A note on the maximization of matrix valued Hankel determinants with applications

Holger Dette
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum
Germany
email: holger.dette@ruhr-uni-bochum.de
FAX: +49 2 34 70 94 559

W. J. Studden
Dept. of Statistics
Purdue University
West-Lafayette IN 47907-1399
USA
email: studden@stat.purdue.edu

June 4, 2003

Abstract

In this note we consider the problem of maximizing the determinant of moment matrices of matrix measures. The maximizing matrix measure can be characterized explicitly by having equal (matrix valued) weights at the zeros of classical (one dimensional) orthogonal polynomials. The results generalize classical work of Schoenberg (1959) to the case of matrix measures. As a statistical application we consider several optimal design problems in linear models, which generalize the classical weighing design problems.

AMS Subject Classification: 42C05, 33C45, 62K05

Keywords and Phrases: Matrix measures, Hankel matrix, orthogonal polynomials, approximate optimal designs, spring balance weighing designs

1 Introduction

In recent years considerable interest has been shown in the area of matrix measures and matrix polynomials [see Rodman (1990), Sinap and Van Assche (1994), Duran and Van Assche (1995), Duran (1995, 1996, 1999), Duran and Lopez-Rodriguez (1996, 1997), Zygmunt (2001, 2002) among many others]. A matrix measure μ is a $p \times p$ matrix $\mu = \{\mu_{ij}\}$ of finite signed measures μ_{ij} on the Borel field of the real line $\mathbb{R}$ or of an appropriate subset. It will be assumed here that for each Borel set $A \subset \mathbb{R}$ the matrix $\mu(A) = \{\mu_{ij}(A)\}$ is symmetric and nonnegative definite, i.e. $\mu(A) \geq 0$. The moments of the matrix measure μ are given by the $p \times p$ matrices

$$(1.1) \qquad S_k = S_k(\mu) = \int t^k d\mu(t) \qquad k = 0, 1, \cdots$$

If it is not stated otherwise, the integrals will usually be over the interval $[0, 1]$. In the present note we are interested in the maximum of the determinant of the Hankel matrix

$$(1.2) \qquad \underline{H}_{2m} = \begin{pmatrix} S_0 & \dots & S_m \\ \vdots & & \vdots \\ S_m & \dots & S_{2m} \end{pmatrix},$$

where the maximum is taken over a certain subset

$$(1.3) \qquad \mathcal{N}^* = \{(S_0, \dots, S_{2m}) \in \mathcal{M}_{2m+1} \mid S_0 \in \mathcal{N}\}$$

of the moment space

$$(1.4) \qquad \mathcal{M}_{2m+1} = \{(S_0(\mu), \dots, S_{2m}(\mu)) \mid \mu \text{ matrix measure on } [0, 1]\}$$

where $\mathcal{N}$ is some subset of symmetric positive semi-definite matrices.

For the one dimensional case $p = 1$ problems of this type have been studied by Schoenberg (1959) and the solution of these problems are discrete measures supported on the roots of classical orthogonal polynomials [see Karlin and Studden (1966)]. Our interest in these types of problems stems from both a mathematical and practical viewpoint. On the one hand we are interested in generalizations of Schoenberg's results to the matrix case. On the other hand the optimization of the determinant of the Hankel matrix in (1.2) appears naturally in some areas of mathematical statistics, where optimal designs for linear regression experiments have to be determined. In Section 2 we present some recent facts on matrix measures, matrix orthogonal polynomials and matrix valued canonical moments [see Dette and Studden (2002a,b)]. This methodology simplifies the optimization problem substantially, and it can be shown that the matrix measure maximizing the Hankel determinant $|\underline{H}_{2m}|$ is a uniform distribution on $m+1$ points. These points are the points $0, 1$ and the roots of the derivative of the (one-dimensional) mth Legendre polynomial and do not depend on the particular choice of the set $\mathcal{N}$ in the definition of $\mathcal{N}^*$. The common weight matrix at these points is the solution of the optimization problem

$$(1.5) \qquad A^* = \text{ arg max}\{|S_0| \mid S_0 \in \mathcal{M}_1 \cap \mathcal{N}\}$$

[here and throughout this paper we assume that the maximum in (1.5) exists and is unique]. Finally, some statistical applications of the general results are discussed in Section 3.

2 Canonical moments and the maximum of the Hankel determinant

Let $\mathcal{N}$ denote a subset of the nonnegative definite matrices such that the maximum in (1.5) exists and is unique. In order to find

$$(2.1) \qquad \max\{|\underline{H}_{2m}(\mu)| \mid \mu \text{ matrix measure on } [0, 1]; S_0(\mu) \in \mathcal{N}\}$$

we need some basic facts on matrix measures and orthogonal matrix polynomials, which have recently been established by Dette and Studden (2002a,b) and will be briefly summarized here for the sake of completeness.

For a matrix measure μ on the interval $[0,1]$ with moments $S_j = \int_0^1 x^j d\mu(t)$ define the "Hankel" matrices

$$\underline{H}_{2m} = \begin{pmatrix} S_0 & \cdots & S_m \\ \vdots & & \vdots \\ S_m & \dots & S_{2m} \end{pmatrix}, \qquad \overline{H}_{2m} = \begin{pmatrix} S_1 - S_2 & \cdots & S_m - S_{m+1} \\ \vdots & & \vdots \\ S_m - S_{m+1} & \dots & S_{2m-1} - S_{2m} \end{pmatrix}, \tag{2.2}$$

and

$$\underline{H}_{2m+1} = \begin{pmatrix} S_1 & \cdots & S_{m+1} \\ \vdots & & \vdots \\ S_{m+1} & \dots & S_{2m+1} \end{pmatrix}, \qquad \overline{H}_{2m+1} = \begin{pmatrix} S_0 - S_1 & \cdots & S_m - S_{m+1} \\ \vdots & & \vdots \\ S_m - S_{m+1} & \dots & S_{2m} - S_{2m+1} \end{pmatrix}, \tag{2.3}$$

then Dette and Studden (2002a) showed that a point $(S_0, \cdots, S_n)$ is in the (interior) moment space M_{n+1} generated by the matrix measures on the interval $[0,1]$, if and only if the matrices $\underline{H}_n$ and $\overline{H}_n$ are nonnegative (positive) definite. The nonnegativity of the matrices $\underline{H}_n$ and $\overline{H}_n$ imposes bounds on the moments S_k as in the one dimensional case [see Dette and Studden (1997), Chapter 1]. To be precise let

$$\underline{h}_{2m}^T = (S_{m+1}, \cdots, S_{2m}) \ , \ \ \underline{h}_{2m-1}^T = (S_m, \cdots, S_{2m-1})$$
$$\bar{h}_{2m}^T = (S_m - S_{m+1}, \cdots, S_{2m-1} - S_{2m}) \ , \ \ \bar{h}_{2m-1}^T = (S_m - S_{m+1}, \cdots, S_{2m-2} - S_{2m-1})$$

and define $S_1^- = 0$ and

$$S_{n+1}^- = \underline{h}_n^T \underline{H}_{n-1}^{-1} \underline{h}_n, \quad n \geq 1 \ , \tag{2.4}$$

and $S_1^+ = S_0$, $S_2^+ = S_1$ and

$$S_{n+1}^+ = S_n - \bar{h}_n^T \bar{H}_{n-1}^{-1} \bar{h}_n, \quad n \geq 2 \ , \tag{2.5}$$

whenever the inverses of the Hankel matrices exist. It is to be noted and stressed that S_n^- and S_n^+ depend on $(S_0, S_1, \cdots, S_{n-1})$ although this is not mentioned explicitly. It follows from a straightforward calculation with partitioned matrices that $(S_0, \dots, S_{n-1})$ is in the interior of the moment space M_n if and only if $S_n^- < S_n^+$ in the sense of Loewner ordering (note that a matrix is positive definite if and only if its main subblock and the corresponding Schur complement are positive definite). Moreover, for $(S_0, \dots, S_n) \in M_{n+1}$ we have

$$S_n^- \leq S_n \leq S_n^+ \tag{2.6}$$

in the sense of Loewner ordering. If $(S_0, S_1, \cdots, S_{n-1})$ is in the interior of the moment space M_n, then we define the kth matrix canonical moment as the matrix

$$U_k = D_k^{-1}(S_k - S_k^-) \ , \qquad 1 \leq k \leq n \ , \tag{2.7}$$

where

$$D_k = S_k^+ - S_k^- . \tag{2.8}$$

These quantities are the analog of the classical canonical moments p_k in the scalar case [see Dette and Studden (1997)]. We will also make use of the quantities

$$V_k = I_p - U_k = D_k^{-1}(S_k^+ - S_k) \ , \qquad 1 \leq k \leq n \ . \tag{2.9}$$

The main results in Dette and Studden (2002a) are the following theorems. The first represents the width D_{n+1} of the moment space M_{n+1} in terms of the matrix canonical moments U_k and V_k and the second shows how canonical moments appear naturally in the three term recurrence relation for the monic matrix orthogonal polynomials.

Theorem 2.1. *If the point* $(S_0, \cdots, S_n)$ *is in the interior of the moment space* M_{n+1} *generated by the matrix measures on the interval* $[0, 1]$, *then*

$$D_{n+1} = S^+_{n+1} - S^-_{n+1} = S_0 U_1 V_1 U_2 V_2 \cdots U_n V_n. \tag{2.10}$$

A $p \times p$ matrix polynomial is a $p \times p$ matrix with polynomial entries. It is of degree n if all the polynomials are of degree less or equal than n and is usually written in the form

$$P(t) = \sum_{i=0}^{n} A_i t^i. \tag{2.11}$$

where the A_i are real $p \times p$ matrices. The matrix polynomial $P(t)$ is called monic if the highest coefficient satisfies $A_n = I_p$ where I_p denotes the $p \times p$ identity matrix. The (pseudo) inner product of two matrix polynomials is defined by

$$< P, Q > = \int P^T(t)\mu(dt)Q(t). \tag{2.12}$$

Sinap and Van Assche (1996) call this the 'right' inner product. The left inner product would put the transpose of the Q polynomial. The orthogonal polynomials are defined by orthogonalizing the sequence $I_p, tI_p, t^2 I_p, \cdots$ with respect to the above inner product. It is easy to see that matrix orthogonal polynomials satisfy a three term recurrence realtionship. The following result expresses the coefficients in this recurrence relation for the monic orthogonal polynomials in terms of canonical moments.

Theorem 2.2.
Let μ *denote a matrix measure on the interval* $[0, 1]$ *with matrix canonical moments* U_n, $n \in I\!N$.

1) *The sequence of monic orthogonal polynomials* $\{\underline{P}_k(x)\}_{k\geq 0}$ *with respect to the matrix measure* μ *satisfies the recurrence formula* $\underline{P}_0(x) = I_p, \underline{P}_{-1}(x) = 0$ *and for* $m \geq 0$

$$x\underline{P}_m(x) = \underline{P}_{m+1}(x) + \underline{P}_m(x)(\zeta_{2m+1} + \zeta_{2m}) + \underline{P}_{m-1}(x)\zeta_{2m-1}\zeta_{2m}, \tag{2.13}$$

where the quantities $\zeta_j \in \mathbb{R}^{p\times p}$ *are defined by* $\zeta_0 = 0$, $\zeta_1 = U_1$, $\zeta_j = V_{j-1}U_j$ *if* $j \geq 2$ *and the sequences* $\{U_j\}$ *and* $\{V_j\}$ *are given in (2.7) and (2.9).*

2) *The sequence of monic orthogonal polynomials* $\{\overline{Q}_k(x)\}_{k\geq 0}$ *with respect to the matrix measure* $x(1-x)d\mu(x)$ *satisfies the recurrence formula* $\overline{Q}_0(x) = I_p, \overline{Q}_{-1}(x) = 0$ *and for* $m \geq 0$

$$x\overline{Q}_m(x) = \overline{Q}_{m+1}(x) + \overline{Q}_m(x)(\gamma_{2m+2} + \gamma_{2m+3}) + \overline{Q}_{m-1}(x)\gamma_{2m+1}\gamma_{2m+2} \ . \tag{2.14}$$

where the quantities $\gamma_j \in \mathbb{R}^{p\times p}$ *are defined by* $\gamma_1 = V_1$, $\gamma_j = U_{j-1}V_j$ *if* $j \geq 2$ *and the sequences* $\{U_j\}$ *and* $\{V_j\}$ *are given in (2.7) and (2.9).*

In the following discussion moment points with $S_n = S_n^+$ for some $n \in \mathbb{N}$ will be of importance. Note that these points satisfy $U_n = I_p$ for some $n \in \mathbb{N}$.

Theorem 2.3. *Assume that $(S_0, \ldots, S_{2m-1}) \in Int(M_{2m})$, then the probability measure corresponding to the point $(S_0, \ldots, S_{2m-1}, S_{2m}^+)$ is uniquely determined. The roots are the different zeros of the polynomial*

$$x(1-x) det\, \overline{Q}_{m-1}(x),$$

where $\overline{Q}_{m-1}(x)$ is the $(m-1)$th monic orthogonal polynomial with respect to the matrix measure $x(1-x)d\mu(x)$. The weights are determined as the unique solution of the system

$$\begin{bmatrix} S_0 \\ S_1 \\ \vdots \\ S_m \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} I_p & I_p & \ldots & I_p & I_p \\ 0 & x_2 I_p & \ldots & x_{k-1} I_p & I_p \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & x_2^m I_p & \ldots & x_{k-1}^m I_p & I_p \\ O & \overline{Q}_{m-1}(x_2) & \ldots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & & \overline{Q}_{m-1}(x_{k-1}) & 0 \end{bmatrix} \begin{bmatrix} \Lambda_1 \\ \Lambda_2 \\ \vdots \\ \Lambda_{k-1} \\ \Lambda_k \end{bmatrix} \tag{2.15}$$

where $rank(\Lambda_i) = p \quad if\ x_i \in \{0,1\}$, $rank(\Lambda_i) = \ell_i \quad if\ \det \overline{Q}_{m-1}(x_i) = 0$, and ℓ_i is the multiplicity of x_i as a root of the polynomial $\det \overline{Q}_{m-1}(x)$.

For a proof of this result see Dette and Studden (2002b). We now present some new properties of matrix valued canonical moments, which turn out to be useful for the maximization of the determinant of the Hankel matrix.

Lemma 2.4. *The eigenvalues of the matrix valued canonical moments $U_1, U_2, \ldots$ are all real and located in the interval $[0,1]$, whenever the canonical moments are defined.*

Proof. Recall the definition of the canonical moments in (2.7) and consider the transformation

$$\begin{aligned} \tilde{U}_k &= (S_k^+ - S_k^-)^{1/2} U_k (S_k^+ - S_k^-)^{-1/2} \\ &= (S_k^+ - S_k^-)^{-1/2} (S_k - S_k^-)(S_k^+ - S_k^-)^{-1/2}. \end{aligned}$$

Obviously $\tilde{U}_k$ is nonnegative definite und symmetric. On the other hand

$$I_p - \tilde{U}_k = (S_k^+ - S_k^-)^{-1/2} (S_k^+ - S_k)(S_k^+ - S_k^-)^{-1/2}$$

is also nonnegative definite and symmetric. Therefore $\tilde{U}_k$ has real eigenvalues located in the interval $[0,1]$ and the assertion follows, because U_k and $\tilde{U}_k$ are similar.

□

Lemma 2.5. *Let μ denote a matrix measure on the interval $[0,1]$ with canonical moments $U_1, U_2, \ldots,$ then the determinant of the Hankel matrix can be represented as*

$$|H_{2m}(\mu)| = |S_0(\mu)|^{m+1} \cdot \prod_{j=1}^{m} \{|V_{2j-2}| \cdot |U_{2j-1}| \cdot |V_{2j-1}| \cdot |U_{2j}|\}^{m-j+1} \tag{2.16}$$

where $|V_0| = 1$.

Proof. From (2.7) and Theorem 2.1 we have $S_k - S_k^- = S_0 \prod_{j=1}^{k} V_{j-1} U_j$ with $V_0 = I_p$. Now a well known result on partitioned matrices [see e.g. Muirhead (1982), 581-582] and the definition of S_{2m}^- in (2.4) shows that

$$\begin{aligned} |\underline{\mathrm{H}}_{2m}(\mu)| &= |S_{2m} - S_{2m}^-| \cdot |\underline{\mathrm{H}}_{2m-2}(\mu)| \\ &= \prod_{j=0}^{m} |S_{2j} - S_{2j}^-| \\ &= |S_0|^{m+1} \prod_{j=1}^{m} \{|V_{2j-2}| \cdot |U_{2j-1}| \cdot |V_{2j-1}| \cdot |U_{2j}|\}^{m-j+1} . \end{aligned}$$

□

Theorem 2.6. *The determinant of the Hankel matrix defined in (1.2) attains its maximum in the set $\mathcal{N}^*$ defined in (1.3) if and only if the corresponding matrix measure μ has equal weight $\frac{1}{m+1} A^*$ at the roots of the polynomial*

$$x(1-x)P_m'(x), \tag{2.17}$$

where $P_m(x)$ denotes the mth (univariate) Legendre polynomial on the interval $[0,1]$ and the matrix A^* is defined in (1.5).

Proof. We determine the solution of the maximization problem in two steps. In a first step we use Lemma 2.5 to find the canonical moments of the maximizing measure and the corresponding moment S_0^*. In the second step we show that the corresponding moment point $(S_0^*, S_1^*, \ldots, S_{2m}^*)$ of the maximizing measure is in fact an element of the set $\mathcal{N}^*$ defined in (1.3).

By Lemma 2.5 the determinant can be maximized using the matrix A^* in (1.5) for S_0 and the canonical moments

$$U_{2k-1}^* = \frac{1}{2} I_p \quad U_{2k}^* = \frac{m-k+1}{2(m-k)+1} I_p \quad k = 1, \ldots, m. \tag{2.18}$$

The first assertion is obvious, while the second follows by maximizing the terms

$$|V_{2j-1}| \cdot |U_{2j-1}|; \quad |V_{2j}|^{m-j} |U_{2j}|^{m-j+1}$$

in (2.16) separately using Lemma 2.4. For example the eigenvalues of U_{2j-1}, say $\lambda_1, \ldots, \lambda_p$, are real and located in the interval $[0,1]$ and this gives

$$|V_{2j-1}||U_{2j-1}| = |(I_p - U_{2j-1})U_{2j-1}| = \prod_{i=1}^{p} \lambda_i (1 - \lambda_i).$$

The product is maximal for $\lambda_1 = \ldots = \lambda_p = 1/2$, which yields $U^*_{2j-1} = \frac{1}{2} I_p$, and the other cases are treated similary.

Now we obtain from (2.18) $U^*_{2m} = I_p$ which shows that the corresponding moment point $(S^*_0, \ldots, S^*_{2m})$ satisfies $S^*_{2m} = S^{*+}_{2m}$. By Theorem 2.3 it therefore follows that the maximizing measure μ^* is supported at the roots of the polynomial $x(1-x)\underline{Q}_{m-1}(x)$, where $\underline{Q}_{m-1}(x)$ is the $(m-1)$th monic orthogonal polynomial with respect to the matrix measure $x(1-x)\mu(dx)$. By Theorem 2.2 this polynomial can be obtained by the recursion $\underline{Q}_{-1}(x) = 0 \in \mathbb{R}^{p\times p}$, $\underline{Q}_0(x) = I_p$,

$$\underline{Q}_{j+1}(x) = (x - \frac{1}{2})\underline{Q}_j(x) - \frac{1}{4}\frac{(m-j-1)(m-j+1)}{(2(m-j)-1)(2(m-j)+1)}\underline{Q}_{j-1}(x).$$

Consequently, all matrix polynomials are diagonal. In particular $\underline{Q}_{m-1}(x) = I_p \cdot \tilde{Q}_{m-1}(x)$, where $x(1-x)\tilde{Q}_{m-1}(x)$ is the supporting polynomial of the (one-dimensional) sequence of canonical moments

$$\frac{1}{2}, \frac{m}{2m-1}, \frac{1}{2}, \frac{m-1}{2m-3}, \ldots, \frac{1}{2}, \frac{1}{3}, 1. \tag{2.19}$$

The results in Dette and Studden (1997), Chap. 4, show that $\tilde{Q}_{m-1}(x)$ is proportional to $P'_m(x)$. Similarly, observing the definition of canonical moments it is easy to see that the corresponding moments satisfy

$$S^*_k = c_k A^* \quad k = 0, 1, 2, \ldots$$

where $c_0, c_1, c_2, \ldots$ are the one-dimensional moments corresponding to the sequence of one dimensional canonical moments defined in (2.19). Therefore the system of equations in (2.15) reduces to

$$\begin{bmatrix} c_0 \\ \vdots \\ c_m \end{bmatrix} \otimes A^* = \left\{ \begin{bmatrix} 1 & 1 & \ldots & 1 & 1 \\ 0 & x_1 & & x_{m-1} & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & x_1^m & \ldots & x_{m-1} & 1 \end{bmatrix} \otimes I_p \right\} \begin{bmatrix} \Lambda_0 \\ \vdots \\ \Lambda_m \end{bmatrix}. \tag{2.20}$$

where $x_1, \ldots, x_{m-1}$ are the roots of the polynomial $P'_m(x)$ and throughtout this paper $\otimes$ denotes the Kronecker product. Because the D-optimal design for the univariate polynomial regression model has equal weights at the roots of the polynomial $P'_m(x)$ it follows that

$$\begin{bmatrix} 1 & 1 & \ldots & 1 & 1 \\ 0 & x_1 & & x_{m-1} & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & x_1^m & \ldots & x_{m-1}^m & 1 \end{bmatrix}^{-1} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_m \end{bmatrix} = \frac{1}{m+1} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix},$$

which implies for the unique solution of (2.20)

$$\Lambda_j = \frac{1}{m+1} A^*,$$

where the matrix A^* is defined by (1.5). Because the optimal canonical moments in (2.18) satisfy $0 < U^*_k < I_p \quad (k = 1, \ldots, 2m-1)$, $U^*_{2m} = I_p$, the corresponding moment point $(S^*_0, \ldots, S^*_{2m})$ is obviously an element of the set $\mathcal{N}^*$ defined in (1.3), which completes the proof of the theorem.

□

The following result is proved in a similar manner and its proof therefore omitted.

Theorem 2.7. *The quantity*

$$\frac{|\underline{H}_{2m}(\mu)|}{|\underline{H}_{2m-2}(\mu)|}$$

attains its maximum in the set $\mathcal{N}^*$ *defined in (1.3) if and only if the corresponding matrix measure* μ *puts weights*

$$\frac{1}{2m}A^*, \frac{1}{m}A^*, \dots, \frac{1}{m}A^*, \frac{1}{2m}A^*$$

at the roots of the polynomial

$$x(1-x)\tilde{U}_{m-1}(x),$$

where $\tilde{U}_{m-1}(x)$ *denotes the (one-dimensional) Chebyshev polynomial of the second kind on the interval* $[0,1]$ *and the matrix* A^* *is defined in (1.5).*

3 A statistical application

Consider a linear regression model of the form

$$Y_{\ell k} = \sum_{i=1}^{p} z_{\ell k i}\Big(\sum_{j=0}^{m} \beta_{ji} x_\ell^j\Big) + \varepsilon_{\ell k} \quad k = 1, \dots, n_\ell; \quad \ell = 1, \dots, s \tag{3.1}$$

where $\varepsilon_{11}, \dots, \varepsilon_{sn_s}$ are i.i.d. random variables with zero mean and variance $\sigma^2 > 0$, x_ℓ varies in the interval $[0,1]$ ($\ell = 1, \dots, s$), and $z_{\ell k i}$ ($k = 1, \dots, n_\ell; i = 1, \dots, p$) are chosen from either the set $\{0,1\}$ or the set $\{-1,0,1\}$. The quantities $z_{\ell k i}$ and x_ℓ can be controlled by the experimenter but the parameters β_{ij} are unknown and have to be estimated from the known data.

The model (3.1) is a generalization of the classical weighing design problem which corresponds to the case $m = 0$. [see Banerjee (1975) or Shah and Sinha (1989)]. For $m = 0$ there are p objects with unknown weights $\beta_{01}, \dots, \beta_{0p}$ which are to be determined using either a spring balance or a chemical balance. A total of n weighings of observations are allowed. For the spring balance any number of the objects can be placed on the single pan and the measurement represents the total weight of the objects. This corresponds to the case where $z_{\ell k i}$ are chosen from $\{0,1\}$. For $m = 0$ the ℓ subscript does not appear and z_{ki} is one or zero depending on whether the ith object is included in the kth weighing or not. If z_{lki} are chosen from $\{-1,0,1\}$ then one has a chemical balance with 2 pans and one can observe the difference in weight of any two subsets of the objects. The model (3.1) generalizes the classical weighing model to the case where the weight of all the objects depends on a variable x_ℓ by a polynomial trend. Note that the degree of the polynomial is the same for each of the objects.

In general, the model (3.1) can be conveniently written as

$$Y = X\beta + \varepsilon \tag{3.2}$$

where $Y = (Y_1, \dots, Y_n)^T$ denotes the vector of observations, ε ist the vector of (unobservable) errors, $\beta = (\beta_{01,}, \dots, \beta_{0p}, \dots, \beta_{m1}, \dots, \beta_{mp})^T$ is the vector of unknown parameters. The matrix X is called the design matrix and is given by

$$X = \begin{bmatrix} (1, x_1, \dots, x_1^m) \otimes Z_1 \\ (1, x_2, \dots, x_2^m) \otimes Z_2 \\ \vdots \\ (1, x_s, \dots, x_s^m) \otimes Z_s \end{bmatrix}, \tag{3.3}$$

where for $\ell = 1, \dots, s$ the matrix $Z_\ell = (z_{\ell k i})_{k=1,\dots,n_\ell}^{i=1,\dots,p}$ has entries 0 or 1 in the spring balance case. The least squares estimate of β is chosen to minmize the n-dimensional Euclidean norm $|Y - X\beta|$. This estimate is then given by $\hat\beta = (X^TX)^{-1}X^TY$ and its covariance matrix can be calculated as

$$\mathrm{Cov}(\hat\beta) \;=\; \sigma^2(X^TX)^{-1} = \sigma^2 \begin{bmatrix} \sum_{\ell=1}^s Z_\ell^T Z_\ell & \sum_{\ell=1}^s x_\ell Z_\ell Z_\ell & \dots & \sum_{\ell=1}^s x_\ell^m Z_\ell^T Z_\ell \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{\ell=1}^s x_\ell^m Z_\ell^T Z_\ell & \sum_{\ell=1}^s x_\ell^{m+1} Z_\ell^T Z_\ell & \dots & \sum_{\ell=1}^s x_\ell^{2m} Z_\ell^T Z_\ell \end{bmatrix} \;=\; \sigma^2 \underline{\mathrm{H}}_{2m}^{-1}(\xi),$$

where $\underline{\mathrm{H}}_{2m}(\xi)$ denotes the "Hankel" matrix of order $2m$ corresponding to the matrix measure ξ with weights $\Lambda_\ell = Z_\ell^T Z_\ell$ at the points $x_\ell \quad (\ell = 1, \dots, s)$. Note that the matrix Λ_ℓ is nonnegative definite (by its construction) and has nonnegative integer valued entries.

Our purpose here is to maximize the determinant of a normalized version of the matrix $\underline{\mathrm{H}}_{2m}(\xi)$ in an approximate sense which will be explained in the next paragraph, since, neither the weighing design problem ($m = 0$) nor the ordinary polynomial regression problem ($p = 1$), are completely solved in the situation above [see e.g. Banerjee (1975), Shah and Sinha (1989), Imhof, Krafft, Schaefer (2000)]. For $m = 0$, in the weighing design, a complete solution for the approximate case is given in Huda and Mukerjee (1988). The ordinary polynomial D-optimal design in the approximate sense is "well-known" and given, for example in Dette and Studden (1997). It will be shown that the two separate solutions can be combined in a simple manner. We give a rather detailed discussion of the spring balance design. The case of the chemical balance is very similar and left to the reader.

Let $\Omega \subset \mathbb{R}^p$ denote the set of all p-dimensional vectors with components in $\{0, 1\}$, which has $t = 2^p$ elements. A probability measure ξ with finite support on the set

$$\mathcal{X} := [0, 1] \times \Omega \tag{3.4}$$

is called a design on $\mathcal{X}$ [see Pukelsheim (1993)]. If n is the total number of observations and ξ puts mass $\xi_{\ell k}$ at the point $(x_\ell, \omega_k) \in \mathcal{X}$, then the experimenter takes approximately $n\xi_{\ell k}$ independent observations under experimental condition x_ℓ using the polynomial regression models corresponding to the nonvanishing components in the vector ω_k. For $(x, \omega) \in \mathcal{X}$ let $f^T(x, \omega) = (1, x, \dots, x^m) \otimes \omega^T$ denote the vector of regression functions of length $p(m + 1)$, then the information matrix is given by

$$M(\xi) \;=\; \int_{\mathcal{X}} f(x,\omega) f^T(x,\omega) d\xi(x,w) = \sum_{\ell=0}^{s}\sum_{k=0}^{t} \xi_{\ell k} \begin{pmatrix} 1 \\ x_\ell \\ \vdots \\ x_\ell^m \end{pmatrix} (1, x_\ell, \ldots, x_\ell^m) \otimes \omega_k \omega_k^T .$$

Without loss of generality we assume that $x_0, \ldots x_r$ are all distinct and define nonnegative matrix weights

(3.5) $$\Lambda_\ell = \sum_{k=0}^{t} \xi_{\ell k} \omega_k \omega_k^T \quad \ell = 0, \ldots, s.$$

If μ is the matrix measure with mass Λ_j at x_j $(j = 0, \ldots, s)$, then it follows that $M(\xi) = \underline{H}_{2m}(\mu)$, and a D-optimal (approximate) design can be obtained by maximizing $\underline{H}_{2m}(\mu)$ over the set of matrix measures

(3.6) $$\mu = \sum_{j=1}^{s} \Lambda_j \delta_{x_j}$$

with weights of the form (3.5). With reference to the previous paragraph we have that Λ_ℓ is approximately equal to $Z_\ell^T Z_\ell / n$, where n is the total number of observations.

Throughout this section we let

(3.7) $$\Xi^{**} = \{\mu \mid \mu \in \Xi \text{ is of the form (3.5) and (3.6)}\} \ ,$$

where Ξ is the set of all matrix measures on the interval $[0, 1]$. Obviously we have $\Xi^{**} \subset \Xi^*$, where

(3.8) $$\Xi^* = \{\mu \mid \mu \in \Xi \ \text{ with } S_0(\mu) = S_0(\nu) \text{ for some measure } \nu \in \Xi^{**} \ \}.$$

Note that the moments $(S_0, \ldots, S_{2m})$ of the matrix measures in Ξ^* define a set of the form (1.3).

Theorem 3.1. *Let $\mathcal{S}_j \subset \{0,1\}^p$ denote the set of all vectors with exactly j components equal to one $(j = 0, \ldots, p)$. A D-optimal (approximate) spring balance design for the regression model (3.1) is the uniform distribution on the set $\{(x, \omega) \mid x(1-x)P_m'(x) = 0, \omega \in \mathcal{S}_{\lfloor \frac{p}{2} \rfloor} \cup \mathcal{S}_{\lfloor \frac{p+1}{2} \rfloor}\}$, where P_m denotes the mth Legendre polynomial on the interval* $[0, 1]$.

Proof. Consider first the maximization of the Hankel determinant $\underline{\mathrm{H}}_{2m}(\mu)$ over the set Ξ^* defined in (3.8). According to Theorem 2.6 the solution of this problem is given by the matrix measure

(3.9) $$\mu^* = \frac{1}{m+1} A^* \sum_{j=0}^{m} \delta_{x_j},$$

where $x_0, \ldots, x_m$ are the roots of the polynomial $x(1-x)P_m'(x)$ and

(3.10) $$A^* = \ \mathrm{argmax}\Big\{|S_0| \Big| S_0 = \int_0^1 \mu(dx); \mu \in \Xi^{**}\Big\}.$$

Note that for the case $m = 0$ we do not hae to distinguish different explanatory variables and consequently the representation (3.5) does not depend on ℓ in this case. Therefore the optimization in (3.10) simplifies to

$$A^* = \operatorname{argmax}\Big\{|S_0|\Big| S_0 = \sum_{k=0}^{t} \xi_k \omega_k \omega_k^T; \sum_{k=0}^{t} \xi_k = 1, \xi_k > 0, \omega_k \in \Omega\Big\}. \tag{3.11}$$

This problem is the problem of finding the D-optimal information matrix for the classical spring balance weighing design (in the approximate case) and has been solved by Huda and Mukerjee (1988) using the classical equivalence theory for approximate designs. The D-optimal approximate spring balance weighing design ξ_0^* puts equal weight at the elements of the set $\mathcal{S}_{\lfloor p/2\rfloor} \cup \mathcal{S}_{\lfloor (p+1)/2\rfloor}$ and the maximizing information matrix is given by

$$A^* = \frac{\lfloor p/2\rfloor + 1}{2(2\lfloor p/2\rfloor + 1)}(I_p + J_p) = \binom{2\lfloor p/2\rfloor + 1}{\lfloor p/2\rfloor}^{-1} \sum_{\omega \in \mathcal{S}_{\lfloor p/2\rfloor} \cup \mathcal{S}_{\lfloor p/2\rfloor + 1}} \omega\omega^T \tag{3.12}$$

where $J_p \in \mathbb{R}^{p\times p}$ denotes the matrix with all entries equal to one. Therefore the measure maximizing $\underline{\mathrm{H}}_{2m}(\mu)$ in the class Ξ^* is of the form (3.9) with A^* given by (3.12). From the second equality in (3.12) it follows that the matrix measure μ^* is also in the class Ξ^{**} defined in (3.7) and therefore corresponds to an approximate design ξ. It is now easy to see that the design ξ_m^* specified in the Theorem 3.1 corresponds to μ^* by the relation (3.5), that is

$$\begin{aligned} |M(\xi_m^*)| &= |\underline{\mathrm{H}}_{2m}(\mu^*)| = \max\{|\underline{\mathrm{H}}_{2m}(\mu)| \mid \mu \in \Xi^*\} \\ &= \max|\{\underline{\mathrm{H}}_{2m}(\mu)| \mid \mu \in \Xi^{**}\} \\ &= \max\{|M(\xi)| \mid \xi \text{ design on } \mathcal{X}\}. \end{aligned}$$

This proves D-optimality of the design ξ_m^* specified in Theorem 3.1. □

Remark 3.2. It is worthwhile to mention that, once the design in Theorem 3.1 has been identified, its D-optimality can also be established by an application of the classical equivalence theorem for the D-optimality criterion [see Pukelsheim (1993), p. 180].

One can use the results of Section 2 to solve other optimization problems in statistics. Exemplarily we consider the case, where the main interest of the experimenter is to discriminate between a regression of degree m and $m-1$ in the model (3.1) and a design which maximizes

$$\frac{\underline{\mathrm{H}}_{2m}(\xi)}{\underline{\mathrm{H}}_{2m-2}(\xi)} \tag{3.13}$$

might be appropriate [see e.g. Studden (1980) or Pukelsheim (1993)]. These designs are called D_1-optimal and can be obtained by similar arguments as given in the proof of Theorem 3.1 using Theorem 2.7.

Theorem 3.3. *A D_1-optimal design ξ_1^* in the regression model (3.1) maximizing the ratio in (3.13) is supported at the set $T_1 = \{(x,\omega) \mid x(1-x)\tilde{U}_{m-1}(x) = 0, \omega \in \mathcal{S}_{\lfloor \frac{p}{2}\rfloor} \cup \mathcal{S}_{\lfloor \frac{p+1}{2}\rfloor}\}$ where $\tilde{U}_{m-1}(x)$*

denotes the $(m-1)th$ (one-dimensional) Chebyshev polynomial of the second kind on the interval $[0,1]$ *orthogonal with respect to the measure* $\sqrt{x(1-x)}dx$. *This design has equal masses*

$$\frac{1}{m}\Big(\binom{2\lfloor p/2\rfloor+1}{\lfloor p/2\rfloor}\Big)^{-1}$$

at the points $\{(x,\omega)\in T_1 \mid x\in(0,1)\}$ *while the masses at the points* $\{(x,\omega)\in T_1 \mid x\in\{0,1\}\}$ *are given by*

$$\frac{1}{2m}\Big(\binom{2\lfloor p/2\rfloor+1}{\lfloor p/2\rfloor}\Big)^{-1}.$$

Acknowledgements. This work was done while H. Dette was visiting the Department of Statistics, Purdue University, and this author would like to thank the department for its hospitality. The work of H. Dette was supported by the Deutsche Forschungsgemeinschaft (De 502/14-1, SFB 475, Komplexitätsreduktion in multivariaten Datenstrukturen). The authors would like to thank Isolde Gottschlich, who typed this paper with considerable technical expertise.

References

K.S. Banerjee (1975). Weighing Designs. Marcel Dekker, N.Y.

H. Dette, W.J. Studden (1997). The Theory of Canonical Moments with Applications in Statistics, Probability, and Analysis, Wiley.

H. Dette, W.J. Studden (2002a). Matrix measures, moment spaces and Favard's theorem on the interval $[0,1]$ and $[0,\infty)$. Linear Algebra Appl. 345, 163 - 193.

H. Dette, W.J. Studden (2002b). Quadrature formulas for matrix measures - a geometric approach. Linear Algebra Appl. 364, 34 - 65.

A.J. Duran (1995). On orthogonal polynomials with respect to a positive definite matrix of measures. Can. J. Math. 47, 88-112.

A. J. Duran (1996). Markov's Theorem for orthogonal matrix polynomials, Can. J. Math, 48, 1180-1195.

A.J. Duran (1999). Ratio asymptotics for orthogonal matrix polynomials. Journal of Approximation Theory 100, 304-344.

A.J. Duran, P. Lopez-Rodriguez (1996). Orthogonal matrix polynomials: zeros and Blumenthal's Theorem. Journal of Approximation Theory 84, 96-118.

A.J. Duran, P. Lopez-Rodriguez (1997). Density questions for the truncated matrix moment problem. Can. J. Math. 49, 708-721.

A. J. Duran, W. Van Assche (1995). Orthogonal matrix polynomials and higher-order recurrence relations. Lin. Alg. and Appl. 219, 261-280.

S. Huda, R. Mukerjee (1988). Optimal weighing designs: approximate theory. Statistics 19, 513-517.

L. Imhof, O. Krafft, M. Schaefer (2000). D-optimal exact designs for parameter estimation in a quadratic model. Sankhya, Ser. B 62, 266-275 (2000).

S. Karlin, W.J. Studden (1966). Tchebycheff Systems: with Applications in Analysis and Statistics. Wiley, New York.

R.J. Muirhead (1982). Aspects of Multivariate Statistical Theory. Wiley, N.Y.

F. Pukelsheim (1993). Optimal Design of Experiments. Wiley, N.Y.

L. Rodman (1990). Orthogonal matrix polynomials. In: P. Nevai (ed.), Orthogonal polynomials: theory and practice. NATO ASI Series C, Vol. 295; Kluwer, Dordrecht.

I.J. Schoenberg (1959). On the maxima of certain Hankel determinants and the zeros of classical orthogonal polynomials. Indag. Math. 62, 282-290.

K.R. Shah, B.K. Sinha (1989). Theory of Optimal Designs. Lecture Notes in Statistics, Springer, N.Y.

A. Sinap, W. Van Assche (1994). Polynomial interpolation and Gaussian quadrature for matrix-valued functions, Linear Algebra and Appl. 207, 71-114.

A. Sinap, W. Van Assche (1996). Orthogonal matrix polynomials and applications. Jour. Computational and Applied Math. 66, 27-52.

W.J. Studden (1980). D_s-optimal designs for polynomial regression using continued fractions. Ann. Stat. 8, 1132-1141.

M.J. Zygmunt (2001). Generalized Chebyshev polynomials and discrete Schrödinger operators. J. Phys. A: Math. Gen. 34, 10613-10618.

M.J. Zygmunt (2002). Matrix Chebyshev polynomials and continued fractions. Linear Algebra Appl. 227, 155-188.