Numerische Stabilität von Vandermonde-Systemen

Thomas Wienecke

geboren am 13.10.1992 in Berlin

Bachelorarbeit Mathematik

Erstgutachterin

Frau Prof. Dr. Gerlind Plonka-Hoch

Zweitgutachter

Herr Prof. Dr. Russell Luke

Abgabedatum

15.10.2014

Fakultät für Mathematik und Informatik der

Georg-August-Universität Göttingen

Inhaltsverzeichnis

1 Einleitung

Schon früh im Studium der numerischen Mathematik trifft man auf das Problem, $n+1$ Punkte in $\mathbb{R}$ durch ein Polynom n-ten Grades zu interpolieren. Bei der Modellierung dieser Aufgabe mit Hilfe der Monom-Basis entsteht ein lineares Gleichungssystem. Die zugehörige Koeffizienten-Matrix ist nur von den reellen Stützstellen abhängig und wird Vandermonde-Matrix genannt. Es zeigt sich schnell, dass diese Vandermonde-Matrix schlecht konditioniert ist und damit bei der Lösung des Interpolationsproblems große numerische Fehler hervorruft.

Anders verhält es sich, wenn man Vandermonde-Matrizen mit Stützstellen in der komplexen Ebene betrachtet. So ist die Vandermonde-Matrix zu den n-ten Einheitswurzeln sogar perfekt konditioniert bezüglich der Spektralnorm. Tatsächlich entstehen in einer Reihe von Anwendungen Vandermonde-Matrizen, die durch Knoten auf dem Einheitskreis definiert sind.

Forschungsergebnisse zur Konditionszahl von Vandermonde-Matrizen mit reellen und komplexen Stützstellen wurden 1990 in [5] zusammengefasst. Seither gab es kaum neue Erkenntnisse auf diesem Gebiet, obgleich diesbezügliche Fragestellungen von hoher Aktualität sind. Neue Erkenntnisse können bei der Entwicklung verbesserter numerischer Algorithmen für die Prony-Methode zur Parameterschätzung und zur Approximation hilfreich sein. Wie etwa verändert sich die Kondition der Vandermonde-Matrix bei Verteilungen, die von den Einheitswurzeln abweichen? Was passiert, wenn ausgehend von der äquidistanten Verteilung auf dem Einheitskreis ein Knoten um einen kleinen Winkel ausgelenkt wird? Diese Fragen sollen in der vorliegenden Arbeit untersucht werden.

Zu Beginn werden einige grundlegende Begriffe eingeführt. Es folgt eine Ungleichung für die Kondition bezüglich der Zeilensummennorm von Vandermonde-Matrizen mit beliebigen Knoten. Diese wird anschließend zur Berechnung der Kondition einiger reellwertiger Vandermonde-Matrizen verwendet. Im letzten Abschnitt wird eine Konfiguration von Stützstellen auf dem Einheitskreis untersucht, welche leicht vom äquidistanten Fall der n-ten Einheitswurzeln abweicht. Die Ergebnisse werden schließlich im Fazit zusammengefasst.

2 Grundlagen

2.1 Norm und Kondition

Zwei Matrixnormen werden im Laufe der Arbeit von entscheidener Bedeutung sein, weshalb wir uns zunächst an deren Definitionen erinnern.

Definition (Frobenius- und Zeilensummennorm)**.** Sei $A = (a_{kj})_{k,j=0}^{n-1} \in \mathbb{C}^{n\times n}$ eine Matrix. Die *Frobeniusnorm* von A ist definiert durch

$$\|A\|_F \coloneqq \sqrt{\sum_{k=0}^{n-1}\sum_{j=0}^{n-1} |a_{kj}|^2}. \tag{1}$$

Die *Zeilensummennorm* von A ist definiert durch

$$\|A\|_\infty \coloneqq \max_{k=0,\dots,n-1} \sum_{j=0}^{n-1} |a_{kj}| \,. \tag{2}$$

Bemerkung. Die Frobeniusnorm ist *unitär invariant*, d.h. für eine unitäre Matrix $U \in \mathbb{C}^{n\times n}$ und eine beliebige Matrix $A \in \mathbb{C}^{n\times n}$ gilt

$$\|AU\|_F = \|UA\|_F = \|A\|_F \,.$$

Im Folgenden wird ähnlich wie in [9, S. 205ff] der Begriff der Kondition einer Matrix motiviert. Dazu seien eine Vektorraumnorm auf $\mathbb{C}^n$ und eine submultiplikative Matrixnorm auf $\mathbb{C}^{n\times n}$ gegeben, die wir jeweils mit $\|\cdot\|$ bezeichnen. Die Matrixnorm sei dabei mit der Vektorraumnorm verträglich, d.h. es gelte $\|Ax\| \le \|A\|\,\|x\|$ für alle $A \in \mathbb{C}^{n\times n}$ und $x \in \mathbb{C}^n$. Wir betrachten ein lineares Gleichungssystem der Form $Ax = b$ mit invertierbarer Matrix $A \in \mathbb{C}^{n\times n}$, $n \in \mathbb{N}$, $b \in \mathbb{C}^n$ und gesuchtem Vektor $x \in \mathbb{C}^n$. Dabei nehmen wir $x \neq 0$ und $b \neq 0$ an. Weiter sei ein verfälschter Eingangsvektor $\tilde{b} = b + \Delta b$, $\Delta b \in \mathbb{C}^n$ mit relativem Fehler

$$\frac{\left\|\tilde{b} - b\right\|}{\|b\|} = \frac{\|\Delta b\|}{\|b\|} \le \delta$$

gegeben. Wir bezeichnen mit $\tilde{x} \in \mathbb{C}^n$ die Lösung des verfälschten Gleichungssystems

$A\tilde{x} = \tilde{b}$. Mit $\Delta x \coloneqq A^{-1}\Delta b$ erhalten wir wegen der Linearität von A^{-1}

$$\tilde{x} = A^{-1}\tilde{b} = A^{-1}b + A^{-1}\Delta b = x + \Delta x.$$

Gesucht ist nun ein Maß des relativen Fehlers der verfälschten Lösung $\tilde{x}$ in Abhängigkeit vom Eingangsfehler δ. Mit $\|\Delta x\| = \|A^{-1}\Delta b\| \leq \|A^{-1}\| \, \|\Delta b\|$ und $\|b\| = \|Ax\| \leq \|A\| \, \|x\|$, folgt für den relativen Fehler

$$\frac{\|\Delta x\|}{\|x\|} \leq \|A\| \left\|A^{-1}\right\| \frac{\|\Delta b\|}{\|b\|} \leq \|A\| \left\|A^{-1}\right\| \delta.$$

Dies motiviert die folgende

Definition (Kondition einer Matrix)**.** Sei $A \in \mathbb{C}^{n\times n}$. Dann ist die *Kondition von A bezüglich der Norm* $\|\cdot\|$ durch $\operatorname{cond}(A) \coloneqq \|A\| \, \|A^{-1}\|$ definiert.

Bemerkung. Mit den Bezeichnungen wie oben gilt dann $\frac{\|\Delta x\|}{\|x\|} \leq \operatorname{cond}(A) \cdot \delta$.

Bemerkung. Auch für den Fall, dass die Matrix A in verfälschter Form $A + \Delta A$ vorliegt und somit das Gleichungssystem $(A + \Delta A)(x + \Delta x) = (b + \Delta b)$ gelöst wird, lässt sich eine Ungleichung mit Hilfe der Konditionszahl herleiten. Dazu sei auf [9, S. 203ff] und [8, S. 54ff] verwiesen.

Bezeichnung. Für die Konditionen bezüglich der Frobenius- und der Zeilensummennorm verwenden wir die Schreibweisen $\operatorname{cond}_F(A)$ bzw. $\operatorname{cond}_\infty(A)$ für Matrizen $A \in \mathbb{C}^{n\times n}$.

2.2 Vandermonde-Matrizen

Definition (Vandermonde-Matrix)**.** Für $n \in \mathbb{N}$ und einen Vektor $z = (z_0, \ldots, z_{n-1}) \in \mathbb{C}^n$ sei die *Vandermonde-Matrix zu den Stützstellen (oder Knoten)* $z_0, \ldots, z_{n-1}$ durch

$$V(z) \coloneqq \begin{pmatrix} 1 & 1 & \ldots & 1 \\ z_0 & z_1 & \ldots & z_{n-1} \\ z_0^2 & z_1^2 & \ldots & z_{n-1}^2 \\ \vdots & \vdots & \ddots & \vdots \\ z_0^{n-1} & z_1^{n-1} & \ldots & z_{n-1}^{n-1} \end{pmatrix} \in \mathbb{C}^{n\times n}$$

definiert. Bezeichnet man die Komponenten der Vandermonde-Matrix mit $v_{kj} \in \mathbb{C}$, so ergibt sich $v_{kj} = z_j^k$ für $k, j = 0, \ldots, n-1$.

Lemma 2.1 ([7]). *Sei $z = (z_0, \ldots, z_{n-1}) \in \mathbb{C}^n$. Es gilt*

$$\det V(z) = \prod_{0 \le k < j \le n-1} (z_j - z_k)\,.$$

Beweis. Bezeichne mit $v_{kj} = z_j^k$ für $k, j = 0, \ldots, n-1$ die Elemente von $V(z)$. Der Beweis erfolgt durch vollständige Induktion nach $n \in \mathbb{N}$.

Induktionsanfang ($n = 1$):
Da das leere Produkt per Definition 1 ergibt, gilt $\det V(z) = 1 = \prod_{0 \le k < j \le 0}(z_j - z_k)$.

Induktionvoraussetzung: Sei die Behauptung für $n - 1 \in \mathbb{N}$ erfüllt.

Induktionsschritt ($n-1 \to n$): Durch Zeilenoperationen ändert sich der Wert der Determinante nicht. Wir ziehen daher in jeder Zeile das z_0-fache der vorherigen Zeile ab und erhalten

$$\det V(z) = \begin{vmatrix} 1 & 1 & \ldots & 1 \\ z_0 & z_1 & \ldots & z_{n-1} \\ z_0^2 & z_1^2 & \ldots & z_{n-1}^2 \\ \vdots & \vdots & \ddots & \vdots \\ z_0^{n-1} & z_1^{n-1} & \ldots & z_{n-1}^{n-1} \end{vmatrix} = \begin{vmatrix} 1 & 1 & \ldots & 1 \\ 0 & (z_1 - z_0) & \ldots & (z_{n-1} - z_0) \\ 0 & (z_1 - z_0) z_1^1 & \ldots & (z_{n-1} - z_0) z_{n-1}^1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & (z_1 - z_0) z_1^{n-2} & \ldots & (z_{n-1} - z_0) z_{n-1}^{n-2} \end{vmatrix}.$$

Die Entwicklung der Determinante nach der ersten Spalte mit Hilfe des Laplace'schen Entwicklungssatzes und unter Ausnutzung der Multilinearität der Determinante liefern nun die Behauptung:

$$\det V(z) = \prod_{r=0}^{n-1} (z_r - z_0) \begin{vmatrix} 1 & \ldots & 1 \\ z_1 & \ldots & z_{n-1} \\ z_1^2 & \ldots & z_{n-1}^2 \\ \vdots & \ddots & \vdots \\ z_1^{n-2} & \ldots & z_{n-1}^{n-2} \end{vmatrix} = \prod_{r=0}^{n-1} (z_r - z_0) \det V(z_1, \ldots, z_{n-1})$$

$$\overset{IV}{=} \prod_{r=0}^{n-1} (z_r - z_0) \prod_{1 \le k < j \le n-1} (z_j - z_k) = \prod_{0 \le k < j \le n-1} (z_j - z_k).$$

□

Korollar 2.2. *Die Vandermonde-Matrix $V(z)$ mit $z = (z_0, \ldots, z_{n-1}) \in \mathbb{C}^n$ ist genau dann invertierbar, wenn $z_k \neq z_j$ für alle $k \neq j$ gilt, d.h. wenn die Stützstellen z_k paarweise verschieden sind.*

Das folgende Lemma benötigen wir im weiteren Verlauf der Arbeit:

Lemma 2.3. *Sei $z = (z_0, \dots, z_{n-1}) \in \mathbb{C}^n$ mit z_k paarweise verschieden und $\alpha \in \mathbb{C}\backslash\{0\}$. Dann gilt*

$$V(\alpha z) = diag\Big(\alpha^0, \dots, \alpha^{n-1}\Big) \cdot V(z) \tag{3a}$$

und entsprechend

$$V(\alpha z)^{-1} = V(z)^{-1} \cdot diag\Big(\alpha^0, \alpha^{-1}, \dots, \alpha^{-n+1}\Big). \tag{3b}$$

Beweis. Wir setzen

$$V = (v_{kj})_{k,j=0}^{n-1} \coloneqq V(z) \text{ und } \tilde{V} = (\tilde{v}_{kj})_{k,j=0}^{n-1} \coloneqq V(\alpha z)\,.$$

Dann gilt für $k, j = 0, \dots, n-1$

$$\tilde{v}_{kj} = (\alpha z_j)^k = \alpha^k z_j^k = \alpha^k v_{kj},$$

d.h. wir können

$$\tilde{V} = \operatorname{diag}\Big(\alpha^0, \dots, \alpha^{n-1}\Big) \cdot V$$

schreiben. Mit $z_0, \dots, z_{n-1} \in \mathbb{C}$ sind auch $\alpha z_0, \dots, \alpha z_{n-1} \in \mathbb{C}$ paarweise verschieden, so dass $\tilde{V}$ nach Korollar 2.2 invertierbar ist und es folgt

$$\tilde{V}^{-1} = \Big(\operatorname{diag}\Big(\alpha^0, \dots, \alpha^{n-1}\Big) \cdot V\Big)^{-1} = V^{-1} \cdot \operatorname{diag}\Big(\alpha^0, \alpha^{-1}, \dots, \alpha^{-n+1}\Big).$$

□

3 Elementarsymmetrische Polynome

In diesem Abschnitt führen wir wie in [2] die *elementarsymmetrischen Polynome* ein. Diese liefern uns später eine explizite Darstellung der inversen Vandermonde-Matrix und sind Grundlage für den Beweis einer oberen Schranke von $\|V^{-1}\|_\infty$.

3.1 Definition und Eigenschaften

Definition (Elementarsymmetrische Polynome)**.** Sei $x = (x_1, \ldots, x_n) \in \mathbb{C}^n$. Wir definieren die *$r$-ten elementarsymmetrischen Polynome* $\sigma_r(x) = \sigma_r(x_1, \ldots, x_n)$ *mit* $r = 0, \ldots, n$ *in den* n *Variablen* x_j *mit* $j = 1, \ldots, n$ durch die Koeffizienten des Polynoms

$$p(z) = \prod_{k=1}^{n} (z + x_k) =: \sum_{r=0}^{n} \sigma_r(x_1, \ldots, x_n) z^{n-r}.$$

Bezeichnung. Als abkürzende Schreibweise setzen wir für $x = (x_1, \ldots, x_n) \in \mathbb{C}^n$

$$\sigma_r^j(x) \coloneqq \sigma_r(x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n).$$

Damit ist $\sigma_r^j(x)$ das r-te elementarsymmetrische Polynom in den $n-1$ Variablen x_k mit $k \in \{1, \ldots, n\} \setminus \{j\}$.

Bemerkung. Ausmultiplizieren des Polynoms und Koeffizientenvergleich liefern

$$\sigma_r(x_1, \ldots, x_n) = \sum_{1 \le j_1 < \cdots < j_r \le n} x_{j_1} \cdots x_{j_r} = \sum_{1 \le j_1 < \cdots < j_r \le n} \left(\prod_{k=1}^{r} x_{j_k} \right) \tag{4}$$

für $r \in \mathbb{N}$. Dabei ist der Fall $r = 0$ nicht als leere Summe, sondern als Summe mit leerem Produkt als einzigem Summanden zu verstehen, so dass $\sigma_0(x) = 1$ gilt. Dies entspricht gerade dem Höchstkoeffizienten des Polynoms aus der Definition.

Lemma 3.1. *Die elementarsymmetrischen Polynome erfüllen die Rekursionsformel*

$$\sigma_r(x_1, \ldots, x_n) = \sigma_r(x_1, \ldots, x_{n-1}) + x_n \sigma_{r-1}(x_1, \ldots, x_{n-1}). \tag{5}$$

Beweis. Wir verwenden Gleichung (4) und teilen die rechte Seite in zwei Gruppen von Summanden auf, je nachdem, ob x_n ein Faktor im Summand ist:

$$\begin{aligned}\sigma_r(x_1,\dots,x_n) &\overset{(4)}{=} \sum_{1\le j_1<\cdots<j_r\le n} x_{j_1}\cdots x_{j_r}\\ &= \sum_{1\le j_1<\cdots<j_r\le n-1} x_{j_1}\cdots x_{j_r} + \sum_{1\le j_1<\cdots<j_{r-1}\le n-1} x_{j_1}\cdots x_{j_r}\cdot x_n\\ &\overset{(4)}{=} \sigma_r(x_1,\dots,x_{n-1}) + x_n\cdot\sigma_{r-1}(x_1,\dots,x_{n-1}).\end{aligned}$$

□

Im Folgenden beweisen wir noch zwei technische Lemmata, die später nützlich sind.

Lemma 3.2. *Für $x=(x_1,\dots,x_n)\in\mathbb{C}^n$ und $\alpha\in\mathbb{C}$ gilt*

$$\sigma_r(\alpha x) = \alpha^r\cdot\sigma_r(x). \tag{6}$$

Beweis. Unter Verwendung von Gleichung (4) erhalten wir

$$\sigma_r(\alpha x) \overset{(4)}{=} \sum_{1\le j_1<\cdots<j_r\le n} \alpha x_{j_1}\cdots\alpha x_{j_r} = \alpha^r\cdot\sum_{1\le j_1<\cdots<j_r\le n} x_{j_1}\cdots x_{j_r} \overset{(4)}{=} \alpha^r\cdot\sigma_r(x).$$

□

Lemma 3.3. *Bezeichnen wir für $n\in\mathbb{N}$ die erste n-te Einheitswurzel mit $\omega_n \coloneqq e^{\frac{2\pi i}{n}}$, so gilt*

$$\left|\sigma_r(\omega_n^1,\dots,\omega_n^{n-1})\right| = 1 \tag{7}$$

für alle $r=0,\dots,n-1$.

Beweis. Nach Lemma 3.2 gilt

$$\left|\sigma_r(-\omega_n^1,\dots,-\omega_n^{n-1})\right| = \left|(-1)^r\sigma_r(\omega_n^1,\dots,\omega_n^{n-1})\right| = \left|\sigma_r(\omega_n^1,\dots,\omega_n^{n-1})\right|.$$

Nach Definition der elementarsymmetrischen Polynome ist $\sigma_r(-\omega_n^1,\dots,-\omega_n^{n-1})$ der $n-r-1$-te Koeffizient des Polynoms vom Grad $n-1$, das eindeutig durch die $n-1$ Nullstellen ω_n^k mit $k=1,\dots,n-1$ bestimmt ist. Wir zeigen, dass dies das Polynom $p(z)=\sum_{r=0}^{n-1} z^r$ ist.
Tatsächlich ist $\omega_n^k = e^{2\pi ik/n}$ für $k=1,\dots,n-1$ eine Nullstelle von $p(z)$, denn wegen

$\omega_n^k \neq 1$ folgt mit Hilfe der geometrischen Reihe

$$p(\omega_n^k) = \sum_{r=0}^{n-1} \left(\omega_n^k\right)^r = \frac{1 - \left(\omega_n^k\right)^n}{1 - (\omega_n^k)} = \frac{1 - (\omega_n^n)^k}{1 - (\omega_n^k)} = 0.$$

Damit ist gezeigt, dass für $r = 0, \ldots, n-1$

$$\left|\sigma_r(\omega_n^1, \ldots, \omega_n^{n-1})\right| = \left|\sigma_r(-\omega_n^1, \ldots, -\omega_n^{n-1})\right| = 1$$

gilt. □

Korollar 3.4. *Mit* $\omega_n := e^{\frac{2\pi i}{n}}$ *gilt*

$$\sum_{r=0}^{n-1} \left|\sigma_r(\omega_n^1, \ldots, \omega_n^{n-1})\right| \overset{(7)}{=} \sum_{r=0}^{n-1} 1 = n. \tag{8}$$

3.2 Eine Abschätzung der Betragssumme elementarsymmetrischer Polynome

Die folgende Abschätzung liefert uns im nächsten Abschnitt eine obere Schranke der Zeilensummennorm inverser Vandermonde-Matrizen.

Lemma 3.5 ([1])**.** *Sei* $x = (x_1, \ldots, x_n) \in \mathbb{C}^n$*. Dann gilt*

$$\sum_{k=0}^{n} |\sigma_k(x)| \leq \prod_{k=1}^{n} (1 + |x_k|)\,. \tag{9}$$

Gleichheit gilt, wenn $x_k = r_k \cdot e^{i\varphi}$ *für ein festes* $\varphi \in \mathbb{R}$ *und beliebige* $r_k \in \mathbb{R}_+$*, d.h. wenn alle* x_k *auf einer Halbgeraden durch den Nullpunkt in der komplexen Ebene liegen.*

Bemerkung. Für die letzte Aussage des Lemmas lässt sich sogar Äquivalenz zeigen, d.h. Gleichheit gilt *genau dann*, wenn $x_k = r_k \cdot e^{i\varphi}$ für ein festes $\varphi \in \mathbb{R}$ und beliebige $r_k \in \mathbb{R}_+$. Da diese Aussage jedoch für die vorliegende Arbeit nicht von Bedeutung ist, beweisen wir nur die schwächere Form des Lemmas.

Beweis des Lemmas 3.5. Zum Beweis verwenden wir vollständige Induktion nach $n \in \mathbb{N}$ und Lemma 3.1.

Induktionsanfang ($n = 1$): Es ist $(z + x_0) = \sigma_0(x_0) \cdot z^1 + \sigma_1(x_0) \cdot z^0$, also $\sigma_0(x_0) = 1$ und $\sigma_1(x_0) = x_0$. Damit ergibt sich: $\sum_{k=0}^{1} |\sigma_k(x_0)| = |1| + |x_0| = 1 + |x_0|$.

Induktionvoraussetzung: Sei die Behauptung für $n-1 \in \mathbb{N}$ erfüllt.

Induktionsschritt $(n-1 \to n)$: Unter Verwendung von $\sigma_0(x_1, \ldots, x_n) = 1$ und $\sigma_n(x_1, \ldots, x_n) = x_1 \cdots x_n$ folgt:

$$\begin{aligned}
\prod_{k=1}^{n} (1 + |x_k|) &= (1 + |x_n|) \cdot \prod_{k=1}^{n-1} (1 + |x_k|) \\
&\overset{IV}{\geq} (1 + |x_n|) \cdot \sum_{k=0}^{n-1} |\sigma_k(x_1, \ldots, x_{n-1})| \\
&= \sum_{k=0}^{n-1} |\sigma_k(x_1, \ldots, x_{n-1})| + \sum_{k=1}^{n} |x_n \cdot \sigma_{k-1}(x_1, \ldots, x_{n-1})| \\
&\geq \sum_{k=1}^{n-1} |\sigma_k(x_1, \ldots, x_{n-1}) + x_n \cdot \sigma_{k-1}(x_1, \ldots, x_{n-1})| \\
&\qquad + |\sigma_0(x_1, \ldots, x_{n-1})| + |x_n \cdot \sigma_{n-1}(x_1, \ldots, x_{n-1})| \\
&\overset{(5)}{=} \sum_{k=1}^{n-1} |\sigma_k(x_1, \ldots, x_n)| + |\sigma_0(x_1, \ldots, x_n)| + |\sigma_n(x_1, \ldots, x_n)| \\
&= \sum_{k=0}^{n} |\sigma_k(x_1, \ldots, x_{n-1})| \,.
\end{aligned}$$

Um den Spezialfall $x_k = r_k \cdot e^{i\varphi}$ der Behauptung zu beweisen, führen wir die Induktion analog durch, allerdings mit der neuen Induktionsbehauptung, dass Gleichheit gelte. Es bleibt nur zu zeigen, dass die zweite obige Abschätzung zur Gleichheit wird, d.h. dass $\sigma_k(x_1, \ldots, x_{n-1})$ und $x_n \cdot \sigma_{k-1}(x_1, \ldots, x_{n-1})$ linear abhängig sind und die Dreiecksungleichung mit Gleichheit

$$\begin{aligned}
&|\sigma_k(x_1, \ldots, x_{n-1})| + |x_n \cdot \sigma_{k-1}(x_1, \ldots, x_{n-1})| \\
&= |\sigma_k(x_1, \ldots, x_{n-1}) + x_n \cdot \sigma_{k-1}(x_1, \ldots, x_{n-1})|
\end{aligned}$$

gilt.
Mit Gleichung (4) sieht man sofort ein, dass $\sigma_k(r_1, \ldots, r_{n-1}) \in \mathbb{R}_+$ für alle $k \in \mathbb{N}$ gilt. Weiter verwenden wir Lemma 3.2 und stellen fest, dass

$$\sigma_k(x_1, \ldots, x_{n-1}) = \sigma_k(e^{i\varphi} \cdot (r_1, \ldots, r_{n-1})) \overset{(6)}{=} e^{ki\varphi} \cdot \underbrace{\sigma_k(r_1, \ldots, r_{n-1})}_{>0}$$

und

$$x_n \cdot \sigma_{k-1}(x_1, \ldots, x_{n-1}) = x_n \cdot \sigma_{k-1}(e^{i\varphi} \cdot (r_1, \ldots, r_{n-1})) \overset{(6)}{=} e^{k \cdot i\varphi} \cdot \underbrace{r_n \sigma_{k-1}(r_1, \ldots, r_{n-1})}_{>0}$$

in die gleiche Richtung $e^{k \cdot i\varphi}$ zeigen, also linear abhängig sind. □

4 Eine Ungleichung für die Kondition von Vandermonde-Matrizen

In diesem Abschnitt sei ein n-elementiger Vektor $z = (z_0, \ldots, z_{n-1}) \in \mathbb{C}^n$ gegeben. Die zugehörige Vandermonde-Matrix sei mit $V \coloneqq V(z)$ bezeichnet. Ziel ist es, eine Abschätzung der Zeilensummennorm inverser Vandermonde-Matrizen und damit eine Abschätzung der Kondition $\text{cond}_\infty(V)$ zu finden. Wir orientieren uns dabei an der Herleitung aus [5].

4.1 Inversion der Vandermonde-Matrix

Es werden zunächst die *Lagrange-Polynome* zu den Knoten $z_0, \ldots, z_{n-1}$ eingeführt, deren Koeffizienten sich als Einträge der inversen Vandermonde-Matrix herausstellen werden.

Definition (Lagrange-Polynome)**.** Für $z = (z_0, \ldots, z_{n-1}) \in \mathbb{C}^n$ definiere die *Lagrange-Polynome* durch

$$l_j(z) \coloneqq \prod_{\substack{r=0 \\ r \neq j}}^{n-1} \frac{z - z_r}{z_j - z_r} \text{ für } j = 0, \ldots, n-1.$$

Bemerkung.

1. Es gilt $l_j \in \Pi_{n-1}$, wobei Π_{n-1} den Raum der Polynome bis Grad $n-1$ bezeichne.
2. Einfaches Nachrechnen liefert $l_j(z_r) = \delta_{jr}$, wobei δ_{jr} das Kronecker-Delta sei.

Wegen $l_j \in \Pi_{n-1}$ können die Lagrange-Polynome ausmultipliziert als

$$l_j(z) = \sum_{r=0}^{n-1} u_{jr} z^r \tag{10}$$

mit den Koeffizienten $u_{jr} \in \mathbb{C}$ geschrieben werden. Im folgenden Lemma zeigt sich, dass diese Koeffizienten genau den Einträgen der inversen Vandermonde-Matrix entsprechen.

Lemma 4.1 ([3])**.** *Sei $z = (z_0, \dots, z_{n-1}) \in \mathbb{C}^n$ und seien l_j die zugehörigen Lagrange-Polynome für $j = 0, \dots, n-1$ mit den Koeffizienten u_{jr} wie in (10). Dann ist die Koeffizienten-Matrix $U = (u_{jr})_{j,r=0}^{n-1}$ die Inverse der Vandermonde-Matrix $V \coloneqq (z_j^k)_{k,j=0}^{n-1}$.*

Beweis. Wir betrachten das Gleichungssystem $V\alpha = y$ mit $\alpha = (\alpha_0, \dots, \alpha_{n-1}) \in \mathbb{C}^n$ und $y = (y_0, \dots, y_{n-1}) \in \mathbb{C}^n$. Zum Beweis muss die Gleichung $Uy = \alpha$ für alle $\alpha, y \in \mathbb{C}^n$ gezeigt werden. Es gilt für $j = 0, \dots, n-1$

$$\begin{aligned}\sum_{k=0}^{n-1} u_{jk} y_k &= \sum_{k=0}^{n-1} u_{jk} \sum_{r=0}^{n-1} z_r^k \alpha_r = \sum_{r=0}^{n-1} \alpha_r \sum_{k=0}^{n-1} u_{jk} z_r^k \\ &= \sum_{r=0}^{n-1} \alpha_r l_j(z_r) = \sum_{r=0}^{n-1} \alpha_r \delta_{jr} = \alpha_j,\end{aligned}$$

wie behauptet. □

Unter Verwendung von Gleichung (10) und mit Hilfe der elementarsymmetrischen Polynome kann eine explizite Darstellung der inversen Vandermonde-Matrix angegeben werden. Dazu schreiben wir für $j = 0, \dots, n-1$

$$\sum_{r=0}^{n-1} u_{jr} z^r = l_j(z) = \prod_{\substack{r=0 \\ r \neq j}}^{n-1} \frac{z - z_r}{z_j - z_r} = \Pi_j \cdot \prod_{\substack{r=0 \\ r \neq j}}^{n-1} (z - z_r)$$

mit

$$\Pi_j \coloneqq \prod_{\substack{r=0 \\ r \neq j}}^{n-1} (z_j - z_r)^{-1}. \tag{11}$$

Nutzen wir Lemma 3.2, so erhalten wir

Lemma 4.2. *Für $j, r = 0, \dots, n-1$ gilt*

$$\begin{aligned}u_{jr} &= \Pi_j \cdot \sigma_{n-1-r}^j(-z_0, \dots, -z_{n-1}) \\ &\overset{(6)}{=} (-1)^{n-1-r} \cdot \Pi_j \cdot \sigma_{n-1-r}(z_0, \dots, z_{j-1}, z_{j+1}, \dots, z_{n-1}),\end{aligned} \tag{12}$$

mit $\Pi_j \coloneqq \prod_{\substack{r=0 \\ r \neq j}}^{n-1} (z_j - z_r)^{-1}$.

4.2 Eine Abschätzung der Zeilensummennorm inverser Vandermonde-Matrizen

Die Grundlage für eine obere Schranke von $\left\|V^{-1}\right\|_\infty$ haben wir bereits im Abschnitt über die elementarsymmetrischen Polynome geschaffen. Mit Hilfe der dort erbrachten Abschätzung der Betragssumme elementarsymmetrischer Polynome können wir den folgenden Satz beweisen. Die Aussage und die Idee des Beweises sind dabei aus [5, S. 196-197] entnommen.

Satz 4.3. *Seien $z_0, \ldots, z_{n-1} \in \mathbb{C}$ paarweise verschieden. Mit $V \coloneqq V(z)$ gilt*

$$\left\|V^{-1}\right\|_\infty \leq \max_{j=0,\ldots,n-1} \left(\prod_{\substack{k=0 \\ k\neq j}}^{n-1} \frac{1+|z_k|}{|z_j - z_k|} \right). \tag{13}$$

Gleichheit gilt, wenn $z_k = r_k \cdot e^{i\varphi}$ für ein festes $\varphi \in \mathbb{R}$ und $r_k \in \mathbb{R}_+$ mit $k = 0, \ldots, n-1$ gilt.

Beweis. Die explizite Darstellung von V^{-1} in Gleichung (12) und die Ungleichung über elementarsymmetrische Polynome (9) aus Lemma 3.5 liefern die Behauptung:

$$\begin{aligned}
\left\|V^{-1}\right\|_\infty &= \max_{j=0,\ldots,n-1} \sum_{r=0}^{n-1} |u_{jr}| \\
&\overset{(12)}{=} \max_{j=0,\ldots,n-1} \sum_{r=0}^{n-1} \left|(-1)^{n-1-r} \Pi_j \sigma^j_{n-1-r}\right| = \max_{j=0,\ldots,n-1} |\Pi_j| \sum_{r=0}^{n-1} \left|\sigma^j_{n-1-r}\right| \\
&\overset{(9)}{\leq} \max_{j=0,\ldots,n-1} |\Pi_j| \prod_{\substack{k=0 \\ k\neq j}}^{n-1} (1+|z_k|) \overset{(11)}{=} \max_{j=0,\ldots,n-1} \prod_{\substack{k=0 \\ k\neq j}}^{n-1} \frac{1+|z_k|}{|z_j - z_k|}.
\end{aligned}$$

□

Für die Herleitung einer unteren Schranke stützen wir uns auf [4]. Anders als dort formuliert, kann die Aussage jedoch nur für Vandermonde-Matrizen mit Stützstellen $z_k \neq 0$ für alle $k = 0, \ldots, n-1$ bewiesen werden. Denn in [4] wird Jensens Formel verwendet, welche eine Aussage über analytische Funktionen f liefert für die $f(0) \neq 0$ gilt. Diese Voraussetzung ist verletzt, wenn 0 als Stützstelle der Vandermonde-Matrix gewählt wird.

Zunächst sei ohne Beweis an Jensens Formel in Integralform erinnert.

Satz 4.4 ([6]). *Sei $\rho \in \mathbb{R}_+$ und $F : \mathbb{C} \mapsto \mathbb{C}$ eine auf $|x| \leq \rho$ analytische Funktion mit $F(0) \neq 0$. Bezeichne mit $\zeta_1, \ldots, \zeta_n \in \mathbb{C}$ die Nullstellen von F, die $|\zeta_j| \leq \rho$ erfüllen. Dabei kommen die Nullstellen entsprechend ihrer Vielfachheit eventuell mehrfach vor. Jensens Formel in Integralform lautet dann*

$$\frac{1}{2\pi} \int_0^{2\pi} \log \left| F\left(\rho e^{i\phi}\right) \right| d\phi = \log |F(0)| + \sum_{j=1}^{n} \log \frac{\rho}{|\zeta_j|}. \tag{14}$$

Unter Verwendung von Satz 4.4 beweisen wir einen Zusammenhang zwischen der Koeffizientensumme eines Polynoms vom Grad n und dessen Nullstellen. Dieser Zusammenhang liefert uns im nächsten Satz schließlich die Abschätzung der Zeilensummennorm inverser Vandermonde-Matrizen.

Lemma 4.5 ([4]). *Sei $p(z) = \sum_{j=0}^{n} a_j z^j$ ein Polynom n-ten Grades mit $a_j \in \mathbb{C}$ für $j = 0, \ldots, n$ und $a_n \neq 0$ sowie Nullstellen $\zeta_j \in \mathbb{C} \setminus \{0\}$ mit $j = 1, \ldots, n$. Dann gilt*

$$\sum_{j=0}^{n} |a_j| \geq |a_n| \prod_{j=1}^{n} \max\left(1, |\zeta_j|\right). \tag{15}$$

Beweis. Ohne Einschränkung können wir annehmen, dass die Nullstellen sortiert vorliegen, so dass

$$|\zeta_1| \leq \cdots \leq |\zeta_r| \leq 1 < |\zeta_{r+1}| \leq \cdots \leq |\zeta_n|$$

gilt. Wir identifizieren p mit F und wählen $\rho = 1$ in Satz 4.4. Mit $\zeta_j \neq 0$ für $j = 1, \ldots, n$ aus der Voraussetzung folgt sofort $p(0) \neq 0$. Weiterhin ist p als Polynom analytisch auf ganz $\mathbb{C}$ und somit insbesondere auch für $|z| \leq 1$. Tatsächlich kann also Satz 4.4 angewendet werden. Zusammen mit $\log \frac{1}{z} = -\log z$ liefert dieser

$$\log |a_0| = \log |p(0)| = \sum_{k=1}^{r} \log |\zeta_k| + \frac{1}{2\pi} \int_0^{2\pi} \log \left| p\left(e^{i\varphi}\right) \right| d\varphi$$

oder äquivalent dazu

$$|a_0| \prod_{k=1}^{r} |\zeta_k|^{-1} = \exp\left(\frac{1}{2\pi} \int_0^{2\pi} \log \left| p\left(e^{i\varphi}\right) \right| d\varphi \right). \tag{16}$$

Mit Hilfe der Nullstellen ζ_j können wir p als Produkt seiner Linearfaktoren darstellen:

$$p(z) = \sum_{j=0}^{n} a_j z^j = a_n \prod_{k=1}^{n} (z - \zeta_k).$$

Die Auswertung an $z = 0$ liefert dann

$$a_0 = p(0) = a_n(-1)^n \prod_{k=1}^{n} \zeta_k,$$

so dass wir (16) vereinfacht darstellen können:

$$|a_n| \prod_{k=r+1}^{n} |\zeta_k| = \exp\left(\frac{1}{2\pi} \int_0^{2\pi} \log\left|p\left(e^{i\varphi}\right)\right| d\varphi\right).$$

Mit

$$\exp\left(\frac{1}{2\pi} \int_0^{2\pi} \log\left|p\left(e^{i\varphi}\right)\right| d\varphi\right) \leq \max_{0\leq\varphi\leq 2\pi} \left|p\left(e^{i\varphi}\right)\right| = \max_{0\leq\varphi\leq 2\pi} \left|\sum_{j=0}^{n} a_j e^{i\varphi}\right| \leq \sum_{j=0}^{n} |a_j| \quad (17)$$

folgt nun wie behauptet

$$\sum_{j=0}^{n} |a_j| \geq |a_n| \prod_{k=1}^{n} \max(1, \zeta_k).$$

□

Nun können wir eine Abschätzung der Zeilensummennorm der inversen Vandermonde-Matrix angeben und beweisen.

Satz 4.6. *Seien $z_0, \ldots, z_{n-1} \in \mathbb{C} \setminus \{0\}$ paarweise verschieden. Mit $V := V(z)$ gilt*

$$\left\|V^{-1}\right\|_\infty \geq \max_{j=0,\ldots,n-1} \left(\prod_{\substack{k=0 \\ k\neq j}}^{n-1} \frac{\max(1, |z_k|)}{|z_j - z_k|} \right). \quad (18)$$

Beweis. Wir wählen ein festes $j \in \{0, \ldots, n-1\}$ und betrachten das Polynom

$$p(z) = \sum_{k=0}^{n-1} a_k z^k := \Pi_j \cdot \prod_{\substack{k=0 \\ k\neq j}}^{n-1} (z - z_k) \overset{(12)}{=} \sum_{k=0}^{n-1} (-1)^{n-1-k} \cdot \Pi_j \cdot \sigma_{n-1-k}^{j}(z_0, \ldots, z_{n-1}) \cdot z^k,$$

d.h. die Koeffizienten von p ergeben sich für $k = 0, \ldots, n-1$ durch

$$a_k = (-1)^{n-1-k} \cdot \Pi_j \cdot \sigma_{n-1-k}^{j}(z_0, \ldots, z_{n-1}).$$

Offensichtlich hat p die $n-1$ Nullstellen $\zeta_r := z_r$ für $r \in \{0, \ldots, n-1\} \setminus \{j\}$. Mit

Ungleichung (15) aus Lemma 4.5 folgt

$$\begin{aligned}\sum_{r=0}^{n-1} |u_{jr}| &\overset{(12)}{=} \sum_{r=0}^{n-1} \left|(-1)^{n-1-r}\Pi_j \sigma_{n-1-r}^j\right| \\ &= \sum_{k=0}^{n-1} |a_k| \overset{(15)}{\geq} |a_n| \prod_{\substack{r=0 \\ r\neq j}}^{n-1} \max\left(1, |\zeta_r|\right) \\ &= |\Pi_j| \underbrace{\left|\sigma_0^j\right|}_{=1} \prod_{\substack{r=0 \\ r\neq j}}^{n-1} \max\left(1, |z_r|\right) = \prod_{\substack{r=0 \\ r\neq j}}^{n-1} \frac{\max\left(1, |z_r|\right)}{|z_j - z_r|}.\end{aligned}$$

Da diese Ungleichung für alle $j \in \{0, \ldots, n-1\}$ erfüllt ist, folgt die Behauptung:

$$\max_{j=0,\ldots,n-1} \sum_{r=0}^{n-1} |u_{jr}| \geq \max_{j=0,\ldots,n-1} \prod_{\substack{r=0 \\ r\neq j}}^{n-1} \frac{\max\left(1, |z_r|\right)}{|z_j - z_r|}.$$

□

5 Vandermonde-Matrizen mit reellen Stützstellen

Als Anwendung von Satz 4.3 werden wir nun die Konditionen einiger Vandermonde-Matrizen zu reellen Stützstellen berechnen. Die Beispiele sind dabei aus [5, S. 197-199] entnommen.

Zunächst beweisen wir ein allgemeines Lemma über die Zeilensummennorm von Vandermonde-Matrizen mit Stützstellen innerhalb des komplexen Einheitskreises. Betrachtet man nämlich nur Stützstellen $z_j \in \mathbb{C}$ mit $|z_j| \leq 1$, so gilt $\|V(z)\|_\infty = n$. Somit muss in diesem Fall zur Bestimmung von $\text{cond}_\infty(V(z))$ nur noch die Norm der Inversen $\left\|V(z)^{-1}\right\|_\infty$ berechnet werden. Dies gilt insbesondere für reelle Stützstellen $x = (x_0, \ldots, x_{n-1}) \in \mathbb{R}^n$ mit $|x_j| \leq 1$ für $j = 0, \ldots, n-1$.

Lemma 5.1. *Seien Stützstellen $z = (z_0, \ldots, z_{n-1}) \in \mathbb{C}^n$ mit $|z_j| \leq 1$ für alle $j = 0, \ldots, n-1$ gegeben. Dann gilt*

$$\|V(z)\|_\infty = n. \tag{19}$$

Beweis. Wegen $\left|z_j^k\right| \leq \left|z_j^r\right|$ für $k > r$ und alle $j = 0, \ldots, n-1$ folgt

$$\|V(z)\|_\infty = \max_{k=0,\ldots,n-1} \left(\sum_{j=0}^{n-1} \left|z_j^k\right| \right) = \sum_{j=0}^{n-1} \left|z_j^0\right| = n.$$

□

5.1 Nicht-negative Stützstellen

Betrachte die Stützstellen $z_j = x_j \in \mathbb{R}_+$ für $j = 0, \ldots, n-1$ mit $x_k \neq x_j$ für $k \neq j$. Diese liegen auf einer Halbgeraden durch den Ursprung und erfüllen damit die Zusatzbedingung von Satz 4.3, so dass bei der oberen Schranke von (13) Gleichheit gilt. Mit Hilfe dieses Satzes können wir also $\left\|V^{-1}\right\|_\infty$ explizit berechnen.

Lemma 5.2. *Ist V die zu den Stützstellen $x_j \in \mathbb{R}$ mit $x_j \geq 0$ gehörige Vandermonde-Matrix und definiert man*

$$p(z) \coloneqq \prod_{j=0}^{n-1} (z - x_j),$$

so folgt

$$\left\| V^{-1} \right\|_\infty = \frac{|p(-1)|}{\min_{j=0,\dots,n-1} \{(1 + x_j) \, |p'(x_j)|\}}.$$

Beweis. Es gilt

$$|p(-1)| = \left| \prod_{j=0}^{n-1} (-1 - x_j) \right| = \prod_{j=0}^{n-1} (1 + x_j).$$

Weiter ist nach der Produktregel

$$p'(z) = \sum_{k=0}^{n-1} \left(\prod_{\substack{j=0 \\ j \neq k}}^{n-1} (z - x_j) \right),$$

also

$$|p'(x_k)| = \prod_{\substack{j=0 \\ j \neq k}}^{n-1} |x_k - x_j|.$$

Insgesamt folgt mit Satz 4.3

$$\begin{aligned} \frac{|p(-1)|}{\min_{j=0,\dots,n-1} \{(1 + x_j) \, |p'(x_j)|\}} &= \frac{\prod_{k=0}^{n-1} (1 + x_k)}{\min_{j=0,\dots,n-1} \{(1 + x_j) \prod_{\substack{k=0 \\ k \neq j}}^{n-1} |x_j - x_k|\}} \\ &= \max_{j=0,\dots,n-1} \left(\prod_{\substack{k=0 \\ k \neq j}}^{n-1} \frac{(1 + x_k)}{|x_j - x_k|} \right) \overset{(13)}{=} \left\| V(x)^{-1} \right\|_\infty. \end{aligned}$$

□

Beispiel 5.3 (Harmonische Stützstellen)**.** Seien die Stützstellen $x_k = \frac{1}{k}$ mit $k = 1, \dots, n$ gegeben. Mit den Bezeichnungen wie in Lemma 5.2 gilt dann

$$|p(-1)| = \prod_{k=1}^{n} \left(1 + \frac{1}{k} \right) = n + 1,$$

wie eine einfache Induktion zeigt. Für $\delta_k \coloneqq (1+x_k)\,|p'(x_k)|$ gilt

$$\begin{aligned}\delta_k &= \left(1+\frac{1}{k}\right)\prod_{\substack{r=1\\r\neq k}}^{n}\left|\frac{1}{k}-\frac{1}{r}\right| = \left(\frac{k+1}{k}\right)\prod_{\substack{r=1\\r\neq k}}^{n}\left|\frac{r-k}{rk}\right| \\ &= \left(\frac{k+1}{k^n}\right)\frac{k}{n!}\prod_{\substack{r=1\\r\neq k}}^{n}|r-k| = \left(\frac{k+1}{k^n}\right)\frac{k}{n!}(n-k)!\,(k-1)! \\ &= \left(\frac{(k+1)!\,(n-k)!}{k^n n!}\right).\end{aligned}$$

Weiter folgt dann

$$\min_{k=1,\dots,n}\delta_k \leq \delta_n = \frac{n+1}{n^n}$$

und schließlich mit Lemma 5.2

$$\operatorname{cond}_\infty(V) = \|V\|_\infty \cdot \left\|V^{-1}\right\|_\infty \geq n\cdot(n+1)\frac{n^n}{n+1} = n^{n+1}.$$

Beispiel 5.4 (Äquidistante Stützstellen [3])**.** Betrachte die Stützstellen $x_k = \frac{k-1}{n-1}$ mit $k = 1,\dots,n$. Ähnliche Untersuchungen wie im vorherigen Beispiel liefern die asymptotische Formel

$$\operatorname{cond}_\infty(V(x)) \sim \frac{\sqrt{2}}{4\pi}\cdot 8^n \text{ für } n\to\infty.$$

Für eine ausführliche Herleitung sei auf [3, S. 344f] verwiesen.

5.2 Symmetrische Stützstellen

Es zeigt sich, dass die Kondition der Vandermonde-Matrix verbessert werden kann, wenn man die Stützstellen symmetrisch um den Ursprung anordnet.

Analog zu Lemma 5.2 lässt sich für den Fall symmetrischer reeller Stützstellen, die folgende Aussage zeigen. Der Beweis kann in [3, S. 341] nachgelesen werden.

Lemma 5.5 ([3], ohne Beweis)**.** *Sei $x = (x_1,\dots,x_n) \in \mathbb{R}^n$ mit $x_j + x_{n+1-j} = 0$ für $j = 1,\dots,n$. Dann gilt mit $p(x) \coloneqq \prod_{k=1}^n (x-x_k)$*

$$\left\|V(x)^{-1}\right\|_\infty = \frac{|p(i)|}{\min_{x_k\geq 0}\left\{\frac{1+x_k^2}{1+x_k}\,|p'(x_k)|\right\}},$$

wobei hier mit i die imaginäre Einheit bezeichnet sei.

Beispiel 5.6 (Äquidistante, symmetrische Stützstellen [5])**.** Seien $x_k = 1 - \frac{2(k-1)}{n-1}$ für $k = 1, \ldots, n$. Dann ergibt sich unter Verwendung von Lemma 5.5

$$\mathrm{cond}_\infty(V(x)) \sim \frac{1}{\pi} e^{-\frac{1}{4}\pi} e^{n\left(\frac{1}{4}\pi + \frac{1}{2}\ln 2\right)} \text{ für } n \to \infty.$$

Anstelle der exponentiellen Wachstumsrate von 8 im nicht-negativen Fall, wird mit symmetrischen äquidistanten Stützstellen eine Wachstumsrate von

$$\exp\left(\frac{1}{4}\pi + \frac{1}{2}\ln 2\right) = 3.1017\ldots$$

erreicht. Die Kondition kann noch weiter verbessert werden, wenn die sogenannten *Tschebycheff Knoten* verwendet werden.

Beispiel 5.7 (Tschebyscheff Stützstellen [3])**.** Wir betrachten die Knoten $x_k = \cos\left(\frac{(2k-1)\pi}{2n}\right)$ für $k = 1, \ldots, n$. Unter Anwendung von Lemma 5.5 lässt sich die asymptotische Formel

$$\mathrm{cond}_\infty(V(x)) \sim \frac{3^{\frac{3}{4}}}{4}\left(1 + \sqrt{2}\right)^n \text{ für } n \to \infty$$

beweisen. Die Wachstumsrate beträgt hier $1 + \sqrt{2} = 2.4142...$

In den vorangegangen Beispielen wurde im besten Fall eine exponentielle Wachstumsrate der Konditionszahl erreicht. Es wird vermutet, dass für Vandermonde-Matrizen zu n reellen Stützstellen stets $\mathrm{cond}_\infty(V) > 2^{\frac{n}{2}}$ gilt, wie in [5, S. 199] ausgeführt ist.

6 Vandermonde-Matrizen mit Stützstellen auf dem Einheitskreis

Wie das vorherige Kapitel zeigt, sind Vandermonde-Matrizen mit reellen Stützstellen schlecht konditioniert und werden daher hauptsächlich für theoretische Betrachtungen verwendet.

Anders verhält es sich, wenn man Knoten in der komplexen Ebene zulässt. Wählt man als Stützstellen die n-ten Einheitswurzeln, so erreicht die Vandermonde-Matrix sogar die perfekte Kondition 1 bezüglich der Spektralnorm. Auch bezüglich der Frobenius- und der Zeilensummennorm wächst die Kondition dieser Vandermonde-Matrizen nur linear in n.

In diesem Abschnitt untersuchen wir Vandermonde-Matrizen mit Knoten auf dem komplexen Einheitskreis, die jedoch von dem perfekten Fall der n-ten Einheitswurzeln abweichen. Für diese speziellen Konfigurationen leiten wir explizite Formeln zur Berechnung der Kondition bezüglich der Zeilensummennorm und der Frobeniusnorm her.

6.1 Vandermonde-Matrizen zu den n-ten Einheitswurzeln

Wir beweisen zunächst die Aussagen über die Frobenius- und die Zeilensummennorm von Vandermonde-Matrizen zu den Knoten der n-ten Einheitswurzeln.

Lemma 6.1. *Für Knoten auf dem Einheitskreis $z = (e^{i\varphi_0}, \dots, e^{i\varphi_{n-1}}) \in \mathbb{C}$ mit $\varphi_j \in \mathbb{R}$ für $j = 0, \dots, n-1$ gilt*

$$\|V(z)\|_F = n. \tag{20}$$

Beweis. Es gilt

$$\|V(z)\|_F^2 = \sum_{k=0}^{n-1} \sum_{j=0}^{n-1} \left| \left(e^{i\varphi_j}\right)^k \right|^2 = \sum_{k=0}^{n-1} \sum_{j=0}^{n-1} 1 = n^2.$$

□

Lemma 6.2. *Sei $n \in \mathbb{N}$ und $z = (\omega_n^0, \omega_n^1, \dots, \omega_n^{n-1}) \in \mathbb{C}^n$ der Vektor der n-ten Einheitswurzeln, d.h. $\omega_n := e^{\frac{2\pi i}{n}}$. Dann gilt für die Kondition bezüglich der Frobeniusnorm*

$$cond_F(V(z)) = n.$$

Beweis. Wir setzen $W = (w_{jr})_{j,r=0}^{n-1} \in \mathbb{C}^{n\times n}$ mit $w_{jr} \coloneqq \frac{1}{n}\omega_n^{-jr}$ und zeigen $V^{-1} = W$. Seien dazu $a_{kr} \in \mathbb{C}$ die Elemente des Matrixproduktes $A = VW \in \mathbb{C}^{n\times n}$. Dann gilt für $k, r = 0, \ldots, n-1$

$$\begin{aligned} a_{kr} &= \sum_{j=0}^{n-1} v_{kj} w_{jr} = \sum_{j=0}^{n-1} \frac{1}{n}\omega_n^{kj}\omega_n^{-jr} \\ &= \frac{1}{n}\sum_{j=0}^{n-1} \omega_n^{j(k-r)} = \frac{1}{n}\sum_{j=0}^{n-1} \left(\omega_n^{k-r}\right)^j \\ &= \frac{1}{n} \cdot \begin{cases} n, \text{ falls } k-r \equiv 0 \mod n \\ 0, \text{ falls } k-r \not\equiv 0 \mod n \end{cases} \\ &= \begin{cases} 1, \text{ falls } k = r \\ 0, \text{ falls } k \neq r. \end{cases} \end{aligned}$$

Damit ist $V^{-1} = W$ gezeigt. Nach Lemma 6.1 gilt $\|V\|_F = n$. Komplexe Konjugation ändert nichts an der Frobeniusnorm einer Matrix, da diese nur von den Beträgen der Matrixelemente abhängt. Wir erhalten damit $\left\|V^{-1}\right\|_F = \|W\|_F = \frac{1}{n}\left\|\overline{V}\right\|_F = 1$, was die Behauptung zeigt. □

Lemma 6.3. *Sei erneut $z = (\omega_n^0, \omega_n^1, \ldots, \omega_n^{n-1}) \in \mathbb{C}^n$ mit $\omega_n = e^{\frac{2\pi i}{n}}$. Dann gilt für die Kondition bezüglich der Zeilensummennorm*

$$cond_\infty(V(z)) = n.$$

Beweis. Wie im Beweis von Lemma 6.2 gezeigt, gilt $V(z)^{-1} = \frac{1}{n}\overline{V(z)}$. Genau wie bei der Frobeniusnorm, ändert komplexe Konjugation nichts an der Zeilensummennorm einer Matrix, da diese nur von den Beträgen der Matrixelemente abhängt. Insbesondere ergibt sich $\left\|\overline{V(z)}\right\|_\infty = \|V(z)\|_\infty$. Mit Lemma 5.1 folgt dann

$$\operatorname{cond}_\infty(V(z)) = \|V(z)\|_\infty \cdot \frac{1}{n} \cdot \left\|\overline{V(z)}\right\|_\infty \overset{(19)}{=} n.$$

□

6.2 Invarianz der Kondition unter Rotation der Knoten

Im Folgenden zeigen wir, dass sich die Kondition einer Vandermonde-Matrix bezüglich der Zeilensummennorm und aller unitär invarianten Normen weder unter Rotation noch unter Spiegelung der Knoten verändert.

Lemma 6.4. *Die Kondition der Vandermonde-Matrix bezüglich der Zeilensummennorm ist invariant unter Multiplikation der Knoten mit einer komplexen Zahl $\alpha = e^{i\varphi} \in \mathbb{C}$ mit $\varphi \in \mathbb{R}$.*

Beweis. Sei $z = (z_0, \ldots, z_{n-1}) \in \mathbb{C}^n$. Wegen $\left|e^{i\varphi}\right| = 1$ für alle $\varphi \in \mathbb{R}$ gilt

$$\begin{aligned}\|V(\alpha z)\|_\infty &= \max_{k=0,\ldots,n-1} \sum_{j=0}^{n-1} \left|\left(e^{i\varphi} z_j\right)^k\right| = \max_{k=0,\ldots,n-1} \sum_{j=0}^{n-1} \left|e^{ik\varphi}\right| \left|z_j^k\right| \\ &= \max_{k=0,\ldots,n-1} \sum_{j=0}^{n-1} \left|z_j^k\right| = \|V(z)\|_\infty .\end{aligned}$$

Für die Zeilensummennorm der inversen Vandermonde-Matrix erinnern wir uns an Gleichung (3b) aus Lemma 2.3: $V(\alpha z)^{-1} = V(z)^{-1} \cdot \mathrm{diag}(\alpha^0, \alpha^{-1}, \ldots, \alpha^{-n+1})$. Bezeichnen wir für $j, r = 0, \ldots, n-1$ mit $u_{jr} \in \mathbb{C}$ erneut die Einträge von V^{-1}, so ist sofort ersichtlich, dass

$$\begin{aligned}\left\|V(\alpha z)^{-1}\right\|_\infty &= \max_{j=0,\ldots,n-1} \sum_{r=0}^{n-1} |u_{jr}| \, |\alpha^{-r}| \\ &= \max_{j=0,\ldots,n-1} \sum_{r=0}^{n-1} |u_{jr}| = \left\|V(\alpha)^{-1}\right\|_\infty .\end{aligned}$$

Wie behauptet folgt insgesamt

$$\begin{aligned}\mathrm{cond}_\infty(V(\alpha z)) &= \|V(\alpha z)\|_\infty \left\|V(\alpha z)^{-1}\right\|_\infty \\ &= \|V(z)\|_\infty \left\|V(z)^{-1}\right\|_\infty = \mathrm{cond}_\infty(V(z)) .\end{aligned}$$

□

Lemma 6.5. *Sei $\|\cdot\|$ eine Matrix-Norm, die invariant unter Multiplikation mit unitären Matrizen ist, d.h. für jede Matrix $A \in \mathbb{C}^{n\times n}$ und jede unitäre Matrix $U \in \mathbb{C}^{n\times n}$ gilt $\|UA\| = \|AU\| = \|A\|$. Weiter seien ein Vektor $z = (z_0, \ldots, z_{n-1}) \in \mathbb{C}^n$ und eine komplexe Zahl $\alpha \in \mathbb{C}$ mit Betrag $|\alpha| = 1$ gegeben. Dann gilt für die Kondition bezüglich der Norm $\|\cdot\|$*

$$\mathit{cond}(V(\alpha z)) = \mathit{cond}(V(z)) ,$$

d.h. die Kondition der Vandermonde-Matrix ist in diesem Fall invariant unter Multiplikation der Knoten mit einer komplexen Zahl α vom Betrag $|\alpha| = 1$.

Beweis. Nach Lemma 2.3 gelten

$$\|V(\alpha z)\| \overset{(3a)}{=} \left\|\operatorname{diag}\left(\alpha^0, \ldots, \alpha^{n-1}\right) \cdot V(z)\right\|$$

und

$$\left\|V(\alpha z)^{-1}\right\| \overset{(3b)}{=} \left\|V(z)^{-1} \cdot \operatorname{diag}\left(\alpha^0, \alpha^{-1}, \ldots, \alpha^{-n+1}\right)\right\|.$$

Mit $e^{i\varphi} \coloneqq \alpha$ für ein $\varphi \in \mathbb{R}$ stellen wir leicht fest, dass $\operatorname{diag}(\alpha^0, \alpha^1, \ldots, \alpha^{n-1})$ eine unitäre Matrix ist:

$$\begin{aligned}
\operatorname{diag}\left(\alpha^0, \alpha^1, \ldots, \alpha^{n-1}\right)^H &= \operatorname{diag}\left(1, e^{i\varphi}, \ldots, e^{(n-1)i\varphi}\right)^H \\
&= \operatorname{diag}\left(1, e^{-i\varphi}, \ldots, e^{-(n-1)i\varphi}\right) \\
&= \operatorname{diag}\left(1, e^{i\varphi}, \ldots, e^{(n-1)i\varphi}\right)^{-1} \\
&= \operatorname{diag}\left(\alpha^0, \alpha^1, \ldots, \alpha^{n-1}\right)^{-1},
\end{aligned}$$

wobei A^H die adjungierte Matrix von A bezeichne. Insgesamt folgt

$$\begin{aligned}
\operatorname{cond}(V(\alpha z)) &= \|V(\alpha z)\| \left\|V(\alpha z)^{-1}\right\| \\
&= \left\|\operatorname{diag}\left(\alpha^0, \ldots, \alpha^{n-1}\right) \cdot V(z)\right\| \left\|V(z)^{-1} \cdot \operatorname{diag}\left(\alpha^0, \alpha^{-1}, \ldots, \alpha^{-n+1}\right)\right\| \\
&= \|V(z)\| \left\|V(z)^{-1}\right\| = \operatorname{cond}(V(z)).
\end{aligned}$$

□

6.3 Vandermonde-Matrizen aus $(n-1)$ Einheitswurzeln und einem Ausreißer

Im Folgenden untersuchen wir die Kondition einer Vandermonde-Matrix zu den Knoten der n-ten Einheitswurzeln, wobei einer der Knoten von seiner ursprünglichen Position um einen Winkel $\varphi \in \left(-\frac{2\pi}{n}, \frac{2\pi}{n}\right)$ auf dem Einheitskreis ausgelenkt wird. Wie bereits in Lemma 6.4 gezeigt, ändern weder Drehungen noch Spiegelungen der Knoten die Kondition der Vandermonde-Matrix. Daher können wir, ohne das Problem zu beschränken, stets den Knoten auslenken, welcher der n-ten Einheitswurzel 1 zugeordnet ist.

Wir betrachten also für $\delta \in (-1, 1)$ den Knotenvektor $z(\delta) = (z_0(\delta), z_1, \ldots, z_{n-1}) \in \mathbb{C}^n$ mit $z_0(\delta) = e^{2\pi i\delta/n}$ und $z_j = e^{2\pi ij/n}$ für $j = 1, \ldots, n-1$. Die dazu gehörende Vandermonde-Matrix bezeichnen wir in diesem Abschnitt mit $V(\delta) \coloneqq V(z(\delta))$. Häufig werden wir auf die explizite Angabe der Abhängigkeit von δ verzichten und beispielsweise nur kurz z_0

anstelle von $z_0(\delta)$ schreiben.

6.3.1 Kondition bezüglich der Zeilensummennorm

Vandermonde-Matrizen mit Stützstellen innerhalb des komplexen Einheitskreises haben die Zeilensummennorm n, wie bereits in Lemma 5.1 gezeigt wurde. Dies trifft offensichtlich auch in unserem Fall zu, so dass wir für die Berechnung der Kondition nur noch die Zeilensummennorm der inversen Vandermonde-Matrix $V(\delta)^{-1}$ ermitteln müssen.

Lemma 6.6. *Sei $z(\delta) = (z_0(\delta), \dots, z_{n-1}) \in \mathbb{C}^n$ mit $z_0(\delta) = e^{2\pi i\delta/n}$ und $z_j = e^{2\pi ij/n}$ für $j = 1, \dots, n-1$. Weiter seien $u_{jr} \in \mathbb{C}$ für $j, r = 0, \dots, n-1$ die Elemente der inversen Vandermonde-Matrix $V(z(\delta))^{-1}$. Dann gilt*

$$\sum_{r=0}^{n-1} |u_{0r}| = n \cdot \prod_{k=1}^{n-1} |z_0 - z_k|^{-1} = n \cdot \prod_{j=1}^{n-1} \frac{1}{2 \cdot \sin\left(\frac{\pi(j-\delta)}{n}\right)}.$$

Beweis. Nach Gleichung (12) gilt

$$u_{0r} = (-1)^{n-1-r} \cdot \Pi_0 \cdot \sigma^0_{n-1-r}(z)$$

mit

$$\Pi_0 = \prod_{k=1}^{n-1} |z_0 - z_k|^{-1}.$$

Sei $k \in \{1, \dots, n-1\}$ und $\varphi \coloneqq \frac{\pi(k-\delta)}{n}$. Dann gilt

$$\begin{aligned} |z_0 - z_k|^2 &= (z_0 - z_k)(\bar{z}_0 - \bar{z}_k) = |z_0|^2 + |z_k|^2 - z_0\bar{z}_k - z_k\bar{z}_0 \\ &= 2 - (e^{2\varphi i} + e^{-2\varphi i}) = 2 - ((e^{\varphi i})^2 + (e^{-\varphi i})^2) \\ &= 2 - ((e^{\varphi i} - e^{-\varphi i})^2 + 2e^{(\varphi-\varphi)i}) = (e^{\varphi i} - e^{-\varphi i})^2 \\ &= 4 \cdot \sin^2(\varphi) = 4 \cdot \sin^2\left(\frac{\pi(k-\delta)}{n}\right), \end{aligned}$$

also

$$\Pi_0 = \prod_{k=1}^{n-1} |z_0 - z_k|^{-1} = \prod_{k=1}^{n-1} \frac{1}{2 \cdot \sin\left(\frac{\pi(k-\delta)}{n}\right)}. \tag{21}$$

Zusammen mit der Aussage des Korollars 3.4 folgt nun die Behauptung:

$$\begin{aligned}\sum_{r=0}^{n-1} |u_{0r}| &= \sum_{r=0}^{n-1} \left|(-1)^{n-1-r} \cdot \Pi_0 \cdot \sigma_{n-1-r}(z_1, \ldots, z_{n-1})\right| \\ &= \left(\sum_{r=0}^{n-1} |\sigma_{n-1-r}(z_1, \ldots, z_{n-1})|\right) \cdot |\Pi_0| \\ &\overset{(8)}{=} n \cdot \prod_{j=1}^{n-1} \frac{1}{2 \cdot \sin\left(\frac{\pi(j-\delta)}{n}\right)}.\end{aligned}$$

□

Der Beweis der folgenden Vermutung kann im Rahmen dieser Arbeit nicht erbracht werden. Numerische Untersuchungen konnten die Aussage jedoch nicht widerlegen.

Vermutung 6.7. *Es gilt*

$$cond_\infty(V(\delta)) = n^2 \prod_{j=1}^{n-1} |z_0(\delta) - z_j|^{-1} = n^2 \prod_{j=1}^{n-1} \frac{1}{2 \cdot \sin\left(\frac{\pi(j-\delta)}{n}\right)}. \tag{22}$$

Bemerkung. Zum Beweis der Vermutung muss gezeigt werden, dass

$$\left\|V(\delta)^{-1}\right\|_\infty = n \cdot \prod_{j=1}^{n-1} \frac{1}{2 \cdot \sin\left(\frac{\pi(j-\delta)}{n}\right)}$$

gilt. Bezeichnen wir erneut mit u_{jr} für $j, r = 0, \ldots, n-1$ die Elemente von $V(\delta)^{-1}$, so bleibt zu zeigen, dass für $j = 1, \ldots, n-1$

$$\sum_{r=0}^{n-1} |u_{jr}| \leq \sum_{r=0}^{n-1} |u_{0r}|$$

gilt. Lemma 6.6 liefert dann die Behauptung.
Der Ansatz, die beiden Faktoren Π_j und $\sum_{r=0}^{n-1} |\sigma_r^j|$ einzeln abzuschätzen, d.h. die Ungleichungen

$$\Pi_j \leq \Pi_0$$

und

$$\sum_{r=0}^{n-1} \left|\sigma_r^j\right| \leq \sum_{r=0}^{n-1} \left|\sigma_r^0\right|$$

für alle $j = 1, \ldots, n-1$ zu beweisen, muss im Allgemeinen scheitern, wie die Figuren 6.1 und 6.2 für den Fall $n = 5$ zeigen. Die schwarzen Graphen, die den Fall $j = 0$ darstellen,

nehmen dort niemals den Maximalwert unter allen Graphen von $j = 0, \ldots, n-1$ an. Die Figur 6.3 legt jedoch nahe, dass die Behauptung des Satzes tatsächlich wahr ist, denn der Graph des Produktes ist stets für $j = 0$ maximal.

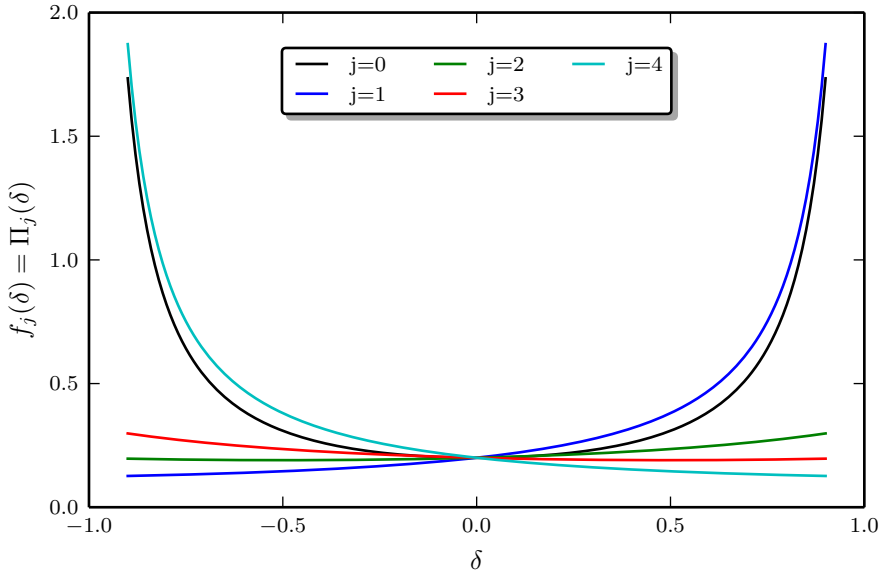

Abbildung 6.1: Erster Faktor der Betragssumme der j-ten Zeile von $V(\delta)^{-1}$ in Abhängigkeit der Auslenkung $\delta \in (-0.9, 0.9)$ für den Fall $n = 5$.

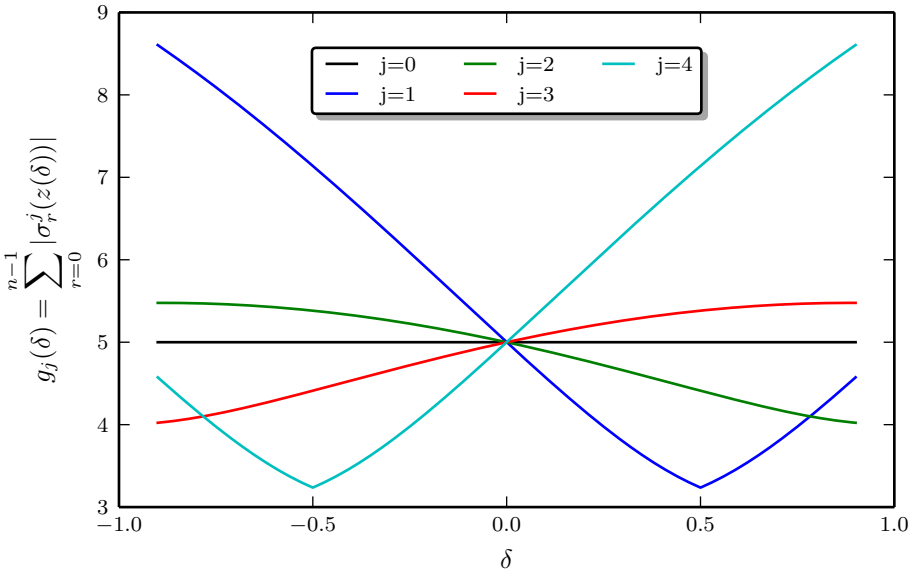

Abbildung 6.2: Zweiter Faktor der Betragssumme der j-ten Zeile von $V(\delta)^{-1}$ in Abhängigkeit der Auslenkung $\delta \in (-0.9, 0.9)$ für den Fall $n = 5$.

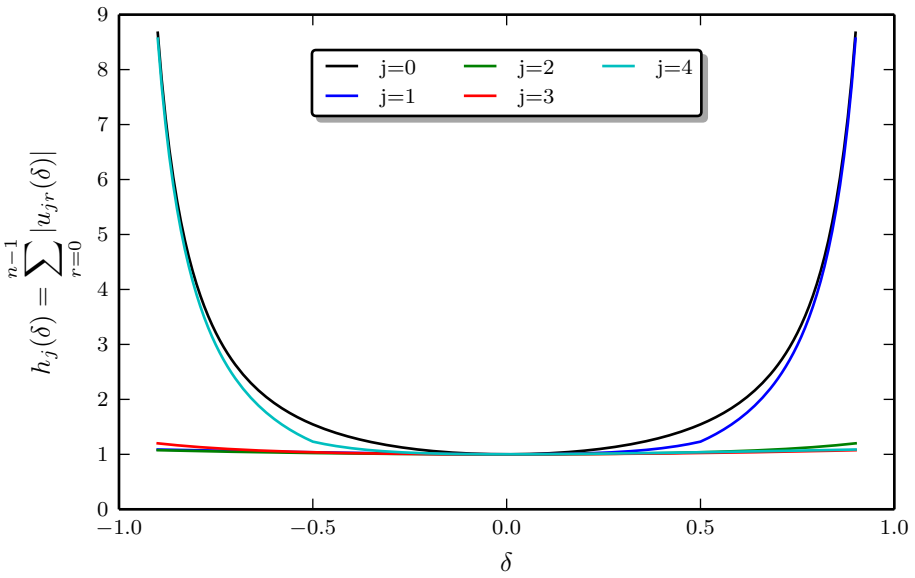

Abbildung 6.3: Betragssumme der j-ten Zeile von $V(\delta)^{-1}$ in Abhängigkeit der Auslenkung $\delta \in (-0.9, 0.9)$ für den Fall $n = 5$.

Der Beweis kann also nur erbracht werden, indem das gesamte Produkt abgeschätzt wird:

$$\Pi_j \sum_{r=0}^{n-1} \left|\sigma_r^j\right| \leq \Pi_0 \sum_{r=0}^{n-1} \left|\sigma_r^0\right| \text{ für } j = 1, \ldots, n-1.$$

In Tabelle 6.1 sind für ausgewählte $n \in \mathbb{N}$ die numerischen Fehler angegeben, welche durch Verwendung der Formel aus der Vermutung entstehen. Für $\delta \in \{0, 0.01, \ldots, 0.99\}$ wurde dabei jeweils die Kondition der Vandermonde-Matrix $V(\delta)$ direkt berechnet und die Differenz mit der Formel aus Vermutung 6.7 gebildet. Die Beträge dieser Differenzen wurden für alle δ aufaddiert und durch n dividiert. Die resultierenden Werte sind in Tabelle 6.1 eingetragen. Die entstandenen Fehler sind in der Größenordnung numerischer Rechenungeauigkeiten bei der Verwendung von Gleitkommazahlen einzuordnen und stützen damit Gleichung (22).

n	Relativer Fehler
2	$3.499422973618493 \cdot 10^{-13}$
5	$7.613465413669474 \cdot 10^{-13}$
10	$1.033839680530946 \cdot 10^{-12}$
20	$2.271782761908981 \cdot 10^{-12}$
30	$2.729076224265251 \cdot 10^{-12}$
40	$3.564437633940542 \cdot 10^{-12}$
50	$3.976907692049281 \cdot 10^{-12}$

Tabelle 6.1: Relativer Fehler bei der Berechnung der Konditionszahlen $\text{cond}_\infty(V(\delta))$ mit Hilfe von Gleichung (22) für verschiedene $n \in \mathbb{N}$.

6.3.2 Kondition bezüglich der Frobeniusnorm

Wir wollen nun eine explizite Formel für die Kondition der Vandermonde-Matrix $V(\delta)$ bezüglich der Frobeniusnorm beweisen. Dazu sei an die Definition der Frobeniusnorm (1) erinnert. Bereits in Lemma 6.1 wurde gezeigt, dass die Frobeniusnorm einer Vandermonde-Matrix mit n Knoten auf dem Einheitskreis n beträgt. Schwierigkeiten kann also nur die Norm der inversen Vandermonde-Matrix bereiten. Wir betrachten im Folgenden die j-te Zeile von $V(\delta)^{-1}$ für $j = 0$ und $j \in \{1, \ldots, n-1\}$ gesondert.

Lemma 6.8. *Bezeichnen wir mit $u_{jr} \in \mathbb{C}$ für $j, r = 0, \ldots, n-1$ die Elemente der inversen Vandermonde-Matrix $V(\delta)^{-1}$, so gilt*

$$\sum_{r=0}^{n-1} |u_{0r}|^2 = n \cdot \prod_{k=1}^{n-1} \frac{1}{4 \cdot \sin^2\left(\pi \frac{k-\delta}{n}\right)}.$$

Beweis. Mit Lemma 3.3 und Gleichung (21) aus dem Beweis von Lemma 6.6 erhalten wir

$$\begin{aligned}\sum_{r=0}^{n-1} |u_{0r}|^2 &= \sum_{r=0}^{n-1} |\Pi_0|^2 \left|\sigma^0_{n-1-r}(z)\right|^2 \\ &\overset{(7)}{=} |\Pi_0|^2 \sum_{r=0}^{n-1} 1^2 = n \cdot |\Pi_0|^2 \\ &\overset{(21)}{=} n \cdot \prod_{k=1}^{n-1} \frac{1}{4 \cdot \sin^2\left(\pi \frac{k-\delta}{n}\right)}.\end{aligned}$$

□

Als Vorbereitung für den Fall $j \in \{1, \dots, n-1\}$ benötigen wir noch zwei Hilfslemmata.

Lemma 6.9. *Sei* $\omega_n \coloneqq e^{\frac{2\pi i}{n}}$. *Definiere* $\tau_j \in \mathbb{C}$ *für* $j = 0, \dots, n-1$ *durch*

$$\tau_j \coloneqq \prod_{\substack{k=0 \\ k \neq j}}^{n-1} \left(\omega_n^j - \omega_n^k\right).$$

Dann gilt $|\tau_j| = n$ *für alle* $j = 0, \dots, n-1$.

Beweis. Wir zeigen $|\tau_j| = |\tau_0|$ für alle $j = 1, \dots, n-1$ und $\tau_0 = n$. Tatsächlich gilt für $j = 1, \dots, n-1$

$$\begin{aligned}|\tau_j| &= \left|\tau_j \cdot \omega_n^{-(n-1)\cdot j}\right| = \prod_{\substack{k=0 \\ k \neq j}}^{n-1} \left|\omega_n^0 - \omega_n^{k-j}\right| \\ &\overset{k'=k-j}{=} \prod_{\substack{k'=-j \\ k' \neq 0}}^{n-1-j} \left|1 - \omega_n^{k'}\right| = \prod_{\substack{k'=0 \\ k' \neq 0}}^{n-1} \left|1 - \omega_n^{k'}\right| = |\tau_0|,\end{aligned}$$

wobei wir im letzten Schritt nur die Reihenfolge der Faktoren verändern und die 2π-Periodizität von $e^{i\varphi}$ ausnutzen.

Mit $p(z) \coloneqq \prod_{k=1}^{n-1}(z - \omega_n^k)$ können wir $\tau_0 = p(1)$ schreiben. Bereits im Beweis von Lemma 3.3 haben wir gesehen, dass

$$p(z) = \prod_{k=1}^{n-1}(z - \omega_n^k) = \sum_{k=0}^{n-1} z^k$$

gilt. Damit folgt sofort

$$\tau_0 = p(1) = \sum_{k=0}^{n-1} 1^k = n.$$

□

Lemma 6.10. *Für $j \in \{1, \dots, n-1\}$ gilt*

$$\sum_{r=0}^{n-1} \cos\left(\frac{2\pi jr}{n} - \varphi\right) = 0.$$

Beweis. Mit $\omega_n \coloneqq e^{\frac{2\pi i}{n}}$ und $p(z) \coloneqq \sum_{k=0}^{n-1} z^n = \prod_{j=1}^{n-1}(z - \omega_n^j)$ gilt

$$\begin{aligned}\sum_{r=0}^{n-1} \cos\left(\frac{2\pi jr}{n} - \varphi\right) &= \sum_{r=0}^{n-1} \Re\left(e^{\frac{2\pi ijr}{n} - i\varphi}\right) = \Re\left(\sum_{r=0}^{n-1} e^{\frac{2\pi ijr}{n} - i\varphi}\right) \\ &= \Re\left(e^{-i\varphi} \sum_{r=0}^{n-1} \left(\omega_n^j\right)^r\right) = \Re\left(e^{-i\varphi} p\left(\omega_n^j\right)\right) = 0,\end{aligned}$$

wobei $\Re(z)$ den Realteil von z bezeichne. □

Lemma 6.11. *Sei $z(\delta) = (z_0(\delta), z_1, \dots, z_{n-1}) \in \mathbb{C}^n$ wie zuvor. Erneut bezeichnen wir mit u_{jr} die Elemente der inversen Vandermonde-Matrix $V(\delta)^{-1}$. Dann gilt für $j = 1, \dots, n-1$*

$$\sum_{r=0}^{n-1} |u_{jr}|^2 = \frac{\rho(\delta)^2 + 1}{n} \tag{23}$$

mit

$$\rho(\delta) \coloneqq \frac{\sin\left(\pi \frac{\delta}{n}\right)}{\sin\left(\pi \frac{j-\delta}{n}\right)}.$$

Beweis. In diesem Beweis werden wir der Übersicht wegen meist auf die explizite Erwähnung der Abhängigkeit von δ verzichten und nur kurz $z_0 = z_0(\delta)$ schreiben.
Sei $j \in \{1, \dots, n-1\}$ fixiert. Wir erinnern uns zunächst daran, dass mit Lemma 3.2 für $r = 0, \dots, n-1$

$$\left|\sigma_r^j(z)\right| = \left|(-1)^r \sigma_r^j(-z)\right| = \left|\sigma_r^j(-z)\right|$$

gilt. Weiterhin ist nach Definition $\sigma_r^j(-z)$ der $n-j-1$-te Koeffizient des Polynoms

$$p(z) = \sum_{k=0}^{n-1} a_k z^k \coloneqq \prod_{\substack{k=0 \\ k \neq j}}^{n-1} (z - z_k) = \left(\prod_{k=1}^{n-1} (z - z_k)\right) \cdot \frac{z - z_0(\delta)}{z - z_j}.$$

Im Beweis von Lemma 3.3 haben wir bereits die Gleichheit

$$\prod_{k=1}^{n-1}(z-z_k)=\sum_{k=0}^{n-1}z^k$$

gezeigt, so dass wir damit

$$p(z)=\sum_{k=0}^{n-1}a_kz^k=\sum_{k=0}^{n-1}z^k\cdot\frac{z-z_0(\delta)}{z-z_j}$$

erhalten. Multiplikation der Gleichung mit $(z-z_j)$ liefert die äquivalente Darstellung

$$(z-z_j)\cdot\sum_{k=0}^{n-1}a_kz^k=(z-z_0(\delta))\cdot\sum_{k=0}^{n-1}z^k$$

oder ausmultipliziert

$$\sum_{k=0}^{n-1}\left(a_kz^{k+1}-a_kz_jz^k\right)=\sum_{k=0}^{n-1}\left(z^{k+1}-z_0z^k\right).$$

Durch Vergleich der Koeffizienten erhalten wir die $n+1$ Gleichungen

$$\begin{aligned}
-a_0z_j&=-z_0\\
a_0-a_1z_j&=1-z_0\\
a_1-a_2z_j&=1-z_0\\
&\vdots\\
a_{n-2}-a_{n-1}z_j&=1-z_0\\
a_{n-1}&=1
\end{aligned}$$

und in etwas kompakterer Form

$$\begin{aligned}
-a_0z_j&=-z_0,\\
a_{k-1}-a_kz_j&=1-z_0\text{ für }k\in\{1,\dots,n-1\},\\
a_{n-1}&=1.
\end{aligned}$$

Es zeigt sich nun, dass für $k=1,\dots,n$

$$a_{n-k}=(1-z_0)\left(\sum_{r=0}^{k-2}z_j^r\right)+z_j^{k-1}\tag{24}$$

gilt. Für $k = 1$ ist dies offensichtlich wahr. Ist die Gleichung für ein $k \in \{1, \dots, n\}$ bereits gezeigt, so folgt mit Hilfe der Gleichung $a_{n-(k+1)} - a_{n-k} z_j = 1 - z_0$, dass unsere Behauptung tatsächlich auch für $a_{n-(k+1)}$ gilt:

$$\begin{aligned} a_{n-k-1} &= 1 - z_0 + z_j a_{n-k} \\ &= (1 - z_0) + z_j \left((1 - z_0) \left(\sum_{r=0}^{k-2} z_j^r \right) + z_j^{k-1} \right) \\ &= (1 - z_0) + (1 - z_0) \left(\sum_{r=0}^{k-2} z_j^{r+1} \right) + z_j^k \\ &= (1 - z_0) \left(\sum_{r=0}^{k-1} z_j^r \right) + z_j^k. \end{aligned}$$

Induktion nach $k = 1, \dots, n$ liefert also die Richtigkeit der Gleichung (24). Wir identifizieren nun die elementarsymmetrischen Polynome gemäß ihrer Definition durch $\sigma_r^j(-z) = a_{n-(r+1)}$ für $r = 0, \dots, n-1$ und formen unter Verwendung der geometrischen Reihe weiter um:

$$\begin{aligned} \sigma_r^j(-z) &= (1 - z_0) \left(\sum_{l=0}^{r-1} z_j^l \right) + z_j^r \\ &= (1 - z_0) \frac{1 - z_j^r}{1 - z_j} + z_j^r = \frac{(1 - z_0)(1 - z_j^r) + (1 - z_j) z_j^r}{1 - z_j} \\ &= \frac{1 + z_0 z_j^r - z_0 - z_j^{r+1}}{1 - z_j} = \frac{(1 - z_0) + (z_0 - z_j) z_j^r}{1 - z_j}. \end{aligned}$$

Damit gilt für alle $r = 0, \dots, n-1$

$$\begin{aligned} |u_{j,n-1-r}| &\overset{(12)}{=} \left| \Pi_j \cdot \sigma_r^j(-z) \right| \\ &= \left| \Pi_j \cdot \left(\frac{(1 - z_0) + (z_0 - z_j) z_j^r}{(1 - z_j)} \right) \right| \\ &= \left| \frac{(1 - z_0) + (z_0 - z_j) z_j^r}{(z_j - z_0) \cdot ((z_j - 1)(z_j - z_1) \dots (z_j - z_{n-1}))} \right| \\ &=: \left| \frac{(1 - z_0) + (z_0 - z_j) z_j^r}{(z_j - z_0) \cdot \tau_j} \right| \\ &= \left| \frac{1}{\tau_j} \left(\frac{1 - z_0}{z_j - z_0} - z_j^r \right) \right|, \end{aligned}$$

mit

$$\tau_j \coloneqq \prod_{\substack{k=0 \\ k\neq j}}^{n-1} \left(\exp\left(\frac{2\pi i j}{n}\right) - \exp\left(\frac{2\pi i k}{n}\right)\right).$$

Wir schreiben nun $\rho e^{i\varphi} \coloneqq \frac{1-z_0}{z_j - z_0}$ mit $\rho, \varphi \in \mathbb{R}$, $\rho \geq 0$ und erhalten zusammen mit $|\tau_j| = n$ aus Lemma 6.9

$$\begin{aligned} |u_{j,n-1-r}|^2 &= \frac{1}{n^2}\left|\rho e^{i\varphi} - z_j^r\right|^2 = \frac{1}{n^2}\left(\rho^2 + 1 - \rho\left(e^{i\cdot\left(\varphi - \frac{2\pi j r}{n}\right)} + e^{i\cdot\left(\frac{2\pi j r}{n} - \varphi\right)}\right)\right) \\ &= \frac{1}{n^2}\left(\rho^2 + 1 - 2\rho\cos\left(\frac{2\pi j r}{n} - \varphi\right)\right). \end{aligned}$$

Mit Hilfe von Lemma 6.10 folgt insgesamt für die j-te Zeile von $V(\delta)^{-1}$

$$\begin{aligned} \sum_{r=0}^{n-1} |u_{j,n-1-r}|^2 &= \sum_{r=0}^{n-1} \frac{1}{n^2}\left(\rho^2 + 1 - 2\rho\cos\left(\frac{2\pi j r}{n} - \varphi\right)\right) \\ &= \frac{\rho^2+1}{n} - \frac{2\rho}{n^2}\sum_{r=0}^{n-1}\cos\left(\frac{2\pi j r}{n} - \varphi\right) = \frac{\rho^2+1}{n}. \end{aligned}$$

Nach Definition ist $\rho = \frac{|1-z_0|}{|z_j - z_0|}$. Mit $|1 - z_0| = 2\cdot\left|\sin\left(\pi\frac{\delta}{n}\right)\right|$ und $|z_j - z_0| = 2\cdot\left|\sin\left(\pi\frac{j-\delta}{n}\right)\right|$ folgt

$$\rho = \left|\frac{\sin\left(\pi\frac{\delta}{n}\right)}{\sin\left(\pi\frac{j-\delta}{n}\right)}\right|,$$

und damit die Behauptung. □

Schließlich können wir die Frobeniusnorm der inversen Vandermonde-Matrix $V(\delta)^{-1}$ explizit angeben und beweisen.

Satz 6.12. *Es gilt*

$$\left\|V(\delta)^{-1}\right\|_F^2 = n\cdot\prod_{k=1}^{n-1}\frac{1}{4\cdot\sin^2\left(\pi\frac{k-\delta}{n}\right)} + \frac{\sin^2\left(\frac{\pi\delta}{n}\right)}{n}\cdot\sum_{k=1}^{n-1}\frac{1}{\sin^2\left(\pi\frac{k-\delta}{n}\right)} + \frac{n-1}{n}. \tag{25}$$

Beweis. Die Behauptung folgt sofort mit Hilfe der Lemmata 6.8 und 6.11:

$$\begin{aligned}\left\|V(\delta)^{-1}\right\|_F^2 &= \sum_{j=0}^{n-1}\sum_{r=0}^{n-1}|u_{jr}|^2 = \sum_{r=0}^{n-1}|u_{0r}|^2 + \sum_{j=1}^{n-1}\sum_{r=0}^{n-1}|u_{jr}|^2 \\ &= n\cdot\prod_{k=1}^{n-1}\frac{1}{4\cdot\sin^2\left(\pi\frac{k-\delta}{n}\right)} + \sum_{j=1}^{n-1}\left(\frac{1}{n}\cdot\frac{\sin^2\left(\pi\frac{\delta}{n}\right)}{\sin^2\left(\pi\frac{j-\delta}{n}\right)} + \frac{1}{n}\right) \\ &= n\cdot\prod_{k=1}^{n-1}\frac{1}{4\cdot\sin^2\left(\pi\frac{k-\delta}{n}\right)} + \frac{\sin^2\left(\pi\frac{\delta}{n}\right)}{n}\cdot\sum_{j=1}^{n-1}\frac{1}{\sin^2\left(\pi\frac{j-\delta}{n}\right)} + \frac{n-1}{n}.\end{aligned}$$

□

Mit Lemma 6.1 folgt nun sofort

Korollar 6.13. *Für die Kondition bezüglich der Frobeniusnorm von $V(\delta)$ gilt*

$$cond_F(V(\delta))^2 = n^3\cdot\prod_{k=1}^{n-1}\frac{1}{4\cdot\sin^2\left(\pi\frac{k-\delta}{n}\right)} + n\cdot\sin^2\left(\frac{\pi\delta}{n}\right)\cdot\sum_{k=1}^{n-1}\frac{1}{\sin^2\left(\pi\frac{k-\delta}{n}\right)} + n\cdot(n-1).$$

7 Fazit

In der vorliegenden Arbeit wurden Vandermonde-Matrizen mit reellen und komplexen Knoten untersucht. Die Betrachtung von Konfigurationen mit Knoten auf der reellen Achse bestätigte den Ruf von Vandermonde-Matrizen, schlecht konditioniert zu sein. So konnten nur Anordnungen von n reellen Knoten gefunden werden, bei denen die Kondition bezüglich der Zeilensummennorm mindestens exponentiell in n wächst.

Im Gegensatz dazu konnte gezeigt werden, dass Vandermonde-Matrizen zur äquidistanten Verteilung der Knoten auf dem komplexen Einheitskreis nahezu perfekt bezüglich der Frobenius- und der Zeilensummennorm konditioniert sind. Die Konditionszahl wächst in diesem Fall nur linear in n.

Abschließend wurden Verteilungen untersucht, bei denen ein Knoten von der Position als n-te Einheitswurzel abweicht. Für diese Konfiguration lieferten theoretische Betrachtungen die Vermutung über eine explizite Formel der Konditionszahl in Bezug auf die Zeilensummennorm. Numerische Untersuchungen konnten diese Vermutung stützen. Schließlich wurde für die Kondition bezüglich der Frobeniusnorm eine explizite Formel aufgestellt und bewiesen werden.

Diese Ergebnisse können nun hilfreich in die Entwicklung effizienter Algorithmen bei der Prony-Methode eingebracht werden.

Literaturverzeichnis

[1] Gautschi, W. [1962]. On inverses of Vandermonde and confluent Vandermonde matrices, *Numer. Math.* **4**: 117–123.

[2] Gautschi, W. [1975a]. Norm Estimates for Inverses of Vandermonde Matrices, *Numer. Math.* **23**: 337–347.

[3] Gautschi, W. [1975b]. Norm Estimates for Inverses of Vandermonde Matrices, *Numer. Math.* **23**: 337–347.

[4] Gautschi, W. [1978]. On inverses of Vandermonde and Confluent Vandermonde Matrices III, *Numer. Math.* **29**: 445–450.

[5] Gautschi, W. [1990]. How (Un)stable Are Vandermonde Systems?, *Lecture Notes in Pure and Appl. Math.* **124**: 193–210.

[6] Lang, S. [2003]. *Complex Analysis*, 4. Aufl., Springer-Verlag.

[7] Marchi, S. D. [1999]. *Generalized Vandermonde Determinants, Toeplitz Matrices and the Polynomial Division Algorithm*, Ergebnisberichte angewandte Mathematik, Univ. Dortmund.

[8] Schaback, R. und Wendland, H. [2005]. *Numerische Mathematik*, Springer-Verlag, Berlin.

[9] Stoer, J. und Bulirsch, R. [1994]. *Numerische Mathematik I*, Springer-Verlag, Berlin.

Eidesstattliche Erklärung

Hiermit erkläre ich, Thomas Wienecke, an Eides statt, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen sind, wurden unter Angabe der Quelle kenntlich gemacht.

Ort, Datum	Unterschrift