Numer. Math. 24, 1—12 (1975)

Optimally Conditioned Vandermonde Matrices*

Walter Gautschi

Received June 20, 1974

Summary. We discuss the problem of selecting the points in an $(n\times n)$ Vandermonde matrix so as to minimize the condition number of the matrix. We give numerical answers for $2\leqq n\leqq 6$ in the case of symmetric point configurations. We also consider points on the non-negative real line, and give numerical results for $n=2$ and $n=3$. For general n, the problem can be formulated as a nonlinear minimax problem with nonlinear constraints or, equivalently, as a nonlinear programming problem.

1. Introduction

We define the Vandermonde matrix of order n by

$$V_n=\begin{bmatrix} 1 & 1 & \dots 1 \\ x_1 & x_2 & \dots x_n \\ \dots & \dots & \dots \\ x_1^{n-1} & x_2^{n-1} & \dots x_n^{n-1}\end{bmatrix},\qquad n>1, \tag{1.1}$$

where the x_ν are real or complex numbers, called "points" or "nodes". We write $V_n(x)$ if we wish to indicate the dependence of V_n on $x=[x_1, x_2, \dots, x_n]$. Systems of linear algebraic equations, whose coefficient matrix is a Vandermonde matrix, or its transposed, occur frequently in numerical analysis, e.g., in polynomial interpolation and in the approximation of linear functionals [1]. Thus, if the polynomial $p(f;x)=\sum_{j=1}^{n} p_j x^{j-1}$ interpolates to function values f_i at n points x_i, then the vector π of the coefficients p_j is related to the vector ϕ of the function values f_i by

$$V_n^T(x)\,\pi=\phi. \tag{1.2}$$

A linear functional L, having moments $m_i=Lx^{i-1}$, $1\leqq i\leqq n$, in turn may be approximated by L^* through $L^*f=Lp(f;x)$. Then $L^*f=\phi^T\varrho$, where

$$V_n(x)\varrho=\mu, \tag{1.3}$$

and μ is the vector of the moments m_i.

In the following we are interested in the condition of such linear systems. We shall use the condition number

$$\operatorname{cond}_\infty V_n=\|V_n\|_\infty\,\|V_n^{-1}\|_\infty, \tag{1.4}$$

which describes the condition of (1.3) in terms of the L_∞-norm, and the condition of (1.2) in terms of the L_1-norm. In particular, we are interested in determining a real point configuration x which minimizes the condition number in (1.4). Such a point configuration will be called *optimal*, and the corresponding Vandermonde matrix (1.1) *optimally conditioned*. We first discuss this problem for sym-

* Paper presented at the GATLINBURG VI Symposium on Numerical Algebra in Munich, December 15–22, 1974 (sponsored by the Stifterverband für die Deutsche Wissenschaft through the German Research Council).

metric (with respect to the origin) point configurations, and give numerical results for $2 \leqq n \leqq 6$. We then consider briefly point configurations on the positive real axis, and obtain optimal ones for $n = 2$ and $n = 3$.

The problem, in both cases, can be formulated either as a nonlinear minimax problem with nonlinear constraints, or, equivalently, as a nonlinear programming problem. The solution of these problems, beyond the values of n treated here, almost certainly requires approximate methods. We shall not attempt to use such methods here, but suggest that interesting test problems for constrained optimization codes may be found below in Sections 3 and 5.

In terms of polynomial interpolation, the condition number (1.4) can be further interpreted as follows. Writing

$$\begin{aligned}\operatorname{cond}_\infty V_n &= \|(V_n^T)^{-1}\|_1 \|V_n^T\|_1 \\ &= \max_{\delta\phi \in \mathbb{R}^n} \frac{\|(V_n^T)^{-1}\delta\phi\|_1}{\|\delta\phi\|_1} \Big/ \min_{\phi \in \mathbb{R}^n} \frac{\|(V_n^T)^{-1}\phi\|_1}{\|\phi\|_1},\end{aligned}$$

one finds, by virtue of (1.2), that

$$\operatorname{cond}_\infty V_n = \max_{\delta f, f} \left\{ \frac{\|p(\delta f; x)\|_\pi}{\|p(f; x)\|_\pi} \Big/ \frac{\|\delta f\|_1}{\|f\|_1} \right\}, \tag{1.5}$$

where

$$\|a_1 + a_2 x + \cdots + a_n x^{n-1}\|_\pi \overset{\text{def}}{=} \|a\|_1 = \sum_{j=1}^n |a_j|.$$

Thus, $\operatorname{cond}_\infty V_n$ is equal to the maximum magnification of relative errors, if the relative error in the data is measured by $\|\delta f\|_1 / \|f\|_1$, and the relative error in the interpolation polynomial by $\|p(\delta f; x)\|_\pi / \|p(f; x)\|_\pi$.

A more common measure of condition is based on absolute errors and the maximum norm. In fact, restricting attention (without loss of generality) to the interval $[-1, 1]$, one has

$$\max_{\delta f} \frac{\|p(\delta f; x)\|_\infty}{\|\delta f\|_\infty} = \Lambda_n, \tag{1.6}$$

where $\Lambda_n = \max_{-1 \leqq x \leqq 1} \sum_{i=1}^n |l_i(x)|$ is the Lebesgue constant [$l_i(x)$ are the fundamental Lagrange interpolation polynomials corresponding to the nodes x_ν]. Optimal nodes $x_\nu \in [-1, 1]$, which minimize Λ_n, are investigated in [3], but remain unknown. (For related work, see also [4, 5].) What seems clear is that the optimal condition number according to (1.6) is substantially smaller than the optimal condition number based on (1.5). The former is known [3] to be $(2/\pi) \log n + O(1)$ as $n \to \infty$; the latter, assuming symmetric optimal nodes, behaves more like 2^n, judging from the limited numerical results available below in Section 4.

2. Condition Number of a Vandermonde Matrix

We begin with expressing the norm of a Vandermonde matrix.

Theorem 2.1. $\|V_n\|_\infty = \max\{n, \sum_{\mu=1}^n |x_\mu|^{n-1}\}$.

Proof. Let

$$v(s) = \sum_{\mu=1}^n |x_\mu|^s, \quad s > 0,$$

and define $v(0) = \lim_{s \downarrow 0} v(s) = n_0$, where n_0 is the number of nonvanishing nodes x_μ. Since $n_0 \leqq n$, and $v(s)$ is convex on $s \geqq 0$, we have

$$\|V_n\|_\infty = \max\{n, \max_{0 \leqq \nu \leqq n-1} v(\nu)\} = \max\{n, v(n-1)\},$$

proving Theorem 2.1.

It is clear that a permutation of the variables $x_1, x_2, \ldots, x_n$ does not affect the value of $\|V_n(x)\|_\infty$. The same is true for $\|V_n^{-1}(x)\|_\infty$ whenever V_n^{-1} exists. We can assume, therefore, that the variables are arranged in some fixed order. If they are all real we shall assume them in decreasing order,

$$x_1 > x_2 > \cdots > x_n. \tag{2.1}$$

The monic polynomial whose zeros are the nodes x_ν, will be denoted by p_n,

$$p_n(x) = \prod_{\nu=1}^{n} (x - x_\nu). \tag{2.2}$$

We recall from [2] the following results.

Theorem 2.2. *If the nodes x_ν in* (2.1) *are located symmetrically with respect to the origin,*

$$x_\nu + x_{n+1-\nu} = 0, \qquad \nu = 1, 2, \ldots, n, \tag{2.3}$$

then

$$\|V_n^{-1}\|_\infty = \max_{1 \leqq \nu \leqq \left[\frac{n+1}{2}\right]} \left\{ \frac{|p_n(i)|}{\frac{1+x_\nu^2}{1+x_\nu} |p_n'(x_\nu)|} \right\}, \qquad i = \sqrt{-1}, \tag{2.4}$$

where $p_n(x)$ is the polynomial in (2.2).

We remark that (2.1) and (2.3) imply that $x_\nu > 0$ for $\nu = 1, 2, \ldots, [n/2]$, and $x_{(n+1)/2} = 0$ if n is odd.

Theorem 2.3. *If the nodes in* (2.1) *are nonnegative, i.e., $x_n \geqq 0$, then*

$$\|V_n^{-1}\|_\infty = \max_{1 \leqq \nu \leqq n} \left\{ \frac{|p_n(-1)|}{(1 + x_\nu) |p_n'(x_\nu)|} \right\}, \tag{2.5}$$

where $p_n(x)$ is the polynomial in (2.2).

Theorem 2.1 can be combined with Theorems 2.2 and 2.3 to yield explicit expressions for the condition number $\operatorname{cond}_\infty V_n$ in the respective cases.

3. Optimally Conditioned Vandermonde Matrices for Real Nodes

We now wish to minimize the condition number

$$\varkappa_n(x) = \operatorname{cond}_\infty V_n(x) \tag{3.1}$$

as a function of the variables $x_1, x_2, \ldots, x_n$. If we allow complex points x_ν we can always achieve $\varkappa_n = n$ by taking x_ν to be the n-th roots of unity. This point configuration, in fact, is optimal for the spectral condition number [2, Example 6.4]. The problem becomes more interesting if we restrict all nodes x_ν to be real. In this case we assume, as before, that

$$x_1 > x_2 > \cdots > x_n. \tag{3.2}$$

The existence of an optimal point x in the domain (3.2) is assured, since $\varkappa_n(x) \to \infty$ both, as x approaches the boundaries of the domain, and as $x \to \infty$ on (3.2). The latter follows from (cf. [2, Theorem 4.2])

$$\varkappa_n(tx) = \max_{1 \le \nu \le n} (|a_{\nu n}| + t|a_{\nu,n-1}| + \cdots + t^{n-1}|a_{\nu 1}|) \sum_{\mu=1}^{n} |x_\mu|^{n-1},$$

$t > 0$ sufficiently large, where $a_{\nu\mu}$ are the coefficients of the ν-th Lagrange polynomial

$$l_\nu(x) = \prod_{\substack{\mu=1 \\ \mu \ne \nu}}^{n} \frac{x - x_\mu}{x_\nu - x_\mu} = a_{\nu n} x^{n-1} + a_{\nu,n-1} x^{n-2} + \cdots + a_{\nu 1}.$$

In fact, $\varkappa_n(tx) \to \infty$ as $t \to \infty$, since $a_{\nu 1}$ cannot vanish for all ν.

Theorem 3.1. *Suppose the problem of minimizing $\varkappa_n(x)$ in (3.1), subject to (3.2), has a unique solution. Then this solution is necessarily symmetric with respect to the origin, i.e.,*

$$x_\nu + x_{n+1-\nu} = 0, \qquad \nu = 1, 2, \ldots, n. \tag{3.3}$$

Proof. Evidently, $\|V_n(-x)\|_\infty = \|V_n(x)\|_\infty$, and also $\|V_n^{-1}(-x)\|_\infty = \|V_n^{-1}(x)\|_\infty$ (see [2, Theorem 4.2]), so that $\varkappa_n(-x) = \varkappa_n(x)$. Hence, if the point x in (3.2) is optimal, so is

$$-x_n > -x_{n-1} > \cdots > -x_1.$$

The assumed uniqueness then gives immediately (3.3). This proves Theorem 3.1.

Although it is not known whether or not the optimal point configuration on the real line is unique, Theorem 3.1 suggests that we constrain the variables to be symmetric. This then leaves us with $[n/2]$ independent variables,

$$x_1 > x_2 > \cdots > x_{[n/2]} > 0, \tag{3.4}$$

which we collect in the vector $x = [x_1, x_2, \ldots, x_{[n/2]}]$. The assumed symmetry has the further advantage that an explicit formula for $\varkappa_n(x)$ is available. Theorems 2.1 and 2.2 in fact give

$$\varkappa_n(x) = \max\left\{\frac{n}{2}, f_n(x)\right\} \cdot \max_{1 \le \nu \le \left[\frac{n+1}{2}\right]} f_{n,\nu}(x), \tag{3.5}$$

where

$$f_n(x) = \sum_{\mu=1}^{[n/2]} x_\mu^{n-1}, \tag{3.6}$$

$$f_{n,\nu}(x) = \left(1 + \frac{1}{x_\nu}\right) \prod_{\substack{\mu=1 \\ \mu \ne \nu}}^{n/2} \frac{1 + x_\mu^2}{|x_\nu^2 - x_\mu^2|}, \qquad \nu = 1, 2, \ldots, n/2 \;\; (n \text{ even}), \tag{3.7}$$

$$\begin{aligned} f_{n,\nu}(x) &= \frac{1 + x_\nu}{x_\nu^2} \prod_{\substack{\mu=1 \\ \mu \ne \nu}}^{(n-1)/2} \frac{1 + x_\mu^2}{|x_\nu^2 - x_\mu^2|}, \quad \nu = 1, 2, \ldots, (n-1)/2 \;\; (n \text{ odd}), \\ f_{n,(n+1)/2}(x) &= 2 \prod_{\mu=1}^{(n-1)/2} \left(1 + \frac{1}{x_\mu^2}\right) \quad (n \text{ odd}). \end{aligned} \tag{3.8}$$

If $n=2$ or $n=3$, the empty products in (3.7), (3.8) are to be understood as having the value 1.

Theorem 3.2. *If $n \geqq 3$ and the nodes x_ν satisfy (3.4), then $f_{n,1}(x) < f_{n,2}(x)$.*

Proof. If n is even, $n \geqq 4$, we obtain from (3.7),

$$\frac{f_{n,1}}{f_{n,2}} = \frac{1+\frac{1}{x_1}}{1+\frac{1}{x_2}} \frac{1+x_2^2}{1+x_1^2} \prod_{\mu=3}^{n/2} \frac{x_2^2-x_\mu^2}{x_1^2-x_\mu^2} < 1.$$

If n is odd, $n \geqq 5$, then from (3.8) we get

$$\frac{f_{n,1}}{f_{n,2}} = \left\{ \frac{1+x_1}{x_1^2(1+x_1^2)} \Big/ \frac{1+x_2}{x_2^2(1+x_2^2)} \right\} \prod_{\mu=3}^{(n-1)/2} \frac{x_2^2-x_\mu^2}{x_1^2-x_\mu^2} < 1,$$

since the function $(1+x)/\left(x^2(1+x^2)\right)$ is decreasing on $(0, \infty)$. The case $n=3$ can be verified directly. This proves Theorem 3.2.

The inequalities (3.4) define a cone in $\mathbb{R}^{[n/2]}$ which we denote by $\mathscr{C}_{[n/2]}$.

Theorem 3.3. *If $a \in \mathscr{C}_{[n/2]}$ is a minimum point of $\varkappa_n$, i.e.,*

$$\varkappa_n(a) \leqq \varkappa_n(x), \qquad \text{all } x \in \mathscr{C}_{[n/2]}, \tag{3.9}$$

then

$$f_n(a) = \frac{n}{2}. \tag{3.10}$$

Proof. Assume first n even. By (3.7) we have for any real t,

$$f_{n,\nu}(ta) = \left(1+\frac{1}{ta_\nu}\right) \prod_{\mu \neq \nu} \frac{\frac{1}{t^2}+a_\mu^2}{|a_\nu^2-a_\mu^2|}, \tag{3.11}$$

$$f_n(ta) f_{n,\nu}(ta) = f_n(a)\left(t+\frac{1}{a_\nu}\right) \prod_{\mu \neq \nu} \frac{1+t^2 a_\mu^2}{|a_\nu^2-a_\mu^2|}. \tag{3.12}$$

Now suppose $f_n(a) < n/2$. Then for $t_+ > 1$ sufficiently close to 1, we have $f_n(ta) < n/2$, $1 \leqq t < t_+$, and so

$$\varkappa_n(ta) = \frac{n}{2} \max_{1 \leqq \nu \leqq n/2} f_{n,\nu}(ta), \qquad 1 \leqq t < t_+.$$

From (3.11) we see, however, that each $f_{n,\nu}(ta)$ decreases as t increases from 1 to t_+. Therefore, $\varkappa_n(a)$ cannot be minimal. Suppose, on the other hand, that $f_n(a) > n/2$. Then

$$\varkappa_n(ta) = f_n(ta) \max_{1 \leqq \nu \leqq n/2} f_{n,\nu}(ta), \qquad t_- < t \leqq 1,$$

for t_- sufficiently close to 1. Eq. (3.12) now shows that $\varkappa_n(ta)$ decreases as t decreases from 1 to t_-, again contradicting minimality of $\varkappa_n(a)$. Consequently, $f_n(a) = n/2$, as asserted.

For n odd, the proof is analogous.

As a consequence of Theorems 3.2 and 3.3, the optimal point configuration $a \in \mathscr{C}_{[n/2]}$, $n \geqq 3$, can be found by solving the constrained minimax problem:

$$\begin{cases} \text{minimize} \quad \max\limits_{2 \leqq \nu \leqq \left[\frac{n+1}{2}\right]} f_{n,\nu}(x) \\ \text{subject to} \\ x \in \mathscr{C}_{[n/2]}, \quad f_n(x) = \dfrac{n}{2}. \end{cases} \tag{3.13}$$

Given a solution $a \in \mathscr{C}_{[n/2]}$ of (3.13), the minimum condition number of V_n is

$$(\varkappa_n)_{\text{opt}} = \frac{n}{2} \cdot \max_{2 \leqq \nu \leqq \left[\frac{n+1}{2}\right]} f_{n,\nu}(a). \tag{3.14}$$

Problem (3.13) can be written equivalently in the form of a nonlinear programming problem, viz.,

$$\begin{aligned} &\text{minimize } u \\ &\text{subject to} \\ &f_{n,\nu}(x) \leqq u, \qquad \nu = 2, 3, \ldots, \left[\frac{n+1}{2}\right], \\ &f_n(x) = \frac{n}{2} \\ &x_\nu - x_{\nu-1} < 0, \qquad \nu = 2, 3, \ldots, \left[\frac{n}{2}\right], \\ &-x_{[n/2]} < 0. \end{aligned} \tag{3.15}$$

To an optimal point x^*, u^* of (3.15) corresponds a solution x^* of (3.13), u^* being the minimum value of the maximum in (3.13), and vice versa.

In the next section we solve (3.13) directly for $2 \leqq n \leqq 6$.

4. Optimal Symmetric Point Configurations for $2 \leqq n \leqq 6$

The case $n = 2$ can be handled directly by (3.5), which gives

$$\varkappa_2(x_1) = \max(1, x_1) \cdot \left(1 + \frac{1}{x_1}\right), \qquad x_1 > 0.$$

The minimum of $\varkappa_2(x_1)$ is clearly assumed at $x_1 = 1$, giving the optimal configuration

$$x_1 = -x_2 = 1, \qquad (\varkappa_2)_{\text{opt}} = 2. \tag{4.1}$$

Problem (3.13) in the case $n = 3$ has the trivial solution

$$\begin{aligned} x_1 &= -x_3 = \sqrt{\tfrac{3}{2}} = 1.2247448714, \\ x_2 &= 0, \\ (\varkappa_3)_{\text{opt}} &= 5. \end{aligned} \tag{4.2}$$

If $n = 4$, the problem becomes

$$\begin{cases} \text{minimize } \left(1 + \dfrac{1}{x_2}\right) \dfrac{1 + x_1^2}{x_1^2 - x_2^2}, \\ x_1 > x_2 > 0, \quad x_1^2 + x_2^2 = 2. \end{cases}$$

This can be reduced to an unconstrained one-dimensional minimization problem viz.,

$$\text{minimize}\left(1+\frac{1}{x_2}\right)\frac{1+(2-x_2^3)^{2/3}}{(2-x_2^3)^{2/3}-x_2^2}, \qquad 0<x_2<1.$$

Setting the derivative equal to zero, and applying the method of bisection, we find

$$\begin{aligned} x_1 &= -x_4 = 1.2228992744,\\ x_2 &= -x_3 = 0.5552395908,\\ (\varkappa_4)_{\text{opt}} &= 11.7755336773. \end{aligned} \tag{4.3}$$

In the case $n=5$, Problem (3.13) becomes

$$\begin{cases}\text{minimize max } \{f_{5,2}, f_{5,3}\}\\ x_1>x_2>0, \qquad x_1^4+x_2^4=\frac{5}{2}.\end{cases}$$

We take x_2 as independent variable and express x_1 in terms of it as $x_1=(\frac{5}{2}-x_2^4)^{1/4}$, $0<x_2<(\frac{5}{4})^{1/4}$. Then the problem again becomes a one-dimensional minimax problem. Note from (3.8) that both $f_{5,2}$ and $f_{5,3}$ become infinite as $x_2\downarrow 0$, and, moreover, $f_{5,3}\sim 2f_{5,2}$. For small x_2, therefore, $f_{5,3}>f_{5,2}$. As $x_2\uparrow(5/4)^{1/4}$, we have $x_1-x_2\to 0$ and $f_{5,2}$ again becomes infinite. The other function, $f_{5,3}$, however, is monotonically decreasing, since

$$\frac{d f_{5,3}}{d x_2}=\frac{4}{x_1^6\, x_2^3}\,(x_2^6+x_2^4-x_1^6-x_1^4)<0$$

by virtue of $x_1>x_2$. It follows that the minimum of max $\{f_{5,2}, f_{5,3}\}$ occurs at the largest x_2 for which $f_{5,2}=f_{5,3}$. We are led, therefore, to solving the transcendental equation

$$x_1^2(1+x_2)-2(x_1^2-x_2^2)(1+x_2^2)=0, \qquad 0<x_2<(\tfrac{5}{4})^{1/4},$$

where $x_1=(\frac{5}{2}-x_2^4)^{1/4}$. This can be done numerically, the result being

$$\begin{aligned} x_1 &= -x_5 = 1.2001030479,\\ x_2 &= -x_4 = 0.8077421768,\\ x_3 &= 0,\\ (\varkappa_5)_{\text{opt}} &= 21.4560069858. \end{aligned} \tag{4.4}$$

Finally, if $n=6$, we must solve

$$\begin{cases}\text{minimize max } \{f_{6,2}, f_{6,3}\}\\ x_1>x_2>x_3>0, \qquad x_1^5+x_2^5+x_3^5=3.\end{cases}$$

Taking x_2, x_3 as independent variables, we consider $f_{6,2}$ and $f_{6,3}$ to be functions of x_2, x_3. Their maximum is to be minimized on the open domain

$$\mathscr{D}\colon 2x_2^5+x_3^5<3, \qquad x_2>x_3>0.$$

Since one or both of the functions $f_{6,2}$, $f_{6,3}$ become infinite on the boundary of $\mathscr{D}$, the optimal point is interior to $\mathscr{D}$. We claim that $f_{6,2}=f_{6,3}$ at the minimum point $x=a\in\mathscr{D}$. Indeed, if we had $f_{6,2}>f_{6,3}$ at $x=a$, the same inequality would hold in some neighborhood $\mathcal{U}(a)$ of a, giving

$$\max\{f_{6,2}, f_{6,3}\} = f_{6,2} \quad \text{for } x \in \mathcal{U}(a).$$

Therefore, $x = a$ being optimal, both partial derivatives of $f_{6,2}$ would have to vanish at $x = a$. A short calculation, however, shows that

$$\frac{\partial f_{6,2}}{\partial x_3} = \frac{2\left(1 + \frac{1}{x_2}\right)(1 + x_2^2)}{x_1^3 (x_1^2 - x_2^2)^2 (x_2^2 - x_3^2)^2} \{x_3^4 (1 + x_3^2)(x_2^2 - x_3^2) + x_1^3 x_3 (1 + x_1^2)(x_1^2 - x_2^2)\} > 0.$$

Thus, $f_{6,2}$ cannot be larger than $f_{6,3}$ at $x = a$. Similarly one invalidates the inequality $f_{6,2} < f_{6,3}$ at $x = a$, since

$$\frac{\partial f_{6,3}}{\partial x_2} = \frac{2\left(1 + \frac{1}{x_3}\right)(1 + x_3^2)}{x_1^3 (x_1^2 - x_3^2)^2 (x_2^2 - x_3^2)^2} \{x_2^4 (1 + x_2^2)(x_2^2 - x_3^2) - x_1^3 x_2 (1 + x_1^2)(x_1^2 - x_3^2)\} < 0.$$

Consequently, $f_{6,2} = f_{6,3}$ at $x = a$, and our problem becomes

$$\begin{aligned} &\text{minimize } f_{6,2}(x) \\ &\text{subject to} \\ &\qquad f_{6,2}(x) = f_{6,3}(x), \\ &\qquad f_6(x) = 3, \\ &\qquad x_1 > x_2 > x_3 > 0. \end{aligned}$$

The first equation in the constraints is quadratic in x_1, hence can readily be solved for x_1. The problem, more explicitly, then becomes

$$\begin{aligned} &\text{minimize} \left(1 + \frac{1}{x_2}\right) \frac{(1 + x_1^2)(1 + x_3^2)}{(x_1^2 - x_2^2)(x_2^2 - x_3^2)} \\ &\text{subject to} \\ &x_1 = \left\{ \frac{x_2^2 \left(1 + \frac{1}{x_3}\right)(1 + x_2^2) - x_3^2 \left(1 + \frac{1}{x_2}\right)(1 + x_3^2)}{\left(1 + \frac{1}{x_3}\right)(1 + x_2^2) - \left(1 + \frac{1}{x_2}\right)(1 + x_3^2)} \right\}^{\frac{1}{2}}, \qquad (4.5) \\ &x_1^5 + x_2^5 + x_3^5 = 3, \\ &x_1 > x_2 > x_3 > 0. \end{aligned}$$

Basically, this is again a one-dimensional minimization problem, if we take x_3 as independent variable. To construct an admissible set of nodes x_1, x_2, x_3, we pick x_3 with $0 < x_3 < 1$, then try to find $x_2 > x_3$ by solving the transcendental equation

$$\phi(x_2, x_3) = x_1^5 + x_2^5 + x_3^5 - 3 = 0, \tag{4.6}$$

where x_1 is the square root expression given in (4.5). Having determined x_2, we obtain x_1 from its explicit formula, which clearly implies that $x_1 > x_2$. Each evaluation of the objective function in (4.5) (considered a function of x_3) thus requires the solution of a transcendental equation.

If x_3 is selected too large, Eq. (4.6) may fail to have a solution $x_2 > x_3$. To examine this point in more detail, we first write the equation for x_1^2 in the alternative form

$$x_1^2 = \frac{(x_3+1)\,x_2^4 + x_3(x_3+1)\,x_2^3 + (x_3+1)(x_3^2+1)\,x_2^2 + x_3(x_3+1)(x_3^2+1)x_2 + x_3^2(x_3^2+1)}{(x_3+1)\,x_2^2 + x_3(x_3+1)\,x_2 + x_3^2+1}$$

(which, incidentally, is the preferred form for computation.) A straightforward calculation then shows that x_1^2 is increasing as a function of x_2 for any fixed $x_3 > 0$. Consequently, the function $\phi(x_2, x_3)$ in (4.6) is increasing unboundedly in x_2, for x_3 fixed, and therefore vanishes at a unique $x_2 > x_3$ if and only if $\phi(x_3, x_3) < 0$. One computes

$$\phi(x, x) = x^5 \left\{\frac{4x^3+5x^2+2x+3}{2x^3+3x^2+1}\right\}^{\frac{5}{2}} + 2x^5 - 3,$$

which on $(0, \infty)$ increases monotonically from -3 to ∞. Therefore, $\phi(x, x)$ has a unique positive zero $x = \xi$, which can be computed to be $\xi = 0.7537718186$ (to 10 decimals), and the equation (4.6) has a unique solution $x_2 > x_3$ if and only if $x_3 < \xi$. A binary search procedure to locate the minimum of the objective function on $0 < x_3 < \xi$, combined with the method of bisection for solving the transcendental equation (4.6), now yields the following optimal point configuration:

$$x_1 = -x_6 = 1.1601101028,$$
$$x_2 = -x_5 = 0.9771502216,$$
$$x_3 = -x_4 = 0.3788765912,$$
$$(\varkappa_6)_{\text{opt}} = 51.3302762899.$$

In Table 4.1 below we compare the optimal condition number with the condition number for equidistant points $x_\nu = 1 - 2(\nu-1)/(n-1)$ and Chebyshev points $x_\nu = \cos\big((2\nu-1)\,\pi/2n\big)$.

Table 4.1. The condition number $\text{cond}_\infty V_n(x)$ for various point configurations x

n	Equidistant points on $[-1, 1]$	Chebyshev points on $(-1, 1)$	Optimal points
2	2.00	2.41	2.00
3	6.00	7.00	5.00
4	18.00	18.94	11.78
5	50.00	41.00	21.46
6	159.375	112.82	51.33

5. Optimally Conditioned Vandermonde Matrices for Nonnegative nodes

In the case of nonnegative nodes,

$$x_1 > x_2 > \cdots > x_n \geqq 0, \tag{5.1}$$

the condition number (3.1) of V_n, by Theorems 2.1 and 2.3, can be expressed as

$$\varkappa_n(x) = \max\{n, g_n(x)\} \cdot \max_{1 \leqq \nu \leqq n} g_{n,\nu}(x), \tag{5.2}$$

where

$$g_n(x) = \sum_{\mu=1}^{n} x_\mu^{n-1}, \tag{5.3}$$

$$g_{n,\nu}(x) = \prod_{\substack{\mu=1 \\ \mu\neq\nu}}^{n} \frac{1+x_\mu}{|x_\nu - x_\mu|}, \qquad \nu = 1, 2, \ldots, n. \tag{5.4}$$

In much the same way as in Section 3 (in fact, somewhat simpler), one proves the following two theorems.

Theorem 5.1. *If* $n \geqq 2$ *and the nodes* x_ν *satisfy* (5.1) *then* $g_{n,1}(x) < g_{n,2}(x)$.

Theorem 5.2. *If* a *is a minimum point of* $\varkappa_n$ *then*

$$g_n(a) = n. \tag{5.5}$$

In addition, we can show the following.

Theorem 5.3. *If* $a = [a_1, a_2, \ldots, a_n]$ *is a minimum point of* $\varkappa_n$ *then* $a_n = 0$.

Proof. Assume first $n \geqq 3$. Suppose that $a_n > 0$. By (5.5), we have

$$a_1^{n-1} + a_2^{n-1} + \cdots + a_n^{n-1} = n. \tag{5.6}$$

We now let x_1 and x_n vary in such a way that

$$x_1^{n-1} + x_n^{n-1} = n - \sum_{\mu=2}^{n-1} a_\mu^{n-1} \tag{5.7}$$

and

$$x_1 > a_2 > \cdots > a_{n-1} > x_n > 0. \tag{5.8}$$

Taking x_n as the independent variable, we then have

$$\frac{dx_1}{dx_n} = -\left(\frac{x_n}{x_1}\right)^{n-2}.$$

We examine the behavior of each $g_{n,\nu}$, $2 \leqq \nu \leqq n$. If $\nu \leqq n-1$, noting that

$$g_{n,\nu} = \frac{(1+x_1)(1+x_n)}{(x_1 - a_\nu)(a_\nu - x_n)} \pi_n, \qquad \pi_n = \prod_{\substack{\mu=2 \\ \mu\neq\nu}}^{n-1} \frac{1+a_\mu}{|a_\nu - a_\mu|},$$

we obtain

$$\frac{dg_{n,\nu}}{dx_n} = \frac{\pi_n}{x_1^{n-2}(x_1 - a_\nu)^2(a_\nu - x_n)^2} \{(x_1 - a_\nu)(a_\nu - x_n)\,[(1+x_1)\,x_1^{n-2} - (1+x_n)x_n^{n-2}]$$
$$+ (1+x_1)(1+x_n)\,[(a_\nu - x_n)\,x_n^{n-2} + (x_1 - a_\nu)\,x_1^{n-2}]\} > 0.$$

Similarly, for $\nu = n$, we have

$$g_{n,n} = \frac{(1+x_1)\,\pi_n^*}{(x_1 - x_n)\prod_{\mu=2}^{n-1}(a_\mu - x_n)}, \qquad \pi_n^* = \prod_{\mu=2}^{n-1}(1+a_\mu),$$

from which we get

$$\frac{dg_{n,n}}{dx_n} = \frac{\pi_n^*}{x_1^{n-2}(x_1-x_n)^2\prod_{\mu=2}^{n-1}(a_\mu-x_n)^2}\Big\{[(1+x_1)\,x_1^{n-2}+(1+x_n)\,x_n^{n-2}]\prod_{\mu=2}^{n-1}(a_\mu-x_n)$$
$$+(1+x_1)(x_1-x_n)\,x_1^{n-2}\sum_{\nu=2}^{n-1}\prod_{\substack{\mu=2\\ \mu\neq\nu}}^{n-1}(a_\mu-x_n)\Big\}>0.$$

In particular,

$$\left.\frac{dg_{n,\nu}}{dx_n}\right|_{x_n=a_n}>0, \qquad \nu=2,3,\ldots,n.$$

It follows that each $g_{n,\nu}(a)$ decreases as a_n is decreased and simultaneously a_1 increased (in such a way that (5.6) remains valid). Consequently, in view of Theorems 5.1 and 5.2, $\varkappa_n(a)$ cannot be minimal, making the assumption $a_n>0$ untenable. This proves the theorem for $n\geqq 3$. If $n=2$, the theorem can be proved directly.

As a result of Theorem 5.3 it suffices to consider $n-1$ variables, $x_1, x_2, \ldots, x_{n-1}$, with

$$x_1>x_2>\cdots>x_{n-1}>x_n=0. \tag{5.9}$$

We denote the cone in $\mathbb{R}^n$, defined by (5.9), by $\mathscr{C}_n^0$. Our minimization problem then becomes

$$\begin{aligned}&\text{minimize} \max_{2\leqq\nu\leqq n} g_{n,\nu}(x)\\ &\text{subject to}\\ &x\in\mathscr{C}_n^0, \qquad g_n(x)=n.\end{aligned} \tag{5.10}$$

If $a=[a_1, a_2, \ldots, a_{n-1}, 0]\in\mathscr{C}_n^0$ is a solution of (5.10) then

$$(\varkappa_n)_{\text{opt}}=n\max_{2\leqq\nu\leqq n} g_{n,\nu}(a). \tag{5.11}$$

As before, in section 3, we can give (5.10) an equivalent formulation as a non-linear programming problem,

$$\begin{aligned}&\text{minimize } u\\ &\text{subject to}\\ &g_{n,\nu}(x)\leqq u, \qquad \nu=2,3,\ldots,n,\\ &g_n(x)=n,\\ &x_\nu-x_{\nu-1}<0, \qquad \nu=2,3,\ldots,n-1,\\ &\quad -x_{n-1}<0.\end{aligned} \tag{5.12}$$

It is possible to solve (5.10) directly for $n=2$ and $n=3$. In the first case, the solution is trivially

$$x_1=2,\ x_2=0, \qquad (\varkappa_2)_{\text{opt}}=3. \tag{5.13}$$

In the second case we have to solve

$$\begin{aligned}&\text{minimize} \max\{g_{3,2}, g_{3,3}\}\\ &\text{subject to}\\ &x_1>x_2>x_3=0, \qquad x_1^2+x_2^2=3,\end{aligned}$$

where

$$g_{3,2}(x) = \frac{1+x_1}{x_2(x_1-x_2)}, \quad g_{3,3}(x) = \frac{(1+x_1)(1+x_2)}{x_1 x_2}.$$

We take x_2 as independent variable, in terms of which $x_1 = \sqrt{3-x_2^2}$, $0 < x_2 < \sqrt{\frac{3}{2}}$. As $x_2 \downarrow 0$, both $g_{3,2}$ and $g_{3,3}$ become infinite, the latter being larger for small x_2 since $x_1 > 1$. As $x_2 \uparrow \sqrt{\frac{3}{2}}$, we have $x_1 - x_2 \to 0$ and $g_{3,2}$ again becomes infinite. The other function, $g_{3,3}$, decreases monotonically, since

$$\frac{dg_{3,3}}{dx_2} = \frac{1}{x_1^3 x_2^2}\{(1+x_2)x_2^2 - (1+x_1)x_1^2\} < 0.$$

Consequently, $\max\{g_{3,2}, g_{3,3}\}$ is minimized at the largest x_2 for which $g_{3,2} = g_{3,3}$. This equation amounts to

$$\frac{1}{x_1 - x_2} = \frac{1+x_2}{x_1} \qquad (x_1 = \sqrt{3-x_2^2}),$$

and has the unique positive root $x_2 = (\sqrt{5}-1)/2$. Our optimal solution, therefore, is given by

$$\begin{aligned} x_1 &= \frac{\sqrt{5}+1}{2} = 1.6180339887, \\ x_2 &= \frac{\sqrt{5}-1}{2} = 0.6180339887, \\ x_3 &= 0, \\ (\varkappa_3)_{\text{opt}} &= 3\left(1+\frac{1}{x_1}\right)\left(1+\frac{1}{x_2}\right) = 12.7082039325. \end{aligned} \tag{5.14}$$

Interestingly enough, x_1 coincides with the ratio of the "golden section".

Acknowledgement. The author is grateful to the referees for pointing out Eq. (1.5), and for providing the short proof of Theorem 2.1.

References

1. Björck, Å., Elfving, T.: Algorithms for confluent Vandermonde systems. Numer. Math. **21**, 130–137 (1973)
2. Gautschi, W.: Norm estimates for inverses of Vandermonde matrices. Numer. Math. **23**, 337–347 (1975)
3. Luttmann, F. W., Rivlin, T. J.: Some numerical experiments in the theory of polynomial interpolation. IBM J. Res. Develop. **9**, 187–191 (1965)
4. McCabe, J. H., Phillips, G. M.: On a certain class of Lebesgue constants. BIT **13**, 434–442 (1973)
5. Powell, M. J. D.: On the maximum errors of polynomial approximations defined by interpolation and by least squares criteria. Comput. J. **9**, 404–407 (1967)

Prof. Dr. W. Gautschi
Purdue University
Computer Sciences
Mathematical Sciences Building
West Lafayette, Indiana 47907
U.S.A.