Analyse spectrale et corrélatoire empiriques

L'ANALYSE SPECTRALE est un outil qui permet de mettre en évidence des périodicités ou des pseudo-périodicités dans un phénomène observé par l'intermédiaire de signaux mesurés. Plus généralement c'est l'outil d'étude du contenu spectral d'un signal.

Par ailleurs, la notion de signaux ou d'images aléatoires est cruciale dans la pratique du traitement de données car elle permet de modéliser des situations incertaines. Il peut s'agir de modéliser des erreurs de mesures ou de modéliser la connaissance partielle d'une quantité d'intérêt ou encore de modéliser un phénomène intrinsèquement aléatoire.

L'objectif de ce travail est de vous familiariser avec ces idées au travers de l'étude de signaux fournis dans divers fichiers.

- Les fichiers `SigU.mat`, `SigG.mat` et `SigB.mat` contiennent chacun une réalisation simulée de bruit blanc, centré, de puissance unité. Le premier possède une densité de probabilité uniforme, le second une densité gaussienne et le troisième une loi binaire.
- Un fichier `Peaks.mat` contenant un signal à identifier noyé dans un bruit gaussien blanc, centré, de puissance unité.
- `SigElq.mat` contient un enregistrement réalisé [1] sur un système électronique de transmission. L'enregistrement peut contenir du *bruit électromagnétique ambiant* en plus d'une composante *bruit électronique du système*.

Préparation : avant la séance répondez aux questions 4 et 6

1 Premiers éléments de visualisation et d'analyse

Récupérez les fichiers fournis et chargez les signaux sous *Matlab* / *Octave* en utilisant la fonction `load` et par exemple les commandes suivantes. Vous pouvez obtenir des informations sur la syntaxe à utiliser en tapant `help load`.

```
% Nettoyage et chargement des données
    clear all, close all
    load ...
```

Il est vivement conseillé de collectionner les commandes dans un fichier ; vous pourrez ainsi facilement modifier les paramètres d'analyse et relancer vos codes.

1. *Recherchez la taille de ces signaux. Sont-ils des vecteurs lignes ou colonnes ? Des signaux réels ou complexes ?*

2. *Calculez la moyenne et la variance empiriques de chaque signal.*

Réalisez un premier tracé sous la forme temporelle et sous la forme d'histogrammes, en utilisant la séquence de commandes suivante.

1. Remerciements à Guillaume Ferré pour ses enregistrements.

```
% Tracés des signaux
    figure(1),clf
    subplot(211)
        plot(Signal),title('...'),xlabel('Temps'), ylabel('...')
    subplot(212)
        hist(Signal,20),title('...'),xlabel('...'), ylabel('...')
```

3. *Commentez chaque tracé (temporels et histogrammes, lien avec les quantités empiriques calculées précédemment,...).*

2 Caractérisation d'ordre deux

Pour un signal déterministe, la transformée de Fourier est l'outil privilégié de l'analyse fréquentielle. Pour un signal aléatoire, la simple transformée de Fourier d'une trajectoire (d'une réalisation) du signal ne représente pas vraiment son contenu spectral. C'est la densité spectrale de puissance (DSP) qui renseigne sur ce comportement spectral. Pour un signal $X = (X_n)_{n\in\mathbb{Z}}$, à valeur complexe, celle-ci est définie comme la transformée de Fourier de la fonction de covariance :

$$S_X(\nu) = \sum_{n=-\infty}^{+\infty} \gamma_X(n)\, e^{-2j\pi\nu n}\,. \tag{1}$$

La fonction de covariance $\gamma = (\gamma_X(k))_{k\in\mathbb{Z}}$ est-elle même définie par :

$$\gamma_X(k) = \mathrm{E}\big[X_n^\star\, X_{n+k}\big]\,.$$

Cette quantité est définie pour des signaux aléatoires d'ordre deux (pour lesquels $\mathrm{E}\big[|X_n|^2\big]$ existe) et elle est indépendante de n pour des signaux stationnaires. Dans la suite de ce document, pour alléger les notations, $\gamma_X(k)$ est simplement noté γ_k et $S_X(\nu)$ est simplement noté $S(\nu)$.

Remarque 1 — *Au sujet de cette covariance, plusieurs choses sont remarquables.*
- *C'est une fonction à symétrie hermitienne : $\gamma_{-k} = \gamma_k^\star$, donc sa* TF *est réelle.*
- *Son module est maximal en $k = 0$.*
- *C'est une fonction définie et positive, donc sa transformée de Fourier est positive (c'est pourquoi on a pu définir la* DSP *comme la* TF *de la covariance).*

Remarque 2 — *Il est important de noter deux choses concernant la* DSP.
- *Comme la* TF *de tous les signaux à temps discret, si elle existe, elle est périodique de période 1.*
- *Bien que le signal soit à temps discret, la variable fréquentielle ν est continue : $\nu \in \mathbb{R}$.*

4. *Démontrez la propriété évoquée au premier point de la première remarque ci-dessus.*

5. *Rappelez la fonction de covariance et la* DSP *dans le cas d'un bruit blanc.*

Dans les problèmes réels on ne connaît évidemment pas la fonction de covariance... On ne dispose que d'un extrait de taille finie N d'une réalisation du signal aléatoire X : $\boldsymbol{x} = [x_0, x_1, \ldots, x_{N-1}]$. Il est alors impossible de reconstruire exactement la DSP ou la fonction de covariance. Le problème de l'analyse spectrale se ramène à celui de l'**estimation de la DSP** ou de manière équivalente à celui de l'**estimation de la fonction de covariance**. Pour estimer ces caractéristiques d'ordre deux, vous allez utiliser deux techniques d'estimation empiriques, l'une de la DSP et l'autre de la covariance.

3 Covariances empiriques

Dans un premier temps, à partir des signaux mesurés $\boldsymbol{x} = [x_0, x_1, \ldots, x_{N-1}]$, on « estime » la fonction de covariance : on « approche » γ_k par $\widehat{\gamma}_k$ en utilisant un estimateur empirique :

$$\widehat{\gamma}_k = \begin{cases} \dfrac{1}{N} \displaystyle\sum_{n=0}^{N-1-|k|} x_n^\star \, x_{n+|k|} & \text{si } |k| \leqslant N-1 \\ 0 & \text{sinon} \end{cases}$$

c'est-à-dire en « moyennant » des $x_n^\star x_{n+k}$ pour avoir la covariance à l'instant k.

6. *Un peu de travail sur l'estimateur…*

6a. *Expliquez pourquoi la relation précédente est limitée à* $|k| \leqslant N-1$.

6b. *Dénombrez les termes que l'on peut moyenner pour avoir* $\widehat{\gamma}_0$. *Et* $\widehat{\gamma}_{N-1}$ *? Et l'un des* $\widehat{\gamma}_k$ *pour un* k *quelconque ?*

6c. *Question plus délicate. Calculez la moyenne (l'espérance) de l'estimateur* $\widehat{\gamma}_k$ *en fonction de* γ_k. *Pour quelles valeurs de* k *peut-on attendre une erreur d'estimation plus faible et plus importante ?*

Pour calculer numériquement la covariance empirique utilisez la fonction qui vous est fournie et les commandes suivantes.

```
% Calcul de la covariance empirique
    [Temps CovarEmpiric]  = CalcCovar(Signal);
```

7. *Tracez les covariances de chacun des signaux, sans omettre de graduer correctement les axes. Reportez-vous également à la remarque 1 concernant les covariances pour vous assurer de la cohérence des résultats obtenus. Prenez le temps de réfléchir, d'analyser et de commenter chacun des résultats observés.*

8. *Que pouvez-vous dire du signal* `Peaks` *? D'après vous, de quoi s'agit-il ?*

9. *Même question à porpos du signal* `SigElq` *?*

4 Périodogramme

On calcule ensuite la transformée de Fourier de cette fonction de covariance estimée :

$$\widehat{S}(\nu) = \sum_{n=-\infty}^{+\infty} \widehat{\gamma}_n \, e^{-2j\pi\nu n} \,,$$

c'est-à-dire qu'on remplace la vraie covariance γ_n par la covariance empirique $\widehat{\gamma}_n$ dans la relation (1). On obtient ainsi un estimateur de la DSP que l'on appelle le périodogramme.

10. *Autre question plus délicate. Calculez la moyenne (l'espérance) de l'estimateur* $\widehat{S}(\nu)$ *et montrez qu'elle s'écrit comme la convolution de la « vraie* DSP *»* $S(\nu)$ *et d'une certaine fenêtre.*

Remarque 3 — *Une expression du type*

$$F(\nu) = \sum_{n=-\infty}^{+\infty} f_n \, e^{-2j\pi\nu n} \, ,$$

est une transformée de Fourier à temps discret. C'est une fonction 1-périodique de la variable ν continue. Pour la connaître entièrement il suffit de la connaître sur $[0, 1]$ ou sur $[-0.5, 0.5]$ par exemple.

En général, on ne peut pas calculer une telle expression, pour deux raisons. Premièrement la somme s'étend sur une infinité de termes et deuxièmement elle nécessite une infinité de valeurs de ν sur $[0, 1]$ ou sur $[-0.5, 0.5]$. En revanche le calcul devient possible

1. *si f_n possède seulement un nombre fini de valeurs non-nulles et*
2. *si on se contente d'un nombre fini de valeurs de ν.*

Par exemple, si f_n est nul pour $n < 0$ et pour $n \geqslant N$, on peut calculer exactement les valeurs de $F(\nu)$ sur une grille fréquentielle régulière de M valeurs de ν entre 0 et 1 : $\boldsymbol{\nu} = [\nu_0, \nu_1, \ldots, \nu_{M-1}]$, avec $\nu_m = m/M$. En pratique on choisira M assez grand pour avoir « l'impression » du continu et ce, indépendamment de N. Ces valeurs s'obtiennent en utilisant un algorithme de transformée de Fourier rapide. Il est crucial de bien comprendre que la longueur de la grille fréquentielle M est indépendante du nombre d'échantillons du signal N. La syntaxe matlab *à utiliser est* `fft(x,M)`.

Utilisez la suite de commande ci-après pour calculez numériquement le périodogramme.

```
% Calcul du périodogramme (DSP empirique)
    [NuGrid Periodo] = CalcPeriodo(Signal);
```

Vous en arrivez ici aux aspects les plus importants de cette partie du TP.

11. *Tracez les périodogrammes de chacun des signaux. N'omettez pas de graduer correctement les axes, notamment l'axe fréquentiel. Prenez le temps de réfléchir, d'analyser et de commenter chacun des résultats.*

12. *Que pouvez-vous dire du signal* `Peaks` *? D'après vous, de quoi s'agit-il ? Proposez un modèle (temporel) pour ce signal. Calculez sa fonction de covariance et sa* DSP. *Tracez chacun d'eux en le superposant aux estimées.*

13. *Que pouvez-vous dire du signal* `SigElq` *?*