENTER WAFFLE*. Il existe généralement deux ensembles de données de ce type, l'un pré-observations et l'autre post-observations. Chacun de ces deux ensembles peu posséder une ou plusieurs acquisitions.

Ces données consistent en des données objets avec 4 points autour du centre de l'objet d'intérêt, qui sont des répliques de la PSF stellaires, respectivement à gauche et à droite. L'idée est donc d'estimer, après calibration de ces données, le centre de ces quatre PSF en ajustant des gaussiennes à deux dimensions. Selon le filtre utilisé, les gaussiennes peuvent se retrouver étirées, il est donc nécessaire de supposer ces gaussiennes anisotropes. Pour procéder à l'estimation, je fais d'abord une estimation manuelle de ces centres ainsi qu'une estimation manuelle de la largeur à mi-hauteur. Une fois que ces quatre centres sont déterminés, je calcule le centre de ces quatre points. Ainsi, je peux calculer le décentrage de l'image de droite par rapport à l'image de gauche et l'inclure dans le modèle à l'aide d'une translation, et de son adjoint. La figure 6.3 montre les résultats obtenus sur la première acquisition du fichier de

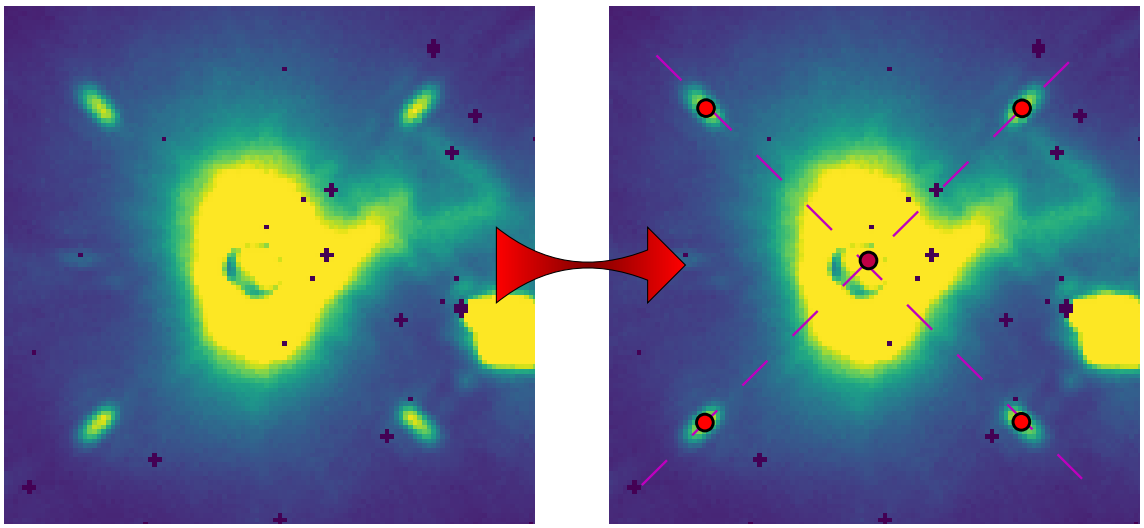


FIGURE 6.3 – Extrait d'une image d'un jeu de données *STAR CENTER WAFFLE* de la cible T Tauri [Kasper et al., 2016], observée en polarimétrie, étudié dans la section 6.2.2. Sont tracés les centres de chaque point estimés et le centre obtenu comme barycentre de ces points.

calibration pré-observations du jeu de données T Tauri dont les résultats sont présentés dans la section 6.2.2.

TABLE 6.1 – Tableau des valeurs obtenues pour les centres droit et gauche pour les jeux pré-observations et post-observations.

		Gauche		Droite	
Données	N° Acquisition	X	Y	X	Y
Pré-obs	1	478.38	521.77	1503.79	511.03
	2	478.42	521.73	1503.81	511.02
	3	478.43	521.72	1503.85	510.98
Post-obs	1	478.38	521.54	1503.74	510.79
	2	478.38	521.57	1503.77	510.84
	3	478.36	521.6	1503.75	510.84

En faisant l'estimation des centres sur toutes les acquisitions des données prises au début et à la fin de la séquence d'observation (cf. Table 6.1), on se rend compte qu'il y a un léger décalage temporel. Sur les jeux de données de T Tauri, cela semble rester stable entre les jeux de données pré et post-observations.

Le décalage temporel au cours de la séquence peut être estimé facilement dans le cas où un compagnon est présent dans le champ. En effet en ajustant le centre d'une gaussienne sur la PSF du compagnon, il est possible d'estimer assez fidèlement le décalage.

Il est possible dans certains jeux de données d'avoir également du *dithering*, c'est-à-dire un décalage artificiel d'un ou plusieurs pixels. La valeur d'un tel décalage est alors indiqué dans la description du fichier de données.

Estimation de la PSF : Afin de pouvoir déconvoluer les données astrophysiques, il est nécessaire d'avoir une estimation correcte de la PSF. Cette estimation se fait à partir des données

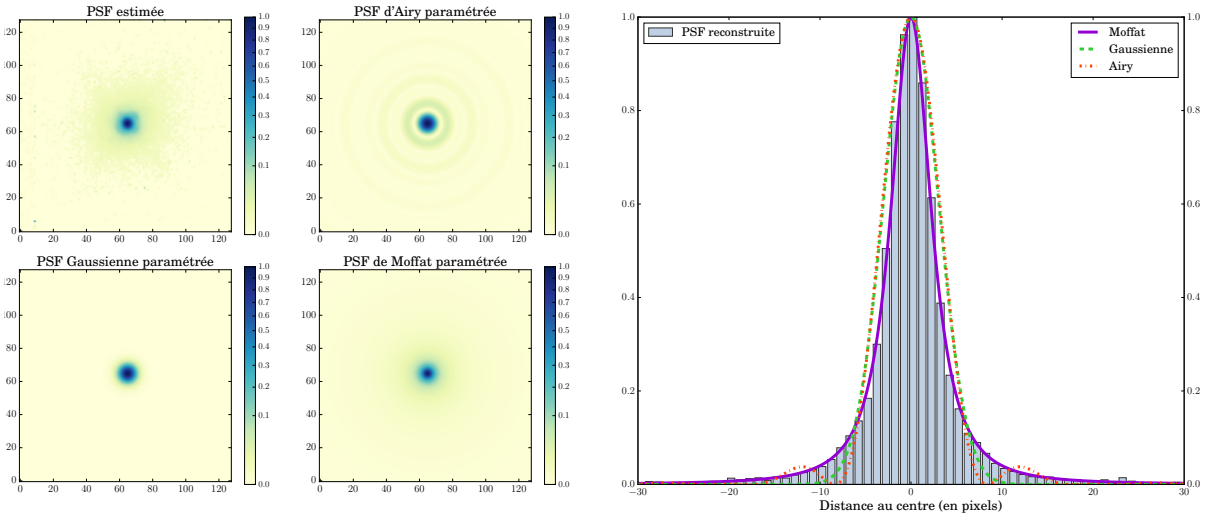


FIGURE 6.4 – Comparaison des différentes formes de PSF paramétrées, pour des paramètres ajustés sur la PSF estimée.

appelées *FLUX CALIB CORO*. Cette estimation est alors faite en deux temps :

— Une première étape de calibration et réduction des données de PSF : les données de PSF sont *calibrées* de la même manière que les données objets. Ensuite, à partir d'un modèle des données *calibrées* non-séparables $(\boldsymbol{d}^{\mathrm{nS}})_{k\in\{1,\dots K\}} \in \mathbb{R}^M$ de la forme

$$\boldsymbol{d}_k^{\mathrm{nS}} = \boldsymbol{T}_k \boldsymbol{x} + \beta_k$$

où $\boldsymbol{T}_k : \mathbb{R}^N \to \mathbb{R}^M$ correspond aux transformations du détecteur gauche et droite permettant de passer au domaine de la reconstruction au domaine des données, et $\beta_k \sim \mathcal{N}(0, \Sigma_k^2)$ un terme de bruit indépendant. On reconstruit alors la carte de PSF $\widehat{\boldsymbol{x}}$ à partir d'un moindre carré, pondéré par la matrice Σ^{-2} multipliée par la carte de pixels valide, et régularisé par la régularisation à préservation de bord TV hyperbolique, donnée par la formulation (1.50). Une telle estimation est présentée sur la figure 6.4, à partir des données de calibration de la cible RXJ 1615.

— Une seconde étape est l'ajustement d'un modèle de PSF sur l'estimation, dont les paramètres sont la limite de diffraction de l'instrument et la taille de l'obstruction centrale du télescope sous forme de rapport à la taille du diamètre. L'ajustement est alors fait par une méthode alternée, présente dans le package Julia `PointSpreadFunctions` [Thiébaut, 2020]. Une PSF d'une pupille avec obstruction centrale correspond en général à une tâche d'Airy. Cependant cela ne tient pas compte des araignées qui peuvent être plus ou moins marquées selon la longueur d'onde. Il est possible de modéliser également la PSF par une Gaussienne ou une fonction de Moffat. Sur la figure 6.4, sont comparées les images des trois paramétrisations de PSF. Sont également tracées des coupes horizontales de chaque PSF. On voit que la paramétrisation qui s'ajuste au mieux sur le modèle est celle de la fonction de Moffat. On la choisit donc comme modèle de PSF lors des reconstructions.

6.1.3 Pré-traitement des données *calibrées*

Afin de pouvoir appliquer les modèles séparables présentés dans le chapitre 2 ainsi que les méthodes de l'état-de-l'art, il est nécessaire de pré-traiter les données. Ce pré-traitement se fait dans l'ordre suivant :

1. Les pixels morts des données sont interpolés en faisant une interpolation des plus proches voisins non-nuls. L'interpolation est également faite sur les cartes de poids.
2. Chaque acquisition $k \in \{1, \dots K\}$ des données $(\boldsymbol{d}_k)_{k\in\{1,\dots K\}}$ est ensuite coupée en deux images $j \in \{1; 2\}$.
3. Chaque donnée résultante $(\boldsymbol{d}_{j,k})_{j\in\{1;2\},k\in\{1,\dots K\}}$ est alors recentrée en faisant une translation des images, afin que le centre de l'étoile hôte coïncide avec le centre de l'étoile hôte sur la première image, c'est-à-dire pour $j = 1$ et $k = 1$.
4. Les données $(\boldsymbol{d}_{j,k})_{j\in\{1;2\},k\in\{1,\dots K\}}$ sont alors tournées de l'inverse additif de l'angle correspondant à l'angle de rotation artificielle du ciel, induit par le dérotateur dans l'instrument, par rapport à l'axe nord.

Il en ressort un cube de taille $N \times K \times 2$ où N est la dimension de l'image (*e.g.* $N = 1024 \times 1024$) et K le nombre total d'acquisitions (*e.g.* $K = 64$).

6.2 Reconstruction et étude de différentes cibles astrophysiques

Dans ce manuscrit, j'ai présenté différentes méthodes de reconstruction des cartes d'intensité non-polarisée I^{u}, d'intensité polarisée I^{p} et d'angle de polarisation θ. L'application de ces méthodes à différents jeux de données astrophysiques est actuellement en cours et je présente dans cette section les résultats sans et avec déconvolution sur la cible RXJ 1615 [de Boer et al., 2016, Avenhaus et al., 2018], ainsi que les premiers résultats pour la cible T TAU [Kasper et al., 2016] sans déconvolution et les résultats sans et avec déconvolution pour une nouvelle cible, découverte récente par les observations SPHERE du temps garanti, encore non étudiée en polarimétrie.

Dans cette section nous sommes plus particulièrement intéressés par l'intensité polarisée I^{p} et l'angle de polarisation θ. Il n'est en effet pas possible à partir de ces jeux de données d'extraire l'intensité non-polarisée du disque, qui est mélangé aux résidus de lumière stellaire contenus dans I^{u}, car il n'y a pas de rotation artificielle du champ. Cependant l'utilisation jointe de l'imagerie différentielle angulaire (ADI) et l'imagerie polarimétrique (DPI), utilisée récemment pour plusieurs jeux de données, permettra ce démélange. Le traitement de tels jeux de données est un travail en cours qui ne sera pas présenté dans ce manuscrit.

Les résultats sont présentés de la manière suivante, sauf indication contraire : à gauche est présentée la carte d'intensité polarisée I^{p}, multipliée en chaque pixel par la distance au carré de celui-ci au centre de l'étoile r_n^2 où si l'on note (x_0, y_0) le centre de l'étoile, alors

$$\forall n \in \{1, \dots N\}, \quad r_n^2 = (x_0 - x_n)^2 + (y_0 - y_n)^2, \tag{6.12}$$

où (x_n, y_n) sont les coordonnées cartésiennes du pixel $n \in \{1, \dots N\}$. Ce rehaussement des intensités est dû à une baisse d'intensité des détails des objets étudiés, proportionnelle au carré de la distance du détail au centre.

Sur l'image de droite, la carte identique est présentée avec, tracés en blancs par-dessus, les angles de polarisation linéaire associés sous forme de vecteur. L'orientation des vecteurs correspond à l'angle par rapport à l'axe nord. Les angles ne sont tracés qu'aux endroits où l'intensité polarisée dépasse un certain seuil de sensibilité fixée manuellement. Les valeurs tracées sur l'axe des abscisses et des ordonnées correspondent à la distance au centre de l'étoile hôte, en secondes d'arc.

6.2.1 RXJ 1615

Il s'agit d'un disque de transition autour de l'étoile RX J1615.3−3255, une étoile de type T Tauri, c'est-à-dire une étoile jeune de faible masse. Une étude complète de ce disque est faite dans les articles [de Boer et al., 2016, Avenhaus et al., 2018]. Il a été observé en polarimétrie à la fois en bande H et en bande J. Il est cependant plus visible en bande H, c'est pourquoi je présente dans la suite les différents résultats que j'ai obtenus en bande H.

La figure 6.5 présente les résultats présentés dans le papier [Avenhaus et al., 2018], où sont tracées les cartes d'intensité projetée orthogonalement, appelée $\mathrm{Q}_\phi = P_\perp$ avec :

Rappel éq. (1.28) $$\forall n \in \{1, \dots N\}, \quad P_{n,\perp} = \mathrm{Q}_n \cos(2\varphi_n) + \mathrm{U}_n \sin(2\varphi_n),$$

où les paramètres de Stokes Q et U sont obtenus par la méthode du Double Ratio, présentée dans la section 1.1.5. Sur la figure, 6.5b sont indiqués les différentes structures d'intérêt.

6.2.1.1 Sans déconvolution

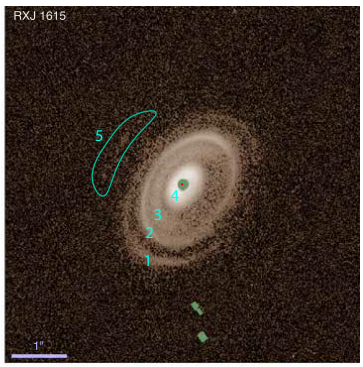
Je présente ici les résultats obtenus avec la méthode de la Double Différence et avec mes méthodes non-linéaires séparables et non-séparable régularisée par a priori de lissage avec préservation des bords.

La figure 6.6 présente les reconstructions obtenues avec les méthodes séparables et non-séparables sans déconvolution. Les figures 6.6a et 6.6b présentent les reconstructions obtenues avec la Double Différence et la MnLS sur données *pré-traitées*. J'ai procédé au pré-traitement de la manière indiquée dans la section 6.1.3. La différence entre les deux méthodes est clairement visible au niveau des pixels interpolés. En dehors de cela les résultats sont assez similaires.

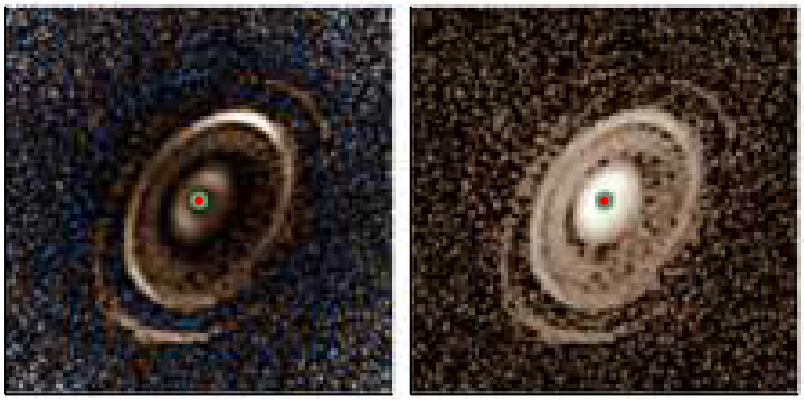
La figure 6.7 présente les cartes d'erreurs minimales en chaque pixel, pour les méthodes séparables de la Double Différence et la MnLS, obtenue de la manière expliquée dans la section 5.2. L'erreur sur l'intensité polarisée est relative à l'intensité polarisée en chaque pixel et est exprimée en pourcentage. L'erreur angulaire est absolue et exprimée en degrés. Pour une étude plus approfondie de ces cartes d'erreurs se référer au chapitre 2.

La figure 6.6c présente les reconstructions obtenues avec la MnLnS. Sur cette reconstruction, le niveau de fond est plus élevé que sur les reconstructions séparables. Cela peut soit provenir d'une mauvaise estimation du fond, soit de la polarisation instrumentale, et est d'autant plus visible avec une reconstruction où la sous-régularisation augmente le bruit. C'est en effet une des difficultés à laquelle j'ai été confrontée avec ce jeu de données, car les structures de faible intensité se retrouvent « noyées »dans le fond. D'autant plus que, du fait de sa brillance, il était alors difficile de trouver un réglage des hyperparamètres optimal, un hyperparamètre un peu trop élevé entrainant rapidement une perte de résolution de toutes les structures d'intensité proches du fond.

Dans le cas d'un problème de calibration du fond thermique, l'utilisation d'une méthode de repondération alternée, permettant l'autocalibration du détecteur jointe à la reconstruction est une piste intéressante pour améliorer la reconstruction des structures de faible intensité. Dans le cas de la polarisation instrumentale, l'utilisation de l'outil de calibration IRDAP [van Holstein et al., 2020] en permet la correction, ou bien en minimisant le paramètre U_ϕ qui représente la polarisation instrumentale, comme présenté dans la section 1.1.5 ou encore principalement dans [Avenhaus et al., 2018].

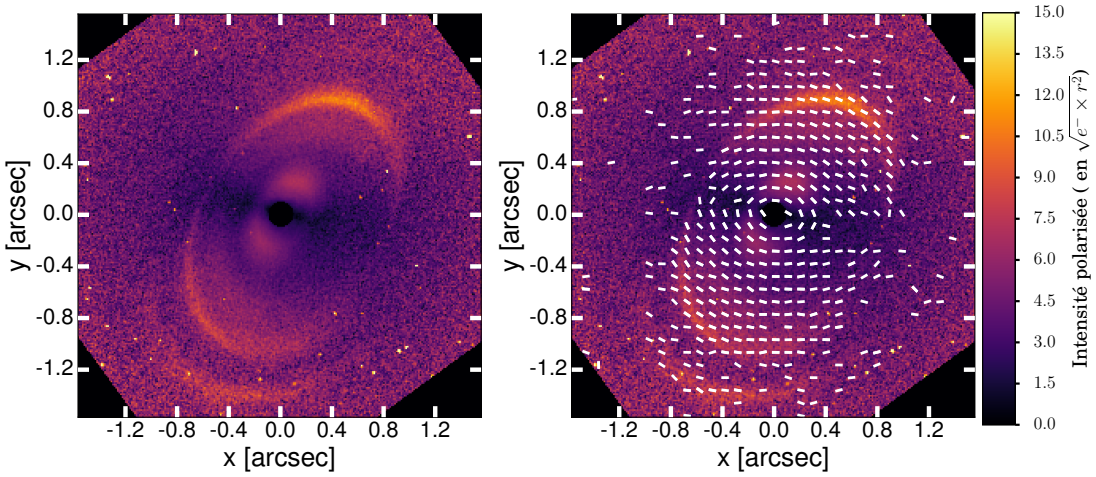


(a) Figure tracée en log, sont marquées les zones d'intérêt, numérotées de 1 à 5.

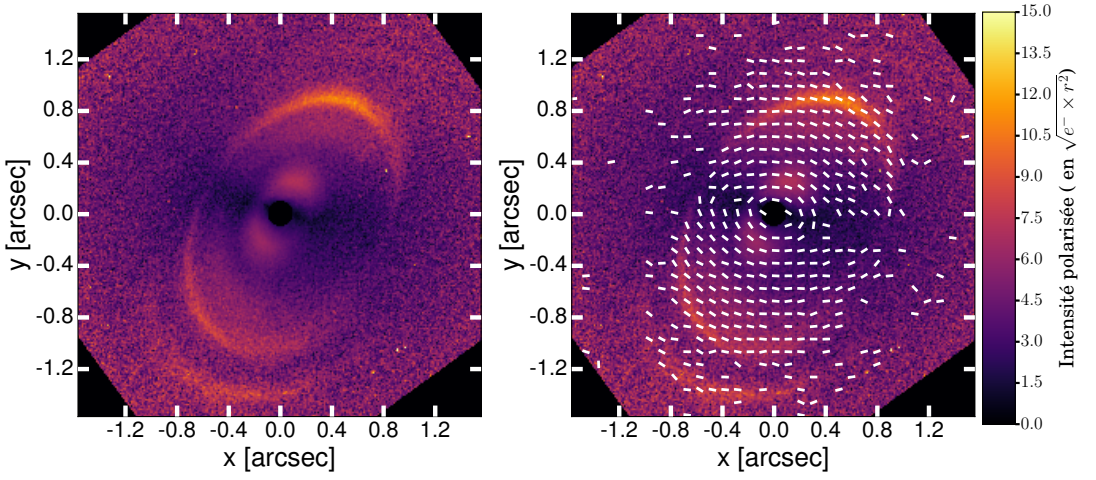


(b) Figure où le paramètre Q_ϕ a été corrigé de la polarisation instrumentale. La seconde est affichée en log.

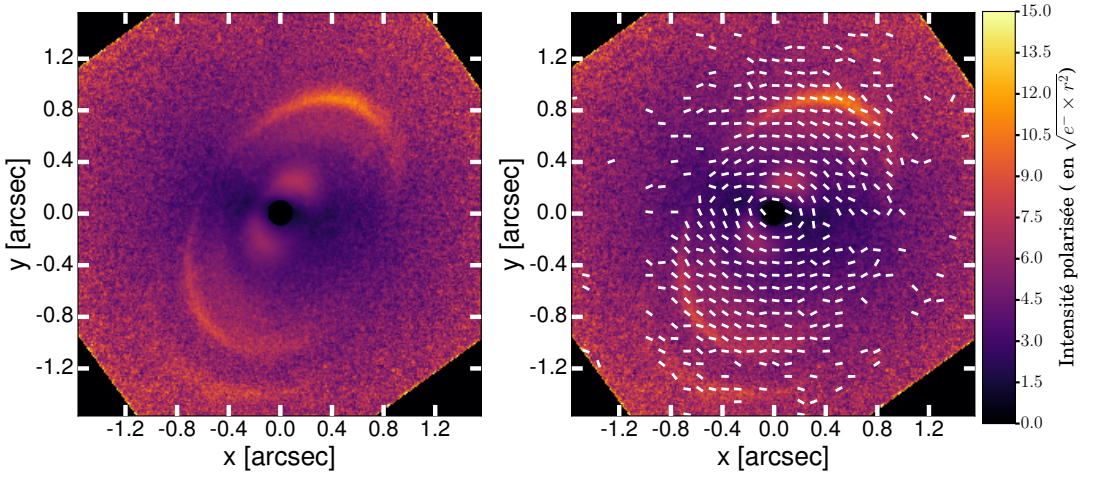
FIGURE 6.5 – Image de Q_ϕ en bande H (source : [Avenhaus et al., 2018]), obtenue par la Double Différence.



(a) Reconstruction obtenue avec la méthode de la Double Différence (cf. section 1.1.5).

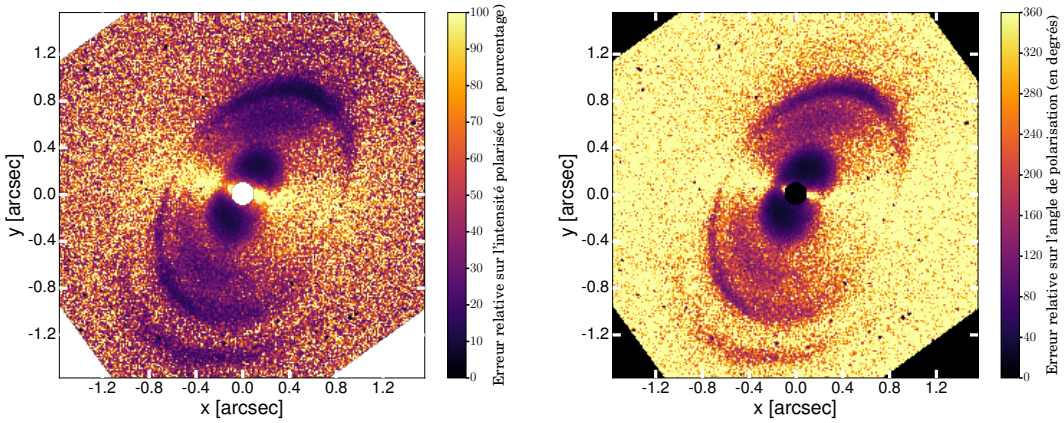


(b) Reconstruction obtenue avec la Méthode non-Linéaire Séparable (MnLS) (cf. section 2.1).

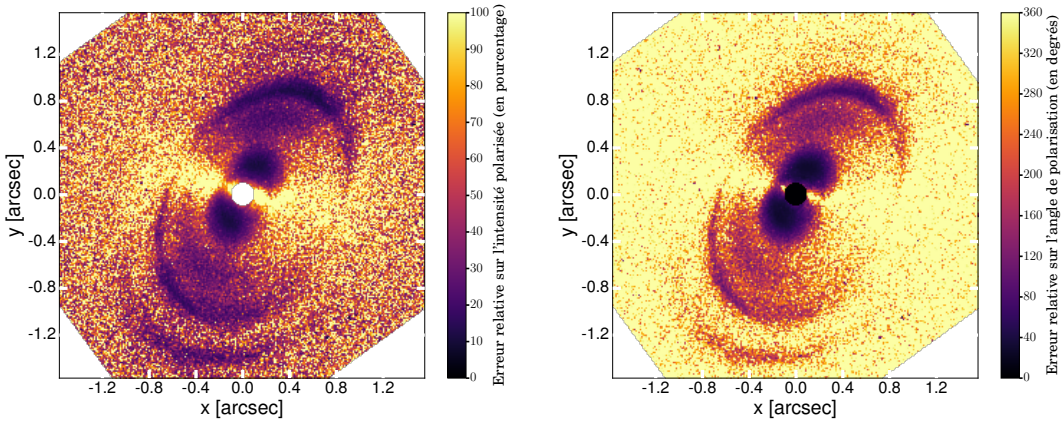


(c) Reconstruction obtenue avec la Méthode non-Linéaire non-Séparable (MnLnS) sans déconvolution (cf. section 3.2). La reconstruction est régularisée avec a priori de lissage à bords francs. Elle est obtenue pour les hyperparamètres $\mu = 10^{0,1}$, $\lambda_{I^{\mathrm{u}}} = 10^2$, et $\lambda_{I^{\mathrm{p}}} = 10^{2,5}$.

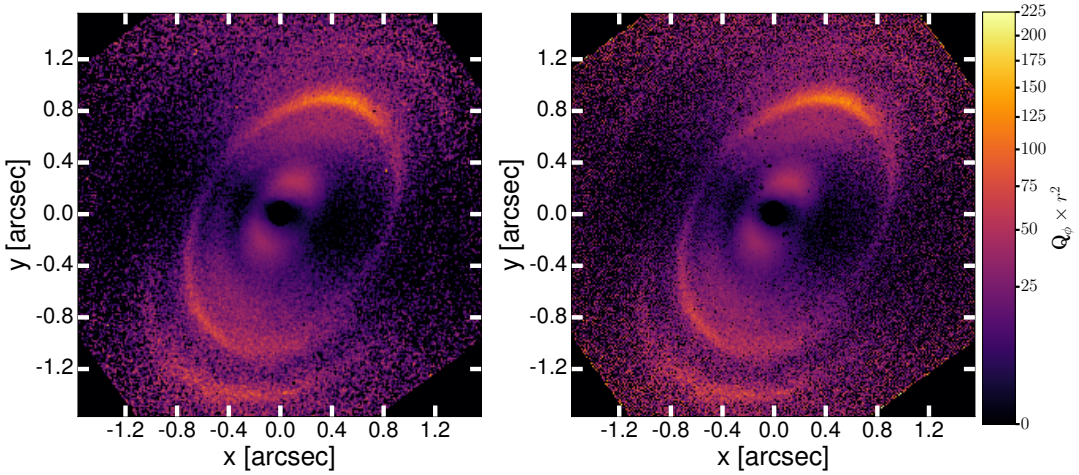
FIGURE 6.6 – Reconstruction du disque autour de l'étoile RX J1615.3–3255 sans déconvolution.



(a) Erreur relative de reconstruction du paramètre I^{p} (carte de gauche) et erreur angulaire du paramètre θ (carte de droite) pour la Double Différence.



(b) Erreur relative de reconstruction du paramètre I^{p} (carte de gauche) et erreur angulaire du paramètre θ (carte de droite) pour la MnLS.



(c) Intensité polarisée projetée Q_ϕ obtenue par la MnLS (carte de gauche) dont l'intensité et l'angle sont présentés sur la figure 6.6b et par la MnLnS (carte de droite) dont l'intensité et l'angle sont présentés sur la figure 6.6a.

FIGURE 6.7 – Erreur relative de reconstruction du disque autour de l'étoile RX J1615.3–3255 avec la méthode de la Double Différence et la MnLS. Projection radial de la MnLS et de la MnLnS (cf. chapitre 3).

La figure 6.7c montre l'intensité polarisée projetée orthogonalement Q_ϕ multipliée par r^2, obtenue d'après la formule (1.28) à partir des paramètres de Stokes Q et U, eux même obtenus par la MnLS et la MnLnS, à partir des cartes Ip et θ selon les formules (1.16). On remarque que le contraste est bien meilleur que sur les cartes de I^{p} présentées sur la figure 6.15. Cela permet de faire apparaître la structure 5 qui, dans les autres reconstructions sans déconvolution, était noyée dans le fond. De tels résultats semblent donc indiquer que le fond brillant n'est pas dû à un problème de calibration du fond thermique mais plutôt un problème de calibration de la polarisation instrumentale. La calibration de la polarisation instrumentale et sa correction, à l'aide de l'outil IRDAP, pourra donc permettre une meilleure estimation.

6.2.1.2 Avec déconvolution

La figure 6.8 présente les résultats déconvolués obtenus à partir de la MLnS-D présentée dans le chapitre 4, pour deux jeux d'hyperparamètres. Le premier jeu, sur la figure 6.8a, correspond à un cas sur-régularisé et laisse apparaitre les structures de faible intensité et une meilleure estimation de leurs angles de polarisation, comme la structure en haut à gauche de l'image. Le second jeu, sur la figure 6.8b, correspond à un cas sous-régularisé, mène à une résolution plus précise des structures brillantes, et permet d'observer le dédoublement de l'anneau intermédiaire, qui est très légèrement visible dans [Avenhaus et al., 2018], sur la figure reproduite en 6.5b, mais qui est bien mieux reconstruit ici.

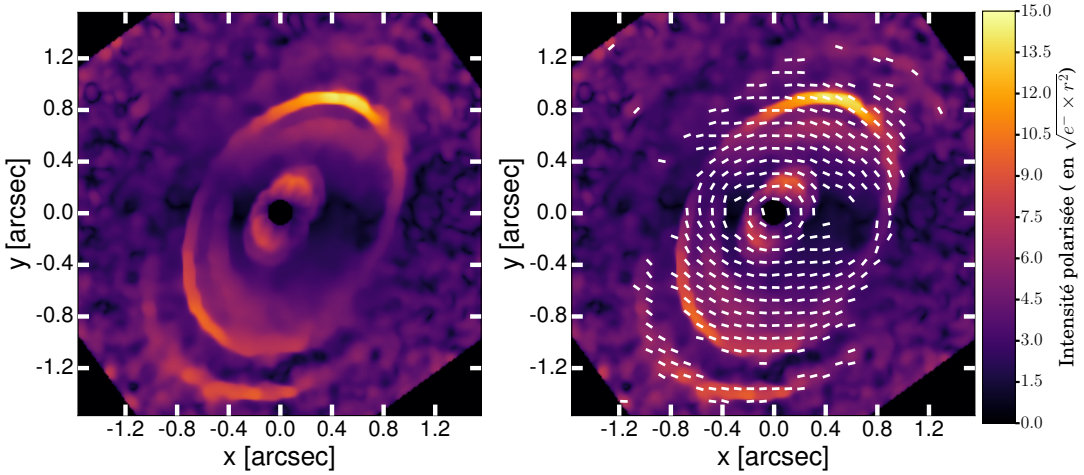
Pour l'application de la déconvolution sur ce jeu de données, j'ai également été confrontée à la difficulté du réglage des hyperparamètres dans ce cas de déconvolution. En effet, comme les détails sont présents à différentes échelles de régularisation, il était délicat de trouver un compromis. C'est pourquoi j'ai conclu qu'il était nécessaire de comparer les résultats pour au moins deux jeux d'hyperparamètres menant à deux régularisations différentes, une pour les régions à faible flux et une qui permet de gagner au maximum en résolution angulaire.

La figure 6.8c présente les cartes de Q_ϕ calculées à partir des cartes déconvoluées de la figure 6.8a d'une part et 6.8b d'autre part. Cette projection permet à nouveau d'améliorer le contraste, mais dans le cas sous-régularisé, tend à augmenter également le bruit. Pour la carte de gauche, la structure n°5 est clairement visible. On peut donc supposer que la reconstruction après correction de la polarisation instrumentale par IRDAP améliorera à nouveau la qualité de la reconstruction.

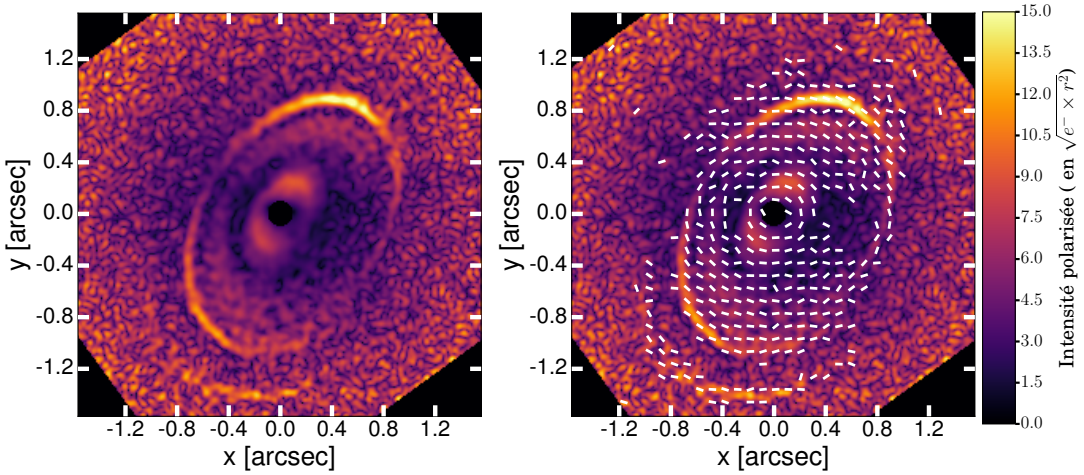
6.2.2 T Tauri

Le système T Tauri est un ensemble d'au moins trois jeunes étoiles qui ont donné leur nom aux étoiles de type T Tauri. Il est localisé dans la région de formation d'étoile Taurus Aurigae. T Tauri N est une de ces étoiles autour de laquelle il a été découvert une structure circumstellaire, observée en imagerie directe avec IRDIS dans [Kasper et al., 2016]. La figure 6.9 présente les résultats en intensité non-polarisée obtenus avec IRDIS par les auteurs en mode imagerie, sur lesquels le résidu stellaire est présent, comme visible sur la figure 6.9.

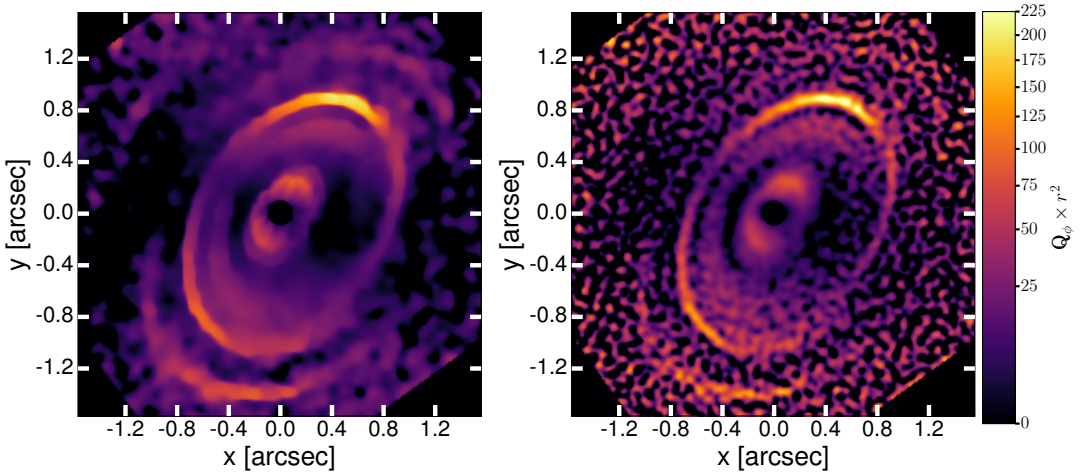
Dans cette section, je présente les résultats obtenus sur la même cible à partir des données DPI de l'instrument IRDIS en bande H, sans déconvolution, pour lesquelles aucune publication n'est encore disponible. La déconvolution de cette cible fait partie des travaux en cours et ne sera pas présentée dans ce manuscrit.



(a) Reconstruction obtenue pour les hyperparamètres $\mu = 10^{-4}$, $\lambda_{\mathrm{I}} = 1$, et $\lambda_{\mathrm{Q,U}} = 10^{-0.5}$.



(b) Reconstruction obtenue pour les d'hyperparamètres $\mu = 10^{-3}$, $\lambda_{\mathrm{I}} = 1$, et $\lambda_{\mathrm{Q,U}} = 10^{-0.4}$.



(c) Intensité projetée Q_ϕ pour les reconstructions par la MLnS-D dont les intensités polarisées et angles sont présentés respectivement sur les figures (a) et (b).

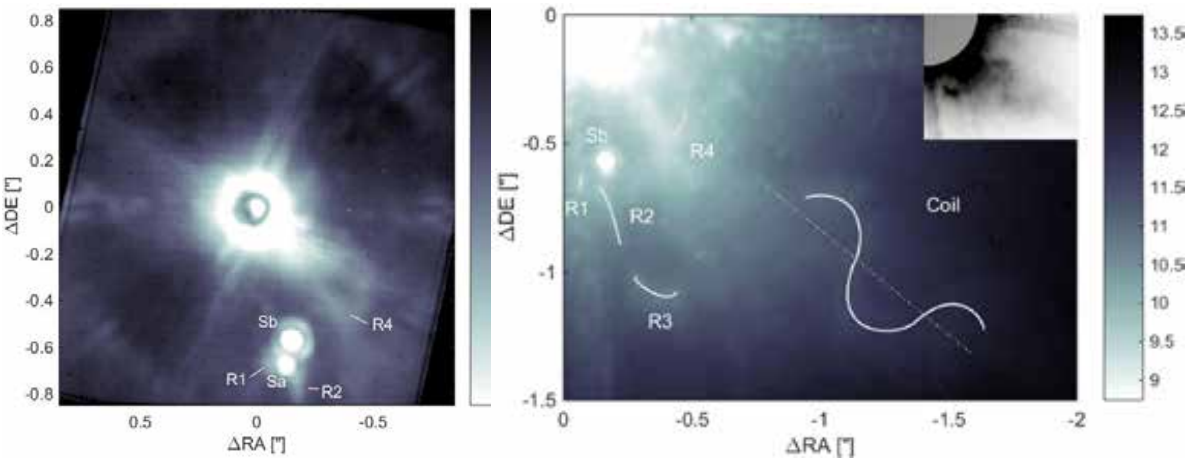
FIGURE 6.8 – Paramètres I^{p} et Q_ϕ de RX J1615.3−3255 reconstruit par la MLnS-D pour différents jeux d'hyperparamètres (cf. chapitre 4).

La figure 6.10 présente les reconstructions obtenues avec les méthodes séparables et non-séparables sans déconvolution. Les figures 6.10a et 6.10b présentent les reconstructions obtenues avec la Double Différence et la MnLS sur données *pré-traitées*. La différence entre les deux reconstructions est assez importante. En effet, sur la reconstruction avec la méthode que je propose, le contraste est bien plus élevé. Il est alors plus évident de différencier du fond, les structures de faible intensité. La figure 6.10c présente les reconstructions obtenues avec la MnLnS. Comme pour la reconstruction de RXJ 1615, le fond est plus brillant, ce qui encore une fois peut être lié à polarisation instrumentale.

La figure 6.11 présente un agrandissement des cartes contenues dans la figures 6.10 dans une région de 4×4 secondes d'arcs autour du centre de l'étoile T TAU N. Cette région est choisie pour coïncider avec les observations de l'état-de-l'art présentées sur la figure 6.9. Sur la reconstruction de la Double Différence sont tracés les marqueurs correspondants. On retrouve les structures R1, R2, R3, R4 et la structure appelée *coil* («ressort »en anglais), présentes dans la figures 6.9b. La reconstruction permet de mieux observer ces structures et d'en révéler de nombreuses nouvelles, telles que des ombres ou d'autres structures de types *coil*.

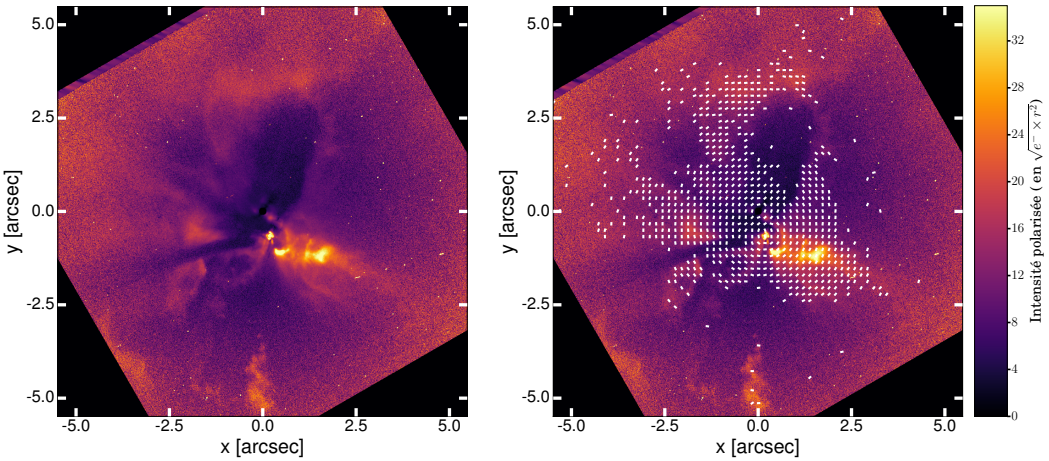
Dans la figure entière 6.10, la différence de qualité entre les méthodes est flagrante, tandis que pour la figure agrandie 6.11, la différence est moins visible, si ce n'est que l'influence des pixels morts est de moindre mesure avec la MnLS par rapport à la reconstruction faite avec la Double Différence.

La différence de contraste entre les reconstructions dans 6.10 peut s'expliquer par l'estimation jointe de I^{u}, I^{p} et θ avec contrainte de positivité sur I^{u} et I^{p} lors de la reconstruction par la MnLS. En effet, pour la méthode de la Double Différence cette contrainte n'est pas imposée, et $I^{\mathrm{p}} \geq 0$ uniquement du fait de sa définition par rapport aux paramètres de Stokes Q et U, c'est-à-dire pour tout $n \in \{1, \dots N\}$, $I^{\mathrm{p}}_n = \sqrt{\mathrm{Q}^2_n + \mathrm{U}^2_n} \geq 0$. Pour la MnLnS, une contrainte de positivité est imposée sur I^{u}, mais de la même façon que pour la Double Différence, la positivité de I^{p} n'est assurée que par définition. La MnLS permet donc à très faible flux (*i.e.* ≈ 0) et à grande distance du centre de contraindre le flux à 0 lors de l'estimation hiérarchique de θ, tandis que l'estimation en Q et U ne le permet pas.

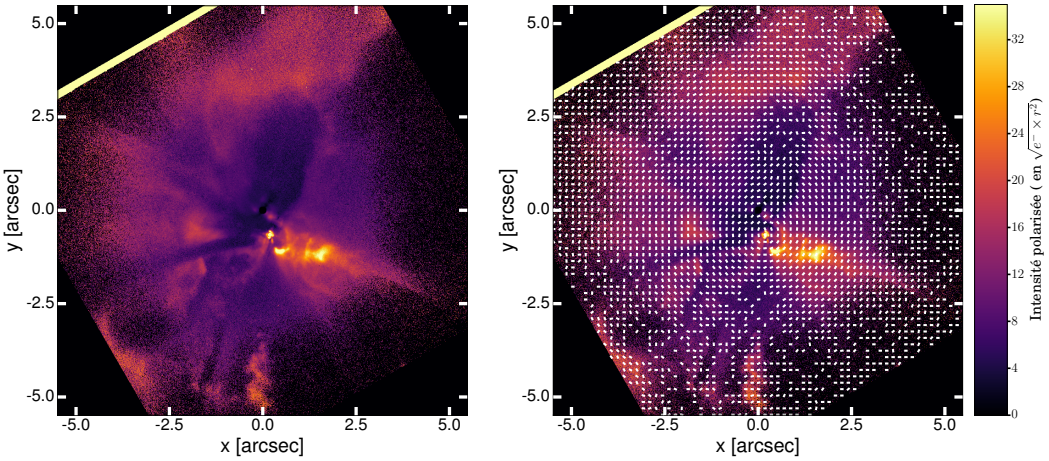


(a) (b)

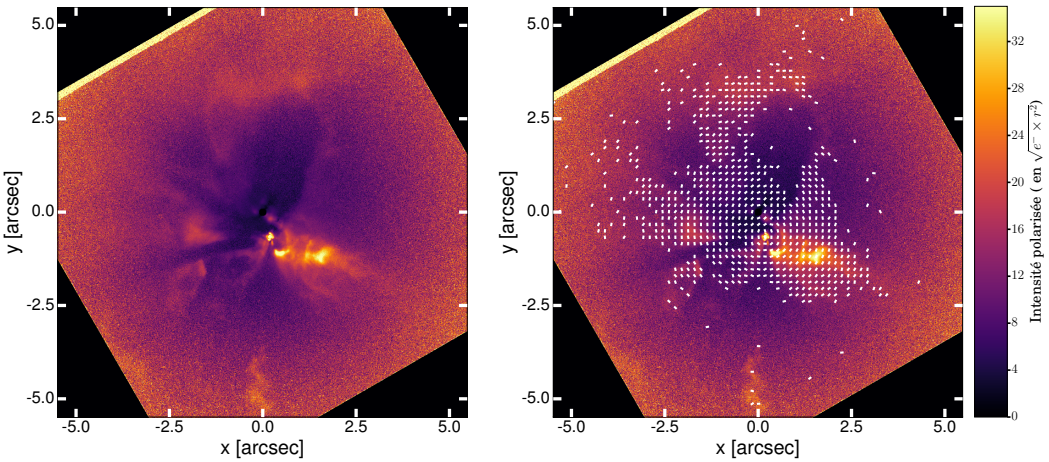
Figure 6.9 – Image coronographique de T Tauri N, pour la figure 6.9a obtenue avec l'IFS en bande H, pour la figure 6.9b avec IRDIS en mode imagerie en bande J (source : [Kasper et al., 2016]).



(a) Reconstruction obtenue avec la méthode de la Double Différence (cf. section 1.1.5).

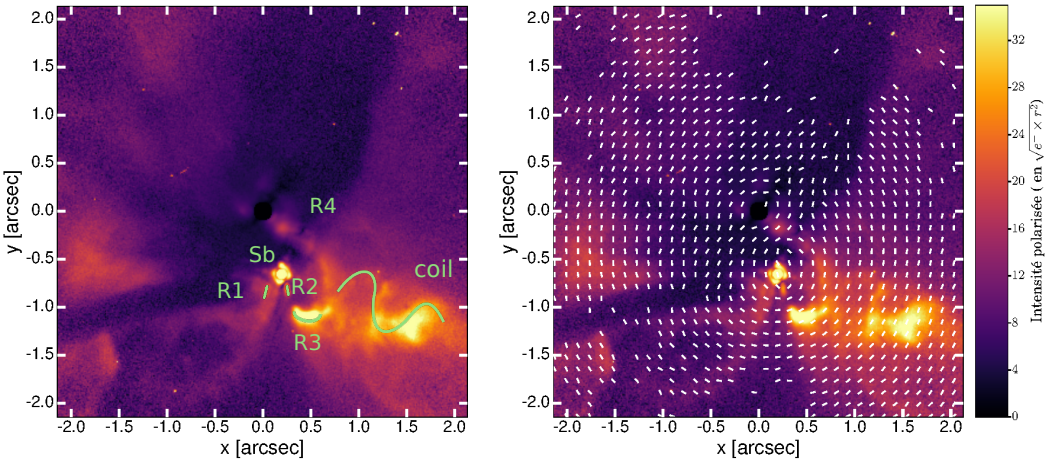


(b) Reconstruction obtenue avec la Méthode non-Linéaire Séparable (MnLS) (cf. section 2.1).

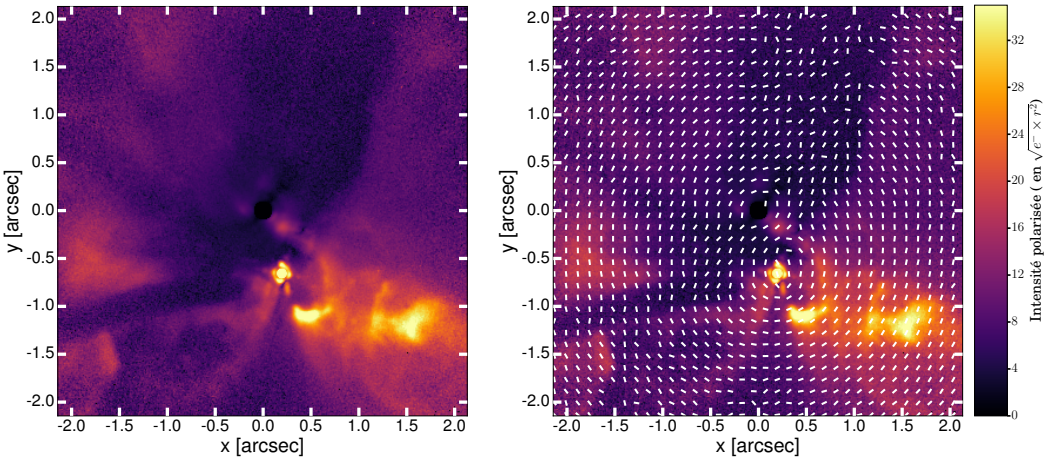


(c) Reconstruction obtenue avec la Méthode non-Linéaire non-Séparable (MnLnS) sans déconvolution (cf. section 3.2). La reconstruction est régularisée avec a priori de lissage à bords francs. Elle est obtenue pour les hyperparamètres $\mu = 10^{0,1}$, $\lambda_{I^u} = 10^2$, et $\lambda_{I^p} = 10^{2,5}$.

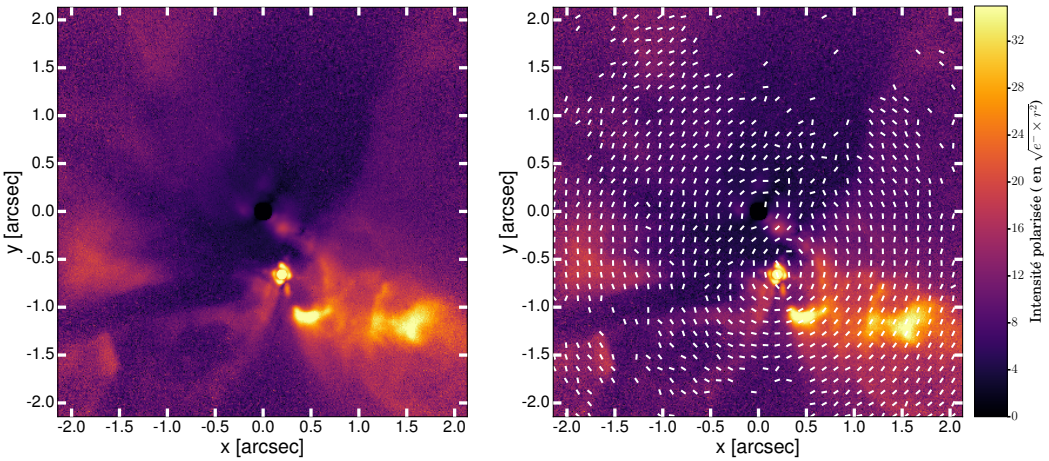
FIGURE 6.10 – Reconstruction de l'environnement circumstellaire autour de l'étoile T Tauri N, sans déconvolution.



(a) Reconstruction obtenue avec la méthode de la Double Différence (cf. section 1.1.5).



(b) Reconstruction obtenue avec la Méthode non-Linéaire Séparable (MnLS) (cf. section 2.1).

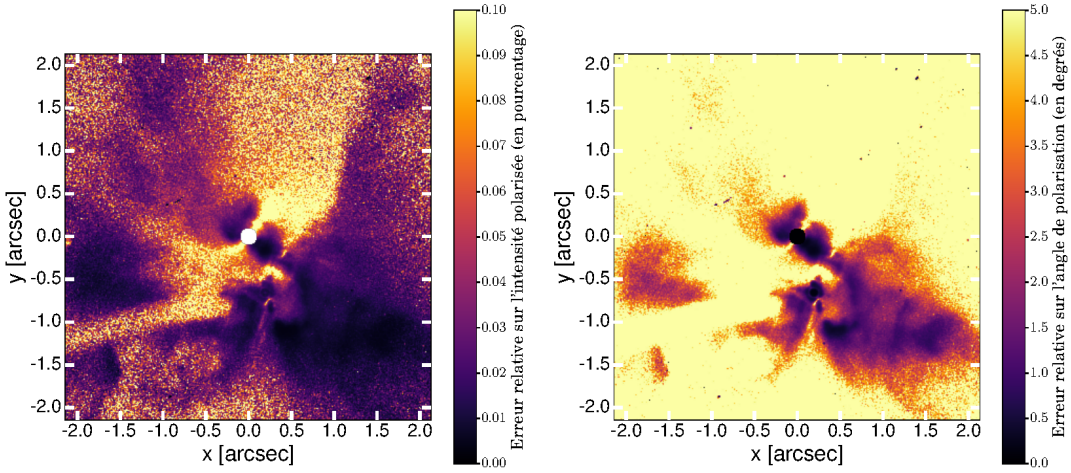


(c) Reconstruction obtenue avec la Méthode non-Linéaire non-Séparable (MnLnS) sans déconvolution (cf. section 3.2). La reconstruction est régularisée avec a priori de lissage à bords francs. Elle est obtenue pour les hyperparamètres $\mu = 10^{0,1}$, $\lambda_{I^{\mathrm{u}}} = 10^2$, et $\lambda_{I^{\mathrm{p}}} = 10^{2,5}$.

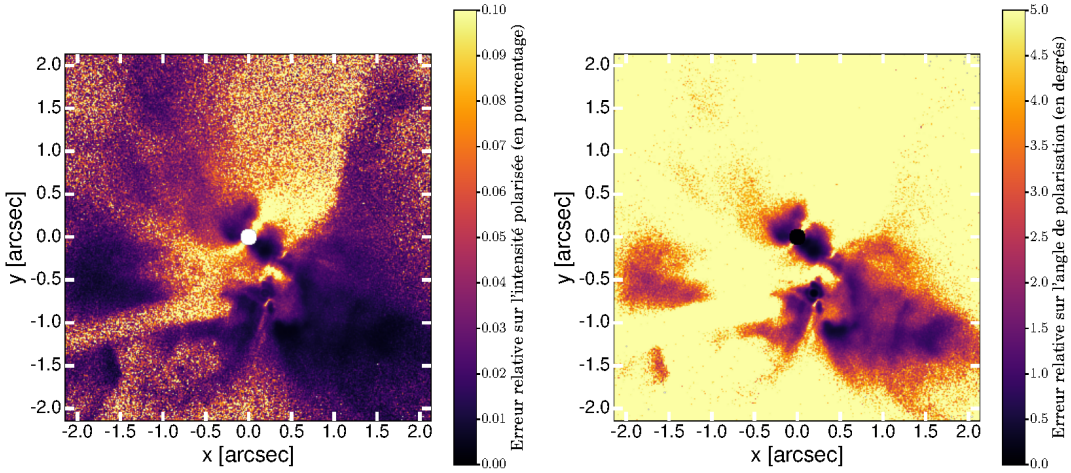
FIGURE 6.11 – Agrandissement de la reconstruction de l'environnement circumstellaire autour de l'étoile T Tauri, sans déconvolution, entre −2 et 2 secondes d'arc.

La figure 6.12 représente les cartes d'erreurs minimales faites en chaque pixel pour les méthodes séparables de la Double Différence et de la MnLS, sur $\widehat{I^{\mathrm{p}}}$ et $\widehat{\theta}$. On remarque que sur les structures très brillantes et observables par imagerie directe sans polarimétrie, l'erreur d'estimation de l'intensité polarisée et de l'angle est très faible, c'est-à-dire de moins de 0.01% pour $\widehat{I^{\mathrm{p}}}$ et moins de 1° pour θ sans correction de la polarisation instrumentale. On voit cependant que l'erreur minimale augmente dès que la brillance de I^{p} diminue et donc que le SNR diminue.

La difficulté principale des reconstructions à partir de ce jeu de données vient de sa taille. De fait, la région d'intérêt dans les données est ici découpée à 900×900 pixels, afin d'éviter les bords du détecteur, ce qui résulte en jeu de donnée de taille $900 \times 1800 \times 64$ où 64 est le nombre d'acquisitions. De ce fait, d'une part la résolution avec la MnLS est assez lente car pour chaque pixel $n \in \{1, \dots N\}$ il est nécessaire de faire plusieurs inversions matricielles, lors de la recherche de θ_n. Il est cependant possible de paralléliser facilement la méthode et la qualité des résultats, par rapport à la méthode de la Double Différence, justifie l'utilisation d'une telle méthode.



(a) Erreur relative de reconstruction du paramètre I^{p} et erreur angulaire du paramètre θ pour la double différ



(b) Erreur relative de reconstruction du paramètre I^{p} et erreur angulaire du paramètre θ pour la MnLS.

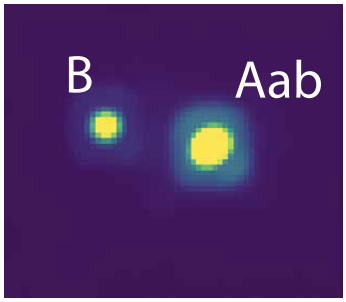
FIGURE 6.12 – Erreur relative de reconstruction de l'environnement circumstellaire du système T TAU avec la Double Différence et la Méthode non-Linéaire Séparable (cf. chapitre 2).

D'autre part, la durée nécessaire à la convergence de l'algorithme VMLM-B, dans le cas de la MnLnS avec régularisation par TV-h, complexifie la recherche des hyperparamètres dans le temps imparti. En effet, le réglage automatique des paramètres à l'aide du critère SURE prend alors plusieurs jours.

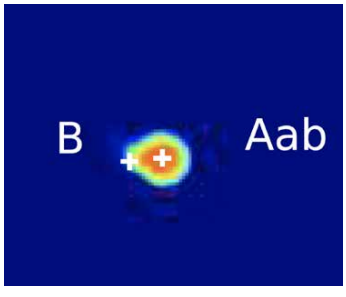
La reconstruction en polarimétrie du système T Tauri avec mes méthodes sans déconvolution, en particulier par la MnLS, permet donc de reconstruire des structures à plus grande séparation angulaire dans le champ, avec un meilleur contraste que par la Double Différence, où celles-ci se fondent dans le fond. Cette reconstruction sur un très grand champ permet de s'interroger sur le choix de la paramétrisation du modèle direct en montrant que l'estimation jointe de I^{p}, I^{u} et θ avec contrainte de positivité sur I^{u} et I^{p} donne de meilleurs résultats à grande distance angulaire. La correction de la polarisation instrumentale et la déconvolution d'une telle cible permettra d'augmenter encore le contraste des reconstructions et une meilleure résolution des structures.

6.2.3 Étude d'un nouvel objet

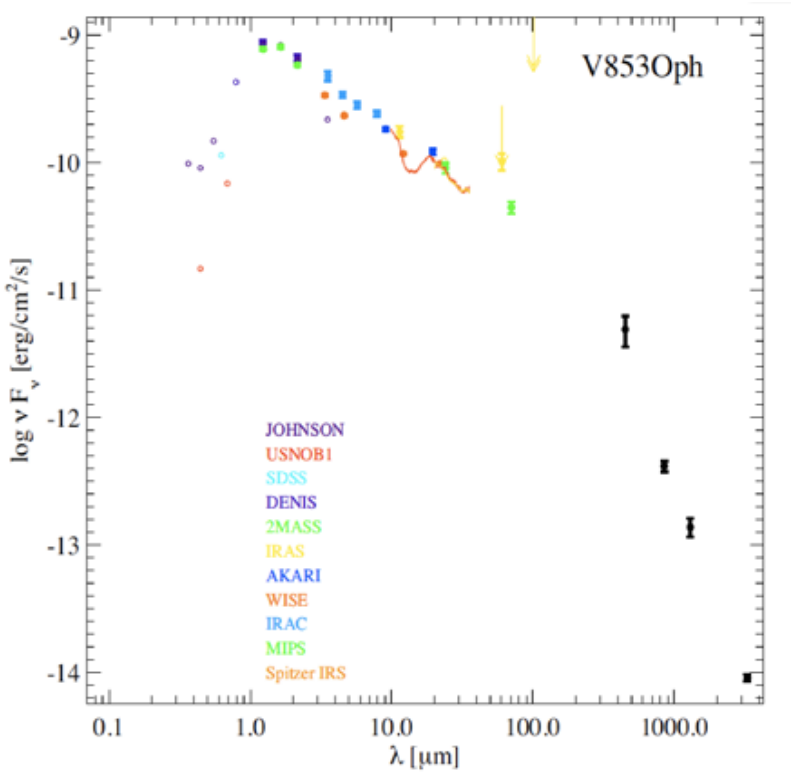
Dans cette section je présente la première observation en polarimétrie d'un disque protoplanétaire autour d'une binaire associée à un troisième compagnon, visbles sur la figure 6.13a, où la PSF allongée témoigne de la présence de deux étoiles convoluées par la PSF. Une perspective lors de mon étude de l'environnement est de déconvoluer ces PSF afin de résoudre les deux étoiles Aa et Ab. Je présente ici les résultats de reconstruction du disque sans et avec déconvolution, à partir des données DPI de l'instrument IRDIS en bande H.



(a) PSF du système d'étoiles obtenues en bande H avec l'instrument IRDIS



(b) Observation du système et du disque avec ALMA (source : [Cox et al., 2017]).



(c) SED de l'environnement.

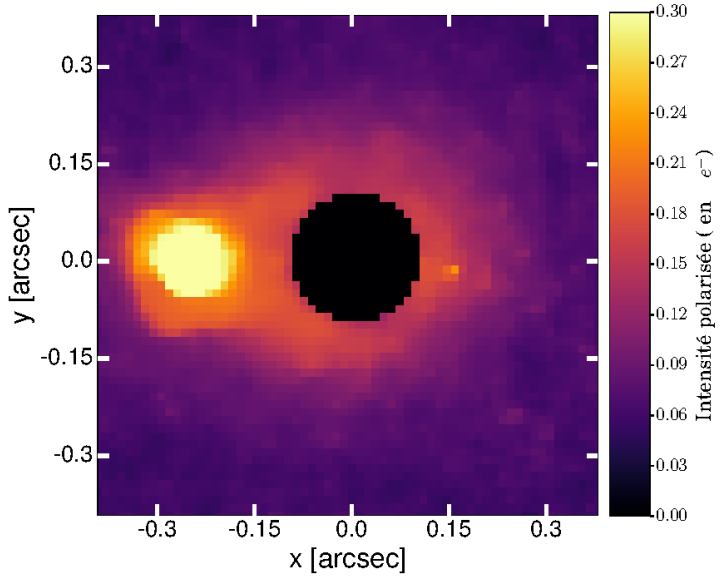
FIGURE 6.13 – PSF, SED et observation avec ALMA de la cible.

6.2.3.1 Sans déconvolution

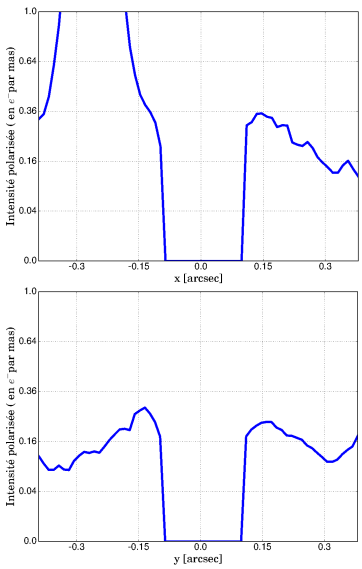
La figure 6.15 présente les reconstructions obtenues avec les méthodes séparables et non-séparables sans déconvolution. Les figures 6.15a et 6.15b présentent les reconstructions obtenues avec la Double Différence et la MnLS sur données *pré-traitées*. Afin de pouvoir observer le disque, la dynamique des images est volontairement saturée au-dessus des intensités maximales du disque. On peut voir que le disque correspond relativement à un *«donut »*, les angles de polarisation coïncident avec une telle morphologie. Dans la reconstruction par la MnLS et la MnLnS, on peut observer une structure plus brillante à droite du disque qui n'est pas observable dans la reconstruction par la double différence. Cette structure est due à un pixel «mort-vivant »(*i.e* qui se comporte tantôt normalement, tantôt comme un pixel mort), non-sélectionné lors de la calibration.

La figure 6.16 présente les cartes d'erreur de reconstruction relative pour l'intensité polarisée $\widehat{I^{\mathrm{P}}}$ et absolue pour l'angle $\widehat{\theta}$, calculées de la manière présentée dans la section 5.2, pour la Double Différence et la MnLS. On voit par ailleurs qu'une toute petite structure en bas à droit apparaît (entourée en vert sur la figure 6.16b). Cette structure provient également d'un ensemble de pixel « morts-vivants »non détectés et, dans le cas des méthodes séparables, interpolés avec leurs voisins. Les cartes d'erreurs semblent elles aussi montrer un aplatissement du disque pour la reconstruction par la MnLS.

Dans la reconstruction par la MnLnS 6.15c, le fond est très brillant et la multiplication par r^2 brouille la différence entre le disque et le fond. On trace donc sur la figure 6.14a, l'intensité polarisée et des coupes horizontales et verticales du disque pour cette reconstruction, sans la multiplication par r^2. On remarque que le disque est moins circulaire car aplati dans la région du bas. Cela peut venir à nouveau de la polarisation instrumentale ou d'une légère inclinaison du disque.

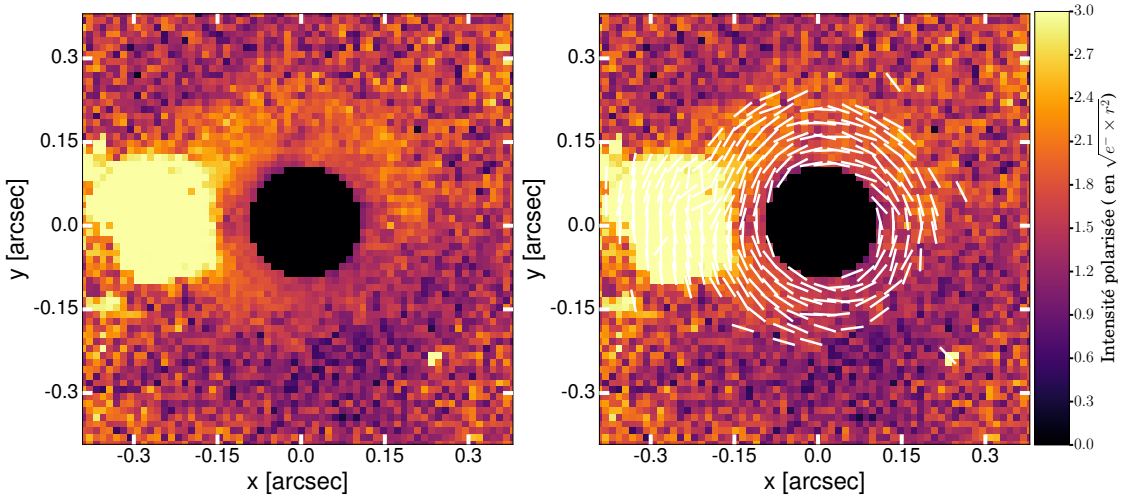


(a) $\widehat{I^{\mathrm{P}}}$ obtenu avec la MnLnS sans déconvolution (cf. section 3.2), sans multiplication par r^2, pour les hyperparamètres $\mu = 10^{0,1}$, $\lambda_{I^{\mathrm{U}}} = 10^2$ et $\lambda_{I^{\mathrm{P}}} = 10^{2,5}$.

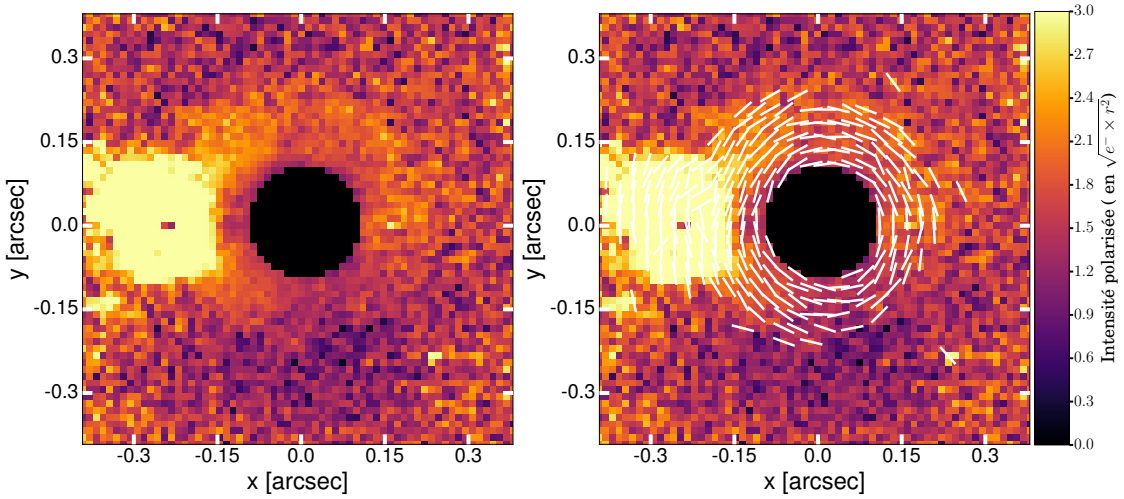


(b) Coupe horizontale et verticale de la figure 6.14a.

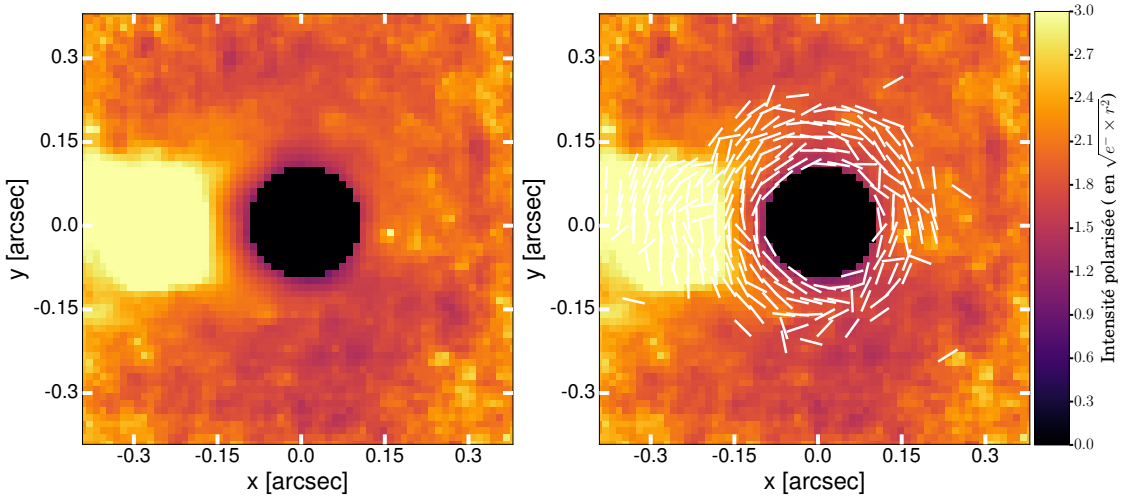
FIGURE 6.14 – Profil de l'intensité polarisée obtenue avec la MnLnS, sans multiplication par r^2.



(a) Reconstruction obtenue avec la méthode de la Double Différence (cf. section 1.1.5).

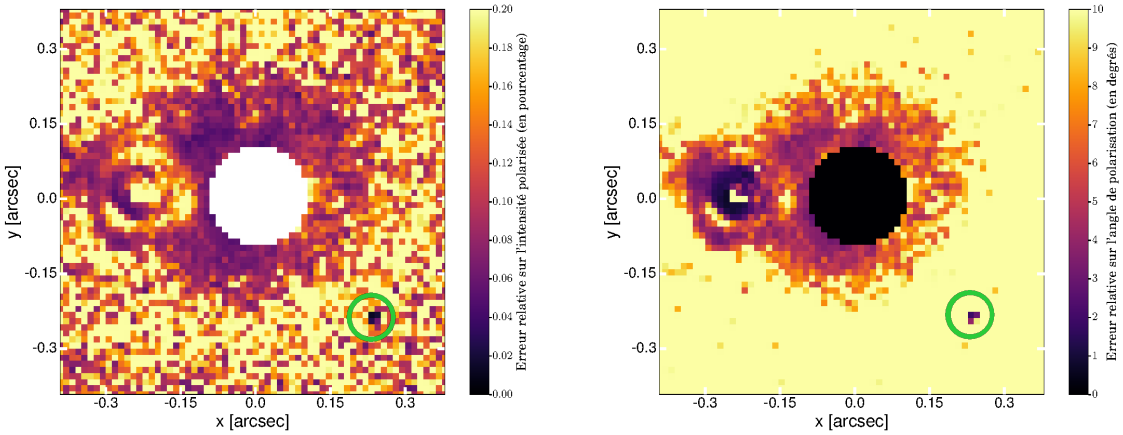


(b) Reconstruction obtenue avec la Méthode non-Linéaire Séparable (MnLS) (cf. section 2.1).

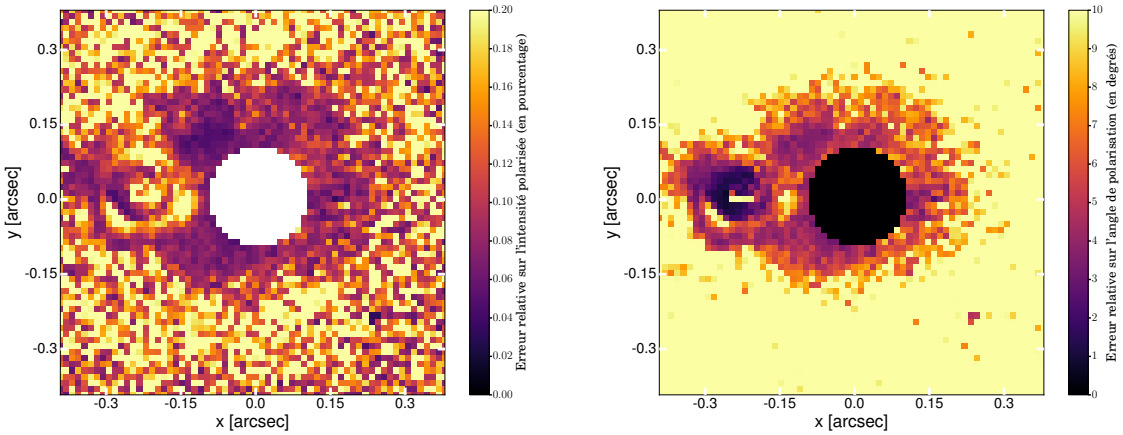


(c) Reconstruction obtenue avec la Méthode non-Linéaire non-Séparable (MnLnS) sans déconvolution (cf. section 3.2). La reconstruction est régularisée avec *a priori* de lissage à bords francs. Elle est obtenue pour les hyperparamètres $\mu = 10^{0,1}$, $\lambda_{I^{\mathrm{u}}} = 10^2$, et $\lambda_{I^{\mathrm{p}}} = 10^{2,5}$.

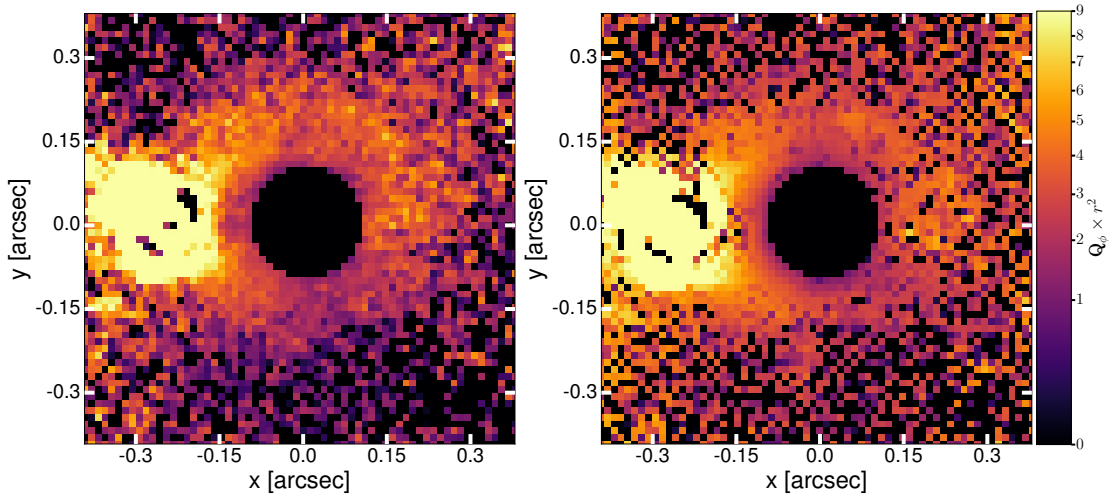
FIGURE 6.15 – Reconstruction du disque sans déconvolution.



(a) Erreur relative de reconstruction du paramètre I^{p} et erreur angulaire du paramètre θ pour la double différence.



(b) Erreur relative de reconstruction du paramètre I^{p} et erreur angulaire du paramètre θ pour la MnLS.



(c) Intensité polarisée projetée Q_ϕ obtenue par la MnLS (à gauche) dont l'intensité et l'angle sont présentés sur la figure 6.15b et par la MnLnS (à droite) dont l'intensité et l'angle sont présentés sur la figure 6.15c.

FIGURE 6.16 – Erreur relative de reconstruction du disque avec la Double Différence et la Méthode non-Linéaire Séparable (cf. chapitre 2). Projection radiale des reconstructions pour la MnLS et la MnLnS.

La figure 6.16c représente l'intensité projetée orthogonalement Q_ϕ pour la MnLS et la MnLnS, multipliée par la distance au centre r^2. On voit que pour la projection de la MnLnS (à droite), le fait que l'angle ne soit pas régularisé bruite fortement la projection. Cependant pour la reconstruction par la MnLS (à gauche), on voit que le contraste est amélioré par rapport aux cartes de $\widehat{I^{\mathrm{p}}}$. Le disque est à nouveau légèrement aplatit en bas, ce qui pourrait donc indiquer que le disque est légèrement incliné. Une reconstruction sans polarisation instrumentale pourra permettre une meilleure estimation de la morphologie.

6.2.3.2 Avec déconvolution

La figure 6.17 présente les résultats de déconvolution pour deux jeux d'hyperparamètres différents, menant d'une part à une solution sous-régularisée, présentée sur la figure 6.17a, et d'autre par à une solution sur-régularisée, présentée sur la figure 6.17b.

Pour estimer la PSF, à partir du jeu de donnée de calibration duquel est extraite l'image 6.13a, j'ai appliqué la méthode présentée dans la section 6.1.2, sur l'étoile B. Afin que ma PSF ne soit pas contaminée par la PSF de la binaire Aab, lors de la reconstruction de l'image de PSF, j'ai masqué la binaire par un masque rectangulaire, en mettant la covariance des pixels à l'infini et donc les poids correspondants à 0.

Dans le jeu de données objet, la PSF de l'étoile B est saturée, ce qui implique qu'en ces pixels, le modèle direct incluant la convolution est faux. Afin d'éviter les artefacts de déconvolution lors de la reconstruction, on met donc les poids $\mathbf{W}_m \in \mathbb{R}^K$ des pixels saturés à 0, pour toutes les acquisitions $k \in \{1, \dots K\}$

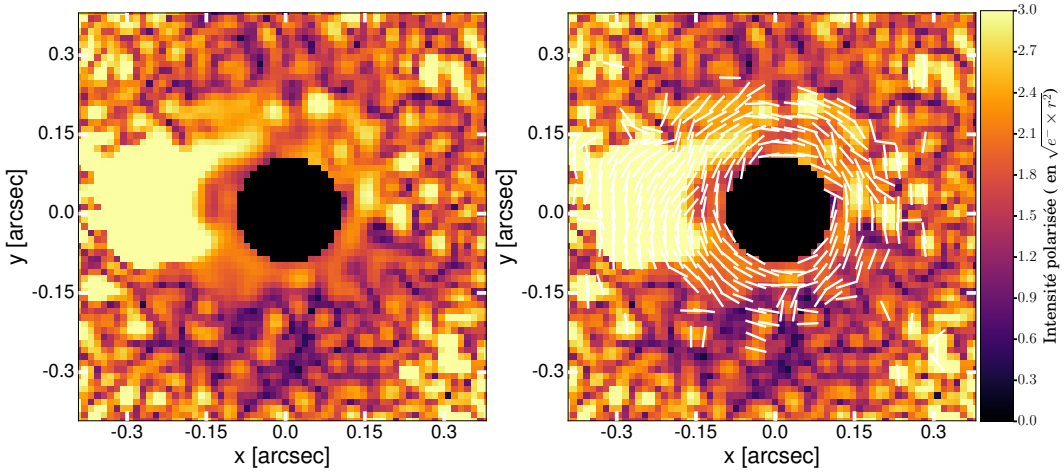
Dans la figure 6.17a, sont présent de nombreux artefacts qui semblent « engloutir »les bords du disque. Malgré le fait que cette reconstruction soit sous-régularisée, le disque semble tout de même sur-régularisé. Le même effet était visible sur RXJ 1615 au bord du coronographe, même dans les cas sous-régularisé (cf. figure 6.8b). Cependant prendre un paramètre plus bas ici entraine une perte totale de l'information dans les artefacts de déconvolution.

Dans la figure 6.17b, l'hyperparamètre μ semble avoir été choisi trop bas, car l'effet « cartoon »est clairement visible, mais la séparation entre le fond et le disque est plus marquée.

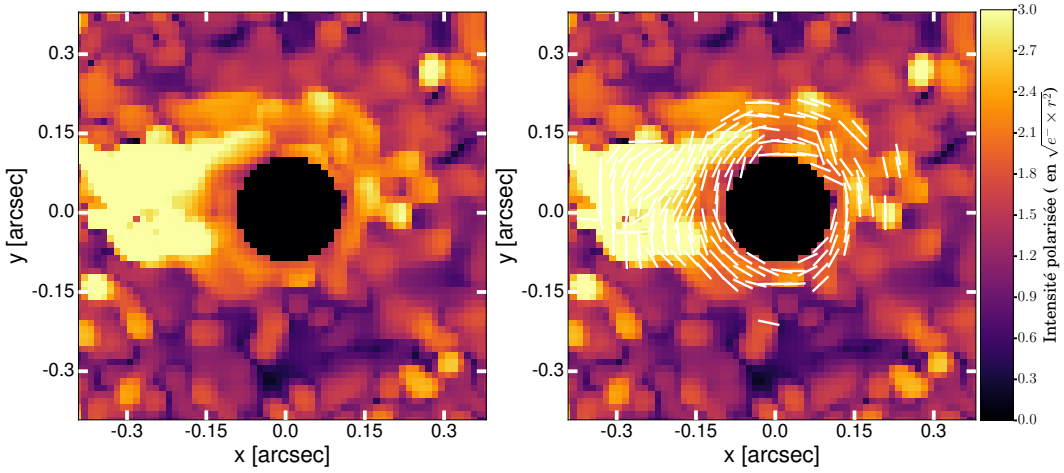
On remarque par ailleurs que dans les deux cas, bien que masquée, comme les ailes étaient toujours présentes, la PSF de l'étoile B a été reconstruite lors de la déconvolution.

La figure 6.17c représente la carte de l'intensité polarisée projetée orthogonalement Q_ϕ obtenue à partir de la MLnS-D pour les deux ensembles d'hyperparamètre. On voit à nouveau que la projection permet d'avoir un meilleur contraste entre le fond et le disque, ce qui implique que la correction de la polarisation instrumentale lors de la calibration avec IRDAP semble majeure, pour diminuer l'erreur d'estimation.

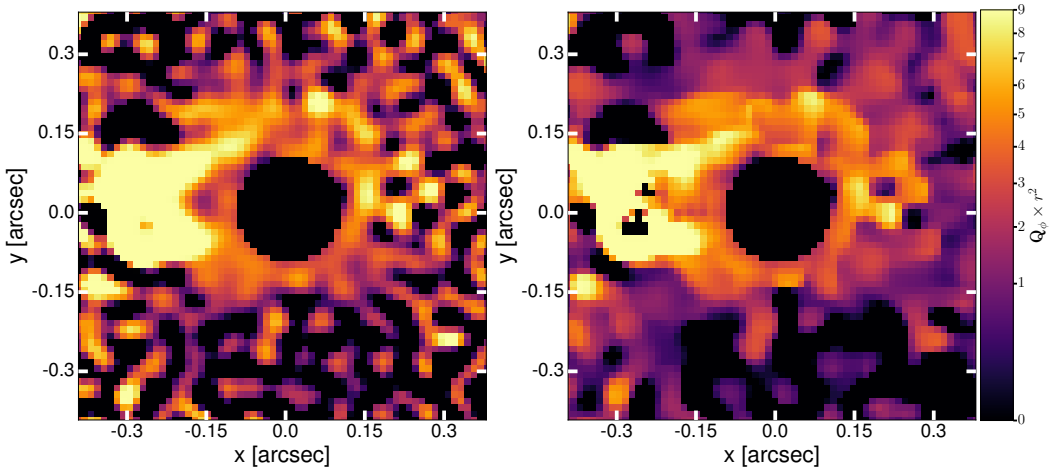
L'étude de cet objet est actuellement en cours et les raffinements nécessaires seront mis en œuvre pour arriver à une meilleure qualité de déconvolution, tels qu'une meilleure calibration et un meilleur choix de modèle, d'hyperparamètres et de régularisation. L'avantage de ce jeu de donnée et qu'il est de taille très réduite, en effet la zone d'intérêt peut se limiter à 64 pixels de large, ce qui permet de pouvoir facilement faire varier les hyperparamètres afin de trouver un compromis plus satisfaisant.



(a) Reconstruction obtenue pour les hyperparamètres $\mu = 10^{-2}$, $\lambda_{\mathrm{I}} = 10^3$, et $\lambda_{\mathrm{Q,U}} = 10^4$.



(b) Reconstruction obtenue pour les hyperparamètres $\mu = 10^{-4}$, $\lambda_{\mathrm{I}} = 10s$, et $\lambda_{\mathrm{Q,U}} = 10^2$.



(c) Intensité projetée Q_ϕ pour les reconstructions par la MnLnS-D dont les intensités polarisées et angles sont présentés respectivement sur les figures (a) (à gauche) et (b) (à droite).

FIGURE 6.17 – Reconstruction du disque autour de l'étoile avec la Méthode non-Séparable Linéaire avec Déconvolution (MLnS-D) (cf. chapitre 4). La reconstruction est régularisée avec *a prioiri* de lissage à bord franc (TV-h).

La difficulté à laquelle j'ai été confrontée avec ce jeu de données et qu'il comporte du *dithering*, c'est-à-dire un décalage volontaire du détecteur, d'un nombre de pixels défini, au cours de la séquence. Le centre de rotation des images bouge de plus d'une image à l'autre et cela peut induire des erreurs d'estimations. La PSF saturée de l'étoile B peut permettre, par ajustement d'une gaussienne, de connaitre le décentrage de manière subpixelique pour chaque acquisition. Cependant, la saturation perturbe la déconvolution.

Ce disque protoplanétaire sera l'objet d'un article actuellement en cours de préparation, dans lequel seront présentées des cartes reconstruites sans et avec déconvolution. La morphologie du disque sera alors comparée à une simulation hydrodynamique.

Conclusion du chapitre

Dans cette section j'ai présenté, dans un premier temps, comment j'estime la calibration du détecteur et de l'instrument, en utilisant l'approche inverse, et comment je calibre et pré-traite mes données.

Dans un second temps, j'ai présenté les reconstructions par les méthodes développées dans cette thèse, de trois cibles astrophysiques dont deux avec déconvolution. Il ressort de tels résultats que mes méthodes de reconstruction permettent d'atteindre un meilleur contraste, surtout lors de la déconvolution où dans le cas de la MnLS à une grande distance angulaire du centre. Mes méthodes permettent aussi de minimiser la polarisation instrumentale dans le cas de l'utilisation de la déconvolution.

La reconstruction des environnements circumstellaires est cependant limitée par la qualité des données, par leur calibration et, plus particulièrement, de la polarisation instrumentale, qui résulte en signal parasite sur les mesures et une perte des structures faibles. La prise en compte de cette calibration à l'aide de l'outil IRDAP [van Holstein et al., 2020] sera appliqué très prochainement afin de pouvoir comparer plus efficacement les méthodes et de pouvoir bénéficier entièrement des performances de mes méthodes, dont l'efficacité a été prouvée sur données simulées, sans polarisation instrumentale.

Chapitre 7

Conclusion

7.1 Contributions

Mon travail de thèse a porté sur le développement de la première méthode de reconstruction d'images d'environnements circumstellaires, basée sur l'approche inverse, incluant la déconvolution par la PSF, à partir des données polarimétriques de l'instrument ESO-VLT/SPHERE IRDIS. Dans cette thèse, j'ai montré, à travers différentes contributions, que l'utilisation d'un modèle complet des données, prenant en compte le pré-traitement, la covariance des données et la convolution par la PSF, permet une reconstruction de bien meilleure qualité que les reconstructions de l'état-de-l'art, permettant à la fois de mieux résoudre les structures brillantes et de mieux distinguer les structures de faible intensité.

Ma première contribution a été de formuler le modèle direct complet des données de l'instrument, adaptable selon le niveau de pré-traitement des données (*i.e. pré-traitées* ou *calibrées*) et selon le niveau de reconstruction souhaité (*i.e.* sans ou avec déconvolution). J'ai étudié divers choix de paramétrisation du modèle direct, c'est-à-dire non-linéaire par les intensités non-polarisée, polarisée et l'angle de polarisation linéaire, qui sont les observables intéressants en astrophysique, linéaire par les paramètres de Stokes, qui sont les paramètres utilisés dans l'état-de-l'art, ou par un mélange intensité/paramètres de Stokes, permettant un compromis entre les deux paramétrisations. À partir de ce modèle direct, j'ai créé un ensemble de données synthétiques qui correspondent à différents niveaux de difficulté d'observation astrophysique, afin de comparer quantitativement la qualité de mes reconstructions.

Ma seconde contribution a alors été de présenter différents critères d'estimation des paramètres d'intérêt, par une formulation de type *«approche inverse »*, basés sur la vraisemblance régularisée des données. La minimisation de ces critères m'a permis d'obtenir des résultats dont la qualité est au moins aussi bonne, et jusqu'à deux fois meilleure, que les reconstructions de l'état-de-l'art, dans le cas de la déconvolution. J'ai présenté d'une part des méthodes séparables (MnLS et MLS), applicables sur données *pré-traitées*, conduisant à une qualité de reconstruction au moins équivalente à celle des méthodes de l'état-de-l'art et bien meilleure dans le cas de données manquantes. D'autre part j'ai présenté des méthodes non-séparables (MLnS, MnLNS, MLnS-D), applicable sur données *calibrées* permettant, à partir de différentes régularisations différentiables et non-différentiables, ainsi qu'une contrainte de positivité pouvant s'écrire sous forme de contrainte épigraphique ou de dynamique, d'obtenir des résultats de bien meilleure qualité que les méthodes de l'état-de-l'art appliquées sur données *pré-traitées* et avec déconvolution *a posteriori*, mais plus coûteuses en temps de calcul selon le choix des

régularisations.

J'ai conclu que la qualité de reconstruction augmentait avec la complexité du modèle utilisé, les méthodes séparables étant les plus simples mais permettant une moins bonne estimation que les méthodes non-séparables, en particulier lors de l'inclusion de la convolution, mais plus coûteuses en temps de calcul. L'apport des régularisations vis-à-vis de leurs complexité m'a permis de conclure que la régularisation différentiable par *a priori* de lissage et préservation des bords (TV-h) était un bon compromis vis-à-vis de la variation totale (TV) et de la norme de Shatten sur la hessienne. L'apport de la contrainte épigraphique est surtout visible pour la reconstruction du paramètre I^{u}, permettant une meilleure régularisation de la carte. Cependant lors de la déconvolution d'objets brillants, elle induit des artefacts dans les paramètres I^{p} et θ, qui consistent en notre objectif principal. De ce fait, pour plus de simplicité, une simple régularisation par TV-h sans contrainte épigraphique a été utilisée sur donnée réelle.

Dans le cas séparable, j'ai proposé une estimation de la covariance en chaque pixel des cartes reconstruites, permettant de juger la qualité de reconstruction. J'ai également proposé une approche de réglage automatique des hyperparamètres de régularisation à partir du critère SURE et j'ai comparé son efficacité vis-à-vis d'un critère sur l'Erreur Quadratique Moyenne des paramètres d'intérêts, qui n'est pas accessible sur données astrophysiques. J'ai pu observer qu'à faible SNR, ce qui est le cas des données réelles, l'estimation des hyperparamètres à partir du critère SURE menait à une sur-régularisation des reconstructions. Ceci permet cependant de trouver un point de départ pour la recherche manuelle des hyperparamètres, qui est très satisfaisante.

Lors de l'application sur données astrophysiques, j'ai observé que la prise en compte du modèle avec convolution, régularisé par TV-h permettait une meilleure résolution de la morphologie des environnements vis-à-vis des méthodes de l'état-de-l'art suivies d'une déconvolution *a posteriori*. Ma méthode permet en outre de s'affranchir des artefacts liés aux mauvais pixels. J'ai montré par ailleurs que la résolution angulaire des détails à partir du modèle complet se faisait à différentes échelles de régularisation, et qu'il était important de procéder à plusieurs reconstructions pour différents jeux d'hyperparamètres afin d'avoir accès à toute l'information dans les images. En effet, une sur-régularisation permet de mettre en avant les structures de faible intensité, mais avec une légère perte de résolution, tandis qu'une sous-régularisation permet une meilleure résolution des structures brillantes mais permet de détecter des structures de faible intensité avec une moins bonne visibilité.

L'application sur données astrophysiques sans et avec déconvolution a cependant fait apparaitre un problème de brillance du fond, dû à la polarisation instrumentale, qui ressort alors plus particulièrement dans les cas sous-régularisés.

Ma troisième contribution a été d'adapter différents algorithmes pour l'optimisation des paramètres d'intérêts, selon la complexité et la différentiabilité des critères, VMLM et VMLM-B dans le cas différentiable et avec contrainte de dynamique, VMFB et VMCV dans le cas non-différentiable. En particulier, j'ai proposé d'ajouter une étape de backtracking dans les algorithmes proximaux à métrique variable VMFB et VMCV, afin d'éviter le calcul de la constante de Lipschitz, qui peut être compliqué dans le cas astrophysique du fait du mauvais conditionnement des opérateurs linéaires. J'ai par ailleurs remarqué que le backtracking permettait d'améliorer le temps de convergence dans le cas de l'utilisation de la régularisation avec *a priori* de lissage et préservation des bords et de la contrainte épigraphique, par rapport à l'approche basée sur la constante de Lipschitz. L'avantage des algorithmes proximaux est leur possibilité d'adaptation à toutes formes de critères, différentiables et non-différentiables. En revanche,

Table 7.1 – Tableau de recensement des méthodes développées dans cette thèse. Dans les noms des méthodes, L signifie Linéaire, nL non-Linéaire, S Séparable, nS non-Séparable, et D déconvolution.

Méthode	Paramètres	Régularisation	Contrainte	Résolution
MnLS	$(I^{\mathrm{u}}, I^{\mathrm{p}}, \theta)$	aucune	Positivité I^{u} et I^{p}	Inversion directe
MLS	(I, Q, U)	aucune	aucune	Inversion directe
MLnS	(I, Q, U)	TV-h sur I, Q et U	aucune	VMLM
	(I, Q, U)	Tikhonov sur I, Q et U	aucune	VMLM
MnLnS	$(I^{\mathrm{u}}$, Q, U)	TV-h sur I^{u} et I^{p}	Positivité sur I^{u}	VMLM-B
	$(I^{\mathrm{u}}$, Q, U)	Tikhonov sur I^{u} et I^{p}	Positivité sur I^{u}	VMLM-B
MLnS-D	(I, Q, U)	Shatten sur I, Q et U	Épigraphique entre I, Q et U	VMCVwB
	(I, Q, U)	TV sur I, Q et U	Épigraphique entre I, Q et U	VMCVwB
	(I, Q, U)	TV-h sur I, Q et U	Épigraphique entre I, Q et U	VMFBwB
	(I, Q, U)	TV-h sur I, Q et U	aucune	VMLM

pour l'application à mon problème de reconstruction, le temps de convergence de ces méthodes est supérieur à ceux de l'algorithme à métrique variable et à mémoire limité VMLM-B, ce qui est trop pénalisant pour le réglage manuel des hyperparamètres de régularisation.

C'est pourquoi l'algorithme VMLM pour l'estimation des paramètres du modèle complet, avec régularisation par TV-h sans contrainte épigraphique m'a paru la plus satisfaisante pour concilier qualité des résultats et temps de calcul.

Le tableau 7.1 recense les méthodes développées dans cette thèse et les classe par ordre de performances croissantes à partir des méthodes ayant des performances au moins égales à celles des méthodes de l'état-de-l'art. Ces performances concilient à la fois la qualité des reconstructions en terme d'Erreur Quadratique Moyenne des paramètres I^{p} et θ (dont l'intérêt est le plus grand en polarimétrie) dans les cas de faible polarisation du disque ($< 25\%$), ce qui correspond aux cas classiques d'observation en astrophysique, et le coût en temps des reconstructions.

7.2 Perspectives

Concernant l'aspect traitement des données, lorsqu'on voit l'apport de la déconvolution avec une régularisation à préservation de bord telle que TV-h, il est dommage qu'une telle reconstruction ne puisse bénéficier que difficilement de régularisations non-différentiables, telles que TV et Shatten, qui ont pourtant fait leurs preuves dans d'autres domaines. Deux perspectives pourraient être envisagées pour pallier cette limitation. La première serait d'utiliser tout comme pour TV-h, des approximations différentiables de ces régularisations, permettant ainsi la résolution des critères associés par la méthode VMLM-B. La seconde serait une optimisation numérique de la résolution des algorithmes proximaux, par parallélisation du code, résolution par bloc ou encore calcul GPU.

Par ailleurs, étant donné les apports majeurs de la MnLnS à la MLnS, il semble important d'envisager une méthode non-Linéaire pour la déconvolution sur les paramètres I^{u}, Q et U. Le raffinement de ce modèle pourrait par ailleurs se faire par l'étude de régularisations structurées sur les paramètres Q et U plutôt que sur I^{p} seul.

D'après les résultats de déconvolution sur données astrophysiques, vis-à-vis des différents

réglages des hyperparamètres, il serait être également intéressant d'étudier si une régularisation multi-échelles adaptée peut permettre un gain significatif.

Pour pallier ces difficultés de calibration du fond thermique, misent en avant par les méthodes non-séparables, il serait souhaitable d'inclure une étape de recalibration du fond, lors de l'estimation des paramètres d'intérêt, à l'aide d'une méthode alternée telle que présentée dans [Berdeu et al., 2020].

Un enjeu majeur de ma thèse est la déconvolution des données, ici faite avec une PSF paramétrée, dont les paramètres ont été estimés à partir de données de calibration. Une autre amélioration de la méthode serait l'utilisation de déconvolution myope ou aveugle, qui revient à estimer conjointement les paramètres et la PSF en utilisant une méthode alternée, à partir de la PSF paramétrée. Par ailleurs, des travaux d'estimation des PSF de l'instrument ESO-VLT SPHERE à l'aide de réseaux de neurones est actuellement en cours. L'utilisation de telles PSF estimées pourrait également être envisagée.

L'amélioration de la déconvolution pourrait également être faite par un choix plus judicieux de régularisation telle qu'une régularisation structurée sur les paramètres liés à l'angle de polarisation. Elle pourrait être également améliorée par une meilleure estimation des hyperparamètres. Plusieurs pistes sont alors envisageables. La première serait l'utilisation d'un critère SURE dans le domaine des paramètres d'intensités et d'angle. Cependant, une telle méthode est difficile car le modèle non-séparable des données en ces paramètres est à la fois non-linéaire et non-inversible. La seconde serait, dans le cas de la déconvolution, d'accélérer le critère SURE afin de pouvoir obtenir les meilleurs hyperparamètres à convergence, en utilisant par exemple des méthodes alternées telles que développées dans [Ammanouil et al., 2019].

Une dernière perspective serait d'étudier la borne de FDCR dans le cas non-séparable et régularisé, afin de pouvoir proposer des cartes d'erreurs sur les valeurs des pixels estimées et pouvoir comparer quantitativement les méthodes régularisées aux méthodes de l'état-de-l'art.

Du côté des résultats astrophysiques, la déconvolution par une régularisation structurée jointe à la calibration de la polarisation instrumentale par la méthode IRDAP développée dans[van Holstein et al., 2020], pourrait aussi permettre une amélioration de la reconstruction des structures de faible SNR en réduisant le fond dans les reconstructions.

La perspective actuelle est, à partir d'un grand nombre de jeux de données astrophysiques, où la polarisation instrumentale a été calibrée, de reconstruire les paramètres d'intérêt I^{p} et θ afin de pouvoir affiner la résolution de leurs morphologies et peut-être découvrir de nouvelles structures jusqu'à présent noyées dans le fond thermique ou dans les fuites stellaires.

La perspective ultime sera la reconstruction jointe des images de la lumière réfléchie par les environnements circumstellaires, polarisée et non-polarisé, par différence polarimétrique et différence angulaire. Une telle étude a été menée à partir de données DPI et ADI de l'instrument IRDIS, pour étudier la surface du disque autour du système AB Aurigae [Boccaletti et al., 2020]. La figure 7.1 montre la carte de lumière polarisée non-déconvoluée reconstruite à partir des données DPI de l'instrument IRDIS à partir des méthodes de l'état-de-l'art. L'utilisation jointe de la différentiation angulaire et polarimétrique pourra permettre d'avoir alors accès à la lumière non-polarisée de la surface du disque, et donc avoir accès à une information complémentaire sur la physique des grains qui composent disque, en particulier en utilisant des méthodes exemptées des artefacts liés à l'ADI, comme la méthode RDI (*Reference Differential Imaging*), ou bien la nouvelle méthode basée sur l'approche inverse développée par [Flasseur et al., 2020].

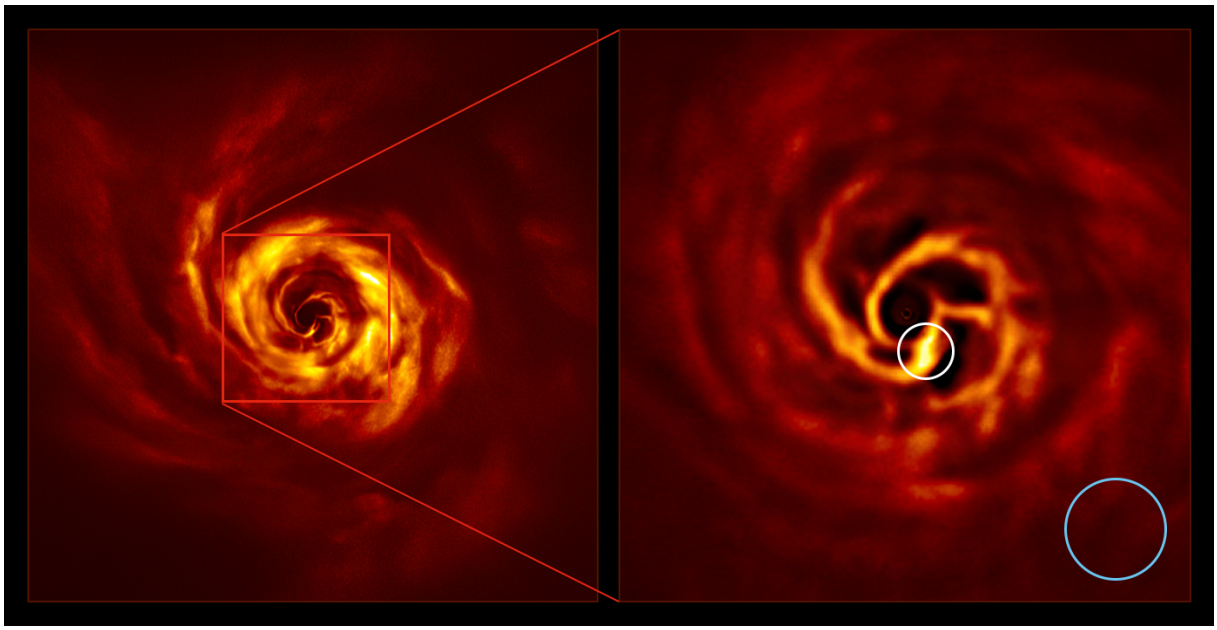


FIGURE 7.1 – Image de la lumière polarisée du disque entourant le système AB Aurigae. Le cercle bleu correspond à la taille de l'orbite de Neptune. Le point brillant entouré en blanc correspondrait à une zone de formation d'une exoplanète (Source : [Boccaletti et al., 2020])
.

Bibliographie

[Adam et al., 2016] Adam, R., Ade, P. A., Aghanim, N., Arnaud, M., Ashdown, M., Aumont, J., Baccigalupi, C., Banday, A., Barreiro, R., Bartolo, N., et al. (2016). Planck 2015 results-viii. high frequency instrument data processing : Calibration and maps. *Astronomy & Astrophysics*, 594 :A8.

[Akiyama et al., 2019] Akiyama, K., Alberdi, A., Alef, W., Asada, K., Azulay, R., Baczko, A.-K., Ball, D., Baloković, M., Barrett, J., Bintley, D., et al. (2019). First m87 event horizon telescope results. iv. imaging the central supermassive black hole. *The Astrophysical Journal Letters*, 875(1) :L4.

[Ammanouil et al., 2019] Ammanouil, R., Ferrari, A., Mary, D., Ferrari, C., and Loi, F. (2019). A parallel and automatically tuned algorithm for multispectral image deconvolution. *Monthly Notices of the Royal Astronomical Society*, 490(1) :37–49.

[Avenhaus et al., 2018] Avenhaus, H., Quanz, S. P., Garufi, A., Perez, S., Casassus, S., Pinte, C., Bertrang, G. H.-M., Caceres, C., Benisty, M., and Dominik, C. (2018). Disks around T Tauri Stars with SPHERE (DARTTS-S). I. SPHERE/IRDIS Polarimetric Imaging of Eight Prominent T Tauri Disks. *ApJ*, 863(1) :44.

[Avenhaus et al., 2014] Avenhaus, H., Quanz, S. P., Schmid, H. M., Meyer, M. R., Garufi, A., Wolf, S., and Dominik, C. (2014). Structures in the protoplanetary disk of HD142527 seen in polarized scattered light. *The Astrophysical Journal*, 781(2) :87. arXiv : 1311.7088.

[Bauschke and Combettes, 2011] Bauschke, H. H. and Combettes, P. L. (2011). *Convex analysis and monotone operator theory in Hilbert spaces*. CMS books in Mathematics. Springer, New York, NY. OCLC : 740935084.

[Beck and Teboulle, 2009] Beck, A. and Teboulle, M. (2009). A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM J. Imaging Sci.*, 2(1) :183–202.

[Benisty et al., 2015] Benisty, M., Juhasz, A., Boccaletti, A., Avenhaus, H., Milli, J., Thalmann, C., Dominik, C., Pinilla, P., Buenzli, E., Pohl, A., et al. (2015). Asymmetric features in the protoplanetary disk mwc 758. *Astronomy & Astrophysics*, 578 :L6.

[Berdeu et al., 2020] Berdeu, A., Soulez, F., Denis, L., Langlois, M., and Thiébaut, r. (2020). PIC : a data reduction algorithm for integral field spectrographs : Application to the SPHERE instrument. *A&A*, 635 :A90.

[Beuzit et al., 2019] Beuzit, J.-L., Vigan, A., Mouillet, D., Dohlen, K., Gratton, R., Boccaletti, A., Sauvage, J.-F., Schmid, H. M., Langlois, M., Petit, C., et al. (2019). Sphere : the exoplanet imager for the very large telescope. *Astronomy & Astrophysics*, 631 :A155.

[Birdi et al., 2019] Birdi, J., Repetti, A., and Wiaux, Y. (2019). Polca SARA - Full polarization, direction-dependent calibration and sparse imaging for radio interferometry. *arXiv :1904.00663 [astro-ph]*. arXiv : 1904.00663.

[Boccaletti et al., 2020] Boccaletti, A., Di Folco, E., Pantin, E., Dutrey, A., Guilloteau, S., Tang, Y., Piétu, V., Habart, E., Milli, J., Beck, T., et al. (2020). Possible evidence of ongoing planet formation in ab aurigae-a showcase of the sphere/alma synergy. *Astronomy & Astrophysics*, 637 :L5.

[Borde and Traub, 2006] Borde, P. J. and Traub, W. A. (2006). High-Contrast Imaging from Space : Speckle Nulling in a Low-Aberration Regime. *ApJ*, 638(1) :488–498.

[Boulanger et al., 2018] Boulanger, J., Pustelnik, N., Condat, L., Sengmanivong, L., and Piolot, T. (2018). Nonsmooth convex optimization for structured illumination microscopy image reconstruction. *Inverse problems*, 34(9) :095004.

[Bredies et al., 2010] Bredies, K., Kunisch, K., and Pock, T. (2010). Total Generalized Variation. *SIAM Journal on Imaging Sciences*, 3(3) :492–526.

[Brent, 1973] Brent, R. P. (1973). *Algorithms for Minimization without Derivatives*. Prentice-Hall, Englewood Cliffs, NJ.

[Carbillet et al., 2011] Carbillet, M., Bendjoya, P., Abe, L., Guerri, G., Boccaletti, A., Daban, J.-B., Dohlen, K., Ferrari, A., Robbe-Dubois, S., Douet, R., et al. (2011). Apodized lyot coronagraph for sphere/vlt. *Experimental Astronomy*, 30(1) :39–58.

[Carrillo et al., 2012] Carrillo, R. E., McEwen, J. D., and Wiaux, Y. (2012). Sparsity Averaging Reweighted Analysis (SARA) : a novel algorithm for radio-interferometric imaging : SARA for radio-interferometric imaging. *Monthly Notices of the Royal Astronomical Society*, 426(2) :1223–1234.

[Charbonnier et al., 1997] Charbonnier, P., Blanc-Féraud, L., Aubert, G., and Barlaud, M. (1997). Deterministic edge-preserving regularization in computed imaging. *IEEE Trans. on Image Process.*, 6(2) :298–311.

[Chierchia et al., 2012] Chierchia, G., Pustelnik, N., Pesquet, J.-C., and Pesquet-Popescu, B. (2012). Epigraphical splitting for solving constrained convex formulations of inverse problems with proximal tools. *arXiv preprint arXiv :1210.5844*.

[Chierchia et al., 2013] Chierchia, G., Pustelnik, N., Pesquet, J.-C., and Pesquet-Popescu, B. (2013). An epigraphical convex optimization approach for multicomponent image restoration using non-local structure tensor. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1359–1363, Vancouver, BC, Canada. IEEE.

[Chierchia et al., 2014] Chierchia, G., Pustelnik, N., Pesquet-Popescu, B., and Pesquet, J.-C. (2014). A Nonlocal Structure Tensor-Based Approach for Multicomponent Image Recovery Problems. *IEEE Transactions on Image Processing*, 23(12) :5531–5544.

[Claudi et al., 2008] Claudi, R. U., Turatto, M., Gratton, R. G., Antichi, J., Bonavita, M., Bruno, P., Cascone, E., De Caprio, V., Desidera, S., Giro, E., et al. (2008). Sphere ifs : the spectro differential imager of the vlt for exoplanets search. In *Ground-based and Airborne Instrumentation for Astronomy II*, volume 7014, page 70143E. International Society for Optics and Photonics.

[Combettes et al., 2014] Combettes, P. L., Condat, L., Pesquet, J.-C., and Vu, B. C. (2014). A forward-backward view of some primal-dual optimization methods in image recovery. In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 4141–4145, Paris, France. IEEE.

[Combettes and Vũ, 2013] Combettes, P. L. and Vũ, B. C. (2013). Variable metric quasi-Fejér monotonicity. *Nonlinear Analysis : Theory, Methods & Applications*, 78 :17–31.

[Combettes and Vũ, 2014] Combettes, P. L. and Vũ, B. C. (2014). Variable metric forward–backward splitting with applications to monotone inclusions in duality. *Optimization*, 63(9) :1289–1318.

[Condat, 2013] Condat, L. (2013). A Primal–Dual Splitting Method for Convex Optimization Involving Lipschitzian, Proximable and Linear Composite Terms. *Journal of Optimization Theory and Applications*, 158(2) :460–479.

[Cox et al., 2017] Cox, E. G., Harris, R. J., Looney, L. W., Chiang, H.-F., Chandler, C., Kratter, K., Li, Z.-Y., Perez, L., and Tobin, J. J. (2017). Protoplanetary disks in ρ ophiuchus as seen from alma. *The Astrophysical Journal*, 851(2) :83.

[de Boer et al., 2020] de Boer, J., Langlois, M., van Holstein, R. G., Girard, J. H., Mouillet, D., Vigan, A., Dohlen, K., Snik, F., Keller, C. U., Ginski, C., Stam, D. M., Milli, J., Wahhaj, Z., Kasper, M., Schmid, H. M., Rabou, P., Gluck, L., Hugot, E., Perret, D., Martinez, P., Weber, L., Pragt, J., Sauvage, J.-F., Boccaletti, A., Le Coroller, H., Dominik, C., Henning, T., Lagadec, E., Ménard, F., Turatto, M., Udry, S., Chauvin, G., Feldt, M., and Beuzit, J.-L. (2020). Polarimetric imaging mode of VLT/SPHERE/IRDIS : I. Description, data reduction, and observing strategy. *A&A*, 633 :A63.

[de Boer et al., 2016] de Boer, J., Salter, G., Benisty, M., Vigan, A., Boccaletti, A., Pinilla, P., Ginski, C., Juhasz, A., Maire, A.-L., Messina, S., et al. (2016). Multiple rings in the transition disk and companion candidates around rx j1615. 3-3255-high contrast imaging with vlt/sphere. *Astronomy & Astrophysics*, 595 :A114.

[Deledalle et al., 2014] Deledalle, C.-A., Vaiter, S., Fadili, J., and Peyré, G. (2014). Stein unbiased gradient estimator of the risk (sugar) for multiple parameter selection. *SIAM Journal on Imaging Sciences*, 7(4) :2448–2487.

[Denneulin et al., 2019] Denneulin, L., Langlois, M., Pustelnik, N., and Thiébaut, É. (2019). Reconstruction polarimétrique d'environnements circumstellaires à partir des données eso/vlt-sphere irdis. In *GRETSI LILLE 2019*.

[Dipierro et al., 2015] Dipierro, G., Price, D., Laibe, G., Hirsh, K., Cerioli, A., and Lodato, G. (2015). On planet formation in HL Tau. *Mon. Not. R. Astron. Soc : Lett.*, 453(1) :L73–L77.

[Eldar, 2008] Eldar, Y. C. (2008). Generalized sure for exponential families : Applications to regularization. *IEEE Transactions on Signal Processing*, 57(2) :471–481.

[Fessler and Hero, 1993] Fessler, J. A. and Hero, A. O. (1993). Cramer-rao lower bounds for biased image reconstruction. In *Proceedings of 36th Midwest Symposium on Circuits and Systems*, pages 253–256. IEEE.

[Flamary, 2017] Flamary, R. (2017). Astronomical image reconstruction with convolutional neural networks. pages 2468–2472.

[Flasseur et al., 2018] Flasseur, O., Denis, L., Thiébaut, É., and Langlois, M. (2018). Exoplanet detection in angular differential imaging by statistical learning of the nonstationary patch covariances : The PACO algorithm. *A&A*, 618 :A138.

[Flasseur et al., 2020] Flasseur, O., Denis, L., Thiébaut, É., and Langlois, M. (2020). Robustness to bad frames in angular differential imaging : a local weighting approach. *Astronomy & Astrophysics*, 634 :A2.

[Flasseur et al., 2019] Flasseur, O., Denis, L., Thiébaut, É., Olivier, T., and Fournier, C. (2019). Expaco : detection of an extended pattern under nonstationary correlated noise by patch

covariance modeling. In *2019 27th European Signal Processing Conference (EUSIPCO)*, pages 1–5. IEEE.

[Foi et al., 2008] Foi, A., Trimeche, M., Katkovnik, V., and Egiazarian, K. (2008). Practical poissonian-gaussian noise modeling and fitting for single-image raw-data. *IEEE Transactions on Image Processing*, 17(10) :1737–1754.

[Fusco et al., 2006] Fusco, T., Rousset, G., Sauvage, J.-F., Petit, C., Beuzit, J.-L., Dohlen, K., Mouillet, D., Charton, J., Nicolle, M., Kasper, M., et al. (2006). High-order adaptive optics requirements for direct detection of extrasolar planets : Application to the sphere instrument. *Optics Express*, 14(17) :7515–7534.

[Haffert et al., 2019] Haffert, S. Y., Bohn, A. J., de Boer, J., Snellen, I. A. G., Brinchmann, J., Girard, J. H., Keller, C. U., and Bacon, R. (2019). Two accreting protoplanets around the young star PDS 70. *Nat Astron*, 3(8) :749–754.

[Högbom, 1974] Högbom, J. (1974). Aperture synthesis with a non-regular distribution of interferometer baselines. *Astronomy and Astrophysics Supplement Series*, 15 :417.

[Huber, 1992] Huber, P. J. (1992). Robust estimation of a location parameter. In *Breakthroughs in statistics*, pages 492–518. Springer.

[Hunziker et al., 2020] Hunziker, S., Schmid, H. M., Mouillet, D., Milli, J., Zurlo, A., Delorme, P., Abe, L., Avenhaus, H., Baruffolo, A., Bazzon, A., et al. (2020). Refplanets : Search for reflected light from extrasolar planets with sphere/zimpol. *Astronomy & Astrophysics*, 634 :A69.

[Jacques et al., 2011] Jacques, L., Duval, L., Chaux, C., and Peyré, G. (2011). A panorama on multiscale geometric representations, intertwining spatial, directional and frequency selectivity. *Signal Processing*, 91(12) :2699–2730.

[Kasper et al., 2016] Kasper, M., Santhakumari, K., Herbst, T., and Köhler, R. (2016). New circumstellar structure in the t tauri system-a near-infrared high-contrast imaging study. *Astronomy & Astrophysics*, 593 :A50.

[Keppler et al., 2018] Keppler, M., Benisty, M., Müller, A., Henning, T., Van Boekel, R., Cantalloube, F., Ginski, C., Van Holstein, R., Maire, A.-L., Pohl, A., et al. (2018). Discovery of a planetary-mass companion within the gap of the transition disk around pds 70. *Astronomy & Astrophysics*, 617 :A44.

[Keppler et al., 2019] Keppler, M., Teague, R., Bae, J., Benisty, M., Henning, T., van Boekel, R., Chapillon, E., Pinilla, P., Williams, J. P., Bertrang, G. H.-M., Facchini, S., Flock, M., Ginski, C., Juhasz, A., Klahr, H., Liu, Y., Müller, A., Pérez, L. M., Pohl, A., Rosotti, G., Samland, M., and Semenov, D. (2019). Highly structured disk around the planet host PDS 70 revealed by high-angular resolution observations with ALMA. *A&A*, 625 :A118.

[Langlois et al., 2014] Langlois, M., Dohlen, K., Vigan, A., Zurlo, A., Moutou, C., Schmid, H., Mili, J., Beuzit, J.-L., Boccaletti, A., Carle, M., et al. (2014). High contrast polarimetry in the infrared with sphere on the vlt. In *Ground-based and Airborne Instrumentation for Astronomy V*, volume 9147, page 91471R. International Society for Optics and Photonics.

[Langlois et al., 2018] Langlois, M., Pohl, A., Lagrange, A.-M., Maire, A.-L., Mesa, D., Boccaletti, A., Gratton, R., Denneulin, L., Klahr, H., Vigan, A., et al. (2018). First scattered light detection of a nearly edge-on transition disk around the t tauri star ry lupi. *A& A*, 614 :A88.

[Lefkimmiatis et al., 2013] Lefkimmiatis, S., Ward, J. P., and Unser, M. (2013). Hessian schatten-norm regularization for linear inverse problems. *IEEE transactions on image processing*, 22(5) :1873–1888.

[Li and Zhang, 2016] Li, Q. and Zhang, N. (2016). Fast proximity-gradient algorithms for structured convex optimization problems. *Applied and Computational Harmonic Analysis*, 41(2) :491–517.

[Liu and Nocedal, 1989] Liu, D. C. and Nocedal, J. (1989). On the limited memory bfgs method for large scale optimization. *Mathematical programming*, 45(1-3) :503–528.

[Lorenz and Pock, 2015] Lorenz, D. A. and Pock, T. (2015). An inertial forward-backward algorithm for monotone inclusions. *Journal of Mathematical Imaging and Vision*, 51(2) :311–325. arXiv : 1403.3522.

[Lynden-Bell and Pringle, 1974] Lynden-Bell, D. and Pringle, J. E. (1974). The evolution of viscous discs and the origin of the nebular variables. *Monthly Notices of the Royal Astronomical Society*, 168(3) :603–637.

[Macintosh et al., 2014] Macintosh, B., Graham, J. R., Ingraham, P., Konopacky, Q., Marois, C., Perrin, M., Poyneer, L., Bauman, B., Barman, T., Burrows, A. S., et al. (2014). First light of the gemini planet imager. *proceedings of the National Academy of Sciences*, 111(35) :12661–12666.

[Mahalanobis, 1936] Mahalanobis, P. C. (1936). On the generalized distance in statistics. National Institute of Science of India.

[Mesa et al., 2016] Mesa, D., Vigan, A., D'Orazi, V., Ginski, C., Desidera, S., Bonnefoy, M., Gratton, R., Langlois, M., Marzari, F., Messina, S., Antichi, J., Biller, B., Bonavita, M., Cascone, E., Chauvin, G., Claudi, R. U., Curtis, I., Fantinel, D., Feldt, M., Garufi, A., Galicher, R., Henning, T., Incorvaia, S., Lagrange, A.-M., Millward, M., Perrot, C., Salasnich, B., Scuderi, S., Sissa, E., Wahhaj, Z., and Zurlo, A. (2016). Characterizing HR 3549 B using SPHERE. *A&A*, 593 :A119.

[Milli et al.,] Milli, J., Engler, N., Schmid, H. M., Olofsson, J., Ménard, F., Kral, Q., Boccaletti, A., Thébault, P., Choquet, E., and Mouillet, D. Optical polarized phase function of the HR 4796A dust ring. page 13.

[Mulders et al., 2013] Mulders, G. D., Min, M., Dominik, C., Debes, J. H., and Schneider, G. (2013). Why circumstellar disks are so faint in scattered light : the case of hd 100546. *Astronomy & Astrophysics*, 549 :A112.

[Nocedal, 1980] Nocedal, J. (1980). Updating quasi-newton matrices with limited storage. *Mathematics of computation*, 35(151) :773–782.

[Nocedal and Wright, 1999] Nocedal, J. and Wright, S. J. (1999). *Numerical optimization*. Springer series in operations research. Springer, New York.

[Pairet et al., 2018a] Pairet, B., Cantalloube, F., and Jacques, L. (2018a). Reference-less algorithm for circumstellar disks imaging. *Proceedings of iTWIST'18*, (23).

[Pairet et al., 2018b] Pairet, B., Cantalloube, F., and Jacques, L. (2018b). Reference-less algorithm for circumstellar disks imaging. *arXiv :1812.01333 [astro-ph]*. arXiv : 1812.01333.

[Pairet et al., 2019] Pairet, B., Jacques, L., and Cantalloube, F. (2019). Iterative low-rank and rotating sparsity promotion for circumstellar disks imaging. *Proceedings of SPARS'19*, 1 :1.

[Pavlov et al., 2008] Pavlov, A., Feldt, M., and Henning, T. (2008). Data reduction and handling for sphere. In *Astronomical Data Analysis Software and Systems XVII*, volume 394, page 581.

[Perrin et al., 2015] Perrin, M. D., Duchene, G., Millar-Blanchaer, M., Fitzgerald, M. P., Graham, J. R., Wiktorowicz, S. J., Kalas, P. G., Macintosh, B., Bauman, B., Cardwell, A., Chilcote,

J., De Rosa, R. J., Dillon, D., Doyon, R., Dunn, J., Erikson, D., Gavel, D., Goodsell, S., Hartung, M., Hibon, P., Ingraham, P., Kerley, D., Konapacky, Q., Larkin, J. E., Maire, J., Marchis, F., Marois, C., Mittal, T., Morzinski, K. M., Oppenheimer, B. R., Palmer, D. W., Patience, J., Poyneer, L., Pueyo, L., Rantakyrö, F. T., Sadakuni, N., Saddlemyer, L., Savransky, D., Soummer, R., Sivaramakrishnan, A., Song, I., Thomas, S., Wallace, J. K., Wang, J. J., and Wolff, S. G. (2015). POLARIMETRY WITH THE GEMINI PLANET IMAGER : METHODS, PERFORMANCE AT FIRST LIGHT, AND THE CIRCUMSTELLAR RING AROUND HR 4796A. *ApJ*, 799(2) :182.

[Petit et al., 2012] Petit, C., Sauvage, J.-F., Sevin, A., Costille, A., Fusco, T., Baudoz, P., Beuzit, J.-L., Buey, T., Charton, J., Dohlen, K., et al. (2012). The sphere xao system saxo : integration, test, and laboratory performance. In *Adaptive Optics Systems III*, volume 8447, page 84471Z. International Society for Optics and Photonics.

[Pinte et al., 2019] Pinte, C., van der Plas, G., Menard, F., Price, D. J., Christiaens, V., Hill, T., Mentiplay, D., Ginski, C., Choquet, E., Boehler, Y., Duchene, G., Perez, S., and Casassus, S. (2019). Kinematic detection of a planet carving a gap in a protoplanetary disc. *Nat Astron*, 3(12) :1109–1114. arXiv : 1907.02538.

[Polyak, 1987] Polyak, B. T. (1987). Introduction to optimization. optimization software. *Inc., Publications Division, New York*, 1.

[Powell, 2006] Powell, M. J. (2006). The newuoa software for unconstrained optimization without derivatives. In *Large-scale nonlinear optimization*, pages 255–297. Springer.

[Price et al., 2018] Price, D. J., Cuello, N., Pinte, C., Mentiplay, D., Casassus, S., Christiaens, V., Kennedy, G. M., Cuadra, J., M., S. P., Marino, S., Armitage, P. J., Zurlo, A., Juhasz, A., Ragusa, E., Laibe, G., and Lodato, G. (2018). Circumbinary, not transitional : On the spiral arms, cavity, shadows, fast radial flows, streamers and horseshoe in the HD142527 disc. *Monthly Notices of the Royal Astronomical Society*, 477(1) :1270–1284. arXiv : 1803.02484.

[Pustelnik et al., 2016] Pustelnik, N., Benazza-Benhayia, A., Zheng, Y., and Pesquet, J.-C. (2016). Wavelet-Based Image Deconvolution and Reconstruction. In *Wiley Encyclopedia of Electrical and Electronics Engineering*, pages 1–34. John Wiley & Sons, Inc., Hoboken, NJ, USA.

[Ramani et al., 2008] Ramani, S., Blu, T., and Unser, M. (2008). Monte-Carlo Sure : A Black-Box Optimization of Regularization Parameters for General Denoising Algorithms. *IEEE Trans. on Image Process.*, 17(9) :1540–1554.

[Repetti et al., 2017] Repetti, A., Birdi, J., Dabbech, A., and Wiaux, Y. (2017). Non-convex optimization for self-calibration of direction-dependent effects in radio interferometric imaging. *Monthly Notices of the Royal Astronomical Society*, 470(4) :3981–4006. arXiv : 1701.03689.

[Rudin et al., 1992] Rudin, L. I., Osher, S., and Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D : Nonlinear Phenomena*, 60(1-4) :259–268.

[Schmid et al., 2018] Schmid, H. M., Bazzon, A., Roelfsema, R., Mouillet, D., Milli, J., Menard, F., Gisler, D., Hunziker, S., Pragt, J., Dominik, C., et al. (2018). Sphere/zimpol high resolution polarimetric imager-i. system overview, psf parameters, coronagraphy, and polarimetry. *Astronomy & Astrophysics*, 619 :A9.

[Sissa et al., 2018] Sissa, E., Olofsson, J., Vigan, A., Augereau, J.-C., D'Orazi, V., Desidera, S., Gratton, R., Langlois, M., Rigliaco, E., Boccaletti, A., et al. (2018). New disk discovered with vlt/sphere around the m star gsc 07396- 00759. *Astronomy & Astrophysics*, 613 :L6.

[Starck et al., 2003] Starck, J.-L., Donoho, D. L., and Candès, E. J. (2003). Astronomical image representation by the curvelet transform. *Astronomy & Astrophysics*, 398(2) :785–800.

[Stein, 1981] Stein, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *The annals of Statistics*, pages 1135–1151.

[Tarantola, 2005] Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. SIAM.

[Theodoridis et al., 2011] Theodoridis, S., Slavakis, K., and Yamada, I. (2011). Adaptive Learning in a World of Projections. *IEEE Signal Process. Mag.*, 28(1) :97–123.

[Thiébaut, 2002] Thiébaut, E. (2002). Optimization issues in blind deconvolution algorithms. In *Astronomical Data Analysis II*, volume 4847, pages 174–183. International Society for Optics and Photonics.

[Thiébaut, 2020] Thiébaut, É. (2020). *PointSpreadFunctions*. Julia Package.

[Thiébaut and Conan, 1995] Thiébaut, E. and Conan, J.-M. (1995). Strict a priori constraints for maximum-likelihood blind deconvolution. *JOSA A*, 12(3) :485–492.

[Tikhonov, 1963] Tikhonov, A. N. (1963). Regularization of incorrectly posed problems. Soviet Mathematics Doklady.

[Tinbergen, 2005] Tinbergen, J. (2005). *Astronomical Polarimetry*. Cambridge University Press. Google-Books-ID : SAS4JzAaMxkC.

[Titterington, 1985] Titterington, D. M. (1985). General structure of regularization procedures in image reconstruction. 144 :381–387.

[Todros and Tabrikian, 2010a] Todros, K. and Tabrikian, J. (2010a). General Classes of Performance Lower Bounds for Parameter Estimation—Part I : Non-Bayesian Bounds for Unbiased Estimators. *IEEE Trans. Inform. Theory*, 56(10) :5045–5063.

[Todros and Tabrikian, 2010b] Todros, K. and Tabrikian, J. (2010b). General Classes of Performance Lower Bounds for Parameter Estimation—Part II : Bayesian Bounds. *IEEE Trans. Inform. Theory*, 56(10) :5064–5082.

[Vũ, 2015] Vũ, B. C. (2015). A Splitting Algorithm for Coupled System of Primal–Dual Monotone Inclusions. *Journal of Optimization Theory and Applications*, 164(3) :993–1025.

[van Holstein et al., 2020] van Holstein, R. G., Girard, J. H., de Boer, J., Snik, F., Milli, J., Stam, D. M., Ginski, C., Mouillet, D., Wahhaj, Z., Schmid, H. M., Keller, C. U., Langlois, M., Dohlen, K., Vigan, A., Pohl, A., Carbillet, M., Fantinel, D., Maurel, D., Origné, A., Petit, C., Ramos, J., Rigal, F., Sevin, A., Boccaletti, A., Le Coroller, H., Dominik, C., Henning, T., Lagadec, E., Ménard, F., Turatto, M., Udry, S., Chauvin, G., Feldt, M., and Beuzit, J.-L. (2020). Polarimetric imaging mode of VLT/SPHERE/IRDIS : II. Characterization and correction of instrumental polarization effects. *A&A*, 633 :A64.

[van Holstein et al., 2017] van Holstein, R. G., Snik, F., Girard, J. H., de Boer, J., Ginski, C., Keller, C. U., Stam, D. M., Beuzit, J.-L., Mouillet, D., Kasper, M., Langlois, M., Zurlo, A., de Kok, R. J., and Vigan, A. (2017). Combining angular differential imaging and accurate polarimetry with SPHERE/IRDIS to characterize young giant exoplanets. *Techniques and Instrumentation for Detection of Exoplanets VIII*, page 38. arXiv : 1709.07519.

[Vigan et al., 2014] Vigan, A., Langlois, M., Dohlen, K., Zurlo, A., Moutou, C., Costille, A., Gry, C., Madec, F., Le Mignant, D., Gluck, L., et al. (2014). Sphere/irdis : final performance assessment of the dual-band imaging and long slit spectroscopy modes. In *Ground-based and*

Airborne Instrumentation for Astronomy V, volume 9147, page 91474T. International Society for Optics and Photonics.

[Vigan et al., 2010] Vigan, A., Moutou, C., Langlois, M., Allard, F., Boccaletti, A., Carbillet, M., Mouillet, D., and Smith, I. (2010). Photometric characterization of exoplanets using angular and spectral differential imaging : Exoplanet characterization using ADI and SDI. *Monthly Notices of the Royal Astronomical Society*, 407(1) :71–82.

Annexe A

Convergences des algorithmes : définitions et théorèmes

A.1 Définitions et propositions

Définition A.1.1 (Fonction semi-continue inférieurement)**.** *Soit $\mathcal{H}$ un espace d'Hilbert. Soit $f : \mathcal{H} \to]-\infty, +\infty]$. On dit que f est semi-continue inférieurement sur$\mathcal{H}$ si et seulement si son épigraphe, donné par :*

$$\text{epi } f \Big\{(\boldsymbol{x}, \xi) \in \text{dom} f \times \mathbb{R} \quad \textit{tels que} \quad f(\boldsymbol{x}) \leq \xi\Big\}, \tag{A.1}$$

est fermé

Ordre de Loewner $\succeq$: Soit $\mathcal{H}$ un espace d'hilbert, deux matrices $\mathbf{A} \in \mathcal{M}(\mathcal{H})$ et $\mathbf{B} \in \mathcal{M}(\mathcal{H})$, alors :

$$\forall \boldsymbol{x} \in \mathcal{H}, \quad \mathbf{A} \succeq \mathbf{B} \Leftrightarrow \langle \boldsymbol{x}, \mathbf{A}\boldsymbol{x} \rangle \geq \langle \boldsymbol{x}, \mathbf{B}\boldsymbol{x} \rangle. \tag{A.2}$$

Soit $\mathcal{H}$ un espace d'hilbert, on note l'ensemble des matrices de $\mathcal{H}$ symétriques ρ-positives comme :

$$\mathcal{P}_\rho(\mathcal{H}) = \Big\{\mathbf{M} \in \mathcal{S}(\mathcal{H}) \quad \text{telles que} \quad \forall \boldsymbol{x} \in \mathcal{H},\ \langle \boldsymbol{x}, \mathbf{M}\boldsymbol{x} \rangle \geq \rho \|\boldsymbol{x}\|^2\Big\} \tag{A.3}$$

où $\mathcal{S}(\mathcal{H})$ est l'ensemble des matrices symétriques de $\mathcal{H}$.

Définition A.1.2. *Soit $D \subset \mathcal{H} \backslash \emptyset$ un ensemble convexe fermé. Soit $(\boldsymbol{x}_n)_{n\in\mathbb{N}} \in \mathcal{H}$ une suite. Alors $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ est Fejér-monotone par rapport à D si :*

$$\forall \boldsymbol{x} \in D \quad \|\boldsymbol{x}_{n+1} - \boldsymbol{x}\|^2 \leq \|\boldsymbol{x}_n - \boldsymbol{x}\|^2.$$

Définition A.1.3. *Soit $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ une suite de $\mathcal{H}$ et $\widehat{\boldsymbol{x}} \in \mathcal{H}$.*

- *Si $\lim_{n\to+\infty} \|\boldsymbol{x}_n - \widehat{\boldsymbol{x}}\| = 0$ alors $\boldsymbol{x}_n \to \widehat{\boldsymbol{x}}$ (convergence forte).*
- *Si $\lim_{n\to+\infty} \langle \boldsymbol{y}, \boldsymbol{x}_n - \widehat{\boldsymbol{x}} \rangle = 0, \forall \boldsymbol{y} \in \mathcal{H}$ alors $\boldsymbol{x}_n \rightharpoonup \widehat{\boldsymbol{x}}$ (convergence faible).*

Remarque : Dans un espace de Hilbert réel fini, forte et faible convergence sont équivalentes.

Proposition A.1.4. *Soit $(\boldsymbol{x}_n)_{n\in\mathbb{N}} \in \mathcal{H}$ une suite Fejér-monotone et $\widehat{\boldsymbol{x}} \in D$ une de ses limites. Alors si toute sous-suite de $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ convergentes vers $\widehat{\boldsymbol{x}}$ appartiennent à D, $(\boldsymbol{x}_n)_{n\in\mathbb{N}} \rightharpoonup \widehat{\boldsymbol{x}}$.*

Définition A.1.5. *Un opérateur* T *est* α*-moyenné si il existe un opérateur* R *non-expansif tel que*

$$\mathrm{T} = (1-\alpha)\mathrm{Id} + \alpha \mathrm{R}.$$

Définition A.1.6. *Soit* $\phi : \mathbb{R}_+ \Rightarrow \mathbb{R}_+$*, soit* $(\mathbf{P}^{[t]})_{n\in\mathbb{N}}$ *une suite dans* $\mathcal{S}_+(\mathcal{H})$ *et soit* $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ *une suite dans* $\mathcal{H}$*. Alors* $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ *est* ϕ*-quasi-Fejér par rapport au sous-ensemble* $D \in \mathcal{H}$*, relativement à* $(\mathbf{P}^{[t]})_{n\in\mathbb{N}}$ *si* $\exists (\eta_n)_{n\in\mathbb{N}} \in \mathbb{R}_+$ *avec* $\sum_n |\eta_n| < +\infty$ *et* $\exists (\rho)_{n\in\mathbb{N}} \in \mathbb{R}_+$ *tel que* $\sum_n |\rho_n| < +\infty$*, tel que :*

$$\forall n \in \mathbb{N},\ \forall x \in D, \quad \phi\left(\|\boldsymbol{x}_{n+1} - \boldsymbol{x}\|_{U_{n+1}}\right) \leq (1+\eta_n)\phi\left(\|\boldsymbol{x}_n - \boldsymbol{x}\|_{U_n}\right) + \rho_n. \tag{A.4}$$

Proposition A.1.7 (Principe de demi-fermeture)**.** *Soit* $\mathcal{H}$ *un espace de Hilbert et* $\mathcal{C} \neq \emptyset$ *un sous ensemble convexe de* $\mathcal{H}$*.*
Soit $\mathbf{T} : \mathcal{C} \to \mathcal{H}$ *un opérateur contractant. Si* $(\boldsymbol{x}^{[t]})_{t\in\mathbb{N}}$ *est une suite de* $\mathcal{C}$ *converge faiblement vers* $\widehat{\boldsymbol{x}}$ *et si* $(\mathbf{T}\boldsymbol{x}^{[t]} - \boldsymbol{x}^{[t]}) \to 0$ *alors* $\widehat{\boldsymbol{x}} \in \mathrm{Fix}\ \mathbf{T}$*.*

Note : Les deux sections suivantes proviennes de notes rédigées en anglais et n'ont pas été traduites en français pour cette thèse.

A.2 On the convergence of Krasnosel'skii-Mann

Théoreme A.2.1. *Let* $\mathcal{H}$ *be a finit Hilbert space and* $C \neq \emptyset$ *a closed convex set of* $\mathcal{H}$*.*
Let $T : C \to C$ *be a non-expansive operator such as* $\mathrm{Fix\ T} \neq \emptyset$*.*
Let $(\lambda_n)_{n\in\mathbb{N}}$ *be a sequence in* $]0,1[$ *such as* $\sum_{n\in\mathbb{N}} \lambda_n(1-\lambda_n) = +\infty$*. Let* $x_0 \in C$ *and* $(\forall n \in \mathbb{N})$ $x_{n+1} = x_n + \lambda_n(\mathrm{T}x_n - x_n)$ *then :*

1. *The sequence* $(x_n)_{n\in\mathbb{N}}$ *is Fejér-monoton compared to* Fix T.
2. $(\mathrm{T}x_n - x_n)_{n\in\mathbb{N}}$ *converges to* 0.
3. *The sequence* $(x_n)_{n\in\mathbb{N}}$ *converges to a fixed point of* Fix T.

Démonstration. The overall idea is to show that the squence x_n converges to a fixed point $\widehat{x}$ of T.

1. We want to show that $(x_n)_{n\in\mathbb{N}}$ is Fejér-monotone with respect to Fix T. Let $\widehat{x} \in$ Fix T, then :

$$\begin{aligned}
\|x_{n+1} - \widehat{x}\|^2 &= \|x_n + \lambda_n(\mathrm{T}x_n - x_n) - \widehat{x}\|^2, \\
&= \|x_n + \lambda_n \mathrm{T}x_n - \lambda_n x_n - \widehat{x} - \lambda_n \widehat{x} + \lambda_n \widehat{x}\|^2, \\
&= \|(1-\lambda_n)(x_n - \widehat{x}) + \lambda_n(\mathrm{T}x_n - \widehat{x})\|^2, \\
&= \|(1-\lambda_n)(x_n - \widehat{x})\|^2 + 2\langle (1-\lambda_n)(x_n - \widehat{x}), \lambda_n(\mathrm{T}x_n - \widehat{x})\rangle + \|\lambda_n(\mathrm{T}x_n - \widehat{x})\|^2, \\
&= (1-\lambda_n)^2\|x_n - \widehat{x}\|^2 + 2\lambda_n(1-\lambda_n)\langle x_n - \widehat{x}, \mathrm{T}x_n - \widehat{x}\rangle + \lambda_n^2\|\mathrm{T}x_n - \widehat{x}\|^2.
\end{aligned}$$

Now we use the fact that $2\langle a-b, a-c\rangle = \|a-b\|^2 + \|a-c\|^2 - \|b-c\|^2$, so :

$$
\begin{aligned}
&\|x_{n+1} - \widehat{x}\|^2 \\
=&(1-\lambda_n)^2\|x_n - \widehat{x}\|^2 + \lambda_n(1-\lambda_n)\left(\|x_n - \widehat{x}\|^2 + \|\mathrm{T}x_n - \widehat{x}\|^2 - \|\mathrm{T}x_n - x_n\|^2\right) + \lambda_n^2\|\mathrm{T}x_n - \widehat{x}\|^2, \\
=&(1-\lambda_n)\|x_n - \widehat{x}\|^2 + \lambda_n\|\mathrm{T}x_n - \widehat{x}\|^2 - \lambda_n(1-\lambda_n)\|\mathrm{T}x_n - x_n\|^2, \\
=&(1-\lambda_n)\|x_n - \widehat{x}\|^2 + \lambda_n\|\mathrm{T}x_n - \mathrm{T}\widehat{x}\|^2 - \lambda_n(1-\lambda_n)\|\mathrm{T}x_n - x_n\|^2, \\
\leq&(1-\lambda_n)\|x_n - \widehat{x}\|^2 + \lambda_n\|x_n - \widehat{x}\|^2 - \lambda_n(1-\lambda_n)\|\mathrm{T}x_n - x_n\|^2 \quad \text{since T is non-expensive,} \\
\leq&\|x_n - \widehat{x}\|^2 - \underbrace{\lambda_n(1-\lambda_n)\|\mathrm{T}x_n - x_n\|^2}_{\geq 0}.
\end{aligned}
$$

In the end, we have $\|x_{n+1} - \widehat{x}\|^2 \leq \|x_n - \widehat{x}\|^2$. The sequence $(x_n)_{n\in\mathbb{N}}$ is thus Fejér-monotone with respect to Fix T.

2. We want to show that $(\mathrm{T}x_n - x_n)_{n\in\mathbb{N}}$ converges to 0.

$$
\begin{aligned}
&\|\mathrm{T}x_{n+1} - x_{n+1}\| \\
=&\|\mathrm{T}x_{n+1} - x_n - \lambda_n(\mathrm{T}x_n - x_n)\|, \\
=&\|\mathrm{T}x_{n+1} - x_n - \lambda_n\mathrm{T}x_n + \lambda_n x_n + \mathrm{T}x_n - \mathrm{T}x_n\|, \\
=&\|\mathrm{T}x_{n+1} - \mathrm{T}x_n + (1-\lambda_n)(\mathrm{T}x_n - x_n)\|, \\
\leq&\|\mathrm{T}x_{n+1} - \mathrm{T}x_n\| + (1-\lambda_n)^2\|\mathrm{T}x_n - x_n\|, \\
\leq&\|x_{n+1} - x_n\|^2 + |1-\lambda_n|\|\mathrm{T}x_n - x_n\|^2 \quad \text{since T is non-expensive,} \\
=&\|x_n + \lambda_n(\mathrm{T}x_n - x_n) - x_n\| + (1-\lambda_n)\|\mathrm{T}x_n - x_n\| \quad \text{since } \lambda_n > 0, \\
=&\lambda_n\|\mathrm{T}x_n - x_n\| + (1-\lambda_n)\|\mathrm{T}x_n - x_n\|, \\
=&\|\mathrm{T}x_n - x_n\|.
\end{aligned}
$$

We have $\|\mathrm{T}x_{n+1} - x_{n+1}\| \leq \|\mathrm{T}x_n - x_n\|$. Consequently, the sequence $(\mathrm{T}x_n - x_n)_{n\in\mathbb{N}}$ converges. Now we want to show that $\lim_{n\to+\infty}\|\mathrm{T}x_n - x_n\| = 0$, because this, according to the definition of the convergence, implies that $(Tx_n - x_n) \to 0$.
Starting from Fejér equations with :

$$
\begin{aligned}
&\|x_{n+1} - \widehat{x}\|^2 \leq \|x_n - \widehat{x}\|^2 - \lambda_n(1-\lambda_n)\|\mathrm{T}x_n - x_n\|^2, \\
\Leftrightarrow&\lambda_n(1-\lambda_n)\|\mathrm{T}x_n - x_n\|^2 \leq \|x_n - \widehat{x}\|^2 - \|x_{n+1} - \widehat{x}\|^2, \\
\Leftrightarrow&\sum_{n\in\mathbb{N}}\lambda_n(1-\lambda_n)\|\mathrm{T}x_n - x_n\|^2 \leq \sum_{n\in\mathbb{N}}\left[\|x_n - \widehat{x}\|^2 - \|x_{n+1} - \widehat{x}\|^2\right], \\
\Leftrightarrow&\sum_{n\in\mathbb{N}}\lambda_n(1-\lambda_n)\|\mathrm{T}x_n - x_n\|^2 \leq \|x_0 - \widehat{x}\|^2 + \|x_1 - \widehat{x}\|^2 + \cdots - \|x_1 - \widehat{x}\|^2 - \ldots, \\
\Leftrightarrow&\sum_{n\in\mathbb{N}}\lambda_n(1-\lambda_n)\|\mathrm{T}x_n - x_n\|^2 \leq \|x_0 - \widehat{x}\|^2 - \|x_{+\infty} - \widehat{x}\|^2, \\
\Leftrightarrow&\sum_{n\in\mathbb{N}}\lambda_n(1-\lambda_n)\|\mathrm{T}x_n - x_n\|^2 \leq \|x_0 - \widehat{x}\|^2 < +\infty.
\end{aligned}
$$

Since $\sum_{n\in\mathbb{N}} \lambda_n(1-\lambda_n) = +\infty$ and $\sum_{n\in\mathbb{N}} \lambda_n(1-\lambda_n)\|\mathrm{T}x_n - x_n\|^2 < +\infty$, then $\lim_{n\to+\infty} \|\mathrm{T}x_n - x_n\| = 0$. We conclude that $(Tx_n - x_n) \to 0$.

3. We proved that the sequence $(x_n)_{n\in\mathbb{N}}$ is Fejér-monotone with respect to Fix T. We also proved that $\mathrm{T}x_n - x_n \to 0$.

   Suppose that a sub-sequence $(x_{n_k})_{k\in\mathbb{N}})$ converges to a point $\widetilde{x}$. Since T is a contraction and $(Tx_n - x_n) \to 0$, according to the demiclosedness principle, $\widetilde{x} \in$ Fix T.

   This means that the sequential clustering point of $(x_n)_{n\in\mathbb{N}}$ are in FixT. Finally, since $(x_n)_{n\in\mathbb{N}}$ is Fejér-monotone with respect to FixT, in conclusion $(x_n)_{n\in\mathbb{N}}$ converges weakly to $\widetilde{x} \in$ FixT.

□

A.3 Forward-Backward

The algorithme Forward-Backward aim to solve :

$$0 \in \mathrm{A} + \mathrm{B}, \tag{A.5}$$

where B is an α-averaged operator, β-Lipshitz (and thus β^{-1}-cocoercitive), and A an other α-averaged operator. The iterations of the algorithm are given by :

Algorithme 10 : Forward-Backward iterations in the specific case $\lambda^{[t]} = 1$ and $\gamma^{[t]} = \gamma$ fixed.

Set $\boldsymbol{x}^{[0]} \in \mathcal{H}$, then :
for $t = 0, 1, \ldots$ **do**
  $\boldsymbol{x}^{[t+1]} = J_{\gamma \mathrm{A}}(\boldsymbol{x}^{[t]} - \gamma \mathrm{B}\boldsymbol{x}^{[t]})$.
end

The iterations of Forward-Backward are a sequence $x_{n+1} = \mathrm{T}x_n$ where $\mathrm{T}\boldsymbol{x} = J_{\gamma \mathrm{A}}(\boldsymbol{x} - \gamma \mathrm{B}\boldsymbol{x})$. We want to prove that $x_n \rightharpoonup \widetilde{x}$ with $\widetilde{x} \in$ FixT. The proof is structured as in the Krasnosel'skii-Mann convergence's proof.

1. We want to show that $(x_n)_{n\in\mathbb{N}}$ is Fejér-monotone with respect to Fix T. Let $\widehat{x} \in$ Fix T, then :

$$\begin{aligned} \|x_{n+1} - \widehat{x}\|^2 &= \|\mathrm{T}x_n - \widehat{x}\|^2, \\ &= \|\mathrm{T}x_n - \mathrm{T}\widehat{x}\|^2. \end{aligned}$$

   Since T is α-averaged :

$$\|x - y\|^2 \geq \|\mathrm{T}x - \mathrm{T}y\|^2 + \frac{1-\alpha}{\alpha}\|(\mathrm{Id} - \mathrm{T})(x - y)\|^2, \tag{A.6}$$

   and then

$$\begin{aligned}\|x_{n+1}-\widehat{x}\|^2 &\leq \|x_n-\widehat{x}\|^2-\frac{1-\alpha}{\alpha}\|x_n-\widehat{x}-\mathrm{T}x_n+\mathrm{T}\widehat{x}\|^2,\\ &= \|x_n-\widehat{x}\|^2-\frac{1-\alpha}{\alpha}\|x_n-\widehat{x}-\mathrm{T}x_n+\widehat{x}\|^2,\\ &= \|x_n-\widehat{x}\|^2-\underbrace{\frac{1-\alpha}{\alpha}\|x_n-\mathrm{T}x_n\|^2}_{\geq 0}.\end{aligned}$$

We have $\|x_{n+1}-\widehat{x}\|^2 \leq \|x_n-\widehat{x}\|^2$. The sequence $(x_n)_{n\in\mathbb{N}}$ is thus Fejér-monotone with respect to Fix T.

2. Now want to show that $(\mathrm{T}x_n-x_n)_{n\in\mathbb{N}}$ converges to 0.

$$\begin{aligned}\|\mathrm{T}x_{n+1}-x_{n+1}\| &= \|\mathrm{T}x_{n+1}-\mathrm{T}x_n\|,\\ &= \|\mathrm{T}(x_{n+1}-x_n)\|.\end{aligned}$$

Using the fact that since T is α-averaged, $\exists$R non-expansive such as $\mathrm{T}=(1-\alpha)\mathrm{Id}+\alpha\mathrm{R}$:

$$\begin{aligned}\|\mathrm{T}x_{n+1}-x_{n+1}\| &= \|\left((1-\alpha)\mathrm{Id}+\alpha\mathrm{R}\right)(x_{n+1}-x_n)\|,\\ &\leq |1-\alpha|\,\|x_{n+1}-x_n\|+\alpha\|\mathrm{R}(x_{n+1}-x_n)\|\\ &\leq (1-\alpha)\,\|x_{n+1}-x_n\|+\alpha\|x_{n+1}-x_n\| \quad \text{since R non-expansive and } \alpha>0,\\ &\leq \|x_{n+1}-x_n\|,\\ &= \|\mathrm{T}x_n-x_n\|.\end{aligned}$$

We have $\|\mathrm{T}x_{n+1}-x_{n+1}\| \leq \|\mathrm{T}x_n-x_n\|$. Consequently, the sequence $(\mathrm{T}x_n-x_n)_{n\in\mathbb{N}}$ converges. Now we want to show that $\lim_{n\to+\infty}\|\mathrm{T}x_n-x_n\|=0$, because this, according to the definition of the convergence, implies that $(Tx_n-x_n)\to 0$.
Starting from Fejér equations with :

$$\begin{aligned}&\|x_{n+1}-\widehat{x}\|^2 \leq \|x_n-\widehat{x}\|^2-\frac{1-\alpha}{\alpha}\|\mathrm{T}x_n-x_n\|^2,\\ \Leftrightarrow&\frac{1-\alpha}{\alpha}\|\mathrm{T}x_n-x_n\|^2 \leq \|x_n-\widehat{x}\|^2-\|x_{n+1}-\widehat{x}\|^2,\\ \Leftrightarrow&\sum_{n\in\mathbb{N}}\frac{1-\alpha}{\alpha}\|\mathrm{T}x_n-x_n\|^2 \leq \sum_{n\in\mathbb{N}}\left[\|x_n-\widehat{x}\|^2-\|x_{n+1}-\widehat{x}\|^2\right],\\ \Leftrightarrow&\sum_{n\in\mathbb{N}}\frac{1-\alpha}{\alpha}\|\mathrm{T}x_n-x_n\|^2 \leq \|x_0-\widehat{x}\|^2+\|x_1-\widehat{x}\|^2+\cdots-\|x_1-\widehat{x}\|^2-\ldots,\\ \Leftrightarrow&\sum_{n\in\mathbb{N}}\frac{1-\alpha}{\alpha}\|\mathrm{T}x_n-x_n\|^2 \leq \|x_0-\widehat{x}\|^2-\|x_{+\infty}-\widehat{x}\|^2,\\ \Leftrightarrow&\sum_{n\in\mathbb{N}}\frac{1-\alpha}{\alpha}\|\mathrm{T}x_n-x_n\|^2 \leq \|x_0-\widehat{x}\|^2<+\infty.\end{aligned}$$

Since $\sum_{n\in\mathbb{N}} \frac{1-\alpha}{\alpha} = +\infty$ and $\sum_{n\in\mathbb{N}} \frac{1-\alpha}{\alpha}\|\mathrm{T}x_n - x_n\|^2 < +\infty$, then $\lim_{n\to+\infty}\|\mathrm{T}x_n - x_n\| = 0$. We conclude that $(Tx_n - x_n) \to 0$.

3. We proved that the sequence $(x_n)_{n\in\mathbb{N}}$ is Fejér-monotone with respect to Fix T. We also proved that $\mathrm{T}x_n - x_n \to 0$.
   We assume that a sub-sequence $(x_{n_k})_{k\in\mathbb{N}})$ converges to a point $\widetilde{x}$. Since T is a contraction and $(Tx_n - x_n) \to 0$, according to the demiclosedness principle, $\widetilde{x} \in$ Fix T.
   This means that the sequential clustering point of $(x_n)_{n\in\mathbb{N}}$ are in FixT. Finally, since $(x_n)_{n\in\mathbb{N}}$ is Fejér-monotone with respect to FixT, in conclusion $(x_n)_{n\in\mathbb{N}}$ converges weakly to $\widetilde{x} \in$ FixT.

A.3.1 Preuve du théorème 1.3.3

Démonstration : *(Fin page 202)* Cet algorithme est un cas spécifique de l'algorithme Variable-Metric Forward-Backward présenté dans [Combettes and Vũ, 2014], et dans le cas où $\lambda^{[t]} = 1$ et $\gamma^{[t]} = \gamma$ fixé.

Posons pour plus de simplicité $\mathbf{A} = \partial\mathcal{G}$ et $\mathbf{B} = \nabla\mathcal{F}$. Posons également $\mathbf{A}^{[t]} = \gamma\mathbf{P}^{[t]}\mathbf{A}$ et $\mathbf{B}^{[t]} = \gamma\mathbf{P}^{[t]}\mathbf{B}$ et la suite

$$\boldsymbol{x}^{[t+1]} = J_{\mathbf{A}^{[t]}}(\boldsymbol{x}^{[t]} - \mathbf{B}^{[t]}\boldsymbol{x}^{[t]}), \tag{A.7}$$

où $J_{\mathbf{A}^{[t]}}(\cdot) = \mathrm{prox}^{\mathbf{P}^{[t]}}_{\gamma\mathcal{G}}(\cdot)$.

Lorsque pour tout $\mathbf{P}^{[t]} = \mathbf{P}$, c'est-à-dire quand la métrique est fixe, la preuve de convergence est simple et reprend les mêmes étapes que la preuve du théorème de Krasnoselskii'Man A.2.1, présenté dans l'annexe A. Dans un tel cas, le but est de montrer que la suite $\boldsymbol{x}^{[t]})_{t\geq 0}$ tend vers un point fixe de $\mathrm{prox}^{\mathbf{P}^{[t]}}_{\gamma\mathcal{G}}(\cdot)$. La difficulté ici est que $\mathbf{P}^{[t]}$ varie à chaque itération $t \geq 0$. Le but de cette preuve est donc de montrer que la suite $\boldsymbol{x}^{[t]})_{t\geq 0}$ tend vers un point $\overline{\overline{\boldsymbol{x}}} \in \mathbf{zer}(\mathbf{A}+\mathbf{B})$.

Soit $\overline{\overline{\boldsymbol{x}}} \in \mathcal{H}$ un point fixe de la suite A.7, c'est-à-dire tel que pour $t \in \mathbb{N}$ on ait $\overline{\overline{\boldsymbol{x}}} = J_{\mathbf{A}^{[t]}}(\overline{\overline{\boldsymbol{x}}} - \mathbf{B}^{[t]}\overline{\overline{\boldsymbol{x}}})$, alors on a $\overline{\overline{\boldsymbol{x}}} \in \mathrm{zer}(\mathbf{A}+\mathbf{B})$. En effet :

$$\begin{aligned}
&0 \in (\mathbf{A}+\mathbf{B})\overline{\overline{\boldsymbol{x}}}\\
\Leftrightarrow &-\mathbf{B}\overline{\overline{\boldsymbol{x}}} \in \mathbf{A}\overline{\overline{\boldsymbol{x}}}\\
\Leftrightarrow &-\gamma\mathbf{B}\overline{\overline{\boldsymbol{x}}} \in \gamma\mathbf{A}\overline{\overline{\boldsymbol{x}}}\\
\Leftrightarrow &\mathbf{P}^{[t]^{-1}}\overline{\overline{\boldsymbol{x}}} - \gamma B\overline{\overline{\boldsymbol{x}}} \in \mathbf{P}^{[t]^{-1}}\overline{\overline{\boldsymbol{x}}} + \gamma\mathbf{A}\overline{\overline{\boldsymbol{x}}}\\
\Leftrightarrow &\overline{\overline{\boldsymbol{x}}} - \gamma\mathbf{P}^{[t]}\mathbf{B}\overline{\overline{\boldsymbol{x}}} \in \overline{\overline{\boldsymbol{x}}} + \gamma\mathbf{P}^{[t]}\mathbf{A}\overline{\overline{\boldsymbol{x}}}\\
\Leftrightarrow &(\mathbf{Id} - \gamma\mathbf{P}^{[t]}\mathbf{B})\overline{\overline{\boldsymbol{x}}} \in (\mathbf{Id} + \gamma\mathbf{P}^{[t]}\mathbf{A})\overline{\overline{\boldsymbol{x}}}\\
\Leftrightarrow &\overline{\overline{\boldsymbol{x}}} = (\mathbf{Id} + \mathbf{A}^{[t]})^{-1}(\mathbf{Id} - \mathbf{B}^{[t]})\overline{\overline{\boldsymbol{x}}}\\
\Leftrightarrow &\overline{\overline{\boldsymbol{x}}} = J_{\mathbf{A}^{[t]}}(\overline{\overline{\boldsymbol{x}}} - \mathbf{B}^{[t]}\overline{\overline{\boldsymbol{x}}})
\end{aligned}$$

L'idée de la preuve est donc de montrer que la suite représentée par l'équation A.7 tend vers ce point fixe, c'est-à-dire $(\boldsymbol{x}^{[t]})_{t\in\mathbb{N}} \rightharpoonup \overline{\overline{\boldsymbol{x}}}$ avec $\overline{\overline{\boldsymbol{x}}} \in \mathrm{zer}(\mathbf{A}+\mathbf{B})$.

La difficulté de cette preuve est que l'opérateur $\mathbf{T}^{[t]} = (\mathbf{Id} + \mathbf{A}^{[t]})^{-1}(\mathbf{Id} - \mathbf{B}^{[t]})$ n'est pas fixe à chaque itérations. Dans le cas où l'opérateur $\mathbf{T}$ est fixe, pour prouver la convergence, il suffit

d'appliquer les mêmes étapes que pour les preuves vues dans les sections 1.3.3. Pour plus de clarté dans cette preuve, nous reprenons la même structure.

1. Dans la preuve de convergence de l'algorithme Forwars-Backward, nous avons montré la Féjer-monotonie de $\boldsymbol{x}^{[t]} \in \mathcal{H}$. Cependant, dans le cas du changement de métrique, il est nécessaire d'introduire une autre définition de la Fejér-monotonie prenant en compte ce changement de métrique. De ce fait nous utilisons la définition de variable-métrique Féjer-monotonie.
   D'après la définition de l'ordre de Loewner (A.2), on a que $\|\boldsymbol{x}\|_{\mathbf{P}^{[t+1]^{-1}}} \leq (1+\eta^{[t]})\|\boldsymbol{x}\|_{\mathbf{P}^{[t]^{-1}}}$. Par conséquence, on a premièrement que :

$$\|\boldsymbol{x}^{[t+1]} - \overline{\overline{\boldsymbol{x}}}\|^2_{\mathbf{P}^{[t+1]^{-1}}} \leq (1+\eta^{[t]})\|\boldsymbol{x}^{[t+1]} - \overline{\overline{\boldsymbol{x}}}\|^2_{\mathbf{P}^{[t]^{-1}}}. \tag{A.8}$$

   On a alors

$$\begin{aligned}
&\|\boldsymbol{x}^{[t+1]} - \overline{\overline{\boldsymbol{x}}}\|^2_{\mathbf{P}^{[t]^{-1}}} \\
=&\|J_{\mathbf{A}^{[t]}}(\boldsymbol{x}^{[t]} - \mathbf{B}^{[t]}\boldsymbol{x}^{[t]}) - J_{\mathbf{A}^{[t]}}(\overline{\overline{\boldsymbol{x}}} - \mathbf{B}^{[t]}\overline{\overline{\boldsymbol{x}}})\|^2_{\mathbf{P}^{[t]^{-1}}} \\
\leq&\|\boldsymbol{x}^{[t]} - \mathbf{B}^{[t]}\boldsymbol{x}^{[t]} - (\overline{\overline{\boldsymbol{x}}} - \mathbf{B}^{[t]}\overline{\overline{\boldsymbol{x}}})\|^2_{\mathbf{P}^{[t]^{-1}}} \\
&- \|\boldsymbol{x}^{[t]} - \mathbf{B}^{[t]}\boldsymbol{x}^{[t]} - (\overline{\overline{\boldsymbol{x}}} - \mathbf{B}^{[t]}\overline{\overline{\boldsymbol{x}}}) - J_{\mathbf{A}^{[t]}}(\boldsymbol{x}^{[t]} - \mathbf{B}^{[t]}\boldsymbol{x}^{[t]}) + J_{\mathbf{A}^{[t]}}(\overline{\overline{\boldsymbol{x}}} - \mathbf{B}^{[t]}\overline{\overline{\boldsymbol{x}}})\|^2_{\mathbf{P}^{[t]^{-1}}} \\
=&\|\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}} - \mathbf{B}^{[t]}(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\|^2_{\mathbf{P}^{[t]^{-1}}} - \|\boldsymbol{x}^{[t]} - \boldsymbol{x}^{[t+1]} - \mathbf{B}^{[t]}(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\|^2_{\mathbf{P}^{[t]^{-1}}} \\
=&\|\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}}\|^2_{\mathbf{P}^{[t]^{-1}}} - 2\langle \boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}}, \mathbf{B}^{[t]}(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\rangle_{\mathbf{P}^{[t]^{-1}}} + \|\mathbf{B}^{[t]}(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}}\|^2_{\mathbf{P}^{[t]^{-1}}} \\
&- \|\boldsymbol{x}^{[t]} - \boldsymbol{x}^{[t+1]} - \mathbf{B}^{[t]}(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\|^2_{\mathbf{P}^{[t]^{-1}}} \\
=&\|\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}}\|^2_{\mathbf{P}^{[t]^{-1}}} - 2\gamma\langle \boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}}, B(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\rangle + \|\mathbf{B}^{[t]}(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\|^2_{\mathbf{P}^{[t]^{-1}}} \\
&- \|\boldsymbol{x}^{[t]} - \boldsymbol{x}^{[t+1]} - \mathbf{B}^{[t]}(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\|^2_{\mathbf{P}^{[t]^{-1}}}
\end{aligned}$$

   Alors, du fait de $\|\mathbf{B}^{[t]}(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\|^2_{\mathbf{P}^{[t]^{-1}}} \leq \gamma^2\|\mathbf{P}^{[t]}\|\,\|B(\boldsymbol{x}^{[t]} - x)\|^2$ et de la β^{-1}-cocoercivité de $\mathbf{B}$, on a :

$$\begin{aligned}
&\|\boldsymbol{x}^{[t+1]} - \overline{\overline{\boldsymbol{x}}}\|^2_{\mathbf{P}^{[t]^{-1}}} \\
\leq&\|\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}}\|^2_{\mathbf{P}^{[t]^{-1}}} - 2\frac{\gamma}{\beta}\|B(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\|^2 + \gamma\|\mathbf{P}^{[t]}\|\,\|B(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\|^2 \\
&- \|\boldsymbol{x}^{[t]} - \boldsymbol{x}^{[t+1]} - \mathbf{B}^{[t]}(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\|^2_{\mathbf{P}^{[t]^{-1}}} \\
=&\|\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}}\|^2_{\mathbf{P}^{[t]^{-1}}} - \gamma\left(\frac{2}{\beta} - \gamma\|\mathbf{P}^{[t]}\|\right)\|B(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\|^2 \\
&- \|\boldsymbol{x}^{[t]} - \boldsymbol{x}^{[t+1]} - \mathbf{B}^{[t]}(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\|^2_{\mathbf{P}^{[t]^{-1}}}
\end{aligned}$$

   Posons maintenant

$$\delta^{[t]} = \gamma\left(\frac{2}{\beta} - \gamma\|\mathbf{P}^{[t]}\|\right) \tag{A.9}$$

en supposant que $\delta^{[t]} \geq 0$, tel que $\sum^{[t]} \delta^{[t]} = +\infty$. La positivité implique une la condition suivante sur la relation entre γ et $\|\mathbf{P}^{[t]}\|$:

$$0 \leq \gamma \leq \frac{2}{\beta \|\mathbf{P}^{[t]}\|}. \tag{A.10}$$

Remarque : Dans la littérature, les auteurs utilisent $\sup_{t\in\mathbb{N}}(\|\mathbf{P}^{[t]}\|)$ pour fixer une borne maximale à γ. Pour être certains que la conditions est respectées, ils réduisent l'intervalle dans lequel se situe γ d'un petit ε, et utilisent cette valeur pour avoir une borne minimale à $\delta^{[t]}$ comme ε^2.

Pour une telle valeur de δ et en remplaçant $\|\boldsymbol{x}^{[t+1]} - \overline{\overline{\boldsymbol{x}}}\|^2_{\mathbf{P}^{[t]-1}}$ dans (A.8), on a :

$$\begin{aligned}
&\|\boldsymbol{x}^{[t+1]} - \overline{\overline{\boldsymbol{x}}}\|^2_{\mathbf{P}^{[t+1]-1}} \\
\leq &(1+\eta^{[t]})\left[\|\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}}\|^2_{\mathbf{P}^{[t]-1}} - \delta^{[t]}\|B(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\|^2 - \|\boldsymbol{x}^{[t]} - \boldsymbol{x}^{[t+1]} - \mathbf{B}^{[t]}(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\|^2_{\mathbf{P}^{[t]-1}}\right], \\
\leq &(1+\eta^{[t]})\|\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}}\|^2_{\mathbf{P}^{[t]-1}} - \underbrace{\left[\delta^{[t]}\|B(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\|^2 + \|\boldsymbol{x}^{[t]} - \boldsymbol{x}^{[t+1]} + \mathbf{B}^{[t]}(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\|^2_{\mathbf{P}^{[t]-1}}\right]}_{\geq 0}.
\end{aligned}$$

Alors, en utilisant [Combettes and Vũ, 2013, Proposition 3.2], on a que la suite $\left(\|\boldsymbol{x}^{[t+1]} - \overline{\overline{\boldsymbol{x}}}\|^2_{\mathbf{P}^{[t]-1}}\right)_{t\in\mathbb{N}}$ converge.

2. Maintenant on veut montrer que $\|\boldsymbol{x}^{[t+1]} - \boldsymbol{x}^{[t]}\| \underset{t\to+\infty}{\to} 0$. Pour ce faire, nous devons montrer que $\sum^{[t]} \|\boldsymbol{x}^{[t+1]} - \boldsymbol{x}^{[t]}\| < +\infty$. On a :

$$\begin{aligned}
\|\boldsymbol{x}^{[t+1]} - \boldsymbol{x}^{[t]}\| &= \|\boldsymbol{x}^{[t+1]} - \boldsymbol{x}^{[t]} + \mathbf{B}^{[t]}\boldsymbol{x}^{[t]} - \mathbf{B}^{[t]}\boldsymbol{x}^{[t]} + \mathbf{B}^{[t]}\overline{\overline{\boldsymbol{x}}} - \mathbf{B}^{[t]}\overline{\overline{\boldsymbol{x}}}\| \\
&\leq \|\mathbf{B}^{[t]}(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\| + \|\boldsymbol{x}^{[t]} - \boldsymbol{x}^{[t+1]} + \mathbf{B}^{[t]}(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\|, \\
&\leq \gamma\|\mathbf{P}^{[t]}\|\,\|\mathbf{B}^{[t]}(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\| + \|\boldsymbol{x}^{[t]} - \boldsymbol{x}^{[t+1]} + \mathbf{B}^{[t]}(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\|, \\
&\leq (\frac{2}{\beta} - \frac{\delta^{[t]}}{\gamma})\|B(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\| + \|\boldsymbol{x}^{[t]} - \boldsymbol{x}^{[t+1]} + \mathbf{B}^{[t]}(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\|.
\end{aligned}$$

On reconnait deux éléments de l'équation de la quasi-Fejér monotonie.

(a) D'une part on a alors :

$$\delta^{[t]}\|B(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\|^2 \leq \|\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}}\|^2_{\mathbf{P}^{[t]-1}} - \|\boldsymbol{x}^{[t+1]} - \overline{\overline{\boldsymbol{x}}}\|^2_{\mathbf{P}^{[t]-1}} - \|\boldsymbol{x}^{[t]} - \boldsymbol{x}^{[t+1]} - \mathbf{B}^{[t]}(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\|^2_{\mathbf{P}^{[t]-1}}. \tag{A.11}$$

Alors :

$$
\begin{aligned}
&\sum^{[t]} \delta^{[t]} \|B(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\|^2 \\
&\leq \sum^{[t]} \left(\|\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}}\|^2_{\mathbf{P}^{[t]^{-1}}} - \|\boldsymbol{x}^{[t+1]} - \overline{\overline{\boldsymbol{x}}}\|^2_{\mathbf{P}^{[t]^{-1}}} - \|\boldsymbol{x}^{[t]} - \boldsymbol{x}^{[t+1]} - \mathbf{B}^{[t]}(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\|^2_{\mathbf{P}^{[t]^{-1}}} \right) \\
&= \|x_0 - \overline{\overline{\boldsymbol{x}}}\|^2_{\mathbf{P}^{[t]^{-1}}} - \left[\|x_{+\infty} - \overline{\overline{\boldsymbol{x}}}\|^2_{\mathbf{P}^{[t]^{-1}}} + \sum^{[t]} \|\boldsymbol{x}^{[t]} - \boldsymbol{x}^{[t+1]} - \mathbf{B}^{[t]}(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\|^2_{\mathbf{P}^{[t]^{-1}}} \right], \\
&\leq \|x_0 - \overline{\overline{\boldsymbol{x}}}\|^2_{\mathbf{P}^{[t]^{-1}}} < +\infty.
\end{aligned}
$$

On a donc $\sum^{[t]} \delta^{[t]} \|B(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\|^2 < +\infty$ et donc $\sum^{[t]} \sqrt{\delta^{[t]}} \|B(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\| < +\infty$. Supposons que $\sum^{[t]} \delta^{[t]} = \sum^{[t]} \sqrt{\delta^{[t]}} = +\infty$, par conséquence $\lim_{t \to +\infty} \|B(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\| \to 0$.

(b) D'autre part on a :

$$
\|\boldsymbol{x}^{[t]} - \boldsymbol{x}^{[t+1]} - \mathbf{B}^{[t]}(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\|^2_{\mathbf{P}^{[t]^{-1}}} \leq \|\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}}\|^2_{\mathbf{P}^{[t]^{-1}}} - \delta^{[t]} \|B(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\|^2 - \|\boldsymbol{x}^{[t+1]} - \overline{\overline{\boldsymbol{x}}}\|^2_{\mathbf{P}^{[t]^{-1}}}. \tag{A.12}
$$

Alors :

$$
\begin{aligned}
&\sum \|\boldsymbol{x}^{[t]} - \boldsymbol{x}^{[t+1]} - \mathbf{B}^{[t]}(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\|^2_{\mathbf{P}^{[t]^{-1}}} \\
&\leq \sum \left(\|\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}}\|^2_{\mathbf{P}^{[t]^{-1}}} - \delta^{[t]} \|B(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\|^2 - \|\boldsymbol{x}^{[t+1]} - \overline{\overline{\boldsymbol{x}}}\|^2_{\mathbf{P}^{[t]^{-1}}} \right), \\
&\leq \|\boldsymbol{x}^{[0]} - \overline{\overline{\boldsymbol{x}}}\|^2_{\mathbf{P}^{[t]^{-1}}} - \left[\|x_{+\infty} - \overline{\overline{\boldsymbol{x}}}\|^2_{\mathbf{P}^{[t]^{-1}}} + \sum^{[t]} \delta^{[t]} \|B(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\|^2 \right], \\
&\leq \|\boldsymbol{x}^{[0]} - \overline{\overline{\boldsymbol{x}}}\|^2_{\mathbf{P}^{[t]^{-1}}} < +\infty.
\end{aligned}
$$

Cependant on veut montrer que $\sum^{[t]} \|\boldsymbol{x}^{[t]} - \boldsymbol{x}^{[t+1]} + \mathbf{B}^{[t]}(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\|$ est fini et non $\sum \|\boldsymbol{x}^{[t]} - \boldsymbol{x}^{[t+1]} + \mathbf{B}^{[t]}(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\|^2_{\mathbf{P}^{[t]^{-1}}}$. En revanche on a :

$$
\begin{aligned}
&\sum^{[t]} \|\boldsymbol{x}^{[t]} - \boldsymbol{x}^{[t+1]} + \mathbf{B}^{[t]}(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\| \\
&\leq \sum^{[t]} \|\sqrt{\mathbf{P}^{[t]}}\| \, \|\boldsymbol{x}^{[t]} - \boldsymbol{x}^{[t+1]} + \mathbf{B}^{[t]}(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\|_{\mathbf{P}^{[t]^{-1}}} \\
&\leq \sup^{[t]} \|\sqrt{\mathbf{P}^{[t]}}\| \sum^{[t]} \|\boldsymbol{x}^{[t]} - \boldsymbol{x}^{[t+1]} + \mathbf{B}^{[t]}(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\|_{\mathbf{P}^{[t]^{-1}}} < +\infty,
\end{aligned}
$$

donc $\|\boldsymbol{x}^{[t]} - \boldsymbol{x}^{[t+1]} + \mathbf{B}^{[t]}(\boldsymbol{x}^{[t]} - \overline{\overline{\boldsymbol{x}}})\| \underset{t \to +\infty}{\to} 0$.

Au final on a :

$$\sum^{[t]} \|\boldsymbol{x}^{[t+1]} - \boldsymbol{x}^{[t]}\| \leq \sum^{[t]} \gamma \|\mathbf{P}^{[t]}\| \, \|B(\boldsymbol{x}^{[t]} - \overline{\boldsymbol{x}})\| + \sum^{[t]} \|\boldsymbol{x}^{[t]} - \boldsymbol{x}^{[t+1]} + \mathbf{B}^{[t]}(\boldsymbol{x}^{[t]} - \overline{\boldsymbol{x}})\| < +\infty.$$

On a donc $\|\boldsymbol{x}^{[t+1]} - \boldsymbol{x}^{[t]}\| \underset{t\to+\infty}{\to} 0$, ce qui implique que $\boldsymbol{x}^{[t+1]} - \boldsymbol{x}^{[t]}$ converge fortement vers 0.

3. On a montré que la suite $\boldsymbol{x}^{[t]}$ est quasi-Fejér-monotone par rapport à $\overline{\boldsymbol{x}} \in \operatorname{zer}(\mathbf{A}+\mathbf{B})$. Nous avons aussi montré que $\boldsymbol{x}^{[t+1]} - \boldsymbol{x}^{[t]} = \mathbf{T}^{[t]}\boldsymbol{x}^{[t]} - \boldsymbol{x}^{[t]} \to 0$.

   Comme énoncé au début de cette preuve, quand $\mathbf{T}^{[t]} = \mathbf{T}$ pour tout $t \in \mathbb{N}$, alors la fin de la preuve de convergence est identique à celle de Forward-Backward. Pour rappelle, dans cette preuve de convergence, on suppose qu'il existe une sous-suite $(\boldsymbol{x}^{[t_k]})_{k\in\mathbb{N}}$ qui converge vers le point fixe $\overline{\boldsymbol{x}}$. Alors, comme $\mathbf{T}$ est une contraction et que $(\mathbf{T}\boldsymbol{x}^{[t]} - \boldsymbol{x}^{[t]}) \to 0$, d'après le principe de demi-fermeture A.1.7 on a $\overline{\boldsymbol{x}} \in \operatorname{Fix}\mathbf{T}$.

   Cela signifie que tous les ensembles de points de la suite $(\boldsymbol{x}^{[t]})_{t\in\mathbb{N}}$ sont dans Fix$\mathbf{T}$. Au final, comme $(\boldsymbol{x}^{[t]})_{t\in\mathbb{N}}$ est Fejér-monotone par rapport à Fix$\mathbf{T}$, on conclut que $(\boldsymbol{x}^{[t]})_{t\in\mathbb{N}}$ converge faiblement vers $\overline{\boldsymbol{x}} \in$ Fix$\mathbf{T}$.

   Dans le cas de la métrique variable, le principe de demi-fermeture ne peut pas être utilisé. Cependant, dans [Combettes and Vũ, 2014], les auteurs utilisent le fait que les opérateurs $\mathbf{A}$ et $\mathbf{B}$ sont maximale montone. En effet, cela implique que leur graphe est fermé au sens des suites dans $\mathcal{H}\times\mathcal{H}$ ([Bauschke and Combettes, 2011], Proposition 20.33). Cela signifie que pour tout couple de suites $(\boldsymbol{a}^{[t]}, \boldsymbol{b}^{[t]})_{t\in\mathbb{N}}$ dans les graphes respectifs de $\mathbf{A}$ et $\mathbf{B}$, et tout couple respectif $(\boldsymbol{a}, \boldsymbol{b}) \in \mathcal{H}\times\mathcal{H}$, si $\boldsymbol{a}^{[t]} \to \boldsymbol{a}$ and $\boldsymbol{b}^{[t]} \to \boldsymbol{b}$, alors le couple $(\boldsymbol{a}, \boldsymbol{b})$ sont dans les graphes respectifs de $\mathbf{A}$ et $\mathbf{B}$. Cette méthode est similaire au principe de demi-fermeture. En utilisant la monotonie maximale de $\mathbf{A}$ et de $\mathbf{B}$, on prouve que $\widetilde{x} \in \operatorname{zer}(\mathbf{A}+\mathbf{B})$.

Ceci conclue la preuve du théorème 1.3.3.

□

A.4 Convergence du primal-dual

Preuve de la proposition 1.3.4

Démonstration : *(Fin page 206)* Cette proposition est un résultat directe de [Vũ, 2015, Theorem 3.1]. Nous énonçons ici les grandes lignes de la preuve.

Une itération de l'algorithme VMPD peut être vu comme une itération de l'algorithme VMFB :

$$\boldsymbol{z}^{[t+1]} = J_{\mathbf{P}^{[t]}\partial g^*}\left(\boldsymbol{z}^{[t]} - \mathbf{P}^{[t]}\nabla \boldsymbol{Phi}(\boldsymbol{z}^{[t]})\right) \tag{A.13}$$

where :

$$\boldsymbol{z}^{[t]} = \begin{pmatrix}\boldsymbol{x}^{[t]}\\ \boldsymbol{y}^{[t]}\end{pmatrix}, \quad \partial \boldsymbol{g}^* = \begin{pmatrix}\partial \iota_{\mathcal{C}} & \mathbf{G}^*\\ -\mathbf{G} & \partial g^*\end{pmatrix}, \quad \nabla\boldsymbol{\Phi} = \begin{pmatrix}\nabla\Phi & 0\\ 0 & 0\end{pmatrix} \quad \text{and} \quad \mathbf{P}^{[t]} = \begin{pmatrix}\mathbf{U}^{[t]^{-1}} & -\mathbf{G}^*\\ -\mathbf{G} & \sigma^{[t]^{-1}}\end{pmatrix}^{-1}, \tag{A.14}$$

Montrer que l'algorithme 4 converge consiste donc à montrer que $\mathbf{P}^{[t]}$ vérifie bien les conditions du théoreme 1.3.3 et que ∂g et $\nabla\mathbf{\Phi}$ sont maximales monotones. En effet, on a :

$$\begin{cases} x^{[t+1]} = \left(\mathbf{Id} + \mathbf{U}^{[t]}\partial\iota_{\mathcal{C}}\right)^{-1}\left(x^{[t]} - \mathbf{U}^{[t]}\left(\nabla\Phi(x^{[t]}) - \mathbf{G}^*y^{[t]}\right)\right), \\ y^{[t+1]} = \left(\mathbf{Id} + \sigma^{[t]}\partial g^*\right)^{-1}\left(y^{[t]} + \sigma^{[t]}G\left(2x^{[t+1]} - x^{[t]}\right)\right) \end{cases} \tag{A.15}$$

$$\Leftrightarrow \begin{cases} \left(\mathbf{Id} + \mathbf{U}^{[t]}\partial\iota_{\mathcal{C}}\right)x^{[t+1]} \ni x^{[t]} - \mathbf{U}^{[t]}\left(\nabla\Phi(x^{[t]}) - \mathbf{G}^*y^{[t]}\right), \\ \left(\mathbf{Id} + \sigma^{[t]}\partial g^*\right)y^{[t+1]} \ni y^{[t]} + \sigma^{[t]}\mathbf{G}\left(2x^{[t+1]} - x^{[t]}\right) \end{cases} \tag{A.16}$$

$$\Leftrightarrow \begin{cases} x^{[t+1]} - \mathbf{U}^{[t]}\partial\iota_{\mathcal{C}}(x^{[t+1]}) \ni x^{[t]} - \mathbf{U}^{[t]}\left(\nabla\Phi(x^{[t]}) - \mathbf{G}^*y^{[t]}\right) \\ y^{[t+1]} + \sigma^{[t]}\partial g^*(y^{[t+1]}) \ni y^{[t]} + \sigma^{[t]}\mathbf{G}\left(2x^{[t+1]} - x^{[t]}\right) \end{cases} \tag{A.17}$$

$$\Leftrightarrow \begin{cases} \mathbf{U}^{[t]^{-1}}\left(x^{[t+1]} - x^{[t]}\right) = -\nabla\Phi(x^{[t]}) + \mathbf{G}^*y^{[t]}, \\ y^{[t+1]} + \sigma^{[t]}\partial g^*(y^{[t+1]}) \ni y^{[t]} + \sigma^{[t]}\mathbf{G}\left(2x^{[t+1]} - x^{[t]}\right) \end{cases} \tag{A.18}$$

$$\Leftrightarrow \begin{cases} -\nabla\Phi(x^{[t]}) = \mathbf{U}^{[t]^{-1}}\left(x^{[t+1]} - x^{[t]}\right) - \mathbf{G}^*y^{[t]}, \\ 0 \in y^{[t+1]} + \sigma^{[t]}\partial g^*(y^{[t+1]}) - y^{[t]} - \sigma^{[t]}\mathbf{G}\left(2x^{[t+1]} - x^{[t]}\right) \end{cases} \tag{A.19}$$

En divisant la seconde ligne par $\sigma^{[t]}$ et en faisant apparaître $G^*y^{[t+1]}$ dans la première ligne on a :

$$\Leftrightarrow \begin{cases} -\nabla\Phi(x^{[t]}) = \mathbf{U}^{[t]^{-1}}\left(x^{[t+1]} - x^{[t]}\right) - \mathbf{G}^*y^{[t]} + \mathbf{G}^*y^{[t+1]} - \mathbf{G}^*y^{[t+1]}, \\ 0 \in \sigma^{[t]^{-1}}\left(y^{[t+1]} - y^{[t]}\right) + \partial g^*(y^{[t+1]}) - \mathbf{G}x^{[t+1]} - \mathbf{G}\left(x^{[t+1]} - x^{[t]}\right) \end{cases} \tag{A.20}$$

Posons

$$\mathbf{M}^{[t]} = \begin{pmatrix} \mathbf{U}^{[t]^{-1}} & -\mathbf{G}^* \\ -\mathbf{G} & \sigma^{[t]^{-1}} \end{pmatrix} \tag{A.21}$$

D'après l'équation (A.14), les equations (A.20) sont équivalentes à :

$$\begin{aligned} -\nabla\mathbf{\Phi}z^{[t]} &\in \partial g z^{[t+1]} + \mathbf{M}^{[t]}\left(z^{[t+1]} - z^{[t]}\right) \\ \Leftrightarrow \mathbf{M}^{[t]}z^{[t]} - \nabla\mathbf{\Phi}z^{[t]} &\in \partial g z^{[t+1]} + \mathbf{M}^{[t]}z^{[t+1]}. \end{aligned} \tag{A.22}$$

Avant de pouvoir aller plus loin il faut s'assurer que $\mathbf{M}^{[t]}$ est bien inversible et donc que $\mathbf{P}^{[t]} = \mathbf{M}^{[t]^{-1}}$ existe. On sait par définition que la matrice $\mathbf{M}^{[t]}$ est auto-adjointe, car $\mathbf{U}^{[t]}$ est symétrique définie positive. De plus si $\mathbf{U}^{[t]}$ et $\sigma^{[t]}$ vérifient la condition (1.92), alors $\mathbf{M}^{[t]}$ est ρ-positive avec $\rho > 0$ (*i.e.* $\langle z, \mathbf{M}^{[t]}z\rangle \geq \rho\|z\|^2$ où $\langle z_1, z_2\rangle = \langle x_1, x_2\rangle + \langle y_1, y_2\rangle$).

En effet :

$$\langle z, \mathbf{M}^{[t]} z\rangle = \left\langle \begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} \mathbf{U}^{[t]^{-1}} & -\mathbf{G}^* \\ -\mathbf{G} & \sigma^{[t]^{-1}} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \right\rangle \tag{A.23}$$

$$= \left\langle \begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} \mathbf{U}^{[t]^{-1}} x - \mathbf{G}^* y \\ -\mathbf{G}x + \sigma^{[t]^{-1}} y \end{pmatrix} \right\rangle \tag{A.24}$$

$$= \langle x, \mathbf{U}^{[t]^{-1}} x\rangle + \langle y, \sigma^{[t]^{-1}} y\rangle - \langle x, \mathbf{G}^* y\rangle - \langle y, \mathbf{G}x\rangle, \tag{A.25}$$

$$= \langle x, \mathbf{U}^{[t]^{-1}} x\rangle + \langle y, \sigma^{[t]^{-1}} y\rangle - 2\langle y, \mathbf{G}x\rangle. \tag{A.26}$$

En faisant apparaître un réel $a \in]1, \|\sigma^{[t]^{\frac{1}{2}}} \mathbf{G} \mathbf{U}^{[t]^{\frac{1}{2}}}\|^{-2}$, afin d'empêcher le terme $\langle y, \sigma^{t]^{-1}} y\rangle$ de disparaître, on a :

$$2\langle \mathbf{G}x, y\rangle = \langle a^{\frac{1}{2}} \sigma^{[t]^{\frac{1}{2}}} \mathbf{G} \mathbf{U}^{[t]^{\frac{1}{2}}} \mathbf{U}^{[t]^{-\frac{1}{2}}} x, a^{-\frac{1}{2}} \sigma^{[t]^{-\frac{1}{2}}} y\rangle, \tag{A.27}$$

$$\leq a \|\sigma^{[t]^{\frac{1}{2}}} \mathbf{G} \mathbf{U}^{[t]^{\frac{1}{2}}} \mathbf{U}^{[t]^{-\frac{1}{2}}} x\|^2 + a^{-1} \|\sigma^{[t]^{-\frac{1}{2}}} y\|^2, \tag{A.28}$$

$$\leq a \|\sigma^{[t]^{-\frac{1}{2}}} \mathbf{G} \mathbf{U}^{[t]^{\frac{1}{2}}}\|^2 \|\mathbf{U}^{[t]^{-\frac{1}{2}}} x\|^2 + a^{-1} \|\sigma^{[t]^{-\frac{1}{2}}} y\|^2, \tag{A.29}$$

$$\leq a \|\sigma^{[t]^{\frac{1}{2}}} \mathbf{G} \mathbf{U}^{[t]^{\frac{1}{2}}}\|^2 \langle x, \mathbf{U}^{[t]^{-\frac{1}{2}}} x\rangle + a^{-1} \langle y, \sigma^{[t]^{-1}} y\rangle. \tag{A.30}$$

Alors, en remplaçant $2\langle y, \mathbf{G}x\rangle$ dans (A.26) par la borne supérieure (A.30), on a :

$$\langle z, \mathbf{M}^{[t]} z\rangle \geq \left(1 - a\|\sigma^{[t]^{\frac{1}{2}}} \mathbf{G} \mathbf{U}^{[t]^{\frac{1}{2}}}\|^2\right) \langle x, \mathbf{U}^{[t]^{-1}} V x\rangle + \left(1 - a^{-1}\right) \langle y, \sigma^{[t]^{-1}} y\rangle. \tag{A.31}$$

Afin de simplifier le calcul, on choisis $a > 0$ tel que :

$$\begin{aligned} 1 - a\|\sigma^{[t]^{\frac{1}{2}}} \mathbf{G} \mathbf{U}^{[t]^{\frac{1}{2}}}\|^2 &= 1 - a^{-1} \\ \Rightarrow a^2 &= \left(\|\sigma^{[t]^{\frac{1}{2}}} \mathbf{G} \mathbf{U}^{[t]^{\frac{1}{2}}}\|^2\right)^{-1} \\ \Rightarrow a &= \|\sigma^{[t]^{\frac{1}{2}}} \mathbf{G} \mathbf{U}^{[t]^{\frac{1}{2}}}\|^{-1}. \end{aligned}$$

Finalement, en remplaçant a dans (A.31) on obtiens que :

$$\begin{aligned} \langle z, \mathbf{M}^{[t]} z\rangle &\geq \left[1 \|\sigma^{[t]^{\frac{1}{2}}} \mathbf{G} \mathbf{U}^{[t]^{\frac{1}{2}}}\|^{-1}\right] \left(\|\mathbf{U}^{[t]}\|^{-1} \|x\|^2 + \sigma^{[t]^{-1}} \|y\|^2\right), \\ &\geq \left[1 - \|\sigma^{[t]^{\frac{1}{2}}} \mathbf{G} \mathbf{U}^{[t]^{\frac{1}{2}}}\|^{-1}\right] \min\left\{\|\mathbf{U}^{[t]}\|^{-1}, \sigma^{[t]^{-1}}\right\} \left(\|x\|^2 + \|y\|^2\right), \\ &= \rho \|z\|^2. \end{aligned}$$

On a donc $\mathbf{M}^{[t]}$ est inversible on a donc :

$$z^{[t]} - \mathbf{M}^{[t]^{-1}} \nabla \mathbf{\Phi} z^{[t]} \in z^{[t+1]} + \mathbf{P}^{[t]^{-1}} \partial g z^{[t+1]}, \tag{A.32}$$

$$\Leftrightarrow \left(\mathbf{Id} - \mathbf{M}^{[t]^{-1}} \nabla \mathbf{\Phi}\right) z^{[t]} \in \left(\mathbf{Id} + \mathbf{M}^{[t]^{-1}} \partial g\right) z^{[t+1]}, \tag{A.33}$$

$$\Leftrightarrow z^{[t+1]} = \left(\mathbf{Id} + \mathbf{P}^{[t]} \partial g\right)^{-1} \left(\mathbf{Id} - \mathbf{P}^{[t]} \nabla \mathbf{\Phi}\right) z^{[t]} \tag{A.34}$$

$$\Leftrightarrow J_{\mathbf{P}^{[t]} \partial g^*} \left(z^{[t]} - \mathbf{P}^{[t]} \nabla \mathbf{\Phi}(z^{[t]})\right) \tag{A.35}$$

On retrouve bien une itération de l'algorithme VMFB dans le cas spécifique où $\gamma \mathbf{P}^{[t]} = \mathbf{M}^{[t]^{-1}}$. Alors, il reste à montrer que $\nabla \mathbf{\Phi}$ est κ -cocoercitive, avec $\kappa > 0$ et donc maximale monotone, et que $\partial \mathbf{g}$ est maximale monotone.

κ-coercitivité et monotonie maximale de $\nabla \mathbf{\Phi}$: On montre tout d'abord que $\nabla \mathbf{\Phi}$ est κ-cocoercitive, car cela implique que $\nabla \mathbf{\Phi}$ est maximale monotone.

Posons $z_1 \in \nabla \mathbf{\Phi}(z_1')$ et $z_2 \in \nabla \mathbf{f}(z_2')$, alors comme $\nabla \mathbf{\Phi}$ est différentiable avec $\nabla \Phi$ β-Lipschitz et β^{-1}-cocoercitive, et donc maximale monotone, on a :

$$\begin{aligned}
\langle z_1 - z_2, z_1' - z_2' \rangle &= \langle z_1 - z_2, \mathbf{B}(z_1) - \mathbf{B}(z_2) \rangle && \text{(A.36)} \\
&= \langle x_1 - x_2, \nabla \Phi(x_1) - \nabla \Phi(x_2) \rangle && \text{(A.37)} \\
&\geq \beta^{-1} \|\nabla \Phi(x_1) - \nabla \Phi(x_2)\|^2 \quad \text{car } \Phi \text{ est cocoercitive} && \text{(A.38)} \\
&= \beta^{-1} \|\mathbf{B}(z_1) - \mathbf{B}(z_2)\|^2 && \text{(A.39)} \\
&\geq 0. && \text{(A.40)}
\end{aligned}$$

On conclu que $\nabla \mathbf{\Phi}$ est β^{-1}-cocoercitive et donc maximale monotone.

Maximale monotonie de $\partial \mathbf{g}$: L'opérateur $\partial \mathbf{g}$ peut être décomposé comme :

$$\partial \mathbf{g}^* = \begin{pmatrix} \partial \iota_{\mathcal{C}} & 0 \\ 0 & \partial g^* \end{pmatrix} + \begin{pmatrix} 0 & \mathbf{G}^* \\ -\mathbf{G} & 0 \end{pmatrix}.$$

D'une part,

$$\partial \mathbf{g}^* = \begin{pmatrix} \partial \iota_{\mathcal{C}} & 0 \\ 0 & \partial g^* \end{pmatrix},$$

comme $g \in \Gamma_0(\mathcal{K})$, alors ∂g est monotone maximale $\mathcal{K}$ [Bauschke and Combettes, 2011, Théorème 20.40]. De plus, d'après [Bauschke and Combettes, 2011, Corrolaire 16.24, Proposition 20.22] ∂g^* est monotone maximale sur $\mathcal{K}$, car $\partial g^* = (\partial g)^{-1}$ et $(\partial g)^{-1}$ est monotone maximale sur $\mathcal{K}$. Enfin, comme ∂g^* est monotone maximale, $\partial \mathbf{g}^*$ est aussi maximale monotone.

Dans le cas particulier où g est une somme de $g_\ell \in \Gamma_0(\mathcal{K}_\ell), \forall \ell = 1, \ldots, L$, alors $\partial g^* = \prod_\ell \partial g_\ell^*$ serait maximale monotone en utilisant le fait que $\prod_\ell \partial g_\ell^*$ est maximale monotone sur $\prod_\ell \mathcal{K}$ [Bauschke and Combettes, 2011, Proposition 20.2]).

D'autre part, en posant

$$\mathbf{G} = \begin{pmatrix} 0 & \mathbf{G}^* \\ -\mathbf{G} & 0 \end{pmatrix},$$

comme $\mathbf{G}^* = -\mathbf{G}$, l'opérateur $\mathbf{G}$ est maximale monotone sur $\mathcal{H}$ [Bauschke and Combettes, 2011, Exemple. 20.30] .

Enfin, comme $\mathrm{dom}\mathbf{G} = \mathcal{H}$, d'après [Bauschke and Combettes, 2011, Corrolaire 14.4 (i)] l'opérateur $\partial\mathbf{g} = \partial\mathbf{g}^* + \mathbf{G}$ est maximale monotone sur $\mathcal{H}$.

De plus, $\mathrm{dom}\nabla\iota_{\mathcal{C}} = \mathrm{dom}\partial\mathbf{g} = \mathcal{H}$, donc $\partial\mathbf{g} + \nabla\mathbf{\Phi}$ est maximale monotone.

Ceci conclue la preuve de la proposition 1.3.4.

□

Annexe B

Papiers de conférences et de colloque acceptés et papier A&A soumis

New inverse method for circumstellar environments reconstruction in polarimetry with the ESO/VLT-SPHERE IRDIS instrument

Laurence Denneulin, Maud Langlois, Éric Thiébaut

Laurence Denneulin, Maud Langlois, Éric Thiébaut, "New inverse method for circumstellar environments reconstruction in polarimetry with the ESO/VLT-SPHERE IRDIS instrument

," Proc. SPIE 10702, Ground-based and Airborne Instrumentation for Astronomy VII, 1070245 (8 July 2018); doi: 10.1117/12.2313901

SPIE.

Event: SPIE Astronomical Telescopes + Instrumentation, 2018, Austin, Texas, United States

New Inverse Method for Circumstellar Environments reconstruction in Polarimetry with the ESO/VLT-SPHERE IRDIS instrument.

Laurence Denneulin[a], Maud Langlois[a], and Éric Thiébaut[a]

[a]CRAL, UMR 5574, CNRS, Univesité Lyon 1, 9 avenue Charles André, 69561 Saint Genis Laval Cedex, France

ABSTRACT

The instrument IRDIS on ESO/VLT-SPHERE allows for observations in polarimetry of circumstellar disks in the near infrared. Since circumstellar disks light is partially linearly polarized by the reflection of the star light on its surface, the DPI mode (Dual Polarimetry Imaging) allows us to recover the intensity and the angle of polarization leading to morphology and dust size studies of these disks. We have developed a new method to reduce the IRDIS-DPI data based on an inverse approach method. This method is based on an optimization using the electric field (Jones Matrices) rather than the intensity (Mueller matrices) and relies on the inverse approach (i.e. fitting a model of the data to the observed dataset) which is significantly less biased and more efficient at minimizing instrumental artefacts. We describe, in this paper, this new method and we compare it to the state of the art methods. Using two IRDIS-DPI datasets we also demonstrate its ability to reconstruct polarized intensities maps and the angle of polarization.

Keywords: Polarimetry, SPHERE/IRDIS, Inverse Approach, circumstellar environments, Data reconstruction

1. INTRODUCTION

Polarimetric differential imaging (DPI) is a powerful technique to suppress the stellar light in scattered-light observations. This technique has been recently successfully applied to allow high contrast detection of several circumstellar disks. It uses the fact that while the light from the central source is largely unpolarized, scattering on the dust grains in the disk produces polarization. DPI allows one to probe the circumstellar environment very close to young stars and achieve the high contrast ratios required to detect the circumstellar disks they can host. With the help of extreme adaptive optics and coronography, the instrument ESO/VLT-SPHERE IRDIS allows very efficient detection of polarized light in the near infrared (NIR) at very small angular resolution and at close separation such as circumstellar disks. Due to the low luminosity of such disks and the stellar glare, the reconstruction of circumstellar disks remains a real challenge. First, their angular size is small requiring high angular resolution. The second problem is that the polarised light they reflect is weak in comparison with the unpolarised stellar light which is much higher and sets limits in the achievable contrast.

The Dual Polarization Imaging (DPI) mode of IRDIS allows to recover the polarized light of the circumstellar disks using their polarized properties. On one hand, the starlight is mainly unpolarized, on the other hand, the light from the disk is partially linearly polarized because of the reflection of the starlight on the dust located in the disk. In order to perform polarimetric measurements a sequence of observation using several Half Wave Plate (HWP) angles is used to reconstruct the unpolarized and polarized signal by combining the obtained dataset.

Several methods have been developped to extract the linearly polarized intensity[1] using the Stoke parameters formalism.[2] The approach proposed in this paper proposes an alternative to recover polarized intensity of the disks, its angle and the unpolarized intensity using inverse methods.

Further author information: (Send correspondence to L. Denneulin)
L. Denneulin: E-mail: laurence.denneulin@univ-lyon1.fr

Ground-based and Airborne Instrumentation for Astronomy VII, edited by Christopher J. Evans, Luc Simard, Hideki Takami,
Proc. of SPIE Vol. 10702, 1070245 · © 2018 SPIE · CCC code: 0277-786X/18/$18 · doi: 10.1117/12.2313901

Polarimetric data By using IRDIS in DPI, it is possible to measure polarized light for any light sources in the FOV. Polarisation measurements are made using a half wave plate to rotate the plane of polarization and 1 pair of wire grid analyzers located inside IRDIS cryostat. The HWP is mounted before the derotator and is rotated to four positions in the beam while the intensity measurements are recorder on the detector at each position angle simultaneously using both crossed polarizer at angle 0° and 90° (called left and right) simultaneously. This method present the advantage of removing instrumental polarization that stays constant between the HWP rotations (i.e. contributions in the optical path located between the HWP and the polarisers) between the measurements. The output of the detector is a sequence of n 1024×2048 images (as set by the HAWAII 2RG detector format and the IRDIS double channel configuration). The left part of the image acquire the light coming through the left polarizer (0°) and the right part from the right one (90°). A sequence of acquisition is composed of different subsequences acquired at different HWP angles. According to the integration time, they can be composed of one or several frames. After pre-processing (including bad pixel removal, Flat and background calibration, alignment of the left and right images), data are saved as a cube of size $1024 \times 1024 \times n \times 2$, where n is the size of the sequence and 2 is for the left and right images.

Inverse approach Method The inverse approach as been developped in the last decades along with the Bayesian methods and convex optimization methods. Given a data set with some additive noise, the inverse approach lean on the known statistic of this noise in order to fit an unknown perfect estimator to this data set. In the case of astrophysical data the quantity acquired by the detector is a number of electrons. Consequently, the noise follows a Poisson distribution with parameters equal to the mean and the variance of the number of electrons. However, given the object of interest in most cases this parameter is quite high. It is thus possible to approximate this Poisson distribution with a Gaussian distribution.

This paper is organized as follows: in the next section the physical model will be introduced and linked to other existing methods. The second section will explain how the physical model is fitted onto the data (inverse approach). Then, using a pixel-based simulation, results of the inverse approach will be studied and compared to other existing methods. In the last section, we will show some results obtained using this new methods on real data from the instrument ESO/VLT-SPHERE/IRDIS.

2. POLARIMETRIC DATA ANALYSIS METHOD

2.1 Data Physical Model

The aim of the physical model is to accurately represent the instrument and to be easy to manipulate. It is based on the study of the transfer of the electric field through the instrument, using Jones representation.

For a fixed wavelength λ, setting $e.\exp\left(\mathrm{i}\left(\omega t - \lambda z\right)\right) \in \mathbb{C}^2$ an electric field where z is the direction of propagation and ω the speed of light in vacuum. This electric field oscillates with a complex amplitude e that is the polarization vector. This polarization vector can be linear, circular or elliptical according to the different patterns it draws in the wavefront (x, y) along the time. The star light is unpolarized: for a given time, amplitudes and phases of the polarization vector components are random. Light emitted by the disk is linearly polarized because it comes from the reflection of the star light. It means that the amplitude is the same for both component of the polarization vector and that phases differ from 180°. We assume that e is the sum of two fields: $u \in \mathbb{C}^2$ witch is unpolarized and $p \in \mathbb{R}^2$ which is linearly polarized. We further assume that u is a zero-mean random vector with independent and identically distributed (i.i.d.) components of variance μ.

The influence of all parts of the instrument can be represented by the Jones matrices (see Ref. 3). The electric field seen by the analyzers is given by $T_2\, J_\alpha\, T_1\, e$ where $T_1 \in \mathcal{M}_2(\mathbb{C})$ and $T_2 \in \mathcal{M}_2(\mathbb{C})$ are Jones matrices which account for the action of mirrors on the polarization before and after the HWP. The set $\mathcal{M}_2(\mathbb{C})$ denotes the set of complex squared 2×2 matrices. The rotation of the polarisation induced by the HWP is given by the Jones matrix $J_\alpha \in \mathcal{S}_2(\mathbb{R})$ with α the orientation of the HWP. The set $\mathcal{S}_2(\mathbb{R})$ denotes the set of real symmetrical 2×2 matrices.

The effect of the analyzer is to project the complex electric field onto a reference plane, which can be represented by the Hermitian product between the vector a_ψ representing the analyzer, oriented with an angle

ψ, and the incoming electric field. Remind that the Hermitian product, for a couple of vectors (x, y), is given by $\langle x, y\rangle_{\mathbb{C}^2} = \sum_k x_k^* y_k$, where x^* denotes the adjoint of x.

Finally, the intensity i seen by the pixels of the detector is the squared modulus of the transmitted electric field. For a given pixel and orientations α and ψ of the HWP and the analyzer, the intensity is given by:

$$i_\psi^\alpha = \mathbb{E}_t \left[|\langle a_\psi, T_2 J_\alpha T_1 e\rangle_{\mathbb{C}^2}|^2 \right] , \tag{1}$$

where $\mathbb{E}_t$ denotes temporal expectation. As shown in Appendix A.1, the intensity can be written as:

$$i_\psi^\alpha = \|v_\psi^\alpha\|_{\mathbb{C}^2}^2 \mu + |\langle v_\psi^\alpha, p\rangle_{\mathbb{C}^2}|^2. \tag{2}$$

where $v_\psi^\alpha = (T_2 J_\alpha T_1)^* a_\psi$, with $(T_2 J_\alpha T_1)^*$ the adjoint of $(T_2 J_\alpha T_1)$.

The parameters of interest are the variance μ of the unpolarized field u wich is the unpolarized intensity, the amplitude $\rho = |p|$ wich is the polarized intensity and the angle θ of the polarized field p wich is the polarization angle.

2.2 The link between this physical model and Malus's law

If the instrument is supposed perfect, then T_1 and T_2 are identity matrices and

$$v_\psi^\alpha = J_\alpha^* a_\psi = J_\alpha^T a_\psi = J_\alpha a_\psi = \begin{pmatrix} \cos 2\alpha & \sin 2\alpha \\ \sin 2\alpha & -\cos 2\alpha \end{pmatrix} \begin{pmatrix} \cos\psi \\ -\sin\psi \end{pmatrix} = \begin{pmatrix} \cos(2\alpha+\psi) \\ \sin(2\alpha+\psi) \end{pmatrix}.$$

Then, $\|v_\psi^\alpha\|_{\mathbb{C}^2}^2 = 1$. Setting $\rho = |p|^2$ and $c_\theta = \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix}$ then:

$$\begin{aligned} i_\psi^\alpha &= \mu + \rho |\langle v_\psi^\alpha, c_\theta\rangle|^2 \\ &= \mu + \rho \cos^2(\theta - (2\alpha + \psi)). \end{aligned}$$

The general physical model considers the input intensity as the sum of unpolarized and polarized intensities: $i = i_\mathrm{p} + i_\mathrm{u}$. When going through a Half-Wave Plate oriented with an angle α, the initial intensity isn't modified, but the angle of polarization is retarded by an angle of 2α. When going through an analyzer oriented with an angle ψ, for one pixel on the detector, Malus's law gives

$$i_\psi^\alpha = i_\mathrm{u}/2 + i_\mathrm{p} \cos^2(\theta - 2\alpha - \psi).$$

According to Malus's law, μ corresponds to the half unpolarized intensity, ρ to the polarized intensity and θ the orientation of the polarization. If $\psi = 0$, the left intensity is recovered (left polarizer). If $\psi = \frac{\pi}{2}$ the right one is recovered.

By setting $\mu = i_\mathrm{u}/2$ and $\rho = i_\mathrm{p}$ we can link both formalisms. In the next section, the new data physical model will be expressed as:

$$I_\psi^\alpha = \|v_\psi^\alpha\|_{\mathbb{C}^2}^2 \frac{I_\mathrm{u}}{2} + I_\mathrm{p} |\langle v_\psi^\alpha, c_\Theta\rangle_{\mathbb{C}^2}|^2, \tag{3}$$

where I_ψ^α, I_u, I_p and Θ are respectively maps of pixels i_ψ^α, i_u, i_p and θ. These maps set the parameters to estimate.

2.3 Sate-of-the-art polarimetric data reduction methods

To reduce data, two classic methods are used: the differential method and the double ratio (see Ref. 1 or Ref. 2). The differential method consists in calculating Stokes parameters by addition and differentiation. If $I_{\psi}^{2\alpha}$ is the image at the angle of HWP α for the orientation of the analyzer ψ, then:

$$\begin{cases} Q = \dfrac{I_{0^\circ}^{0^\circ} - I_{90^\circ}^{0^\circ} - \left(I_{0^\circ}^{90^\circ} - I_{90^\circ}^{90^\circ}\right)}{2}; \\ U = \dfrac{I_{0^\circ}^{22.5^\circ} - I_{90^\circ}^{22.5^\circ} - \left(I_{0^\circ}^{67.5^\circ} - I_{90^\circ}^{67.5^\circ}\right)}{2}. \end{cases} \tag{4}$$

According to Ref. 1, this method is not robust against noise and instrumental artefacts. In this latter paper, the author proposes a second method: the double ratio. First setting:

$$R_Q = \sqrt{\frac{I_{0^\circ}^{0^\circ}/I_{90^\circ}^{0^\circ}}{I_{0^\circ}^{45^\circ}/I_{90^\circ}^{45^\circ}}} \quad \text{and} \quad R_U = \sqrt{\frac{I_{0^\circ}^{22,5^\circ}/I_{90^\circ}^{22,5^\circ}}{I_{0^\circ}^{67,5^\circ}/I_{90^\circ}^{67,5^\circ}}},$$

then:

$$p_Q = \frac{R_Q - 1}{R_Q + 1} \quad \text{and} \quad p_U = \frac{R_U - 1}{R_U + 1}.$$

Finally

$$I_Q = \frac{I_{0^\circ}^{0^\circ} + I_{90^\circ}^{0^\circ} + I_{0^\circ}^{45^\circ} + I_{90^\circ}^{45^\circ}}{2} \quad \text{and} \quad I_U = \frac{I_{0^\circ}^{22,5^\circ} + I_{90^\circ}^{22,5^\circ} + I_{0^\circ}^{67,5^\circ} + I_{90^\circ}^{67,5^\circ}}{2},$$

so

$$\begin{cases} I = I_Q + I_U = I_{\mathrm{p}} + I_{\mathrm{u}}, \\ Q = p_Q I_Q = I_{\mathrm{p}} \cos\left(2\Theta\right), \\ U = p_U I_U = I_{\mathrm{p}} \sin\left(2\Theta\right). \end{cases} \tag{5}$$

Proof. A.2 in the appendix.

Finally, to recover the unpolarized intensity I_{u}, the polarized intensity I_{p} and the angle of polarization Θ, simply calculate $I_{\mathrm{p}} = \sqrt{Q^2 + U^2}$, $I_{\mathrm{u}} = I - \sqrt{Q^2 + U^2}$ and $\Theta = (1/2) * \arctan\left(U/Q\right)$.

One of The limitations of the classical methods is the use of a fixed set of HWP angles and polarizers. the physical model that grounds the inverse approach we propose offers the possibility to attenuate this constraint.

3. INVERSE APPROACH

To simplify the notation let's set $(\alpha_k)_{k=1,\dots,n}$ the angle taken by the HWP for each image in the sequence and $(\psi_\ell)_{\ell=1,2}$ the angle taken by the polarizer for the images left and right. The couple $\{k, \ell\}$ represent a position of HWP α_k and of polarizer ψ_ℓ. Pre-processed data are given by:

$$d_{k,\ell} = \|v_{k,\ell}\|_{\mathbb{C}^2}^2 \frac{I_{\mathrm{u}}}{2} + I_{\mathrm{p}} |\langle v_{k,\ell}, c_\Theta \rangle_{\mathbb{C}^2}|^2 + \beta_{k,\ell}, \quad \beta_{k,\ell]} \sim \mathcal{N}(0, \sigma_{k,\ell}^2). \tag{6}$$

were $\beta_{k,\ell}$ is a zero-mean white noise. Defining $\phi = (I_{\mathrm{u}}, I_{\mathrm{p}}, \Theta)^\top$ and $I_{k,\ell}(\phi) = \|v_{k,\ell}\|_{\mathbb{C}^2}^2 \frac{I_{\mathrm{u}}}{2} + I_{\mathrm{p}} |\langle v_{k,\ell}, c_\Theta \rangle_{\mathbb{C}^2}|^2$, then $d_{k,\ell} \sim \mathcal{N}(I_{k,\ell}(\phi), \sigma_{k,\ell}^2)$. Reminding that the noise follows a Gaussian approximation of a Poisson distribution with parameters equal to the empirical mean and variance, it implies that $\sigma_{k,\ell}^2 = I_{k,\ell} + \sigma_{\mathrm{ro}}^2$, where σ_{ro}^2 is the read out noise of the detector.

The vector ϕ is the parameter to estimate. In order to find the best estimate of this parameter, we maximize the log likelihood. Setting the weights $W_{k,\ell} = \frac{1}{\sigma^2_{k,\ell}}$ then the estimator $\hat{\phi}$ is such that:

$$\hat{\phi} \in \underset{\phi}{\text{Argmin}} \quad \sum_{k,\ell} \frac{W_{k,\ell}}{2} \left\| d_{k,\ell} - I_{k,\ell} \right\|^2 . \tag{7}$$

This criterion is linear in I_{u} and I_{p} however it isn't linear in Θ. It is thus possible to rewrite the problem as:

$$\hat{\Theta} \in \underset{\Theta}{\text{Argmin}} \left(\min_{I_{\text{u}} \geq 0, I_{\text{p}} \geq 0} \quad \sum_{k,\ell} \frac{W_{k,\ell}}{2} \left\| I_{k,\ell}(\phi) - d_{k,\ell} \right\|^2 \right), \tag{8}$$

where $\left(\hat{I}_{\text{u}}, \hat{I}_{\text{p}} \right)^\top = \hat{x}$ is, under the positivity constraint, the solution of the simple system $A(\Theta)\, x = b(\Theta)$ with:

$$A(\Theta) = \begin{pmatrix} \sum_{k,\ell} W_{k,\ell} \| v_{k,\ell} \|^4_{\mathbb{C}^2} & \sum_{k,\ell} W_{k,\ell} \| v_{k,\ell} \|^2_{\mathbb{C}^2} |\langle v_{k,\ell}, c_\Theta \rangle_{\mathbb{C}^2}|^2 \\ \sum_{k,\ell} W_{k,\ell} \| v_{k,\ell} \|^2_{\mathbb{C}^2} |\langle v_{k,\ell}, c_\Theta \rangle_{\mathbb{C}^2}|^2 & \sum_{k,\ell} W_{k,\ell} |\langle v_{k,\ell}, c_\Theta \rangle|^4 \end{pmatrix} \tag{9}$$

and

$$b(\Theta) = \begin{pmatrix} \sum_{k,\ell} W_{k,\ell} \| v_{k,\ell} \|^2_{\mathbb{C}^2} d_{k,\ell} \\ \sum_{k,\ell} W_{k,\ell} |\langle v_{k,\ell}, c_\Theta \rangle_{\mathbb{C}^2}|^2 d_{k,\ell} \end{pmatrix}.$$

with the constraint that the pixels of I_{u} and I_{p} are non-negative, because they are intensities (in electron per seconds). Note that since part of the electric field p is linearly polarized, θ has two solutions in $[0, 2\pi]$ with a difference of π. This is a true ambiguity especially if a regularization of the map of angles is needed.

The implementation of the algorithm is simple. First of all, we calculate $A(\Theta)$ and $b(\Theta)$ for each Θ in a discrete set $[0, 2\pi]$ and solve the linear system $A(\Theta)\, x = b(\Theta)$. Then, if the positivity constraint is not respected, there are three cases to explore for each pixels:

- if $i_{\text{u}} \geq 0$ and $i_{\text{p}} < 0$ then setting $i_{\text{p}} = 0$ and calculate i_{u} for a such value of i_{p}. If $i_{\text{u}} \geq 0$ the criterion in Θ can be calculate, else setting $i_{\text{u}} = 0$ before.
- if $i_{\text{u}} < 0$ and $i_{\text{p}} \geq 0$ then setting $i_{\text{u}} = 0$ and calculate i_{p} for a such value of i_{u}. If $i_{\text{p}} \geq 0$ the criterion in θ can be calculate, else setting $i_{\text{p}} = 0$ before.
- if $i_{\text{p}} < 0$ and $i_{\text{u}} < 0$ then the two precedent cases are to be tested. If both cases work, the one who gives the smallest value of the criterion is kept.

Some existing algorithms permit to find the minimum of a non-convex function, like Powell's methods.[4] Their principle is to take three points in increasing order, from the beginning of a given set, and calculate the criterion for each. If the middle one is lower than the others it is kept as the minimum. Then it takes the two last points and the next one and does the same calculus. At the end of the set, the minimum registered is the global on.

4. SIMULATIONS AND COMPARISONS

An important part of this approach is to know how accurate are estimators and how robust is the method against noises and intensity variations. In order to do this analysis, a pixel will be simulated using th two following parameters: the signal-to-noise ratio (SNR) and the ratio of polarization τ given by:

$$\mathrm{SNR} = \frac{\sqrt{n} i_\mathrm{p}}{\sqrt{\left(i_\mathrm{u} + i_\mathrm{p}\right)/2 + \sigma_\mathrm{ro}^2}} \quad \text{and} \quad \tau = \frac{i_\mathrm{p}}{i_\mathrm{u} + i_\mathrm{p}}, \tag{10}$$

were σ_ro^2 is the read-out noise of the detector.

The choice of a such SNR comes from the fact that the signal of interest is the polarized intensity and its angle of polarization. Moreover, the unpolarized intensity is at least a thousand time greater than the polarized intensity. Finally, since the noise follows a Poisson distribution, the higher the intensity is, the higher is the noise variance.

The SNR expression is estimated using the mean of the polarized intensity for all position of HWP and analyzers over the total variance of the intensity for all position of HWP and analyzers. It is obvious that the SNR will grow with the number of images n. In the simulation it will be fixed to 4 (according to the number of HWP positions). The Fig. 1 shows that when the intensity is low, the SNR is low no matter the polarization ratio. It means that the read out noise odominates the signal. Then, the SNR grows with the total intensity and the polarization ratio.

It is important to notice that not all SNR values depicted on the figure 1 are realistic, especially in cases of high total intensity and high polarization ratio or the opposite, low intensity and low polarization ratio.

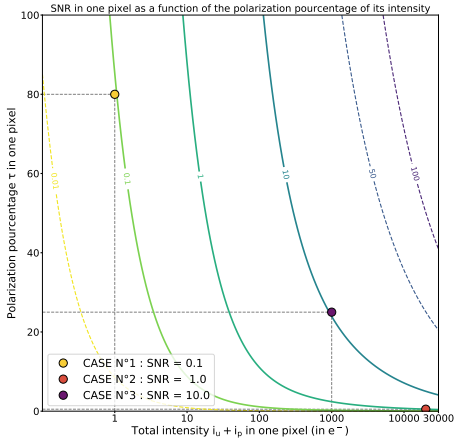


Figure 1. This figure shows the evolution of the SNR of a any pixel according to the total intensity $i_\mathrm{u} + i_\mathrm{p}$ and the polarization ratio τ given in Eq. 10. The maximum of the total intensity is chosen according to the limit of saturation of the IRDIS detector. The three markers depict the cases that will be used in the next section simulations. The first marker (yellow) represent the case of low intensity (1 e^- acquired during the integration time) but with high polarization ratio (80%). It represent the case where almost all the intensity come from the disk and a few from the star residuals). The second marker (orange) represent the inverse case with the intensity of 20 000 e^- and a small ratio of polarization (0.5%). It represent the case where almost all the intensity is set by the star (close to the coronograph) and very small amount of polarized light comes from the disk itself. The last case (purple) is a more realistic case, with a total intensity of 1000 e^- and a ratio of polarization of 25%. It is representative of the data within the adaptive optics correction radius assuming.

First of all, the Fig. 2 shows perfectly well the ambiguity in the angle estimation. Two minima are found with a difference of π. It also shows that the criterion is not globally convex and has some local minima on the upper parts of the criterion. However, the criterion is still locally convex close to each true minimum. Secondly, the figures shows that when the SNR is high (10), the value of θ recovered is almost the same as the true value. When the SNR decreases (1 and 0, 1), the estimated value of θ recovered deviate increasingly from the true value.

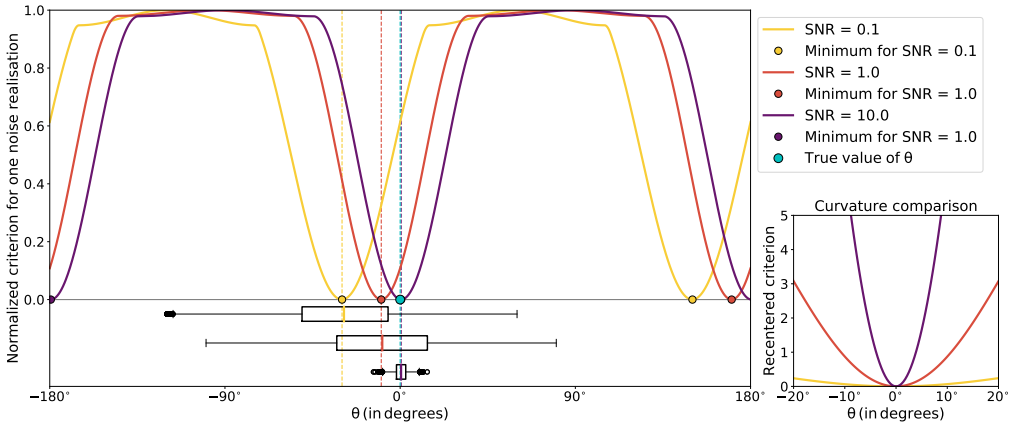


Figure 2. The first figure shows normalized values of the criterion in Eq. 8 as a function of θ for the different cases of SNR described on Fig. 1. Each criterion is scaled to zero by subtracting its mean and normalized. Each minima in $\hat{\theta}$ is marked and the true value of θ is overplotted in blue. Underneath each θ minimum, a boxplot represents the error on each angle estimation. The second figure on the left represent a comparison of the curvature of each criterion around their minimum. These criterion are not normalized but only recentered.

From these results, we conclude that the deviation of the estimate θ grows when the SNR decreases. Fig. 3 shows the linearly polarized electric fields estimated normalized by the simulated amplitude $\sqrt{\delta i_{\rm p}} c_{\hat{\theta}}$, where $\delta i_{\rm p} = \hat{i}_{\rm p}/\bar{i}_{\rm p}$, over 1000 realizations of the noise for these three different SNR values. The first component of the electric field is plotted on the abscissa axis as a function of the second component on the ordinate axis. Moreover, because of the angle ambiguity, for each electric field plotted, its symmetric in relation to the center is also plotted. The Figure also shows the same cloud in the case where images for two different HWP positions (i.e. SNR/$\sqrt{2}$) are missing.

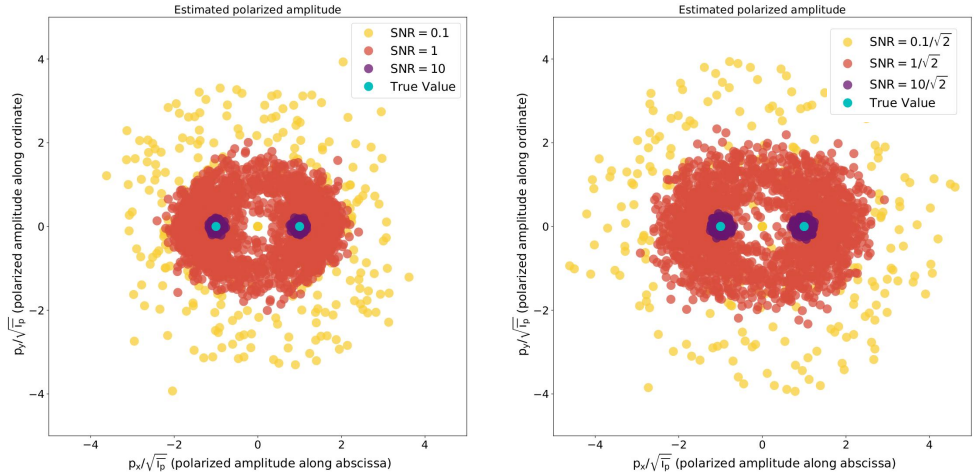


Figure 3. The cloud on the left shows the values of the polarized electric field p recovered normalized by the simulated amplitude $\sqrt{\bar{i}_{\rm p}}$ for the three values of selected SNR cases on Fig. 1 over 1000 realizations of the noise. For a value of SNR$= 0, 1$, the amplitude estimated can be four time the true amplitude. This means that the intensity recovered can be sixteen time the intensity simulated in this peculiar very low SNR case. The cloud on the right also shows the same results when the estimation is based on a partial dataset with images corresponding to two HWP positions are missing. The SNR is thus divided by $\sqrt{2}$.

Fig. 3 shows that When the SNR is high (SNR$= 10$), the dots gathered around the true values of the electric field. This means that the estimated values are very close to the simulated values of both $i_{\rm p}$ and θ. When the signal of interest is as big as the variance of the noises (SNR$= 1$), the dots are spread over a larger area around

the true value of the electric field. The angles estimates wanders the trigonometric circle. They have a biggest density at each pole centered on the true value of θ and its symmetric. The estimated values of i_{p} spreads out and can be twice the value truth. Finally, when the SNR is small (SNR$= 0, 1$), the value of $\sqrt{i_{\mathrm{p}}}$ restored can be three to four time the truth value. On the second figures, the SNR is divided by $\sqrt{2}$ and the spread of the cloud is slightly wider.

The conclusion is that with a SNR smaller than 1 it is, as expected, difficult to restore the true value of the angle and the variance of the estimate becomes very high. Then, higher the SNR is, most precise will be the estimated values of i_{p} and θ. It is worth noticing that our method is very robust against missing data unlike the other methods. The only constraints for our method to work is the need for at least data associated with two positions of HWP i.e at least one stokes $\pm$ Q and $\pm$ U measurements. To study more rigourously how the estimators of i_{u}, i_{p} and θ behave with the SNR, the Mean Squared Error (MSE) and the Bias are calculated. These values give the accuracy of the estimator. For a number n of noise realization, the Bias and the MSE of a given estimate $\hat{\mu}$ are:

$$\mathrm{Bias}\left(\hat{\mu}\right) = \mathbb{E}\left[\hat{\mu}\right] - \mu,$$

where $\mathbb{E}\left[.\right]$ denotes the empirical mean of the estimates, that is $(1/n)\sum_{k=1}^{n}\mu_k$, and

$$\begin{aligned}\mathrm{MSE}\left(\hat{\mu}\right) &= \mathbb{E}\left[\left(\hat{\mu}-\mu\right)^2\right] = \mathbb{E}\left[\hat{\mu}^2\right] - \mathbb{E}\left[\hat{\mu}\right]^2 + \mathbb{E}\left[\hat{\mu}\right]^2 - 2\mathbb{E}\left[\hat{\mu}\right]\mu + \mu^2,\\ &= \mathrm{Var}\left(\hat{\mu}\right) + \mathrm{Bias}\left(\hat{\mu}\right)^2,\end{aligned}$$

were Var defined the empirical variance of $\hat{\mu}$. These values can be calculated for each pixel with the estimators. The expression of the MSE shows the relation between the variance and the bias of an estimate.

The Fréchet-Darmois-Cramér-Rao Lower Bound (FDCRLB) gives a lower born to the MSE for unbiased estimates. The more accurate are estimators, the closer the MSE is to the FDCRLB, demonstrating how much estimators can be improved.

An hypothesis is that estimators are asymptotically unbiased. This means that when SNR increases, the bias decreases. Since the FDCRLB is accurate for unbiased estimators, the CRLB will asymptotically converge to reach the real CRLB when bias reaches zero. In addition, the MSE will decrease with SNR.

Supposing our estimators are unbiased then we can write $\mathrm{Cov}\left(\hat{\phi}\right) \geq \left[\mathcal{I}(\phi)\right]^{-1}$, were $\mathcal{I}(r)$ is the Fisher information matrix. This matrix is easy to calculate and is given by:

$$\mathcal{I}\left(\phi\right) = \begin{pmatrix} A\left(\theta\right) & \begin{matrix}\mathcal{I}_1\\ \mathcal{I}_2\end{matrix} \\ \begin{matrix}\mathcal{I}_1 & \mathcal{I}_2\end{matrix} & \mathcal{I}_3 \end{pmatrix} \tag{11}$$

where $A\left(\theta\right)$ is the matrix given in Eq. 9, and

$$\begin{cases} \mathcal{I}_1 = \sum_{k,\ell} w_{k,\ell}\|v_{k,\ell}\|^2_{\mathbb{C}^2} i_{\mathrm{p}} \dfrac{\partial|\langle v_{k,\ell}, c_\theta\rangle_{\mathbb{C}^2}|^2}{\partial\theta}, \\ \mathcal{I}_2 = \sum_{k,\ell} w_{k,\ell}|\langle v_{k,\ell}, c_\theta\rangle_{\mathbb{C}^2}|^2 i_{\mathrm{p}} \dfrac{\partial|\langle v_{k,\ell}, c_\theta\rangle_{\mathbb{C}^2}|^2}{\partial\theta}, \\ \mathcal{I}_3 = \sum_{k,\ell} w_{k,\ell}\left(i_{\mathrm{p}} \dfrac{\partial|\langle v_{k,\ell}, c_\theta\rangle_{\mathbb{C}^2}|^2}{\partial\theta}\right)^2. \end{cases}$$

See proof. A.3 for more details.

The unpolarized intensity i_{u} scales in the FDCRLB with the weight $w_{k,\ell} = 1/\sqrt{i_{k,\ell} + \sigma_{\mathrm{ro}}^2}$. Replacing the weight by its formula, the expression of the SNR depends on $\mathcal{I}_1$, $\mathcal{I}_3$ and $\mathcal{I}_3$. This means that the SNR value is closely link to the quality of the estimators as well as the polarization ratio. Furthermore, every coefficient of the matrix will depend of θ. Fig. 4 shows the correlation between each parameters. A value close to 1 means that the correlation is strong, a value close to 0 means that the correlation is weak, finally a correlation close to -1 means that values are anti-correlated. Here, the correlation between i_{u} and i_{p} is strong while it is weak between these parameters and θ. Those correlations are not depending on the SNR value.

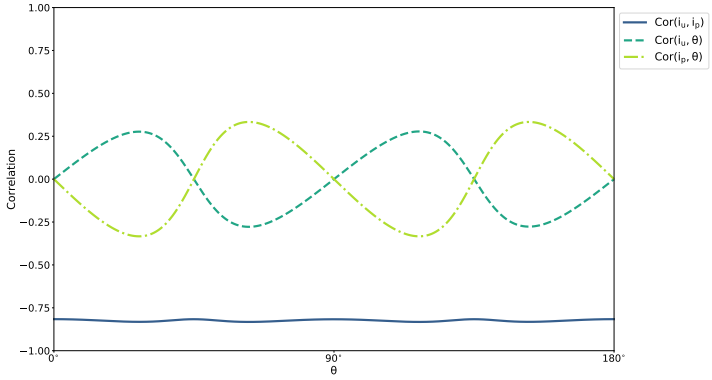


Figure 4. Correlations between each parameters. The parameters i_{u} and i_{p} are strongly anti-correlated. The correlation between the angle and each intensity parameter depends of the angle. If one pair is correlated, the other is anti-correlated.

In order to prove the convergence of the estimators, we simulated data sets for the three SNR cases on Fig: 1. The parameters of interest are estimated with the inverse approach, the differential method and the double ratio method. Fig. 5 shows the square root of the MSE for the estimates of θ and i_{p}, and the squared root of the FDCRLB for all polarized intensity estimator. When SNR increases, the MSE and the FDCRLB decrease. Moreover in case of SNR higher than 1, the three methods are equivalent. However, in case of weak SNR, the accuracy of the estimate is different for each method.

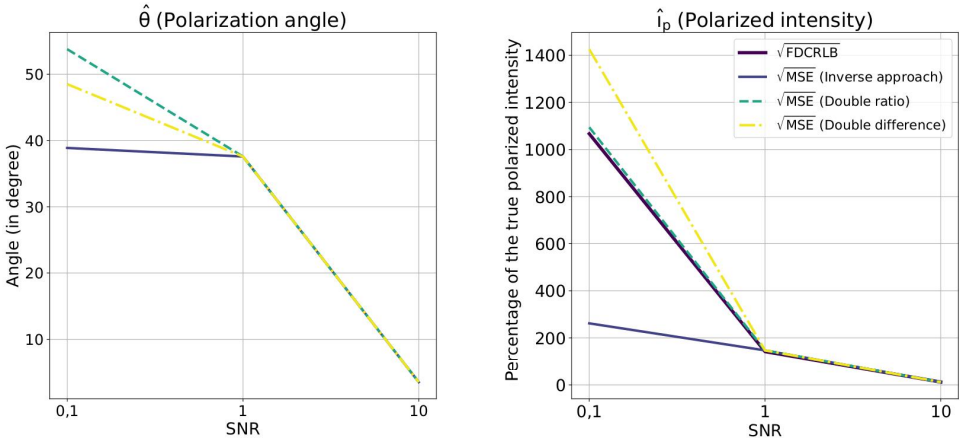


Figure 5. The left figures shows the square root of the MSE of the estimated angle with the three methods as a function of the SNR. The values on the ordinate are in degrees. The figure on the right shows the square root of the CRLB and the square root of the MSE of the estimated polarized intensity as function the SNR for the three methods. The values on the ordinate are given as a percentage of the simulated intensity $\bar{i}_{\mathrm{u}}$. In this exemple, the squared MSE for the inverse approach is 200% of the initial intensity, it corresponds to an error of a factor 2. For the double ratio, the error on the estimate is of a factor 10. For the double difference, the error is of a factor 14.

For the estimate of the polarization angle, the MSE of the estimate obtained by the inverse approach is also

better than the MSE obtained by the other two methods. However its absolute bias is slightly higher leading to the conclusion that the variance of this estimate is even more smaller than the variance of estimates obtained by the other two methods.

For the estimation of the polarized intensity, the MSE obtained by the inverse approach is better than the MSE of estimates obtained by the other two methods. Moreover, its absolute bias is very small compared to the bias of the estimators obtained with the two other methods.

Since the estimates obtained by the inverse approach are perfect under notion of maximum of likelihood, and the three methods are equivalent in case of SNR higher than 1, the double ratio and double difference methods are efficient in such case. However, in case of low SNR, the inverse approach improve consequently the quality of estimates.

The next section shows some results obtained with the inverse approach using real data set taken with the SPHERE/IRDIS DPI instrument.

5. RESULTS ON SPHERE/IRDIS DATA

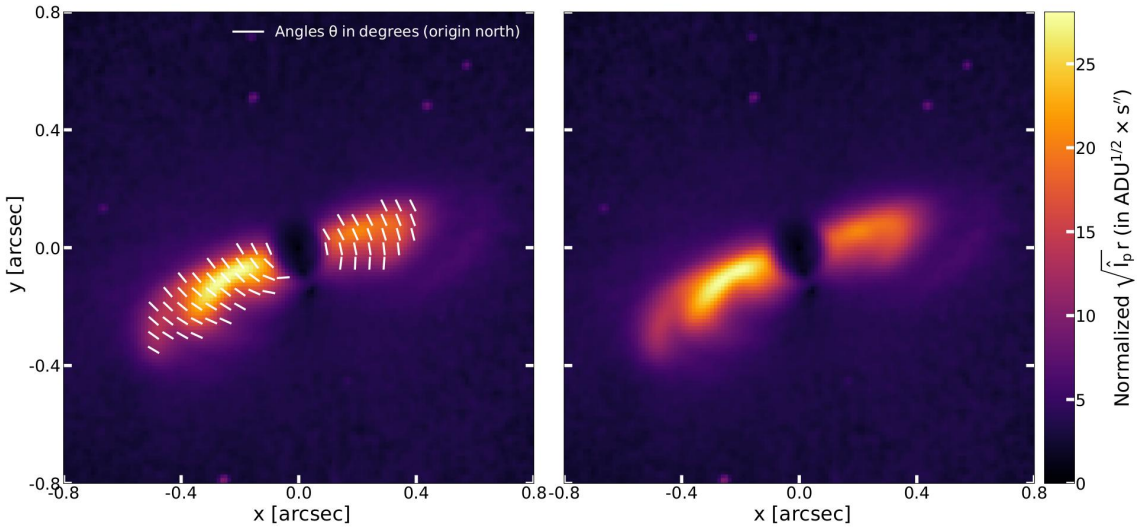


Figure 6. Estimated polarization angle $\hat{\Theta}$ (on the left) and polarized intensity $\hat{I}_p$ (on the right) with the method introduced in this paper of the target RY Lupus.[5]

RY Lup The observations of RY Lup were part of the SPHERE consortium guaranteed time programme (SPHERE High-Contrast Imaging Survey for Exoplanets and SPHERE Disk Survey). The Infra-Red Dual-beam Imager and Spectrograph was used. The IRDIS-DPI observations of RY Lup were carried out in field stabilized mode on 27 May 2016 with the BBH filter (1.625 width= 290 nm) using the apodized pupil Lyot coronagraph (N-ALC-YJH-S). Sixty-four polarimetric cycles were taken, each consisting of one data cube for each of the four half wave plate (HWP) positions (0, 45, 22.5, 67.5 degrees). IRDIS provides a 11 arcseconds square field of view with 12.25 mas/pixel platescale. The true north (TN) and pixel plate scales were measured using the astrometric calibrator NGC3603 observed as part of the standard SPHERE calibration routines performed for the Guaranteed Time Observations (GTO) survey. The outputs include cubes of left and right images (parallel and perpendicular polarized beams, respectively) recentred onto a common origin (the star center) using the satellite spots. This preprocessing also include background subtraction, bad-pixel interpolation, flat-field correction, distortion correction. The data were then reduced following the inverse approach method described in this paper as show on Fig: 6.

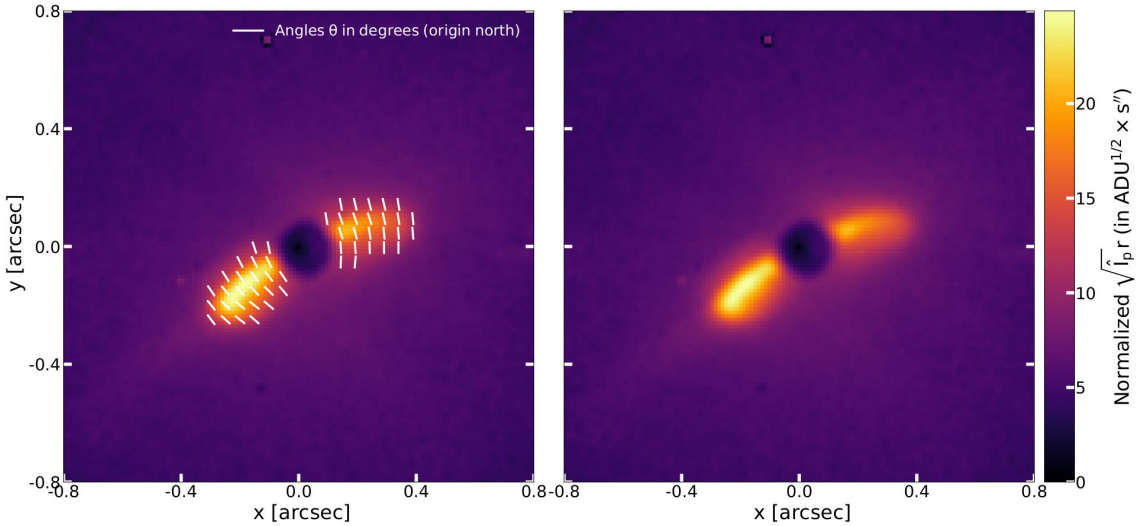


Figure 7. Estimated polarization angle $\hat{\Theta}$ (on the left) and polarized intensity $\hat{I}_{\mathrm{p}}$ (on the right) with the method introduced in this paper of the target TCHA.

TCHA The IRDIS-DPI observations of T Cha were carried out on 19 February 2016 with the BB_H filter using an apodized pupil Lyot coronagraph. Thirty polarimetric cycles were taken, consisting of one data cube for each of the four half wave plate (HWP) positions (0, 45, 22.5 and 67.5 deg.). The first step consists of standard calibration routines, including dark-frame subtraction, flat-fielding and bad-pixel correction. These images are split into two individual frames representing the left and right sides (parallel and perpendicular polarized beams, respectively), and the precise position of the central star is measured using the star center calibration frames on both image sides separately.The data were then reduced following the inverse approach method described in this paper as shown on Fig: 7.

6. CONCLUSION

This paper introduced a new method to reduce polarimetric data based on an inverse approach method. This method is based on an optimization using the electric field (Jones Matrix) rather than the intensity (Mueller matrix). In the first section, a physical model based on Jones formalism has been explained. In the second section, a method to fit this model to the data has been introduced. The parameters of the models are estimated hierarchically with a maximum of likelihood and a Powell's method. A third section present the accuracy of estimates obtained with the inverse approach and shows a comparison with two other state-of-the-art methods (double difference and double ratio).The inverse approach method is more efficient at minimizing instrumental artefacts (bad pixels, dark/background residual) and at retrieving the linearly polarized and unpolarized intensities in particular in low SNR cases. It also enables the use of any set of frames with incomplete polarimetric cycles (with at least one Q and one U measurement). We have successfully demonstrated the efficiency of the method using both simulation and IRDIS-DPI public data.

APPENDIX A. NOTATIONS AND PROOFS

A.1 Proof of Eq. 2

For fixed orientation of the HWP α and of polarizer ψ, the intensity acquired by the detector is given by:

$$i_{\psi}^{\alpha} = \mathbb{E}\left[\left|\langle v_{\psi}^{\alpha}, e\rangle_{\mathbb{C}^2}\right|^2\right]_t .$$

To simplifies the notation, we omit α and ψ. Supposing v constant in time, by linearity of the expectancy:

Table 1. Notations used in the article

N	Size of the images
n	Total number of images in the dataset
$d_{k,\ell}$	data obtained with with $k = 1, \ldots, n$ and $\ell = 1, 2$
α_k	Half-Wave Plate orientation of the data with $k = 1, \ldots, n$
ψ_ℓ	Polarizer orientation of the data with $\ell = 1, 2$
$i_{k,\ell}$, $I_{k,\ell}$	Physical model in one pixel and map of the physical model of the data $d_{k,\ell}$
i_{u}, i_{p}, θ	Parameters of the physical model to estimate in one pixel
I_{u}, I_{p}, Θ	Map of physical model parameters to estimate
$v_{k,\ell}$	Physical model of the instrument for an orientation α_k of HWP and ψ_ℓ of polarizer
e, u, p	Total electric field, unpolarized electric field, polarized electric field in one pixel
t	The timestamp
μ	Half unpolarized intensity $i_{\mathrm{u}}/2$
ρ	Polarized intensity i_{p}
ϕ	$= (I_{\mathrm{u}}, I_{\mathrm{p}}, \Theta)$
x	$= (I_{\mathrm{u}}, I_{\mathrm{p}})$
c_θ, c_Θ	Polarization angle projection vector for one pixel and maps of the projection vector
$\sigma_{k,\ell}$	Noise standard deviation of the data $d_{k,\ell}$
$W_{k,\ell}$	Weights of the data $d_{k,\ell}$
$\mathcal{I}$	Fisher matrix
(I, Q, U, V)	Stokes parameters
$\mathcal{M}(\mathbb{C}^2)$	Set of 2×2 complex matrices
$\mathcal{S}(\mathbb{R}^2)$	Set of 2×2 symmetric matrices

$$\mathbb{E}\left[|\langle v, e\rangle_{\mathbb{C}^2}|^2\right] = |v_x|^2\mathbb{E}\left[|e_x|^2\right] + |v_y|^2\mathbb{E}\left[|e_y|^2\right] + v_x v_y^*\mathbb{E}\left[e_x e_y^*\right] + v_y v_x^*\mathbb{E}\left[e_y e_x^*\right],$$

were * denotes the adjoint. Recalling that $e = u+p$ with u randomn, zero mean, with independent identically distributed component and p real and constant according to time. Using the linearity:

$$\mathbb{E}\left[e_x e_y^*\right] = \mathbb{E}\left[e_x\right]\mathbb{E}\left[e_y^*\right] = \mathbb{E}\left[u_x + p_x\right]\mathbb{E}\left[u_y^* + p_y^*\right] = p_x p_y,$$

The expectancy $\mathbb{E}\left[e_y e_x^*\right]$ can be calculate in the same way. Then:

$$\begin{aligned}
\mathbb{E}\left[|e_x|^2\right] &= \mathbb{E}\left[|u_x + p_x|^2\right], \\
&= \mathbb{E}\left[|u_x|^2\right] + \mathbb{E}\left[u_x\right]p_x + \mathbb{E}\left[u_x^*\right]p_x + p_x^2, \\
&= \mathbb{E}\left[|u_x|^2\right] + p_x^2, \\
&= \mathrm{Cov}\left(u_x, u_x^*\right) + p_x^2, \\
&= \mathrm{Var}\left(\Re(u_x)\right) + \mathrm{Var}\left(\Im(u_x)\right) + p_x^2.
\end{aligned}$$

Same calculus for the y component of e. Setting $\mu = \mathrm{Var}\left(\Re(u_x)\right) + \mathrm{Var}\left(\Im(u_x)\right)$

$$\begin{aligned}
i &= |v_x|^2\left(\mu + p_x^2\right) + |v_y|^2\left(\mu + p_y^2\right) + v_x v_y^* p_x p_y + v_y v_x^* p_x p_y, \\
&= (v_x v_x^* + v_y v_y^*)\mu + v_x v_x^* p_x^2 + v_y v_y^* p_y^2 + \left(v_x v_y^* + v_y v_x^*\right)p_x p_y, \\
&= \langle v, v\rangle_{\mathbb{C}^2}\mu + |\langle v, p\rangle_{\mathbb{C}^2}|^2.
\end{aligned}$$

Finally:

$$i_\psi^\alpha = \|v_\psi^\alpha\|_{\mathbb{C}^2}^2 \mu + |\langle v_\psi^\alpha, p\rangle_{\mathbb{C}^2}|^2.$$

A.2 Proof of Eq. 5

Setting the intensity on the detector $i_\psi^\alpha = i_\mathrm{u}/2 + i_\mathrm{p}\cos^2(\theta - 2\alpha - \psi)$ where α is the angle of the HWP and ψ the angle of the analyzer. Then:

$$R_{Q,U} = \sqrt{\frac{i_\psi^\alpha / i_{\psi+90^\circ}^\alpha}{i_\psi^{\alpha+45^\circ} / i_{\psi+90^\circ}^{\alpha+45^\circ}}} = \sqrt{\frac{i_\psi^\alpha \times i_{\psi+90^\circ}^{\alpha+45^\circ}}{i_{\psi+90^\circ}^\alpha \times i_\psi^{\alpha+45^\circ}}}.$$

Notice that for a given angle θ, $\cos(\theta + 90^\circ) = \sin(\theta)$, then $R_{Q,U}$ is given by:

$$R_{Q,U} = \frac{i_\mathrm{u}/2 + i_\mathrm{p}\cos^2(\theta - 2\alpha - \psi)}{i_\mathrm{u}/2 + i_\mathrm{p}\sin^2(\theta - 2\alpha - \psi)}.$$

Then:

$$\begin{aligned} p_{Q,U} &= \frac{R_{Q,U} - 1}{R_{Q,U} + 1} \\ &= \frac{\frac{i_\mathrm{u}/2 + i_\mathrm{p}\cos^2(\theta-2\alpha-\psi) - i_\mathrm{u}/2 + i_\mathrm{p}\sin^2(\theta-2\alpha-\psi)}{i_\mathrm{u}/2 + i_\mathrm{p}\sin^2(\theta-2\alpha-\psi)}}{\frac{i_\mathrm{u}/2 + i_\mathrm{p}\cos^2(\theta-2\alpha-\psi) + i_\mathrm{u}/2 + i_\mathrm{p}\sin^2(\theta-2\alpha-\psi)}{i_\mathrm{u}/2 + i_\mathrm{p}\sin^2(\theta-2\alpha-\psi)}} \\ &= \frac{i_\mathrm{p}\cos 2(\theta - 2\alpha - \psi)}{i_\mathrm{u} + i_\mathrm{p}}, \end{aligned}$$

and

$$I_{Q,U} = i_\psi^\alpha + i_{\psi+90^\circ}^\alpha + i_\psi^{\alpha+45^\circ} + i_{\psi+90^\circ}^{\alpha+45^\circ} = i_\mathrm{u} + i_\mathrm{p}.$$

Finally, for $\psi = 0$ and $\alpha = 0^\circ$, we get $Q = p_Q I_Q = i_\mathrm{p}\cos 2\theta$ and for $\alpha = 22,5^\circ$ we get $U = i_\mathrm{p}\sin 2\theta$.

A.3 Proof of Eq. 11

Setting $f_{k,\ell} = \frac{1}{2\sigma_{k,\ell}^2}\|d_{k,\ell} - i_{k,\ell}(\phi)\|^2 + \frac{1}{2}\log(2\sigma_{k,\ell}^2)$ and $F = \sum_{k,\ell} f_{k,\ell}$, then:

$$\mathcal{I}(\phi)_{k,\ell} = \mathbb{E}\left[\nabla_\phi F . \nabla_\phi F^\top\right] = -\mathbb{E}\left[\mathcal{H}_\phi F\right], \tag{12}$$

where $\nabla_\phi F$ is the gradient of F and $\mathcal{H}_\phi F$ its Hessian matrix. The first derivative for the parameter ϕ_j gives:

$$\sum_{k,\ell} \frac{\partial f_{k,\ell}}{\partial \phi_j} = \sum_{k,\ell} w_{k,\ell} \frac{\partial i_{k,\ell}(\phi)}{\partial \phi_j} \left(i_{k,\ell}(\phi) - d_{k,\ell}\right),$$

and the second one:

$$\sum_{k,\ell} \frac{\partial^2 f_{k,\ell}}{\partial \phi_{j_1} \partial \phi_{j_2}} = -\sum_{k,\ell} w_{k,\ell} \left(\frac{\partial i_{k,\ell}(\phi)}{\partial \phi_{j_1}} \frac{\partial i_{k,\ell}(\phi)}{\partial \phi_{j_2}} + \frac{\partial^2 i_{k,\ell}(\phi)}{\partial \phi_{j_1} \partial \phi_{j_2}} \left(i_{k,\ell}(\phi) - d_{k,\ell}\right) \right).$$

Yet, since $\mathbb{E}\left[d_{k,\ell}\right] = i_{k,\ell}(\phi)$:

$$-\mathbb{E}\left[\frac{\partial^2 f_{k,\ell}}{\partial\phi_{j_1}\partial\phi_{j_2}}\right] = w_{k,\ell}\mathbb{E}\left[\frac{\partial i_{k,\ell}(\phi)}{\partial\phi_{j_1}}\frac{\partial i_{k,\ell}(\phi)}{\partial\phi_{j_2}}\right] = w_{k,\ell}\frac{\partial i_{k,\ell}(\phi)}{\partial\phi_{j_1}}\frac{\partial i_{k,\ell}(\phi)}{\partial\phi_{j_2}},$$

and then:

$$\mathcal{I}(\phi)_{j_1,j_2} = \sum_{k,\ell} w_{k,\ell}\frac{\partial i_{k,\ell}(\phi)}{\partial\phi_{j_1}}\frac{\partial i_{k,\ell}(\phi)}{\partial\phi_{j_2}}.$$

The gradient of of $i_{k,\ell}(\phi)$ is given by:

$$\nabla i_{k,\ell}\left(\phi\right) = \begin{pmatrix} \|v_{k,\ell}\|^2_{\mathbb{C}^2}, \\ |\langle v_{k,\ell}, c_\theta\rangle_{\mathbb{C}^2}|^2, \\ i_{\mathrm{p}}\dfrac{\partial|\langle v_{k,\ell}, c_\theta\rangle_{\mathbb{C}^2}|^2}{\partial\theta}. \end{pmatrix} \tag{13}$$

Finally, the Fisher information is given by $\mathcal{I}\left(\phi\right) = \nabla i_{k,\ell}\left(\phi\right).\nabla i_{k,\ell}\left(\phi\right)^\top$.

ACKNOWLEDGMENTS

This work was supported by the Programme National de Planétologie (PNP) of CNRS-INSU co-funded by CNES. This work has made use of the SPHERE Data Centre, jointly operated by OSUG/IPAG (Grenoble), PYTHEAS /LAM/CeSAM (Marseille), OCA/Lagrange (Nice) and Observatoire de Paris/LESIA (Paris).

REFERENCES

[1] Avenhaus et al., "Structures in the protoplanetary disk of hd142527 seen in polarized scattered light," *APJ* (2013).
[2] Tinbergen, J., [*Astronomical polarimetry*], Cambridge university press (2005).
[3] Gerrard, A. and Burch, J., [*Introduction to matrix methods in optics*], John Wiley & Sons (1994).
[4] Brent, R., "Algorithms for minimization without derivatives," (1973).
[5] Langlois et al., "First scattered light detection of a nearly edge on circumstellar disk around the t tauri star ry lup," *A & A* (2017).

Reconstruction polarimétrique d'environnements circumstellaires à partir des données ESO/VLT-SPHERE IRDIS.

Laurence Denneulin[1,2], Maud Langlois[1], Nelly Pustelnik[2], Éric Thiébaut[1]

[1] Univ Lyon, Univ Lyon1, Ens de Lyon, CNRS, Centre de Recherche Astrophysique de Lyon UMR5574

[2]Université de Lyon, ENS de Lyon, CNRS, Laboratoire de Physique

laurence.denneulin@univ-lyon1.fr

Résumé – La reconstruction des environnements des étoiles jeunes par imagerie directe est un défi car leur émission est noyée dans les résidus de lumière stellaire malgré l'utilisation d'un coronographe. Il est cependant possible de les démélanger en exploitant leurs différences d'états de polarisation. Dans ce papier, nous présentons une méthode régularisée de reconstruction des environnements circumstellaires polarisés, à partir des données d'IRDIS (*Infrared Dual Imager and Spectrograph*), instrument de SPHERE (*Spectro-Polarimetric High-contrast Exoplanet REsearch*) installé sur l'un des télescopes du VLT (*Very Large Tescope*) de l'ESO (*European Southern Observatory*). Nous discutons des mérites de différentes régularisations selon le cas astrophysique étudié.

Abstract – The circumstellar environments reconstruction by direct imaging is challenging because of the dominating residual star light in spite of the coronograph. Yet, separatation of both signals is possible by using their polarization states differences. In this paper, we present a regularized reconstruction method of the circumstellar environments of ESO/VLT-SPHERE IRDIS's polarimetrical data. We then discuss the merit of various regularizations depending on specific astrophysical cases.

1 Introduction

Ces dernières décennies, l'évolution de l'instrumentation des très grands télescopes a occasionné d'importantes avancées dans l'étude des environnements circumstellaires et la découverte d'exoplanètes. Du fait du grand contraste avec l'étoile requis pour les observer, la reconstruction des environnements reste cependant un défi majeur. En effet, sur le détecteur, le signal d'intérêt est largement dominé par les résidus de lumière stellaire liés aux aberrations de phase. En exploitant les différences d'état de polarisation du signal d'intérêt et des fuites stellaires, l'imagerie en polarimétrie devrait permettre de procéder au nécessaire démélange.

Les méthodes de l'état de l'art de reconstructions polarimétriques à haut contraste (double ratio et double différence [1]), en procédant par combinaison directe des données, sont plus simples que les algorithmes développés pour d'autres modalités d'imagerie polarimétriques (séparation de sources [2, 3] ou approche parcimonieuse [4]). Nous proposons de reconstruire le signal d'intérêt par approche inverse en maximisant la vraisemblance des données polarimétriques connaissant leur modèle sous contraintes. Le modèle des données prenant en compte la convolution par la réponse instrumentale (PSF), une régularisation est nécessaire pour éviter l'amplification du bruit dans la reconstruction.

Le nombre et les performances des algorithmes de déconvolution n'a cessé d'augmenter. Les méthodes de type Quasi-Newton [5] sont très rapides mais ne peuvent être appliquées pour des pénalisations non-lisses. Dans ce cas, les méthodes telles que l'algorithme explicite-implicite (forward-backward) ou Douglas-Rachford sont à même d'obtenir une solution [6, 7]. Les algorithmes de type primaux-duaux [8, 9] permettent de combiner facilement plusieurs termes non-différentiables et des opérateurs linéaires. Nous avons choisi l'algorithme de Condat-Vũ [8] car il prend en compte des fonctions différentiables à gradient Lipschitz auxquelles s'ajoutent autant de contraintes que nécessaire tout en gardant la même structure algorithmique. Ceci rend les comparaisons plus aisées car le même algorithme peut être utilisé pour une grande diversité de régularisations. Compte tenu du type d'objet d'intérêt, nous avons choisi d'exploiter des régularisations telles que la Variation Totale (TV), la norme de Shatten sur l'opérateur hessien [10], la Variation Totale Généralisée (TGV) [11] où une pénalisation hybride TV+ℓ_1 à l'instar de pénalisation utilisées dans l'analyse d'images texture+géométrie [12, 13].

L'article est organisé ainsi : Nous décrivons d'abord notre modélisation des données d'après la physique de l'instrument. Nous énonçons ensuite la fonction objectif que nous minimisons et l'algorithme associé. Nous présentons enfin les résultats obtenus sur des données simulées et sous différentes régularisations, la qualité des résultats étant jugée d'après le taux de polarisation retrouvé et le niveau de contraste.

2 Méthodologie

La polarisation de la lumière décrit le comportement de son champ électrique au cours du temps et peut être représentée

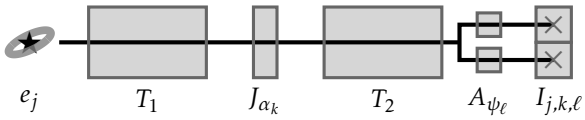


FIGURE 1 – Schéma simplifié de l'instrument ESO/VLT-SPHERE IRDIS. En entrant dans le télescope, le champ électrique e_j est transmis par un premier ensemble de composants optiques, dont l'effet est représenté par la matrice de Jones T_1, puis par une lame demi-onde, représentée par la matrice J_{α_k}, ensuite par un second ensemble de composants optiques, représenté par la matrice T_2, avant d'être divisé en deux. Les champs électriques résultants sont ensuite projetés par deux analyseurs perpendiculaires, représentés par les vecteurs de Jones A_{ψ_ℓ} (pour $\ell \in \{1,2\}$), pour enfin atteindre chacun une partie différente du détecteur qui enregistre l'intensité associée $I_{j,k,\ell}$, en nombre de photo-électrons, au pixel j de l'image numéro $k \in \{1,\dots,n\}$.

par un vecteur complexe, normal à la direction de propagation de l'onde, composé d'une amplitude et d'une phase. La lumière de l'étoile est *non-polarisée*, l'amplitude et la phase de son vecteur de polarisation variant aléatoirement au cours du temps. Une partie la lumière de l'environnement est issue de la réflexion de la lumière de l'étoile sur la poussière qui compose l'environnement. Elle est de ce fait *partiellement polarisée linéairement* : l'amplitude et la phase du vecteur de polarisation de la composante linéairement polarisée sont constantes au cours du temps et ce vecteur peut s'écrire $(a\cos\theta, a\sin\theta)^\top$, où $a \in \mathbb{R}$ est l'amplitude et $\theta \in [0, 2\pi[$ la phase ou l'angle de polarisation linéaire.

2.1 Modèle direct

Soit $(e_j)_{j\in\{1,\dots,N\times N\}}$, la somme de deux champs électriques, non-polarisé et polarisé linéairement, entrant dans l'instrument ESO/VLT-SPHERE IRDIS et atteignant un pixel j du détecteur. Les transformations s'appliquant à e_j dans l'instrument sont représentées par des matrices de Jones, sous-ensemble de $\mathcal{M}_2(\mathbb{C})$ (cf. figure 1). La lame demi-onde module la polarisation selon le cycle d'angles $\alpha_k \in \{0, \pi/2, \pi/4, 3\pi/4\}^{n/4}$ où k est le numéro d'image de la séquence. Lors du passage dans les analyseurs, le champ électrique polarisé ainsi modulé est simplement projeté par $A_{\psi_\ell} = (\cos\psi_\ell, -\sin\psi_\ell)^\top$, où $\psi_1 = 0$ et $\psi_2 = \pi/2$, tandis que le champ électrique non-polarisé est moyenné. Cette moyenne est identique quelque soit le numéro d'image k, permettant ainsi le démélange. L'intensité acquise en un pixel j du détecteur pour une orientation α_k de la lame demi-onde et un angle ψ_ℓ d'analyseur est alors donnée par :

$$I_{j,k,\ell} = \mathbb{E}_t\left[|\langle A_{\psi_\ell}, T_2 J_{\alpha_k} T_1 e_j\rangle_{\mathbb{C}^2}|^2\right], \tag{1}$$

où $\mathbb{E}_t[\dots]$ dénote la moyenne temporelle sur le temps de pose. En posant $v^{k,\ell} = (T_2 J_{\alpha_k} T_1)^* \cdot A_{\psi_\ell}$, où * représente l'opérateur adjoint, et en développant l'expression, on obtient :

$$I_{j,k,\ell} = \frac{1}{2}\|v_{k,\ell}\|^2_{\mathbb{C}^2} I^{\mathrm{u}}_j + |\langle v_{k,\ell}, c_{\theta_j}\rangle_{\mathbb{C}^2}|^2 I^{\mathrm{p}}_j, \tag{2}$$

où $I^{\mathrm{u}} \in \mathbb{R}_+^{N\times N}$ et $I^{\mathrm{p}} \in \mathbb{R}_+^{N\times N}$ sont les cartes d'intensités non-polarisées et polarisées linéairement (soit les moyennes temporelles quadratiques des champs électriques non-polarisés et polarisés pendant la pose) et $c_{\theta_j} \equiv (\cos\theta_j, \sin\theta_j)^\top$ avec $\theta_j \in [0, 2\pi[$ l'angle de polarisation au pixel j.

L'intensité acquise en un pixel j du détecteur suit une loi de Poisson de paramètre $I_{j,k,\ell}$ exprimé en photo-électrons par pixel par image. Ce nombre étant élevé, nous approximons la distribution du bruit par une loi gaussienne indépendante centrée de variance $I_{j,k,\ell}$. Au bruit de Poisson s'ajoute un bruit de lecture supposé centré gaussien de variance $\sigma^2_{\mathrm{ro}\,j}$. D'où le modèle des données suivant :

$$D_{j,k,\ell} = I_{j,k,\ell} + \beta_{j,k,\ell}, \tag{3}$$

où $\beta_{j,k,\ell} \sim \mathcal{N}(0, I_{j,k,\ell} + \sigma^2_{\mathrm{ro}\,j})$ est un terme de bruit. Il est à noter qu'un terme d'attache aux données de type divergence de Kullback semblerait plus approprié même si cette approximation possède des limitations [14]. En pratique, l'approximation gaussienne proposée ici conduit à une pénalisation quadratique donnant de bon résultats en termes de qualité de reconstruction pour des calculs plus rapides [15]. Sans prendre en compte la déconvolution, les paramètres I^{u}_j, I^{p}_j et θ_j pourraient être estimés en maximisant la vraisemblance des données connaissant le modèle ce qui revient à résoudre un problème non-linéaire mais séparable et s'écrivant de façon hiérarchique :

$$\min_{\theta_j}\left\{\min_{I^{\mathrm{u}}_j\geq 0, I^{\mathrm{p}}_j\geq 0} \sum_{k,\ell} w_{j,k,\ell}\left(I_{j,k,\ell}(I^{\mathrm{u}}_j, I^{\mathrm{p}}_j, \theta_j) - D_{j,k,\ell}\right)^2\right\}, \tag{4}$$

pour tout pixel j et avec $w_{j,k,\ell} = \left(I_{j,k,\ell}(I^{\mathrm{u}}_j, I^{\mathrm{p}}_j, \theta_j) + \sigma^2_{\mathrm{ro}\,j}\right)^{-1}$. La possibilité de prendre en compte les données manquantes grâce aux poids $w_{j,k,\ell}$ est un avantage sur les méthodes de l'état de l'art qui ne peuvent fonctionner que sur un cycle complet d'angle de lame demi-onde.

2.2 Modèle linéarisé

La non-linéarité du modèle que nous considérons devient problématique lorsque que la convolution par la PSF est prise en compte. Le problème n'étant alors plus séparable pixel-à-pixel, il est bien plus lourd à résoudre en terme de complexité algorithmique. Nous proposons donc une version linéarisée du modèle en les paramètres de Stokes (I, Q, U) qui s'écrit :

$$I_{j,k,\ell}(\mathrm{I}_j, \mathrm{Q}_j, \mathrm{U}_j) = \mathrm{v}^{\mathrm{norm}}_{k,\ell}\,\mathrm{I}_j + \mathrm{v}^{\mathrm{diff}}_{k,\ell}\,\mathrm{Q}_j + \mathrm{v}^{\mathrm{real}}_{k,\ell}\,\mathrm{U}_j, \tag{5}$$

$$\text{où } \begin{cases}\mathrm{I}_j = I^{\mathrm{u}}_j + I^{\mathrm{p}}_j,\\ \mathrm{Q}_j = I^{\mathrm{p}}_j\cos 2\theta_j,\\ \mathrm{U}_j = I^{\mathrm{p}}_j\sin 2\theta_j,\end{cases} \quad \text{et} \quad \begin{cases}\mathrm{v}^{\mathrm{norm}}_{k,\ell} = (1/2)\|v_{k,\ell}\|^2_{\mathbb{C}^2},\\ \mathrm{v}^{\mathrm{diff}}_{k,\ell} = (1/2)\left(|v^x_{k,\ell}|^2 - |v^y_{k,\ell}|^2\right),\\ \mathrm{v}^{\mathrm{real}}_{k,\ell} = \Re\mathfrak{e}\left(v^x_{k,\ell}v^y_{k,\ell}\right).\end{cases}$$

La relation inverse s'écrit :

$$\begin{cases} I^{\mathrm{p}}_j = \sqrt{\mathrm{Q}^2_j + \mathrm{U}^2_j},\\ \theta_j = (1/2)\arctan(\mathrm{U}_j/\mathrm{Q}_j),\\ I^{\mathrm{u}}_j = \mathrm{I}_j - I^{\mathrm{p}}_j.\end{cases} \tag{6}$$

2.3 Fonction objectif

Afin de passer du domaine des cartes reconstruites au domaine des données, on définit $\mathrm{V}^{\cdots}_k = \mathrm{V}^{\cdots}_{k,1}K_1 + \mathrm{V}^{\cdots}_{k,2}K_2 \in \mathbb{R}^{N\times 2N}$

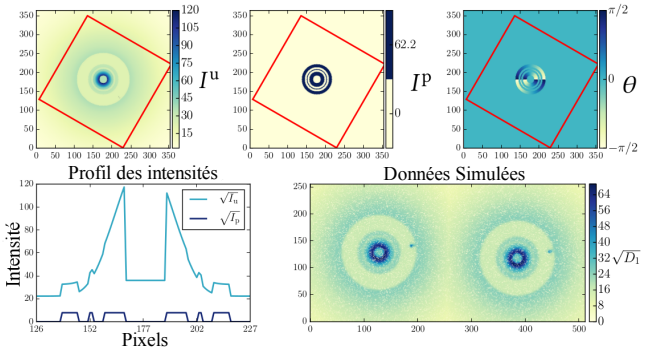


FIGURE 2 – Simulations pour $\tau_{\text{disk}} = 10\%$. Sur la première ligne, de gauche à droite, sont affichées les cartes I^{u}, I^{p} et θ. En dessous sont présentées, à gauche les profils de I^{u} et I^{p}, montrant la différence importante d'intensité entre les deux composantes, à droite les données simulées à partir des cartes.

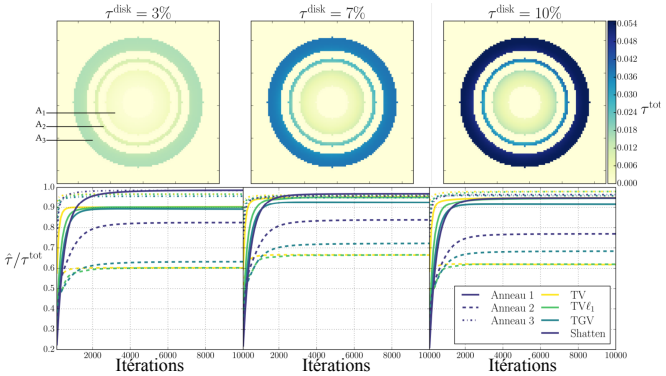


FIGURE 3 – Évolution de $\widehat{\tau}/\tau_{\text{tot}}$ moyens sur chaque anneaux pour les différents τ_{disk}. Sur la première ligne sont affichées les cartes des τ_{tot} pour les différents τ_{disk}. En dessous sont tracés les évolutions des $\widehat{\tau}/\tau_{\text{tot}}$ moyens associés, pour chaque régularisation, en fonction des itérations.

où $K_1 = (1,0)^\top \otimes \mathrm{Id}_N$ et $K_2 = (0,1)^\top \otimes \mathrm{Id}_N$, $\otimes$ représentant le produit de Kronecker. Le terme d'attache aux données correspond alors à la norme de Mahalanobis :

$$\sum_{k=1}^{n} \left\| D_k - \mathcal{H}\left(\mathrm{I}\cdot \mathrm{V}_k^{\text{norm}} + \mathrm{Q}\cdot \mathrm{V}_k^{\text{diff}} + \mathrm{U}\cdot \mathrm{V}_k^{\text{real}}\right)\right\|_{W_k}^2, \qquad (7)$$

où $W_k = w_{k,1}K_1 + w_{k,2}K_2$. L'opérateur $\mathcal{H}$ contient les transformations linéaires s'appliquant à toutes les cartes (translation, rotation du champ et convolution par la PSF).

Quelque soit le pixel j, la contrainte de positivité sur les intensités impose que $I_j^{\mathrm{u}} \geq 0$ et $I_j^{\mathrm{p}} \geq 0$. D'après l'expression de I_j^{p} dans (6), cette dernière inégalité est toujours vraie par construction. Pour que la première inégalité soit vérifiée il faut explicitement imposer que $\mathrm{I}_j \geq I_j^{\mathrm{p}}$ ce qui revient à imposer la contrainte épigraphique $\mathrm{I}_j \geq \left(\mathrm{Q}_j^2 + \mathrm{U}_j^2\right)^{1/2} = \|(\mathrm{Q}_j, \mathrm{U}_j)^\top\|_2$.

Du fait de la nature relativement lisse des disques circumstellaires, nous avons choisi d'utiliser les régularisations TV, Shatten [10], qui évite les aplats, et TGV [11], qui est compromis entre TV et Shatten. Sous ces régularisations, la restitution de points sources très brillants est cependant difficile et peut résulter en une perte d'information autour des points sources ou en une sous-régularisation du problème. Nous proposons donc une pénalisation hybride TV+ℓ_1 qui consiste à séparer les termes à estimer en une composante lisse et une composant parcimonieuse, *e.g.* $X = X_1 + X_2$, et à régulariser par $\mathrm{TV}(X_1) + \|X_2\|_1$, permettant la déconvolution des points sources sans perte d'information.

Finalement, la fonction objectif du problème à résoudre est :

$$\min_{\mathrm{I,Q,U}} \left\{ \begin{array}{l} \frac{1}{2}\sum_{k=1}^{n} \left\| D_k - \mathcal{H}\left(\mathrm{I}\cdot \mathrm{V}_k^{\text{norm}} + \mathrm{Q}\cdot \mathrm{V}_k^{\text{diff}} + \mathrm{U}\cdot \mathrm{V}_k^{\text{real}}\right)\right\|_{W_k}^2 \\ + \lambda_1 \mathcal{L}_1(\mathrm{I}) + \lambda_2 \mathcal{L}_2(\mathrm{Q}) + \lambda_3 \mathcal{L}_3(\mathrm{U}) \\ + \iota_{\mathrm{I} \geq \|(\mathrm{Q,U})\|_2}(\mathrm{I,Q,U}) \end{array} \right. \qquad (8)$$

où $\mathcal{L}_1, \mathcal{L}_2$ et $\mathcal{L}_3$ sont les régularisation appliquées à I, Q et U, de paramètres respectifs λ_1, λ_2 et λ_3. Après résolution de l'équation (8) avec l'algorithme proximal primal-dual de Condat-Vũ [8], retrouver les paramètres d'intérêt revient à appliquer (6).

3 Résultats

L'intérêt de notre approche est double : d'une part, restituer précisément les points sources et, d'autre part, estimer au mieux l'intensité polarisée. Pour ce second critère, nous avons donc choisi de juger l'efficacité de la régularisation d'après le taux de polarisation estimé. Nous avons simulé des données pour un taux de polarisation $\tau_{\text{disk}} = I^{\mathrm{p}}/(I^{\mathrm{u}}_{\text{disk}} + I^{\mathrm{p}})$ donné pour la lumière de l'environnement. Ce taux n'est pas estimable en pratique car $I^{\mathrm{u}}_{\text{disk}}$ est comprise dans I^{u}, aussi nous définissons $\tau_{\text{tot}} = I^{\mathrm{p}}/(I^{\mathrm{u}} + I^{\mathrm{p}})$, estimable sur des données réelles. Un rapport moyen entre $\widehat{\tau} = \widehat{I^{\mathrm{p}}}/(\widehat{I^{\mathrm{u}}} + \widehat{I^{\mathrm{p}}})$ estimé et τ_{tot} implémenté proche de 1 sera notre critère de qualité. Les jeux de données simulés pour τ_{disk} valant 10%, 7% et 3%, correspondent respectivement à un cas simple, moyen et difficile, une intensité polarisée faible étant plus facilement noyée dans le bruit de photons. Les données sont composées de trois anneaux $\{\mathrm{A}_1, \mathrm{A}_2, \mathrm{A}_3\}$, de mêmes intensités mais de τ_{tot} et de taille différentes. Deux points sources sont ajoutés dans I^{u}. Les images simulées sont ensuite combinées d'après (2), convoluées, bruitées et altérées par l'ajout de 10% de pixels non valides pris au hasard. Chaque jeu de données est constitué de $n = 64$ réalisations du bruit. La figure 2 montre les cartes créées et les données simulées associées dans le cas $\tau_{\text{disk}} = 10\%$.

Sur la figure 3, le taux estimé en régularisant par Shatten apparaît plus précis pour les cas difficiles, *i.e.* un taux de polarisation faible ou un anneau fin. À l'inverse, pour les zones étendues avec un taux de polarisation élevé, *e.g.* A_3 avec $\tau_{\text{disk}} = 10\%$, TV et TV+$\ell_1$ donnent les meilleurs résultats. Si TGV donne de bon résultats visuellement au niveau des zones étendues, comme on peut le voir pour I^{p} sur la figure 4, il n'aboutit pas à la meilleure restitution de τ_{tot}. La table 1 permet de confirmer cela. On voit également que les méthodes de l'état de l'art ont tendance à surestimer le taux de polarisation dans les cas les plus simples. Pour ce qui est de la déconvolution des points sources, la figure 4 montre l'efficacité de TV+ℓ_1 par rapport aux autres régularisations. Les points sources sont bien restitués avec des valeurs très proches des valeurs vraies.

TABLE 1 – SNR moyens de chaque anneaux, avec SNR $= I^{\mathrm{p}}(n/(I^{\mathrm{u}} + I^{\mathrm{p}} + \sigma_{\mathrm{ro}}^2))^{1/2}$, et rapports $\widehat{\tau}/\tau_{\mathrm{tot}}$ estimés avec les différentes régularisations, la méthode non linéaire et les méthodes de l'état de l'art.

τ_{disk}		SNR	TV	TGV	Shatten	TV-ℓ_1	Non-Lin.	Dbl. Ratio	Dbl. Diff.
3 %	A_1	1,69	0,904	0,890	**0,986**	0,903	0.886	0.956	0.956
	A_2	3,23	0,602	0,633	**0,825**	0,602	0.492	0.499	0.493
	A_3	4,08	0,967	0,957	**0,985**	0,967	0.860	0.889	0.859
7 %	A_1	4,11	0,955	0,924	**0,968**	0,954	0.941	1.086	1.086
	A_2	7,80	0,665	0,725	**0,839**	0,662	0.540	0.558	0.539
	A_3	9,68	**0,962**	0,950	0,957	0,962	0.855	0.899	0.853
10%	A_1	6,05	**0,950**	0,914	0,948	0,950	1.293	1.570	1.570
	A_2	11,45	0,618	0,685	**0,770**	0,618	0.697	0.720	0.688
	A_3	14,17	**0,980**	0,959	0,964	0,980	0.912	1.037	0.908

4 Conclusions et perspectives

Nous avons donc mis au point un modèle linéaire des données palarimétiques de l'instrument ESO-VLT/SPHERE IRDIS permettant de résoudre les problèmes de reconstruction et de démélange associés avec différents termes de régularisations. Nous avons comparé leurs efficacités en étudiant le rapport entre taux de polarisation estimés et taux simulés. Nous avons remarqué que pour les cas difficiles, les meilleurs reconstructions sont obtenues avec une régularisation par Shatten, tandis que pour les cas faciles, c'est TV qui donne le meilleur résultat. Enfin, nous avons vu qu'une alliance de TV et ℓ_1 permettait une meilleure restitution des points sources.

Dans l'optique d'améliorer la méthode, *e.g.* pour les disques vus de face autour d'une seule étoile, il est possible d'inclure un *a priori* sur l'angle de polarisation se traduisant par l'ajout d'une contrainte angulaire, améliorant ainsi l'estimation du taux de polarisation. Le taux de polarisation estimé dépendant également des paramètres de régularisation, une étude plus fine de cette dépendance doit être menée. Cependant, la convergence étant lente et dépendant des poids de la norme de Mahalanobis, une accélération de la méthode par préconditionnement est également envisagée.

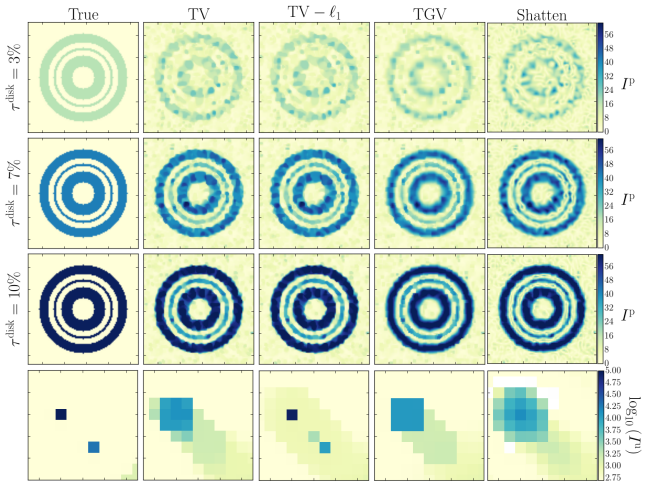


FIGURE 4 – Comparaison des cartes de I^{p} obtenues pour les différentes régularisation en fonction de τ_{disk}. Sur la dernière ligne, comparaison de la restitution des points sources.

Références

[1] H. Avenhaus et al. : Structures in the protoplanetary disk of HD142527 seen in polarized scattered light. *The Astrophysical Journal*, 781(2), p. 87, janv. 2014.

[2] M. Tristram et al. : Iterative destriping and photometric calibration for Planck -HFI, polarized, multi-detector map-making. *Astronomy & Astrophysics*, 534, p. A88, oct. 2011.

[3] Planck Collaboration et al. : Planck 2015 results : VIII. High Frequency Instrument data processing: Calibration and maps. *Astronomy & Astrophysics*, 594, p. A8, oct. 2016.

[4] J. Birdi, A. Repetti, et Y. Wiaux : Sparse interferometric Stokes imaging under polarization constraint (Polarized SARA), *Monthly Notices of the Royal Astronomical Society*, 2018.

[5] S.J. Benson et J.J. Moré : A limited memory variable metric method in subspaces and bound constrained optimization problems. *in Subspaces and Bound Constrained Optimization Problems*, 2001.

[6] H. H. Bauschke et P. L. Combettes : *Convex analysis and monotone operator theory in Hilbert spaces*. New York : Springer, 2011.

[7] N. Pustelnik, A. Benazza-Benhayia, Y. Zheng, J.-C. Pesquet : Wavelet-based Image Deconvolution and Reconstruction. *Wiley Encyclopedia of Electrical and Electronics Engineering*, Feb. 2016.

[8] L. Condat : A Primal–Dual Splitting Method for Convex Optimization Involving Lipschitzian, Proximable and Linear Composite Terms. *Journal of Optimization Theory and Applications*, 158(2), p. 460-479, août 2013.

[9] A. Chambolle et T. Pock, A First-Order Primal-Dual : Algorithm for Convex Problems with Applications to Imaging. *Journal of Mathematical Imaging and Vision*, 40(1), p. 120-145, mai 2011.

[10] S. Lefkimmiatis, J.P. Ward, M. Unser : Hessian Schatten-Norm Regularization for Linear Inverse Problems. *IEEE Transactions on Image Processing*, 22(5), p.1873–1888, mai 2003.

[11] K. Bredies, K. Kunisch, et T. Pock : Total Generalized Variation. *SIAM Journal on Imaging Sciences*, 3(3), p. 492-526, janv. 2010.

[12] L. M. Briceño-Arias, P. L. Combettes, J.-C. Pesquet, and N. Pustelnik : Proximal algorithms for multicomponent image processing. *Journal of Mathematical Imaging and Vision*, 41(1), p. 3-22, Sep. 2011.

[13] Jean-Francois Aujol, Guy Gilboa, Tony Chan, and Stanley Osher : Structure-Texture Image Decomposition - Modeling, Algorithms, and Parameter Selection. *International Journal of Computer Vision*, 67(1), p. 111-136, Avril 2006.

[14] R. Gribonval : Should Penalized Least Squares Regression be Interpreted as Maximum A Posteriori Estimation? *IEEE Transactions on Signal Processing*, 59(5), p. 2405-2410, mai 2011.

[15] J. Boulanger, N. Pustelnik, L. Condat, T. Piolot, L. Sengmanivong : Nonsmooth convex optimization for Structured Illumination Microscopy image reconstruction. *Inverse Problems*, 34(9), 2018.

Primal-dual splitting scheme with backtracking for handling with epigraphic constraint and sparse analysis regularization.

Laurence Denneulin[1,2], Nelly Pustelnik[2,3], Maud Langlois[1], Ignace Loris[4] and Éric Thiébaut[1].

[1]Univ Lyon, Univ Lyon1, ENS de Lyon, CNRS, Centre de Recherche Astrophysique de Lyon UMR5574, F-69230 Saint-Genis-Laval, France

[2]Univ Lyon, ENS de Lyon, Univ Claude Bernard Lyon 1, CNRS, Laboratoire de Physique,F-69342 Lyon, France

[3]ISPGroup & INMA/ICTEAM, UCLouvain, Belgium

[4]Département de Mathématique, Université libre de Bruxelles, Boulevard duTriomphe, 1050 Bruxelles, Belgium

***Abstract*— The convergence of many proximal algorithms involving a gradient descent relies on its Lipschitz constant. To avoid computing it, backtracking rules can be used. While such a rule has already been designed for the forward-backward algorithm (FBwB), this scheme is not flexible enough when a non-differentiable penalization with a linear operator is added to a constraint. In this work, we propose a backtracking rule for the primal-dual scheme (PDwB), and evaluate its performance for the epigraphical constrained high dynamical reconstruction in high contrast polarimetric imaging, under TV penalization.**

1 Introduction

The resolution of inverse problems remains a challenging task in image processing, especially when dealing with a large amount of data, such as in astrophysics (e.g. 10^6 to 10^9 pixels). Important advances have been made for handling non-differentiable objective function, thanks to proximal algorithmic schemes but an important issue is the impact on the convergence behaviour of the Lipschitz constant of the gradient. Yet, the calculus of this constant can be time consuming or difficult. To get round this issue, a backtracking rule can be used. Such a rule has been designed for forward-backard iterations in [1] but for many inverse problems forward-backward iterations are not flexible enough to handle complex regularization terms and/or constraints. We then need to resort to primal-dual schemes [2] for which we propose to design a backtracking rule.

Equipped with a backtracking rule for both forward-backward and primal-dual schemes we propose to evaluate the reconstruction performances of Total Variation (TV) [3] with standard regularization procedure considered in astrophysics that is hyperbolic Total Variation (TV-h) [4] regularization.

To evaluate the performance, we focus on high contrast polarimetric imagery which benefits in considering jointly a TV-based penalization and an epigraphic constraint. Indeed, if epigraphical constraint has been considered in polarimetric radio-interferometry [5], in high contrast polarimetric direct imaging, the state-of-the-art does not take it in account [6].

Section 2 introduces the notations and the objective function we are interested in. Section 3 presents the proposed backtracking rule for primal-dual proximal schemes and convergence results. Section 4 provides the direct model considered in high contrast polarimetric imagery, provides some recalls on TV and TV-h as well as experimental comparisons.

2 Problem formulation

We denote by $\boldsymbol{x} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_L) \in (\mathrm{R}^N)^L$ the L-component signal of interest, each of size N. Our goal is to estimate:

$$\widehat{\boldsymbol{x}} \in \underset{\boldsymbol{x}\in(\mathbb{R}^N)^L}{\mathrm{Argmin}} \{h(\boldsymbol{x}) + \sum_{\ell=1}^{L} g_\ell(\mathrm{D}_\ell \boldsymbol{x}_\ell) + \iota_C(\boldsymbol{x})\}. \quad (1)$$

where $h : (\mathbb{R}^N)^L \to]-\infty, +\infty]$ is a convex and differentiable function with a β-Lipschitz gradient (may denote the data-fidelity term), $\forall \ell = \{1, \dots, L\}$, $\mathrm{D}_\ell \in \mathbb{R}^{K_\ell \times N}$ denotes a linear operator, and $g_\ell : \mathbb{R}^{K_\ell} \to]-\infty, +\infty]$ is a proper, lower semi-continuous (l.s.c.), convex function (may stands for the regularization term, including TV, as well as TV-h in the differentiable case). See [7, 8] for an exhaustive list of penalization choices having this form. Finally, $\iota_C : (\mathbb{R}^N)^L \to \mathbb{R}$ is an epigraphical constraint, written as:

$$C = \{(\boldsymbol{x}_1, \dots, \boldsymbol{x}_L) \in \left(\mathbb{R}^N\right)^L \mid \phi(\boldsymbol{x}_2, \dots, \boldsymbol{x}_L) \leq \boldsymbol{x}_1\} \quad (2)$$

where ϕ proper, l.s.c and convex (cf. e.g. [9]).

3 Backtracking proximal primal-dual

When g_ℓ is differentiable, forward-backward scheme, possibly with backtracking as in [1], can be considered to estimate $\widehat{\boldsymbol{x}}$. When g_ℓ is non-differentiable, a well adapted scheme is the primal-dual algorithm [2], whose main interest is to exploit the differentiability of h and relies on proximal steps for g_ℓ and ι_C. Setting $g(\mathbf{D}\boldsymbol{x}) = \sum_{\ell=1}^{L} g_\ell(\mathrm{D}_\ell \boldsymbol{x}_\ell)$, the iterations are summarized in Algorithm 1. The sequence $(\boldsymbol{x}^{[t]})_{t\in\mathbb{N}}$ is insured to converge to $\widehat{\boldsymbol{x}}$, if the following condition on the parameters $\tau^{[t]}, \sigma^{[t]} \geq 0$ involving the Lipschitz constant $\beta > 0$ holds:

$$1/\tau^{[t]} - \sigma^{[t]}\|\mathbf{D}\|^2 \geq \beta/2. \quad (3)$$

Algorithm 1: Primal-Dual (PD) Condat-Vũ algorithm

Set $\boldsymbol{x}^{[0]} \in (\mathbb{R}^N)^L$ and $\boldsymbol{y}^{[0]} \in \mathbb{R}^{K_1} \times \ldots \times \mathbb{R}^{K_L}$.
for $t = 0, 1, \dots$ **do**
  Set $\tau^{[t]}, \sigma^{[t]} \geq 0$ such that (3) holds.
  $\boldsymbol{x}^{[t+1]} = \mathrm{prox}_{\tau^{[t]}\iota_C}\left(\boldsymbol{x}^{[t]} - \tau^{[t]}\left(\nabla h(\boldsymbol{x}^{[t]}) + \mathbf{D}^*\boldsymbol{y}^{[t]}\right)\right)$;
  $\boldsymbol{y}^{[t+1]} = \mathrm{prox}_{\sigma^{[t]}g^*}\left(\boldsymbol{y}^{[t]} + \sigma^{[t]}\mathbf{D}\left(2\boldsymbol{x}^{[t+1]} - \boldsymbol{x}^{[t]}\right)\right)$.

In the case where β is unknown, we need to resort to backtracking scheme, whose proposed iterations are described in Algorithm 2. The main idea is to start from a small estimate $\mathring{\beta}^{[0]} > 0$ of β. Then at each iterations, to test whether the candidate $\mathring{\boldsymbol{x}}^{[i]}$ yields a reduction of the majorant quadratic approximation of (1) tangent to the current iterate $\boldsymbol{x}^{[t]}$, according to $\mathring{\beta}^{[i]}$. If the condition holds, $\beta^{[t]}$ is updated with $\mathring{\beta}^{[i]}$, else $\mathring{\beta}^{[i]}$ is increased. With such a condition, $\exists t \in \mathbb{N}$ such that $\beta^{[t]} \geq \beta$. The sequence $(\boldsymbol{x}^{[t]})_{t\in\mathbb{N}}$ generated by Algorithm 2 thus converges to $\widehat{\boldsymbol{x}}$.

4 Experiments

High contrast polarimetric imagery – We evaluate the performance of the aglorithm 2 to reconstruct circumstellar environments images using data from the Dual-Polarization Imaging (DPI) [10] modality of the SPHERE/IRDIS instrument [11, 12] installed at the Very Large Telescope (VLT) of the European Southern Observatory (ESO).

Algorithm 2: Primal-Dual with Backtracking (PDwB)

Set $\boldsymbol{x}^{[0]} \in (\mathbb{R}^N)^L$ and $\boldsymbol{y}^{[0]} \in \mathbb{R}^{K_1} \times \ldots \times \mathbb{R}^{K_L}$, $\beta^{[0]} \geq 0$ and $\eta > 1$:

for $t = 0, 1, \ldots$ **do**

  for $i = 0, 1, \ldots$ **do**

    $\mathring{\beta}^{[i]} = \eta^i \beta^{[t]}$ and $\mathring{\tau}^{[i]}, \mathring{\sigma}^{[i]} \geq 0$ such that (3) holds.

    $\mathring{\boldsymbol{x}}^{[i]} = \operatorname{prox}_{\mathring{\tau}^{[i]} \iota_C}\left(\boldsymbol{x}^{[t]} - \mathring{\tau}^{[i]}\left(\nabla h(\boldsymbol{x}^{[t]}) + \mathbf{D}^* \boldsymbol{y}^{[t]}\right)\right)$;

    $\mathring{\boldsymbol{y}}^{[i]} = \operatorname{prox}_{\mathring{\sigma}^{[i]} g^*}\left(\boldsymbol{y}^{[t]} + \mathring{\sigma}^{[i]} \mathbf{D}\left(2\mathring{\boldsymbol{x}}^{[i]} - \boldsymbol{x}^{[t]}\right)\right)$;

    if $h(\mathring{\boldsymbol{x}}^{[i]}) \geq h(\boldsymbol{x}^{[t]}) + \langle \mathring{\boldsymbol{x}}^{[i]} - \boldsymbol{x}^{[t]}, \nabla h(\boldsymbol{x}^{[t]}) \rangle + \frac{\mathring{\beta}^{[i]}}{2} \|\mathring{\boldsymbol{x}}^{[i]} - \boldsymbol{x}^{[t]}\|^2$ **then**

      $\beta^{[t+1]} = \mathring{\beta}^{[i]}$, $\boldsymbol{x}^{[t+1]} = \mathring{\boldsymbol{x}}^{[i]}$ and $\boldsymbol{y}^{[t+1]} = \mathring{\boldsymbol{y}}^{[i]}$.

      break

Direct model – Observations consist in data cubes $\boldsymbol{d} \in (\mathbb{R}^M)^K$ with $M = 1024 \times 2048$ and K a multiple of the four polarisation modulations in the instrument (*e.g.* $K = 64$ to $K > 512$ depending on the object). The $L = 3$ components to estimate (e.g. $\widehat{\boldsymbol{x}}$) corresponds to three Stokes parameters $(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3) = (\mathrm{I}, \mathrm{Q}, \mathrm{U})$, where I is the total intensity while Q and U denote the linearly polarized intensity (resp. horizontal and vertical) [6]. We created a synthetic object $\overline{\boldsymbol{x}}$ (*c.f.* Fig. 1) in order to be able to quantify the algorithmic performance. Synthetic data are created to be similar to real data (see Fig. 1). The dataset is composed of $K = 64$ noise realizations following the direct model:

$$(\forall k \in \{1, \ldots K\}) \quad \boldsymbol{d}_k = \mathcal{B}\left(\sum_\ell \begin{bmatrix} v^1_{k,\ell} \mathbf{A} \overline{\boldsymbol{x}}_\ell \\ v^2_{k,\ell} \mathbf{A} \overline{\boldsymbol{x}}_\ell \end{bmatrix}\right) \tag{4}$$

where $\mathcal{B}(x)$ yields a realization of a Gaussian variable $\mathcal{N}(x, \mathrm{Diag}(x) + \sigma^2_{\mathrm{ro}} \mathbf{Id})$, to approximate Poisson noise plus read out noise of variance σ^2_{ro}, $\mathbf{A} \in \mathbb{R}^{N \times N}$ is the convolution with the PSF and the pairs $(v^1_{k,\ell}, v^2_{k,\ell}) \in \mathbb{R}^2$ represent polarization modulation at the acquisition k on the ℓ-th component.

Data-fidelity term h – It is the following Mahalanobis distance, such that, for every $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_L) \in (\mathrm{R}^N)^L$:

$$h(\boldsymbol{x}) = \sum_k \frac{1}{2} \left\| \boldsymbol{d}_k - \sum_\ell \begin{bmatrix} v^1_{k,\ell} \mathbf{A} \boldsymbol{x}_\ell \\ v^2_{k,\ell} \mathbf{A} \boldsymbol{x}_\ell \end{bmatrix} \right\|^2_{\mathbf{W}_k}, \tag{5}$$

where $\|\boldsymbol{y}\|^2_{\mathbf{W}_k} = \boldsymbol{y}^\top \mathbf{W}_k \boldsymbol{y}$ with $\mathbf{W}_k = \mathrm{Cov}(\boldsymbol{d}_k)^{-1}$. This form of h assumes that the K data frames are mutually independent.

Epigraphical constraint C – The function ϕ in (2) stems from the definition of the Stokes parameters and is given by:

$$\forall n \in \{1, \ldots, N\} \quad \phi(\boldsymbol{x}_2, \ldots, \boldsymbol{x}_L)_n = \sqrt{\sum\nolimits_{\ell=2}^{L} \boldsymbol{x}^2_{n,\ell}}. \tag{6}$$

It is important to avoid strong positive/negative oscillations that may result from the the deconvolution.

Penalisation choice $g_\ell(\mathrm{D}_\ell \cdot)$**: TV or TV-h** – Unless brillant stars are in the field, circumstellar environments can be taken for piecewise constant objects. This motivate the use of edge-preserving penalization. We recall that TV is given $\forall \lambda_\ell \geq 0$ and $\forall \boldsymbol{x} \in \mathcal{H}$, by:

$$\mathrm{TV}_{\lambda_\ell}(\boldsymbol{x}) = \lambda_\ell \|\nabla \boldsymbol{x}\|_{\ell_1}. \tag{7}$$

The formulation of TV-h is, $\forall \lambda_\ell \geq 0$, $\varepsilon > 0$ and $\forall \boldsymbol{x} \in \mathcal{H}$:

$$\mathrm{TV}^{\mathrm{h}}_{\lambda_\ell, \varepsilon}(\boldsymbol{x}) = \lambda_\ell \left(\sqrt{\|\nabla \boldsymbol{x}\|^2_2 + \varepsilon^2} - \varepsilon \right). \tag{8}$$

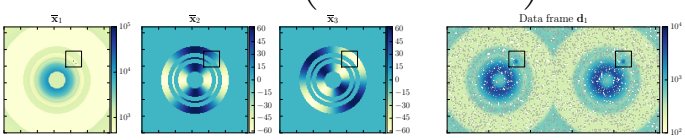


Figure 1: True parameters and synthetics data.

Performance evaluation – Figure 2 shows the convergence of the objective function and of the normalized Mean Squared Error (MSE) of each ℓ-th component, *i.e.* $\|\widehat{\boldsymbol{x}}_\ell - \overline{\boldsymbol{x}}_\ell\|^2 / \|\overline{\boldsymbol{x}}_\ell\|^2$, as a function of the time. We compare the influence of the epigraphical constraint (*i.e.* $\widehat{x} \in C$ and $\widetilde{x} \notin C$) on Fig. 2, on the high dynamical portion of $\boldsymbol{x}$ highlighted on Fig. 1.

Parameter selection – We performed the reconstruction with TV [3] and with TV-h [4] for $\lambda_1 = 0.1$ and $\lambda_2 = \lambda_3 = 0, 03$. For TV-h, we choose ε in $\{10^{-2}, 1, 10^2\}$. We performed the TV-h reconstruction using the algorithm FBwB, with a descent step of $1.99/\beta^{[t]}$. We performed the TV reconstruction using the algorithm 2 with the parameters $\tau^{[t]} = (\beta^{[t]}/\gamma + r\|\mathbf{D}\|^{2-s})^{-1}$ and $\sigma^{[t]} = r\|\mathbf{D}\|^{-2}$, where $r > 0$, $\gamma \in (0, 2)$ and $s \in [0, 2]$, inspired by the diagonal preconditioners proposed by Lorenz and Pock [13, Lemma 10] with $D = \beta \mathrm{Id}$. We fixed $r = 10^{-3}$, $\gamma = 1.99$ and $s = 2$ which seems to gives the fastest convergence. We started with $\beta^{[0]} = 10^{-2}$ and set $\eta = 1, 1$.

Discussion – The epigraphical constraint reduces the oscillations around the two brillant dots (*i.e.* stars) in $\widehat{\boldsymbol{x}}_1$, yet it affects $\widehat{\boldsymbol{x}}_3$. Without the epigraphical constraint, $\widetilde{\boldsymbol{x}}_3$ is not affected by the deconvolution, yet the oscillations in $\widetilde{\boldsymbol{x}}_1$ are amplified. In fact, the pixels of $\widetilde{\boldsymbol{x}}_1$ filled with red on Figure 3 are negatives. When no stars are in the field, the epigraphical constraint has no effects. It could thus be relaxed, in order to use differentiable methods with TV-h. In fact, TV and TV-h give similar results, unless ε is large (*i.e.* TV-h is mostly quadratic). However for the same time of convergence, TV still gives sharper edges than TV-h with $\varepsilon \to 0$. The choice of the method will then depend of the smoothness of the object. Finally, Figure 2 validate numerically the PDwB algorithm. In fact, its convergences behaviour is similar to the convergence of FBwB, with TV-h for small values of ε.

5 Conclusion

In this paper, we designed the PDwB algorithm, to handle both non-smooth TV and the epigraphical constraint. We applied PDwB to perform the reconstruction of simulated high dynamical images of circumstellar environments and compared the performances with FBwB using the TV-h. We observed that the backtracking is effective to achieve the convergence of primal-dual scheme when the Lipchitz constant is unknown, and that it could be applied for more complex reconstructions as texture decomposition. We observed that the epigraphical constraint is not always necessary, allowing the use of differential methods.

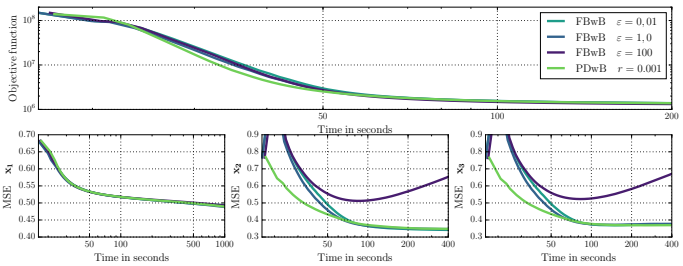


Figure 2: Comparison of the convergence of the objective function and the Mean Square Error (MSE), as a function of the time in seconds.

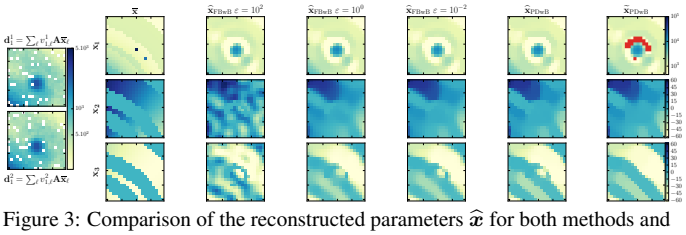


Figure 3: Comparison of the reconstructed parameters $\widehat{\boldsymbol{x}}$ for both methods and $\widetilde{\boldsymbol{x}}$ for the PDwB algorithm. Pixels n such that $\widetilde{\boldsymbol{x}}_n \notin C$ are filled in red on $\widetilde{\boldsymbol{x}}_1$.

References

[1] A. Beck and M. Teboulle " A fast iterative shrinkage-thresholding algorithm for linear inverse problems", SIAM J. Imaging Sci., **2**(1): 183–202, 2009.

[2] L. Condat , "A primal-dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms", Journal of Optimization Theory and Applications, **158**(2):460–479, 2013.

[3] L. I. Rudin, S. Osher and E. Fatemi, "Nonlinear total variation based noise removal algorithms ", Physica D: Nonlinear Phenomena, **60**(1-4): 259–268, 1992.

[4] P. Charbonnier, L. Blanc-Féraud, G. Aubert and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging ", IEEE Transactions on image procesing, **6**(2): 298–311, 1997.

[5] A. Repetti, J. Birdi, A. Dabbech, and Y. Wiaux, "Non-convex optimization for self-calibration of direction-dependent effects in radio interferometric imaging ", Monthly Notices of the Royal Astronomical Society , **470**(4): 3981–4006,2017.

[6] R. G. van Holstein, J.H. Girard, J. de Boer, F. Snik, J. Milli, D. M. Stam, C. Ginski, D. Mouillet, Z. Wahhaj, H. M. Schmid and others, "The polarimetric imaging mode of VLT/SPHERE/IRDIS II: Characterization and correction of instrumental polarization effects ", A&A , **633**: A63, 2020.

[7] N. Pustelnik,A. Benazza-Benhayia, Y. Zheng and J.-C. Pesquet, " Wavelet-based image deconvolution and reconstruction", Wiley Encyclopedia of Electrical and Electronics Engineering, 2016.

[8] L. Denneulin, M. Langlois, N. Pustelnik, and É Thiébaut, " Reconstruction polarimétrique d'environnements circumstellaires à partir des données ESO/VLT-SPHERE IRDIS", GRETSI, 648, 2019.

[9] G. Chierchia, N. Pustelnik, J.-C. Pesquet and B. Pesquet-Popescu, "Epigraphical splitting for solving constrained convex formulations of inverse problems with proximal tools ", Signal, Image and Video Processing, **9**(8): 1737–1749, 2015.

[10] J. de Boer, M. Langlois, R. G. van Holstein, J. H. Girard, D. Mouillet, A. Vigan, K. Dohlen, F. Snik, C. U. Keller, C. Ginski and others, " The polarimetric imaging mode of VLT/SPHERE/IRDIS I: Description, data reduction and observing strategy", A&A , **633**: A63, 2020.

[11] J.-L. Beuzit, A. Vigan, D. Mouillet, K. Dohlen, R. Gratton, A. Boccaletti, J.-F. Sauvage, H. M. Schmid, M. Langlois, C. Petit and others, " SPHERE: the exoplanet imager for the Very Large Telescope", A&A , http://arxiv.org/abs/1902.04080, 2019 (accepted).

[12] Langlois et al., "High contrast polarimetry in the infrared with SPHERE on the VLT ", Ground-based and Airborne Instrumentation for Astronomy V, **9147** , 2014.

[13] D. A. Lorenz and T. Pock, "An inertial forward-backward algorithm for monotone inclusions ", Journal of Mathematical Imaging and Vision , **51**(2): 311–325, 2015.

Astronomy & Astrophysics manuscript no. output
October 8, 2020

RHAPSODIE : Reconstruction of High-contrAst Polarized SOurces and Deconvolution for cIrcumstellar Environments

L. Denneulin[1,2], M. Langlois[1], É. Thiébaut[1], and N. Pustelnik[2]

[1] Univ Lyon, Univ Lyon1, Ens de Lyon, CNRS, Centre de Recherche Astrophysique de Lyon, UMR5574, F-69230 Saint-Genis-Laval, France
[2] Univ Lyon, ENS de Lyon, Univ Claude Bernard Lyon 1, CNRS, Laboratoire de Physique, F-69342 Lyon, France

October 8, 2020

ABSTRACT

Context. Polarimetric imaging is one of the most effective techniques for high-contrast imaging and characterization of circumstellar environments.These environments can be characterized through direct-imaging polarimetry at near-infrared wavelengths. The Spectro-Polarimetric High-contrast Exoplanet REsearch (SPHERE)/IRDIS instrument installed on the Very Large Telescope in its dual-beam polarimetric imaging (DPI) mode, offers the capability to acquire polarimetric images at high contrast and high angular resolution. However dedicated image processing is needed to get rid of the contamination by the stellar light, of instrumental polarization effects and of the blurring by the instrumental point spread function.
Aims. We aim to reconstruct and deconvolve the near-infrared polarization signal from circumstellar environments.
Methods. We use observations of these environments obtained with the high-contrast imaging infrared polarimeter SPHERE-IRDIS at the Very Large Telescope (VLT). We developed a new method to extract the polarimetric signal using an inverse approach method that benefits from the added knowledge of the detected signal formation process. The method includes weighted data fidelity term, smooth penalization, and takes into account instrumental polarization
Results. The method enables to accurately measure the polarized intensity and angle of linear polarization of circumstellar disks by taking into account the noise statistics and the observed objects convolution. It has the capability to use incomplete polarimetry cycles which enhance the sensitivity of the observations. The method improves the overall performances in particular for low SNR/small polarized flux compared to standard methods.
Conclusions. By increasing the sensitivity and including deconvolution, our method will allow for more accurate studies of these disks morphology, especially in the innermost regions. It also will enable more accurate measurements of the angle of linear polarization at low SNR, which would lead to in-depth studies of dust properties. Finally, the method will enable more accurate measurements of the polarized intensity which is critical to construct scattering phase functions.

1. Introduction

With the adaptive-optics-fed high-contrast imaging instruments GPI (Macintosh et al. 2014) and SPHERE-IRDIS (Beuzit et al. 2019; Dohlen et al. 2008), we now have access to the spatial resolution and sensitivity required to observe in the near-infrared (NIR) circumstellar matter at small angular separations. Along with the Integral Field Spectrograph (IFS; Claudi et al. 2008) and the Zürich IMaging POLarimeter (ZIMPOL; Schmid et al. 2018) which can also be used to observe circumstellar environments in polarimetry, the IRDIS instrument is one of the three SPHERE instruments (Beuzit et al. 2019). SPHERE/IRDIS is able to acquire two simultaneous images at two different wavelengths, in a so-called Dual Band Imaging (DBI) mode (Vigan et al. 2014), or for two different polarizations, in a so-called Dual Polarimetry Imaging mode (Langlois et al. 2014; de Boer et al. 2020). Both circumstellar disks and self-luminous giant exoplanets or companion brown dwarfs can be characterized by these new instruments in direct-imaging at these wavelengths.

The NIR polarimetric mode of SPHERE/IRDIS at the Very Large Telescope (VLT), which is described in Beuzit et al. (2019); de Boer et al. (2020); van Holstein et al. (2020), has proven to be very successful for the detection of circumstellar disks in scattered light (Garufi et al. 2017) and shows much promise for the characterization brown dwarfs (van Holstein et al. 2017) and exoplanets (van Holstein 2020, in prep.) when they are surrounded by circumsubstellar disks.

Three particular types of circumstellar disks are studied: protoplanetary disks, transition disks and debris disks. The observations of the protoplanetary and transition disks morphology linked to hydrodynamical simulations allows for the study of their formation scenario as in the study of HD 142527 (Price et al. 2018), IM Lup, RU Lup (Avenhaus et al. 2018), GSC 07396-759 (Sissa et al. 2018). Their observations are valuable because their shapes can be the signposts for the formation of one or several exoplanets. In fact, during their formation, the planets "clean" the dust off their orbits, creating gaps without dust such as for RXJ 1615, MY Lup, PDS 66 (Avenhaus et al. 2018), PDS 70 (Keppler et al. 2018, 2019; Haffert et al. 2019). The planet formation scenario can be explain with hydrodynamical simulations as the cases of HL Tau (Dipierro et al. 2015), HD 163296 (Pinte et al. 2019). Due to gravity, exoplanets can also create spiral arms, as in RY Lup (Langlois et al. 2018) and MWC 758 (Benisty et al. 2015) where the presence of exoplanets is argued with hydrodynamical simulations. Debris disks are the oldest step in the evolution of circumstellar disks, when there is already one or several planets in the system and the gas is almost completely consumed. These disks are composed with dust and grain never accreted into the planets, as in the case of HR 4796A (Perrin et al. 2015; Milli et al. 2019).

These environments can be observed with SPHERE in the near-infrared and in the visible. Yet, such observations are difficult because of the high contrast between the light of the environ-

ment and the residual light from the host star. As a result, when acquiring images, the light of the environment is contaminated by the diffraction stains from the host star. Two methods can be used to disentangle the light of the disk from that of the star: The Angular Differential Imaging (ADI; Marois et al. 2006) and the Differential Polarimetric Imaging (DPI; van Holstein et al. 2020). The ADI technique uses the fact that the stellar residuals are fixed in the pupil plane and the object of interest artificially rotates. This created diversity makes it possible to disentangle the light of the object of interest from the residual light of the star. Yet, such a method does not allow a good reconstruction of the disk morphology which is impacted by artifacts due to self-subtraction. Moreover it failed when the environment is almost rotation invariant. The DPI observations allow to have access to the disks morphology without the artifacts by using the difference of polarization state between the light scattered by the environment and the light of the host star.

The state-of-the-art methods to process datasets in polarimetry, apart from the calibration, are "step-by-step" methods. First, the data are transformed with the required translations and rotations to be easier to process and the bad pixels are interpolated. All these interpolations can lead to a first loss of information. Second, the "clean" data are reduced to the Stokes parameters, which represent the polarization in different directions. These reductions can be done with the double difference or the double ratio (Tinbergen 2005; Avenhaus et al. 2014). If the double ratio takes in account the possibility of multiplicative instrumental effects, none of these methods deal with the photon noise. This results in some limitations in sensitivity in case of low Signal-to-noise Ratio (SNR). Last, if any deconvolution or regularization is needed, it is done *a posteriori*, leading again to a possible loss of information. Still, these state-of-the-art methods have proven over the years to be sufficiently efficient to produce good quality results. However, studies of circumstellar disks are often limited to analyses of the orientation (position angle and inclination) and morphology (rings, gaps, cavities and spiral arms) of the disks (Muto et al. 2012; Quanz et al. 2013; Ginski et al. 2016; de Boer et al. 2016). Quantitative polarimetric measurements of circumstellar disks and substellar companions are currently very challenging, because existing data-reduction methods do not estimate properly the errors sources from both noise and detector calibration. They also rely only on complete polarimetric cycles and do not account for the instrument convolution. For observations of circumstellar disks (van Holstein et al. 2020), calibrating the instrumental polarization effects with a sufficiently high accuracy already yield to multitude of improvements.

Over the last decades in image processing, it has been proven that the reconstruction of parameters of interest benefits from a global inversion, taking into account all the transformations integrated in the instrument, rather than "step-by-step" procedures. The proposed inversion relies on the physical model that expresses the data as a function of the quantities of interest, which are then obtained as a minimum of a cost function. This cost function is generally composed with a data fidelity term and some prior regularizations and/or constraints. Depending on the convexity and the smoothness of the cost function, a wide panel of algorithmic schemes with convergence guarantees may be considered to achieve a minimum of the cost function (Nocedal & Wright 1999; Combettes & Pesquet 2011; Pustelnik et al. 2016). Such methods have been used over decades in astrophysics for the physical parameters estimation (Titterington 1985), mostly in adaptive optics (Borde & Traub 2006) and radio-interferometry with the well known algorithm CLEAN (Högbom 1974). This last method has been the starting point of a wide variety of algorithms, such as the algorithms SARA (Carrillo et al. 2012) and Polca-SARA (Birdi et al. 2019) in polarimetric radio-interferometry using more sofisticated tools as non-smooth penalizations. A non-smooth method was also used for images denoising with curvelets (Starck et al. 2003). The minimization of a co-log-likelihood was also used in the blind deconvolution of images convolved with a PSF stained with aberrations (Thiébaut & Conan 1995). Since half a decade, the use of inverse problem methods has increased in astrophysics and as been used more recently for the estimation of the CMB (Adam et al. 2016) and the imaging of the supermassive black-hole (Akiyama et al. 2019). In high contrast imaging, the use of inverse problem methods is more recent. It has been used to perform auto-calibration of the data (Berdeu et al. 2020) with the IFS/SPHERE. It has also been used to reconstruct extended objects in total intensity by using using ADI data (Pairet et al. 2019; Flasseur et al. 2019) with the SPHERE instrument. Yet, such reconstruction methods have not been used in polarimetric high contrast direct imaging.

In the present work we describe in details the method and the benefits of the use of an *inverse problem* formalism for the reconstruction of circumstellar environments observed in polarimetry with the instrument ESO/VLT SPHERE IRDIS. In Section 2, we first describe the physical model of the data obtained with the ESO/VLT SPHERE IRDIS instrument. This includes the polarimetric parametrisation, the sequences and the convolution by the PSF. In Section 3, we describe RHAPSODIE (Reconstruction of High-contrAst Polarized Sources and Deconvolution for near-Infrared Environments) method we developed. Finally, in Section 5, we present the results obtained with such a method on both simulated and astrophysical data.

2. Modeling polarimetric data

Observations of circumstellar environments in polarimetry are obtained by recording the linearly polarized intensity I^{p}, and its angle of polarization θ, that are produced by the reflection of the star light onto the circumstellar environmental dusts. By modulating the orientation of the polarization, these parameters of interest can be disentangled from the unpolarized light I^{u}, composed of star light and unpolarized circumstellar environment light. Without ADI observations, it is not possible to separate both I^{u} components i.e the total intensity of the disk from the unpolarized stellar intensity.

The estimation of the parameters $(I^{\mathrm{u}}, I^{\mathrm{p}}, \theta)$ is the core of the present contribution. However, to model the effects of the instrument on the observable polarization, Stokes parameters are more suitable, since the data can be expressed as a linear combinaison of them. Stokes parameters describes the total light, the linearly polarized light and the circularly polarized light. Since circular polarization is mostly generated by magnetic interactions and double scattering, in the case of circumstellar environments, it is often negligible and thus not measured by the SPHERE or GPI instruments. In the end, it is possible to reconstruct the parameters of interest I^{u}, I^{p} and θ from a combination of the Stokes parameters.

2.1. Polarization effects

The four Stokes intensity parameters $S = (I, Q, U, V) \in \mathbb{R}^4$ describe the state of polarized light: I is the total intensity accounting for the polarized and unpolarized light, Q and U are the intensities of the light linearly polarized along 2 directions at 45° to each other and V is the intensity of the circularly polarized light.

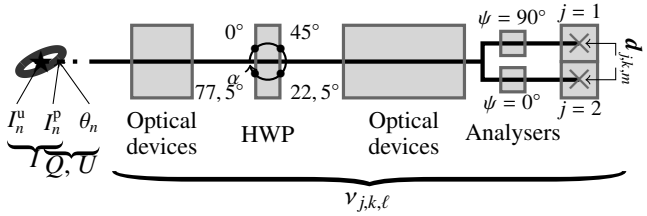


Fig. 1: Scheme of the instrument ESO/VLT-SPHERE IRDIS, showing the modulation of the polarimetry in the instrument at the acquisition $k \in \{1, \dots K\}$.

Under this formalism, polarization effects by an instrument like SPHERE/IRDIS (see Fig. 1) can be modeled by (see Eq. (17) of van Holstein et al. 2020):

$$S_j^{\mathrm{det}} = \mathbf{M}_j^{\mathrm{pol}}\, \mathbf{T}\big(-\theta^{\mathrm{der}}\big)\, \mathbf{M}^{\mathrm{der}}\, \mathbf{T}\big(\theta^{\mathrm{der}}\big)\, \mathbf{T}(-\alpha)\, \mathbf{M}^{\mathrm{HWP}}\, \mathbf{T}(\alpha)\, \mathbf{M}^{\mathrm{M4}}\, \mathbf{T}\big(\theta^{\mathrm{alt}}\big)\, \mathbf{M}^{\mathrm{UT}}\, \mathbf{T}(\theta^{\mathrm{par}})\, S \tag{1}$$

where S_j^{det} are the Stokes parameters on the detector after the left ($j = 1$) or right ($j = 2$) polarizer of the analyser set while S are the Stokes parameters at the entrance of the telescope. In the above equation, $\mathbf{M}$ denotes a Mueller matrix accounting for the polarization effects of a specific part of the instrument: $\mathbf{M}_j^{\mathrm{pol}}$ for the left or right polarizer of the analyzer set, $\mathbf{M}^{\mathrm{der}}$ for the optical derotator, $\mathbf{M}^{\mathrm{HWP}}$ is for the half-wave plate (HWP), $\mathbf{M}^{\mathrm{M4}}$ for the 4-th mirror of the telescope and $\mathbf{M}^{\mathrm{UT}}$ for the 3 mirrors (M1 to M3) constituting the telescope. The term $\mathbf{T}(\theta)$ denotes a rotation matrix of the polarization axes by an angle θ: θ^{der} is the derotator angle, α is the HWP angle, θ^{alt} is the altitude angle and θ^{par} is the parallactic angle of the pointing of the alt-azimuthal telescope.

In SPHERE/IRDIS optical, the unique measured Stokes parameter is the total intensity I_j^{det} (*i.e.* the first component of S_j^{det}). From Eq. (1), this quantity is given by a simple linear combination of the input Stokes parameters:

$$I_j^{\mathrm{det}} = \sum_{\ell=1}^{4} \nu_{j,\ell}(\Theta)\, S_\ell \tag{2}$$

where, for every $\ell \in \{1, \dots 4\}$, S_ℓ denotes the ℓ-th component of the input Stokes parameters S and $\nu_{j,\ell}(\Theta)$, for $j \in \{1, 2\}$, are real coefficients depending on all involved angles $\Theta = (\theta^{\mathrm{der}}, \theta^{\mathrm{alt}}, \theta^{\mathrm{par}}, \alpha)$.

Even though the measure is restricted to the component I_j^{det}, rotating the angle α of the HWP introduces a modulation of the contribution of the Stokes parameters Q and U in I_j^{det}, which can be exploited to disentangle the Stokes parameters I, Q and U. The Stokes parameter V characterizing the circularly polarized light cannot be measured with an instrument such as SPHERE/IRDIS (a modulation by a quarter-wave plate would have been required to do so). In the following, we therefore neglect the circularly polarized light and only consider the unpolarized and linearly polarized light characterized by the Stokes parameters $S = (I, Q, U)$. As a direct simplification, the sum in the right hand side of Eq. (2) is reduced to its first three terms. For a sequence of acquisitions with different angles of the HWP, the detected intensities follow:

$$I_{j,k}^{\mathrm{det}} = \sum_{\ell=1}^{3} \nu_{j,k,\ell}\, S_\ell \tag{3}$$

where $\nu_{j,k,\ell} = \nu_{j,\ell}(\Theta_k)$ with Θ_k the set of angles during the k-th acquisition.

The polarization effects being reduced to the one of the analyzers and the HWP and considering that the rotation due to the altitude and parallactic angles is neglected, the detected intensity writes (van Holstein et al. 2020):

$$I_{j,k}^{\mathrm{det}} = \tfrac{1}{2}\, I + \tfrac{1}{2}\, \cos(4\,\alpha_k + 2\,\psi_j)\, Q + \tfrac{1}{2}\, \sin(4\,\alpha_k + 2\,\psi_j)\, U\,, \tag{4}$$

where ψ_j is the orientation angle of the left/right polarizer while α_k is the HWP angle during the k-th acquisition. Table 1 lists the values of the above linear coefficients for a typical set of HWP angles.

HWP Angle	Left/right Analyzer	$\nu_{j,k,1}$	$\nu_{j,k,2}$	$\nu_{j,k,3}$
$\alpha_k = 0°$	$j = 1$ (left)	$1/2$	$1/2$	0
$\alpha_k = 0°$	$j = 2$ (right)	$1/2$	$-1/2$	0
$\alpha_k = 45°$	$j = 1$ (left)	$1/2$	$-1/2$	0
$\alpha_k = 45°$	$j = 2$ (right)	$1/2$	$1/2$	0
$\alpha_k = 22.5°$	$j = 1$ (left)	$1/2$	0	$1/2$
$\alpha_k = 22.5°$	$j = 2$ (right)	$1/2$	0	$-1/2$
$\alpha_k = 77.5°$	$j = 1$ (left)	$1/2$	0	$-1/2$
$\alpha_k = 77.5°$	$j = 2$ (right)	$1/2$	0	$1/2$

Table 1: Values of the coefficients $\nu_{j,k,\ell}$ for $j \in \{1, 2\}$, $k \in \{1, \dots, K\}$ and $\ell \in \{1, 2, 3\}$ and assuming no instrumental polarization.

2.2. Parameters of interest

The model of the detected intensity given in Eq. (3) is linear in the Stokes parameters $S = (I, Q, U)$, which makes its formulation suited to inverse problem solving. However, to study circumstellar environments in polarimetry, the knowledge of the linearly polarized light I^{p} and the polarization angle θ is crucial. Both set of parameters are related as it follows:

$$\begin{cases} I = I^{\mathrm{u}} + I^{\mathrm{p}} \\ Q = I^{\mathrm{p}}\, \cos(2\,\theta) \\ U = I^{\mathrm{p}}\, \sin(2\,\theta) \end{cases} \tag{5}$$

and conversely by[1]:

$$\begin{cases} I^{\mathrm{p}} = \sqrt{Q^2 + U^2} \\ \theta = (1/2) \arctan\,(U/Q) \mod \pi \\ I^{\mathrm{u}} = I - I^{\mathrm{p}}. \end{cases} \tag{6}$$

These relations are applied independently at any position of the field of view (FOV). The direct model of the detected intensity as a function of the parameters I^{u}, I^{p} and θ is non-linear and, by Eq. (3) and Eq. (5), is reduced to:

$$I_{j,k}^{\mathrm{det}} = \nu_{j,k,1}\, I^{\mathrm{u}} + \big(\nu_{j,k,1} + \nu_{j,k,2}\, \cos(2\,\theta) + \nu_{j,k,3}\, \sin(2\,\theta)\big)\, I^{\mathrm{p}}\,. \tag{7}$$

For a perfect SPHERE/IRDIS-like instrument and not considering field rotation, this simplifies to:

$$I_{j,k}^{\mathrm{det}} = \tfrac{1}{2}\, I^{\mathrm{u}} + I^{\mathrm{p}}\, \cos^2(\theta - 2\,\alpha_k - \psi_j)\,, \tag{8}$$

which is the Malus law.

[1] In this representation there is a ±180° degeneracy for the linear polarization angle.

2.3. Accounting for the instrumental spatial PSF

In polarimetric imaging, each polarimetric parameter is a function of the 2-dimensional FOV. It has been assumed that the Stokes parameters are represented by *images* of N pixels each such as $S_{\ell,n}$ denotes the value of the n-th pixel in the map of the ℓ-th Stokes parameter.

Provided that polarization effects apply uniformly across the FOV and that the instrumental spatial point spread function (PSF) does not depend on the polarization of light, each Stokes parameter of a spatially incoherent source are independently and identically affected by the spatial PSF (e.g Birdi et al. 2018; Smirnov 2011; Denneulin 2020). As the effects of the spatial PSF are linear, we can write the detected intensity for a given detector pixel as:

$$I^{\rm det}_{j,k,m} = \sum_{\ell=1}^{3} \sum_{n=1}^{N} \nu_{j,k,\ell}\, H_{j,k,m,n}\, S_{\ell,n}\,, \tag{9}$$

where $I^{\rm det}_{j,k,m}$ is the measured intensity for the m-th detector pixel in $m \in \{1,\dots M\}$ of the j-th polarizer of the analyzer set during the k-th acquisition in $k \in \{1,\dots K\}$, where the coefficients $\nu_{j,k,\ell}$ are defined in Eq. (3) and $H_{j,k,m,n}$ denotes a given element of the discretized spatial PSF of the instrument.

It follows from our assumption on spatial and polarization effects applying independently, that the discretized spatial PSF does not depend on the polarization index ℓ. Consequently, the spatial and polarization effects in Eq. (9) mutually commute. This may be exploited to factorize expressions and reduce the number of numerical operations to apply the direct model.

Figure 1 and Eq. (1) provide a representation of the instrument from which we build a model of the spatial effects of the instrument. Accordingly, an image representing the spatial distribution of the light as input of the instrument should undergo a succession of image transformations before reaching the detector. These transformations are either geometrical transformations (e.g., the rotation depending on the parallactic angle) or blurring transformations (e.g., by the telescope). Except in the neighborhood of the coronographic mask, the effects of the blurs can be assumed to be shift-invariant and can thus be modeled by convolutions by shift-invariant PSFs. Geometrical transforms and convolutions do not commute but their order may be changed provided the shift-invariant PSFs are appropriately rotated and/or shifted. Thanks to this property and without loss of generality, we can model the spatial effects of the instrument by a single convolution accounting for all shift-invariant blurs followed by a single geometrical transform accounting for all the rotations but also possible geometrical translations and/or spatial (de)magnification (if the pixel size of the polarimetric maps is not chosen to be equal to the angular size of the detector pixels). Following this analysis, our model of the spatial PSF is given by:

$$H_{j,k,m,n} = \sum_{n'} (\mathbf{T}_{j,k})_{m,n'}\, (\mathbf{A}_k)_{n',n}\,, \tag{10}$$

where $\mathbf{A}_k : \mathbb{R}^N \to \mathbb{R}^N$ implements the shift-invariant blur of the input model maps while $\mathbf{T}_{j,k} : \mathbb{R}^N \to \mathbb{R}^{M_j}$ performs the geometrical transform of the blurred model maps for the j-th polarizer of the analyzer set during the k-th acquisition. N is the number of pixels in the model maps and M_j is the number of pixels of the detector (or sub-image) corresponding to the j-th output polarizer.

2.4. Polarimetric data

During a sequence of observations with SPHERE/IRDIS in DPI mode, the HWP is rotated several times along a given cycle of angles $\alpha \in \{0°, 45°, 22.5°, 67.5°\}$. Besides, the two polarizers of the analyzer set of SPHERE/IRDIS are imaged on two disjoint parts of the same detector. The resulting dataset consists in, say K, frames of two, left and right, sub-images, each with a different position of the HWP. After pre-processing of the raw images to compensate for the bias and the uneven sensitivity of the detector and to extract the two sub-images, the available data are modeled by:

$$d_{j,k,m} \approx I^{\rm det}_{j,k,m} \tag{11}$$

where $I^{\rm det}_{j,k,m}$ is given in Eq. (9) for $k \in \{1,\dots,K\}$ the index of the acquisition, $j \in \{1,2\}$ indicating the left/right polarizer of the analyzer, and $m \in \{1,\dots,M\}$ the pixel index in the corresponding left/right sub-image.

The $\approx$ sign in Eq. (11) is to account for an unknown random perturbation term due to the noise. Noise in the pre-processed images can be assumed centered and independent between two different pixels or frames because the pre-processing suppresses the bias and treats pixels separately thus introducing no statistical correlations between pixels. There are a number of sources of noise: photon, background and dark current shot noise, detector read-out noise, etc. For most actual data, the shot noise is the most important contribution and the number of electrons (photo-electrons plus dark current) integrated by a pixel is large enough to approximate the statistics of the data by an independent non-uniform Gaussian distribution whose mean is given by the right-hand-side term of Eq. (11) (i.e., the noise is centered in $I^{\rm det}_{j,k,m}$) and whose variance $\Sigma_{j,k,m} = \mathrm{Var}(d_{j,k,m})$ has to be estimated by the calibration stage (see subsection 4.1).

It is worth noticing that since the proposed model is not valid in the neighborhood of the coronographic mask, the pixel data in this region have to be discarded. Moreover, the detector contains bad pixels (*e.g.* dead pixels, non-linear pixels, saturated pixels) which must be also discarded. It is achieved by supposing that their variance is infinite. This will correspond to set their weights to zero in the data fidelity term of the objective function described in Section 3.2.

3. RHAPSODIE : Reconstruction of High-contrAst Polarized SOurces and Deconvolution for cIrcumstellar Environments

3.1. Inverse problems approach

In polarimetric imaging, one is interested in recovering sampled maps of the polarimetric parameters, which can be the Stokes parameters or the intensities of unpolarized and linearly polarized light and the angle of the linear polarization or some mixture of these parameters. To remain as general as possible, we denote by $\boldsymbol{X} \in \mathbb{R}^{N\times L}$ the set of parameters of interest to be recovered where each of the ℓ-th parameter with $\ell \in \{1,\dots L\}$ is composed with N pixels.

Given the direct model of the pre-processed data developed in the previous section, we propose to recover the parameters of interest $\boldsymbol{X}$ by a penalized maximum likelihood approach. This approach is customary in the solving of inverse problems (Titterington 1985; Tarantola 2005) and amounts to define the estimated parameters $\widehat{\boldsymbol{X}}$ as the ones that minimize a given objective function $f(\boldsymbol{X})$ possibly under constraints expressed as $\boldsymbol{X} \in C$

with C the set of acceptable solutions. The objective function takes the form of the sum of a data-fidelity term $f_{\text{data}}(\boldsymbol{X})$ and of regularization terms $f_\rho(\boldsymbol{X})$:

$$\widehat{\boldsymbol{X}} = \underset{\boldsymbol{X}\in C}{\arg\min}\Big\{f(\boldsymbol{X}) = f_{\text{data}}(\boldsymbol{X}) + \underbrace{\sum\nolimits_\rho \lambda_\rho\, f_\rho(\boldsymbol{X})}_{f_{\text{prior}}(\boldsymbol{X})}\Big\} \tag{12}$$

where $\lambda_\rho \geq 0$ $(\forall \rho)$ are so-called hyper-parameters introduced to tune the relative importance of the regularization terms. The data-fidelity term $f_{\text{data}}(\boldsymbol{X})$ imposes that the direct model be as close as possible to the acquired data while the regularization terms $f_\rho(\boldsymbol{X})$ enforce the components of the model to remain regular (e.g., smooth). Regularization must be introduced to lift degeneracies and avoid artifacts caused by the data noise and the ill-conditioning of the inverse problem. Additional strict constraints may be imposed to the sought parameters via the feasible set C, e.g. to account for the requirement that intensities be non-negative quantities. These different terms and constraints are detailed in the following sub-sections.

3.2. Data fidelity

Knowing the sufficient statistics for the pre-processed data, agreement of the model with the data is best estimated by the co-log-likelihood of the data (Tarantola 2005) or equivalently by the following criterion:

$$f_{\text{data}}(\boldsymbol{X}) = \sum_{j,k} \left\| \boldsymbol{d}_{j,k} - \boldsymbol{I}^{\text{det}}_{j,k}(\boldsymbol{X}) \right\|^2_{\mathbf{W}_{j,k}}, \tag{13}$$

where $\|\cdot\|^2_{\mathbf{W}} = \langle \cdot, \mathbf{W}\cdot \rangle$ denotes Mahalanobis (1936) squared norm, $\boldsymbol{d}_{j,k} = (d_{j,k,1}, \ldots, d_{j,k,M})^\top \in \mathbb{R}^M$ collects all the pixels (e.g., in lexicographic order) of the j-th sub-image in the k-th acquisition as define in Eq. (11). Similarly, $\boldsymbol{I}^{\text{det}}_{j,k}(\boldsymbol{X}) = \boldsymbol{I}^{\text{det}}_{j,k} = (I^{\text{det}}_{j,k,1}, \ldots, I^{\text{det}}_{j,k,m})^\top \in \mathbb{R}^M$ which is given by the model in Eq. (9) applied to the Stokes parameters as a function of the parameters of interest $\boldsymbol{X}$ (*i.e.* $S(\boldsymbol{X}) = S$). For instance if $\boldsymbol{X} = (I^{\text{u}}, I^{\text{p}}, \theta)$, $S(\boldsymbol{X})$ is obtained by Eq. (5). Above, $\mathbf{W}_{j,k}$ is the precision matrix of the data which is diagonal because pixels are considered as mutually independent. To account for invalid data (see 2.4), we define the diagonal entries of the precision matrix as:

$$\forall m \in \{1, \ldots M\}, \quad (\mathbf{W}_{j,k})_{m,m} = \begin{cases} \Sigma^{-1}_{j,k,m} & \text{for valid data;} \\ 0 & \text{for invalid data.} \end{cases} \tag{14}$$

Invalid data include dead pixels, pixels incorrectly modeled by our direct model because of saturation or of the coronograph, missing frames for a given HWP angle and unusable frames because of too strong atmospheric turbulence or unproper coronograph centering.

Note that, in Eq. (13), the Mahalanobis squared norms arise from our Gaussian approximation of the statistics while the simple sum of these squared norms for each sub-image in each frame is justified by the fact that all frames and all sub-images are mutually independent.

In our implementation of the model I^{det}, the geometrical transform $\mathbf{T}_{j,k}$ is performed by means of interpolation by Catmull-Rom splines. The blurring is applied by:

$$\mathbf{A}_k = \mathbf{F}^{-1}\, \text{diag}(\tilde{\boldsymbol{p}}_k)\, \mathbf{F} \tag{15}$$

where $\mathbf{F}$ denotes the FFT operator (of suitable size) and $\text{diag}(\tilde{\boldsymbol{p}}_k)$ implements the frequency-wise multiplication by $\tilde{\boldsymbol{p}}_k = \mathbf{F}\,\boldsymbol{p}_k$ the discrete Fourier transform of $\boldsymbol{p}_k$ the shift-invariant PSF. Note that $\boldsymbol{p}_k$ must be specified in the same reference frame as the FOV.

3.3. Regularization

The problem of recovering the polarimetric parameters from the data is an ill-conditioned inverse problem mainly due to the instrumental blur. Furthermore, the problem may also be ill-posed if there are too many invalid data. In the case of an ill-conditioned inverse problem, the maximum likelihood estimator of the parameters of interest, that is the parameters which minimize the data fidelity term $f_{\text{data}}(\boldsymbol{X})$ defined in Eq. (13) alone, cannot be used because it is too heavily corrupted by noise amplification. Explicitly requiring that the sought parameters be somewhat regular is mandatory to avoid this (Titterington 1985; Tarantola 2005). In practice, this amounts to add one of more regularization terms $f_\rho(\boldsymbol{X})$ to the data-fidelity as assumed by the objective function defined in Eq. (12).

3.3.1. Edge-preserving smoothness

We expect that the light distribution of circumstellar environments be mostly smooth with some sharp edges, hence *edge-preserving smoothness* regularization (Charbonnier et al. 1997) appears to be the most suited choice to this kind of light distribution. When considering the recovering of polarimetric parameters, such a constraint can be directly imposed to the unpolarized intensity I^{u}, to the polarized intensity I^{p} or to the total intensity I by the following regularization terms:

$$f_{I^{\text{u}}}(\boldsymbol{X}) = \sum_n \left(\sqrt{\|\mathbf{D}_n \boldsymbol{I}^{\text{u}}(\boldsymbol{X})\|^2 + \mu^2} - \mu \right), \tag{16}$$

$$f_{I^{\text{p}}}(\boldsymbol{X}) = \sum_n \left(\sqrt{\|\mathbf{D}_n \boldsymbol{I}^{\text{p}}(\boldsymbol{X})\|^2 + \mu^2} - \mu \right), \tag{17}$$

$$f_{I}(\boldsymbol{X}) = \sum_n \left(\sqrt{\|\mathbf{D}_n \boldsymbol{I}(\boldsymbol{X})\|^2 + \mu^2} - \mu \right), \tag{18}$$

where we denote in boldface sampled maps of polarimetric parameters, for instance $\boldsymbol{I}^{\text{u}} \in \mathbb{R}^N$ the image of the unpolarized intensity or $\boldsymbol{I}^{\text{u}}(\boldsymbol{X})$ to make explicit that it is uniquely determined by the sought parameters $\boldsymbol{X}$. In the above expressions, $\mu > 0$ is a threshold level and $\mathbf{D}_n : \mathbb{R}^N \to \mathbb{R}^2$ is a linear operator which yields an approximation of the 2D spatial gradient of its argument around n-th pixel. This operator is implemented by means of finite differences; more specifically applying $\mathbf{D}_n$ to a sampled map $\boldsymbol{u}$ of a given parameter writes:

$$\mathbf{D}_n\, \boldsymbol{u} = \begin{pmatrix} \boldsymbol{u}_{n_1+1,n_2} - \boldsymbol{u}_{n_1,n_2} \\ \boldsymbol{u}_{n_1,n_2+1} - \boldsymbol{u}_{n_1,n_2} \end{pmatrix}. \tag{19}$$

where (n_1, n_2) denotes the row and column indices of the n-th pixel in the map. At the edges of the support of the parameter maps, we simply assume *flat boundary conditions* and set the spatial gradient to zero there.

The regularizations in Eq. (16)–(18) implement an hyperbolic version of a pseudo-norm of the spatial gradient of a given component of the light distribution which behaves as an L_2-norm (i.e., quadratically) for gradients much smaller than μ and as an L_1-norm (i.e., linearly) for gradients much larger than μ. Hence imposing smoothness for flat areas where the spatial gradient is small while avoiding strong penalization for larger spatial gradients at edges of structures.

It has been shown (Lefkimmiatis et al. 2013; Chierchia et al. 2014) that grouping different sets of parameters in regularization terms that are sub-L_2 norm of the gradient like the last one in Eq. (16)–(18) yields solutions in which strong changes tend to

occur at the same locations in the sets of parameters. In order to encourage sharp edges to occur at the same places in the Stokes parameters Q and U, we also consider using the following regularization for these components:

$$f_{Q+U}(\boldsymbol{X}) = \sum_n \left(\sqrt{\|\mathbf{D}_n \boldsymbol{Q}(\boldsymbol{X})\|^2 + \|\mathbf{D}_n \boldsymbol{U}(\boldsymbol{X})\|^2 + \mu^2} - \mu \right). \quad (20)$$

Many other regularizations implementing smoothness constraints can be found in the literature from the simple quadratic one (Tikhonov Tikhonov 1963) to the very popular *total variation* (TV; Rudin et al. 1992). Quadratic regularizations have a tendency to yield strong ripples which, owing to the contrast of the recovered maps, are an unacceptable nuisance while TV yields maps affected by a so-called *cartoon effect* (i.e., piecewise flat images) which is not appropriate for realistic astronomical images. We however note that the hyperbolic edge-preserving regularization with a threshold μ set to a very small level can be seen as a relaxed version of TV and has been widely used as differentiable approximation of this regularization. Our choice of a differentiable regularization is also motivated by the existence of efficient numerical methods to minimize non-quadratic but differentiable objective functions of many (millions or even billions) variables possibly with additional strict constraints. See Denneulin et al. (2019, 2020); Denneulin (2020) for a comparison of possible advanced regularizers.

3.3.2. Tuning of the hyper-parameters

In the regularization function $f_{\text{prior}}(\boldsymbol{X})$, the terms defined in Eq. (16)–(18) and Eq. (20) can be activated (or inhibited) by choosing the corresponding $\lambda_\rho > 0$ (or $\lambda_\rho = 0$). It is also required to tune the threshold level $\mu > 0$ in addition to the λ_ρ multipliers. All these hyper-parameters have an incidence on the recovered solution: the higher a given λ_ρ the smoother the corresponding regularized component and lowering the threshold μ allows more sharp structures to appear. A number of practical methods have been devised to automatically tune the hyper-parameters (SURE, GCV, ℓ-curve Stein 1981; Ramani et al. 2008; Eldar 2008; Deledalle et al. 2014; Golub et al. 1979; Hansen & O'Leary 1993), none of which can be applied straightforwardly to our method. We therefore chose to present results where we tuned the hyper-parameters by hand by inspecting the resulting solutions. Unsupervised tuning of the hyper-parameters is currently under study and will be the subject of a future publication.

To avoid a contamination of the extracted polarized parameters by the unpolarized component, it is better to separate these terms in the regularization. Hence, the regularization of the unpolarized component I^{u} should be done via $f_{I^{\text{u}}}$ defined in Eq. (16) rather than via f_I defined in Eq. (18). For the polarized light, the joint regularization of Q and U by f_{Q+U} defined in Eq. (20) is more effective than the regularization of I^{p} alone by $f_{I^{\text{p}}}$ defined in Eq. (17) which does not constrain the angle θ of the linear polarization. In Eq. (12), we therefore take the Set 1 or Set 2 of hyper-parameters presented in Table 2. The latter combination is preferable as discussed previously.

Hyper-parameters	$\lambda_{I^{\text{p}}}$	$\lambda_{I^{\text{u}}}$	λ_I	λ_{Q+U}
Set 1	$= 0$	≥ 0	$= 0$	≥ 0
Set 2	$= 0$	$= 0$	≥ 0	≥ 0

Table 2: Set of hyper-parameters used in the present work.

3.4. Imposing the positivity of the intensities

Imposing that restored intensities are necessarily non-negative quantities has proven to be a very effective constraint for astronomical imaging where large parts of the images consist in background pixels whose value should be zero (Biraud 1969). Whatever the choice of the parametrization, the intensities I, I^{u} and I^{p} should all be everywhere nonnegative.

Since $I = I^{\text{u}} + I^{\text{p}}$, it is sufficient to require that I^{u} and I^{p} be nonnegative. Hence, for the set of parameters $\boldsymbol{X} = (\boldsymbol{I}^{\text{u}}, \boldsymbol{I}^{\text{p}}, \boldsymbol{\theta})$, the positivity constraint writes:

$$C = \left\{ (\boldsymbol{I}^{\text{u}}, \boldsymbol{I}^{\text{p}}, \boldsymbol{\theta}) \in \mathbb{R}^{N\times 3} \;\middle|\; \forall n \in \{1,\dots,N\},\ I^{\text{u}}_n \geq 0,\ I^{\text{p}}_n \geq 0 \right\}. \quad (21)$$

Expressed for the Stokes parameters $\boldsymbol{X} = (\boldsymbol{I}, \boldsymbol{Q}, \boldsymbol{U})$, the positivity constraint becomes an epigraphical constraint:

$$C = \left\{ (\boldsymbol{I}, \boldsymbol{Q}, \boldsymbol{U}) \in \mathbb{R}^{N\times 3} \;\middle|\; \forall n \in \{1,\dots,N\},\ I_n \geq \sqrt{Q_n^2 + U_n^2} \right\}. \quad (22)$$

Such a constraint can be found in Birdi et al. (2018), but it has not yet been implemented in high contrast polarimetric imaging.

Since $I^{\text{p}}_n = \sqrt{Q_n^2 + U_n^2}$ (for all pixels n), the positivity of the intensity I^{p} of the polarized light automatically holds if the parameters $\boldsymbol{X} = (\boldsymbol{I}^{\text{u}}, \boldsymbol{Q}, \boldsymbol{U})$ are considered. It is then sufficient to impose the positivity of the intensity I^{u} of the unpolarized light as expressed by the following feasible set:

$$C = \left\{ (\boldsymbol{I}^{\text{u}}, \boldsymbol{Q}, \boldsymbol{U}) \in \mathbb{R}^{N\times 3} \;\middle|\; \forall n \in \{1,\dots,N\},\ I^{\text{u}}_n \geq 0 \right\}. \quad (23)$$

3.5. Choice of the polarimetric parameters

Our method expresses the recovered parameters $\widehat{\boldsymbol{X}}$ as the solution of a constrained optimization problem specified in Eq. (12). As explained in Sec. 3.3.2, the kind of imposed regularization is chosen via the values of the multipliers λ_ρ.

Besides the constraints can be implemented by the feasible C for different choices of the parameters $\boldsymbol{X}$. More specifically, $\boldsymbol{X} = (\boldsymbol{I}^{\text{u}}, \boldsymbol{I}^{\text{p}}, \boldsymbol{\theta})$, $\boldsymbol{X} = (\boldsymbol{I}, \boldsymbol{Q}, \boldsymbol{U})$ or $\boldsymbol{X} = (\boldsymbol{I}^{\text{u}}, \boldsymbol{Q}, \boldsymbol{U})$ can be chosen. Whatever the choice for $\boldsymbol{X}$, the relations given in Eq. (5) and Eq. (6) can be used to estimate any parameter of interest given the recovered $\widehat{\boldsymbol{X}}$. These relations can also be used to compute the objective function $f(\boldsymbol{X})$ which require the Stokes parameters needed by the direct model of the data Eq. (9) and various polarimetric component depending on the choice of the regularization.

With $\boldsymbol{X} = (\boldsymbol{I}, \boldsymbol{Q}, \boldsymbol{U})$ the positivity constraints take the form of epigraphic constraints that are more difficult to enforce, because it is not separable in the parameters. To solve the problem in Eq. (12) with such a constraint, an epigraphic projection is required, leading to the use of a forward-backard scheme, reduced in this context to a standard projected gradient descent Combettes & Wajs (2005). A description of the method for such a minimization problem can be found in Denneulin et al. (2020). The choice $\boldsymbol{X} = (\boldsymbol{I}, \boldsymbol{Q}, \boldsymbol{U})$ may avoid some degeneracies because it ensure the convexity of the problem Eq. (12). This setting will be refered as linear RHAPSODIE in the experimental part.

With $\boldsymbol{X} = (\boldsymbol{I}^{\text{u}}, \boldsymbol{I}^{\text{p}}, \boldsymbol{\theta})$ or $\boldsymbol{X} = (\boldsymbol{I}^{\text{u}}, \boldsymbol{Q}, \boldsymbol{U})$, the positivity constraints amounts to applying simple separable bound constraints on some parameters. Since the objective is differentiable, a method such as VMLM-B (Thiébaut 2002) can be used to solve the problem in Eq. (12). The method VMLM-B is a gradient-descent algorithm preconditionned with $\ell - BFGS$. At each iterate of the algorithm, the optimal descent step is computed with a linesearch method. The bounds constraint are propagated in the

gradient and are used at each iteration to constrain the direction of the descent and the length of the step. This setting will be refered as non-linear RHAPSODIE in the experimental part.

In the following, we compare the performances of RHAPSODIE for the polarimetric parameters $\boldsymbol{X} = (\boldsymbol{I}, \boldsymbol{Q}, \boldsymbol{U})$ and $\boldsymbol{X} = (\boldsymbol{I}^{\mathrm{u}}, \boldsymbol{Q}, \boldsymbol{U})$ on simulated synthetic datasets. The best choice of polarimetric parameters is then used to process astrophysical datasets.

4. Data calibrations

4.1. Detector calibration

Before the application of a reconstruction method, the calibration of the data is essential to account for the noise and the artifacts linked to the measurement. It allows for the estimation and correction of any pollution induced by the sky background or the instrument as well as the detector behaviour in term of errors on pixel values.

We use an inverse method to calibrate the raw data from these effects: the quantity required for the calibration are jointly estimated from the likelihood of the calibration data direct model (Thiébaut et al. in prep). In this method, all calibration data are expressed as a function of the different contributions (*i.e.* flux, sky background, instrumental background, gain, noise and quantum efficiency). All these quantity are then jointly estimated by the minimization of the log-likelihood of the data. Calibrated data are raw data corrected from the background and the weights are computed using the data variance. Finally, the bad pixels are estimated by crossing several criteria, such as their linearity, their covariance with the other pixels for the several calibration data or the values of their likelihood. This calibration method produces calibrated data outputs $((d_k)_{k\in\{1,\dots K\}})$ and the weights $((\mathbf{W}_k)_{k\in\{1,\dots K\}})$.

4.2. Instrumental calibration

The instrumental calibration allows for the estimation, from dedicated data, of the instrumental PSF and of the star centers on each side of the detector. Since the star is placed behind the coronagraphic mask, simultaneous PSF estimation is not possible. To estimate the PSF, we use a dedicated flux calibration (`STAR-FLUX`) that is acquired just before and after the science exposure by offsetting the telescope to about 0.5 arcsec with respect the coronagraphic mask by using the SPHERE tip/tilt mirror (Beuzit et al. 2019). When performing this calibration, suitable neutral density filters are inserted to avoid detector saturation. It has been shown in (Beuzit et al. 2019) that these neutral density filters do not affect the PSF shape and thus its calibration. This instrumental calibration leads to the estimation of the PSF given by the operator $\mathbf{A}$.

During the coronographic observing sequence, the star point spread function peak is hidden by the coronagraphic mask and its position was determined using a special calibration (`STAR-CENTER`) where four faint replicas of the star image are created by giving a bi-dimensional sinusoidal profile to the deformable mirror (see Beuzit et al. (2019)). The `STAR-CENTER` calibration was repeated before and after each science observation, and resulting center estimations were averaged. In addition we use the derotator position and the true north calibration from Maire et al. (2016) to extract the angle of rotation of the North Axes. These instrumental calibration steps lead to the estimation of the transformation operators $\mathbf{T}_{j,k}$ which rotate and translate the maps of interest in order to make the centers and the North Axes coincide with those in the data.

4.3. Polarization calibration

When the light is reflected by the optical devices in the instrument, some instrumental polarization is introduced, resulting in a loss of polarized intensity and cross-talk contamination between stokes Q and U.

The classical method to compensate for this instrumental polarization is to employs the azimuthal Q_phi and U_phi parameters to reduce the noise floor in the image (Avenhaus et al. 2014). Because this method is limited to face-on disks, we instead use the method developed by (van Holstein et al. 2020) which rely on the pre-computed calibration of the instrumentation polarization as function of the observational configurations. The pipeline IRDAP (van Holstein et al. 2020) yields the possibility to determine and correct the instrumental polarization in the signal reconstruction, yet this reconstruction requires to estimate first the Q and U parameters, to perform the instrumental polarization correction. After computing the double difference, IRDAP uses a Mueller matrix model of the instrument carefully calibrated using real on sky data to correct for the polarized intensity I^{p} (created upstream of the HWP) and crosstalk of the telescope and instrument to compute the model-corrected Q and U images.

Using IRDAP, we estimate the instrumental transmission parameters $(\nu_{j,k,l})_{j\in\{1,\,2\},k\in\{1,\dots K\},\ell\in\{1,\dots L\}}$ from the Mueller matrix to calibrate the instrumental polarization in our datasets.

IRDAP also determines the corresponding uncertainty by measuring the stellar polarization for each HWP cycle individually and computing the standard error of the mean over the measurements. Finally, IRDAP creates an additional set of Q and U images by subtracting the measured stellar polarization from the model-corrected Q and U images. We use a similar method to correct for the stellar polarization which is responsible for strong light polution at small separation in particular for faint disks such as debris disks. The contribution of this stellar light is estimated as different factors of I^{u} in both Q and U, respectively ε_Q and ε_U. We estimate this stellar contributions by using a pixel annulus Ω located at a separation where there is no disk signal (either near the edge of the coronograph or at the separation corresponding to the adaptive optics cut-off frequency). Both corrections factors ε_Q and ε_U as follow:

$$\begin{cases} \varepsilon_Q = (\sum_{n\in\Omega} Q_n / I^{\mathrm{u}}_n)/N_\Omega \\ \varepsilon_U = (\sum_{n\in\Omega} U_n / I^{\mathrm{u}}_n)/N_\Omega, \end{cases} \tag{24}$$

where N_Ω is the number of pixels of the ring Ω. We then compute $Q^{\mathrm{cor}} = Q - \varepsilon_Q I^{\mathrm{u}}$ and $U^{\mathrm{cor}} = U - \varepsilon_U I^{\mathrm{u}}$ in order to create an additional set of Q^{cor} and U^{cor} images by subtracting the measured stellar polarization from the model-corrected Q and U images.

5. Applications on high contrast polarimetric data

5.1. Application on synthetic data

Prior to the application of the RHAPSODIE methods on astrophysical data, the performance of the linear and non-linear methods were evaluated on synthetic datasets, without (resp. with) the deconvolution displayed on Fig. 2, Fig. 3 and Fig. 4 (resp. Fig. 5, Fig. 6 and Fig. 7). These datasets are composed of unpolarized residual stellar flux, mixed with unpolarized and polarized disk flux. We produce several synthetic datasets following steps given

in Appendix A, for different ratio of the polarization of the disk, called τ^{disk}.

The results of the RHAPSODIE methods are compared to the results obtained with the classical Double Difference method. The Double Difference is applied on recentered and rotated datasets with the bad pixels interpolated. For the comparison with deconvolution, the results of the Double Difference are deconvolved after the reduction, by solving the following problem:

$$(\widehat{Q},\widehat{U}) \in \underset{(Q,U)\in\mathbb{R}^N\times\mathbb{R}^N}{\arg\min} \left[\|\widehat{Q}^{\text{D. D.}} - \mathbf{A}Q\|^2 + \|\widehat{U}^{\text{D. D.}} - \mathbf{A}U\|^2 + \lambda \sum_{n=1}^{N} \left(\sqrt{\|\mathbf{D}_n Q\|^2 + \|\mathbf{D}_n U\|^2 + \mu^2} - \mu \right) \right]. \quad (25)$$

This deconvolution method allows for the deconvolution of stokes Q and U jointly in order to estimate the polarization angle. It differs from standard approaches which perform the deconvolution of stokes Q and U independently (Heikamp & Keller 2019). Such method, sharpen the polarized intensity image, but does not recover the polarization signal lost by averaging over close-by polarization signals with opposite sign and may even introduce artificial structures that are not present in the original source.

The hyperparameters of regularization λ_1, λ_2 and μ are chosen in order to minimize the total Mean Square Error, *i.e.* the sum of the MSE of each map of intensity of interest I^{u}, I^{p}, θ, given by:

$$\begin{aligned} \text{MSE}^{\text{tot}} = &\sum_{n=1}^{N^{I^{\text{u}}}} \mathbb{E}\left[(\widehat{I^{\text{u}}} - \overline{\widehat{I^{\text{u}}}})^2\right] \\ &+ \sum_{n=1}^{N^{I^{\text{p}}}} \mathbb{E}\left[(\widehat{I^{\text{p}}} - \overline{\widehat{I^{\text{p}}}})^2\right] \\ &+ \sum_{n=1}^{N^{\theta}} \mathbb{E}\left[\text{angle}(e^{2i(\widehat{\theta} - \overline{\widehat{\theta}})})^2\right]. \end{aligned} \quad (26)$$

where $N^{I^{\text{u}}}$, $N^{I^{\text{p}}}$ and N^{θ} are the number of pixels with signal of interest.

For the reconstructions without the deconvolution, when the SNR is smaller than 10 (see Fig. A.1), the residuals, as shown on Fig. 3, are larger for the Double Difference than for the RHAPSODIE methods. Yet, on Fig. 2, if the reconstruction with RHAPSODIE methods is smoother than the Double Difference, the thin ring seems to disappear, which is not the case for the Double Difference. The MSE on Fig. 4, however, is smaller for the RHAPSODIE methods than for the Double Difference. It is possible to recover such sharper structure with RHAPSODIE, by reducing the regularization contribution (*i.e.* reducing the hyperparameter λ_2). Yet, if λ_2 is too small, the data will be overfitted and the noise in the reconstruction will be amplified. It is thus necessary to keep a good trade-off between a smooth solution and a solution close to the data. Classically, minimizing the MSE is a good trade-off between underfitting and overfitting.

When RHAPSODIE is used without the deconvolution, the efficiency of the linear and non-linear reconstruction is comparable. When the convolution is included in the model, the non-linear reconstruction is more efficient, because it is not polluted by the deconvolution of the unpolarized point source companion as seen on Fig. 6. The non-linear model give also a better angle reconstruction when $\tau^{\text{disk}} > 10\%$, proving that such a choice of parametrization of the model is pertinent.

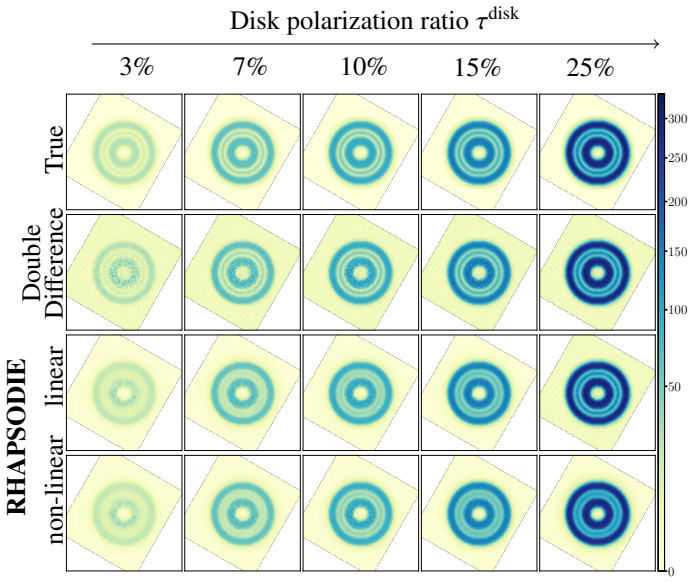


Fig. 2: Visual comparison of the reconstructed polarized intensity I^{p} with the state-of-the-arte Double Difference and the RHAPSODIE methods without deconvolution.

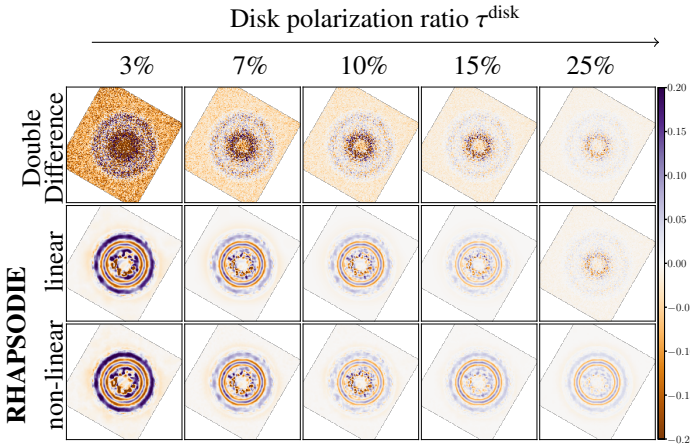


Fig. 3: Maps of residuals of the reconstructions displayed on Fig. 2. These residuals are obtained as the difference between the true and the reconstructed images.

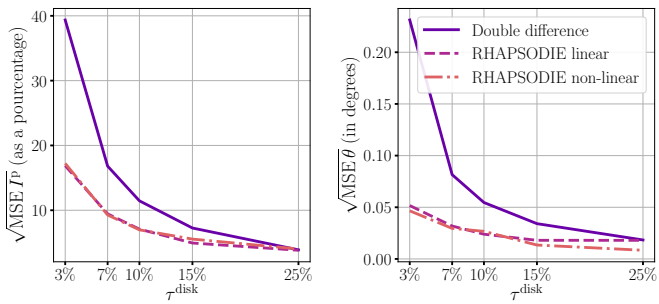


Fig. 4: Comparison of the MSE between the true map and the estimated map of the polarized intensity I^{p} and of the angle of polarization θ for the Double Difference and the linear and non-linear RHAPSODIE methods without deconvolution.

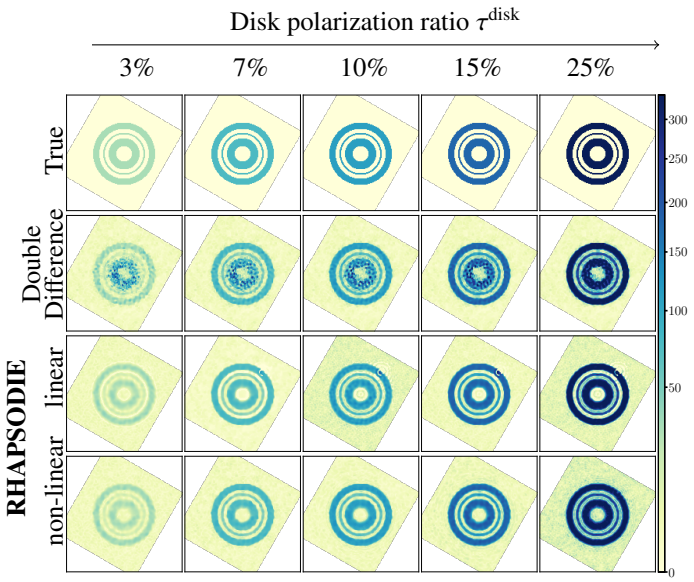


Fig. 5: Visual comparison of the reconstructed polarized intensity I^{p} with the state-of-the-art Double Difference and the RHAPSODIE methods without deconvolution.

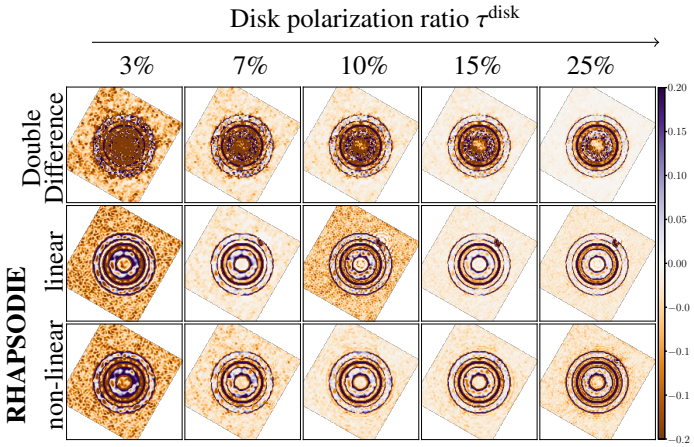


Fig. 6: Maps of residuals of the reconstructions displayed on Fig. 5. These residuals are obtained as the difference between the true and the reconstructed images.

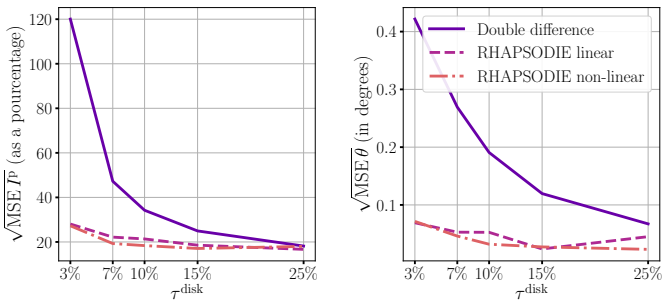


Fig. 7: Comparison of the MSE between the true map and the estimated map of the polarized intensity I^{p} and of the angle of polarization θ for the Double Difference and the linear and non-linear RHAPSODIE methods with deconvolution.

For the reconstructions with the deconvolution, the RHAPSODIE methods give clearly a better reconstruction of the polarized intensity as shown on the reconstructions in Fig. 5, on the residuals in Fig. 6 and on the MSE in Fig. 7. In fact, for $\tau^{\text{disk}} = 3\%$, RHAPSODIE delivers polarized intensity estimation more accurate by a factor 4 based on the simulated dataset. The residuals of the reconstructions, are much larger for the Double Difference than for the RHAPSODIE methods.

The RHAPSODIE methods also allow a better reconstruction of the angle of polarization (see Fig. 4 and Fig. 7), in particular for the non-linear method. The gain on the deconvolved reconstruction with the non-linear RHAPSODIE method of the angle of polarization, compared to that of the Double Difference, is almost always a factor 4, based on the simulated dataset.

According to these results, the RHAPSODIE methods are better than the state-of-the-art methods, in particular the non-linear RHAPSODIE method. This is why in the following section dedicated to the astrophysical data, we select the non-linear RHAPSODIE method with manual selection of the hyperparameters.

5.2. Astrophysical data

RHAPSODIE was applied to several IRDIS datasets dedicated to protoplanetary, transition and debris disks, to test the efficiency of our method and to compare it to the state-of-the-art method. For each reconstruction, we present the map of the projected intensities, by using the standard azimuthal Stokes parameters Q_ϕ and U_ϕ estimated from:

$$\begin{cases} Q_\phi = Q_n \cos(2\psi_n) + U_n \sin(2\psi_n) \\ U_\phi = U_n \cos(2\psi_n) - Q_n \sin(2\psi_n), \end{cases} \tag{27}$$

with $\psi_n = \arctan\left(\frac{n_1 - n_1^{\text{center}}}{n_2 - n_2^{\text{center}}}\right)$ where (n_1, n_2) denotes the row and the column indices of the n-th pixel in the map and $\left(n_1^{\text{center}}, n_2^{\text{center}}\right)$ those of the pixel center. We also present the maps of polarized intensities I^{p} and the angles of polarization θ estimated by the different methods.

First, the reconstructions of the target TW Hydrae with the Double Difference and RHAPSODIE without deconvolution are compared in the Fig.8, without and with the correction of the instrumental polarization. The images are scalled by the square of the separation to account for the drop-off of stellar illumination with distance.

The instrumental polarization in this dataset introduces a polarization rotation and an attenuation (loss of polarization signal) of the intensity with varying time during the observations which can be monitored easilly because the disk is face-on. If uncorrected, the combination of data from multiple polarimetric cycles will result in very poorly constrained polarimetric intensity measurements. When we correct the instrumental polarization (see Fig. 8c) by using the IRDAP method, these effects are compensated and the disk reconstructed is more accurate (Fig. 8c (iii). As a result, the contamination of the Uphi signal from cross-talk is decreased and becomes negligeable as seen on Fig. 8c (iv).

The comparison with van Boekel et al. (2017); de Boer et al. (2020) shows that our method is less impacted by the bad pixels and improves the disk SNR in area where the signal is low. On the other hand it is slightly more sensitive to detector flat calibration accuracy. It is worth noticing that IRDAP has a dedicated correction of the detector response non uniformity (flat field variation between the detector column) for the various amplifiers that is not implemented in RHAPSODIE. However the

Target	Date	Filter	Δ_t (s)	$K^{\rm HWP}$	$K^{\rm miss}$	$\Delta_t^{\rm tot}$ (s)	$\lambda_{I^{\rm u}}^{\rm no\text{-}dec}$	$\lambda_{Q+U}^{\rm no\text{-}dec}$	$\mu^{\rm no\text{-}dec}$	$\lambda_{I^{\rm u}}^{\rm dec}$	$\lambda_{Q+U}^{\rm dec}$	$\mu^{\rm dec}$
TW Hydrae	2015-04-01	H	16	22	13	5424	10^{4}	10^{4}	10^{-4}	10^{4}	10^{4}	10^{-5}
IM Lupus	2016-03-14	H	64	6	7	2624	10^{0}	10^{0}	10^{-4}	10^{5}	$10^{5.5}$	10^{-3}
MY Lupus	2016-03-16	H	64	7	18	2432	10^{0}	10^{-1}	10^{-4}	10^{1}	$10^{0.5}$	10^{-3}
RY Lupus	2016-05-27	H	32	8	0	4096	10^{4}	$10^{5.5}$	10^{-3}	10^{5}	10^{3}	10^{-3}
T Chae	2016-02-20	H	32	30	0	3840	10^{0}	10^{-1}	10^{-3}	$10^{0.5}$	10^{-1}	10^{-2}
RXJ 1615	2016-03-15	H	64	11	7	5184	10^{0}	10^{0}	10^{-4}	10^{1}	$10^{-1.5}$	10^{-3}
HD 106906	2019-01-17/18/20	H	32	42	0	5376	$10^{2.5}$	$10^{1.5}$	10^{-4}	10^{3}	$10^{0.75}$	10^{-3}
HD 61005	2015-05-02	H	16	12	16	2816	10^{5}	10^{5}	10^{-3}	10^{6}	10^{5}	10^{-3}
AU MIC	2017-06-20	J	16	23	0	11776	10^{2}	10^{4}	10^{-3}	10^{5}	10^{4}	10^{-3}

Table 3: Information of the datasets used for the target reconstruction presented in this section: the name of the target, the date of the observation, the filter used, the exposition time Δ_t for one acquisition, the number of cycles of HWP $K^{\rm HWP}$, the number of frame missing or removed $K^{\rm miss}$, the total exposition time of the observation $\Delta_t^{\rm tot}$ considering $K^{\rm miss}$ and the values of the hyperparameters $\lambda_{I^{\rm u}}^{\rm no\text{-}dec}$, $\lambda_{Q+U}^{\rm no\text{-}dec}$, $\mu^{\rm no\text{-}dec}$ (resp. $\lambda_{I^{\rm u}}^{\rm dec}$, $\lambda_{Q+U}^{\rm dec}$, $\mu^{\rm dec}$) used for the reconstruction with RHAPSODIE without deconvolution (resp. with deconvolution).

flat calibration of observations more recent than 2016 have improved and do allow better calibration and as a consequence do not impact our method efficiency anymore.

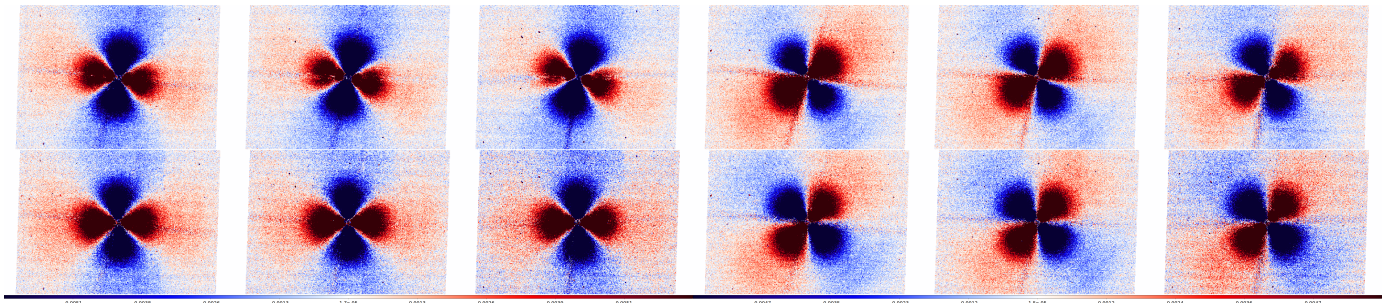
The efficiency of the method we have developed is demonstrated on the figures 9, 10 and 11. Fig. 9 present the RHAPSODIE reconstructions without and with the deconvolution of the protoplanery disks TW Hydrae (van Boekel et al. 2017), IM Lupus (Avenhaus et al. 2018) and MY Lupus (Avenhaus et al. 2018). Fig. 10 presents the RHAPSODIE reconstructions without and with deconvolution of transition disks RY Lupus (Langlois et al. 2018), T Chamaeleontis (Pohl et al. 2017), RXJ 1615 (de Boer et al. 2016). Fig. 11 presents the RHAPSODIE reconstructions without and with the deconvolution of the debris disks HD 106906 (Kalas et al. 2015; Lagrange et al. 2016), HD 61005 (Olofsson et al. 2016), and AU Mic (Boccaletti et al. 2018) compared to the double difference method. The brightness of the disks selected vary by 7 magnitudes in our selection. The contrast between the disk and their host star is also very different and could be more favorable for highly inclined disk such as MY Lup. In such case the star likely shines partially through the disk, which is dimming the star light and thus decreasing the contrast between the star and the disk.

In all cases the method produces high quality restitution of the disk polarized signal and minimizes the artifacts from bad pixels. The U_ϕ signal is as expected for face-on and inclined disks. The instrumental polarization correction has been efficiently applied and the U_ϕ signal remains very low and represents noise for face-on disks. The comparison of these reductions with the state-of-the-art confirms the benefits of our method but it is more difficult to quantify the gain on real data not knowing the truth and should be consolidated by using numerical models of these objects. In addition our method allows to deconvolved these images and to clearly enhance the angular resolution that will benefits the interpretation of the disk morphology and physical properties. The deconvolution sharpen the polarized intensity image, and seems to recover the polarization signal lost by averaging over close-by polarization signals with opposite sign and does not seem to introduce artificial structures that are not present in the original source (see 10 (RY Lup)). As mentioned before the hyperparameters could be further tuned to adjust the smoothing of the deconvolved images according to required noise/angular resolution trad-off.

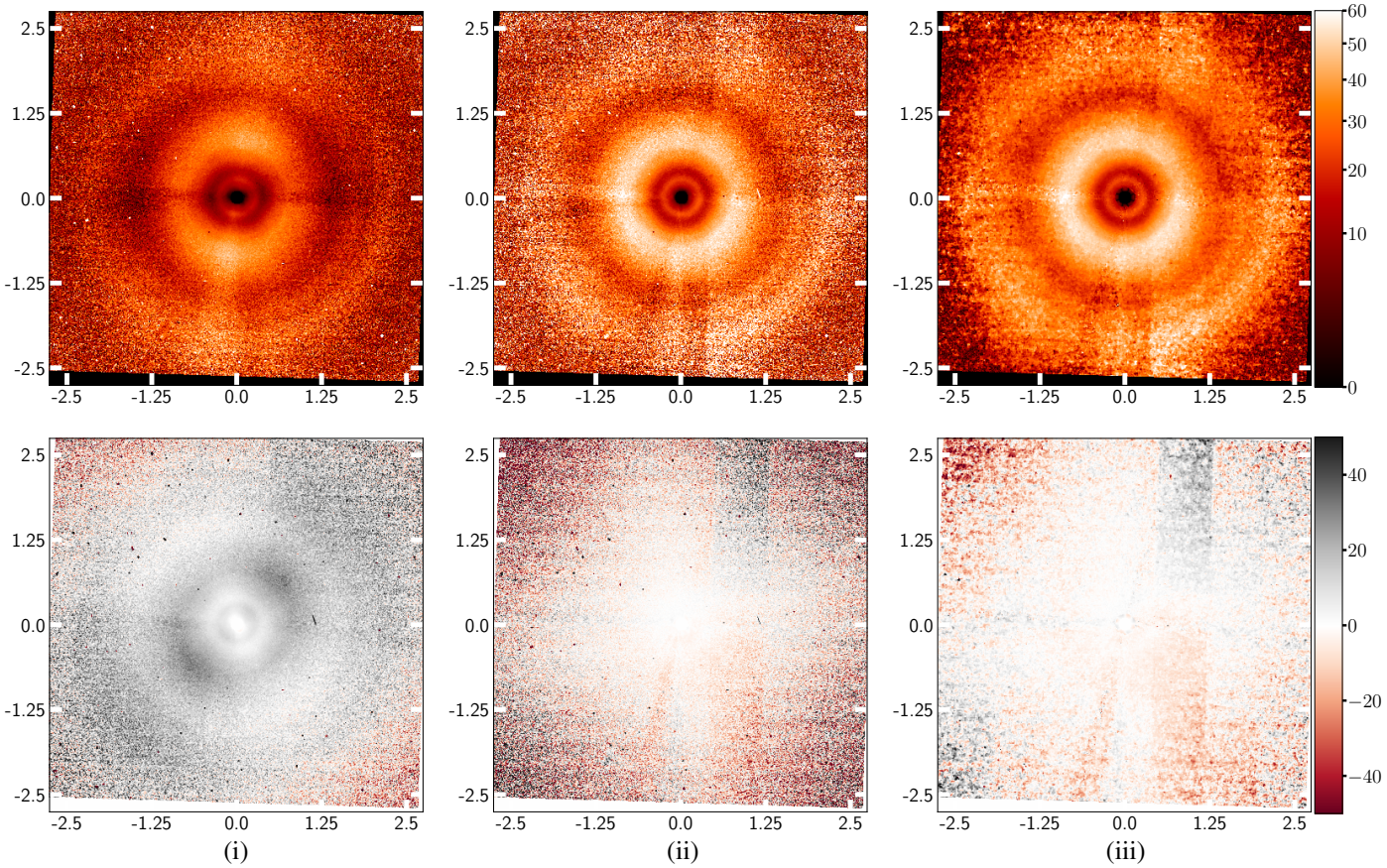
For several of these disks (IM Lup, MY Lup), the outer edge of the disk and thus the lower disk surface have been detected in Avenhaus et al. (2018) and are confirmed by our method. The deconvolution of these datasets allows to highligh further these features, to more clearly see the midplane gaps and to discover the midplane gap in the case of T Cha which was not identified by Pohl et al. (2017). The reason is that without deconvolution, the PSF smears light from the disk upper and lower sides into the visible midplane gap.

Except for Au Mic, The debris disk are fainter than the protoplenetary or the transition disks in polarised intensity. As a consequence these datasets have required carefull stellar polarisation compensation. The HD 106906 debris disk is viewed close to edge-on in polarised light as reported in van Holstein et al. (2020); Esposito et al. (2020). The image clearly shows the known East - West brightness asymmetry of the disk, which was detected in total intensity (Kalas et al. 2015; Lagrange et al. 2016). Thanks to the deep dataset and good reconstruction, we also detect the backward-scattering far side of the disk to the west of the star, just south of the brighter near side of the disk. This feature is further highlighted by the deconvolution. For HD61005 and Au Mic our method also proves to be efficient in recovering the disk polarised signal and to deconvolve this signal despite strong artefacts produced by the rotating spiders (i.e unmasked by the Lyot stop) when observing in field stabilised. The difference in the spider position during the polarimetric cycle results in an artifical polarimetric signal when using standard data reduction techniques. It is worth noticing that the strength of our method to deal with these artefacts comes from its capability to use weighted maps to account for them. For instance the spiders are weighted by their variance during their rotation frame by frame in addition to the inclusion of a static bad pixel ponderation. As a result the spider contributions to the polarimetric signal when using our method is decreased compared to the other methods. The noise which is created by these spiders remains as seen on Fig. 11 and can generate artefacts in the deconvolution as seen for HD 61005. Counteracting the noise from the spiders can be efficiently done by performing DPI observation in pupil tracking as proposed in van Holstein et al. (2017).

Another advantage of our method is its ability to use uncomplete polarimetric cycles which are discarded by the classical methods. This capability of our method leads to an increase in SNR which was quantified more precisely using our model of the data in Denneulin (2020). In order to benefit from this improvement, the instrumental polarization has to remain small because uncomplete polarimetric cycles do not benefit from the instrumental polarisation compensation performed by estimating both couples Q and -Q (and U and -U, respectively).



(a) Stokes parameter Q. (b) Stokes parameter U.



(c) Q_ϕ (above) and U_ϕ (below) displayed in arcseconds. We show the reconstruction with the Double Difference without (i) and with (ii) the instrumental polarization correction. The right column (iii) shows the RHAPSODIE reconstruction with the instrumental polarization correction. The intensities are multiplied in each pixel by the distance to the star r^2.

Fig. 8: Reconstruction of the Q (a) and U (b) parameters for the first three cycle of HWP rotation, without (upper row) and with (lower row) the correction of the polarization. Without the correction, both Q and U are rotated and attenuated. The Q_ϕ and U_ϕ images reconstructed from the entire dataset are presented in (c). All the reconstructions are done without deconvolution in order to demonstrate mainly the efficiency of the instrumental polarization correction and the benefits of RHAPSODIE.

6. Conclusion

We developed a new method to extract the polarimetric signal using an inverse approach method that benefits from the added knowledge of the detected signal formation process. The method includes weighted data fidelity term, smooth penalization, and takes into account instrumental and stellar polarization. This method enables to accurately measure the polarized intensity and angle of linear polarization of circumstellar disks by taking into account the noise propagation and the observed objects convolution. It has the capability to use uncomplete polarimetry cycles (when the instrumental polarization is small) which enhances the sensitivity of these observations. It also takes proper account for bad pixels by using weighted map instead of interpolating them. These bad pixels can cause systematic errors of several tenths of a percent in the polarization measurements as shown by (van Holstein 2020, in prep.) In addition the effect of bad pixel interpolation could also have some impact when reaching 0.1% polarimetric accuracy.

We have validated the method on both simulated and archive data from SPHERE/IRDIS and compared its performances with the state-of-the-art methods. We have implemented the method in a end-to-end data-analysis package called RHAPSODIE. The method we developed improves the overall performances in particular at low SNR/small polarized flux compared to standard

methods. By increasing the sensitivity and including deconvolution, this method will allow for more accurate studies of the orientation and morphology of the disks, especially in the innermost regions. It also will enable more accurate measurements of the angle of linear polarization at low SNR, which would allow for a more in-depth studies of dust properties. Finally, the method will enable more accurate measurements of the polarized intensity which is critical to construct scattering phase functions.

7. Acknowledgements

We acknowledge Rob Van Holstein for his help with the instrumental polarisation correction based on carefull instrumental calibrations he performed which are implemented in the IRDAP tool he developped. This work has made use of the SPHERE Data Centre, jointly operated by OSUG/IPAG (Grenoble), PYTHEAS/LAM/CeSAM (Marseille), OCA/Lagrange (Nice), Observatoire de Paris/LESIA (Paris), and Observatoire de Lyon (OSUL/CRAL). This work is supported by the French National Programms PNP and the Action Spécifique Haute Résolution Angulaire (ASHRA) of CNRS/INSU co-funded by CNES. The authors are grateful to the LABEX Lyon Institute of Origins (ANR-10-LABX-0066) of the Université de Lyon for its financial support within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) of the French government operated by the National Research Agency (ANR). This paper is based on observations made with ESO Telescopes at the La Paranal Observatory under programme ID: 598.C-0359, 095.C-0273, 0102.C-0916, 096.C-0523, 097.C-0523, 096.C-0248.

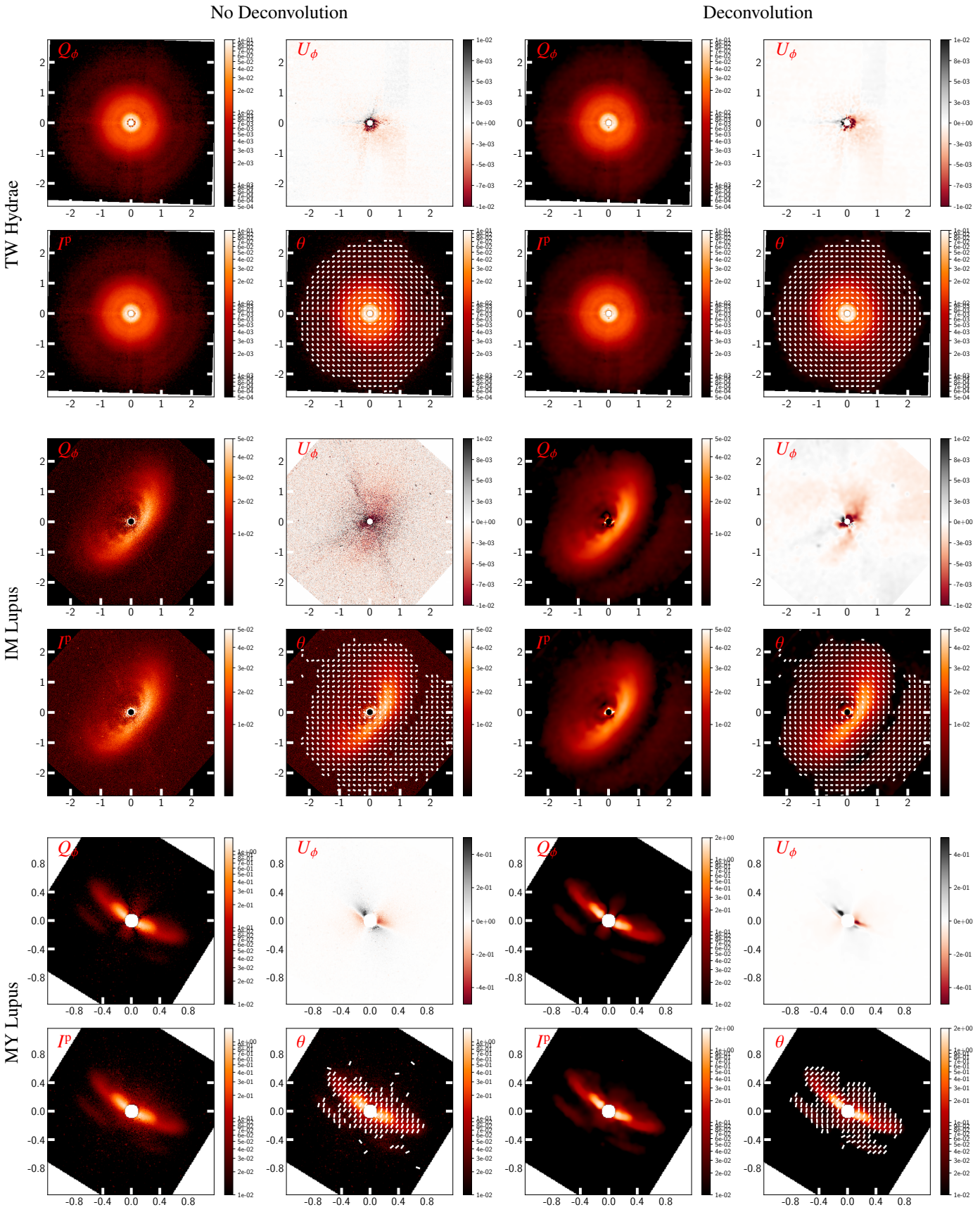


Fig. 9: Reconstruction of several protoplanetary disks with RHAPSODIE without and with the convolution in the data physical model. Are represented in each case: the absolute values of Q_ϕ (upper left), U_ϕ (upper right), I^p (lower left) and I^p with θ displayed over it (lower right). The maps of Q_ϕ and I^p are displayed in arcseconds using a squared root scale for IM Lupus and a logarithmic scale for TW Hydrae and MY Lupus. The maps of U_ϕ are displayed using a linear scale. The pixels lying underneath the coronograph are masked in black for the squared root scale and in white for the logarithmic and linear scales. North is up and East is to the left in all frames.

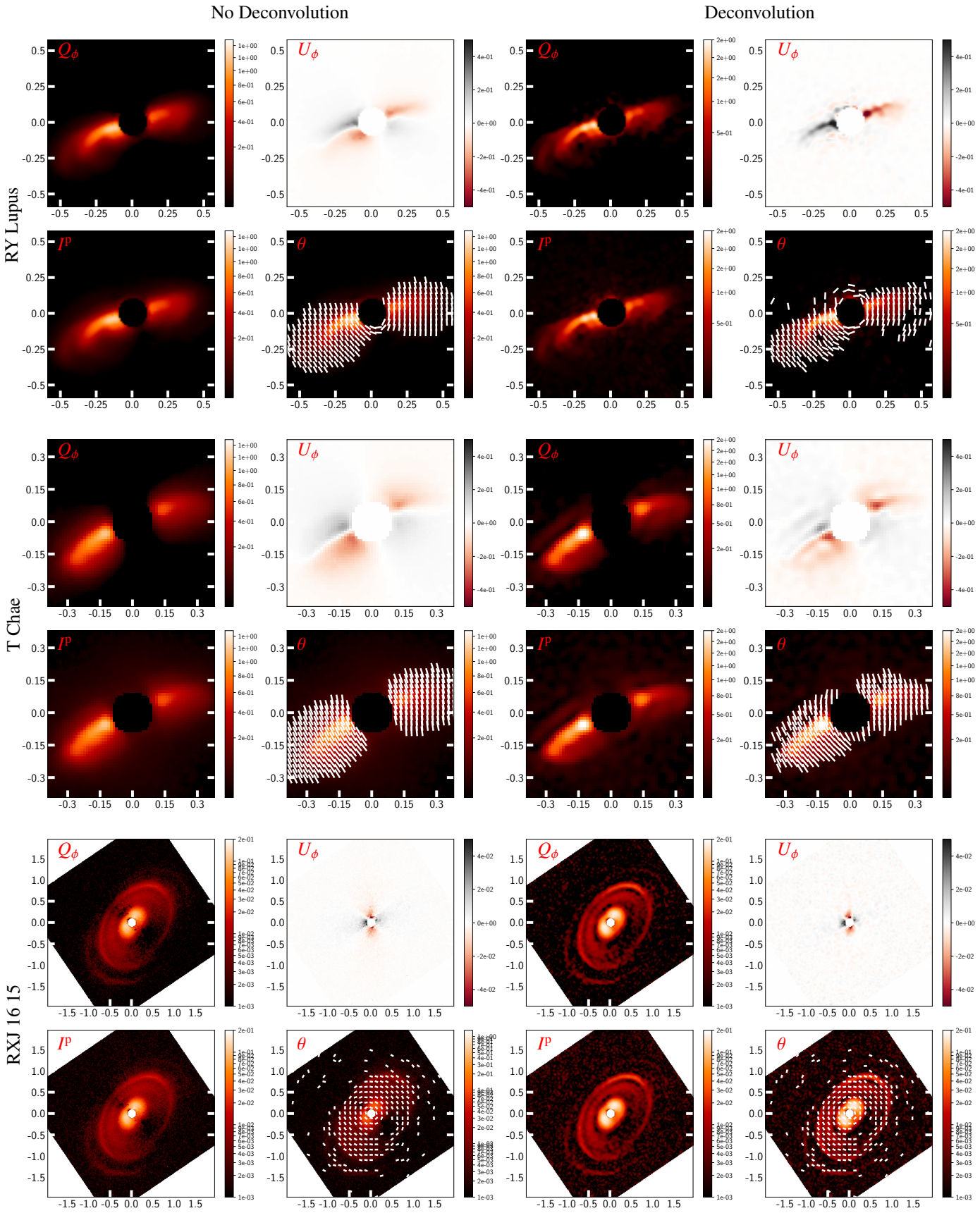


Fig. 10: Reconstruction of several transition disks with RHAPSODIE without and with the convolution in the data physical model. Are represented in each case: Q_ϕ (upper left) (or its absolute values for RXJ 1615), U_ϕ (upper right), I^p (lower left) and I^p with θ drawn over it (lower right). The maps of Q_ϕ and I^p are displayed in arseconds using a squared root scale for RY Lupus and T Chae and logarithmic scale for RXJ 1615. The maps of U_ϕ are displayed with a linear scale. The pixels corresponding to the coronograph are masked in black for the squared root scale and in white for the logarithmic and the linear scales. North is up and East is to the left in all frames.

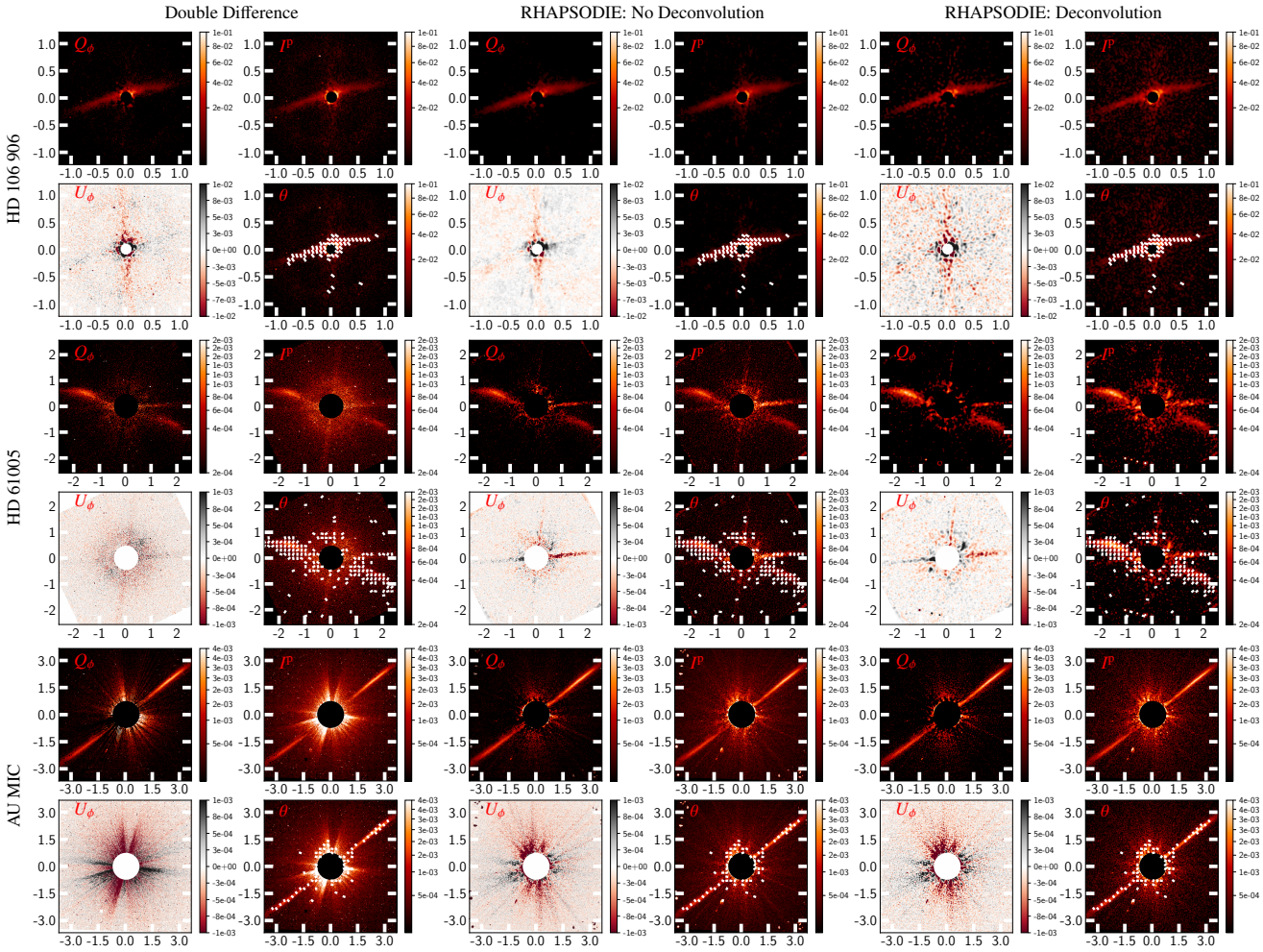

Double Difference
RHAPSODIE: No Deconvolution
RHAPSODIE: Deconvolution
HD 106 906
HD 61 005
AU MIC
Q_ϕ
I^p
U_ϕ
θ

Fig. 11: Reconstruction of several debris disks with the Double Difference and with RHAPSODIE, without and with the convolution in the data physical model (landscape page). The distance to the star center is plotted in arc-seconds. Are represented in each case: the values of Q_ϕ (upper left), the values of U_ϕ (lower left), the polarized intensity I^{p} (upper right) and the polarized intensity I^{p} with the angles θ displayed over it (lower right). The images of Q_ϕ and I^{p} are showed using a squared root scale and U_ϕ using a linear scale. The pixels corresponding to the coronograph are masked in black for the squared root scale and in white for the linear scale. For AU MIC, the mask is extended further because the pixels are saturated up to a radius of 10 pixels around the coronograph, which pollute the reconstruction (as seen also for the Double Difference results).North is up and East is to the left in all frames.

References

Adam, R., Ade, P. A., Aghanim, N., et al. 2016, A&A, 594, A8
Akiyama, K., Alberdi, A., Alef, W., et al. 2019, ApJ Letters, 875, L4
Avenhaus, H., Quanz, S. P., Garufi, A., et al. 2018, ApJ, 863, 44
Avenhaus, H., Quanz, S. P., Schmid, H. M., et al. 2014, ApJ, 781, 87, arXiv: 1311.7088
Benisty, M., Juhasz, A., Boccaletti, A., et al. 2015, A&A, 578, L6
Berdeu, A., Soulez, F., Denis, L., Langlois, M., & Thiébaut, É. 2020, A&A, 635, A90
Beuzit, J.-L., Vigan, A., Mouillet, D., et al. 2019, A&A, 631, A155
Biraud, Y. 1969, A&A, 1, 124
Birdi, J., Repetti, A., & Wiaux, Y. 2018, MNRAS, arXiv: 1801.02417
Birdi, J., Repetti, A., & Wiaux, Y. 2019, arXiv:1904.00663 [astro-ph], arXiv: 1904.00663
Boccaletti, A., Sezestre, E., Lagrange, A. M., et al. 2018, A&A, 614, A52
Borde, P. J. & Traub, W. A. 2006, ApJ, 638, 488
Carrillo, R. E., McEwen, J. D., & Wiaux, Y. 2012, MNRAS, 426, 1223
Charbonnier, P., Blanc-Féraud, L., Aubert, G., & Barlaud, M. 1997, IEEE TIP, 6, 298
Chierchia, G., Pustelnik, N., Pesquet-Popescu, B., & Pesquet, J.-C. 2014, IEEE Transactions on Image Processing, 23, 5531
Claudi, R. U., Turatto, M., Gratton, R. G., et al. 2008, in Ground-based and Airborne Instrumentation for Astronomy II, Vol. 7014, International Society for Optics and Photonics, 70143E
Combettes, P. & Pesquet, J.-C. 2011, 185
Combettes, P. L. & Wajs, V. R. 2005, Multiscale Model. Simul., 4, 1168
de Boer, J., Langlois, M., van Holstein, R. G., et al. 2020, A&A, 633, A63
de Boer, J., Salter, G., Benisty, M., et al. 2016, A&A, 595, A114
de Boer, J., Salter, G., Benisty, M., et al. 2016, A&A, 595, A114
Deledalle, C.-A., Vaiter, S., Fadili, J., & Peyré, G. 2014, SIIMS, 7, 2448
Denneulin, L. 2020, PhD thesis
Denneulin, L., Langlois, M., Pustelnik, N., & Thiébaut, É. 2019, in GRETSI LILLE 2019
Denneulin, L., Pustelnik, N., Langlois, M., Loris, I., & Thiébaut, É. 2020, iTwist, Nantes, France, Dec. 2-4, 2020.
Dipierro, G., Price, D., Laibe, G., et al. 2015, MNRAS Letters, 453, L73
Dohlen, K., Langlois, M., Saisse, M., et al. 2008, in Ground-based and Airborne Instrumentation for Astronomy II, Vol. 7014, International Society for Optics and Photonics, 70143L
Eldar, Y. C. 2008, IEEE Transactions on Signal Processing, 57, 471
Esposito, T. M., Kalas, P., Fitzgerald, M. P., et al. 2020, AJ, 160, 24
Flasseur, O., Denis, L., Thiébaut, É., Olivier, T., & Fournier, C. 2019, in 2019 27th European Signal Processing Conference (EUSIPCO), IEEE, 1–5
Garufi, A., Benisty, M., Stolker, T., et al. 2017, arXiv preprint arXiv:1710.02795
Ginski, C., Stolker, T., Pinilla, P., et al. 2016, A&A, 595, A112
Golub, G. H., Heath, M., & Wahba, G. 1979, Technometrics, 21, 215
Haffert, S. Y., Bohn, A. J., de Boer, J., et al. 2019, Nat Astron, 3, 749
Hansen, P. C. & O'Leary, D. P. 1993, SISC, 14, 1487
Heikamp, S. & Keller, C. U. 2019, A&A, 627, A156
Högbom, J. 1974, A&A Supplement Series, 15, 417
Kalas, P. G., Rajan, A., Wang, J. J., et al. 2015, ApJ, 814, 32
Keppler, M., Benisty, M., Müller, A., et al. 2018, A&A, 617, A44
Keppler, M., Teague, R., Bae, J., et al. 2019, A&A, 625, A118
Lagrange, A. M., Langlois, M., Gratton, R., et al. 2016, A&A, 586, L8
Langlois, M., Dohlen, K., Vigan, A., et al. 2014, in Ground-based and Airborne Instrumentation for Astronomy V, Vol. 9147, International Society for Optics and Photonics, 91471R

Langlois, M., Pohl, A., Lagrange, A.-M., et al. 2018, A& A, 614, A88
Lefkimmiatis, S., Ward, J. P., & Unser, M. 2013, IEEE transactions on image processing, 22, 1873
Mahalanobis, P. C. 1936, National Institute of Science of India
Maire, A.-L., Langlois, M., Dohlen, K., et al. 2016, in Ground-based and Airborne Instrumentation for Astronomy VI, Vol. 9908, International Society for Optics and Photonics, 990834
Marois, C., Lafreniere, D., Doyon, R., Macintosh, B., & Nadeau, D. 2006, ApJ, 641, 556
Milli, J., Engler, N., Schmid, H. M., et al. 2019, 13
Muto, T., Grady, C., Hashimoto, J., et al. 2012, ApJ Letters, 748, L22
Nocedal, J. & Wright, S. J. 1999, Numerical optimization, Springer series in operations research (New York: Springer)
Olofsson, J., Samland, M., Avenhaus, H., et al. 2016, A&A, 591, A108
Pairet, B., Jacques, L., & Cantalloube, F. 2019, Proceedings of SPARS'19, 1, 1
Perrin, M. D., Duchene, G., Millar-Blanchaer, M., et al. 2015, ApJ, 799, 182
Pinte, C., van der Plas, G., Menard, F., et al. 2019, Nat Astron, 3, 1109, arXiv: 1907.02538
Pohl, A., Sissa, E., Langlois, M., et al. 2017, A&A, 605, 17
Price, D. J., Cuello, N., Pinte, C., et al. 2018, MNRAS, 477, 1270, arXiv: 1803.02484
Pustelnik, N., Benazza-Benhayia, A., Zheng, Y., & Pesquet, J.-C. 2016, in Wiley Encyclopedia of Electrical and Electronics Engineering (Hoboken, NJ, USA: John Wiley & Sons, Inc.), 1–34
Quanz, S. P., Avenhaus, H., Buenzli, E., et al. 2013, ApJ Letters, 766, L2
Ramani, S., Blu, T., & Unser, M. 2008, IEEE TIP, 17, 1540
Rudin, L. I., Osher, S., & Fatemi, E. 1992, Physica D: Nonlinear Phenomena, 60, 259
Schmid, H. M., Bazzon, A., Roelfsema, R., et al. 2018, A&A, 619, A9
Sissa, E., Olofsson, J., Vigan, A., et al. 2018, A&A, 613, L6
Smirnov, O. M. 2011, A&A, 527, A106
Starck, J.-L., Donoho, D. L., & Candès, E. J. 2003, A&A, 398, 785
Stein, C. M. 1981, The annals of Statistics, 1135
Tarantola, A. 2005, Inverse problem theory and methods for model parameter estimation (SIAM)
Thiébaut, E. 2002, in Astronomical Data Analysis II, Vol. 4847, International Society for Optics and Photonics, 174–183
Thiébaut, E. & Conan, J.-M. 1995, JOSA A, 12, 485
Tikhonov, A. N. 1963, Soviet Mathematics Doklady
Tinbergen, J. 2005, Astronomical Polarimetry (Cambridge University Press), google-Books-ID: SAS4JzAaMxkC
Titterington, D. M. 1985, 144, 381
van Boekel, R., Henning, T., Menu, J., et al. 2017, ApJ, 837, 132
van Holstein, R. 2020, A&A: accepted
van Holstein, R. G., Girard, J. H., de Boer, J., et al. 2020, A&A, 633, A64
van Holstein, R. G., Snik, F., Girard, J. H., et al. 2017, Techniques and Instrumentation for Detection of Exoplanets VIII, 38, arXiv: 1709.07519
Vigan, A., Langlois, M., Dohlen, K., et al. 2014, in Ground-based and Airborne Instrumentation for Astronomy V, Vol. 9147, International Society for Optics and Photonics, 91474T

Appendix A: Synthetic dataset simulation

In order to evaluate the performance of the RHAPSODIE method, synthetic data have been created. These synthetic data are designed to reproduce astrophysical cases. First, the truth $N = 128 \times 128$ maps $\overline{\overline{I^{\mathrm{u}}}}$, $\overline{\overline{I^{\mathrm{p}}}}$ et $\overline{\overline{\theta}}$ are created. Such a value of N pixels fits the main Region Of Interest (ROI) size. These maps are represented in Fig. A.2.

The hardest disk structures to reconstruct are faint, small and lightly polarized structures, and consequently have high contrast with the unpolarized stellar intensity. This is why the synthetic environment we generated a disk with three rings of equal brightness but with a different contrast with the unpolarized stellar intensity. This disk is partially polarized with a linearly polarization I^{p} and a polarization angle $\theta \in]-\pi, \pi]$ and an unpolarized component I^{udisk}. The disk polarization ratio between both component is given by:

$$\tau^{\mathrm{disk}} = \frac{I^{\mathrm{p}}}{I^{\mathrm{udisk}} + I^{\mathrm{p}}}. \tag{A.1}$$

. These synthetic images are then combined as maps of Stokes parameters I, Q and U.

The I^{udisk} component is mixed with the unpolarized I^{ustar} stellar components and a point source companion (star close to the host star of different brightness). Both maps are presented on Fig. A.2. The unpolaized intensity is represented as $I^{\mathrm{u}} = I^{\mathrm{ustar}} + I^{\mathrm{udisk}}$. It is important to keep in mind that unmixing both disk and stellar unpolarized components is not possible from DPI data without the diversity introduced by ADI. The τ^{disk} value used to synthetise datasets is thus inaccessible in practice from observational polarimetric datasets.

Finally, to generate synthetic calibrated data, the Stokes maps are combined following the expression of the data physical model (11), with $K \times M$ noise realizations from a fixed random seed. Are also introduced 10% of bad pixels choosen randomly. The weights related to each acquisition are simulated at the same time following (14). The datasets are composed of $K_{\alpha_{\mathrm{rot}}} = 8$ HWP cycles, with $K_{\alpha_{\mathrm{acc}}} = 2$ acquisitions per positions in each cycles, giving $K_\alpha = 16$ total images per position of HWP, for a total of $K = 4K_\alpha = 64$ images per dataset.

Several datasets are created with $\tau^{\mathrm{disk}} \in \{3\%, 7\%, 10\%15\%, 25\%\}$, corresponding to difficult cases for $\tau^{\mathrm{disk}} < 10\%$ and less difficult brighter cases above this threshold. Since this polarized ratio is not accessible in practice, to assert the case difficulty, one can use the total polarization, given for all pixel $n \in \{1, \ldots, N\}$ by:

$$\tau^{\mathrm{total}}(\boldsymbol{x}_n) = \frac{I^{\mathrm{p}}_n}{I^{\mathrm{u}}_n + I^{\mathrm{p}}_n}, \tag{A.2}$$

and the Signal-to-Noise Ratio (SNR), given for all pixel $n \in \{1, \ldots, N\}$ by:

$$\mathrm{SNR}(\boldsymbol{x}_n) = \frac{\sqrt{K_\alpha} I^{\mathrm{p}}_n}{\sqrt{(I^{\mathrm{u}}_n + I^{\mathrm{p}}_n)/2 + \sigma^2_{\mathrm{ro}}}}, \tag{A.3}$$

where σ^2_{ro} is the read-out noise variance. The difference between τ^{total} and τ^{disk} is that the last one does not take in account the unpolarized star residuals. The figure A.1 present the SNR maps and the maps of total polarization ratio of the synthetic parameters generated for different τ^{disk}. At the center, where the unpolarized star residual are the brightest, the SNR and the total polarization ratio are the weakest especially in the case of small I^{udisk}. Yet the SNR grows with the separation from the star center (like the stellar contribution or when the polarized contribution of the disk increases.

Before producing $K \times 2N$ noise realization on each dataset, the random seed is reset to the same value. This allow the reproductibility of the results. In fact this realization is obtained by the multiplication of the standard deviation of the pixel to a gaussian, centered and reduced gaussian realization. Since the random seed is the same for each dataset, the realization is the same for the given pixel, only the standard deviation changes.

In order to compare the results of the Double Difference to the results of the RHAPSODIE methods the dataset are preprocessed. The bad pixels are interpolated; then the left and right part of the images are cut, recentered and rotated.

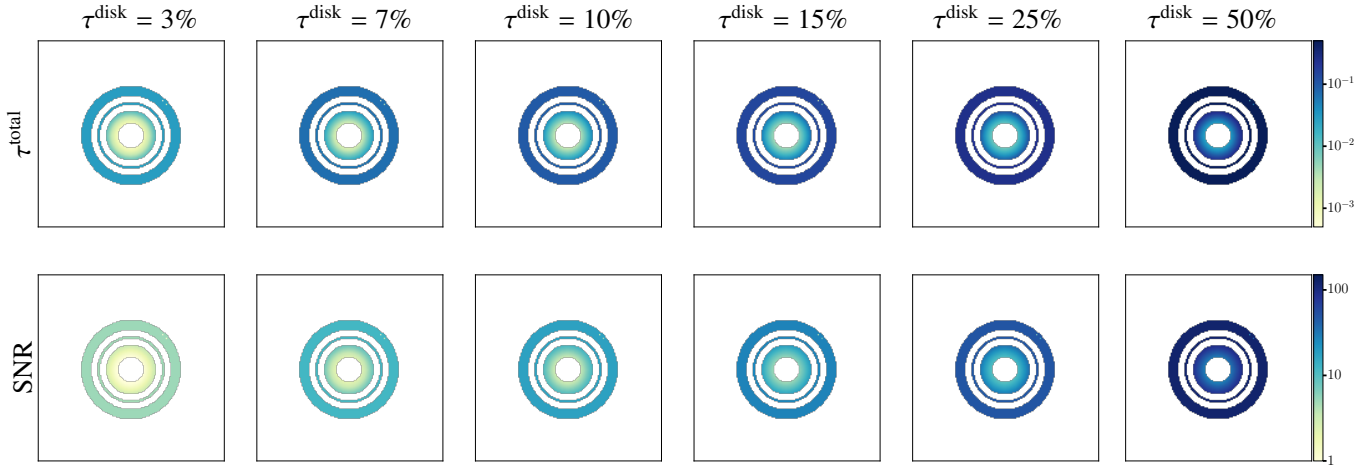


Fig. A.1: τ^{total} and SNR maps for the different values of τ^{disk}.

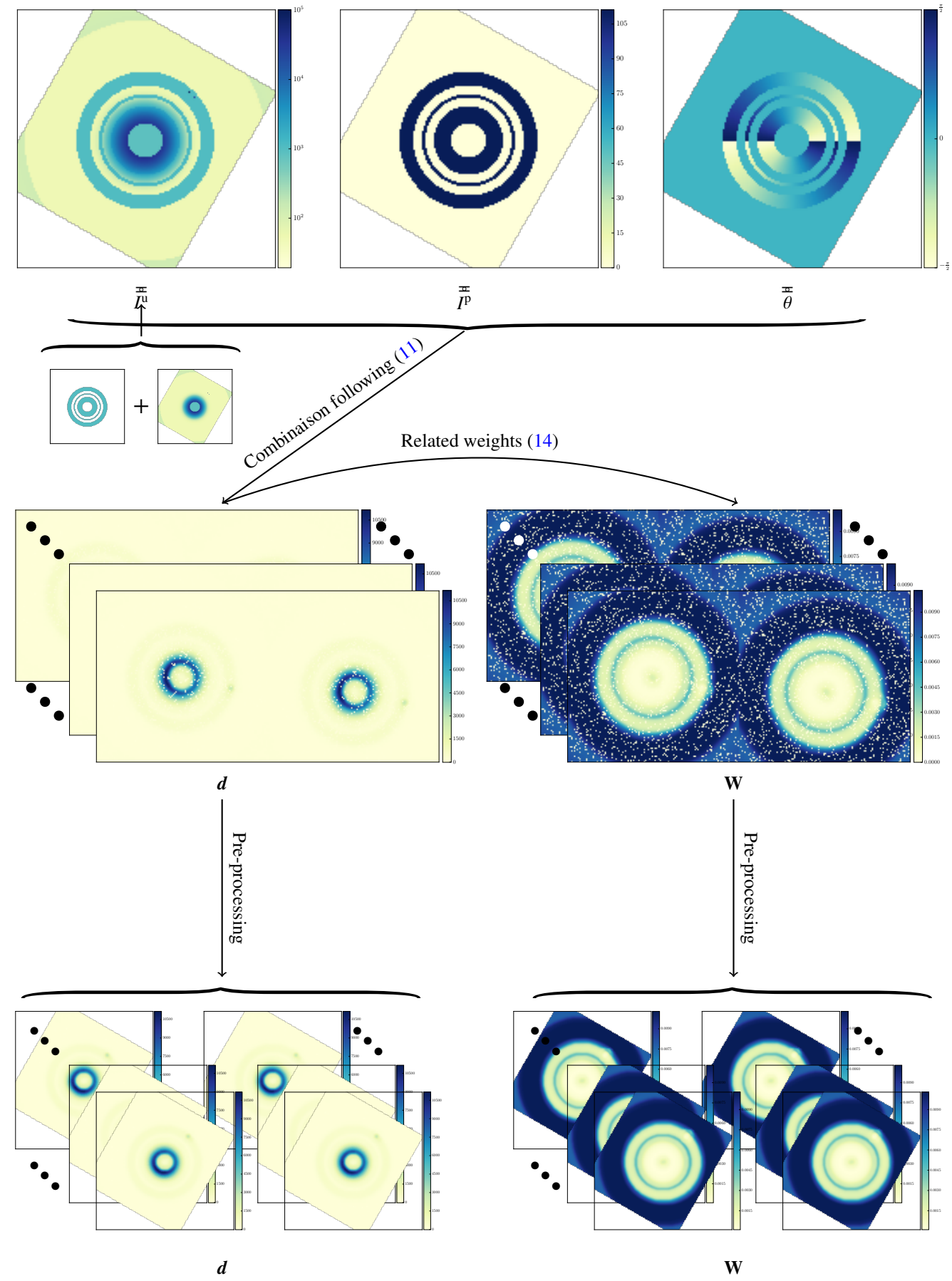


Fig. A.2: Data simulation of calibrated and pre-processed dataset for $\tau^{\mathrm{disk}} = 10\%$.