The Condition of Vandermonde-like Matrices Involving Orthogonal Polynomials*

Walter Gautschi

Department of Computer Sciences
Purdue University
West Lafayette, Indiana 47907

To my teacher, Alexander M. Ostrowski, in gratitude on his 90th birthday

Submitted by Richard A. Brualdi

ABSTRACT

The condition number (relative to the Frobenius norm) of the $n \times n$ matrix

$$P_n = [p_{r-1}(x_j)]_{r, j=1}^n$$

is investigated, where $p_r(\cdot) = p_r(\cdot; d \lambda)$ are orthogonal polynomials with respect to some weight distribution $d \lambda$, and $x_j$ are pairwise distinct real numbers. If the nodes $x_j$ are the zeros of $p_n$, the condition number is either expressed, or estimated from below and above, in terms of the Christoffel numbers for $d \lambda$, depending on whether the $p_r$ are normalized or not. For arbitrary real $x_j$ and normalized $p_r$ a lower bound of the condition number is obtained in terms of the Christoffel function evaluated at the nodes. Numerical results are given for minimizing the condition number as a function of the nodes for selected classical distributions $d \lambda$.

1. INTRODUCTION

Let $p_r(t) = p_r(t; d \lambda)$, $r = 0, 1, 2, \ldots$, denote a sequence of orthogonal polynomials relative to some positive measure $d \lambda(t)$ on the real line. If in the Vandermonde matrix the successive powers $1, t, t^2, \ldots$ are replaced by the successive orthogonal polynomials $p_0(t), p_1(t), p_2(t), \ldots$, there results the matrix

$$P_n = \begin{bmatrix}
p_0(x_1) & p_0(x_2) & \cdots & p_0(x_n) 
p_1(x_1) & p_1(x_2) & \cdots & p_1(x_n) 
\vdots & \vdots & \ddots & \vdots 
p_{n-1}(x_1) & p_{n-1}(x_2) & \cdots & p_{n-1}(x_n)
\end{bmatrix}, \quad p_r(t) = p_r(t; d \lambda),$$

(1.1)

*Work supported in part by the National Science Foundation under grant MCS-7927158.

LINEAR ALGEBRA AND ITS APPLICATIONS 52/53:293–300 (1983) 293


52 Vanderbilt Ave., New York, NY 10017

0024-3795/83/$3.00
which is nonsingular for pairwise distinct nodes \(x_1, x_2, \ldots, x_n\). We shall
assume here that all nodes are real. Our interest is in the condition of \(P_n\). We
find it convenient to consider the condition number

\[
\text{cond}_F(P_n) = \|P_n\|_F\|P_n^{-1}\|_F
\]

with respect to the Frobenius norm \(\|A\|_F = [\text{tr}(A^TA)]^{1/2}\), or the closely
related Turing condition number \(\text{cond}_T(P_n) = n^{-1}\text{cond}_F(P_n)\).

In Section 2 we discuss the case of orthonormal polynomials \(\{p_r(\cdot; d\lambda)\}\)
and nodes at the zeros \(\xi^{(n)}_v\) of \(p_n\). Unnormalized polynomials are considered in
Section 3, and arbitrary real nodes in Section 4. In Section 5 we comment on
the problem of minimizing the condition number in (1.2).

2. ORTHONORMAL POLYNOMIALS—NODES AT ZEROS OF \(p_n\)

**Theorem 2.1.** Let \(p_r(\cdot; d\lambda), r = 0, 1, 2, \ldots,\) be the orthonormal polyno-

mials with respect to the (positive) measure \(d\lambda\), and \(x_v = \xi^{(n)}_v, v = 1, 2, \ldots, n,\)
the zeros of \(p_n(\cdot; d\lambda)\). Let furthermore \(\lambda_v = \lambda^{(n)}_v, v = 1, 2, \ldots, n,\) denote the
Christoffel numbers for \(d\lambda\). Then

\[
\text{cond}_F(P_n) = \left( \sum_{v=1}^n \lambda_v \sum_{v=1}^n \frac{1}{\lambda_v} \right)^{1/2}.
\]

**Remark.** If \(m_A(\lambda), m_H(\lambda)\) denote, respectively, the arithmetic and the
harmonic mean of the (positive) numbers \(\lambda_1, \lambda_2, \ldots, \lambda_n\), the result (2.1) may
be restated in terms of the Turing condition number as

\[
\text{cond}_T(P_n) = \left( \frac{m_A(\lambda)}{m_H(\lambda)} \right)^{1/2}.
\]

Letting \(d\lambda\) vary, for any fixed positive integer \(n\), over all positive measures
which admit orthogonal polynomials of degree \(\leq n\), it follows that \(\text{cond}_T(P_n)\),
hence also \(\text{cond}_F(P_n)\), attains its minimum precisely when \(\lambda_1 = \lambda_2 = \cdots = \lambda_n\).
By a classical result [2] this is the case if and only if \(\{p_r(\cdot; d\lambda)\}\) are the
Chebyshev polynomials of the first kind.
VANDERMONDE-LIKE MATRICES

TABLE 1
THE CONDITION OF $P_n$ FOR SOME CLASSICAL ORTHOGONAL POLYNOMIALS

<table>
<thead>
<tr>
<th>n</th>
<th>Legendre</th>
<th>Chebyshev 2nd kind</th>
<th>Laguerre</th>
<th>Hermite</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>5.362(0)</td>
<td>5.916(0)</td>
<td>2.078(2)</td>
<td>1.373(1)</td>
</tr>
<tr>
<td>10</td>
<td>1.155(1)</td>
<td>1.483(1)</td>
<td>1.005(6)</td>
<td>6.832(2)</td>
</tr>
<tr>
<td>20</td>
<td>2.494(1)</td>
<td>3.924(1)</td>
<td>7.770(13)</td>
<td>3.989(6)</td>
</tr>
<tr>
<td>40</td>
<td>5.367(1)</td>
<td>1.071(2)</td>
<td>1.924(30)</td>
<td>3.699(14)</td>
</tr>
<tr>
<td>80</td>
<td>1.148(2)</td>
<td>2.976(2)</td>
<td>6.607(63)</td>
<td>1.095(31)</td>
</tr>
</tbody>
</table>

Proof of Theorem 2.1. Let $P = P_n$ and $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$. From the discrete orthogonality property of orthonormal polynomials,

$$\sum_{r=1}^{n} \lambda_r p_r(\xi_r)p_s(\xi_r) = \begin{cases} 1 & \text{if } r = s \\ 0 & \text{if } r \neq s \end{cases}, \quad r, s = 0, 1, \ldots, n - 1,$$

it follows that $P\Lambda^{1/2} = Q$ is an orthogonal matrix. Therefore,

$$P^TP = \Lambda^{-1/2}Q^TQ\Lambda^{-1/2} = \Lambda^{-1}, \quad (P^{-1})^TP^{-1} = QQ^T,$$

so that

$$\|P\|_F^2 = \text{tr}(P^TP) = \text{tr}(\Lambda^{-1}),$$
$$\|P^{-1}\|_F^2 = \text{tr}(QQ^T) = \text{tr}(\Lambda).$$

The proof reveals that $1/\lambda_r$ are the squares of the singular values $\sigma_r$ of $P$, from which (2.1) follows also on account of

$$\text{cond}_F(P_n) = \left(\sum_{r=1}^{n} \sigma_r^2 \sum_{r=1}^{n} \frac{1}{\sigma_r^2}\right)^{1/2}, \quad \sigma_r = \sigma_r(P_n). \quad (2.2)$$

The numerical behavior of the condition number in (2.1) is illustrated in Table 1 for some classical orthogonal polynomials. (The numbers in parentheses indicate decimal exponents.)

3. UNNORMALIZED POLYNOMIALS

For unnormalized orthogonal polynomials there seems to be no result comparable in simplicity to (2.1). However, we can prove
THEOREM 3.1. Let \( d_r = \int_{\mathbb{R}} p_{r-1}^2(t; d\lambda) d\lambda(t), \ r = 1, 2, \ldots, \) and \( \Delta_n = \max d_r / \min d_r, \) where the maximum and minimum are taken over \( r = 1, 2, \ldots, n. \) Then, in the notation of Theorem 2.1, if \( x_v = \xi_v^{(n)}, \ v = 1, 2, \ldots, n, \)

\[
\frac{1}{\Delta_n} \leq \frac{\text{cond}_F(P_n)}{\left( \sum_{r=1}^n \lambda_r \sum_{r=1}^n \frac{1}{\lambda_r} \right)^{1/2}} \leq \Delta_n.
\]  

(3.1)

Proof. Letting \( P = P_n \) and \( D = \text{diag}(d_1, d_2, \ldots, d_n), \) we now find that \( D^{-1/2}PA^{1/2} = Q \) is orthogonal. Therefore,

\[
\text{tr}(P^T P) = \text{tr}(\Lambda^{-1/2}Q^T D\Lambda^{-1/2}),
\]

\[
\text{tr}((P^{-1})^T P^{-1}) = \text{tr}(D^{-1/2}Q\Lambda Q^T D^{-1/2}).
\]

With

\[
r_v = e_v^T Q^T DQ e_v, \quad s_v = e_v^T Q\Lambda Q^T e_v,
\]

(3.2)

where \( e_v \) is the \( v \)th coordinate vector, we thus have

\[
(\text{cond}_F P)^2 = \sum_{v=1}^n \frac{r_v}{\lambda_v} \sum_{v=1}^n \frac{s_v}{d_v}.
\]

(3.3)

Since \( \|Qe_v\|_2 = \|Q^T e_v\|_2 = 1, \) the quantities \( r_v, s_v \) in (3.2) are Rayleigh quotients of \( D \) and \( \Lambda, \) respectively; hence, in particular,

\[
\min_r d_r \leq r_v \leq \max_r d_r.
\]

(3.4)

Furthermore

\[
\sum_{v=1}^n s_v = \text{tr}(Q\Lambda Q^T) = \text{tr}(\Lambda).
\]

Therefore, (3.1) follows from (3.3) by replacing \( r_v \) and \( d_r \) by the bounds in (3.4).
4. ARBITRARY REAL NODES

We now consider arbitrary real nodes $x_r$, but assume normalized orthonormal polynomials $p_r(\cdot; d\lambda)$. We recall the definition of the Christoffel function (see, e.g. [1]):

$$\lambda_n(x_0) = \min_{p \in \mathcal{P}_{n-1}, p(x_0) = 1} \int_{\mathbb{R}} p^2(t) \, d\lambda(t), \quad x_0 \in \mathbb{R}, \quad (4.1)$$

or, equivalently,

$$[\lambda_n(x)]^{-1} = \sum_{k=0}^{n-1} p_k^2(x), \quad x \in \mathbb{R}. \quad (4.2)$$

**Theorem 4.1.** Let $x_1, x_2, \ldots, x_n$ be pairwise distinct real numbers and $(p_r(\cdot; d\lambda))$ the orthonormal polynomials with respect to the (positive) measure $d\lambda$. Then

$$\text{cond}_F(P_n) \geq \left( \sum_{r=1}^{n} \frac{1}{\lambda_n(x_r)} \sum_{\nu=1}^{n} \frac{1}{\lambda_n(x_\nu)} \right)^{1/2}. \quad (4.3)$$

**Proof.** Let

$$l_r(t) = \prod_{\mu=1, \mu \neq r}^{n} \frac{t - x_\mu}{x_\nu - x_\mu}, \quad \nu = 1, 2, \ldots, n,$$

be the fundamental Lagrange interpolation polynomials for the nodes $x_1, x_2, \ldots, x_n$, and let

$$l_r(t) = \sum_{\mu=1}^{n} a_{\nu\mu} p_{\nu-1}(t).$$

Then, as is easily seen,

$$P_n^{-1} = [a_{\nu\mu}].$$
Consequently,

$$\int_{\mathbb{R}} \sum_{\nu=1}^{n} p_{\nu}^{2}(t) d\lambda(t) = \int_{\mathbb{R}} \sum_{\nu} \sum_{\mu} a_{\nu \mu} p_{\mu-1}(t) p_{\nu \kappa-1}(t) d\lambda(t)$$

$$= \sum_{\nu} \sum_{\mu, \kappa} a_{\nu \mu} a_{\nu \kappa} \int_{\mathbb{R}} p_{\mu-1}(t) p_{\nu \kappa-1}(t) d\lambda(t)$$

$$= \sum_{\nu} \sum_{\mu} a_{\nu \mu}^{2},$$

and therefore

$$\|P_{n}^{-1}\|_{F}^{2} = \int_{\mathbb{R}} \sum_{\nu=1}^{n} p_{\nu}^{2}(t) d\lambda(t). \quad (4.4)$$

Since $l_{\nu} \in P_{n-1}$ and $l_{\nu}(x_{\nu}) = 1$, it follows from (4.1) that

$$\|P_{n}^{-1}\|_{F}^{2} \geq \sum_{\nu=1}^{n} \lambda_{n}(x_{\nu}). \quad (4.5)$$

On the other hand, using (4.2),

$$\|P_{n}\|_{F}^{2} = \sum_{\nu=1}^{n} \sum_{k=0}^{n-1} \hat{p}^{2}_{k}(x_{\nu}) = \sum_{\nu=1}^{n} \frac{1}{\lambda_{n}(x_{\nu})}. \quad (4.6)$$

The assertion (4.3) now follows immediately from (4.5), (4.6).

We remark that (4.3) holds with equality if $x_{\nu} = \xi_{\nu}^{(n)}$, $\nu = 1, 2, \ldots, n$, as follows from Theorem 2.1 and the fact that $\lambda_{n}(\xi_{\nu}^{(n)}) = \lambda_{n}(\nu)$, $\nu = 1, 2, \ldots, n$. We also remark that Theorem 4.1 remains valid, with essentially the same proof, if the nodes are complex and $\lambda_{n}(\cdot)$ is defined as in (4.1), with $p^{2}(t)$ replaced by $|p(t)|^{2}$.

5. MINIMIZING THE CONDITION NUMBER

An interesting problem is to determine the optimally conditioned matrix $P_{n}$ for any fixed measure $d\lambda$, i.e. to find the nodes $x_{1}, x_{2}, \ldots, x_{n}$ which
minimize the condition number $\text{cond}_F(P_n)$ over all pairwise distinct real nodes. We report here on attempts to solve this problem numerically.

Recall from (2.2) that

$$\text{cond}_F(P_n) = n \left[ \frac{m_A(a^2)}{m_H(a^2)} \right]^{1/2},$$

where $m_A(a^2)$, $m_H(a^2)$ are, respectively, the arithmetic and the harmonic mean of the squares of the singular values $a_i$ of $P_n$. It follows that $\text{cond}_F(P_n) \geq n$, so that the smallest possible condition number (attained for the Chebyshev measure and Chebyshev nodes; cf. Remark to Theorem 2.1) is equal to $n$.

Assuming normalized polynomials $p_r(\cdot; d\lambda)$, the condition number $\text{cond}_F(P_n)$, or rather its square, can be written explicitly as the product of the two expressions in (4.4) and (4.6). Both expressions, including their gradients, can be computed fairly easily, the integral in (4.4) and similar integrals involved in the gradient being evaluated (exactly) by the $n$-point Gauss-Christoffel quadrature rule associated with the measure $d\lambda$. Using this computation in conjunction with a minimization algorithm, for which we selected the procedures in [3], we were able to obtain the results shown in Tables 2 and 3. Although only local extrema can be found in this manner, the closeness of the minimum to the absolute minimum $n$ in some of the examples suggests that the results are indeed optimal to within the precision given.

In Table 2 we show the "optimal" nodes and the minimum condition number for Legendre polynomials ($d\lambda(t) = dt$ on $[-1, 1]$). Table 3 displays only the optimal condition number (without nodes) for some of the other
classical polynomials. In the case $n = 20$ of Laguerre polynomials the minimization algorithm could not be made to converge within a reasonable amount of time. Interestingly, some of the nodes in the Laguerre case turn out to be negative.

For $n = 2$ it can be shown by direct computation that the optimal condition always equals $\text{cond}_d(p_2) = 2$, and that the optimal nodes are the zeros $\xi_1, \xi_2$ of $\rho_2(\cdot; d\lambda)$, provided the measure $d\lambda$ is "symmetric" in the sense $\int_\mathbb{R} t d\lambda(t) = \int_\mathbb{R} t^2 d\lambda(t) = 0$. In the Laguerre case, the optimal nodes are $x_1 = 0, x_2 = 2$.

REFERENCES


Received 21 August 1981; revised 5 October 1981