

# **MULTIVARIATE PREDICTION AND MATRIX SZEGÖ THEORY**

**N. H. BINGHAM**

**WARWICK, 1 December 2011**

This talk arises from joint work with Akihiko  
INOUE (Hiroshima) and Yukio KASAHARA  
(Hokkaido).

## Sources

### *Surveys:*

[Bi] N. H. Bingham, Szegő's theorem and its probabilistic descendants, arXiv:1108.0368v1

[math.PR] 1 Aug 2011 (scalar case);

title as here (matrix case).

### *Orthogonal polynomials on the unit circle (OPUC):*

[Si1] B. Simon, Orthogonal polynomials on the unit circle. Part 1: Classical theory, 2005.

[Si2] B. Simon, Orthogonal polynomials on the unit circle. Part 2: Spectral theory, 2005.

[Si3] B. Simon, Szegő's theorem and its descendants, 2011.

### *Multiple time series:*

[HeLo] H. Helson & D. Lowdenslager, *Acta Math.*, 1958, 1961

[WiMa] N. Wiener & P. Masani, *Acta Math.*, 1957, 1958; *Th. Prob. Appl.*, 1959.

[Ro] Yu. A. Rozanov, *Stationary random processes*, 1967.

[Ha] E. J. Hannan, *Multiple time series*, 1970.

*Matrix OPUC (MOPUC):*

[DPS] D. Damanik, A. Pushnitski & B. Simon, The analytic theory of matrix orthogonal polynomials. *Surveys in Approximation Theory* **4** (2008), 1-85.

[DHKT] M. Derevyagin, O. Holtz, S. Khrushchev & M. Tyaglov, Szegő's theorem for matrix orthogonal polynomials. arXiv:1104.4999v1

[math.CA] 26 April 2011.

*Matrix Hardy classes:*

[N] N. K. NIKOLSKII, *Operators, functions and systems: An easy reading. Vol. 1: Hardy, Hankel and Toeplitz; Vol. 2: Model operators and systems*, AMS, 2002.

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

*Bingham, Inoue and Kasahara:*

[I] Inoue, J. *Analyse Math.* 2000; *Ann. Appl. Prob.*, 2002; *PTRF*, 2008;

[IK] Inoue and Kasahara, *J. Multiv. anal.* 2004; *Ann. Stat.* 2006;

[BIK], to appear, *Stat. Prob. Letters*; [KB], in preparation.

# 1. Kolmogorov Isomorphism Theorem

Stone's theorem: a group  $U = (U_t)$  of unitary transformations has a *spectral representation*

$$U_t = \int e^{i\theta t} E(d\theta),$$

with  $E(\cdot)$  a projection-valued random measure. For  $X = (X_t)$  a stationary  $\ell$ -dimensional time series,  $X_t = U^t X_0$ : spectral rep.

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

On the left – *time domain*; on the right – *frequency domain*. To summarize:

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

the *Kolmogorov Isomorphism Theorem* (Kolmogorov [Ko] in 1941). Hence the spectral representation of the correlation matrix:

$$\gamma_n := E[X_t^T X_{t+n}] = \int e^{in\theta} \mu(d\theta),$$

with  $\mu$  the *spectral measure*: *Herglotz's theorem*.

## §2. Verblunsky's theorem

For  $l = 1$ , the *partial autocorrelation coefficient* (PACF) is

$$a_n := \text{corr}(X_n - P_{[1,n-1]}X_n, X_0 - P_{[1,n-1]}X_0),$$

the correlation between the residuals at times 0,  $n$  resulting from (linear) regression on the intermediate values  $X_1, \dots, X_{n-1}$ . Following Simon [Si1] on OPUC, call the PACF the *Verblunsky coefficients*. Then  $|a_n| < 1$ , and the bijection

$$a \leftrightarrow \mu$$

is *Verblunsky's theorem*. It gives an unrestricted parametrization – very useful in statistics and

prediction theory; see e.g. [Bi] §2.

In the  $\ell$ -dimensional case, using MOPUC rather than OPUC, the spectral measure  $\mu$  is now  $\ell \times \ell$  matrix-valued, and corresponds to the  $\ell \times \ell$  covariance matrix by Herglotz's theorem. Damanik, Pushnitski and Simon [DPS] show 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, and again  $a \leftrightarrow \mu$  is a bijection (Verblunsky's theorem for MOPUC). They use Bernstein-Szegö approximation (cf.[Si1] Th. 1.7.8 in the scalar case, Morf, Vieira and Kailath [MVK] in 1978). The Szegö recursion that leads to OPUC is known in the time-series literature as the *Levinson-Durbin algorithm*, extended to the multivariate case by Whittle [Wh] in 1963.

### §3. Szegő's theorem

Derevyagin, Holtz, Khrushchev and Tyaglov [DHKT], again using Bernstein-Szegő approximation, show that (using  $\dagger$  for the adjoint (conjugate transpose),  $\mu' = w$  for the density of (the abs. continuous component of)  $\mu$ )

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

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  $> -\infty$  *Szegő measures*, and the finiteness of the integral *Szegő's condition*, (*Sz*). They deduce that  $\mu$  is a Szegő measure iff

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

("  $a \in L_2$  "). This is Szegő's theorem for MOPUC. The Wold decomposition extends from the scalar to the matrix case; see e.g. Hannan [Ha], III.

As in the scalar case, one has two components in general, one a moving average, one ‘deterministic’ (random but time-independent). Call the process *non-deterministic* (ND) if the first (‘nice’) component is present (‘some nice’), *purely non-deterministic* (PND) if the second (‘nasty’) component is absent (‘all nice’). In the scalar case, these components correspond to the Lebesgue decomposition  $\mu = \mu_{ac} + \mu_s$ . In the matrix case, the Wold(-Zasuhin) decomposition corresponds to the Lebesgue(-Cramér) decomposition in the full-rank case, but not in general (NASC involves the ‘defect’: the number of superoptimal singular values).

#### **§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; think of  $h$  as the ‘analytic square root’ of  $w$ .

In the matrix case, this gives the *spectral factorization problem*. Wiener and Masani [WM] give the matrix factorization:

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

where again  $G$  is an outer function in the sense of matrix-valued Hardy spaces.

Masani (1966): (i) the Wold(-Zasuhin) decomposition (§4; [WiMas1], Th. 7.11);

(ii) the Kolmogorov Isomorphism Theorem, between the time and spectral domains (§§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); note also

(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.

The matrix case splits between the *full-rank* (rank  $\ell$ ) and *degenerate-rank* (rank  $m < \ell$ ).

Degenerate-rank case: [Ro2], [Mas3], §§11,12, [WiMas3] for  $\ell = 2$ , Matveev in 1959 [Mat] in the general case. Full-rank case (generic, and easier):  $\Gamma$  is positive-definite (cf. regression and multi-collinearity; see e.g. [BiFr], §7.4).

All this is 1950s/60s, pre Fefferman-Stein on

BMO and pre Sarason on VMO. Peller [Pel1] in 1990 considered matrix spectral factorizations

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

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

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

(extending  $\bar{h}/h$  in the scalar case – [Bi], [KaBi]). He showed 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 PND process of full rank. Here a Hardy matrix function is outer iff its determinant is outer ([N]; cf. several papers by the Georgian school (Ephremidze, Janashia and Lagvilada)).

Operator-valued case: non-commutative probability theory, non-commutative Hardy-space theory, non-commutative martingale inequalities etc. See e.g. [RoRo], esp. Ch. 6, Curtain and Zwart [CuZw], Barclay [Bar1], [Bar2], Mei [Mei], 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]), on 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 (given a finite section of the past of length  $n$ ) to the corresponding infinite-predictor coefficients. This depends on *Baxter's inequality* [Bax3], used by Simon [Si1], Ch. 5, in his proof of 'Baxter's theorem' – the Verblunsky coefficients  $a \in \ell_1$  iff the correlation function  $\gamma \in \ell_1$ , the spectral measure  $\mu$  is absolutely continuous, and its density  $\mu' = w$  is continuous and positive.

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 [ChPo] in 1993. For recent developments in the scalar case, see [InKa2]. Approximation by such finite-section operators: 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 (1957): a Hankel operator is a bounded map from  $\ell_2$  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. Finding such a generating sequence is thus a type of *moment problem* (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 and *complete non-determinism* (CND); Bloomfield, Jewell and Hayashi [BlJeHa]. It turns out that (CND) is equivalent to the *intersection of past and future* property (IPF) [IK2]. Forthcoming work [KaBi]: both are equivalent to the *Levinson–McKean property* ([LeMcK] p. 105, in continuous time). Phase functions are crucial here.

Matrix extensions: matrix Nehari problem, Arov and Dym [ArDy1], [ArDy2], [ArDy3] Ch. 4, 7, 10 ('strong regularity').

Related is the Schur (interpolation) problem; matrix case Dubovoj, Fritzsche and Kirstein [DuFrKi]; cf. [ArDy3], §7.6.

## §8. Pure minimality

Scalar case: pure minimality is ( $\mu_s = 0$  and) Kolmogorov's condition  $1/w \in L_1$ . Matrix case: Makagon and Weron [MaWe], [Pou3, Th. 8.10].;

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

We turn now to two stronger conditions – positive angle (§9) and complete regularity (§10). The four conditions in §§7-10, in increasing order of strength, are *intermediate* conditions, between the *weak* conditions (ND), (PND) and the strong conditions (B), (sSz) (Baxter's condition and the strong Szegö condition) – the 'Goldilocks principle'; cf. [Bi].



## 9. Positive angle and the matrix Muckenhoupt condition

*Muckenhoupt condition* ( $A_2$ ) of analysis (see e.g. [Bi] §6.2) – 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: 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:

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

(sup over all intervals  $I$  of the unit circle).

As in [HeSz], [HeSa], ( $PA$ ) (and so also ( $A_2$ ))

is equivalent to a condition on the sequence  $\rho(n)$  of regularity coefficients:

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

## 10. Complete regularity

Strengthening §9, call process *completely regular* if  $\rho(n) \rightarrow 0$  as  $n \rightarrow \infty$ ; see [IbRo], Ch. 4, 5. Treil and Volberg [TV2]: complete regularity is equivalent to the following strengthening of the Muckenhoupt condition ( $A_2$ ):

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

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 Grenander-Szegö [GrSz]), and Toeplitz and Hankel operators have many links in operator theory. So also Hankel operators (for which see Peller [P]) are useful in prediction theory. For connections of Hankel operators with the matrix Muckenhoupt condition and the matricial Nehari problem see Arov and Dym in [ArDy3], Ch. 10, 11.

## §12. Open questions

Q1. Matrix version of Baxter's theorem. As in §6, the matrix version of Baxter's inequality provides a good starting-point.

Q2. Matrix version of [KaBi]. This hinges on solution of the matrix Nehari problem – the step  $\Gamma \rightarrow H$ . Work in progress.