Université Aix-Marseille 2018-2019
Master Mathématiques appliquées et statistiques 2ème année

Représentation parcimonieuse des signaux

Notes de cours et Td : Parcimonie et algorithmes de décomposition parcimonieuse.
Minimisation sous contrainte de parcimonie : algorithme de seuillage itératif.

Notations

— On note $\langle .,. \rangle$ le produit scalaire usuel sur $\mathbb{C}^N$ tel que $\langle x, y \rangle = \sum_{n=0}^{N-1} x_n \overline{y_n}$ pour $x = (x_i)_{i=0,\dots N-1} \in \mathbb{C}^N$ et $y = (y_i)_{i=0,\dots N-1} \in \mathbb{C}^N$. On note $|x|_2^2 = \sum_{n=0}^{N-1} |x_n|^2$ pour $x \in \mathbb{C}^N$.

1 Introduction

Remarque préliminaire : pour simplifier l'étude on se placera dans tout ce travail dans $\mathbb{R}^N$ et on considérera des dictionnaires de $\mathbb{R}^N$. On pourra dans un deuxième temps considérer le cas de $\mathbb{C}^N$ et d'un dictionnaire dans $\mathbb{C}^N$.

Nous avons étudié dans le cours précédent un algorithme, le Matching-Pursuit, capable de retrouver, si elle existe, et sous certaines conditions, une décomposition parcimonieuse du vecteur x dans un dictionnaire $\mathcal{D} = \{\phi^0, \phi^1, \dots, \phi^{K-1}\}$ donné.

Le problème que l'on a cherché à résoudre grâce au Matching-Pursuit peut s'écrire de la manière suivante. Pour fixer les notations on appelle D avec la matrice dont les vecteurs colonnes sont les vecteurs ϕ^ℓ écrits dans la base canonique, et on écrit un vecteur $\alpha \in \mathbb{R}^K$ comme le vecteur colonne $\alpha = (\alpha_0, \dots, \alpha_{K-1})^t$.

On cherche le vecteur $\tilde{\alpha}$ de $\mathbb{R}^K$ tel $D\tilde{\alpha} = x$ et qui a le moins possible de coordonnées non nulles. Autrement dit on cherche

$$\tilde{\alpha} = \underset{\alpha \in \mathbb{R}^K :\, D\alpha = x}{\operatorname{argmin}} \; card\{\ell : \alpha_\ell \neq 0\}$$

En utilisant la notation (classique) $|\alpha|_0 = card\{\ell : \alpha_\ell \neq 0\}$ on se retrouve donc à chercher

$$\tilde{\alpha} = \underset{\alpha \in \mathbb{R}^K :\, D\alpha = x}{\operatorname{argmin}} \; |\alpha|_0$$

Une autre manière d'écrire les choses est de dire que l'on veut résoudre le problème

$$\begin{aligned} &\text{Trouver } \tilde{\alpha} \in \mathbb{R}^K \text{ tel que} \\ |\tilde{\alpha}|_0 \quad &= \min\{|\alpha|_0 : \alpha \in \mathbb{R}^K \text{ et } D\alpha = x\} \end{aligned} \tag{1}$$

Souvent les données mesurées vont être entachées d'erreurs ou de perturbation. Autrement dit au lieu d'avoir à notre disposition un vecteur $x = \sum \alpha_\ell \phi^\ell$ nous avons à notre disposition une mesure $y = x + b$ où b est une perturbation d'amplitude ε, c'est à dire qui vérifie $|b|_2 \leq \varepsilon$. Les perturbations peuvent être par exemple dues à des erreurs de mesure aléatoires.

Le problème que nous pouvons chercher à résoudre est alors le suivant. Il s'agit de trouver $\tilde{\alpha} \in \mathbb{R}^K$ tel que

$$\tilde{\alpha} = \underset{\alpha \in \mathbb{R}^K : |D\alpha - x|_2 \leq \varepsilon}{\operatorname{argmin}} \; |\alpha|_0$$

ou autrement dit

$$\begin{aligned} &\text{Trouver } \tilde{\alpha} \in \mathbb{R}^K \text{ tel que} \\ |\tilde{\alpha}|_0 &= \min\{|\alpha|_0 : \alpha \in \mathbb{R}^K \text{ et } |D\alpha - x|_2 \leq \varepsilon\} \end{aligned} \tag{2}$$

Si on considère le cas général où D est une matrice de taille quelconque $N * K$ et x un vecteur de $\mathbb{R}^N$ un théorème (théorème 2.17 démontré dans [2]), permet de montrer que le problème de trouver $\tilde{\alpha}$ solution de (1) ou encore celui de trouver $\tilde{\alpha}$ solution de (2) est en toute généralité un problème dit NP difficile.

C'est à dire qu'a priori (sauf si une fameuse conjecture toujours irrésolue à ce jour dite "P≠NP" est fausse) il n'existe pas d'algorithme général capable de calculer une solution $\tilde{\alpha}$ pour D une matrice en général et x général.

Une stratégie est alors de chercher à résoudre un problème pour lequel on a plus de chances de trouver un algorithme qui le résolve, et qui donnerait des solutions plutôt satisfaisantes vu ce qu'on veut faire. En particulier un des problèmes posés par (1) et (2) est que la fonction $\alpha \mapsto |\alpha|_0$ n'est pas convexe, ce qui signifie en particulier qu'elle peut admettre des minimas locaux qui ne sont pas des minimas globaux (voir détails dans la partie suivante). Un algorithme qui recherche des minimas peut donc se retrouver à calculer un minimum qui n'est en réalité pas celui que l'on devrait trouver.

On va donc remplacer $\alpha \mapsto |\alpha|_0$ par $\alpha \mapsto |\alpha|_1 = \sum\limits_{i=0}^{K-1} |\alpha_i|$ qui a la propriété d'être convexe et de favoriser la parcimonie comme on le verra dans la suite.

Nous pouvons donc chercher à résoudre le problème suivant :

$$\begin{aligned} &\text{Trouver } \tilde{\alpha} \in \mathbb{R}^K \text{ tel que} \\ |\tilde{\alpha}|_1 &= \min\{|\alpha|_1 : \alpha \in \mathbb{R}^K \text{ et } |D\alpha - x|_2 \leq \varepsilon\} \end{aligned} \tag{3}$$

Le problème (3) est connu dans la littérature sous le nom de « Basis Pursuit denoising ».

Enfin une dernière chose que nous allons faire c'est de chercher à résoudre un problème de minimisation globale plutôt qu'un problème de minimisation sous contrainte, ce qui est en général plus facile. La fonctionnelle suivante va nous être utile.

Soit $N \in N^\star$ et $K \in \mathbb{N}^\star$. Soit $\lambda \in \mathbb{R}^+$ et D une matrice $N \times K$. On note pour $\alpha \in \mathbb{R}^K$ et $x \in \mathbb{R}^N$

$$F_\lambda(\alpha) = \frac{1}{2}|D\alpha - x|_2^2 + \lambda|\alpha|_1$$

On considère donc à $\lambda \in \mathbb{R}_+^*$ fixé le problème

$$\begin{aligned} &\text{Trouver } \tilde{\alpha} \in \mathbb{R}^K \text{ tel que} \\ F_\lambda(\tilde{\alpha}) &= \min\{F_\lambda(\alpha) : \alpha \in \mathbb{R}^K\} \end{aligned} \tag{4}$$

Proposition 1

Avec les notations précédentes si il existe $\tilde{\alpha}^\lambda \in \mathbb{R}^K$ solution de (4) alors il existe $\varepsilon > 0$ tel que $\tilde{\alpha}^\lambda$ est solution de (3).

Démonstration de la proposition 1 : Soit $\tilde{\alpha}^\lambda \in \mathbb{R}^K$ tel que $F_\lambda\left(\tilde{\alpha}^\lambda\right) = \min\limits_{\alpha \in \mathbb{R}^K} F_\lambda(\alpha)$. On note $\varepsilon = |D\tilde{\alpha}^\lambda - x|_2$.

On a donc pour tout $\alpha \in \mathbb{R}^K$ tel que $|D\alpha - x|_2 \leq \varepsilon$

$$\lambda|\tilde{\alpha}^\lambda|_1 + \frac{1}{2}|D\tilde{\alpha}^\lambda - x|_2^2 \leq \lambda|\alpha|_1 + \frac{1}{2}|D\alpha - x|_2^2 \leq \lambda|\alpha|_1 + \frac{1}{2}|D\tilde{\alpha}^\lambda - x|_2^2$$

d'où le résultat.

On admet que la réciproque de la proposition 1 est vraie.

Proposition 2

Avec les notations précédentes, soit $\tilde{\alpha}^\varepsilon$ une solution de (3).

Alors il existe $\lambda \in \mathbb{R}^+$ tel que $\tilde{\alpha}^\varepsilon$ est solution de (4).

Dans toute la suite nous allons chercher à savoir si (4) a une solution, si nous pouvons trouver un algorithme qui la calcule et si c'est le cas nous étudierons les propriétés de cet algorithme.

2 Rappels sur les fonctions convexes.

Les définitions suivantes sont valables dans un espace de Hilbert quelconque. Ici le cas qui nous intéresse est $H = \mathbb{R}^N$.

Définition 1

Soit H un espace de Hilbert.

1. *Soit $f : H \mapsto \mathbb{R}$ une fonction. Cette fonction est convexe si pour tous u, v dans H et $\alpha \in]0,1[$*

$$f(\alpha u + (1-\alpha)v) \leq \alpha f(u) + (1-\alpha) f(v).$$

2. *Soit $f : H \mapsto \mathbb{R}$ une fonction. Cette fonction est strictement convexe si pour tous u, v dans H et $\alpha \in]0,1[$*

$$f(\alpha u + (1-\alpha)v) < \alpha f(u) + (1-\alpha) f(v).$$

Dans le cas où f est une fonction définie sur un intervalle I de $\mathbb{R}$, dérivable et à valeurs dans $\mathbb{R}$ nous avons les critères suivants.

Proposition 3

Soit I un intervalle I de $\mathbb{R}$.
Soit $f : I \to \mathbb{R}$ une fonction dérivable.

1. *f est une fonction convexe si et seulement si f' est une fonction croissante sur I.*
2. *f est une fonction strictement convexe si et seulement si f' est une fonction strictement croissante sur I.*

Exemple : la fonction définie sur $\mathbb{R}$ $x \mapsto x^2$ est strictement convexe. On en déduit que pour tout $\lambda > 0$ F_λ est convexe, et strictement convexe si $Ker(D) = \{0\}$.

Le résultat qui nous intéresse est le suivant.

Proposition 4

1. *Soit $f : H \mapsto \mathbb{R}$ une fonction convexe.*

   Soit $f(\tilde{u})$ un minimum local de f. Alors $f(\tilde{u})$ est un minimum global de f.
2. *Soit $f : H \mapsto \mathbb{R}$ une fonction strictement convexe.*

   Soit $f(\tilde{u})$ un minimum local de f. Alors $f(\tilde{u})$ est l'unique minimum global de f.

Démonstration :

1. il existe une boule $B \subset H$ de centre $\tilde{u}$ et de rayon R tel que pour tout $v \in B$

$$f(\tilde{u}) \leq f(v)$$

Soit $z \in H$ qui n'est pas dans B. On prend α et v tels que $\alpha = \frac{R}{2|\tilde{u}-z|_2}$ et $v = \alpha z + (1-\alpha)\tilde{u}$.

On a que $f(\tilde{u}) \leq f(v)$ ce qui permet de conclure que $f(\tilde{u}) \leq f(z)$ pour tout $z \in H$.

2. on suppose qu'on a deux points $\tilde{u} \neq \tilde{v}$ tels que $f(\tilde{u})$ et $f(\tilde{v})$ sont deux minimas globaux de f. Alors on obtient une contradiction en considérant le point $\frac{1}{2}(\tilde{u}+\tilde{v})$.

3 Minimum de F_λ

Notre objectif ici est de montrer que F_λ admet un minimum sur $\mathbb{R}^N$. Cela vient de la proposition suivante.

Proposition 5

Soit f une fonction continue définie sur $\mathbb{R}^N$ à valeur dans $\mathbb{R}$ non identiquement nulle. On note $\|.\|$ une norme sur $\mathbb{R}^N$.
On suppose que $\lim_{\|\alpha\| \to +\infty} f(\alpha) = +\infty$. Alors f admet un minimum global sur $\mathbb{R}^N$.

On sait donc que $\tilde{\alpha}^\lambda$ existe et du fait que F_λ est strictement convexe on sait que le minimum est unique. Donc le problème (4) a unique solution. Notre objectif est maintenant de trouver un algorithme qui le calcule numériquement.

4 Cas particulier où $\mathcal{D}$ est un dictionnaire constitué d'une base orthonormale.

Dans ce cas on peut calculer directement le minimum de F_λ.

Soit $T \in \mathbb{R}^+$ paramètre fixé. On définit la fonction dite de seuillage « doux » s_T définie sur $\mathbb{R}$ comme suit

$$s_T(a) = \begin{cases} a - T\mathrm{sign}(a) & \text{si } |a| - T \geq 0 \\ 0 & \text{sinon} \end{cases}$$

avec $\mathrm{sign}(a) = 0$ si $a = 0$ et $\mathrm{sign}(a) = \frac{a}{|a|}$ sinon.

On note pour $\alpha \in \mathbb{R}^K$ tel que $\alpha = (\alpha_0, \dots, \alpha_{K-1})$ l'opérateur de $\mathbb{R}^K$ sur $\mathbb{R}^K$ noté $\alpha \mapsto S_T(\alpha) \in \mathbb{R}^K$ tel que pour tout $i \in \{0, \dots, K-1\}$ $(S_T(\alpha))_i = s_T(\alpha_i)$. Autrement dit $S_T(\alpha)$ est le vecteur dont chaque coordonnée i est $s_T(\alpha_i)$.

Proposition 6

Soit $\mathcal{D} = \{\phi^0, \dots, \phi^{N-1}\}$ une base orthonormale. On note D la matrice représentative de D dans la base canonique. Celle-ci est donc une matrice orthogonale.

Soit $x \in \mathbb{R}^N$ et β l'unique vecteur de $\mathbb{R}^N$ tel que $x = D\beta$.
Alors la solution $\tilde{\alpha}^\lambda = (\tilde{\alpha}^\lambda_0, \dots \tilde{\alpha}^\lambda_{N-1})$ de (4) est

$$\tilde{\alpha}^\lambda = S_\lambda(\beta)$$

c'est à dire qu'elle vérifie pour tout $j \in \{0, \dots, N-1\}$

$$(\tilde{\alpha}^\lambda)_j = s_\lambda(\beta_j).$$

Démonstration de la proposition 6 : On a vu, que D est une matrice orthogonale,

$$F_\lambda(\alpha) = \frac{1}{2}|\alpha - \beta|_2^2 + \lambda|\alpha|_1 = \frac{1}{2}\sum_{i=0}^{N-1}(\alpha_i - \beta_i)^2 + \lambda|\alpha_i|$$

la démonstration se conclut facilement une fois qu'on a démontré le lemme suivant.

Lemme 1

*Soit $T > 0$, et $t_0 \in \mathbb{R}$ Soit $f_T : t \mapsto \frac{|t-t_0|^2}{2} + T|t|$ définie sur $\mathbb{R}$.
Le minimum de f_T est atteint pour $t = s_T(t_0)$.*

5 Cas d'un dictionnaire $\mathcal{D}$ quelconque

Dans le cas où on a un dictionnaire $\mathcal{D}$ quelconque avec $\mathcal{D} = \{\phi^0, \ldots, \phi^{K-1}\}$ on ne peut plus appliquer la technique précédente. En effet remarquons que

$$\begin{aligned}
F_\lambda(\alpha) &= \frac{1}{2}|D\alpha - x|_2^2 + \lambda|\alpha|_1 \\
&= \frac{1}{2}\sum_{i=0}^{N-1}((D\alpha)_i - x_i)^2 + \lambda|\alpha_i| \\
&= \frac{1}{2}\sum_{i=0}^{N-1}((D\alpha)_i)^2 + x_i^2 - 2(D\alpha)_i.x_i + \lambda|\alpha_i| \\
&= \frac{1}{2}|D\alpha|_2^2 + \frac{1}{2}|x|_2^2 - 2\langle D\alpha, x\rangle + \lambda|\alpha|_1 \qquad (5)
\end{aligned}$$

On voit qu'on ne peut plus optimiser variable par variable le résultat. En réalité on n'a plus moyen de calculer directement $\tilde{\alpha}^\lambda$. On va écrire un algorithme itératif basé sur la stratégie suivante.

Soit $L > 0$ et

$$G : (\alpha, u) \mapsto \frac{L}{2}|\alpha - u|_2^2 - \frac{1}{2}|D\alpha - Du|_2^2$$

On ajuste L pour que $G(\alpha, u) \geq 0$ pour tout u et α dans $\mathbb{R}^K$.

Proposition 7

Soit L tel que la norme d'opérateur $\|D^\star D\|_2 \leq L$. Alors pour tout u, α dans $\mathbb{R}^K$ on a $G(\alpha, u) \geq 0$.

On choisit à partir de maintenant L tel que $\|D^\star D\|_2 \leq L$. On a donc pour tout $\alpha \in \mathbb{R}^K$ et tout $u \in \mathbb{R}^K$

$$F_\lambda(\alpha) \leq F_\lambda(\alpha) + G(\alpha, u). \qquad (6)$$

Calculons, en utilisant (5), pour tout $\alpha \in \mathbb{R}^K$ et $u \in \mathbb{R}^K$

$$\begin{aligned}
& F_\lambda(\alpha) + G(\alpha, u) \\
= \; & \frac{1}{2}|D\alpha - x|_2^2 + \lambda|\alpha|_1 + \frac{L}{2}|\alpha - u|_2^2 - \frac{1}{2}|D\alpha - Du|_2^2 \qquad (7) \\
= \; & L\left(\frac{1}{2}\left|\alpha - (u + \frac{1}{L}(D^*x - D^*Du))\right|_2^2 + \frac{\lambda}{L}|\alpha|_1\right) \\
& + \frac{1}{2}|x|_2^2 + L\frac{|u|_2^2}{2} - \frac{1}{2}|Du|_2^2 - \frac{1}{2}\left|u + \frac{1}{L}(D^*x - D^*Du)\right|_2^2 \qquad (8)
\end{aligned}$$

Il est possible d'obtenir à u fixé $\min_{\alpha \in \mathbb{R}^K} F_\lambda(\alpha) + G(\alpha, u)$ en un seul calcul.

Proposition 8

Soit $u \in \mathbb{R}^K$ fixé.
$\min_{\alpha \in \mathbb{R}^K} F_\lambda(\alpha) + G(\alpha, u)$ *est atteint pour* $\alpha = S_{\frac{\lambda}{L}}\left(u + \frac{1}{L}(D^*x - D^*Du)\right)$.

On considère maintenant l'algorithme suivant appelé « Algorithme de seuillage itératif »

1. Initialisation : $m := 0$, on choisit α^0 quelconque.
2. Par récurrence sur $m \in \mathbb{N}^*$ on calcule α^m qui minimise $\alpha \mapsto F_\lambda(\alpha) + G(\alpha, \alpha^{m-1})$ c'est à dire
   (a) on calcule $\beta^m = \alpha^{m-1} + \frac{1}{L}(D^*x - D^*D\alpha^{m-1})$.
   (b) on calcule alors α^m tel que $\alpha^m = S_{\frac{\lambda}{L}}(\beta^m)$.

Théorème 1

Avec les notations qui précèdent

1. *On a pour tout $m \in \mathbb{N}^*$*

$$0 \leq F_\lambda(\alpha^m) \leq F_\lambda(\alpha^m) + G(\alpha^m, \alpha^{m-1}) \leq F_\lambda(\alpha^{m-1}) \tag{9}$$

   En particulier la suite $(F_\lambda(\alpha^m))_{m\in\mathbb{N}}$ est décroissante et converge.
2. *La suite $(\alpha^m)_{m\in\mathbb{N}}$ converge vers $\tilde{\alpha}^\lambda$.*

On admet la démonstration du deuxième point du théorème qui peut être travaillée à partir de l'article de référence [1].

Travaux dirigés

Démonstrations des résultats du cours.

Exercice 1

Compléter et détailler entièrement les propositions 1, 4.

Exercice 2

Détailler entièrement l'exemple qui suit la proposition 3.

Exercice 3

Démontrer la proposition 6.

Exercice 4

1. *Détailler les étapes qui permettent de passer de (7) à (8).*
2. *Démontrer la proposition 8.*
3. *Démontrer la première partie du théorème 1.*

Exercices d'approfondissement.

Exercice 5

Démontrer la proposition 5.

Exercice 6

Soit $D \in \mathbb{C}^{N\times K}$ une matrice à coefficients complexes, $x \in \mathbb{C}^N$ et $\lambda > 0$.

Soit $\mathcal{F}_\lambda(\alpha) = |D\alpha - x|_2^2 + \lambda|\alpha|_2^2$
Calculer $\tilde{\alpha}$ tel que $\mathcal{F}_\lambda(\tilde{\alpha}) = \min_{\alpha\in\mathbb{C}^K} \mathcal{F}_\lambda(\alpha)$

Exercice 7

Soit D une matrice de dictionnaire de taille $N \times K$ et un paramètre $\tau > 0$.

On considère le problème LASSO (sélection de variables en statistiques) qui s'écrit

$$\begin{aligned} &\textit{Trouver } \tilde{\alpha} \in \mathbb{R}^K \textit{ tel que} \\ |D\tilde{\alpha} - x|_2 \quad &= \min\{|D\alpha - x|_2 : \alpha \in \mathbb{R}^K \textit{ et } |\alpha|_1 \leq \tau\} \end{aligned} \tag{10}$$

Montrer que si $\tilde{\alpha}$ est l'unique solution du problème (3) pour $\varepsilon > 0$ donné alors il existe $\tau > 0$ tel que $\tilde{\alpha}$ est l'unique solution du problème (10).

Exercice 8

Soit $D = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix}$. Montrer que le vecteur $\tilde{\alpha} = (1, e^{i2\pi/3}, e^{i4\pi/3})^t$ est l'unique minimiseur de $|\alpha|_1$ tel que $D\alpha = D\tilde{\alpha}$. On voit ici qu'un vecteur qui minimise une norme ℓ^1 n'est pas toujours parcimonieux.

Références

[1] Daubechies, I., Defrise, M. et De Mol, C. (2004) An Iterative Thresholding Algorithm for Linear Inverse Problems with a Sparsity Constraint. Communications on Pure and Applied Mathematics, 57, 1413-1457.

[2] S. Foucart et H. Rauhut : A mathematical introduction to compressive sensing. Birkhauser Boston. (2013) (disponible en ligne gratuitement)

[3] Mallat, Stéphane. *A wavelet tour of signal processing.* Academic Press, (2008)