

Petite histoire de $\sigma_\infty^2 = \exp \int \log \gamma(\lambda) d\lambda$ et de la prédiction à un pas

Pierre-Olivier AMBLARD¹

¹GIPSAIab/GAIA, CNRS, Univ. Grenoble Alpes, Saint Martin d'Hères, France

pierre-olivier.amblard@cnrs.fr

Résumé – L'article propose une visite guidée au pas de course des liens existant entre prédiction linéaire à un pas des séries aléatoires stationnaires et théorie des polynômes orthogonaux sur le cercle unité. Il commence par rappeler des aspects historiques du développement de la théorie des séries aléatoires stationnaires qui ont conduit à la formalisation et la résolution du problème de prédiction. Il revient ensuite sur la théorie des polynômes orthogonaux sur le cercle unité avant d'exhiber les liens profonds entre les deux aspects.

Abstract – The paper offers a rapid guided tour of the links between one-step ahead linear prediction of stationary time series and the theory of orthogonal polynomials on the unit circle. It begins by recalling some historical aspects of the building of stationary processes theory that led to the theory of prediction and its practical solutions. The paper then recall some aspects of orthogonal polynomials on the unit circle before showing their fundamental links with prediction theory.

Avertissement. L'idée de proposer ce tour d'horizon m'est venue en lisant le livre de Lasserre, Pauwells et Putinar [10], notamment le passage p. 29 où l'égalité entre $\exp \int \log \gamma(\lambda) d\lambda$ et la fonction de Christoffel en 0 associée à une mesure $\gamma(\lambda) d\lambda$ sur le cercle unité est dévoilée. Ce livre n'aborde pas le problème de la prédiction des processus stationnaires, mais j'ai reconnu avec stupeur dans cette égalité l'erreur quadratique moyenne asymptotique du prédicteur linéaire à un pas d'un signal stationnaire de densité spectrale $\gamma(\lambda)$. J'ai immédiatement fouillé la littérature pour rapidement m'apercevoir que quelques glorieux anciens, dont Thomas Kailath [8], avaient observé et étudié avec bonheur les liens entre prédiction linéaire et polynômes orthogonaux sur le cercle unité.

Ce texte ne comporte donc aucun résultat nouveau, et les analogies qui y sont présentées ont été observées par d'autres par le passé (dont [3, 8, 15], avec de nombreuses sources historiques détaillées en particulier dans [13, 16, 18], voir aussi [12] pour des liens avec les espaces de Hardy), et sont peut-être connues par de nombreux de mes collègues. Mais les liens profonds entre prédiction linéaire et polynômes orthogonaux sont si beaux que je propose ici d'en faire une présentation personnelle, en rappelant quelques éléments historiques. Elle est limitée au cas du temps discret. Des extensions au cas du temps continu existent mais compliquent les développements, et leurs auteurs sont négligés dans ce papier. Les plus importants oubliés sont probablement M. G. Krein, qui fût le premier à remarquer les liens discutés ici et à les étendre au cas continu, et bien sûr N. Wiener, dont le *péril jaune*¹ écrit en 1942 sous forme de rapport classifié, à fait peur pendant plusieurs décennies par la suite aux étudiants. Par ailleurs, outre un bref paragraphe sur le temps discret, l'essentiel est écrit dans le cas continu.

Cette visite au pas de charge est imposée par les contraintes éditoriales de la conférence. Je maintiens toutefois à jour, et je développerai au cours des mois qui viennent, une version plus complète et peut être plus digeste [2].

1 Prédiction linéaire des séries stationnaires.

Kolmogorov et les espaces de Hilbert. Dans [6], Kolmogorov développe la théorie des séries stationnaires à l'aide de la théorie des espaces de Hilbert. Soit $\mathcal{H}$ un tel espace dont les vecteurs sont les signaux d'intérêt. Deux séries x_t et y_t de $\mathcal{H}$ sont conjointement stationnaires si le produit scalaire $\langle \cdot | \cdot \rangle_{\mathcal{H}}$ de $\mathcal{H}$ entre x_t et y_{t+k} est indépendant de $t \in \mathbb{Z}$. On note $\mathcal{H}_{xy}$ le plus petit sous-espace vectoriel fermé de $\mathcal{H}$ contenant l'ensemble des éléments x_t et y_t . On note également $\mathcal{H}_x(t)$ le plus petit sous-espace vectoriel fermé de $\mathcal{H}$ généré par le passé $\{x_s\}_{s \leq t}$.

Soit x_t une série stationnaire ; Kolmogorov montre en utilisant la théorie des représentations spectrales des opérateurs unitaires qu'il existe une fonction positive, non décroissante, continue à droite $\Gamma_x : [-1/2, 1/2] \rightarrow \mathbb{C}$ telle que Γ_x définisse un produit scalaire entre deux fonctions f et g de $[-1/2, 1/2]$ dans $\mathbb{C}$

$$\langle f | g \rangle_{\Gamma_x} = \int_{-1/2}^{1/2} f(\lambda) \bar{g}(\lambda) d\Gamma_x(\lambda)$$

L'ensemble de fonctions pour lequel ce produit scalaire a un sens est l'espace $L^2(\Gamma_x)$ des fonctions de $[-1/2, 1/2]$ dans $\mathbb{C}$ de module carré sommable contre Γ_x . Il faut noter la puissance du formalisme développé par Kolmogorov, puisqu'il inclut la théorie de processus stochastiques stationnaires aussi bien que la théorie des signaux déterministes d'énergie finie ou encore de puissance moyenne finie.

La décomposition de Wold. Travail précédent la théorie de Kolmogorov, le livre (sa thèse) de Wold n'en est probablement pas moins une des inspirations pour [6], puisque la décomposition de Wold [21] est citée par Kolmogorov dès la première page de son article, et est démontrée avec les moyens géométriques des espaces de Hilbert dans son paragraphe 7. Le livre de Wold est très intéressant à lire. Il propose dans le premier chapitre un tour d'horizon sur l'état de l'art au milieu des années 30 de l'analyse des séries temporelles. A l'époque, la théorie des processus stochastiques est naissante, suite à la formalisation moderne de la théorie des probabilités par Kolmogorov (1931), et les travaux de Khintchine (1932-33) et encore de Kolmogorov (1933). Si Khintchine a déjà développé la théorie générale pour

1. N. Wiener, Extrapolation, Interpolation and smoothing of stationary time series, MIT Press, 1949.

des signaux discrets, il revient à Wold d'avoir développé des modèles généraux pour les séries aléatoires. Pour cela, il s'inspire de nombreux travaux de la toute fin du XIXème siècle et des trente premières années du vingtième sur la recherche de *périodicités cachées* dans les séries temporelles. Les auteurs importants ici sont H. Burkhard, A. Schuster, K. Stumpff, E.T. Whittaker et de nombreux autres travaillant à la recherche de périodicités dans des séries géophysiques ou (déjà) financières. Si les approches fonctionnelles reposant sur les séries de Fourier sont déjà largement utilisées, quelques auteurs travaillent sur des équations aux différences. Beaucoup d'approches tentent de représenter et expliquer des raies spectrales d'épaisseur plus large que la résolution donnée par le déjà célèbre périodogramme. Il faut attendre Yule [22], puis Walker [20], pour que des idées d'aléas soient introduites. La discussion de Yule [22] est lumineuse :

If we take a curve representing a simple harmonic function of the time, and superpose on the ordinates small random errors, the only effect is to make the graph somewhat irregular, leaving the suggestion of periodicity still quite clear to the eye.

Yule parle ici de bruit additif et mentionne rapidement le fait que la méthode du périodogramme est efficace pour déterminer la fréquence, à condition toutefois d'observer suffisamment longtemps si le bruit est fort. Il ajoute :

But by improvement of apparatus and automatic methods of recording, let us say, errors of observation are practically eliminated. The recording apparatus is left to itself, and unfortunately boys get into the room and start pelting the pendulum with peas, sometimes from one side and sometimes from the other. The motion is now affected, not by superposed fluctuations but by true disturbances, and effect on the graph will be of an entirely different kind. The graph will remain surprisingly smooth, but amplitude and phase will vary continually.

Les garçons ² entrant dans la salle perturbent non pas l'instrument de mesure mais le processus physique à l'étude (un pendule), générant des perturbations aléatoires continues à la fois de l'amplitude et de la phase des grandeurs mesurées. Partant alors de l'équation aux différences gérant une harmonique pure $x_t = 2 \cos(\theta)x_{t-1} - x_{t-2}$, Yule introduit les « true disturbances » dans l'équation à chaque instant $x_t = 2 \cos(\theta)x_{t-1} - x_{t-2} + \varepsilon_t$, écrivant un modèle AR pour la première fois. Il montre ensuite comment estimer le paramètre à l'aide des corrélations. Sir Walker généralise à tout ordre cette équation et illustre comment en calculer les coefficients, en montrant que les corrélations suivent la même équation aux différences, mais sans le terme de perturbation. Bref, il démontre dans [20] l'équation de Yule-Walker.

Wold reprend ces travaux à vocations pratiques pour faire une théorie complète des processus stationnaires à temps discret. Son livre culmine dans l'énoncé de la décomposition de Wold qui représente tout processus de carré sommable comme la somme d'un processus singulier et d'un processus à moyenne ajustée (mobile) d'ordre infini.

Un signal est dit *singulier* s'il peut s'écrire à chaque instant comme combinaison linéaire de son passé. Autrement dit, x est singulier si $\bigcap_{t \in \mathbb{Z}} \mathcal{H}_x(t) = \mathcal{H}_x$. Pour un signal stationnaire x_t quelconque de $\mathcal{H}$, la projection orthogonale s^x de x sur $\bigcap_{t \in \mathbb{Z}} \mathcal{H}_x(t)$ représente la partie singulière de x . Si s_t^x est la projection orthogonale de x_t sur $\mathcal{H}_x(t-1)$, alors x_t s'écrit de manière unique sous la forme $x_t = s_t^x + \sum_{n \geq 1} c_n^x \varepsilon_{t-n}^x$, où la série $\{\varepsilon_s^x\}_{s \leq t}$ est un système complet et orthonormal de $\mathcal{H}_x(t) \setminus \bigcap_{t \in \mathbb{Z}} \mathcal{H}_x(t)$. Cet argument constitue la décomposition de Wold de x_t en une partie singulière s_t et une partie régulière donnée par une moyenne mobile causale d'une

suite stationnaire d'échantillons indépendants [21, th. 7]. Un signal est alors dit régulier si sa partie singulière est nulle. Dans ce cas la densité spectrale de puissance est strictement positive, et ε_t peut s'écrire comme combinaison linéaire de $\{\varepsilon_s^x\}_{s \leq t}$, autrement dit, x_t peut s'écrire sous forme d'une récursion d'ordre infini attaquée par l'innovation ε_t (modèle AR d'ordre ∞). Cet argument permet à Wold de justifier l'existence de solutions à une équation aux différences d'ordre fini avec une entrée aléatoire, et donc d'introduire rigoureusement la notion de modèle AR d'ordre fini. Par ailleurs, si l'on écrit explicitement $x_t = \sum_{n \in \mathbb{N}^*} b_n x_{t-n} + \varepsilon_t$, le lien avec la prédiction linéaire devient évident.

Prédiction linéaire à un pas des séries stationnaires. Soit x_t une série stationnaire de $\mathcal{H}$. La prédiction linéaire s'intéresse à représenter x_t comme une combinaison linéaire de ses échantillons passés, soit $\hat{x}_t^n = \sum_{i=1}^n a_i^n x_{t-i}$. Les paramètres sont choisis de sorte à minimiser l'erreur quadratique moyenne mesurée à l'aide du produit scalaire $\langle \cdot, \cdot \rangle_{\mathcal{H}}$, et le programme à résoudre est

$$\sigma_n^2(t) = \inf_{a^n} \|x_t - \hat{x}_t^n\|_{\mathcal{H}}^2 = \min_{a^n} \|x_t - \sum_{i=1}^n a_i^n x_{t-i}\|_{\mathcal{H}}^2$$

L'utilisation de la géométrie des espaces de Hilbert permet de trouver l'argument $\hat{x}_0^n$ optimal en utilisant le théorème de la projection orthogonale. Si $\mathcal{H}_x(1:n)$ est le plus petit sous-espace vectoriel de $\mathcal{H}$ contenant les suites $x_{t-1}, \dots, x_{t-n}$, alors $\hat{x}_0^n$ satisfait

$$\langle x_t - \hat{x}_0^n | z_t \rangle_{\mathcal{H}} = 0, \quad \forall z_t \in \mathcal{H}_x(1:n)$$

qui conduit à l'équation de Yule-Walker [20]

$$T_n a^n = \begin{pmatrix} c_0 & c_1 & \cdots & c_{n-1} \\ c_{-1} & c_0 & \ddots & \vdots \\ \vdots & & \ddots & c_1 \\ c_{1-n} & \cdots & c_{-1} & c_0 \end{pmatrix} a^n = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} = R_n$$

où $c_k = \langle x_t | x_{t-k} \rangle_{\mathcal{H}}$. L'erreur quadratique minimale obtenue est indépendante du temps et s'écrit $\sigma_n^2 = c_0 - R_n^{\top} T_n^{-1} R_n$ et l'on note $\sigma_{\infty}^2 = \lim_{n \rightarrow +\infty} \sigma_n^2$ qui représente l'erreur quadratique de prédiction à un pas à passé infini. On note d'emblée que σ_{∞}^2 correspond à la puissance de l'innovation ε_t dans la représentation de Wold d'un signal régulier. Pour calculer cette limite, on peut avoir recours à la théorie du filtrage de Wiener causal à temps discret, appliquée au cas particulier de la prédiction, reposant sur la théorie de la factorisation spectrale forte. Cette route est suivie par exemple dans [5, 14] pour obtenir

$$\sigma_{\infty}^2 = \exp \int_{-1/2}^{1/2} \log \gamma(\lambda) d\lambda$$

Notons que, comme $\log \gamma(\lambda) \leq \gamma(\lambda)$, $-\infty \leq \int_{-1/2}^{1/2} \log \gamma(\lambda) d\lambda < +\infty$, et donc que $\sigma_{\infty}^2 \geq 0$. Le cas $\sigma_{\infty}^2 = 0$ correspond à une série singulière.

Corrélations partielles, algorithme de Levinson-Durbin. Publié en 1947, l'algorithme de Levinson vise à implanter pratiquement le filtre de Wiener, et interroge l'importance de la prise en compte de la taille du filtre dans la qualité d'estimation de la prédiction, débouchant sur une structure récursive naturelle. Levinson mentionne dans l'introduction l'influence sur son travail de deux rapports confidentiels (Phillips & Weiss ; Blackman, Bode & Shannon) sur le lissage et la prédiction des engins de tir. L'argument fondamental utilisé dans la construction de l'algorithme est l'invariance de la fonction de corrélation sous

2. N'oublions pas que Yule écrit en 1927...

l'action du renversement du temps, propriété due à la stationnarité des séries. L'algorithme est le suivant [4, 14]

$$\begin{aligned} a_n^n &= (c_n - \sum_{j=1}^{n-1} a_j^{n-1} c_{n-j})(\sigma_{n-1}^2)^{-1} \\ a_i^n &= a_i^{n-1} - a_n a_{n-i}^{n-1}, \forall i = 1, \dots, n-1 \\ \sigma_n^2 &= \sigma_{n-1}^2 (1 - (a_n^n)^2) \end{aligned}$$

initialisé avec $a_1^1 = c_1/c_0$, $\sigma_0^2 = c_0$. Il est intéressant de noter que a_n^n est la *covariance partielle* entre deux échantillons du signal séparés de $n-1$ autres. Autrement dit, c'est la covariance résiduelle entre ces deux échantillons après soustraction de leur estimation linéaire optimale respective à partir des échantillons du milieu. De plus, si n tend vers l'infini, nous obtenons une deuxième expression pour l'erreur quadratique minimale de prédiction

$$\sigma_\infty^2 = c_0 \prod_{j=1}^{\infty} (1 - (a_j^j)^2) = \exp \int_{-1/2}^{1/2} \log \gamma(\lambda) d\lambda$$

2 Polynômes orthogonaux sur $\partial\mathbb{D}$, le cercle unité.

Les deux tomes [16, 17] de *Orthogonal Polynomial on the Unit Circle* (OPUC) de Barry Simon constituent à ce jour la référence ultime sur le sujet, et beaucoup de ce qui suit s'en inspire.

Récursion de Szegő. Soit Γ une mesure sur le cercle unité ou de manière équivalente sur $[-1/2, 1/2]$. On considère l'espace $L^2(\Gamma)$ des fonctions de carré sommable sur cet intervalle contre la mesure Γ , et le produit scalaire usuel associé $\langle f|g \rangle_\Gamma$ défini plus haut. Les polynômes $\phi_n(z)$ considérés ici sont définis sur le disque unité $|z| \leq 1$, le produit scalaire étant toutefois calculé sur le cercle unité à l'aide de $\langle f|g \rangle_\Gamma$. La suite $\{\phi_n(z)\}_{n \in \mathbb{N}}$ est une suite de polynômes orthogonaux associée à la mesure Γ si $\langle \phi_n | \phi_m \rangle_\Gamma = \kappa_n \delta_{nm}$. Enfin, les $\phi_n(z)$ sont considérés ici dans leur forme monique, à savoir $\phi_n(z) = z^n + \dots$

On considère maintenant l'opérateur anti-unitaire $*$ défini par $\phi_n^*(z) = z^n \overline{\phi_n(1/\bar{z})}$. Donc $\phi_n^*(0) = 1$ (les polynômes considérés sont moniques). La suite $\{\phi_n(z)\}_{n \in \mathbb{N}}$ de polynômes orthogonaux satisfait alors la récursion de Szegő [16, p.56]

$$\begin{pmatrix} \phi_{n+1} \\ \phi_{n+1}^* \end{pmatrix} = \begin{pmatrix} z & -\bar{\alpha}_n \\ -\alpha_n z & 1 \end{pmatrix} \begin{pmatrix} \phi_n \\ \phi_n^* \end{pmatrix} \quad \|\phi_{n+1}\|_\Gamma^2 = (1 - |\alpha_n|^2) \|\phi_n\|_\Gamma^2$$

où la suite des α_n appartient au disque unité ouvert. Cette suite constitue la suite des *coefficients de Verblunsky* de la mesure Γ . Comme $\phi_n^*(0) = 1$, on a $\alpha_n = -\overline{\phi_{n+1}(0)}$. Les coefficients de Verblunsky peuvent être déterminés en fonction des moments de la mesure, c'est-à-dire de la suite $\{c_n\}_{n \in \mathbb{Z}}$ définie par

$$c_n = \int_{-1/2}^{1/2} e^{2i\pi n \lambda} d\Gamma(\lambda)$$

Un lien fondamental entre la matrice Toeplitz $T^n = (c_{j-i})_{1 \leq i, j \leq n}$ et les coefficients α_n peut être établi *via* les déterminants $D^n(d\Gamma) = \det T^{n+1}$:

$$\alpha_n = (-1)^n \frac{D^n(e^{-2i\pi \lambda} d\Gamma)}{D^n(d\Gamma)}$$

([16, p.65]). On a alors le théorème de Szegő

$$\begin{aligned} \sigma_\infty^2 &= \prod_{j=0}^{\infty} (1 - |\alpha_j|^2) = \exp \int_{-1/2}^{1/2} \log \gamma(\lambda) d\lambda \\ \sum_{j \in \mathbb{N}} |\alpha_j|^2 < +\infty &\iff \int_{-1/2}^{1/2} \log \gamma(\lambda) d\lambda > -\infty \end{aligned}$$

où $\gamma(\lambda)$ est la partie absolument continue par rapport à $d\lambda$ de la mesure associée à $\{\alpha_n\}_{n \in \mathbb{N}}$. Le deuxième point de ce théorème est intéressant quand on considère des cas qui le violent. Si par exemple $\int_{-1/2}^{1/2} \log \gamma(\lambda) d\lambda$ diverge, alors $\sigma_\infty = 0$ et la prédiction linéaire est singulière : le présent du processus aléatoire est entièrement déterminé par son passé.

Le théorème de Baxter est aussi d'intérêt : la série des coefficients de Verblunski est sommable si et seulement si la série des moments de la mesure associée l'est également et $\gamma(\lambda) > 0$. Par exemple, les coefficients de corrélation partielles d'un modèle à longue dépendance (FARIMA) décroissent en $1/n$ [4], ce qui en appliquant ce théorème montre que la série des corrélations n'est pas non plus sommable.

Problème variationnel et fonction de Christoffel. La fonction de Christoffel associée au noyau (fonction symétrique définie positive) $k(x, y)$ qui génère un espace de Hilbert à noyau reproduisant (rkHs) $\mathcal{H}$ est définie par

$$\Lambda(x) = \inf_{f \in \mathcal{H}} (\|f\|_{\mathcal{H}}^2 / f(x) = 1)$$

dont on montre facilement qu'elle égale à $1/k(x, x)$.

Considérons l'ensemble $\mathbb{C}_p[z]$ des polynômes complexes de la variable complexe de degré inférieur à p . La forme bilinéaire

$$\begin{aligned} \langle \sum_i \alpha_i z^i | \sum_j \beta_j z^j \rangle &= \sum_{i,j} \alpha_i \bar{\beta}_j \oint_{\partial\mathbb{D}} z^i z^{-j} \gamma(z) \frac{dz}{z} \\ &= \sum_{i,j} \alpha_i \bar{\beta}_j \underbrace{\int_{-1/2}^{1/2} e^{2i\pi(i-j)\lambda} d\Gamma(\lambda)}_{c_{i-j}} \end{aligned}$$

satisfait les axiomes d'un produit hermitien. Elle est définie positive puisque les c_i étant les coefficients de corrélation (ou les moments) de la densité γ , l'expression précédente est positive ou nulle pour toutes suites α, β . $\mathbb{C}_p[z]$ peut donc être vu comme un rkHs avec ce produit hermitien. Si $\{\varphi_n\}_{n \in \mathbb{N}} = \{\phi_n / \|\phi_n\|\}_{n \in \mathbb{N}}$, le noyau s'écrit

$$k_p(z, w) = \sum_{i=0}^p \varphi_i(z) \overline{\varphi_i(w)}$$

La fonction de Christoffel associée en 0 est par suite $\Lambda_p(0) = \inf_{q \in \mathbb{C}_p[z]} (\|q\|^2 / q(0) = 1)$ et, on peut en particulier écrire en imposant $z \in \partial\mathbb{D}$ (ou $z = e^{2i\pi \lambda}$) que

$$\Lambda_p(0) = \inf_{b^p \in \mathbb{C}^p} (\|1 + \sum_j b_j^p e^{2i\pi j \lambda}\|^2)$$

3 OPUC et prédiction linéaire.

L'isomorphisme de Kolmogorov. Le Lemme 4 de [6] établit l'isomorphisme K entre l'espace engendré par la suite $\{x_t\}_{t \in \mathbb{Z}}$ et l'espace $L^2([-1/2; 1/2], \Gamma_x)$ des fonctions de carré intégrable sur $[-1/2; 1/2]$ contre la mesure spectrale de x_t , explicitement par

$$K : \mathcal{H}_x \longrightarrow L^2(\Gamma_x), \quad x_t \longmapsto \exp(2i\pi t \lambda)$$

par lequel $\langle x|y \rangle_{\mathcal{H}} = \langle K(x)|K(y) \rangle_{\Gamma_x}, \forall x, y \in \mathcal{H}_x$. Ce résultat n'est rien d'autre que le fait évident aujourd'hui que fonction de corrélation et densité spectrale de puissance sont transformées de Fourier l'une de l'autre. Ce résultat fondamental met en regard une série stationnaire d'un espace de Hilbert et les fonctions de carré sommable contre une mesure (dépendante de la série) sur le cercle unité.

Bouclons la boucle. Soit x_t une séquence stationnaire, de densité spectrale de puissance $\gamma(\lambda)$. Appliquons l'isomorphisme de Kolmogorov au problème (1). On obtient alors

$$\Lambda_p(0) = \inf_{b^p \in \mathbb{C}^p} (\|1 + \sum_j b_j^p e^{2i\pi j\lambda}\|^2) = \inf_{a^p} \|x_t - \sum_{j=1}^p a_j^p x_{t-j}\|_{\mathcal{H}}^2$$

Autrement dit, calculer la fonction de Christoffel associée à la mesure $\gamma(\lambda)d\lambda$ est équivalent à calculer l'erreur quadratique moyenne de prédiction à un pas pour une série stationnaire de densité spectrale de puissance $\gamma(\lambda)$. On montre en effet [16, p. 118] que

$$\Lambda_p(0) = \prod_{j=0}^{p-1} (1 - |\alpha_j|^2)$$

où $\{\alpha_p\}_p$ est la suite des coefficients de Verblunski, et par suite que $\lim_{p \rightarrow \infty} \Lambda_p(0) = \sigma_{\infty}^2$. Si la fonction de Christoffel est fondamentalement liée à l'erreur quadratique moyenne de prédiction, un lien étroit doit également exister pour les opérations réalisant les minimisations dans les deux programmes précédents. En examinant de près la récursion de Szegő et l'algorithme de Levinson, on s'aperçoit assez aisément qu'il s'agit de deux interprétations de la même récursion, une implantée temporellement, l'autre de manière équivalente par l'isomorphisme dans le domaine transformé (en z ou en fréquences λ). La conclusion est donc l'identification des coefficients de Verblunski α_n aux coefficients de corrélations partielles a_n^n .

Théorème de Verblunsky vs problème des moments. Le problème des moments pose la question de savoir si une suite de nombres c_n peut être considérée comme la suite des moments d'une mesure [1]. En traitement du signal, cette question revient à savoir si une telle suite de nombres peut être considérée comme une fonction de corrélation. Le théorème de Carathéodory-Toeplitz stipule que ces nombres constituent la suite des moments d'une mesure sur le cercle unité si et seulement si les déterminants D^n associés sont positifs pour tout n . Le théorème de Verblunski donne la réponse au problème, non plus pour les coefficients de corrélation c_n mais pour les coefficients de corrélations partielles α_n . Ce théorème stipule qu'à toute suite $\{\alpha_n\}_{n \in \mathbb{N}}$ de nombres appartenant au disque unité est associée une unique mesure sur le cercle unité (et réciproquement).

Approximation de Bernstein-Szegő. Cette approximation peut être vue comme le problème d'identification d'un processus stochastique par un modèle AR. Soit une mesure μ sur le cercle unité, $\{\alpha_n\}_{n \in \mathbb{N}}$ ses coefficients de Verblunski et $\{\phi_n\}_{n \in \mathbb{N}} = \{\phi_n / \|\phi_n\|\}_{n \in \mathbb{N}}$ la suite de ses polynômes orthonormaux. Alors pour tout n , $|\phi_n(\lambda)|^{-2}$ est une densité de probabilité (une densité spectrale de puissance normalisée), la mesure $d\mu_n = |\phi_n(\lambda)|^{-2} d\lambda$ a pour coefficients de Verblunski $\{\alpha_j\}_{0 \leq j < n}$ et $0, \forall j \geq n$, et μ_n converge vers μ (faiblement) quand n tend vers l'infini.

Taux d'entropie d'un processus Gaussien stationnaire. Le taux d'entropie $H(x)$ d'un processus stationnaire x de densité spectrale γ est défini par l'asymptotique de l'entropie d'un vecteur des échantillons de x , $\lim_{n \rightarrow \infty} n^{-1} H(x_1, x_2, \dots, x_n)$. Appliquée à un processus gaussien de coefficients de corrélation c_k , on obtient aisément

$$H(x) = \lim_{n \rightarrow \infty} \frac{1}{n} \log((2\pi e)^{n/2} \sqrt{D_n})$$

Le résultat de Szegő de 1915 concerne l'asymptotique des déterminants de matrices Toeplitz. Il montre que $\lim D_{n+1}/D_n = \sigma_{\infty}^2 = \lim D_n^{1/n}$, qui permet donc de montrer que $H(x) = \log(\sqrt{2\pi e}) + \int \log \gamma(\lambda) d\lambda$.

Quelques repères historiques :

1915 : Szegő résout une conjecture proposée par Pólya sur le comportement asymptotique des déterminants de matrice Toeplitz. Il montre $\sigma_{\infty}^2 = \lim D_n^{1/n}$.

1927 : Yule introduit l'autorégression pour modéliser des séries d'apparence périodique, et l'applique à la série des taches solaires.

1931 : Walker montre que l'autorégression de Yule est satisfaite également par la fonction de corrélation : équation appelée aujourd'hui Yule-Walker. Wiener-Hopf résout l'équation intégrale du même nom.

1931-34 : Kolmogorov formalise la théorie des probabilités et Khintchine celle des processus stochastiques.

1935-36 : Verblunski développe le problème des moments dans le monde des corrélations partielles (évidemment pas dans ces termes).

1938 : Wold défend sa thèse, publiée sous forme de livre [21].

1939 : Dans un papier aux CRAS, Kolmogorov donne l'erreur quadratique moyenne asymptotique pour l'interpolation et la prédiction de signaux aléatoires à temps discrets stationnaires.

1941 : Kolmogorov prouve ses résultats de 1939 dans deux papiers fondamentaux.

1942 : Wiener travaille pour l'armée américaine à des techniques de prédiction pour l'aide au guidage de systèmes de tir.

1947 : Levinson apporte une solution récursive au problème de la prédiction de Wiener en temps discret (redécouverte en 1960 par Durbin).

Remerciements : Les relecteurs du papier ont fait de nombreuses remarques et propositions d'amélioration, pas forcément intégrées, mais très utiles pour la version étendue [2]. Je les remercie sincèrement, en particulier, pour avoir pointé le livre de Pourahmadi [15] que je ne connaissais pas.

Références

- [1] N. I. Akhiezer. *The classical moment problem and some related questions in analysis*. Oliver & Boyd, 1965.
- [2] P. O. Amblard. *Une histoire de la prédiction à un pas*, 2025, en perpétuelle évolution.
- [3] N. H. Bingham. Szegő theorem and its probabilistic descendant. *Probability Surveys*, 9 :287–324, 2012.
- [4] P. J. Brockwell and R. A. Davis. *Time Series : Theory and Methods*. Springer, 1991.
- [5] A. Blanc-Lapierre and B. Picinbono. *Signaux aléatoires*. Masson, Paris, 1983.
- [6] A. N. Kolmogorov. Stationary sequences in Hilbert spaces. *Byull. Moskov. Gos. univ. Mat.*, 2(6) :1–40, 1941.
- [7] A. N. Kolmogorov. Interpolation and extrapolation of stationary random sequences. *Izv. Akad. Nauk SSSR Ser. Mat.*, 5 :3–14, 1941. (Reproduced in Selected works of A. N., vol ii, A. N. Shirayev ed., Springer 1992)
- [8] T. Kailath. A view of three decade of linear filtering theory. *IEEE Trans. Info. Th.*, 20(2) :146–181, 1974.
- [9] T. Kailath. *The legacy of Norbert Wiener : A centennial symposium*, volume 60, chapter Norbert Wiener and the development of mathematical engineering. AMS Proc. Symp. Pure Mathematics, 1997.
- [10] J. B. Lasserre, E. Pauwells, and M. Putinar. *The Christoffel-Darboux kernel for Data Analysis*. Cambridge University Press, 2022.
- [11] N. Levinson. The Wiener rms (root-mean square) error criterion in filter design and prediction. *J. Math. Phys. Mass. Inst. Tech.*, 25 :261–278, 1947.
- [12] N. Nikolski. *Hardy spaces*. Cambridge University Press, 2019.
- [13] N. Nikolski. *Toeplitz matrices and operators*. Cambridge University Press, 2020.
- [14] B. Picinbono. *Random signals and systems*. Prentice Hall, 1993.
- [15] M. Pourahmadi. *Foundations of time series analysis and prediction theory*. Wiley & Sons, 2001.
- [16] B. Simon. *Orthogonal polynomials on the unit circle . Part I : Classical theory*. AMS, colloquium publications, 2004.
- [17] B. Simon. *Orthogonal polynomials on the unit circle . Part II : Spectral theory*. AMS, colloquium publications, 2004.
- [18] B. Simon. *Szegő's theorem and its descendants. Spectral theory for L^2 perturbations of orthogonal polynomials*. Princeton University Press, 2011.
- [19] G. Szegő. *Orthogonal polynomials*. AMS, coll. pub. XXIII, 1939.
- [20] Sir G. Walker. On periodicities in series of related terms. *Proc. Roy. Soc. London*, A, 131(818) :518–532, 1931.
- [21] H. Wold. *A study in the analysis of stationary time series*. Almqvist & Wiksells Boktryckeri-A.-B., Uppsala, 1938.
- [22] G. U. Yule. On a method of investigating periodicities in disturbed series, with special reference to wolfer's sunspot numbers. *Philos. Trans. Roy. Soc., A*, 226 :267–298, 1927.