Exercice numéro 5: Moindres carrés, régularisation, maximum d'entropie

Problème 1 : Dans un système d'imagerie, nous avons établi la relation $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}$ où
$\boldsymbol{g}$ est un vecteur contenant les projections (mesures) $\{g_m,\ m = 1 \cdots, M\}$,
$\boldsymbol{\epsilon}$ est un vecteur représentant les erreurs de mesures $\{\epsilon_m,\ m = 1 \cdots, M\}$,
$\boldsymbol{f}$ est un vecteur représentant les pixels de l'image $\{f_n,\ n = 1 \cdots, N\}$, et
$\boldsymbol{H}$ 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 $\boldsymbol{H}$ soit inversible. Pourquoi la solution $\widehat{\boldsymbol{f}}_0 = \boldsymbol{H}^{-1}\boldsymbol{g}$ n'est, en général, pas une solution satisfaisante ?

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

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.* $\widehat{\boldsymbol{f}}_1$ qui minimise
$$J_1(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2$$
   Montrez que, toute solution de l'équation $\boldsymbol{H}'\boldsymbol{H}\boldsymbol{f} = \boldsymbol{H}'\boldsymbol{g}$ est une solution au sens des moindres carrés du problème et que lorsque $\boldsymbol{H}'\boldsymbol{H}$ est inversible il existe une solution unique donnée par
$$\widehat{\boldsymbol{f}}_1 = [\boldsymbol{H}'\boldsymbol{H}]^{-1}\boldsymbol{H}'\boldsymbol{g}$$
   Quelle relation existe–t–il alors entre $\frac{\|\delta\widehat{\boldsymbol{f}}_1\|}{\|\widehat{\boldsymbol{f}}_1\|}$ et $\frac{\|\delta\boldsymbol{g}\|}{\|\boldsymbol{g}\|}$?

3. Quelle relation existe entre la covariance de $\widehat{\boldsymbol{f}}_1$ et covariance de $\boldsymbol{g}$?

4. Considérons maintenant le cas $M < N$. En évidence, $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f}$ a une infinité de solutions. La solution de norme minimale s'écrit:
$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{H}\boldsymbol{f}=\boldsymbol{g}} \{\|\boldsymbol{f}\|^2\}$$
   Montrez que cette solution s'obtient par la résolution du système d'équations suivante:
$$\begin{bmatrix} \boldsymbol{I} & -\boldsymbol{H}^t \\ \boldsymbol{H} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{f} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{g} \end{bmatrix}$$
   qui donne finalement
$$\widehat{\boldsymbol{f}}_2 = \boldsymbol{H}^t(\boldsymbol{H}\boldsymbol{H}^t)^{-1}\boldsymbol{g}$$
   si $\boldsymbol{H}\boldsymbol{H}^t$ est inversible.

5. Montrez que avec cette solution $\widehat{\boldsymbol{g}} = \boldsymbol{H}\widehat{\boldsymbol{f}}_2 = \boldsymbol{g}$.

6. Quelle relation existe entre la covariance de $\widehat{\boldsymbol{f}}_2$ et covariance de $\boldsymbol{g}$?

7. Revenons au cas général $M \neq N$ et définissons
$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \{J(\boldsymbol{f})\} \quad \text{avec} \quad J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + \lambda\|\boldsymbol{f}\|^2$$
   Montrez que, pour tout $\lambda > 0$, cette solution existe et unique et s'écrit:
$$\widehat{\boldsymbol{f}} = [\boldsymbol{H}'\boldsymbol{H} + lambda\boldsymbol{I}]^{-1}\boldsymbol{H}'\boldsymbol{g}$$

8. Quelle relation existe entre $\widehat{\boldsymbol{g}} = \boldsymbol{H}\widehat{\boldsymbol{f}}$ et $\boldsymbol{g}$?

9. Quelle relation existe entre la covariance de $\widehat{\boldsymbol{f}}$ et covariance de $\boldsymbol{g}$?

10. Une autre solution régularisée $\widehat{\boldsymbol{f}}_2$ à ce problème est de minimiser un critère de la forme
$$J_2(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + \lambda\|\boldsymbol{D}\boldsymbol{f}\|^2,$$
    où $\boldsymbol{D}$ est une matrice approximant un opérateur linéaire de dérivation.
    Montrez que cette solution s'écrit:
$$\widehat{\boldsymbol{f}}_2 = \arg\min_{\boldsymbol{f}} \{J_2(\boldsymbol{f})\} = \left[\boldsymbol{H}'\boldsymbol{H} + \lambda\boldsymbol{D}'\boldsymbol{D}\right]^{-1}\boldsymbol{H}'\boldsymbol{g}$$
    Pourquoi cette solution est-elle préférable à $\widehat{\boldsymbol{f}}_0$ et à $\widehat{\boldsymbol{f}}_1$ et $\widehat{\boldsymbol{f}}_2$?

Problème 2:

Considérons le problème $\boldsymbol{g}=\boldsymbol{H}\boldsymbol{f}$. Nous cherchons une solution $\widehat{\boldsymbol{f}}$ pour le problème inverse telle que $\widehat{\boldsymbol{f}}=\boldsymbol{M}\boldsymbol{g}$, c'est-à-dire qu'elle soit une fonction linéaire des données $\boldsymbol{g}$. Nous cherchons à déterminer la matrice d'inversion $\boldsymbol{M}$ en imposant un certain nombre de contraintes sur elle.

1. Supposons d'abord qu'il existe une solution $\boldsymbol{f}^*$ telle que $\boldsymbol{H}\boldsymbol{f}^*=\boldsymbol{g}$. Alors,

$$\widehat{\boldsymbol{f}}=\boldsymbol{M}\boldsymbol{g}=\boldsymbol{M}\boldsymbol{H}\boldsymbol{f}^*=\boldsymbol{R}\boldsymbol{f}^*$$

   La matrice $\boldsymbol{R}-\boldsymbol{M}\boldsymbol{H}$ mesure le *pouvoir de la résolution dans l'espace des solutions* de l'opérateur inverse $\boldsymbol{M}$. Le cas idéal est $\boldsymbol{R}=\boldsymbol{I}$, ce qui revient à exiger $\boldsymbol{M}=\boldsymbol{H}^{-1}$, ce qui est souvent impossible. Cherchons alors la matrice $\boldsymbol{M}$ telle que

$$J_1(\boldsymbol{M})=\|\boldsymbol{R}-\boldsymbol{I}\|^2=\|\boldsymbol{M}\boldsymbol{H}-\boldsymbol{I}\|^2$$

   soit minimale. Montrez alors que la solution s'écrit

$$\frac{\partial J_1}{\partial \boldsymbol{M}}=[\boldsymbol{M}\boldsymbol{H}-\boldsymbol{I}]\boldsymbol{H}^t=[0]\longrightarrow \boldsymbol{M}=\boldsymbol{H}^t(\boldsymbol{H}\boldsymbol{H}^t)^{-1}$$

2. Une deuxième argumentation est de chercher $\boldsymbol{M}$ telle que $\widehat{\boldsymbol{g}}=\boldsymbol{H}\boldsymbol{f}^*=\boldsymbol{H}\boldsymbol{M}\boldsymbol{g}=\boldsymbol{N}\boldsymbol{g}$ soit la plus proche de $\boldsymbol{g}$. La matrice $\boldsymbol{N}-\boldsymbol{H}\boldsymbol{M}$ mesure le *pouvoir de la résolution dans l'espace des données* de l'opérateur inverse $\boldsymbol{M}$. Le cas idéal est $\boldsymbol{N}=\boldsymbol{I}$, ce qui revient de nouveau à exiger $\boldsymbol{M}=\boldsymbol{H}^{-1}$, ce qui est souvent impossible. Cherchons alors la matrice $\boldsymbol{M}$ telle que

$$J_2(\boldsymbol{M})=\|\boldsymbol{N}-\boldsymbol{I}\|^2=\|\boldsymbol{H}\boldsymbol{M}-\boldsymbol{I}\|^2$$

   soit minimale. Montrez alors que la solution s'écrit Montrez alors que la solution s'écrit

$$\frac{\partial J_2}{\partial \boldsymbol{M}}=[\boldsymbol{H}\boldsymbol{M}-\boldsymbol{I}]\boldsymbol{H}^t=[0]\longrightarrow \boldsymbol{M}=(\boldsymbol{H}^t\boldsymbol{H})^{-1}\boldsymbol{H}^t$$

3. Notons que $\mathrm{cov}[\widehat{\boldsymbol{f}}]=\mathrm{cov}[\boldsymbol{M}\boldsymbol{g}]=\boldsymbol{M}\,\mathrm{cov}[\boldsymbol{g}]\boldsymbol{M}^t$ et si $\mathrm{cov}[\boldsymbol{g}]=\boldsymbol{I}$ on a $\mathrm{cov}[\widehat{\boldsymbol{f}}]=\boldsymbol{M}\boldsymbol{M}^t$. Le cas idéal pour une solution inverse $\widehat{\boldsymbol{f}}$ est avoir une matrice de covariance proche d'identité $\boldsymbol{I}$. On peut donc définir un troisième critère

$$J_3(\boldsymbol{M})=\|\boldsymbol{U}\|^2=\|\boldsymbol{M}\boldsymbol{M}^t\|^2$$

   qui peut servir pour contraindre $\boldsymbol{M}$. Écrivez l'expression de $\frac{\partial J_3}{\partial \boldsymbol{M}}$.

4. Définissons $J(\boldsymbol{M})=\alpha_1 J_1(\boldsymbol{M})+\alpha_2 J_2(\boldsymbol{M})+\alpha_3 J_2(\boldsymbol{M})$. Écrivez l'expression de $\frac{\partial J}{\partial \boldsymbol{M}}$ et l'expression de $\boldsymbol{M}$ qui minimise $J(\boldsymbol{M})$ pour différentes combinaisons de $(\alpha_1,\alpha_2,\alpha_3)$. Dans chaque cas Écrivez l'expression de $\boldsymbol{R}$, $\boldsymbol{N}$ et $\boldsymbol{U}$.

   Vérifier ce tableau :

$\alpha_1\alpha_2\alpha_3$	$\boldsymbol{M}$	$\boldsymbol{N}=\boldsymbol{H}\boldsymbol{M}$	$\boldsymbol{R}=\boldsymbol{M}\boldsymbol{H}$	$\boldsymbol{U}=\boldsymbol{M}\boldsymbol{M}^t$
1 0 0	$\boldsymbol{M}=\boldsymbol{H}^t(\boldsymbol{H}\boldsymbol{H}^t)^{-1}$	$\boldsymbol{I}$	$\boldsymbol{H}^t(\boldsymbol{H}\boldsymbol{H}^t)^{-1}\boldsymbol{H}$	$\boldsymbol{H}^t(\boldsymbol{H}\boldsymbol{H}^t)^{-2}\boldsymbol{H}^t$
0 1 0	$\boldsymbol{M}=(\boldsymbol{H}^t\boldsymbol{H})^{-1}\boldsymbol{H}^t$	$\boldsymbol{H}(\boldsymbol{H}^t\boldsymbol{H})^{-1}\boldsymbol{H}^t$	$\boldsymbol{I}$	$(\boldsymbol{H}^t\boldsymbol{H})^{-1}$
1 0 1	$\boldsymbol{M}=\boldsymbol{H}^t(\boldsymbol{H}\boldsymbol{H}^t+\lambda\boldsymbol{I})^{-1}$	$\boldsymbol{H}\boldsymbol{H}^t(\boldsymbol{H}\boldsymbol{H}^t+\lambda\boldsymbol{I})^{-1}$	$\boldsymbol{H}^t(\boldsymbol{H}\boldsymbol{H}^t+\lambda\boldsymbol{I})^{-1}\boldsymbol{H}$	$\boldsymbol{H}^t(\boldsymbol{H}\boldsymbol{H}^t+\lambda\boldsymbol{I})^{-2}\boldsymbol{H}$
0 1 1	$\boldsymbol{M}=(\boldsymbol{H}^t\boldsymbol{H}+\lambda\boldsymbol{I})^{-1}\boldsymbol{H}^t$	$\boldsymbol{H}(\boldsymbol{H}^t\boldsymbol{H}+\lambda\boldsymbol{I})^{-1}\boldsymbol{H}^t$	$(\boldsymbol{H}^t\boldsymbol{H}+\lambda\boldsymbol{I})^{-1}\boldsymbol{H}^t\boldsymbol{H}$	$(\boldsymbol{H}^t\boldsymbol{H}+\lambda\boldsymbol{I})^{-1}\boldsymbol{H}^t\boldsymbol{H}(\boldsymbol{H}^t\boldsymbol{H}+\lambda\boldsymbol{I})^{-t}$
1 1 0	$\boldsymbol{M}=\boldsymbol{H}^t(\boldsymbol{H}\boldsymbol{H}^t)^{-1}$	$\boldsymbol{I}$	$\boldsymbol{H}^t(\boldsymbol{H}\boldsymbol{H}^t)^{-1}\boldsymbol{H}$	$\boldsymbol{H}^t(\boldsymbol{H}\boldsymbol{H}^t)^{-2}\boldsymbol{H}$
1 1 1	$\boldsymbol{M}=\boldsymbol{H}^t(\boldsymbol{H}\boldsymbol{H}^t+\lambda\boldsymbol{I})^{-1}$	$\boldsymbol{H}\boldsymbol{H}^t(\boldsymbol{H}\boldsymbol{H}^t+\lambda\boldsymbol{I})^{-1}$	$\boldsymbol{H}^t(\boldsymbol{H}\boldsymbol{H}^t+\lambda\boldsymbol{I})^{-1}\boldsymbol{H}$	$\boldsymbol{H}^t(\boldsymbol{H}\boldsymbol{H}^t+\lambda\boldsymbol{I})^{-2}\boldsymbol{H}$

Problème 3:

Considérons le problème $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f}$ et supposons que $\boldsymbol{f}$ représente une image avec $f_n \geq 0$. Supposons que le système $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f}$ est sous-déterminée et que nous cherchons à choisir parmi toutes ses solutions, celle qui maximise l'entropie

$$S = -\sum_j f_j \ln f_j$$

1. Montrez que cette solution existe et s'écrit:

$$f_j = \exp\left[-\lambda_0 - \sum_i H_{ij}\lambda_i\right]$$

   où $\boldsymbol{\lambda} = \{\lambda_i, i = 1, \cdots, M\}$ s'obtient en résolvant le système d'équations suivante:

$$\sum_j H_{ij} \exp\left[-\lambda_0 - \sum_i H_{ij}\lambda_i\right] = g_i$$

   et où λ_0 est une constante qui ne peut être fixée que si on impose par exemple $\sum_j f_j = 1$.

2. Montrez que cette solution peut aussi s'écrire:

$$f_j = \frac{1}{Z(\boldsymbol{\lambda})} \exp\left[-\sum_i H_{ij}\lambda_i\right]$$

   où

$$-\ln Z(\boldsymbol{\lambda}) = \ln f_j + [\boldsymbol{H}^t\boldsymbol{\lambda}]_j$$

   et où $\boldsymbol{\lambda}$ s'obtient en résolvant le système d'équations suivante:

$$\frac{\partial \ln Z(\boldsymbol{\lambda})}{\partial \lambda_i} = g_i$$

3. Montrez aussi que $\boldsymbol{\lambda}$ s'obtient aussi par optimisation du critère dual $D(\boldsymbol{\lambda}) = \|\boldsymbol{g} - \boldsymbol{H}\exp[-\boldsymbol{H}\boldsymbol{f}]\|^2$.

4. Écrivez un algorithme itératif du type gradient ou de Newton qui calcule cette solution en précisant le coût de calcul dans chaque étape. Pour cela, vous pouvez utiliser deux routine `g=direct(H,f)` qui calcule $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f}$ et `f=transp(H,g)` qui calcule $\boldsymbol{f} = \boldsymbol{H}^t\boldsymbol{g}$.

Problème 4:

Considérons le problème $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f}$ et supposons que $\boldsymbol{f}$ représente une image moyenne, *i.e.*chaque f_n correspond à l'espérance d'une grandeur Z_n. Les mesures g_m (supposés sans bruit) peuvent alors être considérées comme une combinaison linéaire de $\mathrm{E}\{Z_n\}$, *i.e.*;

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} \longrightarrow g_m = \sum_{n=1}^{N} a_{m,n} f_n = \sum_{n=1}^{N} a_{m,n} \mathrm{E}\{Z_n\}, \quad m = 1, \cdots, M.$$

Trouvez la forme de la densité de probabilité $p(\boldsymbol{z})$ qui satisfait ces contraintes et qui maximise

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

où $\mu(\boldsymbol{z})$ est une densité de référence connue. Montrez alors 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, *i.e.*;

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

Notant que $f_n = \mathrm{E}\{Z_n\}$ montrez que

1. si $\mu(z_n) \propto \exp\left[-\frac{1}{2}z_n^2\right]$ on a

$$p(z_n) \propto \exp\left[-\frac{1}{2}z_n^2 + \sum_{m=1}^{M} a_{m,n}\lambda_m z_n\right],$$

et par conséquent

$$f_n=\sum_{m=1}^{M} a_{m,n}\lambda_m, \text{ ou encore } \boldsymbol{f}=\boldsymbol{H}'\boldsymbol{\lambda}$$

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

$$g_m=\sum_{n=1}^{N} a_{m,n}\sum_{i=0}^{M} a_{m,n}\lambda_m, \text{ ou encore } \boldsymbol{g}=\boldsymbol{H}\boldsymbol{H}'\boldsymbol{\lambda},$$

ce qui permet d'écrire (supposons que $\boldsymbol{H}\boldsymbol{H}'$ est inversible)

$$\boldsymbol{f}=\boldsymbol{H}'[\boldsymbol{H}\boldsymbol{H}']^{-1}\boldsymbol{g}$$

Interpréter alors ce résultat.

2. si $\mu(z_n)\propto\exp\left[-z_n\right], \quad z_n>0$ on a

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

et par conséquent

$$f_n=1+\sum_{m=1}^{M} a_{m,n}\lambda_m, \text{ ou encore } \boldsymbol{f}=\boldsymbol{1}+\boldsymbol{H}'\boldsymbol{\lambda}$$

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

$$g_m=\sum_{n=1}^{N} a_{m,n}(1+\sum_{i=0}^{M} a_{m,n}\lambda_m), \text{ ou encore } \boldsymbol{g}=\boldsymbol{H}\boldsymbol{1}+\boldsymbol{H}\boldsymbol{H}'\boldsymbol{\lambda},$$

ce qui permet d'écrire (supposons que $\boldsymbol{H}\boldsymbol{H}'$ est inversible)

$$\boldsymbol{f}=\boldsymbol{1}+\boldsymbol{H}'[\boldsymbol{H}\boldsymbol{H}']^{-1}(\boldsymbol{g}-\boldsymbol{H}\boldsymbol{1})=\boldsymbol{H}'[\boldsymbol{H}\boldsymbol{H}']^{-1}\boldsymbol{g}+\left(\boldsymbol{I}-\boldsymbol{H}'[\boldsymbol{H}\boldsymbol{H}']^{-1}\boldsymbol{H}\right)\boldsymbol{1}$$

Interpréter alors ce résultat.

3. si $\mu(z_n)\propto z_n^{(\alpha-1)}=\exp\left[(\alpha-1)\ln z_n\right], \quad z_n>0$ on a

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

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

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

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

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

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

$$\begin{aligned}\boldsymbol{f}&=\frac{\alpha}{\boldsymbol{H}'\boldsymbol{\lambda}} \qquad \text{ou encore} \qquad \boldsymbol{f}=\alpha\boldsymbol{1}./\boldsymbol{H}'\boldsymbol{\lambda}\\ \boldsymbol{g}&=\boldsymbol{H}\frac{\alpha}{\boldsymbol{H}'\boldsymbol{\lambda}} \qquad \text{ou encore} \qquad \boldsymbol{g}=\boldsymbol{H}(\alpha\boldsymbol{1}.\ \boldsymbol{H}'\boldsymbol{\lambda})=(\alpha\boldsymbol{H}\boldsymbol{1})./(\boldsymbol{H}'\boldsymbol{\lambda})\end{aligned}$$

Interpréter alors ce résultat.