BE Numéro 2: Méthodes probabilistes pour les problèmes inverses
Cours: Problèmes inverses   Professeur: A. Mohammad–Djafari

Problème 1 : Dans un système d'imagerie, nous avons établi la relation $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\epsilon}$ où

- $\boldsymbol{y}$ est un vecteur contenant les projections (mesures) $\{y_m,\, m = 1 \cdots, M\}$,
- $\boldsymbol{\epsilon}$ est un vecteur représentant les erreurs de mesures $\{\epsilon_m,\, m = 1 \cdots, M\}$,
- $\boldsymbol{x}$ est un vecteur représentant les pixels de l'image $\{x_n,\, n = 1 \cdots, N\}$, et
- $\boldsymbol{A}$ est une matrice dont les éléments $\{a_{mn}\}$ dépendent de la géométrie du système et sont supposés connus.

1. Supposons d'abord que $M = N$ et que la matrice $\boldsymbol{A}$ soit inversible. Pourquoi la solution $\widehat{\boldsymbol{x}}_0 = \boldsymbol{A}^{-1}\boldsymbol{y}$ n'est, en général, pas une solution satisfaisante ?

   Quelle relation existe–t–il entre $\frac{\|\delta\widehat{\boldsymbol{x}}_0\|}{\|\widehat{\boldsymbol{x}}_0\|}$ et $\frac{\|\delta\boldsymbol{y}\|}{\|\boldsymbol{y}\|}$?

2. Revenons maintenant au cas général $M \neq N$. Montrez alors qu'on peut définir des solutions au sens des moindres carrés (MC), *i.e.* $\widehat{\boldsymbol{x}}_1$ qui minimise

$$J_1(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|^2$$

   Montrer que, toute solution de l'équation $\boldsymbol{A}^t\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^t\boldsymbol{y}$ est une solution au sens des moindres carrés du problème et lorsque $\boldsymbol{A}^t\boldsymbol{A}$ est inversible il existe une solution unique donnée par

$$\widehat{\boldsymbol{x}}_1 = [\boldsymbol{A}^t\boldsymbol{A}]^{-1}\boldsymbol{A}^t\boldsymbol{y}$$

   Quelle relation existe–t–il alors entre $\frac{\|\delta\widehat{\boldsymbol{x}}_1\|}{\|\widehat{\boldsymbol{x}}_1\|}$ et $\frac{\|\delta\boldsymbol{y}\|}{\|\boldsymbol{y}\|}$?

3. Une manière d'obtenir une solution régularisée $\widehat{\boldsymbol{x}}_2$ à ce problème est de définir et de minimiser un critère de la forme

$$J_2(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|^2 + \lambda\|\boldsymbol{D}\boldsymbol{x}\|^2,$$

   où $\boldsymbol{D}$ est une matrice approximant un opérateur linéaire de dérivation.

   Montrer que cette solution s'écrit:

$$\widehat{\boldsymbol{x}}_2 = \arg\min_{\boldsymbol{x}} \{J_2(\boldsymbol{x})\} = \left[\boldsymbol{A}^t\boldsymbol{A} + \lambda\boldsymbol{D}^t\boldsymbol{D}\right]^{-1}\boldsymbol{A}^t\boldsymbol{y}$$

   Pourquoi cette solution est-elle préférable à $\widehat{\boldsymbol{x}}_0$ et à $\widehat{\boldsymbol{x}}_1$?

4. Supposons que $\boldsymbol{A}$ et $\boldsymbol{D}$ soient des matrices circulantes et symétriques. Dans ce cas montrer que la solution régularisée $\widehat{\boldsymbol{x}}_2$ peut être obtenue à l'aide de la TFD (Transformée de Fourier Discrète). Plus précisemment montrer la relation suivante:

$$X(\omega) = \frac{1}{H(\omega)}\,\frac{|H(\omega)|^2}{|H(\omega)|^2 + \lambda|D(\omega)|^2}Y(\omega)$$

   où

   - $H(\omega)$ est la TFD de la première ligne de la matrice $\boldsymbol{A}$,
   - $D(\omega)$ est la TFD de la première ligne de la matrice $\boldsymbol{D}$
   - $X(\omega)$ est la TFD du vecteur solution $\widehat{\boldsymbol{x}}_2$, et
   - $Y(\omega)$ est la TFD du vecteur de mesure $\boldsymbol{y}$.

5. Commenter les expressions de $\widehat{\boldsymbol{x}}_2$ dans la question 3. et de $X(\omega)$ dans la question 4. lorsque $\lambda = 0$ et lorsque $\lambda \longrightarrow \infty$.

6. Dans une approche d'estimation au sens du maximum de vraisemblance (MV) supposons que le bruit $\boldsymbol{\epsilon}$ soit supposé blanc, centré et gaussien. Montrez que l'estimation au sens du MV de $\boldsymbol{x}$, *i.e.*;

$$\widehat{\boldsymbol{x}}_{\text{MV}} = \arg\max_{\boldsymbol{x}} \{p(\boldsymbol{y}|\boldsymbol{x})\}$$

   s'obtient en minimisant

$$J_1(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|^2.$$

7. Quelle est l'expression de cet estimateur si $\boldsymbol{\epsilon}$ est supposé suivre une loi gaussienne généralisée, *i.e.*;

$$\boldsymbol{\epsilon} \sim p(\boldsymbol{\epsilon}) \propto \exp\left[-\|\boldsymbol{\epsilon}\|^\alpha\right] = \exp\left[-\sum_m |\epsilon_m|^\alpha\right], \quad 1 < \alpha \leq 2$$

8. Dans une approche bayésienne pour résoudre ce problème, supposons que l'on puisse attribuer des lois gaussiennes aux deux vecteurs $\boldsymbol{\epsilon}$ et $\boldsymbol{x}$:

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{R}_\epsilon), \qquad \boldsymbol{x} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{R}_X),$$

où $\boldsymbol{R}_\epsilon = \sigma_\epsilon^2 \boldsymbol{I}$ et $\boldsymbol{R}_X = \sigma_x^2 \boldsymbol{P}_0 = \sigma_x^2 (\boldsymbol{D}^t \boldsymbol{D})^{-1}$ sont les matrices de covariance de $\boldsymbol{\epsilon}$ et de $\boldsymbol{x}$.
Ecrivez l'expression des lois $p(\boldsymbol{x}), p(\boldsymbol{\epsilon}), p(\boldsymbol{y}, \boldsymbol{x})$ et $p(\boldsymbol{y}|\boldsymbol{x})$.

9. Montrer que la loi *a posteriori* $p(\boldsymbol{x}|\boldsymbol{y})$ est une loi gaussienne de la forme

$$p(\boldsymbol{x}|\boldsymbol{y}) \propto \exp\left[-\frac{1}{2}\left[[\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}]^t \boldsymbol{R}_\epsilon^{-1} [\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}] + \boldsymbol{x}^t \boldsymbol{R}_X^{-1} \boldsymbol{x}\right]\right]$$

10. Si on note $\widehat{\boldsymbol{x}}_3$ la solution qui maximise $p(\boldsymbol{x}|\boldsymbol{y})$ (l'estimation au sens du maximum *a posteriori* MAP), montrer qu'elle s'obtient par

$$\widehat{\boldsymbol{x}}_3 = \arg\min_{\boldsymbol{x}} \left\{ J_3(\boldsymbol{x}) = [\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}]^t \boldsymbol{R}_\epsilon^{-1} [\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}] + \boldsymbol{x}^t \boldsymbol{R}_X^{-1} \boldsymbol{x} \right\}$$

11. Si $\boldsymbol{R}_\epsilon = \sigma_\epsilon^2 \boldsymbol{I}$ et $\boldsymbol{R}_X = \sigma_x^2 \boldsymbol{P}_0$, montrer que $\widehat{\boldsymbol{x}}_3$ s'obtient par

$$\widehat{\boldsymbol{x}}_3 = \left[\boldsymbol{A}^t \boldsymbol{A} + \lambda \boldsymbol{P}_0^{-1}\right]^{-1} \boldsymbol{A}^t \boldsymbol{y} \quad \text{avec} \quad \lambda = \frac{\sigma_\epsilon^2}{\sigma_x^2}$$

Que peut–on conclure en comparant les solutions $\widehat{\boldsymbol{x}}_2$ et $\widehat{\boldsymbol{x}}_3$?

12. En développant le terme

$$\left[[\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}]^t \boldsymbol{R}_\epsilon^{-1} [\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}] + \boldsymbol{x}^t \boldsymbol{R}_X^{-1} \boldsymbol{x}\right]$$

dans l'expression de $p(\boldsymbol{x}|\boldsymbol{y})$ montrer que l'on peut écrire

$$p(\boldsymbol{x}|\boldsymbol{y}) \propto \exp\left[-\frac{1}{2}[\boldsymbol{x} - \widehat{\boldsymbol{x}}_3]^t \widehat{\boldsymbol{P}}^{-1} [\boldsymbol{x} - \widehat{\boldsymbol{x}}_3]\right]$$

Quelle est alors l'expression de la matrice de covariance *a posteriori* $\widehat{\boldsymbol{P}}$?
Que représentent les éléments diagonaux de cette matrice?

13. Ecrivez l'expression des lois $p(x_n|\boldsymbol{y}), p(y_m|\boldsymbol{y})$ et $p(b_m|\boldsymbol{y})$. A qui peuvent-elle servir ?

14. Supposons maintenant la suite $\{x_1, \cdots, x_N\}$ puissent être modélisée par une chaîne de Markov d'ordre un, c'est à dire :

$$p(x_n|x_1, \cdots, x_N) = p(x_n|x_{n-1})$$

Peut-on calculer $p(\boldsymbol{x})$? Et si on connait de plus $p(x_1)$?
Que devient alors la solution au sens du MAP ?
Etudiez cette solution dans les deux cas suivants:

$$p(x_n|x_1, \cdots, x_N) = p(x_n|x_{n-1}) = \mathcal{N}(x_{n-1}, \sigma_x^2), \quad \text{et} \quad p(x_1) = \mathcal{N}(0, \sigma_x^2)$$

et

$$p(x_n|x_1, \cdots, x_N) = p(x_n|x_{n-1}) \propto \exp\left[-\alpha\phi(x_n - x_{n-1})\right], \quad \text{et} \quad p(x_1) \propto \exp\left[-\alpha\phi(x_1)\right]$$

15. Supposons maintenant que $x_n > 0$ et que nous puissions faire l'hypothèse que les x_n suivent une lois Gamma :

$$p(x_n) = \frac{1}{Z(\alpha, \beta)} x_n^{-\alpha} \exp\left[-\beta x_n\right], \quad x_n > 0, \alpha > -1, \beta > 0, \quad Z(\alpha, \beta) = \frac{\Gamma(\alpha + 1)}{\beta^{(\alpha+1)}}$$

et que x_k et x_n sont indépendantes.
Que devient alors la solution au sens du MAP ?
Peut-on donner une expression explicite à cette solution ?
Comment peut-on calculer cette solution ?

16. Supposons maintenant que l'on nous donne $\boldsymbol{x}$ et on nous demande d'estimer les paramètres (α, β). Considérons d'abord la méthode du maximum de vraisemblance (MV) :

$$(\widehat{\alpha}, \widehat{\beta}) = \arg\max_{(\alpha,\beta)} \{p(\boldsymbol{x}|\alpha, \beta)\}$$

Montrer que dans ce cas nous avons à résoudre le systèmes d'équations suivantes :

$$\left\{ \begin{array}{l} -\frac{\partial \ln Z(\alpha,\beta)}{\partial \alpha} = \frac{1}{N}\sum_{n=1}^N \log x_n \\ -\frac{\partial \ln Z(\alpha,\beta)}{\partial \beta} = \frac{1}{N}\sum_{n=1}^N x_n \end{array} \right.$$

où

$$Z(\alpha, \beta) = \int x^\alpha \exp\left[-\beta x\right] \, \mathrm{d}x,$$

Considérons ensuite la méthode des moments (MM) qui consiste à résoudre le système d'équations :

$$\begin{aligned} \mathrm{E}\{X\} &= \int x \, p(x; \alpha, \beta) \, \mathrm{d}x = \mu = \frac{1}{N}\sum_{n=1}^N x_n \\ \mathrm{E}\left\{(X-m)^2\right\} &= \int (x-m)^2 \, p(x; \alpha, \beta) \, \mathrm{d}x = v = \frac{1}{N}\sum_{n=1}^N (x_n - m)^2 \end{aligned}$$

Montrer qu'il y a une relation analytique entre (μ, v) et (α, β) qui est

$$\left\{ \begin{array}{l} \mu = \frac{(1-\alpha)}{\beta} \\ v = \frac{(1-\alpha)}{\beta^2} \end{array} \right., \left\{ \begin{array}{l} \alpha = \frac{(v-\mu^2)}{v} \\ \beta = \frac{\mu}{v} \end{array} \right.$$

17. Supposons maintenant que nous voulons estimer à la fois $\boldsymbol{x}$ et les paramètres $\boldsymbol{\theta} = (\alpha, \beta)$ à partir des données $\boldsymbol{y}$ par la méthode du maximum de vraisemblance généralisée (MVG) :

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

La maximisation conjointe par rapport à $\boldsymbol{x}$ et $\boldsymbol{\theta}$ étant difficile, on peut envisager une optimisation (sous-optimal) par l'algorithme itératif suivant:

$$\left\{ \begin{array}{lcl} \widehat{\boldsymbol{\theta}}^{(k)} &=& \arg\max_{\boldsymbol{\theta}} \left\{p(\widehat{\boldsymbol{x}}^{(k)}, \boldsymbol{\theta}|\boldsymbol{y})\right\} \\ \widehat{\boldsymbol{x}}^{(k+1)} &=& \arg\max_{\boldsymbol{x}} \left\{p(\boldsymbol{x}, \widehat{\boldsymbol{\theta}}^{(k)}|\boldsymbol{y})\right\} \end{array} \right.$$

Montrer que si on choisit $p(\boldsymbol{\theta})$ uniforme cet algorithm devient:

$$\left\{ \begin{array}{lcl} \widehat{\boldsymbol{\theta}}^{(k)} &=& \arg\max_{\boldsymbol{\theta}} \left\{p(\widehat{\boldsymbol{x}}^{(k)}|\boldsymbol{\theta})\right\} \\ \widehat{\boldsymbol{x}}^{(k+1)} &=& \arg\max_{\boldsymbol{x}} \left\{p(\boldsymbol{x}|\boldsymbol{y}, \widehat{\boldsymbol{\theta}}^{(k)})\right\} \end{array} \right.$$

où dans l'étape 1 pour obtenir $\widehat{\boldsymbol{\theta}}^{(k)} = (\widehat{\alpha}^{(k)}, \widehat{\beta}^{(k)})$ on aura à résoudre

$$\left\{ \begin{array}{l} -\frac{\partial \ln Z(\alpha,\beta)}{\partial \alpha} = \frac{1}{N}\sum_{n=1}^N \log \widehat{x}_n^{(k)} \\ -\frac{\partial \ln Z(\alpha,\beta)}{\partial \beta} = \frac{1}{N}\sum_{n=1}^N \widehat{x}_n^{(k)} \end{array} \right.$$

et dans létape 2 on aura à trouver le minimum du critère

$$J(\boldsymbol{x}) = ||\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}||^2 + \lambda \sum_{n=1}^N \log x_n + \mu \sum_{n=1}^N x_n$$

avec $\lambda = \alpha\sigma_\epsilon^2$ et $\mu = \beta\sigma_\epsilon^2$.

Étudier la structure et les difficultées de la mise en œuvre de cet algorithme.

18. Supposons maintenant que nous voulons estimer d'abord les paramètres $\boldsymbol{\theta} = (\alpha, \beta)$ à partir des données $\boldsymbol{y}$ par la méthode du maximum de vraisemblance :

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \{p(\boldsymbol{y}|\boldsymbol{\theta})\}$$

et ensuite l'utiliser pour l'estimation de $\boldsymbol{x}$.
Étudier l'algorithme EM pour l'estimation $\boldsymbol{\theta}$.

19. Revenons au problème de départ, mais cette foi, nous allons modéliser chaque pixel de l'image par un modèle de mélange de gaussiennes.

$$p(x_n) = \sum_{k=1}^{K} p(z_n = k)\mathcal{N}(m_k, \sigma_k^2) = \sum_{k=1}^{K} \alpha_k \mathcal{N}(m_k, \sigma_k^2)$$

avec z_n une variable qui prend des valeurs discrètes $1, 2, \ldots, K$.
– Étudier, dans un premier temps l'estimation *supervisée* de $\boldsymbol{\theta} = \{(\alpha_k, m_k, \sigma_k^2),\ k = 1, \ldots, K\}$ à partir de l'observation directe de $\boldsymbol{x}$.
– Ensuite étudier l'estimation *non supervisée* de $\boldsymbol{\theta}$ à partir de $\boldsymbol{y}$.

20. Modélisation hirarchique: Dans la modélisation précente, on peut remarquer que

$$x_n|z_n = k \sim \mathcal{N}(m_k, \sigma_k^2)$$

On peut alors construire la structure hirarchique suivante :

$$\begin{array}{lcl} \boldsymbol{y}|\boldsymbol{x} & \sim & \mathcal{N}(\boldsymbol{A}\boldsymbol{x}, \sigma_\epsilon^2 \boldsymbol{I}) \\ x_n|z_n & \sim & \mathcal{N}(m_k, \sigma_k^2) \\ z_n & \sim & P(z_n = k) = \alpha_k \end{array}$$

– Montrez que la loi *a posteriori* $p(x_n|\boldsymbol{y})$ est aussi une loi de mélange de gaussienne

$$p(x_n|\boldsymbol{y}) = \sum_{k=1}^{K} \alpha\prime_k \, \mathcal{N}(m\prime_k, \sigma\prime_k^2)$$

– Exprimer $(\alpha\prime_k, m\prime_k, \sigma\prime_k)$ en fonction de des données $\boldsymbol{y}$ et $(\alpha_k, m_k, \sigma_k)$ et proposez un algorithme d'estimation pour $x_n|\boldsymbol{y}$.
– Trouvez l'expression de $P(z_n = k|\boldsymbol{y})$ et proposer un algorithme de classification (estimation de $z_n|\boldsymbol{y}$).
– Étudiez la structure d'un algorithme d'estimation hirarchique du type itératif

$$\left\{ \begin{array}{lcl} \widehat{\boldsymbol{z}}^{(k)} & = & \arg\max_{\boldsymbol{z}} \left\{ \boldsymbol{z}|\boldsymbol{y}, \widehat{\boldsymbol{x}}^{(k-1)}) \right\} \\ \widehat{\boldsymbol{x}}^{(k+1)} & = & \arg\max_{\boldsymbol{x}} \left\{ p(\boldsymbol{x}|\boldsymbol{y}, \widehat{\boldsymbol{z}}^{(k)}) \right\} \end{array} \right.$$

ou

$$\widehat{\boldsymbol{z}} = \arg\max_{\boldsymbol{z}} \{\boldsymbol{z}|\boldsymbol{y})\} \longrightarrow \widehat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} \{p(\boldsymbol{x}|\boldsymbol{y}, \widehat{\boldsymbol{z}})\}$$

tout en notant que $\widehat{\boldsymbol{z}} = \arg\max_{\boldsymbol{z}} \{\boldsymbol{z}|\boldsymbol{y})\}$ ne peut se faire que d'une manière itérative par l'algorithme EM.
– Comparer le coût de calcul de chacun de ces deux algorithmes.

Problème 2 : Considérons maintenant le problème sans bruit $\quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$.

1. Supposons que nous savons que $\boldsymbol{x}$ représente une quantité positive. Nous cherchons alors une solution positive à ce problème. L'outil "entropie" peut alors être utilisé. Considérons le problème d'optimisation suivant :

$$\max \left(-\sum_{n=1}^{N} x_n \log x_n + \sum_{n=1}^{N} x_n \right) \quad \text{sous les contraintes} \;\; y_m = \sum_{n=1}^{N} a_{m,n} x_n, \quad m = 1, \cdots, M.$$

   En utilisant le Lagrangien

$$\mathcal{L} = -\sum_{n=1}^{N} x_n \log x_n + \sum_{n=1}^{N} x_n + \sum_{m=1}^{M} \lambda_m \left(y_m - \sum_{n=1}^{N} a_{m,n} x_n \right)$$

   montrer que ce problème admet une solution unique (lorsqu'elle existe) qui est de la forme

$$x_n = \exp\left[-\sum_{m=1}^{M} a_{m,n} \lambda_m \right],$$

   où $\{\lambda_m, m = 1, \cdots, M\}$ sont solution du système d'équations

$$y_m = \sum_{n=1}^{N} a_{m,n} \exp\left[-\sum_{m=1}^{M} a_{m,n} \lambda_m \right].$$

   Interpréter cette solution. Proposez un algorithme pour le calculer.

2. Considérons maintenant que $\boldsymbol{x}$ représente une image moyenne, *i.e.* chaque x_n corresspond à l'espérance d'une grandeur $Z_n > 0$. Supposons que le vecteur aléatoire $\boldsymbol{Z}$ admet une densité de probabilité $p_z(\boldsymbol{z})$. Les mesures y_m peuvent alors être considérées comme une combinaison linéaire de $\mathrm{E}\{Z_n\}$, *i.e.*

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \longrightarrow y_m = \sum_{n=1}^{N} a_{m,n} x_n = \sum_{n=1}^{N} a_{m,n} \mathrm{E}\{Z_n\}, \quad m = 1, \cdots, M.$$

   On cherchera alors la densité de probabilité $p(\boldsymbol{z})$ qui satisfait ces contraintes et qui maximise

$$-\int p(\boldsymbol{z}) \log\left(\frac{p(\boldsymbol{z})}{\mu(\boldsymbol{z})} \right) \mathrm{d}\boldsymbol{z}.$$

   Pour cela il faut définir le Lagrangien

$$\mathcal{L} = -\int p(\boldsymbol{z}) \log\left(\frac{p(\boldsymbol{z})}{\mu(\boldsymbol{z})} \right) \mathrm{d}\boldsymbol{z} - \sum_{m=1}^{M} \lambda_m \left(y_m - \sum_{n=1}^{N} a_{m,n} \int z_n \, p(\boldsymbol{z}) \, \mathrm{d}\boldsymbol{z} \right)$$

   et le dériver par rapport à p.

   - Montrer alors que $p(\boldsymbol{x})$ est de la forme

$$p(\boldsymbol{x}) \propto \mu(\boldsymbol{z}) \exp\left[\sum_{m=1}^{M} \lambda_m \sum_{n=1}^{N} a_{m,n} z_n \right]$$

   - Montrer ensuite que lorsque $\mu(\boldsymbol{z})$ est séparable, *i.e.*; $\mu(\boldsymbol{z}) = \prod_{j=1}^{N} \mu(z_n)$, alors $p(\boldsymbol{z})$ l'est aussi et si $\mu(z_n) \propto \exp\left[-\frac{1}{2} z_n^2\right]$ on a

$$p(\boldsymbol{z}) = \prod_{j=1}^{N} p(z_n) \quad \text{avec} \quad p(z_n) \propto \exp\left[-\frac{1}{2} z_n^2 + \sum_{m=1}^{M} a_{m,n} \lambda_m z_n \right],$$

- Notant maintenant que $x_n = \mathrm{E}\{Z_n\}$. Montrez que x_n est de la forme

$$x_n = \sum_{m=1}^{M} a_{m,n}\lambda_m, \quad \text{ou encore} \quad \boldsymbol{x} = \boldsymbol{A}^t\boldsymbol{\lambda}$$

où $\{\lambda_m\}$ sont solution de

$$y_m = \sum_{n=1}^{N} a_{m,n} \sum_{m=1}^{M} a_{m,n}\lambda_m, \quad \text{ou encore} \quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{A}^t\boldsymbol{\lambda}$$

ce qui permet d'écrire

$$\boldsymbol{x} = \boldsymbol{A}^t(\boldsymbol{A}\boldsymbol{A}^t)^{-1}\boldsymbol{y}$$

Interpréter cette solution.

- Que deviennent $\boldsymbol{x}$ si $\mu(z_n) \propto \exp\left[-z_n\right], \quad z_n > 0$?
Pour cela montrer que

$$p(z_n) \propto \exp\left[-z_n - \sum_{m=1}^{M} a_{m,n}\lambda_m z_n\right], \quad z_n > 0$$

et par conséquent

$$x_n = 1 + \sum_{m=1}^{M} a_{m,n}\lambda_m, \; \text{ou encore} \; \boldsymbol{x} = \boldsymbol{1} + \boldsymbol{A}^t\boldsymbol{\lambda}$$

où $\{\lambda_m\}$ sont solution du système d'équations

$$y_m = \sum_{n=1}^{N} a_{m,n}(1 + \sum_{i=0}^{M} a_{m,n}\lambda_m), \; \text{ou encore} \; \boldsymbol{y} = \boldsymbol{A}\boldsymbol{1} + \boldsymbol{A}\boldsymbol{A}^t\boldsymbol{\lambda},$$

ce qui permet d'écrire (supposons que $\boldsymbol{A}\boldsymbol{A}^t$ est inversible)

$$\boldsymbol{x} = \boldsymbol{1} + \boldsymbol{A}^t[\boldsymbol{A}\boldsymbol{A}^t]^{-1}(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{1}) = \boldsymbol{A}^t[\boldsymbol{A}\boldsymbol{A}^t]^{-1}\boldsymbol{y} + \left(\boldsymbol{I} - \boldsymbol{A}^t[\boldsymbol{A}\boldsymbol{A}^t]^{-1}\boldsymbol{A}\right)\boldsymbol{1}$$

Interpréter alors ce resultat.

- Que deviennent $\boldsymbol{x}$ si $\mu(z_n) \propto z_n^{(\alpha-1)}, \quad z_n > 0$?
Pour cela montrer que

$$p(z_n) \propto \exp\left[(\alpha - 1)\log z_n - \sum_{m=1}^{M} a_{m,n}\lambda_m z_n\right], \quad z_n > 0$$

qui est une loi *Gamma*(α, β) avec $\beta = \sum_{m=1}^{M} a_{m,n}\lambda_m$ et par conséquent

$$x_n = \mathrm{E}\{Z_n\} = \frac{\alpha}{\beta} = \frac{\alpha}{\sum_{m=1}^{M} a_{m,n}\lambda_m},$$

où $\{\lambda_m\}$ sont solution du système d'équations

$$y_m = \sum_{n=1}^{N} a_{m,n}\frac{\alpha}{\sum_{m=1}^{M} a_{m,n}\lambda_m}$$

ce qui permet d'écrire (symboliquement ou avec les notations Matlab)

$$\begin{aligned} \boldsymbol{x} &= \frac{\alpha}{\boldsymbol{A}^t\boldsymbol{\lambda}} \qquad \text{ou encore} \qquad \boldsymbol{x} = \alpha\boldsymbol{1}./\boldsymbol{A}^t\boldsymbol{\lambda} \\ \boldsymbol{y} &= \boldsymbol{A}\frac{\alpha}{\boldsymbol{A}^t\boldsymbol{\lambda}} \qquad \text{ou encore} \qquad \boldsymbol{y} = \boldsymbol{A}(\alpha\boldsymbol{1}./\boldsymbol{A}^t\boldsymbol{\lambda}) = \alpha(\boldsymbol{A}\boldsymbol{1})./(\boldsymbol{A}^t\boldsymbol{\lambda}) \end{aligned}$$

Interpréter alors ce resultat.

- Que deviennent $\boldsymbol{x}$ si $\mu(z_n) \propto z_n^{(\alpha-1)}\exp\left[\beta z_n\right], \quad z_n > 0$?