

### 3.2.4 Comportement Asymptotique de l'Estimateur de Malliavin Mancino

**Théorème 4.** *L'estimateur de Malliavin Mancino pour la covariance instantanée et intégrée est consistant.*

**Théorème 5.** *Pour tout  $n, N$  on note :  $\hat{\sigma}_{n,N}^2 = \hat{\Sigma}_{ii}^{n_i,N}$ . On pose  $\rho(n_i) = \max_{0 \leq h \leq n-1} |t_{h+1}^i - t_h^i|$ . Si  $\lim_{n_i \rightarrow \infty} \rho(n_i) = 0$ , alors :*

$$\rho(n)^{-1/2} \left( \hat{\sigma}_{n,N}^2 - \int_0^{2\pi} \sigma^2(t) dt \right) \rightarrow \mathcal{N} \left( 0, 2 \int_0^{2\pi} \sigma^4(t) dt \right),$$

où la convergence se fait en loi.

Les preuves de ces deux théorèmes se trouvent dans [9].

## 3.3 Transformée de Fourier Non Uniforme

### 3.3.1 Définition et types

Les données financières étant par nature asynchrones, il s'agit maintenant de pouvoir effectuer les transformées de Fourier sans pertes d'informations. En effet, utiliser des techniques comme le zero-padding introduit des biais comme l'effet Epps. L'avantage de l'utilisation de la transformée de Fourier non uniforme est sa complexité en  $O(N \log(N) + M \log(\frac{1}{\epsilon}))$  où  $\epsilon$  est la précision souhaitée dans le calcul de la transformée de Fourier,  $N$  le nombre d'observations et  $M$  le nombre de modes de Fourier à calculer [4].

On peut toujours se ramener au cas d'une fonction  $f$   $2\pi$ -périodiques en jouant sur les échelles. On considère dans  $\mathbb{R}^d$  des points quelconques  $\mathbf{x}_j$  pour  $j \in \llbracket 0, N-1 \rrbracket$  et on note  $f_j := f(\mathbf{x}_j)$ . On pose  $K := K_{N_1} \times \cdots \times K_{N_d}$  où :

$$K_{N_i} := \begin{cases} \left\{ -\frac{N_i}{2}, \dots, \frac{N_i}{2} - 1 \right\} & \text{si } N_i \text{ impair,} \\ \left\{ -\frac{N_i-1}{2}, \dots, \frac{N_i-1}{2} - 1 \right\} & \text{sinon.} \end{cases}$$

avec  $N_i$  le nombre de coefficients de Fourier à calculer de l'actif  $i$ . On distingue différents types de transformées de Fourier non uniforme (à  $2\pi$  près en fonction de la convention choisie) [7] :

- **Type I (dit *adjoint*) :**

$$F(\mathbf{k}) := \sum_{j=0}^{N-1} f_j e^{-i\mathbf{k} \cdot \mathbf{x}_j} \quad \text{pour } \mathbf{k} \in K.$$

La transformée de Fourier de  $f$  est calculée sur des fréquences régulières, les  $k \in K$  en des points disposés sur une grille quelconque de  $\mathbb{R}^d$ .

- **Type II (dit *forward*) :**

$$f_j := \sum_{\mathbf{k} \in K} F(\mathbf{k}) e^{i\mathbf{k} \cdot \mathbf{x}_j} \quad \text{pour } j \in \llbracket 0, N-1 \rrbracket.$$

La transformée de Fourier de  $F$  est calculée en des points disposés sur une grille régulière et les fréquences, ici les  $x_j$  sont quelconques.

- **Type III :**

Soit  $\mathbf{s}_k$  des points quelconques de  $\mathbb{R}^d$  :

$$F(k) = \sum_{j=1}^{N-1} f_j e^{-i\mathbf{s}_k \cdot \mathbf{x}_j}.$$

Il n'y a ici aucune restriction de régularité ni sur les fréquences où se calcule la transformée de Fourier ni les points d'échantillonnages de la fonction  $f$ .

### 3.3.2 Cas d'étude

On s'intéresse, dans l'étude, à la **transformée de Fourier non uniforme de type I**. Sans perte de généralité, on peut supposer que les  $x_j$  sont dans l'intervalle  $[0, 2\pi]$ . En dimension 1, il s'agit d'évaluer :

$$F(k) = \frac{1}{N} \sum_{j=0}^{N-1} f_j e^{-ikx_j} \quad \text{pour } k \in \{0, \dots, M-1\}, \quad M \in \mathbb{N}^*.$$

Ce calcul coûte  $O(MN)$  opérations, ce qui est loin de la complexité de la FFT. On va essayer de se ramener à un cas d'application de la FFT.

On peut remarquer que :

$$F(k) = \sum_{n=0}^{N-1} f_n e^{-ikx_n} = \int \left( \sum_{n=0}^{N-1} f_n \delta(t - x_n) \right) e^{-ikt} dt = G(k).$$

C'est donc la transformée de Fourier évaluée en  $k$  de la fonction suivante :

$$g(x) := \sum_{j=0}^{N-1} f_j \delta(x - x_j),$$

où  $\delta$  est la distribution de Dirac.

On ne peut pas se contenter de faire un nouvel échantillonnage régulier en  $x$  puis d'appliquer une FFT à la fonction  $g$  en raison de la présence de la distribution de Dirac. C'est pourquoi on la régularise par convolution avec un noyau (*kernel* en anglais) judicieusement choisi, qu'on notera  $\varphi$ .

En notant  $p$  la période de ce noyau, on pose :

$$\tilde{\varphi}(x) := \sum_{r \in \mathbb{Z}} \varphi(x - rp)$$

La fonction  $\tilde{g} := g * \tilde{\varphi}$  est alors périodique,  $C^\infty$ . On peut maintenant créer une grille plus fine régulière à partir des observations initiales  $x_j$  et appliquer la FFT à la fonction  $\tilde{g}$  sur cette nouvelle grille. Soit  $M_r$  la taille de la grille ainsi définie. Le paramètre  $\sigma = \frac{M_r}{M}$  est le **paramètre de suréchantillonnage (over-sampling)**

où  $M$  est le **nombre de modes de Fourier que l'on souhaite calculer**. En général, il est pris empiriquement égal à 2 [3, 4, 7].

On a alors pour tout  $\xi_\ell = \frac{2\pi}{M_r}\ell$  dans la nouvelle grille suréchantillonnée :

$$\tilde{g}(\xi_\ell) = \sum_{j=1}^{N-1} f_j \tilde{\varphi}(\xi_\ell - x_j), \quad \ell = 0, \dots, M_r - 1. \quad (7)$$

On peut donc maintenant évaluer la transformée de  $\tilde{g}$  en utilisant la FFT sur la grille régulière suréchantillonnée. On ne garde que les  $M$  premiers mode de la transformée de Fourier, étant donné que ce sont eux qui nous intéressent. Par ailleurs, sous l'hypothèse du choix judicieux d'un noyau et d'une période, on a l'approximation suivante :

$$\begin{aligned} \tilde{\Phi}(k) &= \frac{1}{p} \int_{-\frac{p}{2}}^{\frac{p}{2}} \sum_{r \in \mathbb{Z}} \varphi(x - rp) e^{-ik \frac{2\pi}{p} x} dx \\ &\approx \Phi(k). \end{aligned}$$

En effet, si le support des noyaux est suffisamment restreint et la périodisation  $p$  assez grande, le calcul précédent revient à prendre l'intégrale du terme  $r = 0$  dans la somme. Cela montre bien l'importance du choix des paramètres !

Enfin, avec les propriétés de la transformée de Fourier, on retrouve facilement les coefficients cherchés :  $F(k) = \Phi(k)^{-1}G(k)$ .

L'étape la plus coûteuse est le calcul des  $\tilde{g}(\xi_\ell)$ . En effet, les coûts aussi bien de stockage que de calculs augmentent exponentiellement en fonction de la dimension à cause du maillage et de l'interpolation. La solution pour réduire le temps de calcul est de **paramétrer le support du noyau**  $\varphi$  pour que  $\varphi$  soit non nulle sur un intervalle  $[-\alpha, \alpha]$  permettant de réduire le nombre de termes non nuls de la somme (7) ou encore de définir un nombre de points  $M_{sp}$  à prendre en compte dans le calcul de  $\tilde{f}(\xi_\ell)$ . On se restreint alors aux  $M_{sp}$   $x_j$  les plus proches de  $\xi_\ell$  de chaque côté. Le nombre de points pris dans chaque calcul est donc  $\omega = 2M_{sp} + 1$ . Cette approximation est valide étant donné que les noyaux choisis ont des pics centrés autour des  $x_j$  [7].

Empiriquement, l'erreur  $\epsilon$  décroît exponentiellement avec  $\omega$ . On peut donc faire l'approximation  $\omega = c|\log \epsilon|$  [1]. La complexité du calcul (7) est donc en  $O(M\omega) = O(M \log(\frac{1}{\epsilon}))$  tandis que la complexité du calcul des FFT de la fonction convoluée est en  $O(N \log(N))$ . On retrouve bien la complexité énoncée au début de cette partie.

Il est également important de noter que la précision dépendant du paramètre  $M_{sp}$ , il faudra configurer ce paramètre de façon à obtenir la précision souhaitée.

### 3.3.3 Paramétrage des noyaux utilisés

Le choix des noyaux à utiliser pour le calcul de la NUFFT est un choix délicat pour assurer la convergence de l'algorithme. Ces noyaux ont fait l'objet de plusieurs études [1, 7, 4], . Tout comme l'article [3], nous fondons notre étude sur les noyaux suivants :

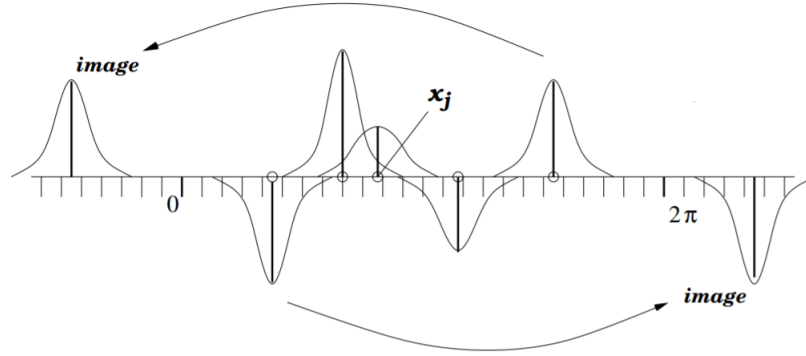


FIGURE 3 – Figure tirée de [7] illustrant la validité de l'approximation de la somme (7)

**Noyau Gaussien** Le noyau gaussien est défini par :

$$\varphi_G(x) = e^{-x^2/4\tau} \text{ pour } x \in \mathbb{R}, \quad \tau = \frac{1}{M^2} \frac{\pi}{\sigma(\sigma - 0.5)} M_{sp}.$$

Il est périodique de période  $2\pi$ . Sa transformée de Fourier est donnée par :

$$\hat{\varphi}_G(k) = 2\sqrt{\tau\pi} e^{-k^2\tau}$$

Selon [3],  $M_{sp} = \left\lfloor \frac{-\ln(\epsilon)(\sigma-1/2)}{(\pi(\sigma-1))} + \frac{1}{2} \right\rfloor$  est un paramétrage empirique de  $M_{sp}$  permettant d'obtenir des résultats satisfaisants.

**Noyau de Kaiser-Bessel** Le noyau de Kaiser-Bessel est défini par :

$$\varphi_{KB}(x) = \frac{1}{\pi} \begin{cases} \frac{\sinh(b\sqrt{M_{sp}^2 - M_r^2 x^2})}{\sqrt{M_{sp}^2 - M_r^2 x^2}} & \text{si } |x| \leq \frac{M_{sp}}{M_r}, \\ \frac{\sin(b\sqrt{M_r^2 x^2 - M_{sp}^2})}{\sqrt{M_r^2 x^2 - M_{sp}^2}} & \text{sinon.} \end{cases}$$

Elle est 1-périodique. Sa transformée de Fourier est donnée par :

$$\hat{\varphi}_{KB}(k) = \frac{1}{M_r} I_0 \left( m \sqrt{b^2 - (2\pi k/M_r)^2} \right),$$

où  $b = \pi \left( 2 - \frac{1}{\sigma} \right)$  et  $I_0(x) = \sum_{k=0}^{\infty} \frac{(\frac{1}{4}x^2)^k}{(k!)^2}$  pour tout  $x \in \mathbb{R}$ .

Encore une fois selon [3],  $M_{sp} = \left\lfloor \frac{1}{2} \left( \lceil \log_{10} \left( \frac{1}{\epsilon} \right) \rceil + 2 \right) \right\rfloor$  donne des résultats empiriquement satisfaisants.

**Exponentielle semi-cercle** L'exponentielle de semi-cercle est définie par :

$$\phi_{ES}(x) = \begin{cases} e^{\beta\sqrt{1-x^2}-1} & \text{si } |x| \leq 1, \\ 0 & \text{sinon.} \end{cases}$$

La transformée de Fourier de l'exponentielle de semi-cercle n'est pas analytiquement connue, mais une approximation efficace de sa transformée peut être retrouvée en [1]. Dans notre cas, nous nous contenterons de l'estimer numériquement.

Empiriquement une valeur satisfaisante de  $\beta$  est  $\beta = 2.3\omega$ .

En pratique, pour paramétrer le support de l'exponentielle de semi-cercle sur  $[-\alpha, \alpha]$  avec  $\alpha = \frac{\pi\omega}{M_r}$ , on utilise le noyau translaté :

$$\varphi_{ES}(x) = \phi_{ES}\left(\frac{x}{\alpha}\right).$$

Un paramétrage satisfaisant de  $M_{sp}$  est  $M_{sp} = \lfloor \frac{1}{2} (\lceil \log_{10} (\frac{1}{\epsilon}) \rceil + 2) \rfloor + 2$ . On déduit  $\omega$  à partir de  $\omega = 2M_{sp} + 1$  [3].