

2. előadás

Lineáris algebra numerikus módszerei

Definíció

Az $m \times n$ típusú (méretű) valós A mátrixon valós a_{ij} számok alábbi táblázatát értjük:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1j} & \dots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{i1} & a_{i2} & \dots & a_{ij} & \dots & a_{in} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mj} & \dots & a_{mn} \end{bmatrix}.$$

Az m és az n értelemszerűen pozitív egész számok.

Beszélünk a mátrix i -edik soráról és j -edik oszlopáról.

A sorok és oszlopok metszéspontjában vannak a mátrix a_{ij} elemei
($i = 1, \dots, m, j = 1, \dots, n$).

Komplex számokból (sőt, más – absztrakt – struktúrák elemeiből)
is felépíthetünk mátrixokat.

Az $m \times n$ típusú valós mátrixok halmazát $\mathbb{R}^{m \times n}$ jelöli, ennek
megfelelően például az $A \in \mathbb{R}^{k \times l}$ azt jelenti, hogy A egy $k \times l$
típusú (méretű) valós mátrix.

Definíció

Az egyetlen sorból vagy egyetlen oszlopból álló mátrixot vektornak nevezzük.

A sorvektor szokásos megadási módja: $x = [x_1, \dots, x_n]$. Az x_i a vektor egy eleme, másképpen komponense. Az oszlopvektorokat az

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$$

formában szoktuk megadni, ahol \mathbb{R}^n az n komponensű oszlopvektorok halmaza tulajdonképpen $\mathbb{R}^n \equiv \mathbb{R}^{n \times 1}$).

Megjegyzés

Ha egyszerűen csak vektort említünk és nem tesszük hozzá, hogy az sor- vagy oszlopvektor-e, akkor azon mindig oszlopvektort értünk. A komponensek kettős indexelése egyrészt felesleges (az egyik mindig 1), másrészt ez is mutatja, hogy valójában a mátrixoktól függetlenül is értelmezhető matematikai objektum.

A definíciókban látható szögletes zárójelek helyett szokás kerek zárójeleket is használni, valamint gyakran célszerű a következő, tömörebb jelölést választani:

$$A = [a_{ij}]_{i,j=1}^{m,n} = (a_{ij})_{i,j=1}^{m,n} \quad q = [q_i]_{i=1}^n = (q_i)_{i=1}^n.$$

Ha a sorok és oszlopok száma megegyezik, akkor a mátrixot ***négyzetesnek*** nevezzük, tömörebb jelölése is némileg egyszerűsödik. Például az $n \times n$ méretű C mátrix esetén:

$$C = [c_{ij}]_{i,j=1}^n, \quad C = (c_{ij})_{i,j=1}^n.$$

I. Összeadás.

Csak azonos méretű mátrixok között értelmezzük, a következőképpen: ha $A, B \in \mathbb{R}^{m \times n}$, akkor

$$C = A + B \in \mathbb{R}^{m \times n} \Leftrightarrow C = (c_{ij})_{i,j=1}^{m,n} = (a_{ij} + b_{ij})_{i,j=1}^{m,n}.$$

Az összeadás két fontos tulajdonsága: kommutatív és asszociatív (a tagok sorrendje felcserélhető és csoportosítható)

$$A + B = B + A \qquad (A + B) + C = A + (B + C).$$

II. Számmal való szorzás.

Legyen $A \in \mathbb{R}^{m \times n}$ és legyen λ valós szám (azaz $\lambda \in \mathbb{R}$). Ekkor

$$C = \lambda A \in \mathbb{R}^{m \times n} \Leftrightarrow C = (c_{ij})_{i,j=1}^{m,n} = (\lambda a_{ij})_{i,j=1}^{m,n}.$$

Nyilvánvaló, hogy $\lambda(\mu A) = (\lambda\mu)A$, továbbá a fenti két művelet értelmezéséből következik az alábbi két – a disztributivitást kimondó – szabály:

$$\lambda(A + B) = \lambda A + \lambda B \quad \text{és} \quad (\lambda + \mu)A = \lambda A + \mu A.$$

Megállapodás szerint számmal jobbról és balról is szorozhatunk:
 $\lambda A = A\lambda$.

Az $\mathbb{R}^{m \times n}$ halmazt (vagy az \mathbb{R}^n halmazt) a fenti két művelettel algebrai struktúrának tekinthetjük és mint ilyen, rendelkezik mindazon tulajdonságokkal, amelyek a lineáris teret (más szóval – itt ugyan szokatlanul hangzik – vektorteret) definiálják; például van zéruselem, az összeadás invertálható, stb. Ezért nem lesz meglepetés, ha az általunk érzékelt háromdimenziós térben definiált vektorok bizonyos jellemzőit (pl. hosszúság) általánosítjuk többdimenzióban is, sőt ezt az általánosítást kiterjesztjük mátrixokra is.

III. Transzponálás (tükrözés).

Az $A \in \mathbb{R}^{m \times n}$ transzponáltját jelölje A^T , amit a következőképpen definiálunk:

$$C = A^T \in \mathbb{R}^{n \times m} \Leftrightarrow C = (c_{ij})_{i,j=1}^{n,m}, \text{ ahol } c_{ij} = a_{ji}$$

Úgy is fogalmazhatjuk: a sorokat és az oszlopokat felcseréljük.

A transzponálás definíciójából adódik, hogy

$$(A^T)^T = A, \quad (A + B)^T = A^T + B^T.$$

Négyzetes mátrixok esetén a transzponálás a főátlóra való tükrözést jelent. (A főátlót azon elemek alkotják, melyek sor- és oszlopindexe megegyezik, azaz az $a_{11}, a_{22}, \dots, a_{nn}$ elemek.)

A transzponálás felhasználásával az oszlopvektorokat meg lehet adni még az $x = [x_1, \dots, x_n]^T$, a sorvektorokat pedig az

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}^T$$

formában is.

IV. Szorzás.

Ha $A \in \mathbb{R}^{m \times k}$, és $B \in \mathbb{R}^{k \times n}$, akkor a szorzatukat a

$$C = (c_{ij})_{i,j=1}^{m,n} = AB \in \mathbb{R}^{m \times n} \Leftrightarrow c_{ij} = \sum_{t=1}^k a_{it} b_{tj}$$

$(i = 1, \dots, m, j = 1, \dots, n).$

előírással definiáljuk.

Látható, hogy amint az összeadásnál, úgy itt is lényeges a műveletben szereplő operandusok (tényezők) mérete, csak itt más a követelmény: az első tényező oszlopainak száma és a második tényező sorainak a száma kell, hogy egyenlő legyen; az eredmény mérete a tényezőkéből következik. Fontos megjegyezni, hogy a szorzás nem kommutatív, tehát általában

A két mátrix között többnyire nem is értelmezhető mindkét sorrendben a szorzás.

Ha $x, y \in \mathbb{R}^n$ akkor az $x^T y$ és az xy^T szorzás végezhető el; előbbi neve **skalár szorzás**, utóbbié **diadikus szorzás**. A skalár szorzat eredménye egy skalár szám (1×1 -es mátrix), szokták külön is definiálni.

Definíció

Az $x, y \in \mathbb{R}^n$ skaláris szorzata a

$$x^T y = \sum_{i=1}^n x_i y_i.$$

A skalár szorzás definíciójából látható, hogy a szorzatmátrix bármelyik (i, j) indexű elemét úgy kapjuk, mint egy skalárszorzatot: az első tényező i -edik sorát (mint sorvektort) szorozzuk a második tényező j -edik oszlopával, azaz

$$c_{ij} = [a_{i1}, \dots, a_{ik}] \begin{bmatrix} b_{1j} \\ \vdots \\ b_{kj} \end{bmatrix}.$$

Gyakran kell alkalmaznunk a mátrixszorzásnak – az értelmezésből közvetlenül adódó – alábbi tulajdonságait:

$$\begin{aligned}(AB)C &= A(BC), \\ A(B + C) &= AB + AC, \\ (A + B)C &= AC + BC, \\ (AB)^T &= B^T A^T.\end{aligned}$$

A továbbiakban a mátrix és mátrix-vektor műveletek felírásánál feltesszük, hogy az ott szereplő mátrixok, ill. vektorok méretei olyanok, amelyek lehetővé teszik az adott műveletet.

Definíció

Az A mátrix szimmetrikus, ha $A^T = A$.

Nyilvánvaló, hogy csak négyzetes mátrix lehet szimmetrikus és ekkor $a_{ij} = a_{ji}$ ($i, j = 1, \dots, n$)

Definíció

A $0 \in \mathbb{R}^{m \times n}$ mátrix zérusmátrix, ha az összes eleme zérus, azaz

$$0 = \begin{bmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{bmatrix}.$$

A zérusmátrix az összeadásra nézve a zéruselem, azaz minden A mátrix esetén

$$A + 0 = A, \quad A0 = 0.$$

Definíció

Az $I \in \mathbb{R}^{n \times n}$ mátrix egységmátrix, ha

$$I = \begin{bmatrix} 1 & 0 & \dots & \dots & 0 \\ 0 & 1 & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & \dots & \dots & 0 & 1 \end{bmatrix}.$$

Az egységmátrix (amit gyakran a magyar elnevezésének kezdőbetűjével, E -vel is jelölünk) a szorzásra nézve egységelem, azaz minden $A \in \mathbb{R}^{n \times n}$ esetén

$$AI = IA = A.$$



Megjegyzés

Szokás az egységmátrix fogalmát a nem négyzetes mátrixokra is kiterjeszteni, a definícióját úgy adva meg, hogy az azonos indexpárú elemei 1-esek, a többi zérus (például

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

Ilyenkor persze az egység elnevezés már nem jogos, félrevezető. Mindenesetre nagyon hasznos kiterjesztés, több matematikai szoftver (a Matlab is) él vele.

Az egységmátrix mellett fontos fogalom az egységvektor.

Definíció

Az $e_i \in \mathbb{R}^n$ vektort (i -edik) egységvektornak nevezzük, ha az i -edik komponense 1-es, a többi pedig zérus.

A transzponáltját felírva, tehát:

$$e_i = [0, \dots, 0, 1, 0, \dots, 0]^T \in \mathbb{R}^n.$$

Definíció

A $D \in \mathbb{R}^{n \times n}$ diagonálmátrix, ha

$$D = \begin{bmatrix} d_1 & 0 & \dots & \dots & 0 \\ 0 & d_2 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & d_{n-1} & 0 \\ 0 & \dots & \dots & 0 & d_n \end{bmatrix}.$$

A diagonális elemek indexei azonosak, ezért jelöljük azokat rendszerint egyetlen indexszel. Magát a D diagonálmátrixot gyakran a $\text{diag}(d_1, \dots, d_n)$ vagy $\text{diag}(d_i)$ ($i = 1, \dots, n$) formában is jelöljük.

Definíció

A $P \in \mathbb{R}^{n \times n}$ mátrix permutációmátrix, ha minden sorában és oszlopában pontosan egy darab 1-es van és a többi elem zérus.

Például az alábbi mátrix egy 4×4 -es permutációmátrix:

$$\begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

Definíció

Az $A = [a_{ij}]_{i,j=1}^n$ mátrix alsó háromszög alakú, ha minden $i < j$ esetén $a_{ij} = 0$ és felső háromszög alakú, ha minden $i > j$ esetén $a_{ij} = 0$.

Az alsó- és felsőháromszögmátrixok alakja sematikususan a következő:

$$\begin{bmatrix} * & 0 & \dots & \dots & 0 \\ * & * & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ \vdots & & & * & 0 \\ * & \dots & \dots & * & * \end{bmatrix}, \begin{bmatrix} * & * & \dots & \dots & * \\ 0 & * & & & \vdots \\ \vdots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & * & * \\ 0 & \dots & \dots & 0 & * \end{bmatrix}.$$

Beszélni szoktunk valamely négyzetes $A \in \mathbb{R}^{n \times n}$ mátrix alsó és felső háromszögrészéről, amelyekre a $\text{tril}(A)$, illetve $\text{triu}(A)$ jelöléssel hivatkozunk. Ezeket értelemszerűen úgy kapjuk, hogy az eredeti mátrix főátlója feletti, illetve alatti elemeit kicseréljük zérusokra.

Például az $A = \begin{bmatrix} 3 & -2 & 2 \\ 9 & 0 & 0 \\ -5 & 3 & 1 \end{bmatrix}$ mátrix esetén

$$\text{tril}(A) = \begin{bmatrix} 3 & 0 & 0 \\ 9 & 0 & 0 \\ -5 & 3 & 1 \end{bmatrix} \quad \text{triu}(A) = \begin{bmatrix} 3 & -2 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

(az l az angol *lower*, u pedig az *upper* szó kezdőbetűje.)

Definíció

Az $A \in \mathbb{R}^{n \times n}$ mátrix sávmátrix p alsó sávszélességgel és q felső sávszélességgel, ha teljesül, hogy

$$a_{ij} = 0, \quad \text{ha } i > j + p, \text{ vagy } i + q < j.$$

Más szóval: a $-p$ -edik diagonálisuk alatti és q -edik diagonálisuk feletti elemeik zérusak. A sávot, amelynek elemeit nem kötelezően zérusokként kezeljük, azon a_{ij} elemek definiálják, amelyek indexeire teljesül, hogy $i - p \leq j \leq i + q$, vagy ekvivalens módon $j - q \leq i \leq j + p$. Sematikusan ábrázolva:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & \dots & a_{1,1+q} & 0 & \dots & \dots & 0 \\ a_{21} & a_{22} & & & & \ddots & & & \vdots \\ \vdots & & \ddots & & & & \ddots & & \vdots \\ a_{1+p,1} & & & \ddots & & & & \ddots & 0 \\ 0 & \ddots & & & \ddots & & & & a_{n-q,n} \\ \vdots & & \ddots & & & \ddots & & & \vdots \\ \vdots & & & \ddots & & & \ddots & & \vdots \\ \vdots & & & & \ddots & & & \ddots & a_{n-1,n} \\ 0 & \dots & \dots & \dots & 0 & a_{n,n-p} & \dots & a_{n,n-1} & a_{nn} \end{bmatrix} \cdot$$

Általában ritka mátrixoknak nevezzük azokat a mátrixokat, amelyek viszonylag sok, ismert pozíciójú zérust tartalmaznak. Ilyenek például a sávmátrixok, vagy lehetnek szabálytalan (de ismert, rögzített) elhelyezkedésű sok zérust tartalmazó mátrixok is. Az ilyen mátrixok tárolása a zérusok figyelmen kívül hagyásával helytakarékosan oldható meg, és a velük való különböző manipulációkat végrehajtó programok is gazdaságosan, művelettakarékosan írhatók meg. (Persze, tudni kell, hogy közben nem változnak-e meg a zérusok, illetve azt, hogy legfeljebb hol változhatnak.) Ebből a szempontból ritkának tekinthetők a háromszögmátrixok is, a nemzérus elemek tárolása történhet vektorban, például oszlopfolytonos sorrendben. A szimmetria is tekintetbe vehető; ilyen esetben elég csak az alsó (vagy a felső) háromszög részt tárolni. (Itt is ügyelni kell a programozás során, hogy hol romlik el esetleg az algoritmus végrehajtása során a szimmetria.)

Definíció

Az $f : \mathbb{R}^{n \times k} \rightarrow \mathbb{R}$ függvényt mátrixnormának (vagy vektornormának, ha $k = 1$) nevezzük, ha

$$f(x) \geq 0 \quad (\forall x \in \mathbb{R}^{n \times k}), \quad f(x) = 0 \Leftrightarrow x = 0,$$

$$f(\lambda x) = |\lambda| f(x) \quad (\forall x \in \mathbb{R}^{n \times k}, \forall \lambda \in \mathbb{R}),$$

$$f(x + y) \leq f(x) + f(y) \quad (\forall x, y \in \mathbb{R}^{n \times k}).$$

A norma szokásos jelölése: $\|x\|$.

Normát nyilván nagyon sokféleképpen adhatunk meg. Bár a vektorokat is (speciális méretű) mátrixokkal azonosítottuk, vannak szempontok, amelyek alapján mégis csak különböző objektumokról van szó, ezért konkrét normákat különböző módon vezetünk be. Vektorok esetén a normák fontos osztályát alkotják, az ún. hatványnormák:

$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}},$$

ahol $p \geq 1$ egész szám.

A leggyakrabban használt vektornormák a következők:

$$\|x\|_1 = \sum_{i=1}^n |x_i| \quad (\text{összeg norma}),$$

$$\|x\|_2 = \left(\sum_{i=1}^n |x_i|^2 \right)^{\frac{1}{2}} \quad (\text{euklideszi norma}),$$

$$\|x\|_\infty = \max_{1 \leq i \leq n} |x_i| \quad (\text{maximum norma}).$$

A legutóbbit a $p \rightarrow \infty$ határátmenettel kapjuk. (Az $\|x\|_2$ norma megfogalmazásánál természetesen a jobboldalon az abszolút érték jel elhagyható valós vektorok esetén, de akkor nem, ha az elemek komplexek.)

Példa

Legyen $v = [1, -7, 0, 3]^T$. Ekkor

$$\|v\|_1 = 1 + 7 + 0 + 3 = 11,$$

$$\|v\|_2 = \sqrt{1 + 49 + 0 + 9} = 7.681 \text{ (kerekítve),}$$

$$\|v\|_\infty = \max\{1, 7, 0, 3\} = 7.$$

A háromkomponensű vektorok által bezárt szög skalár szorzás segítségével való kiszámításának szabályát kiterjesztve, értelmezhetjük az akárhány komponensű vektorok közötti szöget is.

Definíció

Az $x, y \in \mathbb{R}^n$ ($x, y \neq 0$) vektorok szöge θ , amelynek koszinuszát a

$$\cos(\theta) = \frac{x^T y}{\|x\|_2 \|y\|_2}$$

összefüggés definiálja.

Legyen $A = [a_{ij}]_{i,j=1}^{m,n}$. A leggyakrabban használt mátrixnormák a következők:

$$\|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^m |a_{ij}| \quad (\text{oszlopösszeg norma}),$$

$$\|A\|_2 = \{A^T A \text{ legnagyobb sajátértéke}\}^{\frac{1}{2}} \quad (\text{spektrálnorma}),$$

$$\|A\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^n |a_{ij}| \quad (\text{sorösszeg norma}),$$

$$\|A\|_F = \left(\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2 \right)^{\frac{1}{2}} \quad (\text{Frobenius-norma}).$$

(A Frobenius-normánál itt sem kell csak komplex elemek esetén a jobboldalon az abszolút érték jel.)



Példa

Legyen

$$A = \begin{bmatrix} 2 & -1 & 0 \\ 4 & 1 & -1 \end{bmatrix}.$$

Az említett normái:

$$\|A\|_1 = \max\{2 + 4; 1 + 1; 0 + 1\} = 6,$$

$$\|A\|_\infty = \max\{2 + 1 + 0; 4 + 1 + 1\} = 6,$$

$$\|A\|_F = \sqrt{4 + 1 + 0 + 16 + 1 + 1} = 4.796.$$