Projet de restauration d'images : deuxième partie

Charles Soussen, Elisabeth Lahalle

1. Déconvolution d'un signal triangulaire : approche matricielle et approximation circulante

On reprend le problème de déconvolution du signal triangulaire traité dans la première partie. On considère maintenant la version du produit de convolution qui produit un signal $y_{nb} = x \star h$ de même taille que x. Ce produit de convolution s'exprime par :

$$y_{nb}[n] \;=\; \sum_{k=0}^{N_h-1} h[k]\, x[n-k] \;=\; \sum_{k=n-N_h+1}^{n} h[n-k]\, x[k], \tag{1}$$

pour $n = N_h,\, N_h+1, \ldots, N_h+N_x-1$.

Les vecteurs $\boldsymbol{x} = \big[x[1], x[2], \ldots, x[N_x]\big]^T$ et $\boldsymbol{y}_{nb} = \big[y_{nb}[N_h], \ldots, y_{nb}[N_h+N_x-1]\big]^T$ sont donc de même taille N_x, et la formulation matricielle $\boldsymbol{y}_{nb} = \boldsymbol{H}\boldsymbol{x}$ fait intervenir une matrice $\boldsymbol{H}$ de Tœplitz, carrée de taille $N_x \times N_x$, et triangulaire supérieure :

$$\boldsymbol{H} = \begin{bmatrix} h[N_h-1] & h[N_h-2] & \ldots & h[0] & 0 & \ldots & \ldots & 0 \\ 0 & h[N_h-1] & \ldots & \ldots & h[0] & 0 & \ldots & 0 \\ \vdots & & \ddots & & & & & \\ 0 & \ldots & 0 & h[N_h-1] & & \ldots & \ldots & h[0] \\ 0 & \ldots & \ldots & 0 & h[N_h-1] & & \ldots & h[1] \\ \vdots & & & & & \ddots & & \vdots \\ 0 & 0 & \ldots & \ldots & \ldots & 0 & h[N_h-1] & h[N_h-2] \\ 0 & 0 & \ldots & \ldots & \ldots & \ldots & 0 & h[N_h-1] \end{bmatrix}.$$

Dans la suite, on reprend le problème de l'estimation du vecteur $\boldsymbol{x}$ à partir du vecteur de données bruitées $\boldsymbol{y} = \boldsymbol{y}_{nb} + \boldsymbol{w}$, où $\boldsymbol{w}$ est la réalisation d'un bruit blanc gaussien.

1.1 Moindres carrés

On rappelle que le critère des moindres carrés est donné par :

$$\mathcal{F}_0(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|^2$$

Comme $\boldsymbol{H}$ est carrée, le minimum de $\mathcal{F}_0(\boldsymbol{x})$ est donné par $\widehat{\boldsymbol{x}} = \boldsymbol{H}^{-1}\boldsymbol{y}$ (sous réserve que $\boldsymbol{H}$ soit inversible). C'est l'estimée de $\boldsymbol{x}$ au sens des moindres carrés.

Adapter votre programme de filtrage inverse (sans régularisation) en conséquence.

1.2 Résolution par approximation circulante

La matrice $\boldsymbol{H}$ à inverser est de taille $N_x \times N_x$ et si N_x est grand (comme ça peut être le cas en 2D pour une image) l'inversion directe est difficile pour des questions de coût de calcul.

Il existe une méthode rapide pour inverser cette matrice et calculer $\widehat{\boldsymbol{x}}$. Cette méthode est fondée sur les propriétés des matrices circulantes. Ces propriétés permettent d'approcher $\boldsymbol{H}$ par une matrice diagonale $\widetilde{\boldsymbol{H}}$. Le calcul $\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{y}$ devient alors très bon marché en terme de temps de calcul.

Cependant, le calcul de $\boldsymbol{H}^{-1}\boldsymbol{y}$ est approximatif. Il s'agit de remplacer $\boldsymbol{H}$ par

$$\widetilde{\boldsymbol{H}} = \begin{bmatrix} h[N_h-1] & h[N_h-2] & \dots & h[0] & 0 & \dots & \dots & 0 \\ 0 & h[N_h-1] & \dots & \dots & h[0] & 0 & \dots & 0 \\ \vdots & & \ddots & & & & & \\ 0 & \dots & 0 & h[N_h-1] & & \dots & \dots & h[0] \\ h[0] & 0 & \dots & 0 & h[N_h-1] & \dots & \dots & h[1] \\ h[1] & h[0] & 0 & \dots & 0 & h[N_h-1] & \dots & h[2] \\ \vdots & & \ddots & & & & \ddots & \vdots \\ h[N_h-2] & \dots & h[1] & h[0] & \dots & \dots & 0 & h[N_h-1] \end{bmatrix}.$$

Cette approximation consiste à ajouter des termes en bas et à gauche de la matrice $\boldsymbol{H}$ de façon à la rendre circulante. Il faut noter que cette approximation est d'autant meilleure que N_x (la taille du signal) est plus grand que N_h (la longueur de la réponse impulsionnelle).

La matrice $\widetilde{\boldsymbol{H}}$ étant circulante, on peut la diagonaliser dans la base de Fourier (voir cours, $\boldsymbol{W}\boldsymbol{W}^* = N_x\,\boldsymbol{I}$, donc $\boldsymbol{W}^{-1} = \boldsymbol{W}^*/N_x$) :

$$\widetilde{\boldsymbol{H}} = \boldsymbol{W}^{-1}\boldsymbol{\Lambda}_h\boldsymbol{W}\,. \tag{2}$$

La matrice $\boldsymbol{\Lambda}_h$ est diagonale et ses éléments diagonaux sont les valeurs propres $\lambda_h^1, \lambda_h^2, \dots, \lambda_h^{N_x}$ de $\widetilde{\boldsymbol{H}}$. Elles s'obtiennent en calculant la TFD de la première ligne de $\widetilde{\boldsymbol{H}}$, c'est-à-dire en calculant la TFD de la réponse impulsionnelle sur N_x points.

En exploitant la relation de diagonalisation (2) et les relations sur la matrice $\boldsymbol{W}$ contenues en annexe, on obtient $\widehat{\boldsymbol{x}} \approx \widetilde{\boldsymbol{x}}$, avec :

$$\widetilde{\boldsymbol{x}} = \widetilde{\boldsymbol{H}}^{-1}\boldsymbol{y} = \quad \boldsymbol{W}^{-1}\boldsymbol{\Lambda}_h^{-1}\boldsymbol{W}\boldsymbol{y}.$$

On en déduit alors, en multipliant à gauche par $\boldsymbol{W}$:

$$\widetilde{\boldsymbol{X}} = \boldsymbol{W}\widetilde{\boldsymbol{x}} \;=\; \boldsymbol{\Lambda}_h^{-1}\boldsymbol{Y}, \tag{3}$$

où $\widetilde{\boldsymbol{X}}$ et $\boldsymbol{Y}$ sont les FFT de $\widetilde{\boldsymbol{x}}$ et $\boldsymbol{y}$.

1.3 Algorithme

On construit le vecteur $\boldsymbol{g}_{MC}$ des valeurs de la diagonale de $\boldsymbol{\Lambda}_h^{-1}$:

$$\boldsymbol{g}_{MC}[n] = 1/\lambda_h^n \;\text{ pour } n = 1, 2, \dots, N_x. \tag{4}$$

On obtient alors le vecteur $\widetilde{\boldsymbol{X}}$ comme le produit « éléments par éléments » (noté $.*$) des deux vecteurs $\boldsymbol{g}_{MC}$ et $\boldsymbol{Y}$: $\widetilde{\boldsymbol{X}} = \boldsymbol{g}_{MC}\,.*\,\boldsymbol{Y}$.

En résumé, la procédure à suivre est donnée ci-dessous.

1. Construire $\boldsymbol{\lambda}_h$ la FFT de la réponse impulsionnelle sur N_x points.
2. Construire le vecteur $\boldsymbol{g}_{MC}$ du transfert en fréquence selon (4).
3. Construire $\boldsymbol{Y}$ la FFT des données.
4. Calculer $\widetilde{\boldsymbol{X}}$ comme produit du transfert en fréquence $\boldsymbol{g}_{MC}$ par $\boldsymbol{Y}$.
5. Revenir dans le domaine spatial par FFT inverse de $\widetilde{\boldsymbol{X}}$ pour obtenir $\widetilde{\boldsymbol{x}}$.

Remarquez que cette procédure correspond exactement à l'algorithme de « déconvolution naïve » programmé dans la Partie 1.

1.4 Régularisation et douceur

On rappelle la forme du critère pénalisé quadratique :

$$\begin{aligned}\mathcal{F}_\alpha(\boldsymbol{x}) &= \|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|^2 + \alpha \sum_{n=1}^{N_x-1} (x[n+1] - x[n])^2 \\ &= \|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|^2 + \alpha \,\|\boldsymbol{D}_1\boldsymbol{x}\|^2\end{aligned}$$

avec $\boldsymbol{D}_1$ la matrice de différences d'ordre 1, de taille $(N_x - 1) \times N_x$, définie par :

$$\boldsymbol{D}_1 = \begin{bmatrix} 1 & -1 & 0 & 0 & \dots & \dots & 0 \\ 0 & 1 & -1 & 0 & \dots & \dots & 0 \\ \vdots & & \ddots & \ddots & & & \vdots \\ \vdots & & & \ddots & \ddots & & \vdots \\ 0 & \dots & \dots & 0 & 1 & -1 & 0 \\ 0 & \dots & \dots & 0 & 0 & 1 & -1 \end{bmatrix}.$$

Le critère $\mathcal{F}_\alpha(\boldsymbol{x})$ est quadratique en $\boldsymbol{x}$ et on connaît l'expression de son minimum :

$$\widehat{\boldsymbol{x}} = (\boldsymbol{H}^T\boldsymbol{H} + \alpha\boldsymbol{D}_1^T\boldsymbol{D}_1)^{-1}\boldsymbol{H}^T\boldsymbol{y}\,. \tag{5}$$

1.5 Approximation circulante

Pour les mêmes raisons que dans le cas des moindres carrés simples, on accepte une approximation circulante pour $\boldsymbol{H}$ et $\boldsymbol{D}_1$ de manière à « remplacer » les calculs sur des matrices quelconques par des calculs sur des matrices diagonales. En ce qui concerne $\boldsymbol{H}$ on conserve l'approximation par $\widetilde{\boldsymbol{H}}$. En ce qui concerne $\boldsymbol{D}_1$ il suffit d'y ajouter une dernière ligne pour obtenir une approximation circulante $\widetilde{\boldsymbol{D}_1}$, de dimension $N_x \times N_x$:

$$\widetilde{\boldsymbol{D}_1} = \begin{bmatrix} 1 & -1 & 0 & 0 & \dots & \dots & 0 \\ 0 & 1 & -1 & 0 & \dots & \dots & 0 \\ \vdots & & \ddots & \ddots & & & \vdots \\ \vdots & & & \ddots & \ddots & & \vdots \\ 0 & \dots & \dots & 0 & 1 & -1 & 0 \\ 0 & \dots & \dots & 0 & 0 & 1 & -1 \\ -1 & \dots & \dots & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Le minimum approché prend alors la forme suivante :

$$\widetilde{\boldsymbol{x}} = (\widetilde{\boldsymbol{H}}^T\widetilde{\boldsymbol{H}} + \alpha\widetilde{\boldsymbol{D}_1}^T\widetilde{\boldsymbol{D}_1})^{-1}\widetilde{\boldsymbol{H}}^T\boldsymbol{y}.$$

En partant de cette expression et des diagonalisations de $\widetilde{\boldsymbol{H}}$ et $\widetilde{\boldsymbol{D}_1}$, on a :

$$\begin{aligned}\widehat{\boldsymbol{x}} &= \boldsymbol{W}^{-1}(\boldsymbol{\Lambda}_h^*\boldsymbol{\Lambda}_h + \boldsymbol{\Lambda}_d^*\boldsymbol{\Lambda}_d)^{-1}\boldsymbol{\Lambda}_h^*\boldsymbol{W}\boldsymbol{y}, \\ \widehat{\boldsymbol{X}} &= (\boldsymbol{\Lambda}_h^*\boldsymbol{\Lambda}_h + \alpha\boldsymbol{\Lambda}_d^*\boldsymbol{\Lambda}_d)^{-1}\boldsymbol{\Lambda}_h^*\boldsymbol{Y},\end{aligned}$$

en notant $\boldsymbol{\lambda}_h = \left[\lambda_h^1, \lambda_h^2, \dots, \lambda_h^{N_x}\right]^T$ et $\boldsymbol{\lambda}_d = \left[\lambda_d^1, \lambda_d^2, \dots, \lambda_d^{N_x}\right]^T$ les valeurs propres de $\widetilde{\boldsymbol{H}}$ et de $\boldsymbol{D}_1$.

Pour terminer, construisez le vecteur $\boldsymbol{g}_{MCR}$ des valeurs de la diagonale de la matrice $(\boldsymbol{\Lambda}_h^*\boldsymbol{\Lambda}_h + \alpha\boldsymbol{\Lambda}_d^*\boldsymbol{\Lambda}_d)^{-1}\boldsymbol{\Lambda}_h^*$:

$$\boldsymbol{g}_{MCR}[n] = \frac{(\boldsymbol{\lambda}_h^n)^*}{|\boldsymbol{\lambda}_h^n|^2 + \alpha|\boldsymbol{\lambda}_d^n|^2} \quad \text{pour } n = 1, 2, \ldots, N_x. \tag{6}$$

Expliquez alors pourquoi le vecteur $\widetilde{\boldsymbol{X}}$ s'obtient comme le produit éléments par éléments des vecteurs $\boldsymbol{g}_{MCR}$ et $\boldsymbol{Y}$.

En résumé la procédure à suivre est donnée ci-dessous.

1. Construire $\boldsymbol{\lambda}_h$ la FFT de la réponse impulsionnelle sur N_x points.
2. Construire $\boldsymbol{\lambda}_d$ la FFT de $[1; -1]$ sur N_x points.
3. Construire le vecteur $\boldsymbol{g}_{MCR}$ du transfert en fréquence selon (6).
4. Construire $\boldsymbol{Y}$ la FFT des données.
5. Calculer $\widetilde{\boldsymbol{X}}$ comme produit du transfert en fréquence $\boldsymbol{g}_{MCR}$ par $\boldsymbol{Y}$.
6. Revenir dans le domaine spatial par FFT inverse de $\widetilde{\boldsymbol{X}}$ pour obtenir $\widehat{\boldsymbol{x}}$.

2. Restauration d'images

On veut éliminer le « flou » d'une image c'est-à-dire retrouver une image nette à partir d'une version floue de celle-ci. On peut imaginer une photo prise avec une mise au point imparfaite. L'image d'un point n'est pas un point mais une « tâche ». L'image mesurée est alors floue puisqu'elle est la superposition des tâches issues de chaque point de l'image vraie.

Connaissant la forme de la tâche associée à un point, c'est-à-dire la réponse impulsionnelle de l'appareil et connaissant bien sûr l'image floue, le travail à réaliser consiste à retrouver l'image nette.

2.1 Modélisation du problème direct

Le modèle le plus simple pour décrire une telle transformation est une convolution 2D.

$$y[m, n] = \sum_{p=-P}^{P} \sum_{q=-Q}^{Q} h[p, q]\, x[m-p, n-q] + w[m, n]. \tag{7}$$

$x[m, n]$ représente l'image réelle et $y[m, n]$ représente les données disponibles. Afin de prendre en compte les erreurs de mesure et les erreurs de modélisation, on introduit un bruit $w[m, n]$ supposé blanc spatialement.

Notez que le filtre ci-dessus n'est pas causal. En effet, le calcul de $y[m, n]$ fait intervenir le pixel $x[m, n]$ et ses voisins dans les quatre directions (en haut, en bas, à gauche et à droite). Comme dans le cas 1D, il est fondamental de noter qu'il s'agit d'un filtre passe-bas. La plupart des appareils de mesure sont d'ailleurs de ce type : ils ne peuvent pas « suivre de trop hautes fréquences ». Il est alors clair que le système élimine les composantes haute fréquence de l'image, c'est pourquoi le problème inverse est particulièrement difficile : il faut reconstituer les hautes fréquences alors qu'elles ont disparu des données.

2.2 Déconvolution d'image

Le travail à réaliser consiste à construire un estimateur $\widehat{x}[m, n]$ de $x[m, n]$ à partir des mesures $y[m, n]$.

Le critère $\mathcal{F}_\alpha(\boldsymbol{x})$ est maintenant défini par :

$$\mathcal{F}_\alpha(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|_F^2 + \alpha\,\|\boldsymbol{d}_1 * \boldsymbol{x}\|_F^2$$

Justifier et commenter le choix du filtre

$$\boldsymbol{d}_1 \quad = \quad \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}$$

en guise d'opérateur de régularisation.

On peut développer des calculs rigoureux analogues au cas 1D pour établir les résultats du cas 2D. Cependant, ces calculs sont beaucoup plus fastidieux car les matrices ont une structure plus complexe : structure de Tœplitz par bloc, chaque bloc possédant une structure de Tœplitz. Par ailleurs, l'approximation circulante dans les deux directions est plus difficile à manipuler. C'est pourquoi, nous n'abordons le cas 2D que comme une extension du cas 1D. Quelques remarques cependant.

— Les images, la réponse impulsionnelle et les termes de douceur sont bidimensionnels, c'est pourquoi les FFT sont remplacées par des FFT 2D.

— Plus précisément, si l'image à traiter possède N lignes et N colonnes, toutes les FFT 2D doivent être prises sur N lignes et N colonnes.

Les structures des procédures déduites du cas 1D sont indiquées ci-dessous, d'abord dans le cas des moindres carrés simples puis dans le cas des moindres carrés régularisés.

1. Construire $\boldsymbol{\lambda}_h$ la FFT 2D de la réponse impulsionnelle sur $N \times N$ points.
2. Construire la matrice $\boldsymbol{g}_{MC}$ du transfert en fréquence selon (4).
3. Construire $\boldsymbol{Y}$ la FFT 2D des données.
4. Calculer $\widehat{\boldsymbol{X}}$ comme produit de $\boldsymbol{g}_{MC}$ par $\boldsymbol{Y}$.
5. Revenir dans le domaine spatial par FFT 2D inverse de $\widehat{\boldsymbol{X}}$ pour obtenir $\widehat{\boldsymbol{x}}$.

1. Construire $\boldsymbol{\lambda}_h$ la FFT 2D de la réponse impulsionnelle sur $N \times N$ points.
2. Construire $\boldsymbol{\lambda}_d$ la FFT 2D de $\begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}$ sur $N \times N$ points.
3. Construire la matrice $\boldsymbol{g}_{MCR}$ du transfert en fréquence selon (6).
4. Construire $\boldsymbol{Y}$ la FFT 2D des données.
5. Calculer $\widehat{\boldsymbol{X}}$ comme produit de $\boldsymbol{g}_{MCR}$ par $\boldsymbol{Y}$.
6. Revenir dans le domaine spatial par FFT 2D inverse de $\widehat{\boldsymbol{X}}$ pour obtenir $\widehat{\boldsymbol{x}}$.

2.3 Simulation numérique

2.3.1 Simulation du problème direct

Charger l'image `cameraman.tif` dans la matrice `x`, de taille 256×256 pixels. C'est l'image qu'on cherchera à restaurer par la suite.

Filtrer l'image `cameraman.tif` par le filtre moyenneur :

$$\boldsymbol{h}_1 = \frac{1}{15}\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix},$$

(filtrage par `conv2` et en mode `'same'`) et bruiter l'image obtenue avec un bruit additif blanc et gaussien, de façon à assurer un rapport signal sur bruit de 30 dB.

Visualisez l'image nette `x` et l'image floue et bruitée `y`.

2.3.2 Mise en œuvre de la déconvolution

Les programmes à réaliser doivent mettre en œuvre la déconvolution à l'aide de la minimisation des deux critères : les moindres carrés simples et les moindres carrés régularisés.

Notez l'aspect général de l'image obtenue et observez en particulier la netteté des frontières par exemple entre le manteau du photographe et le ciel. Intéressez-vous aussi aux détails du visage ou de l'appareil photo.

Que se passe-t-il avec le filtre inverse et pourquoi ? Avec la méthode régularisée, expliquez qualitativement les résultats obtenus. Quelle est l'influence du paramètre α ? Confrontez les résultats obtenus aux résultats attendus compte tenu de la forme du terme de régularisation.

A suivre.... Ce programme sera complété pour les prochains TD... avec une extension à des fonctions coût non quadratiques...