

L'inversion du filtre de Hodrick-Prescott

Robert Milano - Institut National de Statistiques

16 mars 2009

Le cas asymptotique

Séries infinies uniformément bornées

Opérateurs Linéaires Invariants

Le filtre HP : séries infinies

Inversion du filtre HP : Equations aux différences

Le cas fini

Une base de vecteurs cycliques

Le filtre HP pour une série finie

Périodogramme à l'extrémité série finie

Conclusion

Séries chronologiques infinies indicées sur \mathbb{Z} :

$$\mathbf{X}_{\mathbb{Z}} = (\dots, x_{-3}, x_{-2}, x_{-1}, x_0, x_1, x_2, x_3, \dots)$$

$$\mathbf{Y}_{\mathbb{Z}} = (\dots, y_{-3}, y_{-2}, y_{-1}, y_0, y_1, y_2, y_3, \dots)$$

$$\text{Produit scalaire : } \langle \mathbf{X}_{\mathbb{Z}} | \mathbf{Y}_{\mathbb{Z}} \rangle = \sum_{\ell=-\infty}^{\infty} x_{\ell} \bar{y}_{\ell}.$$

$$\text{On impose Norme finie : } \langle \mathbf{X}_{\mathbb{Z}} | \mathbf{X}_{\mathbb{Z}} \rangle = \sum_{\ell=-\infty}^{\infty} \|x_{\ell}\|^2 = \|\mathbf{X}_{\mathbb{Z}}\|^2 < \infty.$$

Ce qui entraîne :

- (i) $x_k \rightarrow 0$ lorsque $|k| \rightarrow \infty$.
- (ii) Série $\mathbf{X}_{\mathbb{Z}}$ uniformément bornée : $\exists M \in \mathbb{R}_0^+ : \|x_k\| < M \forall k$.

$$\mathbf{X}_{\mathbb{Z}} = (\dots, x_{k-2}, x_{k-1}, x_k, x_{k+1}, x_{k+2}, \dots)$$

Soit opérateur linéaire \mathcal{L} :
$$\mathcal{L}x_k = \sum_{\ell=-\infty}^{\infty} p_{\ell} x_{k+\ell} \quad \forall k \in \mathbb{Z}$$

Invariant : $p_{\ell} \neq p_{\ell}(k) \longrightarrow$ Famille de **vecteurs propres** pour \mathcal{L} :

$$\left\{ \mathbf{B}_{\mathbb{Z}}^{(\omega)}; \omega \in \mathbb{R} \right\} = \left\{ (\dots, e^{-i2\omega}, e^{-i\omega}, 1, e^{i\omega}, e^{i2\omega}, e^{i3\omega}, \dots); \omega \in \mathbb{R} \right\}$$

$$\mathcal{L}\mathbf{B}_{\mathbb{Z}}^{(\omega)} = E_{\omega} \cdot \mathbf{B}_{\mathbb{Z}}^{(\omega)},$$

où $E_{\omega} = \sum_{t=-\infty}^{\infty} p_t e^{i\omega t}$ = valeur propre de \mathcal{L} pour fréquence ω .
 Infinité de vecteurs propres de norme **non** finie.

Le *périodogramme* est la fonction $\|E_{\omega}\|^2$. Pour un série **réelle**, le périodogramme est une fonction **paire** de ω .

Le filtre de Hodrick-Prescott de paramètre λ associe à une série chronologique $\mathbf{X}_{\mathbb{Z}}$ l'unique série $\mathbf{Y}_{\mathbb{Z}}$ vérifiant :

$$\min_{\mathbf{Y}_{\mathbb{Z}}} \sum_{k=-\infty}^{\infty} \left\{ (x_k - y_k)^2 + \lambda [(y_{k+1} - y_k) - (y_k - y_{k-1})]^2 \right\}$$

$$\equiv \min_{\mathbf{Y}_{\mathbb{Z}}} \|\mathbf{X}_{\mathbb{Z}} - \mathbf{Y}_{\mathbb{Z}}\|^2 + \lambda \|\nabla^2 \mathbf{Y}_{\mathbb{Z}}\|^2,$$

où $\lambda > 0$ pondère l'importance du "lissage" (dérivée seconde discrète). Ce qui livre *équation aux différences* d'ordre 4 :

$$\lambda(y_{q-2} - 4y_{q-1} + 6y_q - 4y_{q+1} + y_{q+2}) + y_q = x_q$$

On veut exprimer l'élément y_q ; $q \in \mathbb{Z}$ en fonction de $\{x_0, x_{\pm 1}, x_{\pm 2}, \dots, x_{\pm k}, \dots\}$

→ Résoudre *équation aux différences* :

$$\lambda(y_{q+4} - 4y_{q+3} + 6y_{q+2} - 4y_{q+1} + y_q) + y_{q+2} = x_{q+2}$$

→ A priori : {Solutions} = {Sol. EH} \oplus 1 Sol. Part. = Esp. dim 4

Etape 1 Solution générale *SGH* de équation homogène *EH* :

$$y_q^{(H)} = \rho^q \left(r^{(1)} \cos q\theta + r^{(2)} \sin q\theta \right) + \rho^{-q} \left(r^{(3)} \cos q\theta + r^{(4)} \sin q\theta \right)$$

où ρ et θ fonctions de $\lambda > 0$ (voir doc. p. 9) et où $r^{(1)}, r^{(2)}, r^{(3)}, r^{(4)}$ constantes réelles quelconques.

→ Les solutions **non nulles** de l'*EH* sont **non bornées** sur \mathbb{Z}

Etape 2 Recherche d'une solution particulière de l'ENH

$$ENH \equiv \lambda \Delta^4 y_q + \Delta^2 y_q + 2\Delta y_q + y_q = x_{q+2}$$

→ Méthode *Variation des Constantes* : faire varier les $r^{(i)}$ de *SGH* pour construire **solution particulière** $y_q^{(P)}$ de *ENH*

→ Calculs longs !

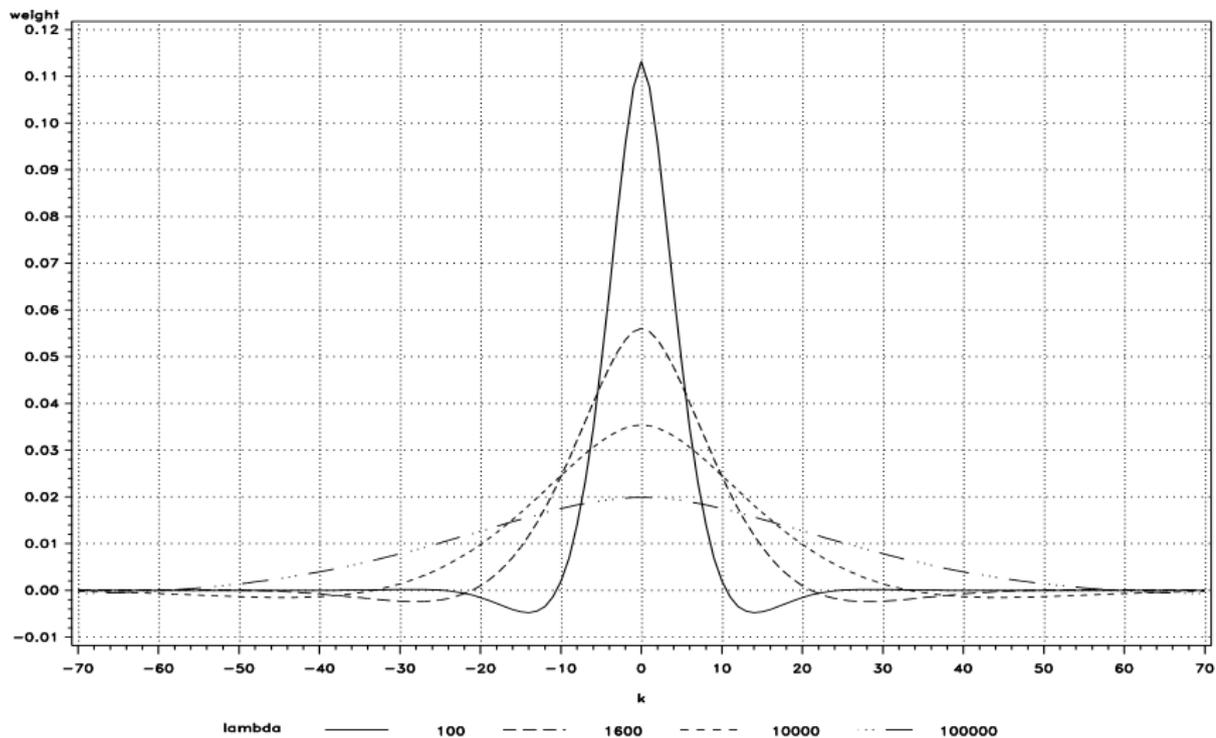
→ Moyenne mobile : $y_q^{(P)} = \sum_{l=-\infty}^{\infty} p_{q+l} x_{q+l}$ **invariante**

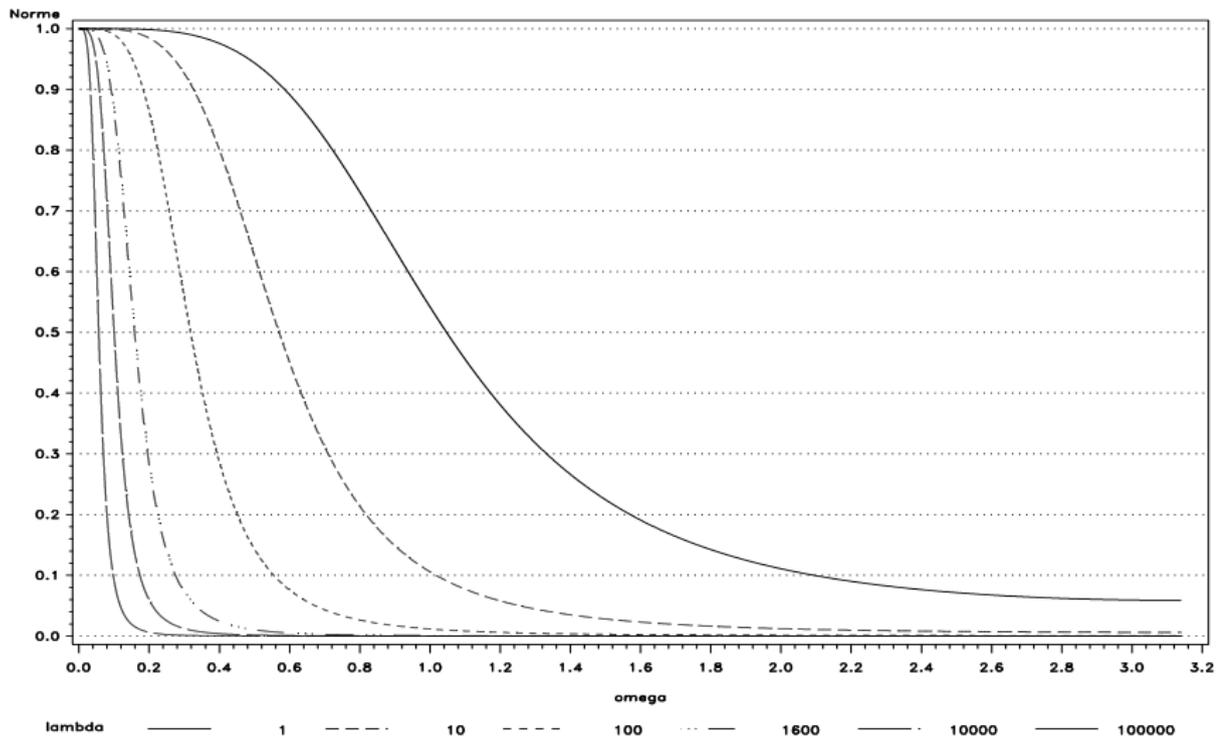
→ Poids de cette moyenne en $q=0$

→ $y_0^{(P)} = \sum_{k=-\infty}^{\infty} p_k x_k$, où poids p_k valent :

$$\begin{cases} \rho^{-k-1} (\rho - \rho^{-1}) (\rho^2 + \rho^{-2} + 6)^{-1} \sin^{-1} \theta [\sin(k-1)\theta - \rho^2 \sin(k+1)\theta] \\ \quad \text{pour } k \leq 1 \\ \rho^{k+1} (\rho - \rho^{-1}) (\rho^2 + \rho^{-2} + 6)^{-1} \sin^{-1} \theta [\sin(k-1)\theta - \rho^{-2} \sin(k+1)\theta] \\ \quad \text{pour } k \geq 2. \end{cases}$$

- Moyenne mobile **symétrique** : $\rho_{-k} = \rho_k \forall k$
- Moyenne mobile **normée** : $\sum_{k=-\infty}^{\infty} \rho_k = 1$
- Opérateur linéaire **invariant** (Indépendt de k)
 - ⇒ Valeurs propres E_{ω} **invariantes** (Fonct. de freq ω seule)
 - ⇒ Normes $\|E_{\omega}\|^2$ **invariantes**
- Résultat analytique ⇒ Calcul exact du périodogramme (Voir doc. p. 27)





Série finie : $\mathbf{X}_T = (x_{-T}, \dots, x_{k-1}, x_k, x_{k+1}, \dots, x_T)$

Série longueur $2T+1 \rightarrow$ Base orthonormée [$\alpha = 2\pi/(2T+1)$] :

$$\left\{ \frac{1}{\sqrt{2T+1}} \left(e^{-inT\alpha}, e^{-in(T-1)\alpha}, \dots, e^{-in\alpha}, 1, e^{in\alpha}, \dots, e^{inT\alpha} \right) \right\}_{q=-T}^T$$

$$x_k = \frac{1}{\sqrt{2T+1}} \sum_{n=-T}^T c_n e^{ink\alpha}$$

$$c_n = \frac{1}{\sqrt{2T+1}} \sum_{k=-T}^T x_k e^{-ink\alpha}$$

- $\mathfrak{L}x_k = \sum_{\ell=-T}^T p_{\ell k} x_{k+\ell} \Rightarrow$ Op. lin. **non** invariant $\Rightarrow p_{\ell k} = f(k)$
- $\mathfrak{L} = \mathfrak{L}_k : \mathfrak{L}_k (e^{-inT\alpha}, \dots, e^{inT\alpha}) = (\mathfrak{L}_{-T} e^{-inT\alpha}, \dots, \mathfrak{L}_T e^{inT\alpha})$
- $\mathfrak{L}_k e^{ink\alpha} = E_{nk} e^{ink\alpha}$ où $E_{nk} = \sum_{\ell=-T}^T p_{\ell k} e^{in\ell\alpha} = f(n, k)$

Soit $\mathbf{X}_T = (x_{-T}, x_{-T+1}, \dots, x_{-1}, x_0, x_1, \dots, x_T)$. On cherche $\mathbf{Y}_T = (y_{-T}, y_{-T+1}, \dots, y_{-1}, y_0, y_1, \dots, y_T)$ vérifiant :

$$\min_{\mathbf{Y}_T} \left\{ \sum_{k=-T}^T (x_k - y_k)^2 + \lambda \sum_{k=-T+1}^{T-1} (y_{k+1} - 2y_k + y_{k-1})^2 \right\},$$

où $\lambda \geq 0$ paramètre réel. La série \mathbf{Y}_T est solution du système de $2T+1$ équations à $2T+1$ inconnues $\{y_{-T}, \dots, y_T\}$:

$$\begin{cases} \lambda(y_{-T} - 2y_{-T+1} + y_{-T+2}) + y_{-T} = x_{-T} & \text{si } k = -T \\ \lambda(-2y_{-T} + 5y_{-T+1} - 4y_{-T+2} + y_{-T+3}) + y_{-T+1} = x_{-T+1} & \text{si } k = -T+1 \\ \lambda(y_{q-2} - 4y_{q-1} + 6y_q - 4y_{q+1} + y_{q+2}) + y_q = x_q & \text{si } -T+2 \leq k \leq T-2 \\ \lambda(y_{T-3} - 4y_{T-2} + 5y_{T-1} - 2y_T) + y_{T-1} = x_{T-1} & \text{si } k = T-1 \\ \lambda(y_{T-2} - 2y_{T-1} + y_T) + y_T = x_T & \text{si } k = T. \end{cases}$$

Résolution numérique \rightarrow Inverser matrice $(2T+1) \times (2T+1)$
 symétrique $F =$

$$\begin{pmatrix}
 \lambda + 1 & -2\lambda & \lambda & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\
 -2\lambda & 5\lambda + 1 & -4\lambda & \lambda & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\
 \lambda & -4\lambda & 6\lambda + 1 & -4\lambda & \lambda & 0 & 0 & 0 & 0 & \dots & 0 \\
 0 & \lambda & -4\lambda & 6\lambda + 1 & -4\lambda & \lambda & 0 & 0 & 0 & \dots & 0 \\
 0 & 0 & \lambda & -4\lambda & 6\lambda + 1 & -4\lambda & \lambda & 0 & 0 & \dots & 0 \\
 0 & 0 & 0 & \lambda & -4\lambda & 6\lambda + 1 & -4\lambda & \lambda & 0 & \dots & 0 \\
 \vdots & \vdots \\
 0 & 0 & 0 & \dots & 0 & 0 & \lambda & -4\lambda & 6\lambda + 1 & -4\lambda & \lambda \\
 0 & 0 & 0 & \dots & 0 & 0 & 0 & \lambda & -4\lambda & 5\lambda + 1 & -2\lambda \\
 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & \lambda & -2\lambda & \lambda + 1
 \end{pmatrix}$$

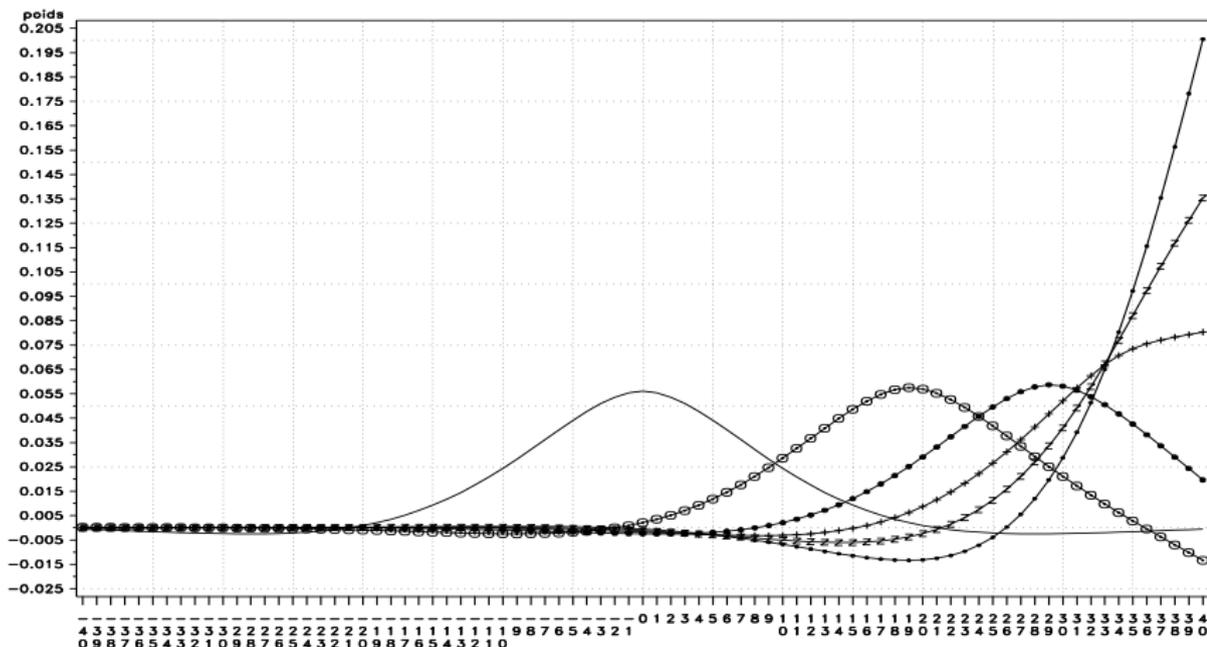
$$F(y_{-T}, \dots, y_k, \dots, y_T)^t = (x_{-T}, \dots, x_k, \dots, x_T)^t$$

$$(y_{-T}, \dots, y_k, \dots, y_T)^t = F^{-1}(x_{-T}, \dots, x_k, \dots, x_T)^t$$

$$y_k = \mathfrak{M}_k x_k = \sum_{k=-T}^T p_k x_k$$

où les poids p_k correspondent à la $k^{\text{ième}}$ colonne de F^{-1} .

Moy. mobile \mathfrak{M}_k **non** invariante. Exemple : série longueur 81
 ($\lambda = 1600$), répartition poids \mathfrak{M}_{41} , \mathfrak{M}_{60} , \mathfrak{M}_{70} , \mathfrak{M}_{75} , \mathfrak{M}_{78} , \mathfrak{M}_{81} :



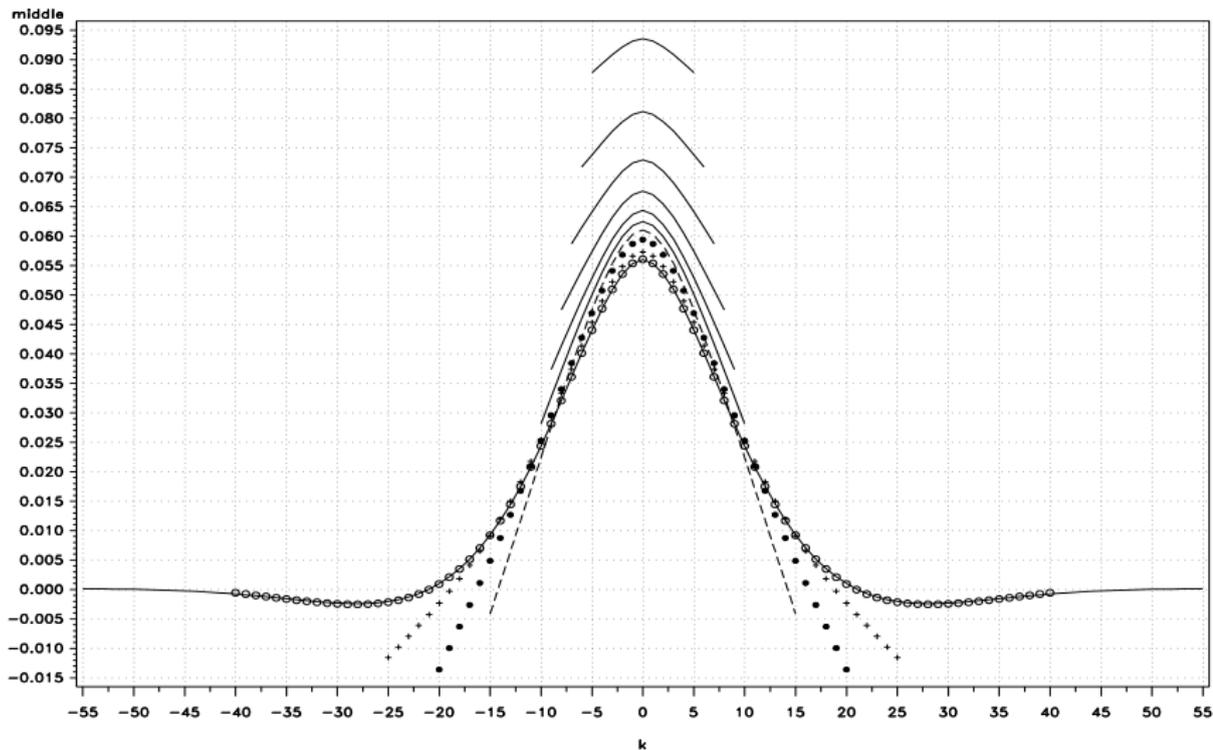
Soit la matrice colonne $\mathbb{I}_{2T+1} = (1, 1, \dots, 1)'$ comportant $2T+1$ éléments tous égaux à 1 :

$$F\mathbb{I}_{2T+1} = \mathbb{I}_{2T+1} \implies \mathbb{I}_{2T+1} = F^{-1}\mathbb{I}_{2T+1}.$$

\implies Somme des poids de chaque moyenne \mathfrak{M}_k vaut 1.

■ Répartition des poids p_k en $y_0 = \sum_{k=-T}^T p_k x_k$
(pour $2T+1 = 11, 13, 15, 17, 19, 21, 31, 41, 51, 81$)

\longrightarrow Si longueur $\sim 70 \implies$ Proche distribution asymptotique



■ Distorsion de \mathfrak{M}_T sur dernière observation de la série \mathbf{X}_T ?
 \Rightarrow Impact de \mathfrak{M}_T sur dernier élément $\mathbf{e}^{inT\alpha} \in \mathbf{B}_T^{(n)}$ de chaque vecteur de base. Pour n fixé, on a :

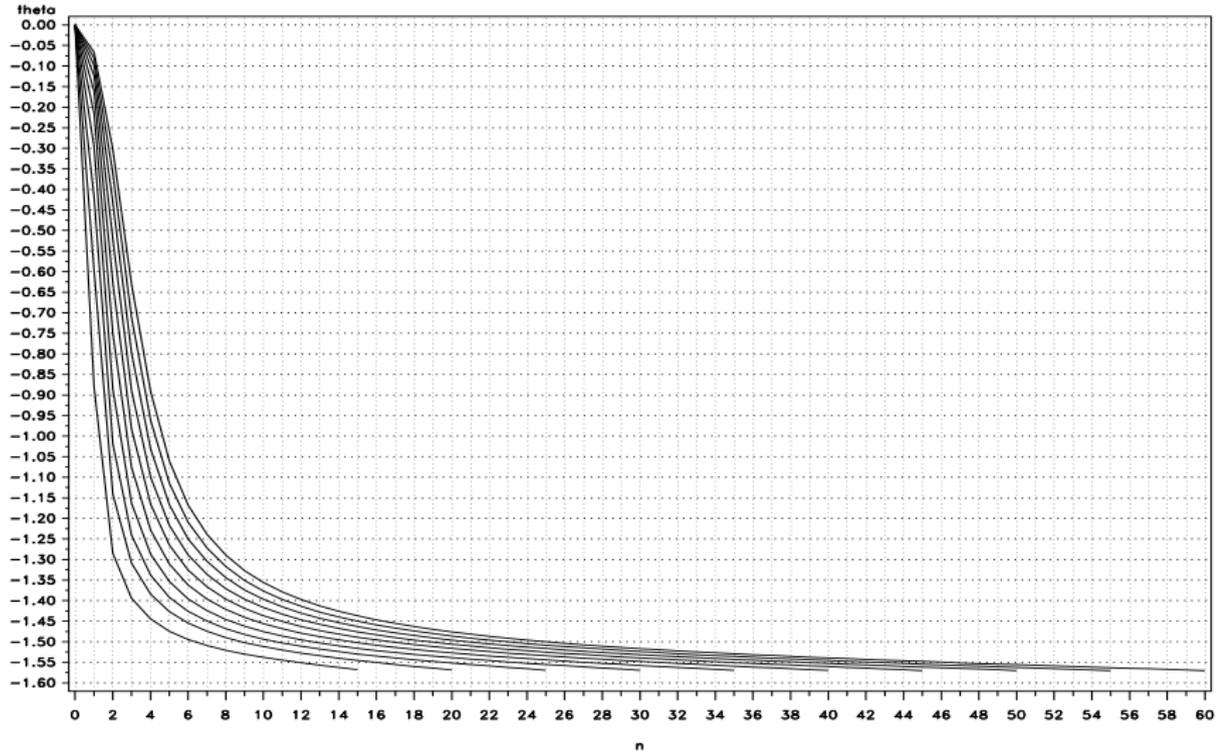
$$\mathfrak{M}_T \mathbf{B}_T^{(n)} = \left(\mathbf{E}_{n(-T)} \mathbf{e}^{-inT\alpha}, \mathbf{E}_{n(-T+1)} \mathbf{e}^{in(-T+1)\alpha}, \dots, \mathbf{E}_{nT} \mathbf{e}^{inT\alpha} \right)$$

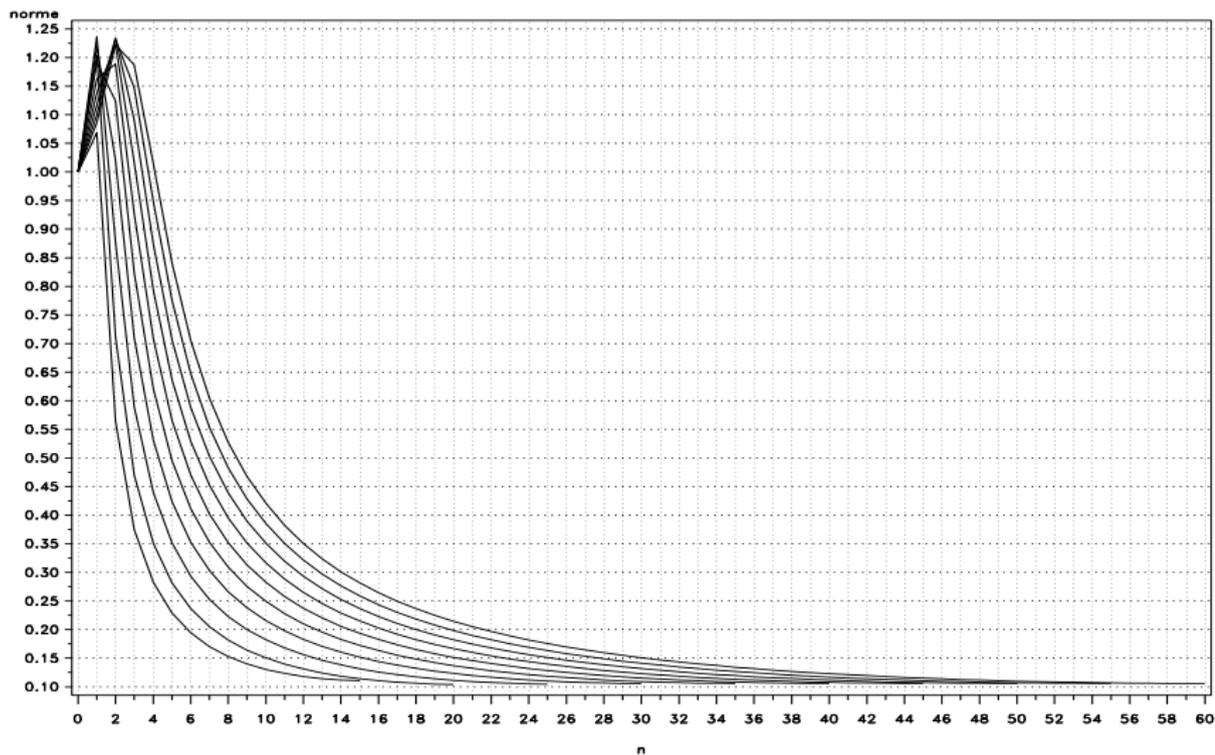
$$\mathbf{E}_{nT} = \sum_{\ell=-T}^T \rho_{\ell T} \mathbf{e}^{in\ell\alpha} = \rho_{nT} \mathbf{e}^{i\theta_{nT}}$$

■ Extrémité série $\mathbf{B}_T^{(n)} \rightarrow$ "coeff dilat." ρ_{nT} et déphasages θ_{nT}
 fonctions de :

- (i) la "fréquence" n [fréquence = $n\alpha = 2n\pi / (2T+1)$]
- (ii) la longueur T

\rightarrow Calculés numériquement \Rightarrow Abaques





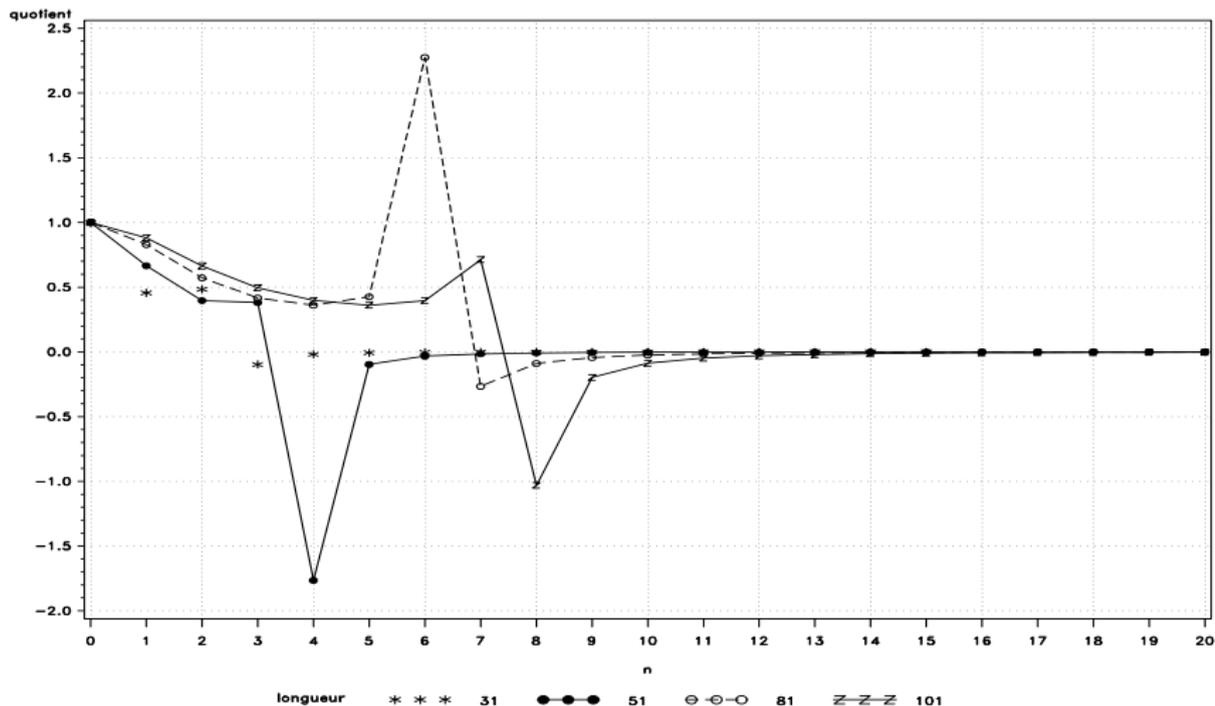
Pour une série finie $e^{ink\alpha}$ périodique de longueur $2T+1$ – ou du moins sa partie réelle $\cos(nk\alpha)$ –, appartenant à la base $\{\mathbf{B}_T^{(n)}; n \in \{-T, \dots, T\}\}$, on veut calculer le rapport entre la valeur $HP_T[\cos(nT\alpha)]$ obtenue par le filtre *HP* au point extrême $k=T$ de cette série et la valeur $HP_\infty[\cos(nT\alpha)]$ obtenue au même point $k=T$ lorsque la série $\cos(nk\alpha)$ est infinie.

→ On calcule, en fonction de n , le quotient

$$\frac{\|E_{\omega=n\alpha}\| \cos(nT\alpha)}{\rho_{nT} \cos(nT\alpha + \theta_{nT})},$$

où la norme $\|E_{\omega=n\alpha}\|$ – calculée en $\omega = n\alpha$ – est celle obtenue dans l'étude des séries infinies.

Quotients : séries longueur 31, 51, 81, 101 ($\lambda = 1600$) :



Ces différents résultats montrent que :

- 1.- Les périodogrammes et les déphasages (radians) de l'opérateur \mathfrak{M}_T sont fonction non seulement du paramètre de lissage λ , mais aussi de la longueur $2T+1$ de la série.
- 2.- Les fréquences les plus basses sont les mieux préservées – voire amplifiées – par l'opérateur \mathfrak{M}_T .
- 3.- Les fréquences les plus basses sont les moins déphasées. Cependant, le déphasage de ces fréquences les plus basses est loin d'être négligeable.
- 4.- Globalement, on constate que le quotient $\|E_{\omega=n\alpha}\| \cos(nT\alpha) / \rho_{nT} \cos(nT\alpha + \theta_{nT})$ est plus élevé pour les fréquences les plus basses, quotient allant grosso modo de 0.5 à 2.5 en valeur absolue.

Le recours à la théorie des *équations aux différences discrètes* permet d'inverser le filtre *HP*, ceci pour les séries infinies. Cette inversion débouche sur une analyse asymptotique précise de l'impact de ce filtre – d'un point de vue fréquentiel – sur les séries chronologiques.

Pour les séries finies, on observe d'importants déphasages pour les éléments de la série situés à proximité des extrémités. Un point important tient au fait que ces déphasages dépendent de la longueur de la série et des fréquences (cycles) composants cette série. Ce dernier point entraîne des distorsions au niveau de la série filtrée. Ces distorsions sont loin d'être négligeables. Pour les fréquences les plus basses, l'erreur commise par le filtre *HP* peut dépasser 100%. Autrement dit, pour les fréquences les plus courantes dans les séries d'origine socio-économique.