Contributions of Issai Schur to Analysis

Harry Dym and Victor Katsnelson[*]

November 1, 2018

The name Schur is associated with many terms and concepts that are widely used in a number of diverse fields of mathematics and engineering. This survey article focuses on Schur's work in analysis. Here too, Schur's name is commonplace: The Schur test and Schur-Hadamard multipliers (in the study of estimates for Hermitian forms), Schur convexity, Schur complements, Schur's results in summation theory for sequences (in particular, the fundamental Kojima-Schur theorem), the Schur-Cohn test, the Schur algorithm, Schur parameters and the Schur interpolation problem for functions that are holomorphic and bounded by one in the unit disk. In this survey, we shall discuss all of the above mentioned topics and then some, as well as some of the generalizations that they inspired. There are nine sections of text, each of which is devoted to a separate theme based on Schur's work. Each of these sections has an independent bibliography. There is very little overlap. A tenth section presents a list of the papers of Schur that focus on topics that are commonly considered to be analysis. We shall begin with a review of Schur's less familiar papers on the theory of commuting differential operators.

Acknowledgement: The authors extend their thanks to Bernd Kirstein for carefully reading the manuscript and spotting a number of misprints.

1. Permutable differential operators and fractional powers of differential operators.

Let

$$P(y) = p_n(x)\frac{d^n y}{dx^n} + p_{n-1}(x)\frac{d^{n-1}y}{dx^{n-1}} + \cdots + p_0(x)y \tag{1.1}$$

and

$$Q(y) = q_m(x)\frac{d^m y}{dx^m} + q_{m-1}(x)\frac{d^{m-1}y}{dx^{m-1}} + \cdots + q_0(x)y, \tag{1.2}$$

---

be formal differential operators, where $n \geq 0$ and $m \geq 0$ are integers, and $p_k(x)$ and $q_k(x)$ are complex valued functions. Then Q commutes with P if $(PQ)(y) = (QP)(y)$. (It is assumed that the coefficients p_k, q_l are are smooth enough, say infinitely differentiable, so that the product of the two differential expressions is defined according to the usual rule for differentiating a product. The commutativity $PQ - QP = 0$ means that the appropriate differential expressions, that are constructed from the coefficients p_k, q_l according to the usual rules for differentiating a product, vanish.)

In [Sch1] Schur proved the following result: *Let P, Q_1 and Q_2 be differential operators of the form (1.1) and (1.2). Assume that each of the operators Q_1 and Q_2 commutes with P: $PQ_1 = Q_1P$ and $PQ_2 = Q_2P$. Then the operators Q_1 and Q_2 commute with each other: $Q_1Q_2 = Q_2Q_1$.*

This result of Schur was forgotten and was rediscovered by S. Amitsur ([Ami], Theorem 1) and by I.M. Krichever ([Kri1], Corollary 1 of Theorem 1.2). (Amitsur does not mention the result of Schur, and Krichever does not mention either the result of Schur, or the result of Amitsur in [Kri1], but does refer to Amitsur in a subsequent paper [Kri2].

The method used by Schur to obtain this result is not less interesting than the result itself. In modern language, Schur developed the calculus of formal pseudodifferential operators in [Sch1]: for every integer n (positive, negative or zero), Schur considers the formal differential "Laurent" series of the form

$$F = \sum_{\substack{-\infty < k \leq n \\ k \text{ an integer}}} f_k(x) \, D^k, \tag{1.3}$$

where the coefficients $f_k(x)$, $-\infty < k \leq n$, are smooth complex-valued functions of x and $D = \frac{d}{dx}$. (He does not discuss the existence of an operator in a space of functions that corresponds to this formal series.) The sum of two formal "Laurent" series and the product of such a series and a complex constant are defined in the usual way. To define the product $F \circ G$ of two such series F and

$$G = \sum_{\substack{-\infty < l \leq m \\ l \text{ an integer}}} g_l(x) \, D^l, \tag{1.4}$$

one needs a rule for commuting powers of the operator D with powers of the operator of multiplication by the function $a(x)$. This rule is defined by the formulas

$$Da = a(x)D + a'(x)I,$$

and

$$D^{-1}a = a(x)D^{-1} - a'(x)D^{-2} + a''(x)D^{-3} + \cdots + (-1)^{k-1}a^{(k-1)}(x)D^{-k} + \cdots,$$

where $a'(x), a''(x), \ldots a^{(k-1)}(x), \ldots$ are the derivatives of the function $a(x)$ of the indicated order. The set of all formal differential "Laurent" series provided with such

operations becomes an associative (but not commutative) ring over the field of complex numbers. If the function $f_n(x)$ is invertible (in which case we can and will assume that $f_n(x) \equiv 1$), then the formal Laurent series (1.3) is invertible, and its inverse is of the form

$$H = \sum_{-\infty < l \le -n} h_l(x) D^l, \tag{1.5}$$

where $h_{-n}(x) = 1$, and the coefficients $h_k(x)$ are polynomials in the functions $f_k(x)$, $k < n$, and their derivatives.

In particular, a differential operator P of the form (1.1) may be considered as a formal "Laurent" series (1.3) whose "positive" part $F_+ = \sum_{0 \le k \le n} f_k(x) D^k$ coincides with P and whose "negative" part $F_- = \sum_{-\infty < k < 0} f_k(x) D^k$ vanishes. In [Sch1], Schur proved that *if each of two formal differential Laurent series F_1 and F_2 commutes with a differential operator P of the form* (1.1): $P \circ F_1 = F_1 \circ P$ and $P \circ F_2 = F_2 \circ P$, *then F_1 and F_2 commute with each other*: $F_1 \circ F_2 = F_2 \circ F_1$. In particular, this result is applicable to polynomial differential operators Q_1 and Q_2 of the form (1.2) commuting with P. (Q_1 and Q_2 are considered as differential formal Laurent series whose "negative" parts are equal to zero.) Schur gives an explicit description of the commutant of the differential operator P (and, even more generally, the description of the commutant of any formal differential Laurent series). The notion of **the fractional power $P^{1/n}$ of the differential operator** P is involved in this description.

Let $n \ge 0$ and let F be a formal differential Laurent series of the form (1.3). The formal differential Laurent series $F^{1/n}$ is defined as the formal differential series R for which the equality

$$\underbrace{R \circ R \circ \cdots \circ R}_{n \text{ times}} = F \tag{1.6}$$

holds. In [Sch1] it is proved that if the function $(f_n(x))^{1/n}$ exists (in which case we can and will assume that $f_n(x) \equiv 1$), then such a series $R = F^{1/n}$ exists and is of the form

$$R = \sum_{\substack{-\infty < \rho \le 1 \\ \rho \text{ an integer}}} r_\rho(x) D^\rho, \tag{1.7}$$

where $r_1(x) \equiv 1$. The coefficients $r_0(x)$, $r_{-1}(x), r_{-2}(x)$, ... can be determined in a recursive manner as polynomials of the functions $f_{n-1}(x), f_{n-2}(x)$, ..., $f_0(x)$, ... and their derivatives. The differential Laurent series $F^{k/n}$ (k an integer) is defined as $F^{k/n} \stackrel{\text{def}}{=} (F^{1/n})^k$.

For example, if

$$L = D^2 + q(x)I \tag{1.8}$$

is a Sturm-Liouville differential operator of second order, then

$$L^{1/2} = D + s_0(x)I + s_{-1}(x)D^{-1} + s_{-2}(x)D^{-2} + s_{-3}(x)D^{-3} + s_{-4}(x)D^{-4} + \cdots , \tag{1.9}$$

3

where

$$s_0(x) = 0, \ s_{-1}(x) = \tfrac{1}{2}\,q(x), \ s_{-2}(x) = -\tfrac{1}{4}\,q', \ s_{-3}(x) = \tfrac{1}{8}\,q''(x) - \tfrac{1}{8}\,q^2(x),$$
$$s_{-4}(x) = -\tfrac{1}{16}\,q'''(x) + \tfrac{3}{8}\,q(x)q'(x), \ \ldots \ . \tag{1.10}$$

Furthermore, $L^{3/2} = (L^{1/2})^3 = L \cdot L^{1/2} = L^{1/2} \cdot L$, and we can calculate

$$L^{3/2} = D^3 + t_2(x)D^2 + t_1(x)D + t_0(x)I + t_{-1}(x)D^{-1} + \cdots, \tag{1.11}$$

where

$$t_2(x) = 0, \ t_1(x) = \frac{3}{2}\,q(x), \ t_0(x) = \frac{3}{4}\,q'(x), \ t_{-1}(x) = \frac{1}{8}\,q''(x) + \frac{3}{8}\,q^2(x), \ \ldots \tag{1.12}$$

In [Sch1] it is proved that *the formal differential Laurent series F commutes with a differential operator P of the form (1.1) (of order n) if and only if F is of the form*

$$F = \sum_{-\infty < k \leq n} c_k P^{k/n}, \tag{1.13}$$

where c_k are complex **constants** *(c_k do not depend on x).* In particular, some series of the form (1.13) can in fact be differential operators (if by some *very special* choice of the c_k the negative part F_- of the series (1.13) vanishes). These and only these differential operators commute with P. Moreover, it is clear that series of the form (1.13) commute with each other.

The results of Schur on fractional powers of differential operators were forgotten. The resurgence of interest in this topic is related to the inverse scattering method for solving non-linear evolution equations. The inverse scattering method was discovered and applied to the Korteweg - de Vries equation by C. Gardner, J. Green, M. Kruskal and R. Miura in their famous paper [GGKM]. This method was then extended to some other important equations towards the end of the sixties. P. Lax [Lax] developed some machinery (that is now commonly known as the method of L - A pairs, or Lax pairs) that allows one to use the inverse scattering formalism in a more organized way. The first step of the Lax method is to express the given evolution equation in the form

$$\frac{\partial L}{\partial t} = [A, \, L], \tag{1.14}$$

where L is a differential operator (with respect to x), some of whose coefficients depends on t, A is a differential operator with respect to x that does not depend on t and (the commutator) $[A, \, L] = AL - LA$. In subsequent developments, the evolution equation (1.14) was investigated using various analytic methods drawn from the theory of inverse spectral and scattering problems and the Riemann-Hilbert problem, among others. In their article [GeDi], I.M. Gel'fand and L.A. Dikii (=L. Dickey) observed that fractional

powers of differential operators can help in a systematic search for pairs L and A whose commutator $[A, L]$ is related to a nonlinear evolution equation. The idea of Gel'fand and Dikii is to consider the "positive" part $(L^\alpha)_+$ of some fractional power L^α as such an operator A. Let us explain how the fractional powers of the Sturm-Liouville operator L of the form (1.8) can be applied to construct the L-A pair for the Korteveg - de Vries equation. Since an operator L of the form (1.8) is of second order, it suffices to consider only integer and half-integer powers of L. Integer powers do not lead to anything useful: the appropriate A just commutes with L. Half-integer powers are more interesting. According to (1.9)-(1.10), $(L^{1/2})_+ = D$. The direct computation of the commutator gives: $[A, L] = q'I$ for $A = D$. The evolution equation (1.14) is of the form $\frac{\partial q}{\partial t} = \frac{\partial q}{\partial x}$ in this case. The case $A = (L^{3/2})_+$ is much more interesting. From (1.11)-(1.12) it follows that

$$A = D^3 + \frac{3}{2} q(x)D + \frac{3}{4} q'(x)I \,. \tag{1.15}$$

The direct calculation of the commutator of the differential expressions A and L of the forms (1.15) and (1.8), respectively, gives

$$[A, L] = \frac{1}{4} q'''(x) + \frac{3}{2} q(x) \, q'(x). \tag{1.16}$$

Thus, the evolution equation (1.14) takes the form

$$\frac{\partial q}{\partial t} = \frac{1}{4} \frac{\partial^3 q}{\partial x^3} + \frac{3}{2} q \frac{\partial q}{\partial x}. \tag{1.17}$$

This is the Korteweg - de Vries equation. In the paper [GeDi] a symplectic structure was introduced and a Hamiltonian formalism was developed. The approach of Gel'fand and Dikii was further developed by M. Adler [Adl] and by B.M. Lebedev and Yu.I. Manin [LebMa]. However, the results of Schur on permutable differential expressions and on fractional powers of differential expressions are not mentioned either in [Adl], or in [LebMa], nor are they mentioned in the well-known surveys [Man], [Tsuj], dedicated to algebraic aspects of non-linear differential equations. The fact that these results of Schur were largely forgotten may be due to the lack of a natural area of application for a long time. We found only one modern source where this aspect of Schur's work is mentioned: *Tata Lectures* by D. Mumford. Mumford cites the paper [Sch1] in Chapter IIIa, §11 of [Mum] (Proposition 11.7).

The paper [Sch1] does not discuss the structure of the set of differential expressions which commute with a given operator P. The answer "the differential expressions which commute with P are those formal Laurent series in $P^{1/n}$ for which "negative part" vanishes is not satisfactory because it just replaces the original question by the question "what is the structure of formal Laurent series in $P^{1/n}$ for which "negative part" vanishes. Of course, if P is a given differential operator and b is a polynomial with constant coefficients then the differential operator $Q = b(P)$ commutes with P. More generally, if Z is any differential operator and a and b are polynomials with constant coefficients then the

operators $P = a(Z)$ and $Q = b(Z)$ commute each with other. However there exist pairs of commuting differential operators P, Q which are not representable in the form $P = a(Z)$, $Q = b(Z)$. (See formula (1) in [BuCh1].) The problem of describing pairs of commuting differential operators was essentially solved by J.L. Burchnall and T.W. Chaundy [BuCh1], [BuCh2], [BuCh3] in the twenties. See also [Bak1]. (The complete answer was obtained for those pairs P, Q whose orders are coprime.) The answer was expressed in terms of Abelian functions. In particular, it was proved that the commuting pair P, Q satisfy the equation

$$r(P, Q) = 0 \,, \tag{1.18}$$

where $r(\lambda, \mu)$ is a (non-zero) polynomial of two variables with constant coefficients. (This result is known as the Burchnall-Chaundy lemma.) The remarkable papers [BuCh1], [BuCh2], [BuCh3], [Bak1] were forgotten. Their results were rediscovered and further developed by I.M. Krichever, [Kri1], [Kri2], [Kri3], [Kri4] in the seventies. (When Krichever started his investigations in this direction, he was not aware of the results of Burchnall and Chaundy. In his paper [Kri1] he mentioned only the relevant recent works of a group of Moscow mathematicians. However, in his subsequent papers he referred to [BuCh1], [BuCh2], [BuCh3] and [Bak1]; see the "Note in Proof" at the end of [Kri2] and references [2-4] in [Kri3].)

Thus, the history of commuting differential expressions, which began with the work of Schur [Sch1], is rich in forgotten and rediscovered results.

REFERENCES

[Adl]   ADLER, M.: *On a trace functional for formal pseudo-differential operators and the symplectic structure of the Korteveg-Devries equations.* Inventiones Math., **50** (1979), pp. 219 - 248.

[Ami]   AMITSUR, S.A. Commutative linear differential operators. Pacif. Journ. of Math. **8** (1958), pp. 1 - 10.

[Bak1]   BAKER, H.F. Note on the foregoing paper "Commutative ordinary differential operators". Proc. Royal Soc. London, **118** (1928), pp. 584 - 593.

[Bak2]   BAKER, H.F. *Abelian Functions.* (Cambridge Mathematical Library). Cambridge Univ. Press, Cambridge 1995.

[BuCh1]   BURCHNALL, J.L. and T.W. CHAUNDY. *Commutative ordinary differential operators.* Proc. London Math. Soc. (Second Ser.), **21** (1922), pp. 420 - 440.

[BuCh2]   BURCHNALL, J.L. and T.W. CHAUNDY. *Commutative Ordinary Differential Operators.* Proc. Royal Soc. London, **118** (1928), pp. 557 - 583.

[BuCh3]   BURCHNALL, J.L. and T.W. CHAUNDY. *Commutative Ordinary Differential Operators. II.— The Identity $P^n = Q^m$.* Proc. Royal Soc. London, **134** (1932), pp. 471 - 485.

[GGKM] GARDNER, C., J. GREEN, M. KRUSKAL, R. MIURA. *A method for solving the Korteweg — de Vries equation.* Phys. Rev. Lett., **19** (1967), pp. 1095 - 1098.

[GeDi] GEL'FAND, I.M. and L.A. DIKII (=L.A. DICKEY). *Drobnye stepeni operatorov i Hamil'tonovy sistemy.* Funkc. Anal. i ego Prilozh., **10**:4, pp. 13 - 29. (Russian). English transl.: *Fractional powers of operators and Hamiltonian systems.* Funct. Anal. and Appl. **10**:4 (1976), pp. 259-273.

[Kri1] KRICHEVER, I.M. *Integrirovanie nelineǐnykh uravneniǐ metodami algebraicheskoǐ geometrii.* Funk. Analiz i ego prilozh., **11**:1 (1977), pp. 15 - 31 (Russian). English transl.: *Integration of nonlinear equations by the methods of algebraic geometry.* Funct. Anal. and Appl. **11**:1 (1977), pp. 12-26.

[Kri2] KRICHEVER, I.M. *Metody algebraicheskoǐ geometrii v teorii nelineǐnykh uravneniǐ* (Russian). Uspeki Matem. Nauk, **32**:6 (1977), pp. 183 - 208. English transl.: *Methods of algebraic geometry in the theory of non-linear equations.* Russ. Math. Surveys, **32**:6 (1977), pp. 185-213.

[Kri3] KRICHEVER, I.M. *Kommutativnye kol'tsa obyknovennykh differentsial'nykh operatorov.* Funkts. Anal. i ego Prilozh., **12**:3 (1978), pp. 21 - 30. (Russian). English transl.: *Commutative rings of ordinary linear differential operators.* Funct. Anal. and Appl. **12**:3 (1977), pp. 175 – 185.

[Kri4] KRICHEVER, I.M. Foreword to the monograph [Bak2], Reedition of 1995, pp. xvii - xxxi in [Bak2].

[Lax] LAX, P. *Integrals of nonlinear equations of evolution and solitary waves.* Comm. Pure Appl. Math. **21** (1968), pp. 467–490.

[LebMa] LEBEDEV B.M. and YU.I. MANIN. *Gamil'tonov operator Gel'fanda-Dikogo i koprisoedinennoe predstavlenie gruppy Vol'terra.* Funkts. Anal. i ego Prilozh. **13**:4, pp. 40-46 (1979) (Russian). English transl.: *Gel'fand-Dikii Hamiltonian operator and the coadjoint representation of the Volterra group,* **13**:4 (1979), pp. 268 - 273.

[Man] MANIN, YU.I. *Algebraicheskie aspecty nelineǐnykh differentsial'nykh uravneniǐ.* Itogi Nauki i Tekhniki (Sovremennye Problemy Matematiki, **11**), VINITI, Moscow 1978, pp. 5 - 152 (Russian). English transl.: *Algebraic aspects of nonlinear differential equations.* Journ. of Soviet. Math. **11**, (1979), pp. 1 - 122.

[Mum] MUMFORD, D. *Tata Lectures on Theta. Vol. II.* (Progress in Mathematics, Vol. **43**) Birkhäuser, Boston·Basel·Stuttgart 1984.

[Sch:Ges] SCHUR, I.: *Gesammelte Abhandlungen [Collected Works]. Vol. I, II, III.* Springer-Verlag, Berlin·Heidelberg·New York 1973.

[Sch1] SCHUR, I. *Über vertauschbare lineare Differentialausdrücke* [On permutable differential expressions - in German]. Sitzungsberichte der Berliner Mathematischen Gesellschaft, **4** (1905), pp. 2 - 8. Reprinted in: [Sch:Ges], Vol. I, pp. 170 - 176.

[Tsuj]   TSUJISHITA, TORO: *Formal geometry of systems of differential equations.* Sugaku
           Expositions, **3**:1 (1990), pp. 25 - 73.

2. Generalized limits of infinite sequences and their matrix transformations.

One of the basic notions of mathematical analysis is the notion of *the limit of a sequence of real or complex numbers.* A sequence $\{x_k\}_{1 \le k < \infty}$ of complex numbers for which the $\lim_{k \to \infty} x_k$ exists is said to be *convergent.* A sequence $\{x_k\}_{1 \le k < \infty}$ of complex numbers for which $\sup_k |x_k| < \infty$ is said to be *bounded.* Let the set of all convergent sequences be denoted by $\mathbf{c}$, the set of all bounded sequences be denoted by $\mathbf{m}$, and the set of all sequences be denoted by $\mathbf{s}$.

It is clear that each of the sets $\mathbf{c}$, $\mathbf{m}$ and $\mathbf{s}$ is a vector space, and that $\mathbf{c} \subset \mathbf{m} \subset \mathbf{s}$.

Sometimes one needs to define a generalized limit $R\text{-}\lim_{k \to \infty} x_k$ (according to some rule R) for some sequences for which the "usual" limit $\lim_{k \to \infty} x_k$ may not exist. Let $\mathbf{c}_R$ denote the set of all sequences $\{x_k\}_{1 \le k < \infty}$ for which the $R\text{-}\lim_{k \to \infty} x_k$ is defined (in other words,"exists"). Usually some natural requirements are imposed on such a rule R. Thus, for example, it is often required that the set $\mathbf{c}_R$ be a vector space. In this case, if the condition

$$\mathbf{c} \subset \mathbf{c}_R \quad \text{and} \quad R\text{-}\lim_{k \to \infty} x_k = \lim_{k \to \infty} x_k \quad \text{for all} \quad \{x_k\}_{1 \le k < \infty} \in \mathbf{c},$$

is satisfied, then the generalized limit $R\text{-}\lim$ is said to be *regular.*

A familiar example of a generalized limit is the well known (Cèsaro) C-limit: Given a sequence $\{x_k\}_{1 \le k < \infty}$, the sequence $\{y_k\}_{1 \le k < \infty}$ is defined as

$$y_n = \frac{x_1 + x_2 + \cdots + x_n}{n} \quad (n = 1, 2, 3, \dots)$$

By definition, the C-limit of the sequence $\{x_k\}$ exists, if usual limit of the sequence $\{y_k\}$ exists, and

$$C\text{-}\lim_{k \to \infty} x_k \overset{\text{def}}{=} \lim_{k \to \infty} y_k.$$

It is not difficult to prove that the C-limit is regular. There are sequences for which the C-limit exists, but the "usual" limit does not exist. For example, the sequence $x_k \overset{\text{def}}{=} \frac{1+(-1)^k}{2}$ does not tend to a limit as $k \to \infty$, but the Cèsaro limit exists, and $C\text{-}\lim_{k \to \infty} x_k = \frac{1}{2}$.

Matrix transformations can be used to define generalized limits. Let A be an infinite

8

matrix:

$$
A = \begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1k} & \cdots \\
a_{21} & a_{22} & \cdots & a_{2k} & \cdots \\
. & . & . & . & . \\
a_{n1} & a_{n2} & \cdots & a_{nk} & \cdots \\
. & . & . & . & .
\end{bmatrix}, \tag{2.1}
$$

where the matrix entries a_{jk} are real or complex numbers. The matrix transformation $x \to y = Ax$ is defined for those sequences $x = \{x_k\}_{1 \leq k < \infty}$ for which all the series $\sum_{1 \leq k < \infty} a_{nk} x_k$, $n = 1, 2, 3, \ldots$, converge. The resulting sequence $y = Ax$, $y = \{y_k\}_{1 \leq k < \infty}$ is defined by $y_n = \sum_{1 \leq k < \infty} a_{nk} x_k$, $(n = 1, 2, 3, \ldots)$. It is clear that the domain of definition $\mathcal{D}_A$ of the matrix transformation generated by the matrix A is a vector space, $\mathcal{D}_A \subset \mathbf{s}$. Moreover, there is a natural generalized limit associated with each such infinite matrix A (that we denote as A-limit and which we shall refer to as *the matrix generalized limit generated by the matrix A*). Namely, by definition, the A-*limit of a sequence* $x = \{x_k\}_{1 \leq k < \infty}$ exists, if $x \in \mathcal{D}_A$ (i.e. the matrix transformation Ax is defined), and the sequence $y = Ax$ is convergent: $y \in \mathbf{c}$. By definition,

$$
A\text{-}\lim_{k \to \infty} x_k = \lim_{k \to \infty} y_k. \tag{2.2}
$$

The Cèsaro generalized limit (C-limit) can be considered as the matrix generalized limit generated by the lower triangular matrix A for which $a_{nk} = \dfrac{1}{n}$ for $k = 1, 2, \ldots, n$ and $a_{nk} = 0$ for $k > n$. The systematic investigation of matrix generalized limits was initiated by O.Toeplitz, [Toep]. A fundamental contribution to the theory of matrix generalized limits was made by Schur. In [Sch16] he introduced three classes of matrix transformations: *convergence preserving, convergence generating* and *regular*.

A matrix transformation $x \to Ax$ is said to be

1. *convergence preserving*, if it is defined for every sequence $x \in \mathbf{c}$, and for $x \in \mathbf{c}$ the sequence $y = Ax$ belongs to $\mathbf{c}$ as well.
2. *convergence generating*, if it is defined for every sequence $x \in \mathbf{m}$, and for $x \in \mathbf{m}$ the sequence $y = Ax$ belongs to $\mathbf{c}$.
3. *regular*, if it is convergence preserving and, moreover, if $x \in \mathbf{c}$ and $y = Ax$, then the equality $\lim_{k \to \infty} y_k = \lim_{k \to \infty} x_k$ holds.

Schur obtained necessary and sufficient conditions for a matrix transformation $x \to Ax$ to belong to each of these three classes. These conditions are presented in the following three theorems that are taken from [Sch16]. They are formulated in terms of the numbers

$$
\sigma_n = \sum_{1 \leq k < \infty} a_{nk} \quad \text{and} \quad \zeta_n = \sum_{1 \leq k < \infty} |a_{nk}| \tag{2.3}
$$

9

for those n, $1 \le n < \infty$, for which these values exist. The values σ_n are said to be the *row sums*; the values ζ_n are said to be the *row norms*.

THEOREM I. *The matrix transformation A is convergence preserving if and only if the following three conditions are satisfied:*

1. *For every k the following limit exists*

$$a_k \overset{\text{def}}{=} \lim_{n \to \infty} a_{nk}. \tag{2.4}$$

2. *The row sums σ_k tend to the finite limit σ:*

$$\sigma = \lim_{k \to \infty} \sigma_k. \tag{2.5}$$

3. *The sequence of row norms is bounded:*

$$\sup_{1 \le n < \infty} \zeta_n < \infty. \tag{2.6}$$

If these conditions are satisfied, then the series $\sum\limits_{1 \le k < \infty} a_k$ *converges absolutely and, if*

$$\alpha \overset{\text{def}}{=} \sum_{1 \le k < \infty} a_k, \tag{2.7}$$

then for every convergent sequence $\{x_k\}$

$$\lim_{n \to \infty} \sum_{1 \le k < \infty} a_{nk} x_k = (\sigma - \alpha) \lim_{k \to \infty} x_k + \sum_{1 \le k < \infty} a_k x_k. \tag{2.8}$$

THEOREM II. *The convergence preserving matrix transformation A is regular if and only if all the column limits a_k defined in (2.4) are equal to zero:*

$$a_k = 0 \qquad (k = 1, 2, 3, \ldots), \tag{2.9}$$

and the limit σ of the row sums σ_n defined in (2.5) is equal to 1:

$$\sigma = 1. \tag{2.10}$$

THEOREM III. *The matrix transformation A is convergence generating if and only the three assumptions of Theorem I are satisfied and the series* $\sum\limits_{1 \le k < \infty} |a_{nk}|$, $n = 1, 2, 3, \ldots,$

converge uniformly with respect to n. In this case

$$\lim_{n \to \infty} \sum_{1 \le k < \infty} a_{nk} x_k = \sum_{1 \le k < \infty} a_k x_k. \tag{2.11}$$

10

Theorem II was formulated and proved by O. Toeplitz in ([Toep]). However, Toeplitz considered only lower-triangular matrices A . Theorem II is commonly known as the Toeplitz theorem or as the Silverman-Toeplitz theorem, since part of Theorem II was obtained also by L.L. Silverman in his PhD thesis, [Silv]. Theorem I is known as the Schur-Kojima theorem. (Part of Theorem I was also obtained by T.Kojima for lower-triangular matrices.) The paper [Koj] by Kojima was published earlier than the paper [Sch16] by Schur. However, in a footnote on the last page of [Sch16], Schur remarks that he only became aware of the paper [Koj] while reading the proofs of his own paper. The matrices A which correspond to convergent generated transformations are called *Schur matrices* in [Pet]. There is a rich literature dedicated to matrix generalized limits and to matrix summation methods. (If a considered sequence is a sequence of partial sums of a series, then the terminology "generalized summation method" is used instead of "generalized limit" or "generalized limitation method".) We mention only the books [Bo], [Coo], [Har], [Pet], [Pey] and [Zel]. In all these books, the sections that deal with the basic theory of generalized limits and generalized summation methods cite the results of Schur and refer to him as one of the founders of this theory.

In a footnote near the beginning of his paper [Sch16], Schur notes that his considerations have many points in common with the considerations of H. Lebesgue and H. Hahn, dedicated to the sequence of integral transformations of the form

$$y_n(r) = \int\limits_a^b A_n(r,\, s)\, ds\,.$$

He also considers some applications of his Theorems I - III to the multiplication of series and to Tauberian theorems. In particular, he derives the Tauberian therem by Tauber (about power series) from Theorem II.

In his other paper [Sch6], Schur consider Hölder and Cèsaro limit methods of r-th order and proves that these limit methods are equivalent.

Given a sequence $x_1,\, x_2,\, x_3,\, \dots$ of real or complex numbers, we form the sequences

$$h_n^{(1)} = \frac{x_1 + x_2 + \cdots + x_n}{n}\,, \qquad h_n^{(2)} = \frac{h_1^{(1)} + h_2^{(1)} + \cdots + h_n^{(1)}}{n}\,,$$

$$h_n^{(3)} = \frac{h_1^{(2)} + h_2^{(2)} + \cdots + h_n^{(2)}}{n}\,, \quad \dots \quad , h_n^{(r)} = \frac{h_1^{(r-1)} + h_2^{(r-1)} + \cdots, + h_n^{(r-1)}}{n}\,.$$

The sequence $h_1^{(r)},\, h_2^{(r)},\, \dots, h_k^{(r)},\, \dots$ is said to be *the sequence of Hölder means of order r* (constructed from the initial sequence $x_1,\, x_2,\, x_3,\, \dots$). Another class of sequences can be constructed as follows. Let

$$s_n^{(1)} = x_1 + x_2 + \cdots + x_n, \quad s_n^{(2)} = s_1^{(1)} + s_2^{(1)} + \cdots + s_n^{(1)},$$

$$s_n^{(3)} = s_1^{(2)} + s_2^{(2)} + \cdots + s_n^{(2)}, \quad \cdots \quad , s_n^{(r)} = s_1^{(r-1)} + s_2^{(r-1)} + \cdots + s_n^{(r-1)},$$

and set

$$c^{(r)} = \frac{s^{(r)}}{\binom{n+r-1}{r}}.$$

The sequence $c_1^{(r)}, c_2^{(r)}, \ldots, c_k^{(r)}, \ldots$ is said to be *the sequence of Cèsaro means of order r* (constructed from the initial sequence $x_1, x_2, x_3, \ldots$). The transformations

$$\{x_1, x_2, \ldots, x_k, \ldots\} \to \{h_1^{(r)}, h_2^{(r)}, \ldots, h_k^{(r)}, \ldots\}$$

and

$$\{x_1, x_2, \ldots, x_k, \ldots\} \to \{c_1^{(r)}, c_2^{(r)}, \ldots, c_k^{(r)}, \ldots\}$$

can be considered as matrix transformations based on appropriately defined matrices that we denote by $\mathsf{H}^{(r)}$ and $\mathsf{C}^{(r)}$, respectively. These matrices are lower-triangular. Both generalized limits $H^{(r)}$-limit and $C^{(r)}$-limit are regular. In [Sch6] it is shown that these generalized limits are equivalent in the following sense:

Let a sequence $x_1, x_2, x_3, \ldots$ and a natural number r be given. Then the sequence of Cèsaro means $\{c_1^{(r)}, c_2^{(r)}, \ldots, c_k^{(r)}, \ldots\}$ tends to a finite limit if and only if the sequence of Hölder means $\{h_1^{(r)}, h_2^{(r)}, \ldots, h_k^{(r)}, \ldots\}$ tends to a finite limit. Moreover, in this case, the two limits must agree.

Schur obtained this result by showing that both the matrices $(\mathsf{H}^{(r)})^{-1} \cdot \mathsf{C}^{(r)}$ and $(\mathsf{C}^{(r)})^{-1} \cdot \mathsf{H}^{(r)}$ satisfy the assumptions of Theorem II (the Toeplitz regularity criterion). Thus, the appropriate matrix transformations are regular.

This result by Schur was not new. At the time that the paper [Sch6] was published proofs of the equivalency of Hölder's and Cèsaro's methods had already been obtained by K. Knopp, by W. Schnee and by W.B. Ford. However, these proofs were very computational, very involved and not very transparent.

REFERENCES

[Bo]   Boos, J. *Classical and modern methods in summability.* (Assisted by Peter Cass.) Oxford Mathematical Monographs. Oxford Science Publications. Oxford University Press, Oxford 2000.

[Coo]   Cooke, R.G. *Infinite Matrices and Sequence Spaces.* MacMillan, London 1950.

[Har]   Hardy, G.H. *Divergent Series.* Clarendon Press, Oxford 1949.

[Koj]   KOJIMA, T. *On generalized Toeplitz's theorems and their applications.* Tôhoku Mathematical Journal, **12** (1917), pp. 291 - 326.

[Pet]   PETERSEN, G.M. *Regular Matrix Transformation.* McGraw-Hill Publishing Compamy, London·New York· Toronto· Sidney 1966.

[Pey]   PEYERIMHOFF, A. *Lectures on summability.* (Lect. Notes in Math, **107**), Spinger Verlag, Berlin·Heidelbrg·New York 1969.

[Sch: Ges] SCHUR, I.: *Gesammelte Abhandlungen [Collected Works]. Vol. I, II, III.* Springer-Verlag, Berlin·Heidelberg·New York 1973.

[Sch6]   SCHUR, I.: *Über die Äquivalenz der Cesàroschen und Hölderschen Mittelwerte. [On the equivalence of Cesàro's and Hölder means - in German].* Mathematische Annalen, **74** (1913), pp. 447 - 458. Reprinted in: [Sch: Ges], Vol. II, pp. 44 - 55.

[Sch16]   SCHUR, I.: *Über lineare Transformationen in der Theorie der unendlichen Reihen. [On linear transformations in the theory of infinite series - in German].* Journ. für die reine und angew. Math., **151** (1921), pp. 79 - 121. Reprinted in: [Sch: Ges], Vol. II, pp. 289 - 321.

[Sch19]   SCHUR, I. *Einige Bemerkungen zur Theorie der unendlichen Reihen. [A remark on the theory of infinite series - in German].* Sitzungsberichte der Berliner mathematisches Gesellschaft, **29**, (1930), pp. 3 - 13. Reprinted in: [Sch: Ges], Vol. III, pp. 216 - 226.

[Silv]   SILVERMAN, L.L. *On the definition of the sum of a divergent series.* University of Missouri Studies, Math. Series I, (1913), pp. 1 - 96.

[Toep]   TOEPLITZ, O.: *Über allgemeine lineare Mittelbildungen.* Prace matematyczno-fizyczne (Warszawa), **22** (1913), pp. 113 - 119.

[Zel]   ZELLER, K. *Theorie der Limitierungsverfahren.* Springer, Berlin 1958.

3. Estimates for matrix and integral operators, bilinear forms and related inequalities.

The terms *Schur test, Schur (or Hadamard-Schur) multiplication of matrices, Schur (or Hadamard-Schur) multipliers* are all related to Schur's contributions to the estimates of operators and bilinear forms, see [Sch4]. In this section we consider the Schur test. Results related to the Schur (or Schur-Hadamard) product will be considered in the next section.

Let $A = \|a_{jk}\|$ be a matrix, finite or infinite, with real or complex entries. This matrix generates the bilinear form

$$A(x, y) = \sum_{j,k} a_{jk} x_k y_j \,, \tag{3.1}$$

where x and y are vectors with entries $\{x_j\}$ and $\{y_k\}$ that are real or complex. The matrix A also generates the linear operator

$$x \to Ax, \quad \text{where} \quad (Ax)_j = \sum_k a_{jk} x_k. \tag{3.2}$$

If the matrix A is not finite, we consider only *finite* vectors x and y, i.e., vectors with only finitely many nonzero entries. This allows us to avoid troubles related to the convergence of infinite sums. If the sets of vectors x and y are provided with norms, then a problem of interest is to estimate the bilinear form (3.1) in terms of the norms of the vectors x and y. In particular, the sets of vectors x and y can be provided with l_2 norms:

$$\|x\|_{l_2} = \left\{ \sum_j |x_j|^2 \right\}^{1/2}, \quad \|y\|_{l_2} = \left\{ \sum_k |y_k|^2 \right\}^{1/2}. \tag{3.3}$$

If the estimate

$$|A(x,y)| \leq C \|x\|_{l^2} \|y\|_{l^2} \tag{3.4}$$

holds for every pair of vectors x and y for some constant $C < \infty$, then the bilinear form (3.1) is said to be *bounded*. The smallest constant C, for which the inequality (3.4) holds, is denoted by C_A and is termed the norm of the bilinear form $A(x,y)$:

$$C_A = \sup_{x \neq 0,\, y \neq 0} \frac{|A(x,y)|}{\|x\|_{l_2} \|y\|_{l_2}}. \tag{3.5}$$

The norm C_A of the bilinear form generated by the matrix A coincides with the norm of the linear operator generated by this matrix, considered as a linear operator acting from l_2 into l_2:

$$\|A\|_{l_2 \to l_2} = \sup_{x \neq 0} \frac{\|Ax\|_{l_2}}{\|x\|_{l_2}}. \tag{3.6}$$

The cases in which it is possible to express the norm C_A in terms of the entries of the matrix A are very rare. Thus, the problem of estimating the value of C_A in terms of the matrix entries is a very important problem. In particular, if the matrix A is infinite, it is important to recognize whether the value C_A is finite or not. Schur made important contributions to this circle of problems.

In [Sch4] (§2, Theorem I), the following estimate was obtained.

THEOREM (The Schur test). *Let* $A = \|a_{jk}\|$ *be a matrix, and let*

$$\zeta(A) = \sup_j \sum_k |a_{jk}|, \quad \kappa(A) = \sup_k \sum_j |a_{jk}|. \tag{3.7}$$

Then

$$C_A \leq \sqrt{\zeta(A)\kappa(A)}. \tag{3.8}$$

14

It is enough to prove the estimate (3.8) for finite matrices A (of arbitrary size). The proof of the estimate (3.8) that was obtained in [Sch4] is based on the fact that

$$C_A = \sqrt{\lambda_{\max}}, \tag{3.9}$$

where $\lambda_{\max}$ is the largest eigenvalue of the matrix $B = A^*A$. Let $\boldsymbol{\xi} = \{\xi_k\}$ be the eigenvector which corresponds to the eigenvalue $\lambda_{\max}$:

$$\lambda_{\max}\boldsymbol{\xi} = B\boldsymbol{\xi} \,.$$

Let $|\xi_p| = \max_k |\xi_k|$. Then, since $\lambda_{\max}|\xi_p| \leq \left(\sum_k |b_{pk}| \right)|\xi_p|$, it is easily seen that

$$\lambda_{\max} \leq \sum_k |b_{pk}| \,,$$

where the $\{b_{jk}\}$ are the entries of the matrix $B = A^*A$: $b_{jk} = \sum_r \overline{a}_{rj}a_{rk}$. Thus,

$$\sum_k |b_{pk}| \leq \sum_k \sum_r |\overline{a}_{rp}||a_{rk}| \leq \left(\sum_r |a_{rp}| \right) \cdot \left(\max_r \sum_k |a_{rk}| \right) \leq \kappa(A) \cdot \zeta(A) \,.$$

This completes the proof.

Another proof, which does not use the equality (3.9), is even shorter:

$$
\begin{aligned}
\left| A(x,y) \right| &\leq \sum_{j,k} |a_{jk}| \cdot |x_k| \cdot |y_j| \\
&= \sum_{j,k} \left(|a_{jk}|^{1/2}|x_k| \right) \cdot \left(|a_{jk}|^{1/2}|y_j| \right) \\
&= \left(\sum_{j,k} |a_{jk}||x_k|^2 \right)^{1/2} \cdot \left(\sum_{j,k} |a_{jk}||y_j|^2 \right)^{1/2} \\
&\leq \left(\sup_k \sum_j |a_{jk}| \cdot \sum_k |x_k|^2 \right)^{1/2} \cdot \left(\sup_j \sum_k |a_{jk}| \cdot \sum_j |y_j|^2 \right)^{1/2} \\
&= \sqrt{\kappa(A)\zeta(A)}\|x\|_{l_2}\|y\|_{l_2} \,.
\end{aligned}
\tag{3.10}
$$

The estimate (3.8) can be considered as a special case of an interpolation theorem that is obtained by introducing the l_1 and l_∞ norms. If $x = \{x_k\}$ is a finite sequence of real or complex numbers, then these norms are defined by the usual rules:

$$\|x\|_{l_1} = \sum_k |x_k| \quad \text{and} \quad \|x\|_{l_\infty} = \sup_k |x_k| \,, \tag{3.11}$$

respectively. If A is a matrix, we can consider the linear operator generated by this matrix as an operator acting in the space l_1 as well as an operator acting in the space l_∞. The corresponding norms $\|A\|_{l_1 \to l_1}$ and $\|A\|_{l_\infty \to l_\infty}$ are defined by the formulas

$$\|A\|_{l_1 \to l_1} = \sup_{x \neq 0} \frac{\|Ax\|_{l_1}}{\|x\|_{l_1}} \quad \text{and} \quad \|A\|_{l_\infty \to l_\infty} = \sup_{x \neq 0} \frac{\|Ax\|_{l_\infty}}{\|x\|_{l_\infty}}.$$

Unlike the norm $\|A\|_{l_2 \to l_2}$, the norms $\|A\|_{l_1 \to l_1}$ and $\|A\|_{l_\infty \to l_\infty}$ can be expressed explicitly in terms of the matrix entries $\{a_{jk}\}$:

$$\|A\|_{l_1 \to l_1} = \kappa(A) \quad \text{and} \quad \|A\|_{l_\infty \to l_\infty} = \zeta(A),$$

where the numbers $\zeta(A)$ and $\kappa(A)$ are defined in (3.7). The estimate (3.8) takes the form

$$\|A\|_{l_2 \to l_2} \leq \sqrt{\|A\|_{l_1 \to l_1} \cdot \|A\|_{l_\infty \to l_\infty}}. \tag{3.12}$$

The inequality (3.12) is a direct consequence of the M. Riesz' Convexity Theorem. To apply this theorem, let $\|A\|_{l_p \to l_q}$ denote the norm of the operator, generated by a matrix A, considered as an operator from l_p into l_q for $1 \leq p \leq \infty$, $1 \leq q \leq \infty$. Then, Riesz' theorem states that $\log \|A\|_{l_p \to l_q}$ is a convex function of the variables $\alpha = 1/p$ and $\beta = 1/q$ in the square $0 \leq \alpha \leq 1$, $0 \leq \beta \leq 1$. This theorem can be found in [HLP], Chapter VIII, sec. **8.13.** G.O. Thorin, [Tho], found a very beautiful and ingenious proof of this theorem using a new method based on Hadamard's Three Circles Theorem from complex analysis. Therefore this theorem is also called the Riesz-Thorin Convexity Theorem. Now this theorem is presented in many sources, and even in textbooks. The Riesz-Thorin Convexity Theorem belongs to a general class of interpolation theorems for linear operators. A typical interpolation theorem for linear operators deals with a linear operator that is defined by a certain analytic expression, for example by a certain matrix or kernel, but is considered not in a fixed space, but in a whole "scale" of spaces. A typical interpolation theorem claims that if the linear operator, generated by the given expression, is bounded in two spaces of the considered "scale of spaces", then it also is bounded in all the "intermediate" spaces. Moreover, the norm of the operator in the "intermediate" spaces is estimated through the norms of the operators in the original two spaces. The Riesz-Thorin theorem states that the spaces l_p with $1 < p < \infty$ are "intermediate" for the pair of spaces l_1 and l_∞.

The estimate (3.8) can also be considered as a special case of another interpolation theorem for linear operators, the so-called interpolation theorem for modular spaces. This theorem is based on quite another circle of ideas that are more geometrical in nature and was partially inspired by Schur's work ([Sch18]). We will discuss this in the next section.

For practical application, the "weighted" version of the Schur estimate (3.8) is useful. In fact, this version was also considered in ([Sch4]) (but not as explicitly, as the "unweighted" version). In the weighted version, a positive sequence $\{r_k\}$, $r_k > 0$, appears and the

"weighted" l_1- and l_∞-norms

$$\|x\|_{l_1,r} = \sum_k |x_k| \cdot r_k \qquad \text{and} \qquad \|x\|_{l_\infty,r^{-1}} = \sup_k \frac{\|x_k\|}{r_k} \tag{3.13}$$

are considered.

THEOREM (The weighted Schur test). *Let $A = [a_{jk}]$ be a matrix and let r_k be a sequence of strictly positive numbers: $r_k > 0$. Let*

$$\zeta_r(A) = \sup_j \frac{1}{r_j} \cdot \sum_k |a_{jk}| \cdot r_k \quad \text{and} \quad \kappa_r(A) = \sup_k \frac{1}{r_k} \sum_j |a_{jk}| \cdot r_j. \tag{3.14}$$

Then the value C_A, defined in (3.5) is subject to the bound

$$C_A \le \sqrt{\zeta_r(A)\kappa_r(A)}. \tag{3.15}$$

It is easy to see that

$$\zeta_r(A) = \sup_{x \ne 0} \frac{\|Ax\|_{l_1,r}}{\|x\|_{l_1,r}} \quad \text{and} \quad \kappa_r(A) = \sup_{x \ne 0} \frac{\|Ax\|_{l_\infty,r^{-1}}}{\|x\|_{l_\infty,r^{-1}}}.$$

Thus, the estimate (3.15) can be presented in the form

$$\|A\|_{l_2 \to l_2} \le \sqrt{\|A\|_{l_1,r \to l_1,r} \cdot \|A\|_{l_\infty,r^{-1} \to l_\infty,r^{-1}}}. \tag{3.16}$$

The inequality (3.16) is also an "interpolation" inequality. It shows that the space l_2 is an "intermediate" space, between the spaces $l_{1,r}$ and $l_{\infty,r^{-1}}$.

The inequality (3.15) can be proved in much the same way as the special case (3.8).

As an example, we consider a Toeplitz matrix, i.e., a matrix A of the form $a_{jk} = w_{j-k}$. The Schur test leads to the estimate

$$C_A \le \sum_l |w_l|.$$

The same bound holds for Hankel matrices, i.e., matrices A of the form $a_{jk} = w_{j+k}$.

As a second example, let us consider the Hilbert matrix $H^+ = \left[\dfrac{1}{j+k-1}\right]_{j,k=1}^\infty$. For this matrix, $\sum_k |h_{jk}^+| = \infty$, so the "unweighted" Schur test does not work. However, if

17

we chose $r_l = l^{-\alpha}$ with a fixed $\alpha \in (0, 1)$, then $\displaystyle\sup_{1 \le j < \infty} \left(j^\alpha \sum_{1 \le k < \infty} \frac{k^{-\alpha}}{j+k} \right) = s(\alpha) < \infty.$

Thus, $\|H^+\| \le s(\alpha)$. Then we can optimize the estimate by choosing the "best" α. In the discrete case the precise value $s(\alpha)$ is unknown. Nevertheless, it is reasonable to choose $\alpha = 1/2$, since this is the optimum value for the continuous analogue of the matrix H^+:

$$x^\alpha \int\limits_0^\infty \frac{1}{x+y} y^{-\alpha} dy = \frac{\pi}{\sin \pi\alpha}, \quad \min_{\alpha \in (0,\,1)} \frac{\pi}{\sin \pi\alpha} = \pi \text{ is attained at the point } \alpha = 1/2\,.$$

(3.17)

Some other applications of the Schur test can be found in [BiSo], Chapter 2, Section **10**.

Schur used the estimate (3.8) in ([Sch4]) to study the infinite Hilbert forms

$$H^- = \sum_{\substack{p,q=1 \\ p \ne q}}^\infty \frac{x_p y_q}{p - q}, \qquad H^+ = \sum_{p,q=1}^\infty \frac{x_p y_q}{p + q - 1}, \tag{3.18}$$

and the generalized Hilbert forms

$$H_\lambda^- = \sum_{p,q=1}^\infty \frac{x_p y_q}{p - q + \lambda}, \qquad H_\lambda^+ = \sum_{p,q=1}^\infty \frac{x_p y_q}{p + q - 1 + \lambda}, \qquad (0 < \lambda < 1).$$

For the Hermitian matrices

$$N = (H^+)^* H^+ + (H^-)^* H^- = [n_{pq}]_{p,q=1}^\infty \quad \text{and} \quad N_\lambda = (H_\lambda^+)^* H_\lambda^+ + (H_\lambda^-)^* H_\lambda^-,$$

the conditions

$$\sum_{1 \le q < \infty} |n_{pq}| < 3 \sum_{\substack{q=-\infty \\ q \ne p}}^\infty \frac{1}{(p-q)^2} = \pi^2, \quad \sum_{1 \le q < \infty} |(n_\lambda)_{pq}| \le \sum_{-\infty < r < \infty} \frac{1}{(r + \lambda)^2} = \frac{\pi^2}{\sin^2 \pi\lambda}$$

are satisfied for every p. According to (3.8), the estimates

$$C_{H^+} \le \pi, \quad C_{H^-} \le \pi, \quad C_{H_\lambda^+} \le \frac{\pi}{\sin \pi\lambda}, \quad C_{H_\lambda^-} \le \frac{\pi}{\sin \pi\lambda} \tag{3.19}$$

hold. It turns out that in fact equality prevails in the first two inequalities in (3.19), i.e., the method of Schur gives the exact values for the norms of the matrices H^+, H^-. (See [HLP], Chapter IX). It should be remarked that an essential part of Chapters VIII and IX of [HLP] is based on results of the paper [Sch4].

In § 6 of [Sch4], the infinite quadratic form

$$F(t) = \sum_{\substack{p,q=1 \\ (p \ne q)}}^\infty \frac{\sin (p-q)t}{p-q} x_p x_q\,, \tag{3.20}$$

18

is considered, where $t, -\pi < t < \pi$, is a parameter. It is shown that the form $F(t)$ is bounded and that

$$- t \sum_{p=1}^{\infty} x_p^2 \leq F(t) \leq (\pi - t) \sum_{p=1}^{\infty} x_p^2. \qquad (3.21)$$

It is also shown that the quadratic form

$$\sum_{\substack{p,q=1 \\ (p \neq q)}}^{\infty} \left| \frac{\sin (p - q)t}{p - q} \right| x_p x_q$$

is unbounded for $t \neq 0$. The form (3.20) is interesting because it is an example of a *symmetric* infinite matrix $[a_{pq}]$ that corresponds to a bounded bilinear form, whereas the form related to the matrix $[\,|a_{pq}|\,]$ is unbounded. The Hilbert matrix $[h_{pq}^-]$, also generates a bounded bilinear form H^- (see (3.18)) and the matrix $[\,|h_{pq}^-|\,]$ also corresponds to an unbounded form. However, the Hilbert matrix H^- is *antisymmetric*.

REFERENCES

[BiSo]   BIRMAN, M.S. and M.Z. SOLOMYAK. *Spektral'naya Teoriya Samosopryazhennykh Operatorov v Gil'bertovom Prostranstve* (Russian), Izdatel'stvo Leningradskogo Universiteta, Leningrad 1980, 264 pp. English transl.: *Spectral Theory of Self-Adjoint Operators in Hilbert Space*. D. Reidel Publishing Company, Dordrecht·Boston·Lancaster·Tokyo 1987, xvi, 301 pp.

[HLP]   HARDY, G.H., J.W. LITTLEWOOD, and G. POLYA. *Inequalities*. 1st ed., 2nd ed. Cambridge Univ. Press, London·New York 1934, 1952.

[Sch: Ges] SCHUR, I.: *Gesammelte Abhandlungen [Collected Works]. Vol. I, II, III.* Springer-Verlag, Berlin·Heidelberg·New York 1973.

[Sch4]   SCHUR, I.: *Bemerkungen zur Theorie der beschränkten Bilinearformen mit unendlich vielen Veränderlichen. [Remarks on the theory of bounded bilinear forms with infinitely many variables - in German]*. Journ. für reine und angew. Math., **140** (1911), pp. 1 - 28. Reprinted in: [Sch: Ges], Vol. I, pp. 464 - 491.

[Sch18] SCHUR, I.: *Über eine Klasse von Mittelbildungen mit Anwendungen auf die Determinantentheorie. [On a class of averaging mappings with applications to the theory of determinants - in German]*. Sitzungsberichte der Berliner Mathematischen Gesellschaft, **22** (1923), pp. 9 - 20. Reprinted in: [Sch: Ges], Vol. II, pp. 416 - 427.

[Tho]   THORIN, G.O. *Convexity theorem generalizing those of M. Riesz and Hadamard with some applications*. Comm. Sem. Math. Univ. Lund [Medd. Lunds Univ. Mat. Sem.] **9**, (1948), pp. 1–58.

4. The Schur product and Schur multipliers.

Let A and B be matrices of the same size whose entries are either real or complex numbers (or even belong to some ring $\mathfrak{R}$): $A = [a_{pq}], B = [b_{pq}]$. The *Schur product* $A \circ B$ of the matrices A and B is the matrix $C = [c_{pq}]$ (of the same size as A and B) for which $c_{pq} = a_{pq} \cdot b_{pq}$.

The term *Schur product* is used because the product $A \circ B$ was introduced in ([Sch4]) for matrices, and some basic results about this product were obtained by Schur in that paper. The most basic of these results states that the cone of positive semidefinite matrices is closed under the Schur product. We recall that a square matrix $M = [m_{pq}]$ (with complex entries) is said to be *positive semidefinite* if the inequality $\sum\limits_{p,q} m_{pq} x_q \overline{x_p} \geq 0$ holds for every sequence $\{x_k\}$ of complex numbers. (In the case of an infinite matrix M, only sequences $\{x_k\}$ with finitely many x_k different from zero are considered.)

THEOREM (The Schur product theorem, Theorem VII, [Sch4]). *If A and B are positive semidefinite matrices (of the same size), then their Schur product $A \circ B$ is a positive semidefinite matrix as well.*

For self-evident reasons, the Schur product is sometimes called *the entrywise product* or *the elementwise product*. It is also often referred to as the *Hadamard product*. The term *Hadamard product* seems to have appeared in print for the first time in the 1948 (first) edition of [Hal1]. This may be due to the well known paper of Hadamard [Had], in which he studied two Maclaurin series $f(z) = \sum_n a_n z^n$ and $g(z) = \sum_n b_n z^n$ with positive radii of convergence and their composition $h(z) = \sum_n a_n b_n z^n$, which he defined as the *coefficientwise* product. Hadamard showed that $h(\cdot)$ can be obtained from $f(\cdot)$ and $g(\cdot)$ by an integral convolution. He proved that any singularity z_1 of $h(\cdot)$ must be of the form $z_1 = z_2 z_3$, where z_2 is a singularity of $f(\cdot)$ and z_3 is a singularity of $g(\cdot)$. (This result is commonly known as *the Hadamard composition theorem*.) Even though Hadamard did not study entrywise products of matrices in this paper, the enduring influence of the cited result as well as his mathematical eminence seems to have linked his name firmly with term-by-term products of all kinds, at least for analysts. (Presentations of the Hadamard composition theorem can be found, for example, in [Bie], Theorem 1.4.1, and in [Tit], Section 4.6.

PROOF of the Schur product theorem. It is enough to prove this theorem for matrices of arbitrary finite size. First we prove the theorem for matrices A and B of rank one. In this case the matrices A and B must be of the form $A = a \cdot a^*$, $B = b \cdot b^*$, where a and b are column vectors. It is evident that the matrix $C = A \circ B$ is of the form $C = c \cdot c^*$ where the column vector c is just the Schur product of the column vectors a and b: $c = a \circ b$. Hence, the matrix C is positive semidefinite. In the general case, we use the spectral decomposition theorem. This theorem states that every finite positive

semidefinite matrix M admits a decomposition of the form $M = \sum\limits_{\lambda \in \sigma(M)} M(\lambda)$, where the summation index λ runs over the spectrum $\sigma(M)$ of the matrix M, and the matrices $M(\lambda)$ are either positive semidefinite matrices of rank one or zero matrices. Decomposing the given matrices A and B in this way: $A = \sum\limits_{\lambda \in \sigma(A)} A(\lambda)$, $B = \sum\limits_{\mu \in \sigma(B)} B(\mu)$, we see that $A \circ B = \sum\limits_{\substack{\lambda \in \sigma(A) \\ \mu \in \sigma(B)}} A(\lambda) \circ B(\mu)$ is a sum of positive semidefinite matrices: The Schur product $A(\lambda) \circ B(\mu)$ of positive definite matrices of rank one is a positive semidefinite matrix, whereas, if at least one of the matrices $A(\lambda)$ or $B(\mu)$ is equal to zero, then their Schur product is equal to zero. Thus, the theorem is proved. $\qquad\square$

Every matrix H, finite or infinite, generates a linear operator $\mathfrak{T}_H$ acting in the space of all matrices of the same size as H:

$$\mathfrak{T}_H : A \to H \circ A, \quad \text{or} \quad \mathfrak{T}_H A = H \circ A.$$

The linear operator $\mathfrak{T}_H$ is said to be *the Schur transformator generated by the matrix H*. (The term *transformator* is borrowed from [GoKr], who used it to designate a linear operator that acts in a space of matrices (operators).) If the Schur transformator $\mathfrak{T}_H$ is a bounded operator in a space of infinite matrices, equipped with a norm, then the matrix H is said to be *the Schur multiplier* (with respect to this norm).

The first basic estimate of the norm of the transformator $\mathfrak{T}_H$ was obtained by Schur in [Sch4] :

THEOREM (The Schur estimate for positive definite Schur transformators). *Let $H = [h_{pq}]$ be a positive semidefinite matrix for which*

$$D_H \overset{\text{def}}{=} \sup_p h_{pp} < \infty. \tag{4.1}$$

Then

$$\|H \circ A\|_{l^2 \to l^2} \leq D_H \|A\|_{l^2 \to l^2}. \tag{4.2}$$

(Here, as before, $\|A\|_{l^2 \to l^2}$ is the operator norm of the matrix A considered in the appropriate space l^2 of sequences).

PROOF of the estimate (4.2). We reproduce here the reasoning of Schur from [Sch4]. It suffices to consider only finite matrices. The proof is based essentially on the fact that a positive semidefinite matrix H admits a factorization of the form

$$H = LL^*, \tag{4.3}$$

where $L = [l_{pq}]$, i.e.,

$$h_{pq} = \sum_r l_{pr} \overline{l_{qr}} \quad (\forall\, p, q). \tag{4.4}$$

21

Therefore, the number $\sum\limits_{p,q} a_{pq}h_{pq}y_q\overline{x_p}$ can be rewritten as

$$\sum_{p,q} a_{pq}h_{pq}y_q\overline{x_p} = \sum_{p,q} a_{pq}\Big(\sum_r l_{pr}\overline{l_{qr}}\Big)y_q\overline{x_p} = \sum_r\sum_{p,q} a_{pq}(l_{pr}\overline{x_p})\overline{(l_{qr}y_q)}\,.$$

Thus,

$$\Big|\sum_{p,q} a_{pq}h_{pq}\overline{x_p}y_q\Big| \le \sum_r \Big|\sum_{p,q} a_{pq}(l_{pr}\overline{x_p})\overline{(l_{qr}y_q)}\Big| \le \sum_r \|A\|\Big(\sum_k |l_{kr}x_k|^2\Big)^{1/2}\Big(\sum_k |l_{kr}y_k|^2\Big)^{1/2}$$

$$= \|A\|\sum_r \Big(\sum_k |l_{kr}x_k|^2\Big)^{1/2}\Big(\sum_k |l_{kr}y_k|^2\Big)^{1/2} \le \|A\|\Big(\sum_r\sum_k |l_{kr}x_k|^2\Big)^{1/2}\Big(\sum_r\sum_k |l_{kr}y_k|^2\Big)^{1/2}$$

$$\le \|A\|\Big(\sum_k \big(\sum_r |l_{kr}|^2\big)|x_k|^2\Big)^{1/2}\Big(\sum_k \big(\sum_r |l_{kr}|^2\big)|y_k|^2\Big)^{1/2}$$

$$\le \|A\|\Big(\sum_k \big(\max_k \sum_r |l_{kr}|^2\big)|x_k|^2\Big)^{1/2}\Big(\sum_k \big(\max_k \sum_r |l_{kr}|^2\big)|y_k|^2\Big)^{1/2}$$

$$= \|A\|\Big(\max_k \sum_r |l_{kr}|^2\Big)\Big(\sum_k |x_k|^2\Big)^{1/2}\Big(\sum_k |y_k|^2\Big)^{1/2}\,.$$

According to (4.4), $\sum\limits_r |l_{kr}|^2 = h_{kk}$. Thus, $\max\limits_k\Big(\sum\limits_r |l_{kr}|^2\Big) = \max\limits_k h_{kk} = D_H$. Finally,

$$\Big|\sum_{p,q} a_{pq}h_{pq}\overline{x_p}y_q\Big| \le \|A\| \cdot D_H \cdot \Big(\sum_k |x_k|^2\Big)^{1/2}\Big(\sum_k |y_k|^2\Big)^{1/2}\,, \tag{4.5}$$

where $\{x_k\}$ and $\{y_k\}$ are *arbitrary* sequences. This is the estimate (4.2). $\qquad\square$

In fact, the reasoning of Schur allows us to prove a slightly more general result:

THEOREM (The Schur factorization estimate for Schur transformators). *Let* $H = [h_{pq}]$ *be a matrix which admits a factorization of the form*

$$H = L \cdot M^*, \quad \text{i.e.,} \quad h_{pq} = \sum_r l_{pr}\overline{m_{qr}} \quad (\forall p, q), \tag{4.6}$$

where the matrices $L = [l_{pr}]$ *and* $M = [m_{rq}]$ *satisfy the conditions*

$$D_L \stackrel{\text{def}}{=} \sup_p \sum_r |l_{pr}|^2 < \infty \quad \text{and} \quad D_M \stackrel{\text{def}}{=} \sup_q \sum_r |m_{qr}|^2 < \infty\,. \tag{4.7}$$

Then for every matrix A *(of the same size as* H*) the following inequality holds:*

$$\|H \circ A\|_{l^2 \to l^2} \le \sqrt{D_L D_M}\, \|A\|_{l^2 \to l^2}\,. \tag{4.8}$$

REMARK. The matrices L, M and H need not be square. The only restriction is that the matrix multiplication $L, M \to L \cdot M^*$ is feasible. In fact, the set over which the summation index r runs in (4.6) need not be a subset of the set of integers. It can be of a much more general nature. Thus, for example, *let $\mathfrak{X}$ be a measurable space carrying a sigma-finite non-negative measure dx. Let $\{l_p(x)\}$ and $\{m_q(x)\}$ be sequences of $\mathfrak{X}$-measurable functions defined on $\mathfrak{X}$ and satisfying the conditions $D_L < \infty$, $D_M < \infty$, where now*

$$D_L = \sup_k \int_{\mathfrak{X}} |l_k(x)|^2 \, dx \quad and \quad D_M = \sup_k \int_{\mathfrak{X}} |m_k(x)|^2 \, dx . \tag{4.9}$$

Let H be a matrix with entries

$$h_{pq} = \int_{\mathfrak{X}} l_p(x)\overline{m_q(x)} \, dx \quad (\forall \, p, q) \tag{4.10}$$

(i.e., the matrix H admits a factorization of the form $H = L \cdot M^*$, where L and M are operators acting from the Hilbert space $L^2(\mathfrak{X}, dx)$ into appropriate spaces of l^∞ sequences). Then the inequality (4.8) holds for an arbitrary matrix A (of the appropriate size), where now D_L and D_M are defined in (4.9).

The last result (with $\mathfrak{X} = (a, b)$, a finite or infinite subinterval of $\mathbb{R}$, and Lebesgue measure dx on (a, b)) appears as Theorem VI in [Sch4].

The matrix

$$H = \left[\frac{1}{\lambda_p + \mu_q} \right]_{1 \le p, q < \infty},$$

where λ_k and μ_k are sequences of positive numbers that are separated from *zero*: $\inf_k \lambda_k > 0$, $\inf_k \mu_k > 0$, serves as an example. Here,

$$h_{pq} = \int_0^\infty e^{-\lambda_p x} \cdot e^{-\mu_q x} dx , \quad i.e., \quad l_p(x) = e^{-\lambda_p x}, m_q(x) = e^{-\mu_q x}, \quad 1 \le p, q < \infty ,$$

and for this H the inequality (4.2):

$$\left\| \left[\frac{a_{pq}}{\lambda_p + \mu_q} \right] \right\|_{l^2 \to l^2} \le D_H \left\| [a_{pq}] \right\|_{l^2 \to l^2}$$

holds with

$$D_H = \frac{1}{2} \cdot \frac{1}{\sqrt{\inf_k \lambda_k}} \cdot \frac{1}{\sqrt{\inf_k \mu_k}}.$$

(This example is adopted from [Sch4]; it appears at the end of §4.)

It is remarkable that the existence of a factorization of the form $H = L \cdot M^*$ for the matrix H is not only sufficient but is also a necessary condition for the operator $A \to H \circ A$

23

to be a bounded operator in the space of all matrices A (equipped with the operator norm in l^2). This *converse* result was proved by G.Bennett in [Ben].

THEOREM (The inversion of the Schur factorization estimate). *Let a given matrix* $H = [h_{pq}]$ *(finite or infinite) satisfy the inequality*

$$\|H \circ A\|_{l^2 \to l^2} \leq D\|A\|_{l^2 \to l^2} \tag{4.11}$$

for all matrices A of the same size as H, for some finite constant D that does not depend on A. Then for every $\epsilon > 0$, the matrix H can be factored in the form $H = L \cdot M^*$, where the matrices $L = [l_{pr}]$ and $M = [m_{rq}]$ act from l^2 to l^∞ and satisfy the inequality $\sqrt{D_L \cdot D_M} < D + \epsilon$, and the values D_L and D_M are defined in (4.7), i.e., $D_L = \|L\|_{l^2 \to l^\infty}$ and $D_M = \|M\|_{l^2 \to l^\infty}$.

This theorem appears as Theorem 6.4 in [Ben]. It shows that the Schur factorization gives a result which is in some sense optimal. The proof of this theorem of G. Bennett is essentially based on results obtained by A. Pietsch on absolute summing operators in Banach spaces, see [Pie1] and [Pie2] (which, in turn, are based on fundamental results of A. Grothendieck, see the references in [Pie1] and [Pie2]).

In [Sch4], Schur considers a new class of functions of matrices, namely, the so called Schur (or Schur-Hadamard) functions of matrices. Let $A = [a_{pq}]$ be an infinite matrix whose entries have a common finite bound: $|a_{pq}| \leq R \ \ (\forall p, q)$, where $R < \infty$. Let $f(\cdot)$ be a function that is defined in the closed disk $\{z : |z| \leq R\}$. The matrix $f^\circ(A)$ is defined "entrywise" as follows:

$$f^\circ(A) \overset{\text{def}}{=} [f(a_{pq})].$$

The following result is proved in [Sch4]: *Let $f(z) = \sum\limits_{k=1}^{\infty} c_k z^k$, where $\sum\limits_{k=1}^{\infty} |c_k| R^k < \infty$ and let the operator generated by the matrix A be bounded, i.e., $\|A\|_{l^2 \to l^2} < \infty$. Then the operator generated by the matrix $f^\circ(A)$ is also bounded: $\|f^\circ(A)\|_{l^2 \to l^2} < \infty$.*
This result appears as Theorem IV in [Sch4].

The concept of the Schur (Schur-Hadamard) product arises in several different areas of analysis (complex function theory, Banach spaces, operator theory, multivariate analysis); see the references in the introduction to [Ben]. The paper [Sty] contains some applications of the Schur product to multivariate analysis as well as a rich bibliography of books and articles related to Schur-Hadamard products. The paper [HorR1] contains a lot of facts about Schur-Hadamard products and Schur-Hadamard functions of matrices as well as a rich bibliography. In particular, it discusses fractional Schur-Hadamard powers of a positive matrix, infinite Schur-Hadamard divisibility of a positive matrix and its relation to the conditional positivity of the logarithmic° matrix. Chapter 5 of the book [HorR2] (about eighty pages) is dedicated to the Schur-Hadamard product of matrices.

A very fruitful generalization of the Schur transformator is the *the Stieltjes double-integral operator*. This notion seems to have appeared first in the papers of Yu.L. Daletskii

and S.G. Krein [DaKr1], [DaKr2], [Da1], [Da2]. Later on, the theory of double-integral operators was elaborated on in great detail by M.S. Birman and M.Z. Solomyak in [BiSo1] – [BiSo4].

Let Λ and M be measurable spaces, i.e., sets provided with sigma-algebras of subsets, and let $E(d\lambda)$ and $F(d\mu)$ be two *orthogonal measures in a separable Hilbert space* $\mathfrak{H}$ that are defined on Λ and M, respectively, i.e., weakly-countably-additive functions taking their values in the set of orthogonal projectors in $\mathfrak{H}$ and satisfying the condition $E(\alpha)E(\beta) = 0$ if $\alpha \cap \beta = \emptyset$ and $F(\gamma)F(\delta) = 0$ if $\gamma \cap \delta = \emptyset$. We assume also that the orthogonal measures $E(d\lambda)$ and $F(d\mu)$ are *spectral measures*, i.e., they also satisfy the conditions $E(\Lambda) = I$ and $F(\mathrm{M}) = I$, where I is the identity operator in $\mathfrak{H}$. If A is a bounded linear operator in $\mathfrak{H}$, then

$$A = \iint\limits_{\mathrm{M} \times \Lambda} F(d\mu)AE(d\lambda)\,, \tag{4.12}$$

where the integral can be understood in any reasonable sense. The equality (4.12) can be considered as a direct generalization of the matrix representation of an operator in a Hilbert space with respect to two orthonormal bases. Namely, let the orthogonal spectral measures $E(d\lambda)$ and $F(d\mu)$ be discrete and let their "atoms" be one-dimensional orthogonal projectors, i.e., the atom of the measure $E(d\lambda)$, located at the point $\lambda \in \Lambda$, is of the form $E(\{\lambda\}) = \langle \, \cdot \, , e_\lambda \rangle e_\lambda$ and the atom of the measure $F(d\mu)$, located at the point $\mu \in \mathrm{M}$, is of the form $F(\{\mu\}) = \langle \, \cdot \, , f_\mu \rangle f_\mu$, where e_λ and f_μ are normalized vectors generating the one-dimensional subspaces $E(\{\lambda\})\mathfrak{H}$ and $F(\{\mu\})\mathfrak{H}$, respectively. The collection of all the vectors $\{e_\lambda\}$ corresponding to all the atoms of the measure $E(d\lambda)$ forms an orthonormal basis of the space $\mathfrak{H}$. Analogously, the collection of all the vectors $\{f_\mu\}$ corresponding to all the atoms of the measure $F(d\mu)$ also forms an orthonormal basis of the space $\mathfrak{H}$. Consequently, the representation (4.12) of the operator A takes the form

$$A = \sum_{\lambda, \mu} f_\mu \, a_{\mu, \lambda} \, \langle \, \cdot \, , e_\lambda \rangle\,, \tag{4.13}$$

where $a_{\mu, \lambda} = \langle Ae_\lambda \, , f_\mu \rangle$. Thus, in the case of discrete orthogonal spectral measures with one-dimensional atoms, the representation (4.12) turns into the matrix representation of a given operator with respect to given orthonormal bases. The matrix $[a_{\mu, \lambda}]$ corresponds to the operator A. If $h(\mu, \lambda)$ is a measurable function defined on $\mathrm{M} \times \Lambda$, then the sum

$$\mathfrak{T}_h A \overset{\text{def}}{=} \sum_{\lambda, \mu} f_\mu \, h_{\mu, \lambda} \cdot a_{\mu, \lambda} \, \langle \, \cdot \, , e_\lambda \rangle \tag{4.14}$$

can be pictured as an application of the Schur transformator corresponding to the matrix $[h_{\mu, \lambda}]$ to the operator A: $A \mapsto \mathfrak{T}_h A$. The sum on the right hand side of the equality (4.14) can be formally written as an integral:

$$\mathfrak{T}_h A = \iint\limits_{\mathrm{M} \times \Lambda} h(\mu, \lambda) F(d\mu)AE(d\lambda)\,. \tag{4.15}$$

However, one can consider integrals of the form (4.15) for arbitrary orthogonal spectral measures $E(d\lambda)$ on Λ and $F(d\mu)$ on M, and more or less arbitrary functions $h(\mu, \lambda)$ on $M \times \Lambda$. If the integral (4.15) exists in a reasonable sense (either as a Lebesgue integral, or a Riemann-Stieltjes integral, or some other integral), it is said to be *a Stieltes double-integral operator*. The problem of establishing the existence of a Stieltes double-integral operator is intimately associated with estimates for it in various norms. In particular, the estimates

$$\left\|\mathfrak{T}_h A\right\|_{\mathfrak{R}\to\mathfrak{R}} \leq C \|A\|_{\mathfrak{R}\to\mathfrak{R}} \tag{4.16}$$

and

$$\left\|\mathfrak{T}_h A\right\|_{\mathfrak{S}_1\to\mathfrak{S}_1} \leq C \|A\|_{\mathfrak{S}_1\to\mathfrak{S}_1} \tag{4.17}$$

are extremely important. Here $\|\Phi\|_{\mathfrak{R}\to\mathfrak{R}}$ is the "uniform" norm of the operator Φ, acting in $\mathfrak{H}$: $\|\Phi\|_{\mathfrak{R}\to\mathfrak{R}} = \sup_{v\in\mathfrak{H}, v\neq 0} \dfrac{\|\Phi v\|_{\mathfrak{H}}}{\|v\|_{\mathfrak{H}}}$, and $\|\Phi\|_{\mathfrak{S}_1\to\mathfrak{S}_1}$ is its "trace" norm.

In [BiSo4] the estimate (4.16) was obtained for functions $h(\,\cdot\,,\,\cdot\,)$ which admit a "factorization" of the form

$$h(\mu, \lambda) = \int_{\mathfrak{X}} m(\mu, x) \cdot l(\lambda, x) dx \,, \tag{4.18}$$

where $\mathfrak{X}$ is a measurable space carrying a non-negative sigma-finite measure dx,

$$C_m = \operatorname*{ess\,sup}_{\mu\in M} \int_{\mathfrak{X}} |m(\mu, x)|^2 dx \,, \quad C_l = \operatorname*{ess\,sup}_{\lambda\in\Lambda} \int_{\mathfrak{X}} |l(\lambda, x)|^2 dx \tag{4.19}$$

and

$$C = \sqrt{C_m \cdot C_l} < \infty \,. \tag{4.20}$$

The inequality (4.16) is then obtained (with the same constant C) by invoking the duality between the set $\mathfrak{R}$ of all bounded operators in $\mathfrak{H}$ and the set $\mathfrak{S}_1$ of all trace class operators. The estimate (4.17) holds with the same constant C (that is given in (4.20)). Unfortunately, the paper [BiSo4] is not translated into English, but some results of this paper, in particular, the estimate (4.16), (4.20), are reproduced in [ABF], Section 2.

The estimate (4.16) is a direct analog of the Schur factorization estimate (4.8), (4.7) and is obtained by the same method that Schur used. However, when Birman and Solomyak started to develop the theory of Stieltjes double-integral operators, they were not aware of the paper [Sch4] by Schur. The close relationship between double-integral operators and the results of Schur was only discovered later. In Section 2 of [Pel], V. Peller obtained a result that "inverts" the estimate (4.16) by Birman and Solomyak in the same sense that Theorem 6.4 of [Ben] (that was stated earlier) inverts the factorization estimate by Schur. Peller proved an even stronger result, a "maximal" version of the inverse result. Namely, he proved that if the function h is such that the estimate (4.16) holds for every bounded operator A in $\mathfrak{H}$ with a finite constant C that is independent of A, then the

function $h(\,\cdot\,,\,\cdot\,)$ admits a factorization of the form (4.18), where the functions $m(\,\cdot\,,\,\cdot\,)$ and $(\,\cdot\,,\,\cdot\,)$ satisfy the conditions

$$\int_{\mathfrak{X}} \left(\operatorname*{ess\,sup}_{\mu \in \mathrm{M}} \left| m(\mu, x) \right| \right)^2 dx < \infty \quad \text{and} \quad \int_{\mathfrak{X}} \left(\operatorname*{ess\,sup}_{\lambda \in \Lambda} \left| l(\lambda, x) \right| \right)^2 dx < \infty. \tag{4.21}$$

The estimate (4.16), (4.20) is "semi-effective": given the function $h(\mu, \lambda)$, it is not so easy to see when it admits a factoraization of the form (4.18). To overcome this difficulty, Birman and Solomyak developed another approach that reduces the study of Stieltes double-integral operators to the study of integral operators of the form

$$u(\lambda) \to v(\mu) = \int_{\Lambda} h(\mu, \lambda) u(\lambda) \rho(d\lambda)\,. \tag{4.22}$$

This reduction is explained in [BiSo2], Theorem 2, and also in [BiSo4], Lemma 1.1. The operator (4.22) acts from the space $L^2(\Lambda, d\rho(\lambda))$ into the space $L^2(\mathrm{M}, d\sigma(\mu))$, where $\rho(d\lambda) = \langle E(d\lambda)\omega, \omega \rangle$, $\sigma(d\mu) = \langle F(d\mu)\theta, \theta \rangle$ and $\omega, \theta \in \mathfrak{H}$. The estimates for the integral operators (4.22) must be carried out for all vectors $\omega, \theta \in \mathfrak{H}$ and must be uniform with respect to the measures $\rho(d\lambda)$ and $\sigma(d\mu)$. To obtain such estimates, Birman and Solomyak developed a method that is based on the approximation of functions from the Sobolev-Slobodetskiĭ classes W_p^α by piecewise-polynomial functions, [BiSo5], [BiSo6], [BiSo7], §§ 8 - 9, [BiSo8], Chapter 3, §§ 5 - 7. In the construction of the approximating functions, a partition of the domain of definition of the approximated function appears. To achieve the desired uniformity of the approximation with respect to the measures $\rho(d\lambda)$ and $\sigma(d\mu)$, this partition must be adapted to these measures.

The approach, based on piecewise-polynomial approximations, allows one to approximate the kernels of the integral operators (4.22) by finite-dimensional kernels, and thus to obtain the needed estimates for the singular values of the Stieltjes double-integral operators. The estimates of the double-integral operators are made not only in the uniform and trace norms, but also in many other norms. These estimates depend upon the smoothness of the function $h(\,\cdot\,,\,\cdot\,)$ (assuming that Λ and M are smooth manifolds).

Double-integral operators appear in the formula for differentiating functions of Hermitian operators with respect to a parameter. Namely, let $\tau \to H(\tau)$ be a function on some open subinterval of the real axis $\mathbb{R}$ whose values are self-adjoint operators in a Hilbert space $\mathfrak{H}$. Let $f : \mathbb{R} \to \mathbb{R}$ be a real-valued function that is defined and bounded on $\mathbb{R}$ and let $E(d\lambda, \tau)$ be the spectral measure of the operator $H(\tau)$. Under appropriate assumptions, Yu. L. Daletskii and S. G. Krein, [DaKr1], obtained the formula

$$\frac{\partial f(H(\tau))}{\partial \tau} = \iint_{\mathbb{R} \times \mathbb{R}} \frac{f(\lambda) - f(\mu)}{\lambda - \mu}\, E(d\mu, \tau)\, \frac{\partial H(\tau)}{\partial \tau}\, E(d\lambda, \tau)\,. \tag{4.23}$$

This formula, which expresses the derivative $\dfrac{\partial f(H(\tau))}{\partial \tau}$ as a Stieltjes double-integral operator, seems to be the first recorded application of Stieltjes double-integral operators.

The paper [Da1] contains a version of Taylor's formula for operator functions. The paper [DaKr2] (and, to some extent, the paper [Da2]) contains a more detailed presentation of the results of the papers [DaKr1] and [Da1] as well as some extensions. Later on, Stieltjes double-integral operators were widely used in scattering theory. M.Sh. Birman, [Bi1], used them to prove the existence of wave operators. (See also [BiSo2], especially the last paragraph of this paper.) Double-integral operators are involved in the study of the so called *spectral shift function* (see [BiSo10] and [BiYa]). The paper [BiSo11] is devoted to the application of double-integral operators to the estimation of perturbations and commutators of functions of self-adjoint operators. It is worth noticing that double-integral operators allow one to make an abstract and symmetric definition of a pseudodifferential operator with prescribed symbol (see item **3** of the paper [BiSo9]).

Thus, the ideas of Issai Schur on the termwise multiplication of matrices, partially forgotten and rediscovered, are seen to lead very far from the original setting.

REFERENCES

[ABF]   ARAZY, J., T.J. BARTON and Y. FRIEDMAN. *Operator differentiable functions.* Integral Equat. and Oper. Theory, **13** (1990), pp. 461–487.

[Ben]   BENNETT, G. *Schur multipliers.* Duke Math. Journ., **44**:3 (1977), pp. 603 - 639.

[Bie]   BIEBERBACH, L. *Analytische Fortsetzung* [*Analytic Continuation*] (in German). (Ergebnisse der Mathematik und ihrer Gerenzgebiete, **3**). Springer Verlag, Berlin·Göttingen·Heidelberg 1955.

[Bi1]   BIRMAN, M.SH. *Ob usloviyakh sushchestvovaniya volnovykh operatorov* (Russian). Izvestiya AN SSSR (Ser. Mat.), **27**:4 (1963), pp. 883 - 906. English transl.: *On conditions for the existence of wave operators*, Amer. Math. Soc. Transl. (Ser. **2**), **54** (1966), pp. 91 – 117.

[Bi2]   BIRMAN, M.SH. *Lokal'nyĭ priznak sushchestvovaniya volnovykh operatorov* (Russian). Dokl. Akad. Nauk SSSR, **159**:3 (1964), pp. 485 - 488. English transl.: *A local criterion for the existence of wave operators* . Soviet Math. Dokl. **5** (1965), pp. 1505 - 1509.

[BiSo1]   BIRMAN, M.SH. and M.Z. SOLOMYAK. *O dvoĭnykh operatornykh integralakh Stil'tyesa* (Russian). Dokl. AN SSSR, **165**:6 (1965), pp. 1223 - 1226. English transl.: Stieltjes double operator integrals. Soviet Math., Dokl. **6**, (1965), 1567-1571.

[BiSo2]   BIRMAN, M.SH. and M.Z. SOLOMYAK. *Dvoĭnye operatornye integraly Stil'tyesa* (in Russian). In "Problemy matematicheskoĭ fiziki", No. **1**. Spektral'naya Teoriya i Volnovye Processy [Spectral Theory and Wave Processes]. (M.Sh. Birman-editor).Izdat.

Lenigradskogo Univ., Leningrad 1966, pp.33 - 67. English transl.: *Stieltjes double-integral operators*, in "Topics in Mathematical Physics", vol.**1**, Consultants Bureau, New York 1967, viii+114 pp.

[BiSo3]   BIRMAN, M.SH. and M.Z. SOLOMYAK. *Dvoĭnye operatornye integraly Stil'tyesa. II* (in Russian). In "Problemy matematicheskoĭ fiziki", No. **2**. Spectral'naya teoriya. Zadachi diffraktsii [Spectral theory. Diffraction problems]. (Edited by M.Sh. Birman). Izdat. Lenigradskogo Univ., Leningrad 1967, pp.26 - 60. English transl.: *Stieltjes double-integral operators.II*, in "Topics in Mathematical Physics", vol.**2**, Consultants Bureau, New York 1968, vii+134 pp.

[BiSo4]   BIRMAN, M.SH. and M.Z. SOLOMYAK. *Dvoĭnye operatornye integraly Stil'tyesa. III* (Russian). In "Problemy matematicheskoĭ fiziki", **6**, Izdat. Lenigradskogo Univ., 1973, pp.27 - 53.

[BiSo5]   BIRMAN, M.SH. and M.Z. SOLOMYAK. *O priblizhenii funktsiĭ klassov W_p^α kusochno-polinomial'nymi funktsiyami* (Russian), Dokl. Akad. Nauk SSSR, **171**:5 (1966), pp. 1015 -1018. English transl.: *Approximation of the classes W_p^α by piceweise polynomial functions.* Sov. Math., Dokl. **7** (1966), pp. 1573 - 1577.

[BiSo6]   BIRMAN, M.SH. and M.Z. SOLOMYAK. *Kusochno-polinomial'nye priblizheniya funktsiĭ klassow W_p^α.* (Russian) Matem. Sbornik (N.S.) **73** (**115**) (1967), pp. 331–355. English transl.: *Piecewise-polynomial approximations of functions of the classes W_p^α.* Math. USSR–Sbornik, **2**:3 (1967), pp. 295 - 317.

[BiSo7]   BIRMAN, M.SH. and M.Z. SOLOMYAK. *Otsenki singulyarnykh chisel integral'nykh operatorov* (Russian). Uspekhi Matem. Nauk **32**:1(193), (1977), pp. 17-84, 271. English transl.:*Estimates of singular numbers of integral operators*, Russ. Math. Surveys **32**:1 (1977), pp. 15-89.

[BiSo8]   BIRMAN, M.SH. and M.Z. SOLOMYAK *Kachestvennyĭ analiz v teoremakh vlozheniya Soboleva i prilozheniya k spektral'noĭ teorii* (Russian). Desyataya matematicheskays shkola. (Letnyaya shkola, Katsiveli/Nal'chik, 1972), pp. 5–189. Izdanie Inst. Matem. Akad. Nauk Ukrain. SSR, Kiev, 1974. English transl.: *Quantitative analysis in Sobolev imbedding theorems and applications to spectral theory.* American Mathematical Society Translations, Ser. **2**, **114**. American Mathematical Society, Providence, R.I., 1980. viii+132 pp.

[BiSo9]   BIRMAN, M.SH. and M.Z. SOLOMYAK. *Dvoĭnye operatornye integraly Stil'tyesa i zadacha o mnozhitekyakh.* Dokl. Akad. Nauk SSSR, **171**:6 (1966), 1251 - 1254. English transl.: *Double Stieltjes operator integrals and problem of multipliers.* Soviet Math. Dokl., **7** (1966), 1618 - 1621.

[BiSo10]   BIRMAN, M.SH. and M.Z. SOLOMYAK. *Zamechaniya o funktsii spektral'nogo sdviga* (Russian). Zapiski Nauchnykh Seminarov LOMI, **27** (1972), pp. 33 – 46. English transl.: *Remarks on the spectral shift function.* Journ. Soviet Math. **3**:4 (1975), pp. 408 - 419.

[BiSo11]   BIRMAN, M.SH. and M.Z. SOLOMYAK. *Operatornoe integrirovanie, vozmushcheniya i kommutatory* (Russian). Zapiski Nauchnykh Seminarov LOMI, vol. **170** (1989). English transl.: *Operator integration, perturbations, and commutators.* Journal of Soviet. Math., **63**:2 (1993), pp. 129 – 148.

[BiYa]   BIRMAN, M.SH. and D.R. YAFAEV. *Funktsiya spektral'nogo sdviga. Raboty M.G. Kreǐna i ikh dal'neǐshee razvitie* (Russian). Algebra i Analiz, **4**:5 (1992), pp. 1 – 44. English transl.: *The spectral shift functioin. The work of M.G.Kreǐn and its further development.* St. Petersburg Math. J. **4**:5 (1993), pp. 833 - 870.

[Da1]   DALETSKII, YU.L. *Pro otsinku zalushkovogo chlenu u formuli Teǐlora dlya funktsiǐ ermitovykh operatoriv* [On an estimate of the remainder term in Taylor's formula for functions of Hermitian operators] (Ukrainian). Dopovidi Akad. Nauk Ukraïn. RSR, **4** (1951), pp. 234–238.

[Da2]   DALETSKII, YU.L. *Integrirovanie i differentsirovanie funktsiǐ ermitovykh operatorov, zavisyashchikh ot parametra* (in Russian). Uspehi Mat. Nauk (N.S.) **12** (1957), no. 1(73), 182–186. English transl. *Integration and differentiation of functions of Hermitian operators depending on a parameter*, Amer. Math. Soc. Transl. (Ser. **2**), **16** (1960), pp. 396–400.

[DaKr1]   DALETSKII, YU.L. and KREIN, S.G. *Formuly differentsirovaniya po parametru funktsiǐ ermitovykh operatorov* (Russian). [Formulas of differentiation according to a parameter of functions of Hermitian operators]. Doklady Akad. Nauk SSSR **76** (1951), pp. 13–16.

[DaKr2]   DALETSKII, YU.L. and KREIN, S.G. *Integrirovanie i differentsirovanie funktsiǐ ermitovykh operatorov i prilozhenie k teorii vozmushchiniǐ.* (in Russian). Voronezh. Trudy seminara po funktsional'nomu analizu, vol. **1** (1956), pp. 81 - 105. English transl. *Integration and differentiation of functions of hermitian operators and applications to the theory of perturbations.* Amer. Math. Soc. Transl. (Ser.**2**), **47** (1965), pp. 1 - 30.

[GoKr]   GOHBERG, I.TS. and M.G. KREIN. *Teoriya Vol'terrovykh Operatorov v Gil'bertovom Prostranstve i ee Prilozheniya* (In Russian).Mauka, Moskow 1977. English transl.: *Theory and applications of Volterra operators in Hilbert space.* (Translations of Mathematical Monographs, **24**). American Mathematical Society, Providence, R.I. 1970 x+430 pp.

[Had]   HADAMARD, J. *Théorèmes sur la séries entières.* Acta Math., **22** (1899), pp. 55 - 63.

[Hal1]   HALMOS, P.R. *Finite-Dimensional Vector Spaces.* (Annals of Math. Studies, **7**), Princeton University Press, Princeton, N.J., 1948,. Second edition: Van Norstand, Princeton, N.J., 1958.

[HorR1]   HORN, R.A. *The Hadamard product.* In: Matrix theory and applications (Phoenix, AZ, 1989), pp. 87 - 169, Proc. Sympos. Appl. Math., **40**, Amer. Math. Soc., Providence, RI, 1990.

[HorR2]   HORN, R.A. *Topics in Matrix Analysis*. Cambridge University Press, Cambridge 1991, i – viii, 607 pp.

[Pel]   PELLER, V.V. *Gankelevy operatory v teorii vozmushcheniĭ samosopryazhennykh operatorov* (Russian). Funktsional'n. Analiz i ego prilozh., **19**:2 (1985), pp. 37-51. English transl.: *Hankel operators in the perturbation theory of unitary and self-adjoint operators*. Funct. Anal. and Appl., **19** (1985), pp.111-123.

[Pie1]   PIETSCH, A. *Absolut p-summierende Abbildungen in normierten Räumen* (In German). Studia Math. **28** (1967), pp. 333 - 353.

[Pie2]   PIETSCH, A. *Operator ideals*. (Mathematische Monographien [Mathematical Monographs], **16**). Deutscher Verlag der Wissenschaften, Berlin 1978. 451 pp.; *Operator ideals*. (North-Holland Mathematical Library, **20**.) North-Holland Publishing Co., Amsterdam·New York, 1980. 451 pp.

[Sch:Ges]   SCHUR, I.: *Gesammelte Abhandlungen [Collected Works]. Vol. I, II, III*. Springer-Verlag, Berlin·Heidelberg·New York 1973.

[Sch4]   SCHUR, I.: *Bemerkungen zur Theorie der beschränkten Bilinearformen mit unendlich vielen Veränderlichen. [Remarks on the theory of bounded bilinear forms with infinitely many variables - in German]*. Journ. für reine und angew. Math., **140** (1911), pp. 1 - 28. Reprinted in: [Sch:Ges], Vol. I, pp. 464 - 491.

[Sty]   STYAN, G. *Hadamard products and multivariate statictical analysis*. Linear Algebra and its Applications, **6** (1973), pp. 217 - 240.

[Tit]   TITCHMARSH, E.C. THE THEORY OF FUNCTIONS. The Clarendon Press, Oxford 1932.

5. The Schur Convexity Theorem.

The well known *Hadamard inequality* states that

$$\det H \leq \prod_{1 \leq k \leq n} h_{kk} \tag{5.1}$$

for every non-negative definite Hermitian matrix $H = [h_{jk}]_{1 \leq j,k \leq n}$. (There are many proofs; see, for example, [HoJo], Section 7.8.) In a short but penetrating paper published in 1923, Issai Schur [Sch18] gave a highly effective method for deriving this inequality. However the importance of the paper [Sch18] rests primarily on the ideas which are contained there and by the impact which the paper had on various areas of mathematics, some of which lie very far from the original setting. This paper has generated and continues to generate many fruitful investigations.

Given a Hermitian matrix $H = [h_{jk}]_{1 \leq j,k \leq n}$, it can be reduced to the diagonal form

$$H = U \operatorname{diag}(\omega_1, \ldots, \omega_n) U^*, \tag{5.2}$$

where $\omega_1, \ldots, \omega_n$ are the eigenvalues of the matrix H, and $U = [u_{jk}]_{1 \leq j,k \leq n}$ is a unitary matrix. (If the Hermitian matrix H is real, then the matrix U can be chosen real also, i.e., if H is real and symmetric, then U is orthogonal.) In particular, the equality (5.2) implies that

$$\begin{bmatrix} h_{11} \\ \vdots \\ h_{nn} \end{bmatrix} = \begin{bmatrix} |u_{11}|^2 & \ldots & |u_{1n}|^2 \\ \ldots & \ldots & \ldots \\ |u_{n1}|^2 & \ldots & |u_{nn}|^2 \end{bmatrix} \begin{bmatrix} \omega_1 \\ \vdots \\ \omega_n \end{bmatrix}. \tag{5.3}$$

Since the matrix U in (5.2) is unitary (orthogonal), the matrix $M = [m_{jk}]_{1 \leq j,k \leq n}$, with

$$m_{jk} = |u_{jk}|^2, \tag{5.4}$$

as in (5.3), possesses the properties

$$\begin{aligned}
&\text{i.} \quad m_{jk} \geq 0, && 1 \leq j, k \leq n; \\
&\text{ii.} \quad \sum_{1 \leq k \leq n} m_{jk} = 1, && 1 \leq j \leq n; \\
&\text{iii.} \quad \sum_{1 \leq j \leq n} m_{jk} = 1, && 1 \leq k \leq n.
\end{aligned} \tag{5.5}$$

It turns out to be fruitful to consider linear transformations whose matrices M satisfy the conditions (5.5), without regard to the relations (5.4).

DEFINITION 1. *A matrix $M = [m_{jk}]_{1 \leq j,k \leq n}$ is said to be **doubly-stochastic** if the conditions (5.5) are fulfilled.*

DEFINITION 2 *A matrix $M = [m_{jk}]_{1 \leq j,k \leq n}$ is said to be **ortho-stochastic** if there exists an orthogonal matrix $U = [u_{jk}]_{1 \leq j,k \leq n}$ such that the matrix entries m_{jk} are representable in the form (5.4), i.e., if M is the Schur product of an orthogonal matrix U with itself.*

REMARK 1. *It is clear that every ortho-stochastic matrix is a doubly-stochastic. However, not every doubly-stochastic matrix is an ortho-stochastic. For example[1], the matrix $P = \dfrac{1}{6} \begin{bmatrix} 0 & 3 & 3 \\ 3 & 1 & 2 \\ 3 & 2 & 1 \end{bmatrix}$ is doubly-stochastic, but not ortho-stochastic.*

Many well known elementary inequalities can be put in the form

$$\Phi(\overline{x}, \ldots, \overline{x}) \leq \Phi(x_1, \ldots, x_n), \tag{5.6}$$

---

[1] This example is adopted from [Sch18].

where $\overline{x} = (x_1 + \cdots + x_n)/\text{n}$ and $x_1, \ldots x_n$ lie in a specified set. For example, the inequality

$$\varphi(\overline{x}) \le \big(\varphi(x_1) + \cdots + \varphi(x_n)\big)/n \tag{5.7}$$

for a *convex function* φ of one variable can be written in the form (5.6), with $\Phi(\xi_1, \ldots, \xi_n) = \varphi(\xi_1) + \cdots + \varphi(\xi_n)$.

We recall, that a *real valued function* φ, *defined on a subinterval* (α, β) *of the real axis, is said to be* **convex** *if* φ *is continuous there and the inequality* $\varphi\big((x_1 + x_2)/2\big) \le \big(\varphi(x_1) + \varphi(x_2)\big)/2$ *holds for every* $x_1, x_2 \in (\alpha, \beta)$. The inequality (5.7) is a special case of the so-called

JENSEN INEQUALITY. *Let φ be a convex function on an interval (α, β), let $x_1, \ldots, x_n$ be points in the interval (α, β), and let the numbers $\lambda_1, \ldots, \lambda_n$ satisfy the conditions*

$$
\begin{array}{lll}
\text{i.} & \lambda_k \ge 0, & 1 \le k \le n; \\
\text{ii.} & \sum_{1 \le k \le n} \lambda_k = 1.
\end{array}
\tag{5.8}
$$

Then

$$\varphi(\lambda_1 x_1 + \cdots + \lambda_n x_n) \le \lambda_1 \varphi(x_1) + \cdots + \lambda_n \varphi(x_n). \tag{5.9}$$

The value $\overline{x} = (x_1 + \ldots + x_n)/\text{n}$ that appears in (5.7), the so called *arithmetic mean* of the values $x_1, \ldots, x_n$, is the most commonly used average value for $x_1, \ldots, x_n$. The value $\lambda_1 x_1 + \ldots + \lambda_n x_n$ that appears in (5.9), the so-called *weighted arithmetic mean*, is a more general average value for $x_1, \ldots, x_n$.

In [Sch18], doubly-stochastic matrices $M = [m_{jk}]_{1 \le j,k \le n}$ are used to construct an *average sequence* $y_1, \ldots, y_n$ from a given sequence of real or complex numbers $x_1, \ldots, x_n$ by the *averaging rule*

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} m_{11} & \ldots & m_{1n} \\ \ldots & \ldots & \ldots \\ m_{n1} & \ldots & m_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = M\mathbf{x}. \tag{5.10}$$

It is intuitively clear that the sequence of "averaged" values $\{y_k\}$ is "less spread out" than the original sequence $\{x_k\}$. In [Sch18], inequalities of the form

$$\Phi(y_1, \ldots, y_n) \le \Phi(x_1, \ldots, x_n), \tag{5.11}$$

are considered for points $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$ in the domain of definition of the function Φ that are related by a doubly stochastic matrix $M = [m_{jk}]_{1 \le j,k \le n}$ by means of the averaging procedure $\mathbf{y} = M\mathbf{x}$ given in (5.10). In particular, the inequality (5.11) is established there for functions Φ of the form $\Phi(\xi_1, \ldots, \xi_n) = \varphi(\xi_1) + \cdots + \varphi(\xi_n)$:

THEOREM I. *Let φ be a convex function defined on a subinterval (α, β) of the real axis, let $x_1, \ldots, x_n$ be arbitrary numbers from (α, β), let $M = [m_{jk}]_{1 \le j,k \le n}$ be a doubly stochastic*

33

matrix and let the numbers $y_1, \ldots, y_n$ be obtained from the averaging procedure $\mathbf{y} = M\mathbf{x}$. Then

$$\varphi(y_1) + \cdots + \varphi(y_n) \leq \varphi(x_1) + \cdots + \varphi(x_n). \tag{5.12}$$

PROOF of Theorem I. In view of the conditions (5.5.i) and (5.5.ii), Jensen's inequality is applicable with $\lambda_k = m_{jk}, k = 1, \ldots, n$, and implies that

$$m_{j1}\varphi(x_1) + \cdots + m_{jn}\varphi(x_n) \leq \varphi(m_{j1}x_1 + \cdots + m_{jn}x_n) = \varphi(y_j).$$

The desired conclusion is now obtained by summing the last inequality over j from $1, \ldots, n$ and invoking the condition (5.5.iii). $\qquad\qquad\qquad\qquad\qquad\qquad\square$

The preceding theorem appears as Theorem V in [Sch18] and is used there to derive the (Hadamard) inequality

$$\prod_{1 \leq k \leq n} \omega_k \leq \prod_{1 \leq k \leq n} h_{kk}$$

for a positive definite Hermitian matrix $H = \left[h_{jk}\right]_{1 \leq j,k \leq n}$ with eigenvalues $\omega_1, \ldots, \omega_n$. The latter is equivalent to the inequality

$$\sum_{1 \leq k \leq n} (-\log h_{kk}) \leq \sum_{1 \leq k \leq n} (-\log \omega_k), \tag{5.13}$$

which is of the form (5.12), with the convex function $\varphi(\xi) = -\log \xi$. In this case, the averaging doubly-stochastic matrix $M = \left[m_{jk}\right]_{1 \leq j,k \leq n}$ is the ortho-stochastic one, with entries m_{jk} of the form (5.4), as in (5.3).

In [Sch18], functions Φ of several variables for which inequalities of the form (5.11) hold are also considered.

DEFINITION 3.   *A function Φ of n variables $x_1, \ldots, x_n$. is said to be **S-convex** (i.e., convex in the sense of Schur) if for every doubly-stochastic matrix M and every pair of points $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = M\mathbf{x}$ in the domain of Φ, the inequality (5.11) holds. The function Φ is said to be **S-concave** if the opposite inequality holds, i.e., if $\Phi(x_1, \ldots, x_n) \leq \Phi(y_1, \ldots, y_n)$, holds for every pair of points $\mathbf{x}$ and $\mathbf{y} = M\mathbf{x}$ in the domain of Φ. A function Φ is S-concave if and only if the function $-\Phi$ is S-convex.*

Let π be a permutation of the set $\{1, \ldots, n\}$. Then the corresponding operator on $\mathbb{R}^n$ that permutes coordinates according to the rule $(x_1, \ldots, x_n) \to (x_{\pi(1)}, \ldots, x_{\pi(n)})$ is linear. Its matrix P_π with respect to the standard basis in $\mathbb{R}^n$ is termed a permutation matrix and is of the form

$$P_\pi = \left[(p_\pi)_{jk}\right]_{1 \leq j,k \leq n}, \quad \text{where, for } k = 1, \ldots, n, \; (p_\pi)_{jk} = \begin{cases} 1, & \text{if } j = \pi(k); \\ 0, & \text{if } j \neq \pi(k). \end{cases} \tag{5.14}$$

There are $n!$ permutation matrices of size $n \times n$. Every permutation matrix is a doubly-stochastic one. The inverse of a permutation matrix is a permutation matrix as well, and hence it is also doubly-stochastic. Therefore,

Every S-convex function Φ of n variables is a symmetric function:

$$\Phi(x_1, \, \ldots \, , x_n) \equiv \Phi(x_{\pi(1)}, \, \ldots \, , x_{\pi(n)}), \text{ for every permutation } \pi. \qquad (5.15)$$

THEOREM II. *Let Φ be a S-convex function of n variables, $n \geq 2$, and let all its partial derivatives of the first order exist and be continuous. Then the function Φ satisfies the condition*

$$\frac{\partial \Phi}{\partial x_1}(x_1, x_2, \, \ldots \, , x_n) - \frac{\partial \Phi}{\partial x_2}(x_1, x_2, \, \ldots \, , x_n) \geq 0, \text{ if } x_1 > x_2. \qquad (5.16)$$

This theorem provides a necessary condition for a symmetric function Φ be S-convex. It appears as Theorem I in [Sch18]. Theorem II in [Sch18] also contains a sufficient condition for a symmetric function Φ be S-convex.

THEOREM III. *Let Φ be a symmetric function of n variables, $n \geq 2$, that satisfies the condition*

$$\left(\frac{\partial^2 \Phi}{\partial x_1^2} + \frac{\partial^2 \Phi}{\partial x_2^2} - 2 \frac{\partial^2 \Phi}{\partial x_1 \partial x_2} \right)(x_1, x_2, \, \ldots \, , x_n) \geq 0 \quad \text{for all} \quad x_1, x_2, \, \ldots \, , x_n. \qquad (5.17)$$

Then the function Φ is S-convex.

However, A. Ostrowski showed that condition (5.16) is both necessary and *sufficient* for a symmetric function Φ to be S-convex; see Theorem VIII in [Ostr]. The reasoning in [Ostr] is based essentially on the the reasoning in [Sch18], but is more precise.

In [Sch18] it is shown that *the elementary symmetric functions* $c_k(x_1, \, \ldots \, , x_n)$, $k = 1, \, \ldots \, , n$, *are S-concave,* and that the functions $\Phi_k(x_1, \, \ldots \, , x_n) = \dfrac{c_{k+1}(x_1, \, \ldots \, , x_n)}{c_k(x_1, \, \ldots \, , x_n)}$, $k = 1, \, \ldots \, , n-1$, *are S-concave.*

To this point, we have reviewed almost all the main results of the short paper [Sch18]. The significance of this paper is not confined to these results, important as they are, but rests primarily on the fact that linear transformations with doubly-stochastic matrices were introduced there. This paper attracted the attention of mathematicians to doubly-stochastic matrices. (In [BeBe] the term "Schur transformation" is used for linear transformations with such matrices; see [BeBe], Chapter I, § 29.) Schur himself did not use the term doubly-stochastic matrix. He just referred to "a matrix M that satisfies the conditions (5.5)." The term "doubly-stochastic matrix" seems to have appeared first in

the first edition of the book [Fel] by W. Feller, in 1950 [2].

Many results were influenced by the paper [Sch18]. We shall begin with the theorems of Hardy-Littlewood-Polya and Birkhoff.

To formulate the Hardy-Littlewood-Polya Theorem, we have to introduce the notion of majorization. Let $\xi_1, \xi_2, \ldots, \xi_n$ be a sequence of *real* numbers. By $\xi_1^*, \xi_2^*, \ldots, \xi_n^*$ we denote the reaarangement of this sequence in non-increasing order:

$$\xi_1^* \geq \xi_2^* \geq \ldots \geq \xi_n^*, \;\; \xi_k^* = \xi_{\pi(k)} \;\text{ for some permutation } \pi \text{ of the set of indices } 1, 2 \ldots, n.$$

DEFINITION 4. *Let $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ and $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$ be two sequences of real numbers. Then we say that the sequence $\mathbf{y}$ is majorized by the sequence $\mathbf{x}$ (or that the sequence $\mathbf{x}$ majorizes the sequence $\mathbf{y}$) if the following conditions are satisfied:*

$$
\begin{aligned}
y_1^* + y_2^* + \ldots + y_k^* \;\; &\leq \;\; x_1^* + x_2^* + \ldots + x_k^*, \qquad (k = 1, 2, \ldots, n-1); \\
y_1^* + y_2^* + \ldots + y_{n-1}^* + y_n^* \;\; &= \;\; x_1^* + x_2^* + \ldots + x_{n-1}^* + x_n^*.
\end{aligned}
$$
(5.18)

*A relation of the form (5.18) is said to be **a majorization** relation and is denoted by the symbol*

$$\boldsymbol{y} \prec \boldsymbol{x}, \;\; or \;\; (y_1, y_2, \ldots, y_n) \prec (x_1, x_2, \ldots, x_n), \qquad (5.19)$$

The relations (5.18) were considered by R.F. Muirhed [Muir] and by M.O. Lorenz [Lor] in the beginning of 20th century. Muirhead introduced these relations (with integer x_k, y_k only) to study inequalities for homogeneous symmetric functions (Muirhead's result is also presented in [HLP], Chapter II, sec. **2.18**). Lorenz used the relations (5.18) to describe the non-uniformity of the distribution of wealth in a population. However, the notation (5.19) and the term "majorization" were introduced by G.H. Hardy, J.W. Littlewood and G. Polya in 1934; see [HLP], Sec. **2.18**. Chapter II of the book [HLP], in which majorization is introduced and discussed, contains a number of references to private communications by Schur.

THEOREM (G.H. Hardy, J.W. Littlewood and G. Polya, [HLP], sec. **2.20**)
I. *Let $\boldsymbol{x} = (x_1, \ldots, x_n)$ and $\boldsymbol{y} = (y_1, \ldots, y_n)$ be two sequences of real numbers and let matrix M be a doubly-stochastic matrix such that $\mathbf{x} = M\mathbf{y}$. Then $\boldsymbol{y} \prec \boldsymbol{x}$.*
II. *Let $\boldsymbol{x} = (x_1, \ldots, x_n)$, and $\boldsymbol{y} = (y_1, \ldots, y_n)$ be two sequences of real numbers such that $\boldsymbol{y} \prec \boldsymbol{x}$. Then there exists a doubly stochastic matrix M such that $\boldsymbol{x} = M\boldsymbol{y}$. (In general such a matrix M is not unique.)*

---

[2] However, the term "stochastic matrix" was used as early as 1931 in [Rom1] (see also [Rom2]) for matrices satisfying the conditions (5.5.i) and (5.5.ii) only (but not necessarily the condition (5.5.iii)). Such matrices play a crucial role in the theory of Markov chains.

Part II of this theorem and the first cited theorem of Schur (which appears as Theorem I in this section) implies the following result:

THEOREM I′. *Let a sequence $\boldsymbol{y} = (y_1, \ldots, y_n)$ be majorized by a sequence $\boldsymbol{x} = (x_1, \ldots, x_n)$, let $x_k, y_k \in (\alpha, \beta) \subset \mathbb{R}$ for $k = 1, \ldots, n)$, and let φ be a convex function on the interval (α, β). Then the inequality (5.12) holds.*

It turns out that the converse statement is true ([HLP1], Theorem 8; [HLP], Theorem **108**):

Let $x_k, y_k \in (\alpha, \beta)$ for $k = 1, \ldots, n$ and ssume that the inequality (5.12) holds for every function φ which is convex on the interval (α, β). Then $\boldsymbol{x} = M\boldsymbol{y}$ for some doubly-stochastic matrix M.

This means that Schur's result (which appears as Theorem I in this section) is sharp in some sense.

In [GoKr1], Chapt. II, Lemma 3.5, a very elementary proof of the following fact is presented: *Let Φ be a symmetric function of n variables which has continuous derivatives of the first order. Assume that the condition (5.16) is satisfied. If a sequence $\boldsymbol{x} = (x_1, \ldots, x_n)$ of real numbers majorizes a sequence $\boldsymbol{y} = (y_1, \ldots, y_n)$, then the inequality (5.11) holds.*

The last result combined with the Hardy-Littlewood-Polya theorem that was discussed earlier yields an independent proof of the fact that *a symmetric function Φ that satisfies the condition (5.16) is S-convex.*

The theorem by G. Birkhoff sheds light on geometric aspects of majorization and Schur averaging. It is clear that the set of all doubly-stochastic matrices is compact and convex. Therefore, it is of interest to find the extreme points of this set. It is clear that permutation matrices are doubly-stochastic and that they are extreme points. It turns out that they are the **only** extreme points.

THEOREM (G. Birkhoff). *Every doubly-stochastic matrix $M = \left[m_{jk}\right]_{1 \leq j,k \leq n}$ is representable as a convex combination of permutation matrices:*

$$M = \sum_{\pi \in \boldsymbol{S}_n} \lambda_\pi P_\pi \,, \tag{5.20}$$

where π runs over the set $\boldsymbol{S}_n$ of all permutations of the set $\{1, \ldots, n\}$, P_π are the corresponding permutation matrices (5.14), and the coefficients $\lambda_\pi = \lambda_\pi(M)$ satisfy the conditions

$$\lambda_\pi \geq 0 \quad (\forall \, \pi \in \boldsymbol{S}_n), \quad \sum_{\pi \in \boldsymbol{S}_n} \lambda_\pi = 1 \,. \tag{5.21}$$

REMARK 2. *In general, the coefficients $\lambda_\pi(M)$ in the representation (5.20) are not*

uniquely determined from the matrix M.

This theorem was formulated and proved in 1946 in the paper [Birk1]. (This formulation also appeared in Example 4* in [Birk2], p.266.) The original proof due to Birkhoff is based on a theorem by Ph. Hall on representatives of subsets, [HalP]. (The latter theorem can also be found in [HalM], sec.**5.1**). G.B. Dantzig [Dan] gives an algorithm for solving a transportation problem, the solution of which leads to Birkhoff's theorem. An independent proof of Birkhoff's theorem was given by J. von Neumann [NeuJ1] in the setting of game theory. "Combinatorial" proofs of Birkhoff's theorem (based on Ph.Hall's theorem), are presented in the books of M. Hall [HalM] (see Theorem 5.1.9), and C. Berge [Ber] (see Theorem 11 in Chapt. 10). A geometric proof (based on a direct investigation of extreme points) is presented in [HoJo], Theorem **8.7.1**. Two different proofs of Birkhoff's theorem are presented in [MaOl], Chapt.2, Sect. **F**. The paper [Mir] is a good survey of doubly-stochastic matrices. In particular, it contains a proof of Birkhoff's theorem. See also the problem book by I.M. Glazman and Yu.I. Lyubich [GlLy], Ch. 7, §4, where Birkhoff's theorem is presented in problem form.

Let $\boldsymbol{x} = (x_1, \dots, x_n)$ be a sequence of real numbers and, for a permutation π of the set $\{1, \dots, n\}$, let $\boldsymbol{x}_\pi = (x_{\pi(1)}, \dots, x_{\pi(n)})$. (Thus, for given $\boldsymbol{x}$ there are $n!$ sequences $\boldsymbol{x}_\pi$, some of which can coincide.) We consider these sequences as vectors in $\mathbb{R}^n$. Let $\boldsymbol{C_x}$ denote the convex hull of all the vectors $\boldsymbol{x}_\pi$ where $\pi \in \boldsymbol{S}_n$.

THEOREM (R. Rado) *Let $\boldsymbol{x} = (x_1, \dots, x_n)$ and $\boldsymbol{y} = (y_1, \dots, y_n)$ be two sequences of real numbers. Then*

$$\boldsymbol{y} \in \boldsymbol{C_x} \quad \Longleftrightarrow \quad \boldsymbol{y} \prec \boldsymbol{x}.$$

PROOF. The implication $\Rightarrow$ is easy. The converse can be obtained by combining the cited theorems of Hardy-Littlewood-Polya and Birkhoff. $\square$

This theorem seems to have been established first by R. Rado [Rad]. His proof was based on a theorem on the separation of convex sets by hyperplanes. A. Horn ([HorA1], Theorem 2) observed it can also be obtained by combining the results of Hardy-Littlewood-Polya and Birkhoff that were cited earlier. A short proof of Rado's theorem, which does not use the Birkhoff theorem, can be found in [Mark] (see Theorem 1.1).

The circle of ideas related to Schur averaging, majorization and Birkhoff's theorem is well represented in the literature. The whole book [MaOl] (of more than 550 pages) is dedicated to this circle. It includes applications to combinatorial analysis, matrix theory, numerical analysis and statistics. The books [ArnB] and [PPT] are also relevant. There are generalizations of Birkhoff's theorem to the infinite dimensional case, see [Mir] and [NeuA].

One generalization of Birkhoff's theorem leads to an interpolation theorem for linear operators. Let B be the linear space $\mathbb{R}^n$ provided with a norm $\| \, . \, \|_B$ such that $\|\boldsymbol{x}\|_B =$

$\|\boldsymbol{x}_\pi\|_{\boldsymbol{B}}$ for every $\boldsymbol{x} \in \mathbb{R}^n$ and for every permutation $\pi \in \boldsymbol{\mathcal{S}}_n$, where, as usual, $\boldsymbol{x}_\pi = (x_{\pi(1)}, \ldots, x_{\pi(n)})$. In other words, this property of the norm $\|.\|_B$ can be expressed as $\|P_\pi\|_{B \to B} = 1$ for every permutation $\pi \in \boldsymbol{\mathcal{S}}_n$ where *the permutation operator* P_π is defined by the permutation matrix P_π, (5.14), in the natural basis of the space $\mathbb{R}^n$. A norm $\|.\|_B$ with this property is said to be *a symmetric norm*. A Banach space B with a symmetric norm is said to be *a symmetric Banach space*.

Let an operator A in the space $\mathbb{R}^n$ be defined by its matrix $A = \left[a_{jk}\right]_{1 \le k \le n}$ in the natural basis of the space $\mathbb{R}^n$ and assume that it satisfies the norm estimates

$$\|A\|_{l^1 \to l^1} \le 1 \quad \text{and} \quad \|A\|_{l^\infty \to l^\infty} \le 1 . \tag{5.22}$$

Then, as noted earlier in Section 4,

$$\sum_{1 \le j \le n} |a_{jk}| \le 1, \quad 1 \le k \le n \quad \text{and} \quad \sum_{1 \le k \le n} |a_{jk}| \le 1, \quad 1 \le j \le n . \tag{5.23}$$

According to one generalization of Birkhoff's theorem, a matrix A satisfying the conditions (5.23) admits a representation of the form

$$A = \sum_{\pi \in \boldsymbol{\mathcal{S}}_n} \lambda_\pi P_\pi$$

where the λ_π are *real* (not necessarily non-negative) numbers satisfying the conditions

$$\sum_{\pi \in \boldsymbol{\mathcal{S}}_n} |\lambda_\pi| \le 1 .$$

Therefore, since $\|P_\pi\|_{B \to B} = 1$, the operator A must be a contraction in this norm:

$$\|A\|_{B \to B} \le 1 . \tag{5.24}$$

Thus, the following result holds:

THEOREM (Interpolation theorem for symmetric Banach spaces). *Let an operator A acting in the space $\mathbb{R}^n$ be a contraction in the l^1 and l^∞ norms, i.e., let the estimates (5.22) hold. Then the operator A is a contraction in every symmetric norm $\| \cdot \|_B$ on $\mathbb{R}^n$, i.e., the estimate (5.24) holds.*

Here we presented the simplest interpolation result for symmetric spaces. A more advanced result can be found in [Mit]. Thus, the development of ideas initiated by Schur leads to interpolation theorems for Banach spaces with symmetric norms.

The last topic which we discuss here is the Schur-Horn convexity theorem. A. Horn ([HorA1], Theorem 4) obtained the following strengthening of the second part of the Hardy-Littlewood-Polya theorem:

THEOREM (A. Horn). *Let* $\boldsymbol{x} = (x_1, \ldots, x_n$ *and* $\boldsymbol{y} = (y_1, \ldots, y_n)$ *be any two points in* $\mathbb{R}^n$ *such that* $\boldsymbol{y} \prec \boldsymbol{x}$. *Then there exists an* **ortho**-*stochastic matrix* M *such that* $\boldsymbol{y} = M\boldsymbol{x}$.

The following result is a direct consequence of the cited theorems of Rado and A. Horn:

Given $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$, *the following two sets are coincide:*

1. *The set* $\boldsymbol{\mathcal{C}_x} \overset{\text{def}}{=}$ *the convex hull of the family of vectors* $\{\boldsymbol{x}_\pi\}_{\pi \in \boldsymbol{\mathcal{S}}_n}$.
2. *The set* $\{M\boldsymbol{x}\}$, *where* M *runs over the set of all* **ortho**-*stochastic matrices*.

In view of the relations (5.2) and (5.3), the last statement can be reformulated in terms of eigenvalues and diagonal entries: Let us associate with every *real symmetric* $n \times n$ matrix $H = \big[h_{jk}\big]_{1 \le j,k \le n}$ the n-tuple $\boldsymbol{h}(H) = (h_{11}, \ldots, h_{nn})$ of its diagonal entries and the n-tuple $\boldsymbol{\omega}(H) = (\omega_1(H), \ldots, \omega_n(H))$ of its eigenvalues arranged in non-increasing order: $\omega_1(H) \ge, \ldots, \ge \omega_n(H)$. We consider these n-tuples as vectors in $\mathbb{R}^n$. Given an n-tuple $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)$ of real numbers, arranged in non-increasing order: $\omega_1 \ge \ldots, \omega_n$, let

$$\boldsymbol{\mathcal{H}_\omega} = \big\{H : \ H \text{ is real symmetric and } \boldsymbol{\omega}(H) = \boldsymbol{\omega}\big\}.$$

THEOREM (Schur-Horn convexity theorem). *Given an* n-*tuple* $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)$ *of real numbers:* $\omega_1 \ge \ldots, \ge \omega_n$, *the set* $\big\{\boldsymbol{h}(H)\big\}_{H \in \boldsymbol{\mathcal{H}_\omega}}$ *of all "diagonals" of matrices from* $\boldsymbol{\mathcal{H}_\omega}$ *is convex. Moreover,*

$$\big\{\boldsymbol{h}(H)\big\}_{H \in \boldsymbol{\mathcal{H}_\omega}} = \boldsymbol{\mathcal{C}_\omega}, \tag{5.25}$$

where $\boldsymbol{\mathcal{C}_\omega}$ *is the convex hull of the family of* $n!$ *vectors* $\boldsymbol{\omega}_\pi = (\omega_{\pi(1)}, \ldots, \omega_{\pi(n)})$, *as* π *runs over the set* $\boldsymbol{\mathcal{C}}_n$ *of all permutations of the set* $\{1, \ldots, n\}$:

$$\boldsymbol{\mathcal{C}_\omega} = \text{Conv}\big\{\boldsymbol{\omega}_\pi : \pi \in \boldsymbol{\mathcal{S}}_n\big\}. \tag{5.26}$$

Schur himself established the formula

$$\big\{\boldsymbol{h}(H)\big\}_{H \in \boldsymbol{\mathcal{H}_\omega}} = \{M\boldsymbol{\omega} : M \ \text{ is ortho-stochastic}\}.$$

He did not described the set on the right geometrically as a convex hull. The term "convex set" does not appear in the paper [Sch18]) at all. The "Schur-Horn convexity theorem" appeared only in the paper by A. Horn [HorA2] (which used in an essential way the cited results by Hardy-Littlewood-Polya and Birkhoff.) However, the influence of Issai Schur on the area was so great that the term "Schur-Horn convexity theorem" is now common.

In the last thirty years, the Schur-Horn convexity theorem has been generalized significantly. In 1973 (fifty years after the publication of [Sch18]) B. Kostant published a seminal paper [Kos] in which he interpreted the Schur-Horn result as a property af adjoint orbits of the unitary group and generalized it to arbitrary compact Lie groups. More precisely, he proved (see especially [Kos], sect. 8) that for an element x in a maximal abelian subspace $\mathfrak{t}$ in the Lie algebra $\mathfrak{k}$ of a compact Lie group K one has

$$pr_{\mathfrak{t}}(\text{Ad}\, K.x) = \text{Conv}\, \mathcal{W}.x \, ,$$

where $pr_{\mathfrak{t}} : \mathfrak{k} \to \mathfrak{t}$ is the orthogonal projection (with respect to the Killing form) and $\mathcal{W}$ is the Weyl group associated with the pair $(\mathfrak{k}_{\mathbb{C}}, \mathfrak{t}_{\mathbb{C}})$. Subsequently, M.F. Atiyah [Ati] and, independently, V. Guillemin and S. Sternberg [GuSt1], [GuSt2] gave an interpretation of Kostant's theorem as a special case of a theorem on the image of the momentum map of a Hamiltonian torus action. Atiyah's proofs depend on some ideas from Morse theory. Subsequently, the results of Kostant, Atiyah, Guillemin and Sternberg were extended to the setting of symmetric spaces. See, for example, the paper [HNP], where more references can be found, the paper [BFR] and the book [HiOl], sections 4.3 and 5.5.

In yet another direction, the relevance of doubly-stochastic matrices and Schur averaging to operator algebras and quantum physics is discussed in the book [AlU].

Thus, once again a relatively short paper of Issai Schur is seen to have had significant influence on the development of a number of diverse areas of mathematics. In particular, [Sch18] paved the way to important results in matrix theory, statistics, the theory of Lie groups and symmetric spaces, symplectic geometry and Hamiltonian mechanics. Many of these areas are very far from the original setting.

REFERENCES

[AlU]   ALBERTI, P.M. and A. UHLMANN *Stochasticity and partial order. Doubly stochastic maps and unitary mixing.* Mathematics and its Applications, **9**. D. Reidel, Dordrecht-Boston, 1982; Mathematische Monographien [Mathematical Monographs], **18**. VEB Deutscher Verlag der Wissenschaften, Berlin, 1981.

[ArnB]   ARNOLD, B.C. *Majorization and Lorenz order: a brief introduction.* (Lect. Notes in Statist., **43**). Springer-Verlag, Berlin 1987, vi+122 pp.

[Ati]   ATIYAH, M.F. *Convexity and commuting Hamiltonians.* Bull. London. Math. Soc., **14**:1 (1982), pp. 1-15.

[BeBe]   BECKENBACH, E.F. and R. BELLMAN. *Inequalities.* (Ergebnisse der Mathem. und ihrer Grenzgeb., Neue Folge, **30**), Springer-Verlag, Berlin·Göttingen·Heidelberg 1961.

[Ber]   BERGE, C. *Théorie des Graphes et ses Applications.* Dunod, Paris 1958 (French). English transl.: *The Theory of Graphs and its Applications.* Methuen & Co, London 1962.

[Birk1]   BIRKHOFF, G. *Tres notas sobre el algebra lineal [Three observations on linear algebra - in Spanish].* Universidad National de Tucumán, Revista, Serie **A**, **5** (1946), pp. 147-151.

[Birk2]   BIRKHOFF, G. *Lattice Theory.* (Amer. Math. Soc. Coll. Publ. **25**.) Amer. Math. Soc., Providence, RI, 1948.

[BFR]   BLOCH, A. M., H. FLASCHKA and T.A. RATIU. *Schur-Horn-Kostant convexity theorem for the diffeomorphism group of the annulus.* Invent. Math. **113**:3 (1993), pp. 511–529.

[Dan]   DANTZIG, G.B. *Application of the symplex method to a transportation problem.* Chapter XXIII in: *Activity Analysis of Production and Allocation.* (KOOPMANS, T.C. - ed.), Wiley, New York 1951.

[Fel]   FELLER, W. *An Introduction to Probability Theory and Its Applications,* 1st ed., Vol. **1**, Wiley, New York 1950.

[GlLy]   GLAZMAN, I.M. and YU.I. LYUBICH. *Konechnomernyĭ lineĭnyĭ analiz v zadachakh* (Russian). Nauka, Moskow 1969, 475 pp. English transl.: *Finite-dimensional linear analysis: a systematic presentation in problem form.* The M.I.T. Press, Cambridge, MA·London, 1974. xvi+520 pp. French transl.: *Analyse linéaire dans les espaces de dimensions finies: manuel en problèmes.* Mir, Moscow, 1972. 400 pp.

[GoKr1]   GOHBERG, I.Ts. and M.G. KREIN. *Vvedenie v Teoriyu Lineĭnykh Nesamoso-pryazhennykh Operatorov* (Russian). Nauka, Moskow 1965, 448 pp. English Transl.: *Introduction to the Theory of Linear Non-Selfadjoint Operators.* (Transl. Math. Monogr. **18**). Amer. Math. Soc., Providence, R.I., 1969.

[GuSt1]   GUILLEMIN, V. and S. STERNBERG. *Convexity properties of the moment mapping.* Invent. Math. **67** (1982), pp. 491-513.

[GuSt2]   GUILLEMIN, V. and S. STERNBERG. *Convexity properties of the moment mapping. II.* Invent. Math. **77** 1984), pp. 533-546.

[HalM]   HALL, M., JUNIOR *Combinatorial Theory.* Blaisdell Publishing Co., Waltham MA ·To- ronto ·London 1967, x+310 pp.

[HalP]   HALL, PH. *On representatives of subsets.* Journ. Lond. Math. Soc, **10** (1935), pp. 26-30.

[Hal2]   HALMOS, P.R. *Bounded Integral Operators in L^2 Spaces.* Springer Verlag, Berlin 1978.

[Har:Col] G.H. HARDY. *Collected Papers. Vol. 2.* Clarendon Press, Oxford 1967.

[HiOl]   HILGERT, J. and G. ÒLAFSSON. *Causal Symmetric Spaces. Geometry and Harmonic Analysis.* (Perspectives in Mathematics, **18**). Academic Press, San Diego·London 1997, i-ivx+286 pp.

[HLP]   HARDY, G.H., J.E. LITTLEWOOD, and G. PÓLYA. *Inequalities.* 1st ed., 2nd ed. Cambridge Univ. Press, London·New York 1934, 1952.

[HLP1]   HARDY, G.H., J.E. LITTLEWOOD and G. PÓLYA. *Some simple inequalities satisfied by convex functions.* Messenger of Mathematics, **58**, pp. 145-152. Reprinted in [Har:Col], pp. 500-508.

[HNP]   Hilgert, J., K.-H. Neeb, and W. Plank. *Symplectic convexity theorems and coadjoint orbits.* Compos. Math., **94** (1994), pp.129-180.

[HoJo]   Horn, R.A., and Ch.R. Johnson. *Matrix Analysis.* Cambrigde University Press, Cambridge·London·New York 1986.

[HorA1]   Horn, A. *Doubly stochastic matrices and the diagonal of the rotation matrix.* Amer. J. Math. **76** (1954), 620 - 630.

[HorA2]   Horn, A. *On the eigenvalues of a matrix with prescribed singular values.* Proc. Amer. Math. Soc., **5**:1 (1954), pp.4-7.

[Kos]   Kostant, B. *On convexity, the Weyl group and the Iwasawa decomposition.* Ann. Sci. École Norm. Sup. (Ser. 4), **6** (1973), pp. 413-455.

[Lor]   Lorenz, M.O. *Methods of measuring concentrations of wealth.* Journ. Amer. Statist. Assoc., **9** (1905), pp. 209-219.

[MaOl]   Marshall, A.W. and I. Olkin. *Inequalities: Majorization and Its Applications.* Academic Press, New York·London·Toronto 1979.

[Mark]   Markus, A.S. *Sobstvennye i singulyarnye chisla summy i proizvedeniya lineǐnykh operatorov* (Russian). Uspekhi Matem. Nauk, **19**:4 (1964), pp. 93-123. English transl.: *The eigen- and singular values of the sum and product of linear operators.* Russian Math. Surveys, **19** (1964), pp. 91-120.

[Mir]   Mirsky, L. *Results and problems in the theory of doubly-stochastic matrices.* Zeitschr. fúr die Wahrscheinlichkeitstheorie und Verw. Gebiete, **1** (1962/1963), 319-334.

[Mit]   Mityagin, B.S. *Interpolyatsionnaya teorema dlya modulyarnykh prostranstv* (Russian). Matem. Sbornik (N.S.), **66** (1965), pp. 473-482. English transl.: *An interpolation theorem for modular spaces.* In: *Interpolation spaces and allied topics in analysis (Lund, 1983)*, Lecture Notes in Math., **1070**, Springer, Berlin, 1984, pp. 10-23.

[Muir]   Muirhead, R.F. *Some methods applicable to identities and inequalities of symmetric algebraic functions of n letters.* Proc. Edinburgh Math. Soc., **21** (1903), pp.144-157.

[NeuA]   Neumann, A. *An infinite dimensional version of the Schur-Horn convexity theorem.* Journ. of Funct. Anal., **161** (1999), pp. 418 - 451.

[NeuJ1]   Neumann, John von. *A certain zero-sum two-person game equivalent to the optimal assignement problem.* In: *Contributions to the Theory of Games. Vol.* II (Kuhn, H.W. and A.W. Tucker - editors). (Annals Math. Study **28**), Princeton Univ. Press, Princeton 1953, pp. 5-12. Reprinted in: [NeuJ2], pp. 44-49.

[NeuJ2]   Neumann, John von. *Collected Works. Vol.VI: Theory of games, astrophysics, hydrodynamics and meteorology.* General editor: A.H. Taub, Pergamon Press, Oxford·London·New York·Paris: 1963, x+ 538 pp.

[Ostr] OSTROWSKI, A. *Sur quelques applications des fonctions convexes and concaves au sens de I. Schur* (French). Journ. de Mathématiques Pures et Appliquées, **31** (1952), pp. 253-292.

[PPT] PEČARIĆ, J.E., F. PROSCHAN and Y.L. TONG. *Convex functions, partial ordering, and statistical applications.* (Mathematics in Science and Engineering, **187**). Academic Press, Boston, MA, 1992, xiv+467 pp.

[Rad] RADO, R. *An inequality.* Journ. Lond. Math. Soc., **27** (1952), pp. 1-6.

[Rom1] ROMANOVSKY, V. *Sur les zèros des matrices stochastiques.* Compt. Rend. Acad. Sci. Paris, **192** (1931), pp. 266-269.

[Rom2] ROMANOVSKY, V. *Recherches sur les chaines de Markoff.* Acta Math., **66** (1930), pp. 147-251.

[Sch18] SCHUR, I.: *Über eine Klasse von Mittelbildungen mit Anwendungen auf die Determinantentheorie.* [*On a class of averaging mappings with applications to the theory of determinants* - in German]. Sitzungsberichte der Berliner Mathematischen Gesellschaft, **22** (1923), pp. 9 - 20. Reprinted in: [Sch: Ges], Vol. II, pp. 416 - 427.

[Sch: Ges] SCHUR, I.: *Gesammelte Abhandlungen [Collected Works]. Vol. I, II, III.* Springer-Verlag, Berlin·Heidelberg·New York, 1973.

6. Inequalities between the eigenvalues and the singular values of a linear operator.

Let $A = \left[a_{jk}\right]_{1 \leq j,k \leq n}$ be an $n \times n$ matrix with eigenvalues $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$. In Theorem II of [Sch2], Schur proved the inequality

$$\sum_{\ell=1}^{n} |\lambda_\ell|^2 \leq \sum_{j,k=1}^{n} |a_{jk}|^2. \tag{6.1}$$

Schur's proof was based on Theorem I of that paper, in which he established the fundamental fact that every square matrix A with complex entries is unitarily equivalent to an upper triangular matrix, i.e., there exists a unitary matrix U such that

$$T = U^*AU = U^{-1}AU \tag{6.2}$$

is upper triangular: $\mathsf{t}_{jk} = 0$ for $j > k$. Therefore, the set of eigenvalues of the matrix A is equal to the set of eigenvalues of the matrix T, which in turn is equal to the set of diagonal entries of T. Thus,

$$\sum_{\ell=1}^{n} |\lambda_\ell|^2 = \sum_{j=1}^{n} |t_{jj}|^2 \leq \sum_{j,k=1}^{n} |t_{jk}|^2 = \operatorname{trace} T^*T = \operatorname{trace} A^*A = \sum_{j,k=1}^{n} |a_{jk}|^2.$$

44

Apart from its use in the proof of the inequality (6.1), Theorem I serves as a model for some important constructions in operator theory that will be discussed below.

In [Sch2], Schur used (6.1) to obtain simple proofs of the estimates

$$|\lambda_l| \leq n \cdot \max_{1 \leq j,k \leq n} |a_{jk}| \qquad (1 \leq l \leq n) \tag{6.3}$$

$$\left|\operatorname{Re}\lambda_l\right| \leq n \cdot \max_{1 \leq j,k \leq n} |b_{jk}| \quad \text{and} \quad \left|\operatorname{Im}\lambda_l\right| \leq n \cdot \max_{1 \leq j,k \leq n} |c_{jk}| \quad (1 \leq l \leq n), \tag{6.4}$$

for the eigenvalues λ_l of a general $n \times n$ matrix $A = \left[a_{jk}\right]$, where, $B = \left[b_{jk}\right] = (A + A^*)/2$ and $C = \left[c_{jk}\right] = (A - A^*)/(2i)$. The estimates (6.4) were first obtained by A.Hirsch [Hir]. They were improved to

$$\left|\operatorname{Im}\lambda_l\right| \leq \sqrt{\frac{n(n-1)}{2}} \cdot \max_{1 \leq j,k \leq n} |c_{jk}| \quad (1 \leq l \leq n) \tag{6.5}$$

for real matrices A by F. Bendixson [Bend] and reproved in [Sch2]. In §7 of ([Sch2]), the interesting inequality

$$\sum_{j<k} |\lambda_j - \lambda_k|^2 \leq \sum_{j<k} |a_{jj} - a_{kk}|^2 + n \sum_{j \neq k} |a_{jk}|^2 \tag{6.6}$$

is derived and then used to obtain the following estimate for the discriminant

$$d = \prod_{j,k} (\lambda_j - \lambda_k)^2,$$

of the characteristic equation $\det\left(\lambda I_n - A\right) = 0$:

$$|d|^{\frac{2}{n(n-1)}} \leq \frac{2}{n(n-1)} \sum_{j<k} |a_{jj} - a_{kk}|^2 + \frac{2}{n-1} \sum_{j \neq k} |a_{jk}|^2. \tag{6.7}$$

In §5 of [Sch2], the well known Hadamard bound

$$|\det A| \leq \left(\max_{1 \leq j,k \leq n} |a_{jk}| \right)^n \cdot n^{n/2}. \tag{6.8}$$

on the maximal value of the determinant of a matrix is derived from the inequality (6.1) with the help of the inequality between the geometric and the arithmetic means:

$$|\det A|^2 = |\lambda_1|^2 \cdot \cdots \cdot |\lambda_n|^2 \leq \left(\frac{|\lambda_1|^2 + \cdots + |\lambda_n|^2}{n} \right)^n \leq \left(\frac{\sum_{j,k=1}^n |a_{jk}|^2}{n} \right)^n.$$

The challenge of obtaining simple new proofs of various Hadamard inequalities seems to have been one of Issai Schur's favorite occupations.

In [Sch2], Schur also considers integral operators $x(t) \to (\boldsymbol{K}x)(t)$ in $L^2(a, b)$,

$$(\boldsymbol{K}x)(t) = \int_a^b K(t, \tau) x(\tau) \, d\tau \quad (a \le t \le b), \tag{6.9}$$

with kernels $K(t, \tau)$ that satisfy the condition

$$\int_a^b \int_a^b |K(t, \tau)|^2 \, dt \, d\tau < \infty. \tag{6.10}$$

Today, such operators are commonly called *Hilbert-Schmidt integral operators*. Schur extended the inequality (6.1) to these operators:

$$\sum_l |\lambda_l(\boldsymbol{K})|^2 \le \int_a^b \int_a^b |K(t, \tau)|^2 \, dt \, d\tau, \tag{6.11}$$

where the summation on the left hand side is extended over the set of all eigenvalues $\lambda_l(\boldsymbol{K})$ of the integral operator $\boldsymbol{K}$. In particular, the series on the left hand side of (6.11) converges.

One of the fundamental results of the Fredholm theory of integral equations [Fred] is the identification of the nonzero eigenvalues $\lambda_l(\boldsymbol{K})$ of an integral operator (6.9) with a continuous kernel as the reciprocals of the zeros of an entire function $D_{\boldsymbol{K}}(\lambda)$ (that is constructed from the kernel $K(t, \tau)$ of this operator). This function is termed *the Fredholm denominator (or the Fredholm determinant) of the operator (6.9)*. It is defined by the Taylor series

$$D_{\boldsymbol{K}}(\lambda) = \sum_{n=0}^{\infty} c_n \lambda^n, \tag{6.12}$$

with coefficients

$$c_n = \frac{(-1)^n}{n!} \int_a^b \ldots \int_a^b \det \begin{bmatrix} K(t_1, t_1) & \cdots & K(t_1, t_n) \\ \cdots & \cdots & \cdots \\ K(t_n, t_1) & \cdots & K(t_n, t_n) \end{bmatrix} dt_1 \cdots dt_n. \tag{6.13}$$

From (6.13) and the Hadamard inequality (6.8), it follows that if

$$\sigma = (b - a)(\max_{a \le t, \tau \le b} |K(t, \tau)|) < \infty,$$

then

$$|c_n| \le \sigma^n \cdot n^{n/2}/n!. \tag{6.14}$$

Consequently, the series (6.12) converges for every complex λ, and its sum $D_{\boldsymbol{K}}(\lambda)$ is an entire function that is subject to the bound

$$\ln |D_{\boldsymbol{K}}(\lambda)| \le \sigma^2 |\lambda|^2 (1 + o(1)) \quad (|\lambda| \to \infty). \tag{6.15}$$

46

Thus, the counting function of the zeros $\mu_1, \mu_2, \ldots$ of $D_{\boldsymbol{K}}(\lambda)$:

$$n_{\boldsymbol{K}}(r) = \#\{\mu_\ell(\boldsymbol{K}) : |\mu_\ell(\boldsymbol{K})| \leq r\} = \#\{\lambda_\ell(\boldsymbol{K}) : |\lambda_\ell(\boldsymbol{K})|^{-1} \leq r\},$$

satisfies the condition

$$n_{\boldsymbol{K}}(r) = O(r^2), \quad \text{as} \quad r \to \infty. \tag{6.16}$$

The estimates (6.15) and its consequence (6.16) were known [Lal] before the Schur paper [Sch2] appeared. However, the estimate (6.11) is stronger than the estimate (6.16). From the convergence of the series $\sum_l |\lambda_l(\boldsymbol{K})|^2$ and from the estimate (6.15) it follows that the Fredholm denominator (6.12)-(6.13) admits the multiplicative decomposition

$$D_{\boldsymbol{K}}(\lambda) = e^{c\lambda + d\lambda^2} \prod_l \left(1 - \lambda\,\lambda_l(\boldsymbol{K})\right) e^{\lambda\,\lambda_l(\boldsymbol{K})}, \tag{6.17}$$

for some choice of constants c and d. The fact that the Fredholm denominator of the integral operator (6.9) with a continuous kernel admits a representation of the form (6.17) was first noted by Schur in $\S 14$ of [Sch2]. (It is important to note that the kernel $K(t, \tau)$ is not assumed to be symmetric or Hermitian.) This result of Schur is sharp in the sense that there exists a continuous kernel K on a finite interval $[a, b]$ whose eigenvalues satisfies the condition

$$\sum_\ell |\lambda_\ell(\boldsymbol{K})|^{2-\epsilon} = \infty \text{ for every } \epsilon > 0.$$

To construct an example, let $K(t, \tau) = \varphi(t - \tau)$ for $0 \leq t, \tau \leq 1$, where $\varphi(t + 1) = \varphi(t)$ is a continuous periodic function on $\mathbb{R}$ with Fourier expansion $\varphi(t) \sim \sum_l c_l e^{2\pi i \ell t}$. Then the functions $e^{2\pi i \ell t}$ are eigenfunctions of the kernel K, and the Fourier coefficients c_ℓ are eigenvalues of this kernel. A kernel with the desired properties is obtained by choosing a continuous periodic function φ whose Fourier coefficients c_ℓ satisfy the condition $\sum_\ell |c_\ell|^{2-\epsilon} = \infty$ for every $\epsilon > 0$. The first example of such a function was constructed by T. Carleman [Carl2]. Other examples can be found in [Bar], Chapt. 4, $\S 16$, or in [Zyg], Chapt. 5., **(4.9)**. In his first publication [Carl1], Carleman proved that in fact $d = 0$ in (6.17). Thus, the scientific career of this outstanding analyst started with an improvement of a result of Issai Schur.

The inequality (6.1) can also be presented in the form

$$\sum_{l=1}^{n} |\lambda_l(A)|^2 \leq \sum_{l=1}^{n} s_l(A)^2, \tag{6.18}$$

where the $\lambda_l(A)$ are the eigenvalues of the matrix A and the numbers $s_l(A)$ are the singular values of A.

The auxiliary inequality

$$\sum_{l=1}^{n} |\lambda_l(A)| \leq \sum_{l=1}^{n} s_l(A) \tag{6.19}$$

can also be proved[3] in an elementary way by using the Schur transformation (6.2) to reduce the matrix A to upper triangular form. In fact, it suffices to prove (6.19) for upper triangular matrices A, since the transformation (6.2) does not change either the eigenvalues or the singular values of the matrix. But then, if $\{e_l\}_{1 \leq l \leq n}$ is the natural basis of the space $\mathbb{C}^n$,

$$a_{ll} = \langle Ae_l, e_l \rangle = \lambda_l(A), \quad l = 1, \ldots, n,$$

up to a reindexing of the eigenvalues, if need be. Now let $A = S \cdot V$ be the polar decomposition of the matrix A: $S \geq 0, V^*V = VV^* = I_n$ and let $h_l = Ve_l$. Then the vectors $\{h_l\}_{1 \leq l \leq n}$ form an orthonormal basis of the space $\mathbb{C}^n$ and, by the Cauchy-Schwarz inequality,

$$|\langle Ae_l, e_l \rangle| = |\langle Sh_l, e_l \rangle| \leq \sqrt{\langle Sh_l, h_l \rangle} \cdot \sqrt{\langle Se_l, e_l \rangle}.$$

Therefore,

$$\sum_{l=1}^n |\lambda_l(A)| = \sum_{l=1}^n |\langle Ae_l, e_l \rangle| \leq \sqrt{\sum_{l=1}^n \langle Sh_l, h_l \rangle} \cdot \sqrt{\sum_{l=1}^n \langle Se_l, e_l \rangle} = \sum_{l=1}^n \lambda_l(S) = \sum_{l=1}^n s_l(A),$$

since

$$\sum_{l=1}^n \langle Sh_l, h_l \rangle = \sum_{l=1}^n \langle Se_l, e_l \rangle = \operatorname{trace} S = \sum_{l=1}^n \lambda_l(S),$$

and, by the definition of singular values, $\{\lambda_l(S)\}_{l=1}^n = \{s_l(A)\}_{l=1}^n$.

The inequalities (6.1), *written in the form* (6.18), and (6.19) were significantly generalized by H. Weyl [Wey] in 1949. The generalization is based on the concept of majorization that was discussed in the previous section. A crucial role is played by the inequalities

$$|\lambda_1(A) \cdot \lambda_2(A) \cdot \cdots \cdot \lambda_k(A)| \leq s_1(A) \cdot s_2(A) \cdot \cdots \cdot s_k(A) \quad (k = 1, 2, \ldots, n-1), \quad (6.20)$$

which are valid when the eigenvalues $\lambda_k(A)$ and the singular values $|s_k(A)|$ are indexed in such a way that $|\lambda_1(A)| \geq |\lambda_2(A)| \geq \cdots \geq |\lambda_n(A)|$ and $s_1(A) \geq s_2(A) \geq \cdots \geq s_n(A)$. The equality

$$|\lambda_1(A) \cdot \lambda_2(A) \cdot \cdots \cdot \lambda_n(A)| = s_1(A) \cdot s_2(A) \cdot \cdots \cdot s_n(A) \qquad (6.21)$$

holds because both sides are equal to $|\det A|$. The relations (6.20) and (6.21) mean that the sequence $\{\ln |\lambda_k(A)|\}_{k=1}^n$ is majorized by the sequence $\{\ln s_k(A)\}_{k=1}^n$:

$$\{\ln |\lambda_k(A)|\}_{k=1}^n \prec \{\ln s_k(A)\}_{k=1}^n. \qquad (6.22)$$

In [Wey], Weyl derived the inequalities (6.20) and then applied the inequality

$$\sum_{k=1}^n \psi(y_k) \leq \sum_{k=1}^n \psi(x_k), \qquad (6.23)$$

---

[3]This proof is adopted from [GoKr1], Chapt. IV, § 8. See Theorem 8.1, especially the footnote 7 on p. 128 of the Russian original or on p. 98 of the English translation.

which holds for any convex function $\psi(\,\cdot\,)$ on $(-\infty, \infty)$ and any pair of sequences $\{y_k\}$ and $\{x_k\}$ such that $\{y_k\} \prec \{x_k\}$, to the sequences $y_k = \ln |\lambda_k(A)|$ and $x_k = \ln s_k(A)$. The inequality (6.23) is a direct consequence of the result[4] by Schur (which states that the inequality (6.23) holds for sequences $\boldsymbol{x}$ and $\boldsymbol{y} = M\boldsymbol{x}$ that are related by a doubly-stochastic matrix M), and of the result[4] by Hardy, Littlewood and Polya, who proved that

$$\boldsymbol{y} \prec \boldsymbol{x} \quad \Longrightarrow \quad \boldsymbol{y} = M\boldsymbol{x}$$

for some doubly-stochastic matrix M. However, Weyl was not aware of these results and gave an independent proof of the implication

$$\boldsymbol{y} \prec \boldsymbol{x} \quad \Longrightarrow \quad (6.23) \tag{6.24}$$

in Lemma 1 of [Wey]. The inequalities (6.20) were known before the paper [Wey] was published. (See, for example, Exercise 17 on page 110 of the book [TuAi].) However, it was Hermann Weyl who first combined the inequalities (6.20) with the implication (6.24) to obtain the following

THEOREM. *Let A be an $n \times n$ matrix with eigenvalues $\{\lambda_k(A)\}_{k=1}^n$ and singular values $\{s_k(A)\}_{k=1}^n$ (counting multiplicities) and let $\varphi(\,\cdot\,)$ be a function on $(0, \infty)$ such that the function $\psi(t) = \varphi(e^t)$ is convex on $(-\infty, \infty)$. Then*

$$\sum_{k=1}^n \varphi(\lambda_k(A)) \leq \sum_{k=1}^n \varphi(s_k(A)) \,. \tag{6.25}$$

Weyl invoked the inequality (6.25) with $\varphi(t) = t^p$ and $p > 0$ to obtain the following generalization of Schur's inequality (6.18):

$$\sum_{l=1}^n |\lambda_l(A)|^p \leq \sum_{l=1}^n s_l(A)^p \qquad (0 < p < \infty). \tag{6.26}$$

Analogous inequalities hold for linear operators A in a Hilbert space $\mathcal{H}$ that belong to the class $\mathfrak{S}_p$, i.e., for which $\sum_l s_l(A)^p < \infty$, where the $s_l(A)$ are the eigenvalues of the operator $\sqrt{A^*A}$. (Usually the singular values $s_l(A)$ are enumerated by the indices $l = 0, 1, 2, \ldots$.) The summation in the last inequality is then extended over all eigenvalues and over all singular values of the operator A. The resulting inequality is very useful in the theory of integral equations. The point is that it is difficult to calculate the eigenvalues and singular values of an integral operator in terms of its kernel. However, the singular values can be effectively estimated from above by approximating the kernel $K(t, \tau)$ by degenerate kernels of the form $K_n(t, \tau) = \sum_{l=1}^n \varphi_l(t)\psi_l(\tau)$ and invoking the fact that

$$s_n(\boldsymbol{K}) = \inf \|\boldsymbol{K} - \boldsymbol{K}_n\|$$

---

[4]These results were discussed in the previous section; see Theorems I, II and I$'$.

49

as $K_n(t, \tau)$ runs over the set of all degenerate kernels of the indicated form. (See [GoKr1], Chapt. 2, § 2, item **3**.) The smoother the kernel K, the more rapid the rate of decay of the sequence $\|\boldsymbol{K} - \boldsymbol{K}_n\|$ and thus, the rate of decay of the sequence $|s_n\boldsymbol{K})|$. The inequality

$$\sum_l |\lambda_l(\boldsymbol{K})|^p \leq \sum_{l=0}^{\infty} s_l(\boldsymbol{K})^p \tag{6.27}$$

is then used to derive the rate of decay of the eigenvalues $\lambda_n(\boldsymbol{K})$. This is the "modern" way to derive the rate of decay of the eigenvalues $\lambda_n(\boldsymbol{K})$ of an integral operator from the smoothness of its kernel $K(t, \tau)$. The theory of spline approximation is often used to construct good approximating kernels. (See, for example, the papers by M. Sh. Birman and M.Z. Solomyak mentioned in Section 4.)

The "classical" approach, which does not exploit the Weyl inequalities, is more complicated and gives weaker results. Chang, in a paper [Chang] that appeared before the paper [Wey], proved that

$$\sum_{l=0}^{\infty} s_l(\boldsymbol{K})^p < \infty \quad \Longrightarrow \quad \sum_l |\lambda_l(\boldsymbol{K})^p < \infty \tag{6.28}$$

for integral operators (6.9) of Hilbert-Schmidt class, i.e., with kernels $K(t, \tau)$ satisfying the condition (6.10). The "classical" methods of the paper [Chang] are involved and rather difficult.

The Weyl inequalities are also useful in "abstract" operator theory. Taking $\varphi(t) = \ln(1 + |\lambda|t)$ (which is admissible, since the function $\psi(t) = \ln(1 + |\lambda|e^t)$ is convex), one can obtain the inequality

$$\left| \prod_l \left(1 - \lambda\,\lambda_l(A)\right) \right| \leq \prod_l \left(1 + |\lambda|\,s_l(A)\right) \qquad (\forall \lambda \in \mathbb{C})\,, \tag{6.29}$$

for linear operators A from the class $\mathfrak{S}_1$ of trace class operators in a Hilbert space. This inequality is useful in the study of the so-called characteristic determinants of trace class operators and related analytic considerations. (See Chapter IV of [GoKr1].) In particular, the inequality (6.29) plays an important role in the proof of a theorem by V.B. Lidskiĭ, which states that the matricial trace and the spectral trace of a trace class operator coincide. (See [Lid], and [GoKr1], Chapt. III, § 8, Theorem 8.1.) This theorem is of principal importance in operator theory.

The Weyl inequality (6.25) is one of the central tools in the toolbox of modern operator theory. However, as Weyl himself wrote [Wey], the first step was taken by Schur:

"*Long ago I. Schur proved* (6.26) *for* [5] $p = 1$. *Recently S.H. Chang showed in his thesis that, in the case of integral equations, the convergence of* $\sum s_l^p$ *implies the convergence of*

---

[5] This reference by Weyl is not accurate. Schur proved the inequality (6.26) for $p = 2$, but not for $p = 1$.

$\sum |\lambda_l|^p$. *These two facts led me to conjecture the relation (6.26), at least for $p \leq 1$. After having conceived a simple idea for the proof, I discussed the matter with C.L. Siegel and J. von Neumann; their remarks have contributed to the final form and generality in which the results are presented here."*

Thus, the paper [Sch2] served as source of inspiration for both T. Carleman and H. Weyl.

REFERENCES

[Bar]   BARI, N.K. *Trigonometricheskie ryady* (Russian). Fiz.-Mat. Giz, Moscow 1961, 936 pp. English transl.: *A treatise on trigonometric series. Vols. I, II.* A Pergamon Press Book The Macmillan Co., New York 1964 , Vol. I: xxiii+553 pp. Vol. II: xix+508 pp.

[Bend]   BENDIXSON, F. *Sur les racines d'une équation fondamentale.* Acta Math., **25** (1902), pp. 359 - 365.

[Carl1]   CARLEMAN, T. *Sur le genre du dénominateur $D(\lambda)$ de Fredholm.* Arkiv för Mathematik, Astronomi och Fysik, **12** (1917), pp. 1 - 5. Reprinted in [Carl3], pp. 1 - 5.

[Carl2]   CARLEMAN, T. *Über die Fourierkoeffizienten einer stetigen Funktion.* Acta Math., **41** (1918), pp. 377 - 384. Reprinted in [Carl3], pp. 15 - 22.

[Carl3]   *Édition Complète des Articles de Torsten Carleman.* Publée par l'Institute Mathématique Mittag-Leffler. Litos Reprotryk, Malmö 1960.

[Chang]   CHANG, SHIH-HSUN. *On the distribution of the characteristic values and singular values of linear integral equations.* Trans. Amer. Math. Soc., **67**:2 (1949), pp. 351 - 367.

[Fred]   FREDHOLM, E.I. *Sur une classe d'équations fonctionnelles.* Acta Math., **27**, pp. 365 - 396.

[GoKr1]   GOHBERG, I.Ts. and M.G. KREIN. *Vvedenie v Teoriyu Lineĭnykh Nesamosopryazhennykh Operatorov* (Russian). Nauka, Moskow 1965, 448 pp. English Transl.: *Introduction to the Theory of Linear Non-Selfadjoint Operators.* (Transl. Math. Monogr. **18**). Amer. Math. Soc., Providence, R.I., 1969.

[Hir]   HIRSCH, A. *Sur les racines d'une équation fondamentale.* Acta Math., **25** (1902), pp. 367 - 370.

[Lal]   LALESCO, T. *Sur l'ordre de la fonction entière $D(\lambda)$ de Fredholm.* Comp. Rend. des Seances de'l Academie des Science, Paris, 25 November (1907), p. 906.

[Lid]   LIDSKIĬ, V.B. *Nesamosopryazhennye operatory, imeyushchie sled* (Russian). Dokl. Akad. Nauk SSSR, **125** (1959), pp. 485 - 488. English transl.: *Non-selfadjoint operators with a trace.* Amer. Math. Soc. Thansl. (**2**) **47** (1965), 43 - 46.

[Sch2]   SCHUR, I.: *Über die charakteristischen Wurzeln einer linearen Substitution mit einer Anwendung auf die Theorie der Integralgleichungen.* [*On the characteristic roots of a linear substitution with an application to the theory of integral equations* - in German]. Mathematische Annalen, **66** (1909), pp. 488 - 510. Reprinted in: [Sch: Ges], Vol. I, pp. 272 - 294.

[Sch: Ges] SCHUR, I.: *Gesammelte Abhandlungen.* [*Collected Works*]. Vol. I, II, III. Springer-Verlag, Berlin·Heidelberg·New York 1973.

[TuAi]   TURNBBULL, H.W. and A.C. AITKEN. *An Introductian to the Theory of Canonoical Matrices.* Blackie & Son, Ltd., London and Glasgow, First printed, 1932; Reprinted 1945, 1948, 1950, 1952.

[Wey]   WEYL, H. *Inequalities between the two kinds of eigenvalues of a linear transformation.* Proc. Nat. Acad. Sci. USA, **35** (1949), pp. 408 - 411. Reprinted in: [Wey:Ges], Band IV, pp. 390 - 393.

[Wey:Ges] WEYL, H. *Gesammelte Abhandlungen* [*Collected Works*]. *Vol. I, II, III, IV.* Springer-Verlag, Berlin·Heidelberg·New York 1968.

[Zyg]   ZYGMUND, A. *Trigonometric series.* 2nd ed. Vols. I, II. Cambridge University Press, New York 1959. Vol. I. xii+383 pp.; Vol. II. vii+354 pp.

7. Triangular representations of matrices and linear operators.

One of the important theorems of Schur that was discussed in the preceding section states that every square matrix is unitarily equivalent to a triangular matrix. In the early fifties, stimulated by this theorem of Schur, Moshe (Mikhail Samulovich) Livšic (= Livshits) obtained an analogue of this result for a class of bounded linear operators in a separable Hilbert space.

To explain his results, let us first recall that every bounded operator A in a Hilbert space $\mathcal{H}$ is representable in the form

$$A = B_A + iC_A, \tag{7.1}$$

where

$$B_A = \operatorname{Re} A = \frac{A + A^*}{2} = (B_A)^* \quad \text{and} \quad C_A = \operatorname{Im} A = \frac{A - A^*}{2i} = (C_A)^*. \tag{7.2}$$

Livšic obtained his conclusions in the class $i\Omega$ of bounded linear operators A for which C_A is of of trace class. In the simplest case of this setting,

$$\operatorname{rank}(C_A) = 1, \tag{7.3}$$

and hence C_A must be be definite: either $C_A \geq 0$, or $C_A \leq 0$. Thus,

$$C_A = j \, |C_A|, \quad \text{where} \quad |C_A| = \sqrt{(C_A)^* C_A}, \quad j = +1 \quad \text{or} \quad j = -1 \tag{7.4}$$

and the imaginary parts β_k of the eigenvalues $\lambda_k = \alpha_k + i\beta_k$ of the operator A are of the form $\beta_k = j|\beta_k|$.

Without loss of generality, we may assume that the the operator A is *completely non-selfadjoint*: *There is no invariant subspace for the operator A on which A induces a self-adjoint operator.* Indeed, if the operator A is not completely non-selfadjoint, then it splits into the orthogonal sum $A = S \oplus A_{\text{cns}}$, where S is a selfadjoint operator and A_{cns} is a completely non-selfadjoint operator. Moreover, $\text{Im}\, A = \text{Im}\, A_{\text{cns}}$.

The eigenvalues $\lambda_k = \alpha_k + i\beta_k$ of a completely non-selfadjoint operator A with non-negative (non-positive) imaginary part are never real: either $\beta_k > 0$ for all k (if $C_A \geq 0$), or $\beta_k < 0$ for all k (if $C_A \leq 0$).

The triangular model T for an operator A satisfying the condition (7.3) acts in the model Hilbert space $\mathcal{H}_{\text{mod}} = l^2 \oplus L^2$ that is the orthogonal sum of the space l^2 of square summable one-sided infinite sequences $(\xi_1, \xi_2, \dots ,)$ of complex numbers of dimension $n \leq \infty$, where n is equal to the number of eigenvalues of the operator A (counting multiplicities), and L^2 is the space of all square summable complex-valued functions on a finite interval $[0, l]$, where the number l is determined uniquely by the operator A. The spaces l^2 and L^2 are equipped with the standard scalar products. The block decomposition of the operator T that corresponds to the decomposition $\mathcal{H}_{\text{mod}} = l^2 \oplus L^2$ of the space $\mathcal{H}_{\text{mod}}$, is of the form

$$T = \begin{bmatrix} T_{\text{dis}} & T_{\text{cou}} \\ 0 & T_{\text{con}} \end{bmatrix}, \tag{7.5}$$

where $T_{\text{dis}} : l^2 \to l^2$, $T_{\text{con}} : L^2 \to L^2$ and $T_{\text{cou}} : L^2 \to l^2$.

The operator T_{dis}, the discrete part of the operator T, is defined by its matrix $[t_{km}]$ in the natural basis of the space l^2. This matrix is upper triangular, i.e., with j as in (7.4),

$$t_{km} = 0 \text{ for } k > m, \quad t_{kk} = \lambda_k, \quad t_{km} = i \, |\beta_k|^{1/2} j \, |\beta_m|^{1/2} \text{ for } k < m. \tag{7.6}$$

The operator T_{dis} is bounded, since A is bounded and $\sum_k |\beta_k| \leq \text{trace} \, |C_A| < \infty$. The operator T_{con}, the continuous part of the operator T, is an integral operator of the form

$$(T_{\text{con}}\xi)(t) = \lambda(t)\xi(t) + i \int_t^l K(t, s) \, \xi(s) \, ds \quad 0 \leq t \leq l, \tag{7.7}$$

where $\lambda(t)$ is a non-decreasing bounded real-valued function on the interval $[0, l]$ which is determined by the operator A. The kernel $K(t, s)$ of the integral operator (7.7) is of the form

$$K(t, s) = 0 \text{ for } 0 \leq s < t \leq l, \quad K(t, s) = i \, j \text{ for } 0 \leq t < s \leq l, \tag{7.8}$$

53

i.e., the operator (7.7) can be considered as upper triangular. The summand $\lambda(t)\xi(t)$ corresponds to the "main diagonal" of this operator. For the operator T_{cou}, the so called coupling operator, an explicit formula can be obtained. Thus, the whole[6] operator T can be naturally considered as an upper triangular operator.

DEFINITION. *Let A be an operator which acts in a Hilbert space $\mathcal{H}$. An operator $\widetilde{A}$ acting in a larger Hilbert space $\widetilde{\mathcal{H}}$, $\widetilde{\mathcal{H}} \supseteq \mathcal{H}$, is said to be an inessential extension of the operator A, if $\widetilde{A} = A \oplus S$, where S is a selfadjoint operator acting in the space $\widetilde{\mathcal{H}} \ominus \mathcal{H}$.*

THEOREM I$'$ (M. Livšic). *Let A be a bounded completely non-selfadjoint linear operator in a Hilbert space $\mathcal{H}$ such that C_A is one-dimensional. Then there exists an inessential extension $\widetilde{A} : \widetilde{\mathcal{H}} \to \widetilde{\mathcal{H}}$ of the operator A that is unitarily equivalent to an "upper triangular" model operator T of the form (7.5): There exists a unitary operator U acting from $\mathcal{H}_{mod} = l^2 \oplus L^2$ onto $\widetilde{\mathcal{H}}$ such that*

$$T = U^* \widetilde{A} U = U^{-1} \widetilde{A} U. \tag{7.9}$$

Triangular models of the same general form (7.5)-(7.6)-(7.7) can also be constructed for bounded linear operators A in a separable Hilbert space $\mathcal{H}$ when C_A is only assumed to be of trace class. They are, however, a bit more complicated.

For an operator A in a separable Hilbert space $\mathcal{H}$, let us introduce the *non-hermitian subspace* $\mathcal{N}_A$ as the closure of the image of its imaginary part C_A:

$$\mathcal{N}_A = \overline{C_A \mathcal{H}}. \tag{7.10}$$

The dimension $\mathfrak{n}_A$ of the non-hermitian subspace $\mathcal{N}_A$ is said to be the *non-hermitian rank* of the operator A:

$$\mathfrak{n}_A = \dim \mathcal{N}_A. \tag{7.11}$$

The restriction $C_A|_{\mathcal{N}_A}$ of the operator C_A on the subspace $\mathcal{N}_A$, considered as an operator in the Hilbert space $\mathcal{N}_A$, is a selfadjoint operator for which the point $\{0\}$ is not an eigenvalue. Therefore, the polar decomposition of this operator is of the form

$$C_A|_{\mathcal{N}_A} = J_A \cdot M_A, \tag{7.12}$$

where

$$J_A : \mathcal{N}_A \to \mathcal{N}_A, \quad J_A = J_A^*, \quad J_A^2 = I_{\mathcal{N}_A} \quad \text{and} \quad M_A : \mathcal{N}_A \to \mathcal{N}_A, \ M_A \geq 0. \tag{7.13}$$

(In this polar decomposition, J_A is the unitary operator and M_A is the operator modulus.)

---

[6]The operator T will not contain a discrete part T_{dis} if the operator A has no eigenvalues. It will not contain a continuous part T_{con} if $l = 0$.

To construct the triangular model of the operator A, let us choose a Hilbert space $\mathcal{E}$ of the same dimension as the non-hermitian subspace $\mathcal{N}_A$: $\dim \mathcal{E} = \dim \mathcal{N}_A$. Let $J_\mathcal{E}$ be an operator in $\mathcal{E}$,

$$J_\mathcal{E} : \mathcal{E} \to \mathcal{E}, \quad J_\mathcal{E} = J_\mathcal{E}^*, \quad J_\mathcal{E}^2 = I_\mathcal{E}, \tag{7.14}$$

of the same signature [7] as that of the operator J_A. For the sake of brevity, we shall restrict our attention to the case of operators with real spectrum only. The model space $\mathcal{H}_{\mathrm{mod}}$ in this case is the space $L_\mathcal{E}^2([0, l])$, i.e., the space of all square-integrable functions on a finite interval $[0, l] \subset \mathbb{R}$, whose values are elements of the Hilbert space $\mathcal{E}$, with the scalar product:

$$\langle \boldsymbol{\xi}, \boldsymbol{\eta} \rangle_{\mathcal{H}_{\mathrm{mod}}} = \int_0^l \langle \xi(t), \eta(t) \rangle_\mathcal{E} \, dt, \quad \text{for} \quad \boldsymbol{\xi} = \xi(t), \, \boldsymbol{\eta} = \eta(t) \in L_\mathcal{E}^2([0, l]).$$

The model operator T acts in the space $\mathcal{H}_{\mathrm{mod}}$ according the rule

$$(T\xi)(t) = \lambda(t)\xi(t) + i \int_t^l \Pi(t) J_\mathcal{E} \Pi(s)^* \xi(s) \, ds, \tag{7.15}$$

where $\lambda(t)$ is a non-decreasing real-valued function on the interval $[0, l]$ and $\Pi(t)$ is a function on the interval $[0, l]$ whose values are Hilbert-Schmidt operators in $\mathcal{E}$ that satisfy the normalization condition

$$\mathrm{trace}_\mathcal{E} \, \Pi(t)^* \Pi(t) \equiv 1, \quad 0 \le t \le l. \tag{7.16}$$

THEOREM I$''$ (M. Livšic). *Let A be a bounded completely non-selfadjoint linear operator in a Hilbert space $\mathcal{H}$ such that C_A is of trace class and the spectrum of A is real. Then there exists an inessential extension $\widetilde{A} : \widetilde{\mathcal{H}} \to \widetilde{\mathcal{H}}$, $\widetilde{\mathcal{H}} \supseteq \mathcal{H}$, of the operator A that is unitarily equivalent to an "upper triangular" model operator T of the form (7.15) : There exists a unitary operator U acting from $\mathcal{H}_{mod} = L_\mathcal{E}^2([0, l])$ onto $\widetilde{\mathcal{H}}$ such that*

$$T = U^* \widetilde{A} U = U^{-1} \widetilde{A} U. \tag{7.17}$$

REMARK. Direct calculation shows that

$$\big((T - T^*)\xi\big)(t) = i \int_0^l \Pi(t) J_\mathcal{E} \Pi(s)^* \, \xi(s) \, ds \tag{7.18}$$

---

[7] The spectrum of an operator J which posses the properties $J = J^*$, $J^2 = I$ can consist of the points $\{+1\}$ and $\{-1\}$ only. These points are eigenvalues of J. Let $\mathfrak{p}_J$ and $\mathfrak{q}_J$ denote the dimensions of the corresponding eigenspaces, $0 \le \mathfrak{p}_J, \mathfrak{q}_J \le \infty$. The signature of the operator J is the pair $(\mathfrak{p}_J, \mathfrak{q}_J)$.

Thus, the model operator (7.15) is of the form

$$(T\xi)(t) = \lambda(t)\xi(t) + 2i \int_0^l \chi(t,s)H(t,s)\xi(s)\,ds\,, \tag{7.19}$$

where

$$\chi(t,s) = 1 \ \text{ for } s > t, \ \ \chi(t,s) = 0 \ \text{ for } s < t\,,$$

and the kernel $H(t,s) = \Pi(t)J_{\mathcal{E}}\Pi(s)^*$ represents the imaginary part of the operator T:

$$\big((T - T^*)\xi\big)(t) = 2i \int_0^l H(t,s)\xi(s)\,ds\,. \tag{7.20}$$

In other words, *the kernel $K(t,s)$ that represents the operator T can be obtained from the kernel $H(t,s)$ that represents the imaginary part of T, by means of " truncation to the upper triangle": $K(t,s) = \chi(t,s)H(t,s)$.*

For operators whose imaginary part is of trace class but whose spectrum is not necessary real, the triangular model has a more complicated form; see e.g., [Liv5] and in [BroLi].

Moshe Livšic introduced the machinery of characteristic functions of linear operators in the mid forties in order to solve a number of problems connected with theory of extensions of linear operators, see [Liv1] and [Liv2]. He then applied this machinery to establish the unitary equivalence of an operator of the class $i\Omega$ to a triangular model in the early fifties. See [Liv3],[Liv4] for the first results and [Liv5] for a detailed presentation.

The characteristic function of a non-selfadjoint linear operator A acting in a Hilbert space $\mathcal{H}$ is defined as follows: Choose a Hilbert space $\mathcal{E}$ of the same dimension as the non-hermitian subspace $\overline{C_A\,\mathcal{H}}$ of the operator A and then factor the operator C_A in the form

$$C_A = \Gamma J_{\mathcal{E}}\Gamma^*\,, \tag{7.21}$$

where Γ, $J_{\mathcal{E}}$ are linear operators,

$$\Gamma : \mathcal{E} \to \mathcal{H}\,, \ \ J_{\mathcal{E}} : \mathcal{E} \to \mathcal{E}\,, \ \ J_{\mathcal{E}}^2 = I_{\mathcal{E}} \ \ (I_{\mathcal{E}} \text{ denotes the identity operator in } \mathcal{E}). \tag{7.22}$$

The characteristic function $W_A(z)$ of the operator A is the operator valued function of the complex variable z that is defined for z out of the spectrum of A by the rule

$$W_A(z) = I_{\mathcal{E}} + 2i\Gamma^*(zI - A)^{-1}\Gamma J_{\mathcal{E}}\,. \tag{7.23}$$

Notice that $W_A(z)$ acts in the Hilbert space $\mathcal{E}$, which, in many problems of interest, is a finite dimensional space. In Livšic's terminology, the space $\mathcal{E}$ is said to be *the channel space* and the operator Γ is said to be *the channel operator*.

Livšic showed that the characteristic function is a unitary invariant of a completely non-selfadjoint operator: *Let A_1 and A_2 be two completely non-selfadjoint operators such that their characteristic functions $W_{A_1}(z)$ and $W_{A_2}(z)$ (with the same channel space $\mathcal{E}$) are equal: $W_{A_1}(z) \equiv W_{A_2}(z)$. Then the operators A_1 and A_2 are unitarily equivalent.* To reduce a non-selfadjoint operator to triangular form, Livšic calculated its characteristic function $W_A(z)$ and then constructed a model operator T in such a way that its characteristic function $W_T(z)$ coincides with $W_A(z)$. Subsequently, triangular models of operators were partially superceded by functional models, see [SzNFo], [Bran], [NiVa].

There is also a very important correspondence between the invariant subspaces of an operator A and certain divisors of its characteristic function. However, the importance of the notion of a characteristic function is not confined to its applications in operator theory. Livšic related *the theory of stationary linear dynamical systems* to the theory of linear non-selfadjoint operators and showed that the characteristic matrix function of a linear operator that serves as an "inner operator" for the dynamical system can be identified with the scattering matrix of this system. Examples are furnished in [8] [Liv6], [Liv7] and [BroLi]. A detailed presentation of the early stages [9] of the theory of open systems (as Livšic termed them) can be found in [Liv8]. In particular, as he noted in the first sentence of Section 2.2 of that source: "The resolution of a system into a chain of elementary systems is closely related to the reduction of the operator ... to triangular form."

The results of Livšic on reducing operators to triangular form are similar in form to the result of Schur. However, the methods that he used are absolutely different from the method of Schur. Schur's result implies that there exists an orthonormal basis $e_1, \ldots, e_n$ of the space $\mathbb{C}^n$ such that the given matrix A is upper-triangular in this basis. Thus, if

$$\mathcal{H}_0 = 0, \quad \mathcal{H}_k = \mathrm{span}\{e_1, \ldots, e_k\}, \quad k = 1, 2, \ldots, n, \tag{7.24}$$

then this collection $\{\mathcal{H}_k\}_{0 \leq k \leq n}$ of subspaces of $\mathbb{C}^n$ possesses the following properties:

   i. $0 = \mathcal{H}_0 \subset \mathcal{H}_1 \subset \mathcal{H}_2 \subset \ldots \subset \mathcal{H}_n = \mathbb{C}^n$,

   ii. $\dim\left(\mathcal{H}_k \ominus \mathcal{H}_{k-1}\right) = 1$, $\hphantom{xxxxxxxxxxxxxxxxxxxxxxxx}$ (7.25)

   iii.   Every subspace $\mathcal{H}_k$ is invariant for the operator A.

Conversely, let an operator A in $\mathbb{C}^n$ and a collection of subspaces $\{\mathcal{H}_k\}_{0 \leq k \leq n}$ satisfying the conditions (7.25.i)–(7.25.iii) be given, and let $e_k \in \mathcal{H}_k \ominus \mathcal{H}_{k-1}$ be unit vectors for $k = 1, 2, \ldots, n$. Then the set of vectors $\{e_k\}$ forms an orthonormal basis of $\mathbb{C}^n$ and the matrix of the operator A in this basis is upper-triangular. It turns out that this strategy

---

[8]For more information on the characteristic function of a linear operator, see also the M.S. Livšic Anniversary Volume [OTSTR], in particular, the Preface and the paper [Kats].

[9] A more elaborate presentation of scattering theory for linear stationary dynamical systems (with emphasis on applications to the wave equation in $\mathbb{R}^n$) was carried out in [LaPhi].

can be adapted to obtain analogues of Schur's theorem in infinite dimensional Hilbert spaces. The first step in this direction was taken by L.A. Sakhnovich [Sakh1] who noticed that although the proof of Schur is based on the fact that every operator A in a finite dimensional linear space over $\mathbb{C}$ has an eigenvector, that really the proof only depended upon following property of the operator A:

Property IS. *For every pair of closed invariant subspaces* $\mathcal{H}_1$ *and* $\mathcal{H}_2$ *of the operator* A *such that* $\mathcal{H}_1 \subset \mathcal{H}_2$ *and* $\dim(\mathcal{H}_2 \ominus \mathcal{H}_1) > 1$, *there exists a third closed invariant subspace* $\mathcal{H}_3$ *of the operator* A *such that* $\mathcal{H}_1 \subset \mathcal{H}_3 \subset \mathcal{H}_2$, $\mathcal{H}_3 \neq \mathcal{H}_1$, $\mathcal{H}_3 \neq \mathcal{H}_2$.

A theorem of J. von Neumann (unpublished) and of N. Aronszajn and K. Smith [ArSm], guarantees that every compact operator in a Hilbert space possesses the property IS. In [Sakh1], Sakhnovich proved that if the imaginary part C_A of the operator A is of Hilbert-Schmidt class, i.e., if

$$\sum_k \big(s_k(C_A)\big)^2 < \infty,\tag{7.26}$$

then the operator A possesses the property IS. From later results of V.I. Matsaev it follows that this condition can be relaxed: if

$$\sum_k \frac{s_k(C_A)}{k+1} < \infty,\tag{7.27}$$

then A possesses the property IS.

Let $\mathcal{H}$ be a Hilbert space and let $\mathfrak{P}_\mathcal{H}$ denote the collection of orthoprojectors onto all possible closed subspaces of $\mathcal{H}$. The set $\mathfrak{P}_\mathcal{H}$ is partially ordered: $P_1 \leq P_2$ if the corresponding ranges are ordered by inclusion, i.e., if $\mathcal{R}_{P_1} \subseteq \mathcal{R}_{P_2}$, and $P_1 < P_2$ if the inclusion of the ranges is proper. A subset $\mathfrak{P}$ of the set $\mathfrak{P}_\mathcal{H}$ that contains at least two orthoprojectors is said to be a *chain* if it is fully ordered, i.e., if the conditions $P_1 \in \mathfrak{P}$, $P_2 \in \mathfrak{P}$, $P_1 \neq P_2$ imply that either $P_1 < P_2$, or $P_2 < P_1$.

If a chain $\mathfrak{P}$ contains orthoprojectors P^- and P^+ ($P^- < P^+$) such that every orthoprojector $P \in \mathfrak{P}$ distinct from them satisfies either the inequality $P < P^-$ or the inequality $P > P^+$, then the pair (P^-, P^+) is said to be a *jump* in the chain $\mathfrak{P}$, and the dimension of the subspace $P^+\mathcal{H} \ominus P^-\mathcal{H}$ is said to be the *dimension of the jump*. A chain without jumps is said to be *continuous*.

The set of all chains in $\mathcal{H}$ can be ordered by inclusion: the chain $\mathfrak{P}_1$ is said to *precede* the chain $\mathfrak{P}_2$ (and we write $\mathfrak{P}_1 \prec \mathfrak{P}_2$) if every orthoprojector in $\mathfrak{P}_1$ also lies in $\mathfrak{P}_2$. A chain $\mathfrak{P}$ is said to be *maximal* with respect to this ordering if there is no chain $\mathfrak{P}'$ satisfying the conditions $\mathfrak{P} \prec \mathfrak{P}'$, $\mathfrak{P} \neq \mathfrak{P}'$.

Let A be a bounded operator in a Hilbert space $\mathcal{H}$ and let $\mathfrak{P}$ be a chain of orthoprojectors in $\mathcal{H}$. Then *the chain* $\mathfrak{P}$ *is said to be an eigenchain for the operator* A *if for every* $P \in \mathfrak{P}$

the subspace $P\mathcal{H}$ is invariant under A, i.e., if the equality $AP = PAP$ holds for every $P \in \mathfrak{P}$.

The following result is established by transfinite induction in [Brod2] and in [Brod4], where it appears as Theorem 15.2.

THEOREM II [M.S. Livšic-M.S. Brodskiĭ; L.A. Sakhnovich]. *Let A be a bounded linear operator in a Hilbert space $\mathcal{H}$ that satisfies the condition IS. Then there exists a maximal chain of orthoprojectors that is an eigenchain for A.*

The idea for the proof of this theorem arose in a conversation between M.S. Livšic and M.S. Brodskii. (See the historical remark in the book [Brod4], p. 278 of the Russian original or p. 234 of the English translation.) L.A. Sakhnovich gave an independent proof in [Sakh1]. This theorem can be considered as a first step in extending the Schur theorem on reducing a matrix to triangular form to the setting of a more general class of operators in Hilbert space. Based on it, Sakhnovich obtained the following result in [Sakh1]:

Every bounded linear operator A in a a separable Hilbert space that satisfies the condition IS has an inessential extension $\widetilde{A}$ which is unitarily equivalent to an integral operator of the form

$$x(t) \to (\boldsymbol{K}x)(t) = \frac{d}{dt} \int\limits_{t}^{1} K(t,s)x(s)\,ds\,, \tag{7.28}$$

acting in a space $L_{\mathcal{E}}^2([0,1])$ of vector functions $x(t)$ defined on the interval $[0,1]$ whose values belong to a Hilbert space $\mathcal{E}$, $\dim \mathcal{E} \leq \infty$, provided with the scalar product:

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle_{L_{\mathcal{E}}^2} = \int\limits_{0}^{1} \langle x(t), y(t) \rangle_{\mathcal{E}}\,dt, \quad \text{for } \boldsymbol{x} = x(t) \text{ and } \boldsymbol{y} = y(t)\,. \tag{7.29}$$

The kernel $K(t,s)$ is a function defined for $0 \leq t, s \leq 1$ whose values are bounded linear operators acting in $\mathcal{E}$.

A limitation of this last result is that the class of kernels $K(t,s)$ is not described. However, starting from this theorem, Sakhnovich obtained the following result in [Sakh2]:

Every bounded operator A in a Hilbert space $\mathcal{H}$ whose spectrum is real and whose imaginary part C_A is of Hilbert-Schmidt class has an inessential extension $\widetilde{A}$ which is unitarily equivalent to an operator of the form

$$x(t) \to H(t)x(t) + \int\limits_{t}^{1} K(t,s)x(s)\,ds\,, \tag{7.30}$$

acting in the space $L_{\mathcal{E}}^2([0,1])$ of functions defined on the interval $[0,1]$ whose values belong to a Hilbert space $\mathcal{E}$, provided with the scalar product (7.29). $H(t)$ is a function defined

on $[0, 1]$ whose values are bounded self-adjoint operators in $\mathcal{E}$: $H(t) = H^*(t)$ for $t \in [0, 1]$. The kernel $K(t, s)$ is a function defined for $0 \le t, s \le 1$ whose values are operators acting in the Hilbert space $\mathcal{E}$ that are of Hilbert-Schmidt class. The kernel K satisfies the condition

$$\int_0^1 \operatorname{trace}_{\mathcal{E}} \{(K^*K)(t, t)\} \, dt < \infty \,. \tag{7.31}$$

This result of Sakhnovich is on the one hand more general than the corresponding result of Livšic (because the condition (7.26) is less restrictive than requiring C_A to be of trace class), but on the other hand it is less concrete, since it provides less information on the form of the kernel K than the other theorem.

Further developments in this area are related to the theory of the abstract triangular representation of operators in a Hilbert space by means of an integral with respect to a chain. This integral appeared in the papers of M.S. Brodskiĭ at the end of the fifties, [Brod1], [Brod2], [Brod3]. In a short time the theory of this new integral and its applications were developed considerably. Important contributions to this theory were made by V.I. Matsaev, [Mats1], and by I.Ts. Gohberg and M.G. Krein, [GoKr3], [GoKr4], [GoKr5]. The development of this theory stimulated new analytic investigations of the spectral properties of both selfadjoint and non-selfadjoint operators.

To explain the definition of this integral, we begin with a finite-dimensional example. Let $\mathcal{H}$ be a complex n-dimensional Hilbert space, $n < \infty$. Let A be an operator in $\mathcal{H}$, and let $\{e_k\}_{1 \le k \le n}$ be an orthonormal basis in $\mathcal{H}$. Then the operator A can be written in the form

$$A = \sum_{j,k=1}^n e_j \, a_{jk} \, \langle \cdot, e_k \rangle \,, \tag{7.32}$$

where $a_{jk} = \langle A \, e_k, e_j \rangle$ are the entries of the matrix of the operator A in this basis. Let the subspaces $\mathcal{H}_k$ be defined by (7.24), and let P_k be the orthoprojector onto $\mathcal{H}_k$. The collection $\mathfrak{P}$ of the orthoprojectors

$$0 = P_0 < P_1 < \cdots < P_{n-1} < P_n = I \tag{7.33}$$

forms a maximal chain in $\mathcal{H}$. This chain is an eigenchain for A. Let

$$\Delta P_k = P_k - P_{k-1}, \quad k = 1, 2, \ldots, n \,. \tag{7.34}$$

Then, since

$$P_k - P_{k-1} = e_k \, \langle \cdot, e_k \rangle \quad \text{and} \quad e_j \, a_{jk} \, \langle ., e_k \rangle = \Delta \, P_j \, A \, \Delta P_k \,, \tag{7.35}$$

formula (7.32) can be written in the form

$$A = \sum_{j,k=1}^n \Delta P_j \, A \, \Delta P_k \,. \tag{7.36}$$

Moreover, if $a_{jk} = 0$ for some choice of j, k, then, by (7.35), $\Delta P_j A \Delta P_k = 0$ in (4.12). Thus, if the matrix a_{jk} is upper triangular, i.e., if $a_{jk} = 0$ for $j > k$, and if $a_{kk} = \lambda_k$ (an eigenvalue of the matrix A), then the representation (7.36) takes the form

$$A = \sum_{k=1}^{n} \lambda_k \, \Delta P_k + \sum_{k=2}^{n} \sum_{j=1}^{k-1} \Delta P_j \, A \Delta \, P_k \, . \tag{7.37}$$

Since $\sum_{j=1}^{k-1} \Delta P_j = P_{k-1}$, (7.37) can be rewritten in the form

$$A = \sum_{k=1}^{n} \lambda_k \, \Delta P_k + \sum_{k=1}^{n} P_{k-1} \, A \, \Delta P_k \, . \tag{7.38}$$

The first sum on the right hand side of (7.38) represents the "diagonal part" of A, the second sum represents the "super-diagonal" part with respect to the Schur basis $\{e_k\}_{1 \leq k \leq n}$. (Everything here depends on the choice of the basis.) Since the matrix of the adjoint operator A^* (with respect to the same orthonormal basis) is lower triangular, i.e., $\Delta P_j \, A^* \Delta P_k = 0$ for $j < k$, and $P_{k-1} \, A^* \, P_k = 0$, the Schur result can be expressed as follows:

For every operator A in a finite-dimensional Hilbert space there exists at least one maximal eigenchain $\mathfrak{P} = \{P_k\}_{0 \leq k \leq n}$. For every such eigenchain, the operator A admits two representations: (7.38) and (with appropriate indexing) the representation

$$A = \sum_{k=1}^{n} \lambda_k \Delta P_k + 2i \sum_{k=1}^{n} P_{k-1} \, C_A \, \Delta P_k \, , \tag{7.39}$$

where ΔP_k is defined by (7.34) and $C_A = \dfrac{A - A^*}{2i} \, .$ $\qquad\qquad\square$

The sums in (7.39) can be considered as "integrals" over the chain $\mathfrak{P}$:

$$A = \int_{\mathfrak{P}} \lambda(P) \, dP + 2i \int_{\mathfrak{P}} P \, C_A \, dP \, . \tag{7.40}$$

In the case of the finite-dimensional $\mathcal{H}$ that was just discussed, the "integrals" in (7.40) are no more than a notation for the finite sums in (7.39). It is not a problem to generalize integrals of the form $\int_{\mathfrak{P}} \lambda(P) \, dP$ to the infinite-dimensional case. This is the usual integral of a scalar function with respect to an orthogonal spectral measure. Integrals of this kind are well understood, because of their connection with needs of the theory of selfadjoint operators. However, integrals of the form

$$\mathfrak{I}(X, \, \mathfrak{P}) \overset{\text{def}}{=} \int_{\mathfrak{P}} P \, X \, dP \, . \tag{7.41}$$

for an arbitrary chain $\mathfrak{P}$ of orthoprojectors and a more or less general bounded linear operator X in an infinite dimensional Hilbert space $\mathcal{H}$ are more difficult to handle.[10]. An integral of the form (7.41) can be *defined* by means of a very natural limiting process that was introduced by M.S. Brodskiĭ [11]: as usual, certain integral sums should be constructed and then the passage to limit should be performed. The condition

$$(P^+ - P^-)X(P^+ - P^-) = 0 \ \text{ for every jump } (P - P^+) \text{ of the chain } \mathfrak{P} \tag{7.42}$$

is an evident necessary condition for the existence of the integral (7.41). However, the problem of obtaining sufficient conditions for the existence of such an integral turned out to be far more difficult. The theory of such integrals, the so called *integral of triangular truncation*, was created mainly in the works of M.S. Brodskiĭ, I.Ts. Gohberg, M.G. Krein and V.I. Matsaev and served to complete a program that was initiated by M.S. Livšic (see the remark to Theorem II'' of this section). A detailed exposition of this theory is presented in [Brod4], [GoKr2] and [GoGoK]. Brodskii proved that under condition (7.42), the integral (7.41) exists, if the operator X is of trace class $\mathfrak{S}_1$. [12] V.I. Matsaev, [Mats1], sharpened this result. He proved, that *under the condition (7.42), the integral (7.41) exists (in the sense of the convergence of integral sums with respect to the uniform operator norm), if the compact operator X belongs to the class $\mathfrak{S}_\omega$, i.e., if the condition* $\sum\limits_{1 \le k < \infty} s_k(X) \cdot k^{-1} < \infty$ *holds*. The latter result is precise in some sense. *If a compact operator X does not belong to the class $\mathfrak{S}_\omega$, then there exists a continuous maximal chain $\mathfrak{P}$ such that the integral (7.41) does not exist even in the sense of weak convergence*; see [Brod4], Lemma 22.2. In any case, *if the operator X is compact and if the integral (7.41) exists (in the sense of the convergence of integral sums with respect to the uniform operator norm), then this integral represents a Volterra operator*. We recall, that a linear operator in a Hilbert space is said to be a *Volterra operator* if it is compact and if its spectrum consists of only one point, the point zero.

The representation (7.40) of an operator A by means of the integral of triangular trunca-

---

[10] Integrals of scalar valued functions with respect to operator valued measures and integrals of operator valued functions with respect to a scalar valued measure are usually much easier to deal with than integrals of operator valued functions with respect to operator valued measures. In the integral (7.41), both the function PX and "the measure" dP are operator valued.

[11] An integral of the form (7.41) can be considered (under appropriate parametrisation of the chain $\mathfrak{P}$) as a special case of a double integral operator of the form

$$\int\limits_0^1 \int\limits_0^1 \chi(t,\, s)\, dP(t)\, X\, dP(s)\,, \ \text{ with } \ \chi(t,\, s) = 1 \ \text{ for } \ s > t, \ \chi(t, s) = 0 \ \text{ for } \ s < t\,,$$

We already met such integrals in Section 4. However, here the function χ is of a very special form, and the results which can be obtained for double operator integrals with *this* function are much more precise than the results which follow from the general theory of double integral operators.

[12] Recall that singular values of a compact operator X are the eigenvalues of the operator $\sqrt{X^*X}$ indexed in such a way that $s_1(X) \ge s_2(X) \ge s_3(X) \ge \ldots$ and that $X \in \mathfrak{S}_1$ if $\sum\limits_{k=1}^\infty s_k(X) < \infty$.

tion can be considered as a coordinate-free representation of A from its maximal eigenchain and its imaginary part. On the one hand, this representation generalizes the results of Livšic (see Theorem I$''$ and the Remark following it that focuses attention on the formulas (7.19) and (7.20)). On the other hand, the representation (7.40) is "coordinate free", i.e., it represents the operator A *itself* in the *original* Hilbert space $\mathcal{H}$, rather than a "model" operator T that acts in the "model" space L^2_{ε} and which is only unitarily equivalent to the original operator A (or even to an inessential extension $\widetilde{A}$ of A acting in a larger space $\widetilde{\mathcal{H}} \supset \mathcal{H}$). In spirit, the representation (7.40) is much closer to the original work of Schur [Sch2] than the triangular model (7.15) of Livšic. The integral representation (7.40) for a bounded linear operator A with imaginary part $C_A \in \mathfrak{S}_1$ was first obtained by Brodskii in [Brod1] using the representation (7.19)-(7.20) as a model. Brodskiĭ just transformed this representation to the coordinate free form (7.40). This proof used the theory of characteristic functions. Later, in [Brod2] and [Brod3], the representation (7.40) was obtained for *arbitrary* Volterra operators A in a Hilbert space (in which case $\lambda(t) \equiv 0$ in (7.40)), and also for bounded linear operators A with real spectrum and $C_A \in \mathfrak{S}_\omega$, independently of the theory of characteristic functions, by methods based on consideration of the eigenchains of the operators A, i.e., by generalizing the reasoning of Issai Schur.

The study of the integral of triangular truncation has led to unexpected and deep connections between the spectra of the real and imaginary components of Volterra operators. In certain cases the clarification of these connections has required the development of new analytic tools, see Chapter III of the book [GoKr2]. As an example of the application of the general results obtained in the setting of the integral of triangular truncation, we consider the Volterra operator $A = B + iC$, $B = B^*$, $C = C^*$ in the Hilbert space $L^2([0,\,1])$ that is defined by the equality

$$Ax(t) = 2i \int_t^1 h(t-s)x(s)\,ds\,,$$

where the function $h(\,\cdot\,)$ is periodic: $h(t+1) = h(t)$, Hermitian: $h(-t) = \overline{h(t)}$, and summable on $[0,\,1]$. It is easily checked that the eigenvalues $\{\xi_j\}_{-\infty}^{\infty}$ and $\{\eta_j\}_{-\infty}^{\infty}$ of the operators B and C (appropriately indexed) are related by the discrete Hilbert transform:

$$\eta_k = \frac{1}{\pi} \sum_{l=-\infty}^{\infty} \frac{\xi_l}{l - k + \frac{1}{2}}\,, \quad -\infty < k < \infty\,. \tag{7.43}$$

Consequently, it is possible to obtain estimates for the discrete Hilbert transform by applying some results on the spectra of the Hermitian components of Volterra operators (7.43).

Thus, the Schur paper [Sch2], which is elementary and purely algebraic, stimulated the creation of several deep and rich analytic theories.

REFERENCES

[ArSm] ARONSZAJN, N. and K.T. SMITH. *Invariant subspaces of completely continuous operators.* Annales of Math., **60** (1954), pp. 345 - 350.

[Bran] BRANGES, LOUIS DE. *Hilbert Spaces of Entire Functions*, Prentice Hall, Englewood Cliffs, NJ, 1968.

[Brod1] BRODSKIĬ, M.S. *Ob integral'nom predstavlenii ogranichennykh nesamosopryazhennykh operatorov s veshchestvennym spektrom* (Russian). [Integral representations of bounded non-selfadjoint operators with a real spectrum]. Dokl. Akad. Nauk SSSR, **126** (1959), 1166–1169.

[Brod2] BRODSKIĬ, M.S. *O treugol'nom predstavlenii vpolne nepreryvnykh operatorov s odnoĭ tochkoĭ spectra* (Russian). Uspehi Mat. Nauk **16**:1 (1961), pp. 135–141. English transl.:On the triangular representation of completely continuous operators with one-point spectra. Amer. Math. Soc. Transl. (2), **47** (1965), pp. 59 - 65.

[Brod3] BRODSKIĬ, M.S. *O treugol'nom predstavlenii nekotorykh operatorov s vpolne nepreryvnoĭ mnimoĭ chast'yu* (Russian). Dokl. Akad. Nauk SSSR, **133** (1960), pp. 1271 - 1274. English transl.: *Triangular representation of some operators with completely continuous imaginary part.* Soviet Math. Dokl., **1** (1960), pp. 952 - 955.

[Brod4] BRODSKIĬ, M.S. *Treugol'nye i Zhordanovy predstavleniya lineĭnykh operatorov.* (Sovremennye problemy matematiki) (Russian). Nauka, Moscow 1969, 287 pp. English translation: *Triangular and Jordan Representations of Linear Operators.* (Transl. of Math. Monographs, **32**). Amer. Math. Soc., Providence, RI, 1971.

[BroBr] BRODSKIĬ, V.M. and M.S. BRODSKIĬ. *Ob abstraktnom treugol'nom predstavlenii lineĭnykh ogranichennykh operatorov i mul'tiplikativnom razlozhenii sootvet-stvuyushchikh im kharakteristicheskikh funktsiĭ* (Russian). Dokl. Akad. Nauk SSSR, **181**:3 (1968), pp. 511 - 514. English transl.: *The abstract triangular representation for bounded linear operators and multiplicative expansions of their respective characteristic functions.* Soviet Math. Dokl., **9** (1968), pp. 846 - 850.

[BroLi] BRODSKIĬ, M.S. and M.S. LIVSHITS (=M.S. LIVŠIC). *Spektral'nyĭ analiz nesamoso-pryazhennykh operatorov i promezhutochnye sistemy*(Russian). Uspehi Mat. Nauk (N.S.) **13**:1 (1958), pp. 3 - 85. English transl.: *Spectral analysis of non-selfadjoint operators and intermediate systems.* Amer. Math. Soc. Transl. (2) **13** (1960), pp. 265 - 346.

[Dav] DAVIDSON, K.R. *Nest Algebras.* Triangular forms for operator algebras on Hilbert space. (Pitman Research Notes in Mathematics Series, **191**). Longman, Harlow & Wiley New York 1988.

[GoGoK] GOHBERG, I.TS., S. GOLDBERG and M. KAASHOEK. *Classes of Linear Operators. Vol.* II. (Operator Theory: Advances and Applications. **OT 63**). Birkhäuser Verlag, Basel 1993.

[GoKr2] GOHBERG, I.TS. and M.G. KREIN. *Theoriya Vol'terrovykh Operatorov v Gil'bertovom Prostranstve i ee Prilozheniya* (Russian), Nauka, Moscow 1967. English transl.: *Theory and Applications of Volterra Operators in Hilbert Space.* (Transl. of Mathem. Monogr., **24**), Amer. Math. Soc., Providence 1970.

[GoKr3] GOHBERG, I.TS. and M.G. KREIN. *O vpolne nepreryvnykh operatorakh so spectrom sosredotochennym v nule* (Russian). [Completely continuous operators with a spectrum conzentrated at zero]. Doklady Akad. Nauk SSSR, **128**:2 (1959), pp. 227 - 230.

[GoKr4] GOHBERG, I.TS. and M.G. KREIN. *K teorii treugol'nykh predstavleniĭ nesamoso-pryazhennykh operatorov* (Russian). Doklady Akad. Nauk SSSR, **137**:5 (1961), pp. 1034 - 1037. English transl.: *On the theory of triangular representations of linear operators.* Soviet Math., Doklady, **2** (1961), pp. 392 - 396.

[GoKr5] GOHBERG, I.TS. and M.G. KREIN. *O vol'terrovykh operatorakh s mnimoĭ komponentoĭ togo ili inogo klassa* (Russian). Doklady Akad. Nauk SSSR, **139**:4 (1961), pp. 779 - 782. English transl.: *Volterra operators with imaginary component in one class ore another.* Soviet Mat., Doklady, **2** (1961), pp. 983 -985.

[Herr] HERRERO, D.A. *Triangular operators.* Bull. Lond. Math. Soc., **23** (1991), pp. 513 - 554.

[Kats] KATSNELSON, V. *Right and left joint system representation of a rational matrix function in general position.*

[LaPhi] LAX, P. and R. PHILLIPS. *Scattering Theory.* Academic Press, New York·London 1967.

[Liv1] LIVŠIC, M.S. (=LIVSHITS, M.S.) *Ob odnom classe lineĭnykh operatorov v gil'bertovom prostranstve* (Russian). Matem. Sbornik, **19** (1946), pp. 239 - 260. English transl.:

[Liv2] LIVŠIC, M.S. (=LIVSHITS, M.S.) *Izometricheskie operatory s ravnymi defektnymi chislami, kvazi-unitarnye operatory* (Russian). Matem. Sbornik, **26** (1950), pp. 247 - 264. English translation:

[Liv3] LIVŠIC, M.S. (=LIVSHITS, M.S.) *O privedenii lineĭnykf ermitovykh operatorov k "treugol'nomu" vidu.* Doklady Akad. Nauk SSSR, **84**:5, pp. 873 - 876.

[Liv4] LIVŠIC, M.S. (=LIVSHITS, M.S.) *O rezol'vente lineĭnogo nesimmetricheskogo operatora.* Doklady Akad. Nauk SSSR, **84**:6, pp. 1131 - 1134.

[Liv5] LIVŠIC, M.S. (=LIVSHITS, M.S.) *O spektral'nom razlozhenii lineĭnykh nesamoso-pryazhennykh operatorov.* Matem. Sbornik, **34** (1954), pp. 145 - 199 (Russian),. English transl.: *On the spectral resolution of linear non-selfadjoint operators.* Amer. Math. Soc. Transl. (Ser.2), **5** (1957), pp. 67 - 114.

[Liv6] LIVŠIC, M.S. (=LIVSHITS, M.S.) Zhurnal Experimental'noĭ i Theor. Phys., **31**:1, pp. 121-131 (Russian). English transl.: *The application of non-self-adjoint operators to scattering theory.* Soviet Physics JETP, **4**:1 (1957), pp. 91 - 98.

[Liv7]   LIVŠIC, M.S. (=LIVSHITS, M.S.) *Metod nesamosopryazhennykh operatorov v teorii rasseyaniya* (Russian). Uspekhi Matem. Nauk, **12**:1 (1957), pp. 212 - 218 (Russian). Engl. transl.: *The method of non-self-adjoint operators in dispersion theory.* Amer. Math. Soc. Transl. (2), **16** (1960), pp.427-434.

[Liv8]   LIVŠIC, M.S.. (= LIVSHITS, M.S.) *Operatory, Kolebanija, Volny (Otkrytye sistemy)*, Nauka, Moscow, 1966 (Russian). English transl.: *Operators, oscillations, waves (open systems).* (Translations of Mathematical Monographs, Vol. **34**.) American Mathematical Society, Providence, R.I., 1973.

[Mats1]   MATSAEV, V.I. (= MACAEV, V.I.) *Ob odnom klasse vpolne nepreryvnykh operatprov* (Russian). Doklady Akad. Nauk SSSR, **139**:3 (1961), pp. 548 - 551. English transl.: *On a class of completely continuous operators.* Soviet Math., Doklady, **2** (1961), pp. 972 - 975.

[Mats2]   MATSAEV, V.I. (= MACAEV, V.I.) *O vol'terrovykh operatorakh, poluchaemykh vozmushcheniem samosopryazhennykh* (Russian). Doklady Akad. Nauk SSSR, **139**:4 (1961), pp. 810 - 813. English transl.: *Volterra operators produced by perturbation of self-adjoint operators.* English thransl.: Soviet Math., Doklady, **2** (1961), pp. 1013 - 1016.

[NiVa]   NIKOLSKIĬ, N.K. and V.I. VASYUNIN. *A unified approach to functional models, and the transcription problem.* The Gohberg anniversary collection, Vol.II. (Operator Theory: Advances and Applications. **OT 41**). Birkhäuser, Basel 1989, pp. 403 - 434.

[OTSTR]   *Operator Theory, System Theory and Related Topics.(The Moshe Livšic Anniversary Volume).* ALPAY, D. and V. VINNIKOV-editors. (Operator Theory: Advances and Applications, **OT 123**), Birkhäuser Verlag, Basel, 2001.

[Sakh1]   SAKHNOVICH, L.A. (= SAHNOVIČ, L.A.) *O privedenii nesamosopryazhennykh operatorov k treugol'nomu vidu [The reduction of non-selfadjoint operators to triangular form - in Russian].* Izvestiya Vysš. Učebn. Zaved. Matematika 1959, no. 1, pp. 180 - 186.

[Sakh2]   SAKHNOVICH, L.A. (= SAHNOVIČ, L.A.) *Issledovanie treugol'noĭ modeli nesamosopryazhennykh operatorov [A study of the triangular form of non-selfadjoint operators - in Russian].* Izvestiya Vysš. Učebn. Zaved. Matematika 1959, no. 4, pp. 141 - 149.

[Sch2]   SCHUR, I.: *Über die charakteristischen Wurzeln einer linearen Substitution mit einer Anwendung auf die Theorie der Integralgleichungen [On the characteristic roots of a linear substitution with an application to the theory of integral equations - in German].* Mathematische Annalen, **66** (1909), pp. 488 - 510. Reprinted in: [Sch: Ges], Vol. I, pp. 272 - 294.

[Sch: Ges] SCHUR, I.: *Gesammelte Abhandlungen [Collected Works]. Vol. I, II, III.* Springer-Verlag, Berlin·Heidelberg·New York 1973.

[SzNFo]   *Sz.-Nagy, B.* and C. Foias. *Analyse Harmonique des Opérateurs de l'espace de Hilbert* (French). Masson and Académiae Kiado, 1967. English transl.:*Harmonic Analysis of Operators in Hilbert Space.* North Holland, Amsterdam 1970.

8. Sequences of multipliers that preserve the class of polynomials with only real zeros, and entire functions of the Laguerre-Polya-Schur class.

Problems related to the distribution of the zeros of polynomials have attracted the attention of mathematicians for a long time. In particular, the following question has generated considerable interest. *How many real zeros does a given a polynomial with real coefficients have?* There are several methods for either estimating or determining precisely the number of zeros of such a polynomial that belong to a given interval (a, b) of the real axis. These include the Descartes' rule of signs, the Budan-Fourier algorithm, the Sturm algorithm and methods based on Hermitian forms. These methods are presented in old books on algebra ([Web], Vol. I, [Kur]), as well as in books devoted to the zeros of polynomials, [Obr], [Mar], [Dieu]. The article [KrNa] contains a detailed survey of the method of Hermitian forms for the separation of the zeros of polynomials. A lot of additional material on the distribution of roots of polynomials can be found in [PoSz], Part V.

In this section we shall focus on a different class of results that deal with transformations that preserve the class of polynomials $P(t) = p_0 + p_1 t + p_2 t^2 \ \cdots \ + p_n t^n$ with real coefficients for which

$$\#_{nr}(P) = \text{ the number of non real roots of the polynomial } P(t)$$

is equal to zero. The simplest result of this kind states that if $P(t)$ is a polynomial with real coefficients and $\alpha \in \mathbb{R}$, then

$$\#_{nr}(P) = 0 \Longrightarrow \#_{nr}(\alpha P + P') = 0.$$

If the roots of $P(t)$ are distinct, then this follows easily from Rolle's theorem applied to $e^{-\alpha t} P(t)$.

THEOREM *Let $P(t) = p_0 + p_1 t + p_2 t^2 \ \cdots \ + p_n t^n$ and $Q(t) = q_0 + q_1 t + q_2 t^2 + \ \cdots \ + q_m t^m$ be polynomials with real coefficients and assume that $\#_{nr}(Q) = 0$. Then the following conclusions hold:*

(1) [Hermite] $\#_{nr}(Q(\frac{d}{dt})P) \leq \#_{nr}(P)$.

(2) [Laguerre] If also the roots of $Q(t)$ fall outside the interval $[0, n]$, then

$$\#_{nr}(Q(0)p_0 + Q(1)p_1 t + Q(2)a_2 t^2 + \ \cdots \ + \ Q(n)p_n t^n) \leq \#_{nr}(P).$$

(3) [Malo] If also the roots of $Q(t)$ are either all positive or all negative and $l = \min(n, m)$, then

$$\#_{nr}(P) = 0 \Longrightarrow \#_{nr}(p_0 q_0 + p_1 q_1 t + \cdots + p_l q_l t^l) = 0.$$

(4) [Schur] If also the roots of $Q(t)$ are either all positive or all negative and $l = \min(n, m)$, then

$$\#_{nr}(P) = 0 \Longrightarrow \#_{nr}(p_0 q_0 + 1! p_1 q_1 t + 2! p_2 q_2 t^2 + \cdots + l! p_l q_l t^l) = 0.$$

The cited results may be found in [Obr], [Lag1]), [Mal] and [Sch7], respectively; see also [PoSz], Part V, Chapter $1, \S 5$, no. **63** and **67** for the first two.

The last three three statements of the theorem deal with an operation of the form

$$p_0 + p_1 t + \cdots + p_n t^n \longrightarrow \gamma_0 p_0 + \gamma_1 p_1 t + \cdots + \gamma_n p_n t^n. \tag{8.1}$$

In the particular case considered by Schur, $\gamma_k = q_k$ for $k \leq m$, and $\gamma_k = 0$ for $k > m$, where the q_k are obtained from the coefficients of a polynomial $Q(t)$ with only negative roots that we now write as $Q(t) = q_0 + \frac{q_1}{1!} t + \frac{q_2}{2!} t^2 + \cdots + \frac{q_m}{m!} t^m$. The importance of this result is that it admits a converse: Every sequences $\{\gamma_k\}_{0 \leq k < \infty}$ for which the operation (8.1) preserves the class of polynomials with real coefficients and $\#_{nr}(P) = 0$ is either generated by a polynomial $Q(t) = q_0 + \frac{q_1}{1!} t + \frac{q_2}{2!} t^2 + \cdots + \frac{q_m}{m!} t^m$ with only negative zeros, or belongs to the *closure* of sequences generated by such polynomials. A full description of this class of sequences $\{\gamma_k\}_{0 \leq k < \infty}$ is presented in the paper [PS] by G. Polya and I. Schur and will now be described briefly below.

Given an infinite unilateral sequence

$$\Gamma = \{\gamma_0, \ \gamma_1, \ \gamma_2 \ \ldots \ \gamma_k, \ \ldots\} \tag{8.2}$$

of real numbers, let $\Phi(t)$ denote the (formal) power series

$$\Phi(t) = \sum_{k=0}^{\infty} \frac{\gamma_k}{k!} t^k \tag{8.3}$$

and, for any polynomial

$$P(t) = p_0 + p_1 t + \cdots + p_n t^n, \tag{8.4}$$

let $\Gamma[P(t)]$ denote the new polynomial:

$$\Gamma[P(t)] = \gamma_0 a_0 + \gamma_1 a_1 t + \gamma_2 a_2 t^2 + \cdots + \gamma_n a_n t^n. \tag{8.5}$$

DEFINITION I ([PS]). I. *The sequence (8.2) is said to be a **sequence of multipliers of the first type** if for every polynomial $P(t)$ with real coefficients (of arbitrary degree n),*

$$\#_{nr}(P) = 0 \Longrightarrow \#_{nr}(\Gamma[P]) = 0.$$

II. *The sequence (8.2) is said to be a **sequence of multipliers of the second type** if for every polynomial $P(t)$ with real coefficients (of arbitrary degree n) and only negative zeros, $\#_{nr}(\Gamma[P]) = 0$.*

DEFINITION II ([PS]).   I. *An entire function $\Phi(t) \not\equiv 0$ is an **entire function of the first type**, if it admits a multiplicative representation of the form*

$$\Phi(t) = c\, t^l\, e^{\alpha t} \prod_k \left(1 + t\, \delta_k\right) , \tag{8.6}$$

where $c \neq 0$ is a real number, l is a non-negative integer, α is a non-negative real number, and the δ_k are non-negative numbers that satisfy the condition $\sum_k \delta_k < \infty$.

II. *An entire function $\Phi(t) \not\equiv 0$ is an **entire function of the second type**, if it admits a multiplicative representation of the form*

$$\Phi(t) = c\, t^l\, e^{-\beta t^2 + \alpha t} \prod_k \left(1 + t\, \delta_k\right) e^{-\delta_k t} , \tag{8.7}$$

where $c \neq 0$ is a real number, l is a non-negative integer, β is a non-negative number, α is a real number, and the δ_k are real numbers that satisfy the condition $\sum_k (\delta_k)^2 < \infty$.

THEOREM (G. Polya and I. Schur, [PS]).   I. *If the sequence $\gamma_0, \gamma_1, \ldots, \gamma_k, \ldots$ is a sequence of multipliers of the first (respectively the second) type, then the series (8.3) converges in the whole complex plane, and the entire function $\Phi(t)$ which is represented by this series is an entire function of the first (respectively the second) type.*

II. *If $\Phi(t)$ is an entire function of the first (respectively the second) type, and (8.3) is its Taylor expansion, then the sequence $\gamma_0, \gamma_1, \ldots, \gamma_k, \ldots$ is a sequence of multipliers of the first (respectively the second) type.*

This theorem gives a full description of the sequences of multipliers of both the first and second type. The appearance of two types of multipliers (and two types of entire functions) corresponds to the fact that in the Schur theorem from [Sch7] that was stated above, the polynomials $P(t)$ and $Q(t)$ appear in a *symmetric way*: if *one* of the polynomials $P(t) = p_0 + p_1 t + p_2 t^2 + \cdots$ or $Q(t) = q_0 + q_1 + q_2 t^2 + \cdots$ has only real zeros, and the *other* has only negative zeros, then all the zeros of the polynomial $p_0 q_0 + 1! p_1 q_1 t + 2! p_2 q_2 t^2 + \cdots$ are real. Thus, roughly speaking, sequences of the first (respectively second) type act on polynomials that are entire functions of the second (respectively first) type. The two types of entire functions arise as limits of the two classes of polynomials:

THEOREM (E. Laguerre, [Lag2]; G. Pólya, [Pol1]).

I. *Let $\Phi(t)$ be an entire function of the first type (respectively the second type). Then there exists a sequence $\{\Phi_n(t)\}_{n=1, 2, \ldots}$ of polynomials such that the zeros of $\{\Phi_n(t)\}$ lie*

in the negative half-axis (respectively the real axis) and $\{\Phi_n(t)\}$ converges to $\Phi(t)$ locally uniformly in the whole complex plane.

II. *If a sequence of polynomials $\{\Phi_n(t)\}_{n=1,2,\ldots}$ converges uniformly in a neighborhood of the origin to a function that is not identically equal to zero and if all the zeros of every polynomial Φ_n lie in the negative half-axis (respectively the real axis), then the sequence $\{\Phi_n\}_{n=1,2,\ldots}$ converges locally uniformly in the whole complex plane and the limit function $\Phi(t)$ is an entire function of the first (respectively second) type.*

Part I of this theorem was obtained by Laguerre, [Lag2]; part II was obtained by Polya, [Pol1]. Laguerre obtained a weak version of part II. Namely, he assumed that the sequences of polynomials $\{\Phi_n\}$ considered above converge in the whole complex plane, not just in a neighborhood of the origin, and deduced the same properties of the limiting function that are stated in part II of the preceding theorem. This result of Laguerre is not strong enough to obtain a description of the multiplier sequences, the stronger result by Pólya is needed. The theorems of Laguerre and Polya, and some generalizations, can be found in [HiWi], Chapter III, § 3, and in [Lev], Chapter VIII.

The paper [PS] by Pólya and Schur served as a source of inspiration for the investigations of I.J. Schoenberg on the representation of totally positive functions and sequences. The notion of total positivity was introduced by Schoenberg in [Scho1].

DEFINITION III. *A real function (or, in other terms, kernel) $K(t,s)$ of two variables ranging over linearly ordered sets $\mathcal{T}$ and $\mathcal{S}$, respectively, is said to be **totally positive** if for every*[13] *m and for every*

$$t_1 < t_2 < \cdots < t_m, \quad s_1 < s_2 < \cdots < s_m \quad t_i \in \mathcal{T}, t_j \in \mathcal{S}, \qquad (8.8)$$

the inequalities

$$K\left(\begin{array}{c} t_1,\, t_2,\, \ldots\, , t_m \\ s_1,\, s_2,\, \ldots\, , s_m \end{array}\right) \geq 0 \qquad (8.9)$$

hold, where

$$K\left(\begin{array}{c} t_1,\, t_2,\, \ldots\, , t_m \\ s_1,\, s_2,\, \ldots\, , s_m \end{array}\right) = \det \begin{bmatrix} K(t_1,\, s_1) & K(t_1,\, s_2) & \cdots & K(t_1,\, s_m) \\ K(t_2,\, s_1) & K(t_2,\, s_2) & \cdots & K(t_2,\, s_m) \\ \vdots & \vdots & & \vdots \\ K(t_m,\, s_1) & K(t_m,\, s_2) & \cdots & K(t_m,\, s_m) \end{bmatrix}. \qquad (8.10)$$

Usually $\mathcal{T}$ and $\mathcal{S}$ are either subintervals of the real axis (that may coincide with the full axis), or countable sets of real numbers such as the set of all integers or the set of all

---

[13]If both sets $\mathcal{T}$ and $\mathcal{S}$ are infinite, then m can be an arbitrary natural number; if at least one of the sets $\mathcal{T}$ or $\mathcal{S}$ is finite then m can be an arbitrary natural number satisfying the restriction $m \leq \min\{|\mathcal{T}|, |\mathcal{S}|\}$, where $|\mathcal{M}|$ denotes the cardinality of the set $\mathcal{M}$.

non-negative integers, or even finite sets of integers. If $\mathcal{T}$ and $\mathcal{S}$ are sets of integers, then K can be viewed as a matrix; in this case, K is referred to as a *totally positive matrix*.

A concept that is more general than total positivity is *sign regularity*.

A function $K(t, s)$ is said to be **sign regular** if there exists a sequence of numbers ε_m, each of which is equal to either $+1$ or -1, such that in the setting of (8.8), the inequalities

$$\varepsilon_m \; K \left(\begin{array}{c} t_1, \; t_2, \; \ldots \; , t_m \\ s_1, \; s_2, \; \ldots \; , s_m \end{array} \right) \geq 0 \tag{8.11}$$

hold.

Totally positive matrices (and kernels) have very interesting spectral properties that were discovered by F.R. Gantmacher and M.G. Krein, [GaKr1], [GaKr2], [GaKr3]. All the eigenvalues of a totally positive matrix are positive and distinct[14]. Moreover, its eigenvectors posses oscillatory properties that are analogous to the oscillatory properties of the eigenfunctions of Sturm-Liouville differential equations. A presentation of the spectral properties of totally positive matrices and kernels can also be found in the survey article [Pink] by A. Pinkus. However, the notion of total positivity was introduced by Schoenberg [Scho1] in his study of *variation-diminishing* kernels. Strictly speaking, in [Scho1], the definitions of total positivity and sign regularity were formulated for the case of *finite matrices*; generalizations to wider settings were developed later by Schoenberg himself and by S. Karlin. (See the book [Kar] for the references and for the history.)

Let $\mathcal{V}[z_1, z_2, \ldots, z_l]$ denote the number of sign changes of a given sequence $[z_1, z_2, \ldots, z_l]$ of real numbers, when the zero terms are discarded. For example, $\mathcal{V}[1, 0, 1, 0, -1] = 1$ and $\mathcal{V}[1, -1, 1, 1, -1, 1] = 3$.

DEFINITION IV. *Let* $K = \begin{bmatrix} k_{ij} \end{bmatrix}$ *be a* $p \times q$ *matrix with real entries* k_{ij}, $1 \leq i \leq p$, $1 \leq j \leq q$; $p, q < \infty$. *The matrix* K *is said to be* **variation-diminishing**, *if for every sequence* $x = [x_1, x_2, \ldots, x_q]$ *of real numbers, the sequence*

$$y_i = \sum_{1 \leq j \leq q} k_{ij} x_j, \quad (1 \leq i \leq p), \tag{8.12}$$

enjoys the property

$$\mathcal{V}[y_1, y_2, \ldots, y_p] \leq \mathcal{V}[x_1, x_2, \ldots, x_q]. \tag{8.13}$$

THEOREM (I.J. Schoenberg, [Scho1]). *Let* K *be a* $p \times q$ *matrix with real entries.*

I. *If the matrix* K *is sign-regular (in particular, if* K *is totally positive), then* K *is variation-diminishing.*

---

[14]Under the assumption that *all* its minors are strictly positive

II. *If the matrix K is variation-diminishing, and*

$$\operatorname{rank} K = q, \qquad (8.14)$$

then K is sign-regular.

Under the additional restriction (8.14), this theorem gives necessary and sufficient conditions for a $p \times q$ matrix K with real entries to be variation-diminishing. A characterization of variation-diminishing matrices without any restrictions was obtained by Th. Motzkin in his PhD Thesis (Basel, 1934). His thesis was published in 1936, [Mot1]; see also [Mot2]. Additional characterizations of matrix variation diminishing transforms can also be found in Chapter 4, § 8 of [HiWi], and Chapter 5, §§ 1, 2 of [Kar]. The latter is a storehouse of wisdom on total positivity, variation diminishing transformations and related issues and applications.

For a function $x(t)$ which is defined on a linearly ordered set $\mathcal{T}$, the number of sign changes $\mathcal{V}[x(t)]$ is defined as $\mathcal{V}[x(t)] = \sup \mathcal{V}[x(t_1), x(t_2), \ldots, x(t_l)]$, where the *supremum* is taken over all $t_1, t_2, \ldots, t_l$ from $\mathcal{T}$ such that $t_1 < t_2 < \ldots < t_l$. (It is possible that $\mathcal{V}[x(t)] = \infty$).

For a real-valued kernel $K(t, s)$ defined for $t \in \mathcal{T}$, $s \in \mathcal{S}$, where $\mathcal{T}$ and $\mathcal{S}$ are subintervals (finite or infinite) of the real axis, the variation diminishing property also can be formulated in the form

$$\mathcal{V}[y(t)] \le \mathcal{V}[x(s)] \quad (t \in \mathcal{T}, \, s \in \mathcal{S}), \qquad (8.15)$$

where

$$y(t) = \int_c^d K(t, s) \, x(s) \, ds, \quad a \le t \le b. \qquad (8.16)$$

Of course some restrictions have to be imposed on the class of functions $x(s)$ and on the kernel $K(t, s)$ to ensure the existence of the transformation (8.16) and the possibility of counting uniquely[15] the number of changes of sign of the functions $x(s)$ and $y(t)$.

The preceding theorem of Schoenberg that characterizes matrices with variation-diminishing properties in terms of their sign-regularity, can be extended to continuous kernels $K(t, s)$,.

It should be mentioned that as early as 1912, in order to estimate the number of real zeros of polynomials with real coefficients, M. Fekete considered formal power series with real coefficients $\sum_{0 \le i < \infty} c_i t^i$ that possess the following property: *for a given natural number r, all the determinants (8.18) with $m = 1, 2, \ldots, r$ are non-negative.* Power series with this property are called *r*-time positive. Multiplication by such a power series $\sum_{0 \le i < \infty} c_i t^i$,

---

[15] Since the number of changes of sign of a function is defined *pointwise*, the functions $x(s)$ and $y(t)$ must be defined *everywhere*, not just *almost everywhere* on the appropriate intervals.

transforms the power series $\sum_{0 \le k \le r} x_k t^k$ into the power series $\sum_{0 \le j < \infty} y_j t^j$ according to the rule

$$\sum_{0 \le j < \infty} y_j t^j = \left(\sum_{0 \le i < \infty} c_i t^i \right) \cdot \left(\sum_{0 \le k < \infty} x_k t^k \right),$$

or, equivalently,

$$y_j = \sum_{0 \le k \le j} c_{j-k} x_k, \quad 0 \le j < \infty.$$

Fekete formulated the following statement (see footnote number six in [Fek]):

Let $\sum_{0 \le i < \infty} c_i t^i$ be an r-time positive (formal) power series, and let $\sum_{0 \le k < \infty} x_k t^k$ be a polynomial of degree r (i.e., $x_k = 0$ for $k > r$) with real coefficients. Then

$$\mathcal{V}[y_0, y_1, y_2, \ \ldots \ , y_j, \ \ldots] \le \mathcal{V}[x_0, x_1, \ \ldots \ , x_r].$$

Totally positive matrices and kernels that depend on the difference of their arguments are of special interest. Let $\{c_i\}_{-\infty < i < \infty}$ be an infinite bilateral sequence and let the infinite Toeplitz matrix K be defined in terms of this sequence by the rule

$$K = \left[k_{p,q} \right]_{0 \le p,q < \infty}, \quad k_{p,q} \overset{\text{def}}{=} c_{p-q}. \tag{8.17}$$

If the sequence c_i is unilateral: c_i are defined only for $i \ge 0$, we first extend the original sequence to the set of all integers by setting $c_i \overset{\text{def}}{=} 0$ for $i < 0$ and then define the Toeplitz matrix K by rule (8.17) applied to the extended sequence.

DEFINITION V (I.J. Schoenberg, [Scho2]). *The real valued infinite sequence $\{c_i\}$, bilateral or unilateral, is said to be* **totally positive** *if the matrix (8.17) is totally positive, i.e., if for every natural m and for every choice of integers $p_1 < p_2 < \ \ldots \ < p_m$, $q_1 < q_2 < \ \ldots \ < q_m$ the inequality*

$$\det \begin{bmatrix} c_{p_1-q_1} & c_{p_1-q_2} & \cdots & c_{p_1-q_m} \\ c_{p_2-q_1} & c_{p_2-q_2} & \cdots & c_{p_2-q_m} \\ \cdot & \cdot & \cdot & \cdot \\ c_{p_m-q_1} & c_{p_m-q_2} & \cdots & c_{p_m-q_m} \end{bmatrix} \ge 0 \tag{8.18}$$

holds.

DEFINITION VI (I.J. Schoenberg, [Scho2]). *A real valued function $\Lambda(t)$ that is defined for all $t \in (-\infty, \infty)$ is said to be* **totally positive** *if it satisfies the following three conditions:*

i. *The kernel $K(t, s) \overset{\text{def}}{=} \Lambda(t - s), \ -\infty < t, s < \infty$, is totally positive, i.e., for every natural number m and for every $t_1 < t_2 < \ \cdots \ < t_m$, $s_1 < s_2 < \ \cdots \ < s_m$ the*

inequality

$$\det \begin{bmatrix} \Lambda(t_1 - s_1) & \Lambda(t_1 - s_2) & \cdots & \Lambda(t_1 - s_m) \\ \Lambda(t_2 - s_1) & \Lambda(t_2 - s_2) & \cdots & \Lambda(t_2 - s_m) \\ . & . & . & . \\ \Lambda(t_m - s_1) & \Lambda(t_m - s_2) & \cdots & \Lambda(t_m - s_m) \end{bmatrix} \geq 0 \qquad (8.19)$$

holds.

ii. *The function $\Lambda(t)$ is measurable.*

iii. *The function $\Lambda(t)$ is positive for at least two distinct values of t.*

$\square$

It is not difficult to prove that if a function $\Lambda(t)$ is defined on $\mathbb{R}$ and is nonnegative there, and if the inequalities (8.19) hold for $m = 2$ and for all $t_1 < t_2$, $s_1 < s_2$, then the function $\psi(t) = -\ln \Lambda(t)$ is convex (in the wide sense [16]) on $\mathbb{R}$. In particular, *if a function $\Lambda(t)$ is totally positive, then the function $-\ln \Lambda(t)$ is convex (in the wide sense) on $\mathbb{R}$.* Therefore, *for every totally positive function $\Lambda(t)$ the limits*

$$\alpha = \lim_{t \to -\infty} \frac{-\ln \Lambda(t)}{t}, \quad \beta = \lim_{t \to +\infty} \frac{-\ln \Lambda(t)}{t}, \qquad (8.20)$$

exist and $-\infty \leq \alpha \leq \beta \leq \infty$. The equality $\alpha = \beta$ holds if and only if $\Lambda(t)$ is of the form

$$\Lambda(t) = e^{kt+l}, \quad -\infty < t < \infty, \text{ for some real constants } k \text{ and } l. \qquad (8.21)$$

(A function of the form (8.21) is easily seen to be totally positive since all the determinants (8.19) vanish, if $m \geq 2$.) Thus, *if a function $\Lambda(t)$ is totally positive, but not of the form (8.21), then $\alpha < \beta$* and hence the two-sided Laplace transform $\int_{\mathbb{R}} \Lambda(t)e^{zt}dt$ exists for all points z in the open strip $\alpha < \operatorname{Re} z < \beta$ of the complex z-plane and represents a holomorphic function there. Moreover, the function, represented by this Laplace transform, takes strictly positive values for $z \in (\alpha, \beta) \subset \mathbb{R}$. Hence, the reciprocal function $\Psi(z)$:

$$\Psi(z) \overset{\text{def}}{=} \left(\int_{\mathbb{R}} \Lambda(t)\, e^{zt}dt \right)^{-1}, \quad \alpha < \operatorname{Re} z < \beta, \qquad (8.22)$$

is meromorphic in the strip $\alpha < \operatorname{Re} z < \beta$, holomorphic in all points of the interval $(\alpha, \beta) \in \mathbb{R}$, and takes strictly positive finite values in this interval: $0 < \Psi(x) < \infty$, $\alpha < x < \beta$.

---

[16] A function defined on $\mathbb{R}$ is said to be convex *in the wide sense* if it is convex in the usual sense on some subinterval of $\mathbb{R}$, which can coincide with $\mathbb{R}$, can be finite or semi-infinite. and is equal to $+\infty$ on the complement of this interval. A non-negative function $\Lambda(t)$ on $\mathbb{R}$ is convex in the wide sense if and only if the inequalities (8.19) hold for $m = 2$ and for all $t_1 < t_2$, $s_1 < s_2$.

THEOREM I (I.J. Schoenberg, Theorem 1 in [Scho2]).

I. *Let $\Lambda(t)$ be a totally positive function that is not of the form (8.21) and let the function $\Psi(z)$ be defined by means of (8.22) as a meromorphic function in the strip $\alpha < \operatorname{Re} z < \beta$ (see (8.20)).*

Then $\Psi(z)$ is holomorphic in this strip and admits an analytic continuation to the whole complex plane $\mathbb{C}$. The continued function (denoted by $\Psi(z)$ as well) is an entire function of the second type,[17] *which is not of the form*

$$\Psi(z) = ce^{az}, \quad \text{where } a \text{ and } c \text{ are real constants, } c \neq 0. \tag{8.23}$$

The function $\Lambda(t)$ can be recovered from $\Psi(z)$ by mean of the inversion formula

$$\Lambda(t) = \frac{1}{2\pi i} \int_{\gamma-i\infty}^{\gamma+i\infty} \frac{1}{\Psi(z)} e^{-zt} dz, \quad -\infty < t < \infty, \tag{8.24}$$

where γ is an arbitrary[18] *real number from (α, β).*

II. *Let $\Psi(z)$ be an entire function of the second type that is not of the form (8.23), and let $\Psi(x)$ be strictly positive on an interval $(\alpha, \beta) \in \mathbb{R}$ (so that the reciprocal function $\dfrac{1}{\Psi(z)}$ is holomorphic in the vertical strip $\alpha < \operatorname{Re} z < \beta$). Let the function $\Lambda(t)$ be defined from this $\Psi(z)$ by means of the integral.*[19] *Then the function $\Lambda(t)$ is totally positive, and if the interval (α, β) is the maximal interval on which the function $\Psi(x)$ is positive,*[20] *then the endpoints α and β of this interval coincide with the limits α and β in (8.20).*

The proof of this theorem is based essentially on methods and results from the paper [PS]. Indeed, the names of Polya and Schur (and Laguerre) appear in the title of [Scho2], and as Schoenberg himself writes *"A proof of Theorem 1 is essentially based on the results and methods developed by Polya and Schur. The only additional element required is a set of sufficient conditions insuring that a linear transformation be variation diminishing."*

For the sake of added perspective, we shall sketch the proof of part II, which is not difficult (once the theorem has been formulated), but shall omit the proof of part I, which is not so simple and straightforward. If the function $\Psi(z)$ is "a linear factor", i.e., if $\Psi(z) = (1 + \delta z)$, $\delta \neq 0$ when $1 + \delta\gamma > 0$ and $\Psi(z) = -(1 + \delta z)$, $\delta \neq 0$ when $1 + \delta\gamma < 0$,

---

[17] In the sense of the paper [PS], see Definition II above in this section.

[18] The value of the integral in (8.24) does not depend on the choice of $\gamma \in (\alpha, \beta)$.

[19] The integral (8.24) converges absolutely for every entire function $\Psi(z)$ of the second type, except when $\Psi(z)$ is of the form $ce^{at}(1 + \delta z)$, where $c \neq 0$, $\delta \neq 0$, c, δ, a are real, in which case the integral (8.24) converges in the sense of principal values.

[20] That is, if either $\Psi(\alpha) = 0$, or $\alpha = -\infty$, and if either $\Psi(\beta) = 0$, or $\beta = \infty$.

then

$$\Lambda(t) = \begin{cases} \delta^{-1}e^{t/\delta}, & t < 0 \\ 0, & t > 0 \end{cases}, \text{ if } \gamma > -1/\delta; \quad \Lambda(t) = \begin{cases} 0, & t < 0 \\ \delta^{-1}e^{t/\delta}, & t > 0 \end{cases}, \text{ if } \gamma < -1/\delta.$$
(8.25)

This function $\Lambda(t)$ is totally positive. If the formula (8.24) is used to construct the function $\Lambda_j(t)$ from $\Psi_j(z)$, $j = 1, 2$, and the function $\Lambda(t)$ from the product $\Psi(z) = \Psi_1(z)\Psi_2(z)$, and if the same γ is used for all three constructions, then

$$\Lambda(t) = \int_{-\infty}^{\infty} \Lambda_1(t - \xi)\Lambda_2(\xi)\, d\xi.$$
(8.26)

Moreover, if Λ_1 and Λ_2 are totally positive, then the function Λ is totally positive as well. Therefore, if $\Psi(z) = \pm \prod_{1 \le k \le n} (1 + \delta_k z)$ is a polynomial with real roots, then the function $\Lambda(t)$ defined by (8.24) is totally positive. Finally, if $\Psi(z)$ is an entire function of the second type, then there exists a sequence of polynomials $\Psi_n(z)$ with real roots such that $\Psi_n(z) \to \Psi(z)$ and, correspondingly, $\Lambda_n(t) \to \Lambda(t)$. Therefore, if $\Psi(z)$ is an entire function of the second type, then the corresponding function $\Lambda(t)$ that is defined by formula (8.24) is totally positive. Thus, part II of the theorem is proved.

The statement that *the difference kernel $K(t, \tau) = \Lambda(y - \tau)$ is variation diminishing if and only if the function $\Lambda(t)$ is of the form (8.24), where $\Psi(z)$ is an entire function of the second type*, was formulated explicitly in [Scho3]. In particular, a difference kernel K is variation diminishing if and only if either the kernel K or the kernel $-K$ is totally positive.[21] Many results related to totally positive and variation diminishing difference kernels can be found in [HiWi], Chapter IV, and especially in [Kar], Chapter 7.

Discrete totally positive difference kernels were first considered in [AESW], [ASW] and [Edr]. The formulations are analogous to the formulations for continuous difference kernels, but the proofs are more difficult and use tools from value distribution theory for meromorphic functions.

THEOREM (A. Aissen, A. Edrei, I.J. Schoenberg, A. Whitney, [AESW], [ASW], [Edr]).

I. *Let $\{s_k\}_{0 \le k < \infty}$ be a totally positive (unilateral) sequence with $s_0 = 1$. Then the series*

$$F(z) = \sum_{0 \le k < \infty} s_k z^k$$
(8.27)

---

[21] This agrees with the results of Schoenberg and Motzkin on general variation diminishing transforms since a difference kernels $K(t, \tau) = \Lambda(t - \tau)$ is sign regular if and only if either the kernel $K(t, \tau)$ or the kernel $-K(t, \tau)$ is totally positive.

converges in a neighborhood of the origin to a function of the form

$$F(z) = e^{\gamma z} \frac{\prod\limits_{k}(1 + \alpha_k z)}{\prod\limits_{k}(1 - \beta_k z)} \quad (\alpha_k \geq 0, \; \beta_k \geq 0, \; \gamma \geq 0, \; \sum(\alpha_k + \beta_k) < \infty). \qquad (8.28)$$

II. *Let $F(z)$ be a function of the form (8.28) and let (8.27) be its Taylor expansion in the vicinity of the origin. Then the sequence $\{s_k\}_{0 \leq k < \infty}$ is totally positive.*

This theorem provides a parametrization of the set of all totally positive unilateral sequences (under the normalizing condition $s_0 = 1$). The sequences α_k, β_k and the number γ serve as independent parameters. Various results related to unilateral and bilateral totally positive sequences can be found in [Kar], Chapter 8.

This parametrization of totally positive sequences plays an essential role in the theory of representations of the infinite symmetric group. It appears in the description of non-decomposable positive definite functions. This was discovered by Elmar Thoma in [Tho], where some earlier results on totally positive sequences were rediscovered. The generating function of the sequences that appear there are of the form (8.28), with $\gamma = 0$, and $\sum\limits_k \alpha_k + \sum\limits_k \beta_k \leq 1$. The theory of totally positive functions and sequences is used in approximation theory, mathematical statistics and in other fields. References can be found in [Kar] and in [GaMic]. Recently, a surprising connection between total positivity and canonical bases for quantum groups was discovered by G. Lusztig; see [FoZe].

The methods and especially the ideology of the paper [PS] underly some of the work of B. Ya. Levin that is considered in Chapter IX of his monograph [Lev]. The famous S.N. Bernstein inequality can be formulated in the following form: *Let $f(z)$ be an entire function of exponential type σ_f. If the inequality $|f(x)| \leq |e^{\sigma x}|$ holds for all real x, $-\infty < x < \infty$ and if $\sigma_f \leq \sigma$, then the derivative $f'(x)$ satisfies the inequality $|f'(x)| \leq |(e^{\sigma x})'| \;\; (-\infty < x < \infty)$.*
In other words, the operator $\frac{d}{dx}$ preserves inequalities on the real axis for some classes of entire functions. Levin has investigated the general form of linear operators which preserve inequalities of this sort. In this investigation, the linear operators that preserve the class of entire functions that is obtained as the closure of polynomials with zeros in the open right half plane play a crucial role. The operators of the form (8.1), which were introduced and investigated in the paper [PS], are precisely those that commute with the operator $z\frac{d}{dz}$.

REFERENCES

[AESW]   AISSEN, M., A. EDREI, I.J. SCHOENBERG and A. WHITNEY. *On the generating functions of totally positive sequences.* Proc. Nat. Àcad. Sci. U.S.A., **37** (1951), pp. 303 - 307.

[ASW]   AISSEN, M., A., I.J. SCHOENBERG and A. WHITNEY. *On the generating functions of totally positive sequences I.* J. d'Anal. Math., Jerusalem, **2** (1952), pp. 93 - 103.

[Dieu]   DIEUDONNE, J.A. *La Theorie analytique des Polynomes d'une variable : (a coefficients Quelconques).* Paris, Gauthier-Villars, 1938.

[Edr]   EDREI, A. *On the generating functions of totally positive sequences II.* J. d'Anal. Math., Jerusalem, **2** (1952), pp. 104 - 109.

[Fek]   FEKETE, M. *Über ein Problem von Laguerre* (German). [On one problem of Laguerre.] Rendiconti del Circolo Matematico di Palermo, **34** (1912), pp. 89 -100. Reprinted in: [PolCo], pp. 1 - 32.

[FoZe]   FOMIN, S. and A. ZELEVINSKI. *Total positivity: tests and parametrizations.* Math. Intelligencer, **22**:1 (2000), pp. 23 - 33.

[GaKr1]   GANTMACHER, F.R. and M.G. KREIN. *Sur les matrices oscillatoires* (French). [On oscillatory matrices]. Compt. Rend. Acad. Sci (Pqris). **201** (1935), pp. 577 - 579.

[GaKr2]   GANTMACHER, F.R and M.G. KREIN. *Sur les Matrices Complèment Non-négatives et Oscillatoires* (French). [On completely non-negative and oscillatory matrices] Compositio Math., **4** (1937), pp. 445 - 476.

[GaKr3]   GANTMACHER, F.R. and M.G. KREIN. *Ostsillyatsionnye Matritsy i Yadra i Malye Kolyebaniya Mekhanicheskikh Sistem* (Russian) [Oscillatory Matrices and Kernels and small Vibrations of Mechanical systems]. GITTL, Moskow 1950. English transl: German transl.:

[GaMic]   GASCA, M. and C.A. MICCHELLI (eds). *Total Positivity and its Applications.* (Mathematics and its Applications, **359**). Kluver Academic Publishers, Dordrecht, 1996. x+518 pp.

[HiWi]   HIRSCHMAN, I.I. and D.V. Widder. *The Convolution Transform.* Princeton Univ. Press, Princeton, NJ, 1955, x + 268 pp. Russian transl.: *Preobrazovaniya Tipa Svertki.* Izdat. Inostrannoĭ Literatury, Moscow 1958, 312 pp.

[Kar]   KARLIN, S. *Total Positivity, Vol. I.* Stanford Univ. Press, Stanford, CA, 1968. xi+576 pp.

[KrNa]   KREIN, M.G. and M.A. NAIMARK. *Metod Simmetricheskikh i Ermitovykh Form v Teorii Otdeleniya Korneĭ Algebraicheskikh Uravneniĭ* (Russian). Nauchno-Issledovatel'skiĭ Institut Matematiki i Mekhaniki pri Khar'kovskom Universitete. DNTVU, Khar'kov 1936, 43 p. English transl.: *The Method of Symmetric and Hermitian Forms in the Theory of the Separation of the Roots of Algebraic Equations*, Linear and Multilinear Algebra, **10** (1981), pp. 265 - 308.

[Kur]   KUROSH, A.G. *Kurs Vyssheĭ Algebry* (Russian). Nauka, Moskow 1965. English transl.: Higher algebra, Mir, Moscow 1972.

[Lag1]   LAGUERRE, E. *Sur quelques points de la théorie des équations numériques.* Acta Math, **4** (1864). Reprinted in [LagO], pp. 184 - 206.

[Lag2]   LAGUERRE, E. *Ser les fonctions du genre zéro et du genre un.* Compt. Rend. Acad. Sciences Paris, **98** (1882), pp. 828 - 831. Reprinted in [LagO], pp. 174 - 177.

[LagO]   LAGUERRE, E. *Euvres. Tome I* (French). Paris. Gauthier Villars, 1898, 1-sr ed.; Chelsea, 1972, 2-nd ed.

[Lev]   LEVIN, B.YA. *Raspredelenie korneĭ tselykh funktsiĭ* (in Russian). Gostekhizdat, Moskow 1956. English transl.: *Distribution of Zeros of Entire Functions.* (Ser.: Transl. of Math. Monogr., **5**, Amer. Math. Soc., Providence, RI, 1-st ed. 1964, viii + 493 pp., 2-nd ed. 1980, xii + 523 pp.

[Mar]   MARDEN, M. *Geometry of polynomials* (2-nd ed.). (Ser.: Mathematical surveys and monographs, **3**). Amer. Math. Sos., Providence, Rhode Island 1966, xiii, 243 p.

[Mal]   MALO, E. *Note sur équations algébriques dont toutes les racines sont réelles* (French). [Note on algebraic equations all roots of which are real]. Journal de Mathématiques spéciales, (ser. 4), t. **4** (1895), p. 7 - 10.

[Mot1]   MOTZKIN, TH. *Beitraege zur Theorie der linearen Ungleichungen* (German). [Contributions to the theory of linear inequalities]. (Doctoral dissertation. Basel, 1934). Azriel, Lerusalem , (1936). 73pp. English transl. in [Mot3], pp. 1 - 80.

[Mot2]   MOTZKIN, TH. *Sur les transformations qui n'augmentent pas le nombre des variations du signe* (French). [On transformations which don't increase the nomber of the variations of signs] C. R. Acad. Sci., Paris **202**, (1936), pp. 894-897.

[Mot3]   MOTZKIN, TH. *Selected papers.* (Contemporary Mathematicians). Birkhäuser, Boston, MA,1983. xxvi+530

[Obr]   OBRESCHKOFF, N. *Verteilung und Berechnung der Nullstellen reeller Polynome.* Deutscher Verlag der Wissenschaften, Berlin 1963, viii, 298 p.

[Pink]   PINKUS, A. *Spectral properties of totally positive kernels and matrices.* In [GaMic], pp. 457 - 511.

[Pol1]   PÓLYA, G. *Über Annäherung durch Polynome mit lauter reellen Wurzeln.* Rend. Circ. Math. Palermo, **36** (1913), pp. 279 - 295.

[PolCo]   PÓLYA, G. *Collected Papers, Vol. II.* MIT Press, Cambridge, MA, and London, England 1974, x + 444 pp.

[PS]   PÓLYA, G. and I. SCHUR: *Über zwei Arten von Faktorenfolgen in der Theorie der algebraischen Gleichungen.* Journ. für reine und angew. Math. **144** (1914), pp. 89 - 113. Reprinted in: [Sch: Ges], Vol. II, pp. 88 - 112. Reprinted also in [PolCo], pp. 100 - 124.

[PoSz]   PÒLYA, G. and G. SZEGÖ. *Aufgaben and Lehrsätze in der Analysis. Band* II (German). (Ser.: Grundlehren der mathematischen Wissenschaften, **20**). Springer-Verlag, Berlin, 1-st ed. 1925, 2-nd ed. 1954. English transl.: *Problems and Theorems in Analysis, Volume* II. (Ser.: Grundlehren der mathematischen Wissenschaften, **216**). Springer-Verlag, Berlin·Heidelberg·New York, 1976, 301

[Scho1]   SCHOENBERG, I.   *Über   variationsvermindernde   lineare   Transformationen.* Math. Zeitschrift, **32** (1930), pp. 321 - 328.

[Scho2]   SCHOENBERG, I. *On totally positive functions, Laplace Integral and entire functions of Laguerre-Polya-Schur type.* Proc. Nat. Acad. Sci. U.S.A., **33** (1947), pp.11 - 17.

[Scho3]   SCHOENBERG, I. *On variation-diminishing integral operators of the convolutions type.* Proc. Nat. Acad. Sci. U.S.A., **34** (1948). pp. 164 - 169.

[Scho4]   SCHOENBERG, I.J. *On Pólya frequency functions. I. The totally positive functions and their Laplace transforms.* J. d'Analyse Math., Jerusalem, **1** (1951), pp. 331 - 374.

[Sch: Ges] SCHUR, I.: *Gesammelte Abhandlungen [Collected Works]. Vol. I, II, III.* Springer-Verlag, Berlin·Heidelberg·New York 1973.

[Sch7]   SCHUR, I.: *Zwei Sätze über algebraische Gleichungen mit lauter reellen Wurzeln [Two theorems on algebraic equations with only real roots - in German].* Journ. für die reine und angew. Math., **144** (1914), pp. 75 - 88. Reprinted in: [Sch: Ges], Vol. II, pp. 56 - 69.

[Tho]   THOMA, E. *Die unzerlegbaren, positiv-definiten Klassenfunktionen der abzählbar unendlichen, symmetrischen Gruppe [The undecomposible positive-definite class functions of the countably infinite symmetric group - in German].* Math. Zeitschr., **85** (1964), pp. 40 - 61.

[Web]   WEBER, H. *Lehrbuch der Algebra.* Vieweg Verlag, Braunschweig 1912 (2-nd ed.), New York, Chelsea 1961 (3-d ed).

9. The Schur class of holomorphic functions, and the Schur algorithm.

The papers [Sch9]-[Sch10] are probably the best known contributions of Issai Schur to analysis. In these papers Schur introduced a new parametrization of functions that are holomorphic and bounded by one in the open unit disk $\mathbb{D}$ and an algorithm for calculating these parameters. These ideas and their subsequent generalizations to matrix and operator valued functions are widely used in a variety of applications that range from signal processing to the study of Pisot and Salem numbers.

DEFINITION 1. *A function $s(z)$ that is holomorphic in the open unit disk $\mathbb{D}$ and satisfies the inequality*

$$|s(z)| \leq 1 \qquad \text{for all points} \quad z \in \mathbb{D} \tag{9.1}$$

is said to belong to the **Schur class** $\mathcal{S}$. *A function $s \in \mathcal{S}$ will be referred to as a* **Schur function**. *A Schur function $s \in \mathcal{S}$ is said to be* **inner** *if the absolute value of the radial limit*

$$s(t) \overset{\text{def}}{=} \lim_{r \to 1-0} s(rt) \tag{9.2}$$

is equal to one a.e. with respect to Lebesgue measure. The set of inner functions will be denoted by the symbol $\mathcal{S}_{in}$. The set of rational inner functions will be denoted $\mathcal{S}_{rin}$.

The simplest inner functions are finite **Blaschke products**:

$$s(z) = cz^{\kappa} \prod_{k=1}^{d} \frac{\alpha_k - z}{1 - \overline{\alpha_k} z}, \tag{9.3}$$

where c is a constant of modulus one, κ is a non-negative integer and $\alpha_k \in \mathbb{D}$.

The **Schur algorithm** was introduced by I. Schur in Section 1 of [Sch9]. It exploits the fact that if $\gamma \in \mathbb{D}$, then the linear fractional transformation

$$\zeta \to \frac{\zeta - \gamma}{1 - \zeta \overline{\gamma}} \tag{9.4}$$

is a one to one mapping of the open unit disk $\mathbb{D}$ onto itself and a one to one mapping of the unit circle $\mathbb{T}$, i.e., the boundary of $\mathbb{D}$, onto itself. If $|\gamma| = 1$ the transformation (9.4) maps the set $\mathbb{C} \setminus \{\gamma\}$ into the point $\{-\gamma\}$ and is *not defined* at the point γ.

Let $f \in \mathcal{S}$ be a Schur function *that is not a constant of modulus one*. Then $|f(0)| < 1$ and hence, in view of the properties of (9.4), the transformation

$$f(z) \to \frac{f(z) - f(0)}{1 - f(z)\,\overline{f(0)}} \tag{9.5}$$

maps $\mathcal{S}$ into $\{s \in \mathcal{S} : s(0) = 0\}$. Therefore, by the Schwarz Lemma, the transformation

$$f(z) \to \frac{f(z) - f(0)}{1 - f(z)\,\overline{f(0)}} \cdot \frac{1}{z} \tag{9.6}$$

maps $\{s \in \mathcal{S} : |s(0)| \neq 1\}$ onto the class $\mathcal{S}$ of all Schur functions. In particular if $f \in \mathcal{S} \setminus \mathcal{S}_{rin}$, then $|f(0)| < 1$ and the transformation (9.6) is well defined. It is easy to see that:

PROPOSITION 1. *The transformation (9.6) maps $f \in \mathcal{S} \setminus \mathcal{S}_{rin}$ into itself.*

PROPOSITION 2. *The transformation (9.6 maps rational inner functions $f \in \mathcal{S}_{rin}$ of degree n, $n \geq 1$, into rational inner functions of degree $n - 1$.*

DESCRIPTION OF THE SCHUR ALGORITHM. The Schur algorithm defines a sequence of Schur functions $\{s_k(z)\}_{0 \leq k < \infty}$ starting from a given Schur function $s(z)$ that is assigned the index zero:

$$s_0(z) \overset{\text{def}}{=} s(z), \tag{9.7}$$

$$s_k(z) \overset{\text{def}}{=} \frac{s_{k-1}(z) - s_{k-1}(0)}{1 - s_{k-1}(z)\,\overline{s_{k-1}(0)}} \cdot \frac{1}{z} \quad (k = 1, 2, 3, \ldots). \tag{9.8}$$

SCHUR PARAMETERS. *Let $s \in \mathcal{S}$ and let $\{s_k\}$ be the sequence (finite or infinite) of functions generated by the Schur algorithm with $s_0(z) = s(z)$. The numbers*

$$\gamma_k \overset{\text{def}}{=} s_k(0) \tag{9.9}$$

are termed the Schur parameters of the function s.

If the starting function $s \notin \mathcal{S}_{rin}$, then, by Proposition 1, the algorithm continues indefinitely and produces infinitely many Schur functions $s_k(z)$, $k = 0, 1, 2, 3, \ldots$ and generates an infinite sequence of Schur parameters $\{\gamma_k\}_{0 \leq k < \infty}$. In this case

$$|\gamma_k| < 1, \quad k = 0, 1, 2, \ldots. \tag{9.10}$$

If the starting function $s \in \mathcal{S}_{rin}$ is is a rational inner function of degree n, then, by Proposition 2, the algorithm terminates after n steps. In this case, it generates a finite sequence of Schur parameters

$$|\gamma_k| < 1, \quad k = 0, 1, 2, \ldots, n-1 \quad \text{and} \quad |\gamma_n| = 1. \tag{9.11}$$

The Schur parameter $\gamma_k(s)$ of a Schur function

$$s(z) = \sum_{k=0}^{\infty} c_k(s) z^k \tag{9.12}$$

depends only on the Taylor coefficients $c_0(s), c_1(s), \ldots, c_k(s)$ of the function s:

$$\gamma_k(s) = \Phi_k(c_0(s), c_1(s), \ldots, c_k(s)), \tag{9.13}$$

where $\Phi_k(c_0, c_1, \ldots, c_k)$ is a rational function of the variables $c_0, \overline{c_0}, c_1, \overline{c_1}, \ldots, c_{k-1}, \overline{c_{k-1}}, c_k$.

Conversely, the Taylor coefficient $c_k(s)$ of a Schur function s depends only on the Schur parameters $\gamma_0(s), \gamma_1(s), \ldots, \gamma_k(s)$ of this function:

$$c_k(s) = \Psi_k(\gamma_0(s), \gamma_1(s), \ldots, \gamma_k(s)), \tag{9.14}$$

where $\Psi_k(\gamma_0, \gamma_1, \ldots, \gamma_k)$ is a polynomial in $\gamma_0, \overline{\gamma_0}, \gamma_1, \overline{\gamma_1}, \ldots, \gamma_{k-1}, \overline{\gamma_{k-1}}, \gamma_k$.

Explicit expressions for Φ_k and Ψ_k are given in [Sch9].

DEFINITION 4. *A sequence $\{\gamma_k\}_{0 \leq k < \infty}$ of complex numbers is said to be* **strictly contractive** *if $|\gamma_k| < 1$ for every k.*

Thus, the sequence of Schur parameters of a Schur function $s \in \mathcal{S} \backslash \mathcal{S}_{rin}$ is strictly contractive. Moreover, every preassigned strictly contractive sequence $\{\gamma_0, \gamma_1, \gamma_2, \ldots, \gamma_k, \ldots\}$ is the sequence of Schur parameters for some unique Schur function $s \in \mathcal{S} \setminus \mathcal{S}_{rin}$. Such a function can be constructed by means of a continued fraction algorithm.

SCHUR CONTINUED FRACTIONS. Given an arbitrary strictly contractive sequence $\{\gamma_0, \gamma_1, \ldots, \gamma_k, \ldots\}$ of complex numbers, one can construct a sequence of rational Schur functions which converge to a Schur function s with Schur parameters $\{\gamma_0(s), \gamma_1(s), \ldots, \gamma_k(s), \ldots\}$ that coincide with the preassigned sequence. The construction is based on the inverse of the transformation

$$f(z) \to \frac{f(z) - \gamma}{1 - f(z)\,\overline{\gamma}} \cdot \frac{1}{z}, \tag{9.15}$$

i.e., on the transformation

$$f(z) \to \frac{\gamma + zf(z)}{1 + \overline{\gamma}zf(z)}, \tag{9.16}$$

which also maps $\mathcal{S}$ into $\mathcal{S}$. We use the 'inverse Schur algorithm' recursively to construct *the n-th Schur approximant*, which (following Schur) we will *denote* by $[z; \gamma_0, \gamma_1, \ldots, \gamma_n]$. Namely, we write

$$
\begin{aligned}
&[z; \gamma_n] = \gamma_n; \\
&[z; \gamma_k, \gamma_{k+1}, \gamma_{k+2}, \ldots, \gamma_n] = \frac{\gamma_k + z \cdot [z; \gamma_{k+1}, \gamma_{k+2}, \ldots, \gamma_n]}{1 + \overline{\gamma_k} \cdot z \cdot [z; \gamma_{k+1}, \gamma_{k+2}, \ldots, \gamma_n]}, \\
&k = n-1, n-2, \ldots, 1, 0.
\end{aligned}
\tag{9.17}
$$

The function $[z; \gamma_0, \gamma_1, \ldots, \gamma_n]$ is a rational Schur function whose Schur parameters $\gamma_k\big([z; \gamma_0, \gamma_1, \ldots, \gamma_n]\big)$ are equal to

$$
\begin{aligned}
\gamma_k\big([z; \gamma_0, \gamma_1, \ldots, \gamma_n]\big) &= \gamma_k \qquad \text{for} \qquad k = 0, 1, \ldots, n; \\
\gamma_k\big([z; \gamma_0, \gamma_1, \ldots, \gamma_n]\big) &= 0 \qquad \text{for} \qquad k > n.
\end{aligned}
$$

Let n_1 and n_2 be two nonnegative integers. Since the Schur parameters with index $k : 0 \leq k \leq \min(n_1, n_2)$ for the functions $[z; \gamma_0, \gamma_1, \ldots, \gamma_{n_1}]$ and $[z; \gamma_0, \gamma_1, \ldots, \gamma_{n_2}]$ coincide, the Taylor coefficients c_k^1 and c_k^2 $(0 \leq k \leq \min(n_1, n_2))$ for these two functions coincide as well. Hence,

$$[z; \gamma_0, \gamma_1, \ldots, \gamma_{n_1}] - [z; \gamma_0, \gamma_1, \ldots, \gamma_{n_2}] = \sum_{\min(n_1,n_2)<k<\infty} (c_k^1 - c_k^2)\, z^k.$$

83

Using the estimates $\left|[z; \gamma_0, \gamma_1, \ldots, \gamma_{n_1}]\right| \le 1$, $\left|[z; \gamma_0, \gamma_1, \ldots, \gamma_{n_2}]\right| \le 1$ for $z \in \mathbb{D}$ and the Schwarz Lemma, we obtain the inequality

$$\left|[z; \gamma_0, \gamma_1, \ldots, \gamma_{n_1}] - [z; \gamma_0, \gamma_1, \ldots, \gamma_{n_2}]\right| \le 2\,|z|^{1+\min(n_1, n_2)} \quad \text{for } z \in \mathbb{D}. \qquad (9.18)$$

From (9.18) it follows that the limit

$$[z; \gamma_0, \gamma_1, \ldots, \gamma_k, \ldots] \overset{\text{def}}{=} \lim_{n \to \infty} [z; \gamma_0, \gamma_1, \ldots, \gamma_n] \qquad (9.19)$$

exists in $\mathbb{D}$. *The function* $[z; \gamma_0, \gamma_1, \ldots, \gamma_k, \ldots]$ *is said to be* **the Schur continued fraction constructed from the sequence** $\{\gamma_0, \gamma_1, \ldots, \gamma_k, \ldots\}$.

The function $[z; \gamma_0, \gamma_1, \ldots, \gamma_k, \ldots] \in \mathcal{S} \backslash \mathcal{S}_{rin}$. Its Schur parameters $\gamma_k\big([z; \gamma_0, \gamma_1, \ldots, \gamma_k, \ldots]\big)$ coincide with the numbers γ_k:

$$\gamma_k\big([z; \gamma_0, \gamma_1, \ldots, \gamma_k, \ldots]\big) = \gamma_k, \quad 0 \le k < \infty.$$

Given $s \in \mathcal{S} \setminus \mathcal{S}_{rin}$, we can form the sequence $\{\gamma_0(s), \gamma_1(s), \ldots, \gamma_k(s), \ldots\}$ of its Schur parameters and then construct the Schur continued fraction $[z; \gamma_0(s), \gamma_1(s), \ldots, \gamma_k(s), \ldots]$. The function represented by this fraction is a Schur function whose Schur parameters coincide with the sequence of Schur parameters of the original function s. Hence, the Taylor coefficients of these two functions coincide as well. Thus, we are led to following result:

THEOREM (I. Schur, [Sch9]).

I. *Every $s \in \mathcal{S} \setminus \mathcal{S}_{rin}$ admits* **the continued fraction expansion**

$$s(z) = [z; \gamma_0(s), \gamma_1(s), \ldots, \gamma_k(s), \ldots]. \qquad (9.20)$$

II. *A Schur function $s \in \mathcal{S}_{rin}$ of degree n admits the representation*

$$s(z) = [z; \gamma_0(s), \gamma_1(s), \ldots, \gamma_n(s)]. \qquad (9.21)$$

DEFINITION 5. *Let $s \in \mathcal{S} \setminus \mathcal{S}_{rin}$, let n be a non-negative integer and let (9.20) be the Schur continued fraction expansion of the function s. Then the function*

$$\mathrm{p}_n(s;\,z) \overset{\text{def}}{=} [z; \gamma_0(s), \gamma_1(s), \ldots, \gamma_n(s)] \qquad (9.22)$$

is said to be **the n-th Schur approximant of the function s.**

REMARK 1. *The n-th Schur approximant is a rational function of z whose numerator and denominator are polynomials of degree not greater than n. In fact the n-th Schur approximant of a $s \in \mathcal{S} \setminus \mathcal{S}_{rin}$ is the n-th convergent of its Schur continued fraction expansion*

(9.20).

The estimate

$$\left| s(z) - \mathrm{p}_n(s;\, z) \right| \leq 2\, |z|^{n+1},\tag{9.23}$$

which follows from the Schwarz Lemma, holds for every $s \in \mathcal{S} \setminus \mathcal{S}_{rin}$ and implies that *the sequence of the Schur approximants of such an s converges to it locally uniformly in the open unit disk $\mathbb{D}$.* This result on the locally uniform convergence of the approximants $\mathrm{p}_n(s;\, z)$ to $s(z)$ in the open unit disc $\mathbb{D}$ appears in [Sch9] (with the rougher estimate $\left| s(z) - \mathrm{p}_n(s;\, z) \right| \leq 2|z|^{n+1}(1-|z|)^{-1}$). The problem of convergence of Schur approximants to s on the unit circle $\mathbb{T}$ is much more difficult. This problem was studied in [Nja] and [Khru2].

The preceding results imply that *the correspondence*

$$\{\gamma_0,\, \gamma_1,\, \dots,\, \gamma_k,\, \dots\} \longleftrightarrow [z;\, \gamma_0,\, \gamma_1,\, \dots,\, \gamma_k,\, \dots]\tag{9.24}$$

is a free parametrization of the class of all $s \in \mathcal{S} \setminus \mathcal{S}_{rin}$ by means of the set of all strictly contractive sequences $\{\gamma_0,\, \gamma_1,\, \gamma_2,\, \dots,\, \gamma_k,\, \dots\}$. where sequences serve as free parameters of this class, This is important because the geometry of the set of all Taylor coefficients of functions of this class is rather complicated, whereas the geometry of the set of all Schur parameters is very simple: it is just the direct product of the open unit disks. This geometry is compatible with probabilistic structures and is well suited for probabilistic study. Some results on Schur functions with random Schur parameters are obtained in [Kats].

REMARK 2.   *It is not easy to express the properties of a concrete Schur function s in terms of its Schur parameters[22] $\gamma_k(s)$. In particular, it is not easy to recognize whether the function s is inner or not. Not much is known about this.*

If $\sum\limits_{0 \leq k < \infty} |\gamma_k(s)| < \infty$, *then the function s is continuous in the closed unit disk $\overline{\mathbb{D}}$, and* $\max\limits_{z \in \mathbb{D}} |s(z)| < 1$. *(Of course, s is not inner.) This result was obtained by I. Schur, [Sch10], §15, Theorem XVIII.*

If $\sum\limits_{0 \leq k < \infty} |\gamma_k(s)|^2 < \infty$, *then again the function s is not inner, as follows from the identity*

$$\prod_{0 \leq k < \infty} (1 - |\gamma_k(s)|^2) = \exp\left\{ \int_{\mathbb{T}} \ln\left(1 - |\, s(t)|^{\,2}\right) m(dt) \right\}.$$

(See [Boy] and also formula (8.14) in [Ger3], which expresses a similar result for polynomials that are orthogonal on $\mathbb{T}$.)

---

[22] However, to express the properties of a Schur function in terms of its Taylor coefficients is, as a rule, even more difficult.

If $\overline{\lim}_{k\to\infty}|\gamma_k(s)| = 1$, then the function s is inner. An equivalent result was obtained by E.A. Rakhmanov in the setting of orthogonal polynomials on the unit circle, [Rakh]. It is known as Rakhmanov's Lemma. A simple function-theoretic proof of the Rakhmanov lemma in the setting of Schur functions can be found in [Kats].

If the sequence of Schur parameters $\{\gamma_k(s)\}_{0\le k<\infty}$ satisfies the Maté–Nevai condition $\lim_{k\to\infty} \gamma_k\gamma_{k+n} = 0$ for $n = 1, 2, 3, \ldots$, but $\overline{\lim}_{k\to\infty}|\gamma_k| > 0$, then s is an inner function. This is Theorem 5 and Corollary 9.1 in [Khru2].

It is also known that there exists infinite Blaschke product s such that $\sum_{0\le k<\infty}|\gamma_k(s)|^p < \infty$ for every $p > 2$. (This is shown in [Khru3].)

The following question is both natural and important:

QUESTION 1: *Given a sequence* $\mathbf{c} = \{c_0, c_1, \ldots\}$, *does there exist a function* $s \in \mathcal{S}$ *such that*

$$\frac{s^{(j)}(0)}{j!} = c_j \quad \text{for} \quad j = 0, 1, \ldots ?$$

Schur obtained an answer to this question by using the algorithm (9.7)-(9.8) starting with

$$s_0(z) = f(z) = \sum_{j=0}^{\infty} c_j z^j , \tag{9.25}$$

to calculate the parameters γ_j. All the series are formal. However, since the Schur parameters $\gamma_0, \ldots, \gamma_k$ only depend upon $c_0, \ldots, c_k$, this does not present a problem. Proceeding this way, Schur obtained the following answer to Question 1.

In order for the series (9.25) to be the Taylor series of a Schur function, it is necessary and sufficient that either $|\gamma_k(f)| < 1$ for every integer $k :\ 0 \le k < \infty$, or $|\gamma_k(f)| < 1$ for $k :\ 0 \le k < n$, and $|\gamma_n(f)| = 1$. In the second case the coefficients c_k of the series (9.25) coincide with the k-th Taylor coefficients of the function $\big[z; \gamma_0(f), \gamma_1(f), \ldots \gamma_n(f)\big]$ for every[23] $k :\ 0 \le k < \infty$.

Schur also considered the following related question:

QUESTION 2. *Given a finite set of complex numbers $\{c_0, \ldots, c_m\}$, does there exist a function $s \in \mathcal{S}$ such that*

$$\frac{s^{(j)}(0)}{j!} = c_j \quad \text{for} \quad j = 0, 1, \ldots, m? \tag{9.26}$$

Moreover, if such functions exist, how can one describe them?

---

[23]For $k = 0, 1, \ldots, n$ this coincidence holds automatically since the Schur parameters $\gamma_0(f), \gamma_1(f), \ldots, \gamma_n(f)$ are built from $c_0, c_1, \ldots, c_n$; the remaining coefficients c_k for $k > n$ are determined by $c_0, c_1, \ldots, c_n$.

Schur answered Question 2 in terms of the Schur parameters generated by the algorithm $(9.7)\text{-}(9.8)$ starting with

$$s_0(z) = g(z) = \sum_{j=0}^{m} c_j z^j \,. \tag{9.27}$$

THEOREM 1 (I. Schur, [Sch9]). *There exists a function $s \in \mathcal{S}$ that meets the interpolation condition (9.26) if and only if: either $|\gamma_k(g)| < 1$ for every integer $k: 0 \le k \le m$, or $|\gamma_k(g)| < 1$ for $k: 0 \le k < n$, and $|\gamma_n(g)| = 1$ for some n, $n \le m$. In the second case, the interpolating Schur function $s(z)$ is unique, namely, $s(z) = \big[z; \gamma_0(g)), \gamma_1(g), \ldots \gamma_n(g)\big]$. In the first case, there are infinitely many interpolating functions $s(z)$. Moreover, the first $m+1$ Schur parameters $\gamma_k(s)$, $k = 0, 1, \ldots, m$, of every interpolant s coincide with the Schur parameters $\gamma_k(g)$ of the given polynomial. The remaining parameters $\gamma_k(s)$ with $k > m$ are either an infinite sequence of arbitrary strictly contractive complex numbers if $s \in \mathcal{S} \setminus \mathcal{S}_{rin}$, or a finite sequence of strictly contractive numbers that terminates with $|\gamma_n(s)| = 1$ for some $m < n$. Moreover, all interpolants are obtained this way.*

Schur also formulated another criterion for the solvability of the interpolation problem (9.26) in terms of an $(m+1) \times (m+1)$ upper triangular Toeplitz matrix based on the the coefficients of the given polynomial:

$$\mathsf{C}_m = \begin{bmatrix} c_0 & c_1 & c_2 & \cdots & c_m \\ 0 & c_0 & c_1 & \cdots & c_{m-1} \\ 0 & 0 & c_0 & \cdots & c_{m-2} \\ . & . & . & \cdots & . \\ 0 & 0 & 0 & \cdots & c_0 \end{bmatrix} \tag{9.28}$$

THEOREM 2 (I. Schur, [Sch9]). *The given polynomial $g(z) = c_0 + c_1 z + \cdots + c_m z^m$ can be interpolated by a function $s \in \mathcal{S}$ if and only the Hermitian form based on the matrix $\mathsf{I} - \mathsf{C}_m^* \mathsf{C}_m$ is non-negative:*

$$\langle (\mathsf{I} - \mathsf{C}_m^* \mathsf{C}_m)\, x,\, x \rangle \ge 0, \quad \forall x \in \mathbb{C}^{m+1}\,, \tag{9.29}$$

where I is the identity matrix in $\mathbb{C}^{m+1}$ and $\langle\ ,\ \rangle$ is the standard scalar product in $\mathbb{C}^{m+1}$. The Hermitian form is strictly positive if and only if there exist more than one interpolating function $s \in \mathcal{S}$.

We remark that the matrix C_m and an analogous matrix D_m based on the coefficients of the reflected polynomial $z^m \overline{g(1/\overline{z})} = \overline{c_m} + \overline{c_{m-1}} z + \cdots + \overline{c_0} z^m$ figure in the well known Schur-Cohn test:

The roots of the polynomial $g(z)$ lie in $\mathbb{D}$ if and only if $\mathsf{D}_m^* \mathsf{D}_m - \mathsf{C}_m^* \mathsf{C}_m > 0$.

A nice proof of this result based on Schur parameters may be found in the first chapter of [FoFr].

Schur derived the criterion (9.29) for the solvability of the interpolation problem (9.26) from the criterion for solvability in terms of Schur parameters (that was formulated as Theorem 1). As a by product of this derivation, he obtained a formula for the factorization of a 2×2 square block matrix

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

with square block diagonal entries A and C (not necessarily of the same size) when the matrix A is invertible, which in turn leads easily to the identity

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix} \cdot \begin{bmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{bmatrix} \cdot \begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix}. \qquad (9.30)$$

The matrix $D - CA^{-1}B$ is termed *the Schur complement of the block entry A with respect to M*. If it is also invertible, then M is invertible and the last formula leads easily to a formula for the inverse matrix M^{-1}, since each of the three factors on the right hand side of (9.30) are easily inverted. The same formula implies that

$$\det M = \det A \cdot \det(D - CA^{-1}B),$$

as was also noted in [Sch9]. Schur complements are widely used in the applied linear algebra and operator theory; see e.g., the long survey (of more then hundred pages) [Oue] and [Sm], respectively.

The k-th step (9.8) of the Schur algorithm can also be presented in the form

$$\begin{bmatrix} s_k(z) \\ 1 \end{bmatrix} = \begin{bmatrix} z^{-1} & -\gamma_{k-1}\, z^{-1} \\ -\overline{\gamma_{k-1}} & 1 \end{bmatrix} \cdot \begin{bmatrix} s_{k-1}(z) \\ 1 \end{bmatrix} \cdot \frac{1}{-\overline{\gamma_{k-1}}\, s_{k-1}(z) + 1},$$

where $\gamma_{k-1} = s_{k-1}(0)$. Thus, it is natural to associate the matrix $\begin{bmatrix} z^{-1} & -\gamma_{k-1}\, z^{-1} \\ -\overline{\gamma_{k-1}} & 1 \end{bmatrix}$ with the k-th step of Schur algorithm. However, it turns out to be more fruitful to deal with the matrix $m_{\gamma_{k-1}}$, where

$$m_\gamma(z) = \begin{bmatrix} z^{-1} & -\gamma \cdot z^{-1} \\ -\overline{\gamma} & 1 \end{bmatrix} \cdot \frac{1}{\sqrt{1 - |\gamma|^2}}, \qquad \text{when} \quad |\gamma| < 1. \qquad (9.31)$$

The matrix m_γ is also a matrix of coefficients of the linear fractional transformation (9.15), which is the basic step (9.6) of the Schur algorithm, whereas the matrix

$$m_\gamma(z)^{-1} = \begin{bmatrix} z & \gamma \\ z \cdot \overline{\gamma} & 1 \end{bmatrix} \cdot \frac{1}{\sqrt{1 - |\gamma|^2}} \qquad (9.32)$$

is the coefficient matrix of the linear fractional transformation (9.16) corresponding to the basic step of the inverse Schur algorithm.

The coefficient matrix of a linear fractional transformation is determined up to a nonzero scalar factor. The matrix of the linear fractional transformation (9.15) is chosen to be of the form (9.31) because then m_γ is j-inner with respect to the signature matrix

$$j = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{9.33}$$

i.e.,

$$(m_\gamma(z)^*)^{-1} j (m_\gamma(z))^{-1} - j = (1 - |z|^2) \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix}. \tag{9.34}$$

DEFINITION. *Let $\omega_m = \{\gamma_k\}_{0 \le k \le m}$ be a strictly contractive sequence of complex numbers and let the entries of the matrix valued function*

$$M_{\omega_m}(z) \stackrel{def}{=} m_{\gamma_m}(z) \cdot m_{\gamma_{m-1}}(z) \cdot \ldots \cdot m_{\gamma_1}(z) \cdot m_{\gamma_0}(z), \quad m = 0, 1, 2, \ldots \tag{9.35}$$

be denoted as

$$M_{\omega_m}(z) = \begin{bmatrix} a_{\omega_m}(z) & b_{\omega_m}(z) \\ c_{\omega_m}(z) & d_{\omega_m}(z) \end{bmatrix}. \tag{9.36}$$

For $s(z) \in \mathcal{S} \setminus \mathcal{S}_{rin}$, let $\omega = \{\gamma_k\}_{0 \le k < \infty}$ be the sequence of its Schur parameters and let $\{s_k(z)\}_{0 \le k < \infty}$ be the sequence of Schur functions generated by the Schur algorithm (so that $\gamma_k = s_k(0)$). Then

$$M_{\omega_m}(z) \begin{bmatrix} s(z) \\ 1 \end{bmatrix} = \begin{bmatrix} s_{m+1}(z) \\ 1 \end{bmatrix} \cdot (c_{\omega_m}(z) s(z) + d_{\omega_m}(z)), \tag{9.37}$$

and hence

$$\frac{a_{\omega_m}(z) s(z) + b_{\omega_m}(z)}{c_{\omega_m}(z) s(z) + d_{\omega_m}(z)} = s_{m+1}(z). \tag{9.38}$$

Moreover,

$$s(z) = \frac{w_{11}(z) s_{m+1}(z) + w_{12}(z)}{w_{21}(z) s_{m+1}(z) + w_{22}(z)}, \tag{9.39}$$

where the matrix $W(z) = \begin{bmatrix} w_{11}(z) & w_{12}(z) \\ w_{21}(z) & w_{22}(z) \end{bmatrix} = M_{\omega_m}^{-1}(z)$ can be expressed in the term of the entries of the matrix $M_{\omega_m}(z)$:

$$W(z) = \begin{bmatrix} d_{\omega_m}(z) & -b_{\omega_m}(z) \\ -c_{\omega_m}(z) & a_{\omega_m}(z) \end{bmatrix} \cdot \frac{1}{a_{\omega_m}(z) d_{\omega_m}(z) - b_{\omega_m}(z) c_{\omega_m}(z)}. \tag{9.40}$$

Furthermore, since

$$W(z) = m_{\gamma_0}^{-1}(z) \cdot m_{\gamma_1}^{-1}(z) \cdot \ \ldots \ \cdot m_{\gamma_{m-1}}^{-1}(z) \cdot m_{\gamma_m}^{-1}(z) , \qquad (9.41)$$

and the matrix m_γ^{-1} is linear with respect to z, the entries of the matrix $W(z)$ are polynomials with respect to z of degree m (or less). In view of (9.34), the matrix function $W(z)$ satisfies the condition

$$W^*(z)jW(z) - j \geq 0 \quad \text{for} \quad z \in \mathbb{D} , \qquad (9.42)$$

$$W^*(t)jW(t) - j = 0 \quad \text{for} \quad t \in \mathbb{T} . \qquad (9.43)$$

Formula (9.39) (in other notation) appears in §14 of [Sch10]. It expresses the Schur function $s(z)$ with Schur parameters $\gamma_k(s)$ that satisfy the condition $|\gamma_k(s)| < 1$, $k = 0, 1, \ldots, n$, as a linear fractional transformation of the function $s_{n+1}(z)$. It is important to note that *the matrix $W(z)$ of this fractional-linear transformation can be constructed from only the first $n + 1$ Schur parameters $\gamma_k(s)$, $k = 0, 1, \ldots, n$*.

The Schur parameters of the functions $s(z)$ and $s_{m+1}(z)$ are related: $\gamma_k(s_{m+1}) = \gamma_{m+k}(s)$, $k = 0, 1, 2, \ldots$. Thus, the following result holds:

THEOREM 3. *The set of all functions $s(z) \in \mathcal{S}$, whose Schur parameters $\gamma_k(s)$ coincide with a prescribed set of numbers γ_k, $|\gamma_k| < 1$, $k = 0, 1, \cdots, m$, can be parametrized by means of the linear fractional transformation*

$$s(z) = \frac{w_{11}(z)\,\omega(z) + w_{12}(z)}{w_{21}(z)\,\omega(z) + w_{22}(z)} , \qquad (9.44)$$

where the coefficient matrix $W(z)$ of this transformation can be constructed from only these γ_k, and the free parameter $\omega(z)$ is an arbitrary function from the class $\mathcal{S}$.

Schur did not mention either formula (9.39) or formula (9.44) explicitly in his description of the solutions of the interpolation problem of the form

$$\gamma_k(s) = \gamma_k, \quad k = 0, 1, \ldots, m . \qquad (9.45)$$

Nor do the properties (9.42) and (9.43) of the matrix $W(z)$ in (9.44) appear in Schur's work. But this was the starting point of subsequent research on interpolation problems with constraints in various classes of analytic functions, particularly that of M. Riesz and R. Nevanlinna. The work [Sch9]-[Sch10] stimulated interest in obtaining matrices of linear fractional transformations that appear in descriptions of the sets of solutions of such problems. The methods described above are recursive and depend essentially upon formulas involving the Schur parameters. V.P. Potapov showed how to obtain an expression for the matrix $W(z)$ that appears in the description of the set of all solutions of the problem (9.45) in the class $\mathcal{S}$ directly in terms of the data $c_0, c_1, \ldots, c_m$, without first calculating the Schur parameters. This method of V.P. Potapov, as applied to the the interpolation problem (9.26), is elaborated on in great detail in the monograph [DFK].

Considerations related to formula (9.39) were used by Schur to obtain the following result:

In order that the function $s(z) = \dfrac{p(z)}{q(z)} \in \mathbb{S}$, where $p(z)$ and $q(z)$ are coprime polynomials, be representable in the form $s(z) = [z; \gamma_0, \gamma_1, \ldots, \gamma_m]$, with $m < \infty$, it is necessary and sufficient that the following two conditions are satisfied:

1. *The polynomial $q(z)$ does not vanish in the closed unit disc $\mathbb{D}$;*

2. *The factorization identity $|q(t)|^2 - |p(t)|^2 = r$, where r is a positive number, holds for $t \in \mathbb{T}$.*

The circle of problems related to Schur functions and the Schur algorithm is closely related to the theory of polynomials that are orthogonal on the unit circle. Note that

$$\zeta \to \frac{1 + \zeta}{1 - \zeta}$$

is one-to-one mapping of the unit disk $\{\zeta : |\zeta| < 1\}$ onto the right half-plane $\{\zeta : \operatorname{Re}\zeta > 0\}$. Therefore, if $s(z)$ is a Schur function, then the function

$$w(z) = \frac{1 + zs(z)}{1 - zs(z)} \tag{9.46}$$

is a Carathéodory function, i.e., a function which is holomorphic and has non-negative real part in the unit disk:

$$\operatorname{Re} w(z) \geq 0 \quad \text{for} \ \ z \in \mathbb{D}. \tag{9.47}$$

The factor z in (9.46) leads to the normalization

$$w(0) = 1. \tag{9.48}$$

Conversely, if $w(z)$ is a Carathéodory function that satisfies the normalization condition (9.48), then it can be uniquely represented in the form (9.46), where $s(z)$ is a Schur function. Every Carathéodory function $w(z)$ which satisfies the normalization condition (9.48) admits the Herglotz representation

$$w(z) = \int_{\mathbb{T}} \frac{t + z}{t - z} \, \sigma(dt), \tag{9.49}$$

where σ is a probability measure on $\mathbb{T}$. Conversely, if σ is a probability measure on $\mathbb{T}$, then formula (9.49) defines a normalized Carathéodory functiom $w(z)$. Thus, the transformation (9.46) together with the representation (9.49), establishes a one-to-one correspondence between Schur functions and probability measures on $\mathbb{T}$. It is easy to see that

$$S \in \mathcal{S}_{rin} \iff \sigma \text{ has finite support} \iff w(z) \text{ is rational with all its poles on } \mathbb{T}.$$

Let σ be a probability measure on $\mathbb{T}$ with infinitely many points of support and let $\{\varphi_k\}_{0 \le k < \infty}$ be a sequence of polynomials that is orthonormal, with respect to σ. Such a sequence can be obtained by applying the Gram-Schmidt orthogonalization procedure to the sequence $\{z^k\}_{0 \le k < \infty}$. Let $\varphi_k^*(z) \overset{\text{def}}{=} z^k \overline{\varphi_k(1/\overline{z})}$ denote the reciprocal polynomial. It turns out that the system of polynomials $\{\varphi_k, \varphi_k^*\}_{0 \le k < \infty}$ satisfy linear recurrence relations that can be written in the form:

$$\begin{bmatrix} \varphi_{k+1}(z) \\ \varphi_{k+1}^*(z) \end{bmatrix} = \frac{1}{\sqrt{1 - |a_k|^2}} \begin{bmatrix} z & -\overline{a}_k \\ -z\,a_k & 1 \end{bmatrix} \cdot \begin{bmatrix} \varphi_k(z) \\ \varphi_k^*(z) \end{bmatrix}, \quad 0 \le k < \infty, \qquad (9.50)$$

with the initial condition

$$\begin{bmatrix} \varphi_0(z) \\ \varphi_0^*(z) \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \qquad (9.51)$$

Here $\{a_k\}_{0 \le k < \infty}$ is a strictly contractive sequence of complex numbers that is determined uniquely by the probability measure σ that generates the sequence $\{\varphi_k\}_{0 \le k < \infty}$ of orthogonal polynomials. The numbers $a_k(\sigma)$ are termed the *reflection coefficients* of the measure σ or of the sequence $\{\varphi_k\}$ of orthogonal polynomials.

THEOREM [Ya.L. Geronimus]. *Let $s \in \mathcal{S} \setminus \mathcal{S}_{rin}$ and let the normalized Carathéodory function $w(z)$ be related to s by (9.46). Then the Schur parameters of s coincide with the reflection coefficients of the measure sigma:*

$$\gamma_k(s) = a_k(\sigma) \quad \text{for} \quad 0 \le k < \infty.$$

This theorem first appears in [Ger1]. It also appears as Theorem 18.2 in [Ger2]. Unfortunately, these papers are not easily accessible. However, an English translation of the second one is available. A simplified presentation of the cited theorem by Geronimus can be found in [Khru1] and in [PiNe].

Many (but not all) properties of a Schur function $s(z)$ can be naturally reformulated in terms of the related function $w(z)$, i.e., in terms of the related sequence of orthogonal polynomials. In particular: *a Schur function $s(z)$ is inner if and only if the related measure $\sigma(dt)$ is singular.* Indeed, $|s(t)| = 1$ if and only if $\operatorname{Re} w(t) = 0$. On the other hand, $\operatorname{Re} w(t) = \sigma'(t)$ for m almost every $t \in \mathbb{T}$. (Here $s(t)$ and $w(t)$ are the boundary values of the appropriate functions and $\sigma'(t)$ is the derivative of the measure σ with respect to the normalized Lebesgue measure m.) Some other connections between Schur functions and orthogonal polynomials can be found in [Gol1], [Gol2] and [Khru2].

There is a rich literature dedicated to the Schur algorithm and related topics. The volume [Ausg] contains a selection of basic early papers on Schur analysis written in German (by G. Herglotz, I. Schur, G. Pick, R. Nevanlinna, H. Weyl), as well as the Afterword written by B. Fritzsche and B. Kirstein, the editors of this volume. See also their survey [FrKi]. The last several years have witnessed an explosion of interest in generalizations of the Schur algorithm and the associated parametrization to matrix and operator

valued functions. In particular, a $p \times q$ matrix valued function $s(z)$ that is holomorphic in $\mathbb{D}$ and satisfies the inequality

$$|\xi^* s(z)\eta| \leq 1 \text{ for every point } z \in \mathbb{D} \text{ and every pair of unit vectors } \xi \in \mathbb{C}^p \text{ and } \eta \in \mathbb{C}^q$$

is said to belong to the Schur class $\mathcal{S}^{p \times q}$. The Schur algorithm is developed in detail for this class of functions in [DeDy]; see also [AlDy] for a reproducing kernel Hilbert space interpretation and additional generalizations to Pontryagin spaces and the references to both of these papers. There are also many papers by the team P. Delsarte, Y. Genin and Y. Kamp that are devoted to generalizations of a number of the the themes discussed in this section; see e.g., [DeGeK] for a start. The books [Alp], [Con], [DFK] and [FoFr] are dedicated to function theoretic questions related to the Schur algorithm and its applications to operator theory. Applications of the Schur algorithm to Pisot and Salem numbers are considered in the book [BDGPS]. The Schur algorithm is also useful in the setting of fast numerical algorithms for systems of linear equations with structured matrices [24] (Toeplitz, Hankel, etc.); see the book [S:Meth] (and in particular the papers [Kail] and [LevKai]). The terminology "I. Schur methods in signal processing" is now widely used in this connection. See the survey [KaSa1] and the volume [KaSa2] for further references in this direction.

REFERENCES

[Alp]   ALPAY, D. *Algorithme de Schur, espaces à noyau reproduisant et théorie des systèmes*. (Panorame and Synthèses, **6**). Socièté Mathèmatique de France, Paris, 1998. viii+189pp. (French). Englisch transl.: *The Schur Algorithm, Reproducing Kernel Spaces and System Theory*. (SMF/AMS Texts and Monographs, **5**). American Mathematical Society, Providence, RI; Socièté Mathèmatique de France, Paris, 2001. viii+150pp.

[AlDy]   ALPAY, D. and H. DYM, *On applications of reproducing kernel spaces to the Schur algorithm and rational J unitary factorization*. In [S:Meth], pp. 89 - 159.

[Ausg]   *Ausgewählte Arbeiten zu den Ursprüngen der Schur-Analysis* (German). [*Selected papers on the origin of the Schur analysis*]. (FRITZSCHE, B. and B. KIRSTEIN - editors) (German). (Series: Teubner-Archiv zur Mathematik - 016). B.G.Teubner Verlagsgesellschaft, Stuttgart·Leipzig 1991, 290pp.

[BDGPS]   BERTIN, M.J., A. DECOMPS-GUILLOUX, M. GRANDET-HUGOT, M.PATHIAUX-DELEFOSSE and J.P. SCHREIBER. *Pisot and Salem Numbers*. Birkhäuser-Verlag, Basel·Boston·Berlin, 1992, xiii+291pp.

[Boy]   BOYD, D.: *Schur's algorithm for bounded holomorphic functions*. Bull. London Math. Soc., **11** (1979), pp. 145 - 150.

---

[24] An algorithm for the fast inversion of Toeplitz matrices was designed be N. Levinson, [Levsn], about fifty five years ago, but it seems that the relationships with Schur algorithm did not play any role there.

[Con] CONSTANTINESCU, T. *Schur Parameters, Factorization and Dilation Problems*. (Operator Theory: Advances and Applications, **OT 82**). Birkhäuser Verlag, Basel, 1996. x+253 pp.

[DeGeK] PH. DELSARTE, Y. GENIN and Y. KAMP, *Schur parametrization of positive definite block-Toeplitz systems*, SIAM J. Appl. Math., **36** (1979), 34-46.

[DeDy] DEWILDE, P. and H. DYM, *Lossless chain scattering matrices and optimum linear prediction: The vector case*, Circuit Theory and Appl., **9** (1981), 135-175.

[DFK] DUBOVOJ, V.K., B. FRITZSCHE and B. KIRSTEIN. *Matricial Version of the Classical Schur Problem*. (Teubner-Texte zur Mathematik, Band **129**). B.G. Teubner Verlagsgesellschaft, Stuttgart·Leipzig, 1992.

[FoFr] C. FOIAS and A. FRAZHO, *The Commutant Lifting Approach to Interpolation Problems*, (Operator Theory: Advances and Applications, **OT 44**). Birkhäuser Verlag, Basel, 1990.

[FrKi] FRITZSCHE, B. and B. KIRSTEIN. *Schuranalysis - Umfassende Entfaltung einer mathematischen Methode* [*The Schur analysis - the comprehensive scope of a mathematical method* - in German.] In: Jahrbuch Überblicke Mathematik 1992. - CHATTERJI, S.D., B. FUCHSSTEINER, U. KULISCH, R. LIEDL und W. PURKERT - editors, Vieweg, Braunschweig, 1992, pp. 117 - 136.

[Ger1] GERONIMUS, YA.L.: *O polinomakh, ortogonal'nykh na kruge, o trigonometricheskoi probleme momentov i ob associirovannykh s neyu funktsiyakh tipa Carathéodory i Schur'a. (On polynomials orthogonal on the circle, on trigonometric moment-problem and on allied Carathéodory and Schur functions*. Russian, Résumé Engl.) Matem. Sbornik. Nov. ser. **15** (**42**), no.1 (1944), pp. 99-130.

[Ger2] GERONIMUS, YA.L.: *Polinomy, ortogonal'nye na kruge, i ikh prilozheniya*. Zapiski nauchno-issledovatel'skogo instituta matematiki i mekhaniki i Khar'kovskogo matem. obshchestva, **19** (1948), pp. 35-120 (Russian). English transl.: *Polynomials orthogonal on a circle and their applications*. Amer. Math. Soc. Transl. (ser. **1**), vol.**3** (1962), pp. 1-78.

[Ger3] GERONIMUS, YA.L.: *Polinomy Ortogonal'nye na Okruzhnosti i na Otrezke*. Fizmatgiz, Moskow, 1958 (In Russian). English transl.: *Polynomials Orthogonal on a Circle and Interval*. Pergamon Press, New York, 1960.

[Gol1] GOLINSKII, L.B: *Schur functions, Schur parameters and orthogonal polynomials on the unit circle*. Zeitschrift für Analysis und ihre Anwendungen **12** (1993), pp. 457 – 469.

[Gol2] GOLINSKII, L.: *On Schur functions and Szegö orthogonal polonomials*. In: *Topics in Interpolation Theory* (Operator Theory: Advances and Applications, **OT 95**). (DYM, H., B. FRITZSCHE, V. KATSNELSON, B. KIRSTEIN - editors.) Birkhäuser Verlag, Basel·Boston·Berlin 1997.

[Kail]   KAILATH, T. *A theorem of I. Schur and its impact on modern signal processing.* In [S:Meth], pp. 9 - 30.

[Kats]   KATSNELSON, V. *A generic Schur function is an inner one.* In: Interpolation Theory, System Theory and Related Topics. The Harry Dym Anniversary Volume. (Operator Theory: Advances and Applications). (ALPAY, D., I. GOHBERG and V. VINNIKOV-editors). Birkhäuser Verlag, Basel·Boston·Berlin, 2002, pp. 249 - 293.

[KaSa1]   KAILATH, T. and A.H. SAYED. *Displacement structure: theory and applications.* SIAM Review, **37**:3, (1005), pp. 297 - 386.

[KaSa2]   (KAILATH, T. and A.H. SAYED-aditors.) *Fast Reliable Algorithms for Matrices with Structure.* Society of Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1999, xvi+342pp.

[Khru1]   KHRUSHCHEV, S.: *Parameters of orthogonal polynomials.* In: "Methods of Apprximation Theory in Complex Analysis and Mathematical Physics. Selected papers from the international seminar held in Leningrad, May 13-26, 1991." GONCHAR, A.A and E. SAFF - editors. Lecture Notes in Math., vol. **1550**. Reissued by Springer-Verlag, Berlin 1993, originally published by "Nauka", Moskow 1993, pp. 185-191.

[Khru2]   KHRUSHCHEV, S.V.: *Schur's algorithm, orthogonal polynomials and convergence of Wall's continued fractions in $L^2(\mathbb{T})$.* Journal of Approximation Theory, vol. **108**, no.2, (2001), pp. 161-248.

[Khru3]   KHRUSHCHEV, S.V.: *A singular Riesz product in the Nevai class and inner functions with the Schur parameters $\bigcap_{p>2} l^p$.* Journal of Approximation Theory, vol. **108**, no.2, (2001), pp. 249-255.

[LevKai]   LEV-ARI, H. and T. KAILATH. *Triangular factorization of structured hermitian matrices.* In [S:Meth], pp. 301 - 323.

[Levsn]   LEVINSON, N. *The Wiener's RMS (root mean square) error criterion in filter desighn and prediction.* Journal of mathematics and physics, **25** (1946), 261 - 278.

[Nja]   NJÅSTAD, O. *Convergence of the Schur algorithm.* Proc. Amer. Math. Soc. **110**, No.4, (1990), pp. 1003-1007.

[Oue]   OUELETTE, D.V. *Schur complements and statistics.* Linear Algebra and its Applications, **36**, (1981), pp. 187 - 295.

[PiNe]   PINTÉR, F. and P. NEVAI. *Schur functions and orthogonal polynomials on the unit circle.* In: *Approximation Theory and Function Series. Budapest, 1995*, pp. 293 - 306. VÉRTESI, P., L. LEINDLER, F. MÓRICZ, SZ. RÉVÉSZ, J. SZABADOS and V. TOTIK - editors. János Bolyai Mathematical Society, Budapest 1996.

[Rakh]   RAHMANOV, E.A.(=RAKHMANOV, E.A.) *Ob asimptotike otnosheniya ortogonal'nykh polinomov. II.* Matem. Sbornik. Nov.ser. **118** (**160**), no. 1 (1982), pp.104-117 (In Russian). English transl.: *On the asymptotics of the ratio of orthogonal polynomials. II.* Math. USSR Sbornik **46** (1983), pp. 105-117.

[Sch: Ges]   SCHUR, I.: *Gesammelte Abhandlungen [Collected Works]. Vol. I, II, III.* Springer-Verlag, Berlin·Heidelberg·New York 1973.

[Sch9]   SCHUR, I.: *Über Potenzreihen, die im Innern des Einheitskreises beschränkt sind, I.* (in German) J. reine und angewandte Math. **147**(1917), 205 - 232. Reprinted in: [Sch: Ges], Vol. II, pp. 137 - 164. Reprinted also in: [Ausg], pp. 22-49. English translation: *On power series which are bounded in the interior of the unit circle. I.*, In: [S:Meth], pp. 31-59.

[Sch10]   SCHUR, I.: *Über Potenzreihen, die im Innern des Einheitskreises beschränkt sind, II.* (in German). J. reine und angewandte Math. **148** (1918), 122 - 145. Reprinted in: [Sch: Ges], Vol. II, pp. 165 - 188. Reprinted also in: [Ausg], pp. 50-73. English translation: *On power series which are bounded in the interior of the unit circle. II.* In: [S:Meth], pp. 61 - 88.

[Sm]   ŠMUL'YAN, YU.L. (=SHMUL'YAN, YU.L.) *Operatornyĭ integral Hellingera* (Russian). Matem. Sbornik **47** (**91**):4 (1959), pp. 381 - 430. English transl.: *A Hellinger operator integral*, Amer. Math.Soc. Transl. (**2**) **22** (1962), 289 - 337.

[TIT]   *Topics in Interpolation Theory.* (DYM, H., B. FRITZSCHE, V. KATSNELSON, B. KIRSTEIN - editors). (Operator Theory: Advances and Applications, **OT 95**), Birkhäuser Verlag, Basel·Boston·Berlin, 1997.

[S:Meth]   *I. Schur Methods in Operator Theory and Signal Processing* (Operator Theory: Advances and Applications, **OT 18**) (GOHBERG, I. - editor). Birkhäuser Verlag, Basel· Boston· Stuttgart, 1986.

10. Issai Schur's papers on analysis.

REFERENCES

[Sch1]   SCHUR, I. *Über vertauschbare lineare Differentialausdrücke* (German) [*On permutable differential expressions*]. Sitzungsberichte der Berliner Mathematischen Gesellschaft, **4** (1905), pp. 2 - 8. Reprinted in: [Sch: Ges], Vol. I, pp. 170 - 176.

[Sch2]   SCHUR, I.: *Über die charakteristischen Wurzeln einer linearen Substitution mit einer Anwendung auf die Theorie der Integralgleichungen* (German)[*On the characteristic roots of a linear substitution with an application to the theory of integral equations*]. Mathematische Annalen, **66** (1909), pp. 488 - 510. Reprinted in: [Sch: Ges], Vol. I, pp. 272 - 294.

[Sch3]   SCHUR, I. *Zur Theorie der linearen homogenen Integralgleichungen*(German) [*On the theory of linear homogeneous integral equations*]. Math. Annalen, **67** (1909), pp. 336 - 339. Reprinted in: [Sch: Ges], Vol. I, pp. 312 - 345.

[Sch4]   SCHUR, I.: *Bemerkungen zur Theorie der beschränkten Bilinearformen mit unendlich vielen Veränderlichen* (German) [*Remarks on the theory of bounded bilinear forms with infinitely many variables*]. Journ. für reine und angew. Math., **140** (1911), pp. 1 - 28. Reprinted in: [Sch: Ges], Vol. I, pp. 464 - 491.

[Sch5]   SCHUR, I.: *Über einen Satz von C. Carathéodory* (German) [*On a theorem of C. Carathéodory*]. Sitzungsberichte der Königl. Preuss. Akad. der Wiss., (1912), pp. 4 - 15. Reprinted in: [Sch: Ges], Vol. II, pp. 12 - 23.

[Sch6]   SCHUR, I.: *Über die Äquivalenz der Cesàroschen und Hölderschen Mittelwerte* (German) [*On the equivalence of Cesàro's and Hölder means*]. Mathematische Annalen, **74** (1913), pp. 447 - 458. Reprinted in: [Sch: Ges], Vol. II, pp. 44 - 55.

[Sch7]   SCHUR, I.: *Zwei Sätze über algebraische Gleichungen mit lauter reellen Wurzeln* (German) [*Two theorems on algebraic equations with only real roots*]. Journ. für die reine und angew. Math., **144** (1914), pp. 75 - 88. Reprinted in: [Sch: Ges], Vol. II, pp. 56 - 69.

[Sch8]   SCHUR, I. *Über die Entwicklung einer gegebenen Funktion nach den Eigenfunktionen eines positiv definiten Kerns* (German) [*On the expansion of a given function in eigenfunctions of a positive definite kernel*]. Schwarz-Festschrift (1914), pp. 392 - 409. Reprinted in: [Sch: Ges], Vol. II, pp. 70 - 87.

[Sch9]   SCHUR, I.: *Über Potenzreihen, die im Innern des Einheitskreises beschränkt sind, I.* (German) J. reine und angewandte Math. **147**(1917), 205 - 232. Reprinted in: [Sch: Ges], Vol. II, pp. 137 - 164. Reprinted also in: [Ausg], pp. 22-49. English translation: *On power series which are bounded in the interior of the unit circle. I.*, In: [S:Meth], pp. 31-59.

[Sch10]   SCHUR, I.: *Über Potenzreihen, die im Innern des Einheitskreises beschränkt sind, II.* (in German). J. reine und angewandte Math. **148** (1918), 122 - 145. Reprinted in: [Sch: Ges], Vol. II, pp. 165 - 188. Reprinted also in: [Ausg], pp. 50-73. English translation: *On power series which are bounded in the interior of the unit circle. II.* In: [S:Meth], pp. 61 - 88.

[Sch11]   SCHUR, I. *Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten* (German) [*On the distribution of the roots of some algebraic equations with integer coefficients*]. Math. Zeitschrift, **1** (1918), pp. 377 - 402. Reprinted in: [Sch: Ges], Vol. II, pp. 213 - 238.

[Sch12]   SCHUR, I.: *Über die Koeffizientensummen einer Potenzreihe mit positivem reellen Teil* (German) [*On the sum of the coefficients of a power series with positive real part*]. Archiv der Math. und Phys. (3) **27** (1918), pp. 126 - 135. Reprinted in: [Sch: Ges], Vol. II, pp. 249 - 258.

[Sch13]   SCHUR, I.: *Über das Maximum des absoluten Betrages eines Polynoms in einem gegebenen Intervall* (German) [*On the maximum of the absolute value of a polynomial on a given interval*]. Math. Zeitschrift, **4** (1919), pp. 271 -287. Reprinted in: [Sch: Ges], Vol. II, pp. 259 - 275.

[Sch14]   SCHUR, I. *Über einen von Herrn L. Lichtenstein benuzten Integralsatz* (German) [*On a theorem on integrals used by L. Lichtenstein*]. Math. Zeitschr. **7** (1920), pp. 232 - 234. Reprinted in: [Sch: Ges], Vol. II, pp. 286 - 288.

[Sch15]   SCHUR, I. *Über algebraische Gleichungen, die nur Wurzeln mit negativen Realteilen besitzen* (German) [*On algebraic equations which possess only roots with negative real parts*]. Zeitschr. für angew. Math. und Mech., **1** (1921), pp. 307 - 311. Reprinted in: [Sch: Ges], Vol. II, pp. 322 - 326.

[Sch16]   SCHUR, I.: *Über lineare Transformationen in der Theorie der unendlichen Reien* (German) [*On linear transformations in the theory of infinite series*]. Journ. für die reine und angew. Math., **151** (1921), pp. 79 - 121. Reprinted in: [Sch: Ges], Vol. II, pp. 289 - 321.

[Sch17]   SCHUR, I.: *Ein Beitrag zur Hilbertschen Theorie vollstetigen quadratischen Formen* (German) [*A contribution to the theory of completely continuous quadratic forms*]. Math. Zeitschr., **12** (1922), 287 - 297. Reprinted in: [Sch: Ges], Vol. II, pp. 402 - 412.

[Sch18]   SCHUR, I.: *Über eine Klasse von Mittelbildungen mit Anwendungen auf die Determinantentheorie* (German) [*On a class of averaging mappings with applications to the theory of determinants*]. Sitzungsberichte der Berliner Mathematischen Gesellschaft, **22** (1923), pp. 9 - 20. Reprinted in: [Sch: Ges], Vol. II, pp. 416 - 427.

[Sch19]   SCHUR, I. *Einige Bemerkungen zur Theorie der unendlichen Reihen* (German) [*A remark on the theory of infinite series*]. Sitzungsberichte der Berliner Mathematischen Gesellschaft, **29**, (1930), pp. 3 - 13. Reprinted in: [Sch: Ges], Vol. III, pp. 216 - 226.

[Sch20]   SCHUR, I. *Affektlose Gleichungen in der Theorie der Laguerreschen und Hermiteschen Polynome* (German) [*On affect free equations in the theory of Laguerre polynomials and Hermite polynomials*]. Journ. für die reine und angewandte Mathematik, **165** (1931), pp. 52 - 58. Reprinted in: [Sch: Ges], Vol. III, pp. 227 - 233.

[Sch21]   SCHUR, I. *On Faber Polynomials*. American Journ. of Math. **67** (1945), pp. 33 - 41. Reprinted in: [Sch: Ges], Vol. III, pp. 361 - 369.

[Sch22]   SCHUR, I. *Identities in the theory of power series*. American Journ. of Math. **69** (1947), pp. 14 - 26. Reprinted in: [Sch: Ges], Vol. III, pp. 379 - 391.

Joint papers:

[KnSch]   KNOPP, K. and I. SCHUR. *Über die Herleitung der Gleichung* $\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$ (German) [*On a derivation of the equation* $\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$]. Archiv der Math. und Phys. (3)**27** (1918), pp. 174 - 176. Reprinted in: [Sch: Ges], Vol. II, pp. 246 - 248.

[PS]   POLYA, G. and I. SCHUR: *Über zwei Arten von Faktorenfolgen in der Theorie der algebraischen Gleichungen* (German) [*On two types of multiplier sequences in the theory of algebraic equations*]. Journ. für reine und angew. Math. **144** (1914), pp. 89 - 113. Reprinted in: [Sch: Ges], Vol. II, pp. 88 - 112.

[SchSz] SCHUR, I. and G. SZEGÖ. *Über die Abschnitte einer im Einheitkreise beschränkten Potenzreihe* (German) [*On the truncation of a power series bounded in the unit disk*]. Sitzungsberichte der Preußischen Akademie der Wissenschaften 1925, Physikalisch-Mathematische Klasse, pp. 545 - 560. Reprinted in: [Sch: Ges], Vol. II, pp. 249 - 258. Reprinted in: [Sch: Ges], Vol. III, pp. 27 - 42.

Sources in which Schur papers were reprinted or translated:

[Sch: Ges] SCHUR, I.: *Gesammelte Abhandlungen* [*Collected Works*]. *Vol. I, II, III.* Springer-Verlag, Berlin· Heidelberg·New York, 1973.

[Ausg] *Ausgewählte Arbeiten zu den Ursprüngen der Schur-Analysis* (German) [*Selected papers on the origin of Schur analysis*]. (FRITZSCHE, B. and B. KIRSTEIN, editors) (German). (Series: Teubner-Archiv zur Mathematik - 016). B.G.Teubner Verlagsgesellschaft, Stuttgart·Leipzig 1991, 290pp.

[S:Meth] *I. Schur Methods in Operator Theory and Signal Processing* (Operator Theory: Advances and Applications, **OT 18**) (GOHBERG, I., editor). Birkhäuser Verlag, Basel· Boston· Stuttgart 1986.

Harry Dym
Department of Mathematics
The Weizmann Institute of Science
Rehovot, 76100, Israel
E-mail: dym@wisdom.weizmann.ac.il

Victor Katsnelson
Department of Mathematics
The Weizmann Institute of Science
Rehovot, 76100, Israel
E-mail: katze@wisdom.weizmann.ac.il