Advanced Signal and Image Processing

Professor: A. Mohammad–Djafari

Exercise number 1: Signal and Image representation and Linear Transforms

Part 1: Signal and Image representation

1. Consider the following signals:

   (a) $f(t) = a\ \sin(\omega t)$
   (b) $f(t) = a\ \cos(\omega t)$
   (c) $f(t) = \sum_{k=1}^{K} [a_k\ \sin(\omega_k t) + b_k \sin(\omega_k t)]$
   (d) $f(t) = \sum_{k=1}^{K} a_k\ \exp\{-j(\omega_k t)\}$
   (e) $f(t) = a\ \exp\left\{-t^2\right\}$
   (f) $f(t) = \sum_{k=1}^{K} a_k\ \exp\left\{-\frac{1}{2}(t - m_k)^2/v_k\right\}$
   (g) $f(t) = a\ \sin(\omega t)/(\omega t)$
   (h) $f(t) = 1, \quad \text{if} \quad |t| < a, \quad 0 \ \text{ elsewhere}$

   For each of these signals, first compute their Fourier Transform $F(\omega)$, then write a Matlab program to plot these signals and their corresponding $|F(\omega)|$.

2. Consider the following signals:

   (a) $f(t) = a\ \exp\{-t/\tau\}, \quad t > 0$
   (b) $f(t) = 0, \quad t \leq 0, \quad 1$

   For each of these signals, compute their Laplace Transform $g(s)$.

3. Consider the following images:

   (a) $f(x, y) = a\ \sin(\omega_x x) + b\ \cos(\omega_y x)$
   (b) $f(x, y) = a\ \exp\left\{-(x^2 + y^2)\right\}$
   (c) $f(x, y) = \sum_{k=1}^{K} a_k\ \exp\left\{-\frac{1}{2}[(x - m_{x_k})^2/v_{x_k} + (y - m_{y_k})^2/v_{y_k}]\right\}$
   (d) $f(x, y) = 1, \quad \text{if} \quad |x| < a \ \ \& \ \ |y| < b, \quad 0 \ \text{ elsewhere}$
   (e) $f(x, y) = 1, \quad \text{if} \quad x^2 + y^2 < a, \quad 0 \ \text{ elsewhere}$

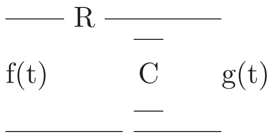
   For each of these images, first compute their Fourier Transform $F(u, v)$, then write a Matlab program to plot these signals and their corresponding $|F(u, v)|$.

4. Consider the following images:

   (a) $f(x, y) = \delta(x - a)\delta(y - b)$
   (b) $f(x, y) = a\ \exp\left\{-(x^2 + y^2)\right\}$
   (c) $f(x, y) = \sum_{k=1}^{K} a_k\ \exp\left\{-\frac{1}{2}[(x - m_{x_k})^2/v_{x_k} + (y - m_{y_k})^2/v_{y_k}]\right\}$
   (d) $f(x, y) = 1, x^2 + y^2 < a, \quad 0 \ \text{ elsewhere}$

   For each of these images, first compute their Radon Transform $g(r, \phi)$, then write a Matlab program to plot these images and their corresponding Radon Transform $g(r, \phi)$.

Part 2: Linear systems: Convolution
Let consider the following system:



1. Write the expression of the transfer function $H(\omega) = \frac{G(\omega)}{F(\omega)}$
2. Write the expression of the impulse response $h(t)$
3. Write the expression of the relation linking the output $g(t)$ to the input $f(t)$ and the impulse response $h(t)$
4. Write the expression of the relation linking the Fourier transforms $G(\omega)$, $F(\omega)$ and $H(\omega)$
5. Write the expression of the relation linking the Laplace transforms $G(s)$, $F(s)$ and $H(s)$
6. Give the expression of the output when the input is $f(t) = \delta(t)$
7. Give the expression of the output when the input is a step function $f(t) = u(t) = \begin{cases} 0 & \forall t < 0, \\ 1 & \forall t \geq 0 \end{cases}$
8. Give the expression of the output when the input is $f(t) = a \sin(\omega_0 t)$
9. Give the expression of the output when the input is $f(t) = \sum_k f_k \, \sin(\omega_k t)$
10. Give the expression of the output when the input ist $f(t) = \sum_j f_j \, \delta(t - t_j)$
11. Suppose $h(t) = \sum_{k=-q}^{p} h_k \delta(t - t_k)$ and the input $f(t) = \sum_{j=0}^{n-1} f_j \delta(t - t_j)$. Note $t_k = k\, T$ and $t_j = j\, T$ with $T = 1$. Then compute the output $g(t_i)$ for $t_i = i\, T$
12. Show that the relation between $\boldsymbol{f} = [f_0, \cdots, f_{n-1}]'$, $\boldsymbol{h} = [h_{-q}, \cdots, h_0, \cdots, h_p]'$ and $\boldsymbol{g} = [g_0, \cdots, g_{m-1}]'$ can be written as $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f}$ or as $\boldsymbol{g} = \boldsymbol{F}\boldsymbol{h}$. give the expressions and the structures of the matrices $\boldsymbol{H}$ and $\boldsymbol{F}$.
13. What do you remark on the structure of these two matrices?
14. Simplify these matrices when $q = 0$?
15. What can we do to transform these matrices to circulantes matrices?
16. Write a Matlab programs which compute $\boldsymbol{g}$ when $\boldsymbol{f}$ and $\boldsymbol{h}$ are given. Let name this program `g=direct(h,f,method)` where method will indicate different methods to use to do the computation. Test it with creating different inputs and different impulse responses and compute the outputs.

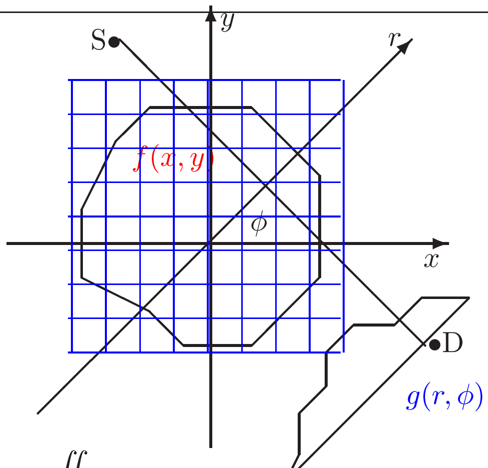
Part 3: Imaging systems: 2D convolution

Consider an imaging system such as a camera which is not well focalised. Let assume that during an experiment, we could measure its Point Spread Function (PSF) $h(x, y)$ which is spread out over a few tens pixels. Assume also that the link between the the photo taken by the camera $g(x, y)$ and the original scene $f(x, y)$ can be modelled as a 2D convolution: $g = h * f$.

1. Write the expression of integral equation linking $g(x, y)$, $f(x, y)$ and $h(x, y)$.
2. Write the expression of the Optical Transfert Function (OTF) $H(\omega_x, \omega_y) = \frac{G(\omega_x,\omega_y)}{F(\omega_x,\omega_y)}$
3. Show that, if we arrange in the vectors $\boldsymbol{f}$, $\boldsymbol{g}$ and $\boldsymbol{h}$ all the pixels of the original image $f(x, y)$, the observed image $g(x, y)$ and the PSF $h(x, y)$ by rasterising column by columns, then we can write the relation between them either as $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f}$ or as $\boldsymbol{g} = \boldsymbol{F}\boldsymbol{h}$. Show then the expression and the structures of the matrices $\boldsymbol{H}$ and $\boldsymbol{F}$.
4. What can we reamrk on the structures of these matrices?
5. What becomes these matrices when the PSF is symmetric?
6. What becomes these matrices when the PSF is separable?
7. What can we do to transform these matrices circulant-bloc-circulant?
8. Write a Matlab programs which computes the image $\boldsymbol{g}$ when the images $\boldsymbol{f}$ and $\boldsymbol{h}$ are given. Let name this program:
   `g=direct(h,f,method)`.
   Test it with creating different inputs and different PSF and compute the outputs.

Part 4: Imaging systems: Computed Tomography (CT)

In X ray tomography, a simple model which links the relative intensity of the rays measured on a detector g to the distribution of absorption coefficient inside the object f by a line integral equation joining the position of the source to the position of the detector. In the following, we will consider the 2D Case, where we can caracterize this line by two variables r and ϕ, and so, we can relate the observed projections $g(r, \phi)$ to the object $f(x, y)$ by the following relations:



$$g(r, \phi) = \iint_D f(x, y)\, \delta(r - x \cos\phi - y \sin\phi)\, \mathrm{d}x\, \mathrm{d}y$$

Fig. 1: X ray Computed Tomography and Radon Transform.

Analytical Image reconstruction methods

Consider the problem of image reconstruction in 2D X ray cOMPUTED tOMOGRAPHY (CT) where the relation between the object $f(x,y)$ and the projections $g(r,\phi)$ is modeled by the Radon Transfrom (RT):

$$g(r,\phi) = \int_{L_{(r,\phi)}} f(x,y)\,\mathrm{d}l = \iint f(x,y)\delta(r - x\cos\phi - y\sin\phi)\,\mathrm{d}x\,\mathrm{d}y,$$

where $L_{(r,\phi)}$ is a line making the angle ϕ with the axis x and positionned at a distance from the origin r.

Using the following operations:

$$\begin{array}{ll}
\text{Derivation } \mathcal{D}: & \overline{g}(r,\phi) = \dfrac{\partial g(r,\phi)}{\partial r} \\
\text{Hilbert Transform } \mathcal{H}: & g_1(r',\phi) = \dfrac{1}{\pi}\displaystyle\int_0^\infty \dfrac{\overline{g}(r,\phi)}{(r-r')}\,\mathrm{d}r \\
\text{Backprojection } \mathcal{B}: & f(x,y) = \dfrac{1}{2\pi}\displaystyle\int_0^\pi g_1(x\cos\phi + y\sin\phi,\phi)\,\mathrm{d}\phi \\
\text{1D Fourier Transform } \mathcal{F}_1: & G(\Omega,\phi) = \displaystyle\int g(r,\phi)\exp\{-j\Omega\, r\}\,\mathrm{d}r \\
\text{2D Fourier Transform } \mathcal{F}_2: & F(u,v) = \displaystyle\iint f(x,y)\exp\{-j(ux+vy)\}\,\mathrm{d}x\,\mathrm{d}y
\end{array}$$

1. Show that we $F(u,v) = G(\Omega,\phi)$ for $u = \Omega\cos\phi, v = \Omega\sin\phi$.

2. Show that, if we define:

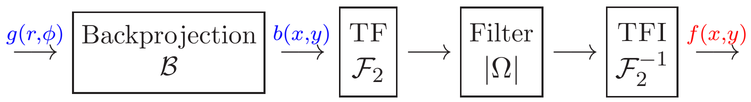
$$b(x,y) = \int_0^\pi g(r,\phi)\,\mathrm{d}\phi = \int_0^\pi g(x\cos\phi + y\sin\phi,\phi)\,\mathrm{d}\phi$$

   we have a relation between $f(x,y)$ and $b(x,y)$ in the following form:

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

   Show then that:

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



   where $|\Omega|^2 = u^2 + v^2$.

3. Show that, if the object f has a property of rotational symmetry, *i.e.*; $f(x,y) = f(\rho)$ with $\rho^2 = x^2 + y^2$, then, it can be reconstructed only from one projection.

4. Show that, if $f(x,y) = f_1(x)\,f_2(y)$ then it is possible to reconstruct it (up to a constant value) only from two projections $\phi = 0$ and $\phi = 90$.

Advanced Signal and Image Processing

Professor: A. Mohammad–Djafari

Exercise number 2: Modeling and parameter estimation

Part 1: Parametric modeling and estimation
Case of sinusoids:
We observed a signal $f(t)$ which we consider to be periodic and model it by $f(t) = a\sin(2\pi t/T + \phi)$. We have samples of this signal every 1 hour $\Delta = 1h$ and have observed it during 4 days ($4 \times 24 = 96$ hours).

1. Assume we know $T = 24$ h. Propose methods to estimate a and ϕ. Test your method using Matlab programming `[a,phi]=estimate_1Sinus_a(f,T);`
2. Now, propose methods to estimate T, a and ϕ. Test your method using Matlab programming `[T,a,phi]=estimate_1sinus(f);`
3. Generalise these two programs for the case where Δ is any value
   `[a,phi]=estimate_1Sinus_a(f,T,Delta);` and
   `[T,a,phi]=estimate_1sinus(f,Delta);`

Now consider the more general model

$$f(t) = \sum_{k=1}^{K} [a_k \cos(k\pi t/T) + b_k \sin(k\pi t/T)]$$

1. Assume we know $T = 24$ h and we know K. Propose methods to estimate a_k and b_k. Test your method using Matlab programming `[a,b]=estimate_KSinus_a(f,K,T);`
2. Assume K is given, propose methods to estimate T, a_k and b_k. Test your method using Matlab programming `[T,a,b]=estimate_Ksinus(f,K);`
3. Now, propose a method to estimate K too.

Case of Gaussian shape signals:
We observed a signal $f(t)$ which we model it by $f(t) = a\mathcal{N}(m, v)$. We have samples of this signal every 1 hour $\Delta = 1h$ and have observed it during 4 days ($4 \times 24 = 96$ hours).

1. Assume we know $m = 48, v = 24$. Propose methods to estimate a. Test your method using Matlab programming `a=estimate_1Gauss_a(f,m,v);`
2. Now, propose methods to estimate m, v and a. Test your method using Matlab programming `[m,v,a]=estimate_1Gauss(f);`

Now consider the more general model

$$f(t) = \sum_{k=1}^{K} a_k \mathcal{N}(m_k, v_k)$$

1. Assume we know K, m_k and v_k. Propose methods to estimate a_k. Test your method using Matlab programming `a=estimate_KGauss_a(f,m,v);`

2. Assume K is given, propose methods to estimate m_k v_k and a_k. Test your method using Matlab programming `[m,v,a]=estimate_KGauss(f,K);`

3. Now, propose a method to estimate K too.

Part 2 Probabilistic Parametric modeling and estimation
Case of MA models:

Consider a signal $f(t)$ which is modeled by the following MA model

$$f(t) = \sum_{k=1}^{K} b_k \epsilon(t - k\Delta)$$

where $\epsilon(t) \sim \mathcal{N}(0, v), \forall t$ and $\delta = 1$.

1. Assume we know K and v. Propose methods to estimate b_k. Test your method using Matlab programming `b=estimate_AR_a(f,K,v);`

2. Assume K is given, propose methods to estimate v and b_k. Test your method using Matlab programming `[b,v]=estimate_AR(f,K);`

3. Now, propose a method to estimate K too.

Case of AR models:

Consider a signal $f(t)$ which is modeled by the following AR model

$$f(t) = \sum_{k=1}^{K} a_k f(t - k\Delta) + \epsilon(t)$$

where $\epsilon(t) \sim \mathcal{N}(0, v), \forall t$ and $\delta = 1$.

1. Assume we know K and v. Propose methods to estimate a_k. Test your method using Matlab programming `a=estimate_AR_a(f,K,v);`

2. Assume K is given, propose methods to estimate v and a_k. Test your method using Matlab programming `[a,v]=estimate_AR(f,K);`

3. Now, propose a method to estimate K too.

Advanced Signal and Image Processing

Professor: A. Mohammad–Djafari

Exercise number 3: Linear models, Deconvolution and Image restoration

Part 1: Identification and Inversion

1D-case: Consider the problem of deconvolution where the measured signal $g(t)$ is related to the input signal $f(t)$ and the impulse response $h(t)$ by $g(t) = h(t) * f(t) + \epsilon(t)$ and where we are looking to estimate $h(t)$ from the knowledge of the input $f(t)$ and output $g(t)$ and to estimate $f(t)$ from the knowledge of the impulse response $h(t)$ and output $g(t)$.

1. Given $f(t)$ and $g(t)$, describe different methods for estimating $h(t)$.
2. Write a Matlab program which can compute $\boldsymbol{h}$ given $\boldsymbol{f}$ and $\boldsymbol{g}$. Let name it: `h=identification(g,f,method)`. Test it by creating different inputs $\boldsymbol{f}$ and outputs $\boldsymbol{g}$. Think also about the noise. Once test your programs without noise and then add some noise on the output $\boldsymbol{g}$ and test them again.
3. Given $g(t)$ and $h(t)$, describe different methods for estimating $f(t)$.
4. Write a Matlab program which can compute $\boldsymbol{f}$ given $\boldsymbol{g}$ and $\boldsymbol{h}$. Let name it: `f=inversion(g,h,method)`. Test it by creating different inputs $\boldsymbol{f}$ and outputs $\boldsymbol{g}$. Think also about the noise. Once test your programs without noise and then add some noise on the output $\boldsymbol{g}$ and test them again.
5. Bring back your experiences and comments.

2D-case: Consider the problem of Image Restoration (Deconvolution) where the observed image $g(x, y)$ is related to the original image $f(x, y)$ and the impulse response (Point Spread Function) $h(x, y)$ by 2D convolution $g(x, y) = h(x, y) * f(x, y) + \epsilon(x, y)$ and where we are looking to estimate $h(x, y)$ from the knowledge of the input $f(x, y)$ and output $g(x, y)$ and to estimate $f(x, y)$ from the knowledge of the impulse response $h(x, y)$ and output $g(x, y)$.

1. Given $f(x, y)$ and $g(x, y)$, describe different methods to estimate $h(x, y)$?

   Write a Matlab program which can compute the PSF $\boldsymbol{h}$ given the input image $\boldsymbol{f}$ and the output image $\boldsymbol{g}$. Let name it:
   `h=identification(g,f,method)`.
   Test it by creating different inputs $\boldsymbol{f}$ and outputs $\boldsymbol{g}$. Think also about the noise. Once test your programs without noise and then add some noise on the output $\boldsymbol{g}$ and test them again.
2. Given $g(x, y)$ and $h(x, y)$, describe different methods for estimating $f(x, y)$.

   Write a Matlab program which can compute $\boldsymbol{f}$ given $\boldsymbol{g}$ and $\boldsymbol{h}$. Let name it:
   `f=inversion(g,h,method)`.
   Test it by creating different inputs $\boldsymbol{f}$ and outputs $\boldsymbol{g}$. Think also about the noise. Once test your programs without noise and then add some noise on the output $\boldsymbol{g}$ and test them again.
3. Bring back your experiences and comments.

Part 2: Least Squares, Generalized inversion and Regularisation

In a measurement system, we have established the following relation: $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}$ where
$\boldsymbol{g}$ is a vector containig the projections (measured data or observations) $\{g_m,\ m = 1\cdots, M\}$,
$\boldsymbol{\epsilon}$ is a vector representing the errors (measurement and modeling) $\{\epsilon_m,\ m = 1\cdots, M\}$,
$\boldsymbol{f}$ is a vector representing the pixels of the image $\{f_n,\ n = 1\cdots, N\}$, and
$\boldsymbol{H}$ is a matrix with the elements $\{a_{mn}\}$ depending on the geometry of the measurement system and assumed to be known.

1. Suppose first $M = N$ and that the matrix $\boldsymbol{H}$ be invertible. Why the solution $\widehat{\boldsymbol{f}}_0 = \boldsymbol{H}^{-1}\boldsymbol{g}$ is not, in general, a satisfactory solution?

   What relation exists between $\frac{\|\delta\widehat{\boldsymbol{f}}_0\|}{\|\widehat{\boldsymbol{f}}_0\|}$ and $\frac{\|\delta\boldsymbol{g}\|}{\|\boldsymbol{g}\|}$?

2. Let come back to the general case $M \neq N$. Show then that the Least Squares (LS) solution, *i.e.* $\widehat{\boldsymbol{f}}_1$ which minimises

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

   is also a solution of equation $\boldsymbol{H}'\boldsymbol{H}\boldsymbol{f} = \boldsymbol{H}'\boldsymbol{g}$ and if $\boldsymbol{H}'\boldsymbol{H}$ is invertible, then we have

   $$\widehat{\boldsymbol{f}}_1 = [\boldsymbol{H}'\boldsymbol{H}]^{-1}\boldsymbol{H}'\boldsymbol{g}$$

   What is the relation between $\frac{\|\delta\widehat{\boldsymbol{f}}_1\|}{\|\widehat{\boldsymbol{f}}_1\|}$ and $\frac{\|\delta\boldsymbol{g}\|}{\|\boldsymbol{g}\|}$?

3. What is the relation between the covarience of $\widehat{\boldsymbol{f}}_1$ and covarience of $\boldsymbol{g}$?

4. Consider now the case $M < N$. Evidently, $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f}$ has infinite number of solutions. The minimum norm solution is:

   $$\widehat{\boldsymbol{f}} = \arg \min_{\boldsymbol{H}\boldsymbol{f}=\boldsymbol{g}} \left\{\|\boldsymbol{f}\|^2\right\}$$

   Show that this solution is obtained via:

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

   which gives:

   $$\widehat{\boldsymbol{f}}_2 = \boldsymbol{H}^t(\boldsymbol{H}\boldsymbol{H}^t)^{-1}\boldsymbol{g}$$

   if $\boldsymbol{H}\boldsymbol{H}^t$ is invertible.

5. Show that with this solution we have: $\widehat{\boldsymbol{g}} = \boldsymbol{H}\widehat{\boldsymbol{f}}_2 = \boldsymbol{g}$.

6. What is the relation between the covarience of $\widehat{\boldsymbol{f}}_2$ and covarience of $\boldsymbol{g}$?

7. Let come back to the general case $M \neq N$ and define

   $$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \{J(\boldsymbol{f})\} \quad \text{with} \quad J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + \lambda\|\boldsymbol{f}\|^2$$

   Show that for any $\lambda > 0$, this solution exists and is unique and is obtained by:

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

8. What relation exists between $\widehat{\boldsymbol{g}} = \boldsymbol{H}\widehat{\boldsymbol{f}}$ and $\boldsymbol{g}$?

9. What is the relation between the covarience of $\widehat{\boldsymbol{f}}$ and covarience of $\boldsymbol{g}$?

10. Another regularised solution $\widehat{\boldsymbol{f}}_2$ to this problem is to minimize a criterion such as:

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

where $\boldsymbol{D}$ is a matrix approximating the operator of derivation.

Show that this solution is given by:

$$\widehat{\boldsymbol{f}}_2 = \arg\min_{\boldsymbol{f}} \{J_2(\boldsymbol{f})\} = \left[\boldsymbol{H}'\boldsymbol{H} + \lambda\boldsymbol{D}'\boldsymbol{D}\right]^{-1}\boldsymbol{H}'\boldsymbol{g}$$

Why this solution is prefered to $\widehat{\boldsymbol{f}}_0$ and to $\widehat{\boldsymbol{f}}_1$ and $\widehat{\boldsymbol{f}}_2$?

11. Suppose that $\boldsymbol{H}$ and $\boldsymbol{D}$ be circulent matrices and symmetric. Then, show that the regularised solution $\widehat{\boldsymbol{f}}_2$ can be written using the DFT by:

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

where

- $H(\omega)$ is the DFT of the first ligne of the matrix $\boldsymbol{H}$,
- $D(\omega)$ is the DFT of the first ligne of the matrix $\boldsymbol{D}$
- $F(\omega)$ is the DFT of the solution vector $\widehat{\boldsymbol{f}}_2$, et
- $G(\omega)$is the DFT of the data measurement vector $\boldsymbol{g}$.

12. Comment the expressions of $\widehat{\boldsymbol{f}}_2$ in the question 3. and $F(\omega)$ in the question 4. when $\lambda = 0$ and when $\lambda \longrightarrow \infty$.

Part 3: Generalized inversion

Consider the problem $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f}$. We are looking the solution $\widehat{\boldsymbol{f}}$ for this inverse problem in such a way that $\widehat{\boldsymbol{f}} = \boldsymbol{M}\boldsymbol{g}$, i.e.; a linear function of the data $\boldsymbol{g}$. We are then looking for the matrix $\boldsymbol{M}$.

1. Suppose first that the solution $\boldsymbol{f}^*$ exists, i.e. $\boldsymbol{H}\boldsymbol{f}^* = \boldsymbol{g}$. Then,

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

The matrix $\boldsymbol{R} - \boldsymbol{M}\boldsymbol{H}$ measures the *resolution power in the space of the solutions* of the inverse operator $\boldsymbol{M}$. The ideal case is $\boldsymbol{R} = \boldsymbol{I}$, i.e.; $\boldsymbol{M} = \boldsymbol{H}^{-1}$, but this almost never possible, and when possible, probably not useful due to the ill-condition nature of $\boldsymbol{H}$. So, let look for $\boldsymbol{M}$ such that:

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

be minimal. Show then that the solution is:

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

2. A second argument is to search for $\boldsymbol{M}$ such that $\widehat{\boldsymbol{g}} = \boldsymbol{H}\boldsymbol{f}^* = \boldsymbol{H}\boldsymbol{M}\boldsymbol{g} = \boldsymbol{N}\boldsymbol{g}$ be as close as possible to $\boldsymbol{g}$. The matrix $\boldsymbol{N} - \boldsymbol{H}\boldsymbol{M}$ measures the *resolution power in the space of the observations* of the operator $\boldsymbol{M}$. The ideal case is $\boldsymbol{N} = \boldsymbol{I}$, i.e.; $\boldsymbol{M} = \boldsymbol{H}^{-1}$, which is again either impossible or not desirable. So, let look for $\boldsymbol{M}$ such that

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

be minimal. Show then the solution is:

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

3. A third argument is based on the fact that: $\text{cov}[\widehat{\boldsymbol{f}}] = \text{cov}[\boldsymbol{M}\boldsymbol{g}] = \boldsymbol{M}\text{cov}[\boldsymbol{g}]\boldsymbol{M}^t$ and if $\text{cov}[\boldsymbol{g}] = \boldsymbol{I}$ we have $\text{cov}[\widehat{\boldsymbol{f}}] = \boldsymbol{M}\boldsymbol{M}^t$. The ideal case for $\widehat{\boldsymbol{f}}$ is that this covarience be close to $\boldsymbol{I}$. We can then would like to define:

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

which can also be used as a constraint for $\boldsymbol{M}$. Write the expression of $\frac{\partial J_3}{\partial \boldsymbol{M}}$.

4. Now, let define $J(\boldsymbol{M}) = \alpha_1 J_1(\boldsymbol{M}) + \alpha_2 J_2(\boldsymbol{M}) + \alpha_3 J_2(\boldsymbol{M})$. Write the expression of $\frac{\partial J}{\partial \boldsymbol{M}}$ and find the matrix $\boldsymbol{M}$ minimizes $J(\boldsymbol{M})$ pour différentes combinaisons de $(\alpha_1, \alpha_2, \alpha_3)$. For each case, give also the expressions of $\boldsymbol{R}$, $\boldsymbol{N}$ and $\boldsymbol{U}$.

Check then the content of the following table:

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

Part 4: Algebraic Reconstruction methods in Computed Tomography

In this part, we consider the problem of Computed Tomography in a simplified and discretized configuration. We consider the problem of image reconstruction from only two projections: horizontal $\phi = 0$ and vertical $\phi = \pi/2$ (See Fig. 3 in Exercise 1).

1. First we assume that the value of f inside a pixel is almost a constant value and that the pixels sizes is juste a unity in both directions. With these assumptions write down the relation between $\boldsymbol{f}$ and $\boldsymbol{g}$ and show that it can be written as $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f}$. What represent the elements of the matrice $\boldsymbol{H}$?

2. Consider now the case where we have only two projections: horizontal and vertical. What becomes the elements of the matrix $\boldsymbol{H}$?

3. Consider now a very small image of (4×4) pixels: $\boldsymbol{f} = [f_1, \cdots, f_{16}]'$ and two projections horizontal and vertical:

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}	

$$\boldsymbol{g}_1 = [g_1, \cdots, g_4]' = [g_{11}, \cdots, g_{14}]',$$
$$\boldsymbol{g}_2 = [g_5, \cdots, g_8]' = [g_{21}, \cdots, g_{24}]'$$

Construct then the two matrices $\boldsymbol{H}_1$ and Hb_2 such that:

$$\boldsymbol{g}_1 = \boldsymbol{H}_1\boldsymbol{f}, \quad \boldsymbol{g}_2 = \boldsymbol{H}_2\boldsymbol{f}, \quad \boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} = \left[\frac{\boldsymbol{H}_1}{\boldsymbol{H}_2}\right]\boldsymbol{f}$$

4. Consider now the following image

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

Compute then its projection. Write a Matlab program to do this computation.
Let name it: `g=direct(f)`.

5. By comparing the analytical relations and discretized algebraic relations, show that the Backprojection operation in continuous case corresponds to the transposition of the matrice $\boldsymbol{H}$ in the discrete case, i.e., the solution $\widehat{\boldsymbol{f}} = \boldsymbol{H}'\boldsymbol{g}$ corresponds to the Backprojected sinograms.

Write a Matlab program to do this computation.
Let name it: `f=transp(g)`.

6. Can you find an expression giving all these solutions ?

Give then the expressions of the matrix: $\boldsymbol{H}' = \left[\ \boldsymbol{H}'_1 \mid \boldsymbol{H}'_2\ \right]$

7. Show that $\widehat{\boldsymbol{f}} = \boldsymbol{H}'\boldsymbol{g} = \boldsymbol{H}'_1\boldsymbol{g}_1 + \boldsymbol{H}'_2\boldsymbol{g}_2$ is the addition of two images. Compute these images. What can we remark ?

8. Consider now the inverse problem: Given the two projections, estimate the image. Why this problem is ill-posed ?

9. Show that this inverse problem has an infinite number of possible solutions and show a few examples.

10. As a tool to try to define the Generalized Inverse solution, compute the two symmetric matrices: $\boldsymbol{H}'\boldsymbol{H}$ and $\boldsymbol{H}\boldsymbol{H}'$ and show that:

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

and

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

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

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

11. Compute the singular values of the matrices $\boldsymbol{H}\boldsymbol{H}'$ and $\boldsymbol{H}'\boldsymbol{H}$ and show that:

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

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

12. Do these matrices are invertible?

13. How then can we define a solution to this problem?

14. Remember that a Least Square solution is defined as:

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

Show that, if the matrix $\boldsymbol{H}^t\boldsymbol{H}$ was invertible, we could write: $\widehat{\boldsymbol{f}} = (\boldsymbol{H}^t\boldsymbol{H})^{-1}\boldsymbol{H}^t\boldsymbol{g}$.

15. In the same way, the minimum norme solution to the equation $\boldsymbol{H}\boldsymbol{f} = \boldsymbol{g}$ is:

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{H}\boldsymbol{f}=\boldsymbol{g}} \left\{\|\boldsymbol{f}\|^2\right\}$$

Show that if the matrix $\boldsymbol{H}\boldsymbol{H}^t$ wa invertible, we could write: $\widehat{\boldsymbol{f}} = \boldsymbol{H}^t(\boldsymbol{H}\boldsymbol{H}^t)^{-1}\boldsymbol{g}$.

16. In this example, we note that matrices $\boldsymbol{H}\boldsymbol{H}^t$ and $\boldsymbol{H}^t\boldsymbol{H}$ are not invertible. However, if we approximate them by their diagonal matrices, then, we have:
`HtH=H’*H; fh=diag(1./diag(HtH))*H’*g; reshape(fh,4,4)`

$$\widehat{\boldsymbol{f}} = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 2 & 2 & 1 \\ 1 & 2 & 2 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

```
HHt=H*H'; fh=H'*diag(1./diag(HHt))*g; reshape(fh,4,4)
```

$$\widehat{\boldsymbol{f}} = \begin{bmatrix} 0 & .5 & .5 & 0 \\ .5 & 1 & 1 & .5 \\ .5 & 1 & 1 & .5 \\ 0 & .5 & .5 & 0 \end{bmatrix}$$

Comment these results.

17. We saw however that we can compute the GI solution which is the mimimum norme solution of $\boldsymbol{H}^t\boldsymbol{H}\boldsymbol{f} = \boldsymbol{H}^t\boldsymbol{g}$ by using the Singular Value Decomposition (SVD):

$$\widehat{\boldsymbol{f}} = \sum_{k=1}^{k} \frac{< \boldsymbol{g}, \boldsymbol{u}_k >}{\lambda_k} \boldsymbol{v}_k$$

where $\boldsymbol{u}_k$ and $\boldsymbol{v}_k$ are, respectively, the eigenvectors of $\boldsymbol{H}\boldsymbol{H}^t$ and $\boldsymbol{H}^t\boldsymbol{H}$ and λ_k are their corresponding eigenvalues:

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

In this example, $K = 7$ and the GI solution can be computed as:

```
fh=svdpca(H,g,.1,7); reshape(fh,4,4)
```

$$\widehat{\boldsymbol{f}} = \begin{bmatrix} -0.2500 & 0.2500 & 0.2500 & -0.2500 \\ 0.2500 & 0.7500 & 0.7500 & 0.2500 \\ 0.2500 & 0.7500 & 0.7500 & 0.2500 \\ -0.2500 & 0.2500 & 0.2500 & -0.2500 \end{bmatrix}$$

Comment this result. .

18. Show that the Kernel of the linear transformation $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f}$, *i.e.* $\{\boldsymbol{f}|\boldsymbol{H}\boldsymbol{f} = 0\}$ is

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

with $\boldsymbol{z}$ any arbitrary vector. This expression can be used to obtain all the possible solutions of the problem.

19. Show that the iterative algorithem:

```
for k=1:100;
fh=fh+.1*H'*(g-H*fh(:));
end;
reshape(fh,4,4)
```

gives the follwing results:

$$\widehat{\boldsymbol{f}} = \begin{bmatrix} -0.2500 & 0.2500 & 0.2500 & -0.2500 \\ 0.2500 & 0.7500 & 0.7500 & 0.2500 \\ 0.2500 & 0.7500 & 0.7500 & 0.2500 \\ -0.2500 & 0.2500 & 0.2500 & -0.2500 \end{bmatrix}$$

Comment this result. Note that, you could also use the programs already written `g=direct(f)` and `f=transp(g)` to do the same computation as follows:

```
for k=1:100;
fh=fh+.1*trans(g-direct(fh));
end;
reshape(fh,4,4)
```

What are the advantages of this writing ?

20. We remark that, in this problem, the data are so poor that, even imposing the minimum norm is not enough to reduce the space of the possible solutions. In many imaging system, another prior information which is very important is the positivity of the solution. Imposing the positivity can then bring a new constraint which can be helpfull to find a unique solution. A simple method is just impose the positivity at each iteration of the previous algorithem:

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

$$\widehat{\boldsymbol{f}} = \begin{bmatrix} 0 & 0.0000 & 0.0000 & 0 \\ 0.0000 & 1.0000 & 1.0000 & 0.0000 \\ 0.0000 & 1.0000 & 1.0000 & 0.0000 \\ 0 & 0.0000 & 0.0000 & 0 \end{bmatrix}$$

Comment this result. .

21. Propose other algorithems to impose this positivity.

Advanced Signal and Image Processing

Professor: A. Mohammad–Djafari

Exercise number 4: Linear Algbera, Generalized inversion, Facor Analysis and Sources separation

Parta A: Linear Algebra, Generalized inversion, Minimum Norme Least Squares inversion

1. We know the sum g_1 and the difference g_2 of two numbers f_1 and f_2. Find those numbers.
   Help: Write the problem in the form of $\boldsymbol{g} = \boldsymbol{A}\boldsymbol{f}$ with $\boldsymbol{g} = [g_1, g_2]'$; $\boldsymbol{f} = [f_1, f_2]'$ and $\boldsymbol{A} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$. Check then if this matrix is invertible and inverse it and find the solution $\widehat{\boldsymbol{f}} = \boldsymbol{A}^{-1}\boldsymbol{g}$. For numerical application, $g_1 = 10, g_2 = 6$.

2. We know that $\boldsymbol{g}$ is a linear function of $\boldsymbol{f}$, *i.e.*$\boldsymbol{g} = \boldsymbol{A}\boldsymbol{f}$ with $\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$.

   - Show that this equation can also be written in the form $\boldsymbol{g} = \boldsymbol{F}\boldsymbol{a}$ with $\boldsymbol{a} = [a_{11}, a_{21}, a_{12}, a_{22}]'$. Give the expression of $\boldsymbol{F}$.
   - If $\boldsymbol{f}$ and $\boldsymbol{g}$ are given, can we determine $\boldsymbol{a}$ or equivalently $\boldsymbol{A}$? Why ?
   - If we fixe the values of $a_{11} = 1$ and $a_{22} = 1$, can we determine a_{12} and a_{21} ? Give their expressions as a function of $\boldsymbol{g}$ and $\boldsymbol{f}$.
   - If between all the possible solutions, we decide to choose the one with minimum norme $\|\boldsymbol{A}\|^2 = \|\boldsymbol{a}\|^2$, can you give the expression of this solution ?
   - If $\boldsymbol{A} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \sin(\theta) \end{bmatrix}$ is a rotation matrix. Can we determine θ?
   - If $\boldsymbol{A} = \begin{bmatrix} a_{11} & 0 \\ 0 & a_{22} \end{bmatrix} \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \sin(\theta) \end{bmatrix}$. Can we determine θ, a_{11} and a_{22}? If we impose $a_{11} = a_{22} = a$, can we determine a and θ?
   - Now, if are only given $\boldsymbol{g}$, can we determine $\boldsymbol{A}$ and $\boldsymbol{f}$? Is the solution unique? What if we impose $\|\boldsymbol{A}\|^2$ to be minimum, find it and then use it to find $\boldsymbol{f}$?
   - Now, consider $\boldsymbol{g}(t) = \boldsymbol{A}\boldsymbol{f}(t), t = 1, \cdots, T$. Define the matrices $\boldsymbol{G} = [\boldsymbol{g}(1); \cdots ; \boldsymbol{g}(T)]$, $\boldsymbol{F} = [\boldsymbol{f}(1); \cdots ; \boldsymbol{f}(T)]$ and show that we can write $\boldsymbol{G} = \boldsymbol{A}\boldsymbol{F}$. If T is great enough, can we determine $\boldsymbol{A}$ from $\boldsymbol{F}$ and $\boldsymbol{G}$?

3. Consider now the general case of $\boldsymbol{g} = \boldsymbol{A}\boldsymbol{f}$ where the matrix $\boldsymbol{A}$ has dimensions $(M \times N)$.

   - Given $\boldsymbol{A}$ and $\boldsymbol{g}$ propose a solution (exact or approximate) for $\boldsymbol{f}$ for the following cases $M = N$, $M < N$ and $M > N$.
   - Given $\boldsymbol{f}$ and $\boldsymbol{g}$ propose a solution (exact or approximate) for $\boldsymbol{A}$ for the following cases $M = N$, $M < N$ and $M > N$.
     Has this problem an unique solution? Can we impose some constraints on elements of $\boldsymbol{A}$ to be able to find a unique solution?
     Which one ?

- If we impose minimum norme $\|\boldsymbol{A}\|^2 = \|\boldsymbol{a}\|^2$, can you give the expression of this solution ?

4. Consider now the following relations $\boldsymbol{g} = \boldsymbol{A}\boldsymbol{f}$ and $\widehat{\boldsymbol{f}} = \boldsymbol{B}\boldsymbol{g}$ where $\boldsymbol{A}$ and $\boldsymbol{B}$ are two invertible matrices of dimensions $(N \times N)$. Interprete the following situations

   - $\boldsymbol{A} = \boldsymbol{B} = \boldsymbol{I}$.
   - The j th column of $\boldsymbol{A}$ is all zero.
   - The j th column of $\boldsymbol{B}$ is all zero.
   - The i th ligne of $\boldsymbol{A}$ is all zero.
   - The i th ligne of $\boldsymbol{B}$ is all zero.
   - $\boldsymbol{A}$ is tri-diagonal, symetric and Toeplitz with $a_{12} = a_{21} = .5$
   - $\boldsymbol{B}$ is tri-diagonal, symetric and Toeplitz with $a_{12} = a_{21} = .5$

5. Consider now the following relations $\boldsymbol{g} = \boldsymbol{A}\boldsymbol{f}$, $\widehat{\boldsymbol{f}} = \boldsymbol{B}\boldsymbol{g}$ and $\widehat{\boldsymbol{g}} = \boldsymbol{A}\widehat{\boldsymbol{f}}$:

$$\boldsymbol{f} \longrightarrow \boxed{\boldsymbol{A}} \longrightarrow \boldsymbol{g} \longrightarrow \boxed{\boldsymbol{B}} \longrightarrow \widehat{\boldsymbol{f}} \longrightarrow \boxed{\boldsymbol{A}} \longrightarrow \widehat{\boldsymbol{g}}$$

   - In which conditions on $\boldsymbol{A}$ and $\boldsymbol{B}$ we have $\widehat{\boldsymbol{f}} = \boldsymbol{f}$?
   - In which conditions $\|\boldsymbol{g} - \boldsymbol{A}\boldsymbol{f}\|^2$ is minimum?
   - In which conditions $\|\widehat{\boldsymbol{f}} - \boldsymbol{B}\boldsymbol{g}\|^2$ is minimum?
   - In which conditions $\|\widehat{\boldsymbol{f}} - \boldsymbol{f}\|^2$ is minimum?
   - In which conditions $\|\widehat{\boldsymbol{g}} - \boldsymbol{g}\|^2$ is minimum?

6. Geometrical interpretation of $\boldsymbol{g} = \boldsymbol{A}\boldsymbol{f}$ and $\widehat{\boldsymbol{f}} = \boldsymbol{B}\boldsymbol{g}$

   - Show that $\boldsymbol{g} = \boldsymbol{A}\boldsymbol{f}$ can be written as $\boldsymbol{g} = f_1\boldsymbol{a}_1 + \cdots + f_N\boldsymbol{a}_N$ where $\boldsymbol{a}_j$ is the j th column of $\boldsymbol{A}$.
   - Show that $\boldsymbol{g} = \boldsymbol{A}\boldsymbol{f}$ can be written as $\boldsymbol{g} = g_1\boldsymbol{e}_1 + \cdots + g_N\boldsymbol{e}_N$ where $\boldsymbol{e}_1 = [1, 0, \cdots, 0]'$, $\boldsymbol{e}_2 = [0, 1, \cdots, 0]'$ and $\boldsymbol{e}_N = [0, 0, \cdots, 1]'$.
   - Comparing these two representations, what can we conclude?

7. Consider now $\boldsymbol{W} = \boldsymbol{B}'$ and so $\widehat{\boldsymbol{f}} = \boldsymbol{B}\boldsymbol{g} = \boldsymbol{W}'\boldsymbol{g} = \boldsymbol{W}'\boldsymbol{A}\boldsymbol{f} = \boldsymbol{Z}'\boldsymbol{f}$ with $\boldsymbol{Z} = \boldsymbol{A}'\boldsymbol{W}$.

   - Write $\widehat{\boldsymbol{f}} = f_1\boldsymbol{z}_1 + \cdots + f_N\boldsymbol{z}_N$ where $\boldsymbol{z}_j$ is the j-th column of $\boldsymbol{Z}'$.
     How to interpret the situation where $\boldsymbol{z}_j$ is composed of only the zeros?

8. Probabilistic interpretation

   - If $p(\boldsymbol{f}) = \mathcal{N}(0, \boldsymbol{I})$, what will be $p(\boldsymbol{g})$ and $p(\widehat{\boldsymbol{f}})$?
   - Find $\boldsymbol{W}$ such that $\text{cov}[\boldsymbol{g}] = \boldsymbol{I}$.
   - Find $\boldsymbol{W}$ such that $\text{cov}[\widehat{\boldsymbol{f}}] = \boldsymbol{I}$.

9. Consider now $\boldsymbol{g}(t) = \boldsymbol{A}\boldsymbol{f}(t), t = 1, \cdots, T$ and focusing on the estimation of $\boldsymbol{A}$ given $\boldsymbol{g}(t)$ and $fb(t)$ for $t = 1, \cdots, T$.

   - If $p(\boldsymbol{f}(t)) = \mathcal{N}(0, \boldsymbol{I}), \forall t$, what will be $p(\boldsymbol{g}(t))$ and $p(\widehat{\boldsymbol{f}}(t))$?
   - Find $\boldsymbol{W}$ such that $\text{cov}[\widehat{\boldsymbol{f}}] = \boldsymbol{I}$.
   - Propose an algorithm which first estimates $\text{cov}[\boldsymbol{g}]$ from the data and then find $\boldsymbol{W}$ such that $\text{cov}[\widehat{\boldsymbol{f}}] = \boldsymbol{I}$.

10. Consider now $\boldsymbol{g}(t) = \boldsymbol{A}\boldsymbol{f}(t), t = 1, \cdots, T$ and focusing on the estimation of $\boldsymbol{A}$ given only $\boldsymbol{g}(t)$ for $t = 1, \cdots, T$. When $\boldsymbol{A}$ is obtained, we can also estimate $\boldsymbol{f}(t)$.

   - First assume $p(\boldsymbol{f}(t)) = \mathcal{N}(0, \boldsymbol{\Sigma}_f), \forall t$, what will be $p(\boldsymbol{g}(t))$ and $p(\widehat{\boldsymbol{f}}(t))$?
   - Propose an algorithm which first estimates $\text{cov}[\boldsymbol{g}]$ from the data. Then by comparing the SVD of it $\text{cov}[\boldsymbol{g}] = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}'$ to the $\text{cov}[\boldsymbol{g}] = \boldsymbol{A}\boldsymbol{\Sigma}_f\boldsymbol{A}'$, find $\boldsymbol{A}$ or equivalently $\boldsymbol{W}$ and $\boldsymbol{\Sigma}_f$.
   - Consider now $\widetilde{\boldsymbol{g}} = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{U}'\boldsymbol{g}$. Show that $\text{cov}[\boldsymbol{g}] = \boldsymbol{I}$. ($\widetilde{\boldsymbol{g}}$ is *whitened data* $\boldsymbol{g}$.)

Parta B: Factor Analysis: from deterministic to probabilistic methods

1. Consider three unknowns f_1, f_2 and f_3. We know the 2 by 2 sums of them, *i.e.* $f_1 + f_2 = g_1$, $f_1 + f_3 = g_2$ and $f_2 + f_3 = g_3$. Determine them.
   Help: Write the problem in the form of $\boldsymbol{g} = \boldsymbol{A}\boldsymbol{f}$ with $\boldsymbol{g} = [g_1, g_2, g_3]'$; $\boldsymbol{f} = [f_1, f_2, f_3]'$ and $\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$.
   Check then if this matrix is invertible and inverse it and find the solution $\widehat{\boldsymbol{f}} = \boldsymbol{A}^{-1}\boldsymbol{g}$.
   For numerical application, $g_1 = 5, g_2 = 7$ and $g_3 = 8$.

2. What if we know their differences, *i.e.*$f_1 - f_2 = g_1$, $f_1 - f_3 = g_2$ and $f_2 - f_3 = g_3$?

3. What if we also know their sums, *i.e.*$f_1 + f_2 + f_3 = g_4$?

4. What if we only know the sums $f_1 + f_2 + f_3 = g_1$ and the $f_1 - f_2 - f3 = g_2$?

5. Consider now the general case $\boldsymbol{g} = \boldsymbol{A}\boldsymbol{f}$ where the matrix $\boldsymbol{A}$ has dimensions $(M \times N)$ and propose solutions for the following cases:

   - $M = N$ and the matrix is invertible and well conditionned (ideal case);
   - $M < N$,
   - $M > N$,
   - $M = N$, but the matrix is not invertible or invertible but ill-conditionned; and
   - a solution which can work for all cases.

6. what happens if the data $\boldsymbol{g}$ has errors $\boldsymbol{g} + \delta\boldsymbol{g}$ or $\boldsymbol{g} + \boldsymbol{\epsilon}$? Can we say something about the estimation error $\delta\boldsymbol{f}$?

7. Consider now the general case $\boldsymbol{g} = \boldsymbol{A}\boldsymbol{f} + \boldsymbol{\epsilon}$.

   - We do not know much about the errors $\boldsymbol{\epsilon}$.
     What are the solutions which correspond to $\boldsymbol{\epsilon}\|^2 = \|\boldsymbol{g} - \boldsymbol{A}\boldsymbol{f}\|^2 = 0$?
     What are the solutions which correspond to $\boldsymbol{\epsilon}\|^2 = \|\boldsymbol{g} - \boldsymbol{A}\boldsymbol{f}\|^2 \leq c$?
     How to find all of them ?
   - Between all these possible solutions which one has minimum norme $\|\boldsymbol{f}\|^2$?
   - If we know that the solution must be positive, can we find it ?
     Propose an algorithm to find that solution if it exists.
   - If we know that many components of the solution $\boldsymbol{f}$ are zero.
     How can we impose this?
     Propose an algorithm to find that solution if it exists.

Advanced Signal and Image Processing

Professor: A. Mohammad–Djafari

Exercise number 4: Parameter estimation

Part 1: Parametric modeling and estimation
Exercise 1: Recursive parameter estimation

Assume $Z_i = X_i + N_i, i = 1, \ldots, n$ where $N_i, i = 1, \ldots, n$ is a sequence of i.i.d. Gaussian random variables, each with zero mean and variance σ^2, and $X_i, i = 1, \ldots, n$ are defined by

$$X_0 = \Theta, \quad X_i = \alpha X_{i-1}, \quad i = 1, \ldots, n$$

where α is known and Θ is a centered Gaussian random parameter with zero mean and variance q^2.

1. Assuming that Θ and N_i are independent, find the MMSE estimation of Θ based on $Z_1, \ldots, Z_n$.
2. For each $n = 1, 2, \ldots$, let $\widehat{\theta}_n$ denote the MMSE estimate of Θ based on $Z_1, \ldots, Z_n$. Show that $\widehat{\theta}_n$ can be computed recursively by

$$\widehat{\theta}_n = K_n^{-1} \left[K_{n-1} \widehat{\theta}_{n-1} + \alpha^n Z_n \right], \quad n = 1, 2, \ldots$$

   where $\widehat{\theta}_0 = 0$ and

$$K_0 = \frac{\sigma^2}{q^2} \quad \text{and} \quad K_n = K_{n-1} + \alpha^{2n}$$

3. Draw a block diagram of this implementation.
4. Find an expression for the MSE $e_n = \mathrm{E}\left\{ (\widehat{\theta}_n - \Theta)^2 \right\}$.
5. For cases $\alpha < 1$; $\alpha = 1$ and $\alpha > 1$, what happens when $n \mapsto \infty$; $q^2 \mapsto \infty$; and $\sigma^2 \mapsto 0$?

Exercise 2: Parameter estimation

Assume $Z_i = A \sin(i\,\pi/2 + \Phi) + N_i, i = 1, \ldots, n$ where $N_i, i = 1, \ldots, n$ is a sequence of i.i.d. Gaussian random variables, each with zero mean and variance σ^2, and n is even.

1. Suppose A and Φ are non random with $A > 0$ and $\Phi \in [-\pi, \pi]$. Find their ML estimates.
2. Suppose A and Φ are random and independent with priors

$$\pi(\phi) = \begin{cases} 1/\pi, & -\pi \leq \phi \leq \pi \\ 0, & \text{otherwise} \end{cases}$$

$$\pi(a) = \begin{cases} \frac{a}{\beta^2} \exp\left\{ -\frac{a^2}{2\beta^2} \right\}, & a \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

   where β is known. Assume also that A and Φ are independent of $\underline{N}$.

2.1 Give the expression of the posterior law $\pi(A, \Phi | Z_1, \dots, Z_n)$;

2.2 Find the joint MAP estimates of A and Φ;

2.3 Give the expression of the posterior laws $\pi(A | Z_1, \dots, Z_n)$ and $\pi(\Phi | Z_1, \dots, Z_n)$, and
find the MAP estimates of A and Φ

3. Under what conditions are the estimates from (1) and (2) approximately equal?

Exercise 3: Discrete periodic deconvolution

Assume $Z_i = S_i + N_i$ where

$$S_i = \sum_{k=0}^{n-1} h_k\, X_{i-k}$$

and where $\underline{h} = [h_0, h_1, \ldots, h_{n-1}]^t$ represents the finite impulse response of a channel. Assume also that X_i and S_i are periodic sequences with known period n, *i.e.* $X_{n\pm k} = X_i$ and $S_{n\pm k} = S_i$ for any integer k and that we observe n samples $\underline{Z} = [Z_0, \ldots, Z_{n-1}]^t$. We want estimate the input sequence $\underline{X} = [X_0, \ldots, X_{n-1}]^t$. Finally, we assume that $\underline{N} \sim \mathcal{N}\left(\underline{0}, \theta \boldsymbol{I}\right)$ and independent of $\underline{X}$.

1. Construct the matrices $\boldsymbol{H}$ and $\boldsymbol{X}$ in such a way that $\underline{Z} = \boldsymbol{H}\underline{X} + \underline{N}$ and $\underline{Z} = \boldsymbol{X}\underline{h} + \underline{N}$.
2. Assume that $\underline{h}$ and $\underline{X}$ are known. Design a Bayesian optimal detector with the uniform prior and the uniform cost coefficients.
3. Assume now that $\underline{h}$ is known and we want to estimate $\underline{X}$. Find the ML estimate $\widehat{\underline{X}}_{ML}(\underline{Z})$ of $\underline{X}$.
4. Find the MAP estimate $\widehat{\underline{X}}_{MAP}(\underline{Z})$ and the MMSE estimate $\widehat{\underline{X}}_{MMSE}(\underline{Z})$ if we assume that $\underline{X} \sim \mathcal{N}\left(\underline{0}, \sigma_x^2 \boldsymbol{I}\right)$.
5. Assume now that the input sequence $\underline{X}$ is known but $\underline{h}$ is unknown. Find the ML estimate $\widehat{\underline{h}}_{ML}(\underline{Z})$ of $\underline{h}$.
6. Find the MAP estimate $\widehat{\underline{h}}_{MAP}(\underline{Z})$ and the MMSE estimate $\widehat{\underline{h}}_{MMSE}(\underline{Z})$ if we assume that $\underline{h} \sim \mathcal{N}\left(\underline{0}, \sigma_h^2 \boldsymbol{I}\right)$.
7. Noting that $\boldsymbol{H}$ and $\boldsymbol{X}$ in (1) are circulant matrices and that we have the following relations:

   $$\begin{array}{llll} \boldsymbol{H} = \boldsymbol{F}^t \boldsymbol{\Lambda}_h \boldsymbol{F}, & \boldsymbol{H}^t\boldsymbol{H} = \boldsymbol{F}^t \boldsymbol{\Lambda}_h^2 \boldsymbol{F}, & \boldsymbol{\Lambda}_h = \operatorname{diag}\left[\mathrm{DFT}[h_0, \cdots, h_{n-1}]\right] = \operatorname{diag}\left[\underline{\tilde{h}}(\omega)\right] & \\ \boldsymbol{X} = \boldsymbol{F}^t \boldsymbol{\Lambda}_x \boldsymbol{F}, & \boldsymbol{X}^t\boldsymbol{X} = \boldsymbol{F}^t \boldsymbol{\Lambda}_x^2 \boldsymbol{F}, & \boldsymbol{\Lambda}_x = \operatorname{diag}\left[\mathrm{DFT}[X_0, \cdots, X_{n-1}]\right] = \operatorname{diag}\left[\underline{\tilde{X}}(\omega)\right] & \\ \boldsymbol{F}^t\boldsymbol{F} = \boldsymbol{F}\boldsymbol{F}^t = \boldsymbol{I}, & \underline{\tilde{h}}(\omega) = \boldsymbol{F}\underline{h}, & \underline{\tilde{X}}(\omega) = \boldsymbol{F}\underline{X}, \quad \underline{h} = \boldsymbol{F}^t\underline{\tilde{h}}(\omega), \quad \underline{X} = \boldsymbol{F}^t\underline{\tilde{X}}(\omega) & \end{array}$$

   find the expressions of $\underline{\tilde{X}}(\omega)_{ML}$ and $\underline{\tilde{h}}(\omega)_{ML}$ in (3) and (5) and $\underline{\tilde{X}}(\omega)_{MAP}$ and $\underline{\tilde{h}}(\omega)_{MAP}$ in (4) and (6).
8. Now assume that $\underline{h}$ and $\underline{X}$ are both unknown. Find the MAP and the MMSE estimates $\widehat{\underline{X}}_{MAP}(\underline{Z})$ and $\widehat{\underline{X}}_{MMSE}(\underline{Z})$ of $\underline{X}$ and $\widehat{\underline{h}}_{MAP}(\underline{Z})$ and $\widehat{\underline{h}}_{MMSE}(\underline{Z})$ of $\underline{h}$ if we assume that $\underline{X} \sim \mathcal{N}\left(\underline{0}, \sigma_x^2 \boldsymbol{I}\right)$ and $\underline{h} \sim \mathcal{N}\left(\underline{0}, \sigma_h^2 \boldsymbol{I}\right)$.
9. Assume now that X_i can only take the values $\{0, 1\}$. with $P(X_i = 0) = \pi_0$ and $P(X_i = 1) = 1 - \pi_0$, with known π_0, and that X_i are independent. Design an optimal detector for X_i assuming $\boldsymbol{h}$ to be known.
10. Assume now that $X_i, i = 1, \ldots, n$ can be modelled as a first order Markov chain with transition probabilites

    $$\begin{array}{ll} P(X_i = 0, X_{i-1} = 0) = s & P(X_i = 0, X_{i-1} = 1) = 1 - s \\ P(X_i = 1, X_{i-1} = 1) = t & P(X_i = 1, X_{i-1} = 0) = 1 - t \end{array}$$

    Design an optimal detector for X_i.

Advanced Signal and Image Processing

Professor: A. Mohammad–Djafari

Exercise number 5: State space and Kalman Filtering

Part 1: State space and Input-Output models
Exercise 1:

$$\begin{cases} \dot{x}(t) & = a\,x(t) + b\,u(t) \quad \text{state equation} \\ y(t) & = c\,x(t) \quad \text{observation equation} \end{cases}$$

where a,b, and c are the scalar values.

1. Give the expression of the impulse response $h(t)$, the transfert functions $H(p)$ and $H(\omega)$ of this system.
2. Give the expression of the output $y(t)$ when the input is $u(t) = 0, t < 0; u(t) = 1, t \geq 0$.
3. Give the expression of the output $y(t)$ when the input is $u(t) = \sum_k \sin(kt)$.
4. Give the expression of the output $y(t)$ for any arbitrary input $u(t)$.
5. Give the expression of the autocorrlation function of the output $C_y(\tau)$ when the input is a Gaussian centered random signal with variance σ^2.
6. Give the expression of the power spectral density (psd) of the output $S_y(\omega)$ as a function of the psd of the input $S_u(\omega)$.

Exercise 2 : Assume now that we have access to the input $u(t)$ and the output $y(t)$ for $t = 1, \cdots, N$.

1. We want to determine $h(t)$, $H(p)$, $H(\omega)$ and the parameters a, b and c of the system. For each of the following cases, propose a method and give all the needed hypothesis and limitations:
   - Finite Impulse Response (FIR) model:
   $$y(t) = \sum_{k=0}^{K-1} h_k\, u(t-k) + \epsilon(t)$$
   with the parameters $\boldsymbol{\theta} = [h_0, \cdots, h_{K-1}]'$.
   - Auto Regressive (ARX):
   $$y(t) = a_1\, y(t-1) + a_2\, y(t-2) + b_1\, u(t-1) + \epsilon(t)$$
   with the parameters $\boldsymbol{\theta} = [a_1, a_2, b_1]'$.
   - Auto Regressive Moving Average (ARMAX):
   $$y(t) = a_1\, y(t-1) + a_2\, y(t-2) + b_1\, u(t-1) + c_1\, \epsilon(t-1) + \epsilon(t)$$
   with the parameters $\boldsymbol{\theta} = [a_1, a_2, b_1, c_1]'$.

2. Show that, for all these models we can write:

$$y(t) = \boldsymbol{\phi}'(t)\boldsymbol{\theta} + \epsilon(t), \qquad \text{and} \qquad \boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{\theta} + \boldsymbol{\epsilon}$$

   (For each case give the expressions of $\boldsymbol{\phi}'(t)$ and $\boldsymbol{\Phi}$.)

3. For each case, give the expression of the LS estimator $\widehat{\boldsymbol{\theta}}$.

4. For each case, give the expression of the Likelihood $\mathcal{L}(\boldsymbol{\theta}) = -\ln p(\boldsymbol{y}|\boldsymbol{\theta})$ and the ML estimator $\widehat{\boldsymbol{\theta}}$.

5. Can we always find an analytical expression of the likelihood $\mathcal{L}(\boldsymbol{\theta}) = -\ln p(\boldsymbol{y}|\boldsymbol{\theta})$ and the ML estimator $\widehat{\boldsymbol{\theta}}$?

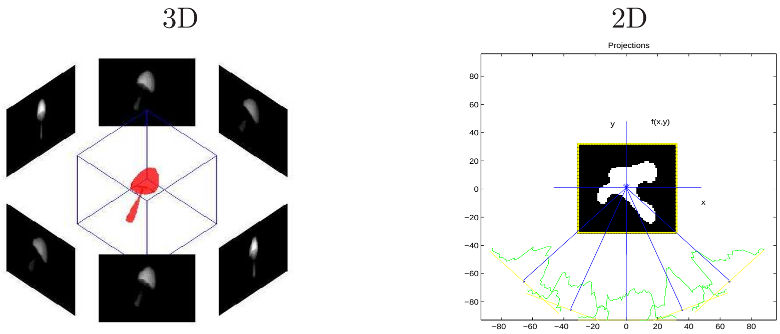
6. Under which conditions the two estimators $\widehat{\boldsymbol{\theta}}_{LS}$ and $\widehat{\boldsymbol{\theta}}_{ML}$ are identical ?

7. Consider now the FIR model. Describe all the steps to obtain the LS estimator $\boldsymbol{\theta}$ in a recursive way.

Inverse Problems in Signal Processing, Imaging Systems and Computer Vision

Professor: A. Mohammad–Djafari
Exercice number 3: Tomography and Image Reconstruction

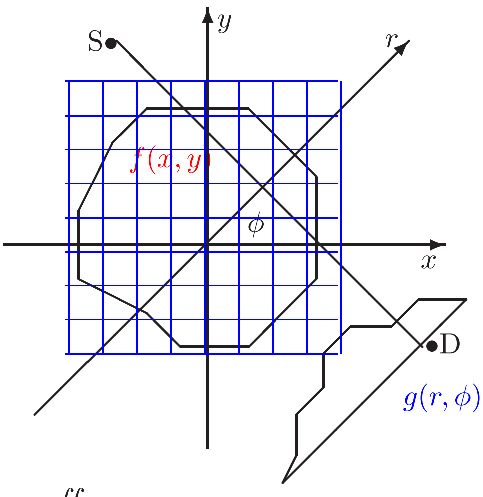
In X ray tomography, a simple model which links the relative intensity of the rays measured on a detector g to the distribution of absorption coefficient inside the object f by a line integral equation joining the position of the source to the position of the detector. The following figure shows this relation:



$$g_\phi(r_1,r_2) = \int_{\mathcal{L}_{r_1,r_2,\phi}} f(x,y,z)\,\mathrm{d}l \qquad g_\phi(r) = \int_{\mathcal{L}_{r,\phi}} f(x,y)\,\mathrm{d}l$$

Fig. 1: X ray Computed Tomography (CT)

In the following, we will consider the 2D Case, where we can caracterize this line by two variables r and ϕ, and so, we can relate the observed projections $g(r,\phi)$ to the object $f(x,y)$ by the following relations:



$$g(r,\phi) = \iint_D f(x,y)\,\delta(r - x\cos\phi - y\sin\phi)\,\mathrm{d}x\,\mathrm{d}y$$

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

Fig. 2: X ray Computed Tomography and Radon Transform.

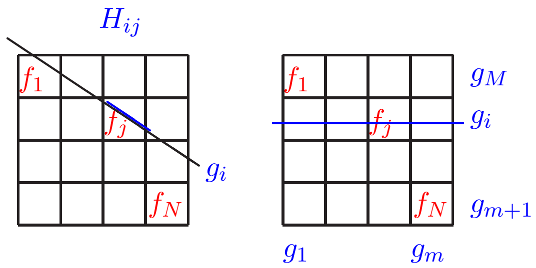


Fig. 3: Discretisation of the CT problem

Part 1: Analytical Image reconstruction methods

Consider the problem of image reconstructionin 2D X ray CT where the relation between the object $f(x,y)$ and the projections $g(r,\phi)$ is modeled by the Radon Transfrom (RT):

$$g(r,\phi) = \int_{L_{(r,\phi)}} f(x,y)\,\mathrm{d}l = \iint f(x,y)\delta(r - x\cos\phi - y\sin\phi)\,\mathrm{d}x\,\mathrm{d}y,$$

where $L_{(r,\phi)}$ is a line making the angle ϕ with the axis x and positionned at a distance from the origin r.

Starting by the expression of inverse Radon Transform, *i.e.*;

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

and using the following operations:

Derivation $\mathcal{D}$: $\quad \overline{g}(r,\phi) = \dfrac{\partial g(r,\phi)}{\partial r}$

Hilbert Transform $\mathcal{H}$: $\quad g_1(r',\phi) = \dfrac{1}{\pi}\displaystyle\int_0^{\infty} \frac{\overline{g}(r,\phi)}{(r-r')}\,\mathrm{d}r$

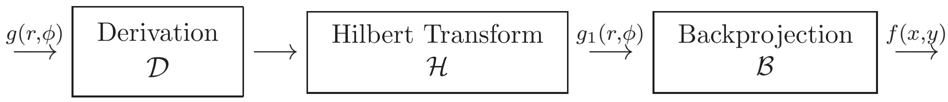
Backprojection $\mathcal{B}$: $\quad f(x,y) = \dfrac{1}{2\pi}\displaystyle\int_0^{\pi} g_1(x\cos\phi + y\sin\phi,\phi)\,\mathrm{d}\phi$

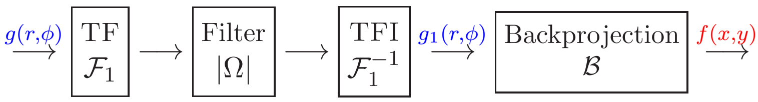
1D Fourier Transform $\mathcal{F}_1$: $\quad G(\Omega,\phi) = \displaystyle\int g(r,\phi)\exp\{-j\Omega r\}\,\mathrm{d}r$

2D Fourier Transform $\mathcal{F}_2$: $\quad F(u,v) = \displaystyle\iint f(x,y)\exp\{-j(ux+vy)\}\,\mathrm{d}x\,\mathrm{d}y$

1. Show that we can determine $f(x,y)$ by any of the following schemes:

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





2. Show that, if we define:

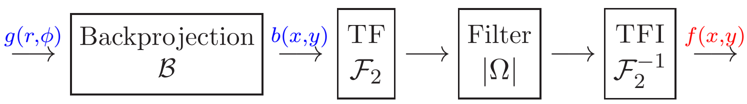
$$b(x,y) = \int_0^{\pi} g(r,\phi)\,\mathrm{d}\phi = \int_0^{\pi} g(x\cos\phi + y\sin\phi,\phi)\,\mathrm{d}\phi$$

we have a relation between $f(x,y)$ and $b(x,y)$ in the following form:

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

Show then that:

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



where $|\Omega|^2 = u^2 + v^2$.

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

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

4. Show that, if the object f has a property of rotational symmetry, *i.e.*; $f(x, y) = f(\rho)$ with $\rho^2 = x^2 + y^2$, then, it can be reconstructed only from one projection.

5. Show that, if $f(x, y) = f_1(x)\, f_2(y)$ then it is possible to reconstruct it (up to a constant value) only from two projections $\phi = 0$ and $\phi = 90$.

Inverse Problems in Signal Processing, Imaging Systems and Computer Vision

Professor: A. Mohammad–Djafari

Exercice number 4: Deconvolution par Least Squares and regularisation: Continuous case

Part 1: Deconvolution

Consider the problem of Deconvolution where the measured signal $g(t)$ is related to the input signal $f(t)$ and the impulse response $h(t)$ by $g(t) = h(t) * f(t) + \epsilon(t)$ and where we are looking to estimate $h(t)$ and $f(t)$. from $g(t)$.

1. First assume $h(t)$ to be known (Simple Deconvolution). Let define the solution $\widehat{f}(t)$ by
$$\widehat{f}(t) = \arg\min_{f} \left\{ \|g - h * f\|^2 + \lambda_1 \|d_f * f\|^2 \right\},$$
where $d_f(t)$ and λ_1 are known and fixed and where the norme $\|z\|^2$ is defined as:
$$\|z\|^2 = \int |z(t)|^2 \, \mathrm{d}t.$$
Show that this solution can be computed by:
$$F(\nu) = \frac{H^*(\nu)}{|H(\nu)|^2 + \lambda_1 |D_f(\nu)|^2} G(\nu),$$
where $F(\nu)$ and $G(\nu)$ are the spectrale density functions (spd) of $f(t)$ and $g(t)$ and $H(\nu)$ and $D_f(\nu)$ are the FT of $h(t)$ and $d_f(t)$.
2. What becomes this solution when: $\lambda_1 = 0$?
3. What is the role of d_f ou $D_f(\nu)$? How to choose it?
4. Now, let try to estimate $h(t)$ from $g(t)$ and $f(t)$ (Identification). Let define the solution $\widehat{h}(t)$ by
$$\widehat{h}(t) = \arg\min_{h} \left\{ \|g - h * f\|^2 + \lambda_2 \|d_h * h\|^2 \right\}.$$
Show that this solution can be computed by:
$$H(\nu) = \frac{F^*(\nu)}{|F(\nu)|^2 + \lambda_2 |D_h(\nu)|^2} G(\nu).$$
5. What becomes this solution when: $\lambda_2 = 0$?
6. What is the role of d_h or $D_h(\nu)$? How to choose it?
7. Let assume that we want to estimate both $h(t)$ and $f(t)$ from only the output $g(t)$ (Blind Deconvolution). Can we suggest to define a solution by:
$$\left(\widehat{f}(t), \widehat{h}(t)\right) = \arg\min_{(f,h)} \left\{ \|g - h * f\|^2 + \lambda_1 \|d_f * f\|^2 + \lambda_2 \|d_h * h\|^2 \right\} \quad ?$$
Why? Does this criterion has a unique solution? Comment your answer.

Part 2: Deconvolution – Discrete case

Consider now the same problem assuming that the system is causal and the signals are causal too, and the impulse response of the system is of finite support. Then, using: $\boldsymbol{h} = [h_0, \cdots, h_p]^t$, $\boldsymbol{f} = [f_0, \cdots, f_M]^t$ and $\boldsymbol{g} = [g_0, \cdots, g_M]^t$.

1. Find the matrix $\boldsymbol{H}$ in such a way that we can write $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{b}$.
2. What is the structure of this matrix?
3. How can we transform it to a circulent matrix?
4. Find the matrix $\boldsymbol{F}$ in such a way that we can write $\boldsymbol{g} = \boldsymbol{F}\boldsymbol{h} + \boldsymbol{b}$.
5. What is the structure of this matrix?
6. How can we transform it to a circulent matrix?
7. Suppose first $\boldsymbol{h}$ and $\boldsymbol{g}$ are known. Let define the solution $\widehat{\boldsymbol{f}}$ by
$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \left\{ |\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}|^2 + \lambda_f |\boldsymbol{D}_f \boldsymbol{f}|^2 \right\},$$
where $\boldsymbol{D}_f$ is the matrix of finite difference (approximation of the zero order derivation).
8. Write the expression of this solution.
9. What becomes this solution when: $\lambda_f = 0$?
10. Propose a method to compute this solution and comment your choice.
11. Suppose that we had transformed the matrices $\boldsymbol{H}$ and $\boldsymbol{D}_f$ circulent matrices: ($\boldsymbol{H} = \text{circ}(\boldsymbol{h})$ and $\boldsymbol{D}_f = \text{circ}(\boldsymbol{d}_f)$). Show then that this solution can be computed by:
$$F(\nu) = \frac{H^*(\nu)}{|H(\nu)|^2 + \lambda_f |D_f(\nu)|^2} G(\nu),$$
where $F(\nu)$ and $G(\nu)$ are the DFT of $\boldsymbol{f}$ and $\boldsymbol{g}$ and $H(\nu)$ and $D_f(\nu)$ are the DFT of $\boldsymbol{h}$ and $\boldsymbol{d}_f$.
12. Suppose now $\boldsymbol{f}$ and $\boldsymbol{g}$ be known. Let define the solution $\widehat{\boldsymbol{h}}$:
$$\widehat{\boldsymbol{h}} = \arg\min_{\boldsymbol{h}} \left\{ |\boldsymbol{g} - \boldsymbol{F}\boldsymbol{h}|^2 + \lambda_h |\boldsymbol{D}_h \boldsymbol{h}|^2 \right\}$$
where $\boldsymbol{D}_h$ isa known matrix.
13. Write the expression of this solution.
14. What becomes this solution when: $\lambda_h = 0$?
15. Propose a method to compute this solution and comment your choice.
16. Suppose that we had transformed the matrices $\boldsymbol{F}$ and $\boldsymbol{D}_h$ to circulant matrices. ($\boldsymbol{F} = \text{circ}(\boldsymbol{f})$ and $\boldsymbol{D}_h = \text{circ}(\boldsymbol{d}_h)$). Show then that this solution can be computed by:
$$H(\nu) = \frac{F^*(\nu)}{|F(\nu)|^2 + \lambda_h |D_h(\nu)|^2} G(\nu).$$
where $F(\nu)$ and $G(\nu)$ are the DFT of $\boldsymbol{f}$ and $\boldsymbol{g}$ and $H(\nu)$ and $D_h(\nu)$ are the DFT of $\boldsymbol{h}$ and $\boldsymbol{d}_h$.

17. Let now assume that we want to estimate both $\boldsymbol{h}$ and $\boldsymbol{f}$ from $\boldsymbol{g}$ (Blind Deconvolution). Can we suggest to define a solution by:

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

Why? Does this criterion has a unique solution?. Comment your answer.

Inverse Problems in Signal Processing, Imaging Systems and Computer Vision

Professor: A. Mohammad–Djafari

Exercice number 5: Least Squares, Generalized inversion and Regularisation

Part 1:

In a measurement system, we have established the following relation: $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}$ where
$\boldsymbol{g}$ is a vector containig the projections (measured data or observations) $\{g_m,\, m = 1\cdots, M\}$,
$\boldsymbol{\epsilon}$ is a vector representing the errors (measurement and modeling) $\{\epsilon_m,\, m = 1\cdots, M\}$,
$\boldsymbol{f}$ is a vector representing the pixels of the image $\{f_n,\, n = 1\cdots, N\}$, and
$\boldsymbol{H}$ is a matrix with the elements $\{a_{mn}\}$ depending on the geometry of the measurement system and assumed to be known.

1. Suppose first $M = N$ and that the matrix $\boldsymbol{H}$ be invertible. Why the solution $\widehat{\boldsymbol{f}}_0 = \boldsymbol{H}^{-1}\boldsymbol{g}$ is not, in general, a satisfactory solution?

   What relation exists between $\frac{\|\delta\widehat{\boldsymbol{f}}_0\|}{\|\widehat{\boldsymbol{f}}_0\|}$ and $\frac{\|\delta\boldsymbol{g}\|}{\|\boldsymbol{g}\|}$?

2. Let come back to the general case $M \neq N$. Show then that the Least Squares (LS) solution, *i.e.* $\widehat{\boldsymbol{f}}_1$ which minimises

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

   is also a solution of equation $\boldsymbol{H}'\boldsymbol{H}\boldsymbol{f} = \boldsymbol{H}'\boldsymbol{g}$ and if $\boldsymbol{H}'\boldsymbol{H}$ is invertible, then we have

   $$\widehat{\boldsymbol{f}}_1 = [\boldsymbol{H}'\boldsymbol{H}]^{-1}\boldsymbol{H}'\boldsymbol{g}$$

   What is the relation between $\frac{\|\delta\widehat{\boldsymbol{f}}_1\|}{\|\widehat{\boldsymbol{f}}_1\|}$ and $\frac{\|\delta\boldsymbol{g}\|}{\|\boldsymbol{g}\|}$?

3. What is the relation between the covarience of $\widehat{\boldsymbol{f}}_1$ and covarience of $\boldsymbol{g}$?

4. Consider now the case $M < N$. Evidently, $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f}$ has infinite number of solutions. The minimum norm solution is:

   $$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{H}\boldsymbol{f}=\boldsymbol{g}} \left\{\|\boldsymbol{f}\|^2\right\}$$

   Show that this solution is obtained via:

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

   which gives:

   $$\widehat{\boldsymbol{f}}_2 = \boldsymbol{H}^t(\boldsymbol{H}\boldsymbol{H}^t)^{-1}\boldsymbol{g}$$

   if $\boldsymbol{H}\boldsymbol{H}^t$ is invertible.

5. Show that with this solution we have: $\widehat{\boldsymbol{g}} = \boldsymbol{H}\widehat{\boldsymbol{f}}_2 = \boldsymbol{g}$.

6. What is the relation between the covarience of $\widehat{\boldsymbol{f}}_2$ and covarience of $\boldsymbol{g}$?

7. Let come back to the general case $M \neq N$ and define

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \{J(\boldsymbol{f})\} \quad \text{with} \quad J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + \lambda\|\boldsymbol{f}\|^2$$

Show that for any $\lambda > 0$, this solution exists and is unique and is obtained by:

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

8. What relation exists between $\widehat{\boldsymbol{g}} = \boldsymbol{H}\widehat{\boldsymbol{f}}$ and $\boldsymbol{g}$?

9. What is the relation between the covarience of $\widehat{\boldsymbol{f}}$ and covarience of $\boldsymbol{g}$?

10. Another regularised solution $\widehat{\boldsymbol{f}}_2$ to this problem is to minimize a criterion such as:

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

where $\boldsymbol{D}$ is a matrix approximating the operator of derivation.

Show that this solution is given by:

$$\widehat{\boldsymbol{f}}_2 = \arg\min_{\boldsymbol{f}} \{J_2(\boldsymbol{f})\} = \left[\boldsymbol{H}'\boldsymbol{H} + \lambda\boldsymbol{D}'\boldsymbol{D}\right]^{-1}\boldsymbol{H}'\boldsymbol{g}$$

Why this solution is prefered to $\widehat{\boldsymbol{f}}_0$ and to $\widehat{\boldsymbol{f}}_1$ and $\widehat{\boldsymbol{f}}_2$?

11. Suppose that $\boldsymbol{H}$ and $\boldsymbol{D}$ be circulent matrices and symmetric. Then, show that the regularised solution $\widehat{\boldsymbol{f}}_2$ can be written using the DFT by:

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

where

- $H(\omega)$ is the DFT of the first ligne of the matrix $\boldsymbol{H}$,
- $D(\omega)$ is the DFT of the first ligne of the matrix $\boldsymbol{D}$
- $F(\omega)$ is the DFT of the solution vector $\widehat{\boldsymbol{f}}_2$, et
- $G(\omega)$is the DFT of the data measurement vector $\boldsymbol{g}$.

12. Comment the expressions of $\widehat{\boldsymbol{f}}_2$ in the question 3. and $F(\omega)$ in the question 4. when $\lambda = 0$ and when $\lambda \longrightarrow \infty$.

Part 2:

Consider the problem $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f}$. We are looking the solution $\widehat{\boldsymbol{f}}$ for this inverse problem in such a way that $\widehat{\boldsymbol{f}} = \boldsymbol{M}\boldsymbol{g}$, i.e.; a linear function of the data $\boldsymbol{g}$. We are then looking for the matrix $\boldsymbol{M}$.

1. Suppose first that the solution $\boldsymbol{f}^*$ exists, i.e. $\boldsymbol{H}\boldsymbol{f}^* = \boldsymbol{g}$. Then,

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

   The matrix $\boldsymbol{R} - \boldsymbol{M}\boldsymbol{H}$ measures the *resolution power in the space of the solutions* of the inverse operator $\boldsymbol{M}$. The ideal case is $\boldsymbol{R} = \boldsymbol{I}$, i.e.; $\boldsymbol{M} = \boldsymbol{H}^{-1}$, but this almost never possible, and when possible, probably not useful due to the ill-condition nature of $\boldsymbol{H}$. So, let look for $\boldsymbol{M}$ such that:

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

   be minimal. Show then that the solution is:

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

2. A second argument is to search for $\boldsymbol{M}$ such that $\widehat{\boldsymbol{g}} = \boldsymbol{H}\boldsymbol{f}^* = \boldsymbol{H}\boldsymbol{M}\boldsymbol{g} = \boldsymbol{N}\boldsymbol{g}$ be as close as possible to $\boldsymbol{g}$. The matrix $\boldsymbol{N} - \boldsymbol{H}\boldsymbol{M}$ measures the *resolution power in the space of the observations* of the operator $\boldsymbol{M}$. The ideal case is $\boldsymbol{N} = \boldsymbol{I}$, i.e.; $\boldsymbol{M} = \boldsymbol{H}^{-1}$, which is again either impossible or not desirable. So, let look for $\boldsymbol{M}$ such that

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

   be minimal. Show then the solution is:

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

3. A third argument is based on the fact that: $\text{cov}[\widehat{\boldsymbol{f}}] = \text{cov}[\boldsymbol{M}\boldsymbol{g}] = \boldsymbol{M}\text{cov}[\boldsymbol{g}]\boldsymbol{M}^t$ and if $\text{cov}[\boldsymbol{g}] = \boldsymbol{I}$ we have $\text{cov}[\widehat{\boldsymbol{f}}] = \boldsymbol{M}\boldsymbol{M}^t$. The ideal case for $\widehat{\boldsymbol{f}}$ is that this covarience be close to $\boldsymbol{I}$. We can then would like to define:

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

   which can also be used as a constraint for $\boldsymbol{M}$. Write the expression of $\frac{\partial J_3}{\partial \boldsymbol{M}}$.

4. Now, let define $J(\boldsymbol{M}) = \alpha_1 J_1(\boldsymbol{M}) + \alpha_2 J_2(\boldsymbol{M}) + \alpha_3 J_2(\boldsymbol{M})$. Write the expression of $\frac{\partial J}{\partial \boldsymbol{M}}$ and find the matrix $\boldsymbol{M}$ minimizes $J(\boldsymbol{M})$ pour différentes combinaisons de $(\alpha_1, \alpha_2, \alpha_3)$. For each case, give also the expressions of $\boldsymbol{R}$, $\boldsymbol{N}$ and $\boldsymbol{U}$.

   Check then the content of the following table:

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

Inverse Problems in Signal Processing, Imaging Systems and Computer Vision

Professor: A. Mohammad–Djafari

Exercice number 6: Maximum Entropy

Part 1:

Consider the problem $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f}$ and Suppose that $\boldsymbol{f}$ represents an image with $f_j \geq 0$ and $\sum_j f_j = 1$. Suppose that the system $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f}$ is underdetermined and that we are looking to choose one solution between all the possible solutions, the one which maximizes the Shannon Entropy

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

1. Show that this solution, if exists, is given by:

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

   where $\boldsymbol{\lambda} = \{\lambda_i, i = 1, \cdots, M\}$ is obtained by solving the following system of equations:

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

   and where λ_0 is a constant which can be determined if we impose $\sum_j f_j = 1$.

2. Show that this solution can also be written as::

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

   where

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

   which results to:

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

   and where $\boldsymbol{\lambda}$ is obtained by:

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

3. Show also that: $\boldsymbol{\lambda}$ is obtained by optimizing the dual criterion $D(\boldsymbol{\lambda}) = \|\boldsymbol{g} - \boldsymbol{H}\exp\left\{-\boldsymbol{H}^t\boldsymbol{\lambda}\right\}\|^2$.

4. Write an iterative algorithem (Gradient based or Newton) which can compute this solution. Give the details and the cost of the computations at each iteration. For this, I suggest you to use two programs `g=direct(H,f)` which aims to compute $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f}$ and `f=transp(H,g)` which aims to compute $\boldsymbol{f} = \boldsymbol{H}^t\boldsymbol{g}$.

Part 2:

Consider the problem $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f}$ and suppose that $\boldsymbol{f}$ is an expected image, *i.e.* each f_j corresponds to the expected value of some quantity Z_j. The data g_i can then be considered as a linear combination of $\mathrm{E}\{Z_j\}$, *i.e.*;

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

Give the expression of the probability law $p(\boldsymbol{z})$ which satisfies the constraintes and maximizes

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

where $\mu(\boldsymbol{z})$ is a reference probability density function (a priori).

Show then that when $\mu(\boldsymbol{z})$ est separable, *i.e.*; $\mu(\boldsymbol{z}) = \prod_{j=1}^{N} \mu(z_j)$, then $p(\boldsymbol{z})$ is also separable, *i.e.*;

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

Noting that $f_j = \mathrm{E}\{Z_j\}$, show that:

1. If $\mu(z_j) \propto \exp\left\{-\frac{1}{2} z_j^2\right\}$, then we have

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

   and consequently

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

   where $\{\lambda_i\}$ are the solution of

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

   where assuming $\boldsymbol{H}\boldsymbol{H}'$ is invertible, we get:

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

   Give your interpretation of this result.

2. If $\mu(z_j) \propto \exp\{-z_j\}, \quad z_j > 0$, then we have:

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

   and consequently

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

where $\{\lambda_i\}$ are given by

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

and if $\boldsymbol{H}\boldsymbol{H}'$ is invertible, we get:

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

Give your interpretation of this result.

3. If $\mu(z_j) \propto z_j^{(\alpha-1)} = \exp\{(\alpha-1)\ln z_j\}, \quad z_j > 0$, then

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

which is the Gamma distribution: $Gamma(\alpha,\beta)$ with $\beta = \sum_{i=1}^{M} H_{i,j}\lambda_i$ and consequently

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

where $\{\lambda_i\}$ are given by:

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

where, using the Matlab notation, we can write:

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

Give your interpretation of this result.

Inverse Problems in Signal Processing, Imaging Systems and Computer Vision

Professor: A. Mohammad–Djafari

Exercice number 7: Maximum Likelihood Estimation and Bayesian inference

Consider the system $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}$

1. Assume that the noise $\boldsymbol{\epsilon}$ can be modelled by a centered, white Gaussian with fixed varience σ_ϵ^2. Show that the ML estimate of $\boldsymbol{f}$, *i.e.*;

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

can be obtained by minimizing:

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

2. What would be the expression of this estimator if $\boldsymbol{\epsilon}$ was following a Generalized Gaussian law, *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. To apply the Bayesian inference approach, assume that our prior knowledge on $\boldsymbol{\epsilon}$ and $\boldsymbol{f}$ could be translated by:

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

where $\boldsymbol{R}_\epsilon = \sigma_\epsilon^2 \boldsymbol{I}$ and $\boldsymbol{R}_f = \sigma_f^2 \boldsymbol{P}_0 = \sigma_f^2 (\boldsymbol{D}'\boldsymbol{D})^{-1}$ the covarience matrices of $\boldsymbol{\epsilon}$ and $\boldsymbol{f}$. Write the expressions of $p(\boldsymbol{f}), p(\boldsymbol{\epsilon}), p(\boldsymbol{g}, \boldsymbol{f})$ and $p(\boldsymbol{g}|\boldsymbol{f})$.

4. Show that the *a posteriori* law $p(\boldsymbol{f}|\boldsymbol{g})$ is a Gaussian in the form:

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

5. If we note by $\widehat{\boldsymbol{f}}_3$ the solution which maximizes $p(\boldsymbol{f}|\boldsymbol{g})$ (Maximum A Posteriori (MAP) estimate), then show that it can obtained by:

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

6. If $\boldsymbol{R}_\epsilon = \sigma_\epsilon^2 \boldsymbol{I}$ and $\boldsymbol{R}_f = \sigma_f^2 \boldsymbol{P}_0$, then show that $\widehat{\boldsymbol{f}}_3$ can be obtained by:

$$\widehat{\boldsymbol{f}}_3 = \left[\boldsymbol{H}'\boldsymbol{H} + \lambda \boldsymbol{P}_0^{-1}\right]^{-1} \boldsymbol{H}'\boldsymbol{g} \quad \text{with} \quad \lambda = \frac{\sigma_\epsilon^2}{\sigma_f^2}$$

What can we conclude if we compare the solutions $\widehat{\boldsymbol{f}}_2$ and $\widehat{\boldsymbol{f}}_3$?

7. Developping the terms:

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

in the expression of $p(\boldsymbol{f}|\boldsymbol{g})$, show that:

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

hat is then the expression of the *a posteriori* covarience matrix $\widehat{\boldsymbol{P}}$?
What represent the diagonal elements of this matrix ?

8. Write the expressions of $p(f_n|\boldsymbol{g}), p(g_m|\boldsymbol{g})$ and $p(b_m|\boldsymbol{g})$. What can be used for?

9. Consider now that the elements $\{f_1,\cdots,f_N\}$ can be modelled as first order Markov chain:

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

Can always calculate $p(\boldsymbol{f})$? and if indeed we know $p(f_1)$?
What becomes then MAP estimate in this case?
Study this solution in the following cases:

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

and

$$p(f_n|f_1,\cdots,f_{n-1},f_{n+1},\cdots,f_N)=p(f_n|f_{n-1})\propto\exp\left\{-\alpha\phi(f_n-f_{n-1})\right\}, \quad \text{and} \quad p(f_1)\propto\exp\left\{-\alpha\phi(f_1)\right\}$$

Inverse Problems in Signal Processing, Imaging Systems and Computer Vision

Professor: A. Mohammad–Djafari

Exercice number 8: Bayesian Computation in the case of Linear model and Gaussian laws:

First, remember the following relations:

$$\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma}) = (2\pi)^{-d/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}$$

$$\int (\boldsymbol{x}-\boldsymbol{\mu})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\mathcal{N}(\boldsymbol{m},\boldsymbol{S})\,\mathrm{d}\boldsymbol{x} = (\boldsymbol{\mu}-\boldsymbol{m})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\boldsymbol{m}) + \mathrm{Tr}\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\right]$$

$$\int (\boldsymbol{W}\boldsymbol{x}-\boldsymbol{\mu})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{W}\boldsymbol{x}-\boldsymbol{\mu})\mathcal{N}(\boldsymbol{m},\boldsymbol{S})\,\mathrm{d}\boldsymbol{x} = (\boldsymbol{\mu}-\boldsymbol{W}\boldsymbol{m})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\boldsymbol{W}\boldsymbol{m}) + \mathrm{Tr}\left[\boldsymbol{W}^t\boldsymbol{\Sigma}^{-1}\boldsymbol{W}\boldsymbol{S}\right]$$

$$\mathcal{N}(\boldsymbol{a},\boldsymbol{A}) * \mathcal{N}(\boldsymbol{b},\boldsymbol{B}) \propto \mathcal{N}(\boldsymbol{a}+\boldsymbol{b},\boldsymbol{A}+\boldsymbol{B})$$

$$\mathcal{N}(\boldsymbol{a},\boldsymbol{A})\mathcal{N}(\boldsymbol{b},\boldsymbol{B}) \propto \mathcal{N}(\boldsymbol{c},\boldsymbol{C}) \quad \text{with} \quad \boldsymbol{C} = (\boldsymbol{A}^{-1}+\boldsymbol{B}^{-1})^{-1} \quad \text{and} \quad \boldsymbol{c} = \boldsymbol{C}[\boldsymbol{A}^{-1}\boldsymbol{a}+\boldsymbol{B}^{-1}\boldsymbol{b}]$$

If

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

then

$$\left\{\begin{array}{l}\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{a},\boldsymbol{A})\\ \boldsymbol{y} \sim \mathcal{N}(\boldsymbol{b},\boldsymbol{B})\end{array}\right. \quad \text{and} \quad \left\{\begin{array}{l}\boldsymbol{x}|\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{a}+\boldsymbol{C}\boldsymbol{B}^{-1}(\boldsymbol{y}-\boldsymbol{b}),\boldsymbol{A}-\boldsymbol{C}\boldsymbol{B}^{-1}\boldsymbol{C}^t)\\ \boldsymbol{y}|\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{b}+\boldsymbol{C}^t\boldsymbol{A}^{-1}(\boldsymbol{x}-\boldsymbol{a}),\boldsymbol{B}-\boldsymbol{C}^t\boldsymbol{A}^{-1}\boldsymbol{C})\end{array}\right.$$

Consider now the vectors $\boldsymbol{g}$, $\boldsymbol{f}$ and $\boldsymbol{\epsilon}$ linked by the linear relation $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f}+\boldsymbol{\epsilon}$ where we assume $\boldsymbol{\epsilon} \sim \mathcal{N}(0,\sigma_\epsilon^2\boldsymbol{I})$ and $\boldsymbol{f} \sim \mathcal{N}(\boldsymbol{f}_0,\boldsymbol{P}_0)$ with $\boldsymbol{P}_0 = (\boldsymbol{D}^t\boldsymbol{D})^{-1}$.

1. Write the expressions of $p(\boldsymbol{\epsilon})$, $p(\boldsymbol{g}|\boldsymbol{f})$, $p(\boldsymbol{f})$ and $p(\boldsymbol{g})$
2. Write the expressions of $p(\boldsymbol{g},\boldsymbol{f})$
3. Write the expressions of $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ in the following relations:

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

   and deduce the expressions of $p(\boldsymbol{f}|\boldsymbol{g})$
4. Write the expressions of la moyenne and de la covarience *a posteriori* $\widehat{\boldsymbol{f}}$ and $\widehat{\boldsymbol{\Sigma}}$.
5. Show that $\widehat{\boldsymbol{f}}$ can be obtained by minimizing: $\quad J(\boldsymbol{f}) = \frac{1}{\sigma_\epsilon^2}\|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\|^2+\|\boldsymbol{D}(\boldsymbol{f}-\boldsymbol{f}_0)\|^2$

Define now two vectors $\boldsymbol{f}_1 = \boldsymbol{f}_0+\boldsymbol{D}^{-1}\boldsymbol{\epsilon}_1$ and $\boldsymbol{g}_1 = \boldsymbol{g}+\sigma_\epsilon\boldsymbol{\epsilon}_2$ where $\boldsymbol{\epsilon}_1$ and $\boldsymbol{\epsilon}_2$ are two Gaussien vectors $\boldsymbol{\epsilon}_1 \sim \mathcal{N}(0,\boldsymbol{I})$ and $\boldsymbol{\epsilon}_2 \sim \mathcal{N}(0,\boldsymbol{I})$ and $\boldsymbol{x} = \widehat{\boldsymbol{\Sigma}}\left(\frac{1}{\sigma_\epsilon^2}\boldsymbol{H}^t\boldsymbol{g}_1+\boldsymbol{D}^t\boldsymbol{D}\boldsymbol{f}_1\right)$.

1. Show that $\boldsymbol{x} = \widehat{\boldsymbol{f}}+\widehat{\boldsymbol{\Sigma}}\left(\frac{1}{\sigma_\epsilon^2}\boldsymbol{H}^t\boldsymbol{\epsilon}_2+\boldsymbol{D}^t\boldsymbol{D}\boldsymbol{\epsilon}_1\right)$.
2. Show that $\mathrm{E}\{\boldsymbol{x}\} = \widehat{\boldsymbol{f}}$ and $\mathrm{Cov}[\boldsymbol{x}] = \widehat{\boldsymbol{\Sigma}}$.

3. Show that to generate a sample of the *a posteriori* $p(\boldsymbol{f}|\boldsymbol{g})$, we can use the following algorithems:
   - Algorithem 1:
     - Decompose the covarience matrix *a posteriori* $\widehat{\boldsymbol{\Sigma}} = \boldsymbol{W}\boldsymbol{W}^t$
     - Generate a Gaussian sample $\boldsymbol{\epsilon}_1 \sim \mathcal{N}(0, \boldsymbol{I})$
     - Compute $\widehat{\boldsymbol{f}} = \boldsymbol{W}\boldsymbol{\epsilon}_1$
   - Algorithem 2:
     - Generate two independent Gaussian vectors $\boldsymbol{\epsilon}_1 \sim \mathcal{N}(0, \boldsymbol{I})$ and $\boldsymbol{\epsilon}_2 \sim \mathcal{N}(0, \boldsymbol{I})$
     - Generate two vectors $\boldsymbol{f}_1 = \boldsymbol{f}_0 + \boldsymbol{D}^{-1}\boldsymbol{\epsilon}_1$ and $\boldsymbol{g}_1 = \boldsymbol{g} + \sigma_\epsilon \boldsymbol{\epsilon}_2$
     - Compute $\widehat{\boldsymbol{f}}$ by minimizing: $J(\boldsymbol{f}) = \frac{1}{\sigma_\epsilon^2}\|\boldsymbol{g}_1 - \boldsymbol{H}\boldsymbol{f}_1\|^2 + \|\boldsymbol{D}\boldsymbol{f}_1\|^2$
4. Compare the computational cost and memory needs of these two algorithems.

Inverse Problems in Signal Processing, Imaging Systems and Computer Vision

Professor: A. Mohammad–Djafari

Exercice number 9: Modelisation of the signals and Bayesian MAP computation

Part 1: Simple Models
We have seen that the computation of the MAP solution of $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}$ becomes an Optimization of the criterion: $J(\boldsymbol{f}) = \sigma_\epsilon^{-2}\|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + \Omega(\boldsymbol{f})$. Then, when assuming that the prior is separable and i.i.d. (which means: $p(\boldsymbol{f}) = \prod_j p(f_j), \forall j$), the second term becomes: $\Omega(\boldsymbol{f}) = -\sum_j \ln p(f_j)$.

1. Check then the following relations:
   Gaussian: $p(f_j|\lambda) = \mathcal{N}(0, \frac{1}{2\lambda})$
   $$p(f_j|\lambda) = \sqrt{\frac{\lambda}{\pi}} \exp\left\{-\lambda|f_j|^2\right\} \quad \longrightarrow \quad \Omega(\boldsymbol{f}) = \lambda \sum_j |f_j|^2$$
   $\Omega(\boldsymbol{f})$ is quadratic in $\boldsymbol{f}$.
   Double Exponential (Laplace) law: $p(f_j) = \mathcal{E}(\lambda)$
   $$p(f_j|\lambda) = \frac{\lambda}{2} \exp\left\{-\lambda|f_j|\right\} \quad \longrightarrow \quad \Omega(\boldsymbol{f}) = \lambda \sum_j |f_j|$$
   $\Omega(\boldsymbol{f})$ is no more a continuous function of $\boldsymbol{f}$.
   Generalized Exponential: $p(f_j) = \mathcal{GE}(\lambda, \beta)$
   $$p(f_j|\lambda, \beta) \propto \exp\left\{-\lambda|f_j|^\beta\right\} \quad \longrightarrow \quad \Omega(\boldsymbol{f}) = \lambda \sum_j |f_j|^\beta$$
   $\Omega(\boldsymbol{f})$ is continuous function of $\boldsymbol{f}$ if $1 < \beta \leq 2$.

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

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

$\Omega(\boldsymbol{f})$ is not a convex function of $\boldsymbol{f}$.

2. In each of the above cases, give the expression of the gradient of $\Omega(\boldsymbol{f})$ when possible and propose an appropriate optimization algorithem for computing the solution $\widehat{\boldsymbol{f}}$ which optimizes $J(\boldsymbol{f})$.

3. Consider now the signal denoising problem $g(t) = f(t) + \epsilon(t)$ which can also be written as $\boldsymbol{g} = \boldsymbol{f} + \boldsymbol{\epsilon}$ and consider the criterion:$J(\boldsymbol{f}) = \sigma_\epsilon^{-2}\|\boldsymbol{g} - \boldsymbol{f}\|^2 + \Omega(\boldsymbol{f})$. Show that, with the above prior laws, we have: $\widehat{f}(t) = \phi\left(g(t)\right)$ ou $\widehat{f_j} = \phi(g_j)$. Give then the expression of the function $\phi(.)$ and plot the curve $\widehat{f_j} = \phi(g_j)$.

4. Now, suppose that, in place of modeling the signal samples f_j by $p(f_j)$, we model $\Delta_j = (f_j - f_{j-1})$ by $p(\Delta_j)$ using the above probability laws, and indeed assume that f_0 follows the same class of prior law $p(f_0)$. Show then that, this time, we will have: $\widehat{f_j} = \phi(g_j - g_{j-1})$.

5. Suppose now that f_j are modeled by a Moving Average (MA) process of ordre K: $f_j = \sum_{k=0}^{K-1} h_k \eta_{j-k}$ where we suppose that η_j are i.i.d. and follow one of the above mentionned prior laws. Give then, the expressions of $p(\boldsymbol{f})$ and the corresponding MAP estimate in each case.

6. Suppose now f_j is modelled by an Auto-Regressif (AR) process of ordre K: $f_j = \sum_{k=1}^{K} a_k f_{j-k} + \eta_j$ where η_j are assumed i.i.d. and ag ain following one of the above-mentionned priors. une des lois proposées. Give then, the expressions of $p(\boldsymbol{f})$ and the corresponding MAP estimate in each case. Start by the simple case of $K = 1$ and $a_1 = 1$.

7. Simulate a signal $f(t) = 0,\ t \in [0, 100]$ and $f(t) = 2,\ \in t = [100, 200]$ and to it a Gaussian noise with varience $\sigma_\epsilon^2 = 1$ to obtain $g(t)$. Apply then the signal denoising methods we just proposed with different prior models and compare the obtained resultats. Comment them.

Part 2: Modelling using hidden variables

We have seen in the course that the introduction of the hidden variables give more possibilities for modeling signals and images. In these models, we introduce a hidden variable, noted $\boldsymbol{c}$, $\boldsymbol{d}$, $\boldsymbol{q}$ or $\boldsymbol{z}$ as functions of their nature ($\boldsymbol{d}$ for real valued variables, $\boldsymbol{q}$ for binary valued,$\boldsymbol{c}$ for ternary valued, $\boldsymbol{z}$ for finite discrete valued). When introduced, we have to estimate them jointly with $\boldsymbol{f}$. This is done, in general, via an iterative algorithem which consists in estimation iteratively $\boldsymbol{f}$ given $\boldsymbol{c}$ and $\boldsymbol{c}$ given $\boldsymbol{f}$.

Modulated varience models:

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

- Check the following relations::

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

$$p(\boldsymbol{g}|\boldsymbol{f}) \propto \exp\left\{\sigma_\epsilon^{-2}\|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\|^2\right\} \quad \longrightarrow \quad p(\boldsymbol{f},\boldsymbol{d}|\boldsymbol{g}) \propto \exp\left\{-J(\boldsymbol{f},\boldsymbol{d})\right\}$$

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

- Optimizing $\boldsymbol{f}$ given $\boldsymbol{d}$:

$$\widehat{\boldsymbol{f}}=(\sigma_\epsilon^{-2}\boldsymbol{H}^t\boldsymbol{H}+2\lambda\boldsymbol{D})^{-1}\boldsymbol{H}^t\boldsymbol{g} \quad \text{with} \quad \boldsymbol{D}=\operatorname{diag}\left[1/(4d_j), j=1,\cdots,n\right]$$

- Optimizing $\boldsymbol{d}$ given $\boldsymbol{f}$:

$$\widehat{d_j}=f_j/2$$

- Check the following relations:

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

Modulated mean Gaussian model:

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

- Check the following relations::

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

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

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

$$p(\boldsymbol{g}|\boldsymbol{f}) \propto \exp\left\{\sigma_\epsilon^{-2}\|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\|^2\right\} \quad \longrightarrow \quad p(\boldsymbol{f},\boldsymbol{z}|\boldsymbol{g}) \propto \exp\left\{-J(\boldsymbol{f},\boldsymbol{z})\right\}$$

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

where $\boldsymbol{z}=[z_1,\cdots,z_N]'$.

- Optimization of $\boldsymbol{f}$ given $\boldsymbol{z}$:

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

- Optimization of $\boldsymbol{z}$ given $\boldsymbol{f}$:

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

- Give the expression of a as a function function of λ. For this, plot the expression of $p(z_j|f_j, \lambda)$ as a function of f_j.

- Generalize this model for the case:

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

Inverse Problems in Signal Processing, Imaging Systems and Computer Vision

Professor: A. Mohammad–Djafari

Exercice number 10: Bayesian Computation

Part 1: Basics of Variational Computation
Let assume to have a model $\mathcal{M}$, a distribution $P(\boldsymbol{X})$ over the variables $\boldsymbol{X}$ which is divided in three parts: observed variables $\boldsymbol{g}$, hidden variables $\boldsymbol{f}$ and parameters $\boldsymbol{\theta}$. During the learning step (identification), we assume to have $\boldsymbol{D} = \{\boldsymbol{g}, \boldsymbol{f}\}$ and we assume that the model is valid and want to estimate $\boldsymbol{\theta}$ using the Bayes rule:

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

or still:

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

During the inversion, we assume $\boldsymbol{g}$ is observed, the model $\mathcal{M}$ is valid, the parameters $\boldsymbol{\theta}$ are estimated and we want to infer on the hidden variables $\boldsymbol{f}$. Using again the Bayes rule, we get:

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

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

  $$P(\mathcal{M}_i|\boldsymbol{D}) = \frac{P(\boldsymbol{D}|\mathcal{M}_i)\ p(\mathcal{M}_i)}{p(\boldsymbol{D})}$$

  For model comparison, we need the expression of the evidence $P(\mathcal{M}|\boldsymbol{D})$. Obtaining an analytical expression for it is often very difficult. We then try to approximate it by a simpler expression $Q(\mathcal{M})$, but as close as possible. We may use the Kulback-Leibler (KL) divergence:

  $$KL(Q, P) = \int_{\mathcal{M}} Q(\mathcal{M}) \ln \frac{Q(\mathcal{M})}{P(\mathcal{M}|\boldsymbol{D})}\, \mathrm{d}\mathcal{M}$$

  as a criterion for measuring their closedness.

  1. Show that:

     $$KL(Q, P) = \int_{\mathcal{M}} Q(\mathcal{M}) \ln \frac{Q(\mathcal{M})}{P(\mathcal{M}, \boldsymbol{D})}\, \mathrm{d}\mathcal{M} + \ln P(\boldsymbol{D})$$

  2. Noting:

     $$\mathcal{F}(Q) = \int_{\mathcal{M}} Q(\mathcal{M}) \ln \frac{P(\mathcal{M}, \boldsymbol{D})}{Q(\mathcal{M})}\, \mathrm{d}\mathcal{M}$$

     show that:

     $$\mathcal{F}(Q) = < \ln P(\mathcal{M}, \boldsymbol{D}) >_{Q(\mathcal{M})} + H(Q)$$

     where

     $$H(Q) = -\int_{\mathcal{M}} Q(\mathcal{M}) \ln Q(\mathcal{M})\, \mathrm{d}\mathcal{M}$$

     and

     $$< \ln P(\mathcal{M}, \boldsymbol{D}) >_{Q(\mathcal{M})} = \int_{\mathcal{M}} Q(\mathcal{M}) \ln P(\mathcal{M}, \boldsymbol{D})\, \mathrm{d}\mathcal{M}$$

3. Show that:

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

which means that $\mathcal{F}(Q)$ is an inferior limit of $\ln P(\boldsymbol{D})$.

4. Show that at the optimum $P(\boldsymbol{D}|\mathcal{M}_i)$ can be approximated by:

$$P(\boldsymbol{D}|\mathcal{M}_i) \approx \frac{\exp\{\mathcal{F}(Q^*)\}\; p(\mathcal{M}_i)}{p(\boldsymbol{D})}$$

5. Show that, if we choose: $Q(\mathcal{M}) = \prod_j Q_i(\mathcal{M}_i)$ we get

$$\mathcal{F}(Q) = \int_{\mathcal{M}_j} Q_j(\mathcal{M}_j) < \ln P(\mathcal{M},\boldsymbol{D}) >_{\prod_{i\neq j} Q_i(\mathcal{M}_i)} + H(Q_j) + \sum_{i\neq j} H(Q_i)$$

and if we optimize $\mathcal{F}(Q)$ with respect to Q_j we get

$$Q_j(\mathcal{M}_j) \propto \exp\left\{< \ln P(\mathcal{M},\boldsymbol{D}) >_{\prod_{i\neq j}} Q_i(\mathcal{M}_i)\right\}$$

Part 2: Application to the case: $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}$ when we want to estimate $\boldsymbol{f}$ and $\boldsymbol{\theta}$.

- Writing the joint posterior:

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

and trying to approximate it by a separable $q(\boldsymbol{f},\boldsymbol{\theta}|\boldsymbol{g}) = q_1(\boldsymbol{f})\; q_2(\boldsymbol{\theta})$:

1. Assume:

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

Then, write the expressions of $p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{\theta})$, $p(\boldsymbol{f},\boldsymbol{\theta}|\boldsymbol{g})$, $p(\boldsymbol{f}|\boldsymbol{g},\boldsymbol{\theta})$, and $p(\boldsymbol{\theta}|\boldsymbol{g},\boldsymbol{f})$.

2. Write the expressions of $KL(q,p)$, $\mathcal{F}(q)$, $\ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{\theta})$, $< \ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{\theta}) >_{q_1(\boldsymbol{f})}$ and $< \ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{\theta}) >_{q_2(\boldsymbol{\theta})}$.

3. Choosing q_1 and q_2 such that: $q_1(\boldsymbol{f}) = \delta(\boldsymbol{f} - \widehat{\boldsymbol{f}})$ and $q_2(\boldsymbol{\theta}) = \delta(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})$, Write the expressions of $< \ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{\theta}) >_{q_1(\boldsymbol{f})}$ and $< \ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{\theta}) >_{q_2(\boldsymbol{\theta})}$.

   Give then the expressions of $\widehat{\boldsymbol{f}}$ and $\widehat{\boldsymbol{\theta}}$ during the iterations (Link with Joint MAP).

4. Choosing $q_1(\boldsymbol{f})$ in the same family than $p(\boldsymbol{f}|\boldsymbol{g},\boldsymbol{\theta})$ and $q_2(\boldsymbol{\theta}) = \delta(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})$, write the expressions of $< \ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{\theta}) >_{q_1(\boldsymbol{f})}$ and $< \ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{\theta}) >_{q_2(\boldsymbol{\theta})}$.

   Give then the expressions of $\widehat{\boldsymbol{f}}$ and $\widehat{\boldsymbol{\theta}}$ during the iterations (Link with Expectation-Maximization algorithem).

5. Choosing $q_1(\boldsymbol{f})$ in the same family than $p(\boldsymbol{f}|\boldsymbol{g},\boldsymbol{\theta})$ and $q_2(\boldsymbol{\theta})$ in the same family than $p(\boldsymbol{\theta}|\boldsymbol{g},\boldsymbol{f})$, write the expressions of $<\ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{\theta})>_{q_1(\boldsymbol{f})}$ and $<\ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{\theta})>_{q_2(\boldsymbol{\theta})}$.
   Give then the expressions of $q_1(\boldsymbol{f})$ and $q_2(\boldsymbol{\theta})$ during the iterations (Link with EM algorithem).
6. Choosing $q_1(\boldsymbol{f}) = \mathcal{N}(\widehat{\boldsymbol{f}},\widehat{\boldsymbol{\Sigma}})$ and $q_2(\boldsymbol{\theta}) = \mathcal{G}(\widehat{\alpha}_{10},\widehat{\beta}_{10})\mathcal{G}(\widehat{\alpha}_{20},\widehat{\beta}_{20})$, write the expressions of $(\widehat{\boldsymbol{f}},\widehat{\boldsymbol{\Sigma}}),(\widehat{\alpha}_{10},\widehat{\beta}_{10}),(\widehat{\alpha}_{20},\widehat{\beta}_{20})$ and en déduire $<\boldsymbol{f}>_{q_1}$, $<\theta_1>_{q_1}$, $<\theta_2>_{q_1}$,

- In the previous example, we model now $\boldsymbol{f}$ by a Gaussian field with modulated varience explained in previous exercises:

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

  The objective now is to estimate $\boldsymbol{f}$, $\boldsymbol{z}$ and $\boldsymbol{\theta}$.

  First, we write the expression of the Joint posterior:

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

  and then, we try to approximate it by $q(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta}|\boldsymbol{g}) = q_1(\boldsymbol{f})\; q_2(\boldsymbol{z})\; q_3(\boldsymbol{\theta})$.

  1. Write the expressions of $p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{d},\boldsymbol{\theta})$, $p(\boldsymbol{f},\boldsymbol{d},\boldsymbol{\theta}|\boldsymbol{g})$, $p(\boldsymbol{f}|\boldsymbol{g},\boldsymbol{d},\boldsymbol{\theta})$, $p(\boldsymbol{d}|\boldsymbol{g},\boldsymbol{f},\boldsymbol{\theta})$ and $p(\boldsymbol{\theta}|\boldsymbol{g},\boldsymbol{f},\boldsymbol{d})$.
  2. Write the expressions of $KL(q,p)$, $\mathcal{F}(q)$, $\ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{\theta})$, $<\ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{d},\boldsymbol{\theta})>_{q_1(\boldsymbol{f})}$ and $<\ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{d},\boldsymbol{\theta})>_{q_2(\boldsymbol{d})}$ and $<\ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{\theta})>_{q_3(\boldsymbol{\theta})}$.
  3. Choosing $q_1(\boldsymbol{f}) = \delta(\boldsymbol{f}-\widehat{\boldsymbol{f}})$, $q_2(\boldsymbol{d}) = \delta(\boldsymbol{d}-\widehat{\boldsymbol{d}})$ and $q_3(\boldsymbol{\theta}) = \delta(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}})$, write the expressions of $<\ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{d},\boldsymbol{\theta})>_{q_1(\boldsymbol{f})}$, $<\ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{d},\boldsymbol{\theta})>_{q_2(\boldsymbol{d})}$ and $<\ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{d},\boldsymbol{\theta})>_{q_3(\boldsymbol{\theta})}$.
     Give the expressions of $\widehat{\boldsymbol{f}}$, $\widehat{\boldsymbol{d}}$ and $\widehat{\boldsymbol{\theta}}$ during the iterations (Link with Joint MAP).
  4. Choosing $q_1(\boldsymbol{f})$ in the same family than $p(\boldsymbol{f}|\boldsymbol{g},\boldsymbol{d},\boldsymbol{\theta})$, $q_2(\boldsymbol{d})$ in the same family than $p(\boldsymbol{d}|\boldsymbol{g},\boldsymbol{f},\boldsymbol{\theta})$ and $q_3(\boldsymbol{\theta}) = \delta(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}})$, write the expressions of $<\ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{d},\boldsymbol{\theta})>_{q_1(\boldsymbol{f})}$, $<\ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{d},\boldsymbol{\theta})>_{q_2(\boldsymbol{d})}$ and $<\ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{d},\boldsymbol{\theta})>_{q_3(\boldsymbol{\theta})}$.
     Give the expressions of $\widehat{\boldsymbol{f}}$, $\widehat{\boldsymbol{d}}$ and $\widehat{\boldsymbol{\theta}}$ during the iterations (Link with EM algorithem).
  5. Choosing $q_1(\boldsymbol{f})$ in the same family than $p(\boldsymbol{f}|\boldsymbol{g},\boldsymbol{d},\boldsymbol{\theta})$, $q_2(\boldsymbol{d})$ in the same family than $p(\boldsymbol{d}|\boldsymbol{g},\boldsymbol{f},\boldsymbol{\theta})$ and $q_3(\boldsymbol{\theta})$ in the same family than $p(\boldsymbol{\theta}|\boldsymbol{g},\boldsymbol{f},\boldsymbol{d})$, write the expressions of $<\ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{d},\boldsymbol{\theta})>_{q_1(\boldsymbol{f})}$, $<\ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{d},\boldsymbol{\theta})>_{q_2(\boldsymbol{d})}$ and $<\ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{d},\boldsymbol{\theta})>_{q_3(\boldsymbol{\theta})}$.
     Give the expressions of $q_1(\boldsymbol{f})$, $q_2(\boldsymbol{d})$ and $q_3(\boldsymbol{\theta})$ during the iterations.
     Compare this algorithem with EM algorithem.

- Dans l'exemple précédent, nous avons modélisé $\boldsymbol{f}$ par un champs gaussien. Prenons now le modèle gaussien à moyenne modulable de l'exercise précédent:

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

The objective is then to estimate $\boldsymbol{f}$, $\boldsymbol{z}$ and $\boldsymbol{\theta}$. Writing the joint posterior:

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

and trying to approximate it by $q(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta}|\boldsymbol{g}) = q_1(\boldsymbol{f})\; q_2(\boldsymbol{z})\; q_3(\boldsymbol{\theta})$.

1. Write the expressions of $p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta})$, $p(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta}|\boldsymbol{g})$, $p(\boldsymbol{f}|\boldsymbol{g},\boldsymbol{z},\boldsymbol{\theta})$, $p(\boldsymbol{z}|\boldsymbol{g},\boldsymbol{f},\boldsymbol{\theta})$ and $p(\boldsymbol{\theta}|\boldsymbol{g},\boldsymbol{f},\boldsymbol{z})$.
2. Write the expressions of $KL(q,p)$, $\mathcal{F}(q)$, $\ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{\theta})$, $< \ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta}) >_{q_1(\boldsymbol{f})}$ and
   $< \ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta}) >_{q_2(\boldsymbol{z})}$ and $< \ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{\theta}) >_{q_3(\boldsymbol{\theta})}$.
3. Choosing $q_1(\boldsymbol{f}) = \delta(\boldsymbol{f}-\widehat{\boldsymbol{f}})$, $q_2(\boldsymbol{z}) = \delta(\boldsymbol{z}-\widehat{\boldsymbol{z}})$ and $q_3(\boldsymbol{\theta}) = \delta(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}})$, write the expressions of $< \ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta}) >_{q_1(\boldsymbol{f})}$, $< \ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta}) >_{q_2(\boldsymbol{z})}$ and $< \ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta}) >_{q_3(\boldsymbol{\theta})}$.

   Give the expressions of $\widehat{\boldsymbol{f}}$, $\widehat{\boldsymbol{z}}$ and $\widehat{\boldsymbol{\theta}}$ during the iterations (Link with Joint MAP).
4. Choosing $q_1(\boldsymbol{f})$ in the same family than $p(\boldsymbol{f}|\boldsymbol{g},\boldsymbol{z},\boldsymbol{\theta})$, $q_2(\boldsymbol{z})$ in the same family than $p(\boldsymbol{z}|\boldsymbol{g},\boldsymbol{f},\boldsymbol{\theta})$ and $q_3(\boldsymbol{\theta}) = \delta(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}})$. Write the expressions of $< \ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta}) >_{q_1(\boldsymbol{f})}$, $< \ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta}) >_{q_2(\boldsymbol{z})}$ and $< \ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta}) >_{q_3(\boldsymbol{\theta})}$.

   Give the expressions of $\widehat{\boldsymbol{f}}$, $\widehat{\boldsymbol{z}}$ and $\widehat{\boldsymbol{\theta}}$ during the iterations (Link with EM algorithem).
5. Choosing $q_1(\boldsymbol{f})$ in the same family than $p(\boldsymbol{f}|\boldsymbol{g},\boldsymbol{z},\boldsymbol{\theta})$, $q_2(\boldsymbol{z})$ in the same family than $p(\boldsymbol{z}|\boldsymbol{g},\boldsymbol{f},\boldsymbol{\theta})$ and $q_3(\boldsymbol{\theta})$ in the same family than $p(\boldsymbol{\theta}|\boldsymbol{g},\boldsymbol{f},\boldsymbol{z})$, write the expressions of $< \ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta}) >_{q_1(\boldsymbol{f})}$, $< \ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta}) >_{q_2(\boldsymbol{z})}$ and $< \ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta}) >_{q_3(\boldsymbol{\theta})}$.

   Give the expressions of $q_1(\boldsymbol{f})$, $q_2(\boldsymbol{z})$ and $q_3(\boldsymbol{\theta})$ during the iterations.

   Compare this with the EM algorithem.

Inverse Problems in Signal Processing, Imaging Systems and Computer Vision

Professor: A. Mohammad–Djafari

Exercice number 11: Bayesian Computation for Blind Deconvolution

Part 1: Blind Deconvolution
Consider the problem of Deconvolution $g(t) = h(t) * f(t) + \epsilon(t)$ when we want to estimate jointly $h(t)$ and $f(t)$. Assume that $g(t)$, $h(t)$ and $f(t)$ are causal and note by $\boldsymbol{g} = [g(0), \cdots, g(M-1)]'$,
$\boldsymbol{\epsilon} = [\epsilon(0), \cdots, \epsilon(M-1)]'$, $\boldsymbol{f} = [f(0), \cdots, f(N-1)]'$ and $\boldsymbol{h} = [h(0), \cdots, g(K-1)]'$ with $M > K$.

1. Show that: $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon} = \boldsymbol{F}\boldsymbol{h} + \boldsymbol{\epsilon}$ giving the expressions of $\boldsymbol{H}$ and $\boldsymbol{F}$.

2. Suppose $p(\boldsymbol{\epsilon}) = \mathcal{N}(0, (1/\beta)\boldsymbol{I})$. Write the expressions of $p(\boldsymbol{g}|\boldsymbol{H}, \boldsymbol{f}, \beta)$, $p(\boldsymbol{g}|\boldsymbol{F}, \boldsymbol{h}, \beta)$ and the expressions of the ML estimators $\widehat{\boldsymbol{f}}_{MV}$ of $\boldsymbol{f}$ when $\boldsymbol{h}$ is known and the expression of $\widehat{\boldsymbol{h}}_{MV}$ when $\boldsymbol{f}$ is known.

3. Suppose $p(f(t)|f(t-1)) = \mathcal{N}(f(t-1), (1/\alpha_f))$ and $p(h(t)|h(t-1)) = \mathcal{N}(h(t-1), (1/\alpha_h))$, $\forall t \geq 0$. Show then that:

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

   and precise the nature and structure of $\boldsymbol{D}$ and $\boldsymbol{C}$.

4. Write the expressions of $p(\boldsymbol{f}|\boldsymbol{H}, \boldsymbol{g}, \beta, \alpha_f, \alpha_h)$, $p(\boldsymbol{h}|\boldsymbol{F}, \boldsymbol{g}, \beta, \alpha_f, \alpha_h)$ and the Bayesian estimators $\widehat{\boldsymbol{f}}_{PM}$ of $\boldsymbol{f}$ when $\boldsymbol{h}$ is known and $\boldsymbol{h}$ when $\boldsymbol{f}$ is known.

5. Noting $\boldsymbol{\theta} = (\beta, \alpha_f, \alpha_h)$, give the expression of $p(\boldsymbol{f}, \boldsymbol{h}|\boldsymbol{g}, \boldsymbol{\theta})$ and $p(\boldsymbol{f}, \boldsymbol{h}, \boldsymbol{\theta}|\boldsymbol{g})$ when $p(\boldsymbol{\theta})$ is known.

6. In a first step, we suppose that $\boldsymbol{h}$ and $\boldsymbol{\theta}$ are known and we want to estimate $\boldsymbol{f}$. Show that: $p(\boldsymbol{f}|\boldsymbol{g}, \boldsymbol{h}, \boldsymbol{\theta})$ can be written as

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

   where

$$\begin{aligned} J(\boldsymbol{f}) &= \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + \lambda_f \|\boldsymbol{C}\boldsymbol{f}\|^2 \\ &= (\boldsymbol{f} - \widehat{\boldsymbol{f}})' \widehat{\boldsymbol{\Sigma}}_f^{-1} (\boldsymbol{f} - \widehat{\boldsymbol{f}}) + c \quad \text{with} \quad \widehat{\boldsymbol{\Sigma}}_f = (\boldsymbol{H}'\boldsymbol{H} + \lambda_f \boldsymbol{C}'\boldsymbol{C})^{-1} \quad \text{and} \quad \widehat{\boldsymbol{f}} = \widehat{\boldsymbol{\Sigma}}_f \boldsymbol{H}'\boldsymbol{g} \end{aligned}$$

   where c does not depend on $\boldsymbol{f}$.

7. Now, we suppose $\boldsymbol{f}$ and $\boldsymbol{\theta}$ are known and we want to estimate $\boldsymbol{h}$. Show that: $p(\boldsymbol{h}|\boldsymbol{g}, \boldsymbol{f}, \boldsymbol{\theta})$ can be written as:

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

where

$$\begin{aligned} J(\boldsymbol{h}) &= \|\boldsymbol{g}-\boldsymbol{F}\boldsymbol{h}\|^2+\lambda_h\|\boldsymbol{C}\boldsymbol{h}\|^2 \\ &= (\boldsymbol{h}-\widehat{\boldsymbol{h}})'\widehat{\boldsymbol{\Sigma}}_h^{-1}(\boldsymbol{h}-\widehat{\boldsymbol{h}})+c \quad \text{with} \quad \widehat{\boldsymbol{\Sigma}}_f=(\boldsymbol{F}'\boldsymbol{F}+\lambda_h\boldsymbol{C}'\boldsymbol{C})^{-1} \quad \text{and} \quad \widehat{\boldsymbol{h}}=\widehat{\boldsymbol{\Sigma}}_f\boldsymbol{F}'\boldsymbol{g} \end{aligned}$$

where c does not depend on $\boldsymbol{h}$.

8. Now, we suppose that only $\boldsymbol{\theta}$ is known and we want to estimate both $\boldsymbol{f}$ and $\boldsymbol{h}$. Show that: $p(\boldsymbol{f},\boldsymbol{h}|\boldsymbol{g},\boldsymbol{\theta})$ writes:

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

where

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

where c_1 does not depend on $\boldsymbol{f}$ and c_2 does not depend on $\boldsymbol{h}$.

We then want to compute the Joint MAP (JAMP) solution::

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

in an iterative way:

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

which becomes:

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

Discuss the convergency issues of this algorithem.

9. Now, we want to compute the posterior means of $\boldsymbol{f}$ and $\boldsymbol{h}$:

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

which need the integration with respect to $\boldsymbol{f}$ and $\boldsymbol{h}$ of $p(\boldsymbol{f},\boldsymbol{h}|\boldsymbol{g},\boldsymbol{\theta})$.

Again, we do not have analytical expressions for these. Then, we try to approximate $p(\boldsymbol{f},\boldsymbol{h}|\boldsymbol{g},\boldsymbol{\theta})$ by $q(\boldsymbol{f},\boldsymbol{h})=q_1(\boldsymbol{f})\,q_2(\boldsymbol{h})$ using the KL divergence:

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

Optimizing $K(q_1q_2:p)$ iteratively:

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

and noting that

$$
\begin{aligned}
K(q_1q_2:p) &= \int\!\!\int q_1q_2\ln(q_1q_2/p)=\int\!\!\int q_1q_2[\ln q_1+\ln q_2-\ln p]\\
&= \int q_1\int q_2[\ln q_2+\ln q_1-\ln p]=\int q_2\ln q_2+<[\ln q_1-\ln p]>_{q_1}\\
&= \int q_2\int q_1[\ln q_1+\ln q_2-\ln p]=\int q_1\ln q_1+<[\ln q_2-\ln p]>_{q_2}
\end{aligned}
$$

show that:

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

10. Now, we may choose specific families for $q_1(\boldsymbol{f})$ and $q_2(\boldsymbol{h})$. We may have two options: parametric and non parametric:

11. The first choice is: $q_1(\boldsymbol{f})=\delta(\boldsymbol{f}-\widehat{\boldsymbol{f}})$ and $q_2(\boldsymbol{h})=\delta(\boldsymbol{h}-\widehat{\boldsymbol{h}})$ where $\widehat{\boldsymbol{f}}$ and $\widehat{\boldsymbol{h}}$ are the parameters to determine. This case is sometimes is called a degenerate case. à déterminer. Then, show that:

$$
\begin{aligned}
q_1^{(k+1)}(\boldsymbol{f}) &= \delta(\boldsymbol{f}-\widehat{\boldsymbol{f}}^{(k+1)})\;\text{ with }\;\widehat{\boldsymbol{f}}^{(k+1)}=\arg\max_{\boldsymbol{f}}\left\{p(\boldsymbol{f}|\boldsymbol{g},\widehat{\boldsymbol{h}}^{(k)},\boldsymbol{\theta})\right\}\\
q_2^{(k+1)}(\boldsymbol{h}) &= \delta(\boldsymbol{h}-\widehat{\boldsymbol{h}}^{(k+1)})\;\text{ with }\;\widehat{\boldsymbol{h}}^{(k+1)}=\arg\max_{\boldsymbol{h}}\left\{p(\boldsymbol{h}|\boldsymbol{g},\widehat{\boldsymbol{f}}^{(k)},\boldsymbol{\theta})\right\}
\end{aligned}
$$

and comment this algorithem.

12. The second choice is: $q_1(\boldsymbol{f}) = \mathcal{N}(\boldsymbol{f}|\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{\Sigma}}_f)$ and $q_2(\boldsymbol{h}) = \mathcal{N}(\boldsymbol{h}|\widehat{\boldsymbol{h}}, \widehat{\boldsymbol{\Sigma}}_h)$. Then, show that::

$$\begin{array}{rcl} q_1^{(k+1)}(\boldsymbol{f}) & = & \mathcal{N}(\widehat{\boldsymbol{f}}^{(k+1)}, \widehat{\boldsymbol{\Sigma}}_f^{(k+1)}) \\ \text{with} & & \widehat{\boldsymbol{f}}^{(k+1)} =<\boldsymbol{f}>_{p(\boldsymbol{f}|\boldsymbol{g},\widehat{\boldsymbol{h}}^{(k)},\boldsymbol{\theta})} = \mathrm{E}\left\{\boldsymbol{f}|\boldsymbol{g},\widehat{\boldsymbol{h}}^{(k)},\boldsymbol{\theta}\right\} \text{ and } \widehat{\boldsymbol{\Sigma}}_f^{(k+1)} = \mathrm{cov}[\boldsymbol{f}|\boldsymbol{g},\widehat{\boldsymbol{h}}^{(k)},\boldsymbol{\theta}] \\ q_2^{(k+1)}(\boldsymbol{h}) & = & \mathcal{N}(\widehat{\boldsymbol{h}}^{(k+1)}, \widehat{\boldsymbol{\Sigma}}_h^{(k+1)}) \\ \text{with} & & \widehat{\boldsymbol{h}}^{(k+1)} =<\boldsymbol{h}>_{p(\boldsymbol{h}|\boldsymbol{g},\widehat{\boldsymbol{f}}^{(k)},\boldsymbol{\theta})} = \mathrm{E}\left\{\boldsymbol{h}|\boldsymbol{g},\widehat{\boldsymbol{f}}^{(k)},\boldsymbol{\theta}\right\} \text{ and } \widehat{\boldsymbol{\Sigma}}_h^{(k+1)} = \mathrm{cov}[\boldsymbol{h}|\boldsymbol{g},\widehat{\boldsymbol{f}}^{(k)},\boldsymbol{\theta}] \end{array}$$

Give then the expressions of $\widehat{\boldsymbol{\Sigma}}_f^{(k+1)}$, $\widehat{\boldsymbol{f}}^{(k+1)}$, $\widehat{\boldsymbol{\Sigma}}_h^{(k+1)}$ and $\widehat{\boldsymbol{h}}^{(k+1)}$ and comment this algorithem.

13. The third choice is: $q_1(\boldsymbol{f}) = p(\boldsymbol{f}|\boldsymbol{g}, \widehat{\boldsymbol{h}}, \boldsymbol{\theta})$ and $q_2(\boldsymbol{h}) = p(\boldsymbol{h}|\boldsymbol{g}, \widehat{\boldsymbol{f}}, \boldsymbol{\theta})$:

$$q_1^{(k+1)}(\boldsymbol{f}) = p(\boldsymbol{f}|\boldsymbol{g}, \widehat{\boldsymbol{h}}^{(k)}, \boldsymbol{\theta})$$
$$q_2^{(k+1)}(\boldsymbol{h}) = p(\boldsymbol{h}|\boldsymbol{g}, \widehat{\boldsymbol{f}}^{(k)}, \boldsymbol{\theta})$$

Then, show that::

$$\begin{array}{rcl} q_1^{(k+1)}(\boldsymbol{f}) & \propto & \exp\left\{<\ln p(\boldsymbol{f},\boldsymbol{h}|\boldsymbol{g},\boldsymbol{\theta}))>_{q_2^{(k)}}\right\} \\ & \propto & p(\boldsymbol{f}|\alpha)\exp\left\{<\ln p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{h},\boldsymbol{\theta})>q_2\right\} \\ q_2^{(k+1)}(\boldsymbol{h}) & \propto & \exp\left\{<\ln p(\boldsymbol{f},\boldsymbol{h}|\boldsymbol{g},\boldsymbol{\theta}))>_{q_1^{(k)}}\right\} \\ & \propto & p(\boldsymbol{h}|\alpha)\exp\left\{<\ln p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{h},\boldsymbol{\theta})>q_1\right\} \end{array}$$

and comment this algorithem.

14. When $q_1^{(k+1)}(\boldsymbol{f})$ and $q_2^{(k+1)}(\boldsymbol{h})$ are Gaussians: $q_1^{(k+1)}(\boldsymbol{f}) = \mathcal{N}(\widehat{\boldsymbol{f}}^{(k+1)}, \widehat{\boldsymbol{\Sigma}}_f^{(k+1)})$ and $q_2^{(k+1)}(\boldsymbol{h}) = \mathcal{N}(\widehat{\boldsymbol{h}}^{(k+1)}, \widehat{\boldsymbol{\Sigma}}_h^{(k+1)})$, give the expressions of $(\widehat{\boldsymbol{\Sigma}}_f^{(k+1)}, \widehat{\boldsymbol{f}}^{(k+1)})$ and $(\widehat{\boldsymbol{\Sigma}}_h^{(k+1)}, \widehat{\boldsymbol{h}}^{(k+1)})$.

Part 2: Non supervised Blind Deconvolution

In the previous problem, we know want also to estimate $\boldsymbol{\theta} = (\beta, \alpha_f, \alpha_h)$

1. Choosing the conjugate priors for $p(\beta)$, $p(\alpha_f)$ and $p(\alpha_h)$, give the expression of $p(\boldsymbol{f}, \boldsymbol{h}, \boldsymbol{\theta}|\boldsymbol{g})$. Does the expression of this posterior separable Cette loi est-elle séparable in $\boldsymbol{f}$, $\boldsymbol{h}$ and $\boldsymbol{\theta}$?

2. Trying to approximate it with $q(\boldsymbol{f}, \boldsymbol{h}, \boldsymbol{\theta}|\boldsymbol{g}) = q_1(\boldsymbol{f})\, q_2(\boldsymbol{h})\, q_3(\boldsymbol{\theta})$ by using the KL divergence:

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

and minimizing this expression iteratively, show that we obtain:

$$\begin{array}{rcl} q_1(\boldsymbol{f}) & \propto & p(\boldsymbol{f}|\alpha)\exp\left\{<\ln p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{h},\boldsymbol{\theta})>q_2q_3\right\} \\ q_2(\boldsymbol{h}) & \propto & p(\boldsymbol{h}|\alpha)\exp\left\{<\ln p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{h},\boldsymbol{\theta})>q_1q_3\right\} \\ q_3(\boldsymbol{\theta}) & \propto & p(\boldsymbol{\theta})\exp\left\{<\ln p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{h},\boldsymbol{\theta})>q_1q_2\right\} \end{array}$$