

**Introduction aux problèmes inverses.  
Applications à l'imagerie optique à haute  
résolution en Astronomie.**

D'après « Des données à l'objet : le problème inverse »,  
chap. 9, sec. 6 de *L'observation en astrophysique*,  
P. Léna et coll.,  
EDP Sciences / CNRS ÉDITIONS,  
3<sup>e</sup> édition, 2008

Laurent Mugnier

Décembre 2008

# Table des matières

9.6	Des données à l'objet : le problème inverse . . . . .	1
9.6.1	Introduction . . . . .	1
9.6.2	Notion de problème bien posé . . . . .	2
9.6.3	Les méthodes d'inversion classiques . . . . .	4
9.6.3.1	Méthode des moindres carrés . . . . .	5
9.6.3.2	Relation entre moindres carrés et filtre inverse . . . . .	5
9.6.3.3	Méthode du maximum de vraisemblance . . . . .	6
9.6.3.4	Exemple : reconstruction de front d'onde par MV . . . . .	7
9.6.3.5	Interprétation de l'échec des méthodes non régularisées . . . . .	8
9.6.4	Inversion régularisée . . . . .	9
9.6.4.1	Estimation bayésienne et estimateur MAP . . . . .	10
9.6.4.2	Équivalence avec la minimisation d'un critère régularisé . . . . .	10
9.6.4.3	Le cas linéaire et gaussien. Lien avec le filtre de Wiener . . . . .	11
9.6.4.4	Application à la reconstruction de front d'onde . . . . .	12
9.6.5	Une application à l'imagerie par optique adaptative . . . . .	13
9.6.5.1	Ingrédients de la déconvolution . . . . .	13
9.6.5.2	Restauration à partir de données astronomiques . . . . .	14
9.6.6	Une application à l'interférométrie coronographique . . . . .	15

## 9.6 Des données à l'objet : le problème inverse

### 9.6.1 Introduction

Généralement les données fournies par l'observation ne correspondent pas directement à la grandeur physique d'intérêt, mais lui sont liées par des lois physiques connues. Un exemple représentatif est l'analyse de front d'onde, c'est-à-dire la mesure d'aberrations (propres à un télescope ou dues à la turbulence), à l'aide d'un analyseur de Hartmann-Shack : la grandeur physique d'intérêt est la phase  $\varphi$  dans la pupille du télescope, qui représente les aberrations. Les données sont les pentes moyennes de la phase sur chaque sous-pupille du Hartmann-Shack selon deux directions perpendiculaires, que l'on peut concaténer dans un vecteur  $i$ . Le calcul des pentes  $i$  connaissant la phase  $\varphi$  est un *problème direct*, classique en Physique, qui nécessite le choix d'un modèle de formation des données<sup>1</sup> ou *modèle direct*.

Estimer ou « remonter à » la phase  $\varphi$  à partir des pentes  $i$  est le *problème inverse* correspondant, qui nécessite d'inverser le modèle de formation des données. Le choix de ce modèle direct n'est pas toujours uniquement dicté par la Physique, il peut résulter d'un compromis entre le souhait d'une modélisation fine des données et le coût lié à l'inversion. Dans l'exemple ci-dessus, on peut envisager d'utiliser un modèle de propagation diffractif lors de la conception de l'analyseur mais on se restreindra à un modèle d'optique géométrique pour l'estimation des aberrations dans le calculateur temps-réel d'une optique adaptative, car celui-ci conduit à un modèle de données linéaire (en fonction des aberrations à mesurer), d'inversion aisée.

On le verra dans la suite, l'inversion peut souvent tirer parti des connaissances statistiques sur les incertitudes de mesures, que l'on modélise généralement comme des bruits. Un modèle direct complet inclut donc la modélisation de la formation des données jusqu'à la détection, voire jusqu'au support de stockage dans le cas de transmission des données avec compression. Il prend ainsi en compte le bruit photonique, le bruit du détecteur, le bruit de quantification du numériseur, le bruit éventuel de compression, etc.

Le traitement de données expérimentales, en Physique et en particulier en Astronomie, consiste donc essentiellement à résoudre un problème inverse, en pratique après une étape de « réduction » ou pré-traitement des données, dont le but est de corriger les défauts instrumentaux de façon à ce que les données puissent être correctement décrites par le modèle direct adopté.

Les méthodes d'inversion naïves ont souvent la caractéristique d'être « instables » au sens où l'inévitable bruit de mesure est amplifié de manière non contrôlée lors de l'inversion, et conduit à une solution inacceptable. Dans ce cas où les données seules ne suffisent pas à obtenir une solution acceptable, il est nécessaire de développer des méthodes d'inversion plus sophistiquées dites *régularisées* qui incorporent des contraintes supplémentaires pour imposer à la solution une certaine régularité compatible avec nos connaissances *a priori* sur celle-ci.

Concevoir explicitement le traitement de données comme l'inversion d'un problème direct est généralement très fructueux. Cela oblige à modéliser l'ensemble du processus de formation de données pour le prendre en compte dans l'inversion. Cela permet aussi d'analyser telle ou telle méthode existante et en particulier d'en expliciter les hypothèses sous-jacentes. Cela permet enfin de concevoir des méthodes tirant parti à la fois des connaissances sur le processus de formation des données et de celles

---

<sup>1</sup>l'analogie pour des mesures quelconques du modèle de formation des images pour un instrument imageur.

que l'on a *a priori*, c'est-à-dire avant de faire les mesures, sur la grandeur physique d'intérêt.

Dans la suite, nous présentons des notions de base et outils pour aborder la résolution d'un problème inverse. Des articles de référence pour le lecteur intéressé par les problèmes inverses sont [1, 2]. Pour aller plus loin un ouvrage pédagogique très complet est [3] ou [4], dont cette introduction s'inspire en partie. Le champ d'application de ces outils est très vaste, comme en témoignent par exemple les chapitres applicatifs (10 à 14) de ces références. Dans le cadre du présent ouvrage, les outils présentés sont illustrés sur quelques problèmes inverses « simples » (c'est-à-dire linéaires) que l'on rencontre en astronomie : reconstruction de front d'onde à partir de mesures Hartmann-Shack, restauration d'images corrigées par optique adaptative, et reconstruction d'images multi-spectrales en interférométrie à frange noire. Ces outils peuvent également s'appliquer à des problèmes inverses non linéaires comme l'analyse de front d'onde au voisinage du plan focal [5, 6] ou la reconstruction d'images en interférométrie optique [7, 8] à partir de modules carrés de visibilité et de clôtures de phase.

### 9.6.2 Notion de problème bien posé

Appelons  $o$  la grandeur physique d'intérêt, qu'on appellera dans la suite l'« objet » observé (phase aberrante dans le cas d'un analyseur de front d'onde, distribution de luminance de l'objet astrophysique observé dans le cas d'une image au plan focal d'un télescope, spectre d'un objet dans le cas d'un spectrographe, etc). Appelons  $i$  les données mesurées (pentes de Hartmann-Shack, image, spectre mesuré, etc). Considérons pour l'instant que  $o$  et  $i$  sont des grandeurs continues, c'est-à-dire des fonctions (de l'espace ou du temps) appartenant à des espaces de Hilbert notés respectivement  $X$  et  $Y$ .

Le modèle direct, déduit de la Physique, permet de calculer un modèle des données pour un objet connu. C'est celui utilisé dans une simulation de données :

$$i = H(o). \quad (9.1)$$

Nous nous limitons ici à un modèle direct *linéaire*, c'est-à-dire :

$$i = Ho, \quad (9.2)$$

où  $H$  est un opérateur linéaire continu de  $X$  dans  $Y$ . C'est dans ce cadre que Hadamard<sup>2</sup> a introduit le concept de *problème bien posé*.

À titre d'exemple et pour fixer les idées, dans le cas où ce modèle direct est linéaire et invariant par translation (imagerie de l'Espace depuis le sol à l'intérieur du domaine isoplanétique par exemple),  $H$  est un opérateur de convolution, et il existe une fonction  $h$  appelée réponse impulsionnelle (RI) ou fonction d'étalement de point (FEP) telle que

$$i = Ho = h \star o. \quad (9.3)$$

On cherche à inverser l'équation (9.2) c'est-à-dire à trouver  $o$  pour  $i$  donné. On dit que le problème est *bien posé* (au sens de Hadamard) si la solution  $o$  vérifie à la fois les conditions habituelles d'existence et d'unicité mais aussi la condition moins classique de stabilité, c'est-à-dire si la solution dépend continûment des données  $i$ . En d'autres termes, une petite variation des données – en pratique, une autre réalisation du bruit aléatoire – ne doit entraîner qu'une petite variation de la solution. Ces trois conditions, dites conditions de Hadamard, se traduisent respectivement par [2] :

<sup>2</sup> Jacques Hadamard (1865-1963), mathématicien français, a contribué à la théorie des nombres et à l'étude des problèmes *bien posés*.

- **existence** : il existe  $o \in X$  tel que  $i = Ho$ , soit  $i \in \text{Im}(H)$  ( $i$  est dans l'espace image de l'opérateur  $H$ );
- **unicité** : le noyau de  $H$  est réduit à zéro, soit  $\text{Ker}(H) = \{0\}$ ;
- **stabilité** : l'inverse<sup>3</sup> de  $H$  sur  $\text{Im}(H)$  est continu

Cette condition de stabilité est équivalente à la condition  $\text{Im}(H)$  fermé, soit encore  $\text{Im}(H) = \overline{\text{Im}(H)}$ .

Pour nombre de problèmes inverses, même les deux premières de ces conditions ne sont pas remplies, sans parler de la troisième. En effet, d'une part  $\text{Im}(H)$  est l'ensemble des images possibles en l'absence de bruit, ensemble qui est plus petit que l'espace des données bruitées auquel appartient  $i$ . En imagerie par exemple,  $\text{Im}(H)$  est un espace vectoriel qui ne contient aucune fréquence supérieure à la fréquence de coupure optique du télescope ( $D/\lambda$ ), alors que le bruit en contient. En général, l'existence de la solution n'est donc pas assurée. D'autre part, le noyau de  $H$  contient tous les objets « non vus » par l'instrument. Ainsi, pour un analyseur de Hartmann-Shack, ce sont les modes spatiaux (par exemple décrits en termes de polynômes de Zernike) non vus comme le piston ( $Z_1$ ) ou le mode dit gaufre; pour un imageur ce sont les fréquences spatiales supérieures à la fréquence de coupure optique du télescope. Le noyau n'est donc généralement pas réduit à  $\{0\}$  et l'unicité n'est pas assurée.

Le mathématicien Nashed [9] a introduit le concept de problème bien posé au sens des moindres carrés, qui permet à la fois de répondre positivement aux questions d'existence – en pratique – et d'unicité, puis de montrer qu'une inversion faite au sens des moindres carrés ne conduit pas à une bonne solution du problème inverse à cause de son instabilité, aussi appelée *non robustesse*<sup>4</sup> au bruit. Il demeure donc mal posé.

On dit que  $\hat{o}_{\text{MC}}$  est une solution aux moindres carrés du problème 9.2 si :

$$\|i - H\hat{o}_{\text{MC}}\|^2 = \inf_o \|i - Ho\|^2, \quad (9.4)$$

où  $\|\cdot\|$  est la norme euclidienne ou  $L_2$ . Nashed montre alors ceci :

- **existence** : une solution de type moindres carrés existe si et seulement si  $i \in \text{Im}(H) + \text{Im}(H)^\perp$ . Cette condition est toujours vérifiée si  $\text{Im}(H)$  est fermé;
- **unicité** : si plusieurs solutions existent, on choisit celle, unique, de norme (d'« énergie ») minimale, c'est-à-dire qu'on projette la solution sur l'orthogonal du noyau,  $\text{Ker}(H)^\perp$ . On la note  $H^\dagger i$ . On appelle « inverse généralisé » et on note  $H^\dagger$  cet opérateur, qui à tout  $i \in \text{Im}(H) + \text{Im}(H)^\perp$  associe la solution aux moindres carrés de norme minimale de  $i = Ho$ , c'est-à-dire la seule solution dans  $\text{Ker}(H)^\perp$ .

On montre de plus que  $H^\dagger$  est continu si et seulement si  $\text{Im}(H)$  est fermé [9].

On dit que le problème 9.2 est bien posé au sens des moindres carrés si une solution aux moindres carrés (de norme minimale) existe, est unique, et est stable (*i.e.*, dépend continûment des données  $i$ ). On voit alors que le problème 9.2 est bien posé au sens des moindres carrés si et seulement si  $H^\dagger$  est continu, c'est-à-dire si et seulement si  $\text{Im}(H)$  est fermé. Et pour nombre d'opérateurs, par exemple lorsque  $H$  est la convolution par une réponse  $h$  de carré intégrable, cette condition n'est pas vérifiée [9].

Une caractérisation des solutions aux moindres carrés permet de bien comprendre intuitivement ce qu'est une telle solution : soit  $H$  un opérateur linéaire continu d'un espace de Hilbert  $X$  vers l'espace de Hilbert  $Y$ , alors les trois propositions suivantes

<sup>3</sup>L'inverse de  $H$  sur  $\text{Im}(H)$ , noté  $H^{-1}$ , est bien défini car on suppose  $\text{Ker}(H) = \{0\}$ .

<sup>4</sup>La qualification de *robustesse au bruit*, souvent employée à propos des méthodes d'inversion, signifie que celle-ci est telle qu'elle n'amplifie pas exagérément la puissance du bruit.

sont équivalentes :

- (a)  $\|i - H\hat{o}_{\text{MC}}\|^2 = \inf_{o \in X} \|i - Ho\|^2$ ;
- (b)  $H^*H\hat{o}_{\text{MC}} = H^*i$ , où  $H^*$  est l'adjoint de  $H$  (équation normale); (9.5)
- (c)  $H\hat{o}_{\text{MC}} = Pi$ , où  $P$  est l'opérateur de projection (orthogonale) de  $Y$  sur  $\overline{\text{Im}(H)}$ .

En particulier, la caractérisation (c) nous apprend que la solution aux moindres carrés résout *exactement* l'équation de départ (9.2) quand on projette les données  $i$  sur (la fermeture de) l'ensemble des données possibles en l'absence de bruit,  $\overline{\text{Im}(H)}$ .

Dans le cas de la dimension finie, c'est-à-dire pour les problèmes pratiques, *i.e.*, discrétisés, un sous-espace vectoriel est toujours fermé. Par conséquent on est assuré à la fois de l'existence et de la continuité de l'inverse généralisé. Cependant, le caractère mal posé du problème continu ne disparaît pas par discrétisation, il change simplement d'aspect : l'instabilité mathématique du problème continu, traduite par la non-continuité de l'inverse généralisé en dimension infinie, se traduit en une instabilité numérique du problème discrétisé : le problème inverse discrétisé, en dimension finie, est *mal conditionné*, comme nous le détaillons plus loin. Le conditionnement d'un problème inverse discrétisé caractérise la robustesse au bruit lors de l'inversion ; il est lié à la dynamique des valeurs propres de  $H^*H$  (matrice en dimension finie), et est d'autant plus mauvais que cette dynamique est grande. Nous renvoyons le lecteur à [3, chap. 1] pour un traitement bien plus complet.

### 9.6.3 Les méthodes d'inversion classiques

Dans ce qui suit, nous considérons que les données, qui ont été numérisées, sont discrètes, finies et rassemblées dans un vecteur  $i$ . En imagerie, pour une image de taille  $M \times M$ ,  $i$  est un vecteur de taille  $M^2$  qui concatène les lignes ou les colonnes de l'image. Notons que cette concaténation n'est pas à faire dans le calcul, nous l'utilisons ici d'un point de vue conceptuel. La première étape de la résolution du problème inverse consiste à discrétiser également l'objet recherché  $o$  en le décomposant sur une base finie (base de pixels ou de sinus cardinaux pour une image, base de polynômes de Zernike pour une phase aberrante). Le modèle qui relie  $o$  à  $i$  est donc une approximation du modèle direct continu des équations (9.1) ou (9.2). En y incorporant explicitement les incertitudes de mesures sous forme d'un bruit additif  $b$  (vecteur composé de variables aléatoires) il s'écrit :

$$i = H(o) + b \quad (9.6)$$

dans le cas général et

$$i = \mathbf{H}o + b \quad (9.7)$$

dans le cas linéaire,  $\mathbf{H}$  étant une matrice. Dans le cas particulier où  $\mathbf{H}$  représente une convolution discrète, le modèle direct s'écrit :

$$i = h \star o + b, \quad (9.8)$$

où  $h$  est la réponse impulsionnelle du système et  $\star$  désigne la convolution discrète.

Notons que, dans le cas d'un bruit de photons, le bruit n'est pas additif, au sens où il dépend des mesures non bruitées  $\mathbf{H}o$  : la notation de l'équation (9.6) est alors quelque peu abusive<sup>5</sup>.

<sup>5</sup>Certains auteurs, par ex. [3, chap. 3] ou [10], notent  $i = \mathbf{H}o \diamond b$  le modèle direct bruité pour signifier une opération de bruitage qui peut dépendre, en chaque valeur de la donnée, de la valeur de la donnée non bruitée.

### 9.6.3.1 Méthode des moindres carrés

La méthode la plus universellement utilisée pour estimer des paramètres  $\mathbf{o}$  à partir de données  $\mathbf{i}$  est la méthode des moindres carrés. Cette méthode consiste à rechercher  $\hat{\mathbf{o}}_{\text{MC}}$  qui minimise l'écart, en moyenne quadratique, entre les données  $\mathbf{i}$  et le modèle de données  $H(\mathbf{o})$  :

$$\hat{\mathbf{o}}_{\text{MC}} = \arg \min_{\mathbf{o}} \|\mathbf{i} - H(\mathbf{o})\|^2, \quad (9.9)$$

où  $\arg \min$  signifie argument du minimum et  $\|\cdot\|$  désigne la norme euclidienne. Cette méthode a été publiée pour la première fois par Legendre<sup>6</sup> en 1805, et vraisemblablement découverte par Gauss<sup>7</sup> quelques années plus tôt mais sans être publiée – voir [11] pour une analyse détaillée de cette fameuse querelle. Legendre utilisa la méthode des moindres carrés pour estimer l'ellipticité de la Terre à partir de mesures d'arc, dans le but de définir le mètre. Dans le cas où le modèle des mesures est linéaire et donné par (9.7), la solution est analytique et s'obtient en annulant le gradient du critère (9.9) :

$$\mathbf{H}^t \mathbf{H} \hat{\mathbf{o}}_{\text{MC}} = \mathbf{H}^t \mathbf{i}. \quad (9.10)$$

Si  $\mathbf{H}^t \mathbf{H}$  est inversible – *i.e.*, si le rang de la matrice  $\mathbf{H}$  est égal à la dimension du vecteur inconnu  $\mathbf{o}$  – alors la solution est unique et s'exprime comme :

$$\hat{\mathbf{o}}_{\text{MC}} = (\mathbf{H}^t \mathbf{H})^{-1} \mathbf{H}^t \mathbf{i}. \quad (9.11)$$

Sinon, comme en dimension infinie (*cf* 9.6.2), il existe une infinité de solutions mais une seule de ces solutions est de norme minimale. C'est l'*inverse généralisé*, que l'on note  $\mathbf{H}^\dagger \mathbf{i}$ .

### 9.6.3.2 Relation entre moindres carrés et filtre inverse

Quand le processus de formation d'image peut être modélisé par une convolution, le caractère invariant par translation de l'imagerie se traduit par une structure particulière de la matrice  $\mathbf{H}$  ; cette structure est approximativement celle d'une matrice circulante (pour une convolution à une dimension), ou circulante par blocs, à blocs circulants (pour une convolution à deux dimensions). Sous cette approximation, qui revient à rendre périodique la réponse impulsionnelle  $\mathbf{h}$ , la matrice  $\mathbf{H}$  se diagonalise par transformée de Fourier discrète (TFD, calculable par un algorithme de FFT), et ses valeurs propres sont les valeurs de la fonction de transfert  $\tilde{\mathbf{h}}$  (définie comme la TFD de  $\mathbf{h}$ ). On peut alors écrire la solution aux moindres carrés de norme minimale du paragraphe précédent dans le domaine de Fourier discret<sup>8</sup> :

$$\tilde{\mathbf{o}}_{\text{MC}}(\mathbf{u}) = \frac{\tilde{\mathbf{h}}^*(\mathbf{u}) \tilde{\mathbf{i}}(\mathbf{u})}{|\tilde{\mathbf{h}}(\mathbf{u})|^2} = \frac{\tilde{\mathbf{i}}}{\tilde{\mathbf{h}}}(\mathbf{u}) \quad \forall \mathbf{u} \text{ tel que } \tilde{\mathbf{h}}(\mathbf{u}) \neq 0, \text{ et } 0 \text{ si } \tilde{\mathbf{h}}(\mathbf{u}) = 0. \quad (9.12)$$

où  $\tilde{\cdot}$  représente la transformée de Fourier discrète. Dans le cas d'un modèle de données convolutif, la solution aux moindres carrés est donc identique au *filtre inverse* (à l'approximation sus-mentionnée près).

où  $\tilde{\cdot}$  représente la transformée de Fourier discrète. Dans le cas d'un modèle de données convolutif, la solution aux moindres carrés est donc identique au *filtre inverse*, à l'approximation sus-mentionnée près.

<sup>6</sup> Adrien-Marie Legendre (1752-1833), mathématicien français, dont les contributions furent importantes en statistique, en algèbre et en analyse.

<sup>7</sup> Carl Gauss (1777-1855), appelé le *prince des mathématiciens*, astronome et physicien allemand, dont les contributions furent aussi diverses que profondes dans ces domaines.

<sup>8</sup> À la matrice  $\mathbf{H}^t$  correspond dans le domaine de Fourier discret la fonction de transfert  $\tilde{\mathbf{h}}^*$  et les inversions matricielles donnent des inverses simples.

### 9.6.3.3 Méthode du maximum de vraisemblance

Dans la méthode des moindres carrés, le choix d'une mesure quadratique de l'écart entre données  $\mathbf{i}$  et modèle de données  $H(\mathbf{o})$  n'est pas justifié, si ce n'est par le fait que l'on peut mener analytiquement le calcul de la solution. En outre, cette méthode n'utilise aucunement notre éventuelle connaissance des propriétés statistiques du bruit. Or cette information sur le bruit permet d'interpréter la méthode des moindres carrés et surtout de l'étendre.

Modélisons les incertitudes de mesure  $\mathbf{b}$  comme un bruit de loi de probabilité  $p_b(\mathbf{b})$ . D'après l'équation (9.6), la loi des données  $\mathbf{i}$  conditionnellement à l'objet, c'est-à-dire à objet  $\mathbf{o}$  donné (supposé connu), est alors<sup>9</sup> :

$$p(\mathbf{i} | \mathbf{o}) = p_b(\mathbf{i} - H(\mathbf{o})) \quad (9.13)$$

L'équation (9.13) permet de tirer des réalisations de données bruitées connaissant l'objet, c'est-à-dire de simuler des données. Dans un problème inverse au contraire, on n'a qu'une réalisation des données (celles effectivement mesurées) et on cherche à estimer l'objet. La méthode du maximum de vraisemblance (MV) consiste justement à renverser le point de vue sur  $p(\mathbf{i} | \mathbf{o})$  en considérant  $\mathbf{o}$  comme une variable,  $\mathbf{i}$  fixe égal aux données, et à rechercher l'objet  $\mathbf{o}$  qui maximise  $p(\mathbf{i} | \mathbf{o})$ . La quantité  $p(\mathbf{i} | \mathbf{o})$  vue comme une fonction de  $\mathbf{o}$  est alors appelée la *vraisemblance* des données, et l'objet  $\hat{\mathbf{o}}_{\text{MV}}$  qui la maximise est celui qui rend les données effectivement observées les plus probables<sup>10</sup> :

$$\hat{\mathbf{o}}_{\text{MV}} = \arg \max_{\mathbf{o}} p(\mathbf{i} | \mathbf{o}). \quad (9.14)$$

Le modèle de bruit le plus utilisé est sans conteste le modèle gaussien centré (c'est-à-dire de moyenne nulle) caractérisé par sa matrice de covariance  $\mathbf{C}_b$  :

$$p(\mathbf{i} | \mathbf{o}) \propto \exp \left\{ -\frac{1}{2} [\mathbf{i} - H(\mathbf{o})]^t \mathbf{C}_b^{-1} [\mathbf{i} - H(\mathbf{o})] \right\}. \quad (9.15)$$

Le bruit est dit blanc si sa matrice de covariance est diagonale. Si de plus cette matrice est proportionnelle à l'identité alors le bruit est dit blanc stationnaire ou homogène. Le bruit de lecture d'un détecteur CCD est souvent modélisé par un tel bruit gaussien centré stationnaire. Le bruit de photons est blanc mais suit, lui, une loi de Poisson, dont on peut montrer qu'à fort flux elle tend vers une loi gaussienne non stationnaire de variance égale au signal détecté en chaque pixel.

La maximisation de la vraisemblance est évidemment équivalente à la minimisation d'un critère défini comme l'opposé du logarithme de celle-ci et appelé anti-log-vraisemblance :

$$J_i(\mathbf{o}) = -\ln p(\mathbf{i} | \mathbf{o}). \quad (9.16)$$

Dans le cas d'un bruit gaussien, l'anti-log-vraisemblance s'écrit :

$$J_i(\mathbf{o}) = \frac{1}{2} [\mathbf{i} - H(\mathbf{o})]^t \mathbf{C}_b^{-1} [\mathbf{i} - H(\mathbf{o})]. \quad (9.17)$$

Si en outre le bruit est blanc, alors :

$$J_i(\mathbf{o}) = \frac{1}{2} \sum_k \frac{|\mathbf{i}(k) - H(\mathbf{o})(k)|^2}{\sigma_b^2(k)}, \quad (9.18)$$

<sup>9</sup>Cette équation se lit : « la probabilité de  $\mathbf{i}$  sachant  $\mathbf{o}$  est la loi du bruit en  $\mathbf{i} - H(\mathbf{o})$ . »

<sup>10</sup>La notation  $\arg \max$  se lit « argument du maximum » et désigne la valeur de la variable qui maximise la quantité qui suit (ici  $p(\mathbf{i} | \mathbf{o})$ ).

où les  $\sigma_b^2(k)$  sont les éléments de la diagonale de  $\mathbf{C}_b$ .  $J_i(\mathbf{o})$  est un critère dit de *moindres carrés pondérés*. Si le bruit est de plus stationnaire de variance  $\sigma_b^2$  alors  $J_i(\mathbf{o}) = 1/(2\sigma_b^2)\|\mathbf{i} - H(\mathbf{o})\|^2$  est exactement le critère de moindres carrés, dits *ordinaires* par opposition à pondérés, du paragraphe précédent.

La méthode des moindres carrés peut donc s'interpréter comme un maximum de vraisemblance si le bruit est blanc gaussien stationnaire. Inversement, si la loi du bruit est connue mais est différente, le maximum de vraisemblance permet de prendre en compte cette connaissance sur le bruit et généralise alors les moindres carrés.

#### 9.6.3.4 Exemple : reconstruction de front d'onde par maximum de vraisemblance

Considérons un analyseur de Hartmann-Shack qui mesure les aberrations dues à la turbulence atmosphérique sur un télescope observant l'Espace depuis le sol. On décompose la phase dans la pupille sur une base (nécessairement tronquée en pratique) de polynômes de Zernike  $Z_k$  :

$$\varphi(x, y) = \sum_{k=1}^{k_{\max}} o_k Z_k(x, y) \quad (9.19)$$

On recherche donc  $\mathbf{o}$ , la collection des coefficients  $o_k$  de cette décomposition, à partir de la mesure des pentes moyennes, concaténées dans un vecteur de mesures  $\mathbf{i}$ . Le modèle de données est identique à l'équation (9.7) :  $\mathbf{i} = \mathbf{H}\mathbf{o} + \mathbf{b}$ , où  $\mathbf{H}$  est essentiellement un opérateur de dérivation appelé matrice d'interaction. Dans la simulation présentée ci-après, l'analyseur de Hartmann-Shack contient  $20 \times 20$  sous-pupilles dont 276 reçoivent de la lumière, du fait d'une occultation centrale (miroir secondaire du télescope) de 33%, ce qui donne 552 mesures de pentes. La phase turbulente vraie  $\varphi$  est une combinaison linéaire de  $k_{\max} = 861$  polynômes de Zernike tirée selon une statistique de Kolmogorov. La matrice  $\mathbf{H}$  est donc de taille  $552 \times 861$ . Le bruit ajouté sur les pentes est blanc gaussien stationnaire et de variance égale à la variance des pentes non bruitées, soit un rapport signal à bruit de 1 sur chaque mesure. Dans ces conditions, l'estimée de la phase, au sens du maximum de vraisemblance, est identique à la solution au sens des moindres carrés (cf paragraphe précédent). La matrice  $\mathbf{H}^t\mathbf{H}$  de l'équation (9.11) est de taille  $k_{\max} \times k_{\max}$  et n'est pas inversible (car on n'a que 552 mesures). La solution inverse généralisée est inexploitable car complètement dominée par le bruit. Le remède souvent retenu est de réduire, lors de la reconstruction de front d'onde, la dimension  $k_{\max}^{\text{rec}}$  de l'espace des inconnues  $\mathbf{o}$ . Ce remède est une des méthodes de *régularisation* connues et est appelé *régularisation par contrôle de la dimension* (cf [2] ou [3, chap. 2] par exemple). Un exemple de reconstruction pour différentes valeurs de  $k_{\max}^{\text{rec}}$  est présenté Fig. 9.1.

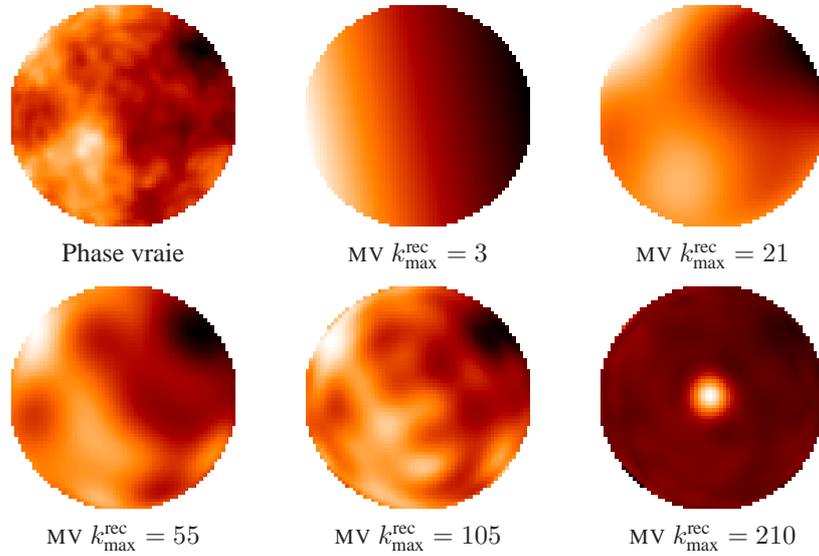


FIG. 9.1 – Reconstruction de front d'onde. Phase simulée (appelée phase vraie, en haut à gauche) et phase reconstruite par maximum de vraisemblance pour différents nombres de modes de Zernike considérés.

Pour  $k_{\max}^{\text{rec}} = 210$ , valeur bien inférieure au nombre de mesures, la phase reconstruite est déjà inacceptable. Sa forme particulière est due au fait que le télescope a une obstruction centrale et que l'on n'a donc pas de mesures au centre de la pupille. Cet exemple illustre bien les cas où la matrice  $\mathbf{H}^t \mathbf{H}$ , bien qu'inversible, est mal conditionnée, c'est-à-dire a des valeurs propres très proches de 0, ce qui conduit à une amplification non contrôlée du bruit. La troncature de l'espace des solutions à des valeurs plus faibles de  $k_{\max}^{\text{rec}}$  produit des solutions plus raisonnables, mais d'une part le choix optimal de  $k_{\max}^{\text{rec}}$  dépend de la force de la turbulence et du niveau de bruit et ce réglage est difficile à faire en pratique, d'autre part cette troncature réalise une erreur de modélisation puisqu'elle revient à négliger toutes les composantes de la phase turbulente au delà de  $k_{\max}^{\text{rec}}$ . Nous voyons plus loin comment l'introduction de connaissances *a priori* sur la régularité spatiale de la phase (spectre de Kolmogorov) permet d'obtenir une solution meilleure sans cette réduction *ad hoc* de l'espace des inconnues.

### 9.6.3.5 Interprétation de l'échec des méthodes non régularisées

L'échec de la méthode du maximum de vraisemblance, illustrée par l'exemple précédent, peut surprendre au vu des bonnes propriétés statistiques du maximum de vraisemblance, qui est un estimateur convergent vers la vraie valeur des paramètres lorsque le nombre de données tend vers l'infini, qui est asymptotiquement efficace, etc. Mais ces bonnes propriétés ne sont qu'asymptotiques, c'est-à-dire qu'elles ne concernent que les situations où l'on a un bon *contraste statistique*, défini simplement comme le rapport du nombre de mesures au nombre d'inconnues.

Dans la pratique et en particulier dans l'exemple précédent, on a souvent à estimer un nombre de paramètres qui est du même ordre de grandeur que le nombre de mesures, parfois plus grand, auquel cas ces propriétés asymptotiques ne sont d'aucun secours. Dans cette situation, fréquente, d'un contraste statistique défavorable, l'inversion est

instable, c'est-à-dire très sensible au bruit, ce qui peut souvent s'interpréter comme provenant du caractère mal posé du problème en dimension infinie sous-jacent.

La restauration d'image est un autre cas typique de cette situation : pour une image de  $M \times M$  pixels, on cherche évidemment à reconstruire un objet de même taille, et si l'on augmente la taille de l'image on augmente de la même manière le nombre de paramètres inconnus liés à l'objet, sans améliorer le contraste statistique, qui reste de l'ordre de 1. On a vu que la solution aux moindres carrés était dans ce cas donnée par le filtre inverse, dont on sait justement qu'il est très instable vis-à-vis du bruit. Cette instabilité se comprend aisément en réinjectant l'équation de mesure (9.8), dans le domaine de Fourier discret, dans la solution (9.12) :

$$\tilde{\mathbf{o}}_{\text{MC}} = \frac{\tilde{\mathbf{i}}}{\tilde{\mathbf{h}}} = \tilde{\mathbf{o}} + \frac{\tilde{\mathbf{b}}}{\tilde{\mathbf{h}}} . \quad (9.20)$$

Il est clair d'après cette équation que le bruit est très amplifié pour toutes les fréquences pour lesquelles la valeur de la fonction de transfert  $\tilde{\mathbf{h}}$  est proche de zéro. On peut penser, pour réduire l'amplification du bruit, à modifier (9.12) et ne réaliser la division de  $\tilde{\mathbf{i}}$  par  $\tilde{\mathbf{h}}$  que pour les fréquences  $\mathbf{u}$  où la fonction de transfert  $\tilde{\mathbf{h}}$  n'est pas « trop petite ». Ceci constitue également une régularisation par contrôle de la dimension de la solution, très similaire au choix d'un  $k_{\text{max}}^{\text{rec}}$  pour la reconstruction de front d'onde, et souffre du même caractère *ad hoc* ainsi que de la même difficulté de réglage.

En résumé, les méthodes d'inversion simples telles que les moindres carrés ou le maximum de vraisemblance ne donnent des résultats satisfaisants que si l'on a un bon contraste statistique. Ainsi, le maximum de vraisemblance peut s'appliquer avec succès à des problèmes tels que le recalage d'image [12], où l'on recherche un vecteur à deux dimensions, afin de recaler entre elles deux images constituées d'un grand nombre de pixels. Plus généralement la recherche des quelques variables d'un modèle paramétrique parcimonieux à partir de données en assez grand nombre : estimation du diamètre d'une étoile à partir de visibilités en interférométrie optique, estimation de l'ellipticité de la Terre à partir de mesures d'arc, etc.

Dans nombre de problèmes où le contraste statistique n'est pas favorable, le problème est mal conditionné et la régularisation c'est-à-dire l'ajout, lors de l'inversion, de contraintes et de connaissances *a priori* sur la solution, est très profitable, comme nous le montrons maintenant.

## 9.6.4 Inversion régularisée

La régularisation d'un problème inverse correspond à l'idée que les données seules ne permettent pas d'obtenir une solution acceptable et qu'il faut donc introduire une information *a priori* sur la régularité de l'objet à estimer. Nous entendons ici par le terme *régularité* le fait que l'objet, pour des raisons physiques tenant à sa nature même, doit posséder certaines propriétés, ou encore obéir à certaines règles (de signe, de taille, de fréquences par exemple). La solution résulte alors d'un compromis entre l'exigence de fidélité aux données et celle de la régularité postulée de l'objet.

En effet, plusieurs solutions très différentes, certaines très mauvaises et d'autres très bonnes, peuvent être compatibles avec les données ; ainsi dans l'exemple précédent de reconstruction d'un front d'onde, la phase vraie et la phase reconstruite pour  $k_{\text{max}}^{\text{rec}} = 210$  (Fig. 9.1) donnent des valeurs de vraisemblance très voisines, c'est-à-dire sont toutes les deux fidèles aux données. De plus, la solution plus « douce », plus régulière, obtenue pour  $k_{\text{max}}^{\text{rec}} = 55$  est moins bien ajustée, c'est-à-dire est moins fidèle aux données que celle pour  $k_{\text{max}}^{\text{rec}} = 210$  – car elle a été obtenue par une optimisation sur moins de degrés de liberté – et pourtant elle est bien plus proche de la phase vraie.

### 9.6.4.1 Estimation bayésienne et estimateur du maximum *a posteriori*

L'estimation bayésienne, que nous présentons brièvement ici, permet de combiner naturellement l'information apportée par les mesures et l'information disponible *a priori*. Supposons avoir exprimé notre connaissance *a priori* de l'objet observé par une distribution de probabilité, dite loi *a priori*  $p(\mathbf{o})$ . Il ne s'agit pas de supposer que l'objet observé est *réellement* la réalisation d'un phénomène aléatoire de loi  $p(\mathbf{o})$  ; cette loi doit simplement être représentative de nos informations *a priori*, c'est-à-dire prendre des valeurs faibles pour les objets reconstruits qui seraient peu compatibles avec celles-ci, et des valeurs élevées pour les objets très compatibles.

La règle de Bayes<sup>11</sup> permet d'exprimer la probabilité de l'objet  $\mathbf{o}$ , conditionnellement aux mesures  $\mathbf{i}$ , dite probabilité *a posteriori* – c'est-à-dire probabilité de l'objet après avoir fait les mesures – en fonction de la probabilité *a priori* de l'objet  $p(\mathbf{o})$  et de la probabilité des mesures conditionnellement à l'objet,  $p(\mathbf{i} | \mathbf{o})$ . Cette dernière probabilité contient nos connaissances sur le modèle de formation de données, y compris le modèle de bruit. La règle de Bayes s'exprime ainsi :

$$p(\mathbf{o} | \mathbf{i}) = \frac{p(\mathbf{i} | \mathbf{o}) \times p(\mathbf{o})}{p(\mathbf{i})} \propto p(\mathbf{i} | \mathbf{o}) \times p(\mathbf{o}). \quad (9.21)$$

Muni de l'expression précédente, comment doit-on choisir l'objet reconstruit  $\hat{\mathbf{o}}$  qui estimerait au mieux le véritable objet ? Un choix couramment fait est l'estimateur du maximum *a posteriori* ou MAP. Il consiste à définir comme solution l'objet qui maximise la probabilité *a posteriori*  $p(\mathbf{o} | \mathbf{i})$  :

$$\hat{\mathbf{o}}_{\text{MAP}} = \arg \max_{\mathbf{o}} p(\mathbf{o} | \mathbf{i}) = \arg \max_{\mathbf{o}} p(\mathbf{i} | \mathbf{o}) \times p(\mathbf{o}). \quad (9.22)$$

C'est l'objet le plus probable, sachant les données et nos connaissances *a priori*. Cette probabilité *a posteriori*, vue comme une fonction de l'objet, est appelée *vraisemblance a posteriori*. Elle prend en compte, *via* l'équation (9.21), à la fois les mesures  $\mathbf{i}$ , le modèle de formation de données et les connaissances *a priori*. En particulier, il est remarquable qu'avec un *a priori* uniforme ( $p(\mathbf{o}) = \text{constante}$ ), c'est-à-dire une absence d'information sur l'objet, le MAP redonne la méthode du maximum de vraisemblance.

On peut montrer que le choix de l'estimateur MAP correspond à minimiser le risque moyen pour une fonction de coût particulière, dite en tout-ou-rien, mais ceci dépasse le cadre de cette introduction aux problèmes inverses. D'autres choix de fonction de coût, tels que le choix d'un  $\hat{\mathbf{o}}$  minimisant l'erreur quadratique moyenne à l'objet vrai  $\mathbf{o}$  sous la loi *a posteriori*, peuvent être envisagés mais conduisent à des coûts plus importants en temps de calcul de la solution. Le lecteur intéressé pourra consulter un ouvrage de référence sur la théorie de l'estimation tel que [13].

### 9.6.4.2 Équivalence avec la minimisation d'un critère régularisé. Lien avec les moindres carrés

La maximisation de la vraisemblance *a posteriori* est équivalente à la minimisation d'un critère  $J_{\text{MAP}}(\mathbf{o})$  défini comme l'opposé du logarithme de celle-ci. D'après (9.22) ce critère peut s'écrire comme la somme de deux composantes :

$$J_{\text{MAP}}(\mathbf{o}) = -\ln p(\mathbf{i} | \mathbf{o}) - \ln p(\mathbf{o}) = J_i(\mathbf{o}) + J_o(\mathbf{o}), \quad (9.23)$$

<sup>11</sup> Thomas Bayes (1762-1761), mathématicien britannique dont le principal travail, portant sur les probabilités inverses, ne fut publié que de façon posthume.

où  $J_i$  est le critère de fidélité aux données déduit de la vraisemblance (cf Équ. (9.16)) qui est souvent un critère de moindres carrés et où  $J_o(\mathbf{o}) \triangleq -\ln p(\mathbf{o})$  est un critère dit de *régularisation* ou de *pénalisation* (de la vraisemblance) qui traduit la fidélité à l'*a priori*.

L'écriture de la solution MAP comme l'objet qui minimise le critère (9.23) fait clairement apparaître qu'elle réalise le compromis entre fidélité aux données et fidélité à l'*a priori* annoncé au début de la section 9.6.4.

Lorsque  $\mathbf{o}$  n'est pas la réalisation d'un phénomène aléatoire, de loi  $p(\mathbf{o})$  (par exemple en restauration d'image),  $J_o(\mathbf{o})$  comprend généralement un facteur multiplicatif appelé coefficient de régularisation ou hyper-paramètre<sup>12</sup>, dont la valeur règle le point de compromis. Des méthodes d'ajustement non supervisé (c'est-à-dire automatique) de ce coefficient existent mais dépassent le cadre de cet ouvrage. Le lecteur intéressé pourra consulter les chapitres 2 et 3 de [3] pour une synthèse sur cette question.

### 9.6.4.3 Le cas linéaire et gaussien. Lien avec le filtre de Wiener

Nous considérons ici le cas où le modèle de données est linéaire (Équ. (9.7)), le bruit est supposé gaussien et l'*a priori* adopté pour l'objet est également gaussien<sup>13</sup>, de moyenne  $\bar{\mathbf{o}}$  et de matrice de covariance  $\mathbf{C}_o$  :

$$p(\mathbf{o}) \propto \exp \left\{ -\frac{1}{2}(\mathbf{o} - \bar{\mathbf{o}})^t \mathbf{C}_o^{-1}(\mathbf{o} - \bar{\mathbf{o}}) \right\}. \quad (9.24)$$

En utilisant l'équation (9.17), on voit que le critère à minimiser pour obtenir la solution MAP est :

$$J_{\text{MAP}}(\mathbf{o}) = \frac{1}{2}[\mathbf{i} - \mathbf{H}\mathbf{o}]^t \mathbf{C}_b^{-1}[\mathbf{i} - \mathbf{H}\mathbf{o}] + \frac{1}{2}(\mathbf{o} - \bar{\mathbf{o}})^t \mathbf{C}_o^{-1}(\mathbf{o} - \bar{\mathbf{o}}). \quad (9.25)$$

Ce critère est quadratique en  $\mathbf{o}$  et son minimum est donc analytique ; celui-ci s'obtient en annulant le gradient du critère :

$$\hat{\mathbf{o}}_{\text{MAP}} = (\mathbf{H}^t \mathbf{C}_b^{-1} \mathbf{H} + \mathbf{C}_o^{-1})^{-1} (\mathbf{H}^t \mathbf{C}_b^{-1} \mathbf{i} + \mathbf{C}_o^{-1} \bar{\mathbf{o}}). \quad (9.26)$$

Quelques commentaires permettent d'éclairer ce résultat d'apparence un brin rébarbative. Tout d'abord, la solution maximum de vraisemblance s'obtient en prenant  $\mathbf{C}_o^{-1} = 0$  dans cette équation. Incidemment, ceci montre que la solution MV correspond, dans le cadre bayésien, à supposer une énergie infinie pour  $\mathbf{o}$ . Ensuite, en prenant de plus  $\mathbf{C}_b$  proportionnelle à la matrice identité, on voit que l'on retrouve exactement la solution moindres carrés de l'équation (9.11). Enfin, le cas de la déconvolution est particulièrement éclairant : on suppose que le bruit est stationnaire de densité spectrale de puissance (DSP)  $S_b$ , que la loi *a priori* sur l'objet est également stationnaire avec une DSP  $S_o$ , et on fait pour toutes les grandeurs en jeu la même approximation de périodicité qu'au paragraphe 9.6.3.2. Toutes les matrices de l'équation (9.26) se diagonalisent alors dans la même base de Fourier discrète, et la solution peut alors s'écrire dans cette base, avec des multiplications et des inverses simples et non plus matricielles :

$$\tilde{\mathbf{o}}_{\text{MAP}}(\mathbf{u}) = \frac{\tilde{\mathbf{h}}^*(\mathbf{u})}{|\tilde{\mathbf{h}}|^2(\mathbf{u}) + S_b/S_o(\mathbf{u})} \tilde{\mathbf{i}}(\mathbf{u}) + \frac{S_b/S_o(\mathbf{u})}{|\tilde{\mathbf{h}}|^2(\mathbf{u}) + S_b/S_o(\mathbf{u})} \tilde{\mathbf{o}}(\mathbf{u}). \quad (9.27)$$

<sup>12</sup> On désigne par *hyper-paramètre* l'ensemble des paramètres du modèle des données (par ex. la variance du bruit) ou du modèle *a priori* (par exemple la turbulence atmosphérique et donc le paramètre  $r_0$ ), qui sont fixés lors de l'inversion.

<sup>13</sup> Nous ne supposons pas du tout ici que la *forme* de l'objet est gaussien, mais que sa *distribution de probabilité* est gaussienne. C'est le cas par exemple lorsque  $\mathbf{o}$  est une phase turbulente qui suit la statistique de Kolmogorov.

Dans cette expression,  $S_b/S_o(\mathbf{u})$  est l'inverse d'un rapport signal à bruit (RSB) à la fréquence spatiale  $\mathbf{u}$  et  $\tilde{\mathbf{o}}$  est la TF de l'objet *a priori*, généralement pris nul ou égal à un objet de valeur constante. Cette expression est celle du *filtre de Wiener* pour le cas où l'objet *a priori* n'est pas nul. Pour les fréquences où le rapport signal à bruit est fort, cette solution tend vers le filtre inverse ; pour les fréquences où ce rapport est faible, la solution tend vers l'objet *a priori*. On peut même voir qu'à chaque fréquence spatiale  $\mathbf{u}$ , la solution est sur un segment qui relie la solution du maximum de vraisemblance (filtre inverse) à la solution *a priori*, la position sur le segment étant donnée par le rapport signal à bruit :

$$\tilde{\mathbf{o}}_{\text{MAP}}(\mathbf{u}) = \alpha \frac{\tilde{\mathbf{i}}(\mathbf{u})}{\tilde{\mathbf{h}}} + (1 - \alpha) \tilde{\mathbf{o}}(\mathbf{u}), \quad (9.28)$$

en ayant posé :

$$\alpha = \frac{|\tilde{\mathbf{h}}|^2(\mathbf{u})}{|\tilde{\mathbf{h}}|^2(\mathbf{u}) + S_b/S_o(\mathbf{u})}. \quad (9.29)$$

Cette expression illustre bien le fait que la régularisation réalise un *compromis biais-variance* qui permet de réduire l'erreur d'estimation : l'erreur quadratique moyenne sur l'objet estimé est la racine de la somme du biais, élevé au carré, et de la variance de l'estimateur utilisé. La solution filtre inverse (obtenue pour  $S_b/S_o = 0$ ) est à biais nul mais amplifie le bruit de manière non contrôlée, c'est-à-dire a une variance non bornée. La solution (9.27) est, par rapport à la solution du maximum de vraisemblance, biaisée vers la solution *a priori*  $\tilde{\mathbf{o}}$  ; accepter ce biais permet de réduire la variance très notablement et, globalement, de réduire l'erreur quadratique moyenne sur l'objet estimé.

#### 9.6.4.4 Application à la reconstruction de front d'onde

Reprenons l'exemple de la reconstruction de front d'onde à partir de données d'un analyseur Hartmann-Shack du paragraphe 9.6.3.4. Le bruit est toujours supposé gaussien, blanc, de matrice de covariance  $C_b = \sigma^2 I$ . On suppose que la phase suit une statistique de Kolmogorov ; elle est donc gaussienne, de matrice de covariance  $C_o$  connue et ne dépendant que du diamètre de Fried  $r_0$  qui quantifie la force de la turbulence. La phase vraie possède une variance spatiale  $\sigma_\varphi^2 = 3,0 \text{ rd}^2$ . La phase la plus probable, connaissant les mesures et cette information *a priori* est la solution MAP donnée par l'équation (9.26). Cette solution est présentée sur la figure 9.2 ; l'erreur d'estimation correspond à une variance spatiale  $\sigma_{err}^2 = 0,7 \text{ rd}^2$ , qui est plus faible que celle de la meilleure des solutions obtenues précédemment en tronquant la représentation de la phase à un petit nombre de polynômes de Zernike ( $\sigma_{err}^2 = 0,8 \text{ rd}^2$ , obtenue pour  $k_{\text{max}}^{\text{rec}} = 55$ ). De plus, la solution MAP ne nécessite pas l'ajustement, délicat en pratique, de la valeur de  $k_{\text{max}}^{\text{rec}}$ .

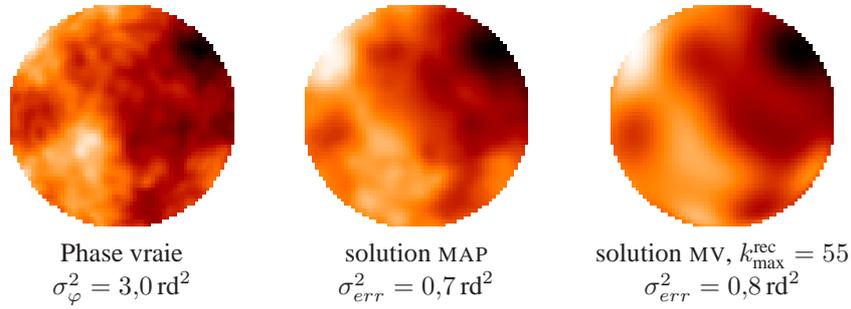


FIG. 9.2 – Phase turbulente à estimer (gauche). Phase estimée par MAP (centre). Meilleure phase estimée par MV sur une base tronquée (droite).

La solution MAP tire parti des connaissances *a priori* sur la statistique spatiale de la turbulence. Pour une utilisation en optique adaptative (OA), où la fréquence d'échantillonnage est généralement bien supérieure à  $1/\tau_0$ , il est judicieux d'opter pour une extension naturelle de cet estimateur qui utilise également les connaissances *a priori* sur la statistique temporelle de la turbulence ; cette extension est l'estimateur optimal du filtrage de Kalman [14].

### 9.6.5 Une application à l'imagerie par optique adaptative

Nous appliquons ici ces outils de résolution du problème inverse à un cas particulier, celui de l'imagerie par optique adaptative, dont nous présentons quelques illustrations.

#### 9.6.5.1 Ingrédients de la déconvolution

Les images longue pose corrigées par OA doivent être déconvoluées, car la correction réalisée n'est que partielle. En considérant la réponse impulsionnelle  $h$  connue, l'objet estimé au sens du MAP, noté  $\hat{o}_{\text{MAP}}$ , est donné par l'équation (9.22) c'est-à-dire est celui qui minimise le critère (9.23). Voyons comment définir les deux composantes de ce critère.

Afin de pouvoir restaurer des objets à grande dynamique, fréquents en astronomie, le terme de fidélité aux données  $J_i$  doit incorporer une modélisation fine du bruit prenant en compte à la fois le bruit de photons et le bruit électronique. Ceci peut être réalisé en approximant le bruit de photons comme un bruit gaussien non stationnaire et aboutit à un critère  $J_i$  du type moindres carrés pondérés (équation (9.18)) plutôt qu'à un terme de moindres carrés ordinaires (cf [15] par ex.).

Pour des objets à bords francs comme des satellites artificiels, des astéroïdes ou des planètes, un *a priori* gaussien (comme celui qui aboutit au filtre de Wiener, cf paragraphe 9.6.4.3), ou de manière équivalente un critère de régularisation quadratique, a tendance à lisser les bords et à introduire près de ceux-ci des oscillations parasites ou *ringing*. Une interprétation de cet effet est qu'une régularisation quadratique donne à l'apparition d'un saut (bord franc sur l'objet, *i.e.*, différence de valeur importante entre pixels adjacents), lors de la minimisation du critère  $J_{\text{MAP}}(o)$ , un coût proportionnel au carré de la valeur du saut. Une solution est alors d'utiliser un critère préservant les bords francs (ou *edge-preserving*) comme les critères dits quadratiques-linéaires, ou  $L_2L_1$ , qui sont quadratiques pour les faibles sauts et linéaires pour les forts sauts. La partie

quadratique assure un bon lissage du bruit et la partie linéaire annule la pénalisation des bords – voir les chapitres 6 et 10 de [3] pour plus de détails.

Par ailleurs, pour de nombreuses raisons, on est souvent amené à considérer que la réponse impulsionnelle  $\mathbf{h}$  est imparfaitement connue. Effectuer une déconvolution « classique » c'est-à-dire en supposant la réponse impulsionnelle connue mais avec une réponse impulsionnelle fautive peut conduire à des résultats catastrophiques. À l'inverse, la déconvolution dite aveugle, où l'on minimise le même critère (9.23) mais en recherchant simultanément  $\mathbf{o}$  et  $\mathbf{h}$ , est très instable, à l'instar des méthodes non régularisées. La *déconvolution myope* consiste à estimer conjointement  $\mathbf{o}$  et  $\mathbf{h}$  dans un cadre bayésien avec une régularisation naturelle pour la réponse impulsionnelle et sans avoir à régler d'hyper-paramètre supplémentaire. L'estimateur MAP conjoint est donné par :

$$\begin{aligned} (\hat{\mathbf{o}}, \hat{\mathbf{h}}) &= \arg \max_{\mathbf{o}, \mathbf{h}} p(\mathbf{o}, \mathbf{h} | \mathbf{i}) = \arg \max_{\mathbf{o}, \mathbf{h}} p(\mathbf{i} | \mathbf{o}, \mathbf{h}) \times p(\mathbf{o}) \times p(\mathbf{h}) \\ &= \arg \min_{\mathbf{o}, \mathbf{h}} (J_i(\mathbf{o}, \mathbf{h}) + J_o(\mathbf{o}) + J_h(\mathbf{h})), \end{aligned}$$

où  $J_h$  est un critère de régularisation sur  $\mathbf{h}$ , qui introduit des contraintes sur la variabilité possible de la réponse impulsionnelle (voir [16] pour plus de détails).

Le paragraphe suivant présente des résultats sur données expérimentales obtenus avec la méthode de restauration MISTRAL décrite dans [10], qui combine les trois ingrédients évoqués ci-dessus : la modélisation fine du bruit, la régularisation non quadratique et l'aspect myope.

### 9.6.5.2 Restauration à partir de données astronomiques expérimentales

L'image de la figure 9.3a montre une longue pose corrigée par OA de Ganymède, satellite de Jupiter. Cette image a été enregistrée le 28/09/1997 sur le banc d'OA de l'ONERA installé sur le télescope de 1,52 m de l'Observatoire de Haute-Provence. La longueur d'onde d'imagerie est  $\lambda = 0,85 \mu\text{m}$  et le temps de pose 100 sec. Le flux total estimé est  $8 \times 10^7$  photons et le rapport  $D/r_0$  estimé est 23. Le champ total est de 7,9 arcsec, dont seulement la moitié est montrée ici. La réponse impulsionnelle moyenne et sa variabilité ont été estimées à partir de l'enregistrement de cinquante images d'une étoile brillante située à proximité. Les figures 9.3b et c montrent les restaurations obtenues par l'algorithme de Richardson-Lucy (MV pour un bruit de Poisson), interrompu à 200 et 3 000 itérations respectivement<sup>14</sup>. Dans le premier cas, similaire à une restauration avec régularisation quadratique, l'image restaurée est assez floue et présente un léger « ringing », et dans le second cas, très similaire au résultat d'un filtrage inverse, le bruit domine la restauration.

<sup>14</sup>L'arrêt d'un algorithme non régularisé avant convergence est une méthode de régularisation encore répandue mais très *ad hoc*, cf [3, chap. 2].

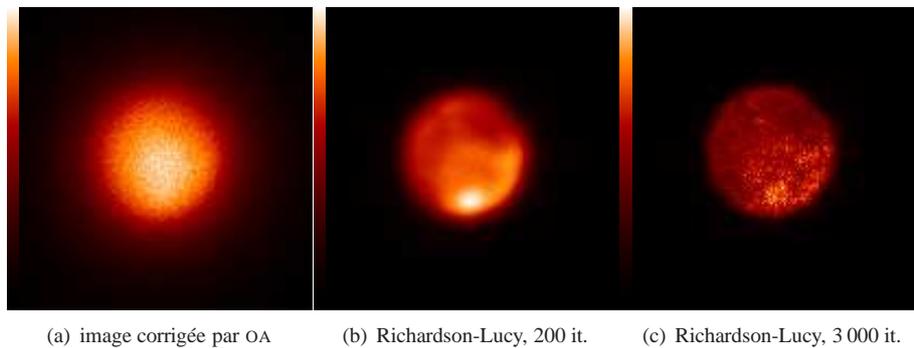


FIG. 9.3 – (a) Image corrigée de Ganymède, obtenue avec le banc d'OA de l'ONERA, le 28 septembre 1997. (b) Restauration par Richardson-Lucy interrompue à 200 itérations ; (c) idem à 3 000 itérations.

L'image figure 9.4a illustre la déconvolution myope [10] avec *a priori* préservant les bords. La figure figure 9.4b montre une image synthétique large bande obtenue à partir de clichés d'une sonde spatiale NASA/JPL (voir <http://space.jpl.nasa.gov/>) passée à proximité de Ganymède. La comparaison montre que de nombreuses caractéristiques de Ganymède sont correctement restaurées. Une comparaison plus équitable consiste à examiner conjointement la déconvolution myope effectuée par MISTRAL avec l'image de la figure 9.4b convoluée par la réponse impulsionnelle parfaite d'un télescope de 1,52 m, présentée figure 9.4c.

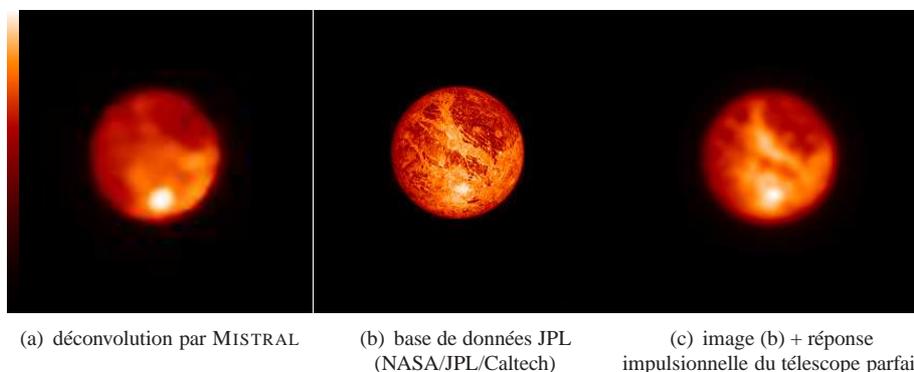


FIG. 9.4 – (a) Déconvolution par MISTRAL de l'image de Ganymède de la figure 9.3. (b) En comparaison, une image synthétique large bande obtenue grâce à la base de données NASA/JPL. (c) Même image synthétique convoluée par la réponse impulsionnelle parfaite d'un télescope de 1,52 m de diamètre.

### 9.6.6 Une application à l'interférométrie coronographique

Nous développons un second exemple de problème d'inversion, relatif ici à la détection des planètes extrasolaires à l'aide d'un interféromètre à frange noire, que l'on peut aussi appeler interféromètre coronographique. Avec l'instrument Darwin, à l'étude en Europe, ou *Terrestrial Planet Finder-Interferometer* (mission NASA à l'étude au cours des années 2000), les données se présenteront sous une forme très différente

de celle d'une image au sens classique du terme. Elles nécessiteront la mise en œuvre d'un processus de reconstruction pour leur exploitation. Elles consisteront, à chaque instant  $t$  de mesure, en une intensité dans chaque canal spectral  $\lambda$ . Cette intensité peut être modélisée comme l'intégrale, sur un certain domaine angulaire, de la carte de transmission instantanée de l'interféromètre  $R_{t,\lambda}(\boldsymbol{\theta})$  par la distribution d'intensité de l'objet observé  $o_\lambda(\boldsymbol{\theta})$ . Le modèle de données est donc linéaire mais foncièrement non convolutif, c'est-à-dire très différent de celui rencontré en imagerie. La carte de transmission est une simple sinusoïde dans le cas d'un interféromètre de Bracewell ; elle devient plus complexe lorsque plus de deux télescopes interfèrent simultanément. En recombinant les données astucieusement et avec des cartes de transmission asymétriques, il est possible d'éliminer la contribution au signal mesuré des composantes de l'objet observé qui ont une distribution spatiale paire : fuites stellaires, lumière exozodiacale, et *a fortiori* lumière zodiacale et émission thermique de l'instrument (qui ont un niveau constant dans le champ). On peut alors ne rechercher, lors de la reconstruction d'image, que les planètes, ce qui correspond au modèle d'objet suivant :

$$o_\lambda(\boldsymbol{\theta}) = \sum_{k=1}^{N_{\text{src}}} F_{k,\lambda} \delta(\boldsymbol{\theta} - \boldsymbol{\theta}_k), \quad (9.30)$$

où  $N_{\text{src}}$  est le nombre de planètes (supposé ici connu) et  $F_{k,\lambda}$  est le spectre de la  $k$ -ième planète dans un intervalle spectral  $[\lambda_{\text{min}}, \lambda_{\text{max}}]$  fixé par l'instrument. Ce modèle paramétré permet de contraindre notablement l'inversion pour contrebalancer le fait que les données sont nettement plus pauvres qu'une image.

Avec ce modèle d'objet, le modèle de formation de données s'écrit :

$$i_{t,\lambda} = \sum_{k=1}^{N_{\text{src}}} R_{t,\lambda}(\boldsymbol{\theta}_k) F_{k,\lambda} + b_{t,\lambda}, \quad (9.31)$$

où  $b_{t,\lambda}$  est un bruit supposé gaussien blanc, dont la variance  $\sigma_{t,\lambda}^2$  peut être estimée à partir des données et est supposée connue ici. Le problème inverse à résoudre est d'estimer les positions  $\boldsymbol{\theta}_k$  et les spectres  $F_{k,\lambda}$  des planètes, que l'on regroupe dans deux vecteurs notés  $(\boldsymbol{\theta}, \mathbf{F})$ . La solution MV est celle qui minimise le critère suivant [17], de type moindres carrés pondérés au vu des hypothèses faites sur le bruit :

$$J_i(\boldsymbol{\theta}, \mathbf{F}) = \sum_{t,\lambda} \frac{1}{\sigma_{t,\lambda}^2} \left[ i_{t,\lambda} - \sum_{k=1}^{N_{\text{src}}} R_{t,\lambda}(\boldsymbol{\theta}_k) F_{k,\lambda} \right]^2 \quad (9.32)$$

Comme on le verra sur les résultats de reconstruction, l'inversion reste difficile dans les conditions de fort bruit envisagées. Le modèle d'objet adopté (Équ. (9.30)), séparable entre variables spatiales et spectrales, contient déjà toutes nos informations *a priori* spatiales sur l'objet. Il est néanmoins possible de contraindre encore plus l'inversion en ajoutant la connaissance supplémentaire que les spectres recherchés sont des grandeurs positives (à toutes les longueurs d'ondes) et de plus sont des fonctions relativement « douces » de la longueur d'onde. Cette dernière information est prise en compte en incorporant au critère à minimiser une régularisation spectrale, qui mesure la rugosité du spectre :

$$J_o(\mathbf{F}) = \sum_{k=1}^{N_{\text{src}}} \mu_k \sum_{\lambda=\lambda_{\text{min}}}^{\lambda_{\text{max}}} \left( \frac{\partial^m F_{k,\lambda}}{\partial \lambda^m} \right)^2, \quad (9.33)$$

où la dérivée  $m$ -ième du spectre ( $m = 1$  ou  $2$  en pratique) est calculée par différences finies et où les  $\mu_k$  sont des hyper-paramètres qui permettent de régler le poids accordé à

la régularisation (voir [18] pour plus de détails). La solution MAP est celle qui minimise le critère composite  $J_{\text{MAP}}(\boldsymbol{\theta}, \mathbf{F}) = J_i(\boldsymbol{\theta}, \mathbf{F}) + J_o(\mathbf{F})$ . La mise en œuvre de cette minimisation est assez délicate du fait de nombreux minima locaux du critère. Elle utilise le fait que, pour chaque position  $\boldsymbol{\theta}$  supposée des planètes, on peut obtenir simplement l'estimée MAP du spectre  $\hat{\mathbf{F}}(\boldsymbol{\theta})$  car le critère  $J_{\text{MAP}}$  est quadratique vis-à-vis des spectres  $\mathbf{F}$ . Si l'on remplace ceux-ci par  $\hat{\mathbf{F}}(\boldsymbol{\theta})$  dans  $J_{\text{MAP}}$ , on obtient une fonction à minimiser partiellement optimisée, qui ne dépend plus (explicitement) que des positions :

$$J_{\text{MAP}}^\dagger(\boldsymbol{\theta}) = J_{\text{MAP}}(\boldsymbol{\theta}, \hat{\mathbf{F}}(\boldsymbol{\theta})) . \quad (9.34)$$

La minimisation de ce critère est fondée sur une recherche séquentielle des planètes, à la manière de l'algorithme CLEAN. La figure 9.5 montre les cartes du critère  $J_{\text{MAP}}^\dagger$  obtenues dans le cas d'une seule planète, en fonction des informations *a priori* utilisées. Il apparaît clairement que les contraintes de positivité et de douceur ajoutées sur les spectres améliorent notablement l'estimation de la position de la planète, en discréditant (Fig. 9.5 droite) des positions compatibles avec les données (Fig. 9.5 gauche et centre) mais correspondant à des spectres très chaotiques.

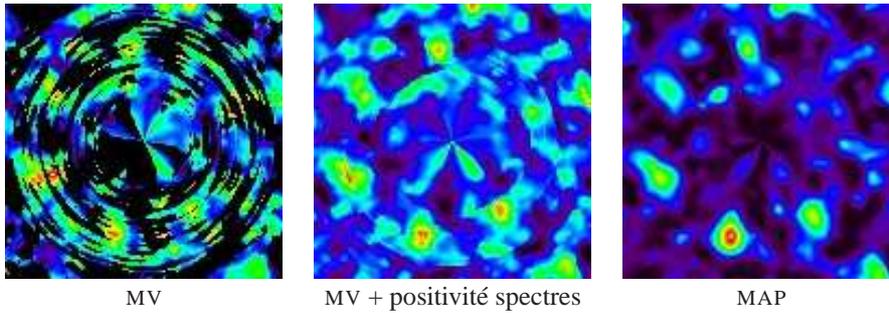


FIG. 9.5 – Cartes de vraisemblance pour la position de la planète (d'après [17] et [18]). À gauche, vraisemblance seule ; au centre, vraisemblance sous contrainte de positivité des spectres ; à droite, MAP c'est-à-dire vraisemblance pénalisée par un critère de régularisation spectrale. Code couleur : noir correspond à une vraisemblance nulle, rouge à une vraisemblance maximale. La vraie position de la planète est en bas, légèrement à gauche, bien visible sur l'image de droite.

La figure 9.6 montre le spectre estimé d'une Terre ; comme attendu, la régularisation spectrale évite l'amplification du bruit et a un effet bénéfique sur l'estimation.

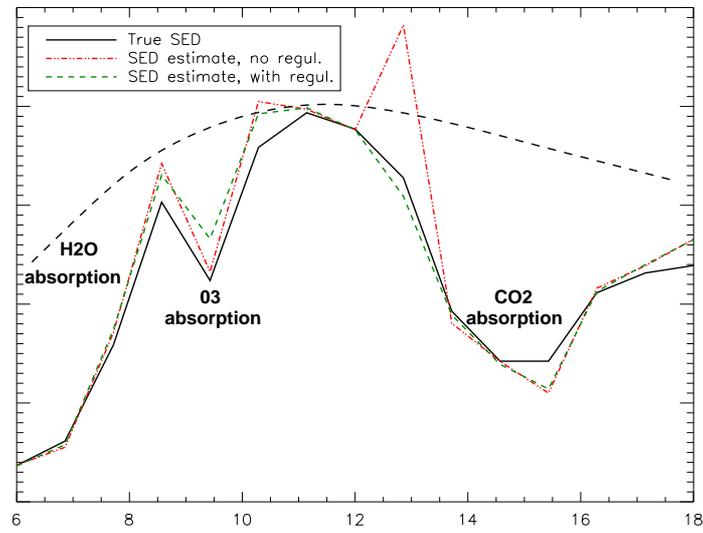


FIG. 9.6 – Spectre reconstruit à la position estimée de la planète : sans régularisation (pointillés) et avec régularisation (tirets), à comparer au spectre vrai d'une Terre (trait plein), d'après [17].

# Bibliographie

- [1] D. M. Titterton. General structure of regularization procedures in image reconstruction. *Astron. Astrophys.*, 144 :381–387, 1985.
- [2] G. Demoment. Image reconstruction and restoration : Overview of common estimation structures and problems. *IEEE Trans. Acoust. Speech Signal Process.*, 37(12) :2024–2036, December 1989.
- [3] Jérôme Idier, editor. *Approche bayésienne pour les problèmes inverses*. Hermès, Paris, 2001.
- [4] Jérôme Idier, editor. *Bayesian Approach for Inverse Problems*. ISTE, London, 2008.
- [5] A. Blanc, L. M. Mugnier, and J. Idier. Marginal estimation of aberrations and image restoration by use of phase diversity. *J. Opt. Soc. Am. A*, 20(6) :1035–1045, 2003.
- [6] L. M. Mugnier, A. Blanc, and J. Idier. Phase diversity : a technique for wave-front sensing and for diffraction-limited imaging. In Peter Hawkes, editor, *Advances in Imaging and Electron Physics*, volume 141, chapter 1, pages 1–76. Elsevier, 2006.
- [7] E. Thiébaud, P. J. V. Garcia, and R. Foy. Imaging with Amber/VLTI : the case of microjets. *Astrophys. Space. Sci.*, 286 :171–176, 2003.
- [8] L. M. Mugnier, G. Le Besnerais, and S. Meimon. Inversion in optical imaging through atmospheric turbulence. In Jérôme Idier, editor, *Bayesian Approach to Inverse Problems*, Digital Signal and Image Processing Series, chapter 10, pages 243–283. ISTE / John Wiley, London, 2008.
- [9] M. Zuhair Nashed. Operator-theoretic and computational approaches to ill-posed problems with applications to antenna theory. *IEEE Trans. Antennas Propag.*, AP-29(2) :220–231, March 1981.
- [10] L. M. Mugnier, T. Fusco, and J.-M. Conan. MISTRAL : a myopic edge-preserving image restoration method, with application to astronomical adaptive-optics-corrected long-exposure images. *J. Opt. Soc. Am. A*, 21(10) :1841–1854, October 2004.
- [11] S. M. Stigler. Gauss and the invention of least squares. *The Annals of Statistics*, 9(3) :465–474, May 1981.
- [12] D. Gratadour, L. M. Mugnier, and D. Rouan. Sub-pixel image registration with a maximum likelihood estimator. *Astron. Astrophys.*, 443 :357–365, November 2005.
- [13] E. Lehmann. *Theory of point estimation*. John Wiley, New York, NY, 1983.

- [14] B. Le Roux, J.-M. Conan, C. Kulcsár, H.-F. Raynaud, L. M. Mugnier, and T. Fusco. Optimal control law for classical and multiconjugate adaptive optics. *J. Opt. Soc. Am. A*, 21(7), July 2004.
- [15] L. Mugnier and G. Le Besnerais. Problèmes inverses en imagerie optique à travers la turbulence. In Jérôme Idier, editor, *Approche bayésienne pour les problèmes inverses*, chapter 10, pages 241–270. Hermès, Paris, 2001.
- [16] J.-M. Conan, L. M. Mugnier, T. Fusco, V. Michau, and G. Rousset. Myopic deconvolution of adaptive optics images by use of object and point spread function power spectra. *Appl. Opt.*, 37(21) :4614–4622, July 1998.
- [17] L. M. Mugnier, E. Thiébaud, and A. Belu. Data processing in nulling interferometry : case of the Darwin mission. In C. Aime, M. Carillet, and A. Ferrari, editors, *Astronomy with High Contrast Imaging III*, EAS Publications Series, pages 69–84. EDP Sciences, Les Ulis, France, 2006.
- [18] E. Thiébaud and L. Mugnier. Maximum a posteriori planet detection and characterization with a nulling interferometer. In *IAUC 200, Direct Imaging of Exoplanets : Science & Techniques*, 2006. Date conférence : October 2005, Nice, France.