Rozdział 3

Normy wektorów i macierzy

W tym rozdziale zakładamy, że

$$\mathbf{K} \subseteq \mathbf{C}.$$

3.1 Ogólna definicja normy

Niech $\psi : \mathbf{K}^{m,n} \to [0, +\infty)$ będzie przekształceniem spełniającym warunki:

(i) $\forall A \in \mathbf{K}^{m,n} \quad \psi(A) = 0 \iff A = 0$,

(ii) $\forall A \in \mathbf{K}^{m,n}\, \forall u \in \mathbf{K} \quad \psi(u * A) = |u| \cdot \psi(A)$,

(iii) $\forall A, B \in \mathbf{K}^{m,n} \quad \psi(A + B) \leq \psi(A) + \psi(B)$
(*nierówność trójkąta* albo *subaddytywność*).

Każde takie przekształcenie ψ nazywamy *normą* w $\mathbf{K}^{m,n}$ i oznaczamy

$$\psi(A) = \|A\|.$$

Norma jest miarą "wielkości" macierzy. Dlatego

$$\|A - B\|$$

uznajemy za miarę odległości między macierzami A i B.

Powiemy, że norma jest *monotoniczna* gdy warunek $|A| \leq |B|$ (tzn. gdy $|a_{i,j}| \leq |b_{i,j}|\ \forall i, j$) implikuje $\|A\| \leq \|B\|$. Jeśli norma w $\mathbf{K}^{n,n}$ spełnia

$$\|A * B\| \leq \|A\| \cdot \|B\|, \qquad \forall A, B \in \mathbf{K}^{n,n},$$

to mówimy, że norma jest *submultiplikatywna*.

3.2 Normy wektorów

3.2.1 Normy p-te

Wektory w $\mathbf{K}^n$ są szczególnymi macierzami. W tym przypadku, ważnymi przykładami norm są *normy Schura*, zdefiniowane dla danej p, $1 \le p \le \infty$, jako

$$\begin{aligned} \|\vec{x}\|_p &= \left(\sum_{i=1}^n |x_i|^p\right)^{1/p} \qquad \text{dla} \quad 1 \le p < \infty, \\ \|\vec{x}\|_\infty &= \max_{1\le i\le n} |x_i|. \end{aligned}$$

Nietrudno zauważyć, że $\|\vec{x}\|_\infty = \lim_{p\to\infty} \|\vec{x}\|_p$, $\forall \vec{x} \in \mathbf{K}^n$.

Warunki (i) i (ii) normy są trywialnie spełnione przez normy Schura. Warunek (iii) łatwo sprawdzić dla $p = 1, \infty$. Dla $p = 1$ mamy bowiem

$$\|\vec{x}+\vec{y}\|_1 = \sum_{i=1}^n |x_i + y_i| \le \sum_{i=1}^n |x_i| + \sum_{i=1}^n |y_i| = \|\vec{x}\|_1 + \|\vec{y}\|_1,$$

a dla $p = \infty$

$$\|\vec{x}+\vec{y}\|_\infty = \max_{1\le i\le n} |x_i + y_i| \le \max_{1\le i\le n} |x_i| + \max_{1\le i\le n} |y_i| = \|\vec{x}\|_\infty + \|\vec{y}\|_\infty.$$

(W obu przypadkach zastosowaliśmy nierówność trójkąta $|u + v| \le |u| + |v|$ dla liczb zespolonych u i v.) Dla innych wartości p warunek (iii) jest dużo trudniej pokazać. Dlatego ograniczymy się tu jedynie do przypadku $p = 2$.

Lemat 3.1 (Nierówność Schwarza)
Dla dowolnych $\vec{u}, \vec{v} \in \mathbf{K}^n$ mamy

$$|\vec{u}^H * \vec{v}| \le \|\vec{u}\|_2 \cdot \|\vec{v}\|_2.$$

Dowód. Dla $t \in \mathbf{K}$ mamy

$$\begin{aligned} 0 \le \|\vec{u} + \vec{v} * t\|_2^2 &= (\vec{u} + \vec{v} * t)^H \cdot (\vec{u} + \vec{v} * t) \\ &= \vec{u}^H * \vec{u} + \bar{t} \cdot t * \vec{v}^H * \vec{v} + \vec{u}^H * \vec{v} * t + \vec{v}^H * \vec{u} * \bar{t} \\ &= \|\vec{u}\|_2^2 + |t|^2 \cdot \|\vec{v}\|_2^2 + |t| \cdot |\vec{u}^H * \vec{v}| \cdot \left(\omega^{(\varphi+\psi)} + \omega^{-(\varphi+\psi)}\right), \end{aligned}$$

gdzie $t = |t| \cdot \omega^\psi$, $\vec{u}^H * \vec{v} = |\vec{u}^H * \vec{v}| \cdot \omega^\varphi$, $\omega = \cos 1 + \imath \cdot \sin 1$.

Biorąc teraz $\psi = -\varphi$ mamy

$$0 \leq \|\vec{u}\|_2^2 + 2|t| \cdot |\vec{u}^H * \vec{v}| + |t|^2 \cdot \|\vec{v}\|_2^2,$$

a biorąc $\psi = \pi - \varphi$ mamy

$$0 \leq \|\vec{u}\|_2^2 - 2|t| \cdot |\vec{u}^H * \vec{v}| + |t|^2 \cdot \|\vec{v}\|_2^2.$$

Stąd dla dowolnej $\tau \in \mathbf{R}$ otrzymujemy

$$0 \leq \|\vec{u}\|_2^2 + 2\tau|\vec{u}^H * \vec{v}| + \tau^2\|\vec{v}\|_2^2.$$

Ponieważ prawa strona ostatniej nierówności jest, jako funkcja τ, trójmianem kwadratowym o wartościach nieujemnych, to

$$0 \geq \Delta = 4\left(|\vec{u} * \vec{v}|^2 - \|\vec{u}\|_2^2 \cdot \|\vec{v}\|_2^2\right),$$

co implikuje $|\vec{u}^H * \vec{v}| \leq \|\vec{u}\|_2 \cdot \|\vec{v}\|_2$ i kończy dowód.

Na podstawie nierówności Schwarza mamy teraz

$$\begin{array}{rcl} \|\vec{u}+\vec{v}\|_2^2 & = & \|\vec{u}\|_2^2 + \|\vec{v}\|_2^2 + \vec{u}^H * \vec{v} + \vec{v}^H * \vec{u} \\ & = & \|\vec{u}\|_2^2 + \|\vec{v}\|_2^2 + 2\Re(\vec{u}^H * \vec{v}) \\ & \leq & \|\vec{u}\|_2^2 + \|\vec{v}\|_2^2 + 2|\vec{u}^H * \vec{v}| \\ & \leq & \|\vec{u}\|_2^2 + \|\vec{v}\|_2^2 + 2\|\vec{u}\|_2\|\vec{v}\|_2 \\ & = & (\|\vec{u}\|_2 + \|\vec{v}\|_2)^2, \end{array}$$

czyli nierówność trójkąta dla $\|\cdot\|_2$.

3.2.2 Pożyteczne (nie)równości

Nietrudno pokazać następujące nierówności łączące normy p-te Schura dla $p = 1, 2, \infty$. Mianowicie, dla każdego $\vec{u} \in \mathbf{K}^n$ mamy

$$\begin{array}{rcccl} \|\vec{u}\|_\infty & \leq & \|\vec{u}\|_1 & \leq & n \cdot \|\vec{u}\|_\infty, \\ \|\vec{u}\|_\infty & \leq & \|\vec{u}\|_2 & \leq & \sqrt{n} \cdot \|\vec{u}\|_\infty, \\ \|\vec{u}\|_2 & \leq & \|\vec{u}\|_1 & \leq & \sqrt{n} \cdot \|\vec{u}\|_2, \end{array}$$

przy czym ostatnia z tych nierówności jest konsekwencją nierówności Schwarza,

$$\|\vec{u}\|_1 = \sum_{i=1}^{n} |u_i| = \sum_{i=1}^{n} |u_i| \cdot |1| \leq \left(\sum_{i=1}^{n} |u_i|^2 \right)^{1/2} \left(\sum_{i=1}^{n} 1^2 \right)^{1/2} = \sqrt{n} \cdot \|\vec{u}\|_2.$$

Dodatkowo zauważamy, że nierówności tych nie można poprawić. Na przykład, dla pierwszego wersora $\vec{e}_1$ mamy $\|\vec{e}_1\|_p = 1 \ \forall p$, a dla $\vec{1} = [1, 1, \dots, 1] \in \mathbf{K}^n$ mamy $\|\vec{1}\|_1 = \sqrt{n}\|\vec{1}\|_2 = n\|\vec{1}\|_\infty$.

Kulą jednostkową w $\mathbf{K}^n$ (ze względu na normę $\|\cdot\|$) nazywamy zbiór wektorów

$$K = \{\, \vec{u} \in \mathbf{K}^n : \ \|\vec{u}\| \leq 1 \,\}.$$

Z podanych powyżej nierówności wynika w szczególności, że

$$K_1 \subset K_2 \subset K_\infty,$$

gdzie K_p jest kulą jednostkową w normie p-tej Schura.

Zauważmy jeszcze, że normy p-te są monotoniczne oraz, że dla dowolnej macierzy permutacji $P \in \mathbf{K}^{n,n}$ i wektora $\vec{x} \in \mathbf{K}^n$

$$\|P * \vec{x}\|_p = \|\vec{x}\|_p,$$

tzn. norma p-ta wektora jest *niezmiennicza* ze względu na przestawienia kolejności jego współrzędnych.

3.3 Normy macierzy

3.3.1 Normy p-te

Normy p-te macierzy są definiowane (indukowane) przez normy p-te wektorów w następujący sposób:

$$\begin{aligned} \|A\|_p &= \sup_{\vec{0} \neq \vec{x} \in \mathbf{K}^n} \frac{\|A * \vec{x}\|_p}{\|\vec{x}\|_p} \\ &= \sup \{\, \|A * \vec{x}\|_p \ : \ \vec{x} \in \mathbf{K}^n, \|\vec{x}\|_p = 1 \,\}. \end{aligned}$$

Zauważmy, że używamy tego samego oznaczenia dla norm wektora jak i macierzy. Jest to uzasadnione, gdyż norma p-ta macierzy jest uogólnieniem normy p-tej wektora. Dla $A = [u_1, \ldots, u_m]^T \in \mathbf{K}^{m,1} = \mathbf{K}^m$ mamy bowiem $\|A\|_p = \sup_{|t|=1} \|A * t\|_p = (\sum_{i=1}^m |u_i|^p)^{1/p}$. (Tutaj $t \in \mathbf{K}$!)

Wprost z definicji wynika, że normy indukowane macierzy spełniają warunek *zgodności* (z normą wektorową), tzn.

$$\forall A \in \mathbf{K}^{m,n}\, \forall \vec{x} \in \mathbf{K}^n \quad \|A * \vec{x}\|_p \leq \|A\|_p \cdot \|\vec{x}\|_p.$$

Normy te są również submultiplikatywne,

$$\forall A \in \mathbf{K}^{m,l}\, \forall B \in \mathbf{K}^{l,n} \qquad \|A * B\|_p \leq \|A\|_p \cdot \|B\|_p.$$

Rzeczywiście, dla $\vec{x} \in \mathbf{K}^l$ mamy

$$\begin{aligned} \|(A * B) * \vec{x}\|_p &= \|A * (B * \vec{x})\|_p \leq \|A\|_p \cdot \|B * \vec{x}\|_p \\ &\leq \|A\|_p \cdot \|B\|_p \cdot \|\vec{x}\|_p, \end{aligned}$$

skąd

$$\sup_{\vec{x} \neq \vec{0}} \frac{\|(A * B) * \vec{x}\|_p}{\|\vec{x}\|_p} \leq \|A\|_p \cdot \|B\|_p.$$

Dla macierzy permutacji $P \in \mathbf{K}^{m,m}$ i $Q \in \mathbf{K}^{n,n}$ mamy

$$\|P * A * Q^T\|_p = \|A\|_p,$$

co oznacza, że przestawienie kolumn i wierszy macierzy nie zmienia jej p-tej normy. Rzeczywiście, ponieważ przestawienie współrzędnych nie zmienia normy p-tej wektora, mamy

$$\sup_{\vec{x} \neq \vec{0}} \frac{\|P * A * Q^T * \vec{x}\|_p}{\|\vec{x}\|_p} = \sup_{\vec{x} \neq \vec{0}} \frac{\|A * Q^T * \vec{x}\|_p}{\|Q^T * \vec{x}\|_p} = \sup_{\vec{y} \neq \vec{0}} \frac{\|A * \vec{y}\|_p}{\|\vec{y}\|_p}.$$

3.3.2 Pożyteczne (nie)równości

Dla niektórych p, normę można wyrazić w sposób pozwalający ją łatwo obliczyć.

Lemat 3.2 *Dla dowolnej macierzy* $A = (a_{i,j}) \in \mathbf{K}^{m,n}$

(a) $\|A\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^n |a_{i,j}|$,

(b) $\|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^m |a_{i,j}|$.

Dowód. (a) Dla $\vec{x} = [x_1, \ldots, x_n]^T \in \mathbf{K}^n$ mamy

$$\begin{aligned} \|A * \vec{x}\|_\infty &= \max_{1 \le i \le m} \left| \sum_{j=1}^n a_{i,j} \cdot x_j \right| \le \max_{1 \le i \le m} \sum_{j=1}^n |a_{i,j}| \cdot |x_j| \\ &\le \|\vec{x}\|_\infty \cdot \left(\max_{1 \le i \le m} \sum_{j=1}^n |a_{i,j}| \right). \end{aligned}$$

Z drugiej strony, weźmy $\vec{x}^* = (x_j^*)$ taki, że $x_j^* = \omega^{-\varphi_j}$, $1 \le j \le n$, gdzie φ_j jest argumentem liczby $a_{s,j}$, tzn. $a_{s,j} = |a_{s,j}|\omega^{\varphi_j}$, a s jest tym indeksem i, dla którego suma $\sum_{j=1}^n |a_{i,j}|$ jest największa. Wtedy $\|\vec{x}^*\|_\infty = 1$ oraz

$$\|A * \vec{x}^*\|_\infty \ge \left| \sum_{j=1}^n a_{s,j} \cdot x_j^* \right| = \left| \sum_{j=1}^n |a_{s,j}| \omega^{\varphi_j} \omega^{-\varphi_j} \right| = \sum_{j=1}^n |a_{s,j}|,$$

a stąd $\|A\|_\infty \ge \max_{1 \le i \le m} \sum_{j=1}^n |a_{i,j}|$.
(b) Dla dowolnego $\vec{x}$ mamy

$$\begin{aligned} \|A * \vec{x}\|_1 &= \sum_{i=1}^m \left| \sum_{j=1}^n a_{i,j} \cdot x_j \right| \le \sum_{i=1}^m \sum_{j=1}^n |a_{i,j}| \cdot |x_j| \\ &= \sum_{j=1}^n |x_j| \cdot \sum_{i=1}^m |a_{i,j}| \le \left(\max_{1 \le j \le n} \sum_{i=1}^m |a_{i,j}| \right) \cdot \|\vec{x}\|_1. \end{aligned}$$

Z drugiej strony, dla $\vec{x}^*$ takiego, że $x_j^* = 0$ dla $j \ne s$, $x_j^* = 1$ dla $j = s$, gdzie s jest tym indeksem j dla którego suma $\sum_{i=1}^m |a_{i,j}|$ jest największa, mamy $\|\vec{x}^*\|_1 = 1$ oraz $\|A * \vec{x}\|_1 = \sum_{i=1}^m |a_{i,s}|$, a stąd $\|A\|_1 \ge \max_{1 \le j \le n} \sum_{i=1}^m |a_{i,j}|$.

Z powyższego lematu łatwo widać, że

$$\begin{aligned} \|A^T\|_\infty &= \|A^H\|_\infty = \|A\|_1, \\ \|A^T\|_1 &= \|A^H\|_1 = \|A\|_\infty. \end{aligned}$$

Szczególną rolę odgrywa norma druga $\| \cdot \|_2$, ze względów, które będą jasne później. Niestety, nie wyraża się ona w tak prosty sposób jak $\| \cdot \|_1$ i $\| \cdot \|_\infty$. W odróżnieniu od tych ostatnich, norma druga ma jednak dodatkową ważną własność; mianowicie, dla dowolnej $A \in \mathbf{K}^{m,n}$

$$\|A^T\|_2 = \|A^H\|_2 = \|A\|_2.$$

Równość ta wynika bezpośrednio z faktu, że

$$\|A\|_2 = \sup_{\vec{z}} \sup_{\vec{y}} \left|\vec{y}^H * A * \vec{z}\right|,$$

gdzie suprema wzięte są po $\vec{z} \in \mathbf{K}^n$ i $\vec{y} \in \mathbf{K}^m$ takich, że $\|\vec{z}\|_2 = 1 = \|\vec{y}\|_2$.

Rzeczywiście, dla dowolnych $\vec{y}$ i $\vec{z}$ o jednostkowych normach mamy

$$|\vec{y}^H * A * \vec{z}| \leq \|\vec{y}\|_2 \cdot \|A * \vec{z}\|_2 = \|A * \vec{z}\|_2 \leq \|A\|_2,$$

przy czym w pierwszej nierówności zastosowaliśmy nierówność Schwarza. Z drugiej strony, dla $\vec{z}$ o jednostkowej normie i takiego, że $A * \vec{z} \neq \vec{0}$ mamy

$$\|A * \vec{z}\|_2 \;=\; \frac{\|A * \vec{z}\|_2^2}{\|A * \vec{z}\|_2} \;=\; \frac{(A * \vec{z})^H * A * \vec{z}}{\|A * \vec{z}\|_2} \;\leq\; \sup_{\|\vec{y}\|_2 = 1} \left|\vec{y}^H * A * \vec{z}\right|,$$

gdzie podstawiliśmy $\vec{y} = A * \vec{z} / \|A * \vec{z}\|_2$.

3.3.3 Norma Frobeniusa

Normę *Frobeniusa* (albo *Euklidesową*) macierzy $A \in \mathbf{K}^{m,n}$ definiujemy jako

$$\|A\|_F \;=\; \left(\sum_{i=1}^m \sum_{j=1}^n |a_{i,j}|^2 \right)^{1/2}.$$

Zaletą normy $\|\cdot\|_F$ jest jej łatwa "obliczalność", natomiast wadą, że nie jest to norma indukowana przez żadną normę wektorową.

Związek między normą Frobeniusa i normą drugą pokazuje następujący lemat.

Lemat 3.3 *Dla dowolnej* $A \in \mathbf{K}^{m,n}$ *mamy*

$$\|A\|_2 \;\leq\; \|A\|_F \;\leq\; \sqrt{\min(m,n)} \cdot \|A\|_2.$$

Dowód. Wykorzystując nierówność Schwarza, dla dowolnego $\vec{x} \in \mathbf{K}^n$ o jednostkowej normie drugiej mamy

$$\begin{aligned} \|A * \vec{x}\|_2^2 \;&=\; \sum_{i=1}^m \left| \sum_{j=1}^n a_{i,j} \cdot x_j \right|^2 \;\leq\; \sum_{i=1}^m \left(\sum_{j=1}^n |a_{i,j}| \cdot |x_j| \right)^2 \\ &\leq\; \sum_{i=1}^m \left(\sum_{j=1}^n |a_{i,j}|^2 \right) \left(\sum_{j=1}^n |x_j|^2 \right) \;=\; \|A\|_F^2, \end{aligned}$$

a stąd $\|A\|_2 \le \|A\|_F$.

Z drugiej strony, przedstawiając A jako

$$A = [\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n]\,, \qquad \vec{a}_j \in \mathbf{K}^m,$$

mamy $\|A\|_2 \ge \|A * \vec{e}_j\|_2 = \|\vec{a}_j\|_2$, gdzie $\vec{e}_j$ jest j-tym wersorem. Stąd

$$\|A\|_2^2 \;\ge\; \frac{1}{n} \cdot \sum_{j=1}^{n} \|a_j\|_2^2 \;=\; \frac{1}{n} \cdot \|A\|_F^2,$$

czyli $\|A\|_F \le \sqrt{n} \cdot \|A\|_2$. Ale również

$$\|A\|_F = \|A^T\|_F \le \sqrt{m} \cdot \|A^T\|_2 = \sqrt{m} \cdot \|A\|_2,$$

co kończy dowód.

Zauważymy jeszcze jedną własność norm p-tych macierzy. Niech macierz A będzie dana w postaci blokowej,

$$A = [A_1, A_2, \ldots, A_s]\,.$$

Wtedy

$$\begin{aligned} \|A_k\|_p \;&=\; \sup_{\|\vec{x}_k\|_p = 1} \|A_k * \vec{x}_k\|_p \;=\; \sup_{\|\vec{x}_k\|_p = 1, \vec{x}_j = \vec{0}, j \ne k} \left\| \sum_{j=1}^{s} A_j * \vec{x}_j \right\|_p \\ &\le\; \sup_{\|\vec{x}\|_p = 1} \|A * \vec{x}\|_p \;=\; \|A\|_p. \end{aligned}$$

Podobnie, jeśli

$$A = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_t \end{bmatrix}$$

to

$$\begin{aligned} \|A_k\|_p^p \;&=\; \sup_{\|\vec{x}\|_p = 1} \|A_k * \vec{x}\|_p^p \;\le\; \sup_{\|\vec{x}\|_p = 1} \sum_{j=1}^{t} \|A_j * \vec{x}\|_p^p \\ &=\; \sup_{\|\vec{x}\|_p = 1} \|A * \vec{x}\|_p^p \;=\; \|A\|_p^p. \end{aligned}$$

Stąd dostajemy wniosek, że jeśli A jest w postaci blokowej to dla każdego bloku $A_{i,j}$ mamy

$$\|A_{i,j}\|_p \leq \|A\|_p, \qquad 1 \leq p \leq \infty.$$

Oczywiście, ta własność zachodzi również dla normy Frobeniusa.