Chapitre 8

Approches probabilistes

Nous avons distingué quatre approches probabilistes différentes pour la résolution des problèmes inverses. L'objet de ce chapitre est de détailler un peu plus ces approches.

Dans ce chapitre, nous aborderons les trois premières approches, en montrant par des exemples simples leur limites. Nous réservons une place privilégiée à l'approche bayésienne qui sera détaillée dans un chapitre séparé.

8.1 Approche n'utilisant que les moments

Dans cette approche, on considère que les grandeurs observées et inconnues sont bien des fonctions aléatoires, on se limite à caractériser les lois de probabilités de ces grandeurs que par leurs moments, souvent jusqu'à l'ordre deux. Ensuite on établit un lien entre les moments des grandeurs observées et ceux des grandeurs inconnues puis on cherche à l'inverser.

Mais, un point qui est souvent laissé pour compte est : l'estimation des moments des grandeurs observées. Souvent on fait l'hypothèse que l'on peut estimer ces moments par les moyennes empiriques. Ceci peut être raisonnable lorsqu'on se limite aux moments d'ordre un et deux, mais cela devient inefficace pour les moments d'ordre supérieur. C'est d'ailleurs dans le premier cas que ces méthodes trouvent particulièrement leur intérêt : c'est l'estimation en moyenne quadratique.

Pour illustrer cette approche, nous allons considérer le cas du filtrage optimal de Wiener pour la résolution du problème de déconvolution de signaux.

Filtrage optimal de WIENER **:**

Considérons le problème de la déconvolution des signaux:

$$g(t) = h(t) * f(t) + b(t)$$

supposons que $f(t)$, $b(t)$ et $g(t)$ sont des fonctions aléatoires et cherchons à lier les moments d'ordre un :

$$\mathrm{E}[g(t)], \quad \mathrm{E}[b(t)] \quad \text{et} \quad \mathrm{E}[f(t)]$$

et les moments d'ordre deux :

$$\begin{aligned} R_{gg}(\tau) &= \mathrm{E}[g(t)g(t+\tau)], \\ R_{ff}(\tau) &= \mathrm{E}[f(t)f(t+\tau)], \\ R_{bf}(\tau) &= R_{fb}(-\tau) = \mathrm{E}[b(t)f(t+\tau)] \\ R_{gf}(\tau) &= R_{fg}(-\tau) = \mathrm{E}[g(t)f(t+\tau)] \end{aligned}$$

de ces grandeurs. Faisant l'hypothèse que $f(t)$ et $b(t)$ sont indépendantes. On obtient :

$$\begin{aligned} \mathrm{E}[g(t)] &= h(t) * \mathrm{E}[f(t)] + \mathrm{E}[b(t)] \\ R_{gg}(\tau) &= h(t) * h(t) * R_{ff}(\tau) + R_{bb}(\tau) \\ R_{gf}(\tau) &= h(t) * R_{ff}(\tau) \end{aligned}$$

Passant dans le domaine de FOURIER on a :

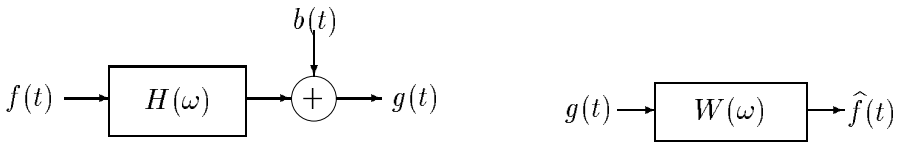
$$\begin{aligned} S_{gg}(\omega) &= |H(\omega)|^2 S_{ff}(\omega) + R_{bb}(\omega) \\ S_{gf}(\omega) &= H(\omega) S_{ff}(\omega), \\ S_{fg}(\omega) &= H^*(\omega) S_{ff}(\omega), \end{aligned}$$

L'objectif du filtrage optimal est de fournir une solution $\widehat{f}(t)$ qui s'obtient par un filtrage linéaire de $g(t)$:

$$\widehat{f}(t) = w(t) * g(t)$$

et la réponse impulsionnelle du filtre $w(t)$ est telle que l'erreur quadratique moyenne

$$EQM = \mathrm{E}\left[\left(f(t) - \widehat{f}(t)\right)^2\right]$$



Modèle d'observation Déconvolution par filtre de Wiener.

FIG. 8.1 - *Filtrage de* WIENER.

soit minimale.

En utilisant le principe d'orthogonalité:

$$\mathrm{E}[(f(t) - w(t) * g(t))\, g(t-\tau)] = 0 \quad \forall t, \tau \longrightarrow R_{fg}(\tau) = w(t) * R_{gg}(\tau).$$

Passant dans le domaine de FOURIER on a:

$$W(\omega) = \frac{S_{fg}(\omega)}{S_{gg}(\omega)} = \frac{H^*(\omega) S_{ff}(\omega)}{|H(\omega)|^2 S_{ff}(\omega) + S_{bb}(\omega)}.$$

Remplaçant pour les expressions de $S_{fg}(\omega)$ et $S_{gg}(\omega)$ on obtient:

$$W(\omega) = \frac{H^*(\omega) S_{ff}(\omega)}{|H(\omega)|^2 S_{ff}(\omega) + S_{bb}(\omega)} = \frac{1}{H(\omega)} \frac{|H(\omega)|^2}{|H(\omega)|^2 + \frac{S_{bb}(\omega)}{S_{ff}(\omega)}}$$

La difficulté dans cette approche commence lors de la mise en œuvre réelle de la méthode. En effet, il faut pouvoir estimer $S_{gg}(\omega)$ et $S_{fg}(\omega)$ ce qui nécessite la connaissance *a priori* de $S_{ff}(\omega)$ et de $S_{bb}(\omega)$.

Souvent, en pratique on fait l'hypothèse que

$$\frac{S_{bb}(\omega)}{S_{ff}(\omega)} = r,$$

avec r une constante représentant l'inverse du rapport signal à bruit. Avec cette hypothèse on a

$$W(\omega) = \frac{1}{H(\omega)} \frac{|H(\omega)|^2}{|H(\omega)|^2 + r}.$$

Cette hypothèse n'est cependant pas très réaliste car elle considère que le signal et le bruit ont la même forme de spectre, ce qui est rarement le cas. En effet, on peut souvent considére le bruit blanc, mais le signal que l'on cherche n'a pas le même spectre.

Notons que lorsque $r \mapsto 0$ on retrouve le filtrage inverse.

Une autre hypothèse plus réaliste consiste à supposer

$$\frac{S_{bb}(\omega)}{S_{ff}(\omega)} = |D(\omega)|^2,$$

où $D(\omega)$ représente la fonction de transfert d'un filtre passe-haut. Avec cette hypothèse on a

$$W(\omega) = \frac{1}{H(\omega)} \frac{|H(\omega)|^2}{|H(\omega)|^2 + |D(\omega)|^2}.$$

Nous verrons dans le chapitre suivant qu'il existe un lien entre cette approche et celle de l'estimation bayésienne dans le cas linéaire avec des lois gaussiennes.

8.2 Estimation au sens du maximum de vraisemblance

Dans cette approche on prend en compte explicitement le caractère incertain des mesures en les caractérisant par une loi de probabilité $p(\boldsymbol{g}|\boldsymbol{f})$.

L'idée de base ensuite est de considérer $p(\boldsymbol{g}|\boldsymbol{f})$ comme une fonction $V(\boldsymbol{f}) = p(\boldsymbol{g}|\boldsymbol{f})$ appelée fonction *vraisemblance*. On définit alors la solution $\widehat{\boldsymbol{f}}$ du problème inverse comme l'argument qui maximise cette fonction :

$$\widehat{\boldsymbol{f}} = \arg\max_{\boldsymbol{f}} \{p(\boldsymbol{g}|\boldsymbol{f})\} = \arg\min_{\boldsymbol{f}} \{-\ln p(\boldsymbol{g}|\boldsymbol{f})\}$$

Exemple 1: Loi Gaussienne :
Considérons le modèle $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{b}$ et supposons que $\boldsymbol{b}$ puisse être modélisé par un vecteur aléatoire centrée, blanc et gaussien :

$$b_i \sim \mathcal{N}(0, \sigma_b^2) \longrightarrow \boldsymbol{b} \sim \mathcal{N}(\boldsymbol{0}, \sigma_b^2 \boldsymbol{I}) \longrightarrow \boldsymbol{g}|\boldsymbol{f} \sim \mathcal{N}(\boldsymbol{H}\boldsymbol{f}, \sigma_b^2 \boldsymbol{I}).$$

On en déduit alors :

$$p(\boldsymbol{g}|\boldsymbol{f}) = K \exp\left[\frac{-1}{2\sigma_b^2}[\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}]^t[\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}]\right].$$

L'estimé au sens du MV devient alors :

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \{-\ln p(\boldsymbol{g}|\boldsymbol{f})\} = \arg\min_{\boldsymbol{f}} \left\{\|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2\right\}$$

et on retrouve l'estimation au sens des moindres carrés.

Si nous connaissons un peu plus sur les caractéristiques des b_i, par exemple si les b_i ont des variances différentes, alors

$$b_i \sim \mathcal{N}(0, \sigma_i^2)$$

et il n'est pas difficile de montrer que

$$p(\boldsymbol{g}|\boldsymbol{f}) = K \exp\left[\frac{-1}{2}(\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f})^t \boldsymbol{W}^{-1}(\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f})\right]$$

avec $\boldsymbol{W} = \text{diag}\{\sigma_1^2, \cdots, \sigma_m^2\}$. Alors,

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \{-\ln p(\boldsymbol{g}|\boldsymbol{f})\} = \arg\min_{\boldsymbol{f}} \left\{\|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2_{\boldsymbol{W}}\right\}$$

et on retrouve l'estimation au sens des moindres carrés généralisés.

L'extension au cas d'un modèle non linéaire se fait aussi très facilement.

Exemple 2: Loi Gaussienne généralisée :
Supposons maintenant que les b_i suivent une loi gaussienne généralisée :

$$b_i \sim p(b_i) \propto \exp\left[-\beta|b_i|^p\right], \quad \beta > 0, \quad 1 < p \leq 2.$$

Si on fait l'hypothèse que $\boldsymbol{b}$ est blanc alors on a :

$$\boldsymbol{b} \sim p(\boldsymbol{b}) \propto \exp\left[-\beta \sum_{i=1}^{M} |b_i|^p\right] \longrightarrow -\ln p(\boldsymbol{b}) = K + \beta \sum_{i=1}^{M} |b_i|^p$$

On en déduit alors :

$$-\ln p(\boldsymbol{g}|\boldsymbol{f}) = K + \beta \sum_{i=1}^{M} |g_i - [\boldsymbol{H}\boldsymbol{f}]_i|^p.$$

L'estimé au sens du MV devient alors :

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \{-\ln p(\boldsymbol{g}|\boldsymbol{f})\} = \arg\min_{\boldsymbol{f}} \left\{ \sum_{i=1}^{M} |g_i - [\boldsymbol{H}\boldsymbol{f}]_i|^p \right\} = \arg\min_{\boldsymbol{f}} \{\|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^p\}$$

et on retrouve l'estimation au sens des moindres carrés.

Exemple 3: Loi Gamma :
Supposons maintenant que les b_i suivent une loi de gamma :

$$b_i \sim \mathcal{G}(\alpha, \beta) \longrightarrow p(b_i) = K_i b_i^{-\alpha} \exp[-\beta b_i]$$

et si on fait l'hypothèse que $\boldsymbol{b}$ est blanc (b_i et b_j) indépendantes $\forall i, j, i \neq j$, on a

$$p(\boldsymbol{b}) = \prod_{i=1}^{M} p(b_i) \longrightarrow \ln p(\boldsymbol{b}) = K + \sum_{i=1}^{M} \alpha \log(b_i) + \beta b_i.$$

On en déduit :

$$\ln p(\boldsymbol{g}|\boldsymbol{f}) = K + \sum_{i=1}^{M} \alpha \log(g_i - [\boldsymbol{H}\boldsymbol{f}]_i) + \beta(g_i - [\boldsymbol{H}\boldsymbol{f}]_i)$$

et

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \{-\ln p(\boldsymbol{g}|\boldsymbol{f})\}.$$

Il n'y a pas d'expression analytique pour cette solution, mais elle peut être calculé numériquement par un algorithme itératif.

En conclusion :

- L'approche MV permet de mieux prendre en compte la nature du bruit en lui attribuant une loi de probabilité $p(\boldsymbol{b})$, ou, d'une manière plus générale, l'incertaine sur les mesures en leur attribuant une loi de probabilité $p(\boldsymbol{g}|\boldsymbol{f})$;
- Dans le cas gaussien on retrouve l'estimation au sens des moindres carrés ;
- Cette approche peut fournir des résultats satisfaisants lorsqu'elle est utilisée en combinaison avec une modélisation paramétrique des inconnues avec un nombre de paramètres très inférieure au nombre de données indépendantes. Malheureusement, cette approche donne rarement des résultats satisfaisants pour la résolution des problèmes inverses dans le cadre algébrique où le nombre d'inconnues est du mêmes ordre de grandeur voir plus grand que le nombre des mesures. C'est pourquoi, certains auteurs ont proposé une approche dite *Maximum de vraisemblance pénalisée* qui consiste à définir la solution par :

  $$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \{-\ln p(\boldsymbol{g}|\boldsymbol{f}) + \phi(\boldsymbol{f})\}$$

  où $\phi(\boldsymbol{f})$ est une fonction de pénalisation choisie d'une manière adhoc pour permettre d'obtenir une solution satisfaisante. Nous verrons alors que cette approche peut mieux être interprétée dans le cadre de l'inférence bayésienne où $\phi(\boldsymbol{f}) = -\ln p(\boldsymbol{f})$ avec $p(\boldsymbol{f})$ la loi *a priori*.

8.3 Approche du maximum d'entropie

Avant de voir comment le principe du maximum d'entropie peut être utilisé pour la résolution des problèmes inverses, nous allons rappeler un certain nombre de définitions et de notions qui seront essentielles pour la compréhension de cette approche.

8.3.1 Définition de l'entropie

Le principe du ME peut être approché de différentes manières. L'approche de la théorie de l'information [1] est sans doute la mieux adaptée à notre problème. JAYNES ([2, 3, 4]) est parmi les premiers auteurs modernes à avoir introduit le formalisme du ME par l'approche de la théorie de l'information. Cette notion d'entropie est introduite de la manière suivante : considérons une variable aléatoire discrète X produisant des réalisations $\boldsymbol{x} = \{x_1, x_2, \cdots, x_n\}$ et attribuons les probabilités $\boldsymbol{p} = \{p_1, p_2, \cdots, p_n\}$ à ces réalisations pour représenter notre information partielle sur cette variable. On définit la quantité $I_i = \ln(1/p_i)$ comme la quantité d'information apportée par la réalisation x_i .

Le raisonnement intuitif conduisant à cette expression est le suivant : plus un événement est rare, plus le gain d'information obtenu par sa réalisation est grand. L'utilisation du logarithme rend additif le gain total d'information obtenu par la réalisation de plusieur événements indépendants. On définit alors l'entropie d'un processus par la somme pondérée des informations individuelles de chaque réalisation. C'est la définition de l'entropie donnée par SHANNON :

$$S(\boldsymbol{p}) = \sum_{i=1}^{n} p_i \ln \frac{1}{p_i} = -\sum_{i=1}^{n} p_i \ln p_i. \tag{8.1}$$

L'entropie S est une mesure d'incertitude de la distribution $\boldsymbol{p} = \{p_1, p_2, \cdots, p_n\}$, déterminée uniquement par certaines règles élémentaires de cohérence logique et d'additivité ([3, 4, 5]).

Généralisant ce concept, on définit $-\ln(q_i/p_i)$ comme le gain d'information, sur une probabilité *a priori* p_i, apporté par la connaissance de la probabilité q_i de réalisation de l'événement x_i. On définit alors :

$$S(\boldsymbol{q}, \boldsymbol{p}) = -\sum_{i=1}^{n} q_i \ln \frac{p_i}{q_i} = \sum_{i=1}^{n} q_i \ln \frac{q_i}{p_i}, \tag{8.2}$$

appelée entropie croisée ou entropie relative de la distribution q_i par rapport à la distribution p_i. Notons ici qu'il y a un changement de signe pour des raisons historiques et il est clair que la minimisation de l'entropie croisée $S(\boldsymbol{q}, \boldsymbol{p})$ se réduit à la maximisation de l'entropie $S(\boldsymbol{q})$ si l'*a priori* $\boldsymbol{p}$ est uniforme.

Sous réserve de quelques précautions, on peut généraliser ce qui précède au cas de distributions continues, et définir l'entropie [1] par :

$$S(p) = -\int p(x) \ln p(x) \, \mathrm{d}x, \tag{8.3}$$

et l'entropie croisée par :

$$S(q, p) = -\int q(x) \ln \frac{p(x)}{q(x)} \, \mathrm{d}x. \tag{8.4}$$

8.3.2 Lois à maximum d'entropie

Voyons maintenant ce que signifie le choix d'une distribution de probabilité à maximum d'entropie contenant une information *a priori*, ou qui soit compatible avec des contraintes connues sur cette distribution. Pour cela, précisons tout d'abord la signification de l'information contenue dans une distribution de probabilité $\{p_1, p_2, \cdots, p_n\}$. À l'évidence il faut que l'on puisse extraire cette information de cette distribution. Considérons une variable aléatoire discrète X qui peut prendre des valeurs $\boldsymbol{x} = \{x_1, x_2, \cdots, x_n\}$ avec une distribution de probabilité $\boldsymbol{p} = \{p_1, p_2, \cdots, p_n\}$. Maintenant si on nous demande quelle est la meilleure estimée $\widehat{\phi}$ d'une fonction $\phi(X)$ au sens du minimum de l'erreur quadratique moyenne. La solution est immédiate :

$$\widehat{\phi} = \mathrm{E}\,[\phi] = \sum_{i=1}^{n} p_i\, \phi(x_i).$$

C'est un problème direct et bien-posé. Inversement, si l'on se pose la question d'ajuster une distribution $\boldsymbol{p}$ pour incorporer une information donnée sur la fonction $\phi(X)$, il faut se poser la question suivante :

> connaissant une règle d'estimation précise, par exemple la minimisation de l'erreur quadratique moyenne, comment choisir $\boldsymbol{p}$ pour que l'on ait $\widehat{\phi} = \mathrm{E}\,[\phi]$?

Le problème est qu'il existe, en général, beaucoup de distributions qui satisfont cette contrainte. Il s'agit d'un problème inverse mal posé au sens où la solution n'est pas unique. Le principe du maximum d'entropie nous permet alors de choisir une solution.

Soit maintenant le cas plus général où nous considérons les m fonctions $\{\phi_1(X), \phi_2(X), \cdots, \phi_m(X)\}$ pour lesquelles nous avons un ensemble de données $\{d_1, d_2, \cdots, d_m\}$ pouvant s'exprimer sous la forme des m contraintes simultanées suivantes :

$$\mathrm{E}\,[\phi_k(X)] = \sum_{i=1}^{n} p_i\, \phi_k(x_i) = d_k, \quad k = 1, \cdots, m.$$

Dans chaque problème, ces données $\{d_1, d_2, \cdots, d_m\}$ peuvent avoir des interprétations physiques différentes et la difficulté consiste à incorporer ces données (contraintes) dans notre distribution de probabilité. Nous voulons ajuster la distribution de probabilité $\{p_1, p_2, \cdots, p_n\}$ à nos données. En général, le nombre des contraintes m est inférieur à n et il y a une infinité de solutions à ce problème. En d'autres termes, on peut trouver une infinité de distributions de probabilité $\{p_1, p_2, \cdots, p_n\}$ qui satisfont ces contraintes. C'est ici que le principe du maximum d'entropie est mis en œuvre pour choisir, parmi ces solutions possibles, celle qui a l'entropie maximale, c'est-à-dire la loi qui satisfait toutes les contraintes (toute l'information connue) et qui est la moins compromettante vis-à-vis de toute autre information inconnue. Le mot *information* devant être pris au sens de la définition de l'information moyenne ou de l'entropie de Shannon (équation 3).

Le problème se formule ainsi :

Problème P1 :

$$\text{maximiser} \quad S(\boldsymbol{p}) = -\sum_{i=1}^{n} p_i \ln p_i,$$

$$\text{sous les contraintes} \quad \sum_{i=1}^{n} p_i\, \phi_k(x_i) = d_k, \quad k = 1, \cdots, m.$$

La solution est obtenue par une technique variationnelle de multiplicateurs de Lagrange et est donnée par :

$$p_i = \frac{1}{Z(\lambda_1, \cdots, \lambda_m)} \exp\left[-\sum_{k=1}^{m} \lambda_k \phi_k(x_i)\right], \quad i = 1, \cdots, n, \tag{8.5}$$

où

$$Z(\lambda_1, \cdots, \lambda_m) = \sum_{i=1}^{n} \exp\left[-\sum_{k=1}^{m} \lambda_k \phi_k(x_i)\right] \tag{8.6}$$

est la fonction de partition, et les $\{\lambda_1, \lambda_2, \cdots, \lambda_n\}$ sont déterminés par le système d'équations :

$$-\frac{\partial \ln Z(\lambda_1, \cdots, \lambda_m)}{\partial \lambda_k} = d_k, \quad k = 1, \cdots, m, \tag{8.7}$$

avec m données d_k et m inconnues λ_k. Notons que ce système d'équations n'est autre que le système d'équations des contraintes constituées par les données.

La valeur maximale de l'entropie est :

$$S_{\max} = \ln Z + \sum_{k=1}^{m} \lambda_k d_k.$$

Si on note $\lambda_0 = \ln Z$ on peut vérifier facilement les propriétés suivantes :

$$-\frac{\partial \lambda_0}{\partial \lambda_k} = \mathrm{E}\left[\phi_k(x_i)\right] = d_k, \tag{8.8}$$

$$-\frac{\partial^2 \lambda_0}{\partial {\lambda_k}^2} = \mathrm{E}\left[\phi_k^2(x_i)\right] - d_k^2 = \mathrm{Var}\left[\phi_k(x_i)\right], \tag{8.9}$$

$$-\frac{\partial^2 \lambda_0}{\partial \lambda_k \partial \lambda_l} = \mathrm{E}\left[\phi_k(x_i)\phi_l(x_i)\right] - d_k d_l = \mathrm{Cov}\left\{\phi_k(x_i), \phi_l(x_i)\right\}. \tag{8.10}$$

8.3.3 Lois à minimum d'entropie relative

L'extension de ce qui précède au cas de l'entropie croisée se formule ainsi :

Problème P2 :
Étant donné une distribution de probabilité *a priori* $\boldsymbol{p} = \{p_1, p_2, \cdots, p_n\}$, déterminer la distribution de probabilité *a posteriori* $\boldsymbol{q} = \{q_1, q_2, \cdots, q_n\}$ qui minimise

$$S(\boldsymbol{q}, \boldsymbol{p}) = -\sum_i q_i \ln \frac{p_i}{q_i} = \sum_i q_i \ln \frac{q_i}{p_i},$$

et qui satisfait les m contraintes

$$\sum_{i=1}^{n} q_i \, \phi_k(x_i) = d_k, \quad k = 1, \cdots, m.$$

La solution est obtenue, là aussi, par une technique variationnelle de multiplicateurs de Lagrange et donnée par :

$$q_i = \frac{1}{Z} p_i \exp\left[-\sum_{k=1}^{m} \lambda_k \phi_k(x_i)\right], \quad i = 1, \cdots, n, \tag{8.11}$$

où

$$Z(\lambda_1, \cdots, \lambda_m) = \sum_{i=1}^{n} p_i \, \exp\left[- \sum_{k=1}^{m} \lambda_k \phi_k(x_i) \right] \tag{8.12}$$

est la fonction de partition, et les $\{\lambda_1, \lambda_2, \cdots, \lambda_n\}$ sont déterminés par le système d'équations (9).

8.3.4 Extension au cas continu

Dans le cas continu, les définitions (8.3) et (8.4) conduisent à :

Problème P3 :

$$\text{maximiser} \quad S(p) = - \int p(x) \, \ln p(x) \, \mathrm{d}x,$$

$$\text{sous les contraintes} \quad \int \phi_k(x) \, p(x) \, \mathrm{d}x = d_k, \quad k = 1, \cdots, m.$$

La solution est

$$p(x) = \frac{1}{Z} \, \exp\left[- \sum_{k=1}^{m} \lambda_k \phi_k(x) \right], \tag{8.13}$$

où

$$Z(\lambda_1, \cdots, \lambda_m) = \int \exp\left[- \sum_{k=1}^{m} \lambda_k \phi_k(x) \right] \, \mathrm{d}x \tag{8.14}$$

est la fonction de partition, et les $\{\lambda_1, \lambda_2, \cdots, \lambda_n\}$ sont déterminés par le système d'équations (9).

Problème P4 :
Étant donné une densité de probabilité *a priori* $p(x)$, déterminer la densité de probabilité *a posteriori* $q(x)$ qui minimise l'entropie croisée

$$S(q, p) = - \int q(x) \, \ln \frac{p(x)}{q(x)} \, \mathrm{d}x,$$

et qui satisfait les m contraintes

$$\int \phi_k(x) \, q(x) \, \mathrm{d}x = d_k, \quad k = 1, \cdots, m.$$

La solution est

$$q(x) = \frac{1}{Z} \, p(x) \, \exp\left[- \sum_{k=1}^{m} \lambda_k \phi_k(x) \right], \tag{8.15}$$

où

$$Z(\lambda_1, \cdots, \lambda_m) = \int p(x) \, \exp\left[- \sum_{k=1}^{m} \lambda_k \phi_k(x) \right] \, \mathrm{d}x \tag{8.16}$$

est la fonction de partition, et les $\{\lambda_1, \lambda_2, \cdots, \lambda_n\}$ sont déterminés par (9).

Notons également que le vecteur $\widehat{\boldsymbol{\lambda}} = (\lambda_1, \cdots, \lambda_m)$ dans tous ces problèmes peut être considéré comme la solution du problème d'optimisation suivant :

Problème dual des problèmes P1-P4 :

$$\widehat{\boldsymbol{\lambda}} = \arg\min_{\boldsymbol{\lambda}} \Big\{ D(\boldsymbol{\lambda}) = \boldsymbol{\lambda}^t \boldsymbol{d} + \ln Z(\boldsymbol{\lambda}) \Big\} \tag{8.17}$$

où $\boldsymbol{d} = (d_1, \cdots, d_m)$. Ce problème d'optimisation est appelé *problème dual* des problèmes P1-P4. Notons que l'expression de $Z(\boldsymbol{\lambda})$ est différente suivant le problème.

Une présentation légèrement différente de ces relations peut être obtenue si on définit

$$\phi_0(x) = 1, \quad \text{et} \quad d_0 = 1,$$

ce qui permet d'inclure la contrainte de normalisation de $q(x)$ et de formuler le problème précédent de la manière suivante :

$$\text{minimiser} \qquad S(q,p) = -\int q(x)\, \ln \frac{p(x)}{q(x)}\, \mathrm{d}x,$$

$$\text{sous les contraintes} \qquad \int \phi_k(x)\, q(x)\, \mathrm{d}x = d_k, \quad k = 0, \cdots, m.$$

La solution peut être écrite sous la forme

$$q(x) = p(x)\, \exp\left[-\sum_{k=0}^{m} \lambda_k \phi_k(x)\right], \tag{8.18}$$

et les $\{\lambda_0, \lambda_1, \lambda_2, \cdots, \lambda_n\}$ sont déterminés par :

$$G(\lambda_0, \lambda_1, \lambda_2, \cdots, \lambda_n) = -\int \phi_k(x)\, p(x)\, \exp\left[-\sum_{l=0}^{m} \lambda_l \phi_l(x)\right] \mathrm{d}x = d_k, \quad k = 0, \cdots, m. \tag{8.19}$$

On remarque que λ_0 est relié à la fonction de partition Z par $Z = -\exp[\lambda_0]$, et on a :

$$\lambda_0 = -\ln Z = -\ln \int p(x) \exp\left[-\sum_{k=1}^{m} \lambda_k \phi_k(x)\right] \mathrm{d}x. \tag{8.20}$$

Les autres coefficients λ_k sont déterminés par le système d'équations :

$$-\frac{\partial}{\partial \lambda_k} \lambda_0(\lambda_1, \cdots, \lambda_n) = d_k, \quad k = 1, \cdots, m. \tag{8.21}$$

Notons aussi que la valeur minimale $S_{\min}(q,p)$ peut être exprimée en fonction des λ_k par :

$$S_{\min}(q,p) = -\lambda_0 - \sum_{k=1}^{m} \lambda_k \phi_k(x). \tag{8.22}$$

Il n'est en général pas possible d'obtenir une relation explicite pour les coefficients λ_k. On résoud alors le système d'équations (8.19) numériquement par des méthodes itératives. Cependant, dans certaines situations simples on peut résoudre le problème d'une façon analytique [3, 6]. Le tableau 1 montre quelques exemples de lois, que l'on obtient sous forme analytique, en résolvant le problème P3 :

$$\text{maximiser} \qquad S(p) = -\int p(x)\, \ln p(x)\, \mathrm{d}x,$$

$$\text{sous les contraintes} \qquad \int \phi_k(x)\, p(x)\, \mathrm{d}x = d_k, \quad k = 1, \cdots, m.$$

Tableau 1: Exemples de lois à maximum d'entropie.

contraintes	domaine de x	loi de probabilité
$\phi_1(x) = x$	$x \in \mathbb{R}_+$	$p(x) = \frac{1}{\mu} \exp[-\mu x]$ loi exponentielle
$\phi_1(x) = \lvert x \rvert$	$x \in \mathbb{R}$	$p(x) = \frac{1}{2\mu} \exp[-\mu \lvert x \rvert]$ loi de LAPLACE
$\begin{cases} \phi_1(x) = x \\ \phi_2(x) = x^2 \end{cases}$	$x \in \mathbb{R}$	$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-m)^2}{2\sigma^2}\right]$ loi de Gausse
$\begin{cases} \phi_1(x) = x \\ \phi_2(x) = \ln x \end{cases}$	$x \in \mathbb{R}_+$	$p(x) \propto \exp[-\lambda \ln x - \mu x]$ $\propto x^{-\lambda} \exp[-\mu x]$ loi Gamma
$\begin{cases} \phi_1(x) = \ln x \\ \phi_2(x) = \ln(1-x) \end{cases}$	$x \in]0, 1[$	$p(x) \propto \exp[-\lambda \ln x - \mu \ln(1-x)]$ $\propto x^{-\lambda}(1-x)^{-\mu}$ loi Béta
$\begin{cases} \phi_1(x) = \ln x \\ \phi_2(x) = x^2 \end{cases}$	$x > 0$	$p(x) \propto \exp[-\lambda \ln x - \mu x^2)]$ $\propto x^{-\lambda} \exp[-\mu x^2]$ loi de Rayleigh

8.3.5 Extension au cas multivariable

Toutes ces relations peuvent être généralisées au cas multivariable. Par exemple le problème P4 devient :

Problème P5 :
Étant donné une densité de probabilité *a priori* $p(\boldsymbol{x})$, déterminer la densité de probabilité *a posteriori* $q(\boldsymbol{x})$ qui minimise l'entropie croisée

$$S(q,p) = -\int q(\boldsymbol{x}) \ln \frac{p(\boldsymbol{x})}{q(\boldsymbol{x})} \, \mathrm{d}\boldsymbol{x},$$

et qui satisfait les m contraintes

$$\int \phi_k(\boldsymbol{x}) \, q(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} = d_k, \quad k = 1, \cdots, m.$$

La solution est

$$q(\boldsymbol{x}) = \frac{1}{Z} p(\boldsymbol{x}) \exp\left[-\sum_{k=1}^{m} \lambda_k \phi_k(\boldsymbol{x})\right], \tag{8.23}$$

où

$$Z(\lambda_1, \cdots, \lambda_m) = \int p(\boldsymbol{x}) \exp\left[-\sum_{k=1}^{m} \lambda_k \phi_k(\boldsymbol{x})\right] \mathrm{d}\boldsymbol{x} \tag{8.24}$$

et les $\{\lambda_1, \lambda_2, \cdots, \lambda_n\}$ sont déterminés par :

$$-\frac{\partial \ln Z(\lambda_1, \cdots, \lambda_m)}{\partial \lambda_k} = d_k, \quad k = 1, \cdots, m. \tag{8.25}$$

Ici aussi, dans certains cas on peut obtenir une relation explicite pour la solution. Par exemple, si $p(\boldsymbol{x})$ est une distribution exponentielle multivariée de la forme séparable

$$p(\boldsymbol{x}) = \prod_{i=1}^{n} \frac{1}{\alpha_i} \exp\left[-\frac{x_i}{\alpha_i}\right],$$

et si les contraintes sont des contraintes sur les moments d'ordre un

$$\mathrm{E}\left[X_i\right] = \int x_i \, q(x_i) \, \mathrm{d}x_i = m_i, \quad i = 1, \cdots, n,$$

alors la solution $q(\boldsymbol{x})$ reste une fonction exponentielle multivariée séparable :

$$q(\boldsymbol{x}) = \prod_{i=1}^{n} \frac{1}{m_i} \exp\left[-\frac{x_i}{m_i}\right].$$

De même si $p(\boldsymbol{x})$ est une fonction gaussienne multivariée séparable :

$$p(\boldsymbol{x}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\alpha_i} \exp\left[-\frac{1}{2}\left(\frac{x_i - m_i}{\alpha_i}\right)^2\right],$$

et si les contraintes sont des contraintes du second ordre :

$$\mathrm{E}\left[(X_i - m_i)^2\right] = \int (x_i - m_i)^2 \, q(x_i) \, \mathrm{d}x_i = \sigma_i^2,$$

alors la solution $q(\boldsymbol{x})$ reste une fonction gaussienne multivariée séparable :

$$q(\boldsymbol{x}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{1}{2}\left(\frac{x_i - m_i}{\sigma_i}\right)^2\right].$$

Ce dernier résultat peut être généralisé au cas où $p(\boldsymbol{x})$ est une fonction gaussienne multivariée non dégénérée :

$$p(\boldsymbol{x}) = \frac{|\boldsymbol{\Sigma}|^{-1/2}}{(2\pi)^{n/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{m})^t \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{m})\right],$$

et si les contraintes sont des contraintes sur les moments d'ordre un et deux non singulières ($\boldsymbol{\Sigma}'$ inversible) :

$$\begin{aligned} \mathrm{E}\left[\boldsymbol{x}\right] &= \int \boldsymbol{x} q(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} = \boldsymbol{m}', \\ \mathrm{E}\left[(\boldsymbol{x} - \boldsymbol{m}')(\boldsymbol{x} - \boldsymbol{m}')^t\right] &= \int (\boldsymbol{x} - \boldsymbol{m}')(\boldsymbol{x} - \boldsymbol{m}')^t q(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} = \boldsymbol{\Sigma}', \end{aligned}$$

alors la solution $q(\boldsymbol{x})$ reste une fonction gaussienne multivariée non dégénérée :

$$q(\boldsymbol{x}) = \frac{|\boldsymbol{\Sigma}'|^{-1/2}}{(2\pi)^{n/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{m}')^t \boldsymbol{\Sigma}'^{-1} (\boldsymbol{x} - \boldsymbol{m}')\right].$$

8.4 Utilisation pour les problèmes inverses

Arrivés à ce stade, nous avons les outils nécessaires pour comprendre comment le principe du maximum d'entropie peut être utilisé dans la résolution des problèmes inverses.

Il existe au moins deux catégories de méthodes qui utilisent le principe du maximum d'entropie pour la résolution des problèmes inverses :

8.4.1 Maximum d'entropie classique

Dans cette approche qui ne s'applique qu'aux objets positifs, l'idée de base consiste à assimiler l'objet $\boldsymbol{f}$ à une distribution de probabilité et les données $\boldsymbol{g}$ à des contraintes linéaires sur celle-ci. Sachant qu'en général, ces contraintes ne sont pas suffisantes pour définir une unique solution au problème, on utilise le PME pour choisir une solution parmi l'ensemble des solutions admissibles qui peuvent être définies soit par

$$\{\boldsymbol{f} : \boldsymbol{H}\,\boldsymbol{f} = \boldsymbol{g}\},$$

ou par

$$\{\boldsymbol{f} : \|\boldsymbol{g} - \boldsymbol{H}\,\boldsymbol{f}\|^2 \leq c\}.$$

On choisit alors dans cet ensemble la solution qui maximise l'entropie

$$-\sum_{j=1}^{N} f_j \log f_j,$$

ou d'une manière plus générale celle qui minimise

$$-\sum_{j=1}^{N} \left[f_j \log\left(\frac{f_j}{m_j}\right) + (f_j - m_j) \right],$$

où $\boldsymbol{m}$ est une solution par défaut (*a priori*).

Notons que dans cette approche on assimile $\boldsymbol{f}$ à une distribution de probabilité et les données $\boldsymbol{g}$ à des contraintes linéaires sur celle-ci.

8.4.2 Maximum d'entropie sur la moyenne

Dans cette approche

- On fait l'hypothèse que $\boldsymbol{f} \in \mathcal{C}$ où $\mathcal{C}$ est un ensemble convexe muni d'une mesure de référence $\mu(\boldsymbol{f})$ (ou d'une densité de probabilité *a priori* $\mathrm{d}\mu(\boldsymbol{f}) = p_0(\boldsymbol{f})\,\mathrm{d}\boldsymbol{f}$ et on suppose que

$$\begin{aligned} \boldsymbol{m} &= \int_{\mathcal{C}} \boldsymbol{f}\,\mathrm{d}\mu(\boldsymbol{f}) \\ &= \int_{\mathcal{C}} \boldsymbol{f}\,p_0(\boldsymbol{f})\,\mathrm{d}\boldsymbol{f}. \end{aligned}$$

- On considère $\boldsymbol{f}$ comme la moyenne d'un vecteur aléatoire $\boldsymbol{F}$ pour lequel on définit une loi de probabilité $P(\boldsymbol{f})$ (ou une densité de probabilité $\mathrm{d}P(\boldsymbol{f}) = p(\boldsymbol{f})\,\mathrm{d}\boldsymbol{f}$:

$$\begin{aligned} \boldsymbol{f} = \mathrm{E}\,[\boldsymbol{F}] &= \int_{\mathcal{C}} \boldsymbol{f}\,\mathrm{d}P(\boldsymbol{f}) \\ &= \int_{\mathcal{C}} \boldsymbol{f}\,p(\boldsymbol{f})\,\mathrm{d}\boldsymbol{f}, \end{aligned}$$

et les données $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f}$ comme des contraintes linéaires sur la loi de probabilité $p(\boldsymbol{f})$:

$$\begin{aligned} \boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} = \boldsymbol{H}\mathrm{E}[\boldsymbol{F}] = \mathrm{E}[\boldsymbol{H}\boldsymbol{F}] &= \int_{\mathcal{C}} \boldsymbol{H}\boldsymbol{f}\, \mathrm{d}P(\boldsymbol{f}) \\ &= \int_{\mathcal{C}} \boldsymbol{H}\boldsymbol{f}\, p(\boldsymbol{f})\, \mathrm{d}\boldsymbol{f}. \end{aligned}$$

– On cherche alors parmi l'ensemble des lois de probabilités satisfaisant ces contraintes celle qui minimise la distance de Kulback entre $\mathrm{d}P(\boldsymbol{f})$ et $\mathrm{d}\mu(\boldsymbol{f})$:

$$-\int \log \frac{\mathrm{d}P(\boldsymbol{f})}{\mathrm{d}\mu(\boldsymbol{f})}\, \mathrm{d}P(\boldsymbol{f}),$$

ou, d'une manière équivalente, celle qui minimise l'entropie relative

$$-\int p(\boldsymbol{f}) \log \left(\frac{p(\boldsymbol{f})}{p_0(\boldsymbol{f})}\right)\, \mathrm{d}\boldsymbol{f}.$$

La solution, nous l'avons vu, peut être obtenue via le Lagrangien :

$$\begin{aligned} \mathcal{L}(\boldsymbol{f}, \boldsymbol{\lambda}) &= \int_{\mathcal{C}} \left[\log \frac{\mathrm{d}P(\boldsymbol{f})}{\mathrm{d}\mu(\boldsymbol{f})} - \sum_{i=1}^{M} \lambda_i (g_i - [\boldsymbol{H}\boldsymbol{f}]_i)\right]\, \mathrm{d}P(\boldsymbol{f}) \\ &= \int_{\mathcal{C}} \left[\log \frac{\mathrm{d}P(\boldsymbol{f})}{\mathrm{d}\mu(\boldsymbol{f})} - \boldsymbol{\lambda}^t(\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f})\right]\, \mathrm{d}P(\boldsymbol{f}) \\ &= \int_{\mathcal{C}} \left[\log \frac{p(\boldsymbol{f})}{p_0(\boldsymbol{f})} - \boldsymbol{\lambda}^t(\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f})\right] p(\boldsymbol{f})\, \mathrm{d}\boldsymbol{f} \end{aligned}$$

et donnée par

$$\mathrm{d}P(\boldsymbol{f}, \boldsymbol{\lambda}) = \exp\left[\boldsymbol{\lambda}^t[\boldsymbol{H}\boldsymbol{f}] - \log Z(\boldsymbol{\lambda})\right]\, \mathrm{d}\mu(\boldsymbol{f}),$$

ou d'une manière équivalente

$$p(\boldsymbol{f}, \boldsymbol{\lambda}) = \exp\left[\boldsymbol{\lambda}^t[\boldsymbol{H}\boldsymbol{f}] - \log Z(\boldsymbol{\lambda})\right] p_0(\boldsymbol{f})\, \mathrm{d}\boldsymbol{f},$$

où

$$Z(\boldsymbol{\lambda}) = \int_{\mathcal{C}} \exp\left[\boldsymbol{\lambda}^t[\boldsymbol{H}\boldsymbol{f}]\right]\, \mathrm{d}\mu(\boldsymbol{f}) = \int_{\mathcal{C}} \exp\left[\boldsymbol{\lambda}^t[\boldsymbol{H}\boldsymbol{f}]\right]\, p_0(\boldsymbol{f})\, \mathrm{d}\boldsymbol{f}.$$

Les paramètres de Lagrange sont obtenus en cherchant la solution du système d'équations non linéaires suivant :

$$\frac{\partial \log Z(\boldsymbol{\lambda})}{\partial \lambda_i} = g_i, \quad i = 1, \cdots, M.$$

– Une fois $p(\boldsymbol{f})$ déterminée on définit la solution du problème par

$$\widehat{\boldsymbol{f}} = \mathrm{E}[\boldsymbol{f}] = \int \boldsymbol{f}\, \mathrm{d}P(\boldsymbol{f}) = \int \boldsymbol{f}\, p(\boldsymbol{f})\, \mathrm{d}\boldsymbol{f}.$$

On peut montrer que, cette solution $\widehat{\boldsymbol{f}}$, qui dépend ainsi de $\boldsymbol{\lambda}$, peut être obtenue directement en minimisant un critère $\Omega(\boldsymbol{f}, \boldsymbol{m})$ sous les contraintes $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f}$.

En effet, en notant

$$\boldsymbol{s} = \boldsymbol{H}^t\boldsymbol{\lambda}, \quad G^*(\boldsymbol{s}) = \log Z(\boldsymbol{s}) = \log \int_{\mathcal{C}} \exp\left[\boldsymbol{s}^t\boldsymbol{f}\right] \, \mathrm{d}\mu(\boldsymbol{f}),$$

et

$$\Omega(\boldsymbol{f}) = \inf_{\boldsymbol{s}}\{\boldsymbol{s}^t\boldsymbol{f} - G^*(\boldsymbol{s})\}, \quad D(\boldsymbol{\lambda}) = \boldsymbol{\lambda}^t\boldsymbol{g} - G^*(\boldsymbol{H}^t\boldsymbol{\lambda}),$$

on peut alors montrer que :

- Le vecteur de $\boldsymbol{\lambda}$ optimal peut être calculé en cherchant l'optimum de $D(\boldsymbol{\lambda})$:

$$\widehat{\boldsymbol{\lambda}} = \arg\max_{\boldsymbol{\lambda}} \{D(\boldsymbol{\lambda})\}$$

  Le critère $D(\boldsymbol{\lambda})$ est appelé *Critère dual* ;

- La solution du problème $\widehat{\boldsymbol{f}}$ peut être calculée en cherchant l'optimum d'un critère $\Omega(\boldsymbol{f}, \boldsymbol{m})$ appelé *critère primal* :

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}\in\mathcal{C}} \{H(\boldsymbol{f}, \boldsymbol{m})\} \quad \text{sous les contraintes} \quad \boldsymbol{g} = \boldsymbol{H}\boldsymbol{f};$$

- Il existe une relation explicite pour calculer la solution en fonction de $G^*(\boldsymbol{s})$:

$$\widehat{\boldsymbol{f}}(\boldsymbol{s}) = \frac{\mathrm{d}G^*(\boldsymbol{s})}{\mathrm{d}\boldsymbol{s}},$$

  et où :

- les fonctions G et Ω dépendent de la mesure de référence $\mu(\boldsymbol{f})$;

- $D(\boldsymbol{\lambda})$ est appelé le *critère dual* et dépend des données $\boldsymbol{g}$ et de la fonction G ;

- $\Omega(\boldsymbol{f}, \boldsymbol{m})$ est appelé le *critère primal* et peut être considérer comme une mesure de distance entre $\boldsymbol{f}$ et $\boldsymbol{m}$, ce qui signifie qu'elle satisfait les conditions suivantes :

  - $\Omega(\boldsymbol{f}, \boldsymbol{m}) \geq 0, \quad \text{et} \quad \Omega(\boldsymbol{f}, \boldsymbol{m}) = 0 \quad \text{ssi} \quad \boldsymbol{f} = \boldsymbol{m}$;
  - $\Omega(\boldsymbol{f}, \boldsymbol{m})$ est continue et convexe sur $\mathcal{C}$;
  - $\Omega(\boldsymbol{f}, \boldsymbol{m}) = \infty \quad \text{si } \boldsymbol{f} \notin \mathcal{C}$.

Lorsque la mesure de référence est séparable

$$\mu(\boldsymbol{f}) = \prod_{j=1}^{N} \mu_j(f_j),$$

on a aussi :

$$\mathrm{d}P(\boldsymbol{f}, \boldsymbol{\lambda}) = \prod_{j=1}^{N} \mathrm{d}P_j(f_j, \boldsymbol{\lambda}),$$

et

$$G(\boldsymbol{s}) = \sum_j g_j(s_j), \quad \Omega(\boldsymbol{f}, \boldsymbol{m}) = \sum_j h_j(f_j, m_j), \quad \widehat{x}_j = g_j'(s_j).$$

En remplaçant $\boldsymbol{s} = \boldsymbol{H}^t\boldsymbol{\lambda}$ on obtient :

$$G(\boldsymbol{\lambda}) = \sum_j g_j\left([\boldsymbol{H}^t\boldsymbol{\lambda}]_j\right), \quad \Omega(\boldsymbol{f},\boldsymbol{m}) = \sum_j h_j(f_j,m_j), \quad \widehat{x}_j = g_j'\left([\boldsymbol{H}^t\widehat{\boldsymbol{\lambda}}]_j\right).$$

On peut voir que $\Omega(\boldsymbol{f},\boldsymbol{m})$ est aussi séparable et que h_j et g_j dépendent de la mesure de référence $\mu_j(f_j)$:

- g_j est la transformée de Cramer (log de la transformée de LAPLACE) de μ_j:

$$g_j(s) = \log \int \exp\left[sf_j\right] \mathrm{d}\mu_j(f_j);$$

- h_j est la convexe conjugée de g_j: $\quad h_j(f) = \inf_s\{sf - g_j(s)\}$.

Prenons quelques exemples :

	$\mu_j(f)$	$g_j(s)$	$h_j(f,m)$
Gaussien :	$\exp\left[-\frac{1}{2}(f-m)^2\right]$	$\frac{1}{2}(s-m)^2$	$\frac{1}{2}(f-m)^2$
Poisson :	$\frac{m^f}{f!}\exp\left[-m\right]$	$\exp\left[m-s\right]$	$-f\log\frac{f}{m}+m-f$
Gamma :	$f^{\alpha-1}\exp\left[-\frac{f}{m}\right]$	$\log(s-m)$	$-\log\frac{f}{m}+\frac{f}{m}-1$

Ainsi, si on considère le critère primal :

$$\text{maximiser} \quad \Omega(\boldsymbol{f},\boldsymbol{m}) = \sum_{j=1}^{N} h_j(f_j,m_j) \quad \text{sous contraintes} \quad \boldsymbol{H}\boldsymbol{f} = \boldsymbol{g},$$

on peut par exemple remarquer que lorsque la mesure de référence est une mesure poissonnienne on a $h_j(f_j,m_j) = -f_j \ln f_j/m_j + (m_j - f_j)$. On peut ainsi faire le parallèle avec la méthode du ME classique.

Cependant, lorsque la mesure de référence n'est pas séparable ou, lorsqu'il y a de l'incertitude sur les données $\boldsymbol{g}$, il n'est plus aussi facile d'utiliser cette approche. Mentionnons quand même que cette approche est récente, et des travaux tout à fait récents ont essayé de prendre en compte le bruit. Ici nous en présentons une approche qui consiste à remplacer $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{n}$ par

$$\boldsymbol{g} = [\boldsymbol{H}|\boldsymbol{I}]\begin{bmatrix}\boldsymbol{f}\\ \boldsymbol{n}\end{bmatrix} \longrightarrow \boldsymbol{g} = \tilde{\boldsymbol{H}}\,\tilde{\boldsymbol{f}}$$

et faire l'hypothèse d'indépendance entre $\boldsymbol{b}$ et $\boldsymbol{f}$:

$$\mu(\tilde{\boldsymbol{f}}) = \mu_x(\boldsymbol{f})\mu_n(\boldsymbol{n})$$

Ensuite, suivant l'approche, on montre que la solution de l'optimisation du critère primal devient :

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}\in\mathcal{C}}\{\mathcal{Q}(\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}) + \alpha\Omega(\boldsymbol{f},\boldsymbol{m})\}$$

avec

$$\Omega(\boldsymbol{f},\boldsymbol{m}) = \sum_{j=1}^{N} h_j(f_j,m_j), \qquad \text{et} \qquad \mathcal{Q}(\boldsymbol{z}) = \sum_{i=1}^{M} q_i(z_j).$$

Là encore, les fonctions $h_j(f_j)$ et $q_i(z_i)$ dépendent des mesures de références $\mu_f(\boldsymbol{f})$ et $\mu_n(\boldsymbol{f})$.