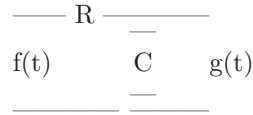


Exercice numéro 1: Convolution et Déconvolution

Problème 1 : Modélisation

Considérons le système suivant:



1. crivez l'expression de la fonction de transfert $H(\omega) = \frac{G(\omega)}{F(\omega)}$
2. crivez l'expression de la réponse impulsionnelle $h(t)$
3. crivez la relation liant la sortie $g(t)$ à l'entrée $f(t)$ et la réponse impulsionnelle $h(t)$
4. crivez la relation liant $G(\omega)$ à l'entrée $F(\omega)$ et $H(\omega)$
5. Calculez la sortie lorsque l'entrée est une impulsion $f(t) = \delta(t)$
6. Calculez la sortie lorsque l'entrée est un échelon $f(t) = u(t)$
7. Calculez la sortie lorsque l'entrée est une sinusoïde $f(t) = a \sin(\omega_0 t)$
8. Calculez la sortie lorsque l'entrée est $f(t) = \sum_k f_k \sin(\omega_k t)$
9. Calculez la sortie lorsque l'entrée est $f(t) = \sum_j f_j \delta(t - t_j)$
10. Supposons $h(t) = \sum_{k=-q}^p h_k \delta(t - t_k)$, alors calculez la sortie lorsque l'entrée est $f(t) = \sum_{j=0}^{n-1} f_j \delta(t - t_j)$.
On suppose $t_k = k T$ et $t_j = j T$ avec $T = 1$.
11. Montrez que le lien entre les coefficients $\mathbf{f} = [f_0, \dots, f_{n-1}]'$, $\mathbf{h} = [h_{-q}, \dots, h_0, \dots, h_p]'$ et $\mathbf{g} = [g_0, \dots, g_{m-1}]'$ peut s'écrire soit sous la forme $\mathbf{g} = \mathbf{H}\mathbf{f}$ ou sous la forme $\mathbf{g} = \mathbf{F}\mathbf{h}$. crivez alors les expressions des matrices \mathbf{H} et \mathbf{F} .
12. Que remarque-t-on sur la structure de ces deux matrices ?
13. Que deviennent ces matrices pour $q = 0$?
14. Que peut-on faire pour que ces matrices deviennent circulantes ?
15. crivez une routine Matlab qui calcule \mathbf{g} en fonction de \mathbf{f} et \mathbf{h} . On peut appeler cette routine `g=direct(h,f)`. Créer ensuite les deux vecteurs \mathbf{h} et \mathbf{f} et tester votre routine. Vous pouvez écrire deux routines qui créent différentes réponse impulsionnelle $h(t)$ et différentes entrées $f(t)$, ce qui vous permet de tester votre routines pour différent cas.

Problème 2 : Identification et Inversion

1. tant donnée l'entrée $f(t)$ et la sortie $g(t)$, peut-on estimer $h(t)$? numérez les différentes méthodes.
2. crivez une routine Matlab qui calcule \mathbf{h} en fonction de \mathbf{f} et \mathbf{g} . On peut appeler cette routine `h=identification(g,f,method)`. Créer ensuite les deux vecteurs \mathbf{g} et \mathbf{f} et tester votre routine. Pensez aussi à rajouter un peu de bruit sur vos données \mathbf{g} .
3. tant donnée la sortie $g(t)$ et la réponse impulsionnelle $h(t)$, peut-on estimer l'entrée $f(t)$? numérez les différentes méthodes.
4. crivez une routine Matlab qui calcule \mathbf{f} en fonction de \mathbf{g} et \mathbf{h} . On peut appeler cette routine `f=inversion(g,h,method)`. Créer ensuite les deux vecteurs \mathbf{h} et \mathbf{g} et tester votre routine. Pensez aussi à rajouter un peu de bruit sur vos données \mathbf{g} .
5. Rapportez vos expériences avec vos commentaires.

Exercice numéro 2: Convolution 2D et Restauration d'image

Problème 1 : Modélisation

Considérons un système d'imagerie telle qu'un appareil photo mal focalisé. Supposons que lors d'une expérience, nous avons pu mesurer sa tache focale (réponse impulsionnelle) $h(x, y)$ qui s'étend sur quelque dizaines de pixels. Supposons aussi que le lien entre l'image parfaite $f(x, y)$ et l'image observée $g(x, y)$ puisse se modéliser par une convolution $g = h * f$.

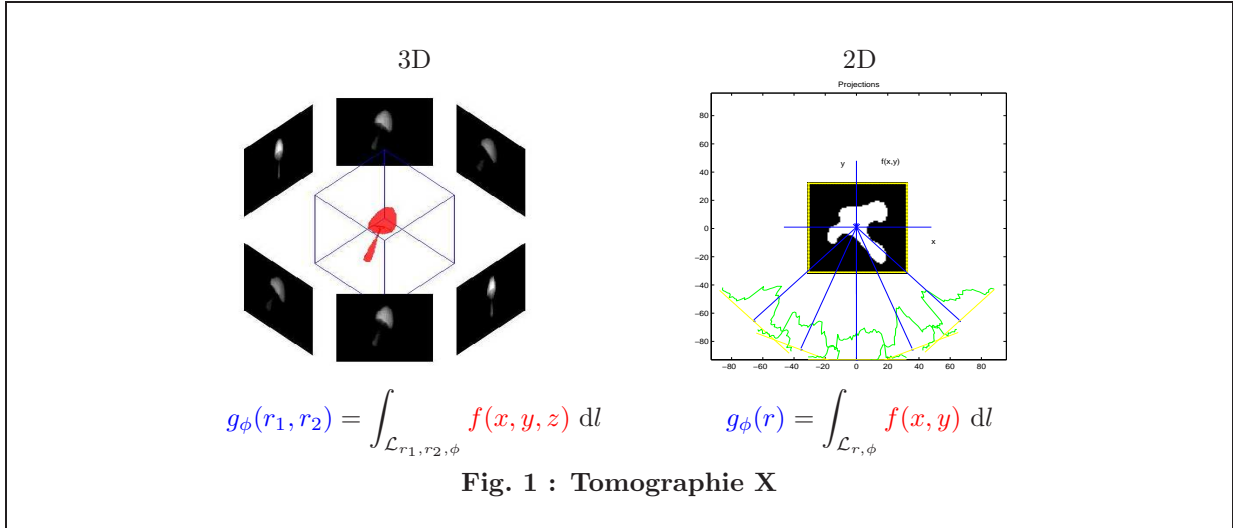
1. écrivez l'expression intégrale liant la sortie $g(x, y)$ à l'entrée $f(x, y)$ et la réponse impulsionnelle $h(x, y)$
2. écrivez l'expression de la fonction de transfert $H(\omega_x, \omega_y) = \frac{G(\omega_x, \omega_y)}{F(\omega_x, \omega_y)}$
3. Montrez que le lien entre les pixels de l'image parfaite organisés comme un vecteur \mathbf{f} en concaténant ses colonnes, les pixels de l'image observée organisés comme un vecteur \mathbf{g} en concaténant ses colonnes et le vecteur \mathbf{h} obtenu en concaténant les colonnes de tache focale, peut s'écrire soit sous la forme $\mathbf{g} = \mathbf{H}\mathbf{f}$ ou sous la forme $\mathbf{g} = \mathbf{F}\mathbf{h}$. écrivez alors les expressions des matrices \mathbf{H} et \mathbf{F} .
4. Que remarque-t-on sur la structure de ces deux matrices ?
5. Que deviennent ces matrices lorsque la tâche focale est symétrique ?
6. Que peut-on faire pour que ces matrices deviennent circulantes-bloc-circulantes ?
7. écrivez une routine Matlab qui calcule \mathbf{g} en fonction de \mathbf{f} et de \mathbf{h} , ou plutôt une routine qui fournit l'image g en fonction des deux images h et f . On peut appeler cette routine `g=direct(h,f)`. Créez ensuite deux images h et f et testez votre routine. Vous pouvez écrire deux routines qui créent différentes réponses impulsionnelles $h(x, y)$ comme un carré ou un disque ou encore une gaussienne et différentes images $f(x, y)$ comme juste un point brillant ou un carré ou un disque ou un échiquier, ou encore une photo, ce qui vous permet de tester vos routines pour différents cas.

Problème 2 : Identification et inversion

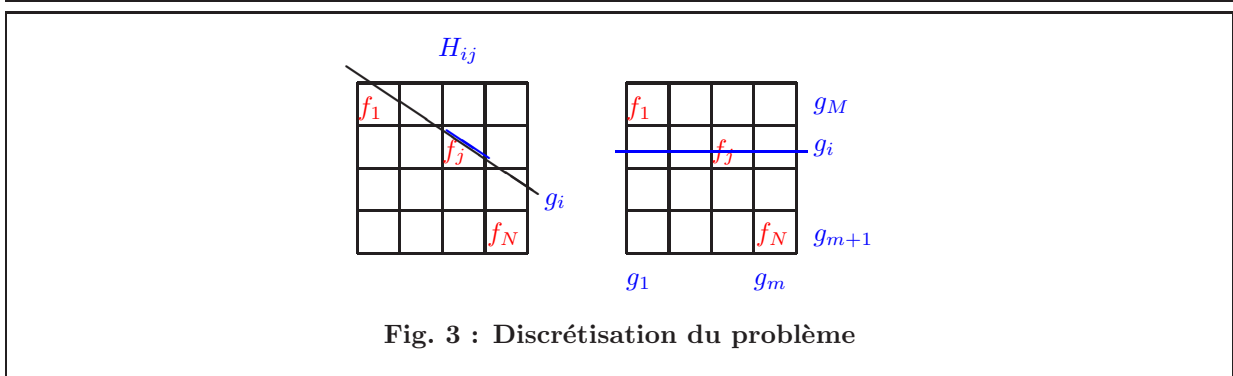
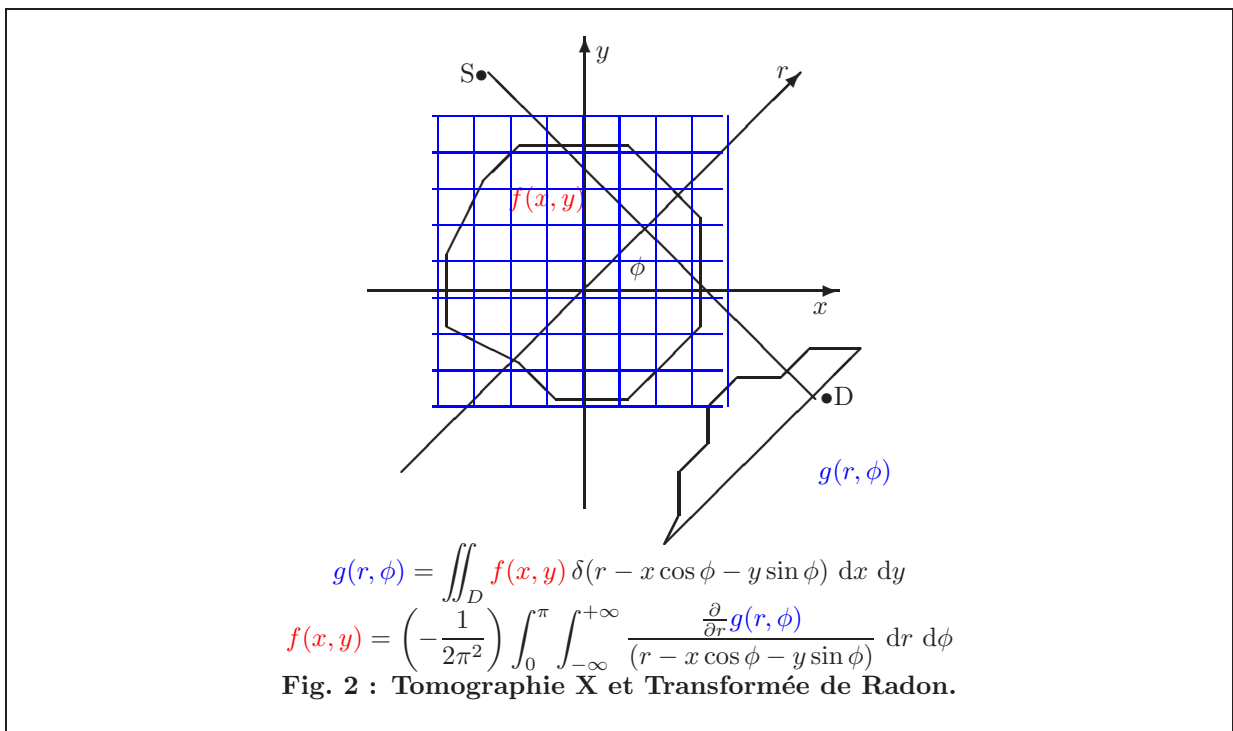
1. tant donnée l'entrée $f(x, y)$ et la sortie $g(x, y)$, peut-on estimer $h(x, y)$? numérotez les différentes méthodes.
2. écrivez une routine Matlab qui calcule h en fonction de f et g . On peut appeler cette routine `h=identification(g,f,method)`. Créez ensuite les deux images g et f et testez votre routine. Pensez aussi à rajouter un peu de bruit sur vos données g .
3. tant donnée la sortie g et la réponse impulsionnelle h , peut-on estimer l'entrée f ? numérotez les différentes méthodes.
4. écrivez une routine Matlab qui calcule f en fonction de g et h . On peut appeler cette routine `f=inversion(g,h,method)`. Créez ensuite les deux images h et g et testez votre routine. Pensez aussi à rajouter un peu de bruit sur vos données g .
5. Rapportez vos expériences avec vos commentaires.

Exercice numéro 3: Tomographie et Reconstruction d'image

En tomographie à rayons X, le modèle qui lie l'intensité des rayons reçus sur un capteur g est liée à la distribution des coefficients d'absorption de l'intérieur d'un objet f par une équation intégral. La figure suivante montre cette relation.



Par la suite, nous allons considérer le cas 2D. Dans ce cas, comme nous l'avons vu dans le cours, il y a des relations suivantes entre $g(r, \phi)$ et $f(x, y)$:



Problème 1 : Méthodes de reconstruction analytiques

Considérons le problème de la reconstruction d'image en tomographie X où la relation entre l'image $f(x, y)$ et les projections $g(r, \phi)$ est modélisée par la transformée de Radon (TR):

$$g(r, \phi) = \int_{L(r, \phi)} f(x, y) dl = \iint f(x, y) \delta(r - x \cos \phi - y \sin \phi) dx dy,$$

où $L(r, \phi)$ est une droite faisant un angle ϕ avec l'axe des y et située à une distance r de l'origine.

Partant de l'équation de l'inversion de la TR, *i.e.*;

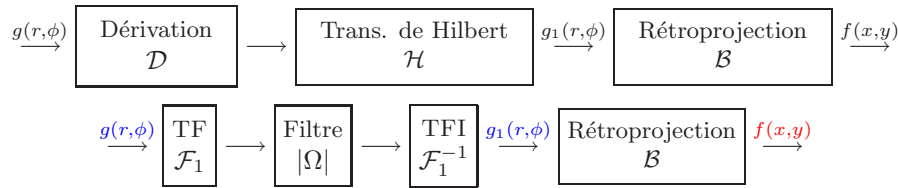
$$f(x, y) = \left(-\frac{1}{2\pi^2}\right) \int_0^\pi \int_0^\infty \frac{\frac{\partial g(r, \phi)}{\partial r}}{(r - x \cos \phi - y \sin \phi)} dr d\phi$$

et en utilisant les définitions suivantes :

$$\begin{aligned} \text{Dérivation } \mathcal{D}: & \quad \bar{g}(r, \phi) = \frac{\partial g(r, \phi)}{\partial r} \\ \text{Transformée de Hilbert } \mathcal{H}: & \quad g_1(r', \phi) = \frac{1}{\pi} \int_0^\infty \frac{\bar{g}(r, \phi)}{(r - r')} dr \\ \text{Rétroprojection } \mathcal{B}: & \quad f(x, y) = \frac{1}{2\pi} \int_0^\pi g_1(x \cos \phi + y \sin \phi, \phi) d\phi \\ \text{Transformée de Fourier 1D } \mathcal{F}_1: & \quad G(\Omega, \phi) = \int g(r, \phi) \exp[-j\Omega r] dr \\ \text{Transformée de Fourier 2D } \mathcal{F}_2: & \quad F(u, v) = \iint f(x, y) \exp[-j(ux + vy)] dx dy \end{aligned}$$

1. Montrez que l'on peut reconstruire $f(x, y)$ par les schémas suivants :

$$f(x, y) = \mathcal{B} \mathcal{H} \mathcal{D} g(r, \phi) = \mathcal{B} \mathcal{F}_1^{-1} |\Omega| \mathcal{F}_1 g(r, \phi)$$



2. Montrez que si on définit

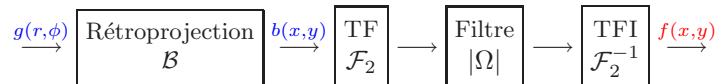
$$b(x, y) = \int_0^\pi g(r, \phi) d\phi = \int_0^\pi g(x \cos \phi + y \sin \phi, \phi) d\phi$$

on a une relation entre $f(x, y)$ et $b(x, y)$ de la forme

$$b(x, y) = f(x, y) * \frac{1}{[x^2 + y^2]^{1/2}}$$

Montrez alors que

$$f(x, y) = \mathcal{F}_2^{-1} |\Omega| \mathcal{F}_2 \mathcal{B} g(r, \phi)$$



où $|\Omega|^2 = u^2 + v^2$.

3. Si on note $G(\Omega, \phi)$ la TF1D par rapport à la variable r de $g(r, \phi)$ pour un angle fixé ϕ et $F(u, v)$ la TF2D de $f(x, y)$, montrez que

$$F(\Omega \cos \phi, \Omega \sin \phi) = G(\Omega, \phi).$$

4. Montrez que si l'image a une propriété de symétrie de révolution, *i.e.*; $f(x, y) = f(\rho)$ avec $\rho^2 = x^2 + y^2$, alors on peut la reconstruire seulement à partir d'une seule projection.
5. Montrez que si $f(x, y) = f_1(x) f_2(y)$ alors on peut la reconstruire (à une constante près) seulement à partir de deux projections $\phi = 0$ et $\phi = 90$

Problème 2 : Méthodes de reconstruction algébriques

Dans cette partie, afin de simplifier les relations, nous allons considérer le problème de la reconstruction d'image à partir de seulement deux projections ($\phi = 0$ et $\phi = \pi/2$ (Voir Fig. 3)).

1. En supposant que la valeur de f à l'intérieur de chaque pixel est constante, formez des vecteur \mathbf{f} et \mathbf{g} et montrez que $\mathbf{g} = \mathbf{H}\mathbf{f}$. Que représentent alors les éléments de la matrice \mathbf{H} ?
2. Considérons le cas où on a deux projections: horizontale et verticale. Que deviennent les éléments de la matrice \mathbf{H} ?
3. Considérons maintenant une image avec seulement (4×4) pixels: $\mathbf{f} = [f_1, \dots, f_{16}]'$ et deux projections horizontale et verticale:

f_1	f_5	f_9	f_{13}	g_8
f_2	f_6	f_{10}	f_{14}	g_7
f_3	f_7	f_{11}	f_{15}	g_6
f_4	f_8	f_{12}	f_{16}	g_5
g_1	g_2	g_3	g_4	

f_{11}	f_{12}	f_{13}	f_{14}	g_{24}
f_{21}	f_{22}	f_{23}	f_{24}	g_{23}
f_{31}	f_{32}	f_{33}	f_{34}	g_{22}
f_{41}	f_{42}	f_{43}	f_{44}	g_{21}
g_{11}	g_{12}	g_{13}	g_{14}	

$$\mathbf{g}_1 = [g_1, \dots, g_4]' = [g_{11}, \dots, g_{14}]'$$

$$\mathbf{g}_2 = [g_5, \dots, g_8]' = [g_{21}, \dots, g_{24}]'$$

Formez alors les matrices \mathbf{H}_1 , \mathbf{H}_2 et \mathbf{H} telle qu'on puisse écrire

$$\mathbf{g}_1 = \mathbf{H}_1\mathbf{f}, \quad \mathbf{g}_2 = \mathbf{H}_2\mathbf{f}, \quad \mathbf{g} = \mathbf{H}\mathbf{f} = \left[\begin{array}{c} \mathbf{H}_1 \\ \mathbf{H}_2 \end{array} \right] \mathbf{f}$$

4. Considérons l'image

$$\mathbf{f} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Calculez alors sa projection.

5. Considérons maintenant le problème inverse: tant donnée les deux projections trouvez l'image. Pourquoi ce problème est mal-posé ?
6. Montrez que ce problème a une infinité de solutions possibles. Montrez quelque exemples.
7. Trouvez l'expression qui permet de trouver toutes ces solutions.
8. En comparant les relation continue et discrète et les opérateurs direct et adjoint, montrez que l'opération de rétroprojection en continue correspond à l'opérateur de transposition de matrice en discret, autrement dit, la solution $\hat{\mathbf{f}} = \mathbf{H}'\mathbf{g}$ correspond à la solution au sens de la la rétro-projection. Exprimer alors la matrice $\mathbf{H}' = [\mathbf{H}'_1 \mid \mathbf{H}'_2]$
9. Montrez que $\hat{\mathbf{f}} = \mathbf{H}'\mathbf{g} = \mathbf{H}'_1\mathbf{g}_1 + \mathbf{H}'_2\mathbf{g}_2$ est l'addition de deux images. Calculez ces deux images. Que remarque-t-on ?
10. En vue de la définition d'une solution au sens d'inversion généralisée, Calculez les matrices $\mathbf{H}'\mathbf{H}$ et $\mathbf{H}\mathbf{H}'$ et montrez que

$$\mathbf{H}\mathbf{H}' = \left[\begin{array}{c|c} \mathbf{H}_1\mathbf{H}'_1 & \mathbf{H}_1\mathbf{H}'_2 \\ \hline \mathbf{H}_2\mathbf{H}'_1 & \mathbf{H}_2\mathbf{H}'_2 \end{array} \right] = \left[\begin{array}{c|c} 4\mathbf{I} & \mathbf{1} \\ \hline \mathbf{1} & 4\mathbf{I} \end{array} \right]$$

et

$$\mathbf{H}'\mathbf{H} = \left[\begin{array}{c|c} \mathbf{H}'_1\mathbf{H}_1 & \mathbf{H}'_2\mathbf{H}_1 \\ \hline \mathbf{H}'_1\mathbf{H}_2 & \mathbf{H}'_2\mathbf{H}_2 \end{array} \right]$$

$$\mathbf{H}'_1\mathbf{H}_1 = \mathbf{H}'_2\mathbf{H}_2 = \left[\begin{array}{c|c} \mathbf{1} + \mathbf{I} & \mathbf{I} \\ \hline \mathbf{I} & \mathbf{1} + \mathbf{I} \end{array} \right]$$

$$\mathbf{H}'_1\mathbf{H}_2 = \mathbf{H}'_2\mathbf{H}_1 = \left[\begin{array}{c|c} \mathbf{I} & \mathbf{I} \\ \hline \mathbf{I} & \mathbf{I} \end{array} \right]$$

11. Calculez les valeurs singulières des matrices $\mathbf{H}\mathbf{H}'$ et $\mathbf{H}'\mathbf{H}$ et montrez que

$$\text{svd}(\mathbf{H}\mathbf{H}') = [8\ 4\ 4\ 4\ 4\ 4\ 4\ 0]$$

$$\text{svd}(\mathbf{H}'\mathbf{H}) = [8\ 4\ 4\ 4\ 4\ 4\ 4\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0]$$

12. Est-ce que ces matrices sont inversibles ?

13. Comment alors peut-on définir une solution à ce problème?

14. Rappelons qu'une solution au sens des moindres carrés s'écrit:

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f}} \{\|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2\},$$

Montrez que si la matrice $\mathbf{H}^t\mathbf{H}$ est inversible on obtient : $\hat{\mathbf{f}} = (\mathbf{H}^t\mathbf{H})^{-1}\mathbf{H}^t\mathbf{g}$.

15. De même, une solution de norme minimale est

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{H}\mathbf{f}=\mathbf{g}} \{\|\mathbf{f}\|^2\}$$

Montrez que si la matrice $\mathbf{H}\mathbf{H}^t$ est inversible on obtient : $\hat{\mathbf{f}} = \mathbf{H}^t(\mathbf{H}\mathbf{H}^t)^{-1}\mathbf{g}$.

16. Notez que nous ne pouvons pas Calculer aucune de ces deux solutions car aucune des deux matrices $\mathbf{H}\mathbf{H}^t$ et $\mathbf{H}^t\mathbf{H}$ est inversible. Notons cependant que si on ne garde que les éléments diagonaux de ces deux matrices, on obtient des résultats suivants :

`fh=diag(1./diag(AtA))*A'*p;reshape(fh,4,4)`

$$\hat{\mathbf{f}} = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 2 & 2 & 1 \\ 1 & 2 & 2 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

`fh=A'*diag(1./diag(AAt))*p;reshape(fh,4,4)`

$$\hat{\mathbf{f}} = \begin{bmatrix} 0 & .5 & .5 & 0 \\ .5 & 1 & 1 & .5 \\ .5 & 1 & 1 & .5 \\ 0 & .5 & .5 & 0 \end{bmatrix}$$

Commenter ces deux résultats.

17. Nous avons vu dans le cours qu'il est cependant possible de calculer la solution inverse généralisée qui est la solution de norme minimale de $\mathbf{H}^t\mathbf{H}\mathbf{f} = \mathbf{H}^t\mathbf{g}$ en utilisant la décomposition tronquée des valeurs singulières (DTVS):

$$\hat{\mathbf{f}} = \sum_{k=1}^K \frac{\langle \mathbf{g}, \mathbf{u}_k \rangle}{\lambda_k} \mathbf{v}_k$$

où \mathbf{u}_k et \mathbf{v}_k sont, respectivement, des vecteurs propres de $\mathbf{H}\mathbf{H}^t$ et de $\mathbf{H}^t\mathbf{H}$ et λ_k sont des valeurs singulières associées.

`[U,S,V]=svd(A);`
`s=diag(S);s1=[1./s(1:7);zeros(1,1)];`
`S1=[diag(s1);zeros(8,8)];`
`fh=V*S1*U'*p;reshape(fh,4,4)`

Dans cet exemple $K = 7$ et la solution IG peut être calculée par :

`fh=svdpca(A,p,.1,7);reshape(fh,4,4)`

$$\hat{\mathbf{f}} = \begin{bmatrix} -0.2500 & 0.2500 & 0.2500 & -0.2500 \\ 0.2500 & 0.7500 & 0.7500 & 0.2500 \\ 0.2500 & 0.7500 & 0.7500 & 0.2500 \\ -0.2500 & 0.2500 & 0.2500 & -0.2500 \end{bmatrix}$$

Commenter ce résultat.

18. Montrez que le noyau de la transformation linéaire $\mathbf{g} = \mathbf{H}\mathbf{f}$, i.e. $\{\mathbf{f} | \mathbf{H}\mathbf{f} = 0\}$ est

$$\mathbf{V}(\mathbf{I} - \mathbf{S}^+\mathbf{S})\mathbf{z} = \sum_{k=K+1}^N z_k \mathbf{v}_k$$

avec \mathbf{z} un vecteur arbitraire. Ceci nous permet de trouver toutes les solutions possibles du problème en rajoutant ce terme arbitraire à la solutions IG.

19. Montrez que l'algorithme itératif suivant:

```
for k=1:100;
fh=fh+.1*A'*(p-A*fh(:));
end;
reshape(fh,4,4)  fourni ce résultat:
```

$$\hat{\mathbf{f}} = \begin{bmatrix} -0.2500 & 0.2500 & 0.2500 & -0.2500 \\ 0.2500 & 0.7500 & 0.7500 & 0.2500 \\ 0.2500 & 0.7500 & 0.7500 & 0.2500 \\ -0.2500 & 0.2500 & 0.2500 & -0.2500 \end{bmatrix}$$

Commenter ce résultat.

20. On peut remarquer que la manque d'information dans les données est telle que la contrainte de norme minimal ne restreint pas suffisamment l'espace des solutions possibles. Dans les problèmes inverses en imagerie, une information qui est souvent disponible est la positivité de la solution. Imposer alors à la solution d'être positive est alors une technique souvent utilisée. Une approche simple dans les méthodes itérative pour imposer cette contraintes est simplement l'imposer à chaque itération :

```
for k=1:100
fh=fh+.1*A'*(p-A*fh(:));
fh=fh.*(fh>0);
end
reshape(fh,4,4);
```

$$fh = \begin{bmatrix} 0 & 0.0000 & 0.0000 & 0 \\ 0.0000 & 1.0000 & 1.0000 & 0.0000 \\ 0.0000 & 1.0000 & 1.0000 & 0.0000 \\ 0 & 0.0000 & 0.0000 & 0 \end{bmatrix}$$

Commenter ce résultat.

21. Proposer d'autres méthodes.

Exercice numéro 4: Déconvolution par moindres carrés et régularisation: Cas continu

Problème 1: Déconvolution

Considérons le problème de la déconvolution où le signal

mesuré $g(t)$ est relié au signal d'entrée $f(t)$ et à la réponse impulsionnelle $h(t)$ par $g(t) = h(t)*f(t) + \epsilon(t)$ et où on cherche à estimer $h(t)$ et $f(t)$ à partir de mesure de $g(t)$.

1. Supposons d'abord $h(t)$ connue (déconvolution simple). Définissons la solution $\hat{f}(t)$ par

$$\hat{f}(t) = \arg \min_f \{ \|g - h * f\|^2 + \lambda_1 \|d_f * f\|^2 \},$$

où $d_f(t)$ et λ_1 sont connues et fixées et où la norme $\|z\|^2$ signifie

$$\|z\|^2 = \int |z(t)|^2 dt.$$

Montrez que cette solution peut être calculée par

$$F(\nu) = \frac{H^*(\nu)}{|H(\nu)|^2 + \lambda_1 |D_f(\nu)|^2} G(\nu),$$

où $F(\nu)$ et $G(\nu)$ sont les fonctions densité spectrale de puissance (dsp) de $f(t)$ et de $g(t)$ et $H(\nu)$ et $D_f(\nu)$ sont les TF de $h(t)$ et de $d_f(t)$.

2. Que devient cette solution lorsque $\lambda_1 = 0$?
 3. Quel est le rôle de d_f ou $D_f(\nu)$? Comment le choisir ?
 4. Cherchons maintenant à estimer $h(t)$ à partir de la mesure de $g(t)$ et de $f(t)$ (identification). Définissons la solution $\hat{h}(t)$ par

$$\hat{h}(t) = \arg \min_h \{ \|g - h * f\|^2 + \lambda_2 \|d_h * h\|^2 \}.$$

Montrez que cette solution peut être calculée par

$$H(\nu) = \frac{F^*(\nu)}{|F(\nu)|^2 + \lambda_2 |D_h(\nu)|^2} G(\nu).$$

5. Que devient cette solution lorsque $\lambda_2 = 0$?
 6. Quel est le rôle de d_h ou $D_h(\nu)$? Comment le choisir ?
 7. Supposons maintenant qu'on souhaite estimer $h(t)$ et $f(t)$ à partir de la seule mesure de $g(t)$ (déconvolution aveugle). Peut-on envisager de définir la solution par

$$(\hat{f}(t), \hat{h}(t)) = \arg \min_{(f,h)} \{ \|g - h * f\|^2 + \lambda_1 \|d_f * f\|^2 + \lambda_2 \|d_h * h\|^2 \} \quad ?$$

Pourquoi? Ce critère a-t-il une seule solution ? Commenter votre réponse.

Problème 2: Déconvolution – Cas discret

Considérons maintenant le même problème dans le cas discret et faisons l'hypothèse que le système est causal et que les signaux sont causaux et à durée limitées et notons $\mathbf{h} = [h_0, \dots, h_p]^t$, $\mathbf{f} = [f_0, \dots, f_M]^t$ et $\mathbf{g} = [g_0, \dots, g_M]^t$.

1. Trouvez la matrice \mathbf{H} de telle sorte que l'on puisse écrire $\mathbf{g} = \mathbf{H}\mathbf{f} + \mathbf{b}$.
2. Quelle est la structure de cette matrice ?
3. Comment peut-on rendre cette matrice circulante ?
4. Trouvez la matrice \mathbf{F} de telle sorte que l'on puisse écrire $\mathbf{g} = \mathbf{F}\mathbf{h} + \mathbf{b}$.
5. Quelle est la structure de cette matrice ?
6. Comment peut-on rendre cette matrice circulante ?
7. Supposons d'abord \mathbf{h} et \mathbf{g} connus. Définissons la solution $\hat{\mathbf{f}}$ par

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f}} \{ |\mathbf{g} - \mathbf{H}\mathbf{f}|^2 + \lambda_f |\mathbf{D}_f \mathbf{f}|^2 \},$$

où \mathbf{D}_f est la matrice de la différence finie (approximation de la dérivation d'ordre un).

8. écrivez l'expression de cette solution.
9. Que devient cette solution lorsque $\lambda_f = 0$?
10. Proposez une méthode pour calculer cette solution et commenter votre choix.
11. Supposons que nous avons rendu les matrices \mathbf{H} et \mathbf{D}_f circulantes ($\mathbf{H} = \text{circ}(\mathbf{h})$ et $\mathbf{D}_f = \text{circ}(\mathbf{d}_f)$). Montrez alors que cette solution peut être calculée par

$$F(\nu) = \frac{H^*(\nu)}{|H(\nu)|^2 + \lambda_f |D_f(\nu)|^2} G(\nu),$$

où $F(\nu)$ et $G(\nu)$ sont les TFD de \mathbf{f} et de \mathbf{g} et $H(\nu)$ et $D_f(\nu)$ sont les TFD de \mathbf{h} et de \mathbf{d}_f .

12. Supposons maintenant \mathbf{f} et \mathbf{g} connus. Définissons la solution $\hat{\mathbf{h}}$:

$$\hat{\mathbf{h}} = \arg \min_{\mathbf{h}} \{ |\mathbf{g} - \mathbf{F}\mathbf{h}|^2 + \lambda_h |\mathbf{D}_h \mathbf{h}|^2 \}$$

où \mathbf{D}_h est une matrice connue.

13. écrivez l'expression de cette solution.
14. Que devient cette solution lorsque $\lambda_h = 0$?
15. Proposez une méthode pour calculer cette solution et commenter votre choix.
16. Supposons que nous avons rendu les matrices \mathbf{F} et \mathbf{D}_h circulantes ($\mathbf{F} = \text{circ}(\mathbf{f})$ et $\mathbf{D}_h = \text{circ}(\mathbf{d}_h)$). Montrez alors que cette solution peut être calculée par

$$H(\nu) = \frac{F^*(\nu)}{|F(\nu)|^2 + \lambda_h |D_h(\nu)|^2} G(\nu).$$

où $F(\nu)$ et $G(\nu)$ sont les TF de \mathbf{f} et de \mathbf{g} et $H(\nu)$ et $D_h(\nu)$ sont les TF de \mathbf{h} et de \mathbf{d}_h .

17. Supposons maintenant qu'on souhaite estimer \mathbf{h} et \mathbf{f} à partir de la seule mesure de \mathbf{g} (déconvolution aveugle). Peut-on envisager de définir la solution par

$$\begin{aligned} (\hat{\mathbf{f}}, \hat{\mathbf{h}}) &= \arg \min_{(\mathbf{f}, \mathbf{h})} \{ |\mathbf{g} - \mathbf{F}\mathbf{h}|^2 + \lambda_f |\mathbf{D}_f \mathbf{f}|^2 + \lambda_h |\mathbf{D}_h \mathbf{h}|^2 \} \\ &= \arg \min_{(\mathbf{f}, \mathbf{h})} \{ |\mathbf{g} - \mathbf{H}\mathbf{f}|^2 + \lambda_f |\mathbf{D}_f \mathbf{f}|^2 + \lambda_h |\mathbf{D}_h \mathbf{h}|^2 \} \end{aligned}$$

Pourquoi? Ce critère a-t-il une seule solution. Commenter votre réponse.

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

Problème 1 : Dans un système d'imagerie, nous avons établi la relation $\mathbf{g} = \mathbf{H}\mathbf{f} + \boldsymbol{\epsilon}$ où \mathbf{g} est un vecteur contenant les projections (mesures) $\{g_m, m = 1 \cdots, M\}$, $\boldsymbol{\epsilon}$ est un vecteur représentant les erreurs de mesures $\{\epsilon_m, m = 1 \cdots, M\}$, \mathbf{f} est un vecteur représentant les pixels de l'image $\{f_n, n = 1 \cdots, N\}$, et \mathbf{H} est une matrice dont les éléments $\{a_{mn}\}$ dépendent de la géométrie du système et sont supposés connus.

- Supposons d'abord que $M = N$ et que la matrice \mathbf{H} soit inversible. Pourquoi la solution $\hat{\mathbf{f}}_0 = \mathbf{H}^{-1}\mathbf{g}$ n'est, en général, pas une solution satisfaisante ?

Quelle relation existe-t-il entre $\frac{\|\delta\hat{\mathbf{f}}_0\|}{\|\hat{\mathbf{f}}_0\|}$ et $\frac{\|\delta\mathbf{g}\|}{\|\mathbf{g}\|}$?

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

$$J_1(\mathbf{f}) = \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2$$

Montrez que, toute solution de l'équation $\mathbf{H}'\mathbf{H}\mathbf{f} = \mathbf{H}'\mathbf{g}$ est une solution au sens des moindres carrés du problème et que lorsque $\mathbf{H}'\mathbf{H}$ est inversible il existe une solution unique donnée par

$$\hat{\mathbf{f}}_1 = [\mathbf{H}'\mathbf{H}]^{-1}\mathbf{H}'\mathbf{g}$$

Quelle relation existe-t-il alors entre $\frac{\|\delta\hat{\mathbf{f}}_1\|}{\|\hat{\mathbf{f}}_1\|}$ et $\frac{\|\delta\mathbf{g}\|}{\|\mathbf{g}\|}$?

- Quelle relation existe entre la covariance de $\hat{\mathbf{f}}_1$ et covariance de \mathbf{g} ?
- Considérons maintenant le cas $M < N$. En évidence, $\mathbf{g} = \mathbf{H}\mathbf{f}$ a une infinité de solutions. La solution de norme minimale s'écrit:

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{H}\mathbf{f}=\mathbf{g}} \{\|\mathbf{f}\|^2\}$$

Montrez que cette solution s'obtient par la résolution du système d'équations suivante:

$$\begin{bmatrix} \mathbf{I} & -\mathbf{H}^t \\ \mathbf{H} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{f} \\ \lambda \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{g} \end{bmatrix}$$

qui donne finalement

$$\hat{\mathbf{f}}_2 = \mathbf{H}^t(\mathbf{H}\mathbf{H}^t)^{-1}\mathbf{g}$$

si $\mathbf{H}\mathbf{H}^t$ est inversible.

- Montrez que avec cette solution $\hat{\mathbf{g}} = \mathbf{H}\hat{\mathbf{f}}_2 = \mathbf{g}$.
- Quelle relation existe entre la covariance de $\hat{\mathbf{f}}_2$ et covariance de \mathbf{g} ?
- Revenons au cas général $M \neq N$ et définissons

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f}} \{J(\mathbf{f})\} \quad \text{avec} \quad J(\mathbf{f}) = \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2 + \lambda\|\mathbf{f}\|^2$$

Montrez que, pour tout $\lambda > 0$, cette solution existe et unique et s'écrit:

$$\hat{\mathbf{f}} = [\mathbf{H}'\mathbf{H} + \lambda\mathbf{I}]^{-1}\mathbf{H}'\mathbf{g}$$

- Quelle relation existe entre $\hat{\mathbf{g}} = \mathbf{H}\hat{\mathbf{f}}$ et \mathbf{g} ?
- Quelle relation existe entre la covariance de $\hat{\mathbf{f}}$ et covariance de \mathbf{g} ?

10. Une autre solution régularisée $\hat{\mathbf{f}}_2$ à ce problème est de minimiser un critère de la forme

$$J_2(\mathbf{f}) = \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2 + \lambda\|\mathbf{D}\mathbf{f}\|^2,$$

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

Montrez que cette solution s'écrit:

$$\hat{\mathbf{f}}_2 = \arg \min_{\mathbf{f}} \{J_2(\mathbf{f})\} = [\mathbf{H}'\mathbf{H} + \lambda\mathbf{D}'\mathbf{D}]^{-1} \mathbf{H}'\mathbf{g}$$

Pourquoi cette solution est-elle préférable à $\hat{\mathbf{f}}_0$ et à $\hat{\mathbf{f}}_1$ et $\hat{\mathbf{f}}_2$?

11. Supposons que \mathbf{H} et \mathbf{D} soient des matrices circulantes et symétriques. Dans ce cas montrez que la solution régularisée $\hat{\mathbf{f}}_2$ peut être obtenue à l'aide de la TFD (Transformée de Fourier Discrète). Plus précisément montrez la relation suivante:

$$F(\omega) = \frac{1}{H(\omega)} \frac{|H(\omega)|^2}{|H(\omega)|^2 + \lambda|D(\omega)|^2} G(\omega)$$

où

- $H(\omega)$ est la TFD de la première ligne de la matrice \mathbf{H} ,
- $D(\omega)$ est la TFD de la première ligne de la matrice \mathbf{D}
- $F(\omega)$ est la TFD du vecteur solution $\hat{\mathbf{f}}_2$, et
- $G(\omega)$ est la TFD du vecteur de mesure \mathbf{g} .

12. Commenter les expressions de $\hat{\mathbf{f}}_2$ dans la question 3. et de $F(\omega)$ dans la question 4. lorsque $\lambda = 0$ et lorsque $\lambda \rightarrow \infty$.

Problème 2:

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

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

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

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

$$J_1(\mathbf{M}) = \|\mathbf{R} - \mathbf{I}\|^2 = \|\mathbf{MH} - \mathbf{I}\|^2$$

soit minimale. Montrez alors que la solution s'écrit

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

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

$$J_2(\mathbf{M}) = \|\mathbf{N} - \mathbf{I}\|^2 = \|\mathbf{HM} - \mathbf{I}\|^2$$

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

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

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

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

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

- Définissons $J(\mathbf{M}) = \alpha_1 J_1(\mathbf{M}) + \alpha_2 J_2(\mathbf{M}) + \alpha_3 J_3(\mathbf{M})$. crivez l'expression de $\frac{\partial J}{\partial \mathbf{M}}$ et l'expression de \mathbf{M} qui minimise $J(\mathbf{M})$ pour différentes combinaisons de $(\alpha_1, \alpha_2, \alpha_3)$. Dans chaque cas crivez l'expression de \mathbf{R} , \mathbf{N} et \mathbf{U} .

Vérifier ce tableau :

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

Problème 3:

Considérons le problème $\mathbf{g} = \mathbf{H}\mathbf{f}$ et supposons que \mathbf{f} représente une image avec $f_j \geq 0$ et $\sum_j f_j = 1$. Supposons que le système $\mathbf{g} = \mathbf{H}\mathbf{f}$ 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 $\boldsymbol{\lambda} = \{\lambda_i, i = 1, \dots, M\}$ 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

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

ce qui permet d'écrire

$$- \ln Z(\boldsymbol{\lambda}) = \frac{1}{n} \sum_j [\ln f_j + [\mathbf{H}^t \boldsymbol{\lambda}]_j]$$

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 $D(\boldsymbol{\lambda}) = \|\mathbf{g} - \mathbf{H} \exp[-\mathbf{H}^t \boldsymbol{\lambda}]\|^2$.
4. Crivez 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 routines $\mathbf{g} = \text{direct}(\mathbf{H}, \mathbf{f})$ qui calcule $\mathbf{g} = \mathbf{H}\mathbf{f}$ et $\mathbf{f} = \text{transp}(\mathbf{H}, \mathbf{g})$ qui calcule $\mathbf{f} = \mathbf{H}^t \mathbf{g}$.

Problème 4:

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

$$\mathbf{g} = \mathbf{H}\mathbf{f} \longrightarrow g_i = \sum_{j=1}^N H_{i,j} f_j = \sum_{j=1}^N H_{i,j} E\{Z_j\}, \quad i = 1, \dots, M.$$

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

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

où $\mu(\mathbf{z})$ est une densité de référence connue. Montrez alors que lorsque $\mu(\mathbf{z})$ est séparable, *i.e.*; $\mu(\mathbf{z}) = \prod_{j=1}^N \mu(z_j)$, alors $p(\mathbf{z})$ l'est aussi, *i.e.*;

$$p(\mathbf{z}) = \prod_{j=1}^N p(z_j) \quad \text{avec} \quad p(z_j) \propto \mu(z_j) \exp \left[\sum_{i=1}^M H_{i,j} \lambda_i z_j \right],$$

Notant que $f_j = E\{Z_j\}$ montrez que

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

$$p(z_j) \propto \exp\left[-\frac{1}{2}z_j^2 + \sum_{i=1}^M H_{i,j} \lambda_i z_j\right],$$

et par conséquent

$$f_j = \sum_{i=1}^M H_{i,j} \lambda_i, \text{ ou encore } \mathbf{f} = \mathbf{H}' \boldsymbol{\lambda}$$

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

$$g_i = \sum_{j=1}^N H_{i,j} \sum_{m=0}^M H_{i,j} \lambda_i, \text{ ou encore } \mathbf{g} = \mathbf{H} \mathbf{H}' \boldsymbol{\lambda},$$

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

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

Interpréter alors ce résultat.

2. si $\mu(z_j) \propto \exp[-z_j]$, $z_j > 0$ on a

$$p(z_j) \propto \exp\left[-z_j - \sum_{i=1}^M H_{i,j} \lambda_i z_j\right], \quad z_j > 0$$

et par conséquent

$$f_j = 1 + \sum_{i=1}^M H_{i,j} \lambda_i, \text{ ou encore } \mathbf{f} = \mathbf{1} + \mathbf{H}' \boldsymbol{\lambda}$$

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

$$g_i = \sum_{j=1}^N H_{i,j} \left(1 + \sum_{i=0}^M H_{i,j} \lambda_i\right), \text{ ou encore } \mathbf{g} = \mathbf{H} \mathbf{1} + \mathbf{H} \mathbf{H}' \boldsymbol{\lambda},$$

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

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

Interpréter alors ce résultat.

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

$$p(z_j) \propto \exp\left[(\alpha-1) \ln z_j - \sum_{i=1}^M H_{i,j} \lambda_i z_j\right], \quad z_j > 0$$

qui est une loi *Gamma*(α, β) avec $\beta = \sum_{i=1}^M H_{i,j} \lambda_i$ et par conséquent

$$f_j = \mathbb{E}\{Z_j\} = \frac{\alpha}{\beta} = \frac{\alpha}{\sum_{i=1}^M H_{i,j} \lambda_i},$$

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

$$g_i = \sum_{j=1}^N H_{i,j} \frac{\alpha}{\sum_{i=1}^M H_{i,j} \lambda_i}$$

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

$$\begin{aligned} \mathbf{f} &= \alpha \mathbf{1} ./ \mathbf{H}' \boldsymbol{\lambda} \\ \mathbf{g} &= \mathbf{H} (\alpha \mathbf{1} ./ \mathbf{H}' \boldsymbol{\lambda}) = (\alpha \mathbf{H} \mathbf{1}) ./ (\mathbf{H}' \boldsymbol{\lambda}) \end{aligned}$$

Interpréter alors ce résultat.

Exercice numéro 6: Estimation au sens du MV et estimation bayésienne

Considérons le système $\mathbf{g} = \mathbf{H}\mathbf{f} + \boldsymbol{\epsilon}$

1. Dans une approche d'estimation au sens du maximum de vraisemblance (MV) supposons que le bruit $\boldsymbol{\epsilon}$ soit supposé centré, blanc, gaussien et de variance σ_ϵ^2 fixée. Montrez que l'estimation au sens du MV de \mathbf{f} , *i.e.*;

$$\hat{\mathbf{f}}_{\text{MV}} = \arg \max_{\mathbf{f}} \{p(\mathbf{g}|\mathbf{f})\}$$

s'obtient en minimisant

$$J_1(\mathbf{f}) = \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2.$$

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

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

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

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_\epsilon), \quad \mathbf{f} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_f),$$

où $\mathbf{R}_\epsilon = \sigma_\epsilon^2 \mathbf{I}$ et $\mathbf{R}_f = \sigma_f^2 \mathbf{P}_0 = \sigma_f^2 (\mathbf{D}'\mathbf{D})^{-1}$ sont les matrices de covariance de $\boldsymbol{\epsilon}$ et de \mathbf{f} .
crivez l'expression des lois $p(\mathbf{f})$, $p(\boldsymbol{\epsilon})$, $p(\mathbf{g}, \mathbf{f})$ et $p(\mathbf{g}|\mathbf{f})$.

4. Montrez que la loi *a posteriori* $p(\mathbf{f}|\mathbf{g})$ est une loi gaussienne de la forme

$$p(\mathbf{f}|\mathbf{g}) \propto \exp\left[-\frac{1}{2} \left[[\mathbf{g} - \mathbf{H}\mathbf{f}]' \mathbf{R}_\epsilon^{-1} [\mathbf{g} - \mathbf{H}\mathbf{f}] + \mathbf{f}' \mathbf{R}_f^{-1} \mathbf{f} \right]\right]$$

5. Si on note $\hat{\mathbf{f}}_3$ la solution qui maximise $p(\mathbf{f}|\mathbf{g})$ (l'estimation au sens du maximum *a posteriori* MAP), montrez qu'elle s'obtient par

$$\hat{\mathbf{f}}_3 = \arg \min_{\mathbf{f}} \left\{ J_3(\mathbf{f}) = [\mathbf{g} - \mathbf{H}\mathbf{f}]' \mathbf{R}_\epsilon^{-1} [\mathbf{g} - \mathbf{H}\mathbf{f}] + \mathbf{f}' \mathbf{R}_f^{-1} \mathbf{f} \right\}$$

6. Si $\mathbf{R}_\epsilon = \sigma_\epsilon^2 \mathbf{I}$ et $\mathbf{R}_f = \sigma_f^2 \mathbf{P}_0$, Montrez que $\hat{\mathbf{f}}_3$ s'obtient par

$$\hat{\mathbf{f}}_3 = [\mathbf{H}'\mathbf{H} + \lambda \mathbf{P}_0^{-1}]^{-1} \mathbf{H}'\mathbf{g} \quad \text{avec} \quad \lambda = \frac{\sigma_\epsilon^2}{\sigma_f^2}$$

Que peut-on conclure en comparant les solutions $\hat{\mathbf{f}}_2$ et $\hat{\mathbf{f}}_3$?

7. En développant le terme

$$[\mathbf{g} - \mathbf{H}\mathbf{f}]' \mathbf{R}_\epsilon^{-1} [\mathbf{g} - \mathbf{H}\mathbf{f}] + \mathbf{f}' \mathbf{R}_f^{-1} \mathbf{f}$$

dans l'expression de $p(\mathbf{f}|\mathbf{g})$ montrez que l'on peut écrire

$$p(\mathbf{f}|\mathbf{g}) \propto \exp\left[-\frac{1}{2} [\mathbf{f} - \hat{\mathbf{f}}_3]' \hat{\mathbf{P}}^{-1} [\mathbf{f} - \hat{\mathbf{f}}_3]\right]$$

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

8. crivez l'expression des lois $p(f_n|\mathbf{g})$, $p(g_m|\mathbf{g})$ et $p(b_m|\mathbf{g})$. A quoi peuvent-elle servir ?

9. Supposons maintenant la suite $\{f_1, \dots, f_N\}$ puissent être modélisée par une chaîne de Markov d'ordre un, c'est à dire :

$$p(f_n | f_1, \dots, f_N) = p(f_n | f_{n-1})$$

Peut-on calculer $p(\mathbf{f})$? Et si on connaît de plus $p(f_1)$?

Que devient alors la solution au sens du MAP ?

Étudiez cette solution dans les deux cas suivants:

$$p(f_n | f_1, \dots, f_{n-1}, f_{n+1}, \dots, f_N) = p(f_n | f_{n-1}) = \mathcal{N}(f_n - f_{n-1}, \sigma_f^2), \quad \text{et} \quad p(f_1) = \mathcal{N}(0, \sigma_f^2)$$

et

$$p(f_n | f_1, \dots, f_{n-1}, f_{n+1}, \dots, f_N) = p(f_n | f_{n-1}) \propto \exp[-\alpha \phi(f_n - f_{n-1})], \quad \text{et} \quad p(f_1) \propto \exp[-\alpha \phi(f_1)]$$

Exercice numéro 7: Calcul bayésien dans le cas d'un modèle linéair et des lois Gaussiennes

Tout d'abord rappelons les relations suivantes:

$$\begin{aligned} \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= (2\pi)^{-d/2} |\boldsymbol{\Sigma}|^{-1/2} \exp \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right] \\ \int (\mathbf{x} - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \mathcal{N}(\mathbf{m}, \mathbf{S}) \, d\mathbf{x} &= (\boldsymbol{\mu} - \mathbf{m})^t \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \mathbf{m}) + \text{Tr} [\boldsymbol{\Sigma}^{-1} \mathbf{S}] \\ \int (\mathbf{W}\mathbf{x} - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\mathbf{W}\mathbf{x} - \boldsymbol{\mu}) \mathcal{N}(\mathbf{m}, \mathbf{S}) \, d\mathbf{x} &= (\boldsymbol{\mu} - \mathbf{W}\mathbf{m})^t \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \mathbf{W}\mathbf{m}) + \text{Tr} [\mathbf{W}^t \boldsymbol{\Sigma}^{-1} \mathbf{W}\mathbf{S}] \\ \mathcal{N}(\mathbf{a}, \mathbf{A}) * \mathcal{N}(\mathbf{b}, \mathbf{B}) &\propto \mathcal{N}(\mathbf{a} + \mathbf{b}, \mathbf{A} + \mathbf{B}) \\ \mathcal{N}(\mathbf{a}, \mathbf{A}) \mathcal{N}(\mathbf{b}, \mathbf{B}) &\propto \mathcal{N}(\mathbf{c}, \mathbf{C}) \quad \text{avec} \quad \mathbf{C} = (\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1} \quad \text{et} \quad \mathbf{c} = \mathbf{C}[\mathbf{A}^{-1}\mathbf{a} + \mathbf{B}^{-1}\mathbf{b}] \end{aligned}$$

Si

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{C}^t & \mathbf{B} \end{bmatrix} \right)$$

alors

$$\begin{cases} \mathbf{x} \sim \mathcal{N}(\mathbf{a}, \mathbf{A}) \\ \mathbf{y} \sim \mathcal{N}(\mathbf{b}, \mathbf{B}) \end{cases} \quad \text{et} \quad \begin{cases} \mathbf{x}|\mathbf{y} \sim \mathcal{N}(\mathbf{a} + \mathbf{C}\mathbf{B}^{-1}(\mathbf{y} - \mathbf{b}), \mathbf{A} - \mathbf{C}\mathbf{B}^{-1}\mathbf{C}^t) \\ \mathbf{y}|\mathbf{x} \sim \mathcal{N}(\mathbf{b} + \mathbf{C}^t\mathbf{A}^{-1}(\mathbf{x} - \mathbf{a}), \mathbf{B} - \mathbf{C}^t\mathbf{A}^{-1}\mathbf{C}) \end{cases}$$

Considérons maintenant les vecteurs \mathbf{g} , \mathbf{f} et $\boldsymbol{\epsilon}$ liés par la relation linéaire $\mathbf{g} = \mathbf{H}\mathbf{f} + \boldsymbol{\epsilon}$ où nous supposons $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I})$ et $\mathbf{f} \sim \mathcal{N}(\mathbf{f}_0, \mathbf{P}_0)$ avec $\mathbf{P}_0 = (\mathbf{D}^t \mathbf{D})^{-1}$.

1. Ecrivez les expressions de $p(\boldsymbol{\epsilon})$, $p(\mathbf{g}|\mathbf{f})$, $p(\mathbf{f})$ et $p(\mathbf{g})$
2. Ecrivez les expressions de $p(\mathbf{g}, \mathbf{f})$
3. Trouvez les expressions de \mathbf{a} , \mathbf{b} , \mathbf{A} , \mathbf{B} et \mathbf{C} dans la relation suivante:

$$\begin{bmatrix} \mathbf{g} \\ \mathbf{f} \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{C}^t & \mathbf{B} \end{bmatrix} \right)$$

et en déduire les expressions de $p(\mathbf{f}|\mathbf{g})$

4. Donner les expressions de la moyenne et de la covariance *a posteriori* $\hat{\mathbf{f}}$ et $\hat{\boldsymbol{\Sigma}}$.
5. Montrer que $\hat{\mathbf{f}}$ peut être obtenu en minimisant $J(\mathbf{f}) = \frac{1}{\sigma_\epsilon^2} \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2 + \|\mathbf{D}(\mathbf{f} - \mathbf{f}_0)\|^2$

Définissons maintenant deux vecteurs $\mathbf{f}_1 = \mathbf{f}_0 + \mathbf{D}^{-1}\boldsymbol{\epsilon}_1$ et $\mathbf{g}_1 = \mathbf{g} + \sigma_\epsilon \boldsymbol{\epsilon}_2$ où $\boldsymbol{\epsilon}_1$ et $\boldsymbol{\epsilon}_2$ sont deux vecteurs aléatoires gaussiens $\boldsymbol{\epsilon}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ et $\boldsymbol{\epsilon}_2 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ et $\mathbf{x} = \hat{\boldsymbol{\Sigma}} \left(\frac{1}{\sigma_\epsilon^2} \mathbf{H}^t \mathbf{g}_1 + \mathbf{D}^t \mathbf{D} \mathbf{f}_1 \right)$.

1. Montrer que $\mathbf{x} = \hat{\mathbf{f}} + \hat{\boldsymbol{\Sigma}} \left(\frac{1}{\sigma_\epsilon^2} \mathbf{H}^t \boldsymbol{\epsilon}_2 + \mathbf{D}^t \mathbf{D} \boldsymbol{\epsilon}_1 \right)$.
2. Montrer que $E\{\mathbf{x}\} = \hat{\mathbf{f}}$ et $\text{Cov}[\mathbf{x}] = \hat{\boldsymbol{\Sigma}}$.
3. Montrer que pour générer un échantillon de la loi *a posteriori* $p(\mathbf{f}|\mathbf{g})$ on peut utiliser un des deux schémas suivants:
 - Schéma 1:
 - Décomposer la matrice de covariance *a posteriori* $\hat{\boldsymbol{\Sigma}} = \mathbf{W}\mathbf{W}^t$
 - Générer un vecteur gaussien $\boldsymbol{\epsilon}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
 - Calculer $\hat{\mathbf{f}} = \mathbf{W}\boldsymbol{\epsilon}_1$
 - Schéma 2:
 - Générer deux vecteurs gaussiens indépendants $\boldsymbol{\epsilon}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ et $\boldsymbol{\epsilon}_2 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
 - Former deux vecteurs $\mathbf{f}_1 = \mathbf{f}_0 + \mathbf{D}^{-1}\boldsymbol{\epsilon}_1$ et $\mathbf{g}_1 = \mathbf{g} + \sigma_\epsilon \boldsymbol{\epsilon}_2$
 - Obtenir $\hat{\mathbf{f}}$ en minimisant $J(\mathbf{f}) = \frac{1}{\sigma_\epsilon^2} \|\mathbf{g}_1 - \mathbf{H}\mathbf{f}_1\|^2 + \|\mathbf{D}\mathbf{f}_1\|^2$
4. Comparer les coûts de calculs et de mémoires de ces deux schémas.

Exercice numéro 8: Modélisation des signaux et estimation au sens du MAP

Problème 1 : Modélisations simples

Nous avons vu que le calcul de l'estimateur au sens du MAP du $\mathbf{g} = \mathbf{H}\mathbf{f} + \boldsymbol{\epsilon}$ se ramène à l'optimisation du critère $J(\mathbf{f}) = \sigma_\epsilon^{-2} \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2 + \Omega(\mathbf{f})$ où dans le cas d'une hypothèse *a priori* i.i.d. pour f_j est donnée par $\Omega(\mathbf{f}) = -\sum_j \ln p(f_j)$.

1. Vérifiez alors les relations suivantes :

Gaussienne : $p(f_j|\lambda) = \mathcal{N}(0, \frac{1}{2\lambda})$

$$p(f_j|\lambda) = \sqrt{\frac{\lambda}{\pi}} \exp[-\lambda|f_j|^2] \quad \longrightarrow \quad \Omega(\mathbf{f}) = \lambda \sum_j |f_j|^2$$

$\Omega(\mathbf{f})$ est quadratique en \mathbf{f} .

Exponentielle double (Laplace) : $p(f_j) = \mathcal{E}(\lambda)$

$$p(f_j|\lambda) = \frac{\lambda}{2} \exp[-\lambda|f_j|] \quad \longrightarrow \quad \Omega(\mathbf{f}) = \lambda \sum_j |f_j|$$

$\Omega(\mathbf{f})$ n'est pas continuellement dérivable en \mathbf{f} .

Exponentielle généralisée : $p(f_j) = \mathcal{GE}(\lambda, \beta)$

$$p(f_j|\lambda, \beta) \propto \exp[-\lambda|f_j|^\beta] \quad \longrightarrow \quad \Omega(\mathbf{f}) = \lambda \sum_j |f_j|^\beta$$

$\Omega(\mathbf{f})$ est continuellement dérivable en \mathbf{f} si $1 < \beta \leq 2$.

Cauchy : $p(f_j) = \mathcal{C}(\lambda)$

$$\widehat{p}(f_j|\lambda) = \frac{\pi\sqrt{\lambda}}{1 + \lambda|f_j|^2} \quad \longrightarrow \quad \Omega(\mathbf{f}) = \lambda \sum_j \ln(1 + \lambda|f_j|^2)$$

$\Omega(\mathbf{f})$ n'est pas convexe en \mathbf{f} .

2. Dans chaque cas, donner l'expression du gradient de $\Omega(\mathbf{f})$ lorsque possible et proposez un algorithme approprié pour calculer la solution $\widehat{\mathbf{f}}$ qui optimise $J(\mathbf{f})$.
3. Considérons maintenant un problème de débruitage de signaux $g(t) = f(t) + \epsilon(t)$ qui s'écrit aussi $\mathbf{g} = \mathbf{f} + \boldsymbol{\epsilon}$ et considérons le critère $J(\mathbf{f}) = \sigma_\epsilon^{-2} \|\mathbf{g} - \mathbf{f}\|^2 + \Omega(\mathbf{f})$. Montrez alors que, avec ces lois *a priori*, on obtient $\widehat{f}(t) = \phi(g(t))$ ou $\widehat{f}_j = \phi(g_j)$. Donnez alors l'expression de la fonction $\phi(\cdot)$ et tracer les courbes $\widehat{f}_j = \phi(g_j)$.
4. Supposons maintenant qu'au lieu de modéliser f_j par $p(f_j)$ on modélise $\Delta_j = (f_j - f_{j-1})$ par $p(\Delta_j)$ par ces lois en supposant que f_0 suit aussi la même loi $p(f_0)$. Montrez alors qu'on obtient $\widehat{f}_j = \phi(g_j - g_{j-1})$.
5. Supposons maintenant f_j est modélisée par un processus à moyenne mobile (MA) d'ordre K : $f_j = \sum_{k=0}^{K-1} h_k \eta_{j-k}$ où on suppose que les η_j sont iid et suivent une des lois proposées. Ecrivez alors les expressions de $p(\mathbf{f})$ et l'estimateur MAP correspondant dans chaque cas.
6. Supposons maintenant f_j est modélisée par un processus autoregressif (AR) d'ordre K : $f_j = \sum_{k=1}^K a_k f_{j-k} + \eta_j$ où on suppose que les η_j sont iid et suivent une des lois proposées. Ecrivez alors les expressions de $p(\mathbf{f})$ et l'estimateur MAP correspondant dans chaque cas. Commencer par le cas $K = 1$ et $a_1 = 1$.
7. Simuler un signal $f(t) = 0$, $t \in [0, 100]$ et $f(t) = 2$, $t \in [100, 200]$ et rajouter un bruit gaussien de variance $\sigma_\epsilon^2 = 1$ pour obtenir $g(t)$. Appliquer ensuite la méthode de débruitage proposée avec ces différents modèles *a priori* et comparez les résultats. Commentez ces résultats.

Problème 2 : Modélisations avec des variables cachées

Nous avons vu que l'introduction d'une variable cachée donne plus de souplesse et de généralité pour la modélisation des signaux et des images. Dans ces modèles, on introduit une variable cachée \mathbf{c} , notée \mathbf{d} , \mathbf{q} ou \mathbf{z} en fonction de leurs natures (\mathbf{d} pour des variables réelles, \mathbf{q} pour des variables binaires et \mathbf{z} pour des variables discrètes). Il faut évidemment estimer ces variables cachées en même temps que \mathbf{f} . En général, on est amené à utiliser un algorithme itératif qui consiste à calculer itérativement \mathbf{f} à \mathbf{c} fixé et \mathbf{c} à \mathbf{f} fixé.

Modèle gaussien à variance modulable :

$$p(f_j|d_j, \lambda) = \mathcal{N}(0, 2d_j/\lambda) \quad \text{et} \quad p(d_j|\lambda) = \mathcal{G}(3/2, \lambda)$$

- Vérifiez alors les relations suivantes :

$$p(\mathbf{f}, \mathbf{d}) \propto \exp \left[-\lambda \sum_j \left(\frac{f_j^2}{4d_j} + d_j \right) \right]$$

$$p(\mathbf{g}|\mathbf{f}) \propto \exp [\sigma_\epsilon^{-2} \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2] \quad \longrightarrow \quad p(\mathbf{f}, \mathbf{d}|\mathbf{g}) \propto \exp [-J(\mathbf{f}, \mathbf{d})]$$

$$\text{avec} \quad J(\mathbf{f}, \mathbf{d}) = \sigma_\epsilon^{-2} \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2 + \lambda \sum_j \left(\frac{f_j^2}{4d_j} + d_j \right)$$

- Optimisation par rapport à \mathbf{f} à \mathbf{d} fixé :

$$\hat{\mathbf{f}} = (\sigma_\epsilon^{-2} \mathbf{H}^t \mathbf{H} + 2\lambda \mathbf{D})^{-1} \mathbf{H}^t \mathbf{g} \quad \text{avec} \quad \mathbf{D} = \text{diag} [1/(4d_j), j = 1, \dots, n]$$

- Optimisation par rapport à \mathbf{d} à \mathbf{f} fixé :

$$\hat{d}_j = f_j/2$$

- Vérifier les relations suivantes :

$$d_j = \inf_{f_j} \left(\frac{f_j^2}{4d_j} + d_j \right) \quad \text{et} \quad f_j/2 = \sup_{d_j} \left(\frac{f_j^2}{4d_j} + d_j \right)$$

Modèle gaussien à moyenne modulable :

$$p(f_j|z_j, \lambda) = \mathcal{N}(z_j, 2/\lambda) \quad \text{et} \quad z_j \in \{m_1 = -1, m_2 = 0, m_3 = +1\}, \quad P(z_j = m_k) = (1/3), k = 1, \dots, K = 3$$

- Vérifiez alors les relations suivantes :

$$p(f_j|\lambda) = (1/3)[\mathcal{N}(0, 2/\lambda) + \mathcal{N}(-1, 2/\lambda) + \mathcal{N}(+1, 2/\lambda)]$$

$$p(\mathbf{f}|\mathbf{z}) \propto \exp \left[-\lambda \sum_j (f_j - z_j)^2 \right]$$

$$p(\mathbf{z}|\mathbf{f}) \propto \exp \left[-\lambda \sum_j (z_j - f_j)^2 \right]$$

$$p(\mathbf{g}|\mathbf{f}) \propto \exp [\sigma_\epsilon^{-2} \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2] \quad \longrightarrow \quad p(\mathbf{f}, \mathbf{z}|\mathbf{g}) \propto \exp [-J(\mathbf{f}, \mathbf{z})]$$

$$\text{avec} \quad J(\mathbf{f}, \mathbf{z}) = \sigma_\epsilon^{-2} \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2 + \lambda \sum_j (f_j - z_j)^2 = \sigma_\epsilon^{-2} \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2 + \lambda \|\mathbf{f} - \mathbf{z}\|^2 + \ln(1/3) \sum_k \sum_j \delta(z_j - m_k)$$

où $\mathbf{z} = [z_1, \dots, z_N]'$.

- Optimisation par rapport à \mathbf{f} à \mathbf{z} fixé :

$$\hat{\mathbf{f}} = (\sigma_\epsilon^{-2} \mathbf{H}^t \mathbf{H} + \lambda \mathbf{I})^{-1} [\mathbf{H}^t \mathbf{g} + \lambda \mathbf{z}]$$

- Optimisation par rapport à \mathbf{z} à \mathbf{f} fixé :

$$\hat{z}_j = \begin{cases} +1 & f_j > a \\ -1 & f_j < -a \\ 0 & -a < f_j < a \end{cases}$$

- Donner l'expression de a en fonction de λ . Pour cela, il est utile de tracer la courbe de $p(z_j|f_j, \lambda)$ en fonction de f_j .

- Généraliser cette modélisation pour le cas :

$$p(f_j|z_j = k, \lambda) = \mathcal{N}(m_k, v_k = 2/\lambda_k) \quad \text{et} \quad z_j \in \{1, \dots, K\}, \quad P(z_j = k) = \pi_k$$

Exercice numéro 9: Calcul bayésien

Problème 1 : Calcul variationnel

Supposons avoir un modèle \mathcal{M} , une distribution $P(\mathbf{X})$ sur un ensemble de variables \mathbf{X} qui est divisé en trois parties : variables observées \mathbf{g} , les variables cachées \mathbf{f} et les paramètres $\boldsymbol{\theta}$. Lors de l'étape de l'apprentissage (identification), on suppose avoir observé l'ensembles $\mathbf{D} = \{\mathbf{g}, \mathbf{f}\}$ et on suppose que le modèle est vrai et on cherche à apprendre $\boldsymbol{\theta}$ en utilisant la règle de Bayes:

$$P(\boldsymbol{\theta}|\mathbf{D}, \mathcal{M}) = \frac{P(\mathbf{D}|\boldsymbol{\theta}, \mathcal{M}) P(\boldsymbol{\theta}|\mathcal{M})}{P(\mathbf{D}|\mathcal{M})} \quad (1)$$

ou encore

$$p(\boldsymbol{\theta}|\mathbf{g}, \mathbf{f}, \mathcal{M}) = \frac{p(\mathbf{g}, \mathbf{f}|\boldsymbol{\theta}, \mathcal{M}) p(\boldsymbol{\theta}|\mathcal{M})}{p(\mathbf{g}, \mathbf{f}|\mathcal{M})} \quad (2)$$

Supposons maintenant que seul \mathbf{g} est observé, le modèle \mathcal{M} est vrais, les paramètres $\boldsymbol{\theta}$ sont estimés, et on cherche à inférer les variables cachées \mathbf{f} . On peut de nouveau utiliser le règle de Bayes :

$$p(\mathbf{f}|\mathbf{g}, \boldsymbol{\theta}, \mathcal{M}) = \frac{p(\mathbf{g}|\mathbf{f}, \boldsymbol{\theta}, \mathcal{M}) p(\mathbf{f}|\boldsymbol{\theta}, \mathcal{M})}{p(\mathbf{g}|\boldsymbol{\theta}, \mathcal{M})} \quad (3)$$

- Sélection de modèle: On souhaite inférer sur le modèle \mathcal{M}_i :

$$P(\mathcal{M}_i|\mathbf{D}) = \frac{P(\mathbf{D}|\mathcal{M}_i) p(\mathcal{M}_i)}{p(\mathbf{D})} \quad (4)$$

Le calcul de $P(\mathcal{M}|\mathbf{D})$ est souvent difficile. On cherchera alors à l'approximer par une lois plus simple $Q(\mathcal{M})$, mais qu'il soit la plus proche de $P(\mathcal{M}|\mathbf{D})$ au sense de la distance de Kulback-Leibler :

$$KL(Q, P) = \int_{\mathcal{M}} Q(\mathcal{M}) \ln \frac{Q(\mathcal{M})}{P(\mathcal{M}|\mathbf{D})} d\mathcal{M} \quad (5)$$

1. Montrez que

$$KL(Q, P) = \int_{\mathcal{M}} Q(\mathcal{M}) \ln \frac{Q(\mathcal{M})}{P(\mathcal{M}, \mathbf{D})} d\mathcal{M} + \ln P(\mathbf{D}) \quad (6)$$

2. Notant

$$\mathcal{F}(Q) = \int_{\mathcal{M}} Q(\mathcal{M}) \ln \frac{P(\mathcal{M}, \mathbf{D})}{Q(\mathcal{M})} d\mathcal{M} \quad (7)$$

montrez que

$$\mathcal{F}(Q) = \langle \ln P(\mathcal{M}, \mathbf{D}) \rangle_{Q(\mathcal{M})} + H(Q) \quad (8)$$

où

$$H(Q) = - \int_{\mathcal{M}} Q(\mathcal{M}) \ln Q(\mathcal{M}) d\mathcal{M} \quad (9)$$

et

$$\langle \ln P(\mathcal{M}, \mathbf{D}) \rangle_{Q(\mathcal{M})} = \int_{\mathcal{M}} Q(\mathcal{M}) \ln P(\mathcal{M}, \mathbf{D}) d\mathcal{M} \quad (10)$$

3. Montrez que

$$\ln P(\mathbf{D}) = KL(Q, P) + \mathcal{F}(Q) \longrightarrow \mathcal{F}(Q) \leq \ln P(\mathbf{D}|\mathcal{M}) \quad (11)$$

ce qui signifie que $\mathcal{F}(Q)$ est une borne inférieure de $\ln P(\mathbf{D})$.

4. Montrez qu'à l'optimum $P(\mathbf{D}|\mathcal{M}_i)$ peut être approximée par :

$$P(\mathbf{D}|\mathcal{M}_i) \approx \frac{\exp[\mathcal{F}(Q^*)] p(\mathcal{M}_i)}{p(\mathbf{D})} \quad (12)$$

5. Montrez que si on choisi $Q(\mathcal{M}) = \prod_j Q_j(\mathcal{M}_j)$ on a

$$\mathcal{F}(Q) = \int_{\mathcal{M}_j} Q_j(\mathcal{M}_j) \langle \ln P(\mathcal{M}, \mathbf{D}) \rangle_{\prod_{i \neq j} Q_i(\mathcal{M}_i)} + H(Q_j) + \sum_{i \neq j} H(Q_i) \quad (13)$$

et si optimise $\mathcal{F}(Q)$ par rapport à Q_j on obtient

$$Q_j(\mathcal{M}_j) \propto \exp \left[\langle \ln P(\mathcal{M}, \mathbf{D}) \rangle_{\prod_{i \neq j} Q_i(\mathcal{M}_i)} \right] \quad (14)$$

- Application au cas $\mathbf{g} = \mathbf{H}\mathbf{f} + \boldsymbol{\epsilon}$ lorsqu'on souhaite estimer \mathbf{f} et $\boldsymbol{\theta}$. Si on écrit la loi *a posteriori* jointe :

$$p(\mathbf{f}, \boldsymbol{\theta} | \mathbf{g}) = \frac{p(\mathbf{g} | \mathbf{f}, \boldsymbol{\theta}) p(\mathbf{f} | \boldsymbol{\theta})}{p(\mathbf{g} | \boldsymbol{\theta})} \quad (15)$$

et on cherche à l'approximer par une loi séparable $q(\mathbf{f}, \boldsymbol{\theta} | \mathbf{g}) = q_1(\mathbf{f}) q_2(\boldsymbol{\theta})$.

1. En supposant

$$\begin{aligned} p(\mathbf{g} | \mathbf{f}, \sigma_\epsilon) &= \mathcal{N}(\mathbf{H}\mathbf{f}, \sigma_\epsilon^2 \mathbf{I}), \\ p(\mathbf{f} | \sigma_f^2 \mathbf{I}) &= \mathcal{N}(\mathbf{H}\mathbf{f}, \sigma_f^2 \mathbf{I}), \\ \boldsymbol{\theta} &= (\theta_1 = 1/\sigma_\epsilon^2, \theta_2 = 1/\sigma_f^2) \\ p(\theta_1) &= \mathcal{G}(\alpha_{10}, \beta_{10}) \\ p(\theta_2) &= \mathcal{G}(\alpha_{20}, \beta_{20}), \end{aligned}$$

donner les expressions de $p(\mathbf{g}, \mathbf{f}, \boldsymbol{\theta})$, $p(\mathbf{f}, \boldsymbol{\theta} | \mathbf{g})$, $p(\mathbf{f} | \mathbf{g}, \boldsymbol{\theta})$, et $p(\boldsymbol{\theta} | \mathbf{g}, \mathbf{f})$.

2. Donner les expressions de $KL(q, p)$, $\mathcal{F}(q)$, $\ln p(\mathbf{g}, \mathbf{f}, \boldsymbol{\theta})$, $\langle \ln p(\mathbf{g}, \mathbf{f}, \boldsymbol{\theta}) \rangle_{q_1(\mathbf{f})}$ et $\langle \ln p(\mathbf{g}, \mathbf{f}, \boldsymbol{\theta}) \rangle_{q_2(\boldsymbol{\theta})}$.
 3. Choisissons les familles $q_1(\mathbf{f}) = \delta(\mathbf{f} - \hat{\mathbf{f}})$ et $q_2(\boldsymbol{\theta}) = \delta(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$. Donner les expressions de $\langle \ln p(\mathbf{g}, \mathbf{f}, \boldsymbol{\theta}) \rangle_{q_1(\mathbf{f})}$ et $\langle \ln p(\mathbf{g}, \mathbf{f}, \boldsymbol{\theta}) \rangle_{q_2(\boldsymbol{\theta})}$. Que deviennent alors $\hat{\mathbf{f}}$ et $\hat{\boldsymbol{\theta}}$ au cours des itérations? (Lien avec MAP jointe)
 4. Choisissons $q_1(\mathbf{f})$ dans la même famille que $p(\mathbf{f} | \mathbf{g}, \boldsymbol{\theta})$ et $q_2(\boldsymbol{\theta}) = \delta(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$. Donner les expressions de $\langle \ln p(\mathbf{g}, \mathbf{f}, \boldsymbol{\theta}) \rangle_{q_1(\mathbf{f})}$ et $\langle \ln p(\mathbf{g}, \mathbf{f}, \boldsymbol{\theta}) \rangle_{q_2(\boldsymbol{\theta})}$. Que deviennent alors $\hat{\mathbf{f}}$ et $\hat{\boldsymbol{\theta}}$ au cours des itérations? (Lien avec l'algorithme EM)
 5. Choisissons $q_1(\mathbf{f})$ dans la même famille que $p(\mathbf{f} | \mathbf{g}, \boldsymbol{\theta})$ et $q_2(\boldsymbol{\theta})$ dans la même famille que $p(\boldsymbol{\theta} | \mathbf{g}, \mathbf{f})$. Donner les expressions de $\langle \ln p(\mathbf{g}, \mathbf{f}, \boldsymbol{\theta}) \rangle_{q_1(\mathbf{f})}$ et $\langle \ln p(\mathbf{g}, \mathbf{f}, \boldsymbol{\theta}) \rangle_{q_2(\boldsymbol{\theta})}$. Que deviennent alors $q_1(\mathbf{f})$ et $q_2(\boldsymbol{\theta})$ au cours des itérations? Comparer cet algorithme avec l'algorithme EM.
 6. Choisissons les familles $q_1(\mathbf{f}) = \mathcal{N}(\hat{\mathbf{f}}, \hat{\boldsymbol{\Sigma}})$ et $q_2(\boldsymbol{\theta}) = \mathcal{G}(\hat{\alpha}_{10}, \hat{\beta}_{10}) \mathcal{G}(\hat{\alpha}_{20}, \hat{\beta}_{20})$. Donner les expressions de $(\hat{\mathbf{f}}, \hat{\boldsymbol{\Sigma}})$, $(\hat{\alpha}_{10}, \hat{\beta}_{10})$, $(\hat{\alpha}_{20}, \hat{\beta}_{20})$ et en déduire $\langle \mathbf{f} \rangle_{q_1}$, $\langle \theta_1 \rangle_{q_1}$, $\langle \theta_2 \rangle_{q_1}$,
- Dans l'exemple précédent, nous avons modélisé \mathbf{f} par un champs gaussien. Prenons maintenant le modèle gaussien à variance modulable de l'exercice précédent :

$$\begin{aligned} p(\mathbf{g} | \mathbf{f}, \sigma_\epsilon) &= \mathcal{N}(\mathbf{H}\mathbf{f}, \sigma_\epsilon^2 \mathbf{I}), \\ p(f_j | d_j, \lambda) &= \mathcal{N}(0, 2d_j/\lambda) \quad \text{et} \quad p(d_j | \lambda) = \mathcal{G}(3/2, \lambda) \\ p(\mathbf{f} | \mathbf{d}, \lambda) &= \mathcal{N}(\mathbf{z}, (1/\lambda)\mathbf{D}), \\ p(\mathbf{d} | \lambda) &= \prod_j \mathcal{G}(3/2, \lambda), \\ \boldsymbol{\theta} &= (\theta_1 = 1/\sigma_\epsilon^2, \theta_2 = 1/\sigma_f^2, \boldsymbol{\alpha} = \{\alpha_k, k = 1, \dots, K\}) \\ p(\theta_1) &= \mathcal{G}(\alpha_{10}, \beta_{10}) \\ p(\theta_2) &= \mathcal{G}(\alpha_{20}, \beta_{20}) \\ p(\boldsymbol{\alpha}) &= \mathcal{D}(\boldsymbol{\alpha} | \boldsymbol{\alpha}_0) \end{aligned}$$

L'objectif est donc d'estimer \mathbf{f} , \mathbf{z} et $\boldsymbol{\theta}$. Si on écrit la loi *a posteriori* jointe :

$$p(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta} | \mathbf{g}) = \frac{p(\mathbf{g} | \mathbf{f}, \boldsymbol{\theta}) p(\mathbf{f} | \mathbf{z}, \boldsymbol{\theta}) p(\mathbf{z})}{p(\mathbf{g} | \boldsymbol{\theta})} \quad (16)$$

et on cherche à l'approximer par une loi séparable $q(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta} | \mathbf{g}) = q_1(\mathbf{f}) q_2(\mathbf{z}) q_3(\boldsymbol{\theta})$.

1. Donner les expressions de $p(\mathbf{g}, \mathbf{f}, \mathbf{d}, \boldsymbol{\theta})$, $p(\mathbf{f}, \mathbf{d}, \boldsymbol{\theta} | \mathbf{g})$, $p(\mathbf{f} | \mathbf{g}, \mathbf{d}, \boldsymbol{\theta})$, $p(\mathbf{d} | \mathbf{g}, \mathbf{f}, \boldsymbol{\theta})$ et $p(\boldsymbol{\theta} | \mathbf{g}, \mathbf{f}, \mathbf{d})$.
2. Donner les expressions de $KL(q, p)$, $\mathcal{F}(q)$, $\ln p(\mathbf{g}, \mathbf{f}, \boldsymbol{\theta})$, $\langle \ln p(\mathbf{g}, \mathbf{f}, \mathbf{d}, \boldsymbol{\theta}) \rangle_{q_1(\mathbf{f})}$ et $\langle \ln p(\mathbf{g}, \mathbf{f}, \mathbf{d}, \boldsymbol{\theta}) \rangle_{q_2(\mathbf{d})}$ et $\langle \ln p(\mathbf{g}, \mathbf{f}, \boldsymbol{\theta}) \rangle_{q_3(\boldsymbol{\theta})}$.
3. Choisissons les familles $q_1(\mathbf{f}) = \delta(\mathbf{f} - \hat{\mathbf{f}})$, $q_2(\mathbf{d}) = \delta(\mathbf{d} - \hat{\mathbf{d}})$ et $q_3(\boldsymbol{\theta}) = \delta(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$. Donner les expressions de $\langle \ln p(\mathbf{g}, \mathbf{f}, \mathbf{d}, \boldsymbol{\theta}) \rangle_{q_1(\mathbf{f})}$, $\langle \ln p(\mathbf{g}, \mathbf{f}, \mathbf{d}, \boldsymbol{\theta}) \rangle_{q_2(\mathbf{d})}$ et $\langle \ln p(\mathbf{g}, \mathbf{f}, \mathbf{d}, \boldsymbol{\theta}) \rangle_{q_3(\boldsymbol{\theta})}$. Que deviennent alors $\hat{\mathbf{f}}$, $\hat{\mathbf{d}}$ et $\hat{\boldsymbol{\theta}}$ au cours des itérations? (Lien avec MAP jointe)

4. Choisissons $q_1(\mathbf{f})$ dans la même famille que $p(\mathbf{f}|\mathbf{g}, \mathbf{d}, \boldsymbol{\theta})$, $q_2(\mathbf{d})$ dans la même famille que $p(\mathbf{d}|\mathbf{g}, \mathbf{f}, \boldsymbol{\theta})$ et $q_3(\boldsymbol{\theta}) = \delta(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})$. Donner les expressions de $\langle \ln p(\mathbf{g}, \mathbf{f}, \mathbf{d}, \boldsymbol{\theta}) \rangle_{q_1(\mathbf{f})}$, $\langle \ln p(\mathbf{g}, \mathbf{f}, \mathbf{d}, \boldsymbol{\theta}) \rangle_{q_2(\mathbf{d})}$ et $\langle \ln p(\mathbf{g}, \mathbf{f}, \mathbf{d}, \boldsymbol{\theta}) \rangle_{q_3(\boldsymbol{\theta})}$. Que deviennent alors $\widehat{\mathbf{f}}$, $\widehat{\mathbf{d}}$ et $\widehat{\boldsymbol{\theta}}$ au cours des itérations? (Lien avec l'algorithme EM)
 5. Choisissons $q_1(\mathbf{f})$ dans la même famille que $p(\mathbf{f}|\mathbf{g}, \mathbf{d}, \boldsymbol{\theta})$, $q_2(\mathbf{d})$ dans la même famille que $p(\mathbf{d}|\mathbf{g}, \mathbf{f}, \boldsymbol{\theta})$ et $q_3(\boldsymbol{\theta})$ dans la même famille que $p(\boldsymbol{\theta}|\mathbf{g}, \mathbf{f}, \mathbf{d})$. Donner les expressions de $\langle \ln p(\mathbf{g}, \mathbf{f}, \mathbf{d}, \boldsymbol{\theta}) \rangle_{q_1(\mathbf{f})}$, $\langle \ln p(\mathbf{g}, \mathbf{f}, \mathbf{d}, \boldsymbol{\theta}) \rangle_{q_2(\mathbf{d})}$ et $\langle \ln p(\mathbf{g}, \mathbf{f}, \mathbf{d}, \boldsymbol{\theta}) \rangle_{q_3(\boldsymbol{\theta})}$. Que deviennent alors $q_1(\mathbf{f})$, $q_2(\mathbf{d})$ et $q_3(\boldsymbol{\theta})$ au cours des itérations? Comparer cet algorithme avec l'algorithme EM.
- Dans l'exemple précédent, nous avons modélisé \mathbf{f} par un champs gaussien. Prenons maintenant le modèle gaussien à moyenne modulable de l'exercice précédent:

$$\begin{aligned}
p(\mathbf{g}|\mathbf{f}, \sigma_\epsilon) &= \mathcal{N}(\mathbf{H}\mathbf{f}, \sigma_\epsilon^2 \mathbf{I}), \\
p(f_j|z_j, \lambda) &= \mathcal{N}(z_j, \sigma_f^2) \\
z_j \in \{m_1 = -1, m_2 = 0, m_3 = +1\}, & \quad P(z_j = m_k) = \alpha_k, k = 1, \dots, K = 3, \sum_k \alpha_k = 1 \\
p(\mathbf{f}|\mathbf{z}, \sigma_f^2 \mathbf{I}) &= \mathcal{N}(\mathbf{z}, \sigma_f^2 \mathbf{I}), \\
p(\mathbf{z}) &= \prod_j P(z_j = m_k) = \prod_k \alpha_k^{\sum_j \delta(z_j - m_k)}, \\
\boldsymbol{\theta} &= (\theta_1 = 1/\sigma_\epsilon^2, \theta_2 = 1/\sigma_f^2, \boldsymbol{\alpha} = \{\alpha_k, k = 1, \dots, K\}) \\
p(\theta_1) &= \mathcal{G}(\alpha_{10}, \beta_{10}) \\
p(\theta_2) &= \mathcal{G}(\alpha_{20}, \beta_{20}) \\
p(\boldsymbol{\alpha}) &= \mathcal{D}(\boldsymbol{\alpha}|\boldsymbol{\alpha}_0)
\end{aligned}$$

L'objectif est donc d'estimer \mathbf{f} , \mathbf{z} et $\boldsymbol{\theta}$. Si on écrit la loi *a posteriori* jointe :

$$p(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta}|\mathbf{g}) = \frac{p(\mathbf{g}|\mathbf{f}, \boldsymbol{\theta}) p(\mathbf{f}|\mathbf{z}, \boldsymbol{\theta}) p(\mathbf{z})}{p(\mathbf{g}|\boldsymbol{\theta})} \quad (17)$$

et on cherche à l'approximer par une loi séparable $q(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta}|\mathbf{g}) = q_1(\mathbf{f}) q_2(\mathbf{z}) q_3(\boldsymbol{\theta})$.

1. Donner les expressions de $p(\mathbf{g}, \mathbf{f}, \mathbf{z}, \boldsymbol{\theta})$, $p(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta}|\mathbf{g})$, $p(\mathbf{f}|\mathbf{g}, \mathbf{z}, \boldsymbol{\theta})$, $p(\mathbf{z}|\mathbf{g}, \mathbf{f}, \boldsymbol{\theta})$ et $p(\boldsymbol{\theta}|\mathbf{g}, \mathbf{f}, \mathbf{z})$.
2. Donner les expressions de $KL(q, p)$, $\mathcal{F}(q)$, $\ln p(\mathbf{g}, \mathbf{f}, \boldsymbol{\theta})$, $\langle \ln p(\mathbf{g}, \mathbf{f}, \mathbf{z}, \boldsymbol{\theta}) \rangle_{q_1(\mathbf{f})}$ et $\langle \ln p(\mathbf{g}, \mathbf{f}, \mathbf{z}, \boldsymbol{\theta}) \rangle_{q_2(\mathbf{z})}$ et $\langle \ln p(\mathbf{g}, \mathbf{f}, \boldsymbol{\theta}) \rangle_{q_3(\boldsymbol{\theta})}$.
3. Choisissons les familles $q_1(\mathbf{f}) = \delta(\mathbf{f} - \widehat{\mathbf{f}})$, $q_2(\mathbf{z}) = \delta(\mathbf{z} - \widehat{\mathbf{z}})$ et $q_3(\boldsymbol{\theta}) = \delta(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})$. Donner les expressions de $\langle \ln p(\mathbf{g}, \mathbf{f}, \mathbf{z}, \boldsymbol{\theta}) \rangle_{q_1(\mathbf{f})}$, $\langle \ln p(\mathbf{g}, \mathbf{f}, \mathbf{z}, \boldsymbol{\theta}) \rangle_{q_2(\mathbf{z})}$ et $\langle \ln p(\mathbf{g}, \mathbf{f}, \mathbf{z}, \boldsymbol{\theta}) \rangle_{q_3(\boldsymbol{\theta})}$. Que deviennent alors $\widehat{\mathbf{f}}$, $\widehat{\mathbf{z}}$ et $\widehat{\boldsymbol{\theta}}$ au cours des itérations? (Lien avec MAP jointe)
4. Choisissons $q_1(\mathbf{f})$ dans la même famille que $p(\mathbf{f}|\mathbf{g}, \mathbf{z}, \boldsymbol{\theta})$, $q_2(\mathbf{z})$ dans la même famille que $p(\mathbf{z}|\mathbf{g}, \mathbf{f}, \boldsymbol{\theta})$ et $q_3(\boldsymbol{\theta}) = \delta(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})$. Donner les expressions de $\langle \ln p(\mathbf{g}, \mathbf{f}, \mathbf{z}, \boldsymbol{\theta}) \rangle_{q_1(\mathbf{f})}$, $\langle \ln p(\mathbf{g}, \mathbf{f}, \mathbf{z}, \boldsymbol{\theta}) \rangle_{q_2(\mathbf{z})}$ et $\langle \ln p(\mathbf{g}, \mathbf{f}, \mathbf{z}, \boldsymbol{\theta}) \rangle_{q_3(\boldsymbol{\theta})}$. Que deviennent alors $\widehat{\mathbf{f}}$, $\widehat{\mathbf{z}}$ et $\widehat{\boldsymbol{\theta}}$ au cours des itérations? (Lien avec l'algorithme EM)
5. Choisissons $q_1(\mathbf{f})$ dans la même famille que $p(\mathbf{f}|\mathbf{g}, \mathbf{z}, \boldsymbol{\theta})$, $q_2(\mathbf{z})$ dans la même famille que $p(\mathbf{z}|\mathbf{g}, \mathbf{f}, \boldsymbol{\theta})$ et $q_3(\boldsymbol{\theta})$ dans la même famille que $p(\boldsymbol{\theta}|\mathbf{g}, \mathbf{f}, \mathbf{z})$. Donner les expressions de $\langle \ln p(\mathbf{g}, \mathbf{f}, \mathbf{z}, \boldsymbol{\theta}) \rangle_{q_1(\mathbf{f})}$, $\langle \ln p(\mathbf{g}, \mathbf{f}, \mathbf{z}, \boldsymbol{\theta}) \rangle_{q_2(\mathbf{z})}$ et $\langle \ln p(\mathbf{g}, \mathbf{f}, \mathbf{z}, \boldsymbol{\theta}) \rangle_{q_3(\boldsymbol{\theta})}$. Que deviennent alors $q_1(\mathbf{f})$, $q_2(\mathbf{z})$ et $q_3(\boldsymbol{\theta})$ au cours des itérations? Comparer cet algorithme avec l'algorithme EM.

Exercice numéro 10: Calcul bayésien

Problème 1 : Déconvolution aveugle

Considérons le problème de la déconvolution $g(t) = h(t) * f(t) + \epsilon(t)$ lorsque nous cherchons à estimer $h(t)$ et $f(t)$. Faisons l'hypothèse que $g(t)$, $h(t)$ et $f(t)$ sont causaux et notons par $\mathbf{g} = [g(0), \dots, g(M-1)]'$, $\boldsymbol{\epsilon} = [\epsilon(0), \dots, \epsilon(M-1)]'$, $\mathbf{f} = [f(0), \dots, f(N-1)]'$ et $\mathbf{h} = [h(0), \dots, h(K-1)]'$ avec $M > K$.

1. Montrez qu'on peut écrire: $\mathbf{g} = \mathbf{H}\mathbf{f} + \boldsymbol{\epsilon} = \mathbf{F}\mathbf{h} + \boldsymbol{\epsilon}$ en précisant les deux matrices \mathbf{H} et \mathbf{F} .
2. Supposons $p(\boldsymbol{\epsilon}) = \mathcal{N}(\mathbf{0}, (1/\beta)\mathbf{I})$. Donner les expressions de $p(\mathbf{g}|\mathbf{H}, \mathbf{f}, \beta)$, $p(\mathbf{g}|\mathbf{F}, \mathbf{h}, \beta)$ et des estimateurs au sens du MV de $\hat{\mathbf{f}}_{MV}$ de \mathbf{f} lorsque \mathbf{h} est connue et de $\hat{\mathbf{h}}_{MV}$ de \mathbf{h} lorsque \mathbf{f} est connue.
3. Supposons $p(f(t)|f(t-1)) = \mathcal{N}(f(t-1), (1/\alpha_f))$ et $p(h(t)|h(t-1)) = \mathcal{N}(h(t-1), (1/\alpha_h))$, $\forall t \geq 0$. Montrez alors que

$$\begin{aligned} p(\mathbf{f}|\alpha_f) &= \mathcal{N}(\mathbf{D}\mathbf{f}, (1/\alpha_f)\mathbf{I}) \propto \alpha_f^{N/2} \exp[-\alpha_f \|\mathbf{C}\mathbf{f}\|^2] \\ p(\mathbf{h}|\alpha_h) &= \mathcal{N}(\mathbf{D}\mathbf{h}, (1/\alpha_h)\mathbf{I}) \propto \alpha_h^{K/2} \exp[-\alpha_h \|\mathbf{C}\mathbf{h}\|^2] \end{aligned}$$

en précisant les formes des matrices \mathbf{D} et \mathbf{C} .

4. Donner les expressions de $p(\mathbf{f}|\mathbf{H}, \mathbf{g}, \beta, \alpha_f, \alpha_h)$, $p(\mathbf{h}|\mathbf{F}, \mathbf{g}, \beta, \alpha_f, \alpha_h)$ et des estimateurs bayésiens de $\hat{\mathbf{f}}_{PM}$ de \mathbf{f} lorsque \mathbf{h} est connue et de $\hat{\mathbf{h}}_{PM}$ de \mathbf{h} lorsque \mathbf{f} est connue.
5. Notant $\boldsymbol{\theta} = (\beta, \alpha_f, \alpha_h)$, donner l'expression de $p(\mathbf{f}, \mathbf{h}|\mathbf{g}, \boldsymbol{\theta})$ et $p(\mathbf{f}, \mathbf{h}, \boldsymbol{\theta}|\mathbf{g})$ lorsque $p(\boldsymbol{\theta})$ est fourni.
6. Dans un premier temps, on suppose \mathbf{h} et $\boldsymbol{\theta}$ connu et on cherche à estimer \mathbf{f} . Montrez que $p(\mathbf{f}|\mathbf{g}, \mathbf{h}, \boldsymbol{\theta})$ peut s'écrire sous la forme de

$$p(\mathbf{f}|\mathbf{g}, \mathbf{h}, \boldsymbol{\theta}) = \mathcal{N}(\hat{\mathbf{f}}, \hat{\boldsymbol{\Sigma}}_f) \propto \exp[-J(\mathbf{f})]$$

où

$$\begin{aligned} J(\mathbf{f}) &= \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2 + \lambda_f \|\mathbf{C}\mathbf{f}\|^2 \\ &= (\mathbf{f} - \hat{\mathbf{f}})' \hat{\boldsymbol{\Sigma}}_f^{-1} (\mathbf{f} - \hat{\mathbf{f}}) + c \text{ avec } \hat{\boldsymbol{\Sigma}}_f = (\mathbf{H}'\mathbf{H} + \lambda_f \mathbf{C}'\mathbf{C})^{-1} \text{ et } \hat{\mathbf{f}} = \hat{\boldsymbol{\Sigma}}_f \mathbf{H}'\mathbf{g} \end{aligned}$$

où c ne dépend pas de \mathbf{f} .

7. Maintenant, on suppose \mathbf{f} et $\boldsymbol{\theta}$ connu et on cherche à estimer \mathbf{h} . Montrez que $p(\mathbf{h}|\mathbf{g}, \mathbf{f}, \boldsymbol{\theta})$ peut s'écrire sous la forme de

$$p(\mathbf{h}|\mathbf{g}, \mathbf{f}, \boldsymbol{\theta}) = \mathcal{N}(\hat{\mathbf{h}}, \hat{\boldsymbol{\Sigma}}_h) \propto \exp[-J(\mathbf{h})]$$

où

$$\begin{aligned} J(\mathbf{h}) &= \|\mathbf{g} - \mathbf{F}\mathbf{h}\|^2 + \lambda_h \|\mathbf{C}\mathbf{h}\|^2 \\ &= (\mathbf{h} - \hat{\mathbf{h}})' \hat{\boldsymbol{\Sigma}}_h^{-1} (\mathbf{h} - \hat{\mathbf{h}}) + c \text{ avec } \hat{\boldsymbol{\Sigma}}_h = (\mathbf{F}'\mathbf{F} + \lambda_h \mathbf{C}'\mathbf{C})^{-1} \text{ et } \hat{\mathbf{h}} = \hat{\boldsymbol{\Sigma}}_h \mathbf{F}'\mathbf{g} \end{aligned}$$

où c ne dépend pas de \mathbf{h} .

8. Maintenant, on suppose seulement $\boldsymbol{\theta}$ connu et on cherche à estimer \mathbf{f} et \mathbf{h} . Montrez que $p(\mathbf{f}, \mathbf{h}|\mathbf{g}, \boldsymbol{\theta})$ s'écrit

$$p(\mathbf{f}, \mathbf{h}|\mathbf{g}, \boldsymbol{\theta}) \propto \exp[-J(\mathbf{f}, \mathbf{h})]$$

où

$$\begin{aligned} J(\mathbf{f}, \mathbf{h}) &= \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2 + \lambda_f \|\mathbf{C}\mathbf{f}\|^2 + \lambda_h \|\mathbf{C}\mathbf{h}\|^2 \\ &= (\mathbf{f} - \hat{\mathbf{f}})' \hat{\boldsymbol{\Sigma}}_f^{-1} (\mathbf{f} - \hat{\mathbf{f}}) + c_1 \text{ avec } \hat{\boldsymbol{\Sigma}}_f = (\mathbf{H}'\mathbf{H} + \lambda_f \mathbf{C}'\mathbf{C})^{-1} \text{ et } \hat{\mathbf{f}} = \hat{\boldsymbol{\Sigma}}_f \mathbf{H}'\mathbf{g} \\ &= \|\mathbf{g} - \mathbf{F}\mathbf{h}\|^2 + \lambda_f \|\mathbf{C}\mathbf{f}\|^2 + \lambda_h \|\mathbf{C}\mathbf{h}\|^2 \\ &= (\mathbf{h} - \hat{\mathbf{h}})' \hat{\boldsymbol{\Sigma}}_h^{-1} (\mathbf{h} - \hat{\mathbf{h}}) + c_2 \text{ avec } \hat{\boldsymbol{\Sigma}}_h = (\mathbf{F}'\mathbf{F} + \lambda_h \mathbf{C}'\mathbf{C})^{-1} \text{ et } \hat{\mathbf{h}} = \hat{\boldsymbol{\Sigma}}_h \mathbf{F}'\mathbf{g} \end{aligned}$$

où c_1 ne dépend pas de \mathbf{f} et c_2 ne dépend pas de \mathbf{h} .

On peut alors vouloir calculer une solution MAP jointe:

$$(\hat{\mathbf{f}}, \hat{\mathbf{h}}) = \arg \max_{(\mathbf{f}, \mathbf{h})} \{p(\mathbf{f}, \mathbf{h} | \mathbf{g}, \boldsymbol{\theta})\} = \arg \min_{(\mathbf{f}, \mathbf{h})} \{J(\mathbf{f}, \mathbf{h})\}$$

d'une manière itérative:

$$\begin{aligned} \hat{\mathbf{f}}^{(k+1)} &= \arg \max_{\mathbf{f}} \{p(\mathbf{f}, \mathbf{h}^{(k)} | \mathbf{g}, \boldsymbol{\theta})\} = \arg \min_{\mathbf{f}} \{J(\mathbf{f}, \mathbf{h}^{(k)})\} \\ \hat{\mathbf{h}}^{(k+1)} &= \arg \max_{\mathbf{h}} \{p(\mathbf{f}^{(k)}, \mathbf{h} | \mathbf{g}, \boldsymbol{\theta})\} = \arg \min_{\mathbf{h}} \{J(\mathbf{f}^{(k)}, \mathbf{h})\} \end{aligned}$$

ce qui devient:

$$\begin{aligned} \hat{\mathbf{f}}^{(k+1)} &= (\mathbf{H}'^{(k)} \mathbf{H}^{(k)} + \lambda_f \mathbf{C}' \mathbf{C})^{-1} \mathbf{H}'^{(k)} \mathbf{g} \\ \hat{\mathbf{h}}^{(k+1)} &= (\mathbf{F}'^{(k)} \mathbf{F}^{(k)} + \lambda_h \mathbf{C}' \mathbf{C})^{-1} \mathbf{F}'^{(k)} \mathbf{g} \end{aligned}$$

Discuter sur la pertinence de ce critère et sur la convergence de cet algorithme.

9. Nous allons maintenant nous intéresser au calcul de la moyenne *a posteriori* de \mathbf{f} et de \mathbf{h} :

$$\begin{aligned} \hat{\mathbf{f}} &= \int \int \mathbf{f} p(\mathbf{f}, \mathbf{h} | \mathbf{g}, \boldsymbol{\theta}) d\mathbf{h} d\mathbf{f} \\ \hat{\mathbf{h}} &= \int \int \mathbf{h} p(\mathbf{f}, \mathbf{h} | \mathbf{g}, \boldsymbol{\theta}) d\mathbf{f} d\mathbf{h} \end{aligned}$$

qui nécessite l'intégration par rapport à \mathbf{f} et \mathbf{h} de $p(\mathbf{f}, \mathbf{h} | \mathbf{g}, \boldsymbol{\theta})$. Il n'existe pas, en général, solutions analytiques pour ces intégrales. Pour pouvoir faire des calculs plus simples, on va approximer la loi $p(\mathbf{f}, \mathbf{h} | \mathbf{g}, \boldsymbol{\theta})$ par une loi séparable $q(\mathbf{f}, \mathbf{h}) = q_1(\mathbf{f}) q_2(\mathbf{h})$ qui lui soit le plus proche au sens de distance de Kullback

$$\begin{aligned} K(q : p) &= \int \int q \ln(q/p) = \int \int q_1(\mathbf{f}) q_2(\mathbf{h}) \ln \left(\frac{q_1(\mathbf{f}) q_2(\mathbf{h})}{p(\mathbf{f}, \mathbf{h} | \mathbf{g}, \boldsymbol{\theta})} \right) d\mathbf{f} d\mathbf{h} \\ &= \int \int q_1(\mathbf{f}) q_2(\mathbf{h}) \ln \left(\frac{q_1(\mathbf{f}) q_2(\mathbf{h})}{p(\mathbf{f}, \mathbf{h}, \mathbf{g} | \boldsymbol{\theta})} \right) d\mathbf{f} d\mathbf{h} + \ln p(\mathbf{g} | \boldsymbol{\theta}) \\ &= \int \int q_1(\mathbf{f}) q_2(\mathbf{h}) \ln \left(\frac{q_1(\mathbf{f}) q_2(\mathbf{h})}{p(\mathbf{f} | \alpha_f) p(\mathbf{h} | \alpha_h) p(\mathbf{g} | \mathbf{f}, \mathbf{h}, \boldsymbol{\theta})} \right) d\mathbf{f} d\mathbf{h} + \ln p(\mathbf{g} | \boldsymbol{\theta}) \\ &= - \int \int q_1(\mathbf{f}) q_2(\mathbf{h}) \ln \left(\frac{p(\mathbf{f} | \alpha_f) p(\mathbf{h} | \alpha_h) p(\mathbf{g} | \mathbf{f}, \mathbf{h}, \boldsymbol{\theta})}{q_1(\mathbf{f}) q_2(\mathbf{h})} \right) d\mathbf{f} d\mathbf{h} + \ln p(\mathbf{g} | \boldsymbol{\theta}) \\ &= -H(q_1) - H(q_2) - \langle \ln p(\mathbf{g}, \mathbf{h} | \mathbf{f}, \boldsymbol{\theta}) \rangle_{q_2} - \langle \ln p(\mathbf{f} | \alpha_f) \rangle_{q_1} + \ln p(\mathbf{g} | \boldsymbol{\theta}) \\ &= -H(q_1) - H(q_2) - \langle \ln p(\mathbf{g}, \mathbf{f} | \mathbf{h}, \boldsymbol{\theta}) \rangle_{q_1} - \langle \ln p(\mathbf{h} | \alpha_h) \rangle_{q_2} + \ln p(\mathbf{g} | \boldsymbol{\theta}) \end{aligned}$$

On cherche alors les expressions de ces deux lois, d'une manière itérative, chaque fois optimisant $K(q_1 q_2 : p)$ par rapport à la lois concernée en fixant l'autre:

$$\begin{aligned} q_1^{(k+1)} &= \arg \min_{q_1} \{K(q_1, q_2^{(k)} : p)\} \\ q_2^{(k+1)} &= \arg \min_{q_2} \{K(q_1, q_2^{(k)} : p)\} \end{aligned}$$

Notant que

$$\begin{aligned} K(q_1 q_2 : p) &= \int \int q_1 q_2 \ln(q_1 q_2 / p) = \int \int q_1 q_2 [\ln q_1 + \ln q_2 - \ln p] \\ &= \int q_1 \int q_2 [\ln q_2 + \ln q_1 - \ln p] = \int q_2 \ln q_2 + \langle [\ln q_1 - \ln p] \rangle_{q_1} \\ &= \int q_2 \int q_1 [\ln q_1 + \ln q_2 - \ln p] = \int q_1 \ln q_1 + \langle [\ln q_2 - \ln p] \rangle_{q_2} \end{aligned}$$

montrez

$$\begin{aligned} q_1^{(k+1)} &= \arg \min_{q_1} \{K(q_1, q_2^{(k)} : p)\} \propto \exp \left[\langle [\ln q_2^{(k)} - \ln p] \rangle_{q_2^{(k)}} \right] \\ q_2^{(k+1)} &= \arg \min_{q_2} \{K(q_1, q_2^{(k)} : p)\} \propto \exp \left[\langle [\ln q_1^{(k)} - \ln p] \rangle_{q_1^{(k)}} \right] \end{aligned}$$

10. Nous devons maintenant choisir une famille pour $q_1(\mathbf{f})$ et $q_2(\mathbf{h})$. Deux options se présentent: famille paramétrique et non paramétrique. Un choix très particulier est $q_1(\mathbf{f}) = \delta(\mathbf{f} - \hat{\mathbf{f}})$ et $q_2(\mathbf{h}) = \delta(\mathbf{h} - \hat{\mathbf{h}})$ où $\hat{\mathbf{f}}$ et $\hat{\mathbf{h}}$ sont des paramètres à déterminer. Il s'agit du cas dite "dégénéré". Montrez qu'avec ce choix, on obtient:

$$q_1^{(k+1)}(\mathbf{f}) = \delta(\mathbf{f} - \hat{\mathbf{f}}^{(k+1)}) \text{ avec } \hat{\mathbf{f}}^{(k+1)} = \arg \max_{\mathbf{f}} \left\{ p(\mathbf{f}|\mathbf{g}, \hat{\mathbf{h}}^{(k)}, \boldsymbol{\theta}) \right\}$$

$$q_2^{(k+1)}(\mathbf{h}) = \delta(\mathbf{h} - \hat{\mathbf{h}}^{(k+1)}) \text{ avec } \hat{\mathbf{h}}^{(k+1)} = \arg \max_{\mathbf{h}} \left\{ p(\mathbf{h}|\mathbf{g}, \hat{\mathbf{f}}^{(k)}, \boldsymbol{\theta}) \right\}$$

Commenter cet algorithm.

11. Un deuxième choix est $q_1(\mathbf{f}) = \mathcal{N}(\mathbf{f}|\hat{\mathbf{f}}, \hat{\boldsymbol{\Sigma}}_f)$ et $q_2(\mathbf{h}) = \mathcal{N}(\mathbf{h}|\hat{\mathbf{h}}, \hat{\boldsymbol{\Sigma}}_h)$. Montrez alors qu'avec ce choix, on obtient:

$$q_1^{(k+1)}(\mathbf{f}) = \mathcal{N}(\hat{\mathbf{f}}^{(k+1)}, \hat{\boldsymbol{\Sigma}}_f^{(k+1)})$$

avec $\hat{\mathbf{f}}^{(k+1)} = \langle \mathbf{f} \rangle_{p(\mathbf{f}|\mathbf{g}, \hat{\mathbf{h}}^{(k)}, \boldsymbol{\theta})} = \mathbb{E} \left\{ \mathbf{f}|\mathbf{g}, \hat{\mathbf{h}}^{(k)}, \boldsymbol{\theta} \right\}$ et $\hat{\boldsymbol{\Sigma}}_f^{(k+1)} = \text{cov}[\mathbf{f}|\mathbf{g}, \hat{\mathbf{h}}^{(k)}, \boldsymbol{\theta}]$

$$q_2^{(k+1)}(\mathbf{h}) = \mathcal{N}(\hat{\mathbf{h}}^{(k+1)}, \hat{\boldsymbol{\Sigma}}_h^{(k+1)})$$

avec $\hat{\mathbf{h}}^{(k+1)} = \langle \mathbf{h} \rangle_{p(\mathbf{h}|\mathbf{g}, \hat{\mathbf{f}}^{(k)}, \boldsymbol{\theta})} = \mathbb{E} \left\{ \mathbf{h}|\mathbf{g}, \hat{\mathbf{f}}^{(k)}, \boldsymbol{\theta} \right\}$ et $\hat{\boldsymbol{\Sigma}}_h^{(k+1)} = \text{cov}[\mathbf{h}|\mathbf{g}, \hat{\mathbf{f}}^{(k)}, \boldsymbol{\theta}]$

Donner alors les expressions de $\hat{\boldsymbol{\Sigma}}_f^{(k+1)}$, $\hat{\mathbf{f}}^{(k+1)}$, $\hat{\boldsymbol{\Sigma}}_h^{(k+1)}$ et $\hat{\mathbf{h}}^{(k+1)}$ et commenter cet algorithm.

12. Un troisième choix est $q_1(\mathbf{f}) = p(\mathbf{f}|\mathbf{g}, \hat{\mathbf{h}}, \boldsymbol{\theta})$ et $q_2(\mathbf{h}) = p(\mathbf{h}|\mathbf{g}, \hat{\mathbf{f}}, \boldsymbol{\theta})$:

$$q_1^{(k+1)}(\mathbf{f}) = p(\mathbf{f}|\mathbf{g}, \hat{\mathbf{h}}^{(k)}, \boldsymbol{\theta})$$

$$q_2^{(k+1)}(\mathbf{h}) = p(\mathbf{h}|\mathbf{g}, \hat{\mathbf{f}}^{(k)}, \boldsymbol{\theta})$$

Montrez alors qu'avec ce choix, on obtient:

$$q_1^{(k+1)}(\mathbf{f}) \propto \exp \left[\langle \ln p(\mathbf{f}, \mathbf{h}|\mathbf{g}, \boldsymbol{\theta}) \rangle_{q_2^{(k)}} \right]$$

$$\propto p(\mathbf{f}|\alpha) \exp \left[\langle \ln p(\mathbf{g}|\mathbf{f}, \mathbf{h}, \boldsymbol{\theta}) \rangle_{q_2} \right]$$

$$q_2^{(k+1)}(\mathbf{h}) \propto \exp \left[\langle \ln p(\mathbf{f}, \mathbf{h}|\mathbf{g}, \boldsymbol{\theta}) \rangle_{q_1^{(k)}} \right]$$

$$\propto p(\mathbf{h}|\alpha) \exp \left[\langle \ln p(\mathbf{g}|\mathbf{f}, \mathbf{h}, \boldsymbol{\theta}) \rangle_{q_1} \right]$$

Commenter cet algorithm.

13. Montrer alors que $q_1^{(k+1)}(\mathbf{f})$ et $q_2^{(k+1)}(\mathbf{h})$ sont des gaussiennes: $q_1^{(k+1)}(\mathbf{f}) = \mathcal{N}(\hat{\mathbf{f}}^{(k+1)}, \hat{\boldsymbol{\Sigma}}_f^{(k+1)})$ et $q_2^{(k+1)}(\mathbf{h}) = \mathcal{N}(\hat{\mathbf{h}}^{(k+1)}, \hat{\boldsymbol{\Sigma}}_h^{(k+1)})$. Donner alors les expressions de $(\hat{\boldsymbol{\Sigma}}_f^{(k+1)}, \hat{\mathbf{f}}^{(k+1)})$ et $(\hat{\boldsymbol{\Sigma}}_h^{(k+1)}, \hat{\mathbf{h}}^{(k+1)})$.

Problème 2 : Déconvolution aveugle non supervisée

Dans le problème précédent, on souhaite maintenant estimer aussi les hyperparamètres $\theta = (\beta, \alpha_f, \alpha_h)$

1. En choisissant des lois *a priori* conjuguées pour $p(\beta)$, $p(\alpha_f)$ et $p(\alpha_h)$, donner l'expression de $p(\mathbf{f}, \mathbf{h}, \theta | \mathbf{g})$. Cette loi est-elle séparable en \mathbf{f} , \mathbf{h} et θ ?
2. Pour pouvoir faire des calculs plus simples, on va approximer cette loi par une loi séparable $q(\mathbf{f}, \mathbf{h}, \theta | \mathbf{g}) = q_1(\mathbf{f}) q_2(\mathbf{h}) q_3(\theta)$ qui lui soit le plus proche au sens de distance de Kullback

$$\begin{aligned} K(q : p) &= \int q \ln(q/p) = \int \int q_1(\mathbf{f}) q_2(\mathbf{h}) q_3(\theta) \ln \left(\frac{q_1(\mathbf{f}) q_2(\mathbf{h}) q_3(\theta)}{p(\mathbf{f}, \mathbf{h}, \theta | \mathbf{g})} \right) \\ &= \int \int q_1(\mathbf{f}) q_2(\mathbf{h}) q_3(\theta) \ln \left(\frac{q_1(\mathbf{f}) q_2(\mathbf{h}) q_3(\theta)}{p(\mathbf{f} | \alpha_f) p(\mathbf{h} | \alpha_h) p(\theta) p(\mathbf{g} | \mathbf{f}, \mathbf{h}, \theta)} \right) + \ln p(\mathbf{g} | \theta) \end{aligned}$$

On cherche alors les expressions de ces trois lois, d'une manière itérative, chaque fois optimisant $K(q : p)$ par rapport à la lois concernée en fixant les deux autres. Montrer alors les relations suivantes:

$$\begin{aligned} q_1(\mathbf{f}) &\propto p(\mathbf{f} | \alpha) \exp [\langle \ln p(\mathbf{g} | \mathbf{f}, \mathbf{h}, \theta) \rangle q_2 q_3] \\ q_2(\mathbf{h}) &\propto p(\mathbf{h} | \alpha) \exp [\langle \ln p(\mathbf{g} | \mathbf{f}, \mathbf{h}, \theta) \rangle q_1 q_3] \\ q_3(\theta) &\propto p(\theta) \exp [\langle \ln p(\mathbf{g} | \mathbf{f}, \mathbf{h}, \theta) \rangle q_1 q_2] \end{aligned}$$

Exercice numéro 11: Séparation de sources

Model:

$$\begin{aligned}
 \mathbf{x}(t) &= \mathbf{A}\mathbf{s}(t) + \boldsymbol{\epsilon}(t), \quad t = 1, \dots, T \\
 \mathbf{x}(\mathbf{r}) &= \mathbf{A}\mathbf{s}(\mathbf{r}) + \boldsymbol{\epsilon}(\mathbf{r}), \quad \mathbf{r} \in \mathcal{R} \\
 \mathbf{x}_i &= [\mathbf{A}\mathbf{s}]_i + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, T \\
 \mathbf{X} &= [\mathbf{x}_1 | \dots | \mathbf{x}_m] = [x_1(1) \dots x_1(m)] \quad \mathbf{S} = [s_1, \dots, s_n]^t
 \end{aligned} \tag{18}$$

Priors:

$$\begin{aligned}
 p(\boldsymbol{\epsilon}_i | \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_i}) &= \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_i}) \quad \forall i, \quad \boldsymbol{\epsilon}_i \text{ iid} \\
 p(\mathbf{x}_i | \mathbf{A}, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_i}) &= \mathcal{N}(\mathbf{A}\mathbf{s}_i, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_i}) \\
 p(\mathbf{s}_j | \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_j}) &= \mathcal{N}(\mathbf{0}, \mathbf{I}) \quad \forall j, \quad \mathbf{s}_j \text{ iid}
 \end{aligned} \tag{19}$$

Posteriors:

$$\begin{aligned}
 p(\mathbf{x}_i | \mathbf{A}, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_i}) &= \mathcal{N}(\mathbf{A}\mathbf{s}_i, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_i}) \\
 p(\mathbf{s}_j | \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_j}) &= \mathcal{N}(\mathbf{0}, \mathbf{I})
 \end{aligned} \tag{20}$$