An Elementary Proof of the Exponential Conditioning of Real Vandermonde Matrices

Unione Matematica Italiana

<http://www.bdim.eu/item?id=BUMI_2007_8_10B_3_761_0>
An Elementary Proof of the Exponential Conditioning of Real Vandermonde Matrices.

Stefano Serra-Capizzano

**Sunto.** – Si fornisce e si discute una dimostrazione elementare, proponibile in un corso di Matematica Numerica della Triennale, del condizionamento esponenziale di matrici di Vandermonde: si impegna esclusivamente la definizione di condizionamento e l’espressione esplicita della norma infinito su [−1, 1] dei polinomi di Chebyshev di prima specie. La stessa idea dimostrativa funziona nel caso della ben nota matrice di Hilbert.

**Summary.** – We provide and discuss an elementary proof of the exponential conditioning of real Vandermonde matrices which can be easily given in undergraduate courses: we exclusively use the definition of conditioning and the sup-norm formula on [−1, 1] for Chebyshev polynomials of first kind. The same proof idea works virtually unchanged for the famous Hilbert matrix.

1. – Introduction.

Given the grid \( G_n = \{x_{0}^{(n)}, x_{1}^{(n)}, \ldots, x_{n}^{(n)}\} \) of pairwise distinct nodes and the set of linearly independent functions \( \phi_{0}, \phi_{1}, \ldots, \phi_{n} \), we define the (generalized) Vandermonde matrix \( V_n \) of size \( n + 1 \) as

\[
(V_n)_{i,j} = \phi_{j}(x_{i}), \quad x_{i} \equiv x_{i}^{(n)}, \quad i,j = 0, \ldots, n.
\]

The classical Vandermonde matrix is obtained by putting \( \phi_{j}(t) = t^{j}, j = 0, \ldots, n \), so that \( (V_{n})_{i,j} = x_{i}^{j}, i,j = 0, \ldots, n \), and is invertible whenever the nodes are pairwise distinct (see Subsection 1.1).

We wish to show that this choice of the basis functions (analogously to the case of the Hilbert matrix) is delicate if we restrict our attention to the case of real nodes i.e. \( x_{i} \equiv x_{i}^{(n)} \in R, i = 0, \ldots, n \). Indeed we prove that the conditioning of \( V_{n} \) i.e. \( \mu(V_{n}) = ||V_{n}|| ||V_{n}^{-1}|| \) is exponential as \( n \) and this tells one that the use of such a basis makes the inherent error (related to e.g. the solution of a linear system \( V_{n}x = b \)) huge, already for moderate size of \( n \), and therefore the Numerical Analysts must be very careful when using such a kind of matrices (see [1]). Here for a generic \( m \)-by-\( m \) matrix \( A, m \in N^{+} \), we consider the spectral norm

\[
||A|| = \sup_{x \neq 0} \frac{||Ax||}{||x||},
\]

related to the Euclidean vector norm \( ||x|| = \sqrt{\sum_{i=1}^{m} |x_{i}|^{2}} \).
In fact, in [1, Theorem 4.1] Beckermann proved that

$$\mu(V_n) \geq \sqrt{\frac{2}{n+1} \left( \sqrt{2} + 1 \right)^{n-1}},$$

and the latter lower-bound may be attained up to a factor \((n + 1)^{3/2}\); in addition, very recently Li [3] slightly improved both these bounds. However, the proofs are quite tricky and involve somehow sophisticate tools (for further discussions see Remark 2.3 and Subsection 2.1). In Section 2 the conditioning of classical real Vandermonde matrices is discussed, by employing only tools that could be presented in detail in a classroom and whose prerequisites are reported below.

1.1 – Prerequisites and notations.

Let \(G_n = \{x_0^{(n)}, x_1^{(n)}, \ldots, x_n^{(n)}\}\) be a grid of pairwise distinct nodes. The polynomial \(p\) of degree at most \(n\) which solves the interpolation problem \(p(x_i) = g_i, g_i\) given values, \(x_i = x_i^{(n)}, i = 0, \ldots, n\), exists and is unique: the existence comes from the explicit formula \(p(x) = \sum_{i=0}^{n} g_i L_i(x), \) with \(L_i(x) = \prod_{j=0, j \neq i}^{n} \frac{x - x_j}{x_i - x_j};\) the unicity follows from the fundamental theorem of algebra. Therefore the matrix \(V_n\) is invertible since the given interpolation problem is equivalent to the solution of the linear system \(V_n a = g,\) with \(p(x) = a_0 + a_1 x + \cdots + a_n x^n\) (i.e. \(p\) is represented in the canonical basis of polynomials, \(1, x, \ldots, x^n\), in place of the Lagrange basis formed by \(L_0(x), L_1(x), \ldots, L_n(x)\)).

Concerning the Chebyshev polynomials, their definition can be given as \(q_{k+1}(x) = 2 x q_k(x) - q_{k-1}(x), k \geq 1, q_0(x) = 1, q_1(x) = x.\) By direct inspection it is clear that \(q_n(x) = 2^{n-1} x^n + \tilde{q}_{n-1}(x)\) with \(q_{n-1}\) of degree at most \(n - 1\) and \(n \geq 1.\) Moreover if \(x \in [-1, 1]\) we consider the transformation \(x = \cos(\theta), \theta \in [0, \pi]\) and therefore a simple trigonometric identity shows that \(q_n(x) = \cos(n \theta).\) As a consequence \(\|q_n\|_{\infty, [-1, 1]} = 1\) and \(p_n^* = 2^{-(n-1)} q_n\) is such that \(p_n^*\) is monic and the famous identity

\[
\|p_n^*\|_{\infty, [-1, 1]} = 2^{-(n-1)}.
\]

is true for \(n \geq 1.\)

2. – Main results.

Here we furnish an elementary proof of the exponential conditioning as \(n\) of any classical Vandermonde matrix \(V_n\) with real nodes.

**Theorem 2.1.** – Let \(V_n\) be a classical real Vandermonde matrix with pairwise distinct nodes \(x_0^{(n)}, x_1^{(n)}, \ldots, x_n^{(n)}\) and let \(\mu(V_n)\) be the (spectral) condition
number of $V_n$. Then there exists $c > 1$ \( (\text{independent of } n) \) such that, for every $n$ large enough, it holds

\[
\mu(V_n) \geq c^n,
\]

where we have assumed that there exists the limit $l$ as $n$ tends to infinity of

\[
s_n = \max_{j=0,\ldots,n} |x_j|, \quad x_j \equiv x_j^{(n)}, \quad j = 0, \ldots, n.
\]

**Proof.** – Setting $f_0 = V_ne_0$ with \((e_k)_i = 1\) if \(k = i\) and zero otherwise, \(k, i = 0, \ldots, n,\) from the definition of the spectral condition number, we can easily see that \((f_0)_i = 1\) for every \(i = 0, \ldots, n,\) and

\[
\|V_n\| = \sup_{x \neq 0, x \in C^{n+1}} \frac{\|V_nx\|}{\|x\|} \geq \frac{\|V_ne_0\|}{\|e_0\|} = \frac{\|f_0\|}{\|e_0\|} = \frac{\sqrt{n+1}}{\sqrt{n+1}},
\]

\[
\|V_n^{-1}\| = \sup_{x \neq 0, x \in C^{n+1}} \frac{\|V_n^{-1}x\|}{\|x\|} \geq \frac{\|V_n^{-1}f_0\|}{\|f_0\|} = \frac{1}{\|f_0\|} = \frac{1}{\sqrt{n+1}}.
\]

Let $s_n = \max_{j=0,\ldots,n} |x_j|$, \( l = \lim_{n \to \infty} s_n \) and \( f_n = V_ne_n = (x_i^{(n)})_{i=0}^n \). We have

\[
\|V_ne_n\| = \|f_n\| \geq s_n^n \quad \text{and, if } l > 1, \text{ for } n \text{ large enough we infer}
\]

\[
\|V_n\| = \sup_{x \neq 0, x \in C^{n+1}} \frac{\|V_nx\|}{\|x\|} \geq \frac{\|V_ne_n\|}{\|e_n\|} = \frac{\|f_n\|}{\|e_n\|} \geq s_n^n \geq (l - \varepsilon)^n
\]

with \( l - 2\varepsilon > 1 \). Hence, by invoking the second inequality in (2.1), we deduce

\[
\mu(V_n) \geq \frac{(l - \varepsilon)^n}{\sqrt{n+1}} \geq (l - 2\varepsilon)^n
\]

definitely that is the desired result with \( c = l - 2\varepsilon > 1 \) and independent of \( n \). For \( l < 1 \) the reasoning is similar since

\[
\|V_n^{-1}\| = \sup_{x \neq 0, x \in C^{n+1}} \frac{\|V_n^{-1}x\|}{\|x\|} \geq \frac{\|V_n^{-1}f_n\|}{\|f_n\|} = \frac{1}{\|f_n\|} \geq \frac{s_n^{-n}}{\sqrt{n+1}} \geq \frac{(l + \varepsilon)^{-n}}{\sqrt{n+1}}
\]

with \( l + \varepsilon < 1 \). Consequently, the first inequality in (2.1) implies

\[
\mu(V_n) \geq (l + \varepsilon)^{-n}
\]

for \( n \) large enough, and thesis follows with \( c = (l + \varepsilon)^{-1} \).

If \( l = 1 \) then we use a different idea. First we consider the auxiliary Vandermonde matrix \( W_n \) with nodes \( x'_j = x_j/s_n \). In such a way we have \( x'_j \in [-1, 1] \) and

\[
W_n = V_nD_n, \quad D_n = \text{diag}_{j=0,\ldots,n}(s_n^{-j}), \quad s_n = \max_{j=0,\ldots,n} |x_j|.
\]

From the sub-multiplicativity of the (matrix) norms (i.e. \( \|AB\| \leq \|A\| \cdot \|B\| \) for any choice of \( m \times m \) matrices \( A, B \)) we obtain \( \mu(W_n) \leq \mu(V_n)\mu(D_n) \) with
\( \mu(D_n) = s_n^n \) if \( s_n \geq 1 \) and \( \mu(D_n) = s_n^{-n} \) otherwise. Therefore

(2.2) \[ \mu(V_n) \geq \mu(W_n) \mu^{-1}(D_n) \]

where, by virtue of the relation \( l = 1 \), for every \( \varepsilon > 0 \) there exists \( \bar{n} \) for which

(2.3) \[ s_{\bar{n}}^n \geq (1 - \varepsilon)^n, \quad \forall n \geq \bar{n}. \]

Since \( \varepsilon > 0 \) is arbitrary, it suffices to prove the exponential conditioning of \( W_n \)
which means that we are assuming all the nodes in the interval \([-1, 1]\).

The inequalities in (2.1) still hold with \( W_n \) in place of \( V_n \) since they do not depend on \( l \). In this new context the key bound is from below and concerns \( \|W_n\| \geq \sqrt{n + 1}. \) We use the monic Chebyshev polynomials of first kind \( p^*_n \) over
\([-1, 1]\) having degree \( n \geq 1 \) (see [4] and Subsection 1.1). Hence, by calling \( x^* \) the vector of the coefficients of \( p^*_n \) in the canonical ascending polynomial basis, by monicity of \( p^*_n \), we deduce \( \|x^*\| \geq 1 \) and then

\[
\|W_n^{-1}\| = \sup_{x \neq 0, x \in \mathbb{C}^{n+1}} \frac{\|W_n^{-1}x\|}{\|x\|} \geq \frac{\|W_n^{-1}(W_n x^*)\|}{\|W_n x^*\|} \geq \frac{\|x^*\|}{\|W_n x^*\|} \geq_{\text{monicity}} \frac{1}{\|W_n x^*\|}.
\]

Now, by direct inspection, the expression of the vector \( W_n x^* \) is such that its \( i \)-th entry is given by \( (W_n x^*)_i = \sum_{j=0}^{n} (\alpha^*_i)^j (x^*)_j = p^*_n(\alpha^*_i) \) and consequently, by (2.1) and by the latter inequality on \( \|W_n^{-1}\| \), we deduce

\[
\mu(W_n) = \|W_n\| \|W_n^{-1}\| \geq \sqrt{n + 1} \frac{1}{\|W_n x^*\|} \geq \sqrt{n + 1} \frac{1}{\sqrt{\sum_{i=0}^{n} \left| p^*_n(\alpha^*_i) \right|^2}} \geq \sqrt{n + 1} \frac{1}{\sqrt{(n + 1) \|p^*_n\|^2_{\infty,[-1,1]}}} \geq (1.2) 2^{n-1}
\]

and finally the claimed result follows from (2.2) and (2.3). \( \Box \)
Remark 2.2. – Theorem 2.1 is still valid under the assumption that \( \max_{j=0, \ldots, n} |x_j^{(n)}| \) has no limit as \( n \) tends to infinity. Indeed, if the above sequence has \( k \) accumulation points, \( k < \infty \), then \( \max_{j=0, \ldots, n} |x_j^{(n)}| \) can be partitioned in exactly \( k \) subsequences having limit: in other words \( N = \bigcup_{j=1}^{k} (n_j^{(q)} : q \in \mathbb{N}) \) and on every Vandermonde subsequence \( V_{n_j^{(q)}} \) we can directly apply Theorem 2.1 (since for every \( j = 1, \ldots, k \), the subsequence \( \max_{j=0, \ldots, n_j^{(q)}} |x_j^{(n_j^{(q)})}| \) has limit as \( q \) tends to infinity).

Therefore there exist positive numbers \( c_j > 1, j = 1, \ldots, k \), for which \( \mu(V_{n_j^{(q)}}) \geq c_j^{q_j} \) with \( q \) large enough. Putting together the partial information it is true that

\[
\mu(V_n) \geq c^n
\]

with \( n \) sufficiently large and \( c = \min_{j=1, \ldots, k} c_j > 1 \). If the set of accumulation points is not finite then the preceding reduction to Theorem 2.1 cannot be done in the same manner since \( \inf_{j \in J} c_j, \#J = \infty \), may equal 1 although every \( c_j \) is strictly larger than 1.

In such a situation we can reduce the reasoning again to Theorem 2.1, by exploiting a contradiction argument. Indeed, if the desired result does not hold then we can define a subsequence \( n_q \) for which

\[
\mu(V_{n_q}) \leq c^{n_q}
\]

for every \( c > 1 \) and for every \( q \geq q_c \). By Theorem 2.1 it has to be true that \( \max_{j=0, \ldots, n_q} |x_j^{(n_q)}| \) has no limit as \( q \) tends to infinity. However we can extract a subsequence \( n_{q_t} \) from \( n_q \) for which \( \max_{j=0, \ldots, n_{q_t}} |x_j^{(n_{q_t})}| \) has limit as \( t \) diverges infinity.

Therefore, again by Theorem 2.1, we have \( \mu(V_{n_{q_t}}) \geq d^{n_{q_t}}, d > 1 \), definitely: it is clear that the last inequality contradicts the relation in (2.4).

2.1 – A further comparison with the literature.

We recall that on this topic there exist many contributions where the exponential conditioning is proven, and also with best constants (see Beckermann [1]). Other results can be found in [2, 5, 7, 8]. In all these works the findings are more general, but the approach is more sophisticated and is not elementary: here for not elementary we mean that advanced knowledge is required and therefore the reasoning is not easily understandable e.g. by undergraduate students. For instance Krylov vectors, Gram matrices, Hankel matrices, expansion coefficients of orthogonal polynomials, and nontrivial inequalities involving these objects are
basic tools in [1, 2, 3, 5, 6, 7, 8]. On the other hand, also some of these more
delicate proofs can be simplified, by only considering further polynomials asso-
ciated to the Chebyshev ones. As an example, let \( s_n \) be the maximum of the
moduli of the nodes like in Theorem 2.1. Since the spectral norm of \( V_n \) cannot be
less than the Euclidean norm of any of its rows, it follows that \( \|V_n\| \geq s_n^\mu \). In
analogy with our proof, take \( q_n = 2^{n-1} p_n \) as the normalized Chebyshev poly-
nomial of degree \( n \) with unitary infinity norm on \([-1, 1]\) and leading coefficient
equal to \( 2^{n-1} \), see Subsection 1.1. Furthermore, arrange the coefficients of
\( \tilde{q}_n(x) = q_n(x/b) \) in the basis of the ascending monomials as a column vector \( y \) (see
[1, Proof of Theorem 4.1]): as a consequence, \( \|y\| \geq (2/s_n)^{n}/2 \) (the leading
coefficient of the new polynomial \( \tilde{q}_n \)) and \( V_n y \) is the column vector containing the
samplings of \( \tilde{q}_n \) at the given nodes. Hence \( \|V_n\| \geq \sqrt{n+1} \) so that

\[
\mu(V_n) \geq s_n^\mu \|y\| \geq \frac{2^{n-1}}{\sqrt{n+1}}.
\]

We remark that similar bounds were obtained by Gautschi and Inglese [2] and by
Tyrtyshnikov [7].

Now if we require that our students know also the explicit expression of the
Chebyshev polynomials [4] and the associated recurrence coefficients (see
Subsection 1.1), then we are allowed to make a slightly more careful analysis.
Following [1, Proof of Theorem 4.1] we infer

\[
\|V_n\| \geq \|(1, s_n, s_n^2, \ldots, s_n^n)\|
\]

and thus, setting \( i^2 = -1 \), the use of Cauchy-Schwartz leads to the following estimates

\[
\mu(V_n) \geq \|(1, s_n, s_n^2, \ldots, s_n^n)\| \frac{\|y\|}{\|V_n y\|}
\]

\[
\geq \frac{|(1, is_n, (is_n)^2, \ldots, (is_n)^n)y|}{\sqrt{n+1}}
\]

\[
= \frac{|q_n(i)|}{\sqrt{n+1}} \geq \frac{(1 + \sqrt{2})^n - (1 + \sqrt{2})^{-n}}{2\sqrt{n+1}}
\]

\[
\geq \frac{(1 + \sqrt{2})^n - (1 + \sqrt{2})^{n-1}}{2\sqrt{n+1}} = \frac{(1 + \sqrt{2})^{n-1}}{2\sqrt{n+1}}.
\]

Remark 2.3. – A philosophical (or may be qualitative) motivation for this bad
behavior of the polynomial canonical basis (irrespective of the choice of \( G_n \subset R \))
can be traced in the “quasi” linear dependence of the functions \( 1, x, \ldots, x^n \) over
the domain \( R \) with large \( n \). In fact, while the notion of linear dependence is theo-
retically boolean (on a given set of vectors, this property holds or does not hold
and only one of these facts is true), in Numerical Analysis we can talk of “quasi” linear dependence when a vector of the set can be approximated sufficiently well by a linear combination of the others. In this respect we should observe that the conditioning of a matrix gives a measure of the degree of linear dependence of the row or column vectors, which compose the given matrix. In our case, e.g. on [0, 1], when k is large the functions $x^k$ and $x^{k+1}$ are “quasi” linearly dependent because $\|x^k - x^{k+1}\|_{\infty,[0,1]} = (1 - 1/(k + 1))^k/(k + 1)$ is infinitesimal as k diverges but $\|x^j\|_{\infty,[0,1]} = 1$ for every $j \in \mathbb{N}$. In fact, with regard to a generic Haar system $\{\phi_0, \phi_1, \ldots, \phi_n\}$ on $I \subset \mathbb{R}$ and in order to deduce an exponential conditioning of the associated Vandermonde sequence $\{(\phi_j(x_i))_{i,j=0}^n\}_{n \in \mathbb{N}}$, the precise formal condition is that

$$\forall b > 0, \quad \frac{\min_{a_0, a_1, \ldots, a_{n-1}} \| \hat{\phi}_n - \sum_{j=0}^{n-1} a_j \hat{\phi}_j \|_{\infty,[-b,b] \cap I}}{\max_{j=0, \ldots, n} \| \hat{\phi}_j \|_{\infty,[-b,b] \cap I}}$$

is exponentially decreasing as a function of n. We notice that this condition is satisfied for $\hat{\phi}_j(x) = x^j$ and for a nontrivial real interval I.

As already observed, the latter bad functional behavior can be read directly in the conditioning of the matrices $V_n$ and also in the conditioning of the famous Hilbert matrix $H_n$ [8], which comes from the same set of polynomial monomials $\phi_j(x) = x^j$ via integrals, and for which a version of Theorem 2.1 can be stated and proved in the same way (but working directly on $H_n$ which is symmetric positive definite). For generalizations of the Hilbert matrix on unbounded intervals, the analysis can be much more involved and in that case it is probably much easier to follow Todd [6] and to prove first an explicit formula for the inverse of a Cauchy matrix; an other alternative is the use of the techniques by Taylor [5] or finally the application of the tools in [1].

A this point we should observe that the pathological behavior of the polynomial canonical basis $\hat{\phi}_j(x) = x^j$, $j = 0, \ldots, n$, is in actuality related to the domain only. Indeed if we shift to the complex field and we consider $\mathcal{G}_n$ formed by the $(n + 1)$-th roots of the unity, then the corresponding Vandermonde matrix is the Fourier matrix $F_n$, which is perfectly conditioned since $F_n$ is unitary and therefore $\mu(F_n) = 1$: indeed on the unit complex circle (where each node of $\mathcal{G}_n$ lies) the functions $1, z, \ldots, z^n$ are substantially linearly independent and more precisely they are orthonormal with respect to the $L^2$ scalar product expressed in the variable $\theta$ with $z^j = \exp(\text{i}j\theta)$, $\theta \in [0, 2\pi)$.

**REFERENCES**


Stefano Serra-Capizzano: Dipartimento di Fisica e Matematica, Università dell’Insubria - Sede di Como, Via Valleggio 11, 22100 Como, Italy (stefano.serrac@uninsubria.it, serra@mail.dm.unipi.it).

---

*Pervenuta in Redazione*

*il 17 luglio 2007*