Chapitre 11

Choix des hyperparamètres

Dans ce chapitre nous aborderons le problème de l'estimation des hyperparamètres, et en particulier, celui de la détermination du paramètre de la régularisation. Dans un prmier temps, nous décrirons brièvement les méthodes classiques du type : adéquation au données, la courbe en L, la validation croisée, et la validation croisée généralisée. Ensuite, nous présenterons le problème général de l'estimation des hyperparamètres dans le cadre bayésien.

11.1 Position du problème

Dans les chapitres précédents, nous avons vu que la résolution d'un problème inverse ne peut se faire d'une manière satisfaisante que qu'avec l'apport d'une information *a priori* appropriée. L'idée de combiner l'information contenue dans les données avec celle de l'*a priori* est à la base de la notion de la régularisation.

Par exemple dans l'approche classique de la régularisation (au sens de Phillips, Towmay et Tikhonov), nous avons vu que la solution est définie par :

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \left\{ \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + \lambda \, \|\boldsymbol{D}\boldsymbol{f}\|^2 \right\}$$

où λ est le paramètre de régularisation qui a pour role de faire une compromis entre la fidélité aux données et l'information *a priori* qui est ici la variation douce de la grandeur inconnue.

D'une manière plus générale on peut définir une solution régularisée par :

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \left\{ \Delta_1(\boldsymbol{g}, \boldsymbol{H}\boldsymbol{f}) + \lambda \Delta_2(\boldsymbol{f}, \widehat{\boldsymbol{f}}_\infty) \right\}$$

où Δ_1 et Δ_2 sont deux distances, la première dans l'espace des mesures $\boldsymbol{g}$ et la deuxième dans l'espace des inconnues $\boldsymbol{f}$.

Notant, la solution pour $\lambda = 0$ par $\widehat{\boldsymbol{f}}_0$ et la solution pour $\lambda = \infty$ par $\widehat{\boldsymbol{f}}_\infty$ on peut encore généraliser (ou reécrire la définition de la solution régularisée par :

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \left\{ \Delta_1(\boldsymbol{f}, \widehat{\boldsymbol{f}}_0) + \lambda \Delta_2(\boldsymbol{f}, \widehat{\boldsymbol{f}}_\infty) \right\}$$

où Δ_1 et Δ_2 sont deux distances euclidiennes, $\widehat{\boldsymbol{f}}_\infty$ est la solution que l'on obtient lorsque $\lambda \longrightarrow \infty$ (la solution a priori) et $\widehat{\boldsymbol{f}}_0$ est la solution que l'on obtient lorsque $\lambda \longrightarrow 0$ (la solution inverse généralisée par exemple).

Arrivé à ce stade deux questions se sonts posées :

- Régulariser de quelle façon ou comment choisir Δ_1 et Δ_2 ?
- Régulariser jusqu'à quel point ou comment choisir le paramètre de la régularisation ?

Les figures de la page suivante montrent, sur un exemple du problème de déconvolution, l'effet du paramètre λ sur la solution.

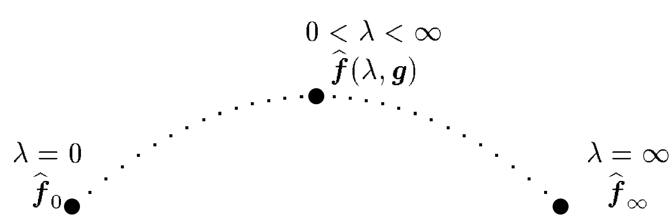

FIG. 11.1 - *Choix du paramètre de régularisation*

11.2.1 Adéquation aux données

Le problème de la minimisation d'un critère composite

$$\min_{\boldsymbol{f}} \Delta_1(\boldsymbol{g}, \boldsymbol{H}\boldsymbol{f}) + \lambda \Delta_2(\boldsymbol{f}, \widehat{\boldsymbol{f}}_\infty)$$

peut être interprété comme un problème d'optimisation sous contrainte

$$\min_{\boldsymbol{f}} \Delta_2(\boldsymbol{f}, \widehat{\boldsymbol{f}}_\infty) \quad \text{s.c.} \quad \Delta_1(\boldsymbol{g}, \boldsymbol{H}\boldsymbol{f}) \leq c_1$$

ou encore

$$\min_{\boldsymbol{f}} \Delta_1(\boldsymbol{g}, \boldsymbol{H}\boldsymbol{f}) \quad \text{s.c.} \quad \Delta_2(\boldsymbol{f}, \widehat{\boldsymbol{f}}_\infty) \leq c_2$$

où c_1 et c_2 sont deux constantes qu'il faudra déterminer.

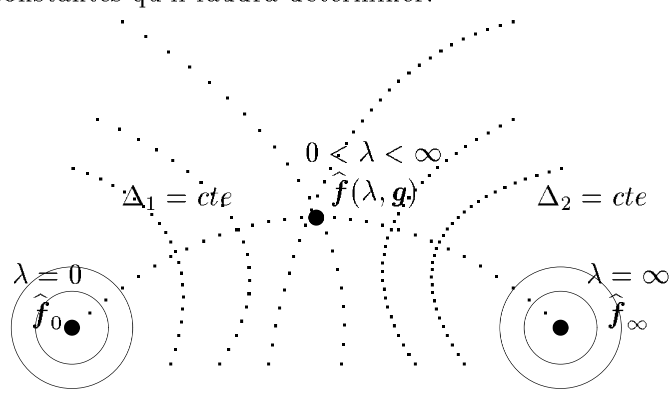

FIG. 11.4 - *Choix du paramètre de régularisation par adéquation aux données*

Pour mieu voir la limitation de cette approche, considérons le deuxième cas. c_2 peut alors être considéré comme une statistique dont sa loi découle de la loi du bruit ou d'une façon plus générale de la loi $p(\boldsymbol{g}|\boldsymbol{f})$.

Supposons par exemple que $b_i \quad v.a.i.\mathcal{N}(0, \sigma_b^2)$. Alors Δ_1 devient une distance quadratique. Supposons aussi que Δ_2 est choisi quadratqiue et égale à $\|\boldsymbol{D}\boldsymbol{f}\|^2$. On a alors :

$$\widehat{\boldsymbol{f}}(\lambda) = \arg\min_{\boldsymbol{f}} \left\{ \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + \lambda \|\boldsymbol{D}\boldsymbol{f}\|^2 \right\} = \left[\boldsymbol{H}^t \boldsymbol{H} + \lambda \boldsymbol{D}^t \boldsymbol{D} \right]^{-1} \boldsymbol{H}^t \boldsymbol{g}$$

$$\mathrm{RSS}(\lambda) = \left[\boldsymbol{g} - \boldsymbol{H}\widehat{\boldsymbol{f}}(\lambda) \right]^t \left[\boldsymbol{g} - \boldsymbol{H}\widehat{\boldsymbol{f}}(\lambda) \right] = \left\| \boldsymbol{g} - \boldsymbol{H}\widehat{\boldsymbol{f}}(\lambda) \right\|^2$$

$$c \ : \quad \chi^2(N) \quad \longrightarrow c = N \text{ ou } N + 2\sqrt{N}$$

$$\mathrm{RSS}(\lambda) = N\sigma_b^2 \quad \longrightarrow \quad \lambda_\chi$$

- Il faut connaître σ_b^2
- $\left[\boldsymbol{g} - \boldsymbol{H}\widehat{\boldsymbol{f}}(\lambda) \right] \neq [\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}]$
- sur-régularisation

11.2.2 Risque moyen

$$\mathrm{RM} \stackrel{\text{def}}{=} \mathrm{E}\left[\|\boldsymbol{f}-\widehat{\boldsymbol{f}}(\lambda,\boldsymbol{g})\|^2\right] = \int \left[\boldsymbol{f}-\widehat{\boldsymbol{f}}(\lambda,\boldsymbol{g})\right]^2 p(\boldsymbol{g}|\boldsymbol{f},\lambda)\,\mathrm{d}\boldsymbol{g}$$

$$\widehat{\boldsymbol{f}}(\lambda) = \left[\boldsymbol{H}^t\boldsymbol{H}+\lambda\boldsymbol{D}^t\boldsymbol{D}\right]^{-1}\boldsymbol{H}^t\boldsymbol{g}$$

$$\begin{aligned}\hat{\lambda} &= \arg\min_{\lambda}\left\{\mathrm{E}\left[\|\boldsymbol{f}-\widehat{\boldsymbol{f}}(\lambda)\|^2\right]\right\}\\ &= \|[\boldsymbol{I}-\boldsymbol{A}(\lambda)]\boldsymbol{H}\boldsymbol{f}\|^2+\sigma_b^2\mathrm{Tr}\left\{\boldsymbol{A}(\lambda)\right\}^2\end{aligned}$$

où

$$\boldsymbol{A}(\lambda) = \boldsymbol{H}\left[\boldsymbol{H}^t\boldsymbol{H}+\lambda\boldsymbol{D}^t\boldsymbol{D}\right]^{-1}\boldsymbol{H}^t$$

et si $\boldsymbol{H}$ n'est pas singulière

$$\boldsymbol{A}(\lambda) = [\boldsymbol{I}-\lambda\boldsymbol{Q}]^{-1} \quad \text{avec} \quad \boldsymbol{Q} = [\boldsymbol{H}^t]^{-1}\boldsymbol{D}^t\boldsymbol{D}\boldsymbol{H}^{-1}$$

$$\lambda_{\mathrm{RM}} = \arg\min_{\lambda}\left\{\mathrm{RM}(\lambda,\boldsymbol{f})\right\}$$

- Il faut connaître σ_b^2
- $\mathrm{RM}(\lambda,\boldsymbol{f})$ dépend du vrai $\boldsymbol{f}$ qui est inconnu

11.2.3 Validation croisée

L'idée de base dans les techniques de validations croisées est de partionner les données en deux sous ensembles. On utiliserais alors les données $\boldsymbol{g}_1$ dans le première sous ensemble pour calculer la solution $\widehat{\boldsymbol{f}}(\lambda)$ et en déduire (ou prédire) celles du deuxième ; $\widehat{\boldsymbol{g}}_2 = \boldsymbol{H}\widehat{\boldsymbol{f}}(\lambda)$. On peut alors déterminer la valeur optimale (au sens de la VC) du λ par :

$$\widehat{\lambda} = \arg\min_{\lambda} \left\{ \mathrm{E}\left[\| \boldsymbol{g}_2 - \widehat{\boldsymbol{g}}_2(\lambda)] \|^2 \right\}$$

Une partition la plus simple possible est d'éliminer à chaque étape une seule donnée g_k, c'est à dire :

$$P_1 = \{g_1, \ldots, g_{k-1}, g_{k+1}, \ldots, g_M\}, \quad P_2 = \{g_k\}$$

Notons par $\boldsymbol{g}^{(-k)}$ $[g_1, \ldots, g_{k-1}, g_{k+1}, \ldots, g_N]$. Le critère de VC peut alors s'écrire :

$$\mathrm{VC}(\lambda) \stackrel{\text{def}}{=} \sum_{k=1}^{M} \left[g_k - \boldsymbol{h}_k^t \widehat{\boldsymbol{f}}\left(\lambda, \boldsymbol{g}^{(-k)}\right) \right]^2$$

et

$$\widehat{\lambda}_{VC} = \arg\min_{\lambda} \{\mathrm{VC}(\lambda)\}$$

Pour aller un peu plus dans le détail, considérons le cas de la régularisation quadratique :

$$\widehat{\boldsymbol{f}}\left(\lambda, \boldsymbol{g}^{(-k)}\right) = \left[\boldsymbol{H}^t\boldsymbol{H} - \boldsymbol{h}_k\boldsymbol{h}_k^t + \lambda \boldsymbol{P}_0^{-1} \right]^{-1} \left[\boldsymbol{H}^t\boldsymbol{g} - \boldsymbol{h}_k\boldsymbol{g}^{(-k)} \right]$$

$$+$$

lemme d'inversion des matrices

$$\Downarrow$$

$$\mathrm{VC}(\lambda) = [\boldsymbol{B}(\lambda)[\boldsymbol{I} - \boldsymbol{A}(\lambda)]\boldsymbol{g}]^2$$

- $\boldsymbol{B}(\lambda) = \mathrm{diag}\left\{(1 - a_{kk})^{-1}\right\}$
- Le calcul de $\mathrm{VC}(\lambda)$ se ramène à celui des éléments diagonaux de $\boldsymbol{A}(\lambda)$
- Si $\boldsymbol{A}(\lambda)$ diagonale, $\mathrm{VC}(\lambda)$ n'a pas de minimum

11.2.4 Validation croisée généralisée

$$\mathrm{VCG}(\lambda) = \frac{\mathrm{VC}(\lambda)}{\mathrm{Tr}\left\{\boldsymbol{I} - \boldsymbol{A}(\lambda)\right\}^2}$$

$$\mathrm{VCG}(\lambda) \stackrel{\text{def}}{=} \sum_{k=1}^{M} w_k(\lambda)^2 \left[g_k - \boldsymbol{h}_k^t \widehat{\boldsymbol{f}}\left(\lambda, \boldsymbol{g}^{(-k)}\right)\right]^2$$

coeffs de pondération : $w_k(\lambda) = (1 - a_{kk})(1 - \mathrm{Tr}\left\{\boldsymbol{I} - \boldsymbol{A}(\lambda)\right\})^{-1}$

$$\widehat{\lambda}_{VCG} = \arg\min_{\lambda} \left\{\mathrm{VCG}(\lambda)\right\}$$

Calcul de VCG(λ)

Ici aussi, comme dans le cas précédent, en utilisant la lemme d'inversion des matrices, il est possible d'exprimer le critère en fonctions des termes calculable d'une manière récursive :

$$\mathrm{VCG}(\lambda) = \sum_{k=1}^{M} \left(\frac{u_k}{v_k}\right)^2$$

- $u_k = g_k - \boldsymbol{h}_k^t \widehat{\boldsymbol{f}}\left(\lambda, \boldsymbol{g}^{(-k)}\right)$ résidu
- $v_k = a_{kk} - 1$
- Auxiliaire de calcul :

$$\begin{array}{rcll} \boldsymbol{P}_{M|M} & = & (\boldsymbol{H}^t \boldsymbol{H} + \lambda \boldsymbol{P}_0^t)^{-1} & \text{Covariance de } (\widehat{\boldsymbol{f}} - \boldsymbol{f}) \\ \boldsymbol{A}(\lambda) & = & \boldsymbol{H} \boldsymbol{P}_{M|M} \boldsymbol{H}^t & \end{array}$$

11.3 Un autre regard

$$\boldsymbol{f} \sim \mathcal{N}(0, \sigma_x^2 \boldsymbol{P}_0)$$

$$\boldsymbol{b} \sim \mathcal{N}(0, \sigma_b^2 \boldsymbol{I})$$

$$\boldsymbol{g}|\boldsymbol{f} \sim \mathcal{N}(\boldsymbol{H}\boldsymbol{f}, \sigma_b^2 \boldsymbol{I})$$

Problème 1 :
Étant donné $\boldsymbol{H}$, σ_x^2, $\boldsymbol{P}_0$, σ_b^2 et $\boldsymbol{g}$, estimer $\boldsymbol{f}$

Solution :

$$\boldsymbol{f}|\boldsymbol{g} \sim \mathcal{N}(\widehat{\boldsymbol{f}}, \boldsymbol{P}) \quad \left\{ \begin{array}{lcl} \widehat{\boldsymbol{f}} & = & \left(\boldsymbol{H}^t\boldsymbol{H} + \lambda \boldsymbol{P}_0^{-1}\right)^{-1} \boldsymbol{H}^t\boldsymbol{g} \\ \boldsymbol{P} & = & \sigma_b^2 \left(\boldsymbol{H}^t\boldsymbol{H} + \lambda \boldsymbol{P}_0^{-1}\right)^{-1} \end{array} \right. , \quad \lambda = \frac{\sigma_b^2}{\sigma_x^2}$$

Problème 2 :
Étant donné $\boldsymbol{H}$, $\boldsymbol{P}_0, \sigma_b^2$ et $\boldsymbol{g}$, estimer σ_x^2 et $\boldsymbol{f}$

ou d'une manière équivalente :

Étant donné $\boldsymbol{H}$, $\boldsymbol{P}_0, \sigma_b^2$, et $\boldsymbol{g}$, estimer λ et $\boldsymbol{f}$

Problème 3 :
Étant donné $\boldsymbol{H}$, $\boldsymbol{P}_0$ et $\boldsymbol{g}$, estimer σ_x^2, σ_b^2 et $\boldsymbol{f}$

ou d'une manière équivalente :

Étant donné $\boldsymbol{H}$, $\boldsymbol{P}_0$ et $\boldsymbol{g}$, estimer λ, σ_b^2 et $\boldsymbol{f}$

Notons : $\boldsymbol{\theta} = [\sigma_x^2, \sigma_b^2]$ ou $\boldsymbol{\theta} = [\lambda, \sigma_b^2]$

Adéquation aux données

$$Z = \frac{1}{\sigma_b^2}\|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 \sim \text{loi de } \chi^2(N)$$

$$\mathrm{E}\left[\|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2\right] = N\sigma_b^2$$

- Première approximation :

$$\mathrm{E}\left[\|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2\right] \approx \mathrm{RSS}(\lambda) = \|\boldsymbol{g} - \boldsymbol{H}\widehat{\boldsymbol{f}}(\lambda)\|^2$$

$$\|\boldsymbol{g} - \boldsymbol{H}\widehat{\boldsymbol{f}}(\lambda)\|^2 = N\sigma_b^2 \quad \longrightarrow \quad \widehat{\lambda} \tag{11.1}$$

- Deuxième approximation :

$$\mathrm{E}\left[\|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2\right] \approx \frac{\mathrm{RSS}(\lambda)}{\mathrm{Tr}\left\{\boldsymbol{I} - \boldsymbol{A}(\lambda)\right\}^2} = \frac{\|\boldsymbol{g} - \boldsymbol{H}\widehat{\boldsymbol{f}}(\lambda)\|^2}{\mathrm{Tr}\left\{\boldsymbol{I} - \boldsymbol{A}(\lambda)\right\}^2}$$

  où

$$\boldsymbol{A}(\lambda) = \boldsymbol{H}\left[\boldsymbol{H}^t\boldsymbol{H} + \lambda\boldsymbol{D}^t\boldsymbol{D}\right]^{-1}\boldsymbol{H}^t$$

$$\frac{\|\boldsymbol{g} - \boldsymbol{H}\widehat{\boldsymbol{f}}(\lambda)\|^2}{\mathrm{Tr}\left\{\boldsymbol{I} - \boldsymbol{A}(\lambda)\right\}^2} = N\sigma_b^2 \quad \longrightarrow \quad \widehat{\lambda} \tag{11.2}$$

- Les deux équations (11.1) et (11.2) sont non linéaires
- Dans le cadre des approximations circulantes des matrices, il est possible d'obtenir des relation explicites pour leurs solutions.

Risque moyen

$$\mathrm{E}\left[\|\boldsymbol{f} - \widehat{\boldsymbol{f}}(\boldsymbol{\theta}, \boldsymbol{g})\|^2\right] = \int \left[\boldsymbol{f} - \widehat{\boldsymbol{f}}(\boldsymbol{\theta}, \boldsymbol{g})\right]^2 p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{\theta})\,\mathrm{d}\boldsymbol{g}$$

Validation croisée

$$\mathrm{E}\left[\|\boldsymbol{g} - \boldsymbol{H}\widehat{\boldsymbol{f}}(\boldsymbol{\theta}, \boldsymbol{g})\|^2\right] = \int \left[\boldsymbol{g} - \boldsymbol{H}\widehat{\boldsymbol{f}}(\boldsymbol{\theta}, \boldsymbol{g})\right]^2 p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{\theta})\,\mathrm{d}\boldsymbol{g}$$

Validation croisée généralisée

$$\mathrm{E}\left[\|\boldsymbol{H}\boldsymbol{f} - \widehat{\boldsymbol{g}}(\boldsymbol{\theta}, \boldsymbol{g})\|^2\right] = \int \left[\boldsymbol{H}\boldsymbol{f} - \widehat{\boldsymbol{g}}(\boldsymbol{\theta}, \boldsymbol{g})\right]^2 p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{\theta})\,\mathrm{d}\boldsymbol{g}$$

Vraisemblance généralisée

$$(\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{\theta}}) = \arg\max_{(\boldsymbol{f},\boldsymbol{\theta})} \{p(\boldsymbol{f}, \boldsymbol{g}; \boldsymbol{\theta})\} = \arg\max_{(\boldsymbol{f},\boldsymbol{\theta})} \{p(\boldsymbol{g}|\boldsymbol{f}; \boldsymbol{\theta}) p(\boldsymbol{f}; \boldsymbol{\theta})\}$$

Vraisemblance marginale

$$\left\{ \begin{array}{lcl} \widehat{\boldsymbol{\theta}} & = & \arg\max_{\boldsymbol{\theta}} \{p(\boldsymbol{g}; \boldsymbol{\theta})\} = \arg\max_{\boldsymbol{\theta}} \left\{ \int p(\boldsymbol{g}|\boldsymbol{f}; \boldsymbol{\theta}) p(\boldsymbol{f}; \boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{f} \right\} \\ \widehat{\boldsymbol{f}} & = & \arg\max_{\boldsymbol{f}} \left\{ p(\boldsymbol{f}|\boldsymbol{g}; \widehat{\boldsymbol{\theta}}) \right\} \end{array} \right.$$

La principale difficultée réside dans le calcul de $p(\boldsymbol{g}; \boldsymbol{\theta})$. C'est pourquoi, en général, on utilise l'algorithme EM pour calculer la solution $\widehat{\boldsymbol{\theta}}$.

Pseudo-vraisemblance

Lorsque $\boldsymbol{\theta}$ constitue les paramètres de la loi *a priori*, $p(\boldsymbol{f}; \boldsymbol{\theta})$ et lorsque cette loi est une loi markovienne, on peut envisager d'approximer la vraisemblance par

$$p(\boldsymbol{g}; \boldsymbol{\theta}) \approx p(\boldsymbol{g}|\boldsymbol{f}; \boldsymbol{\theta}) \sum_i p\left(f_i | f_j; \boldsymbol{\theta}; j \in v_i\right)$$

Contexte complètement bayésien

$$(\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{\theta}}) = \arg\max_{(\boldsymbol{f},\boldsymbol{\theta})} \{p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g})\} = \arg\max_{(\boldsymbol{f},\boldsymbol{\theta})} \{p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{\theta}) p(\boldsymbol{f}|\boldsymbol{\theta}) p(\boldsymbol{\theta})\}$$

Ici, la principale difficulté est le choix de la loi $p(\boldsymbol{\theta})$ de telle sorte que $p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g})$ soit unimodale et que la solution puisse exister et soit unique.

11.4 Algorithme EM

(Expectation-Maximization)

Rechercher la solution à maximum de vraisemblance
Exemple :

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \{p(\boldsymbol{g};\boldsymbol{\theta})\} = \arg\max_{\boldsymbol{\theta}} \{\ln p(\boldsymbol{g};\boldsymbol{\theta})\}$$

- vraisemblance $l(\boldsymbol{\theta}) = p(\boldsymbol{g};\boldsymbol{\theta})$
- log-vraisemblance $L(\boldsymbol{\theta}) = \ln p(\boldsymbol{g};\boldsymbol{\theta})$
- Calcul de $p(\boldsymbol{g};\boldsymbol{\theta})$ n'est souvent pas aisé
- Algorithme EM permet d'atteindre le maximum sans calculer explicitement $p(\boldsymbol{g};\boldsymbol{\theta})$:

Le principe de l'algorithme EM consiste à :

- Introduire une variable auxiliaire $\boldsymbol{z}$ reliée à $\boldsymbol{g}$
- $\boldsymbol{z}$ données complètes
- $\boldsymbol{g}$ données incomplètes
- $\{\boldsymbol{z}\} \mapsto \{\boldsymbol{g}\}$ Projection : $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{z}$

$$p(\boldsymbol{z}|\boldsymbol{g};\boldsymbol{\theta}) = \frac{p(\boldsymbol{z},\boldsymbol{g};\boldsymbol{\theta})}{p(\boldsymbol{g};\boldsymbol{\theta})} = \frac{p(\boldsymbol{z};\boldsymbol{\theta})p(\boldsymbol{g}|\boldsymbol{z})}{p(\boldsymbol{g};\boldsymbol{\theta})} = \frac{p(\boldsymbol{z};\boldsymbol{\theta})I_{\boldsymbol{g}}(\boldsymbol{z})}{p(\boldsymbol{g};\boldsymbol{\theta})}$$

où $I_{\boldsymbol{g}}(\boldsymbol{z})$ est la fonction indicatrice : $I_{\boldsymbol{g}}(\boldsymbol{z}) = \begin{cases} 1 & \text{si } \boldsymbol{g} = \boldsymbol{H}\boldsymbol{z} \\ 0 & \text{sinon} \end{cases}$

$$p(\boldsymbol{g};\boldsymbol{\theta}) = \int_{I_{\boldsymbol{g}}(\boldsymbol{z})} p(\boldsymbol{z};\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{z}$$

- Algorithme iteratif :

$$\begin{cases} \text{(E)} & \text{Calculer} & Q(\boldsymbol{\theta};\widehat{\boldsymbol{\theta}}^{(k)}) & = \mathrm{E}_{Z|Y;\widehat{\boldsymbol{\theta}}^{(k)}} \{\ln p(\boldsymbol{z};\boldsymbol{\theta})\} \\ \text{(M)} & \text{Choisir} & \widehat{\boldsymbol{\theta}}^{(k+1)} & = \arg\max_{\boldsymbol{\theta}} \left\{ Q(\boldsymbol{\theta};\widehat{\boldsymbol{\theta}}^{(k)}) \right\} \end{cases}$$

- $\widehat{\boldsymbol{\theta}}^{k}$ converge vers $\widehat{\boldsymbol{\theta}}$

$$\begin{aligned} Q(\boldsymbol{\theta};\widehat{\boldsymbol{\theta}}^{(k)}) &= \mathrm{E}\left[\ln p(\boldsymbol{z};\boldsymbol{\theta})|\boldsymbol{g};\widehat{\boldsymbol{\theta}}^{(k)}\right] \\ &= \mathrm{E}_{Z|Y;\widehat{\boldsymbol{\theta}}^{(k)}} \{\ln p(\boldsymbol{z};\boldsymbol{\theta})\} \\ &= \int_{I_{\boldsymbol{g}}(\boldsymbol{z})} \ln p(\boldsymbol{z};\boldsymbol{\theta})\, p(\boldsymbol{z}|\boldsymbol{g};\widehat{\boldsymbol{\theta}}^{(k)})\,\mathrm{d}\boldsymbol{z} \end{aligned}$$

11.5 Exemples et exercises :

Exercise 1 :

$$y = x + b, \qquad x \sim \mathcal{N}\left(0, \sigma_x^2\right), \qquad b \sim \mathcal{N}\left(0, \sigma_b^2\right)$$

Vérifier les relations suivantes :

$$\begin{aligned}
y|x &\sim \mathcal{N}\left(x, \sigma_b^2\right), \\
y &\sim \mathcal{N}\left(0, \sigma_x^2 + \sigma_b^2\right) \\
\begin{pmatrix} x \\ y \end{pmatrix} &\sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & \sigma_x^2 \\ \sigma_x^2 & \sigma_x^2 + \sigma_b^2 \end{pmatrix}\right) \\
x|y &\sim \mathcal{N}\left(\widehat{x}, \frac{\sigma_x^2 \sigma_b^2}{\sigma_x^2 + \sigma_b^2}\right), \quad \text{avec} \quad \widehat{x} = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_b^2} y
\end{aligned}$$

Exercise 2 :

Étant donné y et $\sigma_b^2 = 1$, estimer x et $\theta = \sigma_x^2$

Vérifier les relations suivantes :

$$\begin{aligned}
x \sim \mathcal{N}(0, \theta) &\longrightarrow p(x; \theta) = \frac{1}{\sqrt{2\pi\theta}} \exp\left[-\frac{x^2}{2\theta}\right] \\
b \sim \mathcal{N}(0, 1) &\longrightarrow p(b) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{b^2}{2}\right] \\
y|x \sim \mathcal{N}(x, 1) &\longrightarrow p(y|x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(y-x)^2}{2}\right] \\
y \sim \mathcal{N}(0, \theta + 1) &\longrightarrow p(y; \theta) = \frac{1}{\sqrt{2\pi(\theta + 1)}} \exp\left[-\frac{y^2}{2(\theta + 1)}\right]
\end{aligned}$$

$$\Sigma = \begin{pmatrix} \theta & \theta \\ \theta & \theta + 1 \end{pmatrix}, \qquad \det|\Sigma| = \theta$$

$$\Sigma^{-1} = \frac{1}{\theta} \begin{pmatrix} \theta + 1 & -\theta \\ -\theta & \theta \end{pmatrix}, \qquad \det\left|\Sigma^{-1}\right| = 1$$

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma\right] \longrightarrow p(x, y; \theta) = \frac{\det|\Sigma|^{-\frac{1}{2}}}{2\pi} \exp\left[-\frac{1}{2}(x, y)\Sigma^{-1}\begin{pmatrix} x \\ y \end{pmatrix}\right]$$

$$x|y \sim \mathcal{N}(\widehat{x}, \frac{\theta}{\theta + 1}), \quad \widehat{x} = \frac{\theta}{\theta + 1} y$$

$$p(x|y; \theta) = \frac{1}{\sqrt{2\pi\frac{\theta}{\theta+1}}} \exp\left[-\frac{(\theta + 1)(x - \widehat{x})^2}{2\theta}\right]$$

Vraisemblance généralisée :

$$\begin{aligned}
\mathrm{VG}(x,\theta|y) &= p(x,y;\theta) = p(y|x)p(x;\theta) \\
&= \frac{1}{\sqrt{2\pi}}\exp\left[-\frac{(y-x)^2}{2}\right]\frac{1}{\sqrt{2\pi\theta}}\exp\left[-\frac{x^2}{2\theta}\right] \\
&= \frac{1}{2\pi\sqrt{\theta}}\exp\left[-\frac{1}{2}(y^2+\frac{\theta+1}{\theta}x^2-2xy)\right]
\end{aligned}$$

Vraisemblance marginale :

$$\begin{aligned}
\mathrm{VM}(\theta|y) &= \int p(x,y;\theta)\,\mathrm{d}x = p(y;\theta) \\
&= \frac{1}{\sqrt{2\pi(\theta+1)}}\exp\left[-\frac{y^2}{2(\theta+1)}\right]
\end{aligned}$$

Estimation au sens du MAP **de** x **si** θ **connu :**

$$\begin{aligned}
\widehat{x} &= \arg\max_x\left\{p(x|y;\theta)\right\} = \arg\max_x\left\{p(y|x;\theta)p(x;\theta)\right\} \\
&= \arg\max_x\left\{\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{(y-x)^2}{2}\right]\frac{1}{\sqrt{2\pi\theta}}\exp\left[-\frac{x^2}{2\theta}\right]\right\} \\
&= \arg\max_x\left\{\frac{1}{2\pi\sqrt{\theta}}\exp\left[-\frac{1}{2}(y^2+\frac{\theta}{\theta+1}x^2-2xy)\right]\right\} \\
&= \frac{\theta}{\theta+1}y \quad\longrightarrow \widehat{x}\sim\mathcal{N}\left(\frac{\theta}{\theta+1}y,\frac{\theta}{\theta+1}\right)
\end{aligned}$$

Estimation au sens du MV de x **si** θ **connu :**

$$\begin{aligned}
\widehat{x} &= \arg\max_x\left\{p(y|x;\theta)\right\} \\
&= \arg\max_x\left\{\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{(y-x)^2}{2}\right]\right\} \\
&= y \quad\longrightarrow \widehat{x}\sim\mathcal{N}\left(y,1\right)
\end{aligned}$$

Vraisemblance généralisée :

$$(\widehat{x},\widehat{\theta}) = \arg\max_{(x,\theta)} \{p(x,y;\theta)\} = \arg\min_{(x,\theta)} \{-\ln p(x,y;\theta)\}$$

$$\begin{aligned} p(x,y;\theta) &= p(y|x)p(x;\theta) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(y-x)^2}{2}\right] \frac{1}{\sqrt{2\pi\theta}} \exp\left[-\frac{x^2}{2\theta}\right] \\ &= \frac{1}{2\pi\sqrt{\theta}} \exp\left[-\frac{1}{2}\left[(y-x)^2 + \frac{x^2}{\theta}\right]\right] \end{aligned}$$

$$L(x,\theta) = -\ln p(x,y;\theta) = \ln(2\pi) + \frac{1}{2}\left[\ln(\theta) + y^2 + \frac{\theta+1}{\theta}x^2 - 2xy\right]$$

$$\left\{ \begin{array}{rcl} \dfrac{\partial L(x,\theta)}{\partial x} &=& \left(\dfrac{\theta+1}{\theta}x - y\right) = 0 \quad \longrightarrow x = \dfrac{\theta}{\theta+1}y \\ \dfrac{\partial L(x,\theta)}{\partial \theta} &=& \dfrac{1}{2}\left[\dfrac{1}{\theta} - \dfrac{x^2}{\theta^2}\right] = 0 \quad \longrightarrow \theta = x^2 \end{array} \right.$$

$$\left\{ \begin{array}{rcl} \widehat{x} &=& \dfrac{\theta}{\theta+1}y \\ \widehat{\theta} &=& \widehat{x}^2 = \left(\dfrac{\widehat{\theta}}{\widehat{\theta}+1}\right)^2 y^2 \end{array} \right.$$

Condition aux limites :

$$\theta = \left(\frac{\widehat{\theta}}{\widehat{\theta}+1}\right)^2 y^2 \longrightarrow \theta^2 + 2\theta + 1 - \theta y^2 = 0 \longrightarrow \Delta = (2-y^2)^2 - 4 >= 0 \longrightarrow 0 > y^2 > 4$$

$$\left\{ \begin{array}{rcl} \widehat{x} &=& \arg\max_{x} \left\{p(x|y;\widehat{\theta})\right\} \\ \widehat{\theta} &=& \arg\max_{\theta} \{p(\widehat{x};\theta)\} \end{array} \right.$$

Vraisemblance marginale :

$$\left\{\begin{array}{lcl}\widehat{\theta} & = & \arg\max\limits_{\theta}\{p(y;\theta)\} \\ \widehat{x} & = & \arg\max\limits_{x}\left\{p(x|y;\widehat{\theta})\right\}\end{array}\right.$$

$$\left\{\begin{array}{lcl}p(y;\theta) & = & \dfrac{1}{\sqrt{2\pi(\theta+1)}}\exp\left[-\dfrac{y^2}{2(\theta+1)}\right] \\ p(x|y;\theta) & = & \dfrac{1}{\sqrt{2\pi\frac{\theta}{\theta+1}}}\exp\left[-\dfrac{(\theta+1)(x-\widehat{x})^2}{2\theta}\right]\end{array}\right.$$

$$\left\{\begin{array}{lcl}\widehat{\theta} & = & y^2-1 \text{ si } y^2>1 \\ \widehat{x} & = & \dfrac{\widehat{\theta}}{\widehat{\theta}+1}y = y-\dfrac{1}{y}\end{array}\right.$$

Contexte complètement bayésien :

$$(\widehat{x},\widehat{\theta}) = \arg\max_{(x,\theta)}\{p(x,\theta|y)\} = \arg\max_{(x,\theta)}\{p(y|x)p(x|\theta)p(\theta)\} = \arg\min_{(x,\theta)}\{-lnp(y,x,\theta)\}$$

$$\begin{array}{lcl}L(x,\theta) & = & \ln p(x,y;\theta) \\ & = & -\ln(2\pi)-\dfrac{1}{2}\left[\ln(\theta)+y^2+\dfrac{\theta+1}{\theta}x^2-2xy\right]+\ln p(\theta)\end{array}$$

$$\left\{\begin{array}{lcl}\dfrac{\mathrm{d}L(x,\theta)}{\mathrm{d}x} & = & -\dfrac{\theta+1}{\theta}x+y=0 \quad\longrightarrow x=\dfrac{\theta}{\theta+1}y \\ \dfrac{\mathrm{d}L(x,\theta)}{\mathrm{d}\theta} & = & -\dfrac{1}{2}\left[\dfrac{3}{\theta}-\dfrac{x^2}{\theta^2}\right]=0 \quad\longrightarrow \theta=\dfrac{1}{3}x^2\end{array}\right.$$

Avec un choix de $p(\theta)\propto\frac{1}{\theta}$ on a :

$$p(\theta)\propto\frac{1}{\theta} \quad\longrightarrow\quad \left\{\begin{array}{lcl}\widehat{x} & = & \dfrac{\theta}{\theta+1}y \\ \widehat{\theta} & = & \dfrac{1}{3}\widehat{x}^2\end{array}\right.$$

$$p(\theta)\propto\frac{1}{\sqrt{\theta}} \quad\longrightarrow\quad \left\{\begin{array}{lcl}\widehat{x} & = & \dfrac{\theta}{\theta+1}y \\ \widehat{\theta} & = & \dfrac{1}{2}\widehat{x}^2\end{array}\right.$$

Algorithme EM :

$$z = \begin{pmatrix} x \\ y \end{pmatrix} \quad \longrightarrow p(z;\theta) = p(x,y;\theta) = \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma \right)$$

$$\Sigma = \begin{pmatrix} \theta & \theta \\ \theta & \theta+1 \end{pmatrix}$$

$$-\frac{1}{2}(x,y)\Sigma^{-1}\begin{pmatrix} x \\ y \end{pmatrix} = -\frac{1}{2}\left(\frac{\theta+1}{\theta}x^2 - 2xy + y^2 + \ln\det|\Sigma| + 2\ln 2\pi\right)$$

$$\ln p(z;\theta) \propto -\ln\theta - \frac{\theta+1}{\theta}x^2 + 2xy - y^2$$

$$p(z|y;\widehat{\theta}) = p(x,y|y;\widehat{\theta}) = p(x|y;\widehat{\theta})$$

$$\mathrm{E}\left[\ln p(z;\theta)|y;\widehat{\theta}\right] = \ln\theta - \frac{\theta+1}{\theta}\mathrm{E}\left[x^2|y;\widehat{\theta}\right] + 2y\mathrm{E}\left[x|y;\widehat{\theta}\right] - y^2$$

$$Q(\theta,\widehat{\theta}) = \ln\theta - \frac{\theta+1}{\theta}\mathrm{E}\left[x^2|y;\widehat{\theta}\right] + 2y\mathrm{E}\left[x|y;\widehat{\theta}\right] + \mathrm{cte}$$

$$\frac{\partial Q(\theta,\widehat{\theta}^{(k)})}{\partial\theta} = 0 \quad \longrightarrow \widehat{\theta}^{(k+1)} \begin{aligned}[t] &= \mathrm{E}\left[x^2|y;\widehat{\theta}^{(k)}\right] \\ &= \frac{\theta^{(k)}}{\theta^{(k)}+1} + \left(\frac{\theta^{(k)}}{\theta^{(k)}+1}y\right)^2 \end{aligned}$$

$$\widehat{\theta}^{(k+1)} = -\frac{\theta^{(k)}}{\theta^{(k)}+1} - \left(\frac{\theta^{(k)}}{\theta^{(k)}+1}y\right)^2$$

Exercise 3 :
Extension au cas : $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{b}$

$$\boldsymbol{f} \sim \mathcal{N}(0, \boldsymbol{R}_X = \sigma_x^2 \boldsymbol{P}_0), \qquad \boldsymbol{b} \sim \mathcal{N}(0, \boldsymbol{R}_B = \sigma_b^2 \boldsymbol{I})$$

Vérifier les relations suivantes :

$$\begin{array}{rcl}
\boldsymbol{g}|\boldsymbol{f} & \sim & \mathcal{N}(\boldsymbol{H}\boldsymbol{f}, \boldsymbol{R}_B = \sigma_b^2 \boldsymbol{I}) \\
\boldsymbol{g} & \sim & \mathcal{N}(0, \boldsymbol{R}_Y) \quad \text{avec} \quad \boldsymbol{R}_Y = \boldsymbol{H}\boldsymbol{R}_X\boldsymbol{H}^t + \boldsymbol{R}_B \\
\begin{pmatrix} \boldsymbol{f} \\ \boldsymbol{g} \end{pmatrix} & \sim & \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{\Sigma}\right) \quad \text{avec} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{R}_X & \boldsymbol{R}_{XY} \\ \boldsymbol{R}_{YX} & \boldsymbol{R}_Y \end{pmatrix} \\
 & & \text{où} \quad \boldsymbol{R}_{YX} = \boldsymbol{R}_{XY}^t = \boldsymbol{H}\boldsymbol{R}_X \\
\boldsymbol{f}|\boldsymbol{g} & \sim & \mathcal{N}(\widehat{\boldsymbol{f}}, \boldsymbol{P})
\end{array}$$

avec

$$\left\{ \begin{array}{rcl}
\widehat{\boldsymbol{f}} & = & \boldsymbol{R}_X\boldsymbol{H}^t(\boldsymbol{H}\boldsymbol{R}_X\boldsymbol{H}^t + \boldsymbol{R}_B)^{-1}\boldsymbol{g} = \boldsymbol{P}\boldsymbol{H}^t\boldsymbol{R}_B^{-1}\boldsymbol{g} \\
\boldsymbol{P} & = & \boldsymbol{R}_X - \boldsymbol{R}_X\boldsymbol{H}^t(\boldsymbol{H}\boldsymbol{R}_X\boldsymbol{H}^t + \boldsymbol{R}_B)^{-1}\boldsymbol{H}\boldsymbol{R}_X = (\boldsymbol{R}_X^{-1} + \boldsymbol{H}^t\boldsymbol{R}_B^{-1}\boldsymbol{H})^{-1}
\end{array} \right.$$

Avec $\boldsymbol{R}_B = \sigma_b^2 \boldsymbol{I}$ et $\boldsymbol{R}_X = \sigma_x^2 \boldsymbol{P}_0$ on obtient :

$$\left\{ \begin{array}{rcl}
\widehat{\boldsymbol{f}} & = & \left(\boldsymbol{H}^t\boldsymbol{H} + \boldsymbol{\lambda}\boldsymbol{P}_0^{-1}\right)^{-1}\boldsymbol{H}^t\boldsymbol{g} \quad \text{avec} \quad \lambda = \dfrac{\sigma_b^2}{\sigma_x^2} = \dfrac{1}{\theta} \\
\boldsymbol{P} & = & \sigma_b^2\left(\boldsymbol{H}^t\boldsymbol{H} + \lambda\boldsymbol{P}_0^{-1}\right)^{-1}
\end{array} \right.$$

Exercise 4 :
Étant donnée $\boldsymbol{H}, \boldsymbol{g}, \sigma_b^2 = 1$ et $\boldsymbol{P}_0$ estimer $\boldsymbol{f}$ et $\theta = \sigma_x^2$.

Vérifier et compléter les relations suivantes :

$$\begin{array}{rcl} \boldsymbol{f} & \sim & \mathcal{N}(0, \theta \boldsymbol{P}_0) \\ \boldsymbol{b} & \sim & \mathcal{N}(0, \boldsymbol{I}) \\ \boldsymbol{g}|\boldsymbol{f} & \sim & \mathcal{N}(\boldsymbol{H}\boldsymbol{f}, \boldsymbol{I}) \\ \boldsymbol{g} & \sim & \mathcal{N}(0, \boldsymbol{R}_Y) \end{array}$$

$$\boldsymbol{\Sigma} = \theta \begin{pmatrix} \boldsymbol{P}_0 & \boldsymbol{P}_0^t \boldsymbol{H}^t \\ \boldsymbol{H}\boldsymbol{P}_0 & \boldsymbol{H}\boldsymbol{P}_0\boldsymbol{H}^t + \boldsymbol{I} \end{pmatrix},$$

$$\det|\boldsymbol{\Sigma}| = \det\left|\boldsymbol{R}_X - \boldsymbol{R}_{XY}\boldsymbol{R}_Y^{-1}\boldsymbol{R}_{YX}\right| \det|\boldsymbol{R}_Y|$$

$$\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \boldsymbol{P}^{-1} & -\boldsymbol{P}^{-1}\boldsymbol{R}_{XY}\boldsymbol{R}_Y^{-1} \\ -\boldsymbol{R}_Y^{-1}\boldsymbol{R}_{YX}\boldsymbol{P}^{-1} & \boldsymbol{R}_Y^{-1} + \boldsymbol{R}_Y^{-1}\boldsymbol{R}_{YX}\boldsymbol{P}^{-1}\boldsymbol{R}_{XY}\boldsymbol{R}_Y^{-1} \end{pmatrix}$$

$$\det\left|\boldsymbol{\Sigma}^{-1}\right| =$$

$$\begin{array}{rcl} \begin{pmatrix} \boldsymbol{f} \\ \boldsymbol{g} \end{pmatrix} & \sim \mathcal{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{\Sigma}\right] & \\ & = & \dfrac{\det|\boldsymbol{\Sigma}|^{-\frac{1}{2}}}{(2\pi)^{\frac{N+M}{2}}} \exp\left[-\dfrac{1}{2}[\boldsymbol{f}, \boldsymbol{g}]^t \boldsymbol{\Sigma}^{-1} \begin{pmatrix} \boldsymbol{f} \\ \boldsymbol{g} \end{pmatrix}\right] \end{array}$$

$$\boldsymbol{f}|\boldsymbol{g} \sim \mathcal{N}(\widehat{\boldsymbol{f}}, \boldsymbol{P}) = (2\pi)^{-\frac{N}{2}} \det|\boldsymbol{P}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\left[[\boldsymbol{f} - \widehat{\boldsymbol{f}}]^t \boldsymbol{P}^{-1} [\boldsymbol{f} - \widehat{\boldsymbol{f}}]\right]\right]$$

avec

$$\left\{ \begin{array}{rcl} \widehat{\boldsymbol{f}} & = & \left(\boldsymbol{H}^t\boldsymbol{H} + \lambda \boldsymbol{P}_0^{-1}\right)^{-1} \boldsymbol{H}^t \boldsymbol{g} \quad \text{avec} \quad \lambda = \dfrac{1}{\theta} \\ \boldsymbol{P} & = & (\boldsymbol{H}^t\boldsymbol{H} + \lambda \boldsymbol{P}_0^{-1})^{-1} \end{array} \right.$$

Vraisemblance généralisée :

$$\mathrm{VG}(\boldsymbol{f},\theta|\boldsymbol{g}) = p(\boldsymbol{f},\boldsymbol{g};\theta) = p(\boldsymbol{g}|\boldsymbol{f})p(\boldsymbol{f};\theta)$$

$$\begin{aligned}
&= (2\pi)^{-\frac{M}{2}}\det|\boldsymbol{R}_B|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}\left[[\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}]^t\boldsymbol{R}_B^{-1}[\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}]\right]\right]\times \\
&\quad \times(2\pi)^{-\frac{N}{2}}\det|\boldsymbol{R}_X|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}\left[\boldsymbol{f}^t\boldsymbol{R}_X^{-1}\boldsymbol{f}\right]\right] \\
&= (2\pi)^{-\frac{N}{2}}(\theta\det|\boldsymbol{P}_0|)^{-\frac{1}{2}}\exp\left[-\frac{1}{2\theta}\left[\theta[\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}]^t[\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}]+\boldsymbol{f}^t\boldsymbol{P}_0^{-1}\boldsymbol{f}\right]\right]
\end{aligned}$$

Vraisemblance marginale :

$$\begin{aligned}
\mathrm{VM}(\theta|\boldsymbol{g}) &= \int p(\boldsymbol{f},\boldsymbol{g};\theta)\,\mathrm{d}\boldsymbol{f} = p(\boldsymbol{g};\theta) \\
&= (2\pi)^{-\frac{M}{2}}\det|\boldsymbol{R}_Y|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}\left[\boldsymbol{g}^t\boldsymbol{R}_Y^{-1}\boldsymbol{g}\right]\right]
\end{aligned}$$

$$\boldsymbol{R}_Y = \theta\left(\boldsymbol{H}\boldsymbol{P}_0\boldsymbol{H}^t+\frac{1}{\theta}\right)$$

Estimation au sens du MAP **de $\boldsymbol{f}$ si θ connu :**

$$\begin{aligned}
\widehat{\boldsymbol{f}} &= \arg\max_{\boldsymbol{f}}\{p(\boldsymbol{f}|\boldsymbol{g};\theta)\} = \arg\max_{\boldsymbol{f}}\{p(\boldsymbol{g}|\boldsymbol{f};\theta)p(\boldsymbol{f};\theta)\} \\
&= \arg\min_{\boldsymbol{f}}\left\{[\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}]^t\boldsymbol{R}_B^{-1}[\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}]+\boldsymbol{f}^t\boldsymbol{R}_X^{-1}\boldsymbol{f}\right\} \\
&= \left(\boldsymbol{H}^t\boldsymbol{H}+\lambda\boldsymbol{P}_0^{-1}\right)^{-1}\boldsymbol{H}^t\boldsymbol{g}, \quad \lambda=\frac{1}{\theta}
\end{aligned}$$

Estimation au sens du MV de $\boldsymbol{f}$ si θ connu :

$$\begin{aligned}
\widehat{\boldsymbol{f}} &= \arg\max_{\boldsymbol{f}}\{p(\boldsymbol{g}|\boldsymbol{f};\theta)\} \\
&= \arg\max_{\boldsymbol{f}}\left\{(2\pi)^{-\frac{M}{2}}\det|\boldsymbol{R}_B|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}\left[[\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}]^t\boldsymbol{R}_B^{-1}[\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}]\right]\right]\right\} \\
&= \arg\min_{\boldsymbol{f}}\left\{[\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}]^t\boldsymbol{R}_B^{-1}[\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}]\right\} \\
&= (\boldsymbol{H}^t\boldsymbol{H})^{-1}\boldsymbol{H}^t\boldsymbol{g}
\end{aligned}$$

Vraisemblance généralisée

$$(\widehat{\boldsymbol{f}},\widehat{\theta}) = \arg\max_{(\boldsymbol{f},\theta)} \{p(\boldsymbol{f},\boldsymbol{g};\theta)\} = \arg\max_{(\boldsymbol{f},\theta)} \{\ln p(\boldsymbol{f},\boldsymbol{g};\theta)\}$$

$$\begin{aligned}
p(\boldsymbol{f},\boldsymbol{g};\theta) &= p(\boldsymbol{g}|\boldsymbol{f})p(\boldsymbol{f};\theta) \\
&= (2\pi)^{-\frac{M}{2}}\det|\boldsymbol{R}_B|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\left[[\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}]^t\boldsymbol{R}_B^{-1}[\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}]\right]\right] \times \\
&\quad \times(2\pi)^{-\frac{N}{2}}\det|\boldsymbol{R}_X|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\left[\boldsymbol{f}^t\boldsymbol{R}_X^{-1}\boldsymbol{f}\right]\right] \\
&= K\,\theta^{-\frac{1}{2}} \exp\left[-\frac{1}{2}[\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}]^t[\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}] - \frac{1}{2\theta}\boldsymbol{f}^t\boldsymbol{P}_0^{-1}\boldsymbol{f}\right]
\end{aligned}$$

$$L(x,\theta) = \ln p(\boldsymbol{f},\boldsymbol{g};\theta) \propto \left[\ln\theta + [\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}]^t[\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}] + \frac{1}{\theta}\boldsymbol{f}^t\boldsymbol{P}_0^{-1}\boldsymbol{f}\right]$$

$$\left\{\begin{array}{lcl}
\dfrac{\mathrm{d}L(\boldsymbol{f},\theta)}{\mathrm{d}\boldsymbol{f}} & = & 0 \quad \longrightarrow \boldsymbol{f} = \left(\boldsymbol{H}^t\boldsymbol{H}+\lambda\boldsymbol{P}_0^{-1}\right)^{-1}\boldsymbol{H}^t\boldsymbol{g} \\
\dfrac{\mathrm{d}L(\boldsymbol{f},\theta)}{\mathrm{d}\theta} & = & 0 \quad \longrightarrow \theta = \boldsymbol{f}^t\boldsymbol{P}_0^{-1}\boldsymbol{f}
\end{array}\right.$$

$$\left\{\begin{array}{lcl}
\widehat{\boldsymbol{f}} & = & \left(\boldsymbol{H}^t\boldsymbol{H}+\lambda\boldsymbol{P}_0\right)^{-1}\boldsymbol{H}^t\boldsymbol{g} \\
\widehat{\theta} & = & \widehat{\boldsymbol{f}}^t\boldsymbol{P}_0^{-1}\widehat{\boldsymbol{f}}
\end{array}\right.$$

$$\left\{\begin{array}{lcl}
\widehat{\boldsymbol{f}} & = & \arg\max_{\boldsymbol{f}}\left\{p(\boldsymbol{f}|\boldsymbol{g};\widehat{\theta})\right\} \\
\widehat{\theta} & = & \arg\max_{\theta}\left\{p(\widehat{\boldsymbol{f}};\theta)\right\}
\end{array}\right.$$

Vraisemblance marginale :

$$\left\{ \begin{array}{rcl} \widehat{\theta} & = & \arg\max_{\theta} \left\{ p(\boldsymbol{g};\theta) \right\} = (2\pi)^{-\frac{M}{2}} \det |\boldsymbol{R}_Y|^{-\frac{1}{2}} \exp\left[-\frac{1}{2} \left[\boldsymbol{g}^t \boldsymbol{R}_Y^{-1} \boldsymbol{g} \right] \right] \\ \widehat{\boldsymbol{f}} & = & \arg\max_{\boldsymbol{f}} \left\{ p(\boldsymbol{f}|\boldsymbol{g};\widehat{\theta}) \right\} \end{array} \right.$$

$$\boldsymbol{R}_Y = \theta \boldsymbol{H} \boldsymbol{P}_0 \boldsymbol{H}^t + \boldsymbol{I}$$

$$\begin{array}{rcl} \widehat{\theta} & = & \arg\min_{\theta} \left\{ \ln\det \left| \theta \boldsymbol{H} \boldsymbol{P}_0 \boldsymbol{H}^t + \boldsymbol{I} \right| - \boldsymbol{g}^t \left(\theta \boldsymbol{H} \boldsymbol{P}_0 \boldsymbol{H}^t + \boldsymbol{I} \right)^{-1} \boldsymbol{g} \right\} \\ \widehat{\boldsymbol{f}} & = & \arg\max_{\boldsymbol{f}} \left\{ p(\boldsymbol{f}|\boldsymbol{g};\widehat{\theta} \right\} \end{array}$$

Contexte complètement bayésien :

$$(\widehat{\boldsymbol{f}}, \widehat{\theta}) = \arg\max_{(\boldsymbol{f},\theta)} \left\{ p(\boldsymbol{f},\theta|\boldsymbol{g}) \right\} = \arg\max_{(\boldsymbol{f},\theta)} \left\{ p(\boldsymbol{g}|\boldsymbol{f}) p(x|\theta) p(\theta) \right\}$$

$$p(\theta) \propto \frac{1}{\theta}$$

$$L(x,\theta) = \ln p(\boldsymbol{f},\boldsymbol{g};\theta) \propto \left[\ln\theta + [\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}]^t [\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}] + \frac{1}{\theta} \boldsymbol{f}^t \boldsymbol{P}_0^{-1} \boldsymbol{f} \right] - \ln\theta$$

$$\left\{ \begin{array}{rcl} \dfrac{\mathrm{d}L(\boldsymbol{f},\theta)}{\mathrm{d}\boldsymbol{f}} & = & 0 \quad \longrightarrow \boldsymbol{f} = \left(\boldsymbol{H}^t \boldsymbol{H} + \lambda \boldsymbol{P}_0 \right)^{-1} \boldsymbol{H}^t \boldsymbol{g} \\ \dfrac{\mathrm{d}L(\boldsymbol{f},\theta)}{\mathrm{d}\theta} & = & 0 \quad \longrightarrow \theta = \boldsymbol{f}^t \boldsymbol{P}_0^{-1} \boldsymbol{f} \end{array} \right.$$

$$\left\{ \begin{array}{rcl} \widehat{\boldsymbol{f}} & = & \left(\boldsymbol{H}^t \boldsymbol{H} + \lambda \boldsymbol{P}_0 \right)^{-1} \boldsymbol{H}^t \boldsymbol{g} \\ \widehat{\theta} & = & \widehat{\boldsymbol{f}}^t \boldsymbol{P}_0^{-1} \widehat{\boldsymbol{f}} \end{array} \right.$$

Algorithme EM

$$\boldsymbol{z} = \begin{pmatrix} \boldsymbol{f} \\ \boldsymbol{g} \end{pmatrix} \quad \longrightarrow p(\boldsymbol{z};\theta) = p(\boldsymbol{f},\boldsymbol{g};\theta) = \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{\Sigma}\right)$$

$$\ln p(\boldsymbol{z};\theta) \propto \ln\det|\boldsymbol{\Sigma}| - [\boldsymbol{f},\boldsymbol{g}]\boldsymbol{\Sigma}^{-1}\begin{pmatrix} \boldsymbol{f} \\ \boldsymbol{g} \end{pmatrix}$$

$$p(\boldsymbol{z}|\boldsymbol{g};\widehat{\theta}) = p(\boldsymbol{f},\boldsymbol{g}|\boldsymbol{g};\widehat{\theta}) = p(\boldsymbol{f}|\boldsymbol{g};\widehat{\theta})$$

$$\mathrm{E}\left[\ln p(\boldsymbol{z};\theta)|\boldsymbol{g};\widehat{\theta}\right] = \ln\det|\boldsymbol{\Sigma}| - \mathrm{E}\left[[\boldsymbol{f},\boldsymbol{g}]\boldsymbol{\Sigma}^{-1}\begin{pmatrix} \boldsymbol{f} \\ \boldsymbol{g} \end{pmatrix}\right]$$

$$Q(\theta,\widehat{\theta}) = \ln\theta - \frac{\theta+1}{\theta}\mathrm{E}\left[\boldsymbol{f}^2|\boldsymbol{g};\widehat{\theta}\right] + 2\boldsymbol{g}\mathrm{E}\left[\boldsymbol{f}|\boldsymbol{g};\widehat{\theta}\right] + \mathrm{cte}$$

$$\frac{\mathrm{d}Q(\theta,\widehat{\theta}^{(k)})}{\mathrm{d}\theta} = 0 \quad \longrightarrow \widehat{\theta}^{(k+1)} = \mathrm{E}\left[\boldsymbol{f}^2|\boldsymbol{g};\widehat{\theta}^{(k)}\right] = \frac{\theta^{(k)}}{\theta^{(k)}+1} + \left(\frac{\theta^{(k)}}{\theta^{(k)}+1}\boldsymbol{g}\right)^2$$