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

$$P_n = \begin{bmatrix} p_0(x_1) & p_0(x_2) & \cdots & p_0(x_n) \\ p_1(x_1) & p_1(x_2) & \cdots & p_1(x_n) \\ \cdots & \cdots & \cdots & \cdots \\ p_{n-1}(x_1) & p_{n-1}(x_2) & \cdots & p_{n-1}(x_n) \end{bmatrix}, \qquad p_r(t) = p_r(t; d\lambda), \tag{1.1}$$

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

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

$$\operatorname{cond}_F(P_n) = \|P_n\|_F \|P_n^{-1}\|_F \tag{1.2}$$

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

In Section 2 we discuss the case of ortho*normal* polynomials $\{p_r(\cdot\,; d\lambda)\}$ and nodes at the zeros $\xi_\nu^{(n)}$ 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 polynomials with respect to the (positive) measure $d\lambda$, and $x_\nu = \xi_\nu^{(n)}$, $\nu = 1, 2, \ldots, n$, the zeros of $p_n(\cdot\,; d\lambda)$. Let furthermore $\lambda_\nu = \lambda_\nu^{(n)}$, $\nu = 1, 2, \ldots, n$, denote the Christoffel numbers for $d\lambda$. Then*

$$\operatorname{cond}_F(P_n) = \left(\sum_{\nu=1}^{n} \lambda_\nu \sum_{\nu=1}^{n} \frac{1}{\lambda_\nu} \right)^{1/2}. \tag{2.1}$$

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

$$\operatorname{cond}_T(P_n) = \left(\frac{m_A(\lambda)}{m_H(\lambda)} \right)^{1/2}. \tag{2.1$'$}$$

Letting $d\lambda$ vary, for any fixed positive integer n, over all positive measures which admit orthogonal polynomials of degree $\leqslant n$, it follows that $\operatorname{cond}_T(P_n)$, hence also $\operatorname{cond}_F(P_n)$, 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.

TABLE 1

THE CONDITION OF P_n FOR SOME CLASSICAL ORTHOGONAL POLYNOMIALS

n	Legendre	Chebyshev 2nd kind	Laguerre	Hermite
5	5.362(0)	5.916(0)	2.076(2)	1.373(1)
10	1.155(1)	1.483(1)	1.005(6)	6.832(2)
20	2.494(1)	3.924(1)	7.770(13)	3.989(6)
40	5.367(1)	1.071(2)	1.924(30)	3.699(14)
80	1.148(2)	2.976(2)	6.607(63)	1.095(31)

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

$$\sum_{\nu=1}^{n} \lambda_\nu p_r(\xi_\nu) p_s(\xi_\nu) = \begin{Bmatrix} 1 & \text{if} & r = s \\ 0 & \text{if} & r \neq s \end{Bmatrix}, \qquad r, s = 0, 1, \dots, n-1,$$

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

$$P^T P = \Lambda^{-1/2} Q^T Q \Lambda^{-1/2} = \Lambda^{-1}, \qquad (P^{-1})^T P^{-1} = Q \Lambda Q^T,$$

so that

$$\|P\|_F^2 = \mathrm{tr}(P^T P) = \mathrm{tr}(\Lambda^{-1}),$$
$$\|P^{-1}\|_F^2 = \mathrm{tr}(Q \Lambda Q^T) = \mathrm{tr}(\Lambda). \qquad \blacksquare$$

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

$$\mathrm{cond}_F(P_n) = \left(\sum_{\nu=1}^{n} \sigma_\nu^2 \sum_{\nu=1}^{n} \frac{1}{\sigma_\nu^2} \right)^{1/2}, \qquad \sigma_\nu = \sigma_\nu(P_n). \tag{2.2}$$

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* $x_\nu = \xi_\nu^{(n)}$, $\nu = 1,2,\ldots,n$,

$$\frac{1}{\Delta_n} \leqslant \frac{\operatorname{cond}_F(P_n)}{\left(\sum_{\nu=1}^{n} \lambda_\nu \sum_{\nu=1}^{n} \frac{1}{\lambda_\nu}\right)^{1/2}} \leqslant \Delta_n. \tag{3.1}$$

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

$$\operatorname{tr}(P^T P) = \operatorname{tr}(\Lambda^{-1/2} Q^T D Q \Lambda^{-1/2}),$$

$$\operatorname{tr}\big((P^{-1})^T P^{-1}\big) = \operatorname{tr}(D^{-1/2} Q \Lambda Q^T D^{-1/2}).$$

With

$$r_\nu = e_\nu^T Q^T D Q e_\nu, \qquad s_\nu = e_\nu^T Q \Lambda Q^T e_\nu, \tag{3.2}$$

where e_ν is the νth coordinate vector, we thus have

$$(\operatorname{cond}_F P)^2 = \sum_{\nu=1}^{n} \frac{r_\nu}{\lambda_\nu} \sum_{\nu=1}^{n} \frac{s_\nu}{d_\nu}. \tag{3.3}$$

Since $\|Qe_\nu\|_2 = \|Q^T e_\nu\|_2 = 1$, the quantities r_ν, s_ν in (3.2) are Rayleigh quotients of D and Λ, respectively; hence, in particular,

$$\min_r d_r \leqslant r_\nu \leqslant \max_r d_r. \tag{3.4}$$

Furthermore

$$\sum_{\nu=1}^{n} s_\nu = \operatorname{tr}(Q \Lambda Q^T) = \operatorname{tr}(\Lambda).$$

Therefore, (3.1) follows from (3.3) by replacing r_ν and d_ν by the bounds in (3.4). ■

4. ARBITRARY REAL NODES

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

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

or, equivalently,

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

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

$$\operatorname{cond}_F(P_n) \geqslant \left(\sum_{\nu=1}^{n} \lambda_n(x_\nu) \sum_{\nu=1}^{n} \frac{1}{\lambda_n(x_\nu)} \right)^{1/2}. \tag{4.3}$$

Proof. Let

$$l_\nu(t) = \prod_{\substack{\mu=1 \\ \mu \neq \nu}}^{n} \frac{t - x_\mu}{x_\nu - x_\mu}, \qquad \nu = 1, 2, \dots, n,$$

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

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

Then, as is easily seen,

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

Consequently,

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

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

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

and therefore

$$\|P_n^{-1}\|_F^2 = \int_{\mathbb{R}} \sum_{\nu=1}^{n} l_\nu^2(t)\,d\lambda(t). \tag{4.4}$$

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

$$\|P_n^{-1}\|_F^2 \geqslant \sum_{\nu=1}^{n} \lambda_n(x_\nu). \tag{4.5}$$

On the other hand, using (4.2),

$$\|P_n\|_F^2 = \sum_{\nu=1}^{n} \sum_{k=0}^{n-1} p_k^2(x_\nu) = \sum_{\nu=1}^{n} \frac{1}{\lambda_n(x_\nu)}. \tag{4.6}$$

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

We remark that (4.3) holds with equality if $x_\nu = \xi_\nu^{(n)}$, $\nu = 1,2,\dots,n$, as follows from Theorem 2.1 and the fact that $\lambda_n(\xi_\nu^{(n)}) = \lambda_\nu^{(n)}$, $\nu = 1,2,\dots,n$. 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 $x_1, x_2, \dots, x_n$ which minimize the condition number $\operatorname{cond}_F(P_n)$ over all pairwise distinct real nodes. We report here on attempts to solve this problem numerically.

Recall from (2.2) that

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

where $m_A(\sigma^2)$, $m_H(\sigma^2)$ are, respectively, the arithmetic and the harmonic mean of the squares of the singular values σ_ν of P_n. It follows that $\operatorname{cond}_F(P_n) \geqslant n$, 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 $\operatorname{cond}_F(P_n)$, 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

TABLE 2

OPTIMALLY CONDITIONED MATRIX P_n FOR LEGENDRE POLYNOMIALS

n	x_ν	$\operatorname{cond}_F(P_n)$	n	x_ν	$\operatorname{cond}_F(P_n)$
2	$\pm$.5773502692	2.0	20	$\pm$.9885188046	23.46822182
5	$\pm$.8780893894	5.229550605		$\pm$.9585058326	
	$\pm$.5336883454			$\pm$.9083037137	
	.0			$\pm$.8372548421	
10	$\pm$.9610897501	11.01832471		$\pm$.7462848447	
	$\pm$.8560330091			$\pm$.6372598211	
	$\pm$.6772857139			$\pm$.5126743680	
	$\pm$.4346101969			$\pm$.3755003472	
	$\pm$.1497603704			$\pm$.2290741509	
				$\pm$.0769922707	

TABLE 3

OPTIMAL CONDITION OF P_n FOR SOME CLASSICAL ORTHOGONAL POLYNOMIALS

n	Chebyshev 2nd kind	Laguerre	Hermite
2	2.0	2.0	2.0
5	5.544624008	2.106683223(1)	8.393706126
10	1.295377448(1)	1.340933671(3)	8.275431133(1)
20	3.240814863(1)	< 6.4040073(6)	7.476911820(3)

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 $\operatorname{cond}_F(P_2)=2$, and that the optimal nodes are the zeros ξ_1, ξ_2 of $p_2(\cdot\,; d\lambda)$, provided the measure $d\lambda$ is "symmetric" in the sense $\int_{\mathbf{R}} t\, d\lambda(t) = \int_{\mathbf{R}} t^3\, d\lambda(t) = 0$. In the Laguerre case, the optimal nodes are $x_1 = 0, x_2 = 2$.

REFERENCES

1 P. G. Nevai, Orthogonal polynomials, *Mem. Amer. Math. Soc.* 18, No. 213, Amer. Math. Soc., Providence, R.I., 1979.

2 C. Posse, Sur les quadratures, *Nouv. Ann. Math. (2)* 14:49–62 (1875).

3 D. F. Shanno and K. H. Phua, Remark on Algorithm 500, *ACM Trans. Math. Software* 6: 618–622 (1980).

Received 21 August 1981; revised 5 October 1981