BIT Numer Math (2011) 51: 103–125
DOI 10.1007/s10543-010-0293-1

BIT

Optimally scaled and optimally conditioned Vandermonde and Vandermonde-like matrices

Walter Gautschi

Received: 1 July 2010 / Accepted: 16 October 2010 / Published online: 6 November 2010
© Springer Science + Business Media B.V. 2010

Abstract Vandermonde matrices with real nodes are known to be severely ill-conditioned. We investigate numerically the extent to which the condition number of such matrices can be reduced, either by row-scaling or by optimal configurations of nodes. In the latter case we find empirically the condition of the optimally conditioned $n \times n$ Vandermonde matrix to grow exponentially at a rate slightly less than $(1+\sqrt{2})^n$. Much slower growth—essentially linear—is observed for optimally conditioned Vandermonde-Jacobi matrices. We also comment on the computational challenges involved in determining condition numbers of highly ill-conditioned matrices.

Keywords Singular value decomposition · Condition numbers · Vandermonde matrices · Optimal scaling · Optimal conditioning

Mathematics Subject Classification (2000) 15A18 · 15Bxx · 65F35

1 Introduction

The problem of optimally conditioned Vandermonde matrices,[1] that is, of determining a configuration of real nodes in a Vandermonde matrix that minimizes its condition number, has been addressed by us some time ago [3]. A related problem is that of

[1] As is well known, the attribution to Vandermonde of these matrices is incorrect. While it is true that Vandermonde made notable contributions to the theory of determinants, which he founded, and to the theory of equations and combinatorial analysis, there is no trace of "Vandermonde determinants", let alone "Vandermonde matrices", in any of his four mathematical papers. See Lebesgue [10, p. 207], who suggests that mistaking upper indices for powers may have been the source of this error.

Communicated by Lothar Reichel.

W. Gautschi (✉)
Department of Computer Sciences, Purdue University, West Lafayette, IN 47907-2066, USA
e-mail: wxg@cs.purdue.edu

optimal scaling of a Vandermonde matrix with given real nodes, which seems to have received less attention. There are indeed serious computational challenges brought on by the highly ill-conditioned matrices involved. In the case of condition numbers based on the Frobenius matrix norm, for example, the singular value decomposition of Vandermonde matrices can produce results that, at first sight, look reasonable, but in fact are wrong by many orders of magnitude. An example of this is exhibited in Sect. 2. To obtain reliable answers requires the use of variable-(high)precision computation. Further examples of optimal scaling are presented in Sect. 3, including one that is close to optimal conditioning. In Sect. 4 we first draw attention to, and correct, a small error in our example [3, (5.14)] of an optimally conditioned third-order Vandermonde matrix with real symmetric nodes. We then use high-precision calculations in combination with constrained and unconstrained optimization to compute optimally conditioned Vandermonde matrices of orders n up to 10, with nonnegative nodes in Sect. 4.1, with symmetric nodes in Sect. 4.2, and, for $n \le 13$, with unconstrained real nodes in Sect. 4.3. In the last case we find numerically the optimal Frobenius condition number to grow exponentially at a rate somewhat less than $(1+\sqrt{2})^n$. In Sect. 5 we give an analogous discussion of the conditioning of Vandermonde-like matrices and show that optimally conditioned Vandermonde-Chebyshev matrices are perfectly conditioned in the Frobenius norm.

The implications of conditioning, based on condition numbers as here defined, to the sensitivity to errors in the data of Vandermonde systems, and its relevance to the numerical stability of specific solution algorithms, will not be considered in this paper. For an interesting discussion of this we refer to the work of N.J. Higham [8] for ordinary Vandermonde systems and [9] for Vandermonde-like systems.

2 Singular value decomposition: an instance of computational deception

In studying condition numbers of optimally scaled Vandermonde matrices, using the Frobenius norm, we had occasion to employ the singular value decomposition to compute

$$\operatorname{cond}_F(\boldsymbol{V}) = \sqrt{\sum_{\nu=1}^{n} \sigma_\nu^2 \sum_{\nu=1}^{n} \sigma_\nu^{-2}}, \tag{1}$$

where σ_ν are the singular values of the matrix $\boldsymbol{V} \in \mathbb{R}^{n\times n}$. In the course of these computations, we observed substantial discrepancies between computed results and those expected from theory, although the patterns of the former, at first sight, seemed rather reasonable. The trouble can be traced to an unreliable behavior in the case of highly ill-conditioned matrices of the singular value decomposition, unless executed in appropriately high precision.

Consider, for example, the Vandermonde matrix

$$\boldsymbol{V}(\boldsymbol{x}) = [v_{\nu,\mu}] \in \mathbb{R}^{n\times n}, \quad v_{\nu,\mu} = x_\mu^{\nu-1}, \tag{2}$$

where $n = 50$, and $x_\mu = 4(n+1-\mu)/n$, $\mu = 1, 2, \ldots, n$. Row-wise scaling, $\boldsymbol{D}\boldsymbol{V}(\boldsymbol{x})$, with $\boldsymbol{D} = \operatorname{diag}(1, R^{-1}, \ldots, R^{-(n-1)})$, when $R = x_1$, produces an equilibrated matrix $\boldsymbol{V}_{\mathrm{sc}}(R)$ (again a Vandermonde matrix), each row having the same ∞-norm

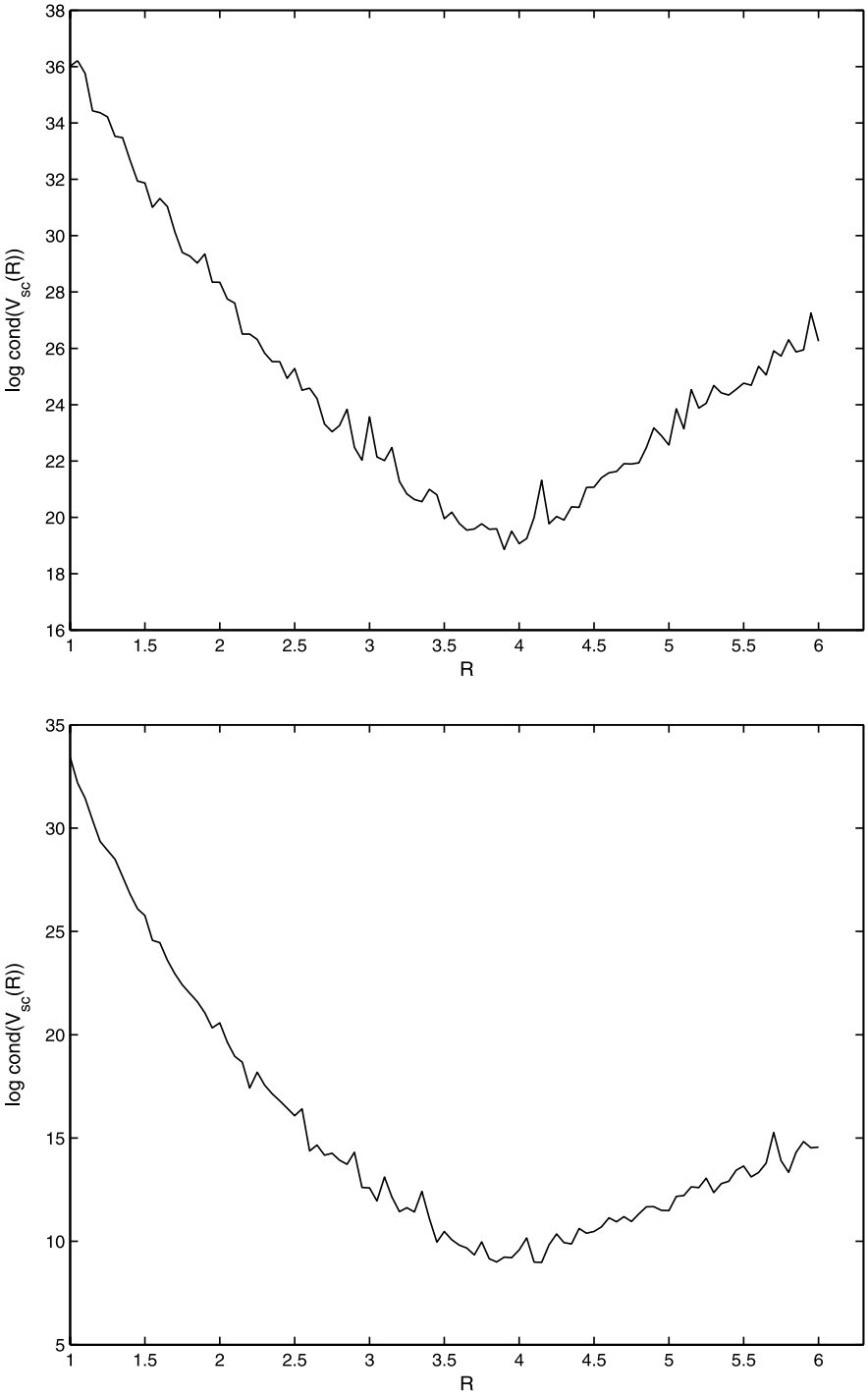


Fig. 1 Frobenius condition of scaled Vandermonde matrices (purported)

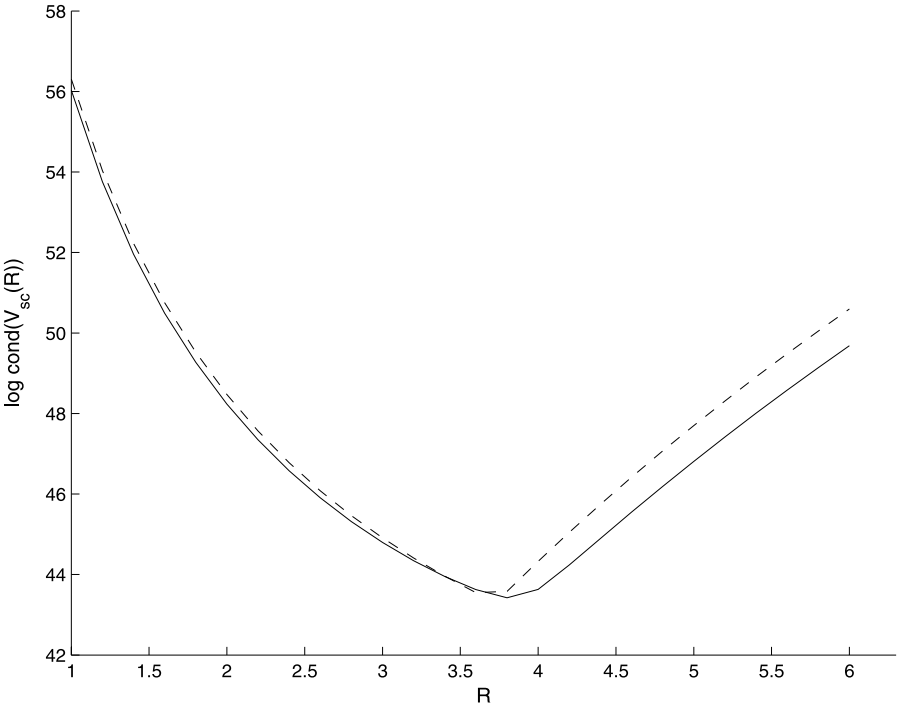


Fig. 2 Condition of scaled Vandermonde matrices (correct) *solid line*: using $\|\cdot\|_F$; *dashed line*: using $\|\cdot\|_\infty$

(= 1). This in turn minimizes the condition(-like) number $\max_{\nu,\mu}|a_{\nu,\mu}|\cdot\|\boldsymbol{A}^{-1}\|^*$, where $\|\cdot\|^*$ denotes any p-norm, or the Frobenius norm, of a matrix [13, Theorem 2.5(b)]. Wanting to illustrate this, we computed the Frobenius condition number of $\boldsymbol{V}_{\rm sc}(R)$, using the Matlab singular-value-decomposition routine `svd.m`, both in ordinary double-precision arithmetic and in symbolic/variable-precision arithmetic (with `digits` = 16), for $R = 1 : .05 : 6$. The results are depicted respectively on the top and bottom of Fig. 1. (The numerical data for the figure on the top are generated by the routine `Fig1a.m`[2], using `VdMsc.m`, and for the figure on the bottom by `sFig1b.m`, using `sVdMsc.m` with the test for complex singular values commented out. The curves themselves are produced respectively by the routines `plotFig1a.m` and `plotFig1b.m`.) Both figures look reasonable, showing a minimum near $R = 4$ ($= x_1$), as expected, but the graph on the top is higher by 1–10 decimal orders than the one on the right, and both exhibit suspicious wiggles. Which one is more trustworthy?

The plain answer is: neither. The correct graph is shown in Fig. 2, another 20 decimal orders higher! Note that, according to this graph, the condition number of the matrix $\boldsymbol{V}$ itself ($R = 1$) is of the order 10^{56}.

[2]All Matlab routines (other than the standard ones) referred to in this paper can be found on the website http://www.cs.purdue.edu/archives/2002/wxg/codes/OCVdM.html.

How can these enormous discrepancies be explained? We discovered, in the case of the symbolic routine `svd.m`, that quite a few of the singular values computed are complex, in fact purely imaginary, some with rather large imaginary parts. For these imaginary σ_ν, the squares σ_ν^2 and σ_ν^{-2} are real again, but negative, and therefore in (1) have the effect of substantially lowering the Frobenius condition number or even making it complex. In the double-precision routine `svd.m`, all computed singular values appear to be positive, and it is not entirely clear what mechanism it is that caused the substantial underestimation of the condition number in that case.

Why do we think that the solid graph in Fig. 2 is indeed the correct one? First of all, we made sure that all computed singular values σ_ν are positive. This required us to carry out the computation in as much as 112-digit arithmetic. Secondly, the ∞-condition number

$$\operatorname{cond}_\infty(\boldsymbol{V}) = \|\boldsymbol{V}\|_\infty \|\boldsymbol{V}^{-1}\|_\infty \tag{3}$$

of a Vandermonde matrix with nonnegative node vector can be computed directly, without requiring matrix inversion (cf. [2, (4.1)]),

$$\operatorname{cond}_\infty(\boldsymbol{V}_n(\boldsymbol{x})) = \max_{1\le\nu\le n} \sum_{\mu=1}^{n} x_\mu^{\nu-1} \cdot \max_{1\le\nu\le n} \prod_{\substack{\mu=1\\ \mu\ne\nu}}^{n} \frac{1+x_\mu}{|x_\nu - x_\mu|}, \quad \boldsymbol{x} \ge \boldsymbol{0}. \tag{4}$$

(This is implemented in the Matlab routine `condVp.m`.) Using the ∞-norm instead of the Frobenius norm, we obtain the dashed curve in Fig. 2, which indeed is very much within the expected range from the curve for the Frobenius norm. (The routines that produced the numerical data for the two curves are `sFig2a.m` and `Fig2b.m`, the former using the routine `sVdMsc.m` and the latter using the routine `condVsc.m`. The curves themselves are produced by the routine `plotFig2.m`.)

The exact location of the minimum, together with the minimum value, can be determined with the help of the routine `fminbnd.m` of the Matlab Optimization Toolbox. Minimizing the functions defined by the routines `fFig2a.m` and `fFig2b.m`, one finds

$$\min_R \log_{10} \operatorname{cond}_F(\boldsymbol{V}_{\mathrm{sc}}(R)) = 43.416 \quad \text{attained at } R = 3.8340,$$
$$\min_R \log_{10} \operatorname{cond}_\infty(\boldsymbol{V}_{\mathrm{sc}}(R)) = 43.305 \quad \text{attained at } R = 3.7277.$$

Thus, while optimal scaling of the matrix $\boldsymbol{V}(\boldsymbol{x})$ reduces its condition number by more than ten decimal orders, the scaled matrix produced is still very much ill-conditioned. This is a pervasive phenomenon for Vandermonde matrices of this type, since any Vandermonde matrix of order n with nonnegative nodes cannot have an ∞-condition number less than [6, Theorem 2.2]

$$(n-1)\left\{1 + \left(1 - \frac{1}{n}\right)^{-1/(n-1)}\right\}^{n-1}, \tag{5}$$

which for $n = 50$ is $2.786\ldots \times 10^{16}$.

3 Equilibrated Vandermonde matrices

As already mentioned in Sect. 2, equilibration of a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ with respect to the infinity vector norm, i.e., scaling it in such a way that all rows have the same infinity norm, has the effect of minimizing a condition(-like) number of the matrix. Similar facts hold for other vector norms. Thus, in the case of the ℓ_1-norm, equilibration minimizes the ∞-condition number $\mathrm{cond}_\infty(\boldsymbol{A}) = \|\boldsymbol{A}\|_\infty \|\boldsymbol{A}^{-1}\|_\infty$ [13, Theorem 2.5(a)], while in the ℓ_2-norm, it approximately minimizes $\mathrm{cond}(\boldsymbol{A}) = \|\boldsymbol{A}\|_2 \|\boldsymbol{A}^{-1}\|^*$ as well as $\mathrm{cond}(\boldsymbol{A}) = \|\boldsymbol{A}\|_F \|\boldsymbol{A}^{-1}\|^*$ within a factor $\sqrt{n}$, where $\|\cdot\|^*$ is any p-norm or the Frobenius norm (*ibid.*, Theorem 3.5(a)). In the case of Vandermonde matrices $\boldsymbol{A} = \boldsymbol{V}$ and the ℓ_p-norm, the scaled matrices, of course, are no longer Vandermonde matrices, unless $p = \infty$. To compute condition numbers, therefore, requires matrix inversion.

Since we are dealing with highly ill-conditioned matrices, we will use the singular value decomposition $\boldsymbol{V} = \boldsymbol{U}_\ell \boldsymbol{\Sigma} \boldsymbol{U}_r^T$ of the matrix $\boldsymbol{V}$, in combination with the computational precautions, in particular variable-(high)precision arithmetic, mentioned in Sect. 2, to obtain

$$\boldsymbol{V}^{-1} = \boldsymbol{U}_r \boldsymbol{\Sigma}^{-1} \boldsymbol{U}_\ell^T, \quad \boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n), \quad \boldsymbol{U}_r \text{ and } \boldsymbol{U}_\ell \text{ orthogonal}, \tag{6}$$

where

$$\sigma_1 > \sigma_2 > \cdots > \sigma_n > 0 \tag{7}$$

are the singular values of $\boldsymbol{V}$. In the examples to be presented, we shall display condition numbers consistently in the Frobenius norm, computed in variable-precision arithmetic from (1), and implemented in the Matlab routine `sVdMsc.m`.

We begin with the example discussed in Sect. 2.

Example 3.1 The Vandermonde matrix (2) with

$$x_\mu = 4(n+1-\mu)/n, \quad \mu = 1, 2, \ldots, n. \tag{8}$$

We compute the condition numbers of the equilibrated matrices relative to the ℓ_p-norms for $p = 1, 2, \infty$, display them as a function of n for $n \le 50$ and compare them with the condition number of the original matrix. The results obtained, with the routine `sVdMsc1.m`, in 96-digit arithmetic, are depicted in Fig. 3; cf. Matlab routine `plotFig3_1.m`. It can be seen that equilibration in this case has a notable effect of reducing not only the magnitude of the condition number, but also its rate of growth. Yet, the improved condition is still too high for most applications in practice. It is also evident that the choice of vector norm in equilibration plays a relatively minor role, as the three respective graphs are almost indistinguishable.

Example 3.2 The Vandermonde matrix (2) with

$$x_\mu = \frac{n}{2}\left(-1 + 2\frac{\mu-1}{n-1}\right), \quad \mu = 1, 2, \ldots, n. \tag{9}$$

This is an example of symmetric, and increasingly spread out, nodes. Graphs analogous to those in Example 3.1, but for $n \le 20$, and computed with `sVdMsc2.m` in

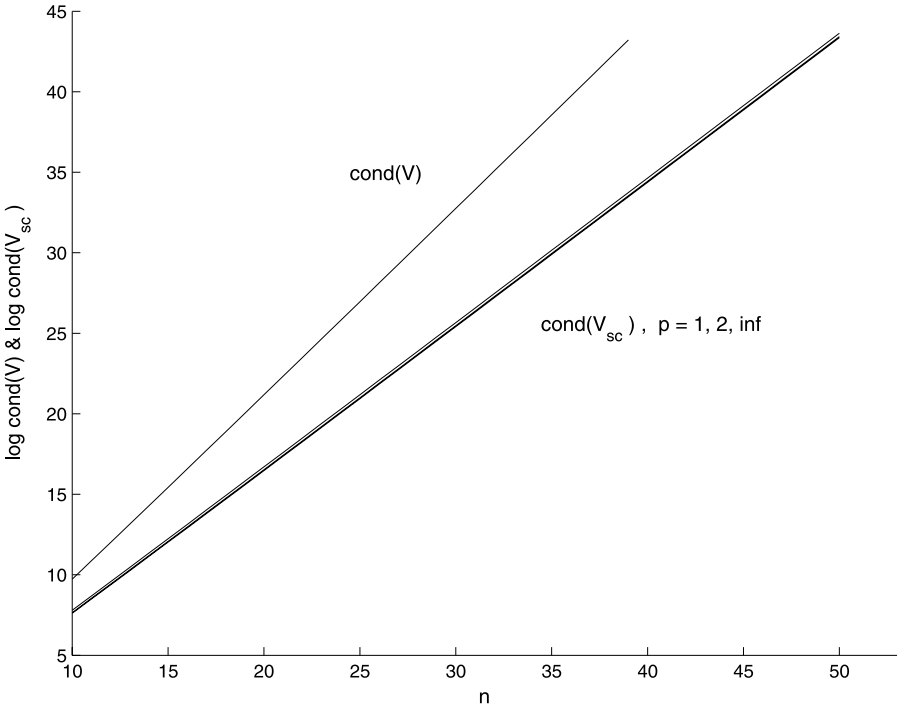


Fig. 3 Frobenius condition numbers of equilibrated and original matrix in Example 3.1

48-digit arithmetic, are shown in Fig. 4; cf. `plotFig3_2.m`. We have a behavior of the condition number of equilibrated matrices vs that of the original matrix which is similar to the one in Fig. 2, but showing markedly superior improvement of conditioning.

Example 3.3 Vandermonde matrix with Chebyshev nodes,

$$x_\mu = \cos\frac{2\mu - 1}{2n}\pi, \quad \mu = 1, 2, \ldots, n. \tag{10}$$

Here the graphs in Fig. 5, produced with `sVdMsc3.m` and `plotFig3_3.m` in 48-digit arithmetic, differ from the preceding graphs in a startling way: the graph for cond($\boldsymbol{V}$) seems to have disappeared! In fact, however, it merged with the other graphs for the equilibrated matrices, being practically identical with them. Equilibration, which, as we know, optimizes the condition of the matrix in one sense or another, has little effect in this case, which means that the matrix is already close to optimally conditioned. We will say more about this in Sect. 4.3.

We recall from [2, (6.5)] that the ∞-condition number of the $n \times n$ Vandermonde matrix with Chebyshev nodes has the asymptotic behavior

$$\mathrm{cond}_\infty(\boldsymbol{V}) \sim \frac{3^{3/4}}{4}(1+\sqrt{2})^n, \quad n \to \infty, \tag{11}$$

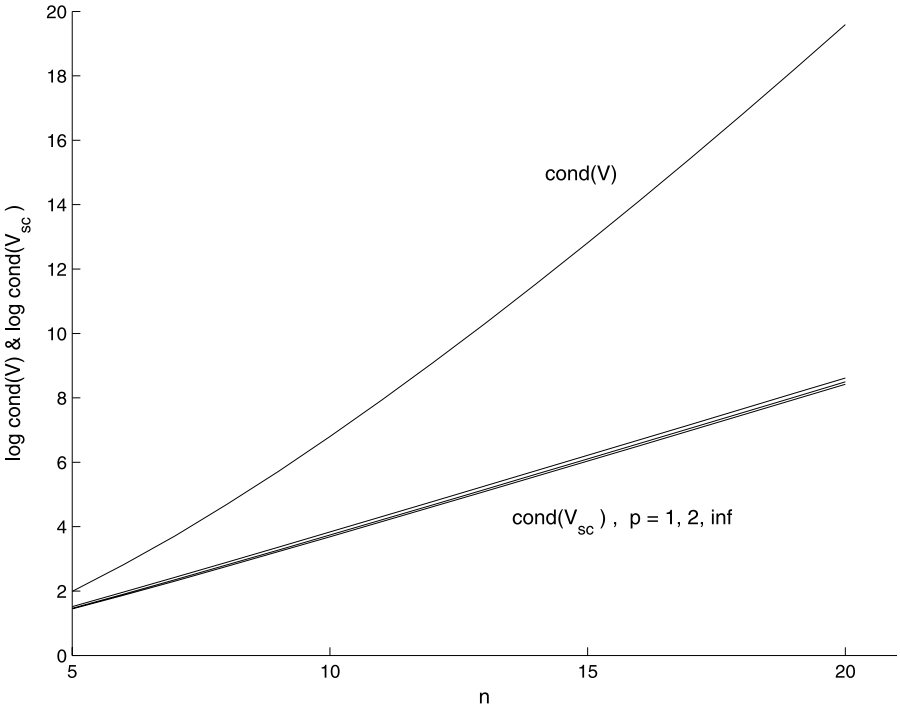


Fig. 4 Frobenius condition numbers of equilibrated and original matrix in Example 3.2

i.e., has exponential growth with rate $(1+\sqrt{2})^n$, which is confirmed in Fig. 5.

It may be worth noting that stretching or shrinking the nodes in (10) (by multiplying them by a constant $a > 1$ resp. $a < 1$) worsens the condition of $\boldsymbol{V}$, but leaves the condition numbers of the scaled matrices unchanged.

Minimality properties of condition numbers similar to those mentioned in Sect. 2 and at the beginning of this section hold also for column equilibration. Thus, the condition(-like) number defined in Sect. 2 is minimized by column equilibration in the ℓ_∞-norm [13, Theorem 2.5(c)], the condition number $\|\boldsymbol{A}\|_1\|\boldsymbol{A}^{-1}\|^*$ (for $\|\cdot\|^*$, see the first paragraph of this section) by column equilibration in the ℓ_1-norm (*ibid.*, Theorem 2.5(d)), and the condition number $\|\boldsymbol{A}\|_2\|\boldsymbol{A}^{-1}\|^*$ or $\|\boldsymbol{A}\|_F\|\boldsymbol{A}^{-1}\|^*$ is minimized approximately by column equilibration in the ℓ_2-norm (*ibid.*, Theorem 3.5(b)).

One could be tempted to equilibrate a matrix twice in succession, first by rows and then by columns, in an attempt to further optimize the condition of the matrix. However, since the first equilibration already minimizes a condition number of sorts, one cannot expect the second one to do more than reduce the condition number by an additional small amount, if at all. This has been confirmed for the Frobenius condition number in the Examples 3.1 and 3.2, where the improvement of the condition by the additional column equilibration is never more than about one decimal order and often much less.

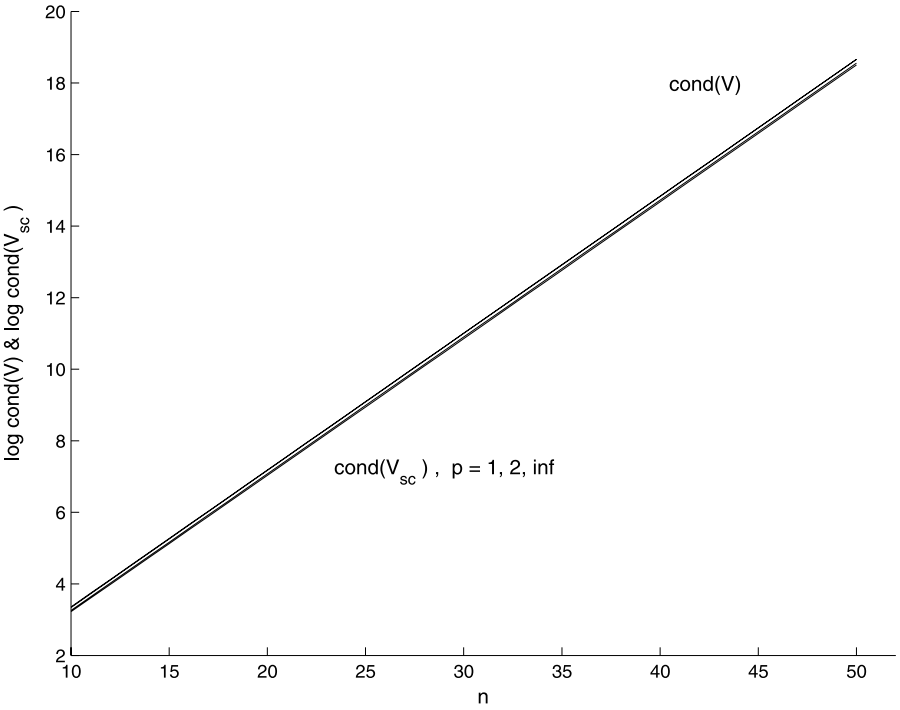


Fig. 5 Frobenius condition numbers of equilibrated and original matrix in Example 3.3

4 Optimally conditioned Vandermonde matrices

The word "optimal", in what follows, is to be understood as meaning "locally optimal". There is no easy way of establishing global optimality, although the possibility of there existing only one optimal point cannot be dismissed entirely.

4.1 Vandermonde matrices with nonnegative nodes

The problem of minimizing the infinity condition number (4) over all nonnegative node configurations

$$x_1 > x_2 > x_3 > \cdots > x_n \geq 0 \tag{12}$$

has been considered in [3, Sect. 5] and solved analytically for $n \leq 3$. For arbitrary n, it was shown (*ibid.*, Theorem 5.3) that the optimal node vector $\boldsymbol{x}^{\text{opt}}$ satisfies $x_n^{\text{opt}} = 0$, and moreover (*ibid.*, Theorem 5.2),

$$\sum_{\nu=1}^{n} (x_\nu^{\text{opt}})^{n-1} = n. \tag{13}$$

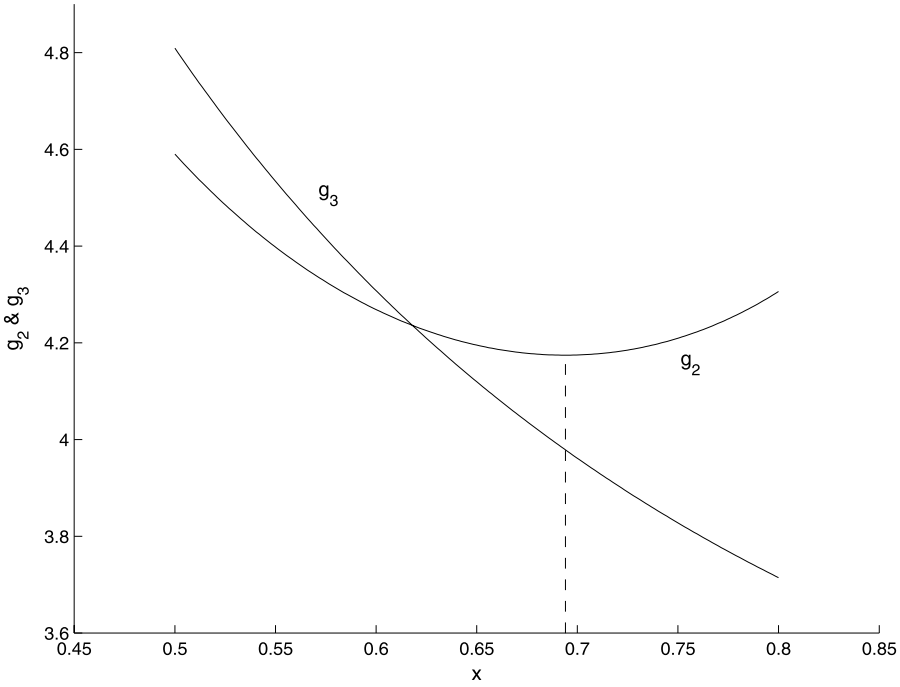


Fig. 6 The graphs of g_2 and g_3

We wish, however, to point out that the analytic solution for $n = 3$ given in the cited reference is slightly in error. The problem, in this case, boils down to solving

$$\max_{x_1>x_2>0}[g_2(x_1, x_2), g_3(x_1, x_2)] = \min$$

subject to $x_1^2 + x_2^2 = 3$, where [3, top of p. 12]

$$g_2(x_1, x_2) = \frac{1 + x_1}{x_2(x_1 - x_2)}, \qquad g_3(x_1, x_2) = \frac{(1 + x_1)(1 + x_2)}{x_1 x_2}.$$

Taking $x = x_2$ as the independent variable, one must solve the one-dimensional minimax problem

$$\max[g_2(\sqrt{3 - x^2}, x), g_3(\sqrt{3 - x^2}, x)] = \min \tag{14}$$

subject to $0 < x < \sqrt{3/2}$. The two graphs for g_2 and g_3 are shown in Fig. 6 in the critical portion of the interval $[0, \sqrt{3/2}]$. In [3], we erroneously assumed that the point of intersection of the two curves yields the minimum point of (14), whereas Fig. 6 shows that the minimum point occurs on the graph for g_2 somewhat to the right of the intersection. By elementary, but tedious, calculations one finds that the correct minimum point is located at the unique positive root of the equation

$$2x^6 - 13x^4 + 6x^3 + 39x^2 - 18 = 0,$$

Table 1 Optimal ∞-condition number and lower bound for nonnegative nodes

n	$\min(\mathrm{cond}_\infty(\boldsymbol{V}p))$	$\mathrm{lb}(\mathrm{cond}_\infty(\boldsymbol{V}p))$
2	3.0000(0)	3.0000(0)
3	1.2524(1)	9.8990(0)
4	7.1725(1)	2.7809(1)
5	3.6102(2)	7.1666(1)
6	1.8844(3)	1.7542(2)
7	9.9422(3)	4.1497(2)
8	5.3511(4)	9.5817(2)
9	2.9009(5)	2.1727(3)
10	1.5997(6)	4.8580(3)

which is x_2^{opt}, and for $\boldsymbol{x}^{\mathrm{opt}}$ yields

$$\begin{aligned} x_1^{\mathrm{opt}} &= 1.586908974551119, \\ x_2^{\mathrm{opt}} &= 0.6940604487284352, \\ x_3^{\mathrm{opt}} &= 0, \end{aligned} \tag{15}$$

giving

$$\mathrm{cond}_\infty(\boldsymbol{V}(\boldsymbol{x}^{\mathrm{opt}})) = 12.52354612417062 \quad (n = 3). \tag{16}$$

To go beyond order 3, we use the routine `fmincon.m` of the Matlab Optimization Toolbox to determine the optimal condition number (4) subject to (12) for $n \le 10$. This is implemented in the routine `optcondVp.m` using `condVp.m`. The required initial approximation $\boldsymbol{x}0$ to $\boldsymbol{x}^{\mathrm{opt}}$ for each n is found by extrapolating the $\boldsymbol{x}^{\mathrm{opt}}$ for the preceding values of n. The results are then checked against the identity (13). If there is insufficient agreement, the initial approximation is improved on the basis of the approximation of $\boldsymbol{x}^{\mathrm{opt}}$ currently at hand, and the routine is run again. This is repeated until sufficient agreement is achieved. A summary of the results so obtained, to five significant digits, is shown in Table 1. The last column, for comparison, lists the lower bound of the condition number according to (5). The respective optimal nodes x_ν^{opt} can be found on the website cited in footnote 2 in the file `xoptVp`.

4.2 Vandermonde matrices with symmetric nodes

If there is a unique (up to permutation of the nodes) optimally conditioned Vandermonde matrix in the ∞-matrix norm, its nodes must be distributed symmetrically with respect to the origin [3, Theorem 3.1],

$$x_1 > x_2 > \cdots > x_n, \qquad x_\nu + x_{n+1-\nu} = 0, \quad \nu = 1, 2, \ldots, n. \tag{17}$$

We therefore turn now our attention to this case of symmetry. It suffices here to consider the reduced vector

$$\boldsymbol{x}s = [x_1, x_2, \ldots, x_{\lfloor (n+1)/2 \rfloor}]^T \ge \boldsymbol{0}, \tag{18}$$

Table 2 Optimal ∞-condition numbers, those for Chebyshev nodes, and lower bounds for symmetric nodes

n	$\min(\text{cond}_\infty(\boldsymbol{Vs}))$	$\text{cond}_\infty(\boldsymbol{VCh})$	$\text{lb}(\text{cond}_\infty(\boldsymbol{Vs}))$
2	2.0000(0)	2.4142(0)	2.0000(0)
3	5.0000(0)	7.0000(0)	2.8284(0)
4	1.1776(1)	1.8942(1)	4.0000(0)
5	2.1456(1)	4.1000(1)	5.6569(0)
6	5.1330(1)	1.1282(2)	8.0000(0)
7	1.1060(2)	2.5984(2)	1.1314(1)
8	2.4222(2)	6.5152(2)	1.6000(1)
9	5.4541(2)	1.5727(3)	2.2627(1)
10	1.2282(3)	3.7495(3)	3.2000(1)

where $x_{\lfloor (n+1)/2 \rfloor} = 0$ whenever n is odd. In this case the ∞-condition number (3) of $\boldsymbol{V}$ is given explicitly by [2, Theorem 4.3]

$$\text{cond}_\infty(\boldsymbol{V}_n(\boldsymbol{x})) = \max_\nu \sum_\mu x_\mu^{\nu-1} \cdot \begin{cases} \max_\nu [(1+\frac{1}{x_\nu}) \prod_{\mu\neq\nu} \frac{1+x_\mu^2}{|x_\nu^2 - x_\mu^2|}], & n \text{ even}, \\ 2\max_\nu [\varepsilon_\nu (1+x_\nu) \prod_{\mu\neq\nu} \frac{1+x_\mu^2}{|x_\nu^2 - x_\mu^2|}], & n \text{ odd}, \end{cases} \tag{19}$$

where ν and μ vary over all integers for which $x_\nu \geq 0$ and $x_\mu \geq 0$, respectively, and where $\varepsilon_\nu = \frac{1}{2}$ when $x_\nu > 0$, and $\varepsilon_\nu = 1$ when $x_\nu = 0$. (This is implemented in the Matlab routine `condVs.m`.)

The problem of finding the optimal $\boldsymbol{xs}$ has been solved analytically in [3, Sect. 4] for $n \leq 6$. Here we use `fmincon.m`, in `optcondVs.m`, to successfully reproduce these results and to extend them to $n = 10$. The procedure is similar to the one adopted in Sect. 4.1 for nonnegative nodes, the check being provided by the identity [3, Theorem 3.3]

$$\sum_{\nu=1}^{\lfloor (n+1)/2 \rfloor} (x_\nu^{\text{opt}})^{n-1} = \frac{n}{2}. \tag{20}$$

The results are summarized in the second column of Table 2. The third column shows condition numbers for Vandermonde matrices with the Chebyshev nodes (10), and the last column the lower bound $2^{n/2}$ from [6, Theorem 3.1]. Optimal conditioning is now distinctly better than in the case of nonnegative nodes.

The optimal nodes x_ν^{opt} can be found on the website cited in footnote 2 in the file `xoptVs`.

4.3 More on optimal node configurations

We have seen in Example 3.3, and confirmed in Table 2, that Chebyshev nodes are close to optimal. We therefore use them as initial approximations in the unconstrained optimization routine `fminsearch.m` to compute optimally conditioned Vandermonde matrices relative to the Frobenius norm. (There are no longer explicit formulae for the ∞-condition number.) The objective function to be supplied to `fmin-`

Table 3 Optimal Frobenius condition numbers

n	$\min(\text{cond}_F(\boldsymbol{V}))$	n	$\min(\text{cond}_F(\boldsymbol{V}))$
2	2.0000(0)	8	2.9074(2)
3	4.5109(0)	9	6.8528(2)
4	1.0156(1)	10	1.6213(3)
5	2.3101(1)	11	3.8473(3)
6	5.3238(1)	12	9.1502(3)
7	1.2397(2)	13	2.1804(4)

`search.m` is generated in the routine `condV.m`, which uses symbolic/variable-precision tools to evaluate (in the routine `sVdMsc.m`) the Frobenius condition number of a Vandermonde matrix for arbitrary real nodes. The driver routine is `optcondV.m`. We succeeded, with only 16-digit computation, to determine optimally conditioned Vandermonde matrices up to order $n = 13$. (The machine time on our SUN workstation varied from a few seconds for the first few values of n to 55 minutes for $n = 13$. The computation for $n = 14$ failed after about an hour's worth of computing, producing the cryptic error message "integer too large in context". It appears that the message was generated in the symbolic `svd.m` routine.) The resulting condition numbers are shown in Table 3; the optimal nodes are listed in the file `xoptV` on the website cited in footnote 2. They are indeed symmetric with respect to the origin, at least to within the accuracy provided by the routine `fminsearch.m`, and are not much different from those in the file `xoptVs`.

In order to probe into possible alternative extrema, we repeated the computation with the initial approximations deliberately transformed to the interval $[0, 1]$. We found that the routine `fminsearch.m` converged to exactly the same symmetric solutions as obtained previously, and didn't even take any longer. This seems to suggest that for optimal conditioning of Vandermonde matrices we have both uniqueness and symmetry.

It is evident from our computations that the condition numbers of optimally scaled and optimally conditioned Vandermonde matrices grow at an exponential rate with respect to the order n. In the present case of Frobenius-optimal conditioning it appears that the exponential law for the optimal condition number holds not only for large n, but already for $n \geq 2$. Assuming this is the case, we find numerically, since for $n = 2$ the optimal condition number is 2, that

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \text{cond}_F(\boldsymbol{V}_n(\boldsymbol{x})) \approx 2 \times (2.32)^{n-2}. \tag{21}$$

A log-plot of (21) (dashed line) together with actually computed logarithms of the optimal condition numbers (indicated by stars) is shown in Fig. 7 and produced by `plotoptV.m`. The asymptotic law (11) of the ∞-condition number in the case of Chebyshev nodes is shown as a dashdot line. The rate of growth for the optimal condition numbers is seen to be just slightly less than the one, $(1 + \sqrt{2})^n$, for the ∞-condition numbers involving Chebyshev nodes.

For a related result involving the condition number in the Euclidean 2-norm, see also [1, Theorem 4.1].

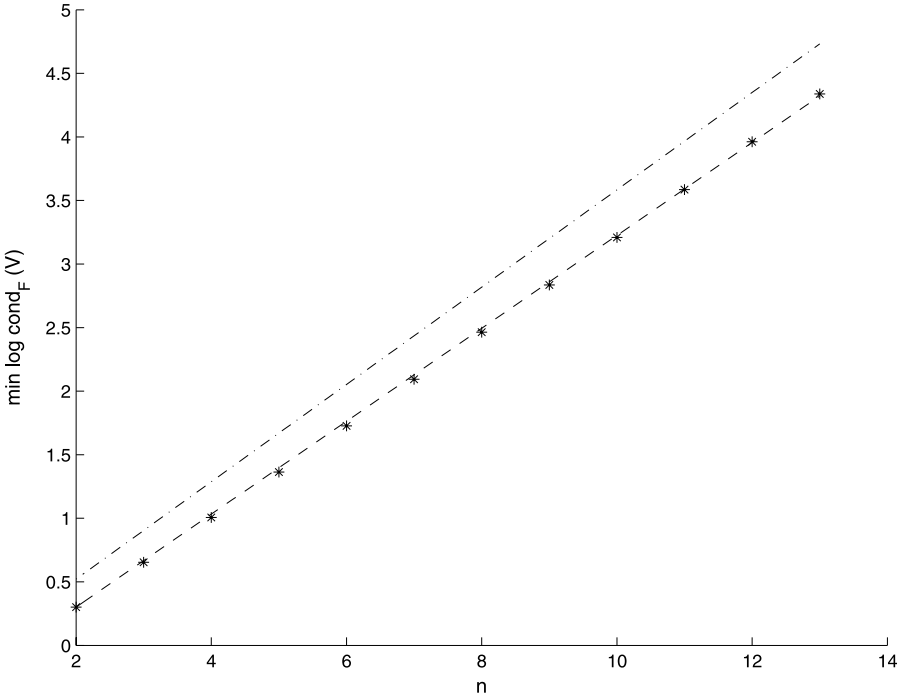


Fig. 7 Optimal Frobenius conditioning of Vandermonde matrices

5 Optimally and perfectly conditioned Vandermonde-like matrices

Vandermonde-like matrices, first considered in [4], are matrices of the form

$$\boldsymbol{W}_n = \boldsymbol{W}_n(\boldsymbol{x}) = \begin{bmatrix} p_0(x_1) & p_0(x_2) & \cdots & p_0(x_n) \\ p_1(x_1) & p_1(x_2) & \cdots & p_1(x_n) \\ p_2(x_1) & p_2(x_2) & \cdots & p_2(x_n) \\ \cdots & \cdots & \cdots & \cdots \\ p_{n-1}(x_1) & p_{n-1}(x_2) & \cdots & p_{n-1}(x_n) \end{bmatrix}, \tag{22}$$

where p_ν, $\nu = 0, 1, \ldots, n-1$, are polynomials of exact degree ν. Thus, the monomials x^ν in a Vandermonde matrix are now replaced by polynomials $p_\nu(x)$. An especially interesting example are polynomials

$$p_\nu(x) = p_\nu(x; \mathrm{d}\lambda), \quad \nu = 0, 1, 2, \ldots, \tag{23}$$

orthogonal with respect to a positive measure $\mathrm{d}\lambda$. Here, row-scaling amounts to renormalizing the orthogonal polynomials. A natural way to do this is to let p_ν in (23) be the ortho*normal* polynomials. We shall henceforth assume this to be the case unless stated otherwise.

If $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]^T$ is the vector of the zeros of $p_n(\cdot; \mathrm{d}\lambda)$—the "Gauss vector" for $\mathrm{d}\lambda$—then (cf. [4, Theorem 2.1])

$$\mathrm{cond}_F(\boldsymbol{W}_n(\boldsymbol{x})) = \sqrt{\sum_{\nu=1}^{n} \lambda_\nu \sum_{\nu=1}^{n} \frac{1}{\lambda_\nu}}, \tag{24}$$

where $\lambda_\nu = \lambda_\nu^{(n)}$, $\nu = 1, 2, \ldots, n$, are the Christoffel numbers for the measure $\mathrm{d}\lambda$.

5.1 General theory

We begin by noting that (cf. (1))

$$\mathrm{cond}_F(\boldsymbol{W}_n) = \sqrt{\sum_{\nu=1}^{n} \sigma_\nu^2 \sum_{\nu=1}^{n} \sigma_\nu^{-2}} = n\sqrt{\frac{m_A(\sigma^2)}{m_H(\sigma^2)}} \geq n, \tag{25}$$

where $m_A(\sigma^2)$ is the arithmetic mean of the quantities σ_ν^2 and $m_H(\sigma^2)$ their harmonic mean. As is well known, the former is larger than, or equal to, the latter and equal if and only if all singular values σ_ν are equal. Any matrix $\boldsymbol{W}_n$ for which equality holds in (25) is called *perfectly conditioned* with respect to the Frobenius norm.

In the case of Gauss vectors $\boldsymbol{x}$ we have similarly, from (24), that

$$\mathrm{cond}_F(\boldsymbol{W}_n(\boldsymbol{x})) = \sqrt{\sum_{\nu=1}^{n} \lambda_\nu \sum_{\nu=1}^{n} \lambda_\nu^{-1}} = n\sqrt{\frac{m_A(\lambda)}{m_H(\lambda)}} \geq n. \tag{26}$$

If equality holds in (26) for all n, then $\lambda_1^{(n)} = \lambda_2^{(n)} = \cdots = \lambda_n^{(n)}$ for all n, which in turn implies, by a classical result of Posse (cf. [5, Example 1.49]), that $\mathrm{d}\lambda$ must be the Chebyshev measure $\mathrm{d}\lambda(x) = \sqrt{1-x^2}$ on $[-1, 1]$, and there is no other measure having the same property. Thus, *Vandermonde–Chebyshev matrices $\boldsymbol{W}_n(\boldsymbol{x})$ with Gauss–Chebyshev vectors $\boldsymbol{x}$ are the only Vandermonde-like matrices that for all n are perfectly conditioned with respect to the Frobenius norm.*

Orthonormalization is a major step toward optimizing the conditioning of $\boldsymbol{W}_n$. Also, it allows us to express the Frobenius condition number, for arbitrary vector $\boldsymbol{x}$, explicitly in the particularly simple form [4, (4.4) and (4.6)]

$$c := \mathrm{cond}_F(\boldsymbol{W}_n(\boldsymbol{x})) = \sqrt{\sum_{\nu=1}^{n} \sum_{k=0}^{n-1} p_k^2(x_\nu) \cdot \int_{\mathbb{R}} \sum_{\nu=1}^{n} \ell_\nu^2(t; \boldsymbol{x}) \mathrm{d}\lambda(t)}, \tag{27}$$

where ℓ_ν are the elementary Lagrange interpolation polynomials for the nodes $x_1, x_2, \ldots, x_n$,

$$\ell_\nu(t; \boldsymbol{x}) = \prod_{\substack{\mu=1 \\ \mu \neq \nu}}^{n} \frac{t - x_\mu}{x_\nu - x_\mu}, \quad \nu = 1, 2, \ldots, n.$$

Moreover, the gradient

$$\boldsymbol{g} := \operatorname{grad} c = [g_1, g_2, \ldots, g_n]^T, \quad g_i = \frac{\partial c}{\partial x_i},$$

computes to

$$g_i = \rho \sum_{k=1}^{n-1} p_k(x_i) p_k'(x_i) + \frac{1}{\rho} \int_{\mathbb{R}} \sum_{\nu=1}^{n} \ell_\nu(t; \boldsymbol{x}) \frac{\partial \ell_\nu}{\partial x_i}(t; \boldsymbol{x}) \mathrm{d}\lambda(t), \tag{28}$$

where

$$\frac{\partial \ell_\nu}{\partial x_i}(t; \boldsymbol{x}) = \begin{cases} \frac{1}{x_\nu - x_i} \frac{t - x_\nu}{t - x_i} \ell_\nu(t; \boldsymbol{x}) & \text{if } i \neq \nu, \\ -\ell_\nu(t; \boldsymbol{x}) \sum_{k \neq \nu} \frac{1}{x_\nu - x_k} & \text{if } i = \nu, \end{cases} \tag{29}$$

and

$$\rho = \sqrt{\frac{\int_{\mathbb{R}} \sum_{\nu=1}^{n} \ell_\nu^2(t; \boldsymbol{x}) \mathrm{d}\lambda(t)}{\sum_{\nu=1}^{n} \sum_{k=0}^{n-1} p_k^2(x_\nu)}}. \tag{30}$$

Both c and $\operatorname{grad} c$ can be computed in a straightforward way, the polynomials p_k and their derivatives by the three-term recurrence relation satisfied by orthonormal polynomials, and the integrals in (27) and (28) exactly by n-point Gauss quadrature relative to the measure $\mathrm{d}\lambda$. This is implemented in the routine `condVl.m`. Using this routine as input to the Matlab optimization routine `fminunc.m` allows us to compute optimally conditioned Vandermonde-like matrices; see the routine `opt-condVl.m`.

If $\boldsymbol{x}$ is the Gauss vector for $\mathrm{d}\lambda$, the formula for g_i can be simplified by noting that the integrand in (28) is a polynomial of degree $2n - 1$, so that n-point Gauss quadrature (for the measure $\mathrm{d}\lambda$) gives exactly

$$\int_{\mathbb{R}} \sum_{\nu=1}^{n} \ell_\nu(t; \boldsymbol{x}) \frac{\partial \ell_\nu}{\partial x_i}(t; \boldsymbol{x}) \mathrm{d}\lambda(t) = \sum_{\nu=1}^{n} \lambda_\nu \frac{\partial \ell_\nu}{\partial x_i}(x_\nu; \boldsymbol{x}) = -\lambda_i \sum_{\substack{k=1 \\ k \neq i}}^{n} \frac{1}{x_i - x_k},$$

where the last equality follows from (29). Likewise, using the well-known formula (see, e.g., [12, (3.4.8)])

$$\sum_{k=0}^{n-1} p_k^2(x_\nu) = \lambda_\nu^{-1},$$

one finds

$$\rho = \sqrt{\frac{\sum_{\nu=1}^{n} \lambda_\nu}{\sum_{\nu=1}^{n} \lambda_\nu^{-1}}}. \tag{31}$$

Table 4 Frobenius condition numbers of Vandermonde–Legendre matrices with monic and normalized polynomials, and optimal condition numbers

n	Monic	Normalized	Optimal	$\|g\|_\infty$	M'time
5	2.0433(01)	5.3624(00)	5.2296(00)	7.5011(–8)	.4
10	6.7459(02)	1.1550(01)	1.1018(01)	2.6139(–4)	5
20	7.0071(05)	2.4941(01)	2.3468(01)	1.5052(–3)	930
35	2.3097(10)	4.6321(01)	4.3348(01)	2.9835(–3)	1,061
50	7.5861(14)	6.8600(01)	6.4128(01)	–	–

Therefore,

$$g_i = \rho \sum_{k=1}^{n-1} p_k(x_i)\, p_k'(x_i) - \frac{\lambda_i}{\rho} \sum_{\substack{k=1 \\ k\neq i}}^{n} \frac{1}{x_i - x_k} \quad (\boldsymbol{x} = \text{Gauss vector for } \mathrm{d}\lambda) \tag{32}$$

with ρ as given by (31).

5.2 Examples

Example 5.1 (Vandermonde–Legendre matrices) These are Vandermonde-like matrices with p_ν the Legendre polynomials. We first compare Frobenius condition numbers when the Legendre polynomials are monic with those for normalized polynomials. In both cases we take $\boldsymbol{x}$ to be the Gauss–Legendre node vector. For selected values of n, the two condition numbers are shown respectively in the second and third column of Table 4. They were computed, with identical results, in two ways, by actual scaling and by the explicit formula (24); see the routine `runVdMlsc.m`. It can be seen that normalization of the polynomials indeed reduces the conditioning dramatically, which is typical for other Jacobi polynomials as well. In the next column we show the optimal condition numbers computed (by `optcondVl.m`) as indicated above. Here the routine `fminunc` was used with, and without, gradient information. For the former, which is considerably slower, we show in the remaining columns the ∞-norm of the gradient[3] upon exiting the optimization routine, and the machine time in seconds (on our SUN workstation) expended to compute the optimal condition number. (When $n = 50$, convergence of the gradient-based routine could not be achieved within a reasonable amount of time.) Figure 8 has a graph of the optimal condition numbers together with the one (dashed line) for normalized Legendre polynomials. It is seen that the condition of Vandermonde–Legendre matrices with normalized polynomials and Gauss–Legendre vector $\boldsymbol{x}$ is nearly optimal.

[3] Another possible indicator of the accuracy is the measure of symmetry $\max_\nu |x_\nu + x_{n+1-\nu}|$ for the computed $\boldsymbol{x}^{\mathrm{opt}}$.

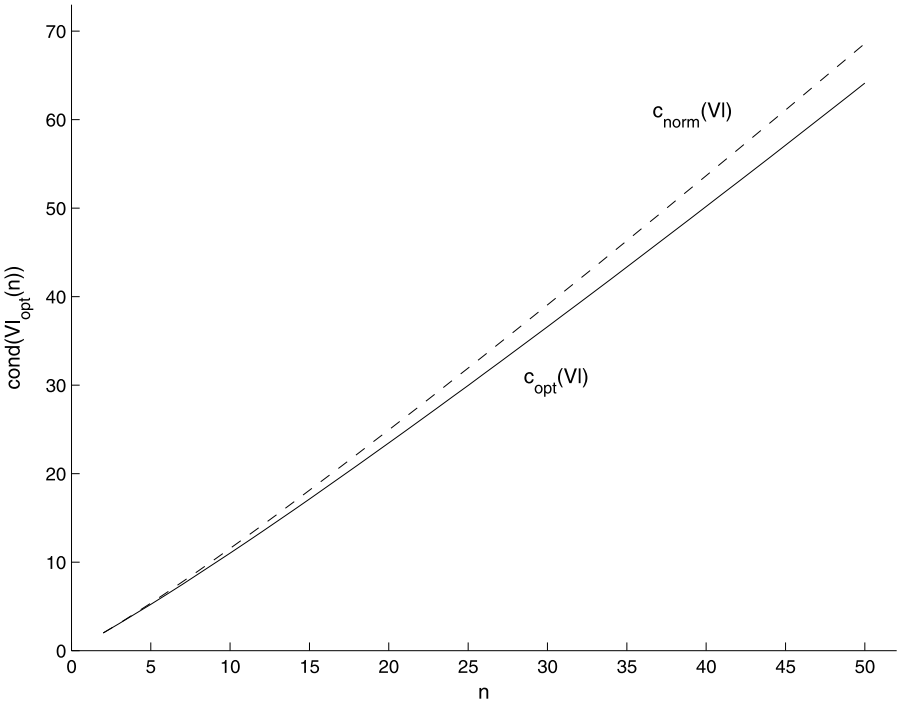


Fig. 8 Frobenius condition of optimal and normalized Vandermonde–Legendre matrices

Example 5.2 (Vandermonde–Chebyshev matrices) Here, p_ν are the Chebyshev polynomials of the first and second kind. For the former, when $\boldsymbol{x}$ is the respective Gauss–Chebyshev vector, the Vandermonde–Chebyshev matrix, as already mentioned, is perfectly conditioned,

$$\operatorname{cond}_F(\boldsymbol{W}_n(\boldsymbol{x})) = n \quad (\boldsymbol{x} = \text{first-kind Gauss–Chebyshev node vector}). \tag{33}$$

For the latter, $\lambda_\nu = \pi \sin^2(\nu\pi/(n+1))/(n+1)$, so that by (24)

$$\operatorname{cond}_F(\boldsymbol{W}_n(\boldsymbol{x})) = \sqrt{\sum_{\nu=1}^{n} \sin^2 \frac{\nu\pi}{n+1} \sum_{\nu=1}^{n} \frac{1}{\sin^2 \frac{\nu\pi}{n+1}}}$$
$$(\boldsymbol{x} = \text{second-kind Gauss–Chebyshev node vector}). \tag{34}$$

Frobenius condition numbers for monic and normalized Chebyshev polynomials, and Gauss–Chebyshev vector $\boldsymbol{x}$, as well as optimal condition numbers for normalized Chebyshev polynomials, are shown in Table 5.

Perfect conditioning (33) in the case of Chebyshev polynomials of the first kind implies zero gradient of $\boldsymbol{W}_n(\boldsymbol{x})$ at the Gauss–Chebyshev point $\boldsymbol{x}$. This can be nicely confirmed analytically. To begin with, $\lambda_\nu = \pi/n$, so that by (31), $\rho = \pi/n$. Furthermore, since $x_i = \cos\theta_i$, $\theta_i = (2i-1)\pi/2n$, one finds for $p_k(x) = \sqrt{2/\pi}T_k(x)$, $k \geq 1$,

Table 5 Frobenius condition numbers of Vandermonde–Chebyshev matrices with monic and normalized polynomials, and optimal condition numbers

n	Monic		Normalized		Optimal	
	Cheb1	Cheb2	Cheb1	Cheb2	Cheb1	Cheb2
5	1.6869(01)	2.4558(01)	5	5.9161	5	5.5446
10	5.3970(02)	8.5144(02)	10	14.832	10	12.954
20	5.5265(05)	9.0818(05)	20	39.243	20	32.408
35	1.8109(10)	3.0289(10)	35	88.148	35	70.299
50	5.9340(14)	9.9957(14)	50	148.66	50	–

that

$$\sum_{k=1}^{n-1} p_k(x_i)\,p_k'(x_i) = \frac{1}{\pi \sin\theta_i} \sum_{k=1}^{n-1} k \sin(2k\theta_i).$$

From the identity [7, (1.352.(1))]

$$\sum_{k=1}^{n-1} k \sin(2k\theta) = \frac{1}{4}\frac{\sin(2n\theta)}{\sin^2\theta} - \frac{n\cos((2n-1)\theta)}{2\sin\theta},$$

which, for $\theta = \theta_i$, yields $nx_i/(2\sqrt{1-x_i^2})$, one gets

$$\sum_{k=1}^{n-1} p_k(x_i)\,p_k'(x_i) = \frac{n}{2\pi}\frac{x_i}{1-x_i^2},$$

hence, by (32),

$$g_i = \frac{x_i}{2(1-x_i^2)} - \sum_{\substack{k=1\\k\neq i}}^{n} \frac{1}{x_i - x_k}. \tag{35}$$

Now,

$$\sum_{\substack{k=1\\k\neq i}}^{n} \frac{1}{x - x_k} = \frac{T_n'(x)}{T_n(x)} - \frac{1}{x-x_i} = \frac{(x-x_i)T_n'(x) - T_n(x)}{(x-x_i)T_n(x)},$$

which, for $x \to x_i$, by applying the rule of Bernoulli-L'Hôpital twice, yields

$$\sum_{\substack{k=1\\k\neq i}}^{n} \frac{1}{x_i - x_k} = \frac{T_n''(x_i)}{2T_n'(x_i)}.$$

On the other hand,

$$(1-x^2)T_n''(x) = xT_n'(x) - n^2 T_n(x),$$

Table 6 Frobenius condition numbers of Vandermonde–Laguerre matrices with monic and normalized polynomials, and optimal condition numbers

n	Monic	Normalized	Optimal
5	3.0362(03)	2.0757(02)	2.1067(01)
10	1.3860(11)	1.0047(06)	1.3409(03)
15	1.9857(20)	7.9047(09)	1.1749(05)
20	2.2351(30)	7.7699(13)	1.8408(08)
25	1.0948(41)	8.7177(17)	1.4895(12)

Table 7 Frobenius condition numbers of Vandermonde–Hermite matrices with monic and normalized polynomials, and optimal condition numbers

n	Monic	Normalized	Optimal
5	1.4185(01)	1.3731(1)	8.3937(0)
10	7.4990(03)	6.8318(2)	8.2754(1)
15	3.1256(07)	4.8288(4)	7.8092(2)
20	4.1560(11)	3.9886(6)	7.4769(3)
25	1.2370(16)	3.6159(8)	7.2162(4)

so that

$$\sum_{\substack{k=1\\k\neq i}}^{n} \frac{1}{x_i - x_k} = \frac{x_i T_n'(x_i)}{(1-x_i^2)\cdot 2T_n'(x_i)} = \frac{x_i}{2(1-x_i^2)},$$

giving $g_i = 0$ by (35).

Example 5.3 (Vandermonde–Laguerre matrices) Monic and normalized Laguerre polynomials and the Gauss–Laguerre node vector $\boldsymbol{x}$ give rise to Frobenius condition numbers shown in Table 6. Because of severe ill-conditioning, the second column (headed "monic") had to be computed in 48-digit arithmetic. Also, in the routine `fminunc.m` for computing optimal condition numbers the default value of 'MaxFunEvals' was increased to 5000. While normalized polynomials do lead to substantially smaller condition numbers than do monic polynomials, the matrices involved are still quite ill-conditioned. Optimal conditioning helps somewhat. A few of the optimal nodes are consistently negative.

Example 5.4 (Vandermonde–Hermite matrices) The condition of Vandermonde–Hermite matrices, with $\boldsymbol{x}$ the Gauss–Hermite vector, is roughly halfway between that for Vandermonde–Legendre and Vandermonde–Laguerre matrices, as is shown in Table 7. The optimal nodes are symmetric with respect to the origin, as in the case of Vandermonde matrices.

As a matter of curiosity, we remark that for $n = 2$, the optimally conditioned Vandermonde-like matrix $\boldsymbol{W}_2(\boldsymbol{x}; \mathrm{d}\lambda)$ is perfectly conditioned with respect to the Frobenius norm, for any (positive) measure $\mathrm{d}\lambda$. The proof, and a formula for $\boldsymbol{x}^{\mathrm{opt}} \in \mathbb{R}^2$, is given in the Appendix.

Acknowledgement The present work was motivated by recent attempts of C.-S. Liu, D.-L. Young, and C.-M. Fan [11] to alleviate ill-conditioning of Vandermonde matrices by suitable scaling.

Appendix

For $n = 2$, the Vandermonde-like matrix for the orthonormal polynomials p_0, p_1 has the form

$$\boldsymbol{W}(\boldsymbol{x}) = \begin{bmatrix} p_0 & p_0 \\ p_1(x_1) & p_1(x_2) \end{bmatrix}, \quad p_0 = \frac{1}{\sqrt{\beta_0}}, \quad \boldsymbol{x} = [x_1, x_2]^T \in \mathbb{R}^2, \tag{A.1}$$

where $p_1(x) = p_1(x; \mathrm{d}\lambda) = c_0 + c_1 x$, $x_1 > x_2$, and $\beta_0 = \int_{\mathbb{R}} \mathrm{d}\lambda(x)$. Its inverse is

$$\boldsymbol{W}^{-1}(\boldsymbol{x}) = \frac{1}{c_1(x_1 - x_2)} \begin{bmatrix} -p_1(x_2)/p_0 & 1 \\ p_1(x_1)/p_0 & -1 \end{bmatrix},$$

and hence

$$F(\boldsymbol{x}) := \mathrm{cond}_F(\boldsymbol{W}(\boldsymbol{x})) = \frac{1}{c_1 p_0 (x_1 - x_2)} [2p_0^2 + p_1^2(x_1) + p_1^2(x_2)]. \tag{A.2}$$

We have

$$\frac{\partial F}{\partial x_1} = \frac{1}{c_1 p_0} \frac{2(x_1 - x_2) p_1(x_1) c_1 - [2p_0^2 + p_1^2(x_1) + p_1^2(x_2)]}{(x_1 - x_2)^2},$$
$$\frac{\partial F}{\partial x_2} = \frac{1}{c_1 p_0} \frac{2(x_1 - x_2) p_1(x_2) c_1 + [2p_0^2 + p_1^2(x_1) + p_1^2(x_2)]}{(x_1 - x_2)^2}.$$

The extremal point $\boldsymbol{x}^{\mathrm{opt}} = [x_1, x_2]^T$ is determined by the equations

$$\begin{aligned} 2(x_1 - x_2) p_1(x_1) c_1 &= 2p_0^2 + p_1^2(x_1) + p_1^2(x_2), \\ 2(x_1 - x_2) p_1(x_2) c_1 &= -[2p_0^2 + p_1^2(x_1) + p_1^2(x_2)], \end{aligned} \tag{A.3}$$

which, subtracting and adding, are equivalent to

$$\begin{aligned} (x_1 - x_2)^2 c_1^2 &= 2p_0^2 + p_1^2(x_1) + p_1^2(x_2), \\ p_1(x_1) + p_1(x_2) &= 0. \end{aligned}$$

The latter equation yields immediately

$$x_1 + x_2 = -2\frac{c_0}{c_1}, \tag{A.4}$$

which, inserted in the former equation, gives, after a little computation,

$$(c_0 + c_1 x_1)^2 = p_0^2,$$

hence

$$x_1 = \frac{\pm p_0 - c_0}{c_1}. \tag{A.5}$$

At the extremal point $\boldsymbol{x}^{\text{opt}}$, by (A.2) and the first equation in (A.3), we get

$$F(\boldsymbol{x}^{\text{opt}}) = \frac{2(x_1 - x_2)p_1(x_1)c_1}{c_1 p_0(x_1 - x_2)} = 2\frac{p_1(x_1)}{p_0}.$$

Since, by (A.5), $p_1(x_1) = c_0 + c_1(\pm p_0 - c_0)/c_1 = \pm p_0$, we must take the plus sign, giving

$$F(\boldsymbol{x}^{\text{opt}}) = 2,$$

as was to be shown. Moreover, from (A.4) and (A.5),

$$x_1^{\text{opt}} = \frac{p_0 - c_0}{c_1}, \qquad x_2^{\text{opt}} = -\frac{p_0 + c_0}{c_1}. \tag{A.6}$$

We may express this more conveniently in terms of the moments μ_k of the measure $\mathrm{d}\lambda$,

$$\mu_k = \int_{\mathbb{R}} x^k \mathrm{d}\lambda(x), \quad k = 0, 1, 2, \ldots.$$

By orthogonality of p_1, we find

$$c_1 = \frac{-\mu_0}{\mu_1} c_0, \tag{A.7}$$

and by orthonormality, after some manipulation,

$$c_0 = \frac{-\mu_1}{\sqrt{\mu_0(\mu_0\mu_2 - \mu_1^2)}}, \tag{A.8}$$

where the radicand is positive by Schwarz's inequality. Insertion into (A.7) yields

$$c_1 = \sqrt{\frac{\mu_0}{\mu_0\mu_2 - \mu_1^2}}. \tag{A.9}$$

Substituting (A.8) and (A.9) in (A.6) and noting that $p_0 = 1/\sqrt{\mu_0}$ finally gives

$$x_1^{\text{opt}} = \frac{\mu_1 + \sqrt{\mu_0\mu_2 - \mu_1^2}}{\mu_0}, \qquad x_2^{\text{opt}} = \frac{\mu_1 - \sqrt{\mu_0\mu_2 - \mu_1^2}}{\mu_0}. \tag{A.10}$$

As expected, for symmetric measures ($\mu_1 = 0$), the optimal point $\boldsymbol{x}^{\text{opt}}$ is also symmetric with respect to the origin.

References

1. Beckermann, B.: The condition number of real Vandermonde, Krylov and positive definite Hankel matrices. Numer. Math. **85**(4), 553–577 (2000)
2. Gautschi, W.: Norm estimates for inverses of Vandermonde matrices. Numer. Math. **23**(4), 337–347 (1975)
3. Gautschi, W.: Optimally conditioned Vandermonde matrices. Numer. Math. **24**(1), 1–12 (1975)
4. Gautschi, W.: The condition of Vandermonde-like matrices involving orthogonal polynomials. Linear Algebra Appl. **52/53**, 293–300 (1983)
5. Gautschi, W.: Orthogonal Polynomials: Computation and Approximation. Oxford University Press, Oxford (2004)
6. Gautschi, W., Gabriele, I.: Lower bounds for the condition number of Vandermonde matrices. Numer. Math. **52**(3), 241–250 (1988)
7. Gradshteyn, I.S., Ryzhik, I.M.: Tables of Integrals, Series, and Products, 7th edn. Elsevier/Academic Press, Amsterdam (2007)
8. Higham, N.J.: Error analysis of the Björck–Pereyra algorithms. Numer. Math. **50**(5), 613–632 (1987)
9. Higham, N.J.: Fast solution of Vandermonde-like systems involving orthogonal polynomials. IMA J. Numer. Anal. **8**(4), 473–486 (1988)
10. Lebesgue, H.: L'œuvre mathématique de Vandermonde. Enseign. Math. (2) **1**, 203–223 (1956) [Originally published in 1940.]
11. Liu, C.-S., Young, D.-L., Fan, C.-M.: A highly accurate multi-scale polynomial interpolation by reducing the condition numbers of Vandermonde matrices. Manuscript
12. Szegö, G.: Orthogonal Polynomials. Colloquium Publ., vol. 23. Am. Math. Soc., Providence (1975)
13. van der Sluis, A.: Condition numbers and equilibration of matrices. Numer. Math. **14**(1), 14–23 (1969)