SIAM J. MATRIX ANAL. APPL.
Vol. 40, No. 4, pp. 1399–1424

OPTIMALLY CONDITIONED VANDERMONDE-LIKE MATRICES*

MYKHAILO KUIAN†, LOTHAR REICHEL†, AND SERGIJ SHIYANOVSKII‡

Dedicated to Walter Gautschi on the occasion of his 90th birthday

Abstract. Vandermonde matrices arise frequently in computational mathematics in problems that require polynomial approximation, differentiation, or integration. These matrices are defined by a set of n distinct nodes $x_1, x_2, \ldots, x_n$ and a monomial basis. A difficulty with Vandermonde matrices is that they typically are quite ill-conditioned when the nodes are real and n is not very small. The ill-conditioning often can be reduced significantly by using a basis of orthonormal polynomials $p_0, p_1, \ldots, p_{n-1}$, with $\deg(p_j) = j$. This was first observed by Gautschi. The matrices so obtained are commonly referred to as Vandermonde-like and are of the form $V_{n,n} = [p_{i-1}(x_j)]_{i,j=1}^n \in \mathbb{R}^{n\times n}$. Gautschi analyzed optimally conditioned and optimally scaled square Vandermonde and Vandermonde-like matrices with real nodes. We extend Gautschi's analysis to rectangular Vandermonde-like matrices with real nodes, as well as to Vandermonde-like matrices with nodes on the unit circle in the complex plane. Additionally, we investigate existence and uniqueness of optimally conditioned Vandermonde-like matrices. Finally, we discuss properties of rectangular Vandermonde and Vandermonde-like matrices $V_{N,n}$ of order $N \times n$, $N \neq n$, with Chebyshev nodes or with equidistant nodes on the unit circle in the complex plane, and show that the condition number of these matrices can be bounded independently of the number of nodes.

Key words. Vandermonde matrix, conditioning, orthogonal polynomials

AMS subject classifications. 65F35, 15A12

DOI. 10.1137/19M1237272

1. Introduction.

Let $x_1, x_2, \ldots, x_N$ be distinct nodes in the complex plane $\mathbb{C}$, and let $p_0, p_1, p_2, \ldots$ be a polynomial family with $\deg(p_j) = j$. Matrices of the form

$$V_{N,n} := \begin{bmatrix} p_0(x_1) & p_1(x_1) & \cdots & p_{n-1}(x_1) \\ p_0(x_2) & p_1(x_2) & \cdots & p_{n-1}(x_2) \\ \vdots & \vdots & \cdots & \vdots \\ p_0(x_N) & p_1(x_N) & \cdots & p_{n-1}(x_N) \end{bmatrix} \in \mathbb{C}^{N\times n} \tag{1.1}$$

are known as Vandermonde-like matrices. When the polynomials p_j are monomials, i.e., $p_j(x) = x^j$, $j = 0, 1, \ldots, n-1$, the Vandermonde-like matrix (1.1), which is rectangular when $N \neq n$, simplifies to a (standard) Vandermonde matrix. Vandermonde-like matrices arise in polynomial interpolation and least-squares approximation, when approximating the derivative of a function known at the nodes $x_1, x_2, \ldots, x_N$ by differentiating the interpolating polynomial or a polynomial least-squares approximant, and when computing the weights of an interpolating quadrature rule with nodes $x_1, x_2, \ldots, x_N$; see, e.g., Gautschi [13, 16] for discussions on applications.

The condition number of a square matrix furnishes a bound for the relative error in the solution of a linear system of equations with the matrix caused by an error in the data vector (right-hand side). A small condition number indicates a small relative

*Received by the editors January 7, 2019; accepted for publication (in revised form) by K. D. Usevich September 16, 2019; published electronically November 21, 2019.
https://doi.org/10.1137/19M1237272

Funding: The work of the authors was supported by the National Science Foundation grants DMS-1729509 and DMS-1720259.

†Department of Mathematical Sciences, Kent State University, Kent, OH, 44242 (mkuian@kent.edu, reichel@math.kent.edu).
‡Liquid Crystal Institute, Kent State University, Kent, OH, 44242 (sshiyano@kent.edu).

error in the data vector only causes a small relative error in the solution. Conversely, a large condition number signals that the computed solution may be very sensitive to an error in the data vector. The condition number also yields bounds for the relative error in the solution of least-squares problems caused by an error in the data vector; see, e.g., [13, 23, 37] for discussions.

We are interested in investigating the conditioning of rectangular Vandermonde and Vandermonde-like matrices. The conditioning is measured by condition numbers defined as

$$\kappa_2(V_{N,n}) := \|V_{N,n}\|_2 \|V_{N,n}^\dagger\|_2, \qquad \kappa_F(V_{N,n}) := \|V_{N,n}\|_F \|V_{N,n}^\dagger\|_F, \tag{1.2}$$

where $V_{N,n}^\dagger$ denotes the Moore–Penrose pseudo-inverse of $V_{N,n}$, $\|\cdot\|_2$ stands for the spectral norm, and $\|\cdot\|_F$ for the Frobenius norm, i.e.,

$$\|V_{N,n}\|_F := \sqrt{\operatorname{trace}(V_{N,n}^* V_{N,n})}. \tag{1.3}$$

The superscript * denotes transposition and complex conjugation when applicable. Bounds for the condition number of (standard) rectangular Vandermonde matrices $V_{N,n}$ with Chebyshev nodes x_j and $1 \le n < N$ have been derived by Li [28], who exploits the structure of the QR factorization of $V_{N,n}$ to bound $\kappa_F(V_{N,n})$.

We will use the singular value decomposition (SVD)

$$V_{N,n} = U \Sigma W^* \tag{1.4}$$

in our analysis, where the $U \in \mathbb{C}^{N\times N}$ and $W \in \mathbb{C}^{n\times n}$ are unitary matrices. The matrix

$$\Sigma = \operatorname{diag}\left[\sigma_1, \sigma_2, \ldots, \sigma_{\min\{N,n\}}\right] \in \mathbb{R}^{N\times n}$$

is diagonal and rectangular when $N \ne n$. Its nontrivial entries are known as singular values. They are nonnegative and ordered according to $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_{\min\{N,n\}} \ge 0$; see, e.g., [37] for details on the SVD. It easily can be shown that when the nodes x_i are distinct, all singular values are positive; see below. We assume this to be the case. Then

$$\|V_{N,n}\|_2 = \sigma_1, \qquad \|V_{N,n}^\dagger\|_2 = \sigma_{\min\{N,n\}}^{-1},$$

and (1.2) yields

$$\kappa_2(V_{N,n}) = \frac{\sigma_1}{\sigma_{\min\{N,n\}}} \ge 1 \tag{1.5}$$

with equality if and only if all singular values are equal. Similarly, substituting (1.4) into (1.3) yields

$$\|V_{N,n}\|_F = \sqrt{\sum_{j=1}^{\min\{N,n\}} \sigma_j^2}, \qquad \|V_{N,n}^\dagger\|_F = \sqrt{\sum_{j=1}^{\min\{N,n\}} \sigma_j^{-2}}.$$

It follows that

$$\kappa_F(V_{N,n}) = \sqrt{\sum_{j=1}^{\min\{N,n\}} \sigma_j^2 \cdot \sum_{j=1}^{\min\{N,n\}} \sigma_j^{-2}} \ge \min\{N,n\}, \tag{1.6}$$

where the lower bound is a consequence of the Cauchy–Schwarz inequality. The bound is attained if and only if all singular values are equal. We say that a matrix is *optimally conditioned* if the lower bounds (1.5) and (1.6) for the condition numbers are achieved. Note that these bounds are achieved simultaneously.

Bazán [3] observed that rectangular Vandermonde matrices with $n \gg N$ can be fairly well-conditioned when the nodes x_i are close to the unit circle in the complex plane and pairwise not too close. However, the situation when the nodes are real is quite different. Gautschi [11] has shown that the inverse of a square Vandermonde matrix is of large norm when the nodes are real. This results in a large condition number. Further investigations by Gautschi [10] and Gautschi and Inglese [15] provide bounds for condition numbers; the latter work shows that the condition number of square Vandermonde matrices with real nodes grows exponentially with the number of nodes. The conditioning of square Vandermonde matrices also is investigated by Beckermann [4], Eisinberg et al. [9], Gautschi [14], Li [27], and Tyrtyshnikov [38].

To circumvent the ill-conditioning of square Vandermonde matrices with real nodes, Gautschi [12] introduced square Vandermonde-like matrices in which the power basis is replaced by a basis of polynomials that are orthonormal with respect to an inner product defined by a non-negative measure with support on the real axis.

It is the purpose of the present paper to generalize results by Gautschi for square Vandermonde and Vandermonde-like matrices with real nodes to rectangular Vandermonde and Vandermonde-like matrices. Moreover, we will discuss square and rectangular Vandermonde and Vandermonde-like matrices with nodes on the unit circle in the complex plane.

This paper is organized as follows. Section 2 reviews results by Gautschi on the conditioning of square Vandermonde and Vandermonde-like matrices, and section 3 extends Gautschi's analysis to rectangular Vandermonde and Vandermonde-like matrices. Using a result by Posse [31], Gautschi [12, 14] showed that square Chebyshev–Vandermonde matrices $V_{n,n} = [T_{i-1}(x_j)]_{i,j=1}^{n}$, where the T_{i-1} are Chebyshev polynomials of the first kind for the interval $[-1, 1]$ and the nodes $x_1, x_2, \ldots, x_n$ are the zeros of T_n, are the only Vandermonde-like matrices that are optimally conditioned for all $n \geq 1$ with respect to the Frobenius and spectral norms. We show this to be the case also for rectangular Vandermonde-like matrices $V_{N,n}$, where $N \neq n$. Section 4 considers square and rectangular Vandermonde and Vandermonde-like matrices defined by polynomials that are orthogonal with respect to an inner product defined by a non-negative measure on the unit circle in the complex plane $\mathbb{C}$ and by nodes that are the abscissas of a Gauss–Szegő quadrature rule associated with this inner product. Section 5 discusses the conditioning of Vandermonde-like matrices determined by a general polynomial basis and Chebyshev nodes and extends results shown by Eisinberg et al. [9]. Specifically, we show that the Frobenius and spectral condition numbers of Vandermonde and Vandermonde-like matrices with Chebyshev nodes can be bounded independently of the number of nodes. An analogous result for Vandermonde matrices with equidistant nodes on the unit circle in $\mathbb{C}$ also is shown. Finally, section 6 contains concluding remarks.

2. Square Vandermonde-like matrices.

This section reviews results shown by Gautschi [12] for square Vandermonde-like matrices. Let $p_0, p_1, p_2, \ldots$ be a family of polynomials, with $\deg(p_j) = j$, that are orthogonal with respect to an inner product determined by a real nonnegative measure $d\lambda$ with support on the real axis at infinitely many points,

$$(f, g)_\lambda := \int f(x)g(x)d\lambda(x). \tag{2.1}$$

Further, assume that the polynomials are normalized to be of unit length with respect to the norm associated with this inner product. Thus,

$$(p_j, p_k)_\lambda = \begin{cases} 0, & j \neq k, \\ 1, & j = k. \end{cases} \tag{2.2}$$

Introduce the N-node Gauss quadrature rule associated with the measure $d\lambda$,

$$\mathcal{G}_N f = \sum_{k=1}^{N} \lambda_k^{(N)} f\left(x_k^{(N)}\right). \tag{2.3}$$

The nodes $x_1^{(N)}, x_2^{(N)}, \ldots, x_N^{(N)}$ are known to be distinct and in the convex hull of the support of $d\lambda$, and the weights $\lambda_1^{(N)}, \lambda_2^{(N)}, \ldots, \lambda_N^{(N)}$ are positive. This quadrature rule can be applied to approximate the integral

$$\mathcal{I} f = \int f(x) d\lambda(x).$$

It is characterized by the property

$$\mathcal{I} f = \mathcal{G}_N f \qquad \forall f \in \mathbb{P}_{2N-1},$$

where $\mathbb{P}_{2N-1}$ denotes the set of polynomials of degree at most $2N-1$; see, e.g., Gautschi [16] and Szegő [36] for discussions on Gauss quadrature.

The Christoffel function associated with the Gauss rule (2.3) can be expressed as

$$\Lambda_N(x) = \left(\sum_{j=0}^{N-1} p_j^2(x) \right)^{-1}; \tag{2.4}$$

see, e.g., Szegő [36, Chapter 2]. Evaluation of this function at the Gaussian nodes yields the Gaussian weights,

$$\lambda_k^{(N)} = \Lambda_N\left(x_k^{(N)}\right), \qquad 1 \leq k \leq N, \tag{2.5}$$

which also are known as Christoffel numbers. Gautschi [12, Theorem 2.1] showed the following result for square Vandermonde-like matrices.

Proposition 1. *Let the Vandermonde-like matrix defined by* (1.1) *with $n = N$ be determined by the orthonormal polynomials that satisfy* (2.2) *and by the Gaussian nodes $x_k := x_k^{(N)}$, $1 \leq k \leq N$. Then*

$$\kappa_F(V_{N,N}) = \left(\sum_{k=1}^{N} \lambda_k^{(N)} \cdot \sum_{k=1}^{N} \left(\lambda_k^{(N)}\right)^{-1} \right)^{1/2}, \tag{2.6}$$

where the $\lambda_k^{(N)}$ are the Gaussian weights (2.5).

The Cauchy–Schwarz inequality applied to the right-hand side of (2.6) shows that

$$\kappa_F(V_{N,N}) \geq N$$

with equality if and only if all Christoffel numbers $\lambda_k^{(N)}$, $k = 1, 2, \ldots, N$, are equal. Gautschi [12] showed that the Chebyshev measure

$$d\lambda(x) = \left(1 - x^2\right)^{-1/2} dx, \qquad -1 < x < 1, \tag{2.7}$$

determines quadrature nodes that give optimally conditioned Vandermonde-like matrices $V_{N,N}$ for all $N \geq 1$. To prove this, Gautschi [12] applied the following result due to Posse [31].

THEOREM 1. *Let $d\lambda$ be a nonnegative measure on the interval $[-1,1]$ with infinitely many points of support, and let $p_0, p_1, p_2, \ldots$ be a sequence of normalized orthogonal polynomials, i.e., they satisfy $\deg(p_j) = j$, for all $j = 0,1,2,\ldots$, and* (2.2)*. Denote the zeros of p_N by $x_1^{(N)}, x_2^{(N)}, \ldots, x_N^{(N)}$. If for all $N = 1,2,3,\ldots$, it holds that*

$$\int_{-1}^{1} p(x) d\lambda(x) = \nu_N \sum_{k=1}^{N} p\left(x_k^{(N)}\right)$$

for some scalar ν_N and all polynomials $p \in \mathbb{P}_{2N-1}$, then $d\lambda$ is the Chebyshev measure (2.7) *or a scaled version thereof.*

We recall that the normalized orthogonal polynomials with respect to this measure are given by

$$p_j(x) = \begin{cases} \sqrt{\frac{1}{\pi}} T_0(x), & j = 0, \\ \sqrt{\frac{2}{\pi}} T_j(x), & j = 1,2,\ldots, \end{cases} \tag{2.8}$$

where the $T_j(x)$ are Chebyshev polynomials of the first kind. They can be defined as

$$T_j(x) = \cos(j \arccos(x)), \qquad j = 0,1,2,\ldots, \tag{2.9}$$

for $-1 \leq x \leq 1$. The Gaussian nodes and weights associated with the measure (2.7) are given by

$$x_k^{(N)} = \cos\left(\frac{2k-1}{2N}\pi\right), \qquad \lambda_k^{(N)} = \frac{\pi}{N}, \qquad 1 \leq k \leq N. \tag{2.10}$$

Obviously, the result of Theorem 1 also holds for an arbitrary compact interval $[a,b]$ provided that the measure $d\lambda$ and the nodes (2.10) are transformed correspondingly.

Square Vandermonde-like matrices with distinct nodes (1.1) are known to be nonsingular; see, e.g., [35, section 3.6]. Let the Vandermonde-like matrix $V_{N,N}$ have the singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_N > 0$. Gautschi's proof [12, Theorem 2.1] of Proposition 1 shows that the singular values are square roots of the reciprocal Christoffel numbers, up to a renumbering.

COROLLARY 1. *All singular values of the Vandermonde-like matrix $V_{N,N}$ are equal if the polynomials p_j are defined by* (2.8) *and the nodes are the Chebyshev points* (2.10)*. Then $V_{N,N}$ is optimally conditioned, i.e., $\kappa_F(V_{N,N}) = N$ and $\kappa_2(V_{N,N}) = 1$.*

Let $p_0, p_1, p_2, \ldots$ be a family of normalized orthogonal polynomials with respect to an inner product (2.1) that is defined by a nonnegative measure $d\lambda$ on the real axis with infinitely many points of support. It is interesting to note that any square Vandermonde-like matrix that is defined by such a family of polynomials $p_0, p_1, p_2, \ldots$ can be made optimally conditioned by an appropriate choice of nodes and row scaling. This can be shown as follows. Let $c \in \mathbb{R}$ be a constant and let the nodes $x_1, x_2, \ldots, x_N$ be the zeros of the polynomial $p_N(x) - c\, p_{N-1}(x)$. It follows from [36, Theorem 3.3.4] that the zeros of $p_N(x) - c\, p_{N-1}(x)$ are distinct and real. The Christoffel–Darboux formula for $i \neq j$ yields

$$\sum_{k=0}^{N-1} p_k(x_i) p_k(x_j) = \frac{\mu_{N-1}}{\mu_N} \frac{p_N(x_i) p_{N-1}(x_j) - p_N(x_j) p_{N-1}(x_i)}{x_i - x_j}, \tag{2.11}$$

where μ_k is the leading coefficient of $p_k(x)$. Since $p_N(x_i) = c\,p_{N-1}(x_i)$, we have

$$p_N(x_i)p_{N-1}(x_j) = c\,p_{N-1}(x_i)p_{N-1}(x_j), \quad p_N(x_j)p_{N-1}(x_i) = c\,p_{N-1}(x_j)p_{N-1}(x_i).$$

Substitution into (2.11) yields

$$\sum_{k=0}^{N-1} p_k(x_i)p_k(x_j) = 0, \qquad i \neq j.$$

Hence, the rows of the matrix $V_{N,N}$ are orthogonal. Normalizing the rows of $V_{N,N}$ makes the matrix orthogonal and, therefore, optimally conditioned.

3. Rectangular Vandermonde-like matrices. We extend the results of the previous section to rectangular Vandermonde-like matrices. The first lemmas are concerned with some basic properties that will be used in the sequel.

Lemma 1. *Let $p_0, p_1, \ldots, p_{n-1}$ be polynomials such that $\deg(p_j) = j$, and assume that the points $x_1, x_2, \ldots, x_N$ are distinct in the complex plane. Then the rectangular Vandermonde-like matrix* (1.1) *is of full rank.*

Proof. Let $m = \min\{n, N\}$. The leading $m \times m$ principal submatrix of the matrix (1.1) is nonsingular by [35, Theorem 3.6.11]. □

The following result will be used in the sequel.

Lemma 2. *An optimally conditioned matrix $A \in \mathbb{C}^{N\times n}$ can be written as $A = \sigma Q$, where σ is positive constant and the matrix Q has orthonormal columns if $N \geq n$ and orthonormal rows if $N \leq n$.*

Proof. The lemma follows from the SVD of $A = U\Sigma W^*$. Here $U \in \mathbb{C}^{N\times N}$ and $W \in \mathbb{C}^{n\times n}$ are unitary matrices. Assume that $N \geq n$. Then $\Sigma = \mathrm{diag}[\sigma_1, \sigma_2, \ldots, \sigma_n] \in \mathbb{R}^{N\times n}$ with $\sigma_1 = \cdots = \sigma_n$. Let the matrix $\hat{U} \in \mathbb{C}^{N\times n}$ be made up of the n first columns of U. Then $A = \sigma_1 \hat{U} W^*$, where the matrix $\hat{U}W^*$ has orthonormal columns. The proof for $N < n$ proceeds similarly. □

Lemma 3. *Let the polynomials $p_0, p_1, \ldots, p_{N-1}$ be Chebyshev polynomials* (2.8), *and let the nodes x_j be the zeros* (2.10) *of p_N. Then, for $1 \leq n \leq N$,*

$$\|V_{N,n}\|_F = \sqrt{\frac{Nn}{\pi}}.$$

Proof. The polynomials (2.8) satisfy for $0 \leq i, j < N$,

$$\sum_{k=1}^{N} p_j(x_k)p_i(x_k) = \begin{cases} 0, & i \neq j, \\ \frac{N}{\pi}, & i = j. \end{cases} \tag{3.1}$$

Therefore,

$$\|V_{n,N}\|_F^2 = \mathrm{trace}(V_{n,N}^* V_{N,n}) = \frac{Nn}{\pi}.$$ □

Let $V_{N,N} \in \mathbb{R}^{N\times n}$ be optimally conditioned. The following theorem shows that rectangular submatrices obtained by removing the last few rows or the last few columns of $V_{N,N}$ also are optimally conditioned.

Theorem 2. *Rectangular Vandermonde-like matrices $V_{N,n}$ of normalized Chebyshev polynomials* (2.8) *with the nodes* (2.10) *are optimally conditioned for all $1 \leq n \leq N$. Also rectangular Vandermonde-like matrices $V_{m,N}$ of normalized Chebyshev polynomials* (2.8) *with the nodes* (2.10) *are optimally conditioned for all $1 \leq m \leq N$.*

Proof. The polynomials $p_0, p_1, \ldots, p_{N-1}$ defined by (2.8) are orthogonal with respect to a discrete inner product (3.1). The matrix $V_{N,N}$ satisfies $V_{N,N}^* V_{N,N} = \frac{N}{\pi} I_N$, where I_N denotes the identity matrix of order N. It follows that all singular values are $\sqrt{N/\pi}$.

We first consider the situation when $V_{N,n}$ has more rows than columns, i.e., $1 \leq n < N$. The matrix $V_{N,n}$ consist of the first n columns of $V_{N,N}$. Therefore, $V_{N,n}^* V_{N,n} = \frac{N}{\pi} I_n$. Hence, all singular values of $V_{N,n}$ are $\sqrt{N/\pi}$, and it follows that the matrix $V_{N,n}$ is optimally conditioned.

Now consider the matrices $V_{m,N}$ with more columns than rows, i.e., $1 \leq m \leq N$. It follows from $V_{N,N} V_{N,N}^* = \frac{N}{\pi} I_N$ that $V_{m,N} V_{m,N}^* = \frac{N}{\pi} I_m$. Hence, the matrix $V_{m,N}$ is optimally conditioned. $\square$

We note that Theorem 2 remains valid also when arbitrary rows or columns are deleted from the matrix $V_{N,N}$.

The Christoffel numbers for the Gauss rule (2.3) associated with the Chebyshev measure (2.7) easily can be determined by using the fact that the matrix $V_{N.N}$ is optimally conditioned. We have $V_{N,N} V_{N,N}^* = \frac{N}{\pi} I$, and it follows from (2.4) and (2.5) that the diagonal entries of this matrix are the reciprocal values of the Christoffel numbers.

The matrix $V_{N,N}$ is closely related to the discrete cosine transform DCT-III matrix, which is important in numerous applications in science and engineering; see [32]. The latter matrix is obtained from $V_{N,N}$ by scaling by the factor $\sqrt{\frac{\pi}{N}}$, using the relation (2.9), and the fact that the nodes $x_k^{(N)}$ are given by (2.10). This yields the orthonormal cosine transform matrix

$$\begin{bmatrix} \sqrt{\frac{1}{N}} & \sqrt{\frac{2}{N}} \cos \frac{\pi}{2N} & \sqrt{\frac{2}{N}} \cos \frac{2\pi}{2N} & \cdots & \sqrt{\frac{2}{N}} \cos \frac{(N-1)\pi}{2N} \\ \sqrt{\frac{1}{N}} & \sqrt{\frac{2}{N}} \cos \frac{3\pi}{2N} & \sqrt{\frac{2}{N}} \cos \frac{2\cdot 3\pi}{2N} & \cdots & \sqrt{\frac{2}{N}} \cos \frac{(N-1)3\pi}{2N} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sqrt{\frac{1}{N}} & \sqrt{\frac{2}{N}} \cos \frac{(2N-1)\pi}{2N} & \sqrt{\frac{2}{N}} \cos \frac{2(2N-1)\pi}{2N} & \cdots & \sqrt{\frac{2}{N}} \cos \frac{(N-1)(2N-1)\pi}{2N} \end{bmatrix}.$$

Since the above matrix is orthogonal, we obtain optimally conditioned submatrices by removing either selected rows or columns.

We would like to explore whether the measure that generates optimally conditioned Vandermonde matrices of Theorem 2 is unique. Towards this end, we first show the following result, which gives a different characterization of the measure (2.7) than Theorem 1. The result is required to show Theorem 4 below.

THEOREM 3. *Let $d\lambda$ be a nonnegative measure on $[-1, 1]$ with infinitely many points of support. Assume that the moments $\mu_j = \int_{-1}^{1} x^j d\lambda(x)$, $j = 0, 1, 2$, satisfy*

$$\mu_0 = 1, \qquad \mu_2 > \mu_1^2. \tag{3.2}$$

Let $p_0(x), p_1(x), p_2(x), \ldots$ be a sequence of associated monic orthogonal polynomials. Denote the zeros of $p_N(x)$ by $x_1^{(N)}, x_2^{(N)}, \ldots, x_N^{(N)}$. If for all $N = 1, 2, 3, \ldots$, it holds that

$$\int_{-1}^{1} p(x) d\lambda(x) = \nu_N \sum_{j=1}^{N} p\left(x_j^{(N)}\right) \tag{3.3}$$

for some scalar ν_N and all polynomials $p \in \mathbb{P}_2$, then $d\lambda$ is a scaled Chebyshev measure (2.7).

Proof. It suffices to show that if for any N, there is a constant ν_N such that

$$\int_{-1}^{1} x^m d\lambda(x) = \nu_N \sum_{k=1}^{N} \left(x_k^{(N)}\right)^m \quad \text{for} \quad m = 0, 1, 2, \tag{3.4}$$

then $d\lambda$ is the measure (2.7) or a scaling thereof. Consider (3.4) for increasing values of m. For $m = 0$, we have

$$\int_{-1}^{1} d\lambda(x) = N\nu_N.$$

We may assume that $\int_{-1}^{1} d\lambda(x) = 1$. Then $\nu_N = 1/N$. Turning to $m = 1$, we get

$$\mu_1 := \int_{-1}^{1} x\, d\lambda(x) = \frac{1}{N} \sum_{k=1}^{N} x_k^{(N)}. \tag{3.5}$$

Express the monic orthogonal polynomial $p_N \in \mathbb{P}_N$ in the form

$$p_N(x) = x^N + a_{N,N-1}x^{N-1} + a_{N,N-2}x^{N-2} + \cdots + a_{N,0}.$$

The relations between the zeros and coefficients of p_N (also known as Vieta's formulas) yield

$$\sum_{k=1}^{N} x_k^{(N)} = -a_{N,N-1}, \tag{3.6}$$

$$\sum_{1 \le k < l \le N} x_k^{(N)} x_l^{(N)} = a_{N,N-2}. \tag{3.7}$$

Any sequence of monic orthogonal polynomials associated with a nonnegative measure with support on a real interval satisfies a recurrence relation of the form

$$p_k(x) = (x - \alpha_k)p_{k-1}(x) - \beta_{k-1}p_{k-2}(x), \qquad k = 2, 3, \ldots; \tag{3.8}$$

see, for example, [16, Theorem 1.27]. Here we assume that the measure has infinitely many points of support.

Comparing coefficients of x^{k-1} in the right-hand side and left-hand side of (3.8) for $k = N, N-1, \ldots, 2$, we obtain the relations

$$\begin{aligned} a_{N,N-1} &= -\alpha_N + a_{N-1,N-2}, \\ a_{N-1,N-2} &= -\alpha_{N-1} + a_{N-2,N-3}, \\ &\cdots \\ a_{2,1} &= -\alpha_2 + a_{1,0}. \end{aligned}$$

Summing these relations yields

$$a_{N,N-1} = -(\alpha_N + \alpha_{N-1} + \cdots + \alpha_2) + a_{1,0}.$$

It follows from $p_1(x) = x + a_{1,0} = x - \alpha_1$ that

$$a_{N,N-1} = -\sum_{k=1}^{N} \alpha_k.$$

Using (3.6) and (3.5) gives

$$\sum_{k=1}^{N} \alpha_k = N\mu_1.$$

Letting $N = 1, 2, \ldots$ in the above sum, we obtain

$$\mu_1 = \alpha_1 = \alpha_2 = \cdots = \alpha_N. \tag{3.9}$$

Hence,

$$a_{N,N-1} = -N\mu_1. \tag{3.10}$$

This relation holds for $N = 1, 2, 3, \ldots$.

Consider the case $m = 2$. We have

$$\mu_2 := \int_{-1}^{1} x^2 d\lambda(x) = \frac{1}{N}\sum_{k=1}^{N}\left(x_k^{(N)}\right)^2.$$

Using

$$\sum_{k=1}^{N}\left(x_k^{(N)}\right)^2 = \left(\sum_{k=1}^{N} x_k^{(N)}\right)^2 - 2\sum_{1\le k<l\le N} x_k^{(N)} x_l^{(N)}$$

together with (3.5) and (3.7) gives

$$a_{N,N-2} = \frac{N}{2}(N\mu_1^2 - \mu_2). \tag{3.11}$$

Comparing coefficients of the power x^{k-2} in the left-hand side and right-hand side of (3.8) for $k = N+2, N+1, \ldots, 2$, gives the relations, in order,

$$\begin{aligned} a_{N+2,N} &= a_{N+1,N-1} - \alpha_{N+2}a_{N+1,N} - \beta_{N+1}, \\ a_{N+1,N-1} &= a_{N,N-2} - \alpha_{N+1}a_{N,N-1} - \beta_N, \\ &\cdots \\ a_{3,1} &= a_{2,0} - \alpha_3 a_{2,1} - \beta_2, \\ a_{2,0} &= -\alpha_2 a_{1,0} - \beta_1. \end{aligned}$$

Summing these relations and using (3.9) yields

$$a_{N,N-2} = -\mu_1\sum_{k=1}^{N-1} a_{k,k-1} - \sum_{k=1}^{N-1}\beta_k.$$

Taking into account that $a_{k,k-1} = -k\mu_1$ for $k = 1, 2, 3, \ldots$ (cf. (3.10)) shows that

$$a_{N,N-2} = \mu_1^2\sum_{k=1}^{N-1} k - \sum_{k=1}^{N-1}\beta_k = \mu_1^2\frac{N(N-1)}{2} - \sum_{k=1}^{N-1}\beta_k.$$

It now follows from (3.11) that

$$\sum_{k=1}^{N-1}\beta_k = \frac{N}{2}\left(\mu_2 - \mu_1^2\right).$$

Letting $N = 2, 3, 4, \dots$ in the above sum, we obtain

$$\frac{\beta_1}{2} = \beta_2 = \beta_3 = \cdots = \beta_N = \frac{1}{2}(\mu_2 - \mu_1^2).$$

To simplify the notation, let

$$\frac{\sigma^2}{4} := \frac{1}{2}(\mu_2 - \mu_1^2).$$

By the assumptions on the moments, we may choose $\sigma = \sqrt{2(\mu_2 - \mu_1^2)}$ positive. Then

$$\beta_1 = \frac{\sigma^2}{2}, \quad \beta_2 = \beta_3 = \cdots = \beta_N = \frac{\sigma^2}{4}.$$

Substituting these values of β_k and the values (3.9) of the α_k into (3.8), we obtain

$$\begin{aligned} p_0(x) &= 1, \quad p_1(x) = x - \mu_1, \\ p_k(x) &= (x - \mu_1)p_{k-1}(x) - \tfrac{\sigma^2}{2} p_{k-2}(x), \quad k = 2, 3, \dots, N+1. \end{aligned} \tag{3.12}$$

Let $\widehat{T}_k$ denote the monic Chebyshev polynomial of the first kind of degree k associated with the measure (2.7). Comparing the recursion relation (3.12) with that for the $\widehat{T}_k$ shows that

$$p_k(x) = \sigma^k \widehat{T}_k\left(\frac{x - \mu_1}{\sigma}\right), \qquad k = 0, 1, 2, \dots. \tag{3.13}$$

It follows from (3.13) that the zeros of the polynomial $p_N(x)$ are

$$x_k^{(N)} = \mu_1 + \sigma \cos\left(\frac{2k-1}{2N}\pi\right), \qquad k = 1, 2, \dots, N. \tag{3.14}$$

They lie in the interval $[-\sigma + \mu_1, \sigma + \mu_1]$. By [36, Theorem 6.1.1] the zeros of any family of orthogonal polynomials are dense on the support of the measure $d\lambda$, which by assumption is $[-1, 1]$. That means $[-\sigma + \mu_1, \sigma + \mu_1] = [-1, 1]$. This implies that $\mu_1 = 0$ and $\sigma = 1$. Hence, $d\lambda$ is $1/\pi$ times the measure (2.7). ∎

We chose the support of the measure $d\lambda$ to live in the interval $[-1, 1]$ in Theorem 3 to emphasize the relation to Theorem 1. However, this property is not required to show (3.14). We therefore have the slightly more general result.

Corollary 2. *Let $d\lambda$ be a nonnegative measure on the real axis with infinitely many points of support, let the moments satisfy* (3.2), *and let $p_0(x), p_1(x), p_2(x), \dots$ be a sequence of monic orthogonal polynomials associated with the measure. Denote the zeros of $p_N(x)$ by $x_1^{(N)}, x_2^{(N)}, \dots, x_N^{(N)}$. If for all $N = 1, 2, 3, \dots,$ it holds that*

$$\int p(x) d\lambda(x) = \nu_N \sum_{j=1}^{N} p\left(x_j^{(N)}\right)$$

for some scalar ν_N and all polynomials $p \in \mathbb{P}_2$, then $d\lambda$ is a Chebyshev measure (2.7) *(possibly scaled and translated).*

Proof. It follows from the assumption of the corollary that the polynomial p_N has the zeros (3.14). These are the zeros of a Chebyshev polynomial of the first kind of degree N for the interval $[-\sigma + \mu_1, \sigma + \mu_1]$, where $\sigma = \sqrt{2(\mu_2 - \mu_1^2)}$, and μ_1 and μ_2 are moments of $d\lambda$. It follows that $d\lambda$ is a scaled Chebyshev measure for the interval $[-\sigma + \mu_1, \sigma + \mu_1]$. ∎

THEOREM 4. *Let $p_0, p_1, \ldots, p_{n-1}$ be a family of orthonormal polynomials with respect to an inner product determined by a nonnegative measure $d\lambda$ on the real axis with infinitely many points of support, i.e., the polynomials satisfy* (2.2). *Assume that the first moments of $d\lambda$ have the properties* (3.2). *Let $x_1^{(N)}, x_2^{(N)}, \ldots, x_N^{(N)}$ denote the nodes of an N-point Gauss quadrature rule associated with the measure $d\lambda$. Consider the real $N \times n$ Vandermonde-like matrix $V_{N,n}$ determined by the polynomials p_j, $0 \leq j < n$, and nodes $x_k = x_k^{(N)}$, $1 \leq k \leq N$. Let $n \geq 3$. Then the matrix $V_{N,n}$ is optimally conditioned for all $N \geq n$ if and only if $d\lambda$ is a scaled or translated Chebyshev measure of the first kind.*

Proof. Suppose that the matrix $V_{N,n}$ of the form (1.1) with $n \geq 3$ is optimally conditioned for any N with respect to the Frobenius norm. This is equivalent to $V_{N,n}$ being optimally conditioned with respect to the spectral norm. The columns of $V_{N,n}$ then are orthogonal and of the same Euclidean norm. In other words,

$$\sum_{k=1}^{N} p_i(x_k)p_j(x_k) = 0, \quad i \neq j, \quad 0 \leq i, j < n, \tag{3.15}$$

and there is a constant c_N such that

$$c_N = \sum_{k=1}^{N} p_i^2(x_k), \qquad 0 \leq i < n.$$

Any polynomial $l \in \mathbb{P}_2$ can be represented in the form

$$l(x) = d_2 p_2(x) + d_1 p_1(x) + d_0 p_0(x)$$

for certain scalar coefficients d_k. We have $\mu_0 = 1$. Therefore $p_0(x) \equiv 1$. Due to the orthogonality of the polynomials p_0, p_1, p_2 with respect to (2.1) and (3.15), we have

$$\int l(x) d\lambda(x) = d_0 (p_0, 1)_\lambda = d_0$$

and

$$\sum_{k=1}^{N} l(x_k) = d_0 \sum_{k=1}^{N} p_0(x_k) = d_0 N.$$

This shows that (3.3) holds for the constant $\nu_N = 1/N$. It follows from Theorem 3 that $d\lambda$ is a scaled and possibly translated Chebyshev measure (2.7). □

The following result shows that if the Vandermonde-like matrix $V_{N,n}$ has sufficiently many more rows than columns, then suitable row scaling of $V_{N,n}$ will render a matrix with orthonormal columns. We will comment on the consequences for computation after the proof.

THEOREM 5. *Let the rectangular Vandermonde-like matrix $V_{N,n}$ be defined by a family of orthonormal polynomials $p_0, p_1, \ldots, p_{n-1}$ with respect to a nonnegative measure $d\lambda$, i.e., the polynomials satisfy* (2.2). *Let $x_1, x_2, \ldots, x_N$ be real distinct nodes, and assume that $n < \left\lfloor \frac{N-1}{2} \right\rfloor$. Then, generally, the columns of $V_{N,n}$ can be made orthonormal by scaling of the rows.*

Proof. Given arbitrary real distinct nodes $x_1, x_2, \ldots, x_N$, one can determine real weights $\lambda_1, \lambda_2, \ldots, \lambda_N$ such that

$$\int p(x) d\lambda(x) = p(x_1)\lambda_1 + p(x_2)\lambda_2 + \cdots + p(x_N)\lambda_N \qquad \forall p \in \mathbb{P}_{N-1}; \tag{3.16}$$

see, e.g., [36, Theorem 3.4.1]. The right-hand side of (3.16) is known as an interpolatory quadrature rule. We will assume that all weights λ_k are nonvanishing. This is the generic situation.

Since $n < \left\lfloor \frac{N-1}{2} \right\rfloor$, we have $p_i p_j \in \mathbb{P}_{N-1}$ for all $0 \leq i, j < n$. Therefore,

$$\delta_{ij} = \int p_i(x) p_j(x) d\lambda(x) = \sum_{k=1}^{N} p_i(x_k) p_j(x_k) \lambda_k,$$

where δ_{ij} is the Kronecker delta.

Let $\Lambda = \mathrm{diag}[\lambda_1^{1/2}, \lambda_2^{1/2}, \ldots, \lambda_N^{1/2}]$. Then the matrix $\Lambda V_{N,n}$ has orthogonal columns of Euclidean norm one. Note that if a weight λ_j is negative, then the corresponding diagonal entry of Λ is purely imaginary. $\square$

We remark that the usefulness of Theorem 5 in computations is limited. The theorem suggests that it may suffice to use about twice as many nodes as the degree of the highest-degree polynomial that defines the Vandermonde-like matrix $V_{N,n}$ to produce a row-scaled matrix $\Lambda V_{N,n}$ with orthonormal columns. Let $f \in \mathbb{R}^N$, and consider the least-squares approximation problem

$$\min_{y \in \mathbb{R}^n} \| V_{N,n} y - f \|_2. \tag{3.17}$$

Replacing this problem by

$$\min_{y \in \mathbb{R}^n} \| \Lambda V_{N,n} y - \Lambda f \|_2 \tag{3.18}$$

may not be appropriate when the diagonal entries of Λ vary considerably in magnitude. Assume that the nodes $x_1, x_2, \ldots, x_N$ that define the matrix $V_{N,n}$ are equidistant on the interval $[-1, 1]$. When $V_{N,n}$ is a standard Vandermonde matrix, then it is known that in order to make the problem not too ill-conditioned, one has to require that, roughly, $n \leq \frac{5}{2}\sqrt{N}$; see Barnard et al. [2]. More general results are shown by Platte et al. [30]. A difficulty that arises when using $n \approx N/2$ is that the weights in (3.18), i.e., the diagonal entries of Λ, may vary considerably in magnitude. This depends on that the quadrature rule (3.16) is a Newton–Cotes rule; see, e.g., [35, Chapter 3] for a discussion of these rules. The magnitude of the weight λ_j of largest magnitude grows quickly when N increases. This can make it difficult to determine the matrix $\Lambda V_{N,n}$ in finite precision arithmetic, as well as to justify the solution of the least-squares problem (3.18).

4. Szegő–Vandermonde matrices. In this section we consider polynomials that are orthogonal with respect to an inner product on the unit circle in $\mathbb{C}$ of the form

$$(f, g)_\lambda := \int_{-\pi}^{\pi} f(z) \overline{g(z)} d\lambda(\theta), \quad z = \exp(i\theta), \quad i = \sqrt{-1}. \tag{4.1}$$

The bar denotes complex conjugation, $d\lambda$ is a nonnegative measure with infinitely many points of support, and the functions f and g are polynomials. Let $p_0, p_1, p_2, \ldots$ be a family of polynomials that are orthonormal with respect to the inner product (4.1), i.e.,

$$(p_j, p_k)_\lambda = \begin{cases} 1, & j = k, \\ 0, & j \neq k, \end{cases}$$

with $\deg(p_j) = j$ for all $j \geq 0$. Polynomials that are orthogonal with respect to the inner product (4.1) are commonly referred to as *Szegő polynomials*. They have many applications in signal processing and frequency analysis; see, e.g., [21, 33].

Integrals of the form

$$\mathcal{I}f := \int_{-\pi}^{\pi} f(\exp(i\theta))d\lambda(\theta)$$

can be approximated by the quadrature formula

$$S_N f := \sum_{k=1}^{N} \lambda_k^{(N)} f\left(z_k^{(N)}\right), \tag{4.2}$$

where the $\lambda_k^{(N)} > 0$ are weights and the $z_k^{(N)}$ are distinct nodes on the unit circle in $\mathbb{C}$. The quadrature rule (4.2) is said to be an N-point Gauss–Szegő rule if

$$\mathcal{I}f = S_N f \qquad \forall f \in \Lambda_{-N+1,N-1}, \tag{4.3}$$

where

$$\Lambda_{-N+1,N-1} := \operatorname{span}\left\{1, z, z^{-1}, z^2, z^{-2}, \ldots, z^{N-1}, z^{-N+1}\right\}.$$

The existence of Gauss–Szegő quadrature rules is shown in, e.g., [25, Theorem 7.1]. The weights of Gauss–Szegő rules are unique and the nodes are unique up to a rotation on the unit circle, i.e., $z_1^{(N)}$ can be chosen arbitrarily on the unit circle. The other nodes, $z_2^{(N)}, \ldots, z_N^{(N)}$, are then uniquely determined; see, e.g., [24] for a discussion and extension.

Let (4.2) be an N-point Gauss–Szegő rule. The Christoffel function associated with the measure $d\lambda$ is given by

$$\Lambda_N(z) := \left(\sum_{k=0}^{N-1} |p_k(z)|^2\right)^{-1}; \tag{4.4}$$

see [29]. This expression is analogous to (2.4). Similarly to (2.5), the weights of the N-point Gauss–Szegő rule (4.2) are given by

$$\lambda_k^{(N)} = \Lambda_N\left(z_k^{(N)}\right), \quad k = 1, 2, \ldots, N.$$

The nodes $z_k^{(N)}$, $k = 1, 2, \ldots, N$, of the Gauss–Szegő rule (4.2) are zeros of a so-called para-orthogonal polynomial,

$$B_N(z, w_N) := p_N(z) + w_N p_N^*(z), \tag{4.5}$$

that satisfies the orthogonality relations

$$\int_{-\pi}^{\pi} z^k B_N(z) d\lambda(\theta) = 0, \quad z = \exp(i\theta), \quad 1 \leq k \leq N-1, \tag{4.6}$$

where the parameter $w_N \in \mathbb{C}$ can be chosen arbitrarily such that $|w_N| = 1$, and $p_N^*(z) := z^N \overline{p_N(1/\overline{z})}$ denotes the reversed polynomial associated with p_N; see Gonzáles et al. [17] or Jones et al. [25] for details. Note that $k > 0$ in (4.6). A different approach to define the nodes $z_k^{(N)}$ is described by Gragg [18]. Efficient algorithms for computing the nodes and weights of Gauss–Szegő rules are considered in [1, 19, 20, 22, 34].

We show a few properties of Vandermonde-like matrices defined by Szegő polynomials. These properties are analogous to those of the Vandermonde-like discussed in sections 2 and 3.

LEMMA 4. *Let $p_0, p_1, \dots, p_{N-1}$ be Szegő polynomials associated with a nonnegative measure $d\lambda$ on the unit circle with infinitely many points of support. In particular, $\deg(p_j) = j$ for all j. Let $z_1^{(N)}, z_2^{(N)}, \dots, z_N^{(N)}$ be nodes of the N-point Gauss–Szegő rule* (4.2)*. Then the Szegő–Vandermonde matrix*

$$V_{N,N} = \begin{bmatrix} p_0(z_1^{(N)}) & p_1(z_1^{(N)}) & \cdots & p_{N-1}(z_1^{(N)}) \\ p_0(z_2^{(N)}) & p_1(z_2^{(N)}) & \cdots & p_{N-1}(z_2^{(N)}) \\ \vdots & \vdots & \cdots & \vdots \\ p_0(z_N^{(N)}) & p_1(z_N^{(N)}) & \cdots & p_{N-1}(z_N^{(N)}) \end{bmatrix}$$

can be row scaled to become optimally conditioned.

Proof. The existence of a Gauss–Szegő rule (4.2) implies discrete orthogonality of the Szegő polynomials. We will use this to show the lemma. Clearly, $p_j(z) \in \mathrm{span}\{1, z, \dots, z^{N-1}\}$ for $0 \le j < N$. For z on the unit circle, we have $\bar{z} = z^{-1}$, and, therefore, $\overline{p_j(z)} \in \mathrm{span}\{1, z^{-1}, \dots, z^{-N+1}\}$ for $0 \le j < N$. It now follows from (4.2) and (4.3) that

$$\sum_{k=1}^{N} \lambda_k^{(N)} p_\ell\left(z_k^{(N)}\right) \overline{p_j\left(z_k^{(N)}\right)} = \int_{-\pi}^{\pi} p_\ell(z)\overline{p_j(z)} d\lambda(\theta) = \begin{cases} 1, & \ell = j, \\ 0, & \ell \neq j. \end{cases}$$

Let

$$D = \mathrm{diag}\left[\sqrt{\lambda_1^{(N)}}, \sqrt{\lambda_2^{(N)}}, \dots, \sqrt{\lambda_N^{(N)}}\right].$$

Then the matrix $Q = DV_{N,N}$ is unitary and, consequently, optimally conditioned. □

THEOREM 6. *Let $z_1, z_2, \dots, z_N$ be distinct complex numbers, and let $p_0, p_1, \dots, p_{N-1}$ be a family of orthonormal polynomials with respect to the inner product* (4.1)*. Then*

$$\kappa_F(V_{N,N}) \ge \left(\sum_{\ell=1}^{N} \Lambda_N(z_\ell) \cdot \sum_{\ell=1}^{N} (\Lambda_N(z_\ell))^{-1} \right)^{1/2},$$

where the matrix $V_{N,N}$ is determined by the values of the polynomials p_j at the nodes z_k, and Λ_N is the Christoffel function (4.4) *associated with the measure $d\lambda$.*

Proof. Consider the Lagrange polynomials

$$l_k(z) = \prod_{\substack{\ell=1 \\ \ell \neq k}}^{N} \frac{z - z_\ell}{z_k - z_\ell}, \quad k = 1, 2, \dots, N, \tag{4.7}$$

which form a basis for $\mathbb{P}_{N-1}$.

The polynomials $p_0, p_1, \dots, p_{N-1}$ are linearly independent. Therefore, the Lagrange polynomials can be expressed as linear combinations of the p_j,

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & \cdots & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NN} \end{bmatrix} \begin{bmatrix} p_0(z) \\ \vdots \\ p_{N-1}(z) \end{bmatrix} = \begin{bmatrix} l_1(z) \\ \vdots \\ l_N(z) \end{bmatrix}. \tag{4.8}$$

Similarly, the polynomials p_k can be expressed as linear combinations of Lagrange polynomials,

$$\begin{bmatrix} p_0(z_1) & p_0(z_2) & \cdots & p_0(z_N) \\ p_1(z_1) & p_1(z_2) & \cdots & p_1(z_N) \\ \vdots & \vdots & \cdots & \vdots \\ p_{N-1}(z_1) & p_{N-1}(z_2) & \cdots & p_{N-1}(z_1) \end{bmatrix} \begin{bmatrix} l_1(z) \\ \vdots \\ \\ l_N(z) \end{bmatrix} = \begin{bmatrix} p_0(z) \\ \vdots \\ \\ p_{N-1}(z) \end{bmatrix}. \tag{4.9}$$

It follows from (4.8) and (4.9) that $V_{N,N}^{-T} = [a_{\ell j}]_{\ell,j=1}^N$. Consequently,

$$\begin{aligned} \int_{-\pi}^{\pi} \sum_{\ell=1}^N |l_\ell(z)|^2 \, d\lambda(\theta) &= \int_{-\pi}^{\pi} \sum_{\ell=1}^N \left(\sum_{j=1}^N a_{\ell j} p_{j-1}(z) \sum_{k=1}^N \overline{a}_{\ell k} \overline{p_{k-1}(z)} \right) d\lambda(\theta) \\ &= \sum_{\ell=1}^N \sum_{j=1}^N \sum_{k=1}^N a_{\ell j} \overline{a}_{\ell k} \int_{-\pi}^{\pi} p_{j-1}(z) \overline{p_{k-1}(z)} d\lambda(\theta) \\ &= \sum_{\ell=1}^N \sum_{j=1}^N |a_{\ell j}|^2 = \left\| V_{N,N}^{-T} \right\|_F^2 = \left\| V_{N,N}^{-1} \right\|_F^2 . \end{aligned}$$

Nevai [29] shows that the Christoffel function associated with a nonnegative measure $d\lambda$ on the unit circle satisfies

$$\Lambda_N(z) = \min_{\substack{p \in \mathbb{P}_{N-1} \\ p(z)=1}} \int_{-\pi}^{\pi} |p(\zeta)|^2 d\lambda(\theta), \quad \zeta = \exp(i\theta). \tag{4.10}$$

The Lagrange polynomials (4.7) satisfy $l_\ell \in \mathbb{P}_{N-1}$ and $l_\ell(z_\ell) = 1$. It follows from (4.10) that

$$\int_{-\pi}^{\pi} |l_\ell(z)|^2 \, d\lambda(\theta) \geq \min_{\substack{p \in \mathbb{P}_{N-1} \\ p(z_\ell)=1}} \int_{-\pi}^{\pi} |p(z)|^2 \, d\lambda(\theta) = \Lambda_N(z_\ell), \quad z = \exp(i\theta).$$

Therefore,

$$\left\| V_{N,N}^{-1} \right\|_F^2 = \int_{-\pi}^{\pi} \sum_{\ell=1}^N |l_\ell(z)|^2 \, d\lambda(\theta) \geq \sum_{\ell=1}^N \Lambda_N(z_\ell), \quad z = \exp(i\theta).$$

In view of (4.4), we obtain

$$\|V_{N,N}\|_F^2 = \sum_{\ell=1}^N (\Lambda_N(z_\ell))^{-1}.$$

Hence,

$$\|V_{N,N}\|_F^2 \left\| V_{N,N}^{-1} \right\|_F^2 \geq \sum_{\ell=1}^N \Lambda_N(z_\ell) \cdot \sum_{\ell=1}^N (\Lambda_N(z_\ell))^{-1},$$

and the theorem follows. ∎

A special type of para-orthogonal polynomials (4.5), sometimes referred to as Delsarte–Genin para-orthogonal polynomials, satisfy the recurrence relations

$$R_N^{(1)}(z) = z p_{N-1}(z) + p_N^*(z), \quad (z-1) R_N^{(2)}(z) = z p_N(z) - p_N^*(z), \tag{4.11}$$

for $N = 0, 1, 2, \ldots$, with $p_{-1}(z) \equiv 0$. These polynomials were first considered in [7]. They are associated with symmetric or skew-symmetric measures on the unit circle in $\mathbb{C}$ and are important because they relate Szegő polynomials to orthogonal polynomials on the interval $[-1, 1]$; see [7, 8, 39]. Applications include frequency analysis; see [5]. The following result for the polynomials (4.11) is analogous to Theorem 3 for polynomials that are orthogonal on a real interval.

THEOREM 7. *Let $\{R_N^{(1)}\}_{N=0}^{\infty}$ and $\{R_N^{(2)}\}_{N=0}^{\infty}$ be families of para-orthogonal polynomials* (4.11) *associated with a symmetric nonnegative measure $d\lambda$ on the unit circle with infinitely many points of support. If for all $N = 2, 3, \ldots$, there is an N-node Gauss–Szegő quadrature rule, whose nodes are the zeros of $R_N^{(1)}$ or $R_N^{(2)}$ with all weights equal, then $d\lambda(\theta)$ is a multiple of $d\theta$. When $d\lambda$ is a skew-symmetric nonnegative measure on the unit circle, there is no such Gauss–Szegő quadrature rule.*

Proof. Assume that for every $N \geq 2$ there is a real scalar ν_N such that

$$\mu_m := \int_{-\pi}^{\pi} z^m d\lambda(\theta) = \nu_N \sum_{k=1}^{N} \left(z_k^{(N)}\right)^m, \quad m = 0, 1, 2, \ldots, \tag{4.12}$$

where $z = \exp(i\theta)$ and the nodes $z_1^{(N)}, z_2^{(N)}, \ldots, z_N^{(N)}$ are the zeros of the polynomials $R_N^{(1)}$ or $R_N^{(2)}$. We may scale the measure $d\lambda$ so that $\mu_0 = \int_{-\pi}^{\pi} d\lambda(\theta) = 1$. Setting $m = 0$ in (4.12) gives

$$1 = \int_{-\pi}^{\pi} d\lambda(\theta) = \nu_N \sum_{k=1}^{N} 1 = \nu_N N,$$

i.e., $\nu_N = \frac{1}{N}$. Turning to $m = 1$, we get

$$\mu_1 = \int_{-\pi}^{\pi} z d\lambda(\theta) = \frac{1}{N} \sum_{k=1}^{N} z_k^{(N)}.$$

Delsarte and Genin [7] showed that both families of para-orthogonal polynomials $R_N^{(1)}$ and $R_N^{(2)}$, defined by (4.11) and associated with a symmetric or skew-symmetric measure $d\lambda$, satisfy a three-term recurrence relation of the form

$$\begin{aligned} &R_N(z) = (z+1)R_{N-1}(z) - 4d_N z R_{N-2}(z), \quad N = 2, 3, \ldots, \\ &R_0(z) = 1, \quad R_1(z) = z + 1; \end{aligned} \tag{4.13}$$

see also [6] for more details. Consider the coefficient $a_{N,N-1}$ of the polynomial

$$R_N(z) = z^N + a_{N,N-1} z^{N-1} + a_{N,N-2} z^{N-2} + \cdots + a_{N,0}. \tag{4.14}$$

We obtain from Vieta's formulas that

$$a_{N,N-1} = -\sum_{k=1}^{N} z_k^{(N)} = -N\mu_1, \quad N \geq 2. \tag{4.15}$$

Comparing the coefficients $a_{N,N-1}$ in the left-hand sides of (4.13) and (4.14) for decreasing degrees N, we obtain

$$
\begin{aligned}
a_{N,N-1} &= a_{N-1,N-2} + (1-4d_N),\\
a_{N-1,N-2} &= a_{N-2,N-3} + (1-4d_{N-1}),\\
&\cdots\\
a_{2,1} &= a_{1,0} + (1-4d_2),\\
a_{1,0} &= 1.
\end{aligned} \tag{4.16}
$$

The relations (4.15) and (4.16) give

$$a_{2,1} = 2 - 4d_2 = -2\mu_1.$$

Hence, $d_2 = \frac{1}{2}(1+\mu_1)$. For $N \geq 3$, we obtain similarly that

$$a_{N,N-1} = -(N-1)\mu_1 + (1+4d_N) = -N\mu_1, \tag{4.17}$$

and, therefore, $d_N = \frac{1}{4}(1+\mu_1)$. It follows that

$$\frac{d_2}{2} = d_3 = \cdots = d_N = \frac{1}{4}(1+\mu_1). \tag{4.18}$$

Bracciali et al. [6] show that the coefficients $d_2, d_3, \ldots, d_N$ are positive. Hence, $\mu_1 > -1$.

Turning to the coefficient $a_{N,N-2}$ in (4.14), we get from Vieta's formulas that

$$a_{N,N-2} = \sum_{1\leq \ell < j \leq N} z_\ell^{(N)} z_j^{(N)}.$$

It now follows from

$$\mu_2 = \int_{-\pi}^{\pi} z^2 d\lambda(\theta) = \frac{1}{N}\left(\left(z_1^{(N)}\right)^2 + \cdots + \left(z_N^{(N)}\right)^2\right)$$

that, for $N \geq 3$,

$$
\begin{aligned}
a_{N,N-2} &= \frac{1}{2}\left(\left(z_1^{(N)} + \cdots + z_N^{(N)}\right)^2 - \left(\left(z_1^{(N)}\right)^2 + \cdots + \left(z_N^{(N)}\right)^2\right)\right)\\
&= \frac{N}{2}(N\mu_1^2 - \mu_2).
\end{aligned} \tag{4.19}
$$

Comparing the coefficients $a_{N,N-2}$ on the left-hand sides of (4.13) and (4.14), we obtain

$$a_{N,N-2} = a_{N-1,N-3} + a_{N-1,N-2} - (1+\mu_1)a_{N-2,N-3}. \tag{4.20}$$

Using the relation (4.17) twice (with N replaced by $N-1$ and $N-2$) and (4.19) (with N replaced by $N-1$) in (4.20) gives

$$\mu_2 = 3\mu_1^2 + 2\mu_1. \tag{4.21}$$

Now consider the coefficient $a_{N,N-3}$. We will derive an expression for the third moment

$$\mu_3 = \int_{-\pi}^{\pi} z^3 d\lambda(\theta) = \frac{1}{N}\left(\left(z_1^{(N)}\right)^3 + \cdots + \left(z_N^{(N)}\right)^3\right), \quad z = \exp(i\theta),$$

in terms of μ_1. Vieta's formula applied to (4.14) yields

$$a_{N,N-3} = -\sum_{1\le \ell<j<k\le N} z_\ell^{(N)} z_j^{(N)} z_k^{(N)}. \tag{4.22}$$

Any set of N complex numbers $z_1, z_2, \ldots, z_N$ satisfies the identity

$$\sum_{1\le \ell<j<k\le N} z_\ell z_j z_k = \frac{1}{6}\left(\left(\sum_{\ell=1}^N z_\ell\right)^3 - 3\left(\sum_{\ell=1}^N z_\ell^2\right)\left(\sum_{\ell=1}^N z_\ell\right) + 2\left(\sum_{\ell=1}^N z_\ell^3\right)\right).$$

Application of this relation to the nodes $z_1^{(N)}, z_2^{(N)}, \ldots, z_N^{(N)}$ in (4.22) and using (4.12) gives

$$a_{N,N-3} = -\frac{1}{6}(N^3\mu_1^3 - 3N^2\mu_2\mu_1 + 2N\mu_3). \tag{4.23}$$

Comparing the coefficients $a_{N,N-3}$ in the left-hand sides of (4.13) and (4.14) yields

$$a_{N,N-3} = a_{N-1,N-4} + a_{N-1,N-3} - (1+\mu_1)a_{N-2,N-4}, \quad N \ge 4,$$

and using the relations (4.19), (4.21), and (4.23) gives

$$\mu_3 = 10\mu_1^3 + 12\mu_1^2 + 3\mu_1. \tag{4.24}$$

From (4.6) the para-orthogonal polynomial R_2 satisfies

$$\int_{-\pi}^{\pi} zR_2(z)d\lambda(\theta) = 0;$$

see, e.g., [25]. Combining the recurrence relation (4.13) and $d_2 = \frac{1}{2}(1+\mu_1)$, we obtain $R_2(z) = z^2 - 2\mu_1 z + 1$. Hence,

$$\int_{-\pi}^{\pi} z(z^2 - 2\mu_1 z + 1)d\lambda(\theta) = \mu_3 - 2\mu_1\mu_2 + \mu_1 = 0.$$

Using (4.21) and (4.24) yields

$$4\mu_1^3 + 8\mu_1^2 + 4\mu_1 = 4\mu_1(\mu_1+1)^2 = 0.$$

Since $\mu_1 > -1$, it follows that $\mu_1 = 0$. Therefore, (4.18) simplifies to

$$\frac{d_2}{2} = d_3 = \cdots = d_N = \frac{1}{4}.$$

We obtain from the recursion formula (4.13) that $R_N(z) = z^N + 1$ for $N \ge 0$. The zeros of R_N are given by

$$z_k^{(N)} = \exp\left(i\frac{2k-1}{N}\pi\right), \quad k = 1, 2, \ldots, N.$$

Thus, they are equidistant on the unit circle. These nodes define an N-node Gauss–Szegő quadrature rule. This rule is exact for all functions in $\Lambda_{-N+1,N+1}$; cf. (4.3). Therefore, by (4.12),

$$\begin{aligned} \mu_j &= \int_{-\pi}^{\pi} z^j d\lambda(\theta) = \nu_N \sum_{k=1}^N \left(z_k^{(N)}\right)^j \\ &= \nu_N \sum_{k=1}^N \exp\left(ij\frac{2k-1}{N}\pi\right) = 0, \end{aligned} \qquad j = \pm1, \pm2, \ldots, \pm(N-1).$$

Thus, all moments except for μ_0 vanish. When $d\lambda(\theta)$ is a symmetric measure, this only leaves one possible solution for $d\lambda(\theta)$, namely, $d\lambda(\theta) = \frac{1}{2\pi} d\theta$, because μ_0 is assumed to be one. If, instead, $d\lambda(\theta)$ is a skew-symmetric measure, then also μ_0 vanish. This contradicts that $\mu_0 = 1$. $\square$

The monic Szegő polynomials associated with the measure $d\lambda(\theta) = \frac{1}{2\pi} d\theta$ are

$$p_N(z) = z^N, \quad N = 0, 1, 2, \ldots.$$

The nodes of an N-point Gauss–Szegő rule associated with this measure are equidistant on the unit circle. They are given by

$$z_k = \exp(i\theta_k), \quad \theta_k = \theta_0 + \frac{2\pi}{N}(k-1), \quad k = 1, 2, \ldots, N, \tag{4.25}$$

where $\theta_0 \in \mathbb{R}$ is arbitrary.

Consider the Vandermonde matrix

$$V_{N,N} = \begin{bmatrix} 1 & z_1 & \cdots & z_1^{N-1} \\ 1 & z_2 & \cdots & z_2^{N-1} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & z_N & \cdots & z_N^{N-1} \end{bmatrix} \tag{4.26}$$

defined by the nodes (4.25). It is easy to show that

$$V_{N,N} V_{N,N}^* = NI.$$

Hence, all singular values of the matrix $V_{N,N}$ equal $N^{1/2}$. It follows that $\kappa_2(V_{N,N}) = 1$ and $\kappa_F(V_{N,N}) = N$. Thus, the matrix (4.26) is optimally conditioned. We will show that this also holds for the corresponding rectangular Vandermonde matrices.

LEMMA 5. *Rectangular $n \times N$ Vandermonde matrices*

$$V_{n,N} = \begin{bmatrix} 1 & z_1 & \cdots & z_1^{N-1} \\ 1 & z_2 & \cdots & z_2^{N-1} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & z_n & \cdots & z_n^{N-1} \end{bmatrix} \tag{4.27}$$

with nodes

$$z_k = \exp(i\theta_k), \quad \theta_k = \theta_0 + \frac{2\pi}{M}(k-1), \quad k = 1, 2, \ldots, n,$$

where $M = max\{n, N\}$ and $\theta_0 \in \mathbb{R}$ is arbitrary, are optimally conditioned; i.e., all singular values of $V_{n,N}$ are equal. Moreover, with $m = min\{n, N\}$,

$$\begin{array}{rclrcl} \|V_{n,N}\|_2 & = & \sqrt{M}, & \|V_{n,N}^\dagger\|_2 & = & 1/\sqrt{M}, \\ \|V_{n,N}\|_F & = & \sqrt{mM}, & \|V_{n,N}^\dagger\|_F & = & \sqrt{m/M}. \end{array} \tag{4.28}$$

Proof. The result follows in the same way as the proof of Theorem 2, and by observing that all singular values $\sigma_1, \sigma_2, \ldots, \sigma_m$ of $V_{n,N}$ are equal to $\sqrt{M}$. Therefore,

$$\begin{aligned}
\|V_{n,N}\|_2 &= \sigma_1 = \sqrt{M}, \\
\|V_{n,N}^{\dagger}\|_2 &= 1/\sigma_1 = 1/\sqrt{M}, \\
\|V_{n,N}\|_F &= \sqrt{\sigma_1^2 + \cdots + \sigma_m^2} = \sqrt{mM}, \\
\|V_{n,N}^{\dagger}\|_F &= \sqrt{\sigma_1^{-2} + \cdots + \sigma_m^{-2}} = \sqrt{m/M}. \qquad \square
\end{aligned}$$

The above lemma is in agreement with the observation by Bazán [3] that matrices of the form (4.27), with $N \geq n$, are fairly well conditioned when the nodes z_j are close to the unit circle and no pair of distinct nodes are very close to each other.

THEOREM 8. *For any N and n, $N \geq n$, matrices of the form* (4.27) *can be row scaled to be optimally conditioned if and only if the arguments θ_k of the nodes $z_k = \rho_k \exp(i\theta_k)$ are of the form $\theta_k = \theta_0 + \frac{2\pi}{n} j_k$ for $1 \leq k \leq n$, where the j_k are integers such that the nodes $z_1, z_2, \ldots, z_n$ are distinct.*

Proof. Consider the row scaling $DV_{n,N}$ of the matrix (4.27), where

$$D = \mathrm{diag}[d_1, d_2, \ldots, d_n], \quad d_k > 0 \ \ \forall k.$$

It follows from Lemma 2 that if the matrix $DV_{n,N}$ is optimally conditioned, then its rows are orthogonal and have all the same Euclidean norm. Let the nodes of the matrix (4.27) be of the form $z_k = \rho_k \exp(i\theta_k)$, $k = 1, 2, \ldots, n$, where $\theta_k \in \mathbb{R}$. Assume that the rows of $DV_{n,N}$ are orthogonal. This implies that the nodes $z_1, z_2, \ldots, z_n$ are distinct. The orthogonality of consecutive rows of the matrix $DV_{n,N}$ yields

$$\begin{aligned}
d_k d_{k+1} + d_k d_{k+1} \bar{z}_k z_{k+1} + d_k d_{k+1} \bar{z}_k^2 z_{k+1}^2 + \cdots \\
+ d_k d_{k+1} \bar{z}_k^{N-1} z_{k+1}^{N-1} = 0, \ \ k = 1, 2, \ldots, n-1.
\end{aligned}$$

Let $\phi_k := \theta_{k+1} - \theta_k$. The above equations can be written as

$$1 + \rho_k \rho_{k+1} \exp(i\phi_k) + \rho_k^2 \rho_{k+1}^2 \exp(i2\phi_k) + \cdots + \rho_k^{N-1} \rho_{k+1}^{N-1} \exp(i(N-1)\phi_k) = 0 \tag{4.29}$$

for $k = 1, 2, \ldots, n-1$. The equations (4.29) imply that

$$\rho_k^N \rho_{k+1}^N \exp(iN\phi_k) = 1, \quad k = 1, 2, \ldots, n-1.$$

It follows that $\phi_k = \frac{2\pi}{N} j_k$ for $k = 1, 2, \ldots, n-1$, where the j_k are integers such that the nodes $z_1, z_2, \ldots, z_n$ are distinct. We choose the scaling factors d_j so that all rows of $DV_{n,N}$ have the same norm. The matrix $DV_{n,N}$ then satisfies the conditions of Lemma 2 and therefore is optimally conditioned. $\square$

The discrete Fourier transform (DFT) matrix can be expressed as $\sqrt{\frac{1}{N}} V_{N,N}$. Denote the nodes by $z_1, z_2, \ldots, z_N$. Theorem 8 shows that the DFT matrix is orthogonal if and only if the nodes are equidistant on the unit circle.

5. General Vandermonde-type matrices. This section considers Vandermonde matrices, whose nodes are zeros of certain orthogonal polynomials on an interval or on the unit circle. Eisinberg et al. [9] showed that the condition number $\kappa_2(V_{n,N})$ of a rectangular (standard) Vandermonde matrix $V_{n,N} \in \mathbb{R}^{N \times n}$, with $N \geq n$ and Chebyshev nodes, is independent of the number of nodes N. We extend this result to Vandermonde-type matrices of the form (1.1), where the polynomials p_i are not required to be orthogonal and $\deg(p_j)$ may be different from j.

THEOREM 9. *Let* $\{p_i\}_{i=0}^{n-1}$ *be a set of linearly independent polynomials in* $\mathbb{P}_{m-1}$, *and let* $\{l_i\}_{i=0}^{n-1}$ *be an arbitrary set of polynomials in* $\mathbb{P}_{m-1}$. *Hence,* $m \geq n$. *The polynomials* $l_0, l_1, \ldots, l_{n-1}$ *may be linearly dependent. Let the nodes* $x_1, x_2, \ldots, x_N$ *be zeros of the Chebyshev polynomial* $T_N(x)$ *with* $N \geq m$, *and define the Vandermonde-type matrices*

$$P_{N,n} = \begin{bmatrix} p_0(x_1) & \cdots & p_{n-1}(x_1) \\ p_0(x_2) & \cdots & p_{n-1}(x_2) \\ \vdots & \vdots & \vdots \\ p_0(x_N) & \cdots & p_{n-1}(x_N) \end{bmatrix}, \quad L_{N,n} = \begin{bmatrix} l_0(x_1) & \cdots & l_{n-1}(x_1) \\ l_0(x_2) & \cdots & l_{n-1}(x_2) \\ \vdots & \vdots & \vdots \\ l_0(x_N) & \cdots & l_{n-1}(x_N) \end{bmatrix}. \tag{5.1}$$

There are constants $\{d_j\}_{j=1}^{6}$ *that can be chosen independently of* N, *such that*

$$\begin{array}{rclrcl} \|P_{N,n}\|_F & = & \sqrt{\frac{N}{\pi}}d_1, & \|P_{N,n}^\dagger\|_F & = & \sqrt{\frac{\pi}{N}}d_2, \\ \|P_{N,n}\|_2 & = & \sqrt{\frac{N}{\pi}}d_3, & \|P_{N,n}^\dagger\|_2 & = & \sqrt{\frac{\pi}{N}}d_4, \\ \|L_{N,n}\|_F & = & \sqrt{\frac{N}{\pi}}d_5, & \|L_{N,n}\|_2 & = & \sqrt{\frac{N}{\pi}}d_6 \end{array} \tag{5.2}$$

for all $N \geq m$.

Proof. The polynomials p_i and l_i, $0 \leq i < n$, can be expressed as

$$p_i(x) = \sum_{k=0}^{m-1} c_{i,k}\tilde{T}_k(x), \quad l_i(x) = \sum_{k=0}^{m-1} \hat{c}_{i,k}\tilde{T}_k(x),$$

where the $\tilde{T}_k$ denote normalized Chebyshev polynomials. We express the matrices (5.1) in factored form,

$$P_{N,n} = T_{N,m}C, \qquad L_{N,n} = T_{N,m}\hat{C},$$

where

$$T_{N,m} = \begin{bmatrix} \tilde{T}_0(x_1) & \cdots & \tilde{T}_{m-1}(x_1) \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \tilde{T}_0(x_N) & \cdots & \tilde{T}_{m-1}(x_N) \end{bmatrix} \tag{5.3}$$

and

$$C = \begin{bmatrix} c_{0,0} & \cdots & c_{0,n-1} \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ c_{m-1,0} & \cdots & c_{m-1,n-1} \end{bmatrix}, \quad \hat{C} = \begin{bmatrix} \hat{c}_{0,0} & \cdots & \hat{c}_{0,n-1} \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \hat{c}_{m-1,0} & \cdots & \hat{c}_{m-1,n-1} \end{bmatrix}.$$

Since $P_{N,n}$ is of full rank, so are the matrices $T_{N,m}$ and C. It follows from Theorem 2 and Lemma 2 that $T_{N,m} = \sqrt{\frac{N}{\pi}}U$, where the matrix U has orthonormal columns. We obtain

$$\|P_{N,n}\|_F = \|T_{N,m}C\|_F = \sqrt{\frac{N}{\pi}}\|C\|_F,$$

$$\|P_{N,n}^\dagger\|_F = \|C^\dagger T_{N,m}^\dagger\|_F = \sqrt{\frac{\pi}{N}}\|C^\dagger\|_F,$$

$$\|L_{N,n}\|_F = \|T_{N,m}\hat{C}\|_F = \sqrt{\frac{\pi}{N}}\|\hat{C}\|_F.$$

Similarly,

$$\|P_{N,n}\|_2 = \sqrt{\frac{N}{\pi}}\|C\|_2, \quad \|P^{\dagger}_{N,n}\|_2 = \sqrt{\frac{\pi}{N}}\|C^{\dagger}\|_2, \quad \|L_{N,n}\|_2 = \sqrt{\frac{\pi}{N}}\|\hat{C}\|_2.$$

Letting

$$d_1 = \|C\|_F, \quad d_2 = \|C^{\dagger}\|_F, \quad d_3 = \|C\|_2, \quad d_4 = \|C^{\dagger}\|_2, \quad d_5 = \|\hat{C}\|_F, \quad d_6 = \|\hat{C}\|_2$$

concludes the proof. ∎

We remark that while the constants d_j, $1 \le j \le 6$, in (5.2) are independent of N, they may be large. Their sizes depend on the choice of polynomials p_j and l_k.

If $\deg(p_i) = i$ for $0 \le i < n$, then $P_{N,n} = T_{N,n}C$ for some nonsingular upper triangular matrix $C \in \mathbb{R}^{n\times n}$. This means that the QR factorization of $P_{N,n}$ is explicitly known. The following result follows directly from Theorem 9 and generalizes [26, Proposition 2.2], which is concerned with Vandermonde matrices with Chebyshev nodes.

COROLLARY 3. *Let the matrix $P_{N,n}$ be given by* (5.1). *Then the condition numbers $\kappa_F(P_{N,n})$ and $\kappa_2(P_{N,n})$ can be bounded independently of N.*

Let for the moment the nodes $x_1, x_2, \ldots, x_N$ in the matrix (5.3) be distinct, but otherwise arbitrary, and let $C \in \mathbb{R}^{m\times n}$. Then

$$\|T_{N,m}C\|_F\|C^{\dagger}T^{\dagger}_{N,m}\|_F \le \|C\|_F\|C^{\dagger}\|_F\|T_{N,m}\|_2\|T^{\dagger}_{N,m}\|_2, \tag{5.4}$$

$$\|T_{N,m}C\|_2\|C^{\dagger}T^{\dagger}_{N,m}\|_2 \le \|C\|_2\|C^{\dagger}\|_2\|T_{N,m}\|_2\|T^{\dagger}_{N,m}\|_2. \tag{5.5}$$

Letting the x_j be Chebyshev nodes yields $\|T_{N,m}\|_2\|T^{\dagger}_{N,m}\|_2 = 1$ (see Theorem 2) and therefore minimizes the upper bounds in (5.4) and (5.5).

Example 5.1. We compare numerically condition numbers of Vandermonde-like and Vandermonde matrices determined by Chebyshev nodes and optimal nodes. The optimal nodes, i.e., the nodes that minimize the condition number $\kappa_F(V_{N,n})$, are determined with Wolfram Mathematica Software using the minimizer NMinimize. We cannot be certain that the nodes so computed indeed are optimal, but the minimization function NMinimize gave the same nodes for many different initial node choices. Computed condition numbers for Vandermonde-like matrices defined by Legendre polynomials and Chebyshev polynomials of the second kind for Chebyshev and optimal nodes are shown in Table 5.1 and Table 5.2, respectively. Corresponding results

TABLE 5.1
Example 5.1. *Frobenius condition numbers of Vandermonde-like matrices $V_{N,n}$ defined by Legendre polynomials and Chebyshev or optimal nodes.*

n	N	Chebyshev nodes	Optimal nodes
5	5	$6.1 \cdot 10^0$	$5.2 \cdot 10^0$
10	10	$1.3 \cdot 10^1$	$1.1 \cdot 10^1$
20	20	$2.9 \cdot 10^1$	$2.3 \cdot 10^1$
35	35	$5.4 \cdot 10^1$	$4.3 \cdot 10^1$
40	40	$6.2 \cdot 10^1$	$5.0 \cdot 10^1$
5	10	$6.1 \cdot 10^0$	$5.0 \cdot 10^0$
10	20	$1.3 \cdot 10^1$	$1.0 \cdot 10^1$
20	40	$2.9 \cdot 10^1$	$2.2 \cdot 10^1$
35	70	$5.4 \cdot 10^1$	$4.1 \cdot 10^1$
40	80	$6.2 \cdot 10^1$	$4.7 \cdot 10^1$

TABLE 5.2
Example 5.1. *Frobenius condition numbers of Vandermonde-like matrices $V_{N,n}$ defined by Chebyshev polynomials of the second kind and Chebyshev or optimal nodes.*

n	N	Chebyshev nodes	Optimal nodes
5	5	$8.2 \cdot 10^0$	$5.5 \cdot 10^0$
10	10	$2.3 \cdot 10^1$	$1.3 \cdot 10^1$
20	20	$6.4 \cdot 10^1$	$3.2 \cdot 10^1$
35	35	$1.5 \cdot 10^2$	$7.0 \cdot 10^1$
40	40	$1.8 \cdot 10^2$	$8.5 \cdot 10^1$
5	10	$8.2 \cdot 10^0$	$5.0 \cdot 10^0$
10	20	$2.3 \cdot 10^1$	$1.1 \cdot 10^1$
20	40	$6.4 \cdot 10^1$	$2.8 \cdot 10^1$
35	70	$1.5 \cdot 10^2$	$6.0 \cdot 10^1$
40	80	$1.8 \cdot 10^2$	$7.3 \cdot 10^1$

TABLE 5.3
Example 5.1. *Frobenius condition numbers of Vandermonde matrices $V_{N,n}$ with Chebyshev and optimal nodes.*

n	N	Chebyshev nodes	Optimal nodes
5	5	$2.8 \cdot 10^1$	$2.3 \cdot 10^1$
10	10	$2.3 \cdot 10^3$	$1.6 \cdot 10^3$
20	20	$1.5 \cdot 10^7$	$9.9 \cdot 10^6$
35	35	$8.3 \cdot 10^{12}$	$5.1 \cdot 10^{12}$
40	40	$3.3 \cdot 10^{13}$	$2.6 \cdot 10^{13}$
5	10	$2.8 \cdot 10^1$	$1.7 \cdot 10^1$
10	20	$2.3 \cdot 10^3$	$1.2 \cdot 10^3$
20	40	$1.5 \cdot 10^7$	$7.5 \cdot 10^6$
35	70	$8.3 \cdot 10^{12}$	$4.0 \cdot 10^{12}$
40	80	$3.3 \cdot 10^{13}$	$1.6 \cdot 10^{13}$

for (standard) Vandermonde matrices are displayed in Table 5.3. As can be expected, the matrices of Table 5.3 have much larger condition numbers than the matrices of Tables 5.1 and 5.2. Chebyshev nodes can be seen to be "near-optimal" in the sense that the condition numbers for the Chebyshev nodes for all examples are larger by only a fairly small factor than the condition numbers for optimal nodes for all combinations of n and N reported.

Numerical experiments suggest that the nodes that minimize the condition number $\kappa_F(V_{N,n})$ are unique. A similar observation for square Vandermonde-like matrices is reported by Gautschi [14]. In fact, our experiments suggest that $\kappa_F(V_{N,n})$ is a locally convex function of the nodes in a neighborhood of the optimal nodes. We remark that it is not hard to show, that for the special case when

$$V_{2,2}(x_1, x_2) = \begin{bmatrix} c_0 & c_1 + c_2 x_1 \\ c_0 & c_1 + c_2 x_2 \end{bmatrix},$$

the condition number $\kappa_F(V_{2,2}(x_1, x_2))$ is strongly convex in the whole region $-\infty < x_1 < x_2 < \infty$. This follows from the fact that the Hessian of $\kappa_F^2(V_{2,2}(x_1, x_2))$ is positive definite. □

The following result is an analogue of Theorem 9 for polynomials with nodes on the unit circle.

THEOREM 10. *Let $\{p_i\}_{i=0}^{n-1}$ be a set of linearly independent polynomials of the form*

$$p_i(z) = c_{i,1} z^{k_1} + c_{i,2} z^{k_1+1} + \cdots + c_{i,m-1} z^{k_1+m-1}$$

with $m \geq n$ and let $k_1 \geq 0$ be an arbitrary integer. Further, let $\{l_i\}_{i=0}^{n-1}$ be a set of arbitrary polynomials of the form

$$l_i(z) = \hat{c}_{i,1} z^{k_2} + \hat{c}_{i,2} z^{k_2+1} + \cdots + \hat{c}_{i,m-1} z^{k_2+m-1}$$

with $k_2 \geq 0$ an arbitrary integer. Let the nodes $z_1, z_2, \ldots, z_N$ be equidistant on the unit circle, i.e., $z_k = \exp(i\theta_k)$, where $\theta_k = \theta_0 + \frac{2\pi k}{N}$ with $\theta_0 \in \mathbb{R}$ arbitrary. Consider the Vandermonde-like matrices

$$P_{N,n} = \begin{bmatrix} p_0(z_1) & \cdots & p_{n-1}(z_1) \\ p_0(z_2) & \cdots & p_{n-1}(z_2) \\ \cdots & \cdots & \cdots \\ p_0(z_N) & \cdots & p_{n-1}(z_N) \end{bmatrix}, \quad L_{N,n} = \begin{bmatrix} l_0(z_1) & \cdots & l_{n-1}(z_1) \\ l_0(z_2) & \cdots & l_{n-1}(z_2) \\ \cdots & \cdots & \cdots \\ l_0(z_N) & \cdots & l_{n-1}(z_N) \end{bmatrix}.$$

There are constants $\{d_j\}_{j=1}^{6}$, that can be chosen independently of N, such that

$$\begin{aligned} \|P_{N,n}\|_F &= \sqrt{N} d_1, \qquad \|P_{N,n}^{\dagger}\|_F = \frac{d_2}{\sqrt{N}}, \\ \|P_{N,n}\|_2 &= \sqrt{N} d_3, \qquad \|P_{N,n}^{\dagger}\|_2 = \frac{d_4}{\sqrt{N}}, \\ \|L_{N,n}\|_F &= \sqrt{N} d_5, \qquad \|L_{N,n}\|_2 = \sqrt{N} d_6 \end{aligned} \tag{5.6}$$

for all $N \geq m$.

Proof. We have the factorizations

$$P_{N,n} = D_1 Z C, \quad L_{N,n} = D_2 Z \hat{C},$$

where

$$D_1 = \operatorname{diag}\left[z_1^{k_1}, z_2^{k_1}, \ldots, z_N^{k_1}\right], \quad D_2 = \operatorname{diag}\left[z_1^{k_2}, z_2^{k_2}, \ldots, z_N^{k_2}\right],$$

and

$$Z = \begin{bmatrix} 1 & z_1 & \cdots & z_1^{m-1} \\ 1 & z_2 & \cdots & z_2^{m-1} \\ \cdots & \cdots & \cdots & \cdots \\ 1 & z_N & \cdots & z_N^{m-1} \end{bmatrix}.$$

Moreover,

$$C = \begin{bmatrix} c_{0,0} & \cdots & c_{n-1,0} \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ c_{0,m-1} & \cdots & c_{n-1,m-1} \end{bmatrix}, \quad \hat{C} = \begin{bmatrix} \hat{c}_{0,0} & \cdots & \hat{c}_{n-1,0} \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \hat{c}_{0,m-1} & \cdots & \hat{c}_{n-1,m-1} \end{bmatrix}.$$

The matrices D_1 and D_2 are unitary. Using the bounds (4.28), we obtain

$$\begin{aligned} \|L_{N,n}\|_F &= \|D_2 Z \hat{C}\|_F = \|Z\|_2 \|\hat{C}\|_F = \sqrt{N} \|\hat{C}\|_F, \\ \|P_{N,n}^{\dagger}\|_F &= \|C^{\dagger} Z^{\dagger} D_1^{\dagger}\|_F = \|Z^{\dagger}\|_2 \|C^{\dagger}\|_F = \tfrac{1}{\sqrt{N}} \|C^{\dagger}\|_F. \end{aligned}$$

The remaining bounds (5.6) follow similarly. ∎

6. Conclusion. Gautschi investigated the conditioning of square Vandermonde and Vandermonde-like matrices determined by orthogonal polynomials with respect to an inner product defined by a measure with support on the real axis. This paper shows results for rectangular Vandermonde and Vandermonde-like matrices. Orthogonal polynomials that define the latter matrices are determined by an inner product that is associated with a measure with support on the real axis or on the unit circle.

Acknowledgment. The authors would like to thank the referees for comments.

REFERENCES

[1] G. S. Ammar, L. Reichel, and D. C. Sorensen, *Algorithm* 730*: An implementation of a divide and conquer algorithm for the unitary eigenproblem*, ACM Trans. Math. Softw., 18 (1992), pp. 292–307, and 20 (1994), p. 161.
[2] R. W. Barnard, G. Dahlquist, K. Pearce, L. Reichel, and K. C. Richards, *Gram polynomials and the Kummer function*, J. Approx. Theory, 94 (1998), pp. 128–143.
[3] F. S. V. Bazán, *Conditioning of rectangular Vandermonde matrices with nodes in the unit disk*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 679–693.
[4] B. Beckermann, *The condition number of real Vandermonde, Krylov, and positive definite Hankel matrices*, Numer. Math., 85 (2000), pp. 553–577.
[5] C. Bracciali, X. Li, and A. Sri Ranga, *Real orthogonal polynomials in frequency analysis*, Math. Comp., 74 (2004), pp. 341–362.
[6] C. Bracciali, A. Sri Ranga, and A. Swaminathan, *Para-orthogonal polynomials on the unit circle satisfying three term recurrence formulas*, Applied Numer. Math., 109 (2016), pp. 19–40.
[7] P. Delsarte and Y. Genin, *The split Levinson algorithm*, IEEE Trans. Acoust. Speech Signal Process., 34 (1986), pp. 470–478.
[8] P. Delsarte and Y. Genin, *The tridiagonal approach to Szegő's orthogonal polynomials, Toeplitz linear system, and related interpolation problems*, SIAM J. Math. Anal., 19 (1988), pp. 718–735.
[9] A. Eisinberg, G. Franzé, and N. Salerno, *Rectangular Vandermonde matrices on Chebyshev nodes*, Linear Algebra Appl., 338 (2001), pp. 27–36.
[10] W. Gautschi, *Optimally conditioned Vandermonde matrices*, Numer. Math., 24 (1975), pp. 1–12.
[11] W. Gautschi, *On inverses of Vandermonde and confluent Vandermonde matrices* III, Numer. Math., 29 (1978), pp. 445–450.
[12] W. Gautschi, *The condition of Vandermonde-like matrices involving orthogonal polynomials*, Linear Algebra Appl., 52/53 (1983), pp. 293–300.
[13] W. Gautschi, *Numerical Analysis*, 2nd ed., Birkhäuser, Basel, 2012.
[14] W. Gautschi, *Optimally scaled and optimally conditioned Vandermonde and Vandermonde-like matrices*, BIT, 51 (2011), pp. 103–125.
[15] W. Gautschi and G. Inglese, *Lower bounds for the condition number of Vandermonde matrices*, Numer. Math., 52 (1988), pp. 241–250.
[16] W. Gautschi, *Orthogonal Polynomials: Approximation and Computation*, Oxford University Press, Oxford, 2004.
[17] P. González-Vera, J. C. Santos-León, and O. Njåstad, *Some results about numerical quadrature on the unit circle*, Adv. Comput. Math., 5 (1996), pp. 297–328.
[18] W. B. Gragg, *Positive definite Toeplitz matrices, the Arnoldi process for isometric operators, and Gaussian quadrature on the unit circle*, J. Comput. Appl. Math., 46 (1993), pp. 183–198. This is a slightly edited version of a paper first published in Russian in *Numerical Methods in Linear Algebra*, ed. E. S. Nikolaev, Moscow University Press, Moscow, 1982, pp. 16–32.
[19] W. B. Gragg, *The QR algorithm for unitary Hessenberg matrices*, J. Comput. Appl. Math., 16 (1986), pp. 1–8.
[20] W. B. Gragg and L. Reichel, *A divide and conquer method for unitary and orthogonal eigenproblems*, Numer. Math., 57 (1990), pp. 695–718.
[21] U. Grenander and G. Szegő, *Toeplitz Forms and Their Applications*, Chelsea, New York, 1984.
[22] M. Gu, R. Guzzo, X.-B. Chi, and X.-Q. Cao, *A stable divide and conquer algorithm for the unitary eigenproblem*, SIAM J. Matrix Anal. Appl., 25 (2003), pp. 385–404.

[23] G. H. Golub and J. H. Wilkinson, *Note on the iterative refinement of least squares solution*, Numer. Math., 9 (1966), pp. 139–148.
[24] C. Jagels and L. Reichel, *Szegő–Lobatto quadrature rules*, J. Comput. Appl. Math., 200 (2007), pp. 116–126.
[25] W. Jones, O. Njåstad, and W. J. Thron, *Moment theory, orthogonal polynomials, quadrature and continued fractions associated with the unit circle*, Bull. London Math. Soc., 21 (1989), pp. 113–152.
[26] M. Kuian, L. Reichel, and S. V. Shiyanovskii, *Fast factorization of rectangular Vandermonde matrices with Chebyshev nodes*, Numer. Algorithms, 81 (2019), pp. 547–559.
[27] R.-C. Li, *Asymptotically optimal lower bounds for the condition number of a real Vandermonde matrix*, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 829–844.
[28] R.-C. Li, *Vandermonde matrices with Chebyshev nodes*, Linear Algebra Appl., 428 (2008), pp. 1803–1832.
[29] P. Nevai, *Géza Freud, orthogonal polynomials and Christoffel functions. A case study*, J. Approx. Theory, 48 (1986), pp. 3–167.
[30] R. Platte, L. N. Trefethen, and A. B. J. Kuijlaars, *Impossibility of fast stable approximation of analytic functions from equispaced samples*, SIAM Rev., 53 (2011), pp. 308–318.
[31] C. Posse, *Sur les quadratures*, Nouv. Ann. Math., 14 (1875), pp. 49–62.
[32] K. R. Rao and P. Yip, *Discrete Cosine Transform*, Academic Press, Boston, 1991.
[33] L. Reichel and G. S. Ammar, *Fast approximation of dominant harmonics by solving an orthogonal eigenvalue problem*, in Mathematics in Signal Processing II, ed. J. G. McWhirter, Oxford University Press, Oxford, 1990, pp. 575–591.
[34] M. Stewart, *An error analysis of a unitary Hessenberg QR algorithm*, SIAM J. Matrix. Anal. Appl., 28 (2006), pp. 40–67.
[35] J. Stoer and R. Bulirsch, *Introduction to Numerical Analysis*, 3rd ed., Springer, New York, 2002.
[36] G. Szegő, *Orthogonal Polynomials*, 4th ed., Amer. Math. Soc., Providence, 1975.
[37] L. N. Trefethen and D. Bau III, *Numerical Linear Algebra*, SIAM, Philadelphia, 1997.
[38] E. E. Tyrtyshnikov, *How bad are Hankel matrices?*, Numer. Math., 67 (1994), pp. 261–269.
[39] A. Zhedanov, *On some classes of polynomials orthogonal on arcs of the unit circle connected with symmetric orthogonal polynomials on an interval*, J. Approx. Theory, 94 (1998), pp. 73–106.