Projet : Analyse de Fourier, ACP et AFC

Consignes :

- **Résoudre** les différents exercices à l'aide du **logiciel R sous R-studio**
- Ecrire les **scripts** pour chaque traitement
- Présenter les résultats dans un **pdf** avec tous les tracés et tableaux de calcul
- **Commenter** et interpréter les résultats lors de la soutenance

1 Série de Fourier (SF)

1.1 Formulation générale :

Une fonction réelle f (ou signal) de période T, C^1 par morceaux sur tout intervalle $[\alpha,\ \alpha + T]$, $\alpha \in \mathbb{R}$ quelconque, se développe en une série de Fourier de la forme :

$$f(t) = a_0 + \sum_{n=1}^{+\infty} a_n \cos(n\omega t) + b_n \sin(n\omega t)$$

avec ω la pulsation du signal.

NB : Aux points de discontinuité, $f(t)$ est remplacée par $\frac{1}{2}(f(t^+) + f(t^-))$

Vocabulaire :

1) Une fonction périodique se répète dans le temps. Si nous mesurons le temps nécessaire au déroulement complet d'un cycle nous obtenons la période de notre signal. Cette période sera mesurée en secondes et s'écrira : T
2) La fréquence d'une fonction péridique est le nombre de cycles complets qui se sont déroulés pendant 1 seconde. On la note F, on a $F = \frac{1}{T}$, (unité en $s^{-1} = Hz$)
3) $\omega = \frac{2\pi}{T} = 2\pi F$: la pulsation du signal (en $rad.s^{-1}$). Elle correspond à la vitesse angulaire. Par exemple si $T = \frac{1}{2}$, en $1s$ le signal fait 2 cycles et a une vitesse angulaire de $4\pi\ rad.s^{-1}$ (2 tours)

On notera :

- f est C^1 par morceaux sur $[a, b]$: f est continue, dérivable, de dérivée continue sur $[a, b]$ sauf sur un nombre fini de points x en lesquels f posséde une limite à gauche et à droite notées $f(x^-)$ et $f(x^+)$

- $a_0 = \dfrac{1}{T}\displaystyle\int_{\alpha}^{\alpha+T} f(t)dt$: valeur moyenne de f sur une longeur de période
- $a_n = \dfrac{2}{T}\displaystyle\int_{\alpha}^{\alpha+T} f(t)\cos(n\omega t)dt$
- $b_n = \dfrac{2}{T}\displaystyle\int_{\alpha}^{\alpha+T} f(t)\sin(n\omega t)dt$

avec $n \geqslant 1$ entier et α : un réel quelconque (on pose souvent $\alpha = 0$ ou $\alpha = \dfrac{-T}{2}$)

Cas particuliers :

1) Si f est **paire** : $a_0 = \dfrac{1}{T}\displaystyle\int_{0}^{\frac{T}{2}} f(t)dt$, $a_n = \dfrac{4}{T}\displaystyle\int_{0}^{\frac{T}{2}} f(t)\cos(n\omega t)dt$, $b_n = 0$
2) Si f est **impaire** : $a_0 = 0$, $a_n = 0$, $b_n = \dfrac{4}{T}\displaystyle\int_{0}^{\frac{T}{2}} f(t)\sin(n\omega t)dt$

1.2 Forme complexe des SF

soit f de période T développable en série de Fourier :

$$f(t) = a_0 + \sum_{n=1}^{+\infty} a_n\cos(n\omega t) + b_n\sin(n\omega t)$$

On peut écrire aussi :

$$f(t) = \sum_{n=-\infty}^{+\infty} c_n e^{in\omega t}, \quad \text{avec } c_n = \frac{1}{T}\int_{\alpha}^{\alpha+T} f(t)e^{-in\omega t}dt$$

NB : pas de parité pour la série complexe (à l'exception de c_0) car une fonction complexe n'est ni paire, ni impaire ($t \to f(t)e^{-in\omega t}$ est complexe)

1.3 Harmonique, spectre de fréquence

1) Pour la série de Fourier réelle :

   La décomposition d'un signal en série de Fourier peut se présentée sous forme d'un diagramme, appel'e spectre en fréquence (ou d'amplitude), dont l'abscisse est $n\omega$ et l'ordonnée $A_n = \sqrt{a_n^2 + b_n^2}$, amplitude de la composante de pulsation $n\omega$, $n \in \mathbb{N}$.

   La composante $n = 1$ s'appelle le fondamental. Pour $n > 1$, les composantes s'appellent les harmoniques de rang n

2) Pour la série de Fourier complexe :

   La décomposition d'un signal en série de Fourier complexe se présente aussi sous forme d'un diagramme mais symétrique par rapport à (Oy), le spectre en fréquence (ou d'amplitude), est obtenue par les abscisses $n\omega$ et les ordonnées $A_n = \dfrac{\sqrt{a_n^2 + b_n^2}}{2} = |c_n| = |c_{-n}|$, amplitude de la composante de pulsation $n\omega$, $n \in \mathbb{Z}$.

1.4 Exemple

Soit le signal f de période T défini par :

$$f(t) = \begin{cases} -A, & t \in [-\frac{T}{2}, 0[\\ A, & t \in [0, \frac{T}{2}[\end{cases}$$

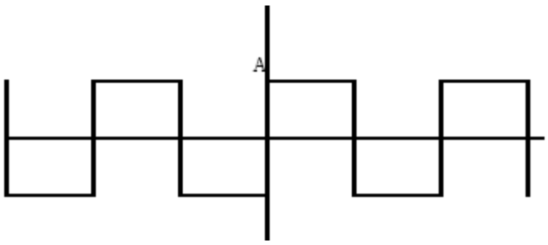


FIGURE 1 – *Signal initial temporel*

On calcule le signal comme série de Fourier, on trouve :

$$f(t) = \frac{4A}{\pi} \sum_{n=0}^{+\infty} \frac{\sin((2n+1)\omega t)}{2n+1}$$

En note que $a_0 = 0$, $a_n = 0$, $b_{2n+1} = \dfrac{4A}{(2n+1)\pi}$ et $b_{2n+2} = 0$

Représentations graphiques sur une période pour $A = 1$ et $T = 2$ de la somme des premiers termes de la série :

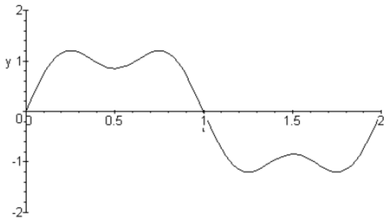


FIGURE 2 – *Somme des deux premiers termes*

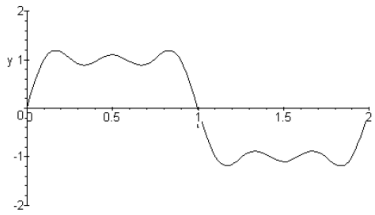


FIGURE 3 – *Somme des trois premiers termes*

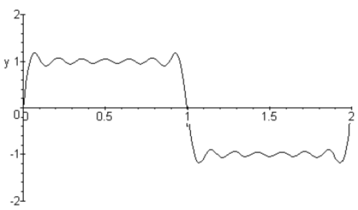


FIGURE 4 – *Somme des six premiers termes*

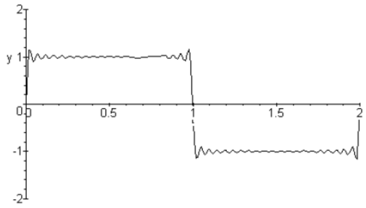


FIGURE 5 – *Somme des vingt premiers termes*

Remarque : On peut écrire

$$f(t) = \frac{4A}{\pi} \sum_{n=0}^{+\infty} \frac{\cos((2n+1)\omega t - \frac{\pi}{2})}{2n+1}$$

La phase est constante pour chaque pulsation

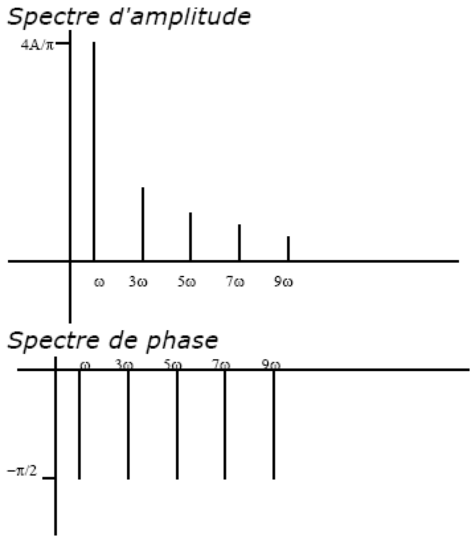


FIGURE 6 – *Spectres d'amplitude et de phase*

L'écriture en série de Fourier complexe du signal donne :

$$f(t) = \frac{-2iA}{\pi} \sum_{n=-\infty}^{+\infty} \frac{e^{(2n+1)i\omega t}}{2n+1}$$

On a la comparaison suivante :

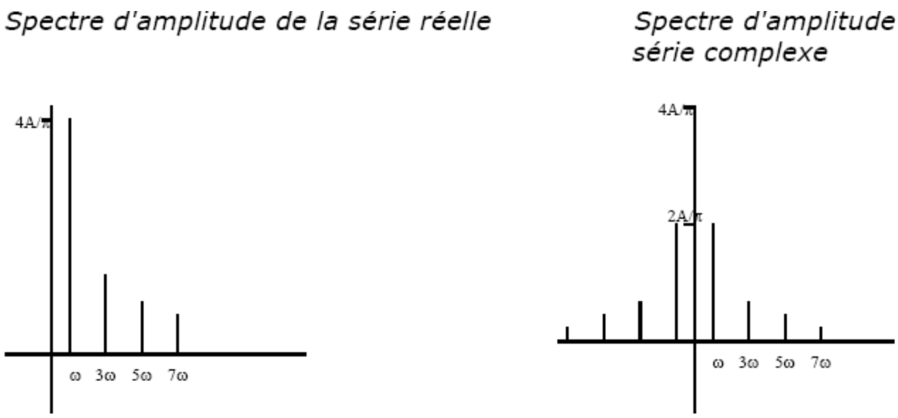


FIGURE 7 – *Spectres d'amplitude*

1.5 Exercice : études de cas

On considère trois signaux f, g et h 2π périodique définis par :

$$f(t) = \pi - |t|,\ \forall t \in]-\pi, \pi], \quad g(t) = t^2 - \pi^2,\ \forall t \in]-\pi, \pi], \quad h(t) = \begin{cases} e^{\frac{-t}{\pi}}, & t \in [0, \pi[\\ 0, & t \in [\pi, 2\pi[\end{cases}$$

1. Représenter graphiquement et séparément ces trois fonctions sur l'intervalle $[-5\pi, 5\pi]$
2. Pour les trois fonctions calculer a_0, puis pour n saisi par l'utilisateur au clavier a_n, b_n et c_n. Utiliser la fonction ***integrate*** de R.
3. Pour chaque signal représenter les sommes des séries réelles et complexes tronquées pour $n = 2, 3, 8$ et 20. Supperposer chaque somme et le signal d'entrée correspondant. Que constate-t-on ?
4. Représenter les spectres d'amplitudes et de phase pour les séries rélles jusqu'à 7 pulsations. Représenter les spectres d'amplitudes pour les séries complexes entre -7 et 7 pulsations. Comparer et interpréter.

2 Transformée de Fourier (TF)

2.1 Intuitivement

Supposons qu'un pianiste joue en même temps les trois notes suivantes : do (132 Hz), mi (165 Hz) et sol (198 Hz). Notre oreille entend la somme des trois notes : l'accord do majeur. Les trois vibrations individuelles et leur somme sont représentées ci-dessous :

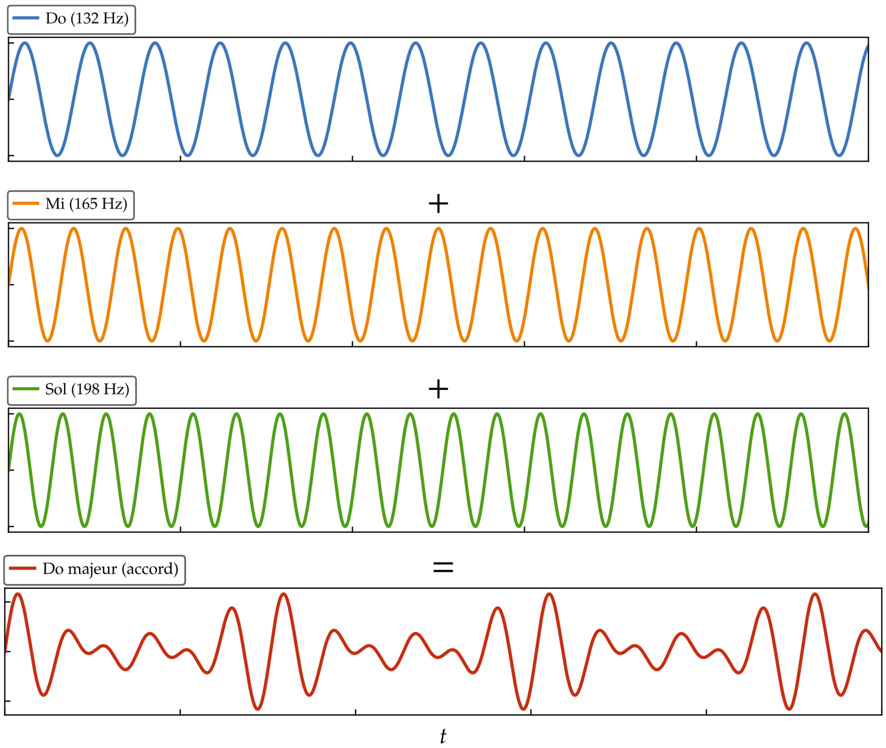


FIGURE 8 – *Ondes sonores qui composent l'accord do majeur.*

A première vue l'accord semble être un patron d'oscillations assez complexe. Pourtant il finit par se répéter, car il est constitué de seulement trois notes simples dont on connait la forme exacte. En termes de transformée de Fourier, les trois notes individuelles sont les composantes de l'accord, qui constitue un signal. Les composantes peuvent avoir différentes amplitudes et différentes fréquences.

La transformée de Fourier est un outil qui permet de décomposer un signal en ses différentes composantes. C'est un peu comme un prisme qui laisse entrer la lumière blanche (la somme de toutes les longueurs d'ondes) et qui sépare ce signal en ses différentes composantes (les couleurs), mais de façon mathématique. La transformée de Fourier permet de passer du domaine temporel/spatial au domaine fréquentiel/spectral. La transformée inverse permet de faire... l'inverse.

- **Domaine temporel/spatial**

  Une oscillation est souvent définie dans le domaine du temps et de l'espace. C'est habituellement la forme de l'onde qu'on peut observer. Par exemple, la note la plus grave sur un piano (la) crée une onde mécanique qui oscille à une fréquence de 27.5 Hz (27.5 fois par seconde) et possède une longueur d'onde spatiale de 12.374 m. Il peut être assez difficile de décrire un signal complexe dans ce domaine (voir l'accord sur la figure précédente).

- **Domaine fréquentiel/spectral**

  Il est plus facile de décrire un signal dans le domaine fréquentiel : c'est-à-dire en décrivant la gamme complète des fréquences et de leurs contributions individuelles relatives dans la composition du signal. La gamme de contribution en fonction de la fréquence constitue le spectre du signal.

  NB : On utilise le terme domaine fréquentiel quand le signal d'entrée est temporel et le terme domaine spectral quand le signal d'entrée est spatial. Les quantités dans le domaine fréquentiel sont exprimées en Hz (équivalent à s^{-1}) et celles dans le domaine spectral en m.

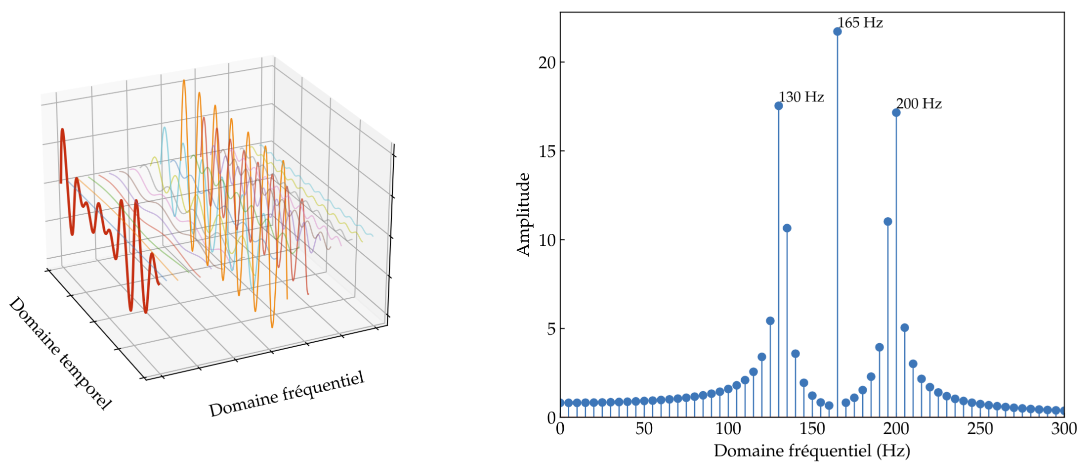


FIGURE 9 – *Domaines temporel et fréquentiel + spectre d'amplitude de l'accord do majeur.*

Une vue transversale du spectre de l'accord do majeur est montré dans la partie droite. On remarque que les composantes qui contribuent le plus au signal total sont les fréquences avoisinantes de 130, 165 et 200 Hz, qui correspondent respectivement aux notes do, mi et sol. Notez que les pics se trouvent à des fréquences approximatives, car le spectre a été discrétisé en faisant des bonds de 5 Hz pour alléger l'illustration. Si on avait discrétisé le spectre avec bonds de 1 Hz on verrait les pics exactement au bon endroit.

2.2 Présentation de l'analyse spectrale

Du développement d'une fonction périodique en série de Fourier, on calcule le spectre de fréquences : diagramme en bâtons représentant l'amplitude du fondamental et des harmoniques en fonction du rang de celles-ci et donc de leur fréquence :

$$f_n = \frac{n\omega}{2\pi}$$

La connaissance de ce spectre permet de prendre des décisions adéquates en matière de traitements des signaux : filtrage, échantillonage,... c'est l'analyse spectrale.

Les signaux non périodiques pourront être traités de façon similaire, en considérant :
Une fonction non périodique peut être considérée comme la limite d'une fonction périodique dont on fait tendre la période T vers $+\infty$, alors $\omega = \dfrac{2\pi}{T}$ tend vers 0.
Dans ces conditions, le spectre de raies d'intervalles de longueur ω deviendra une fonction continue représentée par une courbe

2.3 SF et TF : un peu de maths

- Une fonction x, T périodique développable en SF peut s'écrire :

$$x(t) = \sum_{n=-\infty}^{+\infty} c_n e^{in\omega t}, \quad \text{avec } c_n = \frac{1}{T}\int_{\alpha}^{\alpha+T} x(u)e^{-in\omega u}du$$

  Autre écriture :

  Soient $\alpha = -\dfrac{T}{2}$, $I = [-\dfrac{T}{2}, \dfrac{T}{2}]$ et $\Omega = n\omega$.
  ω étant l'accroissement de Ω entre deux valeurs consécutives $n\omega$ et $(n+1)\omega$, on pose

$$\omega = \Delta\Omega \quad \text{et} \quad c_n = \omega G(\Omega) = \Delta\Omega G(\Omega)$$

  On a alors :

$$x(t) = \sum_{\Omega=-\infty}^{+\infty} \Delta\Omega G(\Omega)e^{i\Omega t}, \quad \text{avec } G(\Omega) = \frac{1}{2\pi}\int_{-\frac{T}{2}}^{\frac{T}{2}} x(u)e^{-i\Omega u}du$$

- Si la fonction x est apériodique, en faisant tendre T vers $+\infty$, $\Delta\Omega$ devient infiniment petit, on la note alors $d\Omega$, différentielle. On aura alors :

$$x(t) = \int_{-\infty}^{+\infty} G(\Omega)e^{i\Omega t}d\Omega$$

  sous réserve que $G(\Omega)$ soit absolument intégrable.

  Finalement, en posant $\varphi(\Omega) = 2\pi G(\Omega)$, on a

$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \varphi(\Omega)e^{i\Omega t}d\Omega \quad \text{et} \quad \varphi(\Omega) = \int_{-\infty}^{\infty} x(u)e^{-i\Omega u}du$$

- **Finish** :

  Soit x un signal réel ou complexe. En utilisant $f = \dfrac{\Omega}{2\pi}$, on a $\Omega = 2\pi f$ et $d\Omega = 2\pi df$, on a

$$x(t) = \int_{-\infty}^{+\infty} X(f)e^{2i\pi ft}df \quad \text{et} \quad X(f) = \int_{-\infty}^{+\infty} x(t)e^{-2i\pi ft}dt$$

  $X(f)$ est la TF de $x(t)$. On peut noter $X(f) = TF(x(t))$

$x(t)$ est la TF inverse de $X(f)$. On peut noter $x(t) = TF^{-1}(X(f))$
Avec si x discontinue en t_0, $TF^{-1}(X(f))(t_0) = \frac{1}{2}(x(t_0^+) + x(t_0^-))$

On remarquera la symétrie des deux relations :

$$\left\{ \begin{array}{ccc} x & \longleftrightarrow & X \\ t & \longleftrightarrow & f \\ i & \longleftrightarrow & -i \end{array} \right\}.$$

2.4 Spectre de fréquence

- Soit x une fonction T périodique. On a

$$c_n = \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} x(u)e^{-in\omega u}du$$

Avec $u = t$ et $n\omega = \Omega = 2\pi f$, on a : c_n devient c_f

$$c_f = \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} x(t)e^{-2i\pi ft}dt$$

Le spectre d'amplitude est donné par $|c_f|$ et le spectre de phase par $Arg(c_f)$, l'argument (l'angle modulo 2π) de c_f.

La densité spectrale de raies est défini par :

$$\Delta f = \frac{c_f}{f_1}, \quad f_1 = \frac{1}{T} : \text{ l'écart entre deux raies adjacentes}$$

Remarque : $f_n = \frac{n\omega}{2\pi}$

...	1	2	3	...	n	...
...	ω	2ω	3ω	...	nω	...
...	f_1	$2f_1$	$3f_1$	...	nf_1	...

- Soit x une fonction non périodique, en faisant tendre T vers $+\infty$ la densité spectrale Δf devient $X(f)$ (qui peut être considéré comme la limite de la densité spectrale de raies). On distinguera $|X(f)|$: spectre d'amplitude et $Arg(X(f))$: spectre de phase

- Si x est une fonction réelle, on a

$$X(f) = \int_{-\infty}^{+\infty} x(t)e^{-2i\pi ft}dt = \int_{-\infty}^{+\infty} x(t)\cos(2i\pi ft)dt - i\int_{-\infty}^{+\infty} x(t)\sin(2i\pi ft)dt = X_{Re}(f) + iX_{Im}(f)$$

Le spectre d'amplitude est $|X(f)| = \sqrt{(X_{Re}(f))^2 + (X_{Im}(f))^2}$ qui est paire.

Le spectre de phase est $Arg(X(f)) = tan^{-1}(\frac{X_{Im}(f)}{X_{Re}(f)})$ qui est impaire.

2.5 Exercices

2.5.1 TF directe et spectre

Soient les signaux suivants :

$$x(t) = rect(\frac{t-\pi}{2\pi}), \text{ avec } rect(u) = \begin{cases} 1, & |u| \leqslant \frac{1}{2} \\ 0, & |u| > \frac{1}{2} \end{cases}$$

$$y(t) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}, \quad \text{et} \quad z(t) = \cos^2(2\pi f_0 t), \text{ tester avec } f_0 = 1, 2 \text{ et } 3$$

1. Tracer les signaux
2. Implémenter une méthode de calcul de TF
3. A l'aide de la TF pour chaque signal tracer les spectres d'amplitude et de phase

2.5.2 TF inverse

Soit $X(f) = \dfrac{1}{1+f^2}$

1. Implémenter une méthode de calcul de TF^{-1} : transformation de Fourier inverse
2. Donner $x(t)$
3. A l'aide des TF calculer dans l'exercice précédent, retrouver les signaux originels

3 Transformée de Fourier d'un signal discrèt (TFTD) et transformée de Fourier discrète (TFD)

NB : les éléments de cours qui suivent sont des résumés. Il est conseillé de lire le pdf : TFTD et TFD

3.1 Contexte TFTD

La TFTD est la TF d'un signal échantilloné (une suite de nombres). Le calcul direct de la TFTD reste le même que pour le TF á la différence que l'on a une somme.
Comme la TFTD est périodique de période 1, le calcul de l'inverse de l'intǵrale de la TFTD est le même que pour la TF inverse mais sur l'intervalle $[-\frac{1}{2}, \frac{1}{2}]$

3.2 Formules

TFTD :

$$X(f) = \sum_{n=-\infty}^{+\infty} x(n)exp(-2i\pi nf),$$

$$x(n) = \int_{-\frac{1}{2}}^{\frac{1}{2}} X(f)exp(2i\pi nf)df$$

3.3 Contexte TFD

La TF ou la TFTD ne sont pas très utiles pour analyser des signaux complexes quand l'intégrale n'a pas de solution analytique fermée pour la TF par exemple. On pourrait cependant utiliser des algorithmes d'approximation de calculs d'intégrales type Simpson, mais cela engendrerait encore plus temps à l'exécution.

La transformation de Fourier discrète est une alternative qui permet d'analyser les signaux numériques discrétisés et de taille finie. La plupart des langages de programmation modernes implémentent un algorithme de transformation de Fourier rapide (FFT).

La TFD ne calcule pas le spectre continu d'un signal continu. Elle permet seulement d'évaluer une représentation spectrale discrète (spectre échantillonné) d'un signal discret (signal échantillonné) sur une fenêtre de temps finie (échantillonnage borné dans le temps).
L'exemple ci-dessous peut laisser croire que la TFD permet de calculer le spectre d'un signal continu, mais cela n'arrive que lorsque la fenêtre d'échantillonnage correspond à un multiple strictement supérieur à deux fois la période du signal échantillonné (dans ce cas on a forcément évité le repliement de spectre, c'est le théorème d'échantillonnage de Nyquist-Shannon).

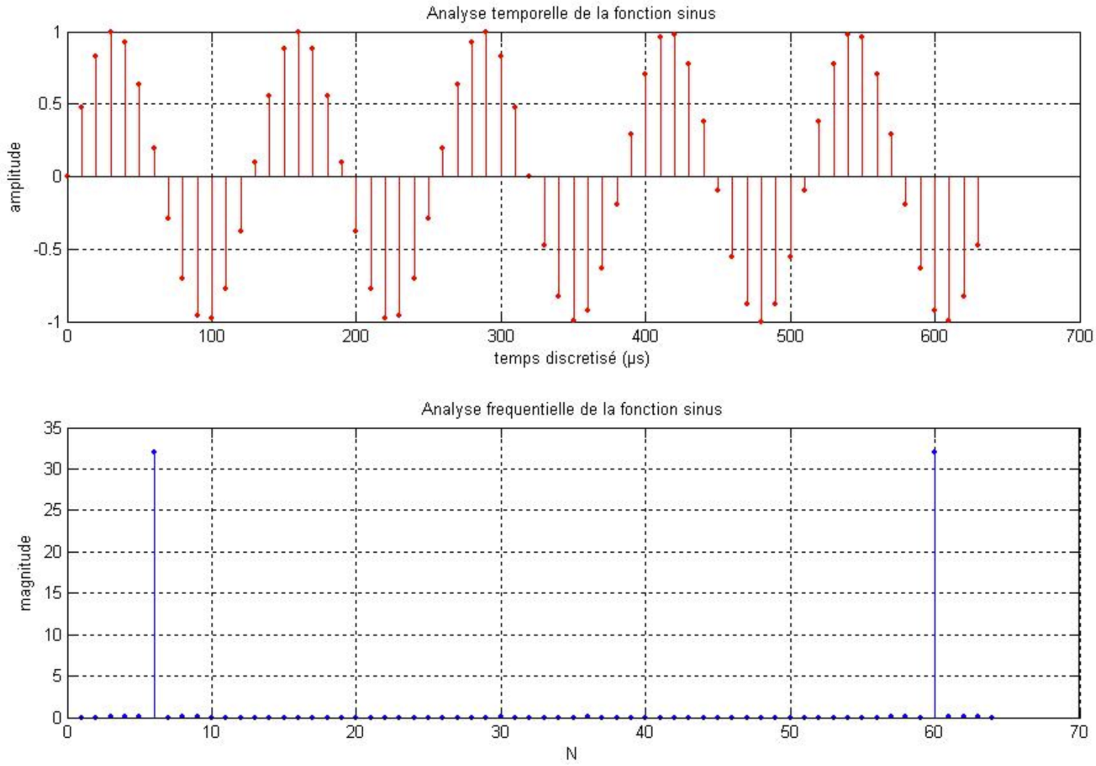


FIGURE 10 – *Transformée de Fourier discrète sur $N = 64$ points d'un sinus de fréquence 7 812,5 Hz échantillonné à 100 000 échantillons par seconde (100 kéch/s).*

3.4 Formules

La transformation de Fourier discrète d'un signal de N échantillons $(x(0), x(1), ..., x(N-1))$ est le vecteur $(X(0), X(1), ..., X(N-1))$ défini par :

$$X(k) = \sum_{n=0}^{N-1} x(n)exp(-2i\pi kn/N), \quad 0 \leqslant k < N$$

A partir de $(X(0), X(1), ..., X(N-1))$, la transformation de Fourier discrète inverse est le vecteur $(x(0), x(1), ..., x(N-1))$ défini par :

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k)exp(2i\pi nk/N), \quad 0 \leqslant n < N$$

3.5 Exercices

Note :

A partir d'un signal continu temporel $t \longrightarrow x(t)$, on peut extraire N termes afin de construire une suite

$$x_n = x(nT_e)$$

Avec T_e la période d'échantillonage. Cet échantillonnage du signal temporel à la période T_e produit une périodisation du spectre à la période $f_e = \frac{1}{T_e}$.
Les N termes de $X(k),\ k = 0, ..., N-1$ sont les N points de fréquences

$$f_k = \frac{kf_e}{N}$$

3.5.1 TFD, cas 1

1. Calculer la TFD de la suite (x_n) formée de $N = 8$ points $(n \in [0, 7])$, obtenue en échantillonnant à la fréquence $f_e = 16Hz$ le signal $x(t)$:

$$x(t) = 2\sin(8\pi t) + 8\cos(4\pi t)$$

2. Représenter le signal et son spectre d'amplitude

3.5.2 TFD, cas 2

1. Calculer la TFD de la suite (x_n) formée de $N = 24$ points $(n \in [0, 7])$, obtenue en échantillonnant à la fréquence $f_e = 16Hz$ le signal $x(t)$:

$$x(t) = 3\sin(8\pi t) + 4\cos(6\pi t)$$

2. Représenter le signal et son spectre d'amplitude

3.5.3 TFD inverse

1. Implémenter la méthode de TFD inverse.

2. Appliquer la TFD inverse aux résultats des deux précd́ents exercices afin de retrouver les signaux initiaux

3.5.4 FFT

1. Implémenter l'algorithme FFT : TFD rapide

2. Trouver sur internet un exemple de calcul que vous implémentrez sous R afin d'illustrer l'algorithme et le temps d'exécution

e

4 Méthode du Chi deux

4.1 Eléments de cours

Le test du χ^2 (prononcer khi deux) fournit une méthode pour déterminer la nature d'une répartition, qui peut être continue ou discrète. Dans la métode présentée on se place dans une répartition uniforme dans le cas discret.

Méthode :

1. On répartit les valeurs de l'échantillon (de taille n) dans k classes distinctes et on calcule les effectifs de ces classes. Il faut vérifier que pour les i de 1 à k, on a $np_i(1-p_i) \geqslant 5$ (éventuellement répartir les valeurs autrement sinon). Pour $i = 1, ..., k$ appelons O_i les effectifs observés et E_i les effectifs théoriques.
2. On calcule $Q = \displaystyle\sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$. La statistique Q donne une mesure de l'écart existant entre les effectifs théoriques attendus et ceux observés dans l'échantillon. En effet, plus Q sera grand, plus le désaccord sera important. La coïncidence sera parfaite si $Q = 0$.
3. On compare ensuite cette valeur Q avec une valeur issue de la table du χ^2 (voir un ex sur internet) à la ligne $k-1$ et à la colonne α. $k-1$ est le nombre de degrés de liberté et α la tolérance.
4. Si $Q > \chi^2_{k-1,\alpha}$, et si n est suffisamment grand, alors l'hypothèse d'avoir effectivement affaire à la répartition théorique voulue est à rejeter avec une probabilité d'erreur d'au plus α.

Exemple :

On a lancé un dé 90 fois et on a obtenu les issues 1 à 6 (k=6) avec les effectifs suivants : 12, 16, 20, 11, 13, 18 (on a vérifié que 90 lancers sont suffisants : $n(1/6)(5/6) \geqslant 5$ implique que $n \geqslant 36$).
Si le dé n'est pas pipé (notre hypothèse), on attend comme effectifs moyens théoriques 15 pour toutes les issues.

$$Q = \frac{(12-15)^2}{15} + \frac{(16-15)^2}{15} + \frac{(20-15)^2}{15} + \frac{(11-15)^2}{15} + \frac{(13-15)^2}{15} + \frac{(18-15)^2}{15} = \frac{64}{15} = 4,27$$

Pour $k-1 = 5$ degrés de liberté et un seuil de tolérance de 5%, la valeur du tableau est 11,1. Cela signifie que la probabilité que Q soit supérieur à 11,1 est de 5%. Comme $4,266 < 11,1$ on accepte l'hypothèse selon laquelle le dé est régulier.

4.2 Ouvrir son restaurant ?

On dresse dans ce tableau en deuxième ligne la distribution théoriques, obtenus par une étude de marché, du pourcentage clients nécessaire du lundi au samedi pour l'ouverture d'un restaurant. En dernière ligne, on répertorie le nombre clients observés.

Jour	Lundi	mardi	mercredi	Jeudi	Vendredi	Samedi
Théoriques (%)	10	10	15	20	30	15
Observés	30	14	34	45	57	20

Avec un seuil de signification $\alpha = 5\%$, la distribution de l'étude de marché est-elle bonne par rapport aux observations ? Peut-on penser à ouvrir son restaurant ?

5 ACP

Voici les données des performances d'athlètes ayant participé à des épreuves du décathlon aux Jeux Olympiques et au Décastar (compétition annuelle d'athlétisme se tenant au Stade de Thouars de Talence, en Gironde)

	c100	long	poids	haut	c400	c110	disq	perche	javel	c1500	RANG	POINTS	COMPET
SEBRLE	11.0	7.58	14.8	2.07	49.8	14.7	43.8	5.02	63.2	292	1	8217	DS
CLAY	10.8	7.40	14.3	1.86	49.4	14.0	50.7	4.92	60.2	302	2	8122	DS
KARPOV	11.0	7.30	14.8	2.04	48.4	14.1	49.0	4.92	50.3	300	3	8099	DS
BERNARD	11.0	7.23	14.2	1.92	48.9	15.0	40.9	5.32	62.8	280	4	8067	DS
YURKOV	11.3	7.09	15.2	2.10	50.4	15.3	46.3	4.72	63.4	276	5	8036	DS
WARNERS	11.1	7.60	14.3	1.98	48.7	14.2	41.1	4.92	51.8	278	6	8030	DS
ZSIVOCZKY	11.1	7.30	13.5	2.01	48.6	14.2	45.7	4.42	55.4	268	7	8004	DS
McMULLEN	10.8	7.31	13.8	2.13	49.9	14.4	44.4	4.42	56.4	285	8	7995	DS
MARTINEAU	11.6	6.81	14.6	1.95	50.1	14.9	47.6	4.92	52.3	262	9	7802	DS
HERNU	11.4	7.56	14.4	1.86	51.1	15.1	45.0	4.82	57.2	285	10	7733	DS
BARRAS	11.3	6.97	14.1	1.95	49.5	14.5	42.1	4.72	55.4	282	11	7708	DS
NOOL	11.3	7.27	12.7	1.98	49.2	15.3	37.9	4.62	57.4	267	12	7651	DS
BOURGUIGNON	11.4	6.80	13.5	1.86	51.2	15.7	40.5	5.02	54.7	292	13	7313	DS
Sebrle	10.8	7.84	16.4	2.12	48.4	14.0	48.7	5.00	70.5	280	1	8893	JO
Clay	10.4	7.96	15.2	2.06	49.2	14.1	50.1	4.90	69.7	282	2	8820	JO
Karpov	10.5	7.81	15.9	2.09	46.8	14.0	51.6	4.60	55.5	278	3	8725	JO
Macey	10.9	7.47	15.7	2.15	49.0	14.6	48.3	4.40	58.5	265	4	8414	JO
Warners	10.6	7.74	14.5	1.97	48.0	14.0	43.7	4.90	55.4	278	5	8343	JO
Zsivoczky	10.9	7.14	15.3	2.12	49.4	15.0	45.6	4.70	63.4	270	6	8287	JO
Hernu	11.0	7.19	14.6	2.03	48.7	14.2	44.7	4.80	57.8	264	7	8237	JO
Nool	10.8	7.53	14.3	1.88	48.8	14.8	42.0	5.40	61.3	276	8	8235	JO
Bernard	10.7	7.48	14.8	2.12	49.1	14.2	44.8	4.40	55.3	276	9	8225	JO
Schwarzl	11.0	7.49	14.0	1.94	49.8	14.2	42.4	5.10	56.3	274	10	8102	JO
Pogorelov	11.0	7.31	15.1	2.06	50.8	14.2	44.6	5.00	53.4	288	11	8084	JO
Schoenbeck	10.9	7.30	14.8	1.88	50.3	14.3	44.4	5.00	60.9	279	12	8077	JO
Barras	11.1	6.99	14.9	1.94	49.4	14.4	44.8	4.60	64.6	267	13	8067	JO
Smith	10.8	6.81	15.2	1.91	49.3	14.0	49.0	4.20	61.5	273	14	8023	JO
Averyanov	10.6	7.34	14.4	1.94	49.7	14.4	39.9	4.80	54.5	271	15	8021	JO
Ojaniemi	10.7	7.50	15.0	1.94	49.1	15.0	40.4	4.60	59.3	276	16	8006	JO
Smirnov	10.9	7.07	13.9	1.94	49.1	14.8	42.5	4.70	60.9	263	17	7993	JO
Qi	11.1	7.34	13.6	1.97	49.6	14.8	45.1	4.50	60.8	273	18	7934	JO
Drews	10.9	7.38	13.1	1.88	48.5	14.0	40.1	5.00	51.5	274	19	7926	JO
Parkhomenko	11.1	6.61	15.7	2.03	51.0	14.9	41.9	4.80	65.8	278	20	7918	JO
Terek	10.9	6.94	15.2	1.94	49.6	15.1	45.6	5.30	50.6	290	21	7893	JO
Gomez	11.1	7.26	14.6	1.85	48.6	14.4	41.0	4.40	60.7	270	22	7865	JO
Turi	11.1	6.91	13.6	2.03	51.7	14.3	39.8	4.80	59.3	290	23	7708	JO
Lorenzo	11.1	7.03	13.2	1.85	49.3	15.4	40.2	4.50	58.4	263	24	7592	JO
Karlivans	11.3	7.26	13.3	1.97	50.5	15.0	43.3	4.50	52.9	279	25	7583	JO
Korkizoglou	10.9	7.07	14.8	1.94	51.2	15.0	46.1	4.70	53.0	317	26	7573	JO
Uldal	11.2	6.99	13.5	1.85	51.0	15.1	43.0	4.50	60.0	282	27	7495	JO
Casarsa	11.4	6.68	14.9	1.94	53.2	15.4	48.7	4.40	58.6	296	28	7404	JO

Les dix épreuves du décathlon :
- course sur 100 m (`c100`),
- saut en longueur (`long`),
- lancer de poids (`poids`),
- saut en hauteur (`haut`),
- course sur 400 m (`c400`),
- course de haies sur 110 m (`c110`),
- lancer de disque (`disq`),
- saut à la perche (`perch`),
- lancer de javelot (`javel`),
- course sur 1500 m (`c1500`)

Autres variables
- rang de classement (`RANG`),
- nombre de points (`POINTS`),
- compétition (`COMPET`),
  - Jeux Olympiques (`JO`),
  - Décastar (`DS`)

Attention ! Les noms des participants sont en majuscule pour le Décastar, afin de permettre de différencier les participations d'un même athlète aux deux épreuves. (exemple : `SERBLE`/`Serble`).

A. Analyse rapide
  1. Récupérer les données du fichier "decathlon" et donner la matrice corrélation des variables quantitatives (ne pas prendre COMPET)
  2. Quelles sont les couples de variables les plus corrélées, les moins correélées, les plus opposées ? Justifier.
  3. Comment se groupent les variables du point de vue des signes de corrélation ? Expliquez pourquoi.

B. ACP : dans cette partie, vous allez procédez à une analyse en composantes principales des performances centrées-réduites, en excluant les variables RANG, POINTS et COMPET.
  4. Donner les valeurs propres de la matrice de corrélation. Trier ces valeurs propres et donner le nombre de vecteurs propres qui expliquent le plus l'inertie du nuage des individus. Quelle

règle peut-on utiliser ? Donner le pourcentage d'inertie totale en conservant les trois premiers vecteurs propres.

5. Déterminer les trois composantes principales (projection des individus sur les trois vecteurs propres), que l'on note C1,C2,C3 dans l'ordre décroisant d'inertie
6. Déterminer le tableau des corrélations des variables par rapport à C1,C2,C3 et donner les deux cercles de corrélation des variables par rapport à (C1,C2) et (C2,C3)
7. Quelles sont les variables qui déterminent les 3 composantes principales ? Proposez un seuil.
8. Expliquez comment les données peuvent être modifiées pour faire apparaître un effet de taille. Comment peut-on alors interpréter les axes principaux de la question 5 ?

6 AFC

Ouvrir le tableau des données "sympathiques" représentant des catégories professionnelles (ex : PAYS pour paysan et LIBE pour libéral) qui ont à choisir les 3 qualités les plus significatives pour qu'une personne soit sympathique, dans une liste de 9 (ex : gene pour généreux, serv pour serviable, comp pour compréhensive et disc pour discrête)

Partie 1 : introduction

1. A l'aide des données, calculer le pourcentage de chaque caractéristique d'une personne sympathique et le pourcentage des catégories professionnelles sur l'ensemble des votes.
2. Combien de personnes ont été sondées ? Donner la proportion des employés pour qui être honnête rend sympathique ? Quelle est la proportion d'employés parmi les gens qui pensent qu'être honnête rend sympathique ?

Partie 2 : Analyse des correspondances

Les correspondances du tableau de contingence donne les valeurs propres suivantes :

0.098 0.022 0.005 0.003 0.001 0.000

On donne les coordonnées sur les axes, les contributions aux axes (en pourcentage) et la qualité de la représentation (en pourcentage aussi) par les axes factoriels pour les profils lignes et les profils colonnes.

	Axis1	Axis2
PAYS	-0.3029	-0.2133
OUVR	-0.3568	-0.1115
VEND	-0.1913	0.1199
COMM	0.0146	0.0463
EMPL	-0.0604	0.2148
TECH	0.0152	0.1473
UNIV	0.4610	0.0006
LIBE	0.4977	-0.1141

	Axis1(%)	Axis2(%)
PAYS	9.6	21.4
OUVR	26.3	11.6
VEND	2.1	3.6
COMM	0.0	1.0
EMPL	0.7	40.8
TECH	0.0	9.7
UNIV	11.4	0.0
LIBE	49.9	11.8

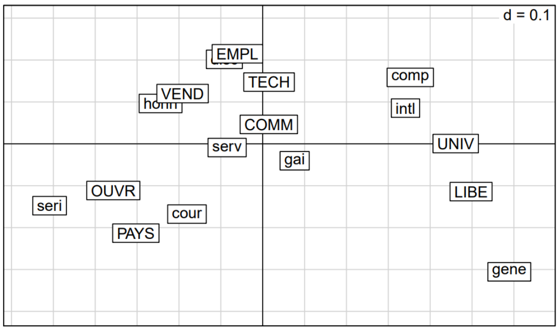
	Axis1	Axis2
PAYS	63.5	31.5
OUVR	89.6	8.7
VEND	35.9	14.1
COMM	1.0	9.9
EMPL	6.6	83.3
TECH	0.6	58.4
UNIV	90.4	0.0
LIBE	94.4	5.0

	Comp1	Comp2
seri	-0.5090	-0.1477
gene	0.5890	-0.3055
gai	0.0763	-0.0404
honn	-0.2438	0.0963
intl	0.3410	0.0852
serv	-0.0850	-0.0089
cour	-0.1808	-0.1677
comp	0.3525	0.1585
disc	-0.0916	0.2016

	Axis1(%)	Axis2(%)
seri	26.4	10.0
gene	28.0	34.0
gai	0.5	0.7
honn	11.1	7.8
intl	16.1	4.5
serv	0.8	0.0
cour	3.8	14.9
comp	12.4	11.3
disc	0.8	16.8

	Axis1	Axis2
seri	85.7	7.2
gene	77.0	20.7
gai	25.4	7.1
honn	82.6	12.9
intl	87.0	5.4
serv	51.6	0.6
cour	48.1	41.4
comp	78.8	15.9
disc	13.1	63.4

Le diagramme ci-dessous est la projection jointe des points-lignes et des points-colonnes sur le premier plan factoriel.



3. Pourquoi 7 valeurs propres ?

4. Quelles sont les modalités qui définissent le premier axe factoriel ? Et le second ? Préciser sur quel(s) critère(s) on se peut se fonder

5. Quelles sont les modalités (lignes et colonnes) qui sont particulièrement mal reprséntés par le premier plan factoriel ?

6. Quelle déduction peut-on faire du fait que OUVR et PAYS sont proches sur le graphique ? Même question pour VEND et honn