

TP - Parcimonie

Corrigé

1.b) Le cours sur les processus stationnaires ne considère que des processus discrets infinis ($X[n]$ pour n dans \mathbb{Z}) mais on peut vérifier que les résultats s'adaptent au cas des processus discrets finis ($X[n]$ pour n dans $\{0, \dots, N-1\}$). En particulier, puisque $X = \epsilon \star f$, on a :

$$R_X = R_\epsilon \star f \star \tilde{f}$$

avec $\tilde{f}[n] = \overline{f[-n]}$.

Puisque ϵ est un bruit blanc, $R_\epsilon[k] = 0$ si $k \neq 0$ et $R_\epsilon[0] = \text{Var}(\epsilon_0)$. La variable aléatoire ϵ_0 est de moyenne nulle et de variance p , donc $R_\epsilon = p\delta_0$ et :

$$R_X = pf \star \tilde{f}$$

Puisque $f = (1, 1/4, 0, \dots, 0, 1/4)$, $f \star \tilde{f} = (9/8, 1/2, 1/16, 0, \dots, 0, 1/16, 1/2)$. Cela donne le résultat.

On a :

$$\hat{R}_X[k] = \hat{R}_\epsilon[k] |\hat{f}[k]|^2 = p \left(1 + \frac{1}{2} \cos(2\pi k/N) \right)^2$$

1.f) D'après le cours (adapté au cas des signaux finis), l'erreur quadratique obtenue en débruitant par la méthode de Wiener est :

$$\mathbb{E}(|X[n] - X_{Wiener}[n]|^2) = \frac{1}{N} \sum_{k=0}^{N-1} \frac{\hat{R}_X[k] \hat{R}_n[k]}{\hat{R}_X[k] + \hat{R}_n[k]}$$

Comme \hat{R}_X ne s'approche pas de zéro, on a pour tout k , lorsque σ est petit :

$$\frac{\hat{R}_X[k] \hat{R}_n[k]}{\hat{R}_X[k] + \hat{R}_n[k]} \approx \hat{R}_n[k] = \sigma^2$$

L'erreur quadratique vaut donc à peu près σ^2 lorsque σ est petit. En échelle logarithmique, on observe donc, aux petites valeurs, une droite de pente 2.

Lorsque σ devient plus grand, la courbe s'infléchit. Pour de grandes valeurs, on a, pour tout k :

$$\frac{\hat{R}_X[k] \hat{R}_n[k]}{\hat{R}_X[k] + \hat{R}_n[k]} \approx \hat{R}_X[k]$$

donc l'espérance quadratique converge vers $\frac{1}{N} \sum_{k=0}^{N-1} \hat{R}_X[k] = R_X[0] = 9p/8$.

1.g) On a $X_n = \epsilon \star f + n$. On va utiliser le fait que, si p est petit, ϵ a beaucoup de coefficients nuls.

On commence par reconstruire une approximation de ϵ en appliquant le filtre inverse de f : on choisit g tel que $\hat{g} = (\hat{f})^{-1}$ et on calcule $X_n \star g = \epsilon \star f \star g + n \star g = \epsilon + n \star g$.

Le bruit $n \star g$ est gaussien. On peut calculer sa variance ; elle vaut :

$$R_{n \star g}[0] = R_n \star g \star \tilde{g}[0] = \sigma^2 g \star \tilde{g}[0]$$

On note $\lambda^2 = R_{n \star g}[0]$.

À partir de $\epsilon + n \star g$, on reconstruit ϵ_d , une approximation de ϵ . On le fait par seuillage : si $|\epsilon[k] + n \star g[k]| < \tau$, on va poser $\epsilon_d[k] = 0$. Sinon, on pose $\epsilon_d[k] = \frac{\epsilon[k] + n \star g[k]}{1 + \lambda^2}$.

La difficulté est de bien choisir le seuil. Il faut que la probabilité d'avoir $|\epsilon[k] + n \star g[k]| < \tau$ soit grande si $\epsilon[k] = 0$ mais faible si $\epsilon[k] \neq 0$. Il faut donc avoir τ grand par rapport à λ mais petit par rapport à 1. Le choix $\tau = 3\lambda$ est empiriquement assez bon. On pourrait faire un calcul plus précis mais ce serait plus compliqué.

Une fois calculée ϵ_d , on renvoie $\epsilon_d \star f$.

2.a)

Lemme 0.1. *À la fin de la k -ème itération de la boucle (pour $k \in \{0, \dots, K\}$), le vecteur r peut s'écrire comme la combinaison linéaire d'au plus $K - k$ éléments de la base (b_1, \dots, b_n) .*

Démonstration. On procède par récurrence sur k . Pour $k = 0$, c'est vrai : $r = x$ et, par hypothèse, x est combinaison linéaire d'au plus K éléments de la base.

Supposons que c'est vrai pour $k < K$ et démontrons-le pour $k + 1$. Le r obtenu après la k -ème itération est, par l'hypothèse de récurrence, combinaison linéaire d'au plus $K - k$ éléments de (b_1, \dots, b_n) , c'est-à-dire qu'on a :

$$r = \lambda_1 b_1 + \dots + \lambda_n b_n$$

avec au plus $K - k$ coefficients non-nuls.

Pour tout i , puisque la base est orthonormale, $|\langle r, b_i \rangle| / \|b_i\| = |\lambda_i|$. Notons i_{max} l'indice sélectionné à la $k + 1$ -ème itération. Il doit vérifier :

$$|\lambda_{i_{max}}| = \max_{i \leq n} |\lambda_i|$$

Si tous les λ_i sont nuls, r est nul ; de plus, $\langle r, b_{i_{max}} \rangle = 0$ donc y n'est pas modifié à la $k + 1$ -ème itération et r non plus. À la fin de la $k + 1$ -ème itération, r est donc toujours nul et il est alors bien combinaison linéaire d'au plus $K - (k + 1)$ éléments de la base (b_1, \dots, b_n) .

Si au moins l'un des λ_i est non-nul, alors $\lambda_{i_{max}} \neq 0$. On remplace, dans y , $y_{i_{max}}$ par $y_{i_{max}} + \lambda_{i_{max}}$. Le nouveau r que l'on obtient est donc $r' = r - \lambda_{i_{max}} b_{i_{max}}$. On a alors :

$$r' = \sum_{i \neq i_{max}} \lambda_i b_i$$

La décomposition de r' sur (b_1, \dots, b_n) a donc un coefficient non-nul de moins que la décomposition de r (puisque $\lambda_{i_{max}}$ a été remplacé par zéro et c'était justement un coefficient non-nul). Donc au plus $K - k - 1 = K - (k + 1)$ coefficients de la décomposition de r' sont non-nuls. \square

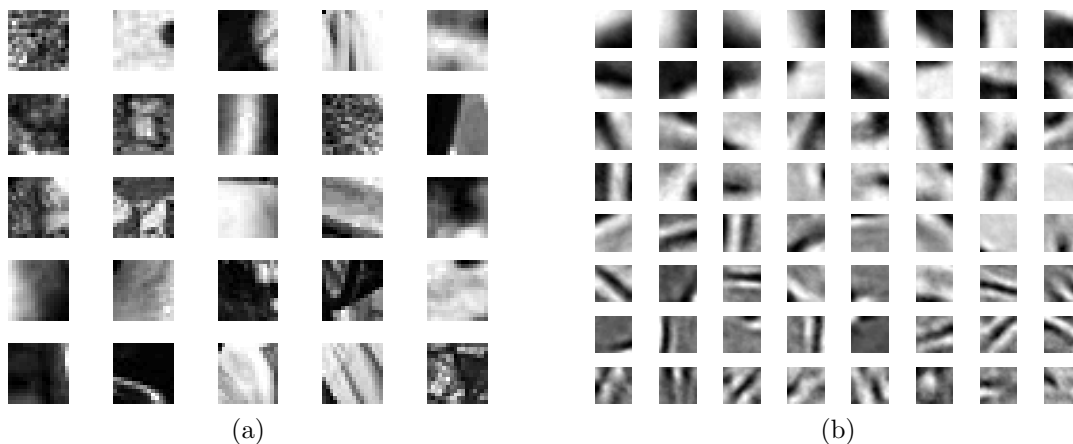


FIGURE 1 – (a) Patches 16×16 extraits de photographies en noir et blanc. (b) Base parcimonieuse correspondante pour $m = 64$ et $K = 6$

Pour $k = K$, le lemme implique qu’après la K -ème itération, $r = 0$. Donc $x = By$. Comme B est une base, y est exactement l’ensemble des coefficients de x dans cette base.

4. Dans toutes les expériences, la valeur de K est assez petite (entre 3 et 8). Le nombre m d’éléments de la base est plus grand. On fait attention à fournir à la fonction `sparsify` beaucoup plus de signaux que le nombre m : de manière très imprécise, la base doit représenter des « points communs » qui reviennent dans un nombre significatif de signaux. Pour les détecter, il faut donc avoir accès à bien davantage de signaux que le nombre de points communs qu’on souhaite trouver.

On commence par des patches 16×16 issus de trois types d’images différentes : des photographies en noir et blanc, des photographies en couleurs et des images de bandes dessinées, en noir et blanc avec des contours très nets. Dans chaque expérience, on utilise 2000 patches. On normalise chaque patch en lui soustrayant sa moyenne puis en le divisant par sa norme.

La base apprise pour les photographies en noir et blanc se trouve sur la figure 1. On voit que les éléments de la base sont des fonctions oscillantes, possédant une direction caractéristique et une fréquence variable. Les fonctions les plus hautes fréquences ont tendance à être un peu localisées en espace. Il s’agit d’une sorte de base d’ondelettes.

Lorsqu’on utilise des patches en couleurs, on retrouve le même genre de fonctions, en un peu moins net. On voit également apparaître des fonctions presque constantes, ou variant assez peu, avec une (éventuellement deux) couleurs caractéristiques.

Sur des images de bandes dessinées, on obtient un résultat similaire mais les contours sont plus nets et les fonctions hautes fréquences sont plus « allongées ». Cela provient probablement du fait que les images de bande dessinées ont une structure proche des images réelles mais avec, justement, des contours beaucoup plus nets (et moins de textures en général).

On a ensuite testé des signaux audio d’une centaine de millisecondes, échantillonnés à 10000 Hz. À nouveau, on a utilisé 2000 signaux pour apprendre la base. On obtient comme éléments de fonctions oscillantes d’amplitude presque constante (figure 4). En calculant leur transformée de Fourier, on voit qu’il s’agit d’un sous-ensemble d’une base de Fourier.

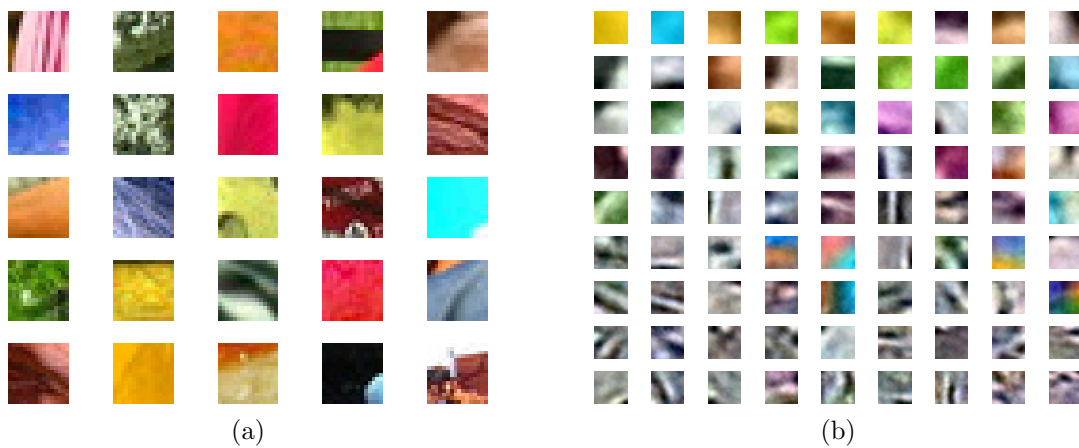


FIGURE 2 – (a) Patches 16×16 extraits de photographies en couleurs. (b) Base parcimonieuse correspondante pour $m = 81$ et $K = 8$

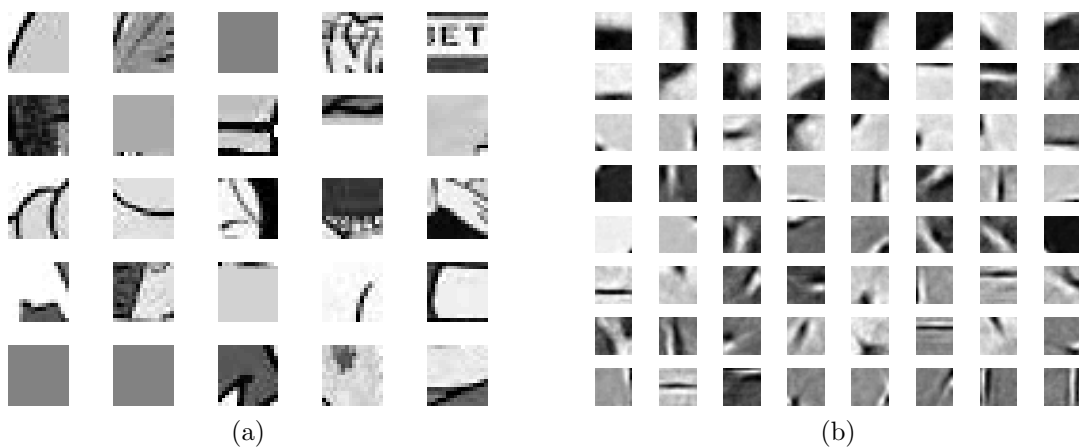


FIGURE 3 – (a) Patches 16×16 extraits d'images de BD en noir et blanc. (b) Base parcimonieuse correspondante pour $m = 64$ et $K = 6$

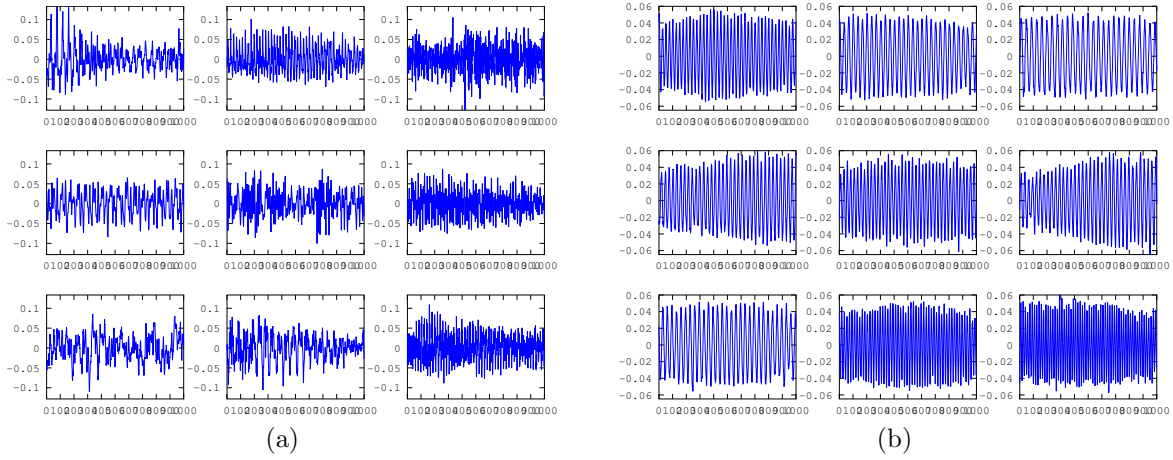


FIGURE 4 – (a) Signaux audio d’environ 100ms (b) Quelques éléments de la base parcimonieuse correspondante pour $m = 64$ et $K = 5$

Enfin, on a fait un essai sur des données météorologiques. Chaque signal, ici, est un élément de \mathbb{R}^{12} , qui contient l’ensemble des températures moyennes par mois mesurées pendant un an dans une station météorologique européenne, au cours d’une année comprise entre environ 1850 et 2015. Il y a 1650 signaux. Le résultat se trouve sur la figure 5.

La base apprise est moins facile à interpréter que dans les cas précédents. De plus, elle dépend fortement de l’initialisation de l’algorithme. On voit généralement apparaître d’abord des courbes en cloche. La combinaison de deux de ces courbes permet d’obtenir une fonction qui semble représenter à peu près la moyenne des températures à l’endroit où est située la station météorologique mais dépend assez peu de l’année.

On voit ensuite apparaître des courbes plus asymétriques qui permettent de décrire les phénomènes météorologiques ponctuels, qui se produisent certaines années mais pas toutes. Par exemple, ici, la sixième courbe permet de représenter un hiver froid suivi, six mois plus tard, d’un automne et d’un hiver plus chauds (ou l’inverse si son coefficient est négatif).

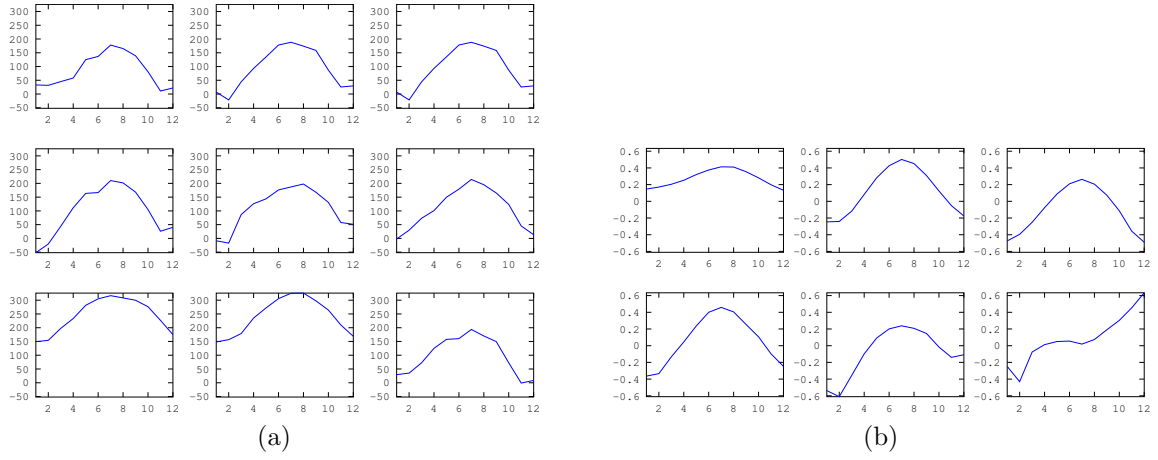


FIGURE 5 – (a) Courbes de températures moyennes par mois (b) Base parcimonieuse correspondante pour $m = 6$ et $K = 3$