

Approche bayésienne pour la résolution de problèmes inverses

Eric Thiébaud¹ Christophe Pichon²

¹Centre de Recherche Astronomique de Lyon
Université Claude Bernard Lyon I
Ecole Normale Supérieure de Lyon

²Institut d'Astrophysique de Paris

28 août – 2 septembre 2006 / VIIIème Ecole de Cosmologie, Cargèse

Introduction

Qu'est-ce qu'un problème inverse ?

- 1 Des observables disponibles et modélisables :

$$\mathbf{y} = \mathbf{m}(\mathbf{x}) + \mathbf{e}$$

- les données $\mathbf{y} \in \mathbb{R}^{N_{\mathbf{y}}}$;
- le modèle $\mathbf{m} : \mathbb{R}^{N_{\mathbf{x}}} \mapsto \mathbb{R}^{N_{\mathbf{y}}}$;
- les paramètres recherchés $\mathbf{x} \in \mathbb{R}^{N_{\mathbf{x}}}$;
- les erreurs $\mathbf{e} \in \mathbb{R}^{N_{\mathbf{y}}}$.

Qu'est-ce qu'un problème inverse ?

- 1 Des observables disponibles et modélisables :

$$\mathbf{y} = \mathbf{m}(\mathbf{x}) + \mathbf{e}$$

- les données $\mathbf{y} \in \mathbb{R}^{N_{\mathbf{y}}}$;
 - le modèle $\mathbf{m} : \mathbb{R}^{N_{\mathbf{x}}} \mapsto \mathbb{R}^{N_{\mathbf{y}}}$;
 - les paramètres recherchés $\mathbf{x} \in \mathbb{R}^{N_{\mathbf{x}}}$;
 - les erreurs $\mathbf{e} \in \mathbb{R}^{N_{\mathbf{y}}}$.
- 2 Quel sont les meilleurs paramètres \mathbf{x} compte tenu des données \mathbf{y} et du modèle \mathbf{m} ?

Qu'est-ce qu'un problème inverse ?

- 1 Des observables disponibles et modélisables :

$$\mathbf{y} = \mathbf{m}(\mathbf{x}) + \mathbf{e}$$

- les données $\mathbf{y} \in \mathbb{R}^{N_{\mathbf{y}}}$;
 - le modèle $\mathbf{m} : \mathbb{R}^{N_{\mathbf{x}}} \mapsto \mathbb{R}^{N_{\mathbf{y}}}$;
 - les paramètres recherchés $\mathbf{x} \in \mathbb{R}^{N_{\mathbf{x}}}$;
 - les erreurs $\mathbf{e} \in \mathbb{R}^{N_{\mathbf{y}}}$.
- 2 Quel sont les meilleurs paramètres \mathbf{x} compte tenu des données \mathbf{y} et du modèle \mathbf{m} ?
- 3 Il peut y avoir beaucoup de paramètres ($N_{\mathbf{x}} \gtrsim 10^6$) et même plus de paramètres que de mesures: $N_{\mathbf{x}} \gtrsim N_{\mathbf{y}}$.

Qu'est-ce qu'un problème inverse ?

- 1 Des observables disponibles et modélisables :

$$\mathbf{y} = \mathbf{m}(\mathbf{x}) + \mathbf{e}$$

- les données $\mathbf{y} \in \mathbb{R}^{N_{\mathbf{y}}}$;
 - le modèle $\mathbf{m} : \mathbb{R}^{N_{\mathbf{x}}} \mapsto \mathbb{R}^{N_{\mathbf{y}}}$;
 - les paramètres recherchés $\mathbf{x} \in \mathbb{R}^{N_{\mathbf{x}}}$;
 - les erreurs $\mathbf{e} \in \mathbb{R}^{N_{\mathbf{y}}}$.
- 2 Quel sont les meilleurs paramètres \mathbf{x} compte tenu des données \mathbf{y} et du modèle \mathbf{m} ?
 - 3 Il peut y avoir beaucoup de paramètres ($N_{\mathbf{x}} \gtrsim 10^6$) et même plus de paramètres que de mesures: $N_{\mathbf{x}} \gtrsim N_{\mathbf{y}}$.
 - 4 **Un simple ajustement de modèle (e.g maximum de vraisemblance) ne marche pas.**

Remarques / restrictions

$$y = m(x) + e$$

Remarques / restrictions

$$\mathbf{y} = \mathbf{m}(\mathbf{x}) + \mathbf{e}$$

- le nombre de mesures est limité : $N_{\mathbf{y}} < \infty$;

Remarques / restrictions

$$\mathbf{y} = \mathbf{m}(\mathbf{x}) + \mathbf{e}$$

- le nombre de mesures est limité : $N_{\mathbf{y}} < \infty$;
- le nombre de paramètres peut être arbitrairement grand : $N_{\mathbf{x}} \leq \infty$ (e.g. base de fonctions) ;

Remarques / restrictions

$$\mathbf{y} = \mathbf{m}(\mathbf{x}) + \mathbf{e}$$

- le nombre de mesures est limité : $N_{\mathbf{y}} < \infty$;
- le nombre de paramètres peut être arbitrairement grand : $N_{\mathbf{x}} \leq \infty$ (e.g. base de fonctions) ;
- les mesures (et les paramètres) peuvent être hétérogènes ;

Remarques / restrictions

$$y = m(x) + e$$

- le nombre de mesures est limité : $N_y < \infty$;
- le nombre de paramètres peut être arbitrairement grand : $N_x \leq \infty$ (e.g. base de fonctions) ;
- les mesures (et les paramètres) peuvent être hétérogènes ;
- le modèle $m(x)$ est connu mais peut être arbitrairement compliqué (e.g. non-linéaire) ;

Remarques / restrictions

$$y = m(x) + e$$

- le nombre de mesures est limité : $N_y < \infty$;
- le nombre de paramètres peut être arbitrairement grand : $N_x \leq \infty$ (e.g. base de fonctions) ;
- les mesures (et les paramètres) peuvent être hétérogènes ;
- le modèle $m(x)$ est connu mais peut être arbitrairement compliqué (e.g. non-linéaire) ;
- les erreurs e prennent en compte le bruit de mesure, erreurs de modélisation, *etc* ;

Remarques / restrictions

$$\mathbf{y} = \mathbf{m}(\mathbf{x}) + \mathbf{e}$$

- le nombre de mesures est limité : $N_{\mathbf{y}} < \infty$;
- le nombre de paramètres peut être arbitrairement grand : $N_{\mathbf{x}} \leq \infty$ (e.g. base de fonctions) ;
- les mesures (et les paramètres) peuvent être hétérogènes ;
- le modèle $\mathbf{m}(\mathbf{x})$ est connu mais peut être arbitrairement compliqué (e.g. non-linéaire) ;
- les erreurs \mathbf{e} prennent en compte le bruit de mesure, erreurs de modélisation, *etc* ;
- les erreurs sont centrées (le biais du modèle est négligeable) :

$$\langle \mathbf{e} \rangle = \mathbf{0} \quad \text{ou, au moins,} \quad \langle e_i \rangle \ll \sqrt{\text{Var}(e_i)} ;$$

Modèle linéaire

- intégrale de Fredholm du 1^{er} ordre (transformée d'Abel, réponse instrumentale linéaire, projection, ...) :

$$y(\mathbf{r}) = \int h(\mathbf{r}, \mathbf{s}) x(\mathbf{s}) d\mathbf{s} + \text{erreur}$$

Modèle linéaire

- intégrale de Fredholm du 1^{er} ordre (transformée d'Abel, réponse instrumentale linéaire, projection, ...) :

$$y(\mathbf{r}) = \int h(\mathbf{r}, \mathbf{s}) x(\mathbf{s}) d\mathbf{s} + \text{erreur}$$

- base de fonctions (discrétisation) :

$$y_i = y(t_i) = \sum_j x_j m_j(t_i) + e_i$$

en notation vectorielle :

$$\mathbf{y} = \sum_j x_j \mathbf{m}_j + \mathbf{e}$$

Modèle linéaire

- intégrale de Fredholm du 1^{er} ordre (transformée d'Abel, réponse instrumentale linéaire, projection, ...) :

$$y(\mathbf{r}) = \int h(\mathbf{r}, \mathbf{s}) x(\mathbf{s}) d\mathbf{s} + \text{erreur}$$

- base de fonctions (discrétisation) :

$$y_i = y(t_i) = \sum_j x_j m_j(t_i) + e_i$$

en notation vectorielle :

$$\mathbf{y} = \sum_j x_j \mathbf{m}_j + \mathbf{e}$$

- si N_x est fini :

$$\mathbf{y} = \mathbf{M} \cdot \mathbf{x} + \mathbf{e}$$

Modèle linéaire / non-linéaire

1 modèle linéaire :

$$y = M \cdot x + e$$

- intégrale de Fredholm du 1^{er} ordre (transformée d'Abel, convolution, ...)
- base de fonctions ;

Modèle linéaire / non-linéaire

1 modèle linéaire :

$$\mathbf{y} = \mathbf{M} \cdot \mathbf{x} + e$$

- intégrale de Fredholm du 1^{er} ordre (transformée d'Abel, convolution, ...)
- base de fonctions ;

2 modèle non-linéaire :

$$\mathbf{y} = m(\mathbf{x}) + e$$

- tout ou partie du noyau $h(\mathbf{r}, \mathbf{s})$ à retrouver : déconvolution aveugle, synthèse spectrale, ...
- mesures de corrélation (à 2 ou 3 points)
- restauration de phase : cristallographie, clôtures de phases (interférométrie),

...

⇒ unicité ?

Formation d'image

- formation d'image (convolution) :

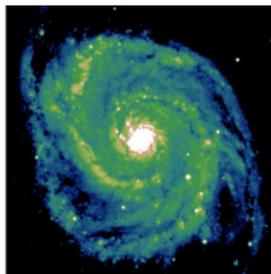
$$y(\omega) = \int h(\omega - \omega') x(\omega') d\omega' + \text{bruit}$$

Formation d'image

- formation d'image (convolution) :

$$y(\omega) = \int h(\omega - \omega') x(\omega') d\omega' + \text{bruit}$$

- objet : $x(\omega')$

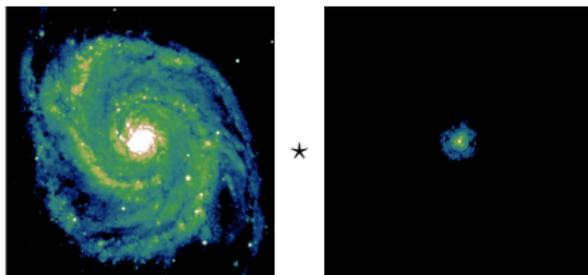


Formation d'image

- formation d'image (convolution) :

$$y(\omega) = \int h(\omega - \omega') x(\omega') d\omega' + \text{bruit}$$

- objet : $x(\omega')$
- réponse impulsionnelle (PSF) : $h(\omega - \omega')$

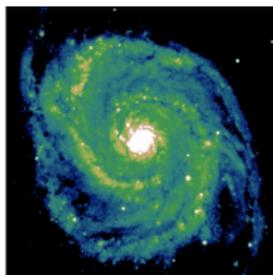


Formation d'image

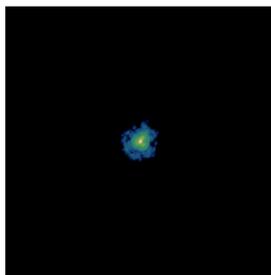
- formation d'image (convolution) :

$$y(\omega) = \int h(\omega - \omega') x(\omega') d\omega' + \text{bruit}$$

- objet : $x(\omega')$
- réponse impulsionnelle (PSF) : $h(\omega - \omega')$



★



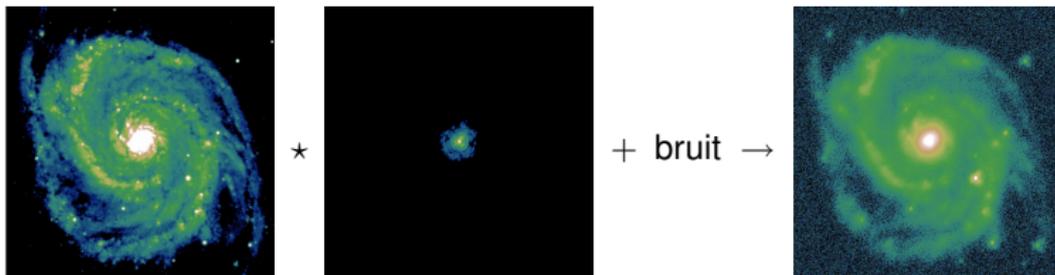
+ bruit

Formation d'image

- formation d'image (convolution) :

$$y(\omega) = \int h(\omega - \omega') x(\omega') d\omega' + \text{bruit}$$

- objet : $x(\omega')$
- réponse impulsionnelle (PSF) : $h(\omega - \omega')$
- image observée : $y(\omega)$



Dynamique galactique

- dynamique disque galactique mince (Transformée d'Abel) :

$$F_{\phi}(R, v_{\phi}) = \sqrt{2} \int_{\frac{1}{2} v_{\phi}^2 - \psi(R)}^0 \frac{f(\varepsilon, R v_{\phi})}{\sqrt{\varepsilon + \psi(R) - \frac{1}{2} v_{\phi}^2}} d\varepsilon$$

Dynamique galactique

- dynamique disque galactique mince (Transformée d'Abel) :

$$F_{\phi}(R, v_{\phi}) = \sqrt{2} \int_{\frac{1}{2} v_{\phi}^2 - \psi(R)}^0 \frac{f(\varepsilon, R v_{\phi})}{\sqrt{\varepsilon + \psi(R) - \frac{1}{2} v_{\phi}^2}} d\varepsilon$$

- observables** : distribution des vitesses $F_{\phi}(R, v_{\phi})$



Dynamique galactique

- dynamique disque galactique mince (Transformée d'Abel) :

$$F_{\phi}(R, v_{\phi}) = \sqrt{2} \int_{\frac{1}{2} v_{\phi}^2 - \psi(R)}^0 \frac{f(\varepsilon, R v_{\phi})}{\sqrt{\varepsilon + \psi(R) - \frac{1}{2} v_{\phi}^2}} d\varepsilon$$

- observables** : distribution des vitesses $F_{\phi}(R, v_{\phi})$



- inconnues** : distribution des orbites $f(\varepsilon, R v_{\phi})$

Dynamique galactique

- dynamique disque galactique mince (Transformée d'Abel) :

$$F_{\phi}(R, v_{\phi}) = \sqrt{2} \int_{\frac{1}{2} v_{\phi}^2 - \psi(R)}^0 \frac{f(\varepsilon, R v_{\phi})}{\sqrt{\varepsilon + \psi(R) - \frac{1}{2} v_{\phi}^2}} d\varepsilon$$

- observables** : distribution des vitesses $F_{\phi}(R, v_{\phi})$



- inconnues** : distribution des orbites $f(\varepsilon, R v_{\phi})$
- cinématique du gaz HI \Rightarrow potentiel $\psi(R)$

Bruit et erreurs de modélisation

- modèle direct :

$$\mathbf{y} = \mathbf{m}(\mathbf{x}) + \mathbf{e}$$

- \mathbf{e} incorpore les erreurs de modélisation et le bruit ;
- la statistique des erreurs est (approximativement) connue :
 - les erreurs sont centrées (modèle non biaisé) :

$$\langle \mathbf{e} \rangle = \langle \mathbf{y} - \mathbf{m}(\mathbf{x}) \rangle = \mathbf{0}$$

- la covariance des erreurs \mathbf{C}_e est la covariance $\mathbf{C}_{\mathbf{y}|\mathbf{x}}$ des mesures étant donné le modèle \mathbf{x} :

$$\begin{aligned} \mathbf{C}_e &= \langle \mathbf{e} \cdot \mathbf{e}^T \rangle && \text{(erreurs centrées)} \\ &= \left\langle [\mathbf{y} - \mathbf{m}(\mathbf{x})] \cdot [\mathbf{y} - \mathbf{m}(\mathbf{x})]^T \right\rangle && \text{(à } \mathbf{x} \text{ donné)} \\ &= \mathbf{C}_{\mathbf{y}|\mathbf{x}} \end{aligned}$$

Approximation gaussienne pour les erreurs

- modèle direct :

$$\mathbf{y} = \mathbf{m}(\mathbf{x}) + \mathbf{e}$$

- \mathbf{e} incorpore les erreurs de modélisation et le bruit ;
- la statistique des erreurs est **normale centrée** :

$$\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_e)$$

c'est-à-dire :

- les erreurs sont centrées (modèle non biaisé) :

$$\langle \mathbf{e} \rangle = \langle \mathbf{y} - \mathbf{m}(\mathbf{x}) \rangle = \mathbf{0}$$

- la fonction de distribution des erreurs est gaussienne :

$$\text{PDF}(\mathbf{y}|\mathbf{x}) = \frac{\exp \left\{ -\frac{1}{2} [\mathbf{y} - \mathbf{m}(\mathbf{x})]^T \cdot \mathbf{C}_{\mathbf{y}|\mathbf{x}}^{-1} \cdot [\mathbf{y} - \mathbf{m}(\mathbf{x})] \right\}}{(2\pi)^{N_y/2} |\mathbf{C}_{\mathbf{y}|\mathbf{x}}|^{\frac{1}{2}}}$$

où N_y est le nombre de mesures et $\mathbf{C}_{\mathbf{y}|\mathbf{x}} = \mathbf{C}_e = \langle \mathbf{e} \cdot \mathbf{e}^T \rangle$;

Qu'est-ce que l'on cherche ?

- modèle direct :

$$\mathbf{y} = \mathbf{m}(\mathbf{x}) + e$$

- on veut les **meilleurs** paramètres $\tilde{\mathbf{x}} \simeq \mathbf{x}$ compte tenu des données, du modèle et de la statistique (supposée) des erreurs ;
- ⇒ il faut un **critère** pour déterminer cette solution ;
- plus ambitieux ; on veut la statistique des paramètres :
 - fonction de distribution des paramètres ;
 - moyenne et variance des paramètres ;
- ⇒ **approche bayésienne**.

Maximum de vraisemblance

Maximum de vraisemblance : Objectif

- Quel est le meilleur modèle ?

⇒ **réponse** : c'est celui qui maximise la probabilité d'avoir observé les données :

$$x_{\text{ML}} = \arg \max_x \Pr(\mathbf{y} | \mathbf{x})$$

où **ML = *Maximum Likelihood***

- ▲ C'est la meilleure solution au sens de la statistiques des erreurs...

Maximum de vraisemblance : Fonction de pénalisation

- solution au sens du maximum de vraisemblance :

$$\begin{aligned}\mathbf{x}_{\text{ML}} &= \arg \max_{\mathbf{x}} \Pr(\mathbf{y}|\mathbf{x}) \\ &= \arg \min_{\mathbf{x}} f_{\text{data}}(\mathbf{x})\end{aligned}$$

- fonction de pénalisation :

$$f_{\text{data}}(\mathbf{x}) = -k \log \Pr(\mathbf{y}|\mathbf{x}) + \text{const}$$

avec $k > 0$;

- condition d'optimalité du 1^{er} ordre :

$$\left. \frac{\partial f_{\text{data}}(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\mathbf{x}_{\text{ML}}} = \mathbf{0}$$

si $f_{\text{data}}(\mathbf{x})$ est convexe cette condition est suffisante.

Maximum de vraisemblance : Approximation gaussienne

- fonction de distribution des erreurs :

$$\text{PDF}(\mathbf{y}|\mathbf{x}) = \frac{\exp \left\{ -\frac{1}{2} [\mathbf{y} - \mathbf{m}(\mathbf{x})]^T \cdot \mathbf{C}_{\mathbf{y}|\mathbf{x}}^{-1} \cdot [\mathbf{y} - \mathbf{m}(\mathbf{x})] \right\}}{(2\pi)^{N_{\mathbf{y}}/2} |\mathbf{C}_{\mathbf{y}|\mathbf{x}}|^{\frac{1}{2}}}$$

- solution pour un modèle non-linéaire :

$$\begin{aligned} \mathbf{x}_{\text{ML}} &= \arg \max_{\mathbf{x}} \Pr(\mathbf{y}|\mathbf{x}) \\ &= \arg \max_{\mathbf{x}} \text{PDF}(\mathbf{y}|\mathbf{x}) && \text{(espace vectoriel)} \\ &= \arg \min_{\mathbf{x}} f_{\text{data}}(\mathbf{x}) \end{aligned}$$

- fonction de pénalisation :

$$\begin{aligned} f_{\text{data}}(\mathbf{x}) &= -2 \log \text{PDF}(\mathbf{y}|\mathbf{x}) + \text{const} \\ &= [\mathbf{y} - \mathbf{m}(\mathbf{x})]^T \cdot \mathbf{C}_{\mathbf{y}|\mathbf{x}}^{-1} \cdot [\mathbf{y} - \mathbf{m}(\mathbf{x})] \end{aligned}$$

- résolution par algorithme de moindres carrés non-linéaires (e.g. Levenberg-Marquardt si pas trop de paramètres) ;

Maximum de vraisemblance : Approximation gaussienne, modèle linéaire

- modèle :

$$\mathbf{y} = \mathbf{M} \cdot \mathbf{x} + \mathbf{e}$$

- fonction de pénalisation :

$$f_{\text{data}}(\mathbf{x}) = (\mathbf{y} - \mathbf{M} \cdot \mathbf{x})^T \cdot \mathbf{C}_{\mathbf{y}|\mathbf{x}}^{-1} \cdot (\mathbf{y} - \mathbf{M} \cdot \mathbf{x})$$

- condition d'optimalité du 1^{er} ordre (**équations normales**) :

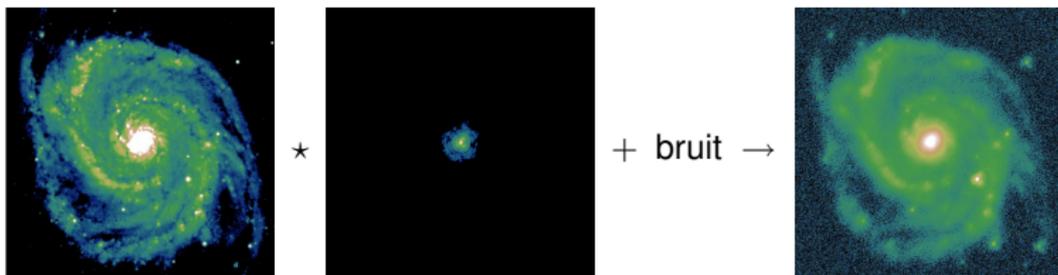
$$\left. \frac{\partial f_{\text{data}}(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\mathbf{x}_{\text{ML}}} = \mathbf{0}$$

$$\Leftrightarrow \mathbf{M}^T \cdot \mathbf{C}_{\mathbf{y}|\mathbf{x}}^{-1} \cdot \mathbf{M} \cdot \mathbf{x}_{\text{ML}} = \mathbf{M}^T \cdot \mathbf{C}_{\mathbf{y}|\mathbf{x}}^{-1} \cdot \mathbf{y}$$

$$\Leftrightarrow \mathbf{x}_{\text{ML}} = \left(\mathbf{M}^T \cdot \mathbf{C}_{\mathbf{y}|\mathbf{x}}^{-1} \cdot \mathbf{M} \right)^{-1} \cdot \mathbf{M}^T \cdot \mathbf{C}_{\mathbf{y}|\mathbf{x}}^{-1} \cdot \mathbf{y}$$

Maximum de vraisemblance : Application à la déconvolution

- formation d'image (convolution) :



$$y(\omega) = \int h(\omega - \omega') x(\omega') d\omega' + e(\omega)$$

- diagonalisation par transformation de Fourier :

$$\hat{y}(\nu) = \underbrace{\hat{h}(\nu) \hat{x}(\nu)}_{\hat{m}(\nu)} + \hat{e}(\nu)$$

Maximum de vraisemblance : Application à la déconvolution

- bruit gaussien blanc :

$$\begin{aligned} f_{\text{data}}(\mathbf{x}) &= \frac{1}{\sigma^2} \int |y(\omega) - m(\omega)|^2 d\omega \\ &= \frac{1}{\sigma^2} \int |\hat{y}(\nu) - \hat{m}(\nu)|^2 d\nu \end{aligned}$$

avec $\sigma^2 = \text{Var}(e)$;

- solution (au sens du maximum de vraisemblance) :

$$\begin{aligned} \frac{\partial f_{\text{data}}}{\partial \hat{\mathbf{x}}(\nu)} \Big|_{\hat{\mathbf{x}}(\nu) = \hat{\mathbf{x}}_{\text{ML}}(\nu)} &= 0 \\ \iff \boxed{\hat{\mathbf{x}}_{\text{ML}}(\nu) = \frac{\hat{y}(\nu)}{\hat{h}(\nu)}} \end{aligned}$$

Maximum de vraisemblance : Application à la déconvolution

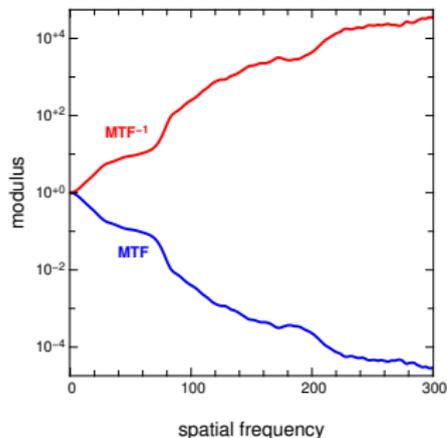
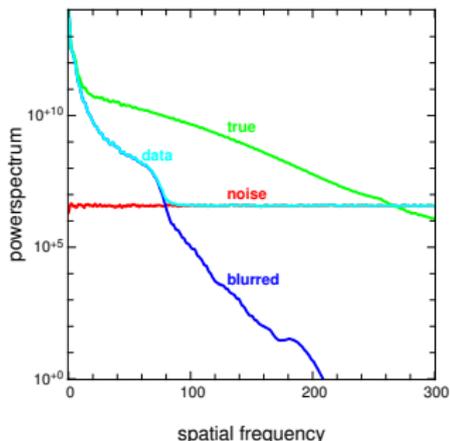
$$\mathbf{x}_{\text{ML}} = \text{FFT}^{-1} \left(\begin{array}{c} \text{FFT} \left(\begin{array}{c} \text{Image convoluée} \end{array} \right) \\ \text{FFT} \left(\begin{array}{c} \text{Image de référence} \end{array} \right) \end{array} \right)$$

Maximum de vraisemblance : Application à la déconvolution

$$\mathbf{x}_{\text{ML}} = \text{FFT}^{-1} \left(\begin{array}{c} \text{FFT} \left(\begin{array}{c} \text{Image convoluee} \end{array} \right) \\ \text{FFT} \left(\begin{array}{c} \text{Image deconvoluee} \end{array} \right) \end{array} \right) = \begin{array}{c} \text{unconstrained maximum likelihood} \end{array}$$

The diagram illustrates the unconstrained maximum likelihood deconvolution process. It shows the inverse FFT of the product of the FFT of a convolved image and the FFT of a deconvolved image, resulting in a noisy image labeled 'unconstrained maximum likelihood'.

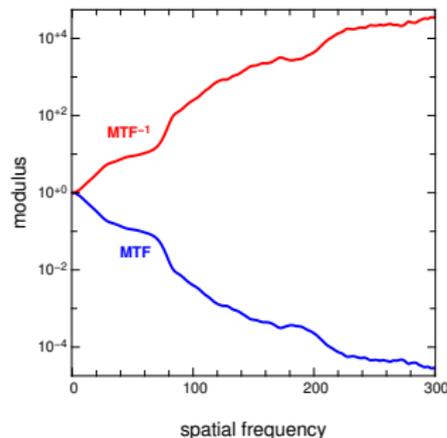
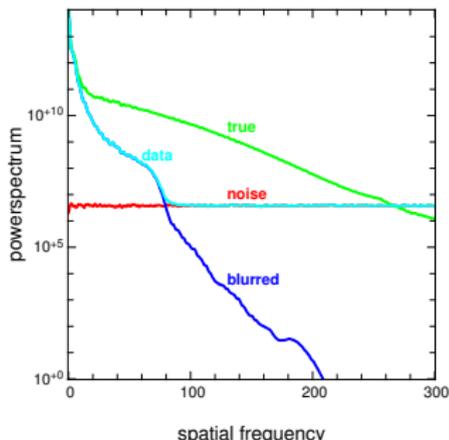
Qu'est-ce qui cloche ?



- inversion directe (*i.e.* maximum de vraisemblance) :

$$\hat{x}_{\text{ML}}(\nu) = \frac{\hat{y}(\nu)}{\hat{h}(\nu)}$$

Qu'est-ce qui cloche ?



- inversion directe (*i.e.* maximum de vraisemblance) :

$$\begin{aligned}\hat{x}_{\text{ML}}(\boldsymbol{\nu}) &= \frac{\hat{y}(\boldsymbol{\nu})}{\hat{h}(\boldsymbol{\nu})} \\ &= \hat{x}(\boldsymbol{\nu}) + \frac{\hat{e}(\boldsymbol{\nu})}{\hat{h}(\boldsymbol{\nu})}\end{aligned}$$

- ▶ amplification du bruit

Troncature fréquentielle

Il est possible d'éviter l'amplification du bruit en imposant une fréquence de coupure :

$$\hat{x}_{\text{cut}}(\nu) = \begin{cases} \begin{cases} \hat{y}(\nu) \\ \hat{h}(\nu) \end{cases} & \text{for } |\nu| < \nu_{\text{cutoff}} \\ 0 & \text{for } |\nu| \geq \nu_{\text{cutoff}} \end{cases}$$

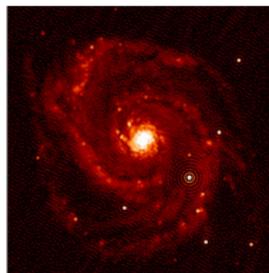
qui donne (avec $\nu_{\text{cutoff}} = 80$ frequels) :

Troncature fréquentielle

Il est possible d'éviter l'amplification du bruit en imposant une fréquence de coupure :

$$\hat{x}_{\text{cut}}(\nu) = \begin{cases} \frac{\hat{y}(\nu)}{\hat{h}(\nu)} & \text{for } |\nu| < \nu_{\text{cutoff}} \\ 0 & \text{for } |\nu| \geq \nu_{\text{cutoff}} \end{cases}$$

qui donne (avec $\nu_{\text{cutoff}} = 80$ frequels) :



Maximum de vraisemblance

Propriétés

- solution (maximum de vraisemblance) :

$$\begin{aligned}
 \mathbf{x}_{\text{ML}} &= \left(\mathbf{M}^T \cdot \mathbf{C}_{\mathbf{y}|\mathbf{x}}^{-1} \cdot \mathbf{M} \right)^{-1} \cdot \mathbf{M}^T \cdot \mathbf{C}_{\mathbf{y}|\mathbf{x}}^{-1} \cdot \mathbf{y} \\
 &= \left(\mathbf{M}^T \cdot \mathbf{C}_{\mathbf{y}|\mathbf{x}}^{-1} \cdot \mathbf{M} \right)^{-1} \cdot \mathbf{M}^T \cdot \mathbf{C}_{\mathbf{y}|\mathbf{x}}^{-1} \cdot (\mathbf{M} \cdot \mathbf{x} + \mathbf{e}) \\
 &= \mathbf{x} + \left(\mathbf{M}^T \cdot \mathbf{C}_{\mathbf{y}|\mathbf{x}}^{-1} \cdot \mathbf{M} \right)^{-1} \cdot \mathbf{M}^T \cdot \mathbf{C}_{\mathbf{y}|\mathbf{x}}^{-1} \cdot \mathbf{e}
 \end{aligned}$$

- non-biaisé au sens des erreurs :

$$\langle \mathbf{x}_{\text{ML}} \rangle_{\mathbf{e}} = \mathbf{x}$$

- “covariance” par rapport aux vrais paramètres :

$$\begin{aligned}
 \mathbf{C}_{\text{ML}} &= \langle (\mathbf{x} - \mathbf{x}_{\text{ML}}) \cdot (\mathbf{x} - \mathbf{x}_{\text{ML}})^T \rangle_{\mathbf{e}} \\
 &= \left(\mathbf{M}^T \cdot \mathbf{C}_{\mathbf{y}|\mathbf{x}}^{-1} \cdot \mathbf{M} \right)^{-1}
 \end{aligned}$$

- ▶ amplification du bruit

Minimum de variance

Critère du minimum de variance

- critère de sélection :

$$\mathbf{x}_{\text{MMSE}} = \mathbf{x}_{\text{MV}} = \arg \min_{\tilde{\mathbf{x}}} \langle \|\mathbf{x} - \tilde{\mathbf{x}}\|^2 \rangle$$

(MMSE = *minimum mean square error*, MV = *minimum variance*) ;

- hypothèses :
 - \mathbf{x} et \mathbf{e} sont deux processus aléatoires indépendants de moyennes et covariances connues :

$$\begin{aligned} \langle \mathbf{x} \rangle &= \bar{\mathbf{x}} & \text{et} & & \mathbf{C}_{\mathbf{x}} &= \langle (\mathbf{x} - \bar{\mathbf{x}}) \cdot (\mathbf{x} - \bar{\mathbf{x}})^T \rangle \\ \langle \mathbf{e} \rangle &= \mathbf{0} & \text{et} & & \mathbf{C}_{\mathbf{e}} &= \mathbf{C}_{\mathbf{y}|\mathbf{x}} = \langle \mathbf{e} \cdot \mathbf{e}^T \rangle \end{aligned}$$

- modèle linéaire : $\mathbf{m}(\mathbf{x}) = \mathbf{M} \cdot \mathbf{x}$;
- on cherche la solution sous la forme :

$$\begin{aligned} \tilde{\mathbf{x}} &= \langle \mathbf{x} \rangle + \mathbf{S} \cdot (\mathbf{y} - \langle \mathbf{y} \rangle) \\ &= \bar{\mathbf{x}} + \mathbf{S} \cdot [\mathbf{M} \cdot (\mathbf{x} - \bar{\mathbf{x}}) + \mathbf{e}] \end{aligned}$$

où \mathbf{S} est l'**opérateur d'estimation** linéaire.

Dérivation du minimum de variance

- pénalisation (MSE = *mean square error*) :

$$\begin{aligned} \text{MSE} &= \langle \|\mathbf{x} - \tilde{\mathbf{x}}\|^2 \rangle \\ &= \langle \|(\mathbf{x} - \bar{\mathbf{x}}) - \mathbf{S} \cdot [\mathbf{M} \cdot (\mathbf{x} - \bar{\mathbf{x}}) + \mathbf{e}]\|^2 \rangle \end{aligned}$$

- MSE est quadratique et convexe en $S_{i,j}$ donc le minimum est atteint en annulant les dérivées partielles de MSE par rapport à \mathbf{S} :

$$\left. \frac{\partial \text{MSE}}{\partial \mathbf{S}} \right|_{\mathbf{S}=\tilde{\mathbf{S}}} = \mathbf{0};$$

en développant :

$$\begin{aligned} \frac{\partial \text{MSE}}{\partial \mathbf{S}} = \mathbf{0} &\Leftrightarrow (\mathbf{S} \cdot \mathbf{M} - \mathbf{I}) \cdot \mathbf{C}_x \cdot \mathbf{M}^T + \mathbf{S} \cdot \mathbf{C}_e = \mathbf{0} \\ &\Leftrightarrow \mathbf{S} \cdot (\mathbf{M} \cdot \mathbf{C}_x \cdot \mathbf{M}^T + \mathbf{C}_e) = \mathbf{C}_x \cdot \mathbf{M}^T \\ &\Leftrightarrow \mathbf{S} = \mathbf{C}_x \cdot \mathbf{M}^T \cdot (\mathbf{M} \cdot \mathbf{C}_x \cdot \mathbf{M}^T + \mathbf{C}_e)^{-1} \end{aligned}$$

$$\begin{aligned}
 \frac{1}{2} \frac{\partial \text{MSE}}{\partial \mathbf{S}} &= \langle \{ \mathbf{S} \cdot [\mathbf{M} \cdot (\mathbf{x} - \bar{\mathbf{x}}) + \mathbf{e}] - (\mathbf{x} - \bar{\mathbf{x}}) \} \cdot [\mathbf{M} \cdot (\mathbf{x} - \bar{\mathbf{x}}) + \mathbf{e}]^T \rangle \\
 &= (\mathbf{S} \cdot \mathbf{M} - \mathbf{I}) \langle (\mathbf{x} - \bar{\mathbf{x}}) \cdot (\mathbf{x} - \bar{\mathbf{x}})^T \rangle \cdot \mathbf{M}^T + \mathbf{S} \cdot \langle \mathbf{e} \cdot \mathbf{e}^T \rangle \\
 &\quad + (\mathbf{S} \cdot \mathbf{M} - \mathbf{I}) \langle (\mathbf{x} - \bar{\mathbf{x}}) \cdot \mathbf{e}^T \rangle + \mathbf{S} \cdot \langle \mathbf{e} \cdot (\mathbf{x} - \bar{\mathbf{x}})^T \rangle \cdot \mathbf{M}^T \\
 &= (\mathbf{S} \cdot \mathbf{M} - \mathbf{I}) \cdot \mathbf{C}_x \cdot \mathbf{M}^T + \mathbf{S} \cdot \mathbf{C}_e
 \end{aligned}$$

en effet

$$\begin{aligned}
 \langle (\mathbf{x} - \bar{\mathbf{x}}) \cdot \mathbf{e}^T \rangle &= \langle \mathbf{x} - \bar{\mathbf{x}} \rangle \cdot \langle \mathbf{e} \rangle^T = \mathbf{0} \cdot \mathbf{0}^T = \mathbf{0} \\
 \langle \mathbf{e} \cdot (\mathbf{x} - \bar{\mathbf{x}})^T \rangle &= \langle (\mathbf{x} - \bar{\mathbf{x}}) \cdot \mathbf{e}^T \rangle^T = \mathbf{0} \\
 \langle (\mathbf{x} - \bar{\mathbf{x}}) \cdot (\mathbf{x} - \bar{\mathbf{x}})^T \rangle &= \mathbf{C}_x \\
 \langle \mathbf{e} \cdot \mathbf{e}^T \rangle &= \mathbf{C}_e
 \end{aligned}$$

Opérateur d'estimation du minimum de variance

On cherche la solution sous la forme :

$$\tilde{\mathbf{x}} = \bar{\mathbf{x}} + \mathbf{S} \cdot (\mathbf{y} - \mathbf{M} \cdot \bar{\mathbf{x}}),$$

où \mathbf{S} est l'**opérateur d'estimation**. Sous le critère de moindre variance :

$$\mathbf{S}_{\text{MV}} = \arg \min_{\mathbf{S}} \langle \|\mathbf{x} - \tilde{\mathbf{x}}\|^2 \rangle$$

on trouve en utilisant les identités matricielles :

$$\begin{aligned} \mathbf{S}_{\text{MV}} &= \mathbf{C}_{\mathbf{x}} \cdot \mathbf{M}^T \cdot (\mathbf{M} \cdot \mathbf{C}_{\mathbf{x}} \cdot \mathbf{M}^T + \mathbf{C}_{\mathbf{e}})^{-1} \\ &= (\mathbf{C}_{\mathbf{x}}^{-1} + \mathbf{M}^T \cdot \mathbf{C}_{\mathbf{e}}^{-1} \cdot \mathbf{M})^{-1} \cdot \mathbf{M}^T \cdot \mathbf{C}_{\mathbf{e}}^{-1} \end{aligned}$$

- 1 pas d'autres hypothèses que : (i) modèle linéaire, (ii) moyenne et covariances connues, et (iii) erreurs centrées
 ► *i.e.* pas forcément gaussien ;
- 2 dimensions du système d'équations à inverser : soit $N_{\mathbf{x}} \times N_{\mathbf{x}}$, soit $N_{\mathbf{y}} \times N_{\mathbf{y}}$;
- 3 \mathbf{S}_{MV} généralise le **filtre de Wiener** ;

Expressions du minimum de variance

solution au sens du minimum de variance :

$$\mathbf{x}_{MV} = \bar{\mathbf{x}} + \mathbf{S}_{MV} \cdot (\mathbf{y} - \mathbf{M} \cdot \bar{\mathbf{x}})$$

en utilisant l'expression de \mathbf{S}_{MV} et les identités matricielles :

$$\begin{aligned} \mathbf{x}_{MV} &= \bar{\mathbf{x}} + \mathbf{C}_x \cdot \mathbf{M}^T \cdot (\mathbf{M} \cdot \mathbf{C}_x \cdot \mathbf{M}^T + \mathbf{C}_e)^{-1} \cdot (\mathbf{y} - \mathbf{M} \cdot \bar{\mathbf{x}}) \\ &= \bar{\mathbf{x}} + (\mathbf{C}_x^{-1} + \mathbf{M}^T \cdot \mathbf{C}_e^{-1} \cdot \mathbf{M})^{-1} \cdot \mathbf{M}^T \cdot \mathbf{C}_e^{-1} \cdot (\mathbf{y} - \mathbf{M} \cdot \bar{\mathbf{x}}) \\ &= (\mathbf{C}_x^{-1} + \mathbf{M}^T \cdot \mathbf{C}_e^{-1} \cdot \mathbf{M})^{-1} \cdot (\mathbf{M}^T \cdot \mathbf{C}_e^{-1} \cdot \mathbf{y} + \mathbf{C}_x^{-1} \cdot \bar{\mathbf{x}}) \end{aligned}$$

cet estimateur est **non-biaisé** au sens où :

$$\langle \mathbf{x}_{MV} \rangle = \bar{\mathbf{x}} = \langle \mathbf{x} \rangle$$

Covariance totale

$$\langle \mathbf{x}_{MV} \rangle = \bar{\mathbf{x}} = \langle \mathbf{x} \rangle$$

donc la covariance totale de \mathbf{x}_{MV} est :

$$\begin{aligned} \text{Cov}(\mathbf{x}_{MV}) &= \langle (\mathbf{x}_{MV} - \bar{\mathbf{x}}) \cdot (\mathbf{x}_{MV} - \bar{\mathbf{x}})^T \rangle \\ &= \langle \mathbf{S}_{MV} \cdot (\mathbf{y} - \mathbf{M} \cdot \bar{\mathbf{x}}) \cdot (\mathbf{y} - \mathbf{M} \cdot \bar{\mathbf{x}})^T \cdot \mathbf{S}_{MV}^T \rangle \\ &= \mathbf{S}_{MV} \cdot \langle [\mathbf{M} \cdot (\mathbf{x} - \bar{\mathbf{x}}) + \mathbf{e}] \cdot [\mathbf{M} \cdot (\mathbf{x} - \bar{\mathbf{x}}) + \mathbf{e}]^T \rangle \cdot \mathbf{S}_{MV}^T \\ &= \mathbf{S}_{MV} \cdot [\mathbf{M} \cdot \langle (\mathbf{x} - \bar{\mathbf{x}}) \cdot (\mathbf{x} - \bar{\mathbf{x}})^T \rangle \cdot \mathbf{M}^T + \langle \mathbf{e} \cdot \mathbf{e}^T \rangle] \cdot \mathbf{S}_{MV}^T \\ &= \mathbf{S}_{MV} \cdot [\mathbf{M} \cdot \mathbf{C}_x \cdot \mathbf{M}^T + \mathbf{C}_e] \cdot \mathbf{S}_{MV}^T \\ &= \mathbf{C}_x \cdot \mathbf{M}^T \cdot (\mathbf{M} \cdot \mathbf{C}_x \cdot \mathbf{M}^T + \mathbf{C}_e)^{-1} \cdot \mathbf{M} \cdot \mathbf{C}_x \\ &= \left(\mathbf{C}_x^{-1} + \mathbf{M}^T \cdot \mathbf{C}_e^{-1} \cdot \mathbf{M} \right)^{-1} \cdot \mathbf{M}^T \cdot \mathbf{C}_e^{-1} \cdot \mathbf{M} \cdot \mathbf{C}_x \\ &= \left[\mathbf{I} - \left(\mathbf{C}_x^{-1} + \mathbf{M}^T \cdot \mathbf{C}_e^{-1} \cdot \mathbf{M} \right)^{-1} \cdot \mathbf{C}_x^{-1} \right] \cdot \mathbf{C}_x \\ &= \mathbf{C}_x - \left(\mathbf{C}_x^{-1} + \mathbf{M}^T \cdot \mathbf{C}_e^{-1} \cdot \mathbf{M} \right)^{-1} \end{aligned}$$

Opérateur de résolution

- Ce qui nous intéresse c'est de trouver les paramètres \mathbf{x} correspondant aux mesures donc il faut comparer \mathbf{x}_{MV} et \mathbf{x} , en moyennant au sens des erreurs e seulement :

$$\langle \mathbf{x}_{MV} \rangle_e = \bar{\mathbf{x}} + \mathbf{S}_{MV} \cdot \mathbf{M} \cdot (\mathbf{x} - \bar{\mathbf{x}})$$

soit

$$\begin{aligned} \langle \mathbf{x}_{MV} \rangle_e - \bar{\mathbf{x}} &= \mathbf{S}_{MV} \cdot \mathbf{M} \cdot (\mathbf{x} - \bar{\mathbf{x}}) \\ &= \mathbf{R}_{MV} \cdot (\mathbf{x} - \bar{\mathbf{x}}) \end{aligned}$$

où

$$\mathbf{R}_{MV} \equiv \mathbf{S}_{MV} \cdot \mathbf{M}$$

est l'**opérateur de résolution** (Backus & Gilbert, 1968) ;

Opérateur de résolution

nous avons vu que :

$$\langle \mathbf{x}_{MV} \rangle_e = \bar{\mathbf{x}} + \mathbf{R}_{MV} \cdot (\mathbf{x} - \bar{\mathbf{x}})$$

avec

$$\begin{aligned} \mathbf{R}_{MV} &= \mathbf{S}_{MV} \cdot \mathbf{M} \\ &= \mathbf{C}_x \cdot \mathbf{M}^T \cdot (\mathbf{M} \cdot \mathbf{C}_x \cdot \mathbf{M}^T + \mathbf{C}_e)^{-1} \cdot \mathbf{M} \\ &= (\mathbf{C}_x^{-1} + \mathbf{M}^T \cdot \mathbf{C}_e^{-1} \cdot \mathbf{M})^{-1} \cdot \mathbf{M}^T \cdot \mathbf{C}_e^{-1} \cdot \mathbf{M} \\ &= \mathbf{I} - (\mathbf{C}_x^{-1} + \mathbf{M}^T \cdot \mathbf{C}_e^{-1} \cdot \mathbf{M})^{-1} \cdot \mathbf{C}_x^{-1} \\ &= \text{Cov}(\mathbf{x}_{MV}) \cdot \mathbf{C}_x^{-1} \\ &= \langle (\mathbf{x}_{MV} - \bar{\mathbf{x}}) \cdot (\mathbf{x}_{MV} - \bar{\mathbf{x}})^T \rangle \cdot \mathbf{C}_x^{-1} \end{aligned}$$

on remarque que :

- 1 $\mathbf{R}_{MV} \neq \mathbf{I}$ donc \mathbf{x}_{MV} est **biaisé** au sens où $\langle \mathbf{x}_{MV} \rangle_e \neq \mathbf{x}$
- 2 le filtrage par \mathbf{R}_{MV} est responsable d'une **atténuation** (i.e. $\mathbf{R}_{MV} \leq \mathbf{I}$) ;

Covariance

- erreur par rapport aux vrais paramètres :

$$\begin{aligned} \mathbf{x} - \mathbf{x}_{MV} &= \mathbf{x} - \bar{\mathbf{x}} - \mathbf{S}_{MV} \cdot (\mathbf{y} - \mathbf{M} \cdot \bar{\mathbf{x}}) \\ &= (\mathbf{I} - \mathbf{R}_{MV}) \cdot (\mathbf{x} - \bar{\mathbf{x}}) - \mathbf{S}_{MV} \cdot \mathbf{e} \end{aligned}$$

- “covariance” par rapport aux vrais paramètres :

$$\begin{aligned} \mathbf{C}_{MV} &= \langle (\mathbf{x} - \mathbf{x}_{MV}) \cdot (\mathbf{x} - \mathbf{x}_{MV})^T \rangle \\ &= (\mathbf{I} - \mathbf{R}_{MV}) \cdot \mathbf{C}_x \cdot (\mathbf{I} - \mathbf{R}_{MV})^T + \mathbf{S}_{MV} \cdot \mathbf{C}_e \cdot \mathbf{S}_{MV}^T \\ &= \left(\mathbf{C}_x^{-1} + \mathbf{M}^T \cdot \mathbf{C}_e^{-1} \cdot \mathbf{M} \right)^{-1} \\ &= \mathbf{C}_x - \underbrace{\mathbf{C}_x \cdot \mathbf{M}^T \left(\mathbf{M} \cdot \mathbf{C}_x \cdot \mathbf{M}^T + \mathbf{C}_e \right)^{-1} \cdot \mathbf{M} \cdot \mathbf{C}_x}_{\text{Cov}(\mathbf{x}_{MV}) = \mathbf{R}_{MV} \cdot \mathbf{C}_x} \\ &= (\mathbf{I} - \mathbf{R}_{MV}) \cdot \mathbf{C}_x \end{aligned}$$

Variance

- erreur par rapport aux vrais paramètres :

$$\mathbf{x} - \mathbf{x}_{\text{MV}} = (\mathbf{I} - \mathbf{R}_{\text{MV}}) \cdot (\mathbf{x} - \bar{\mathbf{x}}) - \mathbf{S}_{\text{MV}} \cdot \mathbf{e}$$

- espérance de la MSE pour la solution \mathbf{x}_{MV} :

$$\begin{aligned} \langle \|\mathbf{x} - \mathbf{x}_{\text{MV}}\|^2 \rangle &= \text{tr} \left(\langle (\mathbf{x} - \mathbf{x}_{\text{MV}}) \cdot (\mathbf{x} - \mathbf{x}_{\text{MV}})^{\text{T}} \rangle \right) \\ &= \text{tr}(\mathbf{C}_{\text{MV}}) \end{aligned}$$

Application à la déconvolution

- formation d'image :
 - réponse linéaire et invariante par translation (isoplanétisme) ;
 - diagonalisation par transformée de Fourier :

$$y(\omega) = \int h(\omega - \omega') x(\omega') d\omega' + e(\omega) \quad \xrightarrow{\text{F.T.}} \quad \hat{y}(\nu) = \underbrace{\hat{h}(\nu) \hat{x}(\nu)}_{\hat{m}(\nu)} + \hat{e}(\nu)$$

- $\langle x \rangle = \bar{x} = \mathbf{0}$ et $\langle e \rangle = \mathbf{0}$;
- covariances C_e et C_x stationnaires
 - ▶ diagonalisation par T.F.
- solution (au sens du minimum de variance) :

$$\begin{aligned} \mathbf{x}_{\text{MV}} &= \bar{\mathbf{x}} + \left(\mathbf{C}_x^{-1} + \mathbf{M}^T \cdot \mathbf{C}_e^{-1} \cdot \mathbf{M} \right)^{-1} \cdot \mathbf{M}^T \cdot \mathbf{C}_e^{-1} \cdot (\mathbf{y} - \mathbf{M} \cdot \bar{\mathbf{x}}) \\ &= \left(\mathbf{C}_x^{-1} + \mathbf{M}^T \cdot \mathbf{C}_e^{-1} \cdot \mathbf{M} \right)^{-1} \cdot \mathbf{M}^T \cdot \mathbf{C}_e^{-1} \cdot \mathbf{y} \quad (\bar{\mathbf{x}} = \mathbf{0}) \\ &= \left(\mathbf{M}^T \cdot \mathbf{M} + \mathbf{C}_e \cdot \mathbf{C}_x^{-1} \right)^{-1} \cdot \mathbf{M}^T \cdot \mathbf{y} \quad (\text{invariance par translation}) \end{aligned}$$

Application à la déconvolution : filtre de Wiener

- formation d'image :

$$y(\omega) = \int h(\omega - \omega') x(\omega') d\omega' + e(\omega) \xrightarrow{\text{F.T.}} \hat{y}(\nu) = \hat{h}(\nu) \hat{x}(\nu) + \hat{e}(\nu)$$

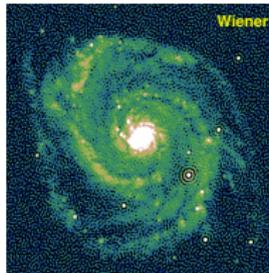
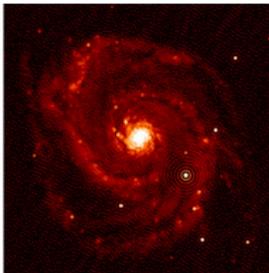
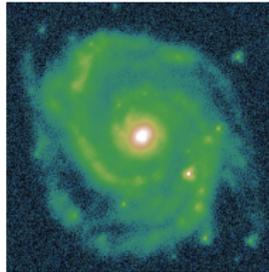
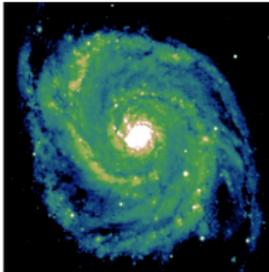
- $\langle x \rangle = \bar{x} = \mathbf{0}$ et $\langle e \rangle = \mathbf{0}$;
- covariances \mathbf{C}_e et \mathbf{C}_x stationnaires ;
- solution (au sens du minimum de variance) = **filtre de Wiener** :

$$\mathbf{x}_{\text{MV}} = \left(\mathbf{M}^T \cdot \mathbf{M} + \mathbf{C}_e \cdot \mathbf{C}_x^{-1} \right)^{-1} \cdot \mathbf{M}^T \cdot \mathbf{y}$$

$\xrightarrow{\text{F.T.}}$

$$\hat{x}_{\text{MV}}(\nu) = \frac{\hat{h}(\nu)^* \hat{y}(\nu)}{|\hat{h}(\nu)|^2 + \frac{\langle |\hat{e}(\nu)|^2 \rangle}{\langle |\hat{x}(\nu)|^2 \rangle}}$$

Filtre de Wiener : Résultats



Minimum de Variance : Résumé

- erreurs (bruit) et signal indépendants
- moyenne et covariance de e et x connues
- $\langle e \rangle = \mathbf{0}$
- estimation linéaire (en fait affine), non-biaisée (au sens où $\langle x_{\text{ML}} \rangle = \bar{x}$)
- généralise le filtre de Wiener
- moindre variance

Maximum a posteriori

Approche bayésienne : Maximum a posteriori

- on maximise la probabilité du modèle étant donné les mesures (MAP = *maximum a posteriori*) :

$$\begin{aligned}
 \mathbf{x}_{\text{MAP}} &= \arg \max_{\mathbf{x}} \Pr(\mathbf{m}(\mathbf{x}) | \mathbf{y}) = \arg \max_{\mathbf{x}} \Pr(\mathbf{x} | \mathbf{y}) \\
 &= \arg \max_{\mathbf{x}} \frac{\Pr(\mathbf{y} | \mathbf{x}) \Pr(\mathbf{x})}{\Pr(\mathbf{y})} \quad (\text{théorème de Bayes}) \\
 &= \arg \min_{\mathbf{x}} \left\{ \underbrace{-\log \Pr(\mathbf{y} | \mathbf{x})}_{f_{\text{data}}(\mathbf{x})} - \underbrace{\log \Pr(\mathbf{x})}_{f_{\text{prior}}(\mathbf{x})} \right\} \\
 &= \arg \min_{\mathbf{x}} f_{\text{post}}(\mathbf{x})
 \end{aligned}$$

- $f_{\text{post}}(\mathbf{x})$ = fonction pénalisante **a posteriori** :

$$f_{\text{post}}(\mathbf{x}) = f_{\text{data}}(\mathbf{x}) + f_{\text{prior}}(\mathbf{x})$$

- $f_{\text{data}}(\mathbf{x})$ = terme de **vraisemblance** (attache aux données)
- $f_{\text{prior}}(\mathbf{x})$ = terme de **régularisation** (a priori)

Approximation gaussienne

- PDF des mesures :

$$\text{PDF}(\mathbf{y}|\mathbf{x}) = \frac{\exp\left\{-\frac{1}{2} [\mathbf{y} - \mathbf{m}(\mathbf{x})]^T \cdot \mathbf{C}_{\mathbf{y}|\mathbf{x}}^{-1} \cdot [\mathbf{y} - \mathbf{m}(\mathbf{x})]\right\}}{(2\pi)^{N_y/2} |\mathbf{C}_{\mathbf{y}|\mathbf{x}}|^{\frac{1}{2}}}$$

- terme de vraisemblance :

$$f_{\text{data}}(\mathbf{x}) = [\mathbf{y} - \mathbf{m}(\mathbf{x})]^T \cdot \mathbf{C}_{\mathbf{y}|\mathbf{x}}^{-1} \cdot [\mathbf{y} - \mathbf{m}(\mathbf{x})]$$

- PDF a priori :

$$\text{PDF}(\mathbf{x}) = \frac{\exp\left\{-\frac{1}{2} (\mathbf{x} - \bar{\mathbf{x}})^T \cdot \mathbf{C}_{\mathbf{x}}^{-1} \cdot (\mathbf{x} - \bar{\mathbf{x}})\right\}}{(2\pi)^{N_x/2} |\mathbf{C}_{\mathbf{x}}|^{\frac{1}{2}}}$$

- terme de régularisation :

$$f_{\text{prior}}(\mathbf{x}) = (\mathbf{x} - \bar{\mathbf{x}})^T \cdot \mathbf{C}_{\mathbf{x}}^{-1} \cdot (\mathbf{x} - \bar{\mathbf{x}})$$

Approximation gaussienne, modèle linéaire

- terme de vraisemblance (modèle linéaire) :

$$f_{\text{data}}(\mathbf{x}) = (\mathbf{y} - \mathbf{M} \cdot \mathbf{x})^T \cdot \mathbf{C}_{\mathbf{y}|\mathbf{x}}^{-1} \cdot (\mathbf{y} - \mathbf{M} \cdot \mathbf{x})$$

- fonction de pénalisation :

$$\begin{aligned} f_{\text{post}}(\mathbf{x}) &= f_{\text{data}}(\mathbf{x}) + f_{\text{prior}}(\mathbf{x}) \\ &= (\mathbf{y} - \mathbf{M} \cdot \mathbf{x})^T \cdot \mathbf{C}_{\mathbf{y}|\mathbf{x}}^{-1} \cdot (\mathbf{y} - \mathbf{M} \cdot \mathbf{x}) + (\mathbf{x} - \bar{\mathbf{x}})^T \cdot \mathbf{C}_{\mathbf{x}}^{-1} \cdot (\mathbf{x} - \bar{\mathbf{x}}) \end{aligned}$$

- gradient :

$$\frac{\partial f_{\text{post}}(\mathbf{x})}{\partial \mathbf{x}} = 2 \mathbf{M}^T \cdot \mathbf{C}_{\mathbf{y}|\mathbf{x}}^{-1} \cdot (\mathbf{M} \cdot \mathbf{x} - \mathbf{y}) + 2 \mathbf{C}_{\mathbf{x}}^{-1} \cdot (\mathbf{x} - \bar{\mathbf{x}})$$

- solution MAP :

$$\begin{aligned} \mathbf{x}_{\text{MAP}} &= \arg \min_{\mathbf{x}} f_{\text{post}}(\mathbf{x}) \\ &= (\mathbf{C}_{\mathbf{x}}^{-1} + \mathbf{M}^T \cdot \mathbf{C}_{\mathbf{y}|\mathbf{x}}^{-1} \cdot \mathbf{M})^{-1} \cdot (\mathbf{M}^T \cdot \mathbf{C}_{\mathbf{y}|\mathbf{x}}^{-1} \cdot \mathbf{y} + \mathbf{C}_{\mathbf{x}}^{-1} \cdot \bar{\mathbf{x}}) \end{aligned}$$

Approximation gaussienne, modèle linéaire

Solution MAP

$$\begin{aligned}
 \mathbf{x}_{\text{MAP}} &= \arg \max_{\mathbf{x}} \Pr(\mathbf{x} | \mathbf{y}) \\
 &= (\mathbf{C}_{\mathbf{x}}^{-1} + \mathbf{M}^T \cdot \mathbf{C}_{\mathbf{y}|\mathbf{x}}^{-1} \cdot \mathbf{M})^{-1} \cdot (\mathbf{M}^T \cdot \mathbf{C}_{\mathbf{y}|\mathbf{x}}^{-1} \cdot \mathbf{y} + \mathbf{C}_{\mathbf{x}}^{-1} \cdot \bar{\mathbf{x}}) \\
 &= \bar{\mathbf{x}} + (\mathbf{C}_{\mathbf{x}}^{-1} + \mathbf{M}^T \cdot \mathbf{C}_{\mathbf{y}|\mathbf{x}}^{-1} \cdot \mathbf{M})^{-1} \cdot \mathbf{M}^T \cdot \mathbf{C}_{\mathbf{y}|\mathbf{x}}^{-1} \cdot (\mathbf{y} - \mathbf{M} \cdot \bar{\mathbf{x}}) \\
 &= \bar{\mathbf{x}} + \mathbf{C}_{\mathbf{x}} \cdot \mathbf{M}^T (\mathbf{M} \cdot \mathbf{C}_{\mathbf{x}} \cdot \mathbf{M}^T + \mathbf{C}_{\mathbf{y}|\mathbf{x}})^{-1} \cdot (\mathbf{y} - \mathbf{M} \cdot \bar{\mathbf{x}}) \\
 &= \mathbf{x}_{\text{MV}} \qquad \qquad \qquad (\text{rappel : } \mathbf{C}_e = \mathbf{C}_{\mathbf{y}|\mathbf{x}}) \\
 &= \arg \min_{\tilde{\mathbf{x}}} \langle \|\mathbf{x} - \tilde{\mathbf{x}}\|^2 \rangle
 \end{aligned}$$

- ▶ les propriétés de MV s'appliquent à MAP
- mais aussi : probabilités, espérances de \neq quantités, ...

Marginalisation

Probabilité des observables

$$\begin{aligned}
 \text{PDF}(\mathbf{y}) &= \int \text{PDF}(\mathbf{y} | \mathbf{x}) \text{PDF}(\mathbf{x}) d\mathbf{x} \\
 &= \frac{\exp \left\{ -\frac{1}{2} (\mathbf{y} - \bar{\mathbf{y}})^T \cdot \mathbf{C}_{\mathbf{y}}^{-1} \cdot (\mathbf{y} - \bar{\mathbf{y}}) \right\}}{(2\pi)^{N_{\mathbf{y}}/2} |\mathbf{C}_{\mathbf{y}}|^{\frac{1}{2}}}
 \end{aligned}$$

avec :

$$\begin{aligned}
 \bar{\mathbf{y}} &= \langle \mathbf{y} \rangle_{\mathbf{x}, e} = \mathbf{M} \cdot \bar{\mathbf{x}} \\
 \mathbf{C}_{\mathbf{y}} &= \mathbf{C}_{\mathbf{y}|\mathbf{x}} + \mathbf{M} \cdot \mathbf{C}_{\mathbf{x}} \cdot \mathbf{M}^T
 \end{aligned}$$

Probabilités

Probabilité *a posteriori* des paramètres

$$\begin{aligned} \text{PDF}(\mathbf{x} | \mathbf{y}) &= \frac{\text{PDF}(\mathbf{y} | \mathbf{x}) \text{PDF}(\mathbf{x})}{\int \text{PDF}(\mathbf{y} | \mathbf{x}') \text{PDF}(\mathbf{x}') d\mathbf{x}'} \\ &= \frac{\exp \left\{ -\frac{1}{2} (\mathbf{x} - \mathbf{x}_{\text{MAP}}(\mathbf{y}))^T \cdot \mathbf{C}_{\mathbf{x}|\mathbf{y}}^{-1} \cdot (\mathbf{x} - \mathbf{x}_{\text{MAP}}(\mathbf{y})) \right\}}{(2\pi)^{N_x/2} |\mathbf{C}_{\mathbf{x}|\mathbf{y}}|^{\frac{1}{2}}} \end{aligned}$$

avec covariance *a posteriori* :

$$\begin{aligned} \mathbf{C}_{\mathbf{x}|\mathbf{y}} &= (\mathbf{C}_{\mathbf{x}}^{-1} + \mathbf{M}^T \cdot \mathbf{C}_{\mathbf{y}|\mathbf{x}}^{-1} \cdot \mathbf{M})^{-1} \\ &\leq \mathbf{C}_{\mathbf{x}} \end{aligned}$$

Contrôle par la régularisation du nombre de paramètres indépendants

Nombre de paramètres résolus par les données

- espérance des termes de log-probabilité :

$$\begin{aligned}
 \langle f_{\text{prior}}(\mathbf{x}_{\text{MAP}}) \rangle &= \text{tr}(\mathbf{M} \cdot \mathbf{C}_{\mathbf{x}} \cdot \mathbf{M}^T \cdot \mathbf{C}_{\mathbf{y}}^{-1}) \\
 &= \text{tr}(\mathbf{M} \cdot \mathbf{S}_{\text{MAP}}) = \text{tr}(\mathbf{S}_{\text{MAP}} \cdot \mathbf{M}) \\
 &= \text{tr}(\mathbf{R}_{\text{MAP}}) \\
 \langle f_{\text{data}}(\mathbf{x}_{\text{MAP}}) \rangle &= \text{tr}(\mathbf{C}_{\mathbf{y}|\mathbf{x}} \cdot \mathbf{C}_{\mathbf{y}}^{-1}) \\
 &= N_{\mathbf{y}} - \text{tr}(\mathbf{R}_{\text{MAP}}) \\
 \langle f_{\text{post}}(\mathbf{x}_{\text{MAP}}) \rangle &= \langle f_{\text{data}}(\mathbf{x}_{\text{MAP}}) \rangle + \langle f_{\text{prior}}(\mathbf{x}_{\text{MAP}}) \rangle \\
 &= N_{\mathbf{y}}
 \end{aligned}$$

- le nombre de paramètres **résolus** par les données est la trace de l'opérateur de résolution :

$$N_{\text{EDF}} = \langle f_{\text{prior}}(\mathbf{x}_{\text{MAP}}) \rangle = \text{tr}(\mathbf{R}_{\text{MAP}})$$

Résolution et propagation des erreurs

- solution *linéaire* :

$$\begin{aligned}\tilde{\mathbf{x}} &= \bar{\mathbf{x}} + \mathbf{S} \cdot (\mathbf{y} - \mathbf{M} \cdot \bar{\mathbf{x}}) \\ &= \bar{\mathbf{x}} + \mathbf{R} \cdot (\mathbf{x} - \bar{\mathbf{x}}) + \mathbf{S} \cdot \mathbf{e}\end{aligned}$$

- ▶ résolution : $\mathbf{R} = \mathbf{S} \cdot \mathbf{M}$;
- ▶ filtrage des erreurs (bruit) par \mathbf{S} ;
- ▶ propagation des erreurs dans la solution :

$$\begin{aligned}\langle \tilde{\mathbf{x}} \rangle_e &= \bar{\mathbf{x}} + \mathbf{R} \cdot (\mathbf{x} - \bar{\mathbf{x}}) \\ \mathbf{E} &\stackrel{\text{def}}{=} \langle (\mathbf{S} \cdot \mathbf{e}) \cdot (\mathbf{S} \cdot \mathbf{e})^T \rangle_e = \mathbf{S} \cdot \mathbf{C}_e \cdot \mathbf{S}^T\end{aligned}$$

Comparaison ML et MAP (ou MV)

Observables

$$\mathbf{y} = \mathbf{M} \cdot \mathbf{x} + \mathbf{e}$$

Solution

$$\mathbf{x}_{\text{ML}} = \mathbf{S}_{\text{ML}} \cdot \mathbf{y}$$

$$\mathbf{x}_{\text{MAP}} = \bar{\mathbf{x}} + \mathbf{S}_{\text{MAP}} \cdot (\mathbf{y} - \mathbf{M} \cdot \bar{\mathbf{x}})$$

Opérateur d'estimation

$$\begin{aligned} \mathbf{S}_{\text{ML}} &= (\mathbf{M}^T \cdot \mathbf{C}_{\mathbf{y}|\mathbf{x}}^{-1} \cdot \mathbf{M})^{-1} \cdot \mathbf{M}^T \cdot \mathbf{C}_{\mathbf{y}|\mathbf{x}}^{-1} \\ &= \mathbf{M}^{-1} \quad (\text{si } \mathbf{M} \text{ inversible}) \end{aligned}$$

$$\begin{aligned} \mathbf{S}_{\text{MAP}} &= \mathbf{C}_{\mathbf{x}} \cdot \mathbf{M}^T \cdot (\mathbf{M} \cdot \mathbf{C}_{\mathbf{x}} \cdot \mathbf{M}^T + \mathbf{C}_{\mathbf{y}|\mathbf{x}})^{-1} \\ &= (\mathbf{C}_{\mathbf{x}}^{-1} + \mathbf{M}^T \cdot \mathbf{C}_{\mathbf{y}|\mathbf{x}}^{-1} \cdot \mathbf{M})^{-1} \cdot \mathbf{M}^T \cdot \mathbf{C}_{\mathbf{y}|\mathbf{x}}^{-1} \end{aligned}$$

Comparaison ML et MAP (ou MV)

Résolution

$$\begin{aligned}\mathbf{R}_{\text{ML}} &= \mathbf{I} \\ \mathbf{R}_{\text{MAP}} &= \mathbf{I} - \mathbf{C}_{x|y} \cdot \mathbf{C}_x^{-1} \\ \mathbf{C}_{x|y} &= (\mathbf{C}_x^{-1} + \mathbf{M}^T \cdot \mathbf{C}_{y|x}^{-1} \cdot \mathbf{M})^{-1}\end{aligned}$$

Propagation des erreurs

$$\begin{aligned}\mathbf{E}_{\text{ML}} &= (\mathbf{M}^T \cdot \mathbf{C}_{y|x}^{-1} \cdot \mathbf{M})^{-1} \\ \mathbf{E}_{\text{MAP}} &= \mathbf{E}_{\text{ML}} \cdot (\mathbf{C}_x^{-1} \cdot \mathbf{E}_{\text{ML}} + \mathbf{I})^{-2} \\ &= (\mathbf{E}_{\text{ML}} \cdot \mathbf{C}_x^{-1} + \mathbf{I})^{-2} \cdot \mathbf{E}_{\text{ML}} \\ &\leq \mathbf{E}_{\text{ML}}\end{aligned}$$

Régularisations quadratiques

- régularisation de Tikhonov ($\mu > 0$) :

$$f_{\text{prior}}(\mathbf{x}) = \mu \|\mathbf{x}\|^2 = \mu \mathbf{x}^T \cdot \mathbf{x}$$

- contrainte de lissage :

$$f_{\text{prior}}(\mathbf{x}) = \mu \sum_{i,j} [x(i+1, j) - x(i, j)]^2 + \mu \sum_{i,j} [x(i, j+1) - x(i, j)]^2$$

- corrélation \sim a priori Gaussien (Tarantola-Valette) :

$$f_{\text{prior}}(\mathbf{x}) = (\mathbf{x} - \mathbf{p})^T \cdot \text{Cov}(\mathbf{x})^{-1} \cdot (\mathbf{x} - \mathbf{p})$$

- régularisation de Tikhonov généralisée :

$$f_{\text{prior}}(\mathbf{x}) = \|\mathbf{D} \cdot \mathbf{x} - \mathbf{p}\|_{\mathbf{Q}}^2 = (\mathbf{D} \cdot \mathbf{x} - \mathbf{p})^T \cdot \mathbf{Q} \cdot (\mathbf{D} \cdot \mathbf{x} - \mathbf{p})$$

\implies solution analytique si linéaire ;

Régularisations non-quadratiques

- maximum d'entropie (avec a priori p):

$$f_{\text{prior}}(\mathbf{x}|\mathbf{p}) = \mu \sum_i \left[p_i - x_i + x_i \log(x_i/p_i) \right]$$

- lissage avec norme $\ell_1 - \ell_2$:

$$f_{\text{prior}}(\mathbf{x}) = \mu \sum_i \mathcal{N}_{1-2}(x_{i+1} - x_i)$$

$$\mathcal{N}_{1-2}(\Delta x) = 2\varepsilon^2 \left[|\Delta x/\varepsilon| - \log(1 + |\Delta x/\varepsilon|) \right] \begin{array}{l} \underset{|\Delta x| \ll \varepsilon}{\simeq} \Delta x^2 \\ \underset{|\Delta x| \gg \varepsilon}{\simeq} 2\varepsilon |\Delta x| \end{array}$$

⇒ pas de solution analytique (en général)

Résultats en déconvolution

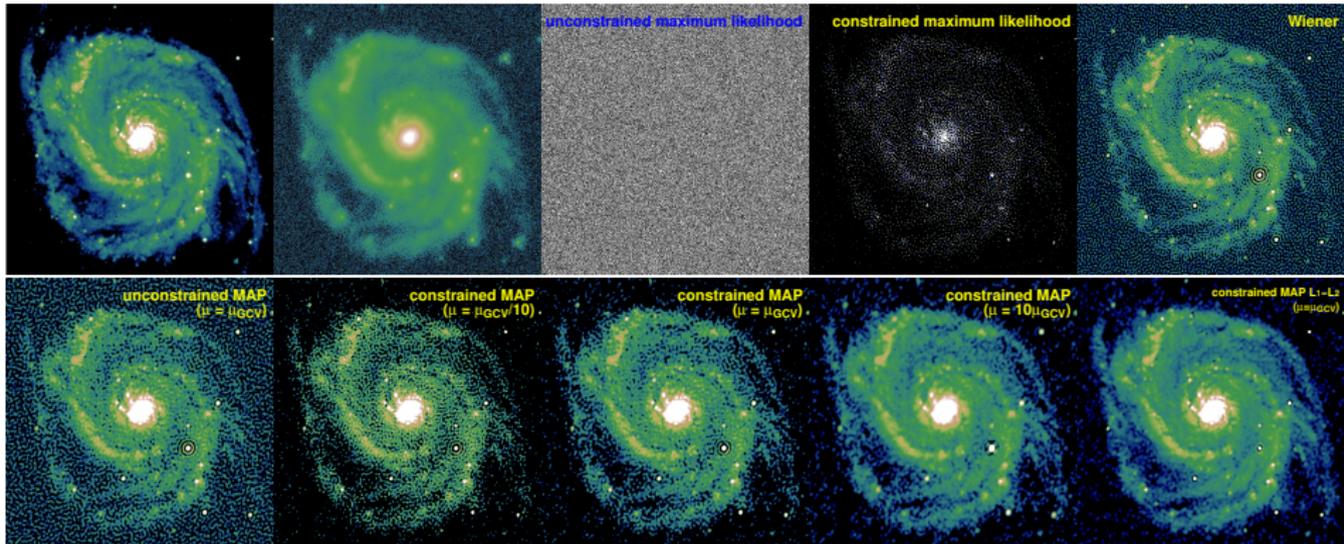
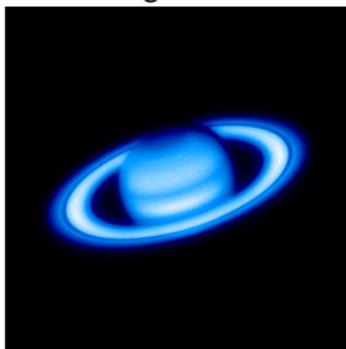
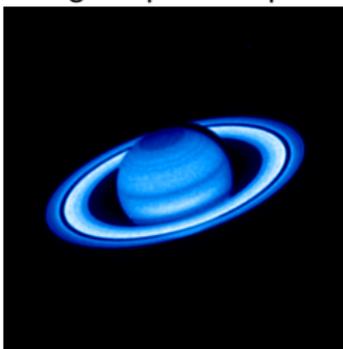
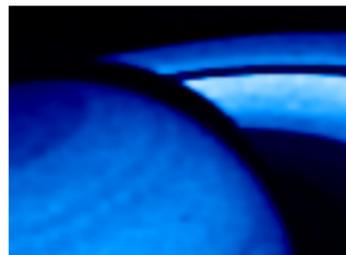
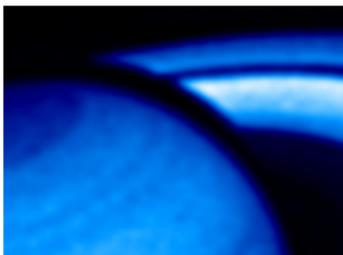
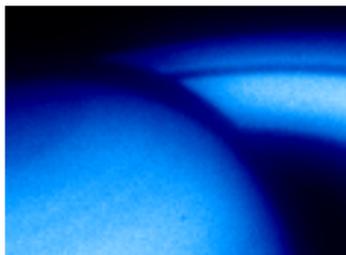
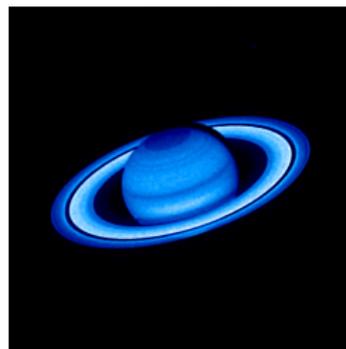


Image HST de Saturne

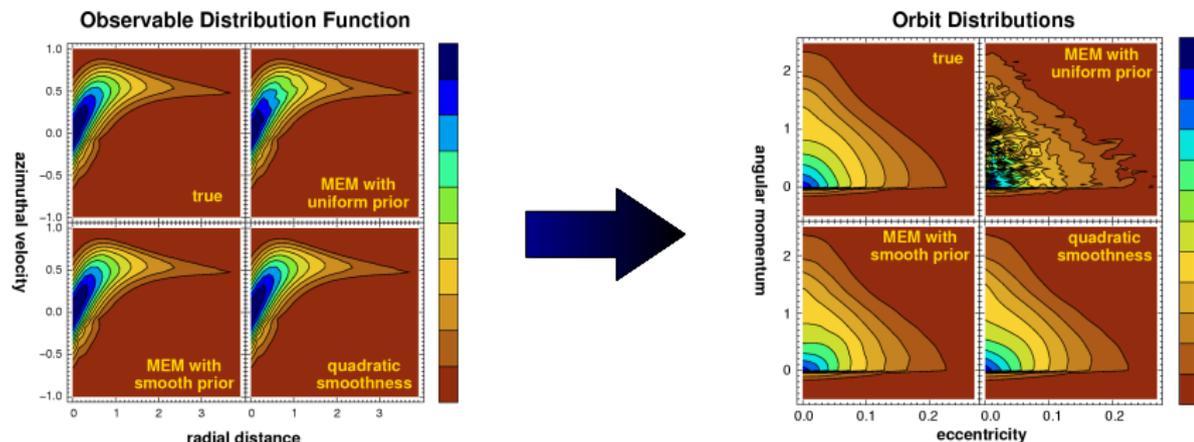
image brute



régul. quadratique

 $l_1 - l_2$ 

Fonction de distribution, disque minces



données : $F_\phi(R, v_\phi)$

paramètres : $f(\varepsilon, R v_\phi)$

$$\text{modèle : } F_\phi(R, v_\phi) = \sqrt{2} \int_0^{\frac{1}{2} v_\phi^2 - \psi(R)} \frac{f(\varepsilon, R v_\phi)}{\sqrt{\varepsilon + \psi(R) - \frac{1}{2} v_\phi^2}} d\varepsilon$$

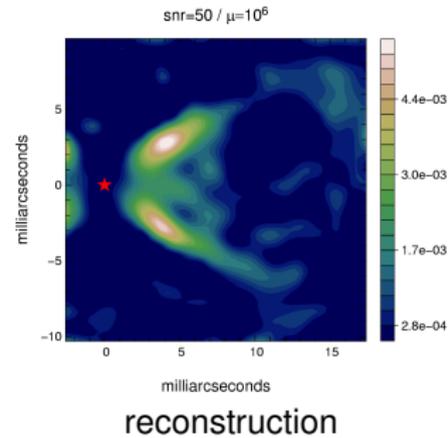
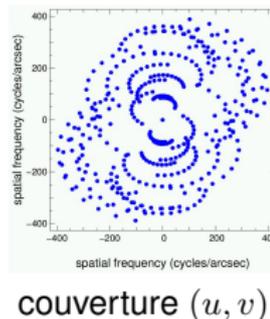
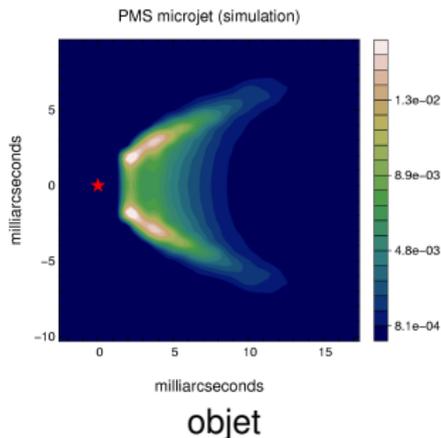
(Pichon & Thiébaud, 1998)

Interférométrie optique

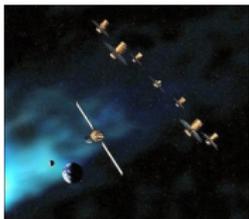
- mesures = spectre de puissance et clôtures de phases (modèle non linéaire)
- peu de mesures et couverture du plan (u, v) inhomogène
- déficit de mesures de phase de Fourier

JMMC

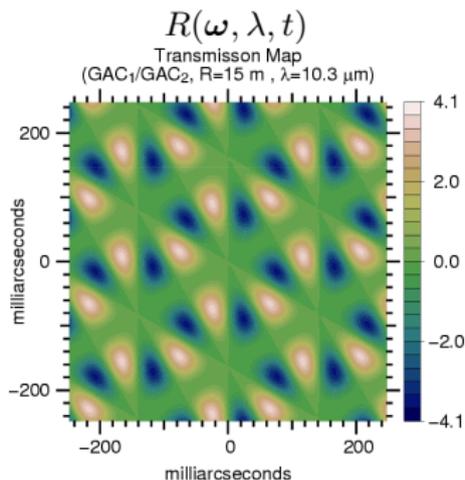
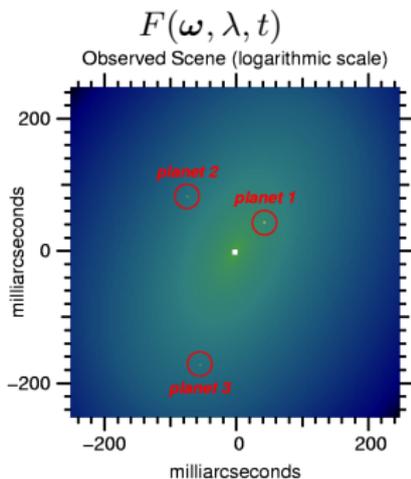
*Multi-aperture Image
Reconstruction
Algorithm*



Darwin : interférométrie à frange noire

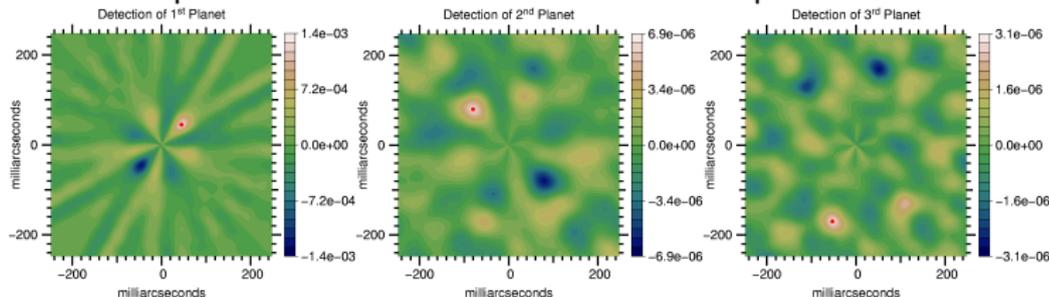


$$A(\lambda, t) = \int R(\omega, \lambda, t) F(\omega, \lambda, t) d\omega$$

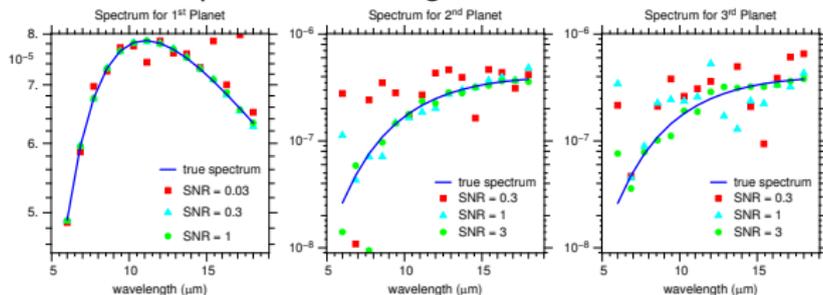


Darwin : détection et caractérisation des planètes

- détection à partir des “cartes” de vraisemblance a posteriori :



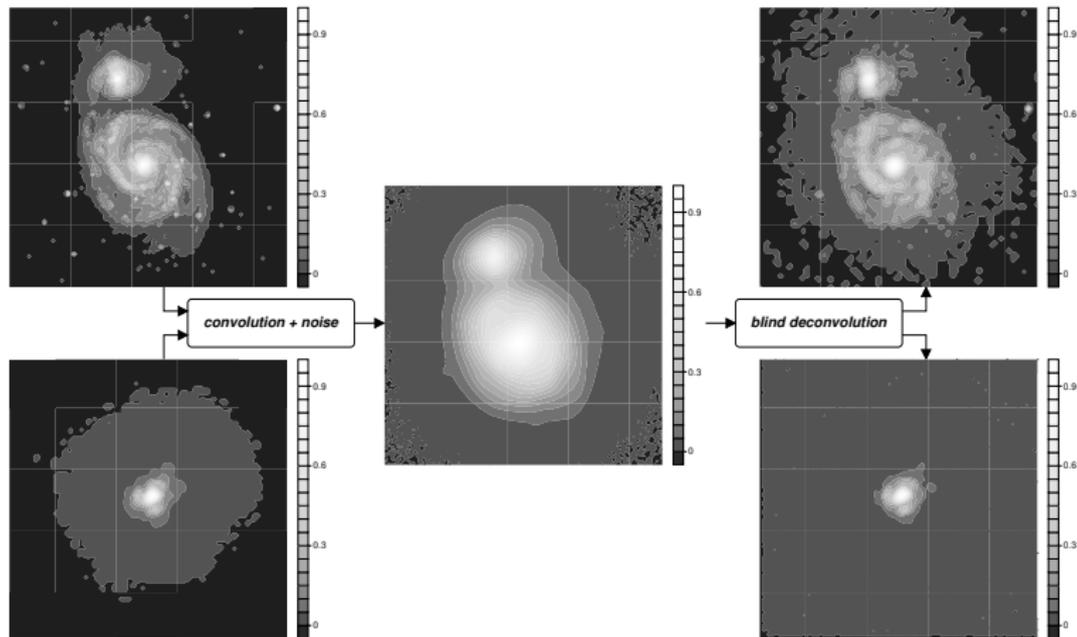
- distribution spectrale d'énergie estimée :



- collaboration avec Laurent Mugnier (ONERA)

Déconvolution aveugle

Que faire si la PSF est mal connue voire inconnue ?



Déconvolution aveugle appliquée en biologie

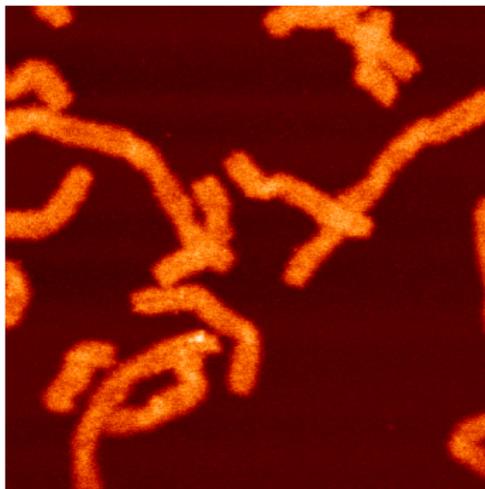


Image de chromosomes par microscopie confocale. Source : Jean-Claude Bernengo (Centre Commun de Quantimétrie, Université Claude Bernard, Lyon, France). Reconstruction par logiciel **MAAD**.

Déconvolution aveugle appliquée en biologie

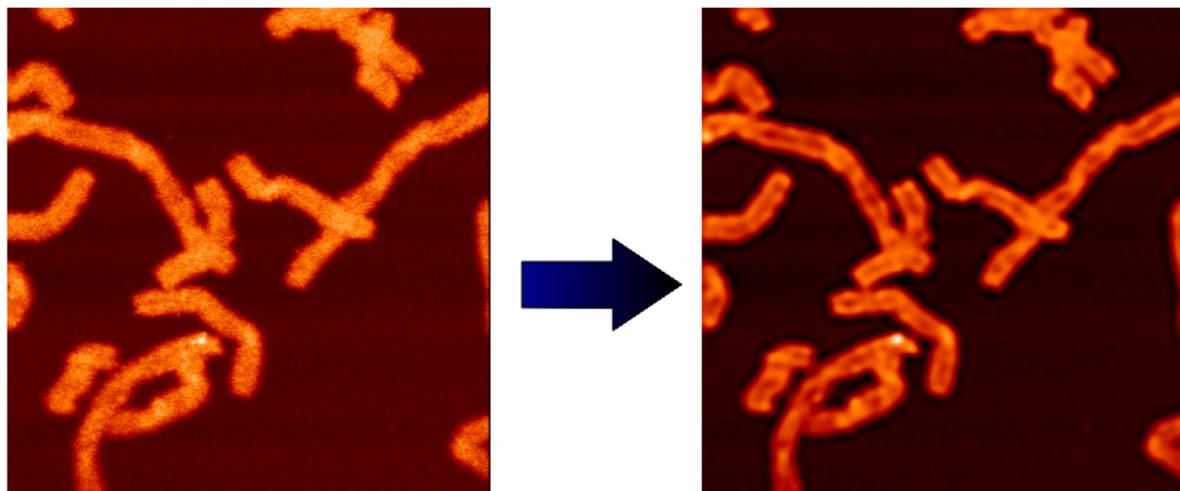
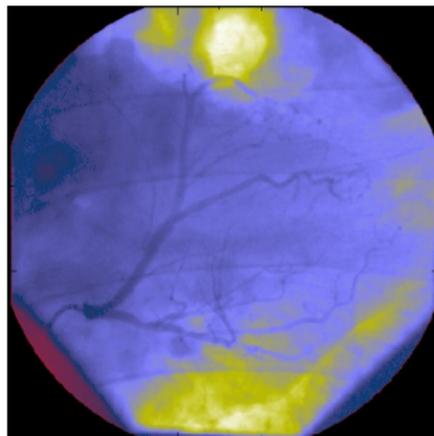


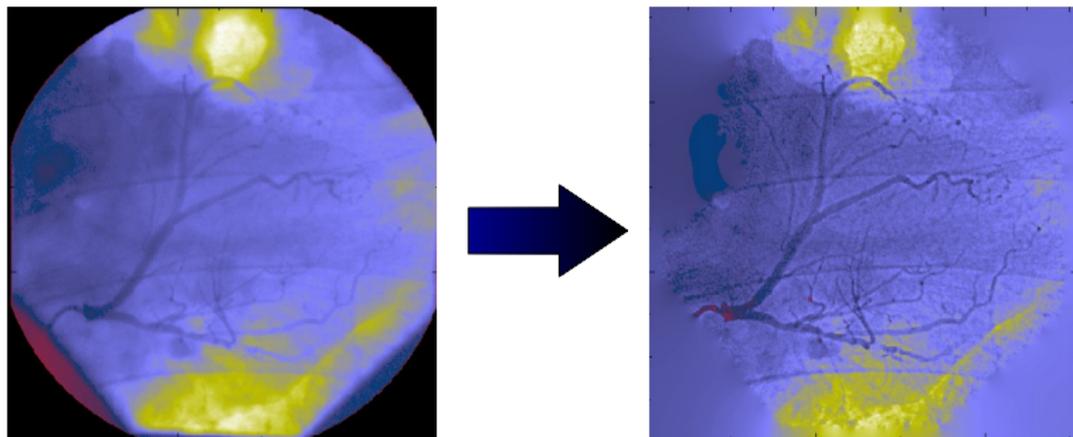
Image de chromosomes par microscopie confocale. Source : Jean-Claude Bernengo (Centre Commun de Quantimétrie, Université Claude Bernard, Lyon, France). Reconstruction par logiciel **MAAD**.

Déconvolution aveugle appliquée en médecine



Vidéo coronarographique obtenue à l'hôpital de la Croix-Rousse (Lyon).
Traitement : **Ferréol Soulez**.

Déconvolution aveugle appliquée en médecine



Vidéo coronarographique obtenue à l'hôpital de la Croix-Rousse (Lyon).
Traitement : **Ferréol Soulez**.

Annexes

Annexe : Dérivées par rapport à une matrice

définition :

$$\left(\frac{\partial \varepsilon}{\partial \mathbf{A}} \right)_{i,j} = \frac{\partial \varepsilon}{\partial A_{i,j}} \quad \text{avec} \quad \varepsilon \in \mathbb{R} \quad \text{et} \quad \mathbf{A} \in \mathbb{R}^{m \times n}$$

par exemple :

$$\begin{aligned} \left(\frac{\partial \|\mathbf{A} \cdot \mathbf{x} - \mathbf{b}\|^2}{\partial \mathbf{A}} \right)_{i,j} &= \frac{\partial \sum_k (\sum_\ell A_{k,\ell} x_\ell - b_k)^2}{\partial A_{i,j}} \\ &= 2 \left(\sum_\ell A_{i,\ell} x_\ell - b_i \right) x_j \end{aligned}$$

d'où :

$$\frac{\partial \|\mathbf{A} \cdot \mathbf{x} - \mathbf{b}\|^2}{\partial \mathbf{A}} = 2 (\mathbf{A} \cdot \mathbf{x} - \mathbf{b}) \cdot \mathbf{x}^T$$

Annexe : Dérivées par rapport à une matrice

$$\begin{aligned} & \frac{\partial(\mathbf{A} \cdot \mathbf{x} - \mathbf{b})^T \cdot \mathbf{W} \cdot (\mathbf{A} \cdot \mathbf{x} - \mathbf{b})}{\partial A_{i,j}} \\ &= \frac{\partial \sum_{k,\ell} (\sum_m A_{k,m} x_m - b_k) W_{k,\ell} (\sum_n A_{\ell,n} x_n - b_\ell)}{\partial A_{i,j}} \\ &= \sum_\ell x_j W_{i,\ell} \left(\sum_n A_{\ell,n} x_n - b_\ell \right) + \sum_k \left(\sum_m A_{k,m} x_m - b_k \right) W_{k,i} x_j \end{aligned}$$

d'où :

$$\boxed{\frac{\partial(\mathbf{A} \cdot \mathbf{x} - \mathbf{b})^T \cdot \mathbf{W} \cdot (\mathbf{A} \cdot \mathbf{x} - \mathbf{b})}{\partial \mathbf{A}} = (\mathbf{W} + \mathbf{W}^T) \cdot (\mathbf{A} \cdot \mathbf{x} - \mathbf{b}) \cdot \mathbf{x}^T}$$

Annexe : Formules de Sherman-Morrison-Woodbury

- pour des matrices $\mathbf{A} \in \mathbb{R}^{m \times m}$ et $\mathbf{B} \in \mathbb{R}^{n \times n}$ **inversibles** :

$$\left(\mathbf{A} + \mathbf{C} \cdot \mathbf{B} \cdot \mathbf{C}^T \right)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \cdot \mathbf{C} \cdot \left(\mathbf{C}^T \cdot \mathbf{A}^{-1} \cdot \mathbf{C} + \mathbf{B}^{-1} \right)^{-1} \cdot \mathbf{C}^T \cdot \mathbf{A}^{-1}$$

où $\mathbf{C} \in \mathbb{R}^{m \times n}$;

- si \mathbf{A} et \mathbf{B} sont **positives définies**, alors :

$$\mathbf{B} \cdot \mathbf{C}^T \cdot \left(\mathbf{A} + \mathbf{C} \cdot \mathbf{B} \cdot \mathbf{C}^T \right)^{-1} = \left(\mathbf{C}^T \cdot \mathbf{A}^{-1} \cdot \mathbf{C} + \mathbf{B}^{-1} \right)^{-1} \cdot \mathbf{C}^T \cdot \mathbf{A}^{-1}$$

Annexe : Espérance de la log-vraisemblance

facile à démontrer :

$$\text{tr}(\mathbf{W} \cdot \mathbf{v} \cdot \mathbf{u}^T) = \mathbf{u}^T \cdot \mathbf{W} \cdot \mathbf{v}$$

d'où il vient :

$$\langle [\mathbf{y} - \mathbf{m}(\mathbf{x})]^T \cdot \mathbf{W} \cdot [\mathbf{y} - \mathbf{m}(\mathbf{x})] \rangle = \text{tr}(\mathbf{W} \cdot \langle [\mathbf{y} - \mathbf{m}(\mathbf{x})] \cdot [\mathbf{y} - \mathbf{m}(\mathbf{x})]^T \rangle)$$

Annexe : Propagation des erreurs

$$\begin{aligned}
 \mathbf{E}_{\text{ML}} &\stackrel{\text{def}}{=} \langle (\mathbf{S}_{\text{ML}} \cdot \mathbf{e}) \cdot (\mathbf{S}_{\text{ML}} \cdot \mathbf{e})^{\text{T}} \rangle_{\mathbf{e}} = \mathbf{S}_{\text{ML}} \cdot \mathbf{C}_{\mathbf{e}} \cdot \mathbf{S}_{\text{ML}}^{\text{T}} \\
 &= \left(\mathbf{C}_{\mathbf{x}}^{-1} + \mathbf{M}^{\text{T}} \cdot \mathbf{C}_{\mathbf{e}}^{-1} \cdot \mathbf{M} \right)^{-1} \\
 \mathbf{E}_{\text{MAP}} &\stackrel{\text{def}}{=} \langle (\mathbf{S}_{\text{MAP}} \cdot \mathbf{e}) \cdot (\mathbf{S}_{\text{MAP}} \cdot \mathbf{e})^{\text{T}} \rangle_{\mathbf{e}} = \mathbf{S}_{\text{MAP}} \cdot \mathbf{C}_{\mathbf{e}} \cdot \mathbf{S}_{\text{MAP}}^{\text{T}} \\
 &= \left(\mathbf{C}_{\mathbf{x}}^{-1} + \mathbf{E}_{\text{ML}}^{-1} \right)^{-1} \cdot \mathbf{E}_{\text{ML}}^{-1} \cdot \left(\mathbf{C}_{\mathbf{x}}^{-1} + \mathbf{E}_{\text{ML}}^{-1} \right)^{-1}
 \end{aligned}$$

or :

$$\begin{aligned}
 \left(\mathbf{C}_{\mathbf{x}}^{-1} + \mathbf{E}_{\text{ML}}^{-1} \right)^{-1} &= \mathbf{E}_{\text{ML}} \cdot \left(\mathbf{C}_{\mathbf{x}}^{-1} \cdot \mathbf{E}_{\text{ML}} + \mathbf{I} \right)^{-1} \\
 &= \left(\mathbf{E}_{\text{ML}} \cdot \mathbf{C}_{\mathbf{x}}^{-1} + \mathbf{I} \right)^{-1} \cdot \mathbf{E}_{\text{ML}}
 \end{aligned}$$

d'où :

$$\begin{aligned}
 \mathbf{E}_{\text{MAP}} &= \mathbf{E}_{\text{ML}} \cdot \left(\mathbf{C}_{\mathbf{x}}^{-1} \cdot \mathbf{E}_{\text{ML}} + \mathbf{I} \right)^{-2} \\
 &= \left(\mathbf{E}_{\text{ML}} \cdot \mathbf{C}_{\mathbf{x}}^{-1} + \mathbf{I} \right)^{-2} \cdot \mathbf{E}_{\text{ML}}
 \end{aligned}$$

donc :

$$\mathbf{E}_{\text{MAP}} \leq \mathbf{E}_{\text{ML}}$$