CentraleSupélec
Campus de Metz
Deuxième année - SG6

Méthodes d'estimations et introduction à la théorie moderne du codage
TD 3 & 4
Analyse spectrale non paramétrique

1 Objectifs

Cette étude permet d'évaluer la compétence C1 (voir la deuxième partie du travail à effectuer ci-dessous).

Les objectifs de cette étude, qui correspond à des « devoirs à la maison » et deux travaux dirigés, sont :

— estimer la DSP de signaux ergodiques avec les méthodes proposées (périodogramme, périodogramme moyenné) et leurs améliorations possibles par fenêtrage ;
— observer l'influence des paramètres (nombre d'échantillons par tranche, nombre de tranches, fenêtre, taux de recouvrement des tranches, ...) ;
— évaluer les performances des estimateurs par des études statistiques ;
— comparer les méthodes d'analyse spectrale non paramétriques ;
— présenter dans un rapport des résultats expérimentaux obtenus en accordant un soin particulier aux points suivants :
  - clarté, précision, cohérence et concision ;
  - logique de la démarche scientifique :
    - pourquoi avez-vous choisi de faire telle mesure ou telle série de mesure ?
    - description *précise* de la mesure,
    - présentation des résultats,
    - analyse des résultats pour répondre (ou non) à la question initiale ;
  - choix et description (échelle — linéaire ou logarithmique —, légende, quelles grandeurs apparaissent sur les axes) des courbes (pour comparer plusieurs courbes il est préférable de les regrouper sur un même graphe, le minimum est d'avoir les mêmes échelles — à moins que seule la forme des courbes ne soit comparée —).

2 Travail à effectuer

!! IMPORTANT !!

Ne pas imprimer, ni donner à visualiser beaucoup de courbes. En choisir quelques unes de *significatives* et résumer les observations. Observer les DSP en échelle linéaire et en échelle log (dB) sur l'axe vertical.

Vous pouvez utiliser le langage de programmation de votre choix (Matlab, Python, ...). Joignez à votre compte-rendu les scripts permettant de tracer les figures de votre rapport. **Vous trouverez beaucoup de scripts Matlab et des indications supplémentaires sur le cours edunao.**

Première partie (estimateurs de la fonction d'autocorrélation)

1. Programmer les deux estimateurs $\hat{\gamma}_{x1}$ et $\hat{\gamma}_{x2}$ présentés à l'annexe B équations (2) et (3), avec la structure suivante.
   — En entrée : le signal $(x(n))_{0\leq n<L}$, doubles codés en binaire.

— En sortie : les coefficients $(\hat{\gamma}_x)_{-M\leq k\leq M}$, doubles codés en binaires.
— paramètres de la ligne de commande :
  — le nombre N d'échantillons pris en compte, entier (int)
  — l'indice i (long) $(0 \leq i < L - N)$ du premier échantillon à prendre en compte,
  — le nombre M (int) spécifiant l'horizon de calcul,
  — le choix estimateur de Blackman-Tukey / estimateur de Bartlett.

En Matlab, on pourra utiliser la fonction `xcorr` avec les options *biased* ou *unbiased*.

Voir une solution possible (en Matlab). Vous pouvez tester ce programme sur un bruit blanc gaussien unitaire (i.e., centré et de variance 1) avec les instructions :

```
» N=1024; i=1; M=1023;choix_estim='bar';
» signal = randn(N,1);
» gamma_bar = question1(signal, M, i, N, choix_estim);
» plot(-M:M,gamma_bar);
```

2. Tester ce programme sur un bruit blanc gaussien centré, de variance 10^{-1}, vérifier que $\hat{\gamma}_x(0)$ vaut 10^{-1}.

3. Mesures du biais et de la variance des deux estimateurs $\hat{\gamma}_{xi}(k)$ (pour $-M \leq k \leq M$ et $i = 1, 2$) sur différentes tranches d'un signal pseudo-aléatoire gaussien centré et de variance 10^{-1}.

   Les statistiques sont faites comme si chaque tranche est le résultat d'une expérience aléatoire. Par exemple, si vous estimez $2M + 1$ valeurs de la fonction d'autocorrélation d'un bruit blanc gaussien à partir d'une tranche de N échantillons, l'expérience aléatoire consiste à choisir une tranche de N échantillons du bruit blanc gaussien, à lui appliquer le programme qui estime les $2M+1$ valeurs de la fonction d'autocorrélation et à observer les $2M+1$ valeurs trouvées (issue de l'expérience aléatoire). Si vous recommencez K fois l'expérience sur K tranches différentes du bruit blanc gaussien, vous obtenez K tableaux de $2M + 1$ réels (doubles codés en binaire). Écrire une fonction qui estime la moyenne, la variance et l'écart-type empiriques du k-ème échantillon $(1 \leq k \leq 2M + 1)$ à partir des K valeurs. Pour avoir des estimations de moyennes qui soient satisfaisantes, il faut $K \geq 16$.

   Voir une solution possible (en Matlab). Vous pouvez tester ce programme sur un bruit blanc gaussien centré de variance 2 avec les instructions :

```
» N=1024;K=32;variance_bb=2;M=N-1;choix_estim='Blackman-Tukey';
» signal = sqrt(variance_bb)*randn(N,K);
» [moy_estim, var_estim] = question3(signal, choix_estim);
» gamma_th = [zeros(M,1);variance_bb;zeros(M,1)]; % fonction d'autocorrélation
    % théorique
» biais_estim = moy_estim-gamma_th;
» eqm_estim = biais_estim.^2+var_estim; % erreur quadratique moyenne estimée
» figure(1), plot(-M:M,biais_estim)
» title(['Estimation du biais de l''estimateur de ',choix_estim])
» figure(2), plot(-M:M,var_estim)
» title(['Estimation de la variance de l''estimateur de ',choix_estim])
» figure(3), plot(-M:M,eqm_estim)
» title(['Estimation de l''eqm de l''estimateur de ',choix_estim])
```

4. Donner l'estimée de la densité spectrale de puissance de ce bruit avec la méthode de Blackman-Tukey, mesurer la variance et la moyenne de la DSP. Pour cela, faîtes une fonction qui calcule $\hat{\Gamma}_x(\nu)$ défini en (4), puis calculez la moyenne et la variance empirique des échantillons obtenus. Attention à bien gérer les valeurs négatives de k quand vous appliquez une fonction `fft` pour calculer le membre de droite de l'égalité (4).

   Voir une solution possible (en Matlab). Vous pouvez tester ce programme sur un bruit blanc gaussien centré de variance 2 avec les instructions :

```
» N=1024;K=32;variance_bb=2;M=512;
» signal = sqrt(variance_bb)*randn(N,K);
» [DSP_estim, DSP_moy_estim, DSP_var_estim] = question4(signal,M);
» figure(4), plot((0:2*M)/(2*M+1),DSP_estim(:,1))
» figure(4), title(['Estimation de la DSP (Blackman-Tukey, N=',num2str(N),', M=',num2str(M),')'])
» figure(4), xlabel('fréquence réduite')
» figure(5), plot((0:2*M)/(2*M+1),DSP_moy_estim)
» figure(5), title(['Moyenne d''estimations de DSP (Blackman-Tukey, ...
    N=',num2str(N),', M=',num2str(M),')'])
» figure(5), xlabel('fréquence réduite')
» figure(6), plot((0:2*M)/(2*M+1),DSP_var_estim)
» figure(6), title(['Variance d''estimations de DSP (Blackman-Tukey, N=',...
   num2str(N),', M=',num2str(M),')'])
» figure(6), xlabel('fréquence réduite')
» figure(6), hold on
» figure(6), plot((0:2*M)/(2*M+1),DSP_moy_estim.^2)
» figure(6), legend('Var($\hat{\Gamma}_x(\nu)$)',...
   '$(E[\hat{\Gamma}_x(\nu)])^2$','Interpreter','latex')
```

Deuxième partie (périodogramme et ses dérivés)
Cette 2ème partie du DM, et en particulier la question 4, permettent d'évaluer la compétence C1.
Générer les échantillons d'un signal, appelé `signal1` dans la suite, issu d'un filtre de Chebyshev d'ordre 3, de fréquence de coupure normalisée $\nu_c = 0,1$ et d'ondulations inférieures à $0,5\,\text{dB}$, attaqué par un bruit blanc gaussien centré et de variance $0,5$. La fonction de transfert du filtre causal est donnée par :

$$G(z) = \frac{0,0154 + 0,0461z^{-1} + 0,0461z^{-2} + 0,0154z^{-3}}{1 - 1,9903z^{-1} + 1,5717z^{-2} - 0,458z^{-3}}.$$

Calculer la réponse impulsionnelle de ce filtre, puis le carré du module de sa TFD qui vous servira de modèle. On ne s'intéressera qu'à la partie du spectre supérieure à $-60\,\text{dB}$. Voir une solution possible en Matlab.

1. *Périodogramme simple.*

   Observez les résultats obtenus avec le périodogramme simple pour différentes valeurs de N. Faire les mesures pour $N = 128$, 256 et 512 sur différentes tranches de `signal1`. Même chose avec un bruit blanc, que l'on appellera `bruit`. Évaluer le biais et la variance de l'estimateur en fonction de N. Pour évaluer le biais, il est nécessaire de soustraire la valeur exacte (modèle) à l'estimation. Commenter les résultats. Voir une solution possible en Matlab.

   Calculer les valeurs théoriques de l'estimée de la DSP par l'estimateur du périodogramme

$$\hat{\Gamma}_x(\nu) = \frac{\left|\sum_{n=0}^{N-1} x(n)e^{-2i\pi n\nu}\right|^2}{N}, \tag{1}$$

   pour un signal sinusoïdal $x(n) = A\sin(2\pi\nu_0 n+\varphi)$ ($\forall n \in \mathbb{Z}$)—où $\varphi \in \mathbb{R}$ est quelconque—de fréquence réduite $\nu_0 = p_0/N$ ($p_0 \in \{1, 2, \ldots, N-1\}$), en utilisant une TFD sur N points[1]. Même

1. On remarquera que dans ce cas, l'estimateur (1) ne correspond pas à la TFD sur N points de l'estimateur de Bartlett.

question pour le périodogramme moyenné sur K tranches de N points. Voir le transparent 15 du cours d'analyse spectrale.

Calculer les valeurs théoriques de l'estimée de la DSP par la méthode du périodogramme (1) pour un signal sinusoïdal $x(n) = A\sin(2\pi\nu_0 n + \varphi)$ ($\forall n \in \mathbb{Z}$)—où $\varphi \in \mathbb{R}$ est quelconque— de fréquence réduite $\nu_0 = p_0/N$ ($p_0 \in \frac{1}{2} + \{1, 2, \ldots, N-1\}$), en utilisant une TFD sur N points. Même question pour le périodogramme moyenné sur K tranches de N points. Voir le transparent 16 du cours d'analyse spectrale.

2. *Périodogramme moyenné.*

   Générer 7500 échantillons d'un signal, appelé `signal2` dans la suite, constitué d'une sinusoïde à 140 Hz d'amplitude $\sqrt{2}$, d'une sinusoïde à 180 Hz d'amplitude $\sqrt{2}/100$ et d'un bruit blanc d'écart-type 0,08, le tout échantillonné à une fréquence d'échantillonnage de 1024 Hz. Voir une solution possible en Matlab.

   Calculer le périodogramme moyenné sur les signaux `signal1`, `bruit` et `signal2` tronqués à 1050 points avec des tranches de 512, 256, 128 et 64 points [2]. Comparer au modèle exact [3] pour chaque courbe et aux résultats de l'estimateur simple. Le tableau `signal2` ayant environ 7500 échantillons, recommencer les opérations sur différents horizons d'observations de 1050 points. Commenter.

3. *Périodogramme modifié.*

   Même chose qu'en 2., en utilisant différentes fenêtres (Hamming, ou autres).

4. **Cette question est importante pour l'évaluation de la compétence C1**. Analyser le signal dans le fichier `signal` (des sinusoïdes dans un bruit rose).

A Rappels

Soit $(x(n))_{n\in\mathbb{Z}}$ un signal aléatoire réel centré ergodique au deuxième ordre — donc stationnaire au sens large —. Il admet une fonction d'autocorrélation :

$$\gamma_x(k) = \mathrm{E}[x(n+k)x(n)] = \lim_{N\to\infty} \frac{1}{2N} \sum_{n=-N}^{N} x(n+k)x(n).$$

La densité spectrale de puissance du signal x vaut :

$$\Gamma_x(\nu) = \sum_{k=-\infty}^{+\infty} \gamma_x(k) e^{-2j\pi k\nu}.$$

Propriétés

$$\gamma_x(0) = \int_{\alpha}^{\alpha+1} \Gamma_x(\nu)\, d\nu = \mathrm{Var}[x(n)], \qquad \text{pour tout } \alpha \in \mathbb{R}\,.$$

Si $A(z)$ est la fonction de transfert d'un filtre linéaire invariant dans le temps, stable EB-SB, attaqué par un signal aléatoire du deuxième ordre $x(n)$, alors la sortie $y(n)$ est un signal aléatoire du deuxième ordre et l'on a la formule des interférences

$$\Gamma_y(\nu) = |G(\nu)|^2 \Gamma_x(\nu),$$

où $G(\nu) = A(e^{2j\pi\nu})$ est la réponse fréquentielle du filtre.

2. Voir une solution possible en Matlab de calcul de périodogramme moyenné appliquée à signal1 et à signal2.

3. Pour signal2 la DSP théorique dépend de la durée de l'horizon d'observation (voir les transparents 13 et 14 du cours d'analyse spectrale et une solution possible pour la calculer en Matlab).

B Estimation de la fonction d'autocorrélation.

Il existe les deux estimateurs suivants.

— Celui de Blackman-Tukey, défini par

$$\hat{\gamma}_{x1}(k) = \frac{1}{N-|k|} \sum_{n=0}^{N-|k|-1} x(n+|k|)x(n) \quad \text{pour } -N < -M \leq k \leq M < N. \tag{2}$$

L'estimateur est sans biais : $\mathrm{E}[\hat{\gamma}_{x1}(k)] = \gamma_x(k)$.

— Celui de Bartlett, défini par

$$\hat{\gamma}_{x2}(k) = \frac{1}{N} \sum_{n=0}^{N-|k|-1} x(n+|k|)x(n) \quad \text{pour } -N < -M \leq k \leq M < N. \tag{3}$$

Dans ce cas,

$$|\mathrm{E}[\hat{\gamma}_{x2}(k)] - \gamma_x(k)| = \frac{|k|}{N}\gamma_x(k),$$

l'estimateur est biaisé pour $k \neq 0$.

G. M. Jenkins et D. G. Watts ont montré que dans de nombreux cas, l'erreur quadratique moyenne de l'estimateur biaisé de Bartlett est inférieure à celle de l'estimateur sans biais de Blackman-Tukey, pour un même nombre N d'échantillons observés. Leurs travaux ont été publiés dans "Spectral Analysis and its applications", Holden-Day, San Francisco 1969.

C Estimation de la densité spectrale de puissance

C.1 Estimateur de Blackman-Tukey

Il est défini par

$$\hat{\Gamma}_x(\nu) = \sum_{k=-M}^{M} \hat{\gamma}_x(k) e^{-2j\pi k\nu}, \tag{4}$$

où $\hat{\gamma}_x(k)$ vaut soit $\hat{\gamma}_{x1}(k)$, soit $\hat{\gamma}_{x2}(k)$. C'est la transformée de Fourier d'un des précédents estimateurs de la fonction d'autocorrélation, prolongé par 0 pour $|k| > M$.

Considérons le cas $M = N-1$

$$\hat{\gamma}_{x2}(k) = \begin{cases} \frac{1}{N}\sum_{n=0}^{N-|k|-1} x(n+|k|)x(n) & \text{pour } -N < k < N \\ 0 \quad \text{ailleurs} \end{cases}.$$

En posant

$$x_t(n) = \begin{cases} x(n) & \text{pour } 0\ n < N \\ 0 & \text{ailleurs} \end{cases} \quad \text{et} \quad \tilde{x}_t(n) = x_t(-n),$$

il est facile de vérifier que

$$\hat{\gamma}_{x2}(k) = \frac{1}{N}\big(x_t * \tilde{x}_t\big)(k) \quad \text{pour } k \in \mathbb{Z}\,.$$

Donc

$$\hat{\Gamma}_{x2}(\nu) \stackrel{\text{def}}{=} \sum_{k=-\infty}^{+\infty} \hat{\gamma}_{x2}(k) e^{-2j\pi k\nu} = \frac{1}{N}|X_t(\nu)|^2,$$

où

$$X_t(\nu) = \sum_{k=0}^{N-1} x(k) e^{-2j\pi k\nu}$$

est la TF du signal tronqué $(x_t(n))_{n\in\mathbb{Z}}$. Cet estimateur de la densité spectrale de puissance est appelé *périodogramme* ou *estimateur simple.*

C.2 Estimateur simple ou périodogramme

Il est défini par

$$\hat{\Gamma}_{x2}(\nu) = \frac{1}{N}|X_t(\nu)|^2.$$

En pratique, on utilise une TFD sur N points. En général N est une puissance de 2 et la TFD est calculée avec un algorithme de TFR (Transformée de Fourier Rapide).

Calcul du biais

D'après ce qui précède,

$$\begin{aligned}
\mathrm{E}[\hat{\Gamma}_{x2}(\nu)] &= \sum_{k=-(N-1)}^{N-1} \mathrm{E}[\hat{\gamma}_{x2}(k)]e^{-2j\pi k\nu} \\
&= \sum_{k=-(N-1)}^{N-1} \left(1 - \frac{|k|}{N}\right) \gamma_x(k)e^{-2j\pi k\nu} \\
&= \Gamma_x(\nu) * W_T(\nu),
\end{aligned}$$

où $W_T(\nu)$ est la TF de la fenêtre triangulaire centrée en 0 et de largeur $2N+1$. Le biais du périodogramme est donc donné par la relation :

$$\mathrm{B}\left(\hat{\Gamma}_{x2}(\nu)\right) = \Gamma_x(\nu) * \big(W_T(\nu) - \delta(\nu)\big),$$

où δ est la distribution de Dirac en 0.

Calcul de la covariance.

On montre que si $(x(n))_{n\in\mathbb{Z}}$ est un bruit blanc gaussien centré et de variance σ^2, alors la covariance de $\hat{\Gamma}_{x2}(\nu_1)$ et $\hat{\Gamma}_{x2}(\nu_2)$ est donnée par la relation :

$$\mathrm{cov}\left(\hat{\Gamma}_{x2}(\nu_1), \hat{\Gamma}_{x2}(\nu_2)\right) = \sigma^4 \left[\left(\frac{\sin N\pi(\nu_1+\nu_2)}{N\sin\pi(\nu_1+\nu_2)}\right)^2 + \left(\frac{\sin N\pi(\nu_1-\nu_2)}{N\sin\pi(\nu_1-\nu_2)}\right)^2\right].$$

Cette relation donne la variance de l'estimateur simple dans le cas d'un bruit blanc gaussien ; il suffit d'y faire tendre ν_2 vers $\nu_1 = \nu$:

$$\mathrm{var}[\hat{\Gamma}_{x2}(\nu)] = \sigma^4 \left[1 + \left(\frac{\sin 2N\pi\nu}{N\sin 2\pi\nu)}\right)^2\right].$$

Il résulte de ces équations deux choses :

— la variance de l'estimateur simple tend par valeurs supérieures vers $\sigma^4 > 0$ pour $\nu \neq 0$ et $\nu \neq 1/2$ et vers $2\sigma^4$ pour $\nu = 0$ ou $\nu = 1/2$.

— Pour $\nu_1 = i/N$ et $\nu_2 = j/N$, correspondant au calcul de la TFD sur N points de $(x(n))_{0\leq n<N}$, la covariance de $\hat{\Gamma}_{x2}(i/N)$ et de $\hat{\Gamma}_{x2}(j/N)$ est nulle pour $i \neq j$. Autrement dit, $\hat{\Gamma}_{x2}(i/N)$ et $\hat{\Gamma}_{x2}(j/N)$ sont non corrélés pour $i \neq j$.

En conclusion, le périodogramme est un estimateur non consistant, donc peu performant. Sa qualité, importante, est qu'il est facile et rapide à calculer, grâce à la TFR.

C.3 Périodogramme moyenné

Pour améliorer les performances du périodogramme en diminuant sa variance, l'idée naturelle est de faire la moyenne de plusieurs résultats du périodogramme, obtenus sur différentes tranches du signal.

Supposons que $x(n)$ soit observé pour $0 \leq n < KN = L$ (L, l'horizon d'observation est découpé en K tranches de N échantillons successifs). Considérons alors, l'estimateur suivant :

$$\hat{\Gamma}_x(\nu) = \frac{1}{K}\sum_{k=0}^{K-1}\frac{|X_{tk}(\nu)|^2}{N}$$

avec

$$X_{tk}(\nu) = \sum_{n=0}^{N-1} x(kN+n)e^{-2j\pi n\nu},$$

c'est le *périodogramme moyenné*. Il est possible également de faire chevaucher les tranches, augmentant ainsi K et N à L fixé. Comme pour le périodogramme simple, on utilise en général une TFD sur N points. Si deux tranches consécutives du signal observé ne sont pas (ou sont peu) corrélées, alors le périodogramme moyenné a une variance K fois plus faible que celle du périodogramme simple. D'un autre côté, plus les tranches sont petites, plus le périodogramme calculé sur une tranche est biaisé. Un compromis est à trouver entre K et N quand L est fixé. Le périodogramme moyenné est un estimateur consistant, sa variance et son biais tendent vers 0 quand N et K tendent simultanément vers l'infini. Pour diminuer le biais du périodogramme simple ou moyenné, dû à la durée finie de l'observation, on peut fenêtrer chaque tranche de N échantillons par une fenêtre adéquate (Hanning, Hamming, Blackman ...), avant d'en calculer le périodogramme simple ou moyenné. Ce nouvel estimateur est appelé *périodogramme modifié* ou *estimateur de Welch*. Toujours dans le but d'améliorer les qualités du périodogramme, on peut filtrer le signal (fréquentiel) obtenu comme résultat du périodogramme, par un filtre passe-bas ; ceci est équivalent à un fenêtrage de l'estimation de la fonction d'autocorrélation. Seules certaines fenêtres sont acceptables : celles dont la TF est positive (comme la fenêtre triangulaire), afin d'assurer une estimée *positive* de la DSP. Ce nouvel estimateur est appelé *périodogramme adouci*, nous ne l'étudierons pas ici en pratique. Remarquons seulement qu'il est trois fois plus coûteux en TFR que les autres périodogrammes (pour chaque tranche, il faut une TFR inverse, la multiplication par une fenêtre et une TFR directe en plus).

C.4 Périodogramme modifié

L'estimateur est le même que le périodogramme moyenné, à la différence près que chaque tranche de N échantillons du signal $(x(n))_{0\leq n<L}$ est fenêtrée par $(w(n))_{0\leq n<N}$ d'horizon N, avant d'en calculer la TF :

$$\hat{\Gamma}_x(\nu) = \frac{1}{K}\sum_{k=0}^{K-1}\frac{|Y_k(\nu)|^2}{N}$$

avec

$$Y_k(\nu) = \sum_{n=0}^{N-1} y(kN+n)e^{-2j\pi n\nu} \quad \text{et} \quad y(kN+n) = x(kN+n).w(n),$$

où $w(n)$ est une fenêtre d'horizon $[0\,;N-1]$ centrée autour de $N/2$ si N est pair ou de $(N-1)/2$ si N est impair (voir F. J. Harris, "On the use of windows for harmonic analysis with the discrete fourier transform", *Proceedings of the IEE*, Vol. 66, No. 1, Jan. 1978).