

1. A vektor és a vektortér fogalma

Célunk: a vektor fogalmának minél tágabb értelmezése. Ez azért hasznos, mert így a síkvektorok körében használatos egyes fogalmak és tételek átvihetők lesznek egyéb objektumokból (pl. függvényekből) álló halmazokra is.

1.1. A síkvektorok halmaza

A vektortér fogalmának általánosításához induljunk ki az összes síkvektor halmazából (jel.: $\mathcal{E}_2$). Gyűjtsük össze ennek a halmaznak néhány tulajdonságát!

- i) Az elemek irányított szakaszok, amelyeket nyíllal szemléltethetünk.
- ii) Az elemeknek van nagyságuk és irányuk.
- iii) Bármely két elemnek értelmezve van az összege, és ez az összeg halmazbeli, azaz annak egy adott eleme.
- iv) Bármely elemnek értelmezve van egy valós számmal való szorzása (a skalárszorosa), és a skalárszoros is a halmaz eleme.
- v) Az elemek között értelmezve van skaláris szorzás. Vigyázat: ez nem ugyanaz, mint az előbbi pontban említett skalárral való szorzás! A skaláris szorzás eredménye nem a halmaz eleme, hanem egy valós szám.

Ezeknek a tulajdonságoknak számos fontos következményük van, pl. az összeadás és a skalárral való szorzás műveletével lineáris kombinációkat képezhetünk, értelmes a síkvektorok lineáris függetlensége stb.

1.2. A vektortér általános definíciója

A vektor fogalmának általánosításához kiválasztjuk a síkvektorok fenti tulajdonságai közül azokat, amelyeket a leglényegesebbnek tartunk, és az általános megfogalmazáshoz elengedhetetlenek. A továbbiakban minden olyan struktúrát, amely ezekkel a tulajdonságokkal rendelkezik, vektortérnek fogunk nevezni. Ezek a kiválasztott tulajdonságok az iii) és iv), azaz a síkvektorokra értelmezett összeadás és skalárral való szorzás művelete. Röviden: ha valamely tetszőleges halmazon értelmezni tudunk a fentiekkel megegyező tulajdonságú két műveletet (amelyeket összeadásnak és skalárral való szorzásnak nevezünk), akkor az így kapott struktúrát (a halmazt a két művelettel ellátva) vektortérnek fogjuk nevezni, és a halmaz elemeit vektoroknak.

Ahhoz, hogy összeadást értelmezhessünk valamilyen nem síkvektorokból álló, hanem

egyéb, pl. függvényekből álló halmazon, tisztáznunk kell, hogy mit is értünk összeadáson. A síkvektorok összeadásakor két síkvektorhoz hozzárendelünk egy harmadikat, azaz az összeadás művelete nem más, mint egy $\mathcal{E}_2 \times \mathcal{E}_2 \rightarrow \mathcal{E}_2$ leképezés. Ha tehát van egy tetszőleges X halmazunk, akkor az összeadás értelmezése azt jelenti, hogy definiálunk egy $X \times X \rightarrow X$ leképezést. Egy ilyen leképezést azonban csak akkor fogunk összeadásnak nevezni, ha tulajdonságaiban hasonlítani fog a síkvektorok összeadásához. A síkvektorok összeadásának négy jól ismert alapvető tulajdonsága van:

1. $\underline{a} + \underline{b} = \underline{b} + \underline{a} \quad \forall \underline{a}, \underline{b} \in \mathcal{E}_2$
2. $(\underline{a} + \underline{b}) + \underline{c} = \underline{a} + (\underline{b} + \underline{c}) \quad \forall \underline{a}, \underline{b}, \underline{c} \in \mathcal{E}_2$
3. A nullvektorra igaz az, hogy $\underline{a} + \underline{0} = \underline{a} \quad \forall \underline{a} \in \mathcal{E}_2$
4. Minden $\underline{a} \in \mathcal{E}_2$ elemhez létezik negatív elem $\mathcal{E}_2$ -ben, azaz olyan elem (jelölje $-\underline{a}$), amelyre $\underline{a} + (-\underline{a}) = \underline{0}$.

Ha tudunk olyan $X \times X \rightarrow X$ leképezést definiálni, amely szintén rendelkezik mindezen négy tulajdonsággal, akkor ezt a műveletet X -beli összeadásnak fogjuk nevezni.

A másik művelet, a síkvektorok skalárral való szorzása során egy síkvektorhoz és egy valós számhoz rendelünk egy síkvektort, ez a művelet tehát egy $\mathcal{E}_2 \times \mathbb{R} \rightarrow \mathcal{E}_2$ leképezés. Tulajdonságai a következők:

1. $\lambda \cdot (\underline{a} + \underline{b}) = \lambda \cdot \underline{a} + \lambda \cdot \underline{b} \quad \forall \underline{a}, \underline{b} \in \mathcal{E}_2 \text{ és } \forall \lambda \in \mathbb{R}$
2. $(\lambda + \mu) \cdot \underline{a} = \lambda \cdot \underline{a} + \mu \cdot \underline{a} \quad \forall \underline{a} \in \mathcal{E}_2 \text{ és } \forall \lambda, \mu \in \mathbb{R}$
3. $\lambda \cdot (\mu \cdot \underline{a}) = (\lambda \cdot \mu) \cdot \underline{a} = \mu \cdot (\lambda \cdot \underline{a}) \quad \forall \underline{a} \in \mathcal{E}_2 \text{ és } \forall \lambda, \mu \in \mathbb{R}$

(Megjegyezzük, hogy a 2. tulajdonságban két különböző típusú összeadás szerepel az egyenlőség két oldalán.) Ha tudunk olyan $X \times \mathbb{R} \rightarrow X$ leképezést definiálni, amely szintén rendelkezik mindezen három tulajdonsággal, akkor ezt a műveletet X -beli skalárral való szorzásnak fogjuk nevezni. Ezek után megfogalmazzuk a vektortér általános definícióját:

1.1 Definíció. Legyen X egy tetszőleges halmaz, amelyen értelmezve van egy $\oplus : X \times X \rightarrow X$ és egy $\odot : X \times \mathbb{R} \rightarrow X$ művelet a következő tulajdonságokkal:

1. $x \oplus y = y \oplus x \quad \forall x, y \in X$
2. $(x \oplus y) \oplus z = x \oplus (y \oplus z) \quad \forall x, y, z \in X$
3. Létezik nullelem (nullvektor), azaz olyan $0_X \in X$ elem, amelyre $x \oplus 0_X = x \quad \forall x \in X$

4. Minden $x \in X$ elemhez létezik negatív elem, azaz olyan elem (jelölje $-x$), amelyre $x \oplus (-x) = 0_X$, ahol 0_X a 3. tulajdonság szerinti nullvektor.

5. $\lambda \odot (x \oplus y) = (\lambda \odot x) \oplus (\lambda \odot y) \quad \forall x, y \in X \text{ és } \forall \lambda \in \mathbb{R}$

6. $(\lambda + \mu) \odot x = (\lambda \odot x) \oplus (\mu \odot x) \quad \forall x \in X \text{ és } \forall \lambda, \mu \in \mathbb{R}$

7. $\lambda \odot (\mu \odot x) = (\lambda \cdot \mu) \odot x = \mu \odot (\lambda \odot x) \quad \forall x \in X \text{ és } \forall \lambda, \mu \in \mathbb{R}$

Ekkor az $(X, \oplus, \odot)$ rendezett hármast vektortérnek, az X elemeit pedig vektoroknak nevezzük.

1.2 Megjegyzés. A műveleti jelek köré írt karika azt hivatott kiemelni, hogy újonnan definiált műveletekről: tetszőleges vektortér elemein értelmezett összeadásról és skalárral való szorzásról van szó. A későbbiekben, amikor már egyértelmű lesz, hogy egy műveleti jel mit fejez ki, a karikákat elhagyjuk. Azt a lazaságot is elkövetjük, hogy „az $(X, \oplus, \odot)$ vektortér” helyett egyszerűen csak „az X vektortér” megfogalmazást illetve jelölést fogjuk használni, amikor nem akarjuk hangsúlyozni a műveleteket. Továbbá, ha összeadás és szorzás is szerepel egy kifejezésben, és nincs kitéve zárójel, akkor mindig a szorzást kell előbb elvégezni, hasonlóan ahhoz, ahogyan a valós számoknál már megszokhattuk.

A fenti definíciós tulajdonságokból levezethető néhány további alapvető tulajdonság. Pl. belátható, hogy minden vektortérben csak egyetlen nullvektor létezik, továbbá minden elemhez csak egyetlen negatív elem tartozik.

Ha $(X, \oplus, \odot)$ egy vektortér, akkor előfordulhat, hogy az X halmaz valamely részhalmaza maga is vektorteret alkot ugyanazon műveletekkel.

1.3 Definíció. Legyen $Y \subset X$. Azt mondjuk, hogy az $(X, \oplus, \odot)$ vektortérnek altere $(Y, \oplus, \odot)$, ha $(Y, \oplus, \odot)$ vektortér.

Belátható, hogy ahhoz, hogy egy vektortér részhalmazáról eldöntsük, hogy maga is vektortér-e ugyanazon műveletekkel, azaz alteret kaptunk-e, elég a két műveletre való zártságot ellenőrizni, mert ha az igaz, akkor a hét műveleti tulajdonság a részhalmazon automatikusan teljesül.

1.3. Példák vektorterekre

Különböző konkrét X halmazokhoz megadunk műveleteket. Vektorteret kapunk-e?

- $X := \mathcal{E}_2$ + a síkvektorok szokásos összeadása és skalárral való szorzása - VT

- $X :=$ az egységnyi hosszúságú síkvektorok halmaza + a síkvektorok szokásos összeadása és skalárral való szorzása - NEM VT
- $X := \mathbb{R}$ + a valós számok szokásos összeadása és szorzása - VT
- $X := \mathbb{C}$ + a komplex számok szokásos összeadása és szorzása
- $X := \mathbb{R}^2$ + a számpárokon már értelmezett összeadás és skalárral való szorzás - VT
- $X := \mathbb{R}^n$, $n \in \mathbb{N}$ tetsz. + a szám-n-eseken már értelmezett összeadás és skalárral való szorzás - VT
- $X := P_n$ (legfeljebb n -edfokú polinomok) + a függvények szokásos összeadása és skalárral való szorzása - VT
- $X := \tilde{P}_n$ (szigorúan n -edfokú polinomok) + a függvények szokásos összeadása és skalárral való szorzása - NEM VT
- $X = C[a, b]$ (az $[a, b] \subset \mathbb{R}$ intervallumon értelmezett folytonos valós függvények) + a függvények szokásos összeadása és skalárral való szorzása - VT
- $X := D(a, b)$ (az $(a, b) \subset \mathbb{R}$ intervallumon értelmezett differenciálható valós függvények) + a függvények szokásos összeadása és skalárral való szorzása - VT
- $X := R[a, b]$ (az $[a, b] \subset \mathbb{R}$ intervallumon értelmezett Riemann-integrálható valós függvények) + a függvények szokásos összeadása és skalárral való szorzása - VT

1.4. Fontos fogalmak vektorterekben

Az általános vektortérfogalom bevezetése után mindazok a fogalmak és a síkvektorok azon tulajdonságai, amelyek az összeadás és a skalárral való szorzás műveletével kapcsolatosak, értelmesek lesznek bármilyen vektortérben. Nézzünk meg néhány alapvető definíciót.

Legyen $(X, \oplus, \odot)$ egy tetszőleges vektortér.

1. Lineáris kombináció

1.4 Definíció. Az $x_1, x_2, \dots, x_n \in X$ vektorok lineáris kombinációjának nevezzük az

$$\alpha_1 \odot x_1 \oplus \alpha_2 \odot x_2 \oplus \dots \oplus \alpha_n \odot x_n, \quad \alpha_i \in \mathbb{R}, \quad i = 1, 2, \dots, n$$

alakú kifejezést.

2. Lineáris függetlenség

1.5 Definíció. Azt mondjuk, hogy az $x_1, x_2, \dots, x_n \in X$ vektorok lineárisan függetlenek, ha

$$\alpha_1 \odot x_1 \oplus \alpha_2 \odot x_2 \oplus \dots \oplus \alpha_n \odot x_n = 0_X \Leftrightarrow \alpha_i = 0, \quad i = 1, 2, \dots, n.$$

A definíció tetszőleges (tehát akár végtelen sok elemű) vektorhalmazra a következőképpen szól:

1.6 Definíció. Azt mondjuk, hogy az $Y \subset X$ halmaz lineárisan független, ha Y bármely véges részhalmaza lineárisan független.

3. Lineáris burok

1.7 Definíció. Az $x_1, x_2, \dots, x_n \in X$ vektorok lineáris burkának nevezzük az $x_1, x_2, \dots, x_n$ összes lineáris kombinációjának a halmazát. Jelölése: $\text{span}[x_1, x_2, \dots, x_n]$.

Megjegyezzük, hogy $\text{span}[x_1, x_2, \dots, x_n]$ mindig altere az $(X, \oplus, \odot)$ vektortérnek.

4. Vektortér generálórendszere:

1.8 Definíció. A $G \subset X$ halmazt az $(X, \oplus, \odot)$ vektortér generálórendszerének nevezzük, ha X minden eleme előállítható G elemeinek lineáris kombinációjával.

5. Vektortér bázisa

1.9 Definíció. A $B \subset X$ halmazt az $(X, \oplus, \odot)$ vektortér bázisának nevezzük, ha B generálórendszere az $(X, \oplus, \odot)$ vektortérnek, és B lineárisan független halmaz.

Belátható, hogy egy vektortérben minden bázis azonos elemszámú (számosságú). A bázist más szóval koordináta-rendszernek nevezzük.

6. Vektortér dimenziója

1.10 Definíció. Az $(X, \oplus, \odot)$ vektortér dimenziójának a benne lévő bázisok elemszámát nevezzük. Jelölése: $\dim X$. Ha $\dim X < +\infty$, akkor a vektorteret véges dimenziós vektortérnek nevezzük.

A bevezetőben tárgyalt síkvektoroknak vannak egyéb speciális tulajdonságai is, amelyeket ott felsoroltunk. Ezekkel nem feltétlenül rendelkezik bármilyen vektortér. Pl. nem biztos, hogy van hosszuk és irányuk az elemeknek, továbbá nem tartozik lényegileg a vektorfogalomhoz, hogy értelmes a skaláris szorzat. A későbbiekben még a hosszúság és a skaláris szorzás fogalmát általánosítjuk.

1.4.1. A véges dimenziós vektorterek kapcsolata az $\mathbb{R}^n$ vektortérrel

Ebben az alfejezetben megmutatjuk, hogy a véges dimenziós vektorterek között kitüntetett szerepe van az $\mathbb{R}^n$ vektortérnek.

Legyen $(X, \oplus, \odot)$ egy tetszőleges n dimenziós vektortér ($n \in \mathbb{N}$ tetsz.) Ekkor X -ben létezik n elemű bázis: $B := \{b_1, \dots, b_n\} \subset X$. A B halmaz generálórendszere az $(X, \oplus, \odot)$ vektortérnek, azaz tetszőleges $x \in X$ elemet elő lehet állítani a B elemeinek lineáris kombinációjával, jelölje az együtthatókat rendre $\alpha_1, \dots, \alpha_n$:

$$x = \alpha_1 b_1 + \dots + \alpha_n b_n.$$

(Itt már elhagytuk a karikákat a műveleti jelekről, sőt, a valós számok szorzásához hasonlóan a szorzást jelölő pontot is elhagyjuk.) Továbbá, B lineárisan független rendszer, amiből következik, hogy a fenti előállítás egyértelmű. Tegyük fel ugyanis, hogy a fenti x elemet más együtthatókkal is megkaphatjuk:

$$x = \tilde{\alpha}_1 b_1 + \dots + \tilde{\alpha}_n b_n.$$

Ekkor a fenti két egyenlőséget kivonva egymásból, a

$$0_X = (\alpha_1 - \tilde{\alpha}_1) b_1 + \dots + (\alpha_n - \tilde{\alpha}_n) b_n$$

egyenlőséget kapjuk. Mivel a $b_1, \dots, b_n$ vektorok lineárisan függetlenek, a jobb oldali lineáris kombináció csak akkor adhatja X nullvektorát, ha mindegyik együttható nulla. Ebből az következik, hogy nincs két különböző előállítás.

A fenti gondolatmenetből az következik, hogy ha X -ben rögzítettünk egy B bázist, akkor ezen bázis segítségével X minden elemének egyértelműen megfeleltethetünk egy szám-nest, nevezetesen, a B bázisbeli előállításában szereplő együtthatókból állót. Azaz pl. a fenti x elemnek megfeleltethetjük az $(\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n$ elemet. Ez a megfeleltetés (bijekció) ráadásul még annyiban is speciális, hogy művelettartó, azaz két X -beli elem összegéhez a megfelelő $\mathbb{R}^n$ -beli elemek összegét rendeli, és bármely X -beli elem skalárszorához a megfelelő $\mathbb{R}^n$ -beli elem skalárszorását rendeli.

Fontos következmény, hogy bármely n (véges) dimenziós X vektortér azonosítható az $\mathbb{R}^n$ vektortérrel, azaz X elemei megfeleltethetők szám-n-eseknek, és ha X elemein összeadást és skalárral való szorzást kell elvégezni, akkor a műveleteket a megfelelő szám-n-esekkel is elvégezhetjük. Ilyen értelemben elegendő a véges dimenziós vektorterek közül csak az $\mathbb{R}^n$ vektortérrel foglalkozni. Ez az azonosíthatóság végtelen dimenziós vektorterekre már

nem igaz, így pl. nem azonosíthatunk minden folytonos függvényt egy szám-n-essel.

2. Lineáris leképezések

Ebben a fejezetben vektorterek között ható leképezésekről (függvényekről) lesz szó. Ezek között a legegyszerűbbek és a gyakorlatban igen fontosak a lineáris leképezések.

2.1 Definíció. Legyen $(X_1, \oplus_1, \odot_1)$ és $(X_2, \oplus_2, \odot_2)$ két vektortér. Az $f : X_1 \rightarrow X_2$ leképezést lineárisnak nevezzük, ha igaz rá a következő két tulajdonság:

1. $f(x_1 \oplus_1 x_2) = f(x_1) \oplus_2 f(x_2) \quad \forall x_1, x_2 \in X_1,$
2. $f(\lambda \odot_1 x) = \lambda \odot_2 f(x) \quad \forall x \in X_1, \forall \lambda \in \mathbb{R}.$

Az $X_1 \rightarrow X_2$ képező lineáris leképezések halmazát $\text{Hom}(X_1, X_2)$ jelöli.

Megjegyezzük, hogy egy lineáris leképezésnek nemcsak az értelmezési tartománya vektortér, de a képtere is:

2.2 Állítás. Ha $f : X_1 \rightarrow X_2$ lineáris leképezés, akkor f képtere $(R(f))$ altere X_2 -nek.

Biz.:

Belátandó, hogy $R(f)$ zárt az összeadásra és a skalárral való szorzásra.

- 1.) Összeadás: Legyenek y_1 és y_2 tetszőleges $R(f)$ -beli elemek. Ekkor létezik $x_1, x_2 \in X_1$, amelyekre $f(x_1) = y_1, f(x_2) = y_2$. Az f linearitása miatt $y_1 + y_2 = f(x_1) + f(x_2) = f(x_1 \oplus_1 x_2)$, ami eleme $R(f)$ -nek, tehát $y_1 + y_2 \in R(f)$.
- 2.) Skalárral való szorzás: Legyen $y \in R(f)$ tetszőleges. Ekkor létezik $x \in X_1 : f(x) = y$. Tetszőleges $\lambda \in \mathbb{R}$ esetén f linearitása miatt $\lambda y = \lambda f(x) = f(\lambda x)$, ami eleme $R(f)$ -nek, tehát $\lambda y \in R(f)$.

A legegyszerűbb esetben X_1 és X_2 is $\mathbb{R}$, azaz a leképezés valós-valós függvény. Matematika 2 órán láthattuk, hogy ezek között a linearitás definíciójának csak az $f(x) = ax$ alakú függvények tesznek eleget, ahol $a \in \mathbb{R}$ rögzített szám. Továbbá, ha $X_1 = \mathbb{R}^n$ és $X_2 = \mathbb{R}^m$, akkor az $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ lineáris leképezés megfelel egy $m \times n$ -es mátrixszal való szorzásnak. A lineáris leképezések és a mátrixok között műveleteket is értelmeztünk: összeadás, skalárral való szorzás, kompozíció (mátrixszorzás). A továbbiakban a lineáris leképezések és mátrixok további érdekes tulajdonságaival foglalkozunk.

2.1. Lineáris leképezések invertálása, az inverz mátrix

A lineáris leképezés függvény, ezért értelmes az a kérdés, hogy egy lineáris leképezésnek mikor létezik inverze. Idézzük fel, hogy mit értettünk általában egy X halmazból Y halmazba képező f függvény inverzén. (Az egyszerűség kedvéért tegyük fel, hogy f az egész X -en értelmezve van.) Láttuk, hogy f akkor invertálható, ha injektív, azaz egy $R(f)$ -beli elemet csak egyetlen X -beli elemhez rendel hozzá. Az f függvény inverze definíció szerint az az $f^{-1} : R(f) \rightarrow X$ függvény, amelyre minden $y \in R(f)$ pontban $f^{-1}(y) = x$, ahol $y = f(x)$. Vagyis ez az f függvény fordítottja. Emlékeztetünk két fontos állításra:

2.3 Állítás. *Ha f invertálható, akkor $f \circ f^{-1} = \text{id}_{R(f)}$ (az $R(f) \rightarrow R(f)$ képező identitásfüggvény), és $f^{-1} \circ f = \text{id}_X$ (az $X \rightarrow X$ képező identitásfüggvény).*

2.4 Állítás. *Egy f függvény inverze mindig invertálható, és $(f^{-1})^{-1} = f$.*

2.5 Állítás. *Az inverz függvény egyértelmű.*

Biz.:

A bizonyítás indirekt. Tegyük fel, hogy az $\tilde{f}^{-1} : R(f) \rightarrow X$ függvény is inverze f -nek, de $\tilde{f}^{-1} \neq f^{-1}$. Ekkor

$$\tilde{f}^{-1} = \tilde{f}^{-1} \circ \text{id}_{R(f)} = \tilde{f}^{-1} \circ (f \circ f^{-1}) = (\tilde{f}^{-1} \circ f) \circ f^{-1} = \text{id}_X \circ f^{-1} = f^{-1}.$$

Tehát a két inverz függvény nem lehet különböző.

Ha X_1, X_2 vektorterek, és $f : X_1 \rightarrow X_2$ lineáris leképezés, akkor láttuk, hogy f képtere altére X_2 -nek. Így ha f invertálható, akkor az f^{-1} inverz leképezés értelmezési tartománya, $D(f^{-1})$ is mindig vektortér (hiszen $D(f^{-1}) = R(f)$). A továbbiakban az $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ típusú lineáris leképezések inverzével foglalkozunk.

2.6 Állítás. *Ha $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ lineáris és invertálható, akkor $f^{-1} : \mathbb{R}^m \supset \rightarrow \mathbb{R}^n$ is lineáris.*

Biz.:

Vizsgáljuk meg a linearitás két tulajdonságát!

1. Igaz-e, hogy $f^{-1}(a + b) = f^{-1}(a) + f^{-1}(b) \quad \forall a, b \in D(f^{-1})$?

Vezessük be az $x := f^{-1}(a)$ és $y := f^{-1}(b)$ jelölést. Mivel f lineáris,

$$f(x) + f(y) = f(x + y),$$

azaz

$$a + b = f(f^{-1}(a) + f^{-1}(b)).$$

Alkalmazzuk mindkét oldalra az f^{-1} függvényt! Ezzel a belátandó tulajdonágot kapjuk:

$$f^{-1}(a + b) = f^{-1}(a) + f^{-1}(b).$$

2. Igaz-e, hogy $f^{-1}(\lambda a) = \lambda f^{-1}(a) \quad \forall a \in D(f^{-1}), \forall \lambda \in \mathbb{R}$?

Mivel f lineáris,

$$f(\lambda x) = \lambda f(x).$$

Az előbb bevezetett jelölést alkalmazva

$$f(\lambda f^{-1}(a)) = \lambda f(f^{-1}(a)) = \lambda a.$$

Alkalmazzuk ismét mindkét oldalra az f^{-1} függvényt! Ezzel

$$f^{-1}(\lambda a) = \lambda f^{-1}(a),$$

ami éppen a belátandó állítás.

2.7 Megjegyzés. Ez az állítás jóval általánosabban is igaz: minden $f : X \rightarrow Y$ (X és Y vektorterek) lineáris függvény inverze is lineáris.

Ismeretes, hogy egy $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ lineáris függvény a $0_n \in \mathbb{R}^n$ nullvektorhoz mindig a $0_m \in \mathbb{R}^m$ nullvektort rendeli hozzá. (Hiszen $f(0_n) = f(0 \cdot x)$, ahol $x \in \mathbb{R}^n$ tetszőleges vektor, és f linearitása miatt $f(0 \cdot x) = 0 \cdot f(x) = 0_m$. Itt kétszer is kihasználtuk azt az egyszerűen belátható tényt, hogy egy vektortér bármely elemét 0-val szorozva a vektortér nullvektorát kapjuk.) Egyszerűen belátható, hogy az invertálható lineáris függvényekre ennél több is igaz:

2.8 Tétel. Az $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ lineáris függvény pontosan akkor invertálható, ha a $0_m \in \mathbb{R}^m$ vektort **csak** a $0_n \in \mathbb{R}^n$ vektorhoz rendeli hozzá.

Biz.:

($\Rightarrow$) Tegyük fel, hogy az f lineáris leképezés invertálható. Ekkor semmilyen értéket nem vehet fel két különböző helyen, így a nullvektort sem $\Rightarrow$ csak a 0_n helyen veszi fel.

($\Leftarrow$) Tegyük fel, hogy f a $0_m \in \mathbb{R}^m$ vektort **csak** a $0_n \in \mathbb{R}^n$ vektorhoz rendeli hozzá. Belátandó, hogy f injektív, azaz $f(x_1) = f(x_2)$ esetén $x_1 = x_2$. Tegyük fel, hogy $f(x_1) = f(x_2)$. Ekkor $f(x_1) - f(x_2) = 0_m$. Mivel f , lineáris, így $f(x_1 - x_2) = 0_m$. De f a 0_m -et csak a 0_n -hez rendeli, így $x_1 - x_2 = 0_n \Rightarrow x_1 = x_2$.

Ennek következménye az alábbi állítás. (A továbbiakban feltesszük, hogy $R(f) = \mathbb{R}^m$, vagyis az egész $\mathbb{R}^m$ -be beleképez.)

2.9 Állítás. *Ha az $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ lineáris függvény invertálható, és $R(f) = \mathbb{R}^m$, akkor $n = m$.*

Bizonyítás helyett csak a legegyszerűbb eseteket gondoljuk végig.

1. Miért nem lehet $m < n$?

Legyen először $f : \mathbb{R}^2 \rightarrow \mathbb{R}$. Tudjuk, hogy ha f lineáris, akkor

$$f(x_1, x_2) = a_1x_1 + a_2x_2$$

alakú. Az $a_1x_1 + a_2x_2$ függvényérték milyen (x_1, x_2) esetén lehet 0? Természetesen nem csak akkor, ha $x_1 = 0$ és $x_2 = 0$. Ezért az előbbi állítás értelmében f nem lehet invertálható. Legyen most $f : \mathbb{R}^3 \rightarrow \mathbb{R}^2$. Az ilyen típusú lineáris függvény

$$f(x_1, x_2, x_3) = (a_1x_1 + a_2x_2 + a_3x_3, b_1x_1 + b_2x_2 + b_3x_3)$$

alakú. A jobb oldali vektor megint nem csak $x_1 = x_2 = x_3 = 0$ esetén lehet nulla, hiszen ez azt jelentené, hogy az

$$\begin{aligned} a_1x_1 + a_2x_2 + a_3x_3 &= 0 \\ b_1x_1 + b_2x_2 + b_3x_3 &= 0 \end{aligned}$$

lineáris egyenletrendszernek egyetlen megoldása van. Ez pedig nem lehetséges, mert ehhez az együtthatómátrix rangjának 3-nak kellene lennie (ennyi az ismeretlenek száma), de csak két sora van.

Általában is belátható, hogy $m < n$ esetén a függvény nem lehet invertálható.

2. Miért nem lehet $m > n$? Mert ekkor az 1. pont értelmében az $f^{-1} : \mathbb{R}^m \rightarrow \mathbb{R}^n$ függvény nem lehetne invertálható.

Ez az oka annak, hogy csak $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ függvények invertálhatóságával foglalkozunk.

Tudjuk, hogy egy $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ lineáris függvény megfelel egy $n \times n$ -es mátrixszal való szorzásnak. Mivel a függvény inverze szintén $\mathbb{R}^n \rightarrow \mathbb{R}^n$ képez, az inverz leképezésnek is megfelel egy $n \times n$ -es mátrix.

2.10 Definíció. *Egy $A \in \mathbb{R}^{n \times n}$ mátrixot invertálhatónak nevezünk, ha a neki megfelelő*

$f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ lineáris leképezés invertálható. Ekkor az $f^{-1} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ lineáris leképezés mátrixát az A mátrix inverzének nevezzük. Jelölése: A^{-1} .

Nyilvánvalóan $A \cdot A^{-1} = I$ és $A^{-1} \cdot A = I$.

Vizsgáljuk meg, hogy milyennek kell lennie egy mátrixnak ahhoz, hogy a mátrix, vagyis a neki megfeleltetett lineáris leképezés invertálható legyen! A 2.8 tételt az $m = n$ esetre alkalmazva, egy $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ lineáris leképezés pontosan akkor invertálható, ha $f(x) = 0_n$ csak az $x = 0_n$ helyen áll fenn. Mivel $f(x) = A \cdot x$, ahol A a lineáris leképezés mátrixa, ezért az a kérdés, hogy $A \cdot x = 0_n$ milyen feltétel mellett áll fenn csak az $x = 0_n$ vektorra. Erre a lineáris egyenletrendszer elméletéből tudjuk a választ: az $A \cdot x = 0_n$ lineáris egyenletrendszernek csak akkor egyedüli megoldása a 0_n vektor, ha $\det A \neq 0$. Vagyis egy A mátrix pontosan akkor invertálható, ha $\det A \neq 0$.

Szükségünk lehet mátrixszorzatok invertálására is. Ezzel kapcsolatos a következő állítás.

2.11 Állítás. *Többtényezős mátrixszorzat inverze az egyes tényezők inverzének fordított sorrendben vett szorzata.*

Ezt elég két mátrixra ellenőriznünk: lássuk be tehát, hogy

$$(AB)^{-1} = B^{-1}A^{-1}.$$

Kihasználva a mátrixszorzás asszociativitását:

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AA^{-1} = I.$$

2.12 Megjegyzés. *Az inverz mátrix kiszámításával az előző félévben foglalkoztunk.*

2.2. A transzponált mátrix

A transzponált mátrix fogalmával az első félévben már találkoztunk. Emlékeztetőül a definíció:

2.13 Definíció. *Az $A \in \mathbb{R}^{m \times n}$ mátrix transzponáltjának nevezzük azt az $A^T \in \mathbb{R}^{n \times m}$ mátrixot, amelynek elemeire*

$$a_{ij}^T := a_{ji}, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, m.$$

Ez tehát az a mátrix, amelyet A sorainak és oszlopainak a felcserélésével kapunk. Mi lehet a transzponált mátrix jelentése, azaz mint lineáris leképezés, mit fejez ki a transzponált

mátrix? Ehhez emlékeztetőnek először felidézzük az $\mathbb{R}^n$ -beli vektorok skaláris szorzatának a fogalmát.

2.14 Definíció. Az $x, y \in \mathbb{R}^n$ vektorok skaláris szorzatán az

$$\langle x, y \rangle := \sum_{i=1}^n x_i y_i$$

számot értjük.

Gondoljuk meg a következőt: vajon igaz-e tetszőleges $A \in \mathbb{R}^{n \times n}$ mátrixra, hogy minden x és $y \in \mathbb{R}^n$ vektor esetén fennáll az

$$\langle x, Ay \rangle = \langle Ax, y \rangle$$

egyenlőség?

Ellenpéldával könnyen megmutatható, hogy nem lehet mindig igaz, hiszen pl. az

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad x = (0, 1), \quad y = (1, 0)$$

megválasztással $\langle x, Ay \rangle = 3 \neq \langle Ax, y \rangle = 2$.

Vizsgáljuk meg, mi annak a feltétele a 2×2 -es mátrixok körében hogy $\langle x, Ay \rangle = \langle Ax, y \rangle$ igaz legyen minden $x, y \in \mathbb{R}^2$ esetén! Kiszámolva a két oldalt:

$$\langle x, Ay \rangle = x_1 a_{11} y_1 + x_1 a_{12} y_2 + x_2 a_{21} y_1 + x_2 a_{22} y_2,$$

és

$$\langle Ax, y \rangle = y_1 a_{11} x_1 + y_1 a_{12} x_2 + y_2 a_{21} x_1 + y_2 a_{22} x_2.$$

Látható, hogy a két kifejezés minden $x, y \in \mathbb{R}^2$ vektorra csak akkor teljesül, ha $a_{12} = a_{21}$, azaz az A mátrix szimmetrikus, másképpen $A = A^T$. Vagyis a mátrix szorzót csak akkor vihetjük át a skaláris szorzatban a másik vektorra, ha az A mátrix egyenlő a saját transzponáltjával.

Mindig igaz viszont a következő: ha $A \in \mathbb{R}^{m \times n}$, akkor az

$$\langle x, Ay \rangle = \langle A^T x, y \rangle$$

egyenlőség fennáll minden $x \in \mathbb{R}^m$ és $y \in \mathbb{R}^n$ esetén. Tehát ha a mátrix szorzót át akarjuk tenni a skaláris szorzatban a másik vektorra, akkor a mátrixot transzponálni is kell.

2.15 Definíció. Ha $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ lineáris leképezés, akkor az $f^* : \mathbb{R}^m \rightarrow \mathbb{R}^n$ lineáris leképezést az f leképezés adjungáltjának nevezzük, ha

$$\langle x, f(y) \rangle = \langle f^*(x), y \rangle \quad \forall x \in \mathbb{R}^m, y \in \mathbb{R}^n.$$

Ha tehát az f lineáris leképezés mátrixa A , akkor az A mátrix transzponáltja az f^* adjungált leképezést reprezentálja.

2.3. Lineáris leképezések tetszőleges véges dimenziós vektorterek között

Láttuk, hogy egy $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ lineáris leképezést megadhatunk egy $m \times n$ -es mátrixszal. Most látni fogjuk, hogy tetszőleges véges dimenziós vektortérből véges dimenziós vektortérbe képező lineáris operátor is mindig reprezentálható mátrixszal.

Legyen $f : X_1 \rightarrow X_2$ lineáris leképezés, ahol X_1, X_2 vektorterek, és $\dim X_1 = n$, $\dim X_2 = m$. Már volt róla szó, hogy egy n dimenzió vektortér azonosítható $\mathbb{R}^n$ -nel, tehát X_1 azonosítható az $\mathbb{R}^n$, X_2 pedig az $\mathbb{R}^m$ vektortérrel. Ez az azonosítás úgy történik, hogy rögzítünk X_1 -ben is és X_2 -ben is egy bázist, és az elemeket előállítjuk a báziselemek lineáris kombinációjával. Ilyen módon az $f : X_1 \rightarrow X_2$ lineáris leképezés azonosítható egy $\tilde{f} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ szintén lineáris leképezéssel, ez pedig megfeleltethető egy $m \times n$ -es mátrixnak. Keressük ezen mátrix elemeit.

Legyen az X_1 vektortérben kiválasztott bázis $B_1 = \{a_1, a_2, \dots, a_n\}$, az X_2 -beli bázis pedig $B_2 = \{b_1, b_2, \dots, b_m\}$. Tegyük fel, hogy az $x \in X_1$ vektor előállítása $x = x_1 a_1 + \dots + x_n a_n$, azaz x -et az $(x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ szám-n-essel azonosíthatjuk. Az x vektor képét jelölje y , azaz $y = f(x)$. Az y az X_2 vektortérben van, így felírható a $b_1, \dots, b_m$ báziselemek lineáris kombinációjaként: $y = y_1 b_1 + \dots + y_m b_m$. Azt keressük, hogy mi a kapcsolat az $(x_1, x_2, \dots, x_n)$ vektor és az $(y_1, \dots, y_m)$ vektor között, pontosabban milyen mátrixszal szorozva kapjuk meg az $(x_1, x_2, \dots, x_n)$ vektorból az $(y_1, \dots, y_m)$ vektort. Az $y = f(x)$ egyenlőségben y és x helyébe is írjuk be a felbontásukat:

$$y_1 b_1 + \dots + y_m b_m = f(x_1 a_1 + \dots + x_n a_n).$$

Az f leképezés lineáris, ezért a jobb oldalon f -et tagonként alkalmazhatjuk, és kihozhatjuk az $x_1, \dots, x_n$ szorzókat:

$$y_1 b_1 + \dots + y_m b_m = x_1 f(a_1) + \dots + x_n f(a_n). \quad (1)$$

Vegyük észre, hogy a jobb oldalon szereplő $f(a_1), \dots, f(a_n)$ vektorok X_2 -ben vannak (ezek az X_1 -beli báziselemek f -képei), tehát mindegyiket felírhatjuk a $b_1, \dots, b_m$ vektorok lineáris kombinációjaként:

$$\begin{aligned} f(a_1) &= \alpha_{11}b_1 + \dots + \alpha_{1m}b_m \\ &\vdots \\ f(a_n) &= \alpha_{n1}b_1 + \dots + \alpha_{nm}b_m \end{aligned}$$

Ezeket az összegeket beírva az (1) egyenlőség jobb oldalába, mindkét oldalon az X_2 -beli báziselemek egy-egy lineáris kombinációja lesz. Mivel a báziselemek lineárisan függetlenek, az egyenlőség pontosan akkor áll fenn, ha az egyes báziselemek együtthatója mindkét oldalon megegyezik. Ebből az

$$\begin{aligned} y_1 &= \alpha_{11}x_1 + \dots + \alpha_{n1}x_n \\ &\vdots \\ y_m &= \alpha_{1m}x_1 + \dots + \alpha_{nm}x_n \end{aligned}$$

összefüggésekhez jutunk. Ebből már látható, hogy a keresett mátrix

$$A = \begin{bmatrix} \alpha_{11} & \alpha_{21} & \dots & \alpha_{n1} \\ \alpha_{12} & \alpha_{22} & \dots & \alpha_{n2} \\ \vdots & & & \\ \alpha_{1m} & \alpha_{2m} & \dots & \alpha_{nm} \end{bmatrix}$$

alakú. Azaz a mátrix i -edik oszlopában az a_i báziselem f -képének a B_2 -bázisbeli koordinátái vannak. A mátrix felírásához tehát csak az X_1 -beli n darab báziselem képét kell ismernünk. A kapott mátrix alakja természetesen függ attól, hogy milyen bázistadtunk meg X_1 -ben és X_2 -ben. Ezért a leképezés (B_1, B_2) bázispárhoz tartozó mátrixrepresentációjáról beszélünk.

2.4. Mátrixok sajátértékei és sajátvektorai

Azt mondjuk, hogy az $A \in \mathbb{R}^{n \times n}$ mátrixnak a $\lambda \in \mathbb{R}$ szám a sajátértéke, ha létezik olyan $x \in \mathbb{R}^n$, $x \neq 0_n$ vektor, amelyre $Ax = \lambda x$. Ekkor az x vektort az A mátrix λ számhoz tartozó sajátvektorának nevezzük.

Megjegyezzük, hogy a sajátérték komplex szám is lehet, és a sajátvektorok is lehetnek komplex elemű vektorok. Ebben a fejezetben azonban mi csak valós sajátértékekkel fogunk foglalkozni, és mindig csak valós elemű sajátvektorokat keresünk.

Pl. 1. Az

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

mátrix minden vektort a kétszeresére nyújt. Azaz A -nak sajátértéke a 2, és sajátvektora minden $\mathbb{R}^2$ -beli nemnulla vektor.

2. Határozzuk meg az

$$A = \begin{bmatrix} 7 & -2 \\ 4 & 1 \end{bmatrix}$$

mátrix sajátértékeit és sajátvektorait!

Megoldás: λ sajátérték $\Leftrightarrow$ létezik $x \neq 0_n$, amelyre $Ax = \lambda x$.

Az $Ax = \lambda x$ egyenlet másképpen úgy is írható: $Ax - \lambda x = 0_n$. A bal oldalon x kiemelhető a következőképpen:

$$(A - \lambda I)x = 0_n,$$

ahol I jelöli a 2×2 -es identitásmátrixot. Ez az egyenlet egy homogén (azaz 0_n jobb oldalú) lineáris egyenletrendszerrel jelent az x sajátvektorra.

Összegezve: a λ szám pontosan akkor sajátértéke az A mátrixnak, ha az $(A - \lambda I)x = 0_n$ homogén lineáris egyenletrendszernek létezik $x \neq 0_n$ megoldása.

2.16 Megjegyzés. • *A Kronecker–Capelli-tételből következik, hogy egy $Bx = 0_n$ homogén lineáris egyenletrendszernek mindig van megoldása.*

- *Világos, hogy mindig megoldása az $x = 0_n$ (triviális megoldás).*
- *Ha $|B| \neq 0$, akkor a megoldás egyértelmű (azaz csak a triviális megoldás van).*
- *Ha $|B| = 0$, akkor végtelen sok megoldás van (azaz csak ekkor van nemtriviális megoldás).*

Bennünket az utolsó eset érdekel: az $(A - \lambda I)\mathbf{x} = 0_n$ egyenletrendszernek pontosan akkor van $x \neq 0_n$ megoldása, amikor $|A - \lambda I| = 0$. A kérdés tehát az, hogy milyen λ esetén lesz $|A - \lambda I| = 0$.

A példában

$$A - \lambda I = \begin{bmatrix} 7 & -2 \\ 4 & 1 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} 7 - \lambda & -2 \\ 4 & 1 - \lambda \end{bmatrix}$$

Ennek determinánsa:

$$|A - \lambda I| = \begin{vmatrix} 7 - \lambda & -2 \\ 4 & 1 - \lambda \end{vmatrix} = (7 - \lambda)(1 - \lambda) + 8 = \lambda^2 - 8\lambda + 15.$$

Azt keressük, hogy ez milyen λ értékekre 0, azaz meg kell oldani a

$$\lambda^2 - 8\lambda + 15 = 0$$

másodfokú egyenletet. A megoldóképletből:

$$\lambda_{1,2} = \frac{8 \pm \sqrt{64 - 60}}{2} \Rightarrow \lambda_1 = 5, \lambda_2 = 3.$$

Ezek tehát a keresett sajátértékek. Most rátérünk a hozzájuk tartozó sajátvektorok meghatározására.

A $\lambda_1 = 5$ -höz tartozó sajátvektorok meghatározása: $\underline{u}$ sajátvektor, ha $A\underline{u} = \lambda_1\underline{u}$, azaz $(A - \lambda_1 I)\underline{u} = \underline{0}$. Ez egy homogén lineáris egyenletrendszer, amelyben keressük az ismeretlen $\underline{u}$ vektor u_1 és u_2 elemét. A megoldandó egyenletrendszer tehát:

$$2u_1 - 2u_2 = 0$$

$$4u_1 - 4u_2 = 0$$

A második egyenlet nem független az elsőtől. (Ennek így is kell lennie, mert végtelen sok megoldást kell kapnunk.) Minden olyan nemnulla (u_1, u_2) számpár megoldás, amely kielégíti az első egyenletet, azaz amelyre $u_1 = u_2$. Így az összes megoldás

$$\underline{u} = \begin{bmatrix} p \\ p \end{bmatrix}, \quad p \in \mathbb{R}, \quad p \neq 0 \text{ tetszőleges.}$$

Ezzel megadtuk a $\lambda_1 = 5$ -höz tartozó sajátvektorokat. Ellenőrzésképpen:

$$\begin{bmatrix} 7 & -2 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} p \\ p \end{bmatrix} = \begin{bmatrix} 5p \\ 5p \end{bmatrix},$$

azaz az A mátrix valóban 5-szörösére nyújtja a (p, p) alakú vektorokat.

Hasonlóan számíthatók ki a $\lambda_2 = 3$ -hoz tartozó $\underline{v}$ sajátvektorok. A megoldás:

$$\underline{v} = \begin{bmatrix} q \\ 2q \end{bmatrix}, \quad q \in \mathbb{R}, \quad q \neq 0 \text{ tetszőleges.}$$

Vegyük észre, hogy a sajátvektorról kikötöttük, hogy az nem lehet a nullvektor, de a sajátérték lehet nulla: nevezetesen, ha $|A| = 0$ (hiszen ekkor $A - \lambda I = A$).

A továbbiakban jelölje $\sigma(A)$ az A mátrix sajátértékeinek a halmazát.

Egyes speciális mátrixok sajátértékeit könnyű meghatározni. Pl. legyen $D \in \mathbb{R}^{n \times n}$ diagonális mátrix, a főátlóban a $d_1, d_2, \dots, d_n$ számokkal:

$$D = \begin{bmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & & \vdots \\ \vdots & & & 0 \\ 0 & \dots & 0 & d_n \end{bmatrix}$$

A D mátrix jelölése röviden: $\text{diag}[d_1, d_2, \dots, d_n]$. Számítsuk ki a sajátértékeit!

Egy λ szám pontosan akkor sajátértéke D -nek, ha $|D - \lambda I| = 0$. A $D - \lambda I$ mátrix a következő diagonálmátrix lesz:

$$D - \lambda I = \begin{bmatrix} d_1 - \lambda & 0 & \dots & 0 \\ 0 & d_2 - \lambda & & \vdots \\ \vdots & & & 0 \\ 0 & \dots & 0 & d_n - \lambda \end{bmatrix}$$

Ismeretes, hogy egy diagonális mátrix determinánsa a főátlóban lévő elemek szorzata, azaz $|D - \lambda I| = (d_1 - \lambda)(d_2 - \lambda) \dots (d_n - \lambda)$. Ez a szorzat pedig λ következő értékeire lesz nulla: $\lambda_1 = d_1, \lambda_2 = d_2, \dots, \lambda_n = d_n$. Azaz D sajátértékei a főátlóban lévő számok: $\sigma(D) = \{d_1, d_2, \dots, d_n\}$.

Mik lesznek a hozzájuk tartozó sajátvektorok, ha nincs két egyforma elem a főátlóban?

A $\lambda_1 = d_1$ sajátértékhez tartozó sajátvektor olyan nemnulla $\underline{u} = (u_1, u_2, \dots, u_n)$ vektor, amelyre $D\underline{u} = d_1\underline{u}$. Ez a következő egyenletrendszerrel jelenti:

$$\begin{aligned} d_1 u_1 &= d_1 u_1 \\ d_2 u_2 &= d_1 u_2 \\ &\vdots \\ d_n u_n &= d_1 u_n \end{aligned}$$

Ez n darab független egyenlet a sajátvektor elemeire. Az első egyenletnek minden $u_1 \in \mathbb{R}$ szám eleget tesz. A többi egyenlet megoldása: $u_2 = 0, u_3 = 0, \dots, u_n = 0$. Azaz: $u_1 = p \in \mathbb{R}$ tetszőleges, $u_i = 0, i = 2, 3, \dots, n$. Így a d_1 sajátértékhez tartozó sajátvektor pl. az $(1, 0, \dots, 0) \in \mathbb{R}^n$ vektor.

Hasonlóan, a λ_i sajátértékhez tartozó sajátvektorok azok a vektorok, amelyek i -edik eleme tetszőleges, és a többi elemük nulla.

Megjegyezzük, hogy a háromszögmátrixokra is igaz, hogy sajátértékeik a főátlóban lévő elemek.

Feladat. Tegyük fel, hogy $\lambda \in \sigma(A)$. Mutassuk meg, hogy ekkor $\lambda^2 \in \sigma(A^2)$!

Megoldás: Legyen x egy λ -hoz tartozó sajátvektora az A -nak. Ekkor

$$A^2x = A(Ax) = A(\lambda x) = \lambda \cdot Ax = \lambda^2 x.$$

Mivel $x \neq 0_n$ (hiszen sajátvektora A -nak), ezért ez pontosan azt jelenti, hogy λ^2 sajátértéke A^2 -nek.

A következő állítás a különböző sajátértékekhez tartozó sajátvektorok egy fontos tulajdonságáról szól.

2.17 Tétel. Az $A \in \mathbb{R}^{n \times n}$ mátrix különböző, $\lambda_1, \lambda_2, \dots, \lambda_k$ sajátértékekhez tartozó sajátvektorai lineárisan függetlenek.

Biz.:

Teljes indukcióval!

- $k = 1$ -re nyilvánvalóan igaz az állítás, hiszen egy darab nemnulla vektor önmagában lineárisan független.
- Tegyük fel, hogy az állítás igaz $k \in \mathbb{N}$ -re, és mutassuk meg, hogy akkor igaz $k + 1$ -re is. A bizonyításnak ez a része indirekt. Tegyük fel, hogy a $k + 1$ darab sajátvektor (amelyeket jelöljön $x_1, x_2, \dots, x_k$) lineárisan összefüggő. Ekkor léteznek olyan $\alpha_1, \alpha_2, \dots, \alpha_{k+1}$ nem csupa nulla számok, hogy

$$\alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_k x_k + \alpha_{k+1} x_{k+1} = 0_n. \quad (2)$$

Tegyük fel, hogy itt $\alpha_1 \neq 0$, és szorozzuk meg az egyenlőség mindkét oldalát balról az A mátrixszal.

$$A\alpha_1 x_1 + A\alpha_2 x_2 + \dots + A\alpha_k x_k + A\alpha_{k+1} x_{k+1} = 0_n.$$

A bal oldali tagokban az α_i szorzókat előrehozzuk, és kihasználjuk, hogy $Ax_i = \lambda_i x_i$:

$$\alpha_1 \lambda_1 x_1 + \alpha_2 \lambda_2 x_2 + \dots + \alpha_k \lambda_k x_k + \alpha_{k+1} \lambda_{k+1} x_{k+1} = 0_n. \quad (3)$$

A (2) egyenlőséget szorozzuk végig λ_{k+1} -gyel:

$$\alpha_1 \lambda k + 1x_1 + \alpha_2 \lambda k + 1x_2 + \dots \alpha_k \lambda k + 1x_k + \alpha_{k+1} \lambda k + 1x_{k+1} = 0_n, \quad (4)$$

majd vonjuk ki (3)-ból (4)-et:

$$\alpha_1(\lambda_1 - \lambda k + 1)x_1 + \alpha_2(\lambda_2 - \lambda k + 1)x_2 + \dots \alpha_k(\lambda_k - \lambda k + 1)x_k = 0_n,$$

az utolsó tag ugyanis kiesett. A kapott egyenlőségben a bal oldal első tagjának együtthatója λ_1 és λ_k különbözősége valamint $\alpha_1 \neq 0$ miatt nem lehet nulla, így viszont a bal oldalon az $x_1, x_2, \dots, x_k$ vektorok egy nemtriviális lineáris kombinációja áll. Ez ellentmond annak a feltevésnek, hogy ezen $x_1, x_2, \dots, x_k$ sajátvektorok lineárisan függetlenek. Így a $k + 1$ darab sajátvektor nem lehet lineárisan összefüggő.

Ebből az állításból következik, hogy ha egy mátrixnak van n darab különböző sajátértéke, akkor az azokhoz tartozó sajátvektorok bázist alkotnak $\mathbb{R}^n$ -ben.

2.4.1. Mátrixok spektrálfelbontása

Megmutatható, hogy ha az $A \in \mathbb{R}^{n \times n}$ mátrixnak van n darab különböző, $\lambda_1, \lambda_2, \dots, \lambda_n$ sajátértéke, akkor A felírható

$$A = S \cdot \text{diag}[\lambda_1, \lambda_2, \dots, \lambda_n] \cdot S^{-1}$$

alakban, ahol S oszlopaiban sorra az A mátrix $\lambda_1, \lambda_2, \dots, \lambda_n$ sajátértékeihez tartozó sajátvektorai vannak. Az A mátrix ilyen alakú felírását spektrálfelbontásnak nevezzük.

Pl. Adjuk meg a már látott

$$A = \begin{bmatrix} 7 & -2 \\ 4 & 1 \end{bmatrix}$$

mátrix spektrálfelbontását, ha lehet.

Megoldás: Láttuk, hogy A -nak van két különböző sajátértéke: $\lambda_1 = 5$ és $\lambda_2 = 3$. $\Rightarrow A$ -nak létezik spektrálfelbontása. Mindkét sajátértékhez keressünk egy-egy sajátvektort!

1. $\lambda_1 = 5$ -höz: Láttuk, hogy minden (p, p) , $p \neq 0$ vektor sajátvektor. Legyen pl. $\underline{u} := (1, 1)$. (Bármelyik sajátvektort választhatjuk.)

2. $\lambda_2 = 3$ -hoz: Láttuk, hogy minden $(q, 2q)$ vektor sajátvektor. Legyen pl. $\underline{v} := (1, 2)$.

Ezzel a keresett X mátrix:

$$S = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

Számítsuk ki X inverzét, pl. Gauss–Jordan-eliminációval:

$$\begin{aligned} \left[\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 1 & 2 & 0 & 1 \end{array} \right] &\xrightarrow{(2.)-(1.)} \left[\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 0 & 1 & -1 & 1 \end{array} \right] \\ &\xrightarrow{(1.)-(2.)} \left[\begin{array}{cc|cc} 1 & 0 & 2 & -1 \\ 0 & 1 & -1 & 1 \end{array} \right] \end{aligned}$$

Az S inverze tehát:

$$S^{-1} = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$$

Ezzel A spektrálfelbontása:

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \cdot \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix} \cdot \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$$

Később visszatérünk arra a kérdésre, hogy mi mindenre jó a fenti mátrixfelbontás.

2.5. Speciális mátrixok

Egyes speciális tulajdonságú mátrixok a gyakorlatban igen fontosak. Az első félévben szó esett a diagonális, szimmetrikus és antiszimmetrikus mátrixokról. A lineáris egyenlet-rendszerek kapcsán megismerkedhettünk a háromszögmátrixokkal. Most további speciális mátrixfajtákról lesz szó. Megjegyezzük, hogy csak valós elemű mátrixokkal foglalkozunk, azonban a definíciók egy része általánosítható komplex elemű mátrixokra is.

2.18 Definíció. Egy $A \in \mathbb{R}^{n \times n}$ mátrixot ortogonális mátrixnak nevezünk, ha sor- (vagy oszlop-)vektoraik egymásra páronként merőleges egységvektorok, azaz skaláris szorzatuk nulla, és hosszúságuk 1.

Belátható, hogy egy mátrix pontosan akkor ortogonális, ha inverze a transzponáltja, azaz

$$A \in \mathbb{R}^{n \times n} \text{ ortogonális mátrix} \iff A^T = A^{-1}, \text{ azaz } A^T A = A A^T = I.$$

Ebből az is egyszerűen következik, hogy egy ortogonális mátrixnak csak 1 vagy -1 lehet a determinánsa, hiszen a

$$\det A = \det A^T, \quad \det A \cdot \det A^T = \det I = 1$$

feltételekből $(\det A)^2 = 1$ következik, azaz $\det A$ értéke 1 vagy -1.

Példák ortogonális mátrixokra:

- Identitásmátrix:

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

- A ϕ szögű elforgatás mátrixa:

$$\begin{bmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{bmatrix}$$

- Az x tengelyre való tükrözés mátrixa:

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

- Tükrözés a $\phi/2$ irányszögű egyenesre:

$$\begin{bmatrix} \cos \phi & \sin \phi \\ \sin \phi & -\cos \phi \end{bmatrix}$$

Vegyük észre, hogy ez a mátrix szimmetrikus, azaz a transzponáltja saját maga. De mivel ortogonális is, a transzponáltja egyben az inverze, amiből következik, hogy az inverze saját maga. Speciálisan, ha $\phi = 90^\circ$, akkor a 45° -os egyenesre tükrözünk, ennek mátrixa:

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

Az ortogonális mátrixok fontos fajtája a permutációs mátrix.

2.19 Definíció. Egy $A \in \mathbb{R}^{n \times n}$ mátrixot permutációs mátrixnak nevezünk, ha minden oszlopában és sorában pontosan egy elem 1, a többi 0.

Az ilyen mátrixok egy vektor elemeit felcserélik, más szóval permutálják. Pl. az

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

mátrix bármely háromdimenziós vektor második és harmadik elemét cseréli fel. Permutációs mátrix pl. az identitásmátrix és a már előbb látott, 45° -os egyenesre való tükrözés mátrixa is.

2.20 Definíció. Egy $A \in \mathbb{R}^{n \times n}$ mátrixot normális mátrixnak nevezünk, ha A kommutál a transzponáltjával, azaz

$$A^T A = A A^T.$$

Nyilvánvalóan, minden ortogonális mátrix normális. (Fordítva nem igaz.) Igaz továbbá, hogy minden szimmetrikus mátrix is normális, mivel ezekre $A = A^T$, és így az $A A^T = A^T A$ feltétel automatikusan teljesül.

2.21 Definíció. Egy $A \in \mathbb{R}^{n \times n}$ mátrixot nilpotens mátrixnak nevezünk, ha létezik olyan $k \in \mathbb{N}$ hatvány, amelyre $A^k = 0$. A legkisebb ilyen tulajdonságú k számot a nilpotens mátrix fokszámának nevezzük.

Megjegyezzük, hogy ha egy mátrixnál $A^n = 0$, akkor $A^m = 0$ is igaz minden $m > n$ esetén.

Pl. az

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

mátrix nilpotens, ugyanis $A^2 = 0$.

Speciálisan, minden olyan tridiagonális mátrix, amelynek nullák vannak a főátlójában, nilpotens. Olyan nilpotens mátrix is létezik azonban, amelynek nincs egyetlen nulla eleme sem, pl.

$$A = \begin{bmatrix} 6 & -9 \\ 4 & -6 \end{bmatrix},$$

hiszen könnyen meggyőződhetünk róla, hogy ennek a mátrixnak a négyzete is a nullmátrix.

Egy $A \in \mathbb{R}^{n \times n}$ mátrixra egyenértékűek a következő állítások:

1. A nilpotens.
2. $\det(A - \lambda I) = \lambda^n$.
3. A egyetlen sajátértéke a nulla.

A gyakorlatban (pl. szélsőérték-számításhoz) fontosak a pozitív (ill. negatív) definit mátrixok is. Ezek bizonyos szempontból a pozitív (negatív) valós számokhoz hasonló tulajdonságokkal rendelkeznek.

2.22 Definíció. Egy $A \in \mathbb{R}^{n \times n}$ mátrixot pozitív (negatív) definit mátrixnak nevezünk, ha minden $x \in \mathbb{R}^n$ vektorra

$$\langle Ax, x \rangle \geq 0 \quad (\text{ill. } \langle Ax, x \rangle \leq 0),$$

és egyenlőség csak akkor van, ha x a nullvektor. Ha az $\langle Ax, x \rangle$ skaláris szorzat egyes x vektorokra pozitív, másokra negatív, akkor az A mátrixot indefinitnek nevezzük.

A definíció alapján általában nehéz megvizsgálni, hogy egy mátrix pl. pozitív definit mátrix-e. Szimmetrikus mátrixra azonban könnyen ellenőrizhető feltétel adható:

2.23 Tétel. (*Sylvester tétele*)

Az $A \in \mathbb{R}^{n \times n}$ szimmetrikus mátrix pontosan akkor pozitív definit, ha összes bal felső sarokaldeterminánsa pozitív.

A tárgyalt mátrixfajták közül különösen kedvező tulajdonságúak a diagonálmátrixok, hiszen ezeknek könnyű megmondani a determinánsát, sajátértékeit, sajátvektorait, valamint egyszerűen invertálhatók és hatványra emelhetők. Ha egy mátrix nem diagonális, akkor gyakran lehetőség van arra, hogy felírjuk egy diagonális mátrixot tartalmazó mátrixszorzat alakjában.

2.24 Definíció. Egy $A \in \mathbb{R}^{n \times n}$ mátrixot diagonalizálható mátrixnak nevezzük, ha A felírható PDP^{-1} alakban, ahol $P \in \mathbb{R}^{n \times n}$ egy invertálható mátrix, és $D \in \mathbb{R}^{n \times n}$ egy diagonálmátrix.

Korábban szó esett a mátrixok spektrálfelbontásáról. Azt tanultuk, hogy ha egy $A \in \mathbb{R}^{n \times n}$ mátrixnak van n db különböző sajátértéke, akkor létezik $A = SDS^{-1}$ alakú felbontása, ahol D olyan diagonálmátrix, amelynek a főátlójában az A mátrix sajátértékei vannak, az S mátrix oszlopvektorai pedig A egy-egy sajátvektora. (Ezek lineárisan független vektorok, mert különböző sajátértékekhez tartozó sajátvektorok mindig lineárisan függetlenek, ezért biztosan létezik S inverze.) Vagyis ekkor A diagonalizálható, és a definíció szerinti P mátrix oszlopvektorai az A sajátvektorai.

A spektrálfelbontás létezéséhez nem szükséges, hogy létezzen n db különböző sajátérték, hanem elegendő, ha van n db lineárisan független sajátvektora a mátrixnak (tehát lehet többszörös sajátértéke!), ekkor is diagonalizálható a mátrix. Az állítás megfordítása is igaz, azaz fennáll a következő

2.25 Tétel. Egy $A \in \mathbb{R}^{n \times n}$ mátrix pontosan akkor diagonalizálható, ha van n db lineárisan független sajátvektora.

Biz.:

($\Rightarrow$) Tegyük fel, hogy A diagonalizálható. Ekkor létezik olyan $P \in \mathbb{R}^{n \times n}$ invertálható mátrix, amelyre $A = PDP^{-1}$, ahol $D = \text{diag}[d_1, d_2, \dots, d_n]$ diagonális mátrix. Szorozzuk meg az egyenlőséget P -vel jobbról:

$$AP = PD.$$

A két oldalon lévő mátrix megfelelő oszlopvektorai egyenlőek, azaz

$$Ap_i = d_i p_i, \quad i = 1, 2, \dots, n.$$

Ez éppen azt jelenti, hogy P oszlopvektorai az A sajátvektorai a $d_1, d_2, \dots, d_n$ sajátértékekkel, és mivel P invertálható, $\det P \neq 0$, ezért ezen oszlopok lineárisan függetlenek (különben a determináns nulla lenne). Következésképpen van n db lineárisan független sajátvektora A -nak.

($\Leftarrow$) Tegyük fel, hogy A -nak van n db lineárisan független sajátvektora. Jelölje P az ezekből mint oszlopvektorokból összeállított mátrixot. Ekkor

$$AP = PD,$$

ahol D a megfelelő sajátértékeket tartalmazza a főátlóban. Szorozzuk meg az egyenlőséget jobbról P inverzével (amely létezik, mert az oszlopai lineárisan függetlenek, tehát $\det P \neq 0$). Ebből

$$A = PDP^{-1},$$

ami éppen A diagonalizálhatóságát jelenti.

Megjegyezzük, hogy a szimmetrikus mátrixok mind diagonalizálhatók. Sőt, minden normális mátrix diagonalizálható.

2.5.1. Mátrixok Jordan-alakja

Egy négyzetes mátrixnak nem mindig van annyi lineárisan független sajátvektora, amekkora a mérete. Láttuk, hogy ekkor a mátrix nem diagonalizálható. Azonban igaz a következő: minden mátrix felírható olyan mátrixszorzat alakjában, amelyben a középső mátrix majdnem diagonális: a főátlón kívül csak közvetlenül a főátló fölött tartalmazhat nemnulla elemet. Ezt nevezzük a mátrix Jordan-alakjának.

2.26 Tétel. Minden $A \in \mathbb{R}^{n \times n}$ négyzetes mátrix felírható $A = PJP^{-1}$ alakban, ahol $P \in \mathbb{R}^{n \times n}$ invertálható mátrix, és J a következő alakú blokkdiagonális mátrix:

$$J = \begin{bmatrix} J_1 & 0 & \dots & 0 \\ 0 & J_2 & \dots & 0 \\ \vdots & & & \\ 0 & \dots & 0 & J_p \end{bmatrix},$$

ahol

$$J_i = \begin{bmatrix} \lambda_i & 1 & \dots & 0 \\ 0 & \lambda_i & 1 & 0 \\ \vdots & & & 1 \\ 0 & \dots & 0 & \lambda_i \end{bmatrix},$$

ha a mátrix i -edik sajátértékét λ_i -vel jelöljük, a különböző sajátértékek számát pedig p -vel. J_i elnevezése: az A mátrix i -edik Jordan-blokkja.

A Jordan-blokkok mérete megegyezik a sajátérték multiplicitásával.

2.6. Mátrixfüggvények

A mátrixfüggvények olyan függvények, amelyek mátrixhoz mátrixot rendelnek. A valós függvények többsége csak olyan mátrixfüggvényekre általánosítható, amelyek négyzetes mátrixokra vannak értelmezve. Ezek közül tekintjük át a legfontosabbakat.

2.6.1. Mátrixpolinomok

A polinomok a legegyszerűbb és legkönnyebben kezelhető függvények. Tekintsük a

$$p(x) = a_k x^k + a_{k-1} x^{k-1} + \dots a_1 x + a_0$$

valós polinomot. Megfogalmazzuk, mit értünk a p polinom A helyen vett helyettesítési értékén, ahol $A \in \mathbb{R}^{n \times n}$. Tudjuk, hogy az $n \times n$ -es mátrixok egész kitevős hatványra emelhetők, és az eredmény szintén $n \times n$ -es mátrix lesz. Ezt a mátrixot tetszőleges valós számmal szorozhatjuk, és az eredményt összeadhatjuk egy másik $n \times n$ -es mátrixszal. Így tetszőleges négyzetes mátrixot behelyettesíthetünk a p polinomban az x helyébe. Az a_0 konstans $a_0 I$ -vel helyettesítjük, ahol I az $n \times n$ -es identitásmátrix:

$$p(A) := a_k A^k + a_{k-1} A^{k-1} + \dots a_1 A + a_0 I.$$

Speciálisan, ha a mátrix diagonális, akkor könnyen belátható, hogy a főátlóban álló számokat kell csak behelyettesíteni a polinomba, azaz a $D = \text{diag}[d_1, d_2, \dots, d_n]$ mátrixra

$$p(D) = \text{diag}[p(d_1), p(d_2), \dots, p(d_n)].$$

Ha a mátrix diagonalizálható, azaz $A = PDP^{-1}$ alakú, akkor szintén egyszerű $p(A)$ kiszámítása:

$$p(A) = Pp(D)P^{-1},$$

ahol a $p(D)$ mátrix a p polinom a D diagonálmátrixra alkalmazva, vagyis a D főátlójában lévő számokat kell csak behelyettesíteni a polinomba. (Számokat behelyettesíteni a polinomba lényegesen egyszerűbb, mint mátrixokat.) Láttuk, hogy ha egy mátrix nem diagonalizálható, akkor is mindig Jordan-alakra hozható ($A = PJP^{-1}$). Ekkor is igaz a fenti összefüggés ($p(A) = Pp(J)P^{-1}$), csak ekkor a középső mátrixban a főátló fölött közvetlenül is lehetnek nemnulla elemek, azonban az ilyen mátrix is viszonylag könnyen behelyettesíthető bármely polinomba.

2.6.2. Mátrixok hatványsorai

A mátrixokat behelyettesíthetjük hatványsorba is. A valós függvények között különösen fontos szerepet tölt be az exponenciális függvény, amelyet az

$$\exp(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

hatványsorral definiálunk. Hatványsorral definiáltuk továbbá a sin és cos függvényt is.

Tekintsük az $f(x) = \sum_{k=0}^{\infty} a_k x^k$ hatványsorral megadott függvényt. Ha $A \in \mathbb{R}^{n \times n}$, akkor az A mátrix hatványsora, $f(A)$ az A mátrixnak az $f(x)$ sorba való behelyettesítését jelenti. Fontos hangsúlyozni, hogy ez nem azt jelenti, hogy a mátrix minden egyes elemét behelyettesítjük a hatványsorba!

Speciális mátrixok hatványsora könnyen kiszámítható. A diagonálmátrixokban csak a főátlóbeli elemeket kell behelyettesíteni a sorba. Ha a mátrix diagonalizálható ($A = PDP^{-1}$), akkor

$$f(A) = Pf(D)P^{-1},$$

ha pedig nem diagonalizálható, akkor Jordan-alakra hozható ($A = PJP^{-1}$), és

$$f(A) = Pf(J)P^{-1}.$$

A mátrix hatványsora nilpotens mátrixokra pontosan kiszámítható, mert egy n -edfokú nilpotens mátrix összes $m > n$ kitevűjű hatványa nullmátrix lesz. Így a végtelen mátrixsor véges polinommal egyszerűsödik.

2.27 Megjegyzés. Könnyen látható, hogy ha A diagonalizálható, akkor az $f(A)$ mátrix sajátértékeit úgy kapjuk, hogy A sajátértékeit behelyettesítjük f -be. Ugyanis ekkor $f(A) = Pf(D)P^{-1}$, azaz $f(A)$ diagonalizálható az $f(D)$ diagonálmátrixszal. Az $f(A)$ sajátértékei tehát az $f(D)$ mátrix főátlójában lévő számok, ezek pedig éppen az $f(\lambda_1), \dots$,

$f(\lambda_n)$ számok, ahol $\lambda_1, \dots, \lambda_n$ az A mátrix sajátértékei.

2.7. Koordináta-transzformációk

Egy vektortérben számtalan módon választhatunk meg bázist a vektorok szám-n-esekkel való reprezentálására. Érdekes tudnunk, hogy az egyik bázisból (más szóval koordináta-rendszerből) milyen képletek segítségével térhetünk át egy másikba.

Legyen adva egy X n -dimenziós vektortérben két bázis: $B_1 = \{a_1, a_2, \dots, a_n\}$ és $B_2 = \{b_1, b_2, \dots, b_n\}$. Azt keressük, hogy milyen összefüggés van egy X -beli vektor B_1 és B_2 bázisbeli koordinátái között. Ezt a függvénykapcsolatot koordináta-transzformációnak nevezzük.

Tegyük fel, hogy az $x \in X$ vektort a B_1 bázisban az $(\alpha_1, \alpha_2, \dots, \alpha_n)$, a B_2 -ben pedig a $(\beta_1, \beta_2, \dots, \beta_n)$ szám-n-essel azonosíthatjuk, azaz

$$x = \alpha_1 a_1 + \alpha_2 a_2 + \dots + \alpha_n a_n = \beta_1 b_1 + \beta_2 b_2 + \dots + \beta_n b_n. \quad (5)$$

A bal oldalon az $a_1, a_2, \dots, a_n$ báziselemeket kifejezhetjük a $b_1, b_2, \dots, b_n$ báziselemek lineáris kombinációjaként:

$$\begin{aligned} a_1 &= \gamma_{11} b_1 + \gamma_{12} b_2 + \dots + \gamma_{1n} b_n \\ a_2 &= \gamma_{21} b_1 + \gamma_{22} b_2 + \dots + \gamma_{2n} b_n \\ &\vdots \\ a_n &= \gamma_{n1} b_1 + \gamma_{n2} b_2 + \dots + \gamma_{nn} b_n \end{aligned}$$

Ezeket az összegeket beírva (5)-ben $a_1, a_2, \dots, a_n$ helyére, mindkét oldalon a $b_1, b_2, \dots, b_n$ vektorok lineáris kombinációja fog szerepelni. Az együtthatók egyeztetéséből az

$$\begin{aligned} \alpha_1 \gamma_{11} + \alpha_2 \gamma_{21} + \dots + \alpha_n \gamma_{n1} &= \beta_1 \\ \alpha_1 \gamma_{12} + \alpha_2 \gamma_{22} + \dots + \alpha_n \gamma_{n2} &= \beta_2 \\ &\vdots \\ \alpha_1 \gamma_{1n} + \alpha_2 \gamma_{2n} + \dots + \alpha_n \gamma_{nn} &= \beta_n \end{aligned}$$

összefüggésekhez jutunk. Látható, hogy az új koordináták lineáris függvényei a régieknek, azaz az $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$, $\beta = (\beta_1, \beta_2, \dots, \beta_n)$ jelöléssel

$$K\alpha = \beta,$$

ahol az $K \in \mathbb{R}^{n \times n}$ mátrix elemei leolvashatók az előbbi egyenletekből:

$$K = \begin{bmatrix} \gamma_{11} & \gamma_{21} & \cdots & \gamma_{n1} \\ \gamma_{12} & \gamma_{22} & \cdots & \gamma_{n2} \\ \vdots & & & \vdots \\ \gamma_{1n} & \gamma_{2n} & \cdots & \gamma_{nn} \end{bmatrix}$$

Vagyis az K mátrix i -edik oszlopába az a_i vektor B_2 -beli koordinátái kerülnek. Ezt a mátrixot a koordináta-transzformáció mátrixának nevezzük.

A gyakorlatban azok a koordináta-transzformációk a legfontosabbak, amelyek megőrzik a vektorok hosszát, azaz amelyekre igaz az, hogy bármely $x \in X$ vektor B_1 és B_2 bázisbeli koordináta-vektora között fennáll a

$$\sqrt{\beta_1^2 + \beta_2^2 + \dots + \beta_n^2} = \sqrt{\alpha_1^2 + \alpha_2^2 + \dots + \alpha_n^2}$$

egyenlőség. Másképpen, ismét az $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$ jelölést használva

$$\langle K\alpha, K\alpha \rangle = \langle \alpha, \alpha \rangle,$$

ahol a zárójel skaláris szorzást jelöl. Ezt az egyenlőséget a mátrix transzponáltjával felírva:

$$\langle \alpha, K^T K \alpha \rangle = \langle \alpha, \alpha \rangle.$$

Belátható, hogy ez minden α vektorra pontosan akkor áll fenn, ha $K^T K = I$, azaz $K^T = K^{-1}$, vagyis a koordináta-transzformáció mátrixa ortogonális. Ezért az ortogonális koordináta-transzformációk a legfontosabbak a gyakorlatban. (Ilyen pl. a síkvektorok esetében, ha mindkét bázisvektort elforgatjuk ugyanazzal a szöggel.)

3. Vektorterek további struktúrákkal

Ebben a fejezetben olyan vektorterekről lesz szó, amelyekben az elemek bizonyos speciális tulajdonságokkal rendelkeznek.

3.1. A vektorok normája, a normált tér

A síkvektorok fontos tulajdonsága a hosszúságuk (nagyságuk). Egy tetszőleges vektortérben nem biztos, hogy van hosszuk az elemeknek, de a gyakorlatban azok a vektorterek a legfontosabbak, amelyekben értelmes ez a fogalom. Az ilyen vektortereket normált terek-

nek fogjuk nevezni.

Ahhoz, hogy egy tetszőleges vektortérben értelmezhezzük az elemek hosszúságát, gondoljuk meg, hogy milyen tulajdonságokkal rendelkezik a síkvektorok hagyományos értelemben vett (Pitagorasz-tétellel számított) hosszúsága (euklideszi hosszúság). Jelöljük ehhez egy $\underline{a} = (a_1, a_2)$ vektor hosszúságát $|\underline{a}|$ -val.

1. A hosszúság mindig nemnegatív, és csak akkor 0, ha a nullvektorról van szó, azaz

$$|\underline{a}| \geq 0 \quad \forall \underline{a} \in \mathcal{E}_2, \text{ és } |\underline{a}| = 0 \Leftrightarrow \underline{a} = \underline{0}.$$

2. Két vektor összege nem lehet hosszabb, mint az összeadandók hosszának az összege, azaz

$$|\underline{a} + \underline{b}| \leq |\underline{a}| + |\underline{b}| \quad \forall \underline{a}, \underline{b} \in \mathcal{E}_2.$$

3. Ha egy vektort $\lambda \in \mathbb{R}$ számmal szorzunk, akkor a vektor hossza $|\lambda|$ -szorosára változik, tehát

$$|\lambda \cdot \underline{a}| = |\lambda| \cdot |\underline{a}| \quad \forall \underline{a} \in \mathcal{E}_2, \forall \lambda \in \mathbb{R}.$$

Úgy gondoljuk, hogy ezek a tulajdonságok jelentik a hosszúság lényegét. Felmerül a kérdés, hogy csak az euklideszi hosszúság rendelkezik-e ezekkel a tulajdonságokkal, vagy más módon is lehetne-e olyan számot a vektorokhoz rendelni, amely hasonló tulajdonságú, és ezért szintén használható lenne a vektorok hosszúságának a jellemzésére. Próbálkozzunk egyéb skalárértékű függvényekkel, és vizsgáljuk meg a fenti három tulajdonságot.

i.) A vektorok irányszöge - nem tesz eleget a fenti tulajdonságoknak, hiszen pl. nem igaz az, hogy csak a nullvektornak nulla az irányszöge. Ez tehát nem használható a hosszúság jellemzésére.

ii.) A vektorokhoz hozzárendelhetjük az x és y tengelyre eső vetületük közül a hosszabbik vetület hosszúságát.

$$\underline{a} \mapsto \max\{|a_1|, |a_2|\}$$

Ez már jó lesz!

iii.) A két vetület hosszúságát összeadva szintén nemnegatív számot kapunk:

$$\underline{a} \mapsto |a_1| + |a_2|$$

Szintén mindhárom tulajdonság teljesül.

Ezért a továbbiakban a hossz fogalmát az előbbiek alapján a következőképpen általánosíthatjuk.

3.1 Definíció. Legyen X vektortér. Ekkor egy $\|\cdot\| : X \rightarrow \mathbb{R}$ függvényt normának nevezzük, ha igaz rá a következő három tulajdonság:

1. $\|x\| \geq 0 \quad \forall x \in X$, és $\|x\| = 0 \Leftrightarrow x = 0_X$.
2. $\|x + y\| \leq \|x\| + \|y\| \quad \forall x, y \in X$.
3. $\|\lambda \cdot x\| = |\lambda| \cdot \|x\| \quad \forall x \in X, \forall \lambda \in \mathbb{R}$.

Azt a számot, amelyet ez a függvény egy $x \in X$ vektorhoz hozzárendel, az x vektor normájának nevezzük, és $\|x\|$ -vel jelöljük. Az $(X, \|\cdot\|)$ rendezett párt normált térnek nevezzük.

Az $\mathcal{E}_2$ vektortéren tehát már három normát is ismerünk. Ezekre a következő jelölések és elnevezések használatosak:

1. $\|\underline{a}\|_1 := |a_1| + |a_2|$ (első norma)
2. $\|\underline{a}\|_2 := \sqrt{a_1^2 + a_2^2}$ (második vagy euklideszi norma)
3. $\|\underline{a}\|_{\max}$ vagy $\|\underline{a}\|_{\infty} := \max\{|a_1|, |a_2|\}$ (maximumnorma)

Ugyanezen háromfajta norma megfelelője $\mathbb{R}^n$ -en is értelmezhető, mindhármát gyakran használjuk. Az $x = (x_1, x_2, \dots, x_n)$ jelöléssel

1. $\|x\|_1 := \sum_{i=1}^n |x_i|$
2. $\|x\|_2 := \sqrt{\sum_{i=1}^n x_i^2}$
3. $\|x\|_{\max}$ vagy $\|x\|_{\infty} := \max\{|x_1|, |x_2|, \dots, |x_n|\}$

3.2. Operátornormák, mátrixnormák

Láttuk, hogy $\text{Hom}(\mathbb{R}^n, \mathbb{R}^m)$, azaz az $\mathbb{R}^n \rightarrow \mathbb{R}^m$ lineáris leképezések halmaza a megfelelő műveletekkel vektortér. Lehetne-e az $X = \text{Hom}(\mathbb{R}^n, \mathbb{R}^m)$ vektortérben is értelmezni az elemek nagyságát?

A legfontosabb speciális eset az, amikor $m = n$ ($\text{Hom}(\mathbb{R}^n, \mathbb{R}^n)$, amit röviden csak a $\text{Hom}(\mathbb{R}^n)$ szimbólummal jelölünk), ezért részletesebben csak ezzel foglalkozunk. Jelölje $\|\cdot\|_{\mathbb{R}^n}$ bármelyik $\mathbb{R}^n$ -beli vektornormát, és 0_n az $\mathbb{R}^n$ nullvektorát. Belátható, hogy

$$\|f\| := \sup_{x \in \mathbb{R}^n, x \neq 0_n} \frac{\|f(x)\|_{\mathbb{R}^n}}{\|x\|_{\mathbb{R}^n}}$$

norma $Hom(\mathbb{R}^n)$ -en. Ezt a normát az $\|\cdot\|_{\mathbb{R}^n}$ vektornorma által indukált operátornormának nevezzük.

Mit fejez ez ki szemléletesen? Az $\frac{\|f(x)\|_{\mathbb{R}^n}}{\|x\|_{\mathbb{R}^n}}$ hányados azt mutatja meg, hogy az f leképezés hányszorosára nyújtja meg az $x \in \mathbb{R}^n$ elemet. Ha az összes x -re vesszük ezt a hányadost, majd ezen hányadosok szupremumát, akkor ez azt mutatja meg, hogy az f leképezés legfeljebb hányszorosára nyújt meg egy elemet.

Belátható, hogy a fenti szuprémum ekvivalens a következőkkel:

$$\begin{aligned} \sup_{x \in \mathbb{R}^n, x \neq 0_n} \frac{\|f(x)\|_{\mathbb{R}^n}}{\|x\|_{\mathbb{R}^n}} &= \sup_{x \in \mathbb{R}^n, \|x\|_{\mathbb{R}^n}=1} \frac{\|f(x)\|_{\mathbb{R}^n}}{\|x\|_{\mathbb{R}^n}} = \sup_{x \in \mathbb{R}^n, \|x\|_{\mathbb{R}^n}=1} \|f(x)\|_{\mathbb{R}^n} = \\ &= \max_{x \in \mathbb{R}^n, \|x\|_{\mathbb{R}^n}=1} \|f(x)\|_{\mathbb{R}^n} \end{aligned}$$

Megjegyezzük, hogy más-más $\mathbb{R}^n$ -beli vektornorma más operátornormát indukál, tehát egy lineáris operátornak nem feltétlenül ugyanannyi pl. a maximumnorma és az euklideszi norma által indukált normája.

Megjegyezzük, hogy a $Hom(\mathbb{R}^n, \mathbb{R}^m)$ vektortéren is értelmezhető a szupremumnorma: ekkor a számlálóban és a nevezőben különböző vektortérbeli normák szerepelnek.

Tudjuk, hogy az $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ lineáris operátor megfelel egy $A \in \mathbb{R}^{n \times n}$ mátrixszal való szorzásnak, azaz $f(x) = A \cdot x$, tehát a sup-norma az A mátrixszal is kifejezhető:

$$\|f\| = \sup_{x \in \mathbb{R}^n, x \neq 0_n} \frac{\|A \cdot x\|_{\mathbb{R}^n}}{\|x\|_{\mathbb{R}^n}}.$$

A szupremumot nem mindig könnyű kiszámítani, ezért érdemes tudnunk, hogy az indukált operátornormákat a leképezés mátrixának elemiből egyszerűbben is ki lehet számolni:

1. Első norma (másképpen oszlopnorma): a leképezés mátrixának minden oszlopában összeadjuk az elemet abszolút értékét, és a legnagyobb ilyen oszlopösszeget választjuk.

$$\|A\|_{\max} := \max_{j \in \{1, 2, \dots, n\}} \sum_{i=1}^n |a_{ij}|$$

2. Második norma (spektrálnorma):

$$\|A\|_2 := \sqrt{|\lambda_{\max}(A^T A)|},$$

ahol A^T az A mátrix transzponáltja, $\lambda_{\max}(A^T A)$ pedig az $A^T A$ mátrix legnagyobb

abszolút értékű sajátértéke.

3. Maximumnorma (sornorma): a legnagyobb sorösszeg.

$$\|A\|_1 := \max_{i \in \{1, 2, \dots, n\}} \sum_{j=1}^n |a_{ij}|$$

Láttuk, hogy az $\mathbb{R}^{n \times n}$ halmaz is vektortér a mátrixokon értelmezett összeadás és skalárral való szorzás műveletével, és ez a vektortér azonosítható a $\text{Hom}(\mathbb{R}^n)$ vektortérrel. A fenti szupremum tehát $\mathbb{R}^{n \times n}$ -en is norma: (az adott $\mathbb{R}^n$ -beli norma által) indukált mátrixnormának nevezzük. Az indukált mátrixnormák rendelkeznek a következő három fontos tulajdonsággal:

- i) Az Ax mátrix-vektor szorzatra mindig igaz, hogy $\|Ax\| \leq \|A\| \cdot \|x\|$ (itt a bal oldali és a jobb oldali második norma valamelyik vektornorma, és az A mátrix normája az ezen vektornorma által indukált mátrixnorma!)
- ii) Az identitásmátrix normája 1, azaz $\|I\| = 1$.
- iii) Két mátrix szorzatának normája nem nagyobb, mint a tényezők normájának a szorzata: $\|AB\| \leq \|A\| \cdot \|B\|$.

Ezen a vektortéren azonban további, nem indukált normák is használatosak. Ezek a mátrixnormák nem írhatók fel egy adott vektornormához tartozó szupremumnormaként. Nem indukált mátrixnormát kapunk pl., ha vesszük a mátrix lehető legnagyobb abszolút értékű elemét:

$$\|A\| := \max_{i,j} |a_{ij}|;$$

vagy ha az összes mátrixelem abszolút értékét összeadjuk:

$$\|A\| := \sum_{i,j=1}^n |a_{ij}|.$$

Ezek tehát mind rendelkeznek a norma három tulajdonságával, de fontos tudni, hogy általában nem rendelkeznek az i)-iii) tulajdonságokkal. Ez az oka, hogy a nem indukált mátrixnormákat ritkán használjuk.

3.3. Az euklideszi tér

Ismeretes, hogy a síkvektorok körben nemcsak az összeadás és a skalárral való szorzás művelete van értelmezve, hanem egy további igen fontos művelet is: a skaláris szorzás.

Ez a művelet azonban nem síkvektort ad eredményül, hanem az

$$\underline{a} \cdot \underline{b} = |\underline{a}| \cdot |\underline{b}| \cdot \cos \gamma$$

definíció szerint (γ a két vektor közrezárt szöge) egy valós számot. Ez a művelet tehát egy $\mathcal{E} \times \mathcal{E} \rightarrow \mathbb{R}$ leképezést jelent.

Idézzük fel a skaláris szorzás műveleti tulajdonságait!

1. $\underline{a} \cdot \underline{b} = \underline{b} \cdot \underline{a} \quad \forall \underline{a}, \underline{b} \in \mathcal{E}_2;$
2. $(\underline{a} + \underline{b}) \cdot \underline{c} = \underline{a} \cdot \underline{c} + \underline{b} \cdot \underline{c} \quad \forall \underline{a}, \underline{b}, \underline{c} \in \mathcal{E}_2$
3. $(\alpha \underline{a}) \cdot \underline{b} = \alpha \cdot (\underline{a} \cdot \underline{b}) \quad \forall \underline{a}, \underline{b} \in \mathcal{E}_2, \forall \alpha \in \mathbb{R}.$

Könnyen meggondolható, hogy a 2. és 3. tulajdonság ekvivalens a következő tulajdonsággal:

$$(\alpha \underline{a} + \beta \underline{b}) \cdot \underline{c} = \alpha \cdot (\underline{a} \cdot \underline{c}) + \beta \cdot (\underline{b} \cdot \underline{c}) \quad \forall \alpha, \beta \in \mathbb{R}, \forall \underline{a}, \underline{b}, \underline{c} \in \mathcal{E}_2.$$

Azt is megállapítottuk továbbá, hogy $\underline{a} \cdot \underline{a} \geq 0$, és egyenlőség akkor és csak akkor van, ha $\underline{a} = \underline{0}$, ugyanis $\underline{a} \cdot \underline{a} = |\underline{a}| \cdot |\underline{a}| \cdot \cos 0 = |\underline{a}|^2$.

A skaláris szorzat használható például arra, hogy segítségével kiszámítsuk két vektor közbezárt szögét. Speciálisan, azt is könnyen megvizsgálhatjuk, hogy merőleges-e egymásra a két vektor, ennek feltétele ugyanis az $\underline{a} \cdot \underline{b} = 0$ egyenlőség.

A továbbiakban szeretnénk a skaláris szorzás fogalmát értelmezni más vektorterekben is.

3.2 Definíció. Legyen X egy tetszőleges vektortér, és legyen értelmezve egy $X \times X \rightarrow \mathbb{R}$ függvény a következő tulajdonságokkal (ezt a függvényt jelölje $\langle \cdot, \cdot \rangle$):

1. $\langle x, y \rangle = \langle y, x \rangle \quad \forall x, y \in X;$
2. $\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle \quad \forall x, y, z \in X, \forall \alpha, \beta \in \mathbb{R};$
3. $\langle x, x \rangle \geq 0 \quad \forall x \in X$ és $\langle x, x \rangle = 0 \Leftrightarrow x = 0_X$.

Ezen $X \times X \rightarrow \mathbb{R}$ függvényt skaláris szorzatnak nevezzük, és az $(X, \langle \cdot, \cdot \rangle)$ rendezett párt (valós) euklideszi térnek nevezzük.

3.3 Megjegyzés. A fenti három műveleti tulajdonság néhány következménye:

- A 2. tulajdonság akkor is igaz, ha az összeg hátul áll, azaz:

$$\langle x, \alpha y + \beta z \rangle = \alpha \langle x, y \rangle + \beta \langle x, z \rangle,$$

ugyanis, felhasználva az 1. és 2. tulajdonságot,

$$\langle x, \alpha y + \beta z \rangle = \langle \alpha y + \beta z, x \rangle = \alpha \langle y, x \rangle + \beta \langle z, x \rangle = \alpha \langle x, y \rangle + \beta \langle x, z \rangle.$$

- $\langle 0_X, y \rangle = 0 \ \forall y \in X$ és $\langle x, 0_X \rangle = 0 \ \forall x \in X$. (Pl. az első állítás igazolása: $\langle 0_X, y \rangle = \langle x - x, y \rangle = \langle x, y \rangle + \langle -x, y \rangle = \langle x, y \rangle + \langle (-1)x, y \rangle = \langle x, y \rangle - 1 \cdot \langle x, y \rangle = 0$.)

Példák euklideszi terekre:

- $X := \mathbb{R}^n, \quad \langle x, y \rangle := \sum_{i=1}^n x_i y_i$.

Lássuk be, hogy ez valóban skaláris szorzat!

1. $\langle x, y \rangle = \sum_{i=1}^n x_i y_i = \sum_{i=1}^n y_i x_i = \langle y, x \rangle$
2. $\langle \alpha x + \beta y, z \rangle = \sum_{i=1}^n (\alpha x_i + \beta y_i) z_i = \alpha \sum_{i=1}^n x_i z_i + \beta \sum_{i=1}^n y_i z_i = \alpha \langle x, z \rangle + \beta \langle y, z \rangle$
3. $\langle x, x \rangle = \sum_{i=1}^n x_i^2 \geq 0$, és $= 0 \Leftrightarrow x_i = 0 \ \forall i = 1, 2, \dots, n$, azaz $x = 0_n$.

- $X := C[a, b], \quad \langle f, g \rangle := \int_a^b f \cdot g$.

3.4. Norma euklideszi terekben

Legyen $(X, \langle \cdot, \cdot \rangle)$ egy euklideszi tér. Ekkor az elemeknek mindig értelmezhető a normája a következőképpen:

$$x \in X, \quad \|x\| := \sqrt{\langle x, x \rangle}.$$

Vegyük észre, hogy ez a norma $\mathbb{R}^n$ -ben a második normát adja (ezt neveztük euklideszi normának). A $C[a, b]$ függvénytérben a

$$\|f\| := \sqrt{\int_a^b f^2}$$

normát kapjuk.

A skaláris szorzat és a norma értelmezése után értelmezhetjük egy tetszőleges euklideszi térben a közbezárt szög fogalmát is:

3.4 Definíció. Egy $(X, \langle \cdot, \cdot \rangle)$ euklideszi térben az $x \in X$ és $y \in X$ vektorok közbezárt szögének azt a γ szöget nevezzük, amelyre

$$\cos \gamma = \frac{\langle x, y \rangle}{\|x\| \cdot \|y\|}.$$

3.5 Definíció. Azt mondjuk, hogy az $(X, \langle \cdot, \cdot \rangle)$ euklideszi térben az $x \in X$ és $y \in X$ vektorok merőlegesek egymásra, ha $\langle x, y \rangle = 0$.

3.5. A Cauchy–Schwarz–Bunyakovszkij-egyenlőtlenség

Az euklideszi terekben fennáll egy fontos egyenlőtlenség, az ún. Cauchy–Schwarz–Bunyakovszkij-egyenlőtlenség:

3.6 Tétel. Egy $(X, \langle \cdot, \cdot \rangle)$ euklideszi térben bármely $a, b \in X$ elemekre fennáll:

$$\langle a, b \rangle^2 \leq \langle a, a \rangle \cdot \langle b, b \rangle.$$

Ezt bizonyítás helyett csak a következő két speciális esetre gondoljuk meg:

1. $X = \mathcal{E}_2$ -ben a CSB-egyenlőtlenség:

$$(\underline{a} \cdot \underline{b})^2 \leq |\underline{a}|^2 \cdot |\underline{b}|^2 \quad \forall \underline{a}, \underline{b} \in \mathcal{E}_2.$$

Biz.:

$$(\underline{a} \cdot \underline{b})^2 = (|\underline{a}| \cdot |\underline{b}| \cdot \cos \gamma)^2 = |\underline{a}|^2 \cdot |\underline{b}|^2 \cos^2 \gamma \leq |\underline{a}|^2 \cdot |\underline{b}|^2,$$

ugyanis $\cos^2 \gamma \in [0, 1]$.

Megjegyzés: Látható, hogy egyenlőség pontosan akkor van, ha $\cos \gamma = 1$ vagy -1 , azaz a két vektor egy egyenesbe esik (lineárisan összefüggő).

2. $X = \mathbb{R}^n$ -ben a CSB-egyenlőtlenség:

$$\left(\sum_{i=1}^n a_i b_i \right)^2 \leq \left(\sum_{i=1}^n a_i^2 \right) \left(\sum_{i=1}^n b_i^2 \right) \quad \forall a, b \in \mathbb{R}^n.$$

Biz.: Defináljuk a következő, x -től függő függvényt:

$$y_i(x) := (a_i x + b_i)^2,$$

és legyen

$$Y(x) := \sum_{i=1}^n y_i(x) = \sum_{i=1}^n (a_i x + b_i)^2.$$

A négyzetösszeget kifejtve

$$Y(x) = \sum_{i=1}^n (a_i^2 x^2 + 2a_i b_i x + b_i^2),$$

amely az

$$Y(x) = Ax^2 + 2Bx + C$$

alakban is felírható, ahol

$$A = \sum_{i=1}^n a_i^2, \quad B = \sum_{i=1}^n a_i b_i \quad \text{és} \quad C = \sum_{i=1}^n b_i^2.$$

Nyilvánvalóan, $Y(x) \geq 0$ minden $x \in \mathbb{R}$ pontban, mivel $y_i(x) \geq 0 \quad \forall i = 1, \dots, n$. Az

$$Ax^2 + 2Bx + C \geq 0$$

egyenlőtlenség azonban minden $x \in \mathbb{R}$ pontban csak akkor állhat fenn, ha a másodfokú egyenlet $D = 4B^2 - 4AC$ diszkriminánsa nem pozitív, azaz

$$4B^2 - 4AC \leq 0.$$

Ez pedig (4-gyel osztva) éppen a belátandó egyenlőtlenséget jelenti.