PREDICTION THEORY FOR STATIONARY FUNCTIONAL TIME SERIES

N. H. BINGHAM

Abstract

We survey aspects of prediction theory in infinitely many dimensions, with a view to the theory and applications of functional time series.

1. Introduction

This paper continues the theme of the author's earlier surveys [Bin1] on prediction theory for one-dimensional time series, [Bin2] on the finite-dimensional case, and touching briefly (with Badr Missaoui, [BinM]) on the *infinite-dimensional case*. This is our theme here. Our motivation is partly mathematical interest and completeness, partly the vigorous development of functional data analysis (FDA; [RamS1], [RamS2], [HorK]) made possible by the explosive growth in computer power, data storage and data handling.

We begin in §2 with the Cramér Representation (CR) and the Kolmogorov Isomorphism Theorem (KIT), on which everything rests. In §3 we turn to Verblunsky coefficients and Schur functions. Szegő's theorem and the Wold decomposition follow in §4, and the Szegő alternative and factorization in §5, all themes familiar from e.g. [Bin1], [Bin2]. Section 6 is on the Beurling-Lax-Halmos theorem and inner functions. Section 7 is on numerical implementation. Complements follow in §8, in particular (§8.1) the deterministic case and (§8.2) model spaces. We close in §9 with some open questions, with which the area abounds.

As with FDA itself, implementation involves discretization ('calculus is continuous, calculation is discrete'), and so in principle (subject to a suitable choice of dimension) reduces the setting to a finite-dimensional one, which can be handled by the finite-dimensional methods of e.g. [Bin2] and the references there (the extensions to higher dimensions, due to Whittle in 1963

and Wiggins and Robinson in 1965, of the Levinson-Durbin algorithm for the scalar case). We prefer to use the infinite-dimensional context suggested by the nature of the data. See e.g. §7.1 below, [RamS1], [DetKA] for more on this. (Ramsay and Silverman remark [RamS1, p.11] 'In general, prediction theory is beyond our scope, and is only considered here and there'.)

The mathematics here involves vectorial integration (see e.g. [Rud2, Ch. 12]).

2. The Cramér Representation and the Kolmogorov Isomorphism Theorem

2.1. The Cramér Representation (CR)

We confine ourselves throughout to *stationary* processes, that is, those whose distributions are invariant under *time-shifts*. One can work in discrete or continuous time, depending on preference, context or the data available. As the term 'time series' indicates, the first is the traditional one, and we shall follow it here. Then the time set is the integers $\mathbb{Z}$, with (Pontryagin-van Kampen) dual [Pon, Ch. 6] the torus $\mathbb{T}$ (equivalently $\mathbb{R}/2\pi\mathbb{Z}$; one can pass between these by $\theta \leftrightarrow e^{i\theta}$, and we shall do this at will). In the second case, the time-set is the real line $\mathbb{R}$, with dual $\mathbb{R}$ also. Stationarity is a strong condition, which yields correspondingly strong conclusions; we turn later to how it may be relaxed (§8.4). For a monograph treatment, see Nikolskii [Nik1], whose sub-title 'Spectral function theory' gives a hint of the mathematics involved (see also [Nik2]).

We write our process (or time series) as $x = \{x_n : n \in \mathbb{Z}\}$, where the x_n are in $\mathbb{R}$, $\mathbb{R}^d$, a Hilbert space $\mathcal{H}$ or a Banach space $\mathcal{B}$, depending on context (one can work more generally; see e.g. [BinM, §5.5, §5.6]). While our data consists of functions, as in our title, these functions will always belong to function spaces, which are at least topological vector spaces, the elements of which we will call vectors as usual. So vectors here are *infinite-dimensional* unless otherwise stated. Covariances are matrices in the finite-dimensional case (as in multivariate analysis in statistics, and in [Bin2]). But here, covariances are (linear) *operators*, hence the crucial role of operator theory in what follows. We will work in a Hilbert space $\mathcal{H}$ unless otherwise stated.

Write the time-shift $n \mapsto n+1$ as U. Then U is unitary, and generates a unitary group, $\mathcal{U}$. Being unitary, U is normal, and so the spectral theorem for the unitary case (Stone's theorem) applies ([Rud2, Ch. 12]; [Sto VIII.2], [RieN, §109], [DunS, X.2]). This involves integration over $\mathbb{T}$ with respect to

a *resolution of the identity* (projection-valued measure, in Mackey's terminology [Mac]) $E = \{E(\theta) : \theta \in \mathbb{T}\}$: $E(\mathbb{T}) = 1$, and

$$U^n = \int_{\mathbb{T}} e^{in\theta} dE(\theta), \qquad n \in \mathbb{Z} \qquad\qquad (U)$$

(as all our integrals will be over the torus $\mathbb{T}$, we omit the $\mathbb{T}$ below). Note that as $|e^{in\theta}| = 1$ and $E(\mathbb{T}) = 1$, for $x \in \mathcal{H}$ (U) gives

$$\|Ux\| = \|\int e^{i\theta} dE(\theta) x\| \leq \|x\| :$$

U is a *contraction*.

By above,

$$x_n = U^n x_0 = \int e^{in\theta} dE(\theta) x_0 = \int e^{in\theta} dY(\theta), \qquad n \in \mathbb{Z},$$

say, giving the *Cramér representation*

$$x_n = \int e^{in\theta} dY(\theta), \qquad n \in \mathbb{Z}. \qquad\qquad (CR)$$

The random measure Y here is the *Cramér measure* [Cra1], [Cra2], [CraL], or (Cramér's terminology) the *spectral process*, of the stationary process x. This is *orthogonally scattered* (Masani, [Mas1]): the masses of disjoint sets are orthogonal (see also Aue and van Delft [AueD], Prop. 2.1).

The Cramér representation holds also in continuous time; see e.g. [Kak3, §2 Th. 1].

2.2. The Kolmogorov Isomorphism Theorem (KIT)

One has the *Kolmogorov Isomorphism Theorem* ([Kak3]; [Kak4, p.104]; [ManS]; [DelE])], as in the scalar and matrix cases (see e.g. [Bin1], [Bin2]). Put succinctly and acronymically,

$$x(n) \leftrightarrow e^{in\cdot} I, \qquad n \in \mathbb{Z}, \qquad\qquad (KIT)$$

with I the identity operator (here the *time domain* is on the left, the *frequency domain* on the right). As with (CR), this holds also in continuous time. Formally, (a convenient source here is [Kak3, §2 Th. 3]):

Theorem (Kolmogorov Isomorphism Theorem). On a probability space $(\Omega, \mathcal{F}, P)$, write $X = L_0^2(\Omega)$ for the Hilbert space of complex-valued zero-mean square-integrable random variables, $\mathcal{H}(x)$ for the closed subspace of X spanned by the set $\{x(t) : t \in \mathbb{R}\}$. For $\{x(t)\}$ an X-valued stationary process on $\mathbb{R}$ with spectral measure μ, the time domain $\mathcal{H}(x)$ and the frequency (spectral) domain $L^2(\mu)$ are isomorphic by a unitary operator $U : \mathcal{H}(x) \to L^2(\mu)$ given by

$$Ux(t) = e^{it\cdot}, \qquad t \in \mathbb{R}.$$

We recall Chung's famous dictum '*The process is the thing*'. A stochastic process is an infinite-dimensional object, characterised by its *distribution*. Both (KIT) and (CR) deal with the distribution of the process in full, in different ways, with the distribution implicit in (KIT) and explicit in (CR). By contrast, we shall meet below results that deal only with the *second-order aspects* of the process: (mean and) covariance. The key result here is *Verblunsky's theorem*, or the *Verblunsky isomorphism* (§3), characterising (or parametrising) the covariance structure. This is fully informative about the process when the process is *Gaussian*; we discuss Gaussianity in §2.4 below (cf. [Bin4, §4.1] in continuous time).

2.3. The Gramian

Take $\mathcal{H}$ a separable Hilbert space, $B(\mathcal{H})$ the algebra of all bounded linear operators on $\mathcal{H}$, $T(\mathcal{H})$ the *trace class* operators in $B(\mathcal{H})$ (cf. §4.1 below). The Kolmogorov Isomorphism Theorem in the Hilbert case is due to Mandrekar and Salehi [ManS]; here we follow Kakihara [Kak1], [Kak2], [Kak3], [Kak4]. Write $X := L_0^2(\Omega, \mathcal{H})$ for the Hilbert space of all $\mathcal{H}$-valued (strong) random variables with mean 0 and finite second moment: with $(\Omega, \mathcal{F}, \mathbb{P})$ the probability space, $\mathbb{E}[x] := \int_\Omega x(\omega) d\mathbb{P}(\omega) = 0$,

$$\|x\|_X^2 := \int_\Omega \|x(\omega)\|_{\mathcal{H}}^2 \, d\mathbb{P}(\omega) = \mathbb{E}\left[\, \|x\|_{\mathcal{H}}^2\right] < \infty,$$

and so with inner product on X

$$(x, y)_X := \int_\Omega (x(\omega), y(\omega))_{\mathcal{H}} \, d\mathbb{P}(\omega) = \mathbb{E}[(x, y)_{\mathcal{H}}].$$

4

The *Gramian operator*, or *Gramian*, is the $T(\mathcal{H})$-valued inner product $[.,.]_X$ defined by

$$([x,y]_X \phi, \psi)_H = \int_\Omega (x(\omega), \phi)_\mathcal{H} (\psi, y(\omega))_\mathcal{H} \, d\mathbb{P}(\omega)$$

$$= \mathbb{E}[(x,\phi)_\mathcal{H} (\psi, y)_\mathcal{H}] \quad (x, y \in X, \phi, \psi \in \mathcal{H}). \qquad (GO)$$

In the finite-dimensional case, this reduces to a doubly-indexed set of scalars, which gives a matrix, the Gramian matrix [HorJ, §7.2], which is positive definite; so too is the Gramian operator likewise. Symbolically,

$$[x,y]_X = \int_\Omega x(\omega) \otimes \overline{y(\omega)} d\mathbb{P}(\omega) = \mathbb{E}[x \otimes \overline{y}],$$

where

$$(\phi \otimes \overline{\psi})\phi' := (\phi', \psi)_\mathcal{H} \phi, \qquad \phi, \phi', \psi \in \mathcal{H}.$$

Now X is both
(i) a left $B(\mathcal{H})$-module under the module action

$$(a, x) \mapsto a.x = ax \qquad (a \in B(\mathcal{H}), x \in \mathcal{H}),$$

(ii) a Hilbert space with Gramian $[.,.]_X$:
X is a *normal Hilbert $B(\mathcal{H})$-module* in the sense of [Kak2]. (For background on Hilbert modules, see Lance [Lan]. Both Hilbert modules and Gramians go back here at least to Masani's comments on Wiener's work on prediction [Mas2].)

We may now without ambiguity abbreviate $[.,.]_X$ (whose values are non-random operators) to $[.,.]$. With it, we can define the *spectral measure*, or *control measure*,

$$F(.) := [Y(.), Y(.)] \qquad (SM)$$

(operator-valued, indeed Gramian-valued), and the *operator covariance function*,

$$\Gamma(m, n) := [x(m), x(n)].$$

The process $x = \{x(n)\}$ is called *operator stationary* (or just *stationary*) if its operator covariance function (or just *covariance function*) is a function of $m - n$ only, $\tilde{\Gamma}(m - n) = \Gamma(m, n)$, say. We assume stationarity unless otherwise stated.

5

As in the matrix case above, the Gramian is *positive definite* (Gretsky [Gret, §3]). One has [Kak3] the *spectral representation*

$$\tilde{\Gamma}(n) = \int e^{in\theta} dF(\theta). \qquad (SR)$$

This is the operator version of Herglotz's theorem (or Bochner's theorem); cf. [Bin1], [Bin2] in finite dimensions and van Delft and Eichler [DelE] in this setting.

Below, we shall need the *spectral density f*, the Radon-Nikodym derivative of the absolutely continuous component of the spectral measure F with respect to normalised Lebesgue measure $d\theta/2\pi$ on $\mathbb{T}$, operator-valued, as F is (take f as 0 if F is singular).

2.4. Gaussianity

In (CR), one has the stationary process $x = (x_n)$ in the time domain represented as a stochastic integral of the orthogonal-increments process $Y = (Y(\theta))$ in the frequency domain. For Gaussians, orthogonality is the same as independence, so if Y is Gaussian, it has independent increments. As sums (and so integrals) of independent Gaussians are Gaussian, x is then Gaussian by (CR). The converse also holds (see e.g. [Cra3], [Horo]). All this is true *regardless of the dimension d*.

This simple and basic fact should thus have been stated (indeed, stressed) in the $d = 1$ case in [Bin1] and the case $1 < d < \infty$ in [Bin2]. Unfortunately, Gaussianity is only touched on in [Bin1] (§2, KIT, §4, Rajchman measures, $ 6.1, \phi$-mixing), and not even mentioned in [Bin2].

3. Verblunsky coefficients; Schur functions

3.1. Verblunsky's theorem

In the scalar case [Bin1], the distribution of a (discrete-time, complex-valued) stationary sequence, given by the spectral measure μ on the unit torus $\mathbb{T}$, may be fully described (encoded, parametrised) by a sequence $\alpha = (\alpha_n)_0^\infty$ of complex numbers $\alpha_n \in \mathbb{D}$, the unit disc of $\mathbb{C}$, the *Verblunsky doefficients* (there are several other names; see [Bin1], [Sim3]). This bijection

$$\mu \leftrightarrow \alpha \qquad (Ver)$$

is:

Theorem (Verblunsky's Theorem). There is a bijection between the sequences $\alpha = (\alpha_n)$ with each $\alpha_n \in \mathbb{D}$ and the probability measures μ on $\mathbb{T}$.

The relevant theory here depends heavily on orthogonal polynomials on the unit circle (OPUC), due originally to Szegő, the theme of Simon's books [Sim3], [Sim4] ([Sim4] deals with orthogonal polynomials on the line and the cirle together; these correspond to continuous and to discrete time).

In the finite-dimensional (ℓ-vector, $\ell \times \ell$ matrix) case, the relevant theory is matrix orthogonal polynomials on the unit circle (MOPUC) [Bin2]. The process also has a Verblunsky parametrisation, but now the α_n are ($\ell \times \ell$) matrices, of norm $\|\alpha_n\| < 1$. These encode the stationary processes, as before: *Verblunsky's theorem.* See Damanik, Pushnitski and Simon in 2008 [DamPS, (3.10), Th. 3.12]. In the infinite-dimensional case, the measures are vector-valued, the α_n are operators; see below.

3.2. Schur functions

In [DamPS], the authors write: 'Among the deepest and most elegant methods in OPUC are those of Khrushchev .... We have not been able to extend them to MOPUC! We regard their extension as an important open question'. The papers they referred to are [Khr1], [Khr2]; Khrushchev gave a monograph account of his work in [Khr3].

In addition to the sequence $\alpha = (\alpha_n)$ of Verblunsky parameters, there are two other ways of encoding the spectral measure μ on $\mathbb{T}$ that have been useful: the *Carathéodory function F* and the *Schur function f*, given by

$$F(z) := \int \Big(\frac{\theta + z}{\theta - z}\Big) d\mu(\theta), \qquad f(z) := z^{-1}(F(z) - 1)(F(z) + 1)^{-1}.$$

Here we use scalar notion, but the formulae extend to the matrix case and to the operator case by replacing 1 in the formula for f by the identity (and of course preserving the order of the factors in f as commutativity is lost). The Schur function f has a *continued-fraction expansion* in terms of the *Schur parameters γ_n*, where $\gamma_n < 1$ unless f is a finite Blaschke product. That the Schur parameters are the same as the Verblunsky parameters is *Geronimus's theorem* [Sim3, Part 1 p.3].

Remarkably enough, work on *quantum random walks* led Grünbaum, Velázquez, Werner and Werner in 2013 [GruV] and Bourgain, Grünbaum, Velázquez and Wilkenin in 2014 [BouG] to extend Khrushchev's work from OPUC to MOPUC, so answering the question raised in [DamPS] above.

7

The work on *operator-valued Schur functions* above extends [DamPS], and so also Verblunsky's theorem, to the operator case.

4. Szegő's theorem; Wold decomposition

4.1. Szegő's theorem

For *stationary* processes, the key result underlying prediction is *Szegő's theorem*, which relates to the influence of the *remote past*. This may be absent, e.g., bathwater forgetting its thermal history as it thermalises; total, e.g. tempered steel, whose thermal history is locked in; or present with a partial influence, e.g. the climate-weather interplay, where (on a time-scale, in years, short enough to neglect climate change) climate is permanent, while weather is (again on a suitable time-scale, in days) temporary and unpredictable, indeed chaotic.

To re-capitulate from [Bin1, Th. 3]: with σ^2 the variance of the least-squares prediction one step ahead from the whole (infinite) past, Szegő's theorem in the scalar case [Bin1, §4 Th. 3] tells us when the influence of the remote past is *not total*, i.e. when $\sigma > 0$, so each time-step adds genuine new randomness. The condition for this is *Szegő's condition* (Sz) below:

Theorem (Szegö's Theorem).
(i) $\sigma > 0$ iff the Szegő condition $\log w \in L_1$ holds, that is,

$$\int -\log w(\theta)d\theta > -\infty. \qquad (Sz)$$

(ii) $\sigma > 0$ iff $\alpha \in \ell_2$.
(iii)

$$\sigma^2 = \prod_1^\infty (1 - |\alpha_n|^2),$$

so $\sigma > 0$ iff the product converges, i.e. iff

$$\sum |\alpha_n|^2 < \infty : \qquad \alpha \in \ell_2;$$

(iv) σ^2 is the geometric mean $G(\mu)$ of μ: for $\sigma > 0$,

$$\sigma^2 = \exp(\frac{1}{2\pi} \int \log w(\theta)d\theta) =: G(\mu) > 0. \qquad (K)$$

Szegő's theorem is extended to the finite-dimensional setting by Derevyagin, Holtz, Khrushchev and Tyaglov in 2012 [DerH, Th. 28, Th. 29]: with

† for the adjoint (following their notation here), det and tr for determinant and trace,

$$\det \prod_0^\infty (\mathbb{I} - \alpha_k \alpha_k^\dagger) = \prod_0^\infty \det(\mathbb{I} - \alpha_k \alpha_k^\dagger) = \exp \int \operatorname{tr}\, \log f(\theta) d\theta/2\pi \quad (KSz)$$

(the *Kolmogorov-Szegő formula*; see e.g. [Bin1, §4]), and Szegő's condition – that the right here is positive – holds iff

$$\sum_0^\infty \|\alpha_k^\dagger \alpha_k\| < \infty \qquad\qquad (Sz)$$

(extending early work of Delsarte, Genin and Kamp [DelGK, Th. 18, 19]).

The *product theorem for determinants* in (KSz) above is simple linear algebra in finitely many dimensions, and holds quite generally. Neither of these is true in infinitely many dimensions! The results of [DerH] do extend to infinitely many dimensions, but as they involve determinants we must restrict to situations where infinite determinants are defined. Recall from functional analysis (see e.g. [Con1]):
(i) the *trace class* operators and the *Hilbert-Schmidt* operators, the two most important classes of operators (they suffice for our purposes here);
(ii) that $det(\mathbb{I} - A)$ is defined when the operator A is trace class (nuclear, in Grothendieck's terminology), see e.g. [Sim2];
(iii) that this is so when A is a product of two Hilbert-Schmidt operators;
(iv) the multiplication theorem for determinants holds here too [BriC], so one may interchange det and $\prod$ in (KSz) as above;
(v) by contrast, for Hilbert-Schmidt operators, the multiplication theorem for determinants involves extra terms, of the kind arising in Fredholm theory [BriC] (see §8.15).
The map $A \mapsto det(\mathbb{I} - A)$ is continuous (indeed, Lipschitz), so no convergence problems arise here [Sim1]. The restriction that the $\alpha_j^\dagger \alpha_j$ be trace class – that the α_j be Hilbert-Schmidt – is the natural one for Szegő's theorem to hold. It is not restrictive in practice; see §7.

Recall also that the condition for A to be trace class is that its eigenvalues λ_j (multiplicity counted) should be summable, $\sum |\lambda_j| < \infty$ ($\sum \lambda_j < \infty$ here as the $A = \alpha_k \alpha_k^\dagger$ are positive, so $\lambda_j > 0$). No such restriction on the eigenstructure is needed in the finite-dimensional case (see §8.15 below).

Payen [Pay] makes a thorough study of the Hilbert-valued case. There

([Pay, II.]; [BinM]) he gives an infinite-dimensional form of Szegő's theorem in terms of factorization (§5.2 below).

To summarise: Szegő's theorem extends to the setting of infinite-dimensional Hilbert space, but only when, with α_k the (operator) Verblunsky coefficients, the α_k are Hilbert-Schmidt, so the $\alpha_k \alpha_k^\dagger$ are trace class, so the $\det(\mathbb{I} - \alpha_k \alpha_k^\dagger)$ in (KSz) are defined.

4.2. The Szegő limit theorems

Strongly related to the work above are the *Szegő limit theorems*. Szegő's first (weak) limit theorem (1915) and much later second (strong) limit theorem (1952) concern the asymptotics of Toeplitz determinants T_n of order n as $n \to \infty$. The first concerns asymptotics of $\log \det T_n$, the condition for which is Szegő's condition, that the geometric mean of the spectral measure satisfies $G(\mu) > 0$. The second gives the asymptotics of $\det T_n$ itself, under a stronger condition, on $E(\mu)$ (in the standard notation). In their modern form, due to Ibragimov, both have the pleasing features that 'they hold whenever they make sense'; see e.g. [Bin1, §6] for details and references.

The proofs of both results were simplified dramatically by Geronimo and Case (1979) and Borodin and Okounkov (2000). The key result here, known as the *Borodin-Okounkov formula*, essentially reduces the limit results to algebra. This work has been further simplified, and extended to matix and operator versions, by a number of authors; we refer to the papers of Böttcher [Bot1], [Bot2], [Bot3] for details and references.

4.3. Baxter's Theorem

The Verblunsky doefficients $\alpha = (\alpha_n)$ are the *partial autocorrelation coefficients* in statistical language. Recall the great advantage of partial autocorrelations over ordinary correlations: the first give an *unrestricted parametrization*, as all values in the unit disk $\mathbb{D}$ can arise, while the second are restricted by complicated nested inequalities.

For convenience, abbreviate (Sz) above to $\alpha \in \ell_2(\mathbb{N})$, an L_2-condition. The result for the corresponding L_1-condition is Baxter's theorem:

Theorem (Baxter's theorem). The following are equivalent:
(i) the Verblunsky coefficients (or PACF) are summable,

$$\alpha \in \ell_1; \qquad\qquad (Bax)$$

(ii) the autocorrelations are summable, and μ is absolutely continuous with

continuous positive density:

$$\min_\theta w(\theta) > 0.$$

See [Sim3, Ch. 5], [Bin1, §5] for the scalar case, extended by Kasahara and Bingham [KasB1, §5]. A thorough study of Baxter's theorem in the matrix case was given by Dym and Kimsey [DymK]; cf. [Bin2], [KasB3]. We raise (§9, Q3) the question of extending Baxter's theorem to infinitely many dimensions.

4.4. Wold decomposition

The *past at time n*, the *remote past* and the *remote future* of the process $x = (x(n))$ are the closed linear subspaces spanned by the random variables below:

$$H(x, n) := \bigvee(x(k) : k < n), \quad H(x, -\infty) := \bigcap \left(\bigvee(x, n) : n \in \mathbb{Z}\right),$$

$$H(x, +\infty) := \bigcap \left(\bigvee(x, n) : n \in \mathbb{Z}\right).$$

The process is called *deterministic* if all three are equal, *purely non-deterministic (pnd)* if the second is trivial. The *Wold decomposition* [Kak2] splits the process $x = (x(n))$ into a deterministic and a purely non-deterministic component,

$$x = x_d + x_p,$$

which are Gramian orthogonal:

$$[x_d(m), x_p(n)] = 0 \quad (m, n \in \mathbb{Z}).$$

The Cramér measure Y of the process also splits, into a sum of absolutely continuous and singular components, Y_a and Y_s, and similarly for the spectral measure:

$$Y = Y_s + Y_a, \qquad F = F_s + F_a.$$

In one dimension, one has *Wold-Cramér concordance*: regarded as processes in their own right, the 'good' process x_p has Cramér and spectral measures Y_a, F_a, while the 'bad' one x_d has Y_s, F_s (see [Bin1] for details and references). In the matrix case, Wold-Cramér concordance holds in the full-rank case but not in general ([Bin2]; see e.g. Payen [Pay, Remarque 8, 376-7]). The vector case is studied in [Kak3], where such concordance is shown to be

preserved under dilation from the stationary case here to the harmonisable case (§ 7.4).

The 'good' component x_p in the Wold decomposition is a *moving average* of the products of the *innovations* (new randomness) at past times and the matrices appearing in the Taylor expansion of the Szegő function ('analytic square root' of the spectral density matrix), as we shall see below. Compare [Bin4, §2.5] in continuous time.

In the Hilbert-valued case, spectral criteria for the process to be purely non-deterministic are given by Kallianpur and Mandrekar [KalM].

5. Szegő alternative; factorization

5.1. Szegő alternative

In one dimension, one has a clean split, the *Szegő alternative*, between 'good' and 'bad' cases. In the first, there is a genuine innovation (input of new randomness) in each time-step from n to $n + 1$. The size of this new input is measured by the *prediction error variance* σ^2, the infimum of the variances of all linear predictors based on the present and (finite sections of) the past of the next future value x_{n+1}. This 'good' case can only happen if there is an absolutely continuous component F_a to F, that is, if the spectral density f is not a.e. zero. When this is so, it happens if and only if the *Szegő condition* (Sz) holds, and then $\sigma^2 > 0$ is the geometric mean of $\log f$, as in Szegő's Theorem. Note that the singular component of F (if present) *plays no role* here.

In the Wold decomposition, x is the sum of a 'bad' part x_d (§3), and the 'good' part x_p, a moving average of the innovations over the past to date; this contains a factor σ, and so is absent if $\sigma = 0$ [Bin1]. In terms of the Verblunsky coefficients,

$$\sigma^2 = \prod_1^\infty (1 - |\alpha_n|^2),$$

so (Sz) holds iff $\sigma > 0$ iff the product converges. Then (and only then), one can define the *Szegő function* $h(z)$:

$$h(z) := \exp\left(\frac{1}{2} \int \left(\frac{e^{i\theta} + z}{e^{i\theta} - z}\right) \log w(\theta) d\theta/2\pi\right) \qquad (z \in D). \qquad (OF)$$

This is an *outer function*, in the Hardy space $H^2(\mathbb{D})$ ([Dur]; [Sim3]; [Bin1]). It is analytic in $\mathbb{D}$, and zero-free there [Sim3, Th. 2.4.1]. As (from Fatou's

theorem: [GarMR, Th. 1.10], [Rud1, Th. 17.10])

$$\int \left(\frac{e^{i\theta} + z}{e^{i\theta} - z}\right) \log w(\theta) \, d\theta/2\pi \to \log f(\phi) \quad a.e. \qquad (z = re^{i\phi}, \quad r \uparrow 1),$$

its square has radial limit f a.e. on $\mathbb{T}$:

$$\lim |h(z)|^2 = f(\theta) \quad a.e. \quad (z = re^{i\theta}, \ r \uparrow 1).$$

It may thus be regarded as the *analytic square root* of the spectral density f.

In the matrix case, we refer to [Bin2, §5] for details and references. Szegő's theorem (the matrix forms of (Sz), (KSz) above) holds, see [DerH, Th. 28], extended to the infinite-dimensional (operator) case under the *trace-class* restriction in §4.1.

5.2. Factorization

In the scalar case, existence of the Szegő function (analytic square root of the spectral density) and the Szegő condition are equivalent. One seeks a matrix and a vector version of this. So one seeks to factorise a matrix spectral density into its analytic square root times its adjoint (denoted by $*$ here, to conform to the sources below),

$$F = \Phi\Phi^*.$$

There is an extensive theory here, due mainly to Wiener and Masani and to Helson and Lowdenslager. We note that the coefficient matrices Φ_n in the Taylor expansion

$$\Phi(z) = \sum_0^\infty \Phi_n z^n$$

appear in the prediction-error matrix (and operator, in the vector case, below). See e.g. Masani [Mas2, p.278], Whittle [Whi] (for the scalar case, see e.g. [GreSz, §10.8]).

For factorizations in the operator case, we refer to [RosR Ch. 6]. We also mention briefly the approach of Power [Pow]. We make the restrictive assumption that the spectral density f is essentially bounded (i.e. is bounded, after excluding some null set on the torus $\mathbb{T}$). Then

$$f = hh^* + g,$$

13

where g is a positive operator and h is analytic and outer. Here g can be taken to be *minimal*. A Wold(-Zasuhin) decomposition is obtained. The prediction-error operator $G(f)$ of f is obtained as

$$G(f) = G(hh^*) = (QhQ)(QhQ)^*,$$

where Q is an orthogonal projection; we refer to [Pow] for details.

As Power points out, his method gives g and h as functions of f, while the methods of Sz.-Nagy and Foiaş ([SzNFBK]; §5 below) do not. But this applies only to the behaviour encoded in the spectral *density*, the absolutely continuous component in the Lebesgue decomposition of the spectral *measure*. We turn now to how to address the 'bad' behaviour encoded in the other two components.

6. The Beurling-Lax-Halmos theorem; inner functions

So far as prediction theory is concerned, it is the 'good' part of the process (or its spectral measure) that matters. Nevertheless, the Wold decomposition shows us that to understand the structure of a stationary process we need to look at the 'bad' part also.

Recall the *outer function* of (OF), §4 above. The term is due to Beurling [Beu], as is the corresponding term *inner function*. These correspond respectively to the 'good' and 'bad' parts above. When we pass from the time domain (Wold decomposition) to the frequency (or spectral) domain by the Kolmogorov Isomorphism Theorem, we obtain the *factorization* into outer and inner factors in Hardy spaces $H^2(\mathcal{H})$. For the scalar case (where the Hardy space is written H^2) there are good accounts in Duren [Dur], Garnett [Garn] and Hoffman [Hof] (see also [Rud1, Ch. 17]). For the vector case, see [SzNF, III, V], [Nik1, I.7, XI.3], [RosR, Ch. 4-6], [Hof, 114-116].

The remote past (the 'bad' part) is invariant under the time-shift. The invariant subspaces are exactly those given by multiplication in $H^2(\mathcal{H})$ by an inner function ([SzNF, V, Th. 3.3]; [Nik1, I.7, XI.3]; [RosR, 1.12]). This is the *Beurling-Lax-Halmos theorem*, due to Beurling (dimension $d = 1$), Lax [Lax] ($1 < d < \infty$) and Halmos [Halm] ($d \leq \infty$) [Con2]. So study of the remote past reduces to study of inner functions, u say. These have a rich structure; they factorise into a unimodular constant, a Blaschke product, and an integral factor as in (OF) but with a singular rather than an absolutely continuous measure. See §7.1,2 below.

Halmos's approach [Halm] uses the concept of a *wandering subspace*; these correspond to the *innovations* (new randomness) in the process. For an isometry V on a Hilbert space $\mathcal{H}$, call a subspace L of $\mathcal{H}$ *wandering* if for distinct integers m, n $V^m L$ and $V^n L$ are orthogonal. Then [SzNF, §1.1] if

$$M_+(L) := \oplus_0^\infty V^n L,$$

one has

$$L = M_+(L) \ominus V M_+(L).$$

Call an isometry V on $\mathcal{H}$ (which will be U of §2 for us) a *unilateral shift* if $\mathcal{H}$ has a wandering subspace L with $M_+(L) = \mathcal{H}$. Then L, called *generating* for V, is uniquely determined by V, indeed $L = \mathcal{H} \ominus V\mathcal{H}$. One has the *Wold decomposition* ([SzN, Th. 1.1], [RosR, §1.3]): $\mathcal{H}$ decomposes into the orthogonal sum $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$ such that the $\mathcal{H}_i$ reduce $\mathcal{H}$ (each is mapped onto itself by V), $V|\mathcal{H}_0$ is unitary and $V|\mathcal{H}_1$ is a unilateral shift. This decomposition is unique; indeed,

$$\mathcal{H}_0 = \bigcap_0^\infty V^n \mathcal{H}, \quad \mathcal{H}_1 = M_+(L), \text{ where } L = \mathcal{H} \ominus V\mathcal{H}.$$

Either $\mathcal{H}_0$ or $\mathcal{H}_1$ may be absent ($= \{0\}$).

7. Implementation

7.1. Theory

As remarked before, calculus is continuous; calculation is discrete. Data is discrete. Our given data on past curves (at times $1, \cdots, n$ say – the distinction between being given the infinite past or a finite part of it is important theoretically but evaporates at the implementation stage) necessarily consists of finitely many points on each, interpolated or smoothed so as to give a curve (continuous, say), with whatever degree of smoothness the statistician chooses (as with density estimation; see e.g. Silverman [Sil]), by whatever means the statistician chooses – splines, wavelets etc. The prediction process then consists of the inevitable three steps:
(i) Discretisation of the data (now curves), into d-vectors, for some $d < \infty$. For choice of d, see e.g. Li and Hsing [LiH].
(ii) Prediction, using e.g. the multidimensional version of the Levinson-Durbin algorithm, as in [Bin2]; see below.

(iii) Interpolation or smoothing of this set of predicted values at time $n + 1$ to give the predicted curve at time $n + 1$.

As one will see from the above, the practical problems involved at the implementation stage are largely *numerical.* This is a familiar phenomenon; see e.g. [BinS2] for a different setting (random fields) where the writer (with Symons) recently encountered such things.

The Levinson-Durbin algorithm

For this classical algorithm for the computation of the best linear predictor based on the last n data points, see e.g. Brockwell and Davis [BroD, §5.2], [Bin1] in the scalar case, [Bin2] in the matrix case. It is worth noting that numerical improvements have been made here (the *split Levinson algorithm*); see Delsarte and Genin [DelG].

7.2. Numerics

The three-step procedure above thus amounts substantially to

(i) *discretization* of the infinite-dimensional process (random curves) in (CR) to random d-vectors for suitably chosen d;

(ii) *prediction* by finite-dimensional methods [Bin2] (e.g., the Levinson-Durbin algorithm);

(iii) *smoothing* (e.g., spline interpolation with a roughness penalty) to return to the infinite-dimensional setting.

See also Hyndman and Shang [HynS1], [HynS2].

One way to implement (i) is to expand the covariance by Mercer's theorem and use an orthogonal eigenexpansion, the *Karhunen-Loève expansion* [Loe2, II, 37.5], then truncate after a suitable number of terms. See e.g. Hall et al. [HalPP], Aue et al. [AueNH] for implementation. In the Gaussian case of §1.4 [MarR, §5.3], the terms are independent (and Gaussian). Such methods can be extremely effective (for background see e.g. [BinS1], [BinS2] and the references there).

Kernel methods

Kernel methods, commonly used in machine learning, have recently been advocated for functional prediction by Hashimoto et al. [Has1], [Has2].

8. Complements

8.1. The deterministic case

When the Szegő condition (Sz) fails (e.g., when the density is absent – $F_a = 0$ in the notation of §3), the entire process is 'bad', and consists entirely

of 'echoes of the remote past'. In the scalar case (for simplicity): the variance σ_n^2 of the best linear predictor based on the last n readings decreases to $\sigma^2 = 0$. The interesting question of how fast was addressed long ago by Rosenblatt [Ros], and more recently by Babayan, Ginovyan and Taqqu [BabGT].

8.2. Model spaces

Again in the scalar case first for simplicity: the (unilateral, forward) *shift* S (time $n \mapsto n + 1$) is represented on the Hardy space H^2 by

$$Sf = zf \qquad (f = f(z) \in H^2).$$

This has adjoint the backward shift S^*:

$$S^*f = \frac{f - f(0)}{z}.$$

This follows from the Taylor series for $f(z) = \sum_0^\infty a_n z^n$ in each case. Note that $f \in H^2$ and $a = (a_n) \in \ell_2(\mathbb{N})$ are equivalent, and give the 'Hardy-Hilbert space' (see de Branges [dBra], Martinez-Avendano and Rosenthal [MartR] for monograph treatments).

By the Beurling(-Lax-Halmos) theorem, the invariant subspaces of S are uH^2 for the inner functions u. Similarly, the invariant subspaces for S^* are the orthogonal complements of these, written

$$\mathcal{K}_u := (uH^2)^\perp.$$

These are called *model spaces* ('model spaces are the invariant subspaces of the backward shift'). To explain the terminology, we quote [GarMR, 105] 'The term *model space* originates in the theory of model operators, developed by Sz.-Nagy and Foiaş, where it is shown that certain types of Hilbert-space contractions are unitarily equivalent to the compressions of the unilateral shift to a model space. This underscores the importance of model spaces in developing concrete, function-theoretic realizations of abstract Hilbert space operators.' For background, see the two classic sources [SzNF] and Nikolskii [Nik1] cited in §5, plus Rosenblum and Rovnyak [RosR] and the two-volume [Nik2], and the two recent treatments by Garcia, Mashreghi and Ross [GarMR] cited above (see also [GarR]) and Agler, McCarthy and Young [AglMY]. (Note that different authors use the term compression differently

[GarMR, Remark 9.2].) See also [Nik2, §4.8.8] and [NikV].

8.3. Compressions and dilations

For A, B operators on spaces $\mathcal{A} \subset \mathcal{B}$, B is a *dilation* of A if

$$A^n = pr\ B^n \quad \text{for all } n \in \mathcal{N},$$

where pr is projection [SzNF, p.10]. Then A is a *compression* of B ([GarMR, Def. 9.1]; [SzNF] does not use the term compression).

The basic result here is the *Sz.-Nagy-Foiaş dilation theorem*: if H is a Hilbert space and T a linear contraction on it, there exists a larger Hilbert space $\mathcal{H}$ and a unitary operator U on $\mathcal{H}$ with U a dilation of T and

$$\mathcal{H} = \bigvee (U^n H : n \in \mathbb{N})$$

(such a dilation is called *minimal*). For the extensive theory here, see e.g. [SzNF, Ch. I], [Nik1, Introductory Lecture, Lecture III], [GarMR, Ch. 9].

8.4. Harmonizability

This concept, due to Loève [Loè] in 1948, addresses the need to relax the strong assumption of stationarity, which one cannot expect to hold exactly in practice. It has been studied and extended by Karhunen, Cramér and others. As the covariance function now needs two arguments, s and t say, rather than one, $s - t$, the spectral measure now becomes a 'bimeasure'. For details and references, see e.g. [Rao1,2], [Kak1-4]. To summarise [Rao2, 292]: strongly harmonisable $\subset$ weakly harmonizable $\subset$ Karhunen $\subset$ Cramér.

Note [Kak3, Th. 2] that a process is weakly harmonizable iff it has a *stationary dilation*. For readers to whom dilation theory is new, we offer the following heuristic for this result. For a population in equilibrium ('zero population growth', ZPG), at population level things look stationary even over long time-scales. At the individual level, things look far from stationary as one progresses through life.

Thus weak harmonizability is the broadest context in which we can hope to bring the powerful tools available in the stationary case to bear. Beyond that, one is more in the realm of the Kalman filter and its extensions, where the dynamics dominate, one predicts using only the immediate past and the present, and what matters is speed and accuracy of reaction (e.g., control of manned spacecraft, or mortar-locating radar).

8.5. Banach spaces and beyond

Such extensions were addressed briefly in [BinM, §5.5]. We refer there for further detail, and to e.g. Dette et al. [DetKA], Chobanyan and Weron [ChoW], Weron [Wer], Miamee and Salehi [MiaS] and Klotz and Riedel [KloR]. For a Banach-space version of the Wold decomposition, see [FauH].

8.6. Operator-valued processes

For an approach via dilation theory and operator models, see Makagon and Salehi [MakS, §2] and the references cited there. The vector- and operator-valued cases are developed together in [RosR, Ch. 4-6].

8.7. The multivariate case

See e.g. [Kak3, §6] for a multivariate vector-valued treatment.

8.8. High- and infinite-dimensional probability and statistics.

For high-dimensional treatments, see e.g. [Ver] for probability, [Wai] for statistics. For infinite-dimensional statistics, see e.g. Giné and Nickl [GinN].

8.9. Specifically infinite-dimensional phenomena.

We have been concerned here with vector-valued Hardy space theory (inner-outer function factorization, etc.), in finitely many dimensions [Bin2], or infinitely many, our main concern here. There are results in the area which do *not* extend to the infinite-dimensional case. For example, see Treil [Tre1], [Tre2] for the (operator) corona problem.

8.10. Cramér-Karhunen-Loève expansion

The Karhunen-Loève expansion ([MarR, §5.3], [GinN, Th. 2.6.10]) may be combined with the Cramér representation to give a dynamicized form of the expansion, which has been used for modelling functional time series. See e.g. Antoniadis and Sapatinas [AntS] (who considered El Niño), [AntPS], Panaretos and Tavakoli [PanT1], [PanT2].

8.11. Functional data analysis; prediction; change-points

For some recent developments in these areas, see e.g. Aue, Norinho and Hörmann [AueNH], Dette, Kokot and Aue [DetKA].

8.12. Smoothness of functions

Our data are functions, drawn from function spaces, often Hilbert spaces

$\mathcal{H}$. These may (and typically do) have some smoothness properties. One of the most basic is *continuity of point evaluation*, $x \mapsto f(x)$ $(f \in \mathcal{H})$. By the Riesz representation theorem, this is the condition for $\mathcal{H}$ to be a *reproducing-kernel Hilbert space* (RKHS): to have a *reproducing kernel* $k(.,.)$ such that, with $k_x(.) := k(., x)$,

$$(f, k_x) = (f(.), k(., x))_{\mathcal{H}} = f(x) \qquad (f \in \mathcal{H}).$$

Such spaces are common in the Hardy-space setting above. For example, the Hardy-Hilbert space H^2 is a RKHS with kernel

$$k_\lambda(z) = 1/(1 - \overline{\lambda}z) \qquad (\lambda, z \in \mathbb{D})$$

[GarMR, Prop. 3.3], and so are the model spaces $\mathcal{K}_u$ above, with kernel (with the u here to be understood in the notation)

$$k_\lambda(z) = \frac{1 - \overline{u(\lambda)}u(z)}{1 - \overline{\lambda}z} \qquad (\lambda, z \in \mathbb{D})$$

[GarMR, §5.5]. For the extensive theory of RKHS, and applications to probability and statistics, see e.g. Berlinet and Thomas-Agnan [BerTA].

8.13. Hankel operators and Nehari sequences.

Toeplitz operators occur frequently in the above; they have many links with Hankel operators (for which see e.g. Peller [Pel]). The *Nehari problem* [Pel, Ch. 5] is: given a sequence $\gamma = (\gamma_n)$ $(\gamma_n \in \mathbb{C})$, find a function ϕ in the unit ball of L_∞ with

$$\gamma_n = \int e^{in\theta} \phi(\theta) d\theta / 2\pi \quad (n = 1, 2, \cdots).$$

Nehari's theorem is that such a solution exists iff the Hankel matrix (γ_{m+n}) of γ acts as a contraction on ℓ_2. There is more than one solution (the *indeterminate case*) iff $\gamma = (\gamma_n)$ is the negatively indexed Fourier coefficients of the phase factor of some function in H_2; in this case γ is called a *Nehari sequence* (compare the determinate and indeterminate cases in the moment problem; see e.g. [Bin3] and the references there).

Nehari sequences also occur in prediction theory, for example in connection with the condition of *complete non-determinacy* for time series (see e.g. [KasB1]). Study of problems of Nehari type has led (among other things)

to extensions of the strong Szegő and Baxter theorems in the scalar case [KasB1, §5] and to matrix forms of Baxter's theorem [KasB2,3].

8.14. Continuous time.
We refer to [Bin4] for the analogous continuous-time case (see §4.4 there for ways of passing between the two).

8.15. Infinite determinants.
Infinite determinants go back to Fredholm's work on integral equations in 1903 (and so pre-date functional analysis, which they helped to motivate); see Smithies [Smi], Ruston [Rus]. We recommend Simon's survey [Sim1]. See also Britz et al. [BriT] for 'regularized' determinants, and Fuglede-Kadison determinants, for which see e.g. [BleL].

9. Questions

We close with some questions arising from the work surveyed above.
Q1. Is there an infinite-dimensional version of Szegő's theory of orthogonal polynomials on the unit circle (OPUC, [Sim3], [Bin1]; MOPUC, [Bin2])? In this regard, see [GarMR, Example 8.3]. This involves the *Szegő* (or *Toeplitz*) *conjugation*, which is familiar from the *Szegő recursion* of OPUC and MOPUC.
Q2. Can one express the trace-class condition of §4.1 in terms of the spectral measure μ itself?
Q3. Can Baxter's theorem (§4.1) be extended to the infinite-dimensional (operator) case, by operator-valued Schur functions or other means?
Q4. Can the assumption of essentially bounded spectral density in [Pow], §4, be relaxed (or dropped)?
Q5. Can the assumption of having a scalar multiple in [SzNF, V.6,7], §4, be relaxed (or dropped)?

References

[AglMY] J. Agler, J. E. McCarthy and N. Young, *Operator analysis: Hilbert space methods in complex analysis*. Cambridge Tracts Math. **219**, Cambridge University Press, 2020.

[AntPS] A. Antoniadis, E. Paroditis and T. Sapatinas, A functional wavelet-kernel approach for time series prediction. *J. Roy. Stat. Soc. B* **68** (2006), 837-857.

[AntS] A. Antoniadis and T. Sapatinas, Wavelet methods for continuous-time prediction using Hilbert-valued autoregressive processes. *J. Multivariate Analysis* **87** (2003), 133-158.

[AueD] A. Aue and A. van Delft, Testing for stationarity of functional time series in the frequency domain. *Ann. Statist.* **48** (2020), 2505-2547; arXiv:1701.01741.

[AueNH] A. Aue, D. D. Norinho and S. Hörmann, On the prediction of stationary functional time series. *J. Amer. Stat. Soc.* **110** (2015), 378-392.

[BabGT] N. K. Babayan, M. S. Ginovyan and M. S. Taqqu, Extensions of Rosenblatt's results on the asymptotic behaviour of the prediction error for deterministic stochastic sequences. Special Issue: Murray Rpsenblatt Memorial, *J. Time Series Analysis* **42**(5-6) (2021), 622-652; arXiv:2006.00430.

[BerTA] A. Berlinet and C. Thomas-Agnan, *Reproducing kernel Hilbert spaces with applications in probability and statistics*. With a preface by Persi Diaconis. Kluwer, 2004.

[Beu] A. Beurling, On two problems concerning linear transformations in Hilbert space. *Acta Math.* **81** (1948), 239-255 (reprinted in *Collected works of Arne Beurling*, Vol. **2**, *Harmonic analysis* (ed. L. Carleson et al.), Birkhäuser, 1989).

[Bin1] N. H. Bingham, Szegő's theorem and its probabilistic descendants. *Probability Surveys* **9** (2012), 287-324.

[Bin2] N. H. Bingham, Multivariate prediction and matrix Szegő theory. *Probability Surveys* **9** (2012), 325-339.

[Bin3] N. H. Bingham, The life, work and legacy of P. L. Chebyshev. *Proceedings: Bicentennial Conference on P. L. Chebyshev (1821-1894)* (ed. A. N. Shiryaev), *Th. Probab. Appl.* **66**(4) (2021), to appear.

[Bin4] N. H. Bingham, Prediction theory in continuous time. arXiv:2111.08560.

[BinM] N. H. Bingham and Badr Missaoui, Aspects of prediction. *J. Applied Probability* **51A** (2014), 189-201.

[BinS1] N. H. Bingham and Tasmin L. Symons, Gaussian random fields on sphere and sphere cross line. *Stochastic Proc. Appl.* (Larry Shepp Memorial

Issue); arXiv:1812.02103; `https://doi.org/10.1016/j.spa.2019.08.007`.

[BinS2] N. H. Bingham and Tasmin L. Symons, Aspects of random fields. *Theory of Probability and Mathematical Statistics* (Special Issue in honour of M. I. Yadrenko, ed. A. Olenko), to appear.

[BleL] D. P. Blecher and L. E. Labuschagne, Applications of the Fuglede-Kadison determinant: Szegő's theorem and outers for noncommutative H^p. *Trans. Amer. Math. Soc.* **360** (2008), 6131-6147.

[Bot1] A. Böttcher, One more proof of the Borodin-Okounkov formula for Toeplitz determinants. *Integral Equations and Operator Theory* **41** (2001), 123-125.

[Bot2] A. Böttcher, On the determinant formulas by Borodin, Okounkov, Baik, Deift and Rains. *Operator Theory Advances and Applications* **135** (2002), 91-99.

[Bot3] A. Böttcher, Borodin-Okounkov and Szegő for Toeplitz operators in model spaces. *Integral Equations and Operator Theory* **78** (2014), 407-414.

[BouG] J. Bourgain, F. A. Grünbaum, L. Velázquez, A. H. Werner and R. F. Werner, Quantum recurrence of a subspace and operator-valued Schur functions. *Comm. Math. Phys.* **329** (2014), 1031-1067.

[dBra] L. de Branges, *Square-summable power series*. Springer, Monographs in Math., 2010 (1st ed., with J. Rovnyak, Holt, Rinehart and Winston, 1966).

[BriC] T. Britz, A. Carey, F. Gesztesy, R. Nichols, F. Sukhochev and D. Zanin, The product formula for regularized Fredholm determinants. *Proc. Amer. Math. Soc. B* **8** (2021), 42-51.

[BroD] P. J. Brockwell and R. A. Davis, *Time series: Theory and methods*, 2nd ed., Springer, 1991 (1st ed. 1987).

[ChoW] S. A. Chobanyan and A. Weron, Banach space valued stationry processes and their linear prediction. *Dissertationes Math.* **125**, 1975.

[Con1] J. B. Conway, *A course in functional analysis*, 2nd ed. Grad. Texts in Math. **96**, Springer, 1990 (1st ed. 1985).

[Con2] J. B. Conway, *A course in operator theory*. Grad. Studies in Math. **21**, Amer. Math. Soc., 2000.

[Cra1] H. Cramér, On the theory of stationary random processes. *Ann. Math.* **41** (1940), 215-230 (reprinted in *Collected Works of Harald Cramér, Volume II*, 925-940, Springer, 1994).

[Cra2] H. Cramér, On harmonic analysis in certain function spaces. *Ark. Mat. Astr. Fys.* **28B** (1942), 1-7 (*Works II*, 941-947).

[Cra3] H. Cramér, A contribution to the theory of stochastic processes. *Proc. Second Berkeley Symposium Mat. Stat. Prob* (ed. J. Neyman) 329-339, U.

California Press, 1951 (*Works II*, 992-1002).

[CraL] H. Cramér and R. Leadbetter, *Stationary and related stochastic processes.* Wiley, 1967.

[DamPS] D. Damanik, A. Pushnitski and B. Simon, The analytic theory of matrix orthogonal polynomials. *Surveys in Approximation Theory* **4** (2008), 1-85; arXiv:0711:2703.

[DelE] A. van Delft and M. Eichler, A note on Herglotz's theorem for time series on function spaces. *Stoch. Proc. Appl.* **130** (2020), 3687-3710.

[DelG] P. Delsarte and Y. Genin, The split Levinson algorithm. *IEEE Trans. Acoust. Speech Signal Proc* ASSP-**34** (1986), 470-478.

[DelGK] P. Delsarte, Y. Genin and Y. G. Kamp, Orthogonal polynomial matrices on the unit circle. *IEEE Trans. Ciruits and Systems* CAS-**25** (1978), 149-160.

[DerH] M. Derevyagin, O. Holtz, S. Khrushchev and M. Tyaglov, Szegő's theorem for matrix orthogonal polynomials. *J. Approx. Th.* **164** (2012), 1238-1261; arXiv:1104.4999.

[DetKA] H. Dette, K. Kokot and A. Aue, Functional data analysis in the Banach space of continuous functions. *Ann. Stat.* **48** (2020), 1168-1192.

[DunS] N. Dunford and J. T. Schwartz, *Linear operators, II: Spectral theory, Self-adjoint operators in Hilbert space*, Wiley, 1963.

[Dur] P. L. Duren, *Theory of H^p spaces.* Academic Press, 1970.

[DymK] H. Dym and D. P. Kimsey, CMV matrices, a matrix version of Baxter's theorem, scattering and de Branges spaces. *EMS Surveys Math. Sci.* **3** (2016), 105p.

[FauH] G. D. Faulkner and J. E. Honeycutt, Orthogonal decompositions of isometries in a Banach space. *Proc. Amer. Math. Soc.* **69** (1978), 125-128.

[GarMR] S. R. Garcia, J. Mashreghi and W. T. Ross, *Introduction to model spaces and their operators.* Cambridge Stud. Adv. Math. **148**, Cambridge University Press, 2016.

[GarR] S. R. Garcia and W. T. Ross, Model spaces: A survey. *Invariant subspaces of the shift operator*, 197-245. *Contemp. Math.* **638**, Amer. Math. Soc., 2015.

[Garn] J. B. Garnett, *Bounded analytic functions.* Academic Press, 1981 (revised 1st ed., Grad. Texts Math. **236**, Springer, 2007).

[GinN] E. Giné and R. Nickl, *Mathematical foundations of infinite-dimensional statistical models.* Cambridge University Press, 2016.

[GreSz] U. Grenander and G. Szegő, *Toeplitz forms and their applications.* U. California Press, 1958.

[Gret] N. E. Gretsky, Operator valued Gramians and inner products in vector valued function spaces. *Ann. Mat. Pura Appl.* **101** (1976), 337-354.

[GruV] F. A. Grünbaum, L. Velázquez, A. H. Werner and R. F. Werner, Recurrence for discrete-time unitary evolution. *Comm. Math. Phys.* **320** (2013), 543-569.

[HalPP] P. Hall, S. S. Poskitt and B. Presnell, A functional data analytic approach to signal detection. *Technometrics* **43** (2001), 1-9.

[Halm] P. R. Halmos, Shifts on Hilbert spaces. *J. Reine Angew. Math.* **208** (1961), 102-112.

[Has1] Y. Hashimoto, I. Ishikawa, M. Ikeda, F. Kumura and Y. Kawahara, Kernel mean embeddings of von Neumann-algebra-valued measures. arXiv:2007.14698.

[Has2] Y. Hashimoto, I. Ishikawa, M. Ikeda, F. Kumura, T. Katsura and Y. Kawahara, Reproducing kernel Hilbert C^*-module and kernel mean embedding. arXiv:2101.11410.

[Hof] K. Hoffman, *Banach spaces of analytic functions.* Prentice-Hall, 1962.

[HorJ] R. A. Horn and C. R. Johnson, *Matix analysis*, Cambridge University Press, 1985.

[HorK] L. Hórvath and P. Kokoszka, *Inference for functional data with applications.* Springer, 2012.

[Horo] J. Horowitz, Gaussian random measures. *Stoch. Proc. Appl.* **22** (1986), 129-133.

[HynS1] R. J. Hyndman and H. L. Shang, Forecasting functional time series (with discussion). *J. Korean Math. Soc.* **38** (2009), 199-221.

[HynS2] R. J. Hyndman and H. L. Shang, The `ftsa` package for R. `https://cran.r-project.org/web/packages/ftsa/ftsa.pdf`.

[Kak1] Y. Kakihara, A classification of vector harmonisable processes. *Stoch. Anal. Appl.* **10** (1992), 277-311.

[Kak2] Y. Kakihara, Vector harmonizable processes: Wold and Cramér decompositions. *Stoch. Anal. Appl.* **13** (1995), 531-541.

[Kak3] Y. Kakihara, The Kolmogorov isomorphism theorem and extensions to some nonstationary processes. *Stochastic processes: Theory and methods, Handbook of Statistics* **19** (ed. D. N. Shanbhag and C. R. Rao), 443-470, North-Holland, 2001.

[Kak4] Y. Kakihara, Spectral domains of vector harmonizable processes. *J. Statistical Planning and Inference* **100** (2002), 93-108.

[KalM] G. Kallianpur and V. Mandrekar, Spectral theory of stationary H-valued processes. *J. Multiv. Analysis* **1** (1971), 1-16.

[KasB1] Y. Kasahara and N. H. Bingham, Verblunsky coefficients and Nehari sequences. *Trans. Amer. Math. Soc.* **366** (2014), 1363-1378.

[KasB2] Y. Kasahara and N. H. Bingham, Coefficient stripping in the matrix Nehari problem. *J. Approx. Th.* **220** (2017), 1-11.

[KasB3] Y. Kasahara and N. H. Bingham, Matricial Baxter's theorem with Nehari sequence. *Math. Nachrichten* **291** (2018), 2590-2598.

[Khr1] S. Khrushchev, Schur's algorithm, orthogonal polynomials, and convergence of Wall's continued fractions in $L^2(\mathbb{T})$. *J. Approx. Th.* **108** (2001), 161-248.

[Khr2] S. Khrushchev, Classification theorems for general orthogonal polynomials on the unit circle. *J. Approx. Th.* **116** (2002), 268-342.

[Khr3] S. Khrushchev, *Orthogonal polynomials and continued fractions, from Euler's point of view.* Encycl. Math. Appl. **122**, Cambridge University Press, 2008.

[KloR] L. Klotz and M. Riedel, Spectral representation and extrapolation of stationary random processes on linear spaces. *Prob. Math. Stat.* **21** (2001), 179-197.

[Lan] E. C. Lance, *Hilbert C^*-modules: A toolkit for operator algebraists.* London Math. Soc. Lect. Notes Ser. **210**, Cambridge Univ. Press, 1995.

[Lax] P. D. Lax, Translation invariant spaces. *Acta Math.* **101** (1959), 163-178.

[LiH] Y. Li and T. Hsing, Deciding the dimension of effective dimension reduction space for functional and high-dimensional data. *Ann. Stat.* **38** (2010), 3028-3062.

[Loe1] M. Loève, Fonctions aléatoires du second ordre. Appendix to P. Lévy, *Processus stochastiques et mouvement brownien*, Gauthier-Villars, 1948.

[Loe2] M. Loève, *Probability theory*, 4th ed. Grad. Texts in Math. **45** (Volume I), **46** (Volume II), Springer, 1977.

[Mac] G. W. Mackey, *Unitary group representations in physics, probability and number theory.* Benjamin, 1978.

[MakS] A. Makagon and H. Salehi, Notes on infinite dimensional stationary sequences. Prob. Th. on Vector Spaces IV, *Lecture Notes in Math.* **1391** (1989), 200-238.

[ManS] V. Mandrekar and H. Salehi, The square-integrability of operator-valued functions with respect to a non-negative operator-valued measure and Kolmogorov's isomorphism theorem. *Indiana U. Math. J.* **20** (1972), 545-563.

[MarR] Marcus, M. B. and Rosen, J., *Markov processes, Gaussian processes*

and local times. Cambridge University Press, 2006.

[MartR] R. A. Martinez-Avendano and P. Rosenthal, *An introduction to operators on the Hardy-Hilbert space.* Grad. Texts Math. **237**, Springer, 2007.

[Mas1] P. Masani, Orthogonally scattered measures. *Adv. Math.* **2** (1968), 61-117.

[Mas2] P. Masani, Comments on the prediction-theoretic papers. *Norbert Wiener: Coll. Works Vol. III* (ed. P. Masani), MIT Press, 1981, p.276-306.

[MiaS] A. G. Miamee and H. Salehi, The factorization problem for positive operator-valued functions in a Banach space and regularity of Banach space valued stationary stochastic processes. *Sanhyā A* **39** (1977), 211-222.

[Nik1] N. K. Nikolskii, *Treatise on the shift operator: Spectral function theory.* Grundl. Math. Wiss. **273**, Springer, 1986.

[Nik2] N. K. Nikolskii, *Operators, functions and systems: an easy reading. Volume 1: Hardy, Hankel and Toeplitz; Volume 2: Model operators and systems.* Math. Surveys and Monographs **92, 93**, Amer. Math. Soc., 2002.

[NikV] N. K. Nikolskii and V. I. Vasyunin, Notes on two function models. *The Bieberbach conjecture: Proceedings of the Symposium on the Occasion of the Proof*, 113-141. Math. Surv. Monog. **21**, Amer. Math. Soc., 1986.

[PanT1] V. Panaretos and S. Tavakoli, Fourier analysis of stationary time series in function space. *Ann. Stat.* **41** (2013), 568-603.

[PanT2] V. Panaretos and S. Tavakoli, Cramér-Karhunen-Loève representation and harmonic principal components analysis of functional time series. *Stoch. Proc. Appl.* **123** (2013), 2779-2807.

[Pay] R. Payen, Fonctions aléatoires du second ordre à valeurs dans un espace de Hilbert. *Ann. Inst. H. Poincaré Prob. Stat.* **3** (1967), 323-396.

[Pel] V. V. Peller, *Hankel operators and their applications.* Springer, 2003.

[Pon] L. S. Pontryagin, *Topological groups*, 2nd ed., Gordon and Breach, 1966.

[Pow] S. C. Power, Spectral characterization of the Wold-Zasuhin decomposition and prediction-error operator. *Math. Proc. Cambridge Phil. Soc.* **110** (1991), 559-567.

[RamS1] J. O. Ramsay and B. W. Silverman, *Functional data analysis*, 2nd ed., Springer, 2005 (1st ed. 1997).

[RamS2] J. O. Ramsay and B. W. Silverman, *Applied functional data analysis: Methods and case studies*, Springer, 2002.

[Rao1] M. M. Rao, Harmonizable processes: structure theory. *Enseignement Math.* **28** (1982), 295-351.

[Rao2] M. M. Rao, Harmonizable, Cramér, and Karhunen classes of process.

Handbook of statistics **5**: *Time series in the time domain* (ed. E. J. Hannan et al.), North-Holland, 1985, 279-310.

[RieN] F. Riesz and B. Sz.-Nagy, *Leçons d'analyse fonctionelle*, 2nd ed., Akadémiai Kiadó, 1953 (1st ed. 1952).

[Ros] M. Rosenblatt, Some purely deterministic processes. *J. Math. Mech.* **6** (1957), 801-810 (reprinted in *Selected works of Murray Rosenblatt* (ed. R. A. Davis et al.), Springer, 2011, 124-133).

[RosR] M. Rosenblum and J. Rovnyak, *Hardy classes and operator theory*. Oxford University Press, 1985; Dover, 1997.

[Rud1] W. Rudin, *Real and complex analysis*. McGraw-Hill, 3rd ed., 1987 (1st ed. 1966, 2nd ed. 1974).

[Rud2] W. Rudin, *Functional analysis*, 2nd ed., McGraw-Hill, 1991 (1st ed. 1973).

[Rus] A. F. Ruston, *Fredholm theory in Banach spaces*. Cambridge Tracts in Mathematics **86**, Cambridge University Press, 1986.

[Sil] B. W. Silverman, *Density estimation for statistics and data analysis*. Monogr. Stat. Appl. Prob. **26**, Chapman & Hall, 1986.

[Sim1] B. Simon, Notes on infinite determinants of Hilbert-space operators. *Adv. Math.* **24** (1977), 244-273.

[Sim2] B. Simon, *Trace ideals and their applications*, 2nd ed. AMS Surveys and Monographs **120**, Amer. Math. Soc., 2005 (1st ed. LMS Lecture Note Series **35**, CUP, 1979).

[Sim3] B. Simon, *Orthogonal polynomials on the unit circle. Part 1: Classical theory, Part 2: Spectral theory*. AMS Colloquium Publications **54.1**, **54.2**, American Math. Soc., Providence RI, 2005.

[Sim4] B. Simon, *Szegő's theorem and its descendants: Spectral theory for L^2 perturbations of orthogonal polynomials*. Princeton University Press, 2011.

[Smi] F. Smithies, *Integral equations*. Cambridge Tracts **49**, Cambridge University Press, 1965.

[Sto] M. H. Stone, *Linear transformations in Hilbert space and their applications to analysis*. Amer. Math. Soc. Colloq. Publ. **XV**, AMS, 1932.

[SzNF] B. Sz.-Nagy, C. Foiaş, H. Bercovici and L. Kérchy, *Harmonic analysis of operators on Hilbert space*, 2nd ed., Springer Universitext, 2010 (1st ed., Sz.-Nagy and Foiaş, North-Holland, 1970).

[Tre1] S. R. Treil, Angles between co-invariant subspaces and the operator corona problem. The Szökefalvi-Nagy problem. *Dokl. Akad. Nauk SSSR* **302** (1988), 1063-1068.

[Tre2] S. R. Treil, An operator corona theorem. *Indiana U. Math. J.* **53**

(2004), 1763-1780.

[Ver] R. Vershynin, *High-dimensional probability. An introduction with applications in data science.* Cambridge University Press, 2018.

[Wai] M. J. Wainwright, *High-dimensional statistics: A non-asymptotic viewpoint.* Cambridge University Press, 2019.

[Wer] A. Weron, Prediction theory in Banach spaces. Proc. Winter Sch. Prob. Karpacz. *Lecture Notes in Math.* **472** (1975), 207-228.

[Whi] P. Whittle, The analysis of multiple time series. *J. Roy. Stat. Soc. B* **15** (1953), 125-139.

Mathematics Department, Imperial College, London SW7 2EZ, UK;
n.bingham@ic.ac.uk

$* \cdots * \cdots * \cdots *$

NEW REFERENCES, Feb 2022 – for the convenience of the referee

[Bot1] A. Böttcher, One more proof of the Borodin-Okounkov formula for Toeplitz determinants. *Integral Equations and Operator Theory* **41** (2001), 123-125.

[Bot2] A. Böttcher, On the determinant formulas by Borodin, Okounkov, Baik, Deift and Rains. *Operator Theory Advances and Applications* **135** (2002), 91-99.

[Bot3] A. Böttcher, Borodin-Okounkov and Szegő for Toeplitz operators in model spaces. *Integral Equations and Operator Theory* **78** (2014), 407-414.

[BriC] T. Britz, A. Carey, F. Gesztesy, R. Nichols, F. Sukhochev and D. Zanin, The product formula for regularized Fredholm determinants. *Proc. Amer. Math. Soc. B* **8** (2021), 42-51.

[Con1] J. B. Conway, *A course in functional analysis*, 2nd ed. Grad. Texts in Math. **96**, Springer, 1990 (1st ed. 1985).

[Con2] J. B. Conway, *A course in operator theory*. Grad. Studies in Math. **21**, Amer. Math. Soc., 2000.

[HalPP] P. Hall, S. S. Poskitt and B. Presnell, A functional data analytic approach to signal detection. *Technometrics* **43** (2001), 1-9.

[Khr1] S. Khrushchev, Schur's algorithm, orthogonal polynomials, and convergence of Wall's continued fractions in $L^2(\mathbb{T})$. *J. Approx. Th.* **108** (2001), 161-248.

[Khr2] S. Khrushchev, Classification theorems for general orthogonal polynomials on the unit circle. *J. Approx. Th.* **116** (2002), 268-342.

[Khr3] S. Khrushchev, *Orthogonal polynomials and continued fractions, from Euler's point of view.* Encycl. Math. Appl. **122**, CUP, 2008.

[Loe2] M. Loève, *Probability theory*, 4th ed. Grad. Texts in Math. **45**, Volume I, **46**, Volume II, Springer, 1977.

[Rus] A. F. Ruston, *Fredholm theory in Banach spaces.* Cambridge Tracts in Mathematics **86**, Cambridge University Press, 1986.

[Sim1] B. Simon, Notes on infinite determinants of Hilbert-space operators. *Adv. Math.* **24** (1977), 244-273.

[Sim2] B. Simon, *Trace ideals and their applications*, 2nd ed. AMS Surveys and Monographs **120**, AMS., 2005 (1st ed. LMS LNS **35**, CUP, 1979).

[Smi] F. Smithies, *Integral equations.* Cambridge Tracts **49**, Cambridge University Press, 1965.