

MULTIVARIATE PREDICTION AND MATRIX SZEGÖ THEORY

N. H. BINGHAM

Abstract

Following the recent survey by the same author of Szegő's theorem and orthogonal polynomials on the unit circle (OPUC) in the scalar case, we survey the corresponding multivariate prediction theory and matrix OPUC (MOPUC).

AMS 2000 subject classifications. Primary 60G10, secondary 60G25

Key words and phrases Stationary process, vector-valued, multivariate prediction theory, multivariate orthogonal polynomials on the unit circle (MOPUC), Kolmogorov isomorphism theorem, Verblunsky's theorem, Szegő's theorem

CONTENTS

§0. Introduction	
§1. The Kolmogorov Isomorphism Theorem	
§2. Verblunsky's theorem	
§3. Szegő's theorem	
§4. Matrix spectral factorisations and matrix Szegő functions	
§5. The strong Szegő theorem	
§6. Baxter's inequality and Baxter's theorem	
§7. Nehari sequences and the Levinson-McKean condition	
§8. Pure minimality	
§9. Positive angle and the matrix muckenhoupt condition	
§10. Complete regularity	
§11. Hankel operators	
§12. Open questions	
References	

§0. Introduction

Prediction theory for discrete-time stationary stochastic processes, a topic in time series, has been long studied and has benefited greatly from the recent work on orthogonal polynomials on the unit circle (OPUC); see e.g. Simon's books [Si1], [Si2], [Si3] for background and references. Here the partial autocorrelation function (PACF) plays an important role, as it provides an unrestricted parametrization of the relevant spectral measure μ ; see e.g. the papers by Inoue [In1], [In2], [In3], Inoue and Kasahara [InKa1], [InKa2] and the survey by Bingham [Bi] for background and references. The PACF is essentially the sequence of Verblunsky coefficients, and the unrestricted parametrization is Verblunsky's theorem, in the language of OPUC; see [Si1], Ch. 1,2.

While this is very satisfactory for univariate time series, one often encounters multivariate time series, particularly in areas such as mathematical finance, where the dimensionality ℓ is the number of risky assets held in a portfolio; ℓ reflects the need to diversify one's portfolio (in the style of Markowitz), and may be large (see e.g. [BiFrKi] for a case in point). Multivariate time series form an important area (see e.g. the books by Hannan [Ha] and Reinsel [Re]). Likewise, multivariate prediction goes back to work by Wiener and Masani [WiMa1], [WiMa2], [WiMa3], Helson and Lowdenslager [HeLo] of 1957-61 (also Masani [Mas1] – [Mas5]). Just as in the univariate case with OPUC, in the multivariate case the matrix theory of OPUC – MOPUC for short below – is crucially relevant. Following the great stimulus to OPUC provided by Simon's books, MOPUC has recently been extensively developed. Our purpose here is to survey these developments with a view to multivariate prediction theory, providing a matrix sequel to [Bi].

1. The Kolmogorov Isomorphism Theorem

There is a process Y on T with orthogonal increments and a probability measure μ on T with

- (i) $X_n = \int e^{in\theta} dY(\theta)$;
- (ii) $E[dY(\theta)^2] = d\mu(\theta)$.
- (iii) The autocorrelation function γ then has the *spectral representation*
 $\gamma_n = \int e^{-in\theta} d\mu(\theta)$.
- (iv) One has the *Kolmogorov isomorphism* between \mathcal{H} (the *time domain*) and

$L_2(\mu)$ (the *frequency domain*) given by

$$X_t \leftrightarrow e^{it}, \quad (KIT)$$

for integer t (as time is discrete).

Parts (i) and (ii) are Cramér representation of 1942 (Doob X.4; Cramér and Leadbetter 7.5). Part (iii) (Herglotz, 1911) follows from (i) and (ii) (Doob; Brockwell & Davis 4.3). Part (iv) is due to Kolmogorov, 1941. This rests on Stone's theorem of 1932 ([Sto], [RiNa] §137, extended to semigroups in [HiPh] XXII: spectral representation of groups of unitary transformations of linear operators on Hilbert space); see Doob 636-7, Dunford & Schwartz X.5 for spectral theory. //

Write U for the shift $t \mapsto t + 1$. Since we can write $X_t = U^t X_0$, this gives us a spectral representation for $X = (X_t)$:

$$X_t = U^t X_0 = \int e^{i\theta t} E(d\theta) X_0$$

(see e.g. [Rozanov [Ro2], Th. I.4.2) – *time domain* on the left, *frequency domain* on the right.

§2. Verblunsky's theorem

In the univariate case, the Verblunsky coefficients $a = (a_n)_{n=1}^\infty$ satisfy $|a_n| < 1$, and the bijection

$$a \leftrightarrow \mu$$

of Verblunsky's theorem provides the unrestricted parametrization so useful in statistics and prediction theory; see e.g. Pourahmadi [Pou1], [Pou2], [Bi] §2 for background here.

In the ℓ -dimensional case, where one uses MOPUC rather than OPUC, the spectral measure μ is now $\ell \times \ell$ matrix-valued, and corresponds to the $\ell \times \ell$ covariance matrix by Herglotz's theorem. The Verblunsky coefficients have been studied by Damanik, Pushnitski and Simon [DaPuSi], §3. They show (3.10) that the Verblunsky coefficients are now $\ell \times \ell$ matrices on the unit circle, satisfying

$$\|a_n\| < 1,$$

that any sequence a of such matrices can arise in this way, and that the map $a \leftrightarrow \mu$ is again a bijection (Verblunsky's theorem for MOPUC – [DaPuSi],

Th. 3.12). Their method is Bernstein-Szegö approximation ([DaPuSi] §3.6; cf. [Si1] Th. 1.7.8 in the scalar case).

Results of this type go back in the statistical literature to Morf, Vieira and Kailath [MoViKa] in 1978. They showed that the a_n have singular values α_n with $|\alpha_n| \leq 1$.

The Szegö recursion that leads to OPUC is known in the time-series literature as the *Levinson-Durbin algorithm*. The Levinson-Durbin algorithm was extended to the multivariate case by Whittle [Wh] in 1963, and by Wiggins and Robinson [WigRo] in 1965. In [MoViKa], the authors use the results of [Wh] and [WigRo], which they call the LWR algorithm, and give a normalized version of it. They show that unless the Levinson-Durbin algorithm stops (corresponding to some correlation matrix failing to be positive-definite), $|\alpha_n| < 1$ for all n .

§3. Szegö's theorem

Derevyagin, Holtz, Khrushchev and Tyaglov [DeHoKhTy] continue this study, again using Bernstein-Szegö approximation (§§5,6). They show (§7) that (using \dagger for the adjoint (conjugate transpose), μ' for the density of (the absolutely continuous component of) μ , denoted w below)

$$\log \Pi_{n=1}^{\infty} \det(1 - a_n^{\dagger} a_n) = \int tr \log \mu' d\theta / 2\pi = \int tr \log w d\theta / 2\pi$$

for any non-trivial (i.e. of infinite support) matrix-valued probability measure on the unit circle. Call those for which the integral on the right is $> -\infty$ *Szegö measures*, and the finiteness of the integral *Szegö's condition*, (Sz). They deduce that μ is a Szegö measure iff

$$\sum \|a_n^{\dagger} a_n\| < \infty$$

(" $a \in L_2$ "). This is Szegö's theorem for MOPUC. They continue (§8) with the matrix version of the Helson-Lowdenslager theorem:

$$\exp \int \frac{1}{\ell} tr \log \mu' d\theta / 2\pi = \inf_{A,P} \int \frac{1}{\ell} tr [(A + P)^{\dagger} d\mu (A + P)],$$

where the infimum is over all matrices A of determinant 1 and all trigonometric polynomials $P(e^{i\theta}) = \sum_{k>0} A_k e^{ik\theta}$.

The Wold decomposition extends from the scalar to the vector case. See

e.g. Masani [Ma1], [Mas4], Hannan [Ha], and in the context of operator theory, Sz.-Nagy and Foias [SzNF], Ch. 1, Nikolskii [Nik] Ch. 1. As in the scalar case, (Sz) is the condition that the deterministic component in the Wold decomposition should vanish, leaving only the moving-average component. This is the condition of *non-determinism*, (ND) (" $(Sz) = (ND)$ ").

As in the scalar case, it is often useful to strengthen the Szegő condition by requiring explicitly that the singular component μ_s of the spectral measure μ should vanish. This is called the condition of *pure non-determinism*, (PND) :

$$(PND) = (ND) + \{\mu_s = 0\} = (Sz) + \{\mu_s = 0\}.$$

§4. Matrix spectral factorizations and matrix Szegő functions

Factorizations are already present in the scalar case. For an analytic function in the Hardy space on the disc, identify the boundary values of the function on the unit circle with the function itself, as usual; then the spectral density w and the Szegő function h are related by

$$w = h\bar{h} = |h|^2.$$

Here h is in the Hardy space H_2 , and is an outer function (see e.g. [Bi] for references and details); one can think of h as the ‘analytic square root’ of w .

In the matrix case, one speaks of the *spectral factorization problem*. The two main traditional approaches to multivariate prediction theory are those of Helson and Lowdenslager [HeLo], based on approximation by polynomials, and of Wiener and Masani [WiMas1] – [WiMas3], based on matrix factorization:

$$W = GG^\dagger, \quad (WM)$$

where again G is an outer function in the sense of matrix-valued Hardy spaces (see e.g. [Pel3], §13.3 and Appendix 2.3, or [RosRov], Ch. 4,5, for definitions and details; cf. Stein [Ste], Ch. III).¹ We refer to factorizations of the form (WM) as *Wiener-Masani factorizations*, and to G , the matrix analogue of the Szegő function h in the scalar case (Δ in the notation of [Si1]), as a (*matrix*) *Szegő function*. Wiener-Masani factorizations are unique up to unitary equivalence ([WiMas2], Th. 8.12).

This work is well summarized in Masani [Mas3]. Here one finds:

¹Wiener and Masani [WiMas3], Def. 3.5 use the term *optimal* in place of *outer*. Rozanov [Ro2], Th. II.4.2 uses the term *maximal*.

- (i) the Wold decomposition (§4; [WiMas1], Th. 7.11). In the scalar case, the deterministic (time-independent) component of X_t corresponds to the singular component μ_s of the spectral measure, the moving-average component to the absolutely continuous component μ_{ac} in the Lebesgue decomposition. In the matrix case, the terms Wold-Zasuhin decomposition and Lebesgue-Cramér decomposition are often used; the correspondence is exact in the full-rank case (below), but not in general [Mas1];
- (ii) the Kolmogorov Isomorphism Theorem, between the time domain and the spectral domain (§§6, 7);
- (iii) Wiener-Masani factorizations (WM) (Th. 9.7 – see also Rozanov [Ro1], [Ro2]);
- (iv) the matrix extension of Kolmogorov’s formula for the one-step prediction error (eq. (10.1), the main result of [WiMas1] (Th. 7.10); see also Whittle [Wh]);
- (v) convergence of the finite-past predictor to the infinite-past predictor (§13) – cf. Baxter’s inequality, §6 below).

To these, we also add

- (vi) Whittle’s multivariate extension of the Levinson-Durbin algorithm, mentioned in §2.

See also [Mas4] for extensive commentary on Wiener’s work in this area.

In the matrix case, one needs to discriminate between the *full-rank* and *degenerate-rank* cases – where the rank of the spectral density matrix W is full (ℓ) or degenerate ($m < \ell$). The degenerate-rank case is considered in [Ro2], [Mas3], §§11,12, [WiMas3] for $\ell = 2$, Matveev [Mat] in the general case; we refer there for details. The generic, and easier, case is the full-rank case, where Γ is positive-definite (the contrast between the full- and degenerate-rank cases is similar to that arising in, e.g., regression, where one encounters multi-collinearity; see e.g. [BiFr], §7.4.

This interesting work dates from the 1960s, before the work of Fefferman and Stein on BMO and of Sarason on VMO. Armed with these, Peller [Pel1] in 1990 considered matrix spectral factorizations

$$w = h^* h = h_{\sharp} h_{\sharp}^*$$

(recall h is determined to within unitary equivalence). He introduced the *phase function*

$$u = h_{\sharp}^* h^{-1},$$

the analogue of the phase function \bar{h}/h of §5 in the scalar case. He showed,

among other results, that the process is completely regular (§10 below) iff

$$u \in VMO.$$

Arov and Dym [ArDy3], §3.16] give matrix factorizations of positive definite functions into factors from the Nevanlinna class.

The simplest, and principal, case is that of a purely non-deterministic process of full rank. See e.g. several papers by Ephremidze, Janashia and Lagvilada [EpJaLa], [EpLa], [JaLaEp]. In particular, a (matrix) function in a Hardy space is outer iff its (scalar) determinant is outer.k,

More general than the matrix-valued case is the operator-valued case. This is important in operator theory, in non-commutative probability theory, non-commutative Hardy-space theory, non-commutative martingale inequalities etc. See e.g. [RosRov], esp. Ch. 6, Curtain and Zwart [CuZw], Barclay [Bar1], [Bar2], Peller [Pel1] – [Pel3].

§5. The strong Szegő theorem

The strong Szegő theorem, as presented in e.g. [Si1] Ch. 6, [Bi] §5, extends in full to the matrix case. For a short proof, see Böttcher [Bo1]; cf. [Bo2], [Bo3], Basor and Widom [BasWi]. A different approach has been given more recently by Chanzy [Cha1], [Cha2].

§6. Baxter's inequality and Baxter's theorem

Baxter used OPUC in a series of probabilistic papers of 1961-63 ([Bax1] – [Bax3]). His results concern, among other things, the weak and strong forms of Szegő's limit theorem (for Toeplitz determinants), finite and infinite Wiener-Hopf equations (in discrete time $n = 0, 1, 2, \dots$: finite with $\sum_{k=0}^n$, infinite with $\sum_{k=0}^{\infty}$), and the convergence of finite-predictor coefficients (in which one is given a finite section of the past of length n) to the corresponding infinite-predictor coefficients. This last depends on *Baxter's inequality* [Bax3]. Baxter's inequality was used by Simon [Si1], Ch. 5, in his proof that the Verblunsky coefficients $a \in \ell_1$ iff the correlation function $\gamma \in \ell_1$, the spectral measure μ is absolutely continuous, and its density $\mu' = w$ is continuous and positive. Simon calls this result *Baxter's theorem* (though Baxter did not formulate the result in this form, and 'the Baxter-Simon theorem' might be better here, but we will follow [Si1]). Perhaps because of this rather

involved history, and the fact that Simon's book [Si1] is still comparatively recent, there is as yet no matrix extension to Baxter's theorem. We raise here the question of obtaining one, and turn now to what is known.

Baxter's inequality and convergence of finite predictors in the matrix case were considered by Masani in 1966 ([Ma3] §13) and by Cheng and Pourahmadi [ChePou] in 1993. Theoretical progress in the area since then has been extensive, and the question arises of weakening the conditions that they impose. For more recent developments in the scalar case, see [InKa2].

Results on approximation by such finite-section operators have been given in great generality by Seidel and Silbermann [SeSi] (see §2.5.4), using Banach-algebra techniques (as did Baxter and Simon).

§7. Nehari sequences and the Levinson–McKean condition

Nehari's theorem of 1957 states that a Hankel operator is a bounded map from ℓ_2 on the natural numbers to itself iff the sequence generating it is the sequence of negative Fourier coefficients of a bounded function. See e.g. [Si1] Th. 6.2.17 ("The modern theory of Hankel operators started with the following result of Nehari"), Peller [Pel3]. Finding such a generating sequence is thus a type of *moment problem*, and as with other moment problems there may be no solution, a unique solution or more than one solution; the moment problem is then called insoluble, determinate or indeterminate. The indeterminate case is particularly important; the generating sequence is then called a *Nehari sequence*. This Nehari moment (or interpolation) problem was considered by Adamjan, Arov and Krein [AdArKr] in 1968; they described the solution set in terms of Sarason's concept of *rigidity* [Sa1]. Rigidity has previously been studied in this area in connection with the concept of *complete non-determinism* (CND); see Bloomfield, Jewell and Hayashi [BlJeHa]. It turns out that (CND) is equivalent to the *intersection of past and future* property (IPF) [IK2]. It has very recently been shown that both are equivalent to the *Levinson–McKean property* (the name comes from work of Levinson and McKean, [LeMcK] p. 105, in continuous time; the name is given by analogy in the discrete-time case). Phase functions (§4) play a crucial role here; see Kasahara and Bingham [KaBi] for details.

The question arises of matrix extensions of these results (work in progress). The matrix Nehari problem has been considered in detail by Arov and Dym [ArDy1], [ArDy2], [ArDy3] Ch. 4, 7, 10 ('strong regularity'). See also Dym's review of Peller's book on Hankel operators [Pel3] (MR1949210

(2004e:47040)). For matrix phase functions, see §4.

Related to this Nehari problem is the Schur (interpolation) problem, the matrix case of which is considered at book length in Dubovoj, Fritzsche and Kirstein [DuFrKi]; cf. [ArDy3], §7.6.

§8. Pure minimality

In the scalar case, pure minimality is characterized by ($\mu_s = 0$ and) Kolmogorov's condition $1/w \in L_1$, using w for the density of the spectral measure μ (now absolutely continuous). This result extends to the multivariate case; see Makagon and Weron [MaWe], [Pou3, Th. 8.10]. The spectral density is now a matrix W , and its inverse W^{-1} is integrable. It is thus natural that the condition

$$W^{-1} \in L_1$$

should be imposed in studying processes subject to stronger regularity conditions than pure minimality. We turn below to two such conditions – positive angle (§9) and complete regularity (§10). We regard the four conditions in §§7-10, which are in increasing order of strength, as *intermediate* conditions, being intermediate between the *weak* conditions (ND), (PND) (Szegö condition + $\mu_s = 0$) and the strong conditions (B), (sSz) (Baxter's condition and the strong Szegö condition). See e.g. [Bi] for the scalar case.

9. Positive angle and the matrix Muckenhoupt condition

The *Muckenhoupt condition* (A_2) of analysis is important in many areas; see e.g. [Bi] §6.2 for background and references. It occurs in connection with the *positive angle* condition, (PA), and the conditions of Helson and Szegö [HeSz] and of Helson and Sarason [HeSa]. Matrix versions of the Muckenhoupt condition are considered at length by Arov and Dym [ArDy1], [ArDy3]. For matrix versions of the Helson-Szegö condition, see Pourahmadi [Pou1].

Treil and Volberg [TrVo1] show that the following matrix Muckenhoupt condition is necessary and sufficient for the positive-angle condition (PA) in the multivariate case:

$$\sup_I \left\| \left(\frac{1}{|I|} \int_I W \right)^{1/2} \left(\frac{1}{|I|} \int_I W^{-1} \right)^{1/2} \right\| < \infty, \quad (A_2)$$

where the supremum is taken over all intervals I of the unit circle. Here the

condition that W be invertible a.e. (which corresponds to the ‘pure’ in pure minimality, see §8 above) has to be imposed explicitly, as noted by Peller in his review of [TrVo2] (MR1428818 (99k:42073)).

As shown in [HeSz], [HeSa], the condition (PA) (and so (A_2) by above) is equivalent to a condition on the sequence $\rho(n)$ of regularity coefficients of the form

$$\rho(\cdot) < 1.$$

We note that in his review of earlier work on this problem by Makagon, Miamee and Schröder [MaMiSc], Pourahmadi says ”Attempts to obtain a similar result for q -variate stationary sequences have been unyielding” (MR1443841 (98g:60074)).

10. Complete regularity

We turn now to a strengthening of the conditions of §9 above. The process is said to be *completely regular* if $\rho(n) \rightarrow 0$ as $n \rightarrow \infty$; see [IbRo], Ch. 4, 5. It was shown by Treil and Volberg [TV2] that complete regularity is equivalent to the following strengthening of the Muckenhoupt condition (A_2) :

$$\limsup_{|I| \rightarrow 0} \left\| \left(\frac{1}{|I|} \int_I W \right)^{1/2} \left(\frac{1}{|I|} \int_I W^{-1} \right)^{1/2} \right\| < \infty$$

(cf. Peller [Pel1], [Pel3]); here as before $W^{-1} \in L_1$ needs to be assumed explicitly. Note the form ($\rho(\cdot) \rightarrow 0$, $\limsup \dots = 1$) of the strengthenings here of the conditions ($\rho(\cdot) < 1$, $\sup \dots < \infty$) of §7 above.

§11. Hankel operators

Prediction theory has always involved Toeplitz operators (as in the book [GrSz] by Grenander and Szegö), and Toeplitz and Hankel operators have many links in operator theory. So it is natural that Hankel operators are useful in prediction theory. A monograph treatment of Hankel operators is given by Peller ([Pe3]; see also the review by Dym cited above in §7). Connections of Hankel operators with the matrix Muckenhoupt condition and with the matricial Nehari problem are considered by Arov and Dym in [ArDy3], Ch. 10, 11.

§12. Open questions

We mention two.

Question 1. Find the matrix version of Baxter's theorem.

As mentioned in §6, the matrix version of Baxter's inequality provides a good starting-point.

Question 2. Find the matrix version of [KaBi].

This hinges on solution of the matrix Nehari problem – the step

$$\Gamma \rightarrow H.$$

We hope to return to this elsewhere.

References

- [AdArKr] V. M. Adamjan, D. Z. Arov and M. G. Krein, Infinite Hankel matrices and generalized problems of Carathéodory–Fejér and I. Schur. *Functional Anal. Appl.* **2** (1968), 269-281.
- [ArDy1] D. Z. Arov and H. Dym, Matricial Nehari problems, J -inner functions and the Muckenhoupt condition. *J. Funct. Anal.* **181** (2001), 227-299.
- [ArDy2] D. Z. Arov and H. Dym, Criteria for the strong regularity of J -inner functions and γ -generating functions. *J. Math. Anal. Appl.* **280** (2003), 387-399.
- [ArDy3] D. Z. Arov and H. Dym, *J -contractive matrix valued functions and related topics*, Encycl. Math. Appl. 116, Cambridge Univ. Press, Cambridge, 2008.
- [Bar1] S. J. Barclay, Continuity of the spectral factorization mapping. *J. London Math. Soc.* **70** (2004), 763-779.
- [Bar2] S. J. Barclay, A solution of the Douglas-Rudin problem for matrix-valued functions. *Proc. London Math. Soc.* **99** (2009), 757-786.
- [BasWi] E. L. Basor and H. Widom, On a Toeplitz determinant identity of Borodin and Okounkov. *Integral Equations and Operator Theory* **37** (2000), 397-401.
- [Bax1] G. Baxter, A convergence equivalence related to polynomials orthogonal on the unit circle. *Trans. Amer. Math. Soc.* **99** (1961), 471-487.
- [Bax2] G. Baxter, An asymptotic result for the finite predictor. *Math. Scand.* **10** (1962), 137-144.

- [Bax3] G. Baxter, A norm inequality for a "finite-section" Wiener-Hopf equation. *Illinois J. Math.* **7** (1963), 97-103.
- [Bi] N. H. Bingham, Szegő's theorem and its probabilistic descendants, arXiv:1108.0368v1 [math.PR] 1 Aug 2011.
- [BiFr] N. H. Bingham and J. M. Fry, *Regression: Linear models in statistics*. SUMS (Springer Undergraduate Mathematics Series), 2010.
- [BiFrKi] N. H. Bingham, J. M. Fry and R. Kiesel, Multivariate elliptic processes. *Statistica Neerlandica* **64** (2010), 352-366.
- [BiInKa] N. H. Bingham, A. Inoue and Y. Kasahara, An explicit representation of Verblunsky coefficients. *Statistics and Probability Letters* **82.2** (2012), 403-410.
- [BlJeHa] P. Bloomfield, N. P. Jewell and E. Hayashi, Characterization of completely non-deterministic stochastic processes. *Pacific J. Math.* **107** (1983), 307-317.
- [Bo1] A. Böttcher, One more proof of the Borodin-Okounkov formula for Toeplitz determinants. *Integral Equations and Operator Theory* **41** (2001), 13 -125.
- [Bo2] A. Böttcher, Featured review of the Borodin-Okounkov and Basor-Widom papers. *Mathematical Reviews* 1790118/9 (2001g:47042a,b).
- [Bo3] A. Böttcher, On the determinant formulas by Borodin, Okounkov, Baik, Deift and Rains. *Operator Th. Adv. Appl.* 135, 91-99, Birkhäuser, Basel, 2002.
- [Cha1] J. Chanzy, Théorèmes-limite de Szëgo dans le cas matriciel. *Proc. Japan Acad. A* **82** (2006), 113-116.
- [Cha2] J. Chanzy, Inversion d'un opérateur de Toeplitz tronqué à symbole matriciel et théorèmes-limite de Szegő. *Ann. Math. Blaise Pascal* **13** (2006), 111-205.
- [ChePou] R. Cheng and M. Pourahmadi, Baxter's inequality and convergence of predictors of multivariate stochastic processes. *Probab. Th. Rel. Fields* **95** (1993), 115-124.
- [CuZw] R. F. Curtain and H. Zwart, *An introduction to infinite-dimensional linear systems*. Springer, 1995.
- [DaPuSi] D. Damanik, A. Pushnitski and B. Simon, The analytic theory of matrix orthogonal polynomials. *Surveys in Approximation Theory* **4** (2008), 1-85.
- [DeHoKhTy] M. Derevyagin, O. Holtz, S. Khrushchev and M. Tyaglov, Szegő's theorem for matrix orthogonal polynomials. arXiv:1104.4999v1 [math.CA] 26 April 2011.

- [Dev] A. Devinatz, The factorization of operator-valued functions. *Ann. Math.* **73** (1961), 458-495.
- [DuFrKi] V. K. Dubovoj, B. Fritzsche and B. Kirstein, *Matricial version of the classical Schur problem*. Teubner, Stuttgart, 1992.
- [EpJaLa] L. Ephremidze, G. Janashia and E. Lagvilava, An analytic proof of the matrix spectral factorization theorem. *Georgian Math. J.* **15** (2008), 241-249.
- [EpLa] L. Ephremidze and E. Lagvilada, Remark on outer analytic matrix functions. *Proc. A. Razmadze Math. Inst.* **152** (2010), 29-32.
- [GrSz] U. Grenander and G. Szegö, *Toeplitz forms and their applications*. Univ. California Press, Berkeley, CA, 1958.
- [Ha] E. J. Hannan, *Multiple time series*, Wiley, 1970
- [HeLo] H. Helson and D. Lowdenslager, Prediction theory and Fourier series in several variables, I, II. *Acta Math.* **99** (1958), 165-202, **106** (1961), 175-213.
- [HeSa] H. Helson and D. Sarason, Past and future. *Math. Scand* **21** (1967), 5-16.
- [HiPh] E. Hille and R. S. Phillips, *Functional analysis and semigroups*. Colloq. Publ. 31, American Math. Soc., 1957.
- [IbRo] I. A. Ibragimov and Yu. A. Rozanov, *Gaussian random processes*. Springer, 1978.
- [In1] A. Inoue, Asymptotics for the partial autocorrelation function of a stationary process. *J. Analyse Math.* **81** (2000), 65-109.
- [In2] A. Inoue, Asymptotic behaviour for partial autocorrelation functions of fractional ARIMA processes. *Ann. Appl. Prob.* **12** (2002), 1471-1491.
- [In3] A. Inoue, AR and MA presentations of partial autocorrelation functions with applications. *Prob. Th. Rel. Fields* **140** (2008), 523-551.
- [InKa1] A. Inoue and Y. Kasahara, Partial autocorrelation functions of the fractional ARIMA process. *J. Multivariate Analysis* **89** (2004), 135-147.
- [InKa2] A. Inoue and Y. Kasahara, Explicit representation of finite predictor coefficients and its applications. *Ann. Statist.* **34** (2006), 973-993.
- [JaLaEp] G. Janashia, E. Lagvilada and L. Ephremidze, A new method of matrix spectral factorization. *IEEE Trans. Info. Th.* **57** (2011), 2318-2326.
- [KaBi] Y. Kasahara and N. H. Bingham, Verblunsky coefficients and Nehari sequences. Preprint, Hokkaido University.
- [Kol] A. N. Kolmogorov, Stationary sequences in Hilbert space. *Bull. Moskov. Gos. Univ. Mat.* **2** (1941), 1-40 (in Russian; reprinted, *Selected works of A. N. Kolmogorov, Vol. 2: Theory of probability and mathematical statistics*,

Nauka, Moskva, 1986, 215-255).

[LeMcK] N. Levinson and H. P. McKean, Weighted trigonometrical approximation on R^1 with application to the germ field of a stationary Gaussian noise. *Acta Math.* **112** (1964), 99-143 (reprinted in *Selected papers of Norman Levinson* Vol. 2 (Birkhäuser, Basel, 1998), IX, 222-266).

[MakMiSc] A. Makagon, A. G. Miamee and B. S. W. Schröder, Recursive condition for positivity of the angle for multivariate stationary sequences. *Proc. Amer. Math. Soc.* **126** (1998), 1821-1825.

[MakWe] A. Makagon and A. Weron, q -variate minimal stationary processes. *Studia Math.* **59** (1976), 41-52.

S [Mas1] P. Masani, Cramér's theorem on monotone matrix functions and the Wold decomposition. *Probability and statistics: The Harald Cramér volume* (ed. U. Grenander) 175-189, Wiley, 1959.

[Mas2] P. Masani, The prediction theory of multivariate stochastic processes, III. *Acta Math.* **104** (1960), 141-162.

[Mas3] P. Masani, Shift-invariant spaces and prediction theory. *Acta Math.* **107** (1962), 275-290.

[Mas4] P. Masani, Recent trends in multivariate prediction theory. *Multivariate Analysis* (Proc. Int. Symp., Dayton OH) 351-382, Academic Press, 1966.

[Mas5] P. Masani, Comments on the prediction-theoretic papers, 276-305 in [Wi].

[Mat] R. F. Matveev, Regularity of multi-dimensional stochastic processes with discrete time. *Dokl. Akad. Nauk SSSR* **126** (1959), 713-715.

[MoViKa] M. Morf, A. Vieira and T. Kailath, Covariance characterization by partial autocorrelation matrices. *Ann. Statist.* **6** (1978), 643-678.

[Nik] 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.

[Pel1] V. V. Peller, Hankel operators and multivariate stochastic processes. *Proc. Symp. Pure Math.* 51, Part 1, 357-371, AMS, Providence, RI, 1990.

[Pel2] V. V. Peller, Factorization and approximation problems for matrix functions. *J. American Math. soc.* **11** (1998), 751-770.

[Pel3] V. V. Peller, *Hankel operators and their applications*, Springer, 2003.

[Pou1] M. Pourahmadi, A matricial extension of the Helson-Szegö theorem and its application in multivariate prediction. *J. Multivariate Analysis* **16** (1985), 265-275.

[Pou2] M. Pourahmadi, Joint mean-covariance models with applications to

- longitudinal data. Unconstrained parametrization. *Biometrika* **86** (1999), 677-690.
- [Pou3] M. Pourahmadi, *Foundations of time series analysis and prediction theory*. Wiley, 2001.
- [Re] G. C. Reinsel, *Elements of multivariate time series analysis*, Springer, 1997.
- [RiNa] F. Riesz and B. Sz.-Nagy, *Leçons d'analyse fonctionnelle*, 2nd ed., Akad. Kiadó, 1953.
- [RosRov] M. Rosenblum and J. Rovnyak, *Hardy classes and operator theory*, Dover, New York, 1997 (1st ed. Oxford University Press, 1985).
- [Ro1] Yu. A. Rozanov, Spectral properties of multivariate stationary processes and boundary properties of analytic functions. *Th. Probab. Appl.* **5** (1960), 362-376.
- [Ro2] Yu. A. Rozanov, *Stationary random processes*, Holden Day, San Francisco, CA, 1967.
- [Sa1] D. Sarason, An addendum to "Past and future". *Math. Scand.* **30** (1972), 62-64.
- [Sa2] D. Sarason, Functions of vanishing mean oscillation. *Trans. Amer. Math. Soc.* **207** (1975), 391-405.
- [Sa3] D. Sarason, *Function theory on the unit circle*. Virginia Polytechnic Institute and State University, Blacksburg, VA, 1979.
- [SeSi] M. Seidel and B. Silbermann, Banach algebras of structured matrix sequences. *Linear Algebra Appl.* **430** (2009), 1243-1281.
- [Si1] B. Simon, *Orthogonal polynomials on the unit circle. Part 1: Classical theory*. AMS Colloq. Publ. 54.1, AMS, Providence, RI, 2005.
- [Si2] B. Simon, *Orthogonal polynomials on the unit circle. Part 2: Spectral theory*. AMS Colloq. Publ. 54.2, AMS, Providence, RI, 2005.
- [Si3] B. Simon, *Szegő's theorem and its descendants. Spectral theory for L_2 perturbations of orthogonal polynomials*. Princeton Univ. Press, Princeton, NJ, 2011.
- [Ste] E. M. Stein, *Harmonic analysis*. Princeton University Press. 1993.
- [Sto] M. H. Stone, *Linear transformations in Hilbert space and their applications to analysis*. Coll. Publ. XV. American Math. Soc., 1932.
- [SzNF] B. Sz.-Nagy and C. Foias, *Harmonic analysis of operators on Hilbert space*, North-Holland, 1970 (2nd ed., with H. Bercovici and L. Kérchy, Springer Universitext, 2010).
- [TrVo1] S. Treil and A. Volberg, Wavelets and the angle between past and future. *J. Functional Analysis* **143** (1997), 269-308.

- [TrVo2] S. Treil and A. Volberg, Completely regular multivariate stochastic processes and the Muckenhoupt condition. *Pacific J. Math.* **190** (1999), 361-382.
- [Wh] P. Whittle, On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix. *Biometrika* **50** (1963), 129-134.
- [Wi] N. Wiener, *Collected works, Volume III* (ed. P. Masani), MIT Press, 1981.
- [WiAk] N. Wiener and E. J. Akutowicz, A factorization of positive Hermitian matrices, *J. Math. Mech.* **8** (1959), 111-120 (reprinted in [Wi] 264-273).
- [WiMa1] N. Wiener and P. Masani, The prediction theory of multivariate stochastic processes, I: The regularity condition. *Acta Math.* **98** (1957), 111-150 (reprinted in [Wi], 164-203).
- [WiMa2] N. Wiener and P. Masani, The prediction theory of multivariate stochastic processes, II: The linear predictor. *Acta Math.* **98 99** (1958), 93-137 (reprinted in [Wi] 164-203).
- [WiMa3] N. Wiener and P. Masani, On bivariate stationary processes and the factorization of matrix-valued functions. *Th. Prob. Appl.* **4** (1959), 300-308 (reprinted in [Wi] 255-263).
- [WigRo] R. A. Wiggins and E. A. Robinson, Recursive solution to the multichannel filtering problem. *J. Geophys. Res.* **70** (1965), 1885-1891.

Mathematics Department, Imperial College, London SW7 2AZ, UK
 n.bingham@ic.ac.uk nick.bingham@btinternet.com