PADOVA UNIVERSITY PRESS

Dolomites Research Notes on Approximation

Special issue dedicated to Mirosław Baran on the occasion of his 60th birthday, Volume 14 · 2021 · Pages 27–39

On the Limit of Optimal Polynomial Prediction Measures

Len Bos [a]

Abstract

Suppose that $K \subset \mathbb{C}$ is compact and that $z_0 \in \mathbb{C}\backslash K$ is an external point. An optimal prediction measure for regression by polynomials of degree at most n, is one for which the variance of the prediction at z_0 is as small as possible. Hoel and Levine ([4]) have considered the case of $K = [-1,1]$ and $z_0 = x_0 \in \mathbb{R}\backslash[-1,1]$, characterizing the optimal measures. More recently, [2] has given the equivalence of the optimal prediction problem with that of finding polynomials of extremal growth. They also study in detail the case of $K = [-1,1]$ and $z_0 = ia \in i\mathbb{R}$, purely imaginary. In this work we find, for these two cases, the limits of the optimal prediction measures as $n \to \infty$ and show that they are the push-forwards via conformal mapping of the Poisson kernel measure for the disk. Moreover, in the case of $z_0 = ia \in i\mathbb{R}$, we show that the optimal prediction measure of degree n is actually the Gauss-Lobatto quadrature formula for this limiting push-forward measure.

1 Introduction

Optimal Experimental Design has a rich history within Statistics. The interested reader may consult the classical book of Karlin and Studden [5] (especially Chapter X) or the more recent monograph of Dette and Studden [3]. A brief description of the statistical motivation of the subject of this work is also available in [2].

We state a generalized version of the problem. Suppose that $A, B \subset \mathbb{C}^d$ are two compact sets. Let

$$\mathcal{M}(A) := \{\mu : \mu \text{ is a probability measure on } A\}$$

and consider, for $\{q_1, \cdots, q_N\} \subset \mathbb{C}_n[z]$, an $L^2(\mu)$-orthonormal basis for $\mathbb{C}_n[z]$,

$$K_n^\mu(z,z) := \sum_{k=1}^N |q_k(z)|^2.$$

Here $\mathbb{C}_n[z]$ denotes the space of polynomials in $z = (z_1, \cdots, z_d)$ of degree at most n; $N := \dim(\mathbb{C}_n[z])$.

In other words, $K_n^\mu(z,z)$ is the diagonal of (reciprocal of) the Christoffel function (also known as the Bergman kernel) for $\mathbb{C}_n[z]$. The probability measure supported on A,

$$\mu_n = \operatorname{argmin}_{\mu \in \mathcal{M}(A)} \max_{z \in B} K_n(z,z)$$

is said to be G-optimal of degree n for A relative to B.

The case of $A = [-1,1]$ and $B = \{x_0\} \subset \mathbb{R}\backslash A$, a single point, was considered by Hoel and Levine [4] and generalized to what we called an optimal prediction measure in [2] for $A \subset \mathbb{C}^d$ and $B = \{z_0\} \in \mathbb{C}^d\backslash A$.

The case of $B = A$ is the classic case of optimal design, in which case the celebrated equivalence theorem of Kiefer and Wolfowitz [6] informs us that G-optimality is equivalent to what is called D-optimality. Moreover, in this case, it was shown in [1] that the weak-* limit of any sequence of optimal μ_n exists and is equal to the equilibrium measure of complex Pluripotential Theory.

In this work we begin a study of the limit of optimal prediction measures, conjecturing that potential theory continues to play an important role. Specifically, we consider three univariate examples of optimal prediction measures, the Hoel Levine case with $A = [-1,1]$ and $B = \{x_0\} \subset \mathbb{R}\backslash A$, the case of $A = \mathbb{D} \subset \mathbb{C}$, the unit disk, and $B = \{z_0\} \subset \mathbb{C}\backslash\mathbb{D}$, and also the case of $A = [-1,1]$, $B = \{z_0\} \subset i\mathbb{R}$, a purely imaginary point. We show, that for the disk, the limit of the optimal measure is the Poisson kernel measure for the reciprocal point $1/z_0$ and that in the other two cases the limit is the push-forward of this Poisson kernel measure under the Joukowski map, $J(z) = (z + 1/z)/2$.

2 The case of $A = \mathbb{D} \subset \mathbb{C}$, the Unit Disk

This case was first discussed in [2], but for the sake of completeness we present it again, in detail.

Consider a point $z_0 = |z_0|e^{i\phi}$ external to $\mathbb{D}$, i.e., such that $|z_0| > 1$. Consider the Poisson Kernel measure for the *reciprocal*, $1/z_0 = re^{-i\phi}$, $r = 1/|z_0| < 1$,

$$d\mu := \frac{1}{2\pi} P_r(\theta + \phi)d\theta = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} r^{|n|} e^{in(\theta+\phi)} d\theta. \tag{1}$$

[a] Department of Computer Science, University of Verona , Verona, Italy

Proposition 2.1. For any degree n, the Poisson kernel measure for $1/z_0$ is an optimal prediction measure for $z_0 \in \mathbb{C}\backslash\mathbb{D}$.
Proof. We begin by calculating the complex moments $m_j(\mu) := \int_{\partial\mathbb{D}} z^j d\mu$ of $d\mu$.
Lemma 2.2. We have

$$m_j(\mu) = \begin{cases} z_0^{-j}, & j \ge 0 \\ \overline{z}_0^{j}, & j < 0 \end{cases}.$$

Proof of the Lemma. First, for $j \ge 0$, we have

$$\begin{aligned} m_j(\mu) &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\sum_{n=-\infty}^{\infty}|z_0|^{-|n|}e^{in(\theta+\phi)}\right)z^j d\theta \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\sum_{n=-\infty}^{\infty}|z_0|^{-|n|}e^{in(\theta+\phi)}\right)e^{ij\theta} d\theta \\ &= \frac{1}{2\pi}\sum_{n=-\infty}^{\infty}|z_0|^{-|n|}e^{in\phi}\int_{-\pi}^{\pi}e^{in\theta}e^{ij\theta}d\theta \\ &= \sum_{n=-\infty}^{\infty}|z_0|^{-|n|}e^{in\phi}\delta_{n,-j} \\ &= |z_0|^{-j}e^{-ij\phi} \\ &= \left(|z_0|e^{i\phi}\right)^{-j} \\ &= z_0^{-j}. \end{aligned}$$

The $j < 0$ case follows from the fact that $m_j(\mu) = \overline{m_{-j}(\mu)}$. ■

Consequently the associated Gram matrix for the basis $\{1, z, z^2, \cdots, z^n\}$ is

$$(G_n(\mu))_{jk} := \int_{\partial\mathbb{D}} \overline{z}^j z^k d\mu = \int_{\partial\mathbb{D}} z^{(k-j)} d\mu = \begin{cases} z_0^{j-k}, & j \le k \\ \overline{z}_0^{k-j}, & j > k \end{cases},$$

$$G_n(\mu) = \begin{pmatrix} 1 & z_0^{-1} & z_0^{-2} & \cdot & \cdot & z_0^{-n} \\ \overline{z}_0^{-1} & 1 & z_0^{-1} & \cdot & \cdot & z_0^{-(n-1)} \\ \cdot & & & & & \cdot \\ \cdot & & & & & z_0^{-1} \\ \overline{z}_0^{-n} & \cdot & \cdot & \cdot & \overline{z}_0^{1} & 1 \end{pmatrix}.$$

It is easily verified by direct calculation that $G^{-1}(\mu)$ is the tri-diagonal matrix

$$\frac{|z_0|^2}{1-|z_0|^2}\begin{pmatrix} -1 & z_0^{-1} & 0 & \cdot & \cdot & 0 \\ \overline{z}_0^{-1} & -(1+1/|z_0|^2) & z_0^{-1} & 0 & \cdot & 0 \\ 0 & \overline{z}_0^{-1} & -(1+1/|z_0|^2) & z_0^{-1} & 0 & 0 \\ 0 & & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & \overline{z}_0^{-1} & -(1+1/|z_0|^2) & z_0^{-1} \\ 0 & \cdot & \cdot & \cdot & \overline{z}_0^{-1} & -1 \end{pmatrix}.$$

Using a well-known formula for $K_n^\mu(z,z)$ it follows that

$$\begin{aligned} K_n^\mu(z_0,z_0) &= \begin{pmatrix} 1 \\ z_0 \\ \cdot \\ \cdot \\ z_0^n \end{pmatrix}^* G^{-1}(\mu)\begin{pmatrix} 1 \\ z_0 \\ \cdot \\ \cdot \\ z_0^n \end{pmatrix} \\ &= |z_0|^{2n} \end{aligned}$$

where $*$ denotes the conjugate transpose.

Indeed, it is easy to verify that

$$G^{-1}(\mu)\begin{pmatrix} 1 \\ z_0 \\ \cdot \\ \cdot \\ z_0^n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \cdot \\ 0 \\ z_0^n \end{pmatrix}$$

from which the formula for $K_n^\mu(z_0,z_0)$ is an immediate consequence.

Now, we claim that, in fact

$$|z_0|^{2n} = \min_{\nu\in\mathcal{M}(\mathbb{D})} K_n^\nu(z_0,z_0)$$

and hence that μ is indeed an optimal prediction measure. To see this we use the variational form of the Christoffel function,

$$K_n^{\nu}(z_0,z_0)=\max_{\deg(p)\le n}\frac{|p(z_0)|^2}{\int_K |p(z)|^2 d\nu}.$$

Then, choosing $p(z)=z^n$ as a candidate in the maximum, we have

$$\begin{aligned} K_n^{\nu}(z_0,z_0) &\ge \frac{|z_0|^{2n}}{\int_{\mathbb{D}}|z|^{2n}d\nu} \\ &\ge |z_0|^{2n} \text{ (as } |z|\le 1) \\ &= K_n^{\mu}(z_0,z_0). \end{aligned}$$

■

3 The Case of $A=[-1,1]$ and $z_0\in\mathbb{R}\backslash[-1,1]$

In [4] Hoel and Levine show that in the univariate case, for $A=[-1,1]$, and *any* $z_0\in\mathbb{R}\backslash A$, a *real* external point, the optimal prediction measure is a discrete measure supported at the $n+1$ extremal points $x_k=\cos(k\pi/n)$, $0\le k\le n$, of $T_n(x)$ the classical Chebyshev polynomial of the first kind with weights given (also for $z_0\in\mathbb{C}\backslash[-1,1]$) by

Lemma 3.1. (Hoel-Levine [4]) Suppose that $-1=x_0<x_1<\cdots<x_n=+1$ are given and that $z_0\in\mathbb{C}\backslash[-1,1]$. Then among all discrete probability measures supported at these points, the measure $\mu=\sum_{i=0}^n w_i\delta_{x_i}$ with

$$w_i:=\frac{|\ell_i(z_0)|}{\sum_{i=0}^n|\ell_i(z_0)|},\ 0\le i\le n \tag{2}$$

with $\ell_i(z)$ the ith fundamental Lagrange interpolating polynomial for these points, minimizes $K_n^{\mu}(z_0,z_0)$.

Proof. For completeness we include a proof. We first note that for such a discrete measure, $\{\ell_i(z)/\sqrt{w_i}\}_{0\le i\le n}$ form an orthonormal basis. Hence

$$K_n^{\mu}(z_0,z_0)=\sum_{i=0}^n\frac{|\ell_i(z_0)|^2}{w_i}. \tag{3}$$

In the case of the weights chosen according to (2) we obtain

$$K_n^{\mu_0}(z_0,z_0)=\left(\sum_{i=0}^n|\ell_i(z_0)|\right)^2. \tag{4}$$

We claim that for any choice of weights K_n given by (3) is at least as large as that given by (4). To see this, just note that by the Cauchy-Schwartz inequality,

$$\begin{aligned} \left(\sum_{i=0}^n|\ell_i(z_0)|\right)^2 &= \left(\sum_{i=0}^n\frac{|\ell_i(z_0)|}{\sqrt{w_i}}\cdot\sqrt{w_i}\right)^2 \\ &\le \left(\sum_{i=0}^n\frac{|\ell_i(z_0)|^2}{w_i}\right)\cdot\left(\sum_{i=0}^n w_i\right) \\ &= \sum_{i=0}^n\frac{|\ell_i(z_0)|^2}{w_i}. \end{aligned}$$

■

We remark that in this case it turns out that

$$K_n^{\mu_0}(z_0,z_0)=T_n^2(z_0). \tag{5}$$

Proposition 3.2. Suppose that μ_n is the optimal prediction measure for the Hoel-Levine case. Then the weak-* limit

$$\lim_{n\to\infty}\mu_n=\frac{\sqrt{z_0^2-1}}{\pi}\frac{1}{|z_0-t|}\frac{dt}{\sqrt{1-t^2}}.$$

Proof. We first compute the Lagrange polynomials for the optimal support points $x_k=\cos(k\pi/n)$, $0\le k\le n$. Now

$$\ell_i(x)=\frac{\omega_n(x)}{(x-x_i)\omega_n'(x_i)}$$

where $\omega_n(x):=c_n\prod_{k=0}^n(x-x_k)$, for any constant $c_n\ne 0$.

Indeed, for these extended Chebyshev nodes, we may take

$$\omega_n(x)=T_{n+1}(x)-T_{n-1}(x)$$

as

$$\begin{aligned} T_{n+1}(x_i)-T_{n-1}(x_i) &= \cos\big((n+1)\frac{i\pi}{n}\big)-\cos\big((n-1)\frac{i\pi}{n}\big)\\ &= \cos\big(i\pi+\frac{i\pi}{n}\big)-\cos\big(i\pi-\frac{i\pi}{n}\big)\\ &= 0. \end{aligned}$$

Now, the formulas for $\omega_n'(x_i)$ turn out to be slightly different for the boundary points, $x_0=+1$ and $x_n=-1$. Consider first $0<i<n$. Then

$$\begin{aligned} \omega_n'(x_i) &= T_{n+1}'(x_i)-T_{n-1}'(x_i)\\ &= (n+1)\frac{\sin\big((n+1)i\pi/n\big)}{\sin(i\pi/n)}-(n-1)\frac{\sin\big((n-1)i\pi/n\big)}{\sin(i\pi/n)}. \end{aligned}$$

But

$$\begin{aligned} \sin\big((n+1)i\pi/n\big) &= \sin\big(i\pi+\frac{i\pi}{n}\big)\\ &= \cos(i\pi)\sin(i\pi/n)\\ &= (-1)^i\sin(i\pi/n) \end{aligned}$$

and similarly,

$$\sin\big((n-1)i\pi/n\big)=-(-1)^i\sin(i\pi/n).$$

Hence, for $0<i<n$,

$$\begin{aligned} \omega_n'(x_i) &= (-1)^i(n+1)\frac{\sin(i\pi/n)}{\sin(i\pi/n)}+(-1)^i(n-1)\frac{\sin(i\pi/n)}{\sin(i\pi/n)}\\ &= (-1)^i\{(n+1)+(n-1)\}\\ &= 2n(-1)^i \end{aligned}$$

while for $i=0$,

$$\begin{aligned} \omega_n'(x_0) &= \omega_n'(1)\\ &= T_{n+1}'(1)-T_{n-1}'(1). \end{aligned}$$

But

$$T_m'(1)=\lim_{\theta\to0} m\frac{\sin(m\theta)}{\sin(\theta)}=m^2$$

and so

$$\begin{aligned} \omega_n'(x_0) &= (n+1)^2-(n-1)^2\\ &= 4n. \end{aligned}$$

Similarly

$$\omega_n'(x_n)=4n(-1)^n.$$

Consequently

$$\ell_i(x)=(T_{n+1}(x)-T_{n-1}(x))\begin{cases}\frac{(-1)^i}{2n(x-x_i)}, & 0<i<n\\ \frac{(-1)^i}{4n(x-x_i)}, & i=0,n\end{cases}.$$

Now, assuming without loss of generality that the external prediction point $z_0>1$, we have that $T_{n+1}(z_0)>T_{n-1}(z_0)$, and so

$$|\ell_i(z_0)|=(T_{n+1}(z_0)-T_{n-1}(z_0))\begin{cases}\frac{1}{2n|z_0-x_i|}, & 0<i<n\\ \frac{1}{4n|z_0-x_i|}, & i=0,n\end{cases}.$$

It follows that, for $0<i<n$,

$$\begin{aligned} w_i &= \frac{(T_{n+1}(z_0)-T_{n-1}(z_0))\frac{1}{(2n)|z_0-x_i|}}{(T_{n+1}(z_0)-T_{n-1}(z_0))\sum_{k=0,\cdots,n}'\frac{1}{(2n)|z_0-x_k|}}\\ &= \frac{\frac{1}{|z_0-x_i|}}{\sum_{k=0,\cdots,n}'\frac{1}{|z_0-x_k|}} \end{aligned}$$

where the apostrophe on the summation means that the first and last terms are halved. For $i=0,n$

$$w_i=\frac{\frac{1/2}{|z_0-x_i|}}{\sum_{k=0,\cdots,n}'\frac{1}{|z_0-x_k|}}.$$

An integral with respect to the corresponding discrete measure, $d\mu_n$, is

$$\begin{aligned}\int_{-1}^{1} f(t)d\mu_n(t) &= \frac{\sum'_{k=0,\cdots,n} f(x_k)/|z_0-x_k|}{\sum'_{k=0,\cdots,n} \frac{1}{|z_0-x_k|}}\\ &= \frac{\sum'_{k=0,\cdots,n} f(\cos(k\pi/n))/|z_0-\cos(k\pi/n)|}{\sum'_{k=0,1,\cdots,n} \frac{1}{|z_0-\cos(k\pi/n)|}}\\ &= \frac{\frac{1}{n+1}\sum'_{k=0,\cdots,n} f(\cos(k\pi/n))/|z_0-\cos(k\pi/n)|}{\frac{1}{n+1}\sum'_{k=0,1,\cdots,n} \frac{1}{|z_0-\cos(k\pi/n)|}}.\end{aligned}$$

And hence

$$\begin{aligned}\lim_{n\to\infty}\int_{-1}^{1} f(t)d\mu_n(t) &= \frac{\int_0^\pi \frac{1}{|z_0-\cos(\theta)|} f(\cos(\theta))d\theta}{\int_0^\pi \frac{1}{|z_0-\cos(\theta)|}d\theta}\\ &= \frac{\sqrt{z_0^2-1}}{\pi}\int_{-1}^{1} f(t)\frac{1}{|z_0-t|}\frac{dt}{\sqrt{1-t^2}}\end{aligned}$$

as the integral in the denominator can be evaluated via a standard residue calculation to be

$$\int_0^\pi \frac{1}{|z_0-\cos(\theta)|}d\theta = \frac{\pi}{\sqrt{z_0^2-1}},\ z_0\in\mathbb{R}\backslash[-1,1].$$

■

3.1 The Limiting Optimal Measure as the Push-Forward of the Poisson Kernel Measure

It turns out that the limit of the optimal measure given by Prop. 3.2 is the push-forward of the Poisson Kernel measure (1) under the Joukowski conformal map

$$J(z) := \frac{1}{2}(z+1/z).$$

Indeed, for an external point $z_0 = |z_0|e^{i\phi}\in\mathbb{C}\backslash\mathbb{D}$, $|z_0|>1$, setting $r = 1/|z_0| < 1$,

$$\begin{aligned} d\mu &= \frac{1}{2\pi}P_r(\theta+\phi)d\theta\\ &= \frac{1}{2\pi}\frac{1-r^2}{1-2r\cos(\theta+\phi)+r^2}d\theta\\ &= \frac{1-r^2}{2\pi}\frac{1}{|z-1/z_0|^2}d\theta,\ z:=e^{i\theta}.\end{aligned}$$

Now for the Hoel-Levine case, let $x_0>1$ be the real external point. The Joukowski map provides a double covering of the interval $[-1,1]$ and has two branches to its inverse:

$$J^{-1}(z) = z\pm\sqrt{z^2-1}.$$

We take

$$z_0 = J^{-1}(x_0) = x_0+\sqrt{x_0^2-1}$$

with the sign chosen so that $|z_0|>1$. Now for $x=\cos(\theta)\in[-1,1]$, let $z = J^{-1}(x) = x+i\sqrt{1-x^2} = e^{i\theta}$ so that

$$\begin{aligned} 1-r^2 &= 1-\left(x_0-\sqrt{x_0^2-1}\right)^2\\ &= 1-\left\{x_0^2-2x_0\sqrt{x_0^2-1}+(x_0^2-1)\right\}\\ &= 2(-(x_0^2-1)+x_0\sqrt{x_0^2-1})\\ &= 2\sqrt{x_0^2-1}(x_0-\sqrt{x_0^2-1})\end{aligned}$$

while

$$\begin{aligned} |z-1/z_0|^2 &= |x+i\sqrt{1-x^2}-(x_0-\sqrt{x_0^2-1})|^2\\ &= (x-x_0+\sqrt{x_0^2-1})^2+(1-x^2)\\ &= (x-x_0)^2+2(x-x_0)\sqrt{x_0^2-1}+(x_0^2-1)+(1-x^2)\\ &= (x-x_0)(x-x_0+2\sqrt{x_0^2-1}-(x+x_0))\\ &= 2(x_0-x)(x_0-\sqrt{x_0^2-1}).\end{aligned}$$

Hence integrals transform as

$$\begin{aligned}\int_{-\pi}^{\pi} f(J(e^{i\theta}))d\mu(\theta) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\cos(\theta))\frac{2\sqrt{x_0^2-1}(x_0-\sqrt{x_0^2-1})}{2|x-x_0|(x_0-\sqrt{x_0^2-1})}d\theta \\ &= \frac{\sqrt{x_0^2-1}}{\pi}\int_0^{\pi} f(\cos(\theta))\frac{1}{|x-x_0|}d\theta \\ &= \frac{\sqrt{x_0^2-1}}{\pi}\int_{-1}^{1} f(x)\frac{1}{|x-x_0|}\frac{1}{\sqrt{1-x^2}}dx,\end{aligned}$$

i.e., the limiting measure for the Hoel-Levine case.

The relationship between the Poisson Kernel measure and the limiting measure for the interval may also be understood in terms of moments. Specifically, consider $f(x) = T_n(x)$, the classical Chebyshev polynomial of the first kind. Then, as is well-known, $f(J(z)) = T_n(J(z)) = J(z^n)$. Hence the Chebyshev moments

$$\begin{aligned}\int_{-1}^{1} T_n(x)d\mu(x) &= \frac{1-r^2}{2\pi}\int_{-\pi}^{\pi}\frac{(z^n+z^{-n})/2}{|z-1/z_0|^2}d\theta \\ &= (m_n(\mu)+m_{-n}(\mu))/2 \text{ (c.f. Lemma 2.2)} \\ &= z_0^{-n} \\ &= (x_0-\sqrt{x_0^2-1})^n.\end{aligned}$$

4 A Purely Imaginary Point External to $[-1, 1]$

Here we take the external point to be ai where, without loss of generality, we assume that $a > 0$.

We will prove

Proposition 4.1. The weak-* limit of the optimal prediction measures of degree n, μ_n, is the push-forward of the Poisson Kernel measure for the point $1/J^{-1}(ia)$, i.e.,

$$\lim_{n\to\infty}\mu_n = \mu.$$

Actually, this is a direct consequence of a stronger, perhaps surprising result.

Theorem 4.2. The optimal prediction measure, μ_n, is the Gauss-Lobatto quadrature measure of degree n for μ.

4.1 The Optimal Prediction Measure in Relation to the Push Forward Measure

The optimal prediction measure for degree n is characterized in [2] by means of two sequences of polynomials.

Definition 4.3. For $a > 0$ we define the sequences of polynomials $Q_n(z)$ and $R_n(z)$ by

$$\begin{aligned}Q_1(z) &= -\frac{az+i}{\sqrt{a^2+1}}, \\ Q_2(z) &= \frac{1}{\sqrt{a^2+1}}\left(-(a+\sqrt{a^2+1})z^2 - iz + \sqrt{a^2+1}\right), \\ Q_{n+1}(z) &= 2zQ_n(z) - Q_{n-1}(z), \quad n = 2, 3, \cdots.\end{aligned}$$

and

$$\begin{aligned}R_0(z) &= \frac{a}{\sqrt{a^2+1}}, \\ R_1(z) &= \frac{a+\sqrt{a^2+1}}{\sqrt{a^2+1}}z, \\ R_{n+1}(z) &= 2zR_n(z) - R_{n-1}(z), \quad n = 1, 2, \cdots.\end{aligned}$$

Since the recursions are both those of the classical Chebyshev polynomials it is not surprising that there are formulas for $Q_n(z)$ and $R_n(z)$ in terms of these.

Lemma 4.4. We have

$$Q_n(z) = \frac{1}{\sqrt{a^2+1}}\left(-(az+i)T_{n-1}(z) + \sqrt{a^2+1}(1-z^2)U_{n-2}(z)\right)$$

where $T_n(z)$ is Chebyshev polynomial of the first kind and $U_n(z) := \frac{1}{n+1}T'_{n+1}(z)$ that of the second kind.

Lemma 4.5. We have

$$\begin{aligned}R_n(z) &= \frac{1}{\sqrt{a^2+1}}\left(\sqrt{a^2+1}zU_{n-1}(z) + aT_n(z)\right) \\ &= \frac{1}{\sqrt{a^2+1}}\left\{\frac{\sqrt{a^2+1}+a}{2}U_n(z) + \frac{\sqrt{a^2+1}-a}{2}U_{n-2}(z)\right\}.\end{aligned}$$

The Chebyshev polynomials satisfy a so-called Pell identity, $T_n^2(x) - (x^2-1)U_{n-1}^2(x) \equiv 1$. The polynomials $Q_n(z)$ and $R_n(z)$ are also related by a Pell identity

Proposition 4.6. For $z = x \in \mathbb{R}$, we have
$$|Q_n(x)|^2 - (x^2-1)R_{n-1}^2(x) \equiv 1.$$

As shown in [2], the optimal prediction measure of degree n, μ_n, is the discrete measure supported on the zeros of $(x^2-1)R_{n-1}(x)$ with weights given by the Hoel-Levine formula (Lemma 3.1)
$$w_i := \frac{|\ell_i(ia)|}{\sum_{i=0}^n |\ell_i(ia)|},\ 0 \le i \le n.$$

We claim that, just as in the case of a real external point, the weak-* limit of the optimal prediction measures is the push-forward of the associated Poisson Kernel measure
$$d\mu = \frac{1-r^2}{2\pi}\frac{1}{|z-1/z_0|^2}d\theta,\ z := e^{i\theta}.$$
under the Joukowski map. We first calculate this push-forward measure.

Let
$$z_0 = J^{-1}(ia) = i(a+\sqrt{a^2+1})$$
where the sign is chosen so that $|z_0| > 1$. The reciprocal is then
$$1/z_0 = i(a-\sqrt{a^2+1}).$$
Hence $r = \sqrt{a^2+1}-a$ and
$$\begin{aligned} 1-r^2 &= 1-\left(\sqrt{a^2+1}-a\right)^2 \\ &= 1-\left(2a^2+1-2a\sqrt{a^2+1}\right) \\ &= 2a\left(\sqrt{a^2+1}-a\right). \end{aligned}$$
Also, for $z = e^{i\theta}$,
$$\begin{aligned} |z-1/z_0|^2 &= |e^{i\theta}-i(a-\sqrt{a^2+1})|^2 \\ &= \cos^2(\theta)+\left(\sin(\theta)-(a-\sqrt{a^2+1})\right)^2 \\ &= \cos^2(\theta)+\sin^2(\theta)+(a-\sqrt{a^2+1})^2-2(a-\sqrt{a^2+1})\sin(\theta) \\ &= 1+2a^2+1-2a\sqrt{a^2+1}-2(a-\sqrt{a^2+1})\sin(\theta) \\ &= 2\{\sqrt{a^2+1}(\sqrt{a^2+1}-a)-(a-\sqrt{a^2+1})\sin(\theta)\} \\ &= 2(\sqrt{a^2+1}-a)(\sqrt{a^2+1}+\sin(\theta)). \end{aligned}$$
Hence integrals transform as
$$\begin{aligned} \int_{-\pi}^{\pi} f(J(z))d\mu(\theta) &= \frac{1}{2\pi}\frac{a\left(\sqrt{a^2+1}-a\right)}{\sqrt{a^2+1}-a}\int_{-\pi}^{\pi} f(\cos(\theta))\frac{1}{\sqrt{a^2+1}+\sin(\theta)}d\theta \\ &= \frac{a}{2\pi}\int_{-\pi}^{\pi} f(\cos(\theta))\frac{1}{\sqrt{a^2+1}+\sin(\theta)}d\theta \\ &= \frac{a}{2\pi}\Big\{\int_0^{\pi} f(\cos(\theta))\frac{1}{\sqrt{a^2+1}+\sin(\theta)}d\theta \\ &\quad + \int_{-\pi}^{0} f(\cos(\theta))\frac{1}{\sqrt{a^2+1}+\sin(\theta)}d\theta\Big\} \\ &= \frac{a}{2\pi}\int_0^{\pi} f(\cos(\theta))\Big\{\frac{1}{\sqrt{a^2+1}+\sin(\theta)}+\frac{1}{\sqrt{a^2+1}-\sin(\theta)}\Big\}d\theta \\ &= \frac{a}{2\pi}\int_0^{\pi} f(\cos(\theta))\frac{2\sqrt{a^2+1}}{a^2+1-\sin^2(\theta)}d\theta \\ &= \frac{a\sqrt{a^2+1}}{\pi}\int_{-1}^{1} f(x)\frac{1}{a^2+1-(1-x^2)}\frac{dx}{\sqrt{1-x^2}} \\ &= \frac{a\sqrt{a^2+1}}{\pi}\int_{-1}^{1} f(x)\frac{1}{a^2+x^2}\frac{dx}{\sqrt{1-x^2}}. \end{aligned}$$
In other words, the push forward measure is
$$d\mu = \frac{a\sqrt{a^2+1}}{\pi}\frac{1}{a^2+x^2}\frac{dx}{\sqrt{1-x^2}}. \tag{6}$$

We may use Lemma 2.2 to note that the Chebyshev moments are given by
$$\int_{-1}^{1} T_n(x)d\mu(x) = \frac{1-r^2}{2\pi}\int_{-\pi}^{\pi}\frac{(z^n+z^{-n})/2}{|z-1/z_0|^2}d\theta$$

$$\begin{aligned} &= (m_n(\mu)+m_{-n}(\mu))/2 \\ &= (z_0^{-n}+\overline{z}_0^{-n})/2 \\ &= \Re(z_0^{-n}) \\ &= \Re\big((i(a-\sqrt{a^2+1}))^n\big) \end{aligned}$$

which we formally express as

Lemma 4.7. For the push-forward measure (6), we have

$$\int_{-1}^{1} T_n(x)d\mu = (\sqrt{a^2+1}-a)^n \begin{cases} 0, & n \text{ odd} \\ (-1)^{n/2}, & n \text{ even} \end{cases}.$$

The integrals of the Chebyshev polynomials of the second kind are also of interest.

Lemma 4.8. We have

$$I_n := \int_{-1}^{1} U_n(x)d\mu = \begin{cases} 0, & n \text{ odd} \\ \frac{(-1)^{n/2}(\sqrt{a^2+1}-a)^{n+1}+a}{\sqrt{a^2+1}}, & n \text{ even} \end{cases}.$$

Proof. If n is odd then the Chebyshev polynomial of the second kind, $U_n(x)$, is an odd function and, from (6), $d\mu$ is symmetric about the origin, and hence the integral is 0. If n is even we may use the classical identity

$$U_n(x) = 2\sum_{j=0}^{n/2} T_{2j}(x) - 1$$

to obtain, by Lemma 4.7,

$$\begin{aligned} I_n &= 2\sum_{j=0}^{n/2}(-1)^j(\sqrt{a^2+1}-a)^{2j} - 1 \\ &= 2\sum_{j=0}^{n/2}(-(\sqrt{a^2+1}-a)^2)^j - 1 \\ &= 2\frac{(-(\sqrt{a^2+1}-a)^2)^{n/2+1}}{-(\sqrt{a^2+1}-a)^2-1} - 1 \\ &= 2\frac{(-1)^{n/2}(\sqrt{a^2+1}-a)^{n+2}+1}{2(a^2+1-a\sqrt{a^2+1})} - 1 \\ &= \frac{(-1)^{n/2}(\sqrt{a^2+1}-a)^{n+2}+1}{\sqrt{a^2+1}(\sqrt{a^2+1}-a)} - 1 \\ &= \frac{(-1)^{n/2}(\sqrt{a^2+1}-a)^{n+1}+a}{\sqrt{a^2+1}}. \end{aligned}$$

■

4.2 Useful Lemmas

Lemma 4.9. The polynomials $R_n(x)$ are orthogonal with respect to the measure

$$d\widetilde{\mu} = (1-x^2)d\mu = \frac{a\sqrt{a^2+1}}{\pi}\frac{\sqrt{1-x^2}}{a^2+x^2}dx.$$

Proof (due to F. Wielonsky [8]). Consider

$$\begin{aligned} P_n(x) &:= \frac{2\sqrt{a^2+1}}{\sqrt{a^2+1}-a}R_n(x) \\ &= (\sqrt{a^2+1}+a)^2U_n(x)+U_{n-2}(x) \quad \text{(by Lemma 4.5)} \end{aligned}$$

and set

$$h(z) := z^2 + (\sqrt{a^2+1}+a)^2.$$

It is easy to verify that

$$P_n(\cos(\theta)) = \frac{1}{\sin(\theta)}\Im\big(e^{i(n+1)\theta}h(e^{-i\theta})\big).$$

Hence by Theorem 2.6 of [7], the P_n are orthogonal on $[-1,1]$ with respect to the measure $w(x)dx$ where

$$w(\cos(\theta)) = \frac{\sin(\theta)}{|h(e^{i\theta})|^2}.$$

But, setting $C = (\sqrt{a^2+1}+a)^2$, it is easy to verify that $(C-1)^2 = 4a^2C$. Hence

$$\begin{aligned} |h(e^{i\theta})|^2 &= (\cos(2\theta)+C)^2+\sin^2(2\theta) \\ &= 1+C^2+2C\cos(2\theta) \\ &= 1+C^2+2C(2x^2-1) \\ &= (C-1)^2+4Cx^2 \\ &= 4a^2C+4Cx^2 \\ &= 4C(a^2+x^2). \end{aligned}$$

It follows that the $R_n(x)$ are orthogonal with respect to the measure

$$\frac{\sqrt{1-x^2}}{a^2+x^2}dx,$$

a constant multiple of $d\widetilde{\mu}$. ■

Lemma 4.10. The norm (squared) of R_n with respect to $\widetilde{\mu}$ is

$$\int_{-1}^{1} R_n^2(x)d\widetilde{\mu} = \frac{a}{2\sqrt{a^2+1}}.$$

Proof. We calculate, using Lemma 4.5,

$$\begin{aligned} \int_{-1}^{1} R_n^2(x)d\widetilde{\mu} &= \frac{1}{4(a^2+1)}\int_{-1}^{1}\left\{(\sqrt{a^2+1}+a)U_n(x)+(\sqrt{a^2+1}-a)U_{n-2}(x)\right)^2 d\widetilde{\mu} \\ &= \frac{1}{4(a^2+1)}\left\{(\sqrt{a^2+1}+a)^2\int_{-1}^{1}U_n^2(x)d\widetilde{\mu}+2\int_{-1}^{1}U_n(x)U_{n-2}(x)d\widetilde{\mu}\right. \\ &\quad \left.+(\sqrt{a^2+1}-a)^2\int_{-1}^{1}U_{n-2}^2(x)d\widetilde{\mu}\right\}. \end{aligned}$$

By Lemmas 4.11 and 4.12 this equals

$$\begin{aligned} &\frac{1}{4(a^2+1)}\left\{(\sqrt{a^2+1}+a)^2\frac{1+(-1)^n(\sqrt{a^2+1}-a)^{2(n+1)}}{2}\right. \\ &-2\frac{(\sqrt{a^2+1}-a)^2+(-1)^n(\sqrt{a^2+1}-a)^{2n}}{2} \\ &+ \quad \left.(\sqrt{a^2+1}-a)^2\frac{1+(-1)^{n-2}(\sqrt{a^2+1}-a)^{2(n-1)}}{2}\right\} \\ =\ &\frac{1}{8(a^2+1)}\left\{(\sqrt{a^2+1}+a)^2-2(\sqrt{a^2+1}-a)^2+(\sqrt{a^2+1}-a)^2+(-1)^n\times 0\right\} \\ =\ &\frac{1}{8(a^2+1)}\left\{(\sqrt{a^2+1}+a)^2-(\sqrt{a^2+1}-a)^2\right\} \\ =\ &\frac{1}{8(a^2+1)}4a\sqrt{a^2+1} \\ =\ &\frac{a}{2\sqrt{a^2+1}}. \end{aligned}$$

■

Lemma 4.11. We have

$$\int_{-1}^{1} U_n^2(x)d\widetilde{\mu} = \frac{1}{2}\left\{1+(-1)^n(\sqrt{a^2+1}-a)^{2(n+1)}\right\}.$$

Proof. Using the Pell identity,

$$T_n^2(x)+(1-x^2)U_{n-1}^2(x)\equiv 1,$$

we calculate, by Lemma 4.7,

$$\begin{aligned} \int_{-1}^{1} U_n^2(x)d\widetilde{\mu} &= \int_{-1}^{1} U_n^2(x)(1-x^2)d\mu \\ &= \int_{-1}^{1}(1-T_{n+1}^2(x))\,d\mu \\ &= \int_{-1}^{1}\frac{1-T_{2(n+1)}(x)}{2}d\mu \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2}\Big\{1-\int_{-1}^{1} T_{2(n+1)}(x)d\mu\Big\}\\
&= \frac{1}{2}\{1-(-1)^{n+1}(\sqrt{a^2+1}-a)^{2(n+1)}\}\\
&= \frac{1}{2}\{1+(-1)^{n}(\sqrt{a^2+1}-a)^{2(n+1)}\}.
\end{aligned}
$$

■

Lemma 4.12. We have

$$\int_{-1}^{1} U_n(x)U_{n-2}(x)d\widetilde{\mu} = -\frac{1}{2}\{(\sqrt{a^2+1}-a)^2+(-1)^n(\sqrt{a^2+1}-a)^{2n}\}.$$

Proof. From the trigonometric identity

$$2\sin((n+1)\theta)\sin((n-1)\theta) = \cos(2\theta)-\cos(2n\theta)$$

it follows that

$$(1-x^2)U_n(x)U_{n-2}(x) = \frac{1}{2}\big(T_2(x)-T_{2n}(x)\big).$$

Hence, by Lemma 4.7,

$$
\begin{aligned}
\int_{-1}^{1} U_n(x)U_{n-2}(x)d\widetilde{\mu} &= \int_{-1}^{1}(1-x^2)U_n(x)U_{n-2}(x)d\mu\\
&= \frac{1}{2}\int_{-1}^{1}(T_2(x)-T_{2n}(x))\,d\mu\\
&= \frac{1}{2}\{-(\sqrt{a^2+1}-a)^2-(-1)^n(\sqrt{a^2+1}-a)^{2n}\}\\
&= -\frac{1}{2}\{(\sqrt{a^2+1}-a)^2+(-1)^n(\sqrt{a^2+1}-a)^{2n}\}.
\end{aligned}
$$

■

The relations

$$
\begin{aligned}
T_n^2(x)-T_{n-1}(x)T_{n+1}(x) &\equiv 1-x^2\\
U_n^2(x)-U_{n-1}(x)U_{n+1}(x) &\equiv 1
\end{aligned}
$$

are due to Turan. A similar identity that is easy to verify is that

$$U_{n-2}(x)U_n(x)-U_{n-3}(x)U_{n+1}(x) \equiv 4x^2-1. \tag{7}$$

It turns out that there is also a Turan-type relation for the $R_n(x)$.

Proposition 4.13. We have

$$R_n^2(x)-R_{n-1}(x)R_{n+1}(x) \equiv \frac{a^2+x^2}{a^2+1}.$$

Proof. We again use the formula of Lemma 4.5

$$R_n(x) = \frac{1}{2\sqrt{a^2+1}}\{(\sqrt{a^2+1}+a)U_n(x)+(\sqrt{a^2+1}-a)U_{n-2}(x)\}.$$

Then

$$
\begin{aligned}
4(a^2+1)R_n^2(x) &= (\sqrt{a^2+1}+a)^2U_n^2(x)+2U_n(x)U_{n-2}(x)\\
&\quad +(\sqrt{a^2+1}-a)^2U_{n-2}^2(x)
\end{aligned}
$$

while

$$
\begin{aligned}
4(a^2+1)R_{n-1}(x)R_{n+1}(x) &= (\sqrt{a^2+1}+a)^2U_{n-1}(x)U_{n+1}(x)\\
&\quad +(\sqrt{a^2+1}-a)^2U_{n-1}(x)U_{n-3}(x)\\
&\quad +U_{n-1}^2(x)+U_{n-3}(x)U_{n+1}(x).
\end{aligned}
$$

Hence

$$
\begin{aligned}
&4(a^2+1)(R_n^2(x)-R_{n-1}(x)R_{n+1}(x))\\
&=(\sqrt{a^2+1}+a)^2(U_n^2(x)-U_{n-1}(x)U_{n+1}(x))
\end{aligned}
$$

$$
\begin{aligned}
&+(\sqrt{a^2+1}-a)^2(U_{n-2}^2(x)-U_{n-3}(x)U_{n-1}(x)) \\
&+2U_{n-2}(x)U_n(x)-U_{n-1}^2(x)-U_{n-3}(x)U_{n+1}(x) \\
=&(\sqrt{a^2+1}+a)^2\times 1 \\
&+(\sqrt{a^2+1}-a)^2\times 1 \\
&-(U_{n-1}^2(x)-U_{n-2}(x)U_n(x)) \\
&+U_{n-2}(x)U_n(x)-U_{n-3}(x)U_{n+1}(x) \\
=&2(2a^2+1)-1+(4x^2-1)\ \text{(by (7))} \\
=&4(a^2+x^2).
\end{aligned}
$$

■

4.3 Proof of Theorem 4.2

The support points for the Gauss-Lobatto quadrature rule for $d\mu$ are the endpoints ± 1 together with the zeros of the orthogonal polynomial of degree $n-1$ with respect to the measure $d\widetilde{\mu}=(1-x^2)d\mu$, in this case $R_{n-1}(x)$. Hence the support of the Gauss-Lobatto rule and the optimal prediction measure of degree n are the same. We need to show that the associated weights are also the same. If we denote the support by

$$-1=x_0<x_1<\cdots<x_n=+1$$

then the weights for the optimal prediction measure are (cf. Lemma 3.1)

$$w_j=|\ell_j(ia)|/\sum_{k=0}^{n}|\ell_k(ia)|.$$

On the other hand the weights for the Gauss-Lobatto rule are

$$\omega_j=\int_{-1}^{1}\ell_j(x)d\mu.$$

We will show that

$$w_j=\omega_j\quad 0\le j\le n.$$

Consider first for the interior points, $1\le j\le n-1$. It is shown in [2] that

$$\sum_{k=0}^{n}|\ell_k(ia)|=|Q_n(ai)|=\sqrt{a^2+1}(\sqrt{a^2+1}+a)^{n-1}.$$

Further, for the interior points, $1\le j\le n-1$,

$$\ell_j(z)=\frac{1-z^2}{1-x_j^2}\widetilde{\ell}_j(z)$$

where $\widetilde{\ell}_j(z)$ is the Lagrange polynomial of degree $n-2$ for the points $x_1,\cdots,x_{n-1}$, i.e.,

$$\ell_j(z)=\frac{1-z^2}{1-x_j^2}\frac{R_{n-1}(z)}{R'_{n-1}(x_j)(z-x_j)}.$$

It follows that

$$|\ell_j(ai)|=\frac{1+a^2}{1-x_j^2}\frac{|R_{n-1}(ai)|}{|R'_{n-1}(ai)|\sqrt{a^2+x_j^2}}.$$

Now, it is easy to verify that

$$R_{n-1}(ai)=\frac{a}{\sqrt{a^2+1}}(i(\sqrt{a^2+1}+a))^{n-1}.$$

Hence

$$
\begin{aligned}
w_j &= \frac{|\ell_j(ai)|}{|Q_n(ai)|} \\
&= \frac{1+a^2}{1-x_j^2}\frac{|R_{n-1}(ai)|}{|Q_n(ai)|}\frac{1}{|R'_{n-1}(ai)|\sqrt{a^2+x_j^2}} \\
&= \frac{1+a^2}{1-x_j^2}\frac{a}{a^2+1}\frac{1}{|R'_{n-1}(ai)|\sqrt{a^2+x_j^2}} \\
&= \frac{a}{1-x_j^2}\frac{1}{|R'_{n-1}(ai)|\sqrt{a^2+x_j^2}}. \qquad (8)
\end{aligned}
$$

Further, the Gauss-Lobatto weight

$$\begin{aligned}\omega_j &= \int_{-1}^{1} \ell_j(x)d\mu \\ &= \int_{-1}^{1} \frac{1-x^2}{1-x_j^2}\widetilde{\ell}_j(x)d\mu \\ &= \frac{1}{1-x_j^2}\int_{-1}^{1} \widetilde{\ell}_j(x)(1-x^2)d\mu \\ &= \frac{1}{1-x_j^2}\int_{-1}^{1} \widetilde{\ell}_j(x)d\widetilde{\mu} \\ &= \frac{1}{1-x_j^2}\widehat{\omega}_j\end{aligned}$$

where $\widehat{\omega}_j$ are the Gauss Quadrature weights for degree $n-2$ and measure $d\widetilde{\mu}$.

Hence

$$\omega_j = \frac{1}{1-x_j^2}\frac{1}{K_{n-2}(x_j,x_j)}$$

where $K_m(x,y)$ is the reproducing kernel for the measure $d\widetilde{\mu}$. But, as the $R_n(x)$ are orthogonal with respect to $d\widetilde{\mu}$, the Christoffel-Darboux formula yields

$$\begin{aligned}K_{n-2}(x_j,x_j) &= \frac{k_{n-2}}{h_{h-2}k_{n-1}}\left[R'_{n-1}(x_j)R_{n-2}(x_j)-R'_{n-2}(x_j)R_{n-1}(x_j)\right] \\ &= \frac{k_{n-2}}{h_{n-2}k_{n-1}}R'_{n-1}(x_j)R_{n-2}(x_j) \quad (\text{since } R_{n-1}(x_j)=0)\end{aligned}$$

where k_m is the leading coefficient of $R_m(x)$ and $h_m = \int_{-1}^{1} R_m^2(x)d\widetilde{\mu}$. Now, by the recurrence formula,

$$\frac{k_{n-2}}{k_{n-1}} = \frac{1}{2}$$

and by Lemma 4.10,

$$h_{n-2} = \frac{a}{2\sqrt{a^2+1}}.$$

Hence

$$K_{n-2}(x_j,x_j) = \frac{\sqrt{a^2+1}}{a}R'_{n-1}(x_j)R_{n-2}(x_j).$$

Further, from the Turan identity, Proposition 4.13, as $R_{n-1}(x_j)=0$,

$$\begin{aligned}\frac{a^2+x_j^2}{a^2+1} &= R_{n-2}^2(x_j)-R_{n-1}(x_j)R_{n-3}(x_j) \\ &= R_{n-2}^2(x_j).\end{aligned}$$

Consequently,

$$\begin{aligned}K_{n-2}(x_j,x_j) &= \frac{\sqrt{a^2+1}}{a}|R'_{n-1}(x_j)|\frac{\sqrt{a^2+x_j^2}}{\sqrt{a^2+1}} \\ &= \frac{\sqrt{a^2+x_j^2}}{a}|R'_{n-1}(x_j)|\end{aligned}$$

and

$$\widehat{\omega}_j = \frac{a}{\sqrt{a^2+x_j^2}|R'_{n-1}(x_j)|}.$$

Consequently

$$\omega_j = \frac{1}{1-x_j^2}\widehat{\omega}_j = \frac{a}{1-x_j^2}\frac{1}{\sqrt{a^2+x_j^2}|R'_{n-1}(x_j)|}$$

which we see, in comparison with (8) is equal to w_j.

The endpoint cases follow easily. Indeed, we have by symmetry that

$$w_0 = w_n \quad \text{and} \quad \omega_0 = \omega_n.$$

Further, as $d\mu$ is a probability measure

$$\sum_{j=0}^{n} w_j = 1 = \sum_{j=0}^{n} \omega_j.$$

Hence

$$\begin{aligned} 2w_0 + \sum_{j=1}^{n-1} w_j &= 2\omega_0 + \sum_{j=1}^{n-1} \omega_j \\ &= 2\omega_0 + \sum_{j=1}^{n-1} w_j \\ \Longrightarrow w_0 &= \omega_0 \quad (= w_n = \omega_n). \end{aligned}$$

■

References

[1] T. Bloom, L. Bos, N. Levenberg and S. Waldron. On the Convergence of Optimal Measures. *Constructive Approx.* 32, 59 – 179, 2010.

[2] L. Bos, N. Levenberg, and J. Ortega-Cerda. Optimal Polynomial Prediction Measures and Extremal Polynomial Growth. *Constr. Approx.* 54, 431 – 453, 2021.

[3] H. Dette W.J. and Studden. The Theory of Canonical Moments with Applications in Statistics, Probability and Analysis. Wiley Interscience, New York, 1997.

[4] P.G. Hoel and A. Levine. Optimal spacing and weighting in polynomial prediction. *Ann. Math. Statist.* 35, 1553 – 1560, 1964.

[5] S. Karlin and W.J. Studden. Tchebycheff Systems with Applications in Analysis and Statistics. Wiley Interscience, New York, 1966.

[6] J. Kiefer and J. Wolfowitz. The equivalence of two extremum problems. *Canad. J. Math.* 12, 363 – 366, 1960.

[7] G. Szego. Orthogonal Polynomials. AMS Colloquium Publications, Vol. 23, 2003.

[8] F. Wielonsky. personal communication, 2021.