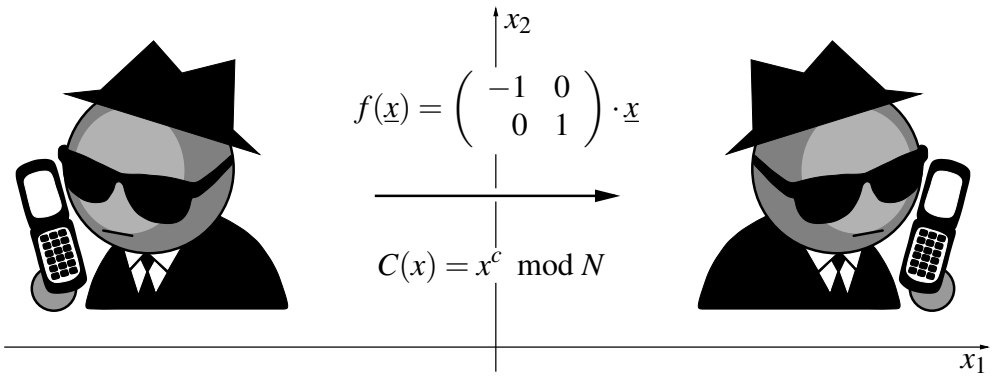
Szeszlér Dávid

BEVEZETÉS A SZÁMÍTÁSELMÉLETBE 1



Készült a

Budapesti Műszaki és Gazdaságtudományi Egyetem
Villamosmérnöki és Informatikai Kar
Számítástudományi és Információelméleti Tanszék

gondozásában.

Lektorálta: Wiener Gábor

Utolsó módosítás: 2018. szeptember 12.

Előszó

Ez a jegyzet a BME Villamosmérnöki és Informatika Karán a mérnök informatikus alapképzés első félévében oktatott *Bevezetés a számításelméletbe 1* című tárgy hallgatói számára készült.

A két féléves Bevezetés a számításelméletbe (avagy BSz) tárgy matematikai alapozótárgy; a feladata az, hogy bevezesse a mérnök informatikus alapszakra járó hallgatókat a matematikának néhány olyan területébe, amelyeket a képzés későbbi tárgyai használni fognak és amelyek nem tartoznak a matematika analízis nevű klasszikus ágába (ezekkel ugyanis az azonos nevű tárgy foglalkozik). Így a BSz anyaga a tárgy jellegéből fakadóan több, egymástól többé-kevésbé elszigetelt fejezetre oszlik, amelyek közül az első félévben kettő kerül terítékre: a lineáris algebra és a számelmélet. Ezekkel foglalkozik ez a jegyzet is.

A jegyzet anyaga

A jegyzet írásakor arra törekedtünk, hogy az minél pontosabban kövesse a BSz1 előadásokon elhangzó és a vizsgára elsajátítandó anyagot. Ennek ellenére, kisebb-nagyobb és előre ki nem számítható eltérések elkerülhetetlenül lesznek az órák és a jegyzet között. (Ezekben az esetekben természetesen az órákon elhangzó anyag – illetve a félév végén a tárgy honlapján közzétett vizsgatételsor – a mérvadó.)

Helyenként azonban a jegyzet szándékosan túlmegy az előadások várható anyagán – ezeket a részeket apró betűs szedéssel különböztettük meg. Ezekben szerepelnek olyan bizonyítások, amelyek az órákról valószínűleg idő hiányában kiszorulnak, néhányszor pedig röviden leírunk olyan alkalmazásokat, amelyek elsősorban a tárgy iránt különösen érdeklődők kíváncsiságát hivatottak kielégíteni. Szintén nekik szól a jegyzet utolsó fejezete, amelynek témája, a végtelen halmazok számossága, már nem része a tárgy anyagának.

A jegyzet mindhárom fejezete a matematika egy kiterjedt területébe kíván bevezetést nyújtani. Mindhárom anyagrészhez bőségesen találhatók (még magyar nyelven is) olyan tankönyvek, amelyek a szóban forgó területet jóval mélyebben tárgyalják – a fejezetek végén ajánlunk is ezek közül néhányat. Ne várja tehát senki ettől a jegyzettől az egyes témák átfogó megalapozását, de még az azokhoz tartozó legalapvetőbb módszerek és eredmények ismertetését sem. Éppen ellenkezőleg: csupán a BSz1 tárgy keretei közé elférő anyagrészek önmagában koherens, szakmailag korrekt felépítése volt a cél.

Hibák, javítások

Mint minden tankönyv, ez is biztosan bőven tartalmaz kisebb-nagyobb hibákat. Lehetnek ezek akár apró elírások, de az is előfordulhat, hogy egy-egy ponton félreérthető, vagy nehezen követhető megfogalmazások szerepelnek. Előre is hálásan köszönjük minden olvasónak, ha egy ilyen hibát jelez a `szeszler@cs.bme.hu` email címen.

A jegyzet ezen jelzések, valamint a tárgyat oktató kollégák tapasztalata és véleménye nyomán folyamatosan változik (remélhetőleg az előnyére). A legfrissebb verzió mindig elérhető a `http://cs.bme.hu/bsz1/jegyzet/` oldalon.

Hogyan használjuk a jegyzetet?

Egy sokat idézett anekdota szerint, amikor I. Ptolemaiosz király megkérdezte Euklideszt, a kor nagy hírű matematikusát és az Elemek című alapmű szerzőjét, hogy vajon nem lehetne-e a matematikát az Elemekben írtaknál könnyebben elsajátítani, Euklidesz ezt válaszolta: „A geometriához nem vezet királyi út.”

Euklidesz kétségkívül a világtörténelem legjelentősebb matematikai tankönyvírója: az Elemeket még a XIX. században, a megjelenése után bő 2100 évvel is használták a nagy európai egyetemeken. Bár a BSz1 jegyzettel kapcsolatos ambícióink ennék sokkal szerényebbek, a neves kolléga tanácsát mégis minden olvasó számára megszívlelendőnek tartjuk. Mai megfogalmazásban ez a következő: a matematikát nem lehet passzívan befogadni, ez csakis a tanuló aktív és kreatív közreműködésével lehetséges.

Arra bátorítunk minden olvasót, hogy semmit ne higgyen el, amit ebben a jegyzetben olvas: győződjön meg róla saját maga! A definíciókat, tételeket, algoritmusokat próbálja ki különféle példákon. Vigye végig, tesztelje ezeken a példákon az oda tartozó bizonyításokat, gondolatmeneteket is!

A jegyzet számos feladatot is tartalmaz, megoldásokkal együtt. Bár ezek a BSz1 zárthelyi dolgozatokra való felkészülésben is segíthetnek és hasonlók a BSz1 gyakorlatokon is előkerülhetnek, a fő funkciójuk itt mégis más. A jegyzetben szereplő feladatok sosem igényelnek ravasz gondolatokat, a tárgyalt anyag ötletes alkalmazását; nem céljuk a BSz1 gyakorlatokon elvégzendő munka kiváltása. Ehelyett a szerepük legtöbbször az, hogy bemutassák, testközelbe hozzák egy fogalom definícióját, egy tétel állítását, vagy egy bizonyítás gondolatmenetét egy konkrét példán keresztül. Így ezek szerves részét képezik a tárgyalt anyagnak; ha az olvasó egy-egy feladat megoldásának elolvasása után úgy érzi, hogy egy ahhoz nagyon hasonlót önállóan még nem tudna megoldani, akkor ezt bátran tekintse arra utaló jelzésnek, hogy a vonatkozó anyagrész megtanulásával kapcsolatban még van tennivalója.

A fent említett anekdotát lejegyző Proklosz szerint Euklidesz így zárta a királynak adott válaszát: „Munka nélkül nincs kenyér, sem geometria.”

Tartalomjegyzék

1. fejezet

Lineáris algebra

„Mátrix”, „determináns”, „sajátérték”, „dimenzió”... Ezek a szavak olyan fogalmakat jelölnek, amelyek lépten-nyomon felbukkanak a legkülönfélébb alkalmazásokban, matematikán belül és kívül. Egy közös matematikai eszköztárhoz tartoznak – ennek a neve lineáris algebra.

A matematikának számos ága (mint például az analízis, a geometria, a valószínűségszámítás) használ lineáris algebrai eszközöket, de a mérnöki tudományoknak is nehéz volna olyan területét találni, ahol ne volna szükség lineáris algebrai ismeretekre.

Fokozottan érvényes ez az informatikára, amelynek szinte lehetetlen felsorolni azokat az alkalmazásait, ahol a lineáris algebra meghatározó szerephez jut. Hogyan jelenítsünk meg és mozgassunk egy tárgyat a képernyőn három dimenzióban? Hogyan továbbítsunk információt megbízhatóan egy megbízhatatlan csatornán? Hogyan tömörítsünk egy képet úgy, hogy a képfájl mérete jelentősen csökkenjen, de ez a kép minőségét alig befolyásolja? Ezek a kérdések csak ízelítőt igyekeznek adni azokból a területekből, ahol az informatikus lineáris algebrai háttér nélkül biztosan nem boldogul.

1.1. Koordinátageometria a térben

Bármennyire természetesnek is tűnik ma, a 17. században forradalmi ötlet volt, hogy a geometriai alakzatok és jelenségek algebrai eszközökkel is kezelhetők. Ennek az – általában Descartes-nak tulajdonított, bár már előtte is használt – gondolatnak az alapja a koordinátarendszer, amelynek a segítségével a pontok a síkban számpároknak, a térben számhármasoknak feleltethetők meg.

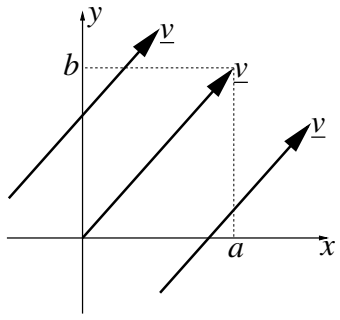
A középiskolai tanulmányokból a síkbeli koordinátageometria elemei mindenki számára ismertek. Ebben a fejezetben rövid bevezetőt adunk a térbeli koordinátageometriába; ezen belül is a térbeli egyenesek és síkok leírásával foglalkozunk.

1.1.1. A koordinátarendszer

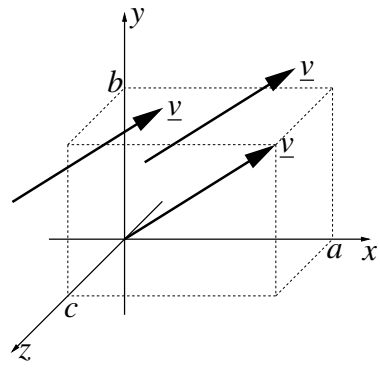
Hasonlóan ahhoz, ahogyan a síkban két merőleges egyenes és ezeken a pozitív irány és az egység kijelölése után minden pont egyértelműen megfeleltethető egy számpárnak, ugyanez a térben is természetes módon megtehető három, páronként merőleges egyenes – az x, y és z tengelyek – segítségével. Így a tér minden P pontja három koordinátával jellemezhető: $P(x,y,z)$. Például az a kijelentés, hogy „a P pont y-koordinátája 5” nyilván annyit jelent, hogy a P-t tartalmazó, az x és z tengelyekkel párhuzamos sík az y tengelyt az origótól pozitív irányban, attól 5 egységre metszi. A térbeli koordinátarendszert szokás úgy felvenni, hogy az x, y és z tengelyek (ebben a sorrendben) úgynevezett *jobbsodrású rendszert* alkossanak; ez azt jelenti, hogy jobb kezünk hüvelykujját az x, mutató ujját az y tengely pozitív irányába kinyújtva középső ujjunk kijelöli a z tengely pozitív irányát (lásd az 1.1b ábrát).

A térbeli vektorok

A síkbeli koordinátageometria kidolgozásának legfőbb segédeszközei a vektorok és a rajtuk végzett műveletek voltak; a térben ezek ugyanilyen hasznosak lesznek. A térbeli vektor továbbra is irányított szakaszt jelent – annak a kiegészítésnek a megtartásával, hogy az egymással párhuzamos, azonos irányú és hosszúságú vektorokat azonosnak tekintjük. Így az 1.1a és az 1.1b ábrán is ugyanannak a $\underline{v}$ vektornak a három–három példányát látjuk. A vektorok továbbra is jellemezhetők koordinátákkal: az (a,b,c) számhármas azt a vektort jelöli, ami (illetve amelynek az egyik példánya) az origóból az (a,b,c) pontba mutat. Fontos tehát észben tartani: függetlenül attól, hogy a $\underline{v}$ vektornak éppen melyik példányával találkozunk (azaz a koordinátarendszerben épp hová toltuk), a koordinátái változatlanok. Így például az 1.1a ábrán a $\underline{v}$ koordinátái (mindhárom esetben) (a,b), az 1.1b ábrán pedig $\underline{v}=(a,b,c)$. A vektoroknak azt a példányát, amelynek a kezdőpontja az origó, helyvektornak is szokás nevezni.



1.1a ábra



1.1b ábra

A valós számokból álló rendezett (vagyis a számok sorrendjét is tekintetbe vevő) számhármasok halmazát $\mathbb{R}^3$-bel jelöljük, hasonlóan ahhoz, ahogyan a rendezett

számpárok halmazát $\mathbb{R}^2$ jelölte. (Itt a „rendezett” jelző természetesen nem a tagok nagysági viszonyára utal: az $(1;2)$ és a $(2;1)$ egyaránt rendezett párok, de különbözők.) A fentiek szellemében tehát $\mathbb{R}^3$, illetve $\mathbb{R}^2$ azonosítható a térvektorok, illetve a síkvektorok halmazával.

Műveletek térvektorokkal

A síkvektorok közötti összeadás és kivonás, valamint a síkvektor skalárral (valós számmal) való szorzásának definíciója mindenki számára ismert. Azonnal látható, hogy ugyanezek a definíciók a térben is működnek, ezeket változtatás nélkül elfogadjuk. Például: $\underline{u}+\underline{v}$ meghatározásához felvesszük $\underline{u}$-t (annak egy tetszőleges példányát), majd $\underline{v}$-t úgy toljuk el, hogy annak kezdőpontja épp $\underline{u}$ végpontjába essen; most az $\underline{u}$ kezdőpontjából $\underline{v}$ végpontjába mutató vektor (és ennek minden eltoltja) lesz $\underline{u}+\underline{v}$. (Ha pedig $\underline{u}$ vagy $\underline{v}$ a nullvektor, akkor az összeg a másikukkal egyezik meg.) Mivel ez a definíció valójában térvektorok esetén is egy síkban (az $\underline{u}$ és $\underline{v}$ által kifeszített, origón átmenő síkban) „zajlik”, ezért változtatás nélkül működik – és ugyanez elmondható $\underline{u}-\underline{v}$ és $\lambda\cdot\underline{v}$ definíciójáról is, ahol $\lambda\in\mathbb{R}$.

A vektorok azért annyira hasznos segédeszközei a koordinátageometriának, mert a köztük végzett műveletek eredményeinek koordinátái nagyon egyszerűen megkaphatók az eredeti vektorok koordinátáiból.

1.1.1. Tétel. *Legyenek* $\underline{u}=(u_1,u_2,u_3)\in\mathbb{R}^3$ *és* $\underline{v}=(v_1,v_2,v_3)\in\mathbb{R}^3$ *térvektorok és* $\lambda\in\mathbb{R}$ *skalár. Ekkor*

(i) $\underline{u}+\underline{v}=(u_1+v_1,u_2+v_2,u_3+v_3)$,
(ii) $\underline{u}-\underline{v}=(u_1-v_1,u_2-v_2,u_3-v_3)$ *és*
(iii) $\lambda\cdot\underline{u}=(\lambda u_1,\lambda u_2,\lambda u_3)$.

A fenti tétel tehát azt mondja ki, hogy a térbeli vektorok közötti műveletek és a koordináták kapcsolata analóg módon működik azzal, ahogyan azt a síkvektorok esetében megszoktuk. A fenti tétel természetesen nem magától értetődő igazság, bizonyításra szorulna. A bizonyítást itt annak ellenére is elhagyjuk, hogy egyáltalán nem volna nehéz (a műveletek egyszerű tulajdonságaiból könnyen levezethető); ehelyett arra hivatkozunk, hogy a síkvektorokra vonatkozó (és így a középiskolai tanulmányokból ismert) analóg tétel bizonyítása itt is akadálymentesen működne.

A skaláris szorzat

Az $\underline{u}$ és $\underline{v}$ vektorok skaláris szorzatán definíció szerint az $\underline{u}\cdot\underline{v}=|\underline{u}|\cdot|\underline{v}|\cdot\cos\varphi$ skalárt értjük, ahol $|\underline{u}|$ és $|\underline{v}|$ a vektorok hosszát, φ pedig a bezárt szögüket jelöli. (Ha pedig $\underline{u}$ és $\underline{v}$ valamelyike $\underline{0}$, akkor $\underline{u}\cdot\underline{v}=0$.) Ismét elmondható, hogy a definíciót minden változtatás nélkül elfogadjuk térvektorokra is.

A skaláris szorzat gyakran a vektorok merőlegességének vizsgálatakor jut szerephez: $\underline{u}\cdot\underline{v}=0$ akkor és csak akkor igaz, ha $\underline{u}$ és $\underline{v}$ merőlegesek (vagy ha valamelyikük $\underline{0}$). Valóban: $\cos\varphi=0$ épp $\varphi=90^\circ$ esetén igaz (feltéve, hogy $0^\circ\le\varphi\le180^\circ$).

A skaláris szorzatot megint az teszi nagyon hasznos segédeszközzé, hogy az értéke a vektorok koordinátáiból a síkvektorok esetével analóg módon könnyen meghatározható.

1.1.2. Tétel. *Legyenek* $\underline{u} = (u_1, u_2, u_3) \in \mathbb{R}^3$ *és* $\underline{v} = (v_1, v_2, v_3) \in \mathbb{R}^3$ *térvektorok. Ekkor* $\underline{u} \cdot \underline{v} = u_1v_1 + u_2v_2 + u_3v_3$.

Ennek a tételnek a bizonyítását ismét elhagyjuk a középiskolai tanulmányokból ismert, síkvektorokra vonatkozó analóg állításra, illetve annak bizonyítására hivatkozva.

1.1.3. Feladat. Határozzuk meg annak a háromszögnek a területét, amelynek csúcsai $A(5;4;8)$, $B(4;-1;4)$ és $C(3;1;2)$.

Megoldás: A $T = \frac{a \cdot b \cdot \sin\gamma}{2}$ területképletet fogjuk használni (ahol a és b a háromszög két oldala és γ ezek közbezárt szöge). Kezdjük a C csúcsból induló oldalak, vagyis a $\overrightarrow{CA}$ és $\overrightarrow{CB}$ vektorok hosszának meghatározásával. Ha $\underline{a}$, $\underline{b}$ és $\underline{c}$ jelöli az origóból a háromszög csúcsaiba mutató vektorokat, akkor $\overrightarrow{CA} = \underline{a} - \underline{c}$ és $\overrightarrow{CB} = \underline{b} - \underline{c}$ következik a vektorok különbségének definíciójából. Mivel $\underline{a}$, $\underline{b}$ és $\underline{c}$ koordinátái megegyeznek a megfelelő csúcsok koordinátáival, az 1.1.1. Tételt használva $\overrightarrow{CA} = (5;4;8) - (3;1;2) = (2;3;6)$ és $\overrightarrow{CB} = (4;-1;4) - (3;1;2) = (1;-2;2)$.

A vektorok hosszának meghatározása (a síkbeli esettel analóg módon) a Pitagorasz-tétel segítségével lehetséges. Például a $\overrightarrow{CA}$ vektor esetében annak a téglatestnek a testátlóját keressük, amelynek két átellenes csúcsa az origó, illetve $(2;3;6)$. A Pitagorasz-tétel szerint az origó távolsága a $(2;3;0)$ ponttól $\sqrt{2^2+3^2}$. Ismét alkalmazva a Pitagorasz-tételt (az origó, a $(2;3;0)$ és a $(2;3;6)$ csúcsok által meghatározott derékszögű háromszögre) kapjuk, hogy $|\overrightarrow{CA}| = \sqrt{2^2+3^2+6^2} = 7$. Hasonlóan adódik, hogy $|\overrightarrow{CB}| = \sqrt{1^2+(-2)^2+2^2} = 3$.

A $\overrightarrow{CA}$ és $\overrightarrow{CB}$ vektorok γ szögének meghatározásához a skaláris szorzatot használjuk: egyrészt a definícióból adódik, hogy $\overrightarrow{CA} \cdot \overrightarrow{CB} = |\overrightarrow{CA}| \cdot |\overrightarrow{CB}| \cdot \cos\gamma = 7 \cdot 3 \cdot \cos\gamma$, másrészt az 1.1.2. Tétel szerint $\overrightarrow{CA} \cdot \overrightarrow{CB} = 2 \cdot 1 + 3 \cdot (-2) + 6 \cdot 2 = 8$. Ezekből tehát $\cos\gamma = \frac{8}{21}$.

Innen a $\sin^2\gamma + \cos^2\gamma = 1$ összefüggésből (felhasználva, hogy γ 180°-nál nem nagyobb, így $\sin\gamma \geq 0$) adódik, hogy $\sin\gamma = \frac{\sqrt{377}}{21}$. Ebből tehát a háromszög területe: $T = \frac{7 \cdot 3 \cdot \sin\gamma}{2} = \frac{\sqrt{377}}{2}$.

(Később kevesebb számolást igénylő módszert is megismerünk ennek a feladatnak a megoldására; lásd az 1.4.18. Feladatot.) □

1.1.2. Az egyenes egyenletrendszere

A síkban megszoktuk, hogy ha egy e egyenesnek már ismerjük egy adott P_0 pontját, akkor e meghatározásához elég megadni akár egy irányvektorát, akár egy normálvektorát – a kettő között az átjárás egyszerű. A térben már más a helyzet: hiába

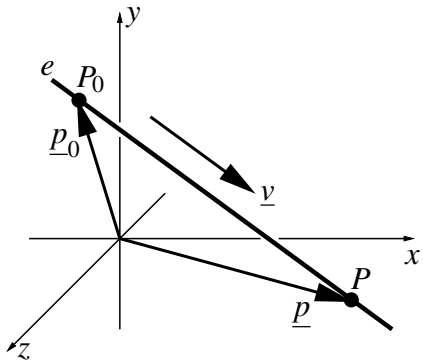
ismerjük e egy normálvektorát (vagyis egy e-re merőleges vektort) és P_0-t, ezáltal e még nincs egyértelműen megadva. Valóban: képzeljük el, hogy egy, az $\underline{n}$ vektorral párhuzamos tengely körül propeller módjára forgatjuk a P_0-on átmenő, $\underline{n}$-re merőleges egyenest; ekkor a propeller minden állásában olyan egyenest kapunk, amely P_0-t tartalmazza és merőleges $\underline{n}$-re.

A térben ezért egy e egyenes megadásához egy adott $P_0(x_0,y_0,z_0)$ pontja mellett egy $\underline{v}=(a,b,c)$ ($\underline{v}\neq\underline{0}$) irányvektorát adhatjuk csak meg (vagyis egy e-vel párhuzamos vektort). A kérdés az, hogy P_0 és $\underline{v}$ ismeretében hogyan kaphatjuk meg e összes többi pontját, illetve hogyan tesztelhetjük, hogy egy tetszőleges $P(x,y,z)$ pont rajta van-e e-n.

Nyilván eljuthatunk e minden P pontjához – és csak ezekhez, – ha P_0-ból felmérjük a $\underline{v}$ irányvektor minden lehetséges skalárszorosát (1.2. ábra). Ha tehát $P(x,y,z)$ egy tetszőleges pont a térben és $\underline{p}_0$, illetve $\underline{p}$ jelöli az origóból P_0-ba, illetve P-be mutató vektorokat, akkor $P\in e$ akkor és csak akkor igaz, ha $\underline{p}=\underline{p}_0+\lambda\cdot\underline{v}$ fennáll egy alkalmas $\lambda\in\mathbb{R}$ skalárra. Mivel $\underline{p}=(x,y,z)$ és az 1.1.1. Tétel szerint $\underline{p}_0+\lambda\cdot\underline{v}=(x_0+\lambda\cdot a,y_0+\lambda\cdot b,z_0+\lambda\cdot c)$, ez az alábbival ekvivalens:

$$\begin{aligned} x &= x_0+\lambda\cdot a\\ y &= y_0+\lambda\cdot b\\ z &= z_0+\lambda\cdot c\\ \lambda &\in \mathbb{R} \end{aligned} \tag{1.1.1}$$

Ezt az egyenletrendszert az egyenes *paraméteres egyenletrendszerének* szokás nevezni. Ha a λ paramétert képzeletben végigfuttatjuk a valós számegyenesen, akkor az 1.1.1. egyenletrendszer által megadott (x,y,z) pont végigfut az e egyenesen.



1.2. ábra

1.1.4. Feladat. Hol döfi a $P_0(8;-8;15)$ ponton átmenő, $\underline{v}=(2;-3;5)$ irányvektorú egyenes az x és y tengelyeket tartalmazó síkot?

Megoldás: Az egyenes paraméteres egyenletrendszere: $x=8+2\lambda$, $y=-8-3\lambda$,

$z = 15+5\lambda$. Az x és y tengelyeket tartalmazó síkon azok a pontok vannak, amelyeknek a z koordinátája 0. A $z = 15+5\lambda = 0$ egyenletből $\lambda = -3$ adódik. Ezt az egyenletrendszerbe visszahelyettesítve: $x = 2, y = 1$. Így a döféspont: $(2;1;0)$. □

Miközben az egyenes paraméteres egyenletrendszere sokszor jól használható, mégsem tölti be azt a feladatot, amit a síkban az egyenes egyenletétől elvártunk: nem válaszolja meg (közvetlenül) azt a kérdést, hogy egy adott pont rajta van-e az egyenesen vagy sem. Például: rajta van-e a $Q(12;-14;20)$ pont az 1.1.4. Feladatban szereplő e egyenesen? A paraméteres egyenletrendszer szerint a válasz akkor lenne igenlő, ha a $12 = 8+2\lambda$, $-14 = -8-3\lambda$, $20 = 15+5\lambda$ egyenletek teljesülnének valamely λ-ra. Ilyen λ azonban nincs, mert az első két egyenletből $\lambda = 2$, az utolsóból $\lambda = 1$ adódna; így $Q \notin e$. Ha ugyanezt a kérdést az $R(6;-5;10)$ pontra tesszük fel, akkor a $6 = 8+2\lambda$, $-5 = -8-3\lambda$, $10 = 15+5\lambda$ egyenletek mindegyikéből a közös $\lambda = -1$ érték adódik; így $R \in e$ (hiszen $\lambda = -1$-et a paraméteres egyenletrendszerbe helyettesítve épp R-et kapjuk). Ha most a kérdést egy általános $P(x,y,z)$ pontra tesszük fel, azt mondhatjuk, hogy $P \in e$ pontosan akkor igaz, ha az e paraméteres egyenletrendszerének mindhárom egyenletéből λ-t kifejezve közös értéket kapunk:

$$\frac{x-8}{2} = \frac{y+8}{-3} = \frac{z-15}{5} \tag{1.1.2}$$

Valóban: ha ez a két egyenlet teljesül, akkor $\frac{x-8}{2}$, $\frac{y+8}{-3}$ és $\frac{z-15}{5}$ közös értékét λ-nak választva a paraméteres egyenletrendszer épp P-t adja. Az 1.1.2. egyenletrendszer pedig már abban az értelemben írja le e-t, ahogyan azt az egyenes síkbeli egyenletétől elvártuk: alkalmas a $P \in e$ állítás igazságának tesztelésére.

Természetesen a fenti gondolatmenet tetszőleges e egyenesre megismételhető, ez következik az alábbi tétel bizonyításában. Azonban sajnos nem minden esetben működik minden pont úgy, ahogyan azt a fenti példában láttuk: az 1.1.2. egyenletrendszerben az irányvektor koordinátái a nevezőbe kerültek, ami nyilván nem történhetett volna meg, ha közülük valamelyik nulla.

1.1.5. Tétel. *Legyen adott az e egyenesnek egy $P_0(x_0,y_0,z_0)$ pontja és egy $\underline{v} = (a,b,c)$, $\underline{v} \neq \underline{0}$ irányvektora. Ekkor egy tetszőleges $P(x,y,z)$ pontra $P \in e$ pontosan akkor igaz, ha*

(i) $\dfrac{x-x_0}{a} = \dfrac{y-y_0}{b} = \dfrac{z-z_0}{c}$, *feltéve hogy $a \neq 0$, $b \neq 0$ és $c \neq 0$;*

(ii) $\dfrac{x-x_0}{a} = \dfrac{y-y_0}{b}$ *és $z = z_0$, feltéve hogy $a \neq 0$ és $b \neq 0$, de $c = 0$ (és ezzel analóg állítás igaz, ha $\underline{v}$ koordinátái közül pontosan egy 0, de az nem c);*

(iii) $x = x_0$, $y = y_0$, *feltéve hogy $a = b = 0$, de $c \neq 0$ (és ezzel analóg állítás igaz, ha $\underline{v}$ koordinátái közül pontosan egy nem nulla, de az nem c).*

Bizonyítás: $P \in e$ pontosan akkor igaz, ha e 1.1.1 paraméteres egyenletrendszere valamely $\lambda \in \mathbb{R}$ értékre P-t adja. Az $a \neq 0$, $b \neq 0$, $c \neq 0$ esetben tehát a három egyenletből λ-t kifejezve közös értéket kell kapnunk, ami épp a tétel *(i)* állítására vezet (összhangban az 1.1.2 példával). Ha $a \neq 0$, $b \neq 0$, de $c = 0$, akkor a megfelelő

λ létezése épp azt jelenti, hogy a $z = z_0$ egyenlet fennáll (ez a paraméteres egyenletrendszer harmadik egyenlete) és az első két egyenletből λ-t kifejezve közös értéket kapunk; ez épp a tétel *(ii)* állítása. Végül az $a = b = 0$, $c \neq 0$ esetben a paraméteres egyenletrendszer első két egyenlete az $x = x_0$, $y = y_0$ alakra egyszerűsödik, míg a harmadik egyenlet mindig kielégíthető a $\lambda = \frac{z-z_0}{c}$ választással. □

Az 1.1.5. Tételben szereplő alakot szokás a térbeli egyenes *(nem paraméteres) egyenletrendszerének* nevezni. Fontos kiemelni, hogy a térbeli egyeneseket csak egyenletrendszerrel lehet leírni, az „egyenes egyenletéről" beszélni a térben értelmetlen.

1.1.6. Feladat. Rajta van-e a $P(8;4;2)$ és $Q(17;4;-1)$ pontokon átmenő e egyenesen az $R(23;4;-3)$ pont?

Megoldás: e-nek irányvektora a $\overrightarrow{PQ}$ vektor, amit a Q-ba és P-be mutató helyvektorok különbségeként kaphatunk meg: $\overrightarrow{PQ} = (17;4;-1) - (8;4;2) = (9;0;-3)$. Kényelmi okokból használhatjuk ennek a harmadát is, a $\underline{v} = (3;0;-1)$ irányvektort. P és $\underline{v}$ alapján felírhatjuk (az 1.1.5. Tétel *(ii)* állítása szerint) e egyenletrendszerét: $\frac{x-8}{3} = \frac{z-2}{-1}$, $y = 4$. Ebbe az R koordinátáit helyettesítve: $\frac{23-8}{3} = \frac{-3-2}{-1}$, $4 = 4$. Mivel az egyenletrendszert R kielégíti, $R \in e$ igaz. (Használhatnánk az egyenletrendszer felírásánál P helyett Q-t is, illetve $\underline{v}$ helyett $\lambda\underline{v}$-t is bármely $\lambda \neq 0$-ra. Az így kapott egyenletrendszerek más-más alakúak volnának, de mindegyikük e-t írná le. De megoldható a feladat az e egyenletrendszerének felírása nélkül is: $\overrightarrow{PR} = (15;0;-5)$, így $\overrightarrow{PR} = \frac{5}{3}\overrightarrow{PQ}$; mivel $\overrightarrow{PR}$ és $\overrightarrow{PQ}$ párhuzamosak, ezért P, Q és R egy egyenesen vannak.) □

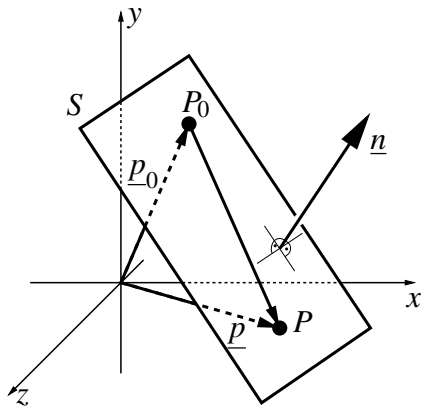
1.1.3. A sík egyenlete

Egy térbeli S síkot már nem lehet megadni egy P_0 pontjával és egy $\underline{v}$ „irányvektorával" (vagyis egy S-sel párhuzamos $\underline{v}$ vektorral), mert P_0 és $\underline{v}$ nem határozná meg egyértelműen S-et. Valóban: ha egy síkot a P_0-on átmenő, $\underline{v}$ irányvektorú egyenes, mint tengely körül forgatnánk (hasonlóan ahhoz, ahogyan egy könyv gerince körül a lapjai forognak), a sík minden állásában $\underline{v}$-vel párhuzamos, P_0-on átmenő síkot kapnánk.

Ezért az S sík leírásához egy P_0 pontján kívül egy rá merőleges $\underline{n}$ normálvektort adunk meg. A kérdés természetesen az, hogy P_0 és $\underline{n}$ ismeretében hogyan írhatjuk le a síkot, hogyan tudjuk egy tetszőleges P pontról eldönteni, hogy S-en van-e. Az alábbi tétel erre ad választ; mind a tétel állítása, mind pedig a bizonyítása analóg lesz a síkbeli egyenes normálvektoros egyenletével kapcsolatos (középiskolában tanult) ismeretekkel. A tételben megjelenő egyenletet a sík (térbeli) egyenletének szokás nevezni.

1.1.7. Tétel. *Legyen adott az S síknak egy $P_0(x_0,y_0,z_0)$ pontja és egy $\underline{n}=(a,b,c)$, $\underline{n}\neq\underline{0}$ normálvektora. Ekkor egy tetszőleges $P(x,y,z)$ pontra $P\in S$ pontosan akkor igaz, ha $ax+by+cz=ax_0+by_0+cz_0$.*

Bizonyítás: $P\in S$ pontosan akkor igaz, ha a $\overrightarrow{P_0P}$ vektor párhuzamos S-sel (lásd az 1.3. ábrát). Itt $\overrightarrow{P_0P}=\underline{p}-\underline{p_0}$, ahol $\underline{p}$ és $\underline{p_0}$ a megfelelő pontokba mutató helyvektorok. Így az 1.1.1. tétel szerint $\overrightarrow{P_0P}=(x-x_0,y-y_0,z-z_0)$.



1.3. ábra

$\overrightarrow{P_0P}$ pedig pontosan akkor párhuzamos S-sel, ha merőleges $\underline{n}$-re. Ez viszont (a skaláris szorzat definíciója szerint) pontosan akkor igaz, ha $\overrightarrow{P_0P}\cdot\underline{n}=0$. Az 1.1.2. Tétel szerint $\overrightarrow{P_0P}\cdot\underline{n}=(x-x_0)a+(y-y_0)b+(z-z_0)c$. Így $\overrightarrow{P_0P}\cdot\underline{n}=0$ beszorzás és átrendezés után épp az $ax+by+cz=ax_0+by_0+cz_0$ egyenletre vezet. □

1.1.8. Feladat. Rajta van-e a $P_0(5;-2;4)$ ponton átmenő, az $\frac{x-9}{2}=\frac{y+5}{6}=z$ egyenletrendszerű e egyenesre merőleges S síkon a $Q(4;-1;1)$ pont?

Megoldás: Az S egy pontját ismerjük, így az egyenletének felírásához egy normálvektorát kell megkeresnünk. Mivel e merőleges S-re, ezért e irányvektorai azonosak S normálvektoraival. e egy $\underline{v_e}$ irányvektorát kiolvashatjuk az egyenletrendszeréből: $\underline{v_e}=(2;6;1)$. (Ehhez a második egyenlet jobb oldalán álló z tagot $\frac{z-0}{1}$ alakban fogtuk fel. Az irányvektor „kiolvasása” részletesebben azt jelenti, hogy ha felírnánk a $\underline{v_e}=(2;6;1)$ irányvektorú, a $(9;-5;0)$ ponton átmenő f egyenes egyenletrendszerét az 1.1.5. Tétel szerint, akkor épp a feladatbeli egyenletrendszert kapnánk. Mivel az egyenletrendszereik azonosak, ezért $e=f$, így $\underline{v_e}$ valóban irányvektora $\underline{e}$-nek.)

Megkaptuk tehát S egy normálvektorát: $\underline{n_s}=\underline{v_e}=(2;6;1)$. Ez és P_0 alapján már felírhatjuk S egyenletét: $2x+6y+z=2\cdot5+6\cdot(-2)+1\cdot4$, vagyis $2x+6y+z=2$.

Ebbe a Q pont koordinátáit behelyettesítve azt kapjuk, hogy az egyenlet nem teljesül ($3 \neq 2$), ezért $Q \notin S$. □

1.1.9. Feladat. Írjuk fel az $A(3;3;1)$, $B(5;2;4)$ és $C(8;5;0)$ pontokra illeszkedő S sík egyenletét.

Megoldás: Elegendő S egy $\underline{n}$ normálvektorát meghatároznunk, hiszen S-nek három adott pontját is ismerjük. Olyan vektort keresünk tehát, ami S-re, és így az $\overrightarrow{AB}$ és $\overrightarrow{AC}$ vektorokra is merőleges.

$\overrightarrow{AB}$-t és $\overrightarrow{AC}$-t a már látott módon, helyvektorok különbségeként kapjuk: $\overrightarrow{AB} = (5;2;4) - (3;3;1) = (2;-1;3)$ és $\overrightarrow{AC} = (8;5;0) - (3;3;1) = (5;2;-1)$.

Legyen a keresett normálvektor $\underline{n} = (a,b,c)$. Mivel $\underline{n}$ merőleges $\overrightarrow{AB}$-re és $\overrightarrow{AC}$-re, ezért $\overrightarrow{AB} \cdot \underline{n} = 0$ és $\overrightarrow{AC} \cdot \underline{n} = 0$. Az 1.1.2. Tétel szerint ez a $2a - b + 3c = 0$, $5a + 2b - c = 0$ egyenletrendszerre vezet. A második egyenletből $c = 5a + 2b$, ezt az elsőbe helyettesítve: $17a + 5b = 0$. Ennek megoldása például az $a = 5$, $b = -17$, amiből $c = 5a + 2b = -9$. Így $\underline{n} = (5;-17;-9)$ normálvektora S-nek. Ez és (például) A segítségével felírva S egyenletét: $5x - 17y - 9z = 5 \cdot 3 - 17 \cdot 3 - 9 \cdot 1$, vagyis $5x - 17y - 9z = -45$.

(Nem szabad meglepődnünk azon, hogy $\underline{n}$ keresésekor egy három ismeretlenes, de csak két egyenletből álló egyenletrendszert kaptunk, hiszen S-nek nyilván végtelen sok különböző normálvektora van.)

Később kevesebb számolást igénylő módszert is megismerünk ennek a feladatnak a megoldására; lásd az 1.4.17. Feladatot. □

Az 1.1.7. tételnek fontos következménye, hogy minden lineáris, vagyis $ax + by + cz = d$ alakú egyenlet (ahol $a,b,c,d \in \mathbb{R}$ rögzített számok és x,y,z változók) a térben síkot határoz meg. Valóban: ha választunk egy tetszőleges x_0, y_0, z_0 számhármast, ami az egyenletet kielégíti (ezt nyilván meg lehet tenni), akkor a $P_0(x_0, y_0, z_0)$ pontra illeszkedő, $\underline{n} = (a,b,c)$ normálvektorú sík egyenlete épp $ax + by + cz = d$ lesz.

Érdemes ennek a megfigyelésnek a birtokában az egyenes 1.1.5. Tételben megismert egyenletrendszerére visszatekinteni, hiszen a rendszert alkotó mindkét egyenlet (átrendezés után) lineáris egyenlet (amelyben legalább egy változó együtthatója 0). Az egyenletrendszer megoldáshalmaza tehát azon pontokból áll, amelyek a két megfelelő síkon rajta vannak – vagyis az egyenletrendszer két sík metszésvonalaként írja le az egyenest. (Az a tény pedig, hogy a síkok egyenletében egy változó együtthatója 0 annak felel meg, hogy a síkok a megfelelő tengellyel párhuzamosak.)

1.2. Az n magas számoszlopok tere

A sci-fi irodalomban nem különösebben jártasak is minden bizonnyal hallották már a „négydimenziós tér” kifejezést és sokakban kelt ez a név borzongó hitetlenkedést:

hogyan is lehetne négy, páronként merőleges egyenes által alkotott koordinátarendszert elképzelni? Ebben a fejezetben megismerjük a kiábrándító választ: sehogyan sem, viszont a négy-, vagy akár n-dimenziós tér fogalma viszonylag egyszerű és a geometriától teljesen független.

1.2.1. $\mathbb{R}^n$ fogalma

Tudjuk, hogy a (köznyelv szerint is két-, illetve háromdimenziós) sík, illetve tér pontjai számpároknak, illetve számhármasoknak feleltethetők meg. Ennek mintájára természetesen adódik a gondolat: bevezethetjük az n darab szám által leírt „pontokat" is, amelyeknek a halmazát $\mathbb{R}^n$ fogja jelölni. Az így alkotott fogalmat igazán hasznossá az fogja tenni, ha a sík- és térvektorok körében megszokott vektorműveleteket is átörökítjük az 1.1.1. Tétel szellemében – és ennek megfelelően $\mathbb{R}^n$ elemeit is inkább vektoroknak, mint pontoknak fogjuk hívni. Míg a koordinátageometriában a vektorokat általában *sorvektorként* írtuk – vagyis a koordináták egymás mellett, egy sorban álltak, – addig a lineáris algebrában általános n esetén inkább használatos az *oszlopvektoros* jelölés, ahol a koordináták egymás alatt, egy oszlopban helyezkednek el. A kettő között nyilván nincs érdemi különbség, az oszlopvektoros jelölés haszna később fog megmutatkozni.

1.2.1. Definíció. *Tetszőleges $n \geq 1$ egész esetén az n darab valós számból álló számoszlopok halmazát $\mathbb{R}^n$ jelöli. Az $\mathbb{R}^n$-en értelmezett, „$+$"-szal jelölt* összeadást *és tetszőleges $\lambda \in \mathbb{R}$ esetén a „$\cdot$"-tal (vagy egyszerűen egymás mellé írással) jelölt* skalárral való szorzást *az alábbi egyenlőségek szerint értelmezzük:*

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1+y_1 \\ x_2+y_2 \\ \vdots \\ x_n+y_n \end{pmatrix} \qquad \textit{és} \qquad \lambda \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \\ \vdots \\ \lambda x_n \end{pmatrix}.$$

$\mathbb{R}^n$ elemeit *vektornak* nevezzük, ezeket aláhúzott latin betű ($\underline{v}$, $\underline{a}$, stb.) jelöli. A vektorokat alkotó számokat a vektor *koordinátáinak* hívjuk. Ha a vektort koordinátáival adjuk meg, akkor ezeket (a fenti definícióban látott módon) gömbölyű zárójelek közé írjuk. Így például legyen $\mathbb{R}^4$-ben

$$\underline{v} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} \text{ és } \underline{w} = \begin{pmatrix} 5 \\ 0 \\ -1 \\ \sqrt{2} \end{pmatrix}, \text{ ekkor } \underline{v}+\underline{w} = \begin{pmatrix} 6 \\ 2 \\ 2 \\ 4+\sqrt{2} \end{pmatrix} \text{ és } (-2)\underline{v} = \begin{pmatrix} -2 \\ -4 \\ -6 \\ -8 \end{pmatrix}.$$

A vektor fogalma tehát ezentúl *nem jelent irányított szakaszt*, csak egy számoszlopot. Az irányított szakaszoknak a középiskolából ismert fogalmára (a félreértések elkerülése végett) a továbbiakban *síkvektorként* vagy *térvektorként* hivatkozunk. Megjegyezzük, hogy $\mathbb{R}^n$-et szándékosan nem neveztük „az n-dimenziós térnek", mert ezt

a fogalmat később ennél általánosabb értelemben fogjuk bevezetni (lásd az 1.2.20. Definíciót), noha $\mathbb{R}^n$-ről is ki fog derülni, hogy n-dimenziós.

A fenti definícióban nem vezettük be külön az $\mathbb{R}^n$-beli kivonás műveletét. Erre nincs is szükség: $\underline{u}-\underline{v}$ alatt az $\underline{u}+(-1)\cdot\underline{v}$ összeget fogjuk érteni (és ez az 1.2.1. Definíció szerint valóban koordinátánkénti kivonást jelent, összhangban a síkban és térben megszokottakkal).

Nullvektornak hívjuk és $\underline{0}$-val jelöljük azt az $\mathbb{R}^n$-beli vektort, amelynek minden koordinátája 0. Ez „örökli" a térbeli nullvektornak azt a tulajdonságát, hogy a vele végzett összeadás azonos a változatlanul hagyással: $\underline{v}+\underline{0}=\underline{v}$ igaz minden $\underline{v}\in\mathbb{R}^n$-re.

Fontos kiemelni, hogy $n\geq 4$ esetén az $\mathbb{R}^n$-beli vektoroknak már nincs közvetlen geometriai jelentése. Persze megpróbálhatjuk algebrai úton megfogalmazni azoknak a fogalmaknak és állításoknak a megfelelőjét tetszőleges n-re is, amelyeket $n\leq 3$ esetén még „látunk" és ez gyakran hasznos és gyümölcsöző lesz – láthatóvá azonban nem tesz semmit. Például: az egyenes 1.1.1. paraméteres egyenletrendszere által megihletve definiálhatjuk az $\mathbb{R}^n$-beli egyenes fogalmát is, mint a $\underline{p}_0+\lambda\cdot\underline{v}$ alakban kifejezhető vektorok halmazát, ahol $\underline{p}_0,\underline{v}\in\mathbb{R}^n$ rögzített vektorok és $\lambda\in\mathbb{R}$ skalár. De még ha el is fogadjuk ezt a definíciót (amely bizonyos alkalmazásokban valóban hasznos, bár ebben a jegyzetben többet nem kerül elő), egy „százdimenziós egyenes" akkor sem lesz más, mint bizonyos száz magasságú számoszlopok halmaza, a vizuális fantáziánk számára értelmezhető tartalma nincs. Ennek megfelelően fontos megérteni az elvi különbséget az 1.1.1. Tétel és az 1.2.1. Definíció között: az például, hogy a térvektorokat koordinátánként kell összeadni egy tétel, amely levezethető a térvektorok (mint irányított szakaszok) összeadásának tulajdonságaiból; ugyanez $\mathbb{R}^n$-ben már semmiből nem levezethető, hanem definíció kérdése.

Természetes gondolat volna a skaláris szorzást is kiterjeszteni $\mathbb{R}^n$-re az 1.1.2. Tétel nyomdokain. Az így kapott fogalom valóban fontos és hasznos, de (ebben a jegyzetben) csak egy általánosabb fogalom speciális eseteként fogunk vele találkozni (lásd az 1.5.5. Definíciót).

A következő tétel azt mondja ki, hogy azok a műveleti tulajdonságok, amelyeket a sík- és térvektorok esetén megszoktunk, igazak maradnak $\mathbb{R}^n$-ben is.

1.2.2. Tétel. *Legyen $\underline{u},\underline{v},\underline{w}\in\mathbb{R}^n$ és $\lambda,\mu\in\mathbb{R}$. Ekkor igazak az alábbiak:*
(i) $\underline{u}+\underline{v}=\underline{v}+\underline{u}$, *(vagyis az $\mathbb{R}^n$-beli összeadás* kommutatív*);*
(ii) $(\underline{u}+\underline{v})+\underline{w}=\underline{u}+(\underline{v}+\underline{w})$, *(vagyis az $\mathbb{R}^n$-beli összeadás* asszociatív*);*
(iii) $\lambda\cdot(\underline{u}+\underline{v})=\lambda\cdot\underline{u}+\lambda\cdot\underline{v}$*;*
(iv) $(\lambda+\mu)\cdot\underline{v}=\lambda\cdot\underline{v}+\mu\cdot\underline{v}$*;*
(v) $\lambda\cdot(\mu\cdot\underline{v})=(\lambda\mu)\cdot\underline{v}$.

Bizonyítás: A felsorolt tulajdonságok mindegyike azonnal következik a valós számok műveleti tulajdonságaiból. Illusztrációképp részletesen leírjuk a *(iii)* tulajdonság bizonyítását, a többit az olvasóra hagyjuk. Legyen $\underline{u}=\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}$ és $\underline{v}=\begin{pmatrix}y_1\\y_2\\\vdots\\y_n\end{pmatrix}$.

Ekkor az egyenlőség bal oldala: $\lambda\cdot(\underline{u}+\underline{v})=\lambda\cdot\begin{pmatrix}x_1+y_1\\x_2+y_2\\\vdots\\x_n+y_n\end{pmatrix}=\begin{pmatrix}\lambda(x_1+y_1)\\\lambda(x_2+y_2)\\\vdots\\\lambda(x_n+y_n)\end{pmatrix}$. Az egyenlet jobb oldala: $\lambda\cdot\underline{u}+\lambda\cdot\underline{v}=\begin{pmatrix}\lambda x_1\\\lambda x_2\\\vdots\\\lambda x_n\end{pmatrix}+\begin{pmatrix}\lambda y_1\\\lambda y_2\\\vdots\\\lambda y_n\end{pmatrix}=\begin{pmatrix}\lambda x_1+\lambda y_1\\\lambda x_2+\lambda y_2\\\vdots\\\lambda x_n+\lambda y_n\end{pmatrix}$. Látható, hogy az egyenlet két oldalán álló vektorok valóban azonosak. □

1.2.3. Feladat. Legyen $\mathbb{R}^4$-ben

$$\underline{a}=\begin{pmatrix}1\\-1\\0\\0\end{pmatrix},\ \underline{b}=\begin{pmatrix}0\\1\\-1\\0\end{pmatrix},\ \underline{c}=\begin{pmatrix}0\\0\\1\\-1\end{pmatrix}\ \text{és}\ \underline{d}=\begin{pmatrix}2\\-9\\4\\3\end{pmatrix}.$$

a) Kifejezhető-e $\underline{d}$ az $\underline{a}$, $\underline{b}$ és $\underline{c}$ vektorokból (az $\mathbb{R}^4$-beli műveletekkel)?

b) Mely $\underline{v}$ vektorok fejezhetők ki az $\underline{a}$, $\underline{b}$ és $\underline{c}$ vektorokból? Adjunk olyan feltételt, amelynek segítségével $\underline{v}$ ismeretében a válasz gyorsan megadható.

Megoldás: Hogyan lehetne áttekinteni, hogy $\underline{a}$, $\underline{b}$ és $\underline{c}$ felhasználásával mely $\underline{v}$-k fejezhetők ki? Elsőre azt hihetnénk, hogy az összes ilyen $\underline{v}$ leírása reménytelenül bonyolult feladat, hiszen olyan komplikált kifejezéseket építhetünk, mint például $\underline{v}=6(7(3\underline{a}-5\underline{c})+2(\underline{b}-4\underline{a}))-\frac{8}{9}\underline{b}$, és ez még a „szelídebbek” közé tartozik. De azonnal látszik, hogy az 1.2.2. Tételbeli tulajdonságokat használva ez a kifejezés jóval egyszerűbb alakra hozható: elvégezve a beszorzásokat és összevonásokat a $\underline{v}=118\underline{a}+\frac{10}{9}\underline{b}-210\underline{c}$ alakot kapjuk. Ebből a tapasztalatból kiindulva nem nehéz meggyőződnünk arról, hogy *minden* olyan $\underline{v}$ vektor, amely $\underline{a}$, $\underline{b}$ és $\underline{c}$ felhasználásával kifejezhető, felírható $\underline{v}=\alpha\underline{a}+\beta\underline{b}+\gamma\underline{c}$ alakban, ahol $\alpha,\beta,\gamma\in\mathbb{R}$.

Az a) feladat megoldása tehát arra egyszerűsödik, hogy olyan α, β és γ skalárokat keressünk, amelyekre $\alpha\underline{a}+\beta\underline{b}+\gamma\underline{c}=\underline{d}$. Behelyettesítve a négy vektor konkrét értékét és elvégezve a műveleteket:

$$\alpha\underline{a}=\begin{pmatrix}\alpha\\-\alpha\\0\\0\end{pmatrix},\beta\underline{b}=\begin{pmatrix}0\\\beta\\-\beta\\0\end{pmatrix},\gamma\underline{c}=\begin{pmatrix}0\\0\\\gamma\\-\gamma\end{pmatrix},\ \text{így}\ \alpha\underline{a}+\beta\underline{b}+\gamma\underline{c}=\begin{pmatrix}\alpha\\-\alpha+\beta\\-\beta+\gamma\\-\gamma\end{pmatrix}.$$

Tehát $\alpha\underline{a}+\beta\underline{b}+\gamma\underline{c}=\underline{d}$ ekvivalens az $\alpha=2$, $-\alpha+\beta=-9$, $-\beta+\gamma=4$, $-\gamma=3$ lineáris egyenletrendszerrel. Az első és utolsó egyenletekből $\alpha=2$ és $\gamma=-3$, ezeket felhasználva pedig a maradék két egyenlet egyaránt $\beta=-7$-et ad. Ezzel tehát az a) feladatot megoldottuk: a válasz igen, $\underline{d}=2\underline{a}-7\underline{b}-3\underline{c}$.

Persze $\underline{d}$ helyett mást választva nemleges válaszra is juthattunk volna: ha a középső két egyenlet más-más értéket adott volna β-ra és így az egyenletrendszer

ellentmondásra vezet. A b) feladat megoldásához tehát vegyünk egy tetszőleges $\underline{v} = \begin{pmatrix} p \\ q \\ r \\ s \end{pmatrix}$ vektort és járjuk végig az a) feladat megoldását újra, de most $\underline{d}$ helyett $\underline{v}$-vel. A gondolatmenet azonos, csak a kapott egyenletrendszer más: $\alpha = p$, $-\alpha + \beta = q$, $-\beta + \gamma = r$, $-\gamma = s$, ahol tehát α, β és γ továbbra is a változók és p, q, r, s paraméterek (vagyis ezek értékének a függvényében kell eldönteni, hogy az egyenletrendszer megoldható-e). Az első és utolsó egyenletből $\alpha = p$, $\gamma = -s$, ezeket a másik kettőbe helyettesítve $\beta = p + q$ és $\beta = -s - r$. Az egyenletrendszer tehát akkor lesz megoldható, ha a β-ra kapott két érték azonos és így nem jutunk ellentmondásra: $p + q = -r - s$. Átrendezés után a $p + q + r + s = 0$ feltételre jutunk, vagyis azt mondhatjuk, hogy azok és csak azok a vektorok fejezhetők ki $\underline{a}$, $\underline{b}$ és $\underline{c}$ felhasználásával, amelyekben a négy koordináta összege 0. (Ez például $\underline{d}$-re is teljesül, összhangban az a) feladat eredményével.) □

1.2.2. $\mathbb{R}^n$ alterei

Geometriai szemléletünk alapján világos, hogy a tér végtelen sok különböző példányban tartalmazza a síkot – hasonlóan ahhoz, ahogyan a sík végtelen sok példányban tartalmazza az „egydimenziós teret”, az egyenest. A következő definíció ezt a jelenséget igyekszik általánosítani $\mathbb{R}^n$-re is – noha fontos már most rögzíteni, hogy a kapott fogalom nem lesz teljesen azonos a szemlélet által sugallt képpel, például nem minden térbeli sík lesz altere a térnek (csak az origón átmenők).

1.2.4. Definíció. *Legyen $\emptyset \neq V \subseteq \mathbb{R}^n$ az $\mathbb{R}^n$ tér egy nemüres részhalmaza. V-t az $\mathbb{R}^n$* alterének *nevezzük, ha az alábbi két feltétel teljesül:*
(i) bármely $\underline{u}, \underline{v} \in V$ esetén $\underline{u} + \underline{v} \in V$ is igaz;
(ii) bármely $\underline{v} \in V$, $\lambda \in \mathbb{R}$ esetén $\lambda \cdot \underline{v} \in V$ is igaz.
Azt a tényt, hogy V altere $\mathbb{R}^n$-nek, így jelöljük: $V \leq \mathbb{R}^n$.

A definícióban szereplő két feltétel teljesülését röviden úgy fejezzük ki, hogy *V zárt az összeadásra és a skalárral szorzásra* (vagyis bármely két V-beli vektor összege, illetve bármely V-beli vektor tetszőleges számszorosa V-beli). $V = \mathbb{R}^n$ és $V = \{\underline{0}\}$ (vagyis a csak a nullvektort tartalmazó részhalmaz) nyilván teljesítik ezeket a feltételeket, így alterek; ezeket *triviális altérnek* hívjuk. A definícióból az is következik, hogy $\underline{0} \in V$ minden $V \leq \mathbb{R}^n$ altérre igaz: valóban, a *(ii)* tulajdonságot $0 \in \mathbb{R}$-re és egy tetszőleges $\underline{v} \in V$-re alkalmazva kapjuk, hogy $\underline{0} = 0 \cdot \underline{v} \in V$.

1.2.5. Feladat. Alteret alkotnak-e az alábbi $\mathbb{R}^n$-beli részhalmazok?

a) $V_1 = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 : x \geq 0 \text{ és } y \geq 0 \right\}$

b) $V_2 = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 : x \cdot y \geq 0 \right\}$

c) $V_3 = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 : x = y \right\}$

d) $V_4 = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 : 12x - 23y + 34z = 46 \right\}$

e) $V_5 = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 : 12x - 23y + 34z = 0 \right\}$

f) $V_6 = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \in \mathbb{R}^4 : x_1 + x_2 + x_3 + x_4 = 0 \right\}$

Megoldás: a) V_1 tehát azokból a síkvektorokból áll, amelyeknek mindkét koordinátája nemnegatív. V_1 zárt az összeadásra (két V_1-belit összeadva mindig V_1-belit kapunk), de nem zárt a skalárral szorzásra: például $\underline{v} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \in V_1$, de $(-1)\underline{v} = \begin{pmatrix} -1 \\ -1 \end{pmatrix} \notin V_1$. Ezért V_1 nem altér.

b) V_2 zárt a skalárral szorzásra, de nem zárt az összeadásra: például $\underline{u} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \in V_2$ és $\underline{v} = \begin{pmatrix} 0 \\ -1 \end{pmatrix} \in V_2$, de $\underline{u} + \underline{v} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \notin V_2$. Ezért V_2 sem altér.

c) Legyen $\underline{u}, \underline{v} \in V_3$, vagyis $\underline{u} = \begin{pmatrix} p \\ p \end{pmatrix}$ és $\underline{v} = \begin{pmatrix} q \\ q \end{pmatrix}$ valamely $p, q \in \mathbb{R}$ értékekre. Ekkor $\underline{u} + \underline{v} = \begin{pmatrix} p+q \\ p+q \end{pmatrix} \in V_3$, vagyis V_3 zárt az összeadásra. Hasonlóan látszik, hogy V_3 a skalárral szorzásra is zárt: ha $\underline{u} = \begin{pmatrix} p \\ p \end{pmatrix} \in V_3$, akkor $\lambda \underline{u} = \begin{pmatrix} \lambda p \\ \lambda p \end{pmatrix} \in V_3$. Ezért $V_3 \leq \mathbb{R}^2$ altér. Ugyanezt geometriai érveléssel is megmutathatjuk: V_3 az $x = y$ egyenletű, (origón átmenő) egyenes vektoraiból áll, így bármely két V_3-beli összege, illetve bármely V_3-beli tetszőleges skalárszorosa is nyilván ugyanezen az egyenesen van, vagyis V_3-beli.

d) Például $\underline{v} = \begin{pmatrix} 0 \\ -2 \\ 0 \end{pmatrix} \in V_4$, de $2\underline{v} = \underline{v} + \underline{v} = \begin{pmatrix} 0 \\ -4 \\ 0 \end{pmatrix} \notin V_4$. Ezért V_4 egyik műveletre sem zárt, így nem altér. Indokolhatjuk ezt azzal is, hogy $\underline{0} \notin V_4$, pedig fentebb megmutattuk, hogy $\underline{0}$ minden altérnek eleme. Geometriai érveléssel is látszik, hogy V_4 nem altér: az 1.1.3. pontból tudjuk, hogy V_4 egy S sík vektoraiból (vagyis az origóból a $12x - 23y + 34z = 46$ egyenletű sík pontjaiba mutató vektorokból) áll; mivel S az origót nem tartalmazza, ezért például két V_4-beli összege már nincs S-en.

e) A feladat annyiban különbözik az előzőtől, hogy V_5 már egy origón átmenő S sík vektoraiból áll. A térvektorokra vonatkozó összeadás és skalárral szorzás

(geometriai) definíciójából világos, hogy két S-re illeszkedő vektor összege, illetve bármely S-re illeszkedő vektor minden skalárszorosa is S-beli. Ezért $V_5 \leq \mathbb{R}^3$ altér.

f) $\mathbb{R}^4$-ben a geometriai szemlélet már nyilván nem segíthet, legyen ezért $\underline{u} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \in V_6$ és $\underline{v} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} \in V_6$. Ekkor $\underline{u}+\underline{v} = \begin{pmatrix} x_1+y_1 \\ x_2+y_2 \\ x_3+y_3 \\ x_4+y_4 \end{pmatrix}$. Mivel $\underline{u},\underline{v} \in V_6$, ezért $x_1+x_2+x_3+x_4 = 0$ és $y_1+y_2+y_3+y_4 = 0$. Ezt a két egyenletet összeadva: $(x_1+y_1)+(x_2+y_2)+(x_3+y_3)+(x_4+y_4) = 0$, ami mutatja, hogy $\underline{u}+\underline{v} \in V_6$. Hasonlóan: az $x_1+x_2+x_3+x_4 = 0$ egyenletet λ-val szorozva kapjuk, hogy $\lambda x_1+\lambda x_2+\lambda x_3+\lambda x_4 = 0$, vagyis $\lambda\underline{u} = \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \\ \lambda x_3 \\ \lambda x_4 \end{pmatrix} \in V_6$. Tehát $V_6 \leq \mathbb{R}^4$ altér. □

A fenti feladat c) és e) részének megoldásával analóg módon lehet megmutatni, hogy $\mathbb{R}^2$-ben minden origón átmenő egyenes, $\mathbb{R}^3$-ben pedig minden origón átmenő egyenes és sík alteret határoz meg. Később ki fog derülni, hogy ezeken és a triviális altereken kívül nincs is más altér $\mathbb{R}^2$-ben, illetve $\mathbb{R}^3$-ben.

1.2.3. Generált altér

Ismert középiskolai gyakorlófeladat, hogy két, nem párhuzamos síkvektorból már a sík bármely vektora kifejezhető. Az alábbi állítás ennek a térbeli megfelelőjét is kimondja.

1.2.6. Állítás.
(i) Legyenek $\underline{a},\underline{b} \in \mathbb{R}^3$ nem párhuzamos, az origón átmenő S síkba eső vektorok. Ekkor minden, az S-re illeszkedő $\underline{v} \in \mathbb{R}^3$ vektor kifejezhető $\underline{v} = \alpha\underline{a}+\beta\underline{b}$ alakban.

(ii) Legyenek $\underline{a},\underline{b},\underline{c} \in \mathbb{R}^3$ olyan térvektorok, amelyek nem illeszkednek közös (origón átmenő) síkra. Ekkor minden $\underline{v} \in \mathbb{R}^3$ vektor kifejezhető $\underline{v} = \alpha\underline{a}+\beta\underline{b}+\gamma\underline{c}$ alakban.

Bizonyítás: Legyen $\underline{v} = \overrightarrow{OP}$, ahol O jelöli az origót. *(i)* bizonyításához legyen az O-n átmenő, $\underline{a}$ irányvektorú egyenes e, a P-n átmenő, $\underline{b}$ irányvektorú egyenes pedig f. Mivel e és f nem párhuzamos, közös síkba eső egyenesek, ezért létezik a Q metszéspontjuk. Ekkor $\underline{v} = \overrightarrow{OQ}+\overrightarrow{QP}$ és mivel $\overrightarrow{OQ}$, illetve $\overrightarrow{QP}$ párhuzamosak $\underline{a}$-val, illetve $\underline{b}$-vel, ezért $\overrightarrow{OQ} = \alpha\underline{a}$ és $\overrightarrow{QP} = \beta\underline{b}$ alkalmas $\alpha,\beta \in \mathbb{R}$ skalárokra.

(ii) bizonyításához legyen S az $\underline{a}$ és $\underline{b}$ által kifeszített (O-n átmenő) sík és messe a P-n átmenő, $\underline{c}$ irányvektorú egyenes S-et az R pontban. Ekkor $\underline{v} = \overrightarrow{OR}+\overrightarrow{RP}$, ahol $\overrightarrow{RP} = \gamma\underline{c}$ alkalmas $\gamma \in \mathbb{R}$-re (mert $\underline{c}$ párhuzamos $\overrightarrow{RP}$-vel) és $\overrightarrow{OR} = \alpha\underline{a}+\beta\underline{b}$ alkalmas $\alpha,\beta \in \mathbb{R}$ skalárokra az *(i)* állítást felhasználva. □

A fenti tételben megjelenő $\alpha\underline{a}+\beta\underline{b}+\gamma\underline{c}$ kifejezést általánosítja az alábbi definíció.

1.2.7. Definíció. *Legyenek $\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_k \in \mathbb{R}^n$ vektorok és $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R}$ skalárok. Ekkor a $\lambda_1\underline{v}_1 + \lambda_2\underline{v}_2 + \ldots + \lambda_k\underline{v}_k$ vektort a $\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_k$ vektorok $\lambda_1, \lambda_2, \ldots, \lambda_k$ skalárokkal vett* lineáris kombinációjának *nevezzük.*

A lineáris kombináció tehát nem más, mint hogy néhány adott vektor mindegyikét megszorozzuk egy-egy skalárral és a kapott vektorokat összeadjuk. Lineáris kombináció akár egyetlen vektorból is készíthető: $\underline{v}$-nek $\lambda\underline{v}$ lineáris kombinációja. Sőt, elfogadjuk azt a megállapodást (mert később kényelmes lesz), hogy még az üres halmazból (tehát „nulla darab vektorból") is készíthető lineáris kombináció: ennek eredménye a $\underline{0}$ nullvektor.

Az 1.2.6. Állítás értelmében tehát elmondhatjuk, hogy ha $\underline{a}$ és $\underline{b}$ nem párhuzamos térvektorok, akkor lineáris kombinációjukként – vagyis $\alpha\underline{a} + \beta\underline{b}$ alakban – egy origón átmenő sík vektorai állnak elő (az ugyanis nyilvánvaló, hogy az $\underline{a}$ és $\underline{b}$ által kifeszített S síkon kívül eső vektorok nem fejezhetők ki ilyen alakban). Ha viszont $\underline{a}$ és $\underline{b}$ párhuzamosak, akkor nyilván a mindkettejükkel párhuzamos, origón átmenő egyenes vektorai fejezhetők ki belőlük lineáris kombinációval. Megfigyelhetjük, hogy a lineáris kombinációként kifejezhető vektorok halmaza mindkét esetben altér; ezt az igen fontos megfigyelést általánosítja az alábbi tétel.

1.2.8. Tétel. *Legyenek $\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_k \in \mathbb{R}^n$ tetszőleges, rögzített vektorok. Jelölje W az összes olyan $\mathbb{R}^n$-beli vektor halmazát, amelyek kifejezhetők a $\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_k$ vektorokból lineáris kombinációval. Ekkor W altér $\mathbb{R}^n$-ben.*

Bizonyítás: Az 1.2.4. Definíció szerint azt kell megmutatnunk, hogy W zárt az összeadásra és a skalárral szorzásra és $W \neq \emptyset$. Legyenek tehát $\underline{w}_1, \underline{w}_2 \in W$ tetszőlegesek. Ekkor $\underline{w}_1 = \alpha_1\underline{v}_1 + \alpha_2\underline{v}_2 + \ldots + \alpha_k\underline{v}_k$ és $\underline{w}_2 = \beta_1\underline{v}_1 + \beta_2\underline{v}_2 + \ldots + \beta_k\underline{v}_k$ valamely $\alpha_1, \alpha_2, \ldots, \alpha_k$ és $\beta_1, \beta_2, \ldots, \beta_k$ skalárokra. Ezekből (az 1.2.2. Tétel szerint átrendezés és kiemelés után): $\underline{w}_1 + \underline{w}_2 = (\alpha_1 + \beta_1)\underline{v}_1 + (\alpha_2 + \beta_2)\underline{v}_2 + \ldots + (\alpha_k + \beta_k)\underline{v}_k$. Így $\underline{w}_1 + \underline{w}_2 \in W$ valóban igaz, hiszen $(\underline{w}_1 + \underline{w}_2)$-t kifejeztük a $\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_k$ vektorok lineáris kombinációjaként. Hasonlóan, $\lambda\underline{w}_1 = (\lambda\alpha_1)\underline{v}_1 + (\lambda\alpha_2)\underline{v}_2 + \ldots + (\lambda\alpha_k)\underline{v}_k$, amiből $\lambda\underline{w}_1 \in W$ következik. Kiegészítve ezt azzal, hogy $\underline{0} \in W$ miatt $W \neq \emptyset$ mindig igaz (hiszen a $\underline{0}$ csupa 0 együtthatókkal tetszőleges vektorrendszerből kifejezhető lineráis kombinációval), a tétel bizonyítása teljes. □

A fenti tételben szereplő W altérnek ad nevet az alábbi definíció.

1.2.9. Definíció. *Legyenek $\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_k \in \mathbb{R}^n$ tetszőleges vektorok. Ekkor a $\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_k$ vektorokból lineáris kombinációval kifejezhető $\mathbb{R}^n$-beli vektorok halmazát $\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_k$* generált alterének *nevezzük és $\langle \underline{v}_1, \underline{v}_2, \ldots, \underline{v}_k \rangle$-val jelöljük. Vagyis*

$$\langle \underline{v}_1, \underline{v}_2, \ldots, \underline{v}_k \rangle = \{\lambda_1\underline{v}_1 + \lambda_2\underline{v}_2 + \ldots + \lambda_k\underline{v}_k : \lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R}\}.$$

Ha pedig a $W \leq \mathbb{R}^n$ altérre $W = \langle \underline{v}_1, \underline{v}_2, \ldots, \underline{v}_k \rangle$, akkor a $\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_k$ vektorhalmazt a W altér generátorrendszerének *nevezzük.*

Ezeket a fogalmakat használva tehát például az 1.2.6. Állítás *(ii)* részét úgy is elmondhatjuk, hogy ha $\underline{a},\underline{b},\underline{c} \in \mathbb{R}^3$ közös síkra nem illeszkedő térvektorok, akkor $\langle \underline{a},\underline{b},\underline{c}\rangle = \mathbb{R}^3$, vagyis $\underline{a},\underline{b},\underline{c}$ generált altere $\mathbb{R}^3$ (amely persze maga is altér). Ugyanezt fejezzük ki akkor is, ha azt mondjuk, hogy $\mathbb{R}^3$-nek generátorrendszere $\underline{a},\underline{b},\underline{c}$.

Az 1.2.7. Definíció után rögzített megállapodásnak megfelelően az üres halmaz is generál alteret: $\langle\emptyset\rangle = \{\underline{0}\}$. Egyszerű, de fontos megfigyelés az is, hogy a $\underline{v}_1,\underline{v}_2,\ldots,\underline{v}_k$ vektorok maguk is elemei a $\langle \underline{v}_1,\underline{v}_2,\ldots,\underline{v}_k\rangle$ generált altérnek; valóban, ha $\underline{v}_i$ kivételével mindegyiküket 0, de $\underline{v}_i$-t 1 együtthatóval szorozzuk, épp $\underline{v}_i$-t kapjuk.

1.2.10. Feladat. Határozzuk meg az alábbi generált altereket.

a) $\left\langle \begin{pmatrix}1\\0\\5\end{pmatrix}, \begin{pmatrix}0\\1\\-2\end{pmatrix} \right\rangle$ b) $\left\langle \begin{pmatrix}1\\6\\1\end{pmatrix}, \begin{pmatrix}3\\4\\-1\end{pmatrix} \right\rangle$

c) $\left\langle \begin{pmatrix}1\\6\\1\end{pmatrix}, \begin{pmatrix}3\\4\\-1\end{pmatrix}, \begin{pmatrix}4\\-11\\-6\end{pmatrix} \right\rangle$ d) $\left\langle \begin{pmatrix}1\\-1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\-1\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\\-1\end{pmatrix} \right\rangle$

e) $\left\langle \begin{pmatrix}1\\-1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\-1\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\\-1\end{pmatrix}, \begin{pmatrix}5\\-3\\2\\-4\end{pmatrix} \right\rangle$

Megoldás: a) Írjuk fel a két vektor egy tetszőleges lineáris kombinációját: $\alpha\begin{pmatrix}1\\0\\5\end{pmatrix} + \beta\begin{pmatrix}0\\1\\-2\end{pmatrix} = \begin{pmatrix}\alpha\\\beta\\5\alpha-2\beta\end{pmatrix}$. Nyilván azokat az $\begin{pmatrix}x\\y\\z\end{pmatrix}$ vektorokat fejezhetjük ki ilyen alakban, amelyekre $z = 5x-2y$; ezek alkotják tehát a generált alteret. Látszik, hogy ez épp az $5x-2y-z=0$ egyenletű (origón átmenő) S sík vektoraiból áll. Nem meglepő, hogy a generált altér egy sík, hiszen két nem párhuzamos térvektorból mindig egy sík vektorai fejezhetők ki. Eredményünket kifejezhetjük úgy is, hogy az $5x-2y-z=0$ egyenletű sík vektoraiból álló altérnek generátorrendszere a feladatban megadott két vektor.

b) Ezt a feladatot is megoldhatnánk a fentihez hasonló módon, de egyszerűbben célhoz érünk, ha geometriai alapon gondolkodunk. Ugyanis itt is két nem párhuzamos térvektor generált alterét keressük, az eredmény tehát itt is egy origón átmenő S sík kell legyen. Ha megtaláljuk S-nek egy $\underline{n}$ normálvektorát, akkor felírhatjuk az egyenletét. $\underline{n}$ pedig nyilván merőleges a feladatban megadott két vektorra (hiszen azok S-re illeszkednek). Így $\underline{n}$-et megtalálhatjuk az 1.1.9. Feladatban már látott módszerrel: ha $\underline{n} = (a,b,c)$, akkor (a skaláris szorzatot használva) az $a+6b+c=0$, $3a+4b-c=0$ egyenletrendszer fejezi ki azt, hogy $\underline{n}$ merőleges a feladatbeli vektorokra. A másodikból c-t kifejezve, azt az elsőbe helyettesítve a $4a+10b=0$ egyenletet kapjuk, amelynek például megoldása $a=5$, $b=-2$; így $\underline{n}=(5,-2,7)$ jó normálvektor. S egyenlete tehát (felhasználva, hogy átmegy az origón) $5x-2y+7z=0$; az erre a síkra illeszkedő vektorok alkotják a generált alteret.

c) Jelölje a feladatban megadott vektorokat (sorrendben) $\underline{a}$, $\underline{b}$ és $\underline{c}$. Az $\langle \underline{a}, \underline{b} \rangle$ generált alteret már meghatároztuk és $\underline{c}$ is illeszkedik az eredményként kapott S síkra (mert a koordinátái kielégítik az $5x - 2y + 7z = 0$ egyenletet). Ezért nyilván $\langle \underline{a}, \underline{b}, \underline{c} \rangle = \langle \underline{a}, \underline{b} \rangle$, hiszen az $\alpha\underline{a} + \beta\underline{b} + \gamma\underline{c}$ lineáris kombináció képzésekor csak S-re illeszkedő vektort kaphatunk. Az S vektorai által alkotott altérnek tehát egyaránt generátorrendszere az $\underline{a}, \underline{b}, \underline{c}$ és az $\underline{a}, \underline{b}$ vektorhalmaz is.

d) Itt elég megfigyelnünk, hogy ezt a feladatot valójában már megoldottuk: az 1.2.3. Feladatban épp az itt felsorolt három vektorból kifejezhető vektorokat kellett megtalálnunk; épp ezt a fogalmat neveztük most el generált altérnek (figyelembe véve az 1.2.3. Feladat megoldásának első bekezdését is, amely szerint a „kifejezhetőség" azonos a „lineáris kombinációval kifejezhetőséggel"). Akkori eredményünk szerint a generált alteret azok az $\mathbb{R}^4$-beli vektorok alkotják, amelyekben a négy koordináta összege 0. (Később, az 1.2.5. Feladatban beláttuk, hogy ez valóban altér; ez persze most már következik az 1.2.8. Tételből is.)

e) Jelölje a feladatban megadott vektorokat (sorrendben) $\underline{a}$, $\underline{b}$, $\underline{c}$ és $\underline{d}$, a keresett $\langle \underline{a}, \underline{b}, \underline{c}, \underline{d} \rangle$ generált alteret V, a d) feladatban már meghatározott $\langle \underline{a}, \underline{b}, \underline{c} \rangle$ generált alteret pedig W. V az $\alpha\underline{a} + \beta\underline{b} + \gamma\underline{c} + \delta\underline{d}$ alakban kifejezhető vektorokból áll. Mivel azonban $\underline{a}, \underline{b}, \underline{c} \in W$ mellett $\underline{d} \in W$ is teljesül (hiszen $\underline{d}$ koordinátáinak összege is 0), ezért $\alpha\underline{a} + \beta\underline{b} + \gamma\underline{c} + \delta\underline{d} \in W$ is igaz, hiszen a W altérből az összeadás és a skalárral szorzás (az 1.2.4. Definíció szerint) nem vezet ki. Így a V-beli, vagyis $\alpha\underline{a} + \beta\underline{b} + \gamma\underline{c} + \delta\underline{d}$ alakban felírható vektorok mind W-beliek, másrészt nyilván minden W-belit megkaphatunk így (még $\delta = 0$ választással is). Tehát $V = W$, amit kifejezhetünk úgy is, hogy W-nek generátorrendszere az $\underline{a}, \underline{b}, \underline{c}$ és az $\underline{a}, \underline{b}, \underline{c}, \underline{d}$ vektorrendszer is. □

Az e) feladatból érdemes levonni azt a következtetést, hogy ha egy W altér egy generátorrendszerét kiegészítjük további, W-hez tartozó vektorokkal, akkor ismét W egy generátorrendszerét kapjuk. Ez az állítás általában is igaz, a bizonyítása kiolvasható az iménti megoldásból.

1.2.4. Lineáris függetlenség

Az 1.2.10. Feladat megoldásának egyik tapasztalata, hogy egy W altér egy generátorrendszerében lehetnek „fölösleges" elemek is: az e) feladatban az $\underline{a}, \underline{b}, \underline{c}, \underline{d}$ generátorrendszerből $\underline{d}$-t elhagyva ismét generátorrendszert kapunk. Ennek oka az, hogy $\underline{d}$ kifejezhető az $\underline{a}, \underline{b}, \underline{c}$ vektorokból lineáris kombinációval (hiszen $\underline{d} \in \langle \underline{a}, \underline{b}, \underline{c} \rangle$), így egy az $\underline{a}, \underline{b}, \underline{c}, \underline{d}$ vektorokból készülő lineáris kombinációban $\underline{d}$-t „kiválthatjuk" és az $\underline{a}, \underline{b}, \underline{c}$ vektorokból készült lineáris kombinációt kapunk. (Részletesebben: ha $\underline{d} = \kappa\underline{a} + \lambda\underline{b} + \mu\underline{c}$ és valamely $\underline{v} \in \langle \underline{a}, \underline{b}, \underline{c}, \underline{d} \rangle$ vektorra $\underline{v} = \alpha\underline{a} + \beta\underline{b} + \gamma\underline{c} + \delta\underline{d}$, akkor behelyettesítés és átrendezés után a $\underline{v} = (\alpha + \delta\kappa)\underline{a} + (\beta + \delta\lambda)\underline{b} + (\gamma + \delta\mu)\underline{c}$ alakot kapjuk.) Hasonló a helyzet az 1.2.10. Feladat c) részében is, ahol $\underline{a}, \underline{b}, \underline{c}$ és $\underline{a}, \underline{b}$ ugyanannak az altérnek a generátorrendszerei. Az alábbi definíció épp annak a jelenségnek ad nevet, amikor egy vektorrendszerben nincs ebben az értelemben „fölösleges" vektor.

1.2.11. Definíció. *A $\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_k \in \mathbb{R}^n$ vektorrendszert akkor nevezzük* lineárisan függetlennek*, ha a $\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_k$ vektorok közül semelyik sem fejezhető ki a többi lineáris kombinációjaként. Ha ez nem teljesül – vagyis a $\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_k$ vektorok között van legalább egy olyan, ami kifejezhető a többi lineáris kombinációjaként –, akkor a $\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_k$ vektorrendszert* lineárisan összefüggőnek *nevezzük.*

Az 1.2.7. Definíció után mondottakkal összhangban elfogadjuk azt a megállapodást, hogy az üres halmaz (vagyis a „nulla darab vektorból álló rendszer") lineárisan független. A $k = 1$ esetben a definíciót úgy kell értelmezni, hogy az egyetlen $\underline{v}_1$ vektorból álló rendszer akkor lineárisan független, ha $\underline{v}_1 \neq \underline{0}$; ha viszont $\underline{v}_1 = \underline{0}$, akkor $\underline{v}_1$ kifejezhető a „többi" vektorból és így lineárisan összefüggő (mert $\underline{0}$ még az üres halmazból is kifejezhető lineáris kombinációval). A definícióból az is következik, hogy lineárisan összefüggő minden olyan vektorrendszer, amely a nullvektort tartalmazza (hiszen a $\underline{0}$ bármely más vektorból kifejezhető annak nullaszorosaként).

Legyenek például $\underline{a}, \underline{b}, \underline{c}$ olyan térvektorok, amelyek nem illeszkednek közös, origón átmenő síkra. Ekkor az $\langle \underline{a}, \underline{b} \rangle$ generált altér egy origón átmenő sík, így $\underline{c}$ nincs benne. Ezért $\underline{c}$ nem fejezhető ki az $\underline{a}, \underline{b}$ vektorokból lineáris kombinációval – és mivel ugyanez elmondható $\underline{a}$ és $\langle \underline{b}, \underline{c} \rangle$, valamint $\underline{b}$ és $\langle \underline{a}, \underline{c} \rangle$ viszonyában is, ezért $\underline{a}, \underline{b}, \underline{c}$ lineárisan függetlenek. Ha viszont az $\underline{a}, \underline{b}, \underline{c}$ térvektorok illeszkednek egy közös, origón átmenő síkra, akkor biztosan lineárisan összefüggők: ha van köztük két nem párhuzamos, akkor ezekből az 1.2.6. Tétel szerint a harmadik kifejezhető; ha mindhárman párhuzamosak akkor vagy van köztük nullvektor vagy bármelyikük skalárszorosa bármelyik másiknak.

A definícióból azonnal következik, de hasznos megemlíteni, hogy lineárisan független rendszer bármely részhalmaza is lineárisan független. Valóban: ha a részhalmaz egyik vektora kifejezhető volna a többi lineáris kombinációjaként, akkor a bővebb rendszerből a részhalmazba be nem került vektorokat 0 együtthatóval a lineáris kombinációhoz fűzve azt az ellentmondást kapnánk, hogy a bővebb rendszer egyik vektora is előáll a többi lineáris kombinációjaként.

Fontos kiemelni, mert gyakori félreértés forrása, hogy az „a $\underline{v}_1, \ldots, \underline{v}_k$ vektorok lineárisan függetlenek" kijelentés nem külön-külön minősíti a $\underline{v}_1, \ldots, \underline{v}_k$ vektorokat, hanem a k darab vektor által alkotott rendszer egészének egy tulajdonságát mondja ki. (Ezt jobban megvilágítja a következő példa. Az „ezek a zoknik koszosak" kijelentés külön-külön minősíti a zoknikat: az egyik is koszos meg a másik is. Ezzel szemben az „ezek a zoknik nem illenek össze" kijelentés már nem azt mondja, hogy „az egyik sem illik össze és a másik sem", hanem a zoknikra együttesen vonatkozik. A lineáris függetlenség tehát az utóbbi esetre hasonlít.) Ennek a félreértésnek egy másik változata az a tévhit, amely szerint a $\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_k$ vektorok lineárisan függetlensége azonos volna azzal, hogy közülük bármely kettő lineárisan független. Ez nem így van: ha például $\underline{a}, \underline{b}, \underline{c}$ közös origón átmenő síkra illeszkedő térvektorok, de nincs köztük két párhuzamos, akkor közülük bármely kettő lineárisan független (hiszen egyik sem skalárszorosa a másiknak), de $\underline{a}, \underline{b}, \underline{c}$ mégis lineárisan összefüggő.

Ha egy adott $\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_k$ vektorrendszerről el kell döntenünk, hogy az lineárisan független-e, akkor az 1.2.11. Definíció értelmében k darab külön ellenőrzést

kell végeznünk: mindegyik $\underline{v}_i$-ről ki kell derítenünk, hogy a többiből kifejezhető-e. A következő tétel azért nagyon hasznos, mert lehetővé teszi, hogy k helyett egyetlen ellenőrzéssel eldöntsük a kérdést.

1.2.12. Tétel. *A $\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_k \in \mathbb{R}^n$ vektorrendszer akkor és csak akkor lineárisan független, ha a $\lambda_1\underline{v}_1 + \lambda_2\underline{v}_2 + \ldots + \lambda_k\underline{v}_k = \underline{0}$ egyenlőség kizárólag abban az esetben teljesül, ha $\lambda_1 = \lambda_2 = \ldots = \lambda_k = 0$.*

Bizonyítás: Kezdjük az „akkor” állítással: tegyük fel, hogy $\lambda_1\underline{v}_1 + \ldots + \lambda_k\underline{v}_k = \underline{0}$ csak a $\lambda_1 = \ldots = \lambda_k = 0$ esetben teljesül; belátjuk, hogy ekkor $\underline{v}_1, \ldots, \underline{v}_k$ lineárisan független. Indirekt bizonyítunk: feltesszük, hogy $\underline{v}_1, \ldots, \underline{v}_k$ mégsem lineárisan független és ebből ellentmondásra jutunk. Ha ugyanis $\underline{v}_1, \ldots, \underline{v}_k$ nem lineárisan független, akkor valamelyikük kifejezhető a többiből lineáris kombinációval; legyen ez például $\underline{v}_1$ (hiszen a vektorok indexelése tetszőleges). Ekkor $\underline{v}_1 = \alpha_2\underline{v}_2 + \alpha_3\underline{v}_3 + \ldots + \alpha_k\underline{v}_k$ valamely $\alpha_2, \alpha_3, \ldots, \alpha_k$ együtthatókra. Átrendezve: $1\underline{v}_1 - \alpha_2\underline{v}_2 - \alpha_3\underline{v}_3 - \ldots - \alpha_k\underline{v}_k = \underline{0}$. Ezzel sikerült ellentmondásra jutnunk: $\lambda_1\underline{v}_1 + \ldots + \lambda_k\underline{v}_k = \underline{0}$ megvalósulhat a $\lambda_1 = 1$, $\lambda_2 = -\alpha_2, \ldots, \lambda_k = -\alpha_k$ választással is, vagyis nem csak csupa 0 együtthatóval.

Most belátjuk a tétel „csak akkor” állítását: feltesszük, hogy $\underline{v}_1, \ldots, \underline{v}_k$ lineárisan független és megmutatjuk, hogy ekkor $\lambda_1\underline{v}_1 + \ldots + \lambda_k\underline{v}_k = \underline{0}$ csak a $\lambda_1 = \ldots = \lambda_k = 0$ esetben teljesül. Ismét indirekt bizonyítunk: tegyük fel, hogy $\lambda_1\underline{v}_1 + \ldots + \lambda_k\underline{v}_k = \underline{0}$, de a λ_i-k között van nemnulla; például $\lambda_1 \neq 0$. Ekkor átrendezés és $\lambda_1 \neq 0$-val való osztás után a

$$\underline{v}_1 = -\frac{\lambda_2}{\lambda_1}\underline{v}_2 - \frac{\lambda_3}{\lambda_1}\underline{v}_3 - \ldots - \frac{\lambda_k}{\lambda_1}\underline{v}_k$$

alakot kapjuk. Ez ellentmondás: $\underline{v}_1, \ldots, \underline{v}_k$ mégsem lineárisan független, mert $\underline{v}_1$ kifejezhető a többiből lineáris kombinációval. □

A tétel állítását a következőképp is szokták fogalmazni: $\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_k$ akkor és csak akkor lineárisan független, ha csak a *triviális lineáris kombinációjuk* adja a nullvektort. Itt a triviális jelző arra utal, hogy a lineáris kombinációban minden együttható 0. Értelemszerűen adódik a tételből, hogy $\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_k$ akkor és csak akkor lineárisan összefüggő, ha a triviális lineáris kombináción kívül más lehetőség is van belőlük a $\underline{0}$ kifejezésére; más szóval, ha $\lambda_1\underline{v}_1 + \lambda_2\underline{v}_2 + \ldots + \lambda_k\underline{v}_k = \underline{0}$ úgy is megvalósulhat, hogy a λ_i-k között van 0-tól különböző.

Az 1.2.12. Tétel állítása alkalmas lehetett volna akár arra is, hogy így definiáljuk a lineáris függetlenség fogalmát. Számos tankönyv így is tesz (és az 1.2.11. Definícióbeli feltételt mondja ki tételként), mert az 1.2.12. Tétel segítségével a gyakorlatban valóban sokszor könnyebb a lineáris függetlenség eldöntése. Valójában mindegy, melyik utat járjuk; ebben a jegyzetben azért a fenti definíciót választottuk, mert így talán jobban érthető a fogalom megalkotásának motivációja.

1.2.13. Feladat. Lineárisan függetlenek-e az alábbi vektorrendszerek?

a) $\begin{pmatrix}1\\2\\2\\0\end{pmatrix}, \begin{pmatrix}1\\2\\2\\5\end{pmatrix}, \begin{pmatrix}0\\0\\3\\1\end{pmatrix}, \begin{pmatrix}0\\4\\0\\1\end{pmatrix}$ b) $\begin{pmatrix}1\\2\\2\\0\end{pmatrix}, \begin{pmatrix}1\\2\\2\\5\end{pmatrix}, \begin{pmatrix}0\\0\\3\\1\end{pmatrix}, \begin{pmatrix}2\\4\\0\\1\end{pmatrix}$

Megoldás: a) Jelölje a vektorokat sorban $\underline{a}$, $\underline{b}$, $\underline{c}$ és $\underline{d}$. Az 1.2.12. Tételt használjuk, vagyis arra vagyunk kíváncsiak, hogy az $\alpha\underline{a}+\beta\underline{b}+\gamma\underline{c}+\delta\underline{d}=\underline{0}$ egyenletnek van-e más megoldása, mint $\alpha=\beta=\gamma=\delta=0$. Behelyettesítve a vektorokat és elvégezve a műveleteket: $\alpha\underline{a}+\beta\underline{b}+\gamma\underline{c}+\delta\underline{d}=\begin{pmatrix}\alpha+\beta\\2\alpha+2\beta+4\delta\\2\alpha+2\beta+3\gamma\\5\beta+\gamma+\delta\end{pmatrix}$. Ebből következik, hogy az $\alpha\underline{a}+\beta\underline{b}+\gamma\underline{c}+\delta\underline{d}=\underline{0}$ egyenlet az

$$\begin{aligned}\alpha+\beta&=0,\\2\alpha+2\beta+4\delta&=0,\\2\alpha+2\beta+3\gamma&=0,\\5\beta+\gamma+\delta&=0\end{aligned}$$

lineáris egyenletrendszerre vezet. A második, illetve a harmadik egyenletből az első kétszeresét levonva kapjuk, hogy $4\delta=0$ és $3\gamma=0$; így $\gamma=\delta=0$. Ezeket a negyedik egyenletbe helyettesítve $\beta=0$, amit az elsőbe helyettesítve $\alpha=0$ adódik. Tehát $\alpha\underline{a}+\beta\underline{b}+\gamma\underline{c}+\delta\underline{d}=\underline{0}$ csak $\alpha=\beta=\gamma=\delta=0$ esetén valósulhat meg, így a négy vektor lineárisan független.

b) A megoldás gondolatmenete azonos az előzővel, a kapott egyenletrendszerben csak az első egyenlet változik: $\alpha+\beta+2\delta=0$. Mivel ennek az egyenletnek épp kétszerese a második, ezért az egyikük elhagyható, az „információtartalmuk” azonos. A megmaradt, három egyenletből álló rendszerben az első két egyenlet különbségéből $3\gamma=4\delta$ adódik. Próbálkozzunk ezért például a $\gamma=4$, $\delta=3$ értékválasztással, ami ezt teljesíti. Ekkor az utolsó egyenletből $\beta=-\frac{7}{5}$ adódik, amit az első kettő bármelyikébe visszahelyettesítve az $\alpha=-\frac{23}{5}$ értéket kapjuk. Ezzel tehát megkaptuk az egyenletrendszer egy lehetséges, nem csupa nulla megoldását (a végtelen sokból). Következésképp a négy vektor lineárisan összefüggő, amit mutat például a $-\frac{23}{5}\underline{a}-\frac{7}{5}\underline{b}+4\underline{c}+3\underline{d}=\underline{0}$ összefüggés. □

Az „újonnan érkező vektor lemmája”

Tegyük fel, hogy egy vektorrendszer lineárisan független, de „érkezik” egy további vektor, amellyel együtt már lineárisan összefüggővé válik. Definíció szerint ekkor van legalább egy olyan a vektorok között, ami kifejezhető a többi lineáris kombinációjaként – az azonban közvetlenül a definícióból már nem derül ki, hogy ez éppen az újonnan érkezett vektorra is teljesül-e. Az alábbi egyszerű, de igen hasznos állítás azt mondja ki, hogy ez mégis igaz.

1.2.14. Lemma. (Az újonnan érkező vektor lemmája.)
Tegyük fel, hogy az $\underline{f}_1, \underline{f}_2, \ldots, \underline{f}_k$ rendszer lineárisan független, de $\underline{f}_1, \underline{f}_2, \ldots, \underline{f}_k, \underline{f}_{k+1}$ lineárisan összefüggő. Ekkor $\underline{f}_{k+1} \in \langle \underline{f}_1, \underline{f}_2, \ldots, \underline{f}_k \rangle$ (vagyis $\underline{f}_{k+1}$ kifejezhető az $\underline{f}_1, \underline{f}_2, \ldots, \underline{f}_k$ lineáris kombinációjaként).

Bizonyítás: Mivel $\underline{f}_1, \underline{f}_2, \ldots, \underline{f}_k, \underline{f}_{k+1}$ lineárisan összefüggő, ezért az 1.2.12. Tétel értelmében léteznek olyan $\lambda_1, \lambda_2, \ldots, \lambda_k, \lambda_{k+1}$ skalárok, amelyek közül nem mind nulla és $\lambda_1 \underline{f}_1 + \ldots + \lambda_k \underline{f}_k + \lambda_{k+1} \underline{f}_{k+1} = \underline{0}$. Ha itt $\lambda_{k+1} = 0$ teljesülne, akkor azt kapnánk, hogy $\lambda_1 \underline{f}_1 + \ldots + \lambda_k \underline{f}_k = \underline{0}$ és a $\lambda_1, \lambda_2, \ldots, \lambda_k$ skalárok között van nemnulla. Ez (ismét az 1.2.12. Tétel szerint) ellentmondana annak, hogy $\underline{f}_1, \underline{f}_2, \ldots, \underline{f}_k$ lineárisan független. Következésképp $\lambda_{k+1} \neq 0$, amiből átrendezés és λ_{k+1}-gyel való osztás után az

$$\underline{f}_{k+1} = -\frac{\lambda_1}{\lambda_{k+1}} \underline{f}_1 - \frac{\lambda_2}{\lambda_{k+1}} \underline{f}_2 - \ldots - \frac{\lambda_k}{\lambda_{k+1}} \underline{f}_k$$

összefüggést kapjuk. Vagyis $\underline{f}_{k+1} \in \langle \underline{f}_1, \underline{f}_2, \ldots, \underline{f}_k \rangle$ valóban igaz. □

1.2.5. Az F-G egyenlőtlenség

Megismertünk két alapvető fogalmat: a V altér generátorrendszerének, illetve a lineárisan független rendszernek a fogalmát. Láttuk például, hogy $\mathbb{R}^3$-ben három vektor már alkothat generátorrendszert, de kettő még nyilván nem; láttuk azt is, hogy $\mathbb{R}^3$-ben három vektor még alkothat lineárisan független rendszert, de könnyű meggondolni, hogy négy már nem. Az alábbi tétel azt mondja ki, hogy ez a megfigyelés általában is érvényes: minden altérben igaz, hogy abban bármely lineárisan független rendszer legföljebb annyi vektorból áll, mint az altér bármely generátorrendszere. A tétel a lineáris algebra elméletében alapvető fontosságú, a továbbiakban „F-G egyenlőtlenségként” hivatkozunk majd rá.

1.2.15. Tétel. (F-G egyenlőtlenség)
Legyen $V \leq \mathbb{R}^n$ altér, $\underline{f}_1, \underline{f}_2, \ldots, \underline{f}_k$ V-beli vektorokból álló lineárisan független rendszer, $\underline{g}_1, \underline{g}_2, \ldots, \underline{g}_m$ pedig generátorrendszer V-ben. Ekkor $k \leq m$.

Bizonyítás: A bizonyítás fő eszköze az alábbi segédtétel lesz, amely azt állítja, hogy minden V-beli lineárisan független rendszer bármely $\underline{f}$ eleme „kicserélhető” V egy tetszőleges generátorrendszerének egy alkalmasan választott $\underline{g}$ elemére úgy, hogy a csere után a rendszer lineáris függetlensége megmaradjon; itt „csere” alatt nyilván azt értjük, hogy $\underline{f}$-et kidobjuk a rendszerből és a helyére $\underline{g}$-t tesszük.

1.2.16. Lemma. (Kicserélési lemma)
Legyen $V \leq \mathbb{R}^n$ altér, $\underline{f}_1, \underline{f}_2, \ldots, \underline{f}_k$ V-beli vektorokból álló lineárisan független rendszer, $\underline{g}_1, \underline{g}_2, \ldots, \underline{g}_m$ pedig generátorrendszer V-ben. Ekkor minden $i \leq k$ esetén található olyan $j \leq m$, hogy az $\underline{f}_1, \underline{f}_2, \ldots, \underline{f}_{i-1}, \underline{g}_j, \underline{f}_{i+1}, \ldots, \underline{f}_k$ vektorrendszer szintén lineárisan független.

A Lemma bizonyítása: Mivel az $\underline{f}_i$ vektorok számozása érdektelen, a bizonyítás leírásának egyszerűsítése érdekében feltehetjük, hogy $i = k$; vagyis az $\underline{f}_k$-t szeretnénk kicserélni egy alkalmas $\underline{g}_j$ vektorra.

Próbálkozzunk először a $\underline{g}_1$-gyel: ha véletlen az $\underline{f}_1, \ldots, \underline{f}_{k-1}, \underline{g}_1$ rendszer lineárisan független, akkor kész is vagyunk a bizonyítással. Ha viszont nem az, akkor az „újonnan érkező vektor" 1.2.14. lemmája miatt $\underline{g}_1 \in \langle \underline{f}_1, \ldots, \underline{f}_{k-1} \rangle$ kell teljesüljön. Valóban: az $\underline{f}_1, \ldots, \underline{f}_{k-1}$ rendszer még lineárisan független (hiszen a lineárisan független $\underline{f}_1, \ldots, \underline{f}_k$ rendszer része), de a $\underline{g}_1$ „érkezése" lineárisan összefüggővé teszi.

A fenti bekezdés gondolatmenetét $\underline{g}_1$ helyett persze bármelyik $\underline{g}_j$ vektorra megismételhetjük. Így ha a lemma állításával ellentétben az $\underline{f}_k$ semelyik $\underline{g}_j$-re nem volna kicserélhető, akkor ebből az következne, hogy a $\underline{g}_1, \underline{g}_2, \ldots, \underline{g}_m$ vektorok mindegyike benne van az $\langle \underline{f}_1, \ldots, \underline{f}_{k-1} \rangle$ generált altérben.

Mivel azonban $\langle \underline{f}_1, \underline{f}_2, \ldots, \underline{f}_{k-1} \rangle$ altér, ezért definíció szerint zárt az összeadásra és a skalárral való szorzásra. Így a $\underline{g}_1, \underline{g}_2, \ldots, \underline{g}_m$ vektorokkal együtt ezek minden lineáris kombinációja is benne van $\langle \underline{f}_1, \underline{f}_2, \ldots, \underline{f}_{k-1} \rangle$-ben. Ebből viszont következik $\underline{f}_k \in \langle \underline{f}_1, \underline{f}_2, \ldots, \underline{f}_{k-1} \rangle$ is – hiszen $\underline{g}_1, \underline{g}_2, \ldots, \underline{g}_m$ generátorrendszer, így minden V-beli vektor, közöttük $\underline{f}_k$ is felírható a lineáris kombinációjukként.

Ezzel ellentmondásra jutottunk: $\underline{f}_k \in \langle \underline{f}_1, \underline{f}_2, \ldots, \underline{f}_{k-1} \rangle$ azt jelenti, hogy $\underline{f}_k$ kifejezhető az $\underline{f}_1, \underline{f}_2, \ldots, \underline{f}_{k-1}$ vektorok lineáris kombinációjaként, ami ellentmond a lineáris függetlenség 1.2.11. definíciójának. Ez az ellentmondás tehát bizonyítja, hogy az indirekt feltevés hamis volt, legalább egy $\underline{g}_j$-vel a csere működni fog. ◇

A Kicserélési lemma ismételt alkalmazásával az F-G egyenlőtlenség állítása már könnyen belátható. Először alkalmazzuk a lemmát $\underline{f}_1$-re: kapjuk a $\underline{g}_j, \underline{f}_2, \underline{f}_3, \ldots, \underline{f}_k$ rendszert (valamilyen $j \leq m$-re). Mivel ez a rendszer továbbra is lineárisan független, rá és a (változatlan) $\underline{g}_1, \underline{g}_2, \ldots, \underline{g}_m$ generátorrendszerre megint alkalmazhatjuk a Kicserélési lemmát: kapjuk a $\underline{g}_j, \underline{g}_\ell, \underline{f}_3, \ldots, \underline{f}_k$ lineárisan független rendszert. Ezt az eljárást folytatva tehát karban tartunk egy állandóan változó lineárisan független rendszert, amelynek az elemszáma a cserék miatt végig változatlan. Az eljárás folytathatóságát mindig a Kicserélési lemma garantálja, amely szerint a csere után a lineáris függetlenség megmarad. Az eljárás nyilván akkor áll le, ha a cserék eredményeképpen kapott lineárisan független rendszer már teljes egészében része a $\underline{g}_1, \underline{g}_2, \ldots, \underline{g}_m$ rendszernek.

Összefoglalva: az eljárás végén kapunk egy k elemű lineárisan független rendszert, amelynek vektorai az (eredeti) $\underline{g}_1, \underline{g}_2, \ldots, \underline{g}_m$ generátorrendszer tagjai közül kerülnek ki. Eközött a k darab vektor között ráadásul nem lehetnek azonosak, hiszen ez a lineáris függetlenség definíciójának ellentmondana (mert két azonos vektor közül az egyik kifejezhető volna a másik 1-szeresének és az összes többi 0-szorosának összegeként). Mivel tehát a $\underline{g}_1, \underline{g}_2, \ldots, \underline{g}_m$ generátorrendszer tagjai közül kiválasztható k darab csupa különböző, ebből $k \leq m$ nyilván következik. □

Az F-G egyenlőtlenség fontosságára tekintettel (és az érdekesség kedvéért) mutatunk a tételre egy másik, teljes indukción alapuló bizonyítást is.

Az 1.2.15. tétel egy alternatív bizonyítása: k-ra vonatkozó teljes indukcióval bizonyítunk. Érdemes ezért a tételt a következő módon átfogalmazni: „ha egy $V \leq \mathbb{R}^n$ altérben van k elemű lineárisan független rendszer, akkor V minden generátorrendszere legalább k elemű". (Ez nyilván valóban ekvivalens a tétel állításával, de a teljes indukció működését követhetőbbé teszi.)

Ha $k = 1$, akkor V-ben van a nullvektortól különböző vektor (mert $\underline{f}_1 \neq \underline{0}$), így valóban minden generátorrendszere legalább 1 elemű (mert az üres halmaz a $\{\underline{0}\}$ alteret generálná). A tétel tehát a $k = 1$ esetben igaz. Legyen ezért a továbbiakban $k \geq 2$ és tegyük fel, hogy a tétel $(k-1)$-re már igaz; célunk belátni, hogy ekkor k-ra is az. Ehhez meg fogunk adni egy $W \leq \mathbb{R}^n$ alteret, ami tartalmaz $k-1$ elemű lineárisan független rendszert és $m-1$ elemű generátorrendszert; mivel ebből az indukciós feltevés szerint $k-1 \leq m-1$ következni fog, ez bizonyítja a $k \leq m$ állítást.

Mivel $\underline{g}_1, \ldots, \underline{g}_m$ generátorrendszer V-ben, ezért minden V-beli vektor, így $\underline{f}_k$ is előáll a $\underline{g}_1, \underline{g}_2, \ldots, \underline{g}_m$ lineáris kombinációjaként: $\underline{f}_k = \gamma_1 \underline{g}_1 + \gamma_2 \underline{g}_2 + \ldots + \gamma_m \underline{g}_m$. A γ_i-k között kell legyen nemnulla (mert $\underline{f}_k \neq \underline{0}$). Legyen például $\gamma_m \neq 0$ és legyen $W = \langle \underline{g}_1, \underline{g}_2, \ldots, \underline{g}_{m-1} \rangle$. Megmutatjuk, hogy minden $1 \leq j \leq k-1$ esetén az $\underline{f}_j$-hez található olyan α_j skalár, hogy $\underline{f}_j + \alpha_j \underline{f}_k \in W$. Ugyanis $\underline{f}_j$ is felírható $\underline{g}_1, \underline{g}_2, \ldots, \underline{g}_m$ lineáris kombinációjaként: $\underline{f}_j = \beta_1 \underline{g}_1 + \beta_2 \underline{g}_2 + \ldots + \beta_m \underline{g}_m$. Ekkor $\alpha_j = -\frac{\beta_m}{\gamma_m}$ megfelel a célnak:

$$\underline{f}_j + \alpha_j \underline{f}_k = (\beta_1 - \frac{\beta_m}{\gamma_m}\gamma_1)\underline{g}_1 + (\beta_2 - \frac{\beta_m}{\gamma_m}\gamma_2)\underline{g}_2 + \ldots + (\beta_m - \frac{\beta_m}{\gamma_m}\gamma_m)\underline{g}_m,$$

ahol tehát $\underline{g}_m$ együtthatója $\beta_m - \frac{\beta_m}{\gamma_m}\gamma_m = 0$, így $\underline{f}_j + \alpha_j \underline{f}_k$ valóban W-beli (mert felírható $\underline{g}_1, \underline{g}_2, \ldots, \underline{g}_{m-1}$ lineáris kombinációjaként).

Most megmutatjuk, hogy az $\underline{f}_j + \alpha_j \underline{f}_k$, $j = 1, 2, \ldots, k-1$ vektorok lineárisan függetlenek (ahol az α_j skalárokat az előző bekezdésben írtak szerint választottuk). Vegyük ugyanis egy $\underline{0}$-t adó lineáris kombinációjukat a $\lambda_1, \lambda_2, \ldots, \lambda_{k-1}$ skalárokkal:

$$\lambda_1(\underline{f}_1 + \alpha_1 \underline{f}_k) + \lambda_2(\underline{f}_2 + \alpha_2 \underline{f}_k) + \ldots + \lambda_{k-1}(\underline{f}_{k-1} + \alpha_{k-1}\underline{f}_k) = \underline{0}.$$

Átrendezve:

$$\lambda_1 \underline{f}_1 + \lambda_2 \underline{f}_2 + \ldots + \lambda_{k-1}\underline{f}_{k-1} + (\lambda_1\alpha_1 + \lambda_2\alpha_2 + \ldots + \lambda_{k-1}\alpha_{k-1})\underline{f}_k = \underline{0}.$$

Ezzel az $\underline{f}_1, \underline{f}_2, \ldots, \underline{f}_k$ egy $\underline{0}$-t adó lineáris kombinációját kaptuk; mivel azonban ezekről tudjuk, hogy lineárisan függetlenek, ezért (az 1.2.12. Tétel szerint) a lineáris kombináció minden együtthatója 0 kell legyen. Vagyis $\lambda_1 = \lambda_2 = \ldots = \lambda_{k-1} = 0$ (és mellesleg $\lambda_1\alpha_1 + \lambda_2\alpha_2 + \ldots + \lambda_{k-1}\alpha_{k-1} = 0$ is igaz). Így (ismét az 1.2.12. Tétel miatt) az $\underline{f}_j + \alpha_j \underline{f}_k$ vektorok valóban lineárisan függetlenek.

Összefoglalva: a W altérben $\underline{f}_1 + \alpha_1 \underline{f}_k, \underline{f}_2 + \alpha_2 \underline{f}_k, \ldots, \underline{f}_{k-1} + \alpha_{k-1}\underline{f}_k$ lineárisan független, $\underline{g}_1, \underline{g}_2, \ldots, \underline{g}_{m-1}$ generátorrendszert alkot. Így az indukciós feltevésből $k-1 \leq m-1$ következik, amivel tehát a $k \leq m$ állítást beláttuk. □

1.2.6. Bázis, dimenzió

Az $\mathbb{R}^3$ térben a közös origón átmenő síkra nem illeszkedő vektorhármasok speciális szereppel bírnak: egyszerre igaz rájuk, hogy lineárisan függetlenek és generátor-

rendszert alkotnak. Általános $V \leq \mathbb{R}^n$ alterek esetén az ilyen tulajdonságú vektorrendszereknek ad nevet az alábbi definíció. Ugyancsak ebben a pontban általánosítjuk azt a jelenséget, hogy $\mathbb{R}^3$-ben csak pontosan három darab vektorból álló rendszer rendelkezhet egyszerre ezzel a két tulajdonsággal.

1.2.17. Definíció. *Legyen $V \leq \mathbb{R}^n$ altér. A V-beli vektorokból álló $\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_k$ rendszert* bázisnak *nevezzük V-ben, ha a rendszer lineárisan független és generátorrendszer V-ben.*

A fentiek szerint tehát $\underline{a}, \underline{b}, \underline{c}$ bázis $\mathbb{R}^3$-ben, ha nem illeszkednek közös origón átmenő síkra. $\mathbb{R}^2$-ben bázist alkot bármely két, nem párhuzamos vektor: lineárisan függetlenek (hiszen egyikből sem fejezhető ki a másik) és generátorrendszert alkotnak $\mathbb{R}^2$-ben (mert minden $\mathbb{R}^2$-beli vektor kifejezhető a lineáris kombinációjukként).

1.2.18. Feladat. Álljon V azokból az $\mathbb{R}^4$-beli vektorokból, amelyekben a négy koordináta összege 0. Adjunk meg egy bázist V-ben.

Megoldás: Az 1.2.5. Feladat f) részében beláttuk, hogy V valóban altér. (Ezt azért fontos megemlíteni, mert bázisa csak altérnek lehet, tetszőleges $\mathbb{R}^n$-beli részhalmaznak nem). Az 1.2.10. Feladat d) részében (illetve korábban az 1.2.3. Feladatban) megmutattuk, hogy az ott megadott három vektor – jelölje ezeket sorban $\underline{a}$, $\underline{b}$ és $\underline{c}$ – generátorrendszer V-ben. Most az 1.2.13. Feladatban már látott módszerrel megmutatjuk, hogy $\underline{a}, \underline{b}, \underline{c}$ lineárisan független. Vegyük ugyanis ezeknek egy $\underline{0}$-t adó lineáris kombinációját: $\alpha\underline{a} + \beta\underline{b} + \gamma\underline{c} = \underline{0}$. Behelyettesítve $\underline{a}, \underline{b}, \underline{c}$ értékét az $\alpha = 0$, $-\alpha + \beta = 0$, $-\beta + \gamma = 0$, $-\gamma = 0$ lineáris egyenletrendszert kapjuk. Az első és a harmadik egyenletből $\alpha = \gamma = 0$, ezt használva mindkét másik egyenletből $\beta = 0$ adódik. Így az 1.2.12. Tétel szerint $\underline{a}, \underline{b}, \underline{c}$ valóban lineárisan független. Összefoglalva: $\underline{a}, \underline{b}, \underline{c}$ lineárisan független és generátorrendszer V-ben, így bázis. (Természetesen nem ez az egyetlen megoldás, V-ben végtelen sok különböző bázis létezik.) □

1.2.19. Tétel. *Tegyük fel, hogy a $V \leq \mathbb{R}^n$ altérben a $\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_k$ rendszer és a $\underline{c}_1, \underline{c}_2, \ldots, \underline{c}_m$ rendszer egyaránt bázisok. Ekkor $k = m$.*

Bizonyítás: Mindkét rendszer bázis, ezért a $\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_k$ rendszer lineárisan független és $\underline{c}_1, \underline{c}_2, \ldots, \underline{c}_m$ generátorrendszer V-ben. Alkalmazva az 1.2.15. F-G egyenlőtlenséget: $k \leq m$. Ugyanezt fordított szereposztásban is elmondhatjuk: $\underline{c}_1, \underline{c}_2, \ldots, \underline{c}_m$ lineárisan független és $\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_k$ generátorrendszer V-ben, ezért $m \leq k$. Ezekből tehát $k = m$ valóban következik. □

A $V \leq \mathbb{R}^n$ alterekre jellemző tulajdonság tehát, hogy a bázisaik hány vektorból állnak – ugyanis ez a szám bármely két bázisra azonos. Ennek ad nevet az alábbi definíció.

1.2.20. Definíció. *Legyen a $V \leq \mathbb{R}^n$ altérben a $\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_k$ rendszer bázis. Ekkor azt mondjuk, hogy a V* dimenziója *k. Ezt a következőképp jelöljük:* $\dim V = k$.

Még egyszer kiemeljük, hogy a dimenzió fogalmának az 1.2.19. Tétel ad létjogosultságot: egyetlen altérnek sem lehet „két különböző dimenziója". Például most már elmondhatjuk, hogy $\mathbb{R}^3$ dimenziója három – ehhez elég visszaidéznünk az $\mathbb{R}^3$-beli bázisokról az 1.2.17. Definíció után mondottakat. Ugyanígy egybevág a korábbi „naív" szóhasználatunkkal az a tény, hogy $\mathbb{R}^2$ dimenziója kettő, hiszen a nem párhuzamos vektorpárok alkotnak benne bázist.

Standard bázis

Az alábbi állításból pedig azt fog következni, hogy $\mathbb{R}^n$ dimenziója – a várakozásainkkal összhangban – valóban n.

1.2.21. Állítás. *Jelölje minden $1 \leq i \leq n$ esetén $\underline{e}_i$ azt az $\mathbb{R}^n$-beli vektort, amelynek (fölülről) az i-edik koordinátája 1, az összes többi koordinátája 0. Ekkor $\underline{e}_1, \underline{e}_2, \ldots, \underline{e}_n$ bázis $\mathbb{R}^n$-ben.*

Bizonyítás: Készítsük el az $\underline{e}_1, \underline{e}_2, \ldots, \underline{e}_n$ vektorok egy tetszőleges lineáris kombinációját az $\alpha_1, \alpha_2, \ldots, \alpha_n$ skalárokkal:

$$\alpha_1\underline{e}_1 + \alpha_2\underline{e}_2 + \ldots + \alpha_n\underline{e}_n = \alpha_1 \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \alpha_2 \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \ldots + \alpha_n \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \vdots \\ \alpha_n \end{pmatrix}$$

Azonnal látszik, hogy $\underline{e}_1, \underline{e}_2, \ldots, \underline{e}_n$ generátorrendszer $\mathbb{R}^n$-ben, hiszen a lineáris kombinációjukként egy tetszőleges vektor előállhat: ha egy adott $\underline{v}$ vektort szeretnénk belőlük kifejezni, akkor $\underline{v}$ i-edik koordinátáját választjuk $\underline{e}_i$ együtthatójának minden i-re – és ez egyben az egyetlen lehetőségünk is.

Ebből következik, hogy ha épp a nullvektort akarjuk kifejezni $\underline{e}_1, \underline{e}_2, \ldots, \underline{e}_n$ lineáris kombinációjaként, akkor minden együtthatót 0-nak kell választanunk. Ez az 1.2.12. Tétel értelmében azt jelenti, hogy $\underline{e}_1, \underline{e}_2, \ldots, \underline{e}_n$ lineárisan független.

Összefoglalva: $\underline{e}_1, \underline{e}_2, \ldots, \underline{e}_n$ generátorrendszer $\mathbb{R}^n$-ben és lineárisan független, így valóban bázis. □

A fenti állításból tehát valóban következik $\dim \mathbb{R}^n = n$, de most már azt is értjük, miért óvakodtunk $\mathbb{R}^n$-et „az n-dimenziós térnek" nevezni: mert $\mathbb{R}^n$ csak egyike az n-dimenziós tereknek. Például $\mathbb{R}^3$-ön kívül az 1.2.18. Feladatban látott V altér is háromdimenziós (mert találtunk benne három elemű bázist) és könnyű megmutatni, hogy minden $m > n$-re $\mathbb{R}^m$-nek van n-dimenziós altere.

Az 1.2.21. Állításban látott $\underline{e}_1, \underline{e}_2, \ldots, \underline{e}_n$ bázis természetesen nem az egyetlen $\mathbb{R}^n$-ben (mert ilyenből végtelen sok van), de speciális szereppel bír $\mathbb{R}^n$ bázisai között. Például $\mathbb{R}^3$-ben és $\mathbb{R}^2$-ben a koordinátarendszer tengely irányú egységvektorai alkotják ezt a bázist.

1.2.22. Definíció. *Az 1.2.21. Állításban definiált $\underline{e}_1, \underline{e}_2, \ldots, \underline{e}_n$ bázist* standard bázisnak *hívjuk $\mathbb{R}^n$-ben. A standard bázis jelölése: E_n (vagy ha n értéke a szövegkörnyezetből egyértelmű, akkor egyszerűen csak E).*

Koordinátavektor

A térben a tengelyirányú egységvektorok (vagyis a standard bázis) szoros összefüggésben állnak a koordinátarendszerrel: minden $\underline{v} = (a, b, c)$ vektor kifejezhető a lineáris kombinációjukként, a kifejezéshez használt együtthatók pedig épp $\underline{v}$ koordinátái: $\underline{v} = a\underline{e}_1 + b\underline{e}_2 + c\underline{e}_3$. A koordinátarendszer működése szempontjából viszont az is alapvető fontosságú, hogy minden $\underline{v}$ vektor csak egyféleképpen legyen kifejezhető – vagyis különböző koordinátahármasok ne felelhessenek meg ugyanannak a vektornak. Az alábbi tétel azt mondja ki, hogy ugyanez igaz minden bázisra.

1.2.23. Tétel. *A $V \leq \mathbb{R}^n$ altérben a $\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_k$ vektorok akkor és csak akkor alkotnak bázist, ha minden $\underline{v} \in V$ egyértelműen (vagyis pontosan egyféleképpen) fejezhető ki a lineáris kombinációjukként.*

Bizonyítás: Az állítás „csak akkor” részéhez meg kell mutatnunk, hogy ha $\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_k$ teljesíti a feltételt, akkor bázis – vagyis generátorrendszer V-ben és lineárisan független. Az előbbi rögtön következik a feltételből (hiszen minden $\underline{v} \in V$ kifejezhető $\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_k$ lineáris kombinációjaként). Az utóbbi pedig az 1.2.12. Tételből: mivel a feltétel $\underline{v} = \underline{0}$-ra is igaz és a triviális lineáris kombináció biztosan a $\underline{0}$-t adja, ezért nemtriviális lineáris kombináció nem adhat $\underline{0}$-t.

Az „akkor” irány bizonyításából az világos, hogy minden $\underline{v} \in V$ kifejezhető a $\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_k$ lineáris kombinációjaként (hiszen a bázis egyben generátorrendszer is). Indirekt tegyük fel, hogy valamely $\underline{v} \in V$ kétféleképpen is kifejezhető: $\underline{v} = \lambda_1\underline{b}_1 + \lambda_2\underline{b}_2 + \ldots + \lambda_k\underline{b}_k$, $\underline{v} = \mu_1\underline{b}_1 + \mu_2\underline{b}_2 + \ldots + \mu_k\underline{b}_k$ és $\lambda_j \neq \mu_j$ valamely j-re. A két kifejezés különbségét véve: $\underline{0} = (\lambda_1 - \mu_1)\underline{b}_1 + (\lambda_2 - \mu_2)\underline{b}_2 + \ldots + (\lambda_k - \mu_k)\underline{b}_k$. Azt kaptuk, hogy a $\underline{0}$ kifejezhető a $\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_k$ nemtriviális lineáris kombinációjaként (hiszen $\lambda_j - \mu_j \neq 0$). Ez az 1.2.12. Tétel szerint ellentmond annak, hogy $\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_k$ lineárisan független. □

A tétel szerint tehát ha egy altérben adott egy bázis, akkor az altér minden $\underline{v}$ vektorának kölcsönösen egyértelműen megfeleltethetők a $\underline{v}$ kifejezéséhez szükséges lineáris kombináció együtthatói – hasonlóan ahhoz, ahogyan a koordinátarendszer működik a síkban és a térben. Ennek ad nevet az alábbi definíció – amely tehát annak

a szemléletes megfogalmazásnak ad precíz jelentést, hogy minden bázis „meghatároz egy koordinátarendszert” a megfelelő V altérben.

1.2.24. Definíció. *Legyen $V \leq \mathbb{R}^n$ altér, $B = \{\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_k\}$ bázis V-ben és $\underline{v} \in V$ tetszőleges vektor. Azt mondjuk, hogy a $\underline{k} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_k \end{pmatrix} \in \mathbb{R}^k$ vektor a $\underline{v}$* vektor B szerinti koordinátavektora*, ha $\underline{v} = \lambda_1\underline{b}_1 + \lambda_2\underline{b}_2 + \ldots + \lambda_k\underline{b}_k$. Ennek jelölése: $\underline{k} = [\underline{v}]_B$.*

Egyszerű, de fontos megfigyelés, hogy ha $V = \mathbb{R}^n$ és $B = E_n$ a standard bázis, akkor $\underline{v} = [\underline{v}]_B$ minden $\underline{v} \in V$-re (ez kiolvasható az 1.2.21. Állítás bizonyításából). Valójában ez az összefüggés emeli speciális szerepbe E_n-t az $\mathbb{R}^n$ bázisai között. Azt is fontos tudatosítani, hogy $[\underline{v}]_B$ nem csak $\underline{v}$-től függ: ugyanannak a vektornak különböző bázisok szerint más és más koordinátavektorok felelnek meg. Erre az alábbi feladat a) része is példát mutat: ott $[\underline{v}]_B \neq \underline{v}$, szemben a $B = E_n$ esetével.

1.2.25. Feladat. Határozzuk meg az alábbi $[\underline{v}]_B$ koordinátavektorokat.

a) $V = \mathbb{R}^3$, $B = \left\{ \underline{b}_1 = \begin{pmatrix} 1 \\ 6 \\ 1 \end{pmatrix}, \underline{b}_2 = \begin{pmatrix} 3 \\ 4 \\ -1 \end{pmatrix}, \underline{b}_3 = \begin{pmatrix} 1 \\ -8 \\ 2 \end{pmatrix} \right\}$, $\underline{v} = \begin{pmatrix} 3 \\ 4 \\ 9 \end{pmatrix}$

b) V és B a (legutóbb) az 1.2.18. Feladatban látott altér és bázis, $\underline{v}$ pedig az 1.2.3. Feladatban megadott $\underline{d}$ vektor.

Megoldás: a) Először is figyeljük meg, hogy B valóban bázis $\mathbb{R}^3$-ben, mert a vektorai nem illeszkednek közös origón átmenő síkra; valóban, a $\underline{b}_1$-re és $\underline{b}_2$-re illeszkedő, origón átmenő sík egyenlete $5x - 2y + 7z = 0$ (lásd az 1.2.10. Feladat b) részét), erre pedig $\underline{b}_3$ nem illeszkedik. A kérdés tehát az, hogy a $\lambda_1\underline{b}_1 + \lambda_2\underline{b}_2 + \lambda_3\underline{b}_3 = \underline{v}$ egyenlet milyen λ_1, λ_2, λ_3 skalárokra teljesül. Behelyettesítve:

$$\lambda_1 \begin{pmatrix} 1 \\ 6 \\ 1 \end{pmatrix} + \lambda_2 \begin{pmatrix} 3 \\ 4 \\ -1 \end{pmatrix} + \lambda_3 \begin{pmatrix} 1 \\ -8 \\ 2 \end{pmatrix} = \begin{pmatrix} \lambda_1 + 3\lambda_2 + \lambda_3 \\ 6\lambda_1 + 4\lambda_2 - 8\lambda_3 \\ \lambda_1 - \lambda_2 + 2\lambda_3 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \\ 9 \end{pmatrix}.$$

Így a $\lambda_1 + 3\lambda_2 + \lambda_3 = 3$, $6\lambda_1 + 4\lambda_2 - 8\lambda_3 = 4$, $\lambda_1 - \lambda_2 + 2\lambda_3 = 9$ lineáris egyenletrendszerre jutunk. A második egyenlethez a harmadik 4-szeresét adva: $10\lambda_1 = 40$, amiből $\lambda_1 = 4$. Ezt az első és (például) a harmadik egyenletbe helyettesítve a $3\lambda_2 + \lambda_3 = -1$, $-\lambda_2 + 2\lambda_3 = 5$ egyenletrendszert kapjuk; ennek a megoldása (például a második egyenlet háromszorosát az elsőhöz adva) $\lambda_2 = -1$, $\lambda_3 = 2$. Így az eredmény: $[\underline{v}]_B = \begin{pmatrix} 4 \\ -1 \\ 2 \end{pmatrix}$.

b) V-ről korábban (az 1.2.5. Feladat f) részében, illetve az 1.2.10. Feladat d) részében) már láttuk, hogy altér, ebben B az 1.2.18. Feladat szerint valóban bázis – jelölje ennek a vektorait most sorban $\underline{b}_1$, $\underline{b}_2$ és $\underline{b}_3$. Az 1.2.3. Feladatban pedig megmutattuk, hogy $\underline{v} = 2\underline{b}_1 - 7\underline{b}_2 - 3\underline{b}_3$. Korábbi eredményeinket tehát most így foglalhatjuk össze: $[\underline{v}]_B = \begin{pmatrix} 2 \\ -7 \\ -3 \end{pmatrix}$. □

Bázis létezése

Az 1.2.20. Definíció szerint egy V altér dimenziója k, ha van benne k elemű bázis. Azt ugyan már tudjuk, hogy ebben az esetben minden V-beli bázis k elemű, de egy fontos kérdés még nyitva maradt: van-e egyáltalán minden altérben legalább egy bázis? A válasz szerencsére igen, ez következni fog az alábbi tételből.

1.2.26. Tétel. *Legyen $V \leq \mathbb{R}^n$ altér, $\underline{f}_1, \underline{f}_2, \ldots, \underline{f}_k$ V-beli vektorokból álló lineárisan független rendszer. Ekkor $\underline{f}_1, \underline{f}_2, \ldots, \underline{f}_k$ kiegészíthető véges sok (esetleg nulla) további vektorral úgy, hogy a kapott rendszer bázis legyen.*

Bizonyítás: Legyen $W = \langle \underline{f}_1, \underline{f}_2, \ldots, \underline{f}_k \rangle$. Nyilván igaz, hogy $W \subseteq V$ (mert V altér, így az $\underline{f}_i$-kből lineáris kombinációval kifejezhető vektorok mind V-beliek kell legyenek). Ha $V = W$, akkor $\underline{f}_1, \underline{f}_2, \ldots, \underline{f}_k$ generátorrendszer és így bázis is V-ben; ekkor tehát a tételt beláttuk (nulla elemmel egészítettük ki a rendszert). Ha $W \neq V$, akkor létezik egy $\underline{v} \in V$, $\underline{v} \notin W$ vektor. Az újonnan érkező vektor 1.2.14. Lemmája szerint ekkor $\underline{f}_1, \underline{f}_2, \ldots, \underline{f}_k, \underline{v}$ lineárisan független (ellenkező esetben ugyanis a lemma szerint $\underline{v} \in W$ teljesülne). Ha $\underline{f}_1, \underline{f}_2, \ldots, \underline{f}_k, \underline{v}$ már generátorrendszer V-ben, akkor a tételt ismét beláttuk. Ha nem, akkor folytatjuk ezt az eljárást (vagyis kiegészítjük a rendszert egy $\langle \underline{f}_1, \underline{f}_2, \ldots, \underline{f}_k, \underline{v} \rangle$-hez nem tartozó vektorral, stb).

Azt kell még megmutatnunk, hogy ez a folyamat egy ponton leáll és szolgáltatja a keresett bázist. Ez azonban következik az 1.2.15. FG-egyenlőtlenségből: mivel $\mathbb{R}^n$-ben van n elemű generátorrendszer (például a standard bázis), ezért nem létezhet benne n-nél nagyobb lineárisan független rendszer. Így az eljárás legföljebb $n-k$ vektor hozzávétele után megáll. □

1.2.27. Következmény. *Minden $V \leq \mathbb{R}^n$ altérben van bázis (és ezért $\dim V$ is létezik).*

Bizonyítás: Ha $V = \{\underline{0}\}$, akkor az üres halmaz bázis V-ben. Ha viszont V tartalmaz egy $\underline{v} \neq \underline{0}$ vektort, akkor $\underline{v}$-re (mint egyetlen elemből álló lineárisan független rendszerre) alkalmazva a fenti tételt kapunk egy V-beli bázist. □

Érdemes visszaemlékeznünk az 1.2.2. pontban az $\mathbb{R}^2$ és $\mathbb{R}^3$ altereiről mondottakra: a triviálisakon kívül csak az origón átmenő egyenesek és síkok ilyenek. Akkor ezt nem bizonyítottuk, de ezt most könnyen pótolhatjuk: egy $V \leq \mathbb{R}^3$ altérre $\dim V$

attól függően 0, 1, 2 vagy 3, hogy az altér a $\{\underline{0}\}$, egy origón átmenő egyenes, egy origón átmenő sík, vagy $\mathbb{R}^3$.

Végül megemlítjük az 1.2.26. Tétel egy további hasznos következményét.

1.2.28. Következmény. *Legyen $V \leq \mathbb{R}^n$ altér, $\underline{f}_1, \underline{f}_2, \ldots, \underline{f}_k$ V-beli vektorokból álló lineárisan független rendszer. Ha $\dim V = k$, akkor $\underline{f}_1, \underline{f}_2, \ldots, \underline{f}_k$ bázis V-ben.*

Bizonyítás: V-ben van k elemű generátorrendszer (V minden bázisa ilyen), így az 1.2.15. FG-egyenlőtlenség szerint minden lineárisan független V-beli rendszer legföljebb k elemű. Ezért ha az 1.2.26. Tételt alkalmazzuk V-re és az $\underline{f}_1, \underline{f}_2, \ldots, \underline{f}_k$ rendszerre, a kapott bázis csak maga $\underline{f}_1, \underline{f}_2, \ldots, \underline{f}_k$ lehet. □

1.2.29. Feladat. Álljon V azokból az $\mathbb{R}^4$-beli vektorokból, amelyekben a négy koordináta összege 0. Egészítsük ki az $\underline{f}_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \\ -6 \end{pmatrix}$ vektort V-beli bázissá.

Megoldás: Az 1.2.26. Tétel bizonyításában leírt módszert követjük. Tekintsük először az $\langle \underline{f}_1 \rangle$ generált alteret, amely tehát $\underline{f}_1$ skalárszorosaiból áll. Nyilván $V \neq \langle \underline{f}_1 \rangle$, ezért választunk egy tetszőleges $\underline{f}_2 \in V$, $\underline{f}_2 \notin \langle \underline{f}_1 \rangle$ vektort: legyen például $\underline{f}_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}$. Most $\langle \underline{f}_1, \underline{f}_2 \rangle$ meghatározásához vegyük egy tetszőleges lineáris kombinációjukat: $\alpha \underline{f}_1 + \beta \underline{f}_2 = \begin{pmatrix} \alpha+\beta \\ 2\alpha \\ 3\alpha \\ -6\alpha-\beta \end{pmatrix}$. Látszik, hogy például $\underline{f}_3 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ -1 \end{pmatrix} \in V$, de $\underline{f}_3 \notin \langle \underline{f}_1, \underline{f}_2 \rangle$ (mert $2\alpha = 1$ és $3\alpha = 0$ egyszerre nem lehet). Az 1.2.18. Feladatban láttuk, hogy $\dim V = 3$, így az 1.2.28. Következmény miatt $\underline{f}_1, \underline{f}_2, \underline{f}_3$ már bázis V-ben. □

1.3. Lineáris egyenletrendszerek

Az eddig megismert fogalmak kapcsán, különböző feladatok megoldásaiban lépten-nyomon lineáris egyenletrendszerekbe botlottunk. Ha például adott $\mathbb{R}^n$-beli $\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_k$ és $\underline{w}$ vektorokra a $\underline{w} \in \langle \underline{v}_1, \underline{v}_2, \ldots, \underline{v}_k \rangle$ állítás igazságát akarjuk tesztelni – más szóval: $\underline{w}$-t kifejezni $\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_k$ lineáris kombinációjaként, – akkor egy

n egyenletből álló, k változós lineáris egyenletrendszerre jutunk, ahol a változók a lineáris kombináció együtthatói; lásd az 1.2.3. és az 1.2.25. Feladatokat. Ha a $\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_k$ lineáris függetlenségét kell eldöntenünk, akkor az 1.2.12. Tétel szerint ismét lineáris egyenletrendszerre jutunk – mégpedig olyanra, ahol a jobb oldalakon álló számok mind nullák; lásd az 1.2.13. Feladatot. Ilyenkor nem az egyenletrendszer megoldhatósága a kérdés – hiszen nyilván megoldást kapunk, ha minden változó értékét nullának választjuk (ez felel meg a triviális lineáris kombinációnak). Ehelyett ebben az esetben arra vagyunk kíváncsiak, hogy az egyenletrendszer *egyértelműen megoldható–e* – vagyis hogy ez az egyetlen megoldása van-e a rendszernek.

Lineáris egyenlet alatt az $a_1x_1 + a_2x_2 + \ldots + a_nx_n = b$ alakú egyenleteket értjük, ahol $x_1, x_2, \ldots, x_n$ változók (ismeretlenek) és az $a_1, a_2, \ldots, a_n$ együtthatók, valamint a b konstans tag adott számok. Egy lineáris egyenletrendszer pedig néhány (véges sok) lineáris egyenletből áll, amelyeket az $x_1, x_2, \ldots, x_n$ változóknak egyszerre kell kielégíteni. Lineáris egyenletrendszerek természetesen nem csak a fentebb említett feladatokban, hanem számos gyakorlati alkalmazásban is felmerülnek.

A lineáris egyenletrendszerek megoldásának egyszerű módszereit – változók kifejezését és másik egyenletekbe helyettesítését, vagy az „egyenlő együtthatók módszerét" – középiskolás tanulmányaiból mindenki ismeri. Bár az alábbiakban szintén lineáris egyenletrendszerek megoldásáról lesz szó, mégis, egy minőségi szinttel feljebb lépünk: nem csak egyes, konkrét lineáris egyenletrendszerek megoldása a célunk, hanem mindegyiké egyszerre: egy hatékony algoritmust ismerünk meg erre a feladatra. A *Gauss-elimináció* nevű eljárás egy adott lineáris egyenletrendszer esetén képes lesz eldönteni, hogy a rendszer megoldható-e és ha igen, áttekinthető módon megadja az *összes* megoldását. Az algoritmus egyrészt a gyakorlati alkalmazások miatt is nagyon fontos, másrészt a lineáris egyenletrendszerekről szóló, általános érvényű állítások bizonyítását is lehetővé teszi.

1.3.1. Ismerkedés a Gauss-eliminációval

A Gauss-elimináció tulajdonképp nem más, mint az „egyenlő együtthatók módszerének" a szisztematikus, minden esetre kiterjedő megvalósítása. Az eljárással való első ismerkedésként tekintsük az alább a bal oldalon látható, három változós és négy egyenletből álló lineáris egyenletrendszert. Az első változó, x_1 kiküszöbölésére (vagyis „eliminálásához") először az első egyenletet elosztjuk 2-vel, hogy x_1 együtthatója 1 legyen; a többi egyenletet egyelőre változatlanul hagyjuk (ez látható középen). Majd az új első egyenlet egy-egy alkalmas többszörösét (2-szeresét, 6-szorosát, illetve 4-szeresét) levonjuk a többiből úgy, hogy a kapott egyenletekben x_1 már ne szerepeljen:

$$\begin{array}{l} 2x_1 - x_2 + 6x_3 = 12 \\ 2x_1 + 2x_2 + 3x_3 = 24 \\ 6x_1 - x_2 + 17x_3 = 46 \\ 4x_1 - x_2 + 13x_3 = 32 \end{array} \quad \rightarrow \quad \begin{array}{l} x_1 - \frac{1}{2}x_2 + 3x_3 = 6 \\ 2x_1 + 2x_2 + 3x_3 = 24 \\ 6x_1 - x_2 + 17x_3 = 46 \\ 4x_1 - x_2 + 13x_3 = 32 \end{array} \quad \rightarrow \quad \begin{array}{l} x_1 - \frac{1}{2}x_2 + 3x_3 = 6 \\ 3x_2 - 3x_3 = 12 \\ 2x_2 - x_3 = 10 \\ x_2 + x_3 = 8 \end{array}$$

A lépéseink hátterében természetesen az a szándék állt, hogy x_1 már csak az első egyenletben szerepeljen; így a maradék, eggyel kevesebb egyenletet tartalmazó rendszerben a változók száma is eggyel kisebb. A folytatásban erre a kisebb rendszerre ismételjük meg az eddigiekkel analóg lépéseket.

Mielőtt azonban továbbmegyünk, megismerjük a lineáris egyenletrendszereknek azt a tömörebb tárolási módját, amelyet a Gauss-elimináció során alkalmazni szoktak. Mivel az algoritmus az egyenletrendszernek számtalan, egyre egyszerűbb változatát készíti el, ezért csak a futáshoz szükséges adatokat tárolja – a változók nevei, a műveleti jelek és az egyenlőségjel fölöslegesek. Ehelyett az úgynevezett *kibővített együtthatómátrixot* használjuk: egy számtáblázatban – vagyis *mátrixban* – sorról sorra leírjuk az egyenletekben a változók együtthatóit, illetve (vonallal elválasztva) az egyenlet jobb oldalán álló konstans tagot. A fenti egyenletrendszer, illetve az elimináció első két, már megtett lépése tehát mátrixos alakban így néz ki:

$$\left(\begin{array}{rrr|r} 2 & -1 & 6 & 12 \\ 2 & 2 & 3 & 24 \\ 6 & -1 & 17 & 46 \\ 4 & -1 & 13 & 32 \end{array}\right) \sim \left(\begin{array}{rrr|r} 1 & -\frac{1}{2} & 3 & 6 \\ 2 & 2 & 3 & 24 \\ 6 & -1 & 17 & 46 \\ 4 & -1 & 13 & 32 \end{array}\right) \sim \left(\begin{array}{rrr|r} 1 & -\frac{1}{2} & 3 & 6 \\ 0 & 3 & -3 & 12 \\ 0 & 2 & -1 & 10 \\ 0 & 1 & 1 & 8 \end{array}\right)$$

Figyeljük meg, hogy az egyenletek beszorzása egy α számmal (vagyis $\frac{1}{\alpha}$-val való leosztása) a kibővített együtthatómátrixban a megfelelő sor α-val való végigszorzásának felelt meg (beleértve a jobb szélső, a vonaltól jobbra álló tagot is). Hasonlóan, amikor az egyik egyenlethez hozzáadtuk egy másik α-szorosát (vagyis levontuk annak a $(-\alpha)$-szorosát), akkor a mátrixban az egyik megfelelő sorhoz adtuk tagonként a másik α-szorosát. Mindkét esetben a mátrix soraival végzett műveletek – α-val szorzás, illetve összeadás – analóg volt azzal, ahogyan $\mathbb{R}^n$-ben az oszlopvektorok közötti műveleteket értelmeztük: tagonként végeztük azokat.

Folytatva a fent elkezdett eliminációt, most a második sort (vagyis egyenletet) osztjuk el 3-mal (azaz megszorozzuk $1/3$-dal), majd a harmadik és a negyedik sorból kivonjuk az új második sor 2, illetve 1-szeresét (bal oldalt, illetve középen). A harmadik sorban most eleve 1-es jött létre, így ezt most nem szükséges beszorozni, rögtön kivonhatjuk a kétszeresét negyedikből (jobb oldalt).

$$\sim \left(\begin{array}{rrr|r} 1 & -\frac{1}{2} & 3 & 6 \\ 0 & 1 & -1 & 4 \\ 0 & 2 & -1 & 10 \\ 0 & 1 & 1 & 8 \end{array}\right) \sim \left(\begin{array}{rrr|r} 1 & -\frac{1}{2} & 3 & 6 \\ 0 & 1 & -1 & 4 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 2 & 4 \end{array}\right) \sim \left(\begin{array}{rrr|r} 1 & -\frac{1}{2} & 3 & 6 \\ 0 & 1 & -1 & 4 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

A kapott mátrix negyedik sora csupa nulla lett, ami a $0x_1+0x_2+0x_3=0$ egyenletnek felel meg. Ez nyilván azonosság, amely a változók minden értékére teljesül – vagyis „nem hordoz információt”, az elhagyása nem változtat az egyenletrendszer megoldásain:

$$\sim \left(\begin{array}{ccc|c} 1 & -\frac{1}{2} & 3 & 6 \\ 0 & 1 & -1 & 4 \\ 0 & 0 & 1 & 2 \end{array}\right)$$

A most kapott alakot *lépcsős alaknak* szokás nevezni (mert a sorokban lépcsőzetesen, egyre beljebb helyezkednek el az 1-esek, amelyektől balra csupa 0 áll). Figyeljük meg, hogy a lépcsős alakból az egyenletrendszer megoldása már kevés számolással kiolvasható. Az utolsó sor a $0x_1 + 0x_2 + 1x_3 = 2$ egyenletnek felel meg – vagyis $x_3 = 2$. A második sor egyenletben kifejezve: $x_2 - x_3 = 4$. Mivel x_3 értékét már tudjuk, ebből $x_2 = 6$ adódik – az első sorból pedig x_1 értéke kapható meg hasonlóan. Azonban ezeket a felfelé haladva történő visszahelyettesítéseket is kiválthatjuk az eddigiekhez hasonló lépésekkel. Folytatva a Gauss-eliminációt, először a harmadik sor 1-szeresét, illetve (-3)-szorosát adjuk a második, illetve az első sorhoz; végül az új második sor $1/2$-szeresét adjuk az elsőhöz:

$$\sim \left(\begin{array}{ccc|c} 1 & -\frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 & 6 \\ 0 & 0 & 1 & 2 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 6 \\ 0 & 0 & 1 & 2 \end{array}\right)$$

A kapott alak – amelynek a neve *redukált lépcsős alak* – már valóban az egyenletrendszer megoldását adja: a mátrix sorai sorban az $x_1 = 3$, $x_2 = 6$, $x_3 = 2$ egyenleteknek felelnek meg. A Gauss-elimináció ezzel véget ér – az egyenletrendszer tehát egyértelműen megoldható.

Érdemes végiggondolni, hogyan változna a fenti megoldás, ha az eredeti egyenletrendszert csak annyiban módosítjuk, hogy a negyedik egyenlet jobb oldalát 33-ra cseréljük: $4x_1 - x_2 + 13x_3 = 33$. Az elimináció során így is a fentiekkel azonos lépéseket tennénk, csak a mátrixok jobb alsó sarkában álló számok változnának – mégpedig mindig 1-gyel nagyobb értéket kapnánk, mint a fenti számolásban. Ez egészen addig a pontig igaz, ahol az imént a csupa nulla sort elhagytuk; ehelyett most a következő alakra jutunk:

$$\left(\begin{array}{ccc|c} 1 & -\frac{1}{2} & 3 & 6 \\ 0 & 1 & -1 & 4 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{array}\right)$$

Itt az utolsó sor a $0x_1 + 0x_2 + 0x_3 = 1$ egyenletnek felel meg, ami nyilván ellentmondás, a változók semmilyen értékére nem teljesül. Az algoritmus tehát „az egyenletrendszer nem megoldható” kijelentéssel zárul. A Gauss-eliminációval kapcsolatban meghonosodott szóhasználat szerint a fenti mátrix utolsó sorát – és minden olyan sort, amelyben a vonaltól balra csupa nulla, de a vonaltól jobbra nemnulla áll – *tilos sornak* nevezik. Az elnevezés annyiban félrevezető, hogy egy ilyen sor keletkezése természetesen nem „tilos”, az algoritmus során előfordulhat; a név arra utal, hogy tilos sor keletkezése esetén az egyenletrendszer nem megoldható.

1.3.2. Még egy példa a Gauss-eliminációra

Nem számíthatunk arra, hogy a Gauss-elimináció minden bemenet esetén a fenti példában látott, akadálymentes módon zajlik – például mert vannak olyan lineáris egyenletrendszerek, amelyeknek végtelen sok megoldása van, az algoritmusnak pedig ezeket is kell tudnia kezelni. Az alábbiakban erre látunk példát:

$$\begin{aligned}
x_1 + \ x_2 + \ 2x_3 + 2x_4 + \ \ x_5 &= -1 \\
4x_1 + 4x_2 + \ 8x_3 + 9x_4 + \ \ x_5 &= -7 \\
2x_1 + 5x_2 + 13x_3 + \ x_4 + 26x_5 &= 10 \\
x_1 + 3x_2 + \ 8x_3 + 2x_4 + 11x_5 &= \ 1 \\
2x_1 + \ x_2 + \ \ x_3 + 2x_4 + \ 3x_5 &= \ 3
\end{aligned}$$

Az algoritmus futtatásához először felírjuk az egyenletrendszer kibővített együtthatómátrixát. A mátrixban látható karikázott elem mindig azt a pozíciót fogja jelölni, ahol az eljárás éppen tart. A karika haladása nyomán – hasonlóan a már látott példához – 1-esek keletkeznek: minden sorban ez lesz az első nemnulla elem. Ezeket az 1-eseket *vezéregyesnek* nevezzük. A bal felső sarokban álló elem eleve is egy 1-es, így a sort nem szükséges skalárral végigszorozni: az első vezéregyes rögtön adott, az alatta álló nemnulla elemeket változtatjuk nullává az első sor megfelelő többszörösének hozzáadásával:

$$\left(\begin{array}{ccccc|c}
\textcircled{1} & 1 & 2 & 2 & 1 & -1 \\
4 & 4 & 8 & 9 & 1 & -7 \\
2 & 5 & 13 & 1 & 26 & 10 \\
1 & 3 & 8 & 2 & 11 & 1 \\
2 & 1 & 1 & 2 & 3 & 3
\end{array}\right) \sim \left(\begin{array}{ccccc|c}
\textcircled{1} & 1 & 2 & 2 & 1 & -1 \\
0 & 0 & 0 & 1 & -3 & -3 \\
0 & 3 & 9 & -3 & 24 & 12 \\
0 & 2 & 6 & 0 & 10 & 2 \\
0 & -1 & -3 & -2 & 1 & 5
\end{array}\right) \sim$$

Az előző példában látottak szerint most egy sorral lejjebb és egy oszloppal jobbra léptetjük a karikát. Azonban most azt látjuk, hogy a karikában álló szám 0 (alább, balra); ezzel nyilván nem lehet végigosztani a második sort, az algoritmus futásán tehát módosítanunk kell. A megoldás nagyon egyszerű: felcserélhetjük a második sort a harmadikkal (alább, jobbra). A sorok felcserélése nyilván a megfelelő egyenletek megcserélésének felel meg, ez pedig nem befolyásolja a megoldást. A második sor helyére cserélhettük volna egyébként a harmadik helyett bármelyik alatta lévő sort is; érdemes viszont megfigyelni, hogy az első sorral cserélni hiba lett volna, mert ezzel az első oszlopban már elért eredményeinket tönkretettük volna.

$$\left(\begin{array}{ccccc|c}
1 & 1 & 2 & 2 & 1 & -1 \\
0 & \textcircled{0} & 0 & 1 & -3 & -3 \\
0 & 3 & 9 & -3 & 24 & 12 \\
0 & 2 & 6 & 0 & 10 & 2 \\
0 & -1 & -3 & -2 & 1 & 5
\end{array}\right) \sim \left(\begin{array}{ccccc|c}
1 & 1 & 2 & 2 & 1 & -1 \\
0 & \textcircled{3} & 9 & -3 & 24 & 12 \\
0 & 0 & 0 & 1 & -3 & -3 \\
0 & 2 & 6 & 0 & 10 & 2 \\
0 & -1 & -3 & -2 & 1 & 5
\end{array}\right) \sim$$

Most a korábban látottakat követve mehetünk tovább: a második sort 3-mal osztva létrehozzuk a második vezéregyest. Majd az alatta lévő sorokból a második megfelelő többszöröseit kivonva elérjük, hogy a vezéregyes alatt minden elem 0 legyen; eközben persze a harmadik sor nem változik, mert ott (az iménti sorcsere nyomán) eleve nulla áll.

$$\left(\begin{array}{ccccc|c} 1 & 1 & 2 & 2 & 1 & -1 \\ 0 & \textcircled{1} & 3 & -1 & 8 & 4 \\ 0 & 0 & 0 & 1 & -3 & -3 \\ 0 & 2 & 6 & 0 & 10 & 2 \\ 0 & -1 & -3 & -2 & 1 & 5 \end{array}\right) \sim \left(\begin{array}{ccccc|c} 1 & 1 & 2 & 2 & 1 & -1 \\ 0 & \textcircled{1} & 3 & -1 & 8 & 4 \\ 0 & 0 & 0 & 1 & -3 & -3 \\ 0 & 0 & 0 & 2 & -6 & -6 \\ 0 & 0 & 0 & -3 & 9 & 9 \end{array}\right) \sim$$

Ismét jobbra és lefelé léptetve a karikát abban megint 0 áll. Ráadásul most a sorcsere sem segít, mert a karika alatti elemek is mind nullák (alább, balra). Első benyomásunk talán az lehet, hogy erre a helyzetre végképp nincs megoldás: ha nem akarjuk elrontani az eddig elért eredményeket, akkor a karika helyén semmilyen trükkel nem tudunk 1-est generálni. Ez ugyan igaz, de később ki fog derülni, hogy valójában ez a helyzet inkább szerencsésnek tekinthető: anélkül keletkeztek nullák, hogy tennünk kellett volna értük – „a véletlen dolgozott helyettünk”. Az algoritmus futását viszont folytatnunk kell: az ilyen esetekre (tehát: a karikában és alatta mindenhol nulla áll) vonatkozó szabály az, hogy a karika eggyel jobbra lép, de marad a jelenlegi sorában (alább, jobbra):

$$\left(\begin{array}{ccccc|c} 1 & 1 & 2 & 2 & 1 & -1 \\ 0 & 1 & 3 & -1 & 8 & 4 \\ 0 & 0 & \textcircled{0} & 1 & -3 & -3 \\ 0 & 0 & 0 & 2 & -6 & -6 \\ 0 & 0 & 0 & -3 & 9 & 9 \end{array}\right) \sim \left(\begin{array}{ccccc|c} 1 & 1 & 2 & 2 & 1 & -1 \\ 0 & 1 & 3 & -1 & 8 & 4 \\ 0 & 0 & 0 & \textcircled{1} & -3 & -3 \\ 0 & 0 & 0 & 2 & -6 & -6 \\ 0 & 0 & 0 & -3 & 9 & 9 \end{array}\right) \sim$$

A folytatás most problémamentes: létrejött a harmadik vezéregyes (ehhez nem is kellett a harmadik sort végigszoroznunk), alatta nullákat generálunk a harmadik sor megfelelő többszöröseink hozzáadásával (alább, balra). Azt tapasztaljuk, hogy ezzel az utolsó két sor csupa nullává változik. Ezeket tehát (az első példában látotthoz hasonlóan) elhagyjuk. Ezzel elértük a *lépcsős alakot* (alább, jobbra):

$$\left(\begin{array}{ccccc|c} 1 & 1 & 2 & 2 & 1 & -1 \\ 0 & 1 & 3 & -1 & 8 & 4 \\ 0 & 0 & 0 & \textcircled{1} & -3 & -3 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right) \sim \left(\begin{array}{ccccc|c} 1 & 1 & 2 & 2 & 1 & -1 \\ 0 & 1 & 3 & -1 & 8 & 4 \\ 0 & 0 & 0 & \textcircled{1} & -3 & -3 \end{array}\right) \sim$$

A Gauss-elimináció futásából már csak a *redukált lépcsős alakig* vezető út van hátra. Ez nem mást jelent, mint hogy a lépcsős alaknak a vezéregyest tartalmazó oszlopaiban a vezéregyesek fölötti nemnulla elemeket nullává változtatjuk. Most tehát három ilyen elem van: a negyedik oszlopban a fölső kettő és a második oszlopban a legfölső. Ezeket a lépéseket érdemes alulról fölfelé végezni: először a legalsó sor vezéregyese fölötti elemeket változtatjuk nullává, majd az utolsó előtti sor vezéregyese fölöttieket, stb. – hasonlóan ahhoz, ahogyan a lépcsős alakig vezető úton a vezéregyesek alatti elemekkel jártunk el. Először tehát a harmadik sor 1-szeresét, illetve (-2)-szeresét adjuk a második, illetve az első sorhoz (lásd alább, balra). Majd a második sor (-1)-szeresét adjuk az első sorhoz (alább, jobbra).

$$\left(\begin{array}{ccccc|c} 1 & 1 & 2 & 0 & 7 & 5 \\ 0 & 1 & 3 & 0 & 5 & 1 \\ 0 & 0 & 0 & 1 & -3 & -3 \end{array}\right) \sim \left(\begin{array}{ccccc|c} 1 & 0 & -1 & 0 & 2 & 4 \\ 0 & 1 & 3 & 0 & 5 & 1 \\ 0 & 0 & 0 & 1 & -3 & -3 \end{array}\right)$$

Elértük a redukált lépcsős alakot, az algoritmus ezzel megáll. A kérdés persze az: mond-e ez bármit az egyenletrendszer megoldásáról? A valóság az, hogy a redukált lépcsős alak nem más, mint az egyenletrendszer összes (jelen esetben: végtelen sok) megoldásának áttekinthető formában való leírása. Ennek illusztrálására először visszaírjuk a redukált lépcsős alakú rendszert a hagyományos formába:

$$\begin{aligned} x_1 \quad - \quad x_3 \quad + 2x_5 &= \quad 4 \\ x_2 + 3x_3 \quad + 5x_5 &= \quad 1 \\ x_4 - 3x_5 &= -3 \end{aligned}$$

Jól látszik, hogy az x_3 és x_5 változóknak bármilyen értéket adhatunk, ezeknek a tetszőleges megválasztása után a maradék három változó már egyértelműen kifejezhető lesz – épp ennek a mikéntjét írja le a redukált lépcsős alak. Az egyenletrendszer összes megoldását tehát a következőképpen adhatjuk meg:

$$\begin{aligned} &x_3 = \alpha \in \mathbb{R},\ x_5 = \beta \in \mathbb{R} \\ &x_1 = 4 - 2\beta + \alpha \\ &x_2 = 1 - 5\beta - 3\alpha \\ &x_4 = -3 + 3\beta \end{aligned}$$

Az x_3 és x_5 változókat itt *szabad paraméternek* szokás nevezni (mert az értékük tetszőlegesen megválasztható). Látható, hogy szabad paraméterek azokból a változókból lettek, amelyeknek megfelelő oszlopokban a redukált lépcsős alak nem tartalmazott vezéregyest.

1.3.3. A Gauss-elimináció

Egy k egyenletből álló és n változós – röviden: $(k \times n)$-es – lineáris egyenletrendszer leírásához be kell vezetnünk a kettős indexelésű együtthatókat: $a_{i,j}$ jelöli az i-edik

egyenletben a j-edik változó együtthatóját minden $1 \leq i \leq k$ és $1 \leq j \leq n$ esetén. Az i-edik egyenlet jobb oldalán álló konstans tagot b_i-vel jelölve a lineáris egyenletrendszer „hagyományos” alakja, illetve *kibővített együtthatómátrixa* a következő:

$$\begin{array}{rcl} a_{1,1}x_1 + a_{1,2}x_2 + \ldots + a_{1,n}x_n & = & b_1 \\ a_{2,1}x_1 + a_{2,2}x_2 + \ldots + a_{2,n}x_n & = & b_2 \\ \vdots & \vdots & \vdots \\ a_{k,1}x_1 + a_{k,2}x_2 + \ldots + a_{k,n}x_n & = & b_k \end{array} \qquad \left(\begin{array}{cccc|c} a_{1,1} & a_{1,2} & \ldots & a_{1,n} & b_1 \\ a_{2,1} & a_{2,2} & \ldots & a_{2,n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{k,1} & a_{k,2} & \ldots & a_{k,n} & b_k \end{array}\right)$$

Elemi sorekvivalens lépések

Az alábbi definíció ad nevet azoknak a lépéseknek, amelyeket a fentebb megoldott két példában végeztünk az egyenletrendszeren.

1.3.1. Definíció. *Kibővített együtthatómátrixával adott lineáris egyenletrendszer esetén* elemi sorekvivalens lépésnek *nevezzük az alábbiakat tetszőleges* $1 \leq i, j \leq k$, $i \neq j$ *és* $\lambda \in \mathbb{R}$, $\lambda \neq 0$ *skalár esetén:*

(i) a mátrix i-edik sorának (tagonként való) megszorzása λ-val;

(ii) a mátrix i-edik sorának helyettesítése sajátmagának és a j-edik sor λ-szorosának (tagonként vett) összegével;

(iii) az i-edik és a j-edik sor felcserélése;

(iv) egy csupa nulla elemeket tartalmazó sor elhagyása.

Szinte magától értetődő, de a Gauss-elimináció működése szempontjából alapvető fontosságú az alábbi állítás – és egyben indokolja a fenti fogalom elnevezését.

1.3.2. Állítás. *Az 1.3.1. Definícióban felsorolt lépések* ekvivalens átalakítások, *vagyis az egyenletrendszer megoldásait nem változtatják meg. (Részletesebben: ha az $x_1, x_2, \ldots, x_n$ számok kielégítik az egyenletrendszert egy lépés megtétele előtt, akkor annak megtétele után is; és fordítva, ha $x_1, x_2, \ldots, x_n$ kielégítik az egyenletrendszert egy lépés megtétele után, akkor előtte is.)*

Bizonyítás: A bizonyítást a *(ii)* lépésre írjuk csak le, a többit az olvasóra hagyjuk. Ha $x_1, x_2, \ldots, x_n$ kielégítik az egyenletrendszert, akkor $a_{i,1}x_1 + a_{i,2}x_2 + \ldots + a_{i,n}x_n = b_i$ és $a_{j,1}x_1 + a_{j,2}x_2 + \ldots + a_{j,n}x_n = b_j$. Az utóbbi egyenlet λ-szorosát az előbbihez adva: $(a_{i,1} + \lambda a_{j,1})x_1 + (a_{i,2} + \lambda a_{j,2})x_2 + \ldots + (a_{i,n} + \lambda a_{j,n})x_n = b_i + \lambda b_j$. Ez épp azt jelenti, hogy a *(ii)* lépés megtétele után az új i-edik egyenlet teljesül (a többi pedig nem is változott).

Megfordítva: ha $x_1, x_2, \ldots, x_n$ megoldása a rendszernek a *(ii)* lépés megtétele után, akkor $(a_{i,1} + \lambda a_{j,1})x_1 + (a_{i,2} + \lambda a_{j,2})x_2 + \ldots + (a_{i,n} + \lambda a_{j,n})x_n = b_i + \lambda b_j$ és $a_{j,1}x_1 + a_{j,2}x_2 + \ldots + a_{j,n}x_n = b_j$ igazak. Az utóbbi egyenlet λ-szorosát az előbbiből

kivonva: $a_{i,1}x_1 + a_{i,2}x_2 + \ldots + a_{i,n}x_n = b_i$. Vagyis az i-edik egyenlet a lépés megtétele előtt is teljesült. □

Fontos kiemelni, hogy – bármennyire egyszerű is a bizonyítása – a fenti állítás a biztosítéka a Gauss-elimináció helyes működésének. Ennek érzékeltetésére érdemes megemlíteni egy feladatmegoldásokban fellépő, tipikus hibát: amikor valaki az elimináció egy pontján egy korábbi mátrix egy sorát adja hozzá az aktuális mátrix egy sorához. Miért is ne lehetne ezt megtenni – kérdezhetnénk –, hiszen bármelyik mátrixban is került elő a két szóban forgó sor, mindkét megfelelő egyenlet teljesül az $x_1, x_2, \ldots, x_n$ változókra, így nyilván az összegük is? A válasz az, hogy ez ugyan igaz, de nem elegendő az algoritmus helyes működéséhez: ez nem ekvivalens lépés, bővítheti az egyenletrendszer megoldáshalmazát. (Például: előfordulhat, hogy ezáltal egy egyenlethez a saját ellentettjét adjuk valamilyen álcázott formában, ezáltal „elveszítjük az egyenlet információtartalmát" – mintha elhagynánk a rendszerből.)

Lépcsős alak, redukált lépcsős alak

A korábban megoldott példákban láttuk, hogy a Gauss-elimináció során a célunk a redukált lépcsős alak elérése. Az alábbi definíció ezt a fogalmat vezeti be.

1.3.3. Definíció. *Egy kibővített együtthatómátrixával adott lineáris egyenletrendszert* lépcsős alakúnak *mondunk, ha az alábbiak teljesülnek:*

(i) A mátrix minden sorában van nemnulla elem és (balról) az első nemnulla elem egy 1-es – az úgynevezett vezéregyes.

(ii) Ha $1 \leq i < j \leq k$, akkor az i-edik sorban álló vezéregyes kisebb sorszámú oszlopban van, mint a j-edik sor vezéregyese.

(iii) A vezéregyesekkel egy oszlopban, azok alatt álló minden elem 0.

Redukált lépcsős alakúnak *mondjuk a mátrixot, ha még az alábbi is teljesül:*

(iv) A vezéregyesekkel egy oszlopban, azok fölött álló minden elem is 0.

(Könnyű végiggondolni, hogy a fenti definíció *(i)* és *(ii)* feltételeiből már következik a *(iii)*, azt csak a jobb érthetőség érdekében vettük be a feltételek közé.) A redukált lépcsős alak egy „eleve megoldott" lineáris egyenletrendszernek tekinthető. Valóban: ahogyan azt a 36. oldalon láttuk, a vezéregyeseket nem tartalmazó oszlopok *szabad paramétereknek* felelnek meg, a mátrix sorai pedig (átrendezés után) azt írják le, hogy a többi változó hogyan fejezhető ki a szabad paraméterekből. Ha pedig minden oszlopban van vezéregyes, akkor a sorok egyszerűen megadják a változók értékeit az egyértelmű megoldásban (lásd a 33. oldalt).

A Gauss-elimináció általános leírása

A Gauss-elimináció leírásánál azt kell pontosan meghatároznunk, hogy egy tetszőleges bemenet esetén az 1.3.1. Definícióban felsorolt lépéseket milyen sorrendben és milyen paraméterválasztásokkal hajtjuk végre. Az algoritmus két fázisból áll: az elsőben vagy elérjük a lépcsős alakot, vagy „tilos sor" keletkezik (lásd a 33. oldalt).

A második fázisban érjük el a redukált lépcsős alakot; ha az első fázisban tilos sor keletkezett, akkor a második fázis természetesen elmarad.

Az eljárás során nyilvántartjuk annak a mátrixelemnek a sor- és oszlopszámát, ahol a következő vezéregyesnek kell keletkeznie; ezeket i és j jelöli. A korábban megoldott, második példában az $a_{i,j}$ elemet bekarikázással jelöltük.

GAUSS-ELIMINÁCIÓ – ELSŐ FÁZIS

Bemenet: Egy k sorú és $n+1$ oszlopú mátrix (egy k egyenletből álló, n változós lineáris egyenletrendszer kibővített együtthatómátrixa, lásd a 37. oldalt).

0. lépés. $i \leftarrow 1$, $j \leftarrow 1$

1. lépés.
- Ha $a_{i,j} = 0$, akkor folytassuk a **2. lépésnél**.
- Szorozzuk meg az i-edik sort $\frac{1}{a_{i,j}}$-vel.
- Ha $i = k$, akkor folytassuk a **3. lépésnél**.
- Minden $i < t \leq k$ esetén adjuk a t-edik sorhoz az (imént módosított) i-edik sor $(-a_{t,j})$-szeresét.
- Ha $j = n$, akkor folytassuk a **3. lépésnél**.
- $i \leftarrow i+1$, $j \leftarrow j+1$
- Folytassuk az **1. lépésnél**.

2. lépés.
- Ha $i < k$ és van olyan $i < t \leq k$, amelyre $a_{t,j} \neq 0$, akkor:
  - ▶ Cseréljük fel az i-edik sort a t-edik sorral.
  - ▶ Folytassuk az **1. lépésnél**.
- Ha $j = n$, akkor:
  - ▶ $i \leftarrow i-1$
  - ▶ Folytassuk a **3. lépésnél**.
- $j \leftarrow j+1$
- Folytassuk az **1. lépésnél**.

3. lépés.
- Ha $i = k$, akkor PRINT „A mátrix lépcsős alakú.”; STOP.
- Ha van olyan $i < t \leq k$, amelyre $b_t \neq 0$, akkor PRINT „Az egyenletrendszer nem megoldható.”; STOP.
- Minden $i < t \leq k$ esetén hagyjuk el a t-edik (csupa nulla) sort.
- PRINT „A mátrix lépcsős alakú.”; STOP.

A fenti pszeudokód valóban azt az eljárást valósítja meg, amit a korábbi két megoldott példában már láttunk. Az 1. lépésben hajtjuk végre az elimináció tipikus műveletét: a „bekarikázott” $a_{i,j}$ elemet 1-essé, az alatta állókat pedig 0-vá változtatjuk (az 1.3.1. Definícióbeli *(i)* és *(ii)* lépésekkel), majd a karikát eggyel jobbra és lefelé mozgatjuk. A 2. lépés kezeli azt az esetet, amikor a karikában 0 áll: ha lehet, akkor egy lentebbi sorral való cserével nemnulla elemet hoz létre a karikában

és újból meghívja az 1. lépést, ha viszont a karika alatt sincs nemnulla elem, akkor eggyel jobbra lép.

A 3. lépésnek akkor adódik át a vezérlés, ha a karika nem léptethető tovább: elértük vagy a mátrix utolsó sorát, vagy az n. (vagyis a vonaltól balra utolsó) oszlopát. Mindkét esetben megáll az algoritmus, de az utóbbi esetben előbb még fontos lépéseket hajt végre. Az algoritmus működésének helyessége szempontjából alapvető fontosságú a következő állítás.

1.3.4. Állítás. *Ha a Gauss-elimináció első fázisában a 3. lépés végrehajtása során $i < k$, akkor minden $i < t \leq k$ esetén a t-edik sorban az utolsó (vonaltól jobbra álló) elem kivételével minden elem 0.*

Bizonyítás: Amikor a vezérlés a 3. lépéshez kerül, akkor $i = k$ vagy $j = n$ (vagy mindkettő) és az i-edik az utolsó olyan sor, ahol az eljárás vezéregyest hozott létre. (Külön figyelmet érdemel a 2. lépésben a $j = n$ eset: ilyenkor i értékét csökkentjük eggyel – vagyis a karikát fölfelé mozgatjuk – mielőtt a 3. lépésre ugrunk. Ennek az oka éppen az, hogy enélkül a technikai beavatkozás nélkül úgy érnénk el a 3. lépést, hogy az i-edik sorban nincs vezéregyes. Például a 36. oldalon látott példában az első fázist az $i = 3$, $j = 4$ ponton fejeztük be mielőtt elhagytuk volna a két csupa nulla sort. A fenti pszeudokód végrehajtása során a karika először egyet lépne jobbra és lefelé, vagyis $i = 4$, $j = 5$ volna; majd a 2. lépésnél írtak szerint a karika egyet lépne fölfelé, a 3. lépéshez az $i = 3$, $j = 5$ értékekkel jutnánk el.)

Az $i < k$ esetben tehát $j = n$. Mivel j értéke 1-től n-ig egyesével nőtt, ezért közben minden értéket felvett, mindig az 1. lépés, vagy a 2. lépés végrehajtásának a végén növeltük. Az előbbi esetben az 1. lépésben írtak szerint az aktuális i-edik sortól lefelé a j-edik oszlopban minden elemet nullává változtattunk. Ha j-t a 2. lépésben növeltük, akkor a j-edik oszlop eleve csupa nulla volt az aktuális i-edik sortól lefelé. A j-edik oszlopban létrehozott vagy ott „talált” nulla értékeket az eljárás a folytatásban végig megőrizte. Így az első fázis végrehajtásának végén érvényes i-re már valóban igaz, hogy az annál nagyobb indexű sorokban az utolsót leszámítva minden elem 0. □

A fenti állítás igazolja, hogy az eljárás 3. lépésében írtak valóban helyesek: ha $i < k$, akkor az $(i+1)$-ediktől a k-adikig terjedő sorok között vagy van tilos sor (ha a jobb oldalon nemnulla áll) és így az egyenletrendszer nem megoldható, vagy mindegyikük elhagyható. Az utóbbi esetben tehát lépcsős alakú kibővített együtthatómátrixszal fejezzük be az első fázist – és természetesen ugyanez igaz, ha a 3. lépést $i = k$ miatt értük el.

GAUSS-ELIMINÁCIÓ – MÁSODIK FÁZIS

Bemenet: Egy k' sorú és $n+1$ oszlopú, lépcsős alakú kibővített együtthatómátrix.

A második fázisban a vezéregyesek fölötti nemnulla elemeket változtatjuk nullává. Minden ilyen $a_{i,j}$ elemhez egyszer kell végrehajtani az 1.3.1. Definícióbeli *(ii)* lépést: ha a j-edik oszlop vezéregyese $a_{t,j} = 1$ és $i < t$, akkor az i-edik sorból kivonjuk

a t-edik sor $(a_{i,j})$-szeresét. A második fázist már nem érdemes a fentihez hasonló pszeudokód formájában részletezni: nincsenek esetszétválasztások, az eljárás akadálymentesen zajlik. Nem szükséges, de érdemes ezeket a lépéseket a mátrixban alulról fölfelé haladva végezni, mert így (a korábbi lépésekben már létrehozott nulláknak köszönhetően) kevesebb számítást végzünk.

Mivel a második fázis végrehajtása fenntartja az 1.3.3. Definíció szerint a lépcsős alakhoz szükséges tulajdonságokat, de a végén a vezéregyesek fölötti elemeket is nullává változtatja, ezért valóban redukált lépcsős alakot hoz létre.

A Gauss-eliminációról eddig mondottakat az alábbi tételben foglaljuk össze.

1.3.5. Tétel. *Tetszőleges, kibővített együtthatómátrixával adott lineáris egyenletrendszer esetén a Gauss-eliminációt futtatva az alábbi esetek közül pontosan az egyik valósul meg:*

(i) Az első fázis 3. lépésének végrehajtásakor az eljárás „tilos sort” talál. Ekkor az egyenletrendszer nem megoldható.

(ii) Az algoritmus redukált lépcsős alakra hozza a kibővített együtthatómátrixot, amelynek minden oszlopában van vezéregyes. Ekkor az egyenletrendszer egyértelműen megoldható.

(iii) Az algoritmus redukált lépcsős alakra hozza a kibővített együtthatómátrixot, de annak nem minden oszlopában van vezéregyes. Ekkor az egyenletrendszernek végtelen sok megoldása van.

A (ii) és (iii) esetekben az egyenletrendszer megoldásai a redukált lépcsős alakból közvetlenül kiolvashatók.

Egy fontos elméleti következmény

Tegyük fel, hogy adott egy k egyenletből álló, n ismeretlenes lineáris egyenletrendszer. Milyen következtetésre juthatunk a megoldhatóságával kapcsolatban, ha csupán k és n nagysági relációját ismerjük? Ezzel a kérdéssel kapcsolatban számos elterjedt tévhit él; így például:

- *nem igaz*, hogy ha $k = n$, akkor biztosan van megoldás;
- *nem igaz*, hogy ha $k < n$, akkor biztosan végtelen sok megoldás van;
- *nem igaz*, hogy ha $k > n$, akkor biztosan nincs megoldás.

Ezekre a hamis állításokra érdemes ellenpéldákat keresni (az utolsóra már láttunk is). Az erről a kérdésről mondható egyetlen igaz állítást az alábbi tétel tartalmazza. A bizonyítás egyben illusztrálja azt a gyakori jelenséget, hogy egy algoritmus elméleti eredmények bizonyítására is alkalmas lehet.

1.3.6. Tétel. *Ha egy k egyenletből álló, n ismeretlenes lineáris egyenletrendszer egyértelműen megoldható, akkor $k \geq n$.*

Bizonyítás: Futtassuk le a Gauss-eliminációt az egyenletrendszerre. Mivel az megoldható, ezért nem keletkezik tilos sor, az algoritmus redukált lépcsős alakot hoz létre; legyen ebben a sorok száma k'. Nyilván $k' \leq k$, mert az algoritmus csökkentheti a sorok számát (az első fázis 3. lépésében), de nem növelheti. Mivel az egyen-

letrendszer egyértelműen megoldható, ezért a redukált lépcsős alak minden oszlopa tartalmaz vezéregyest (lásd a 33. oldal példáját). Ebből $k' = n$ következik. Ezeket összevetve: $k \geq k' = n$, amivel a tételt beláttuk. □

1.3.7. Feladat. A p paraméter minden értékére döntsük el, hogy az alábbi lineáris egyenletrendszer megoldható-e és ha igen, adjuk meg az összes megoldását.

$$\begin{aligned} x_1 + x_3 + 3x_4 + 8x_5 &= 4 \\ 2x_1 + 3x_2 + 8x_3 + 6x_4 + 10x_5 &= 17 \\ 3x_1 + 2x_2 + 7x_3 + 12x_4 + 8x_5 &= 21 \\ 2x_1 + 4x_2 + 10x_3 + 8x_4 + p \cdot x_5 &= 22 \end{aligned}$$

Megoldás: A Gauss-eliminációt alkalmazva a következőket kapjuk:

$$\left(\begin{array}{ccccc|c} 1 & 0 & 1 & 3 & 8 & 4 \\ 2 & 3 & 8 & 6 & 10 & 17 \\ 3 & 2 & 7 & 12 & 8 & 21 \\ 2 & 4 & 10 & 8 & p & 22 \end{array}\right) \sim \left(\begin{array}{ccccc|c} 1 & 0 & 1 & 3 & 8 & 4 \\ 0 & 3 & 6 & 0 & -6 & 9 \\ 0 & 2 & 4 & 3 & -16 & 9 \\ 0 & 4 & 8 & 2 & p-16 & 14 \end{array}\right) \sim$$

$$\left(\begin{array}{ccccc|c} 1 & 0 & 1 & 3 & 8 & 4 \\ 0 & 1 & 2 & 0 & -2 & 3 \\ 0 & 0 & 0 & 3 & -12 & 3 \\ 0 & 0 & 0 & 2 & p-8 & 2 \end{array}\right) \sim \left(\begin{array}{ccccc|c} 1 & 0 & 1 & 3 & 8 & 4 \\ 0 & 1 & 2 & 0 & -2 & 3 \\ 0 & 0 & 0 & 1 & -4 & 1 \\ 0 & 0 & 0 & 0 & p & 0 \end{array}\right)$$

(Itt már a sorok leosztását és a keletkezett vezéregyes alatti elemek „kinullázását” egyszerre végeztük.) Ha $p = 0$, akkor az utolsó sor elhagyható, az első fázis véget ért. A második fázisban egyetlen további lépéssel (a negyedik oszlop 3-asának kinullázásával) kapjuk a redukált lépcsős alakot:

$$\left(\begin{array}{ccccc|c} 1 & 0 & 1 & 0 & 20 & 1 \\ 0 & 1 & 2 & 0 & -2 & 3 \\ 0 & 0 & 0 & 1 & -4 & 1 \end{array}\right)$$

A $p = 0$ esetben tehát végtelen sok megoldás van: $x_3 = \alpha \in \mathbb{R}$, $x_5 = \beta \in \mathbb{R}$, $x_1 = 1 - 20\beta - \alpha$, $x_2 = 3 + 2\beta - 2\alpha$, $x_4 = 1 + 4\beta$.

Ha viszont $p \neq 0$, akkor az utolsó sor p-vel osztásával ér véget az első fázis. A második fázisban most 4 darab vezéregyes fölötti elemet kell kinullázni, az alábbi redukált lépcsős alakot kapjuk:

$$\left(\begin{array}{ccccc|c} 1 & 0 & 1 & 3 & 8 & 4 \\ 0 & 1 & 2 & 0 & -2 & 3 \\ 0 & 0 & 0 & 1 & -4 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{array}\right) \sim \left(\begin{array}{ccccc|c} 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 2 & 0 & 0 & 3 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{array}\right)$$

Így a $p \neq 0$ esetben is végtelen sok megoldás van: $x_3 = \alpha \in \mathbb{R}$, $x_1 = 1 - \alpha$, $x_2 = 3 - 2\alpha$, $x_4 = 1$, $x_5 = 0$. □

1.4. Determináns

Fentebb említettük, hogy egy $(n \times n)$-es (vagyis n ismeretlenes, n egyenletből álló) lineáris egyenletrendszer – szemben az ezzel kapcsolatos esetleges tévhitekkel – nem feltétlen megoldható. Valójában az 1.3.5. Tételben említett mindhárom eset előfordulhat: lehet 0, 1 vagy végtelen sok megoldása is. Mégis: ha valaki elég sok lineáris egyenletrendszert megoldott már, támadhat az az érzése, hogy az $(n \times n)$-es esetben az tekintendő „természetesnek", ha a rendszer egyértelműen megoldható, a másik két lehetőség csak valami „véletlen egybeesés" műve lehet. Az alábbiakban az fog kiderülni, hogy emögött az érzés mögött matematikailag pontosan megfogalmazható tartalom rejlik.

Kezdetnek érdemes a (2×2)-es esetet kipróbálni:

$$\begin{aligned} a_{1,1}x_1 + a_{1,2}x_2 &= b_1 \\ a_{2,1}x_1 + a_{2,2}x_2 &= b_2 \end{aligned}$$

Könnyű gyakorlófeladat megmutatni (később ezt jóval általánosabban is belátjuk), hogy ha $a_{1,1} \cdot a_{2,2} - a_{1,2} \cdot a_{2,1} \neq 0$, akkor a rendszer egyértelműen megoldható; ha viszont $a_{1,1} \cdot a_{2,2} - a_{1,2} \cdot a_{2,1} = 0$, akkor a rendszernek vagy nincs megoldása, vagy végtelen sok van – ez már a jobb oldalakon álló b_1 és b_2 értékétől is függ. Be fogjuk bizonyítani (lásd az 1.4.11. Tételt), hogy ehhez hasonló jelenség minden n-re igaz: az $(n \times n)$-es lineáris egyenletrendszerben a bal oldalakon álló együtthatókból képezhető egy kifejezés, amelynek az értéke pontosan akkor nem nulla, ha a rendszer egyértelműen megoldható. Ezt az értéket hívják az együtthatómátrix *determinánsának* (mert „determinálja", vagyis előre meghatározza a rendszer viselkedését). A determináns fogalma ebből a megfigyelésből ered, de számtalan egyéb alkalmazása is van – ezek egy részét később megismerjük.

1.4.1. Permutációk inverziószáma

Permutáció alatt egy olyan n tagú számsorozatot értünk, amely az $1, 2, \ldots, n$ számok mindegyikét pontosan egyszer tartalmazza valamilyen $n \geq 1$ egész esetén. A permutáció tehát az $1, 2, \ldots, n$ számok egy „összekeverése", valamilyen sorrendben való felsorolása. A permutációkat görög betűkkel szokás jelölni, a π permutációban az i-edik helyen álló számot π_i jelöli. Például $n = 8$ esetén $\pi = (5, 3, 1, 8, 4, 2, 6, 7)$ egy permutáció, itt $\pi_1 = 5$, $\pi_2 = 3$, …, $\pi_8 = 7$. Az alábbi fogalom a determináns definíciójában jut lényeges szerephez.

1.4.1. Definíció. *Azt mondjuk, hogy a $\pi = (\pi_1, \pi_2, \ldots, \pi_n)$ permutációban a π_i és π_j tagok egymással* inverzióban állnak, *ha $i < j$, de $\pi_i > \pi_j$. A π permutáció* inverziószáma *az összes inverzióban álló számpárok száma. Ennek jele: $I(\pi)$.*

A permutáció két tagja tehát akkor áll inverzióban, ha „rossz sorrendben" vannak: a kisebb később következik, mint a nagyobb; az inverziószám pedig a „rossz

sorrendben" álló párok száma. Például a fenti, $\pi = (5,3,1,8,4,2,6,7)$ permutációban a 8 inverzióban áll a 6-tal, de nem áll inverzióban a 3-mal. Ugyanerre a permutációra $I(\pi) = 11$. (Ezt legegyszerűbb úgy összeszámolni, ha minden tagra az utána következő, de nála kisebb tagokat számoljuk meg és adjuk össze. Így például itt az 5-ös után 4 nála kisebb tag áll, a 3-as után 2, stb., végül is $I(\pi) = 4+2+0+4+1+0+0 = 11$.)

Az alábbi állítást szintén a determinánssal kapcsolatban fogjuk használni.

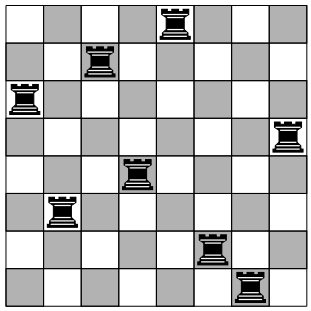
1.4.2. Állítás. *Cseréljük fel a $\pi = (\pi_1, \pi_2, \ldots, \pi_i, \ldots, \pi_j, \ldots, \pi_n)$ permutációban a π_i és π_j tagokat, a kapott permutációt jelölje $\pi' = (\pi_1, \pi_2, \ldots, \pi_j, \ldots, \pi_i, \ldots, \pi_n)$. Ekkor $I(\pi)$ és $I(\pi')$ paritása különböző (vagyis az egyik páros, a másik páratlan). (Azaz: két elem cseréjekor az inverziószám paritást vált.)*

Bizonyítás: Először két szomszédos elem cseréjére látjuk be az állítást – vagyis amikor $j = i+1$. Ekkor π_i és π_j egymáshoz való viszonya megváltozik: ha eddig inverzióban álltak, akkor ezután nem és fordítva, ha eddig nem, akkor a csere után igen. Viszont ez az egyetlen változás az inverziószámban: a π_i, π_j páron kívül minden más elempárra igaz, hogy ha a csere előtt inverzióban álltak, akkor a csere után is és ha előtte nem, akkor utána sem. Ezért az inverziószám pontosan 1-gyel nő vagy csökken, így a paritása megváltozik.

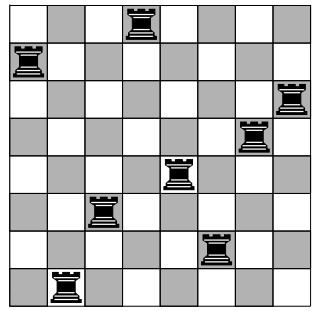
Térjünk most rá az általános eset bizonyítására, amikor π_i és π_j nem feltétlen szomszédosak. Ekkor π_i és π_j cseréjét szomszédos elempárok cseréjének egy sorozatával helyettesíthetjük. Feltéve, hogy $i < j$, először π_j-t felcseréljük π_{j-1}-gyel, majd az így melléje került π_{j-2}-vel, stb. Egészen addig „bugyborékoltatjuk" föl π_j-t, amíg az $(i+1)$-edik pozícióba, π_i mellé kerül. Az eddig végrehajtott szomszédcserék számát jelölje t. (Tudjuk, hogy $t = j-i-1$, de ez a gondolatmenet szempontjából érdektelen.) Folytatva a szomszédcseréket, most π_i és π_j egymás mellé került, így további 1 lépésben felcserélhetők. Végül a π_i elem megteszi visszafelé ugyanazt az utat, amelyen a π_j érkezett: először felcseréljük π_{i+2}-vel, stb. Nyilván π_i is t darab szomszédcsere árán érkezik meg a j-edik pozícióba. Ezzel megkaptuk a π' permutációt, amelyhez összesen $2t+1$ szomszédcserét végeztünk. Mivel minden szomszédcsere (ahogyan azt fentebb beláttuk) megváltoztatja az inverziószám paritását és összesen $(2t+1)$ – vagyis páratlan sok – paritásváltás történt, ezért $I(\pi)$ és $I(\pi')$ valóban ellentétes paritású. □

1.4.2. Bástyaelhelyezések

Régi kombinatorika feladat a következő: hányféleképpen helyezhető el a sakktáblán 8 bástya úgy, hogy semelyik kettő ne üsse egymást? Tudni kell, hogy a sakktábla (8×8)-as és két bástya akkor üti egymást a sakk szabályai szerint, ha egy sorban vagy egy oszlopban vannak. Úgy kell tehát elhelyezni a 8 bástyát, hogy minden sorban és oszlopban pontosan 1 legyen. A két átló nyilván jó megoldás, de az 1.4. ábra mutat két olyan példát is, amik jobban érzékeltetik a lehetőségek széles körét.



1.4a ábra



1.4b ábra

A bástyaelhelyezések megszámlálásához „kódolni” fogjuk azokat. Az ötlet egyszerű: felülről lefelé haladva minden egyes sorhoz leírjuk, hogy abban hányadik mezőn áll a bástya – hiszen minden sorban pontosan egy van. Például az 1.4a ábra esetében az első sorban az 5. mezőn, a másodikban a 3.-on áll bástya, stb. Végül is az 1.4a ábra bástyaelhelyezésének kódja: $(5,3,1,8,4,2,6,7)$. Látszik, hogy a kód egy permutáció; ugyanez nyilván bármely bástyaelhelyezésre igaz, hiszen minden oszlopban is egyetlen bástya van, ezért a kód készítésekor minden 1 és n közötti számot pontosan egyszer írunk le. A kódolás visszafelé is működik: ha például adott a $(4,1,8,7,5,3,6,2)$ permutáció, akkor az első sorban a 4. mezőre, a másodikban az 1.-re teszünk bástyát, stb.; végül az 1.4b ábra bástyaelhelyezését kapjuk.

A tanulság tehát az, hogy a jó bástyaelhelyezések kölcsönösen egyértelmű megfeleltetésben állnak (vagyis „kódolhatók”) az $1,2,\ldots,8$ elemek permutációival. Ezért a bástyaelhelyezések száma azonos a permutációk számával, ami $8! = 1\cdot 2\cdot \ldots \cdot 8 = 40320$. (Valóban: az első tagot nyolcféleképp választhatjuk, minden választást hétféleképp folytathatunk a második tag kiválasztásával; ez eddig $8\cdot 7$ lehetőség, amelyek mindegyikét hatféleképp folytathatjuk a harmadik taggal, stb.)

A determináns definiálásakor a fenti gondolatmenetből a permutációk és a bástyaelhelyezések közötti kölcsönösen egyértelmű megfeleltetés lesz fontos – persze nem a sakktábla, hanem egy $(n\times n)$-es mátrix esetében.

1.4.3. Definíció. *Egy $(n\times n)$-es mátrix (számtáblázat) elemei közül választott n darab elemet* bástyaelhelyezésnek *nevezünk, ha a mátrix minden sorában és oszlopában pontosan egy kiválasztott elem van. Azt mondjuk, hogy az $1,2,\ldots,n$ számok egy π permutációja* megfelel egy bástyaelhelyezésnek*, ha az első sorban a π_1-edik, a másodikban a π_2-edik, stb., az n-edikben a π_n-edik elemet választottuk ki.*

A fentiekből nyilván következik, hogy a bástyaelhelyezések száma egy $(n\times n)$-es mátrix esetén is $n!$.

1.4.3. A determináns definíciója

Ennyi előkészítés után már értelmezhetjük a determináns fogalmát.

1.4.4. Definíció. *Legyen adott egy $(n\times n)$-es A mátrix. Az A minden bástyaelhelyezésére szorozzuk össze az azt alkotó n elemet, majd a szorzatot lássuk el előjellel a következő szabály szerint: ha a bástyaelhelyezésnek megfelelő permutáció inverziószáma páros, akkor az előjel legyen pozitív, ha viszont páratlan az inverziószám, akkor az előjel legyen negatív. Az így kapott $n!$ darab, n tényezős előjelezett szorzat összegét az A* determinánsának *nevezzük. Ennek jele: $|A|$ vagy* $\det A$.

A definíciónak fontos része, hogy A *négyzetes mátrix*, vagyis a sorainak és oszlopainak a száma azonos; nem négyzetes mátrix determinánsáról beszélni értelmetlen. Legyen például A a következő:

$$A=\begin{pmatrix} 2 & 3 & 4 & 5 & 6\\ 7 & 8 & 9 & 10 & 11\\ 12 & 13 & 14 & 15 & 16\\ 17 & 18 & 19 & 20 & 21\\ 22 & 23 & 24 & 25 & 26 \end{pmatrix}$$

Ha $\det A$-t definíció szerint szeretnénk kiszámítani, akkor ez $5!=120$ darab ötténytyezős szorzat kiszámítását, előjelezését és ezek összeadását jelentené. Például a $2\cdot 8\cdot 14\cdot 20\cdot 26$ szorzat, ami a mátrix északnyugat-délkeleti átlójának – az úgynevezett *főátlónak* – felel meg, egyike a 120 szorzatnak. Az ennek megfelelő permutáció $(1,2,3,4,5)$ (mert az elemeket sorban az 1., majd a 2. oszlopból választottuk, stb.) Ennek az inverziószáma 0 (mert semelyik két elem nem áll inverzióban), ez páros szám, így ez a szorzat pozitív előjelet kap. Egy másik szorzat a 120 közül az $5\cdot 8\cdot 12\cdot 21\cdot 24$, mert ez az öt elem szintén bástyaelhelyezést alkot. A megfelelő permutáció a $(4,2,1,5,3)$, aminek az inverziószáma 5; ez páratlan, így a szorzat negatív előjelet kap. Így tehát

$$\det A=+2\cdot 8\cdot 14\cdot 20\cdot 26-5\cdot 8\cdot 12\cdot 21\cdot 24\pm\ldots\pm\ldots$$

A további 118 darab szorzat kiszámításának és előjelezésének részletezése nélkül is látható, hogy a determináns definíció szerinti kiszámítása nagyon fáradságos munka, később ennél jóval hatékonyabb algoritmust is megismerünk.

A fejezet hátralévő részére elfogadjuk azt a jelölést, hogy egy adott A mátrix i-edik sorának és j-edik oszlopának kereszteződésében álló elemet $a_{i,j}$ jelöli – és hasonlóan, a B mátrix megfelelő elemét $b_{i,j}$, stb. Ezt a jelölést használva az 1.4.4. Definíciót az alábbi képlettel is kifejezhetjük: egy $(n\times n)$-es A mátrixra

$$\det A=\sum_{\pi\text{: permutáció}}(-1)^{I(\pi)}\cdot a_{1,\pi_1}\cdot a_{2,\pi_2}\cdot\ldots\cdot a_{n,\pi_n}$$

Itt tehát az összegzés (amelyet $\sum$ jelöl) az $1,2,\ldots,n$ összes π permutációjára fut – ami tartalmilag azonos az összes bástyaelhelyezésre való összegzéssel. Az a_{i,π_i} a

π-nek megfelelő bástyaelhelyezésben az i-edik sorból választott elem, ezeket szorozzuk össze. Végül $(-1)^{I(\pi)}$ is megfelel a definícióban írt előjelezési szabálynak, mert (-1)-nek a páros hatványai $(+1)$-gyel, a páratlanok (-1)-gyel egyenlők.

A (2×2)-es mátrixok determinánsának kiszámítási szabályát érdemes külön megjegyezni. Ha

$$A=\begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix},$$

akkor összesen két szorzatot kell kiszámítanunk: $a_{1,1}\cdot a_{2,2}$ az $(1,2)$, $a_{1,2}\cdot a_{2,1}$ pedig a $(2,1)$ permutációnak felel meg; az előbbi inverziószáma 0, az utóbbié nyilván 1. Következésképp $\det A = a_{1,1}\cdot a_{2,2} - a_{1,2}\cdot a_{2,1}$. Ezek szerint a (2×2)-es mátrix determinánsát a két átló szorzatának különbségeként kapjuk, a főátló kap pozitív, az úgynevezett *mellékátló* pedig negatív előjelet. (Fontos viszont, hogy ez a szabály csak a (2×2)-es mátrixokra érvényes!) A kapott képletet érdemes a 43. oldalon a determináns fogalmának motivációjával kapcsolatban mondottakkal összevetni.

1.4.5. Feladat. Számítsuk ki az alábbi determinánsok értékét.

a) $\begin{vmatrix} 0 & 0 & 0 & 7 & 2 \\ 0 & 0 & 0 & 9 & 3 \\ 0 & 0 & 0 & 5 & 8 \\ 7 & 2 & 9 & 8 & 6 \\ 3 & 1 & 4 & 5 & 5 \end{vmatrix}$ b) $\begin{vmatrix} 0 & 0 & 9 & 0 & 7 \\ 2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 5 & 0 \\ 6 & 0 & 0 & 0 & 4 \\ 0 & 8 & 3 & 0 & 0 \end{vmatrix}$

Megoldás: A definíciót alkalmazva mindkét esetben $5! = 120$ darab ötténye zős szorzat kiszámítása, előjelezése és összegzése vár ránk. Ezt szerencsére nagyban leegyszerűsíti, hogy nem kell figyelembe venni azokat a szorzatokat, amelyeknek legalább egy tényezője 0. Valóban: ilyenkor a szorzat értéke 0, ami (előjeltől függetlenül) nem járul hozzá a determinánst adó összeghez. Így mindkét determináns esetében csak a 0-t nem tartalmazó bástyaelhelyezésekkel foglalkozunk.

Az a) feladatban az első három sor mindegyikében az utolsó két oszlop valamelyikébe kellene a bástyát tennünk, ha 0-ra nem teszünk bástyát. Ez nyilván lehetetlen (mert minden oszlopba csak egy bástya kerülhet). Ezért mind a 120 bástyaelhelyezés tartalmaz 0-t, így a determináns is 0 (mert 120 darab 0 előjeles összege).

A b) feladatban már lehet 0-t nem tartalmazó szorzatot készíteni, de szerencsére nem túl sokat. A 0-kat elkerülve a harmadik sorból csak a 5-öst választhatjuk. Az első sorból 9 vagy 7 választható. Az előbbi esetben az ötödik sorból már csak a 8-ast választhatjuk (mert a harmadik oszlopból már vettünk elemet), emiatt a második sorból csak 2-est, így a negyedikből csak a 4-est. Hasonlóan, ha az első sorból a 7-est vesszük, akkor a negyedikből már csak a 6-ost, a másodikból az 1-est és az ötödikből a 3-ast választhatjuk. Így összesen csak két nemnulla szorzat keletkezik: $9\cdot 2\cdot 5\cdot 4\cdot 8$ és $7\cdot 1\cdot 5\cdot 6\cdot 3$. A determináns kiszámításához már csak az ezekhez tartozó előjeleket kell meghatározni. Az első szorzatnak megfelelő permutáció $(3,1,4,5,2)$ (mert az első sorból a 3. elemet vettük ki, a másodikból az 1.-t, stb). Ennek a permutációnak az inverziószáma 4 (az inverzióban álló elempárok: $(3,1)$,

$(3,2)$, $(4,2)$ és $(5,2)$). Mivel az inverziószám páros, a szorzat előjele pozitív. Hasonlóan, a második szorzathoz tartozó permutáció $(5,2,4,1,3)$, ennek az inverziószáma 7, az előjel negatív. Így a determináns értéke $9\cdot 2\cdot 5\cdot 4\cdot 8 - 7\cdot 1\cdot 5\cdot 6\cdot 3 = 2250$. □

1.4.4. A determináns alaptulajdonságai

Ha egy (40×40)-es mátrix determinánsát a definíció szerint szeretnénk kiszámítani, akkor $40! \approx 8,16\cdot 10^{47}$ előjeles szorzatot kellene összegeznünk; ez még a jelenlegi leggyorsabb szuperszámítógépeknek is tovább tartana, mint az ősrobbanás óta eltelt idő egybilliószorosa. Létezik azonban a determináns kiszámítására a gyakorlatban jól alkalmazható, hatékony eljárás is. Ez azon alapul, hogy bizonyos speciális mátrixok determinánsa könnyen megállapítható, a többi pedig szintén ilyenné alakítható olyan lépésekkel, amelyek vagy nem, vagy nagyon egyszerűen követhető módon befolyásolják a determináns értékét. Az alábbiakban ennek a részleteit ismejük meg.

Egy $(n\times n)$-es A mátrixot *felsőháromszög-mátrixnak* nevezünk, ha a főátlója alatt álló minden elem 0; vagyis minden $1\le i,j\le n$, $i>j$ esetén $a_{i,j}=0$ (lásd 1.5a ábra). Hasonlóan, *alsóháromszög-mátrixnak* akkor nevezzük A-t, ha a főátlója fölött csak 0 áll, azaz $1\le i,j\le n$, $i<j$ esetén $a_{i,j}=0$ (lásd 1.5b ábra).

$$\begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} & \dots & a_{1,n}\\ 0 & a_{2,2} & a_{2,3} & a_{2,4} & \dots & a_{2,n}\\ 0 & 0 & a_{3,3} & a_{3,4} & \dots & a_{3,n}\\ 0 & 0 & 0 & a_{4,4} & \dots & a_{4,n}\\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & 0 & \dots & a_{n,n} \end{pmatrix}$$

1.5a ábra

$$\begin{pmatrix} a_{1,1} & 0 & 0 & 0 & \dots & 0\\ a_{2,1} & a_{2,2} & 0 & 0 & \dots & 0\\ a_{3,1} & a_{3,2} & a_{3,3} & 0 & \dots & 0\\ a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} & \dots & 0\\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots\\ a_{n,1} & a_{n,2} & a_{n,3} & a_{n,4} & \dots & a_{n,n} \end{pmatrix}$$

1.5b ábra

1.4.6. Tétel. *Legyen A egy $(n\times n)$-es mátrix.*

(i) Ha A-nak van csupa 0 elemet tartalmazó sora vagy oszlopa, akkor $\det A = 0$.

(ii) Ha A felsőháromszög-mátrix vagy alsóháromszög-mátrix, akkor a determinánsa a főátlóbeli elemek szorzata: $\det A = a_{1,1}\cdot a_{2,2}\cdot \ldots \cdot a_{n,n}$.

Bizonyítás: Az *(i)* azonnal következik a determináns 1.4.4. Definiciójából: mivel mind az $n!$ darab szorzat tartalmaz elemet abból a sorból vagy oszlopból is, amelyiknek minden tagja 0, ezért mindegyik szorzat értéke és így az (előjeles) összegükként kapott determináns is 0.

A *(ii)* bizonyításához legyen például A felsőháromszög-mátrix. Azokat a bástyaelhelyezéseket kell megkeresnünk, amelyek nem tartalmaznak 0 elemet, mert a többiből készült szorzatok nem befolyásolják a determináns értékét (hasonlóan az 1.4.5. Feladathoz). Így az első oszlopból csak $a_{1,1}$-et választhatjuk, a többi elem 0. A második oszlopból $a_{1,2}$-t már nem választhatjuk, mert az első sorból már vettünk elemet, $a_{3,2}$-től lefelé viszont minden elem 0; így az egyetlen lehetőség $a_{2,2}$.

Hasonlóan, mivel az első két sorból már vettünk elemet és $a_{4,3}$-tól lefelé csak 0 áll, ezért a harmadik oszlopból csak $a_{3,3}$ választható. Folytatva a gondolatmenetet látszik, hogy az egyetlen, 0-t nem tartalmazó szorzat $a_{1,1}\cdot a_{2,2}\cdot\ldots\cdot a_{n,n}$ lesz. Az ennek megfelelő permutáció $(1,2,\ldots,n)$, ennek az inverziószáma 0, így a szorzat pozitív előjelet kap. Az állítást ezzel felsőháromszög-mátrixra beláttuk; ha pedig a bizonyításban a „sor" és „oszlop" szavakat és a mátrixelemek indexeit végig felcseréljük (valamint a „lefelé" helyett a „jobbra" szót használjuk), akkor az alsóháromszög-mátrixra vonatkozó bizonyítást kapjuk. □

Ha tehát egy tetszőleges determinánst hatékonyan szeretnénk kiszámítani, akkor a fenti tétel a célt tűzi ki: ilyenné kell alakítani a mátrixot, hogy egyszerűen leolvashassuk a determináns értékét. Az alábbi tétel az ehhez megtehető lépéseket írja le.

1.4.7. Tétel. *Legyen A $(n\times n)$-es mátrix, $\lambda\in\mathbb{R}$ skalár, $1\le i,j\le n$, $i\neq j$ egészek.*

(i) Ha A egy sorát vagy oszlopát (tagonként) megszorozzuk λ-val, akkor a kapott A' mátrix determinánsa λ-szorosa A-énak: $\det A'=\lambda\cdot\det A$.

(ii) Ha A két sorát vagy két oszlopát felcseréljük, akkor a kapott A' mátrix determinánsa ellentettje A-énak: $\det A'=(-1)\cdot\det A$.

(iii) Ha A i-edik sorát helyettesítjük sajátmagának és a j-edik sor λ-szorosának (tagonként vett) összegével, akkor a kapott A' mátrix determinánsa megegyezik A-éval: $\det A'=\det A$. *Hasonlóan, ha A'-t az i-edik oszlop sajátmagának és a j-edik oszlop λ-szorosának az összegével való helyettesítésével kapjuk, akkor is* $\det A'=\det A$.

Bizonyítás: (i) bizonyításához tegyük fel például, hogy A'-t az i-edik sor λ-val szorzásával kaptuk. Hasonlítsuk össze A és A' determinánsának definíció szerinti kiszámítását: mivel minden bástyaelhelyezés pontosan egy elemet tartalmaz az i-edik sorból, ezért az A kiszámítása közben keletkező szorzatok mindegyikében pontosan egy tényező a λ-szorosára változik, amikor $\det A'$-t számítjuk. Ebből persze következik, hogy maga a szorzat értéke is λ-szoros lesz (az előjele viszont nem módosul, hiszen az azt meghatározó bástyaelhelyezés ugyanaz). Mivel tehát a $\det A'$ kiszámításakor keletkező mindegyik összeadandó a λ-szorosára változik, ezért ezek (előjeles) összege, vagyis a determináns értéke is. A bizonyítás változatlanul érvényes az i-edik sor helyett a j-edik oszlop λ-val szorzására is.

A *(ii)* bizonyítását először egy példán illusztráljuk: a 46. oldalon már látott A mátrix 3. és 5. sorának felcserélésével kapjuk A'-t:

$$A=\begin{pmatrix} 2 & 3 & 4 & \boxed{5} & 6\\ 7 & \boxed{8} & 9 & 10 & 11\\ \boxed{12} & 13 & 14 & 15 & 16\\ 17 & 18 & 19 & 20 & \boxed{21}\\ 22 & 23 & \boxed{24} & 25 & 26 \end{pmatrix}\qquad A'=\begin{pmatrix} 2 & 3 & 4 & \boxed{5} & 6\\ 7 & \boxed{8} & 9 & 10 & 11\\ 22 & 23 & \boxed{24} & 25 & 26\\ 17 & 18 & 19 & 20 & \boxed{21}\\ \boxed{12} & 13 & 14 & 15 & 16 \end{pmatrix}$$

A-ban bekereteztünk egy bástyaelhelyezést: a 46. oldalon már láttuk, hogy az ennek megfelelő permutáció $\pi=(4,2,1,5,3)$, ennek az inverziószáma 5, így a keletkező

szorzat negatív előjelet kap. A' determinánsának definíció szerinti kiszámításakor is megjelenik ugyanez a szorzat, a különbség csak a tényezők sorrendjében van (ami nyilván érdektelen) és a szorzatot adó bástyaelhelyezésben: az ennek (lásd a jobb oldalon) megfelelő permutáció a $\pi' = (4,2,3,5,1)$, aminek az inverziószáma 6, így A' kiszámításakor ugyanez a szorzat már pozitív előjelet kap. Érdemes megfigyelni a szóban forgó két permutáció közötti kapcsolatot is: π-ből a $\pi_3 = 1$ és $\pi_5 = 3$ tagok felcserélésével kapjuk π'-t; ez nem is meglepő, hiszen épp a 3. és 5. sorok felcserélésével kaptuk A-ból A'-t.

Ugyanez a jelenség nem csak a fenti példában, hanem általában is érvényes: ha A-ból az i-edik és a j-edik sor felcserélésével kapjuk A'-t, akkor A és A' bástyaelhelyezései párbaállíthatók úgy, hogy a párt alkotó bástyaelhelyezésekből keletkező szorzatok (sorrendtől) eltekintve azonosak, az előjelük viszont ellentétes. Valóban: ha a $\pi = (\pi_1, \pi_2, \ldots, \pi_i, \ldots, \pi_j, \ldots, \pi_n)$ permutációnak megfelelő A-beli bástyaelhelyezést a $\pi' = (\pi_1, \pi_2, \ldots, \pi_j, \ldots, \pi_i, \ldots, \pi_n)$ permutációnak megfelelő A'-belivel állítjuk párba, akkor a két keletkező szorzat (előjeltől és sorrendtől eltekintve) valóban azonos (mert $a_{i,\pi_i} = a'_{j,\pi_i}$ és $a_{j,\pi_j} = a'_{i,\pi_j}$ a sorcsere miatt). Az 1.4.2. Állításban beláttuk, hogy $I(\pi)$ és $I(\pi')$ különböző paritású, amiből következik, hogy a szorzat A-ban és A'-ben különböző előjelű. Így A definíció szerinti kiszámításában minden tag ellentettjét véve épp A' definíció szerinti kiszámítását kapjuk, amivel az állítást sorcserére beláttuk. Oszlopcserére a bizonyítás a fentivel lényegében azonos (az egyetlen apró különbség az, hogy π-ből nem π_i és π_j, hanem i és j felcserélésével kapjuk π'-t.)

A *(iii)* bizonyítását az alábbi (néha önmagában is hasznos) lemmával kezdjük.

1.4.8. Lemma. *Tegyük fel, hogy az $(n \times n)$-es X, Y és Z mátrixok az i-edik soraiktól eltekintve elemről elemre megegyeznek. Az i-edik soraikra viszont fennáll, hogy $z_{i,j} = x_{i,j} + y_{i,j}$ minden $1 \leq j \leq n$ esetén; vagyis a Z i-edik sora épp az X és az Y i-edik sorának (tagonkénti) összege. Ekkor* $\det Z = \det X + \det Y$. *Ezzel analóg állítás érvényes oszlopokra is (a részleteket mellőzzük).*

A Lemma bizonyítása: Vegyünk egy tetszőleges bástyaelhelyezést Z-ben, feleljen ez meg a π permutációnak; az ebből keletkező szorzat tehát

$$(-1)^{I(\pi)} \cdot z_{1,\pi_1} \cdot \ldots \cdot z_{i,\pi_i} \cdot \ldots \cdot z_{n,\pi_n}. \tag{1.4.1}$$

A $z_{i,\pi_i} = x_{i,\pi_i} + y_{i,\pi_i}$ behelyettesítés után ez az alábbival egyenlő:

$$(-1)^{I(\pi)} \cdot z_{1,\pi_1} \cdot \ldots \cdot (x_{i,\pi_i} + y_{i,\pi_i}) \cdot \ldots \cdot z_{n,\pi_n}.$$

Felbontva a zárójelet és felhasználva, hogy minden $k \neq i$ esetén $z_{k,\pi_k} = x_{k,\pi_k} = y_{k,\pi_k}$, ez tovább egyenlő az alábbival:

$$(-1)^{I(\pi)} \cdot x_{1,\pi_1} \cdot \ldots \cdot x_{i,\pi_i} \cdot \ldots \cdot x_{n,\pi_n} + (-1)^{I(\pi)} \cdot y_{1,\pi_1} \cdot \ldots \cdot y_{i,\pi_i} \cdot \ldots \cdot y_{n,\pi_n}. \tag{1.4.2}$$

Mivel az 1.4.1., illetve az 1.4.2. kifejezéseket minden bástyaelhelyezésre összegezve definíció szerint $\det Z$-t, illetve $(\det X + \det Y)$-t kapjuk, a lemmát ezzel beláttuk. (Oszlopok esetére a bizonyítás ezzel lényegében azonos.) ◇

Rátérve most már a tétel *(iii)* állításának bizonyítására, a fenti lemma alkalmazható az A' mátrixra, hiszen abban az i-edik sor minden eleme egy kéttagú összeg: $a'_{i,k} = a_{i,k} + \lambda \cdot a_{j,k}$ minden k-ra. A lemmát tehát a következő szereposztásban alkalmazzuk: $Z = A'$, $X = A$ és Y pedig az a mátrix, amely az i-edik sorától eltekintve azonos A-val, az i-edik sorában pedig az A j-edik sorának λ-szorosa áll: $y_{i,k} = \lambda \cdot a_{j,k}$ minden k-ra. A lemmát ezekre alkalmazva $\det A' = \det A + \det Y$ következik.

Készen leszünk tehát a bizonyítással, ha belátjuk, hogy $\det Y = 0$. Ehhez először vegyük észre, hogy Y i-edik sorára alkamazható a tétel (már bebizonyított) *(i)* állítása: ha Y' jelöli azt a mátrixot, amely az i-edik sorától eltekintve azonos Y-nal (és így A-val), az i-edik sorában pedig az A j-edik sorának másolata áll (vagyis $y'_{i,k} = a_{j,k}$ minden k-ra), akkor *(i)*-ből $\det Y = \lambda \det Y'$ következik. Y'-re pedig a tétel (szintén már bebizonyított) *(ii)* állítását érdemes alkalmazni: ha Y'-ben felcseréljük az i-edik és a j-edik sort, akkor ettől egyrészt a determináns (*(ii)* szerint) az ellentettjére változik, másrészt viszont nyilván változatlan is marad (hiszen Y'-n a sorcsere „nem látszik”, annak i-edik és j-edik sora azonos). Ebből tehát $\det Y' = -(\det Y')$ és így $\det Y' = 0$ következik. Ezzel a bizonyítás teljes: $\det Y = \lambda \det Y'$ miatt $\det Y = 0$, amiből $\det A' = \det A + \det Y$ miatt $\det A' = \det A$. Oszlopokra a bizonyítás ismét változtatás nélkül elmondható (a különbség csak annyi, hogy *(i)*, *(ii)* és a lemma állításából is az oszlopokra vonatkozó változatot használjuk). □

Megfigyelhető, hogy a determinánsnak az 1.4.6. és az 1.4.7. Tételben felsorolt tulajdonságaiban is a sorok és az oszlopok szerepe azonos. Ezt a jelenséget később az 1.5.4. Tételben pontosabban is megfogalmazzuk.

1.4.9. Feladat. Számítsuk ki az alábbi determináns értékét.

$$\begin{vmatrix} 3 & 12 & -3 & -6 \\ 2 & 8 & 3 & -9 \\ 1 & 5 & -1 & 0 \\ -1 & -3 & 3 & 5 \end{vmatrix}$$

Megoldás: Az 1.4.7. Tételben írt lépésekkel a mátrixot felsőháromszög-mátrixszá alakítjuk, közben nyomon követjük a determináns változásait. A felsőháromszög-mátrix determinánsát viszont az 1.4.6. Tétel szerint már le fogjuk tudni olvasni.

Először az 1.4.7. Tétel *(i)* állítását használjuk az 1. sorra: ha a sort $1/3$-dal megszorozzuk, akkor a determináns értéke is harmadára változik:

$$\begin{vmatrix} 3 & 12 & -3 & -6 \\ 2 & 8 & 3 & -9 \\ 1 & 5 & -1 & 0 \\ -1 & -3 & 3 & 5 \end{vmatrix} = 3 \cdot \begin{vmatrix} 1 & 4 & -1 & -2 \\ 2 & 8 & 3 & -9 \\ 1 & 5 & -1 & 0 \\ -1 & -3 & 3 & 5 \end{vmatrix} =$$

Mivel a jobb oldali determináns értéke harmada a bal oldaliénak, az egyenlőség fennállásához a jobb oldali determináns előtti 3-as szorzó szükséges. (Ezt a lépést úgy is elképzelhetjük, mintha a „determináns első sorából kiemelnénk 3-at”.)

Most az 1.4.7. Tétel *(iii)* állítását használjuk: a 2., 3., illetve 4. sorokhoz adjuk az 1. sor (-2)-szeresét, (-1)-szeresét, illetve 1-szeresét; közben a determináns értéke nem változik (alább, balra). Ezzel elértük, hogy a mátrix első oszlopa már egy felsőháromszög-mátrixéval azonos. Mielőtt hasonlóan folytatnánk, előbb felcseréljük a 2. és a 3. sort, hogy a főátló második pozíciójában is nemnulla álljon; ez az 1.4.7. Tétel *(ii)* állítása szerint a determinánst az ellentettjére változtatja:

$$= 3 \cdot \begin{vmatrix} 1 & 4 & -1 & -2 \\ 0 & 0 & 5 & -5 \\ 0 & 1 & 0 & 2 \\ 0 & 1 & 2 & 3 \end{vmatrix} = (-3) \cdot \begin{vmatrix} 1 & 4 & -1 & -2 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 5 & -5 \\ 0 & 1 & 2 & 3 \end{vmatrix} =$$

Ezek után csak a 4. sorból kell a 2.-at kivonnunk, hogy a 2. oszloppal is készen legyünk (balra). Majd a 3. sorból „emelünk ki 5-öt”:

$$= (-3) \cdot \begin{vmatrix} 1 & 4 & -1 & -2 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 5 & -5 \\ 0 & 0 & 2 & 1 \end{vmatrix} = (-3) \cdot 5 \cdot \begin{vmatrix} 1 & 4 & -1 & -2 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 2 & 1 \end{vmatrix} =$$

Végül a 4. sorból kivonjuk a 3. sor 2-szeresét, amivel már felsőháromszög-mátrixot kapunk. Ennek a determinánsa pedig a főátlóbeli elemek szorzata:

$$= (-3) \cdot 5 \cdot \begin{vmatrix} 1 & 4 & -1 & -2 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 3 \end{vmatrix} = (-3) \cdot 5 \cdot (1 \cdot 1 \cdot 1 \cdot 3) = -45$$

Tehát a feladatbeli determináns értéke: -45. □

1.4.5. A determináns kiszámítása

Az 1.4.9. Feladat egyben példát mutat a determináns hatékony kiszámítására szolgáló algoritmus működésére is. Amint az látható, az eljárás lényegében azonos a Gauss-eliminációval – annak egy variánsa. Ez nem is meglepő, hiszen az 1.4.7. Tételben felsorolt lépések szoros rokonságot mutatnak az 1.3.1. Definícióban látott elemi sorekvivalens átalakításokkal; az 1.4.6. Tétel által célként kitűzött felsőháromszög-mátrix, illetve csupa 0 sor pedig megfelel a lépcsős alaknak, illetve az 1.3.1. Definícióbeli *(iv)* lépésnek.

Ezek alapján elmondható, hogy a Gauss-elimináció – annak is az első fázisa – a lineáris egyenletrendszerek megoldása mellett a determináns hatékony kiszámítására is használható. Valójában az algoritmus még egyszerűbb is ebben az esetben, az alábbi formában írható le.

Az eljárás során nyilvántartjuk annak a sornak (és egyben oszlopnak) a számát, ahol az elimináció épp tart; ezt i jelöli. Ezen kívül nyilvántartunk egy $D \in \mathbb{R}$ számot, ami a determináns értékének a változásait követi nyomon. (Pontosabban: D mindig egy olyan számot jelöl, amivel az aktuálisan tárolt mátrix determinánsát megszorozva a bemenetként kapott mátrix determinánsát kapjuk.)

GAUSS-ELIMINÁCIÓ – A DETERMINÁNS KISZÁMÍTÁSÁRA

Bemenet: Egy $(n \times n)$-es A mátrix.

0. lépés. $i \leftarrow 1$, $D \leftarrow 1$

1. lépés.
- Ha $a_{i,i} = 0$, akkor folytassuk a **2. lépésnél**.
- $D \leftarrow a_{i,i} \cdot D$
- Szorozzuk meg az i-edik sort $\frac{1}{a_{i,i}}$-vel
- Ha $i = n$, akkor PRINT „$\det A =$”, D; STOP.
- Minden $i < t \leq n$ esetén adjuk a t-edik sorhoz az (imént módosított) i-edik sor $(-a_{t,i})$-szeresét.
- $i \leftarrow i + 1$
- Folytassuk az **1. lépésnél**.

2. lépés.
- Ha $i < n$ és van olyan $i < t \leq n$, amelyre $a_{t,i} \neq 0$, akkor:
  - ▶ Cseréljük fel az i-edik sort a t-edik sorral.
  - ▶ $D \leftarrow (-1) \cdot D$
  - ▶ Folytassuk az **1. lépésnél**.
- PRINT „$\det A = 0$”; STOP.

A fenti eljárás tehát a Gauss-elimináció korábbi, a 39. oldalon látott formáját a D értékének nyomon követése mellett annyiban módosítja, hogy a „karika” (vagyis a következő vezéregyes pozíciója) sosem tér ki a főátlóból. Ehelyett amint erre az algoritmus eredeti verziójában (pontosan a 2. lépés $j \leftarrow j + 1$ utasítása miatt) sor kerülne, az algoritmus leáll és közli, hogy a determináns értéke 0. Ennek a döntésnek a helyességét a következő indokolja: ha a főátlóban álló karikában és alatta minden elem 0, akkor ez az oszlop megfelel a felsőháromszög-mátrix definíciójának, a karikát a főátló következő pozíciójába mozgatva folytathatnánk az eliminációt (a fenti, egyszerűsített formájában) egészen a felsőháromszög-mátrix eléréséig. Azonban amikor végül ezt elérjük, a főátló elemeinek szorzata – és így a bemenetként kapott mátrix determinánsa is – mindenképp 0 lesz, mert a főátlóbeli 0 elem az elimináció végéig megmarad. Fölösleges tehát folytatni az eliminációt. (Alternatív indoklásnak mondhatjuk azt is, hogy ha az eliminációt annak az eredeti, a 39. oldalon írt formájában folytatnánk, akkor végül biztosan keletkezne csupa 0 sor. Valóban: a mátrix $(n \times n)$-es, így ha nincs minden oszlopban vezéregyes, akkor minden sorba sem juthat. A csupa 0 sor miatt tehát a determináns mindenképp 0 lesz.)

1.4.10. Feladat. Az $(n\times n)$-es A mátrix főátlóján kívül mindenhol 1-es áll, a főátló minden eleme p, ahol $p\in\mathbb{R}$ paraméter. Határozzuk meg $\det A$ értékét.

Megoldás: Vonjuk ki A első sorát az összes többiből; ezek a lépések (összesen $(n-1)$ darab) az 1.4.7. Tétel szerint a determinánst nem változtatják. A kapott mátrixnak tehát az első sora változatlan, a másodiktól lefelé viszont a főátlóban mindenhol $p-1$, az első oszlopban $1-p$, máshol pedig $1-1=0$ áll:

$$\begin{vmatrix} p & 1 & 1 & \dots & 1 \\ 1 & p & 1 & \dots & 1 \\ 1 & 1 & p & \dots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \dots & p \end{vmatrix} = \begin{vmatrix} p & 1 & 1 & \dots & 1 \\ 1-p & p-1 & 0 & \dots & 0 \\ 1-p & 0 & p-1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1-p & 0 & 0 & \dots & p-1 \end{vmatrix} =$$

Most az első oszlophoz adjuk hozzá az összes többit; ezzel a determináns ismét nem változik. A mátrixnak tehát most csak az első oszlopa módosul, a legfelső eleme $p+(n-1)\cdot 1 = p+n-1$, a többi pedig $(1-p)+(p-1)=0$ lesz:

$$= \begin{vmatrix} p+n-1 & 1 & 1 & \dots & 1 \\ 0 & p-1 & 0 & \dots & 0 \\ 0 & 0 & p-1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & p-1 \end{vmatrix}$$

Ezzel sikerült felsőháromszög-mátrixot elérnünk; ennek a determinánsa a főátlóbeli elemek szorzata, vagyis $(p+n-1)\cdot(p-1)^{n-1}$. Mivel a determináns a megtett lépések során nem változott, ezért $\det A = (p+n-1)\cdot(p-1)^{n-1}$. □

A fenti feladat arra mutat példát, hogy az 1.4.7. Tétel a Gauss-eliminációnál kreatívabban is használható. Az algoritmus előnye, hogy azzal minden konkrét determinánst meg lehet határozni, akár számítógéppel is; ez azonban nem jelenti azt, hogy ravaszabb ötletekkel ne lehetne rövidebb úton célhoz érni.

1.4.6. Determináns és lineáris egyenletrendszerek

A determináns bevezetését az motiválta, hogy az eldönti (vagy „determinálja”), hogy egy $(n\times n)$-es lineáris egyenletrendszer egyértelműen megoldható-e (lásd a 43. oldalt). Most már bebizonyíthatjuk, hogy ez valóban igaz, mert megismertük a két terület közötti kapcsolatot: a Gauss-eliminációt.

1.4.11. Tétel. *Legyen $(A|\underline{b})$ egy n változós, n egyenletből álló lineáris egyenletrendszer kibővített együtthatómátrixa. (A tehát csak a változók együtthatóit tartalmazza, $\underline{b}$ az egyenletek jobb oldalaiból áll. Más szóval: a 37. oldalon látható kibővített együtthatómátrixban a vonaltól balra A, attól jobbra $\underline{b}$ áll.) Ekkor az egyenletrendszer akkor és csak akkor egyértelműen megoldható, ha* $\det A\neq 0$.

Bizonyítás: Futtassuk $(A|b)$-re a Gauss-eliminációt (a 39. oldalon leírt formájában). Az algoritmus által megtett lépések az együtthatómátrix determinánsát megváltoztathatják ugyan, de (az 1.4.7. Tétel szerint) annak a nulla vagy nemnulla mivoltát nem. Más szóval: ha $\det A = 0$, akkor az együtthatómátrix végig 0 determinánsú marad, ha viszont $\det A \neq 0$, akkor végig nemnulla determinánsú együtthatómátrixok keletkeznek. (Pontosabban: ez az állítás megszűnik igaz lenni, ha az algoritmus első fázisának 3. lépésében csupa 0 sor elhagyása történik – utána ugyanis az együtthatómátrix már nem négyzetes, a determinánsáról nem beszélhetünk.)

A Gauss-elimináció (mint minden lineáris egyenletrendszer esetében) az alábbi három lehetőség valamelyikével ér véget:

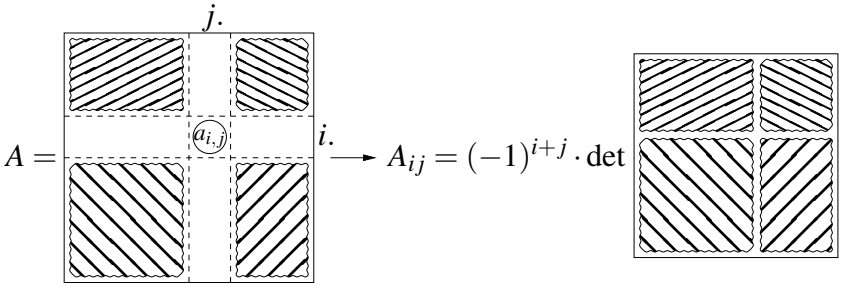
1. *Az egyenletrendszer nem megoldható.* Ez azt jelenti, hogy az első fázis 3. lépésében tilos sor keletkezett. Mivel a tilos sorban a vonaltól balra mindenhol 0 áll, az együtthatómátrix determinánsa ezen a ponton 0 (hiszen csupa 0 sora van), így eredetileg is $\det A = 0$ volt.
2. *Az egyenletrendszernek végtelen sok megoldása van.* Ez azt jelenti, hogy a lépcsős alakban az együtthatómátrixnak már kevesebb sora van, mint oszlopa (így a vezéregyest nem tartalmazó oszlopoknak megfelelő változók szabad paraméterek lesznek). Mivel A eredetileg $(n \times n)$-es, ezért az első fázis 3. lépésében keletkeznie kellett csupa 0 sornak. Ez megint azt jelenti, hogy az együtthatómátrix determinánsa ezen a ponton 0, így eredetileg is $\det A = 0$.
3. *Az egyenletrendszer megoldása egyértelmű.* Ez azt jelenti, hogy a redukált lépcsős alakban az együtthatómátrix determinánsa 1, mert a főátlójában csupa 1-es, mindenhol máshol 0 áll. Mivel az együtthatómátrix determinánsa végül nem nulla, ezért eredetileg is $\det A \neq 0$.

Mindezekből látszik, hogy az egyértelmű megoldhatóságnak valóban szükséges és elégséges feltétele, hogy $\det A \neq 0$. □

1.4.7. A kifejtési tétel

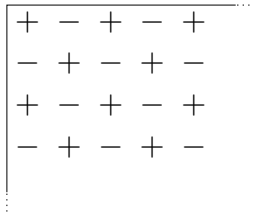
Az alább következő tétel arra ad lehetőséget, hogy egy $(n \times n)$-es mátrix determinánsának kiszámítását visszavezessük n darab $(n-1) \times (n-1)$-es determináns kiszámítására. Ez elsőre nem hangzik „jó üzletnek”: így egyetlen Gauss-elimináció helyett n darab (bár valamivel rövidebb) Gauss-elimináció árán kapjuk meg a determinánst. A tétel haszna nem is abban rejlik, hogy ezzel egy általános négyzetes mátrix determinánsa hatékonyan volna kiszámítható. (Nem nehéz végiggondolni, hogy a tétel ismételt alkalmazásával végeredményben ugyanúgy $n!$ szorzatot kellene kiértékelnünk, mintha a definíció szerint számolnánk a determinánst – erről pedig már láttuk, hogy általában reménytelen.) A kifejtési tétel elsősorban a determinánssal kapcsolatos elméleti állítások bizonyításában válik nélkülözhetetlenné – de látunk később olyan számolási feladatot is, ahol nehezebben boldogulnánk nélküle. A tétel kimondásához először bevezetjük az alábbi fogalmat.

1.4.12. Definíció. *Az $(n \times n)$-es A mátrix $a_{i,j}$ eleméhez tartozó* előjeles aldeterminánst *úgy kapjuk, hogy A-ból elhagyjuk az i-edik sorát és a j-edik oszlopát, majd a kapott $(n-1) \times (n-1)$-es mátrix determinánsát $(-1)^{i+j}$-nel szorozzuk (lásd az 1.6. ábrát). Ennek a jele: $A_{i,j}$.*



1.6. ábra

Az $A_{i,j}$ definíciójában szereplő $(-1)^{i+j}$ előjelet *sakktáblaszabálynak* is szokták nevezni, mert ez i és j függvényében úgy változik, mint a sakktábla mezőinek színei (lásd az 1.7. ábrát).



1.7. ábra

1.4.13. Tétel. (Kifejtési tétel)
Ha az $(n \times n)$-es A mátrix valamelyik sorának, vagy oszlopának minden elemét megszorozzuk a hozzá tartozó előjeles aldeterminánsértékével és a kapott n darab kéttényezős szorzatot összeadjuk, akkor A determinánsának értékét kapjuk.

A tétel állítása képletben, ha azt az i-edik sorra alkalmazzuk (vagyis „az i-edik sor szerint fejtjük ki a determinánst") a következő:

$$\det A = a_{i1}A_{i1} + a_{i2}A_{i2} + \ldots + a_{in}A_{in}.$$

Hasonlóan, ha a j-edik oszlop szerint fejtünk ki, akkor a tétel ezt állítja:

$$\det A = a_{1j}A_{1j} + a_{2j}A_{2j} + \ldots + a_{nj}A_{nj}.$$

Egy konkrét példa:

$$\begin{vmatrix} 2 & 3 & 4 & 5 \\ 6 & 7 & 8 & 9 \\ 10 & 11 & 12 & 13 \\ 14 & 15 & 16 & 17 \end{vmatrix} = -3 \cdot \begin{vmatrix} 6 & 8 & 9 \\ 10 & 12 & 13 \\ 14 & 16 & 17 \end{vmatrix} +$$

$$+7 \cdot \begin{vmatrix} 2 & 4 & 5 \\ 10 & 12 & 13 \\ 14 & 16 & 17 \end{vmatrix} - 11 \cdot \begin{vmatrix} 2 & 4 & 5 \\ 6 & 8 & 9 \\ 14 & 16 & 17 \end{vmatrix} + 15 \cdot \begin{vmatrix} 2 & 4 & 5 \\ 6 & 8 & 9 \\ 10 & 12 & 13 \end{vmatrix}$$

Itt tehát a (4×4)-es determinánst a 2. oszlopa szerint fejtettük ki (és az előjeles aldeterminánsok sakktáblaszabályból fakadó előjeleit kivittük a szorzatok elé).

A kifejtési tétel bizonyítása: Tegyük fel, hogy a tételt például az i-edik sorra alkalmazzuk. Amikor $\det A$ értékét az 1.4.4. definíció szerint kiszámítjuk, minden bástyaelhelyezés pontosan egy elemet tartalmaz az i-edik sorból. Ezért megtehetjük, hogy a definícióban szereplő $n!$ darab szorzatot aszerint csoportosítjuk, hogy az i-edik sorból melyik elemet tartalmazzák. Ha az egyik ilyen csoport tagjai az i-edik sorból az $a_{i,j}$ elemet tartalmazzák, akkor ezekből a szorzatokból kiemelhető az $a_{i,j}$ közös tényező. Ezt mind az n csoportra elvégezve $\det A$ így írható:

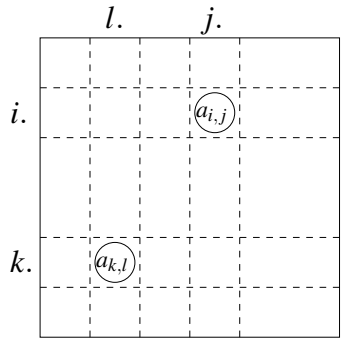
$$\det A = a_{i,1}(\ldots) + a_{i,2}(\ldots) + \ldots + a_{i,n}(\ldots).$$

Gondoljuk most meg, hogy a fenti felírásban mi kerül az $a_{i,j}$ elemmel szorzott $(\ldots)$ zárójelbe. Mivel n-tényezős szorzatokból emeltük ki az $a_{i,j}$ közös tényezőt, ezért a zárójelben $(n-1)$-tényezős szorzatok előjeles összege áll. Másrészt mivel a kiemelés előtt a szorzatok minden sorból és oszlopból egy elemet tartalmaztak, ezért $a_{i,j}$ kiemelése után olyan $(n-1)$-tényezős szorzatok keletkeznek, amelyek az i-edik sor és a j-edik oszlop kivételével az A minden további sorából és oszlopából pontosan egy elemet tartalmaznak. Más megfogalmazásban: az $a_{i,j}$-vel szorzott $(\ldots)$ zárójelben éppen az $A_{i,j}$ értelmezésében szereplő $(n-1) \times (n-1)$-es determináns definíció szerinti kiszámításakor keletkező szorzatok előjelezett összege áll.

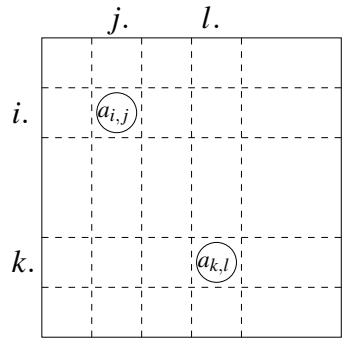
Mindez elmondható akkor is, ha az i-edik sor helyett a j-edik oszlopra alkalmazzuk a tételt. Ezzel pedig a kifejtési tétel állítását majdnem bebizonyítottuk: $\det A$ definíció szerinti kiszámításánál, illetve a kifejtési tétel tetszőleges sorra vagy oszlopra való alkalmazásánál ugyanazt az $n!$ darab előjeles szorzatot adjuk össze. Azt kell még belátni, hogy minden ilyen n-tényezős szorzat ugyanazt az előjelet kapja a kétféle kiszámítás során.

Ehhez érdemes a determináns definíciójában szereplő előjelezési szabályt némileg átfogalmazni. Hogyan kaphatjuk meg a bástyaelhelyezésnek megfelelő permutáció felírása és az inverziószám kiszámítása nélkül a szorzat előjelét? Ha például az $a_{i,j}$ és az $a_{k,l}$ elemek szerepelnek a bástyaelhelyezésben, akkor ez azt jelenti, hogy az annak megfelelő π permutációban az i-edik helyen j, a k-adikon l áll: $\pi = (\pi_1, \ldots, \pi_i = j, \ldots, \pi_k = l, \ldots, \pi_n)$. Ha most j és l inverzióban állnak, az azt

jelenti, hogy $j > l$, vagyis $a_{i,j}$ és $a_{k,l}$ Északkelet-Délnyugat pozícióban vannak egymáshoz képest (mint az 1.8a ábrán). Ha viszont j és l nem állnak inverzióban, akkor $j < l$, így $a_{i,j}$ és $a_{k,l}$ Északnyugat-Délkelet pozícióban vannak (mint az 1.8b ábrán).



1.8a ábra



1.8b ábra

Ezt felhasználva a determináns definíciójában szereplő előjelezési szabály így is fogalmazható: egy n-tényezős szorzathoz tartozó előjel $(-1)^I$, ahol I jelöli a megfelelő bástyaelhelyezésben az elemek közül kiválasztható, egymással ÉK-DNy pozícióban álló párok számát.

Térjünk most vissza a kifejtési tétel bizonyítására. Válasszunk egy

$$a_{1,\pi_1} \cdot a_{2,\pi_2} \cdot \ldots \cdot a_{n,\pi_n} \tag{1.4.3}$$

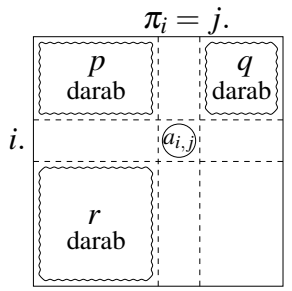
szorzatot és vizsgáljuk meg, hogy valóban ugyanazt az előjelet kapja-e a definíció szerinti, illetve a kifejtési tétel szerinti számításnál. Az imént láttuk, hogy a definíció szerinti számításkor az előjel az ÉK-DNy pozícióban álló elempárok számától függ; jelöljük ezt a számot I-vel. Tegyük fel, hogy a kifejtési tétel szerinti számításnál az 1.4.3 szorzat az $a_{i,j}A_{i,j}$ tagban szerepel (vagyis $\pi_i = j$ és a kifejtési tételt vagy az i-edik sorra, vagy a j-edik oszlopra alkalmaztuk). Ekkor az előjelet két dolog befolyásolja: egyrészt a sakktáblaszabály szerinti $(-1)^{i+j}$ előjel, másrészt az $A_{i,j}$ értelmezésében szereplő determináns definíció szerinti kiszámításában az

$$a_{1,\pi_1} \cdot \ldots \cdot a_{i-1,\pi_{i-1}} \cdot a_{i+1,\pi_{i+1}} \cdot \ldots \cdot a_{n,\pi_n} \tag{1.4.4}$$

szorzathoz tartozó előjel. Az utóbbi előjel pedig $(-1)^J$, ahol J most az 1.4.4 szorzatnak megfelelő bástyaelhelyezésben jelöli az ÉK-DNy pozícióban álló párok számát.

Mivel az 1.4.4 szorzatnak megfelelő bástyaelhelyezés csak az $a_{i,j}$ elemben különbözik az eredeti, az 1.4.3 szorzatnak megfelelő bástyaelhelyezéstől, ezért J annyival kevesebb I-nél, amennyi ÉK-DNy pozícióban álló elempárban $a_{i,j}$ szerepel. Az $a_{i,j}$ sora és oszlopa az A mátrixot négy részre osztja; jelölje p, q, illetve r a négy rész közül a bal felsőben, a jobb felsőben, illetve a bal alsó részben lévő, az 1.4.3 bástyaelhelyezéshez tartozó elemek számát (lásd az 1.9. ábrát). Ekkor $a_{i,j}$ éppen a jobb felső és a bal alsó részben levő, összesen $q + r$ darab elemmel áll ÉK-DNy pozícióban. A fentiek szerint tehát $J = I - (q + r)$.

Vegyük még észre azt is, hogy mivel az első $i - 1$ sor mindegyikében pontosan egy elem szerepel a bástyaelhelyezésben, ezért $p + q = i - 1$. Hasonlóan, az első $j - 1$ oszlop mindegyikében is pontosan egy elem szerepel, így $p + r = j - 1$. Ezek



1.9. ábra

összevetéséből: $q+r=i+j-(2p+2)$. Ezt behelyettesítve az előző bekezdés végén kapott összefüggésbe:

$$J = I-(i+j)+2(p+1).$$

Összegezve tehát az eddig mondottakat, a kifejtési tétel szerinti számításnál az 1.4.3 szorzat előjele:

$$(-1)^{i+j}\cdot(-1)^J = (-1)^{i+j}\cdot(-1)^{I-(i+j)+2(p+1)} = (-1)^I\cdot(-1)^{2(p+1)} = (-1)^I.$$

Ez valóban megegyezik a szorzat 1.4.3 definíció szerinti előjelével, így a tételt beláttuk. □

1.4.14. Feladat. A p paraméter minden értékére számítsuk ki az alábbi determinánst.

$$\begin{vmatrix} p & 2 & 3 \\ 4 & p & 5 \\ 6 & 0 & p \end{vmatrix}$$

Megoldás: Természetesen számolhatnánk a Gauss-eliminációval is, de a paraméter ezt kényelmetlenné teszi. (Különösen akkor, ha leosztunk vele: ilyenkor külön kellene kezelni a $p=0$ esetet.) Ehelyett most a kifejtési tétellel oldjuk meg a feladatot. Bármelyik sor vagy oszlop szerint kifejthetünk, de praktikus olyat választani, amelyik tartalmazza a 0 elemet – így ugyanis eggyel kevesebb előjeles aldeterminánst kell majd kiértékelnünk. Fejtsünk ki például a 3. sor szerint:

$$\begin{vmatrix} p & 2 & 3 \\ 4 & p & 5 \\ 6 & 0 & p \end{vmatrix} = 6\cdot\begin{vmatrix} 2 & 3 \\ p & 5 \end{vmatrix} - 0\cdot\begin{vmatrix} p & 3 \\ 4 & 5 \end{vmatrix} + p\cdot\begin{vmatrix} p & 2 \\ 4 & p \end{vmatrix} =$$

A középső tag tehát 0 (ezért választottuk a 3. sort), a másik két (2×2)-es determinánst a 47. oldalon látott szabály szerint számíthatjuk:

$$= 6(10-3p)+p(p^2-8) = p^3-26p+60.$$ □

1.4.8. Determináns a térgeometriában

Térgeometriai problémák megoldásában a determináns gyakran jut kulcsszerephez; alább a teljesség igénye nélkül mutatunk két ilyen alkalmazást.

A vektoriális szorzat

A térvektorok skaláris szorzatának fogalma a koordinátageometria egyik leghasznosabb segédeszköze. Van azonban egy másik, gyakran alkalmazott szorzatfogalom is a térvektorok körében, amely szintén sok feladat megoldását könnyíti. A művelet elnevezését az indokolja, hogy itt a szorzat eredménye maga is egy térvektor. Fontos kiemelni, hogy a vektoriális szorzat csak térvektorokra értelmezett, ennek a fogalomnak (szemben a skaláris szorzattal) semmilyen $\mathbb{R}^n$-re való kiterjesztése nem használatos.

1.4.15. Definíció. *Az $\underline{u}$ és $\underline{v}$ térvektorok* vektoriális szorzata *az az $\underline{u}\times\underline{v}$-vel jelölt térvektor, amelyre az alábbi feltételek fennállnak:*

- *$\underline{u}\times\underline{v}$ hossza: $|\underline{u}\times\underline{v}| = |\underline{u}|\cdot|\underline{v}|\cdot\sin\varphi$, ahol $|\underline{u}|$ és $|\underline{v}|$ a vektorok hosszát, φ pedig a bezárt szögüket jelöli;*
- *$\underline{u}\times\underline{v}$ merőleges $\underline{u}$-ra és $\underline{v}$-re;*
- *$\underline{u}$, $\underline{v}$ és $\underline{u}\times\underline{v}$ (ebben a sorrendben) jobbsodrású rendszert alkot (lásd a 2. oldalt).*

Ha $\underline{u}=\underline{0}$ vagy $\underline{v}=\underline{0}$, akkor definíció szerint $\underline{u}\times\underline{v}=\underline{0}$.

A definícióból rögtön látszik, hogy ez a művelet nem kommutatív: $\underline{u}\times\underline{v}$ és $\underline{v}\times\underline{u}$ egymás ellentett vektorai. A vektoriális szorzatnak több fizikai alkalmazása is van (például a forgatónyomaték vagy mágneses mezőben a töltésre ható erő meghatározásánál). Térkoordinátageometriában azért hasznos, mert két (nem párhuzamos) vektorra merőleges vektor előállítására sok feladatban van szükség. Ehhez persze az kell, hogy $\underline{u}\times\underline{v}$ könnyen kiszámítható legyen $\underline{u}$ és $\underline{v}$ ismeretében; erről szól az alábbi tétel.

1.4.16. Tétel. *Legyenek $\underline{u}=(u_1,u_2,u_3)$ és $\underline{v}=(v_1,v_2,v_3)$ térvektorok. Ekkor*

$$\underline{u}\times\underline{v}=\left(\begin{vmatrix} u_2 & u_3\\ v_2 & v_3\end{vmatrix},-\begin{vmatrix} u_1 & u_3\\ v_1 & v_3\end{vmatrix},\begin{vmatrix} u_1 & u_2\\ v_1 & v_2\end{vmatrix}\right).$$

A tételt nem bizonyítjuk (ehhez kicsit el kellene mélyednünk a vektoriális szorzat tulajdonságainak vizsgálatában). A tételbeli képlet megjegyzését viszont segíti, ha azt az alábbi alakban írjuk:

$$\underline{u}\times\underline{v}=\begin{vmatrix} \underline{i} & \underline{j} & \underline{k}\\ u_1 & u_2 & u_3\\ v_1 & v_2 & v_3\end{vmatrix},$$

ahol $\underline{i}$, $\underline{j}$, illetve $\underline{k}$ jelölik a koordinátarendszerben az x, y, illetve z tengelyek irányába mutató egységvektorokat. Persze ennek a képletnek az értelmezésekor joggal merül fel a kérdés: mi értelme olyan mátrix determinánsáról beszélni, amelynek bizonyos elemei térvektorok? A válasz általában az, hogy semmi: a determináns definíciója számokból álló négyzetes mátrixokra vonatkozott. Ha viszont a szóban forgó mátrixnak – mint ahogyan a fenti képletben is – pontosan egy sorát (vagy oszlopát) töltik ki térvektorok, akkor az 1.4.4. Definíció alkalmazható: minden bástyaelhelyezés egyetlen térvektort és $(n-1)$ számot tartalmaz, ezek szorzata pedig akadálytalanul képezhető (a szóban forgó vektort szorozzuk az $(n-1)$ darab szám szorzatával). Így a determináns definíció szerinti kiszámításakor a mátrixban álló vektorok egy lineáris kombinációját kapjuk. Végiggondolható, hogy a determinánssal kapcsolatban kimondott tételek bizonyítása változatlanul működik, így a tételek is érvényben maradnak ilyen esetekre is (de például a Gauss-elimináció már nem volna alkalmazható, hiszen egy „vektorral leosztani” nem lehet). A fenti képlet tehát végül is precíz értelemmel is megtölthető – mégis, leginkább úgy érdemes rá tekinteni, mint az 1.4.16. Tétel megjegyzését segítő eszközre: valóban, ha a (3×3)-as determinánst az 1.4.13. Tételt alkalmazva az első sora szerint kifejtjük, akkor az

$$\begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix} \cdot \underline{i} - \begin{vmatrix} u_1 & u_3 \\ v_1 & v_3 \end{vmatrix} \cdot \underline{j} + \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix} \cdot \underline{k}$$

vektort kapjuk; az 1.4.16. Tétel pedig épp erről állítja, hogy az azonos $\underline{u}\times\underline{v}$-vel.

A vektoriális szorzat használatának bemutatására újból megoldjuk az 1.1.9. Feladatot:

1.4.17. Feladat. Írjuk fel az $A(3;3;1)$, $B(5;2;4)$ és $C(8;5;0)$ pontokra illeszkedő S sík egyenletét.

Megoldás: S-nek három pontját is ismerjük, így ismét elegendő egy $\underline{n}$ normálvektorát meghatároznunk. Mivel $\overrightarrow{AC}=(5;2;-1)$ és $\overrightarrow{AB}=(2;-1;3)$ párhuzamosak S-sel, ezért $\underline{n}=\overrightarrow{AC}\times\overrightarrow{AB}$ jó normálvektor lesz. Ezt az 1.4.16. Tétel szerint határozzuk meg:

$$\overrightarrow{AC}\times\overrightarrow{AB}=\begin{vmatrix} \underline{i} & \underline{j} & \underline{k} \\ 5 & 2 & -1 \\ 2 & -1 & 3 \end{vmatrix} = \begin{vmatrix} 2 & -1 \\ -1 & 3 \end{vmatrix}\cdot\underline{i} - \begin{vmatrix} 5 & -1 \\ 2 & 3 \end{vmatrix}\cdot\underline{j} + \begin{vmatrix} 5 & 2 \\ 2 & -1 \end{vmatrix}\cdot\underline{k} =$$

$$= \Big(2\cdot 3-(-1)\cdot(-1)\Big)\cdot\underline{i} - \Big(5\cdot 3-2\cdot(-1)\Big)\cdot\underline{j} + \Big(5\cdot(-1)-2\cdot 2\Big)\cdot\underline{k} = 5\underline{i}-17\underline{j}-9\underline{k}$$

Így $\underline{n}=(5;-17;-9)$ normálvektora S-nek. Ebből (például) A-t használva felírhatjuk S egyenletét: $5x-17y-9z=-45$. □

A fenti feladat megoldásának szempontjából a vektoriális szorzat definíciójának csak egyetlen eleme fontos: hogy $\underline{u}\times\underline{v}$ merőleges $\underline{u}$-ra és $\underline{v}$-re. Vannak azonban a fogalomnak olyan alkalmazásai is, ahol $\underline{u}\times\underline{v}$ hossza játssza a főszerepet: $|\underline{u}|\cdot|\underline{v}|\cdot\sin\varphi$ nem más, mint az $\underline{u}$ és $\underline{v}$ által kifeszített (vagyis $|\underline{u}|$ és $|\underline{v}|$ oldalhosszú) paralelogramma területe (hiszen $|\underline{v}|\cdot\sin\varphi$ az $\underline{u}$ oldalhoz tartozó magasság nagysága). Ennek illusztrálására újból megoldjuk az 1.1.3. Feladatot:

1.4.18. Feladat. Határozzuk meg annak a háromszögnek a területét, amelynek csúcsai $A(5;4;8)$, $B(4;-1;4)$ és $C(3;1;2)$.

Megoldás: A keresett T terület a $\overrightarrow{CA} = (2;3;6)$ és $\overrightarrow{CB} = (1;-2;2)$ vektorok által kifeszített paralelogramma területének a fele – vagyis a $\overrightarrow{CA} \times \overrightarrow{CB}$ vektor hosszának a fele. $\overrightarrow{CA} \times \overrightarrow{CB}$ meghatározásához ismét az 1.4.16. Tételt használjuk:

$$\overrightarrow{CA} \times \overrightarrow{CB} = \begin{vmatrix} \underline{i} & \underline{j} & \underline{k} \\ 2 & 3 & 6 \\ 1 & -2 & 2 \end{vmatrix} = 18\underline{i} + 2\underline{j} - 7\underline{k}$$

A kapott $(18;2;-7)$ vektor hosszát a Pitagorasz-tétellel számítjuk (a részleteket lásd az 1.1.3. Feladat eredeti megoldásában): $|\overrightarrow{CA} \times \overrightarrow{CB}| = \sqrt{18^2 + 2^2 + (-7)^2} = \sqrt{377}$. Így a háromszög területe: $T = \frac{\sqrt{377}}{2}$ □

A vegyesszorzat

A térvektoroknak egy további, hasznos szorzatfogalma is ismert: ez a vegyesszorzat, amely három térvektorhoz rendel egyetlen számot. Ez a fogalom valójában ötvözete a korábban megismert két szorzatnak:

1.4.19. Definíció. *Az $\underline{u}$, $\underline{v}$ és $\underline{w}$ térvektorok* vegyesszorzata *az $(\underline{u} \times \underline{v}) \cdot \underline{w}$ skalár (ahol a „$\times$” a vektoriális, a „$\cdot$” a skaláris szorzatot jelöli). A vegyesszorzat jelölése egyszerűen az egymás mellé írás: $\underline{u}\ \underline{v}\ \underline{w}$.*

Ha az $\underline{u}$, $\underline{v}$ és $\underline{w}$ térvektorokat az origóba állítjuk, akkor egyértelműen meghatároznak – más szóval *kifeszítenek* – egy paralelepipedont. Az alábbi tétel állítása szerint ennek térfogata – előjeltől eltekintve – nem más, mint a vegyesszorzat.

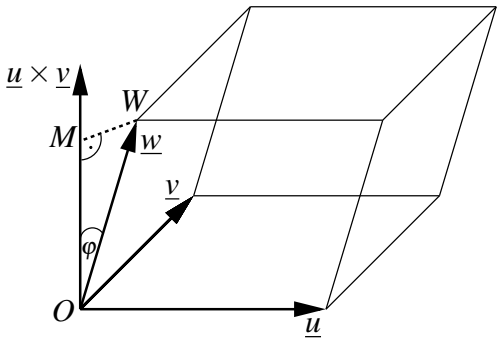
1.4.20. Tétel. *Az $\underline{u}$, $\underline{v}$ és $\underline{w}$ térvektorok által kifeszített paralelepipedon V térfogatára $V = |\underline{u}\ \underline{v}\ \underline{w}|$.*

Bizonyítás: V-t az $\underline{u}$ és $\underline{v}$ által kifeszített pararelogramma T területének és az ehhez az oldalhoz tartozó m magasságnak a szorzataként kapjuk meg. Fentebb már láttuk, hogy $T = |\underline{u} \times \underline{v}|$, $\underline{u} \times \underline{v}$ iránya pedig merőleges az $\underline{u}$ és $\underline{v}$ síkjára. Az m magasság az OMW derékszögű háromszög OM oldala, ahol O az origó, W az (origóba állított) $\underline{w}$ vektor végpontja, M pedig a W-ből az $\underline{u} \times \underline{v}$ egyenesére állított merőleges talppontja (lásd az 1.10. ábrát). Az OMW derékszögű háromszögből $m = OM = |\underline{w}| \cdot \cos\varphi$, ahol φ jelöli $\underline{u} \times \underline{v}$ és $\underline{w}$ bezárt szögét. Ezekből

$$\underline{u}\ \underline{v}\ \underline{w} = (\underline{u} \times \underline{v}) \cdot \underline{w} = |\underline{u} \times \underline{v}| \cdot |\underline{w}| \cdot \cos\varphi = T \cdot m = V$$

valóban következik.

A fenti számítás és az 1.10. ábra feltételezi, hogy $\underline{u} \times \underline{v}$ és $\underline{w}$ az $\underline{u}$ és $\underline{v}$ síkjának azonos oldalára esik és így $0^\circ \leq \varphi \leq 90^\circ$. Ez azonban nem feltétlen igaz: ha az 1.10. ábrán $\underline{u}$-t és $\underline{v}$-t felcseréljük, akkor $\underline{u} \times \underline{v}$ az ellentettjére változik és így $\underline{u} \times \underline{v}$ és $\underline{w}$



1.10. ábra

szöge az ábrán látott φ-nek a $180°$-ra való kiegészítő szöge. Ez azonban az állítás igazságát nem befolyásolja: $\cos(180° - \varphi) = -\cos\varphi$, így $\underline{u}$ és $\underline{v}$ cseréje után $\underline{u}\ \underline{v}\ \underline{w}$ is az ellentettjére változik, de $|\underline{u}\ \underline{v}\ \underline{w}|$ változatlan. □

Koordinátáikkal adott térvektorok vegyesszorzatának kiszámítása elvileg nem ütközik akadályba, hiszen a skaláris, illetve a vektoriális szorzat kiszámítására már ismerünk egy-egy képletet. Az alábbi tétel ennél mégis többet mond.

1.4.21. Tétel. *Legyenek* $\underline{u} = (u_1, u_2, u_3)$, $\underline{v} = (v_1, v_2, v_3)$ *és* $\underline{w} = (w_1, w_2, w_3)$ *térvektorok. Ekkor*

$$\underline{u}\ \underline{v}\ \underline{w} = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}.$$

Bizonyítás: Fejtsük ki a fenti determinánst (lásd az 1.4.13. Tételt) a 3. sora szerint:

$$\begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} = w_1 \cdot \begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix} + w_2 \cdot \left(- \begin{vmatrix} u_1 & u_3 \\ v_1 & v_3 \end{vmatrix} \right) + w_3 \cdot \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix}.$$

Ez pedig az 1.4.16 és az 1.1.2. Tételek szerint valóban nem más, mint $\underline{u} \times \underline{v}$ és $\underline{w}$ skaláris szorzata, vagyis $\underline{u}\ \underline{v}\ \underline{w}$. □

Az 1.4.20. és az 1.4.21. Tételekből levonható közös tanulság az, hogy egy (3×3)-as determináns értéke a sorai, mint térvektorok által kifeszített paralelepipedon előjeles térfogata. Ez persze ebben a formában valóban csak a (3×3)-as esetre vonatkozik, hiszen n darab $\mathbb{R}^n$-beli vektorra ez az állítás nincs értelmezve. A részletek mellőzésével azonban megemlítjük, hogy mind a paralelepipedon, mind pedig a térfogat fogalma általánosítható $\mathbb{R}^n$-re és az 1.4.21. Tétel megfelelő kiterjesztése is igaz marad. Még ha ezt a munkát nem is végezzük el, a determinánssal kapcsolatos feladatok, állítások megértéséhez és használatához segíthet az a szemlélet, ha a determinánsra mint egyfajta előjeles térfogatfogalomra tekintünk.

1.4.22. Feladat. Határozzuk meg annak a tetraédernek a térfogatát, amelynek csúcsai $A(2;3;4)$, $B(2;4;2)$, $C(3;4;9)$ és $D(3;6;3)$.

Megoldás: Az $ABCD$ tetraéder V_t térfogata épp hatoda az $\overrightarrow{AB}$, $\overrightarrow{AC}$ és $\overrightarrow{AD}$ vektorok

által kifeszített paralelepipedon V_p térfogatának. Valóban: $V_t = \frac{T \cdot m}{3}$, ahol T (például) az ABC oldal területe, m pedig az ehhez tartozó magasság nagysága. Itt T nyilván fele az $\overrightarrow{AB}$ és $\overrightarrow{AC}$ vektorok kifeszítette paralelogramma területének, m pedig azonos a paralelepipedon ehhez az oldalhoz tartozó magasságával. Így $V_p = 2T \cdot m$, ami csakugyan indokolja a $V_p = 6 \cdot V_t$ összefüggést.

Így elég meghatározni a V_p térfogatot, ami tehát az $\overrightarrow{AB} = (0;1;-2)$, $\overrightarrow{AC} = (1;1;5)$ és $\overrightarrow{AD} = (1;3;-1)$ vektorok kifeszítette paralelepipedon térfogata. Ez az 1.4.20 és az 1.4.21. Tételek szerint az alábbi determináns értéke (vagy annak ellentettje), amit Gauss-eliminációval határozunk meg:

$$\begin{vmatrix} 0 & 1 & -2 \\ 1 & 1 & 5 \\ 1 & 3 & -1 \end{vmatrix} = - \begin{vmatrix} 1 & 1 & 5 \\ 0 & 1 & -2 \\ 1 & 3 & -1 \end{vmatrix} = - \begin{vmatrix} 1 & 1 & 5 \\ 0 & 1 & -2 \\ 0 & 2 & -6 \end{vmatrix} = - \begin{vmatrix} 1 & 1 & 5 \\ 0 & 1 & -2 \\ 0 & 0 & -2 \end{vmatrix} = 2.$$

Így a paralelepipedon térfogata 2, az $ABCD$ tetraéderé pedig ebből $\frac{1}{3}$. □

1.5. Műveletek mátrixokkal

Mátrixokkal már találkoztunk a lineáris egyenletrendszerek, illetve a determináns kapcsán. Most „saját jogukon” foglalkozunk velük.

1.5.1. Definíció. *Adott $k,n \geq 1$ egészek esetén* $(k \times n)$-es mátrixnak *nevezünk egy k sorból és n oszlopból álló táblázatot, amelynek minden cellájában egy valós szám áll. A $(k \times n)$-es mátrixok halmazát $\mathbb{R}^{k \times n}$ jelöli. Az A mátrix i-edik sorának és j-edik oszlopának kereszteződésében álló elemet (továbbra is) $a_{i,j}$ jelöli (és hasonlóan a $B,C,\ldots$ mátrixok esetében $b_{i,j}$, $c_{i,j}$, stb.). Az $\mathbb{R}^{k \times n}$-en értelmezett, „$+$”-szal jelölt* összeadást *és tetszőleges $\lambda \in \mathbb{R}$ esetén a „$\cdot$”-tal (vagy egyszerűen egymás mellé írással) jelölt* skalárral való szorzást *az alábbi egyenlőségek szerint értelmezzük:*

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \ldots & a_{1,n} \\ a_{2,1} & a_{2,2} & \ldots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k,1} & a_{k,2} & \ldots & a_{k,n} \end{pmatrix} + \begin{pmatrix} b_{1,1} & b_{1,2} & \ldots & b_{1,n} \\ b_{2,1} & b_{2,2} & \ldots & b_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{k,1} & b_{k,2} & \ldots & b_{k,n} \end{pmatrix} =$$

$$= \begin{pmatrix} a_{1,1}+b_{1,1} & a_{1,2}+b_{1,2} & \ldots & a_{1,n}+b_{1,n} \\ a_{2,1}+b_{2,1} & a_{2,2}+b_{2,2} & \ldots & a_{2,n}+b_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k,1}+b_{k,1} & a_{k,2}+b_{k,2} & \ldots & a_{k,n}+b_{k,n} \end{pmatrix},$$

$$\lambda \cdot \begin{pmatrix} a_{1,1} & a_{1,2} & \ldots & a_{1,n} \\ a_{2,1} & a_{2,2} & \ldots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k,1} & a_{k,2} & \ldots & a_{k,n} \end{pmatrix} = \begin{pmatrix} \lambda a_{1,1} & \lambda a_{1,2} & \ldots & \lambda a_{1,n} \\ \lambda a_{2,1} & \lambda a_{2,2} & \ldots & \lambda a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda a_{k,1} & \lambda a_{k,2} & \ldots & \lambda a_{k,n} \end{pmatrix}.$$

A mátrixok összeadása és skalárral szorzása tehát tagonként történik – ugyanúgy, mint az oszlopvektoroké $\mathbb{R}^n$-ben. A definícióból az is következik, hogy az $A+B$ összeg csak akkor értelmezett, ha A és B sorainak és oszlopainak száma is megegyezik. Így például ha

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \text{ és } B = \begin{pmatrix} 5 & 0 \\ -1 & \sqrt{2} \end{pmatrix}, \text{ akkor}$$

$$A+B = \begin{pmatrix} 6 & 2 \\ 2 & 4+\sqrt{2} \end{pmatrix} \text{ és } (-2)\cdot A = \begin{pmatrix} -2 & -4 \\ -6 & -8 \end{pmatrix}$$

Ha $\mathbb{R}^{k\times n}$-et csak az összeadás és skalárral szorzás szempontjából vizsgáljuk, akkor $\mathbb{R}^{k\cdot n}$-hez (vagyis a $k\cdot n$ magas oszlopvektorokhoz) képest a különbség csak jelölésbeli: teljesen mindegy, hogy $k\cdot n$ darab számot táblázatban, egy oszlopban vagy akár csigavonalban írva tárolunk, ha a műveletek egyébként ugyanúgy működnek. Így aztán nem is meglepő, hogy az alábbi tétel igaz: ez az 1.2.2. Tétel állításának csak átfogalmazása.

1.5.2. Tétel. *Legyen $A,B,C \in \mathbb{R}^{k\times n}$ és $\lambda,\mu \in \mathbb{R}$. Ekkor igazak az alábbiak:*
(i) $A+B = B+A$, *(vagyis a mátrixösszeadás kommutatív);*
(ii) $(A+B)+C = A+(B+C)$, *(vagyis a mátrixösszeadás asszociatív);*
(iii) $\lambda\cdot(A+B) = \lambda\cdot A+\lambda\cdot B$;
(iv) $(\lambda+\mu)\cdot A = \lambda\cdot A+\mu\cdot A$;
(v) $\lambda\cdot(\mu\cdot A) = (\lambda\mu)\cdot A$.

Mátrixok kivonását az $\mathbb{R}^n$ esetéhez hasonlóan az 1.5.1. Definíciót használva értelmezzük: az $\mathbb{R}^{k\times n}$-beli A és B mátrixokra $A-B = A+(-1)\cdot B$. Szintén az $\mathbb{R}^n$-beli nullvektorral analóg az $\mathbb{R}^{k\times n}$-beli *nullmátrix* fogalma, amelyet egyszerűen 0 jelöl: ennek minden eleme 0, így $A+0 = A$ minden $A \in \mathbb{R}^{k\times n}$-re igaz.

Érdemes kiemelni azt is, hogy a sorvektorok és az oszlopvektorok valójában speciális mátrixok: egy n hosszú sorvektor $(1\times n)$-es, egy n magas oszlopvektor pedig $(n\times 1)$-es mátrixnak tekinthető. Így az $\mathbb{R}^n$-ben értelmezett műveletek speciális esetei a mátrixokra bevezetett műveleteknek.

1.5.1. Mátrix transzponáltja

Egy mátrix tekinthető úgy is mint egymás mellé írt oszlopvektorok egy sorozata, de úgy is, mint egymás alá írt sorvektorok egy listája. A kettő közötti átjárást teszi lehetővé a következő nagyon egyszerű fogalom.

1.5.3. Definíció. *A $(k\times n)$-es A mátrix* transzponáltjának *nevezzük az $(n\times k)$-as B mátrixot, ha $b_{i,j} = a_{j,i}$ teljesül minden $1\le i\le n$ és $1\le j\le k$ esetén. Ennek a jele: $B = A^T$.*

A transzponálást elképzelhetjük úgy is, hogy A-t „tükrözzük” a bal fölső sarkából induló „45°-os egyenesre” (amit azért nem nevezünk főátlónak, mert A nem

feltétlen négyzetes mátrix). Így A i-edik oszlopából A^T i-edik sora lesz (persze oszlopvektor helyett sorvektorként) és hasonlóan, A j-edik sorának elemei adják A^T j-edik oszlopát. Ha például

$$A = \begin{pmatrix} 2 & 3 & 4 & 5 & 6 \\ 7 & 8 & 9 & 10 & 11 \\ 12 & 13 & 14 & 15 & 16 \end{pmatrix}, \text{ akkor } A^T = \begin{pmatrix} 2 & 7 & 12 \\ 3 & 8 & 13 \\ 4 & 9 & 14 \\ 5 & 10 & 15 \\ 6 & 11 & 16 \end{pmatrix}.$$

A transzponált fogalma természetesen sor- és oszlopvektorokra (mint speciális mátrixokra) is alkalmazható: ha például $\underline{x}$ oszlopvektor, akkor $\underline{x}^T$ az $\underline{x}$-szel azonos elemekből álló sorvektor.

A determináns alaptulajdonságainak megismerésekor fontos volt, hogy azok sorokra és oszlopokra egyaránt érvényesek: így volt ez az 1.4.6., az 1.4.7. és az 1.4.13. Tételek esetében is. A transzponált fogalmának a birtokában ennek a jelenségnek pontos formát tudunk adni – hiszen A^T sorai épp A oszlopai és fordítva.

1.5.4. Tétel. *Minden A négyzetes mátrixra* $\det A^T = \det A$.

Bizonyítás: A bizonyítást először egy példán illusztráljuk. Az alábbi A mátrixban bekeretezett bástyaelhelyezésről a 46. oldalon már láttuk, hogy az annak megfelelő permutáció $\pi = (4,2,1,5,3)$, aminek az inverziószáma 5. Így $\det A$ definíció szerinti kiszámításakor az ebből keletkező szorzat negatív előjelet kap.

$$A = \begin{pmatrix} 2 & 3 & 4 & \boxed{5} & 6 \\ 7 & \boxed{8} & 9 & 10 & 11 \\ \boxed{12} & 13 & 14 & 15 & 16 \\ 17 & 18 & 19 & 20 & \boxed{21} \\ 22 & 23 & \boxed{24} & 25 & 26 \end{pmatrix}, \quad A^T = B = \begin{pmatrix} 2 & 7 & \boxed{12} & 17 & 22 \\ 3 & \boxed{8} & 13 & 18 & 23 \\ 4 & 9 & 14 & 19 & \boxed{24} \\ \boxed{5} & 10 & 15 & 20 & 25 \\ 6 & 11 & 16 & \boxed{21} & 26 \end{pmatrix}$$

A^T determinánsának definíció szerinti kiszámításakor is megjelenik ugyanez a szorzat (lásd fent). Itt a megfelelő permutáció a $\pi' = (3,2,5,1,4)$, aminek az inverziószáma „véletlenül” szintén 5, így az előjel is marad negatív.

Rátérve a tétel bizonyítására, legyen A tetszőleges $(n \times n)$-es mátrix és legyen $B = A^T$. Meg fogjuk mutatni, hogy $\det A$ és $\det B$ definíció szerinti kiszámításakor ugyanazok a szorzatok keletkeznek és mindegyik szorzat ugyanazt az előjelet is kapja a két esetben. Ebből a tétel állítása nyilván következni fog.

Legyen tehát $\pi = (\pi_1, \pi_2, \ldots, \pi_n)$ tetszőleges permutáció. Ennek $\det A$ kiszámításakor a $(-1)^{I(\pi)} \cdot a_{1,\pi_1} \cdot a_{2,\pi_2} \cdot \ldots \cdot a_{n,\pi_n}$ előjelezett szorzat felel meg. Mivel $a_{i,j} = b_{j,i}$ minden $1 \leq i, j \leq n$ esetén, ezért ugyanez a szorzat (egyelőre előjeltől eltekintve) megjelenik B-ben is $b_{\pi_1,1} \cdot b_{\pi_2,2} \cdot \ldots \cdot b_{\pi_n,n}$ alakban.

Legyen ezért π' az a permutáció, amiben az 1 a π_1-edik helyen, a 2 a π_2-edik helyen, stb., az n a π_n-edik helyen áll; ekkor π'-t a π *inverzének* nevezzük. Az elnevezést az indokolja, hogy ha a π permutációt olyan kölcsönösen egyértelmű függvénynek fogjuk fel, ami az $1,2,\ldots,n$ számokhoz rendre a $\pi_1, \pi_2, \ldots, \pi_n$ értékeket

rendeli, akkor π' a π függvénytani értelemben vett inverze. Nyilván igaz, hogy π' is permutáció (és hogy minden permutáció inverze egyértelműen létezik, valamint π' inverze π). A fogalomra fentebb már láttunk példát: $\pi = (4,2,1,5,3)$ inverze a $\pi' = (3,2,5,1,4)$ permutáció (és viszont).

A B elemeiből készített fenti szorzat tehát $b_{1,\pi'_1} \cdot b_{2,\pi'_2} \cdot \ldots \cdot b_{n,\pi'_n}$ alakban írható, így a $(-1)^{I(\pi')}$ előjelet kapja. Készen leszünk ezért a tétel bizonyításával, ha megmutatjuk, hogy $I(\pi) = I(\pi')$ igaz minden π permutációra és annak a π' inverzére. Legyen ezért a π permutációban $\pi_i = k$ és $\pi_j = \ell$, ekkor a π' inverz permutációban $\pi'_k = i$ és $\pi'_\ell = j$. A k és ℓ tagok π-ben definíció szerint akkor állnak inverzióban, ha $i < j$, de $k > \ell$. Ez viszont szintén definíció szerint azt jelenti, hogy π'-ben az i és j tagok állnak inverzióban, hiszen $\ell < k$, de $\pi'_\ell = j > i = \pi'_k$. Összefoglalva: π-ben π_i és π_j akkor és csak akkor állnak inverzióban, ha π'-ben i és j állnak inverzióban. (Ezt a fenti példán illusztrálva: π-ben a $\pi_1 = 4$ és $\pi_3 = 1$ tagok inverzióban állnak, ennek megfelelően π'-ben az 1 és a 3 állnak inverzióban.) Így tehát a π-ben inverzióban álló elempárok kölcsönösen egyértelműen megfeleltethetők a π'-ben inverzióban álló elempároknak, amiből $I(\pi) = I(\pi')$ valóban következik. □

Érdemes felidézni, hogy a determináns 1.4.4. szakaszban bizonyított alaptulajdonságai egyaránt igazak voltak a mátrix soraira és oszlopaira, illetve az ezeken végzett lépésekre. A fenti tétel ennek a jelenségnek a hátterét adja, hiszen A sorai megfelelnek A^T oszlopainak. Sőt, az 1.5.4. Tételre hivatkozva rövidíthetők az 1.4.4. szakaszban adott bizonyítások: a determináns alaptulajdonságait elegendő sorokra bizonyítani, mert ezeket A^T-ra alkalmazva az oszlopokra vonatkozó változatot kapjuk.

1.5.2. Mátrixok szorzása

A mátrixokkal végezhető műveletek közül eddig az összeadást, a skalárral szorzást és a transzponálást ismertük meg. Ezekre korlátozva (ahogy azt már fentebb is említettük) a mátrixok alig jelentenének újdonságot $\mathbb{R}^n$-hez képest. Az alábbi definíció viszont egy olyan, alapvető fontosságú műveletet vezet be, amely oszlopvektorok körében már nem létezik.

1.5.5. Definíció. *A $(k \times n)$-es A és az $(n \times m)$-es B mátrixok* szorzatának *nevezzük és $A \cdot B$-vel jelöljük azt a $(k \times m)$-es C mátrixot, amelyre minden $1 \leq i \leq k$ és $1 \leq j \leq m$ esetén*

$$c_{i,j} = a_{i,1} \cdot b_{1,j} + a_{i,2} \cdot b_{2,j} + \ldots + a_{i,n} \cdot b_{n,j}.$$

A definíció tehát az $A \cdot B$ szorzatmátrixot csak akkor értelmezi, ha A oszlopainak száma megegyezik B sorainak számával (a fenti definícióban mindkettőt n jelölte); ha pedig ez a feltétel fennáll, akkor az $A \cdot B$ szorzat A-tól a sorainak, B-től az oszlopainak a számát „örökli”. Az $A \cdot B$ i-edik sorának és j-edik oszlopának kereszteződésében álló elemet az A i-edik sora és a B j-edik oszlopa határozza meg – mégpedig úgy, hogy ezeknek a „skaláris szorzatát” képezzük (vagyis az egyik

első elemét a másik első elemével, a másodikat a másodikkal, stb. az n-ediket az n-edikkel összeszorozzuk és ezt az n darab kétténye zős szorzatot összeadjuk).

Az $A \cdot B$ szorzat kiszámítását megkönnyíti, ha B-t az 1.11. ábrán látható módon feltoljuk. Ekkor a $C = A \cdot B$ szorzat A-tól jobbra, B alatt keletkezik és a $c_{i,j}$ kiszámításához szükséges A-beli i-edik sor, illetve B-beli j-edik oszlop épp a $c_{i,j}$ elem helyétől balra, illetve attól fölfelé találhatók, ami a definícióbeli képlet alkalmazását kényelmessé teszi.

A mátrixszorzásnak fontos speciális esete az, amikor A egy n hosszúságú $\underline{u}$ sorvektor, B pedig egy $\mathbb{R}^n$-beli $\underline{v}$ oszlopvektor. Ekkor $\underline{u} \cdot \underline{v}$ valóban a skaláris szorzat fogalmát általánosítja: $\underline{u} \cdot \underline{v} = u_1 v_1 + u_2 v_2 + \ldots + u_n v_n$, ahol u_i, illetve v_i az $\underline{u}$, illetve az $\underline{v}$ i-edik koordinátáját jelöli.

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \ldots & a_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{i,1} & a_{i,2} & \ldots & a_{i,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k,1} & a_{k,2} & \ldots & a_{k,n} \end{pmatrix} \qquad \begin{pmatrix} b_{1,1} & \ldots & b_{1,j} & \ldots & b_{1,m} \\ b_{2,1} & \ldots & b_{2,j} & \ldots & b_{2,m} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ b_{n,1} & \ldots & b_{n,j} & \ldots & b_{n,m} \end{pmatrix} = B$$

$$\begin{pmatrix} & & \uparrow & & \\ & & \boxed{c_{i,j}} & & \\ \leftarrow & & & & \end{pmatrix} = C = A \cdot B$$

1.11. ábra

1.5.6. Feladat. Legyen $A = \begin{pmatrix} 2 & -1 & -5 \\ 1 & 4 & -3 \end{pmatrix}$ és $B = \begin{pmatrix} 5 & -4 \\ -2 & 3 \end{pmatrix}$. Elvégezhetők-e az alábbi műveletek? Ha igen, adjuk meg az eredményt:

a) $4A + 9B$ b) $A \cdot B$ c) $B \cdot A$ d) $B \cdot A - 2A$ e) $A^T \cdot B^T$

Megoldás: a) A $4A$ mátrix (2×3)-as, $9B$ viszont (2×2)-es, ezért nem adhatók össze.

b) A-nak 3 oszlopa, B-nek 2 sora van; mivel ezek nem egyenlők, az $A \cdot B$ szorzat nem létezik.

c) B oszlopainak száma már egyezik A sorainak számával, ezért a $B \cdot A$ szorzás elvégezhető:

$$\begin{pmatrix} 2 & -1 & -5 \\ 1 & 4 & -3 \end{pmatrix} = A$$

$$B = \begin{pmatrix} 5 & -4 \\ -2 & 3 \end{pmatrix} \begin{pmatrix} c_{1,1} & c_{1,2} & c_{1,3} \\ c_{2,1} & c_{2,2} & c_{2,3} \end{pmatrix} = C = B \cdot A$$

Itt definíció szerint

$$\begin{aligned}
c_{1,1} &= 5\cdot 2+(-4)\cdot 1 = 6,\\
c_{1,2} &= 5\cdot(-1)+(-4)\cdot 4 = -21,\\
c_{1,3} &= 5\cdot(-5)+(-4)\cdot(-3) = -13,\\
c_{2,1} &= (-2)\cdot 2+3\cdot 1 = -1,\\
c_{2,2} &= (-2)\cdot(-1)+3\cdot 4 = 14 \text{ és}\\
c_{2,3} &= (-2)\cdot(-5)+3\cdot(-3) = 1.
\end{aligned}$$

Így tehát $B\cdot A = \begin{pmatrix} 6 & -21 & -13 \\ -1 & 14 & 1 \end{pmatrix}$.

d) Mivel $B\cdot A$ és $2A$ egyaránt (2×3)-as, a kivonás elvégezhető:

$$B\cdot A - 2A = \begin{pmatrix} 6 & -21 & -13 \\ -1 & 14 & 1 \end{pmatrix} - \begin{pmatrix} 4 & -2 & -10 \\ 2 & 8 & -6 \end{pmatrix} = \begin{pmatrix} 2 & -19 & -3 \\ -3 & 6 & 7 \end{pmatrix}.$$

e) A^T (3×2)-es, B^T (2×2)-es, ezért a szorzás elvégezhető, az $A^T\cdot B^T$ szorzatot most D-vel jelöljük:

$$\begin{pmatrix} 5 & -2 \\ -4 & 3 \end{pmatrix} = B^T$$

$$A^T = \begin{pmatrix} 2 & 1 \\ -1 & 4 \\ -5 & -3 \end{pmatrix}\begin{pmatrix} d_{1,1} & d_{1,2} \\ d_{2,1} & d_{2,2} \\ d_{3,1} & d_{3,2} \end{pmatrix} = D = A^T\cdot B^T$$

$A^T\cdot B^T$ kiszámításakor látható, hogy más sorrendben, de ugyanazokat a számolásokat végezzük, mint a c) feladatban $B\cdot A$ esetében: $d_{1,1} = 2\cdot 5+1\cdot(-4) = 6 = c_{1,1}$, $d_{1,2} = 2\cdot(-2)+1\cdot 3 = -1 = c_{2,1}$, stb. Általában: $d_{i,j}$ és $c_{j,i}$ kiszámítása ugyanazokat a műveleteket igényli, így $d_{j,i} = c_{i,j}$ minden i és j esetén. Ezért a végeredményként kapott D is a c) feladatban kapott C transzponáltja: $A^T\cdot B^T = \begin{pmatrix} 6 & -1 \\ -21 & 14 \\ -13 & 1 \end{pmatrix}$. □

A fenti feladat c) és e) részének megoldása között mutatkozó szoros kapcsolatot általánosítja az alábbi, egyszerűsége dacára is sokszor hasznos állítás.

1.5.7. Állítás. *Ha az A és B mátrixokra az $A\cdot B$ szorzat létezik, akkor $B^T\cdot A^T$ is létezik és $(A\cdot B)^T = B^T\cdot A^T$.*

Bizonyítás: Legyen A $(k\times n)$-es. Ekkor az $A\cdot B$ szorzat létezése miatt B $(n\times m)$-es valamilyen m-re. Ezért B^T $(m\times n)$-es és A^T $(n\times k)$-as, így a $B^T\cdot A^T$ szorzat valóban létezik. Ráadásul $A\cdot B$ $(k\times m)$-es és $B^T\cdot A^T$ $(m\times k)$-as, így $(A\cdot B)^T$ és $B^T\cdot A^T$ azonos méretűek.

Legyen $X = A\cdot B$ és $Y = B^T\cdot A^T$. A mátrixszorzás definíciója szerint $x_{i,j}$ minden $1\le i\le k$ és $1\le j\le m$ esetén az A i-edik sorának és a B j-edik oszlopának a skaláris

szorzata: $x_{i,j} = a_{i,1}b_{1,j} + a_{i,2}b_{2,j} + \ldots + a_{i,n}b_{n,j}$. Mivel B^T j-edik sora elemről elemre azonos B j-edik oszlopával és A^T i-edik oszlopa pedig A i-edik sorával, ezért $y_{j,i}$ definíció szerinti meghatározásakor ugyanazt a számítást végezzük, mint $x_{i,j}$ esetében: $y_{j,i} = b_{1,j}a_{i,1} + b_{2,j}a_{i,2} + \ldots + b_{n,j}a_{i,n}$. Így $x_{i,j} = y_{j,i}$ minden i és j esetén igaz, ami az $X^T = Y$ állítást bizonyítja. □

A mátrixszorzás tulajdonságai

Látni fogjuk, hogy a mátrixokkal sok szempontból hasonlóan lehet „számolni”, mint a valós számokkal – de legalább ilyen lényeges a különbségek tisztázása: vannak olyan, a számok körében alapvetőnek tekintett műveleti tulajdonságok, amelyek mátrixokra már nem igazak. A legfontosabb ezek közül, hogy *a mátrixszorzás nem kommutatív*, vagyis $A \cdot B$ és $B \cdot A$ nem azonosak. Erre már az 1.5.6. Feladatban is láttunk példát: ott az $A \cdot B$ szorzat nem is létezett, míg $B \cdot A$ igen. Előfordulhat az is, hogy mindkét szorzat létezik, de nem azonos méretűek: ha A $(k \times n)$-es, B $(n \times k)$-as és $k \neq n$. De még ha azonos méretűek is, akkor sem igaz (általában), hogy egyenlők; például ha

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \text{ és } B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \text{ akkor } A \cdot B = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \text{ és } B \cdot A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Ez a példa rávilágít egy másik nagyon fontos jelenségre is: a mátrixok körében előfordulhat, hogy $A \cdot B = 0$, miközben $A \neq 0$ és $B \neq 0$, ahol 0 a nullmátrixot jelöli.

A mátrixokkal való számoláskor tehát fontos ügyelni arra, hogy a fenti két „hamis szabályt” ne alkalmazzuk – de ezektől eltekintve (és figyelembe véve, hogy a szorzás nem végezhető el két tetszőleges mátrix között) minden „elvárható” műveleti tulajdonság teljesül; ezt fejezi ki (az 1.5.2. Tétel és) az alábbi tétel.

1.5.8. Tétel. *Az alábbi műveleti tulajdonságok bármely A, B és C mátrixra és $\lambda \in \mathbb{R}$ skalárra fennállnak. (Ezeket úgy kell érteni, hogy ha az egyenlőség egyik oldalán a műveletek elvégezhetők, akkor a másik oldalon is, és a két oldal egyenlő.)*

(i) $(\lambda A) \cdot B = \lambda(A \cdot B) = A \cdot (\lambda B)$*;*

(ii) $A \cdot (B+C) = A \cdot B + A \cdot C$ *és* $(B+C) \cdot A = B \cdot A + C \cdot A$ *(vagyis a mátrixszorzás az összeadásra nézve* disztributív*);*

(iii) $(A \cdot B) \cdot C = A \cdot (B \cdot C)$ *(vagyis a mátrixszorzás asszociatív).*

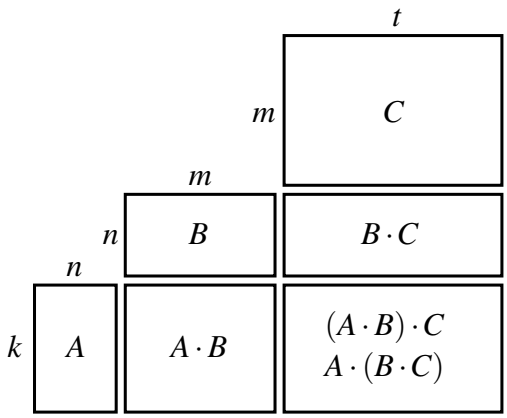
Bizonyítás: Mindhárom állítás a valós számok elemi műveleti tulajdonságaiból következik. Az egyetlen nehézség az lesz, hogy ezeket egyszerre sok (ráadásul kettős indexeléssel jelölt) számra kell alkalmazni.

Kezdjük az *(i)* bizonyításával. Az első egyenlőséget látjuk be, a második ezzel analóg. Legyen A $(k \times n)$-es mátrix; ekkor λA is $(k \times n)$-es, így mindkét oldal elvégezhetősége azzal ekvivalens, hogy B-nek n sora van. Legyen ezért B $(n \times m)$-es és legyen $X = A \cdot B$, $Y = \lambda(A \cdot B)$ és $Z = (\lambda A) \cdot B$. Ekkor a mátrixszorzás definíciója szerint minden $1 \leq i \leq k$ és $1 \leq j \leq m$ esetén $x_{i,j} = a_{i,1} \cdot b_{1,j} + \ldots + a_{i,n} \cdot b_{n,j}$, így $y_{i,j} = \lambda \cdot (a_{i,1} \cdot b_{1,j} + \ldots + a_{i,n} \cdot b_{n,j})$. Ismét a definíciókból következik, hogy

$z_{i,j} = (\lambda a_{i,1}) \cdot b_{1,j} + \ldots + (\lambda a_{i,n}) \cdot b_{n,j}$. Az utóbbiból λ-t kiemelve látszik, hogy $y_{i,j} = z_{i,j}$ minden i és j esetén igaz, amivel *(i)*-et beláttuk.

A *(ii)* állításból is az első egyenlőséget látjuk csak be. Ismét legyen A $(k \times n)$-es, ekkor mindkét oldal elvégezhetősége azzal ekvivalens, hogy B és C is n sorú mátrixok és oszlopaik száma egyenlő. Legyen ezért B és C $(n \times m)$-es és legyen $X = A \cdot B$, $Y = A \cdot C$ és $Z = A \cdot (B + C)$. Most definíció szerint $x_{i,j} = a_{i,1} \cdot b_{1,j} + \ldots + a_{i,n} \cdot b_{n,j}$, $y_{i,j} = a_{i,1} \cdot c_{1,j} + \ldots + a_{i,n} \cdot c_{n,j}$ és $z_{i,j} = a_{i,1} \cdot (b_{1,j} + c_{1,j}) + \ldots + a_{i,n} \cdot (b_{n,j} + c_{n,j})$. Látszik, hogy $z_{i,j} = x_{i,j} + y_{i,j}$ minden i-re és j-re igaz, ami épp a *(ii)* állítás.

(iii) bizonyításához legyen megint A $(k \times n)$-es. A bal oldalon $A \cdot B$ akkor elvégezhető, ha B $(n \times m)$-es valamilyen m-re, ekkor a szorzat $(k \times m)$-es lesz; ezt C-vel akkor tudjuk jobbról megszorozni, ha az $(m \times t)$-es valamilyen t-re (lásd az 1.12. ábrát). A jobb oldal elvégezhetősége azzal ekvivalens, hogy egyrészt a $B \cdot C$ szorzat és így B is n sorú, másrészt hogy B oszlopainak és C sorainak száma megegyezik; az utóbbit m-mel jelölve ismét azt kapjuk, hogy B-nek $(n \times m)$-esnek, C-nek $(m \times t)$-esnek kell lennie. Annyit láttunk eddig tehát be, hogy a két oldal elvégezhetősége ekvivalens (és mindkét oldalon $(k \times t)$-es szorzat keletkezik).



1.12. ábra

Legyen $X = A \cdot B$ és $Y = (A \cdot B) \cdot C$. Ekkor

$$y_{i,j} = x_{i,1} \cdot c_{1,j} + x_{i,2} \cdot c_{2,j} + \ldots + x_{i,m} \cdot c_{m,j}$$

minden i-re és j-re. Itt minden $1 \leq r \leq m$ esetén $x_{i,r}$ helyére behelyettesíthetjük az $X = A \cdot B$ szorzásból fakadó értékét:

$$y_{i,j} = (a_{i,1}b_{1,1} + a_{i,2}b_{2,1} + \ldots + a_{i,n}b_{n,1})c_{1,j} + (a_{i,1}b_{1,2} + a_{i,2}b_{2,2} + \ldots + a_{i,n}b_{n,2})c_{2,j} + \\ + \ldots + (a_{i,1}b_{1,m} + a_{i,2}b_{2,m} + \ldots + a_{i,n}b_{n,m})c_{m,j}.$$

A zárójeleket felbontva azt kapjuk, hogy $y_{i,j}$ végül is $n \cdot m$ darab háromtényezős szorzat összege: az összes $a_{i,r} \cdot b_{r,s} \cdot c_{s,j}$ típusú szorzaté, ahol $1 \leq r \leq n$ és $1 \leq s \leq m$. Egy ezzel teljesen analóg számolás mutatja (ezt nem részletezzük), hogy

az $A \cdot (B \cdot C)$ mátrix megfelelő eleme is ugyanennek az $n \cdot m$ szorzatnak az összege. Ezzel *(iii)*-at is beláttuk. □

Van – a fenti tételben sorolt tulajdonságok mellett – a mátrixok és a számok szorzása között még egy fontos hasonlóság. A számok világában az 1 speciális szereppel bír: a vele végzett szorzás azonos a változatlanul hagyással, vagyis $a \cdot 1 = 1 \cdot a = a$ minden $a \in \mathbb{R}$ esetén igaz. Mátrixokra az ennek megfelelő fogalmat vezeti be az alábbi definíció.

1.5.9. Definíció. Egységmátrixnak *nevezzük azt az $(n \times n)$-es mátrixot, amelynek a (bal felső sarkot a jobb alsóval összekötő) főátlójában minden elem 1, az összes többi eleme pedig 0. Ennek a jele: E_n (vagy ha n értéke a szövegkörnyezetből egyértelmű, akkor egyszerűen csak E).*

Nem fog problémát okozni, hogy az 1.2.22. Definícióban E_n (illetve E) jelölte a standard bázist is, a két fogalom között a kapcsolat egyébként is szoros: az egységmátrix oszlopai épp a standard bázis $\underline{e}_1, \ldots, \underline{e}_n$ vektorai. Az alábbi állítás adja az egységmátrix elnevezésének a hátterét.

1.5.10. Állítás. *Tetszőleges $(k \times n)$-es A mátrixra $A \cdot E_n = A$ és $E_k \cdot A = A$ teljesül. Vagyis: a (megfelelő méretű) egységmátrixszal akár balról, akár jobbról végzett szorzás azonos a változatlanul hagyással.*

Bizonyítás: Az állítás azonnal következik a mátrixszorzás definíciójából: az $A \cdot E_n$ szorzatot C-vel jelölve

$$c_{i,j} = a_{i,1} \cdot 0 + \ldots + a_{i,j-1} \cdot 0 + a_{i,j} \cdot 1 + a_{i,j+1} \cdot 0 + \ldots + a_{i,n} \cdot 0 = a_{i,j}$$

adódik (lásd az 1.13. ábrát). Az $E_k \cdot A = A$ állítás bizonyítása ezzel analóg. □

1.5.11. Feladat. Igazak-e az alábbi egyenlőségek bármely $(n \times n)$-es A, B, C és D mátrixokra? (E az $(n \times n)$-es egységmátrixot jelöli. Egy $(n \times n)$-es X mátrixra X^2 jelöli az $X \cdot X$ szorzatot.)

a) $(A+B)(C+D) = AC + AD + BC + BD$

b) $(A+B)^2 = A^2 + 2AB + B^2$

c) $(A+E)(A-E) = A^2 - E$

Megoldás: a) Legyen $X = C + D$. Alkalmazható a disztributivitás (1.5.8. Tétel, *(ii)* állítás): $(A+B)(C+D) = (A+B)X = AX + BX$. Most X helyére $(C+D)$-t visszahelyettesítve: $(A+B)(C+D) = A(C+D) + B(C+D)$. Végül mindkét zárójelet felbontva: $(A+B)(C+D) = AC + AD + BC + BD$ (itt ismét a disztributivitást alkalmaztuk). Így az egyenlőség igaz.

$$\begin{pmatrix} 1 & 0 & \dots & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & \dots & 1 \end{pmatrix} = E_n$$

$$A = \begin{pmatrix} a_{1,1} & \dots & a_{1,j} & \dots & a_{1,n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{i,1} & \dots & a_{i,j} & \dots & a_{i,n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{k,1} & \dots & a_{k,j} & \dots & a_{k,n} \end{pmatrix} \quad \begin{pmatrix} & & \uparrow & & \\ & & \\ \leftarrow & & \boxed{c_{i,j}} & & \\ & & \\ & & \end{pmatrix} = C = A \cdot E_n$$

1.13. ábra

b) Alkalmazhatjuk a (már bizonyított) a) egyenlőséget $C = A$ és $D = B$ választással: $(A+B)^2 = (A+B)(A+B) = A^2 + AB + BA + B^2$. Ahhoz tehát, hogy a b) egyenlőség igaz legyen, $2AB = AB + BA$, vagyis $AB = BA$ kellene, hogy teljesüljön. Ez viszont (általában) nem igaz – így a b) egyenlet is sérül minden olyan esetben, amikor $AB \neq BA$. Például a 70. oldalon látott A-ra és B-re $(A+B)^2 = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$, de $A^2 + 2AB + B^2 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$.

c) Ismét alkalmazhatjuk az a) egyenlőséget $C = A$, $B = E$ és $D = -E$ választással: $(A+E)(A-E) = A^2 + A(-E) + EA + E(-E)$. Itt $A(-E) = -AE$ és $E(-E) = -E^2$ (ez az 1.5.8. Tétel, *(i)* állításából következik $\lambda = -1$ esetén). Az 1.5.10. Állítás szerint $AE = A$ és $E^2 = E$, így $(A+E)(A-E) = A^2 - A + A - E$. Itt $-A + A = 0$ (ahol 0 az $(n \times n)$-es nullmátrix), a 0-val végzett összeadás pedig azonos a változatlanul hagyással. Tehát $(A+E)(A-E) = A^2 - E$, az egyenlőség igaz. (Érdemes megfigyelni, hogy ennek a számolásnak lényeges eleme volt, hogy $AE = A$ és $EA = A$ egyaránt igaz. Ezért nem okozott problémát a mátrixszorzás kommutativitásának a hiánya – de például az $(A+B)(A-B) = A^2 - B^2$ egyenlőség általában már ugyanúgy hamis, mint a b).) □

1.5.12. Feladat. Legyen $A = \begin{pmatrix} 2 & -1 & 6 \\ 2 & 2 & 3 \\ 6 & -1 & 17 \\ 4 & -1 & 13 \end{pmatrix}$ és $\underline{b} = \begin{pmatrix} 12 \\ 24 \\ 46 \\ 32 \end{pmatrix}$. Oldjuk meg az $A \cdot \underline{x} = \underline{b}$ „mátrixegyenletet” – vagyis keressük meg az összes olyan $\underline{x}$-et, amire ez az egyenlet fennáll.

Megoldás: A-nak 3 oszlopa van, ezért $\underline{x}$-nek 3 sora kell legyen, hogy az $A \cdot \underline{x}$ szorzat létezzen. Mivel a szorzatnak 1 oszlopa van, ezért $\underline{x}$-nek is ennyi kell legyen (mert a

szorzat örökli $\underline{x}$ oszlopainak számát). Így tehát $\underline{x}$ egy $\underline{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$ oszlopvektor kell legyen (ezért is jelöltük eleve így). Elvégezve az $A \cdot \underline{x}$ szorzást:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \underline{x}$$

$$A = \begin{pmatrix} 2 & -1 & 6 \\ 2 & 2 & 3 \\ 6 & -1 & 17 \\ 4 & -1 & 13 \end{pmatrix} \begin{pmatrix} 2x_1 - x_2 + 6x_3 \\ 2x_1 + 2x_2 + 3x_3 \\ 6x_1 - x_2 + 17x_3 \\ 4x_1 - x_2 + 13x_3 \end{pmatrix} = A \cdot \underline{x}$$

Ezért az $A \cdot \underline{x} = \underline{b}$ mátrixegyenlet az alábbi lineáris egyenletrendszerrel egyenértékű:

$$\begin{aligned} 2x_1 - x_2 + 6x_3 &= 12 \\ 2x_1 + 2x_2 + 3x_3 &= 24 \\ 6x_1 - x_2 + 17x_3 &= 46 \\ 4x_1 - x_2 + 13x_3 &= 32 \end{aligned}$$

Ez a lineáris egyenletrendszer természetesen megoldható Gauss-eliminációval – de valójában ez már meg is történt: a 31. oldalon épp ezen az egyenletrendszeren futtattuk először az eliminációt. Így az ott kapott megoldás alapján $\underline{x} = \begin{pmatrix} 3 \\ 6 \\ 2 \end{pmatrix}$ a mátrixegyenlet egyetlen megoldása. □

A mátrixszorzás eddig megismert tulajdonságai mind a számokkal végzett műveletekkel való hasonlóságot – vagy éppen annak a hiányát emelték ki. Az alábbi tétel viszont már egy, a számok körében semmitmondó, a lineáris algebrában viszont nagyon sok alkalmazással bíró állítást mond ki.

1.5.13. Tétel. (A determinánsok szorzástétele)
Bármely A és B $(n \times n)$-es mátrixokra $\det(A \cdot B) = \det A \cdot \det B$ *teljesül.*

A tétel állítása távolról sem magától értetődő, a bizonyítása kicsit komolyabb erőfeszítéseket kíván. Alább következik egy lehetséges bizonyításnak a vázlata, a részletek végiggondolását az érdeklődő olvasóra bízzuk.

Bizonyítás (vázlat): Hajtsuk végre A sorain az 1.4.7. Tételben írt egyik lépést. Ekkor mind $\det(A \cdot B)$, mind pedig $\det A \cdot \det B$ értéke ugyanúgy változik meg, mint amit az 1.4.7. Tétel $\det A$ változásáról állít. Például: ha A i-edik sorát helyettesítjük sajátmagának és a j-edik sor λ-szorosának az összegével, akkor $\det(A \cdot B)$ és $\det A \cdot \det B$ is változatlan marad. Valóban: $\det A \cdot \det B$ esetében ez közvetlenül látszik, $\det(A \cdot B)$ esetében pedig azért igaz, mert A változásának hatására $A \cdot B$ is úgy változik meg, hogy annak az i-edik sora felülíródik sajátmagának és $A \cdot B$ j-edik sorának az összegével. Hasonlóan: ha A két sorát felcseréljük, akkor $\det(A \cdot B)$ és $\det A \cdot \det B$ is az ellentettjére változik, ha pedig A egy sorát megszorozzuk λ-val, akkor $\det(A \cdot B)$ és $\det A \cdot \det B$ is a λ-szorosára változik. Ezekkel analóg állítások mondhatók akkor is, ha B oszlopain (de nem a sorain) hajtjuk végre az 1.4.7. Tételben írt lépéseket.

A sorain, illetve B oszlopain ezeknek a lépéseknek az alkalmazásával elérhető, hogy A is és B is felsőháromszög-mátrixszá váljon. A esetében ez lényegében a Gauss-elimináció a determináns kiszámítására vonatkozó változatának a végrehajtását jelenti (lásd az 53. oldalt) azzal a különbséggel, hogy az elimináció nem áll meg akkor sem, ha valamelyik oszlopban a főátlóban és alatta mindenhol 0-t talál, hanem folytatódik egészen a felsőháromszög-mátrix eléréséig. B esetében pedig egy ezzel analóg, de az oszlopokra vonatkozó eljárásra van szükség: ez a főátlót fordított irányban, a jobb alsó saroktól a bal felsőig járja be és a sorokban hoz létre 0-kat a keletkező vezéregyesektől balra.

Tegyük fel tehát, hogy már A és B is felsőháromszög-mátrix. A mátrixszorzás definíciójából következik, hogy ekkor a $C = A \cdot B$ szorzat is felsőháromszög-mátrix és a főátlójának elemeire $c_{i,i} = a_{i,i} \cdot b_{i,i}$ teljesül. Ezért $\det(A \cdot B)$ az A és a B főátlójában álló elemeknek (összesen tehát $2n$ darabnak) a szorzata – és nyilván ugyanez mondható $\det A \cdot \det B$ értékéről is. Mivel $\det(A \cdot B)$ és $\det A \cdot \det B$ a felsőháromszög-mátrixok eléréséig (A és B esetében is) ugyanúgy változtak és végül egyenlők, ezért ugyanez igaz volt már az eredeti A-ra és B-re is. □

1.5.3. Mátrixszorzás és lineáris egyenletrendszerek

Az 1.5.12. Feladatban látott jelenség külön figyelmet érdemel: az $A \cdot \underline{x} = \underline{b}$ „mátrixegyenletről” kiderült, hogy ekvivalens azzal a lineáris egyenletrendszerrel, amelynek a kibővített együtthatómátrixa $(A|\underline{b})$. A lineáris egyenletrendszerek pedig korábban már a generált altér fogalma kapcsán is előkerültek. Ezek a kapcsolatok a lineáris algebra különböző területei között – egyszerűségük dacára – annyira fontosak, hogy külön tételt szentelünk nekik.

1.5.14. Tétel. *Legyenek $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n, \underline{b} \in \mathbb{R}^k$ vektorok és legyen A az $\underline{a}_i$-k egyesítésével keletkező $(k \times n)$-es mátrix. (A oszlopai tehát $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n$.) Ekkor az alábbi állítások ekvivalensek (vagyis ha bármelyikük igaz, akkor a másik kettő is):*

(i) Megoldható az $A \cdot \underline{x} = \underline{b}$ „mátrixegyenlet”.

(ii) Megoldható az $(A|\underline{b})$ kibővített együtthatómátrixú lineáris egyenletrendszer.

(iii) $\underline{b} \in \langle \underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n \rangle$

Bizonyítás: A *(ii)* és *(iii)* állítások ekvivalenciájára már az 1.2.3. Feladatban is láttunk példát és a 30. oldalon általában is esett róla szó. *(iii)* teljesülése azt jelenti, hogy létezik a $\lambda_1\underline{a}_1 + \lambda_2\underline{a}_2 + \ldots + \lambda_n\underline{a}_n = \underline{b}$ lineáris kombináció. Itt a $\lambda_1\underline{a}_1 + \lambda_2\underline{a}_2 + \ldots + \lambda_n\underline{a}_n$ vektor i-edik koordinátája minden $1 \leq i \leq k$ esetén $a_{i,1}\lambda_1 + a_{i,2}\lambda_2 + \ldots + a_{i,n}\lambda_n$ (mert A az $\underline{a}_j$-k egyesítése, ezért $a_{i,j}$ egyenlő az $\underline{a}_j$ i-edik koordinátájával). Következésképp $\lambda_1\underline{a}_1 + \lambda_2\underline{a}_2 + \ldots + \lambda_n\underline{a}_n = \underline{b}$ ekvivalens azzal, hogy $a_{i,1}\lambda_1 + a_{i,2}\lambda_2 + \ldots + a_{i,n}\lambda_n = b_i$ minden $1 \leq i \leq k$ esetén teljesül. Ezzel épp az $(A|\underline{b})$ lineáris egyenletrendszert kapjuk (az egyetlen különbség, hogy a változókat a szokásos $x_1, \ldots, x_n$ helyett $\lambda_1, \ldots, \lambda_n$ jelöli).

Az *(i)* és *(ii)* ekvivalenciájához először (az 1.5.12. Feladat megoldásának nyomvonalán haladva) vegyük észre, hogy $\underline{x}$ csak $\mathbb{R}^n$-beli oszlopvektor lehet (mert egyrészt n sora van, ha $A \cdot \underline{x}$ elvégezhető, másrészt 1 oszlopa van, ha $A \cdot \underline{x}$ is 1 oszlopú).

Az $\underline{x}$ j-edik koordinátáját minden $1 \leq j \leq n$ esetén x_j-vel jelölve az $A \cdot \underline{x}$ szorzat i-edik koordinátája a mátrixszorzás definíciója szerint $a_{i,1}x_1 + a_{i,2}x_2 + \ldots + a_{i,n}x_n$. Ezért $A \cdot \underline{x} = \underline{b}$ azzal ekvivalens, hogy $a_{i,1}x_1 + a_{i,2}x_2 + \ldots + a_{i,n}x_n = b_i$ teljesül minden $1 \leq i \leq k$ esetén – vagyis ismét az $(A|\underline{b})$ lineáris egyenletrendszert kapjuk. □

Érdemes megjegyezni, hogy a bizonyításából valójában több is kiolvasható, mint amit a tétel állít: nem csak a megoldhatóság ténye ekvivalens a három esetben, hanem a megoldások maguk is lényegében azonosak: *(i)*-ben az $\underline{x}$ vektor koordinátái, *(ii)*-ben az $(A|\underline{b})$ lineáris egyenletrendszer megoldásában a változók értékei, illetve *(iii)*-ban a $\underline{b}$-t az $\underline{a}_i$-kből kifejező lineáris kombináció együtthatói ugyanazok.

A fenti tétel szerint a lineáris egyenletrendszereket a továbbiakban az eddig használt $(A|\underline{b})$ kibővített együtthatómátrixos alak mellett $A \cdot \underline{x} = \underline{b}$ formában is írhatjuk – ez tényleg csak jelölésbeli különbség. Külön érdemes felfigyelni arra a szemléletre, amit a tétel *(i)* és *(iii)* állítása közötti ekvivalencia hordoz: ha az A mátrixot az $\underline{x}$ oszlopvektorral szorozzuk, akkor az eredményként kapott oszlopvektor az A oszlopainak egy lineáris kombinációja – mégpedig épp az $\underline{x}$ koordinátáival, mint együtthatókkal képzett lineáris kombinációja.

Fontossága miatt külön megfogalmazzuk az 1.5.14. Tétel *(ii)* és *(iii)* állítása közötti kapcsolatnak azt a speciális esetét, amikor $\underline{b}$ a nullvektor és megoldhatóság helyett egyértelmű megoldhatóságról beszélünk.

1.5.15. Következmény. *Legyenek $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n \in \mathbb{R}^k$ vektorok és legyen A az ezek egyesítésével keletkező $(k \times n)$-es mátrix. Ekkor az alábbi állítások ekvivalensek:*

(i) Az $A \cdot \underline{x} = \underline{0}$ lineáris egyenletrendszernek az egyetlen megoldása $\underline{x} = \underline{0}$.

(ii) Az $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n$ vektorok lineárisan függetlenek.

Bizonyítás: A következmény állításával az 1.2.13. Feladatban már találkoztunk: ott 4 darab $\mathbb{R}^4$-beli vektor lineáris függetlensége épp attól függött, hogy van-e annak a (4×4)-es lineáris egyenletrendszernek a csupa nullától különböző megoldása, amelynek az együtthatómátrixa a 4 vektor egyesítéséből állt.

Ugyanez általában is elmondható: az 1.2.12. Tétel szerint $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n$ akkor és csak akkor lineárisan független, ha $\lambda_1\underline{a}_1 + \lambda_2\underline{a}_2 + \ldots + \lambda_n\underline{a}_n = \underline{0}$ csak a triviális esetben, vagyis $\lambda_1 = \lambda_2 = \ldots = \lambda_n = 0$ mellett teljesül. Az 1.5.14. Tétel (illetve az utána írt megjegyzés) szerint ez ekvivalens azzal, hogy az $A \cdot \underline{x} = \underline{0}$ lineáris egyenletrendszer egyetlen megoldása az, hogy minden változó értéke 0. □

Érdemes megjegyezni, hogy mivel az $A \cdot \underline{x} = \underline{0}$ rendszernek $\underline{x} = \underline{0}$ biztosan megoldása, ezért a fenti következmény *(i)* állítása úgy is fogalmazható, hogy $A \cdot \underline{x} = \underline{0}$ egyértelműen megoldható. Ha A ráadásul négyzetes mátrix, akkor a következmény állítását továbbgondolva az alábbi, alapvető fontosságú összefüggésekre jutunk.

1.5.16. Tétel. *Legyen A $(n \times n)$-es mátrix. Ekkor az alábbi állítások ekvivalensek:*

(i) A oszlopai, mint $\mathbb{R}^n$-beli vektorok lineárisan függetlenek;

(ii) $\det A \neq 0$;

(iii) A sorai, mint n hosszú sorvektorok lineárisan függetlenek.

Bizonyítás: A oszlopainak lineáris függetlensége az 1.5.15. Következmény (és az utána írt megjegyzés) szerint azzal ekvivalens, hogy az $(A|\underline{0})$ kibővített együtthatómátrixú lineáris egyenletrendszer egyértelműen megoldható. Mivel A négyzetes mátrix, ez az 1.4.11. Tétel szerint valóban akkor és csak akkor teljesül, ha $\det A \neq 0$.

A *(ii)* és *(iii)* közötti ekvivalenciához A transzponáltjára alkalmazzuk az *(i)* és *(ii)* között már bizonyított ekvivalenciát. Mivel A^T oszlopai (lényegében) azonosak A soraival, ezért A sorai akkor és csak akkor lineárisan függetlenek, ha $\det A^T \neq 0$. Mivel azonban az 1.5.4. Tétel szerint $\det A = \det A^T$, ezért ez valóban ekvivalens a $\det A \neq 0$ feltétellel. □

Érdemes a (3×3)-as esetben egy további kapcsolatra rámutatni. Az $\underline{u}$, $\underline{v}$ és $\underline{w}$ térvektorok az 1.5.16 Tétel szerint akkor és csak akkor lineárisan függetlenek, ha nem nulla a determinánsa annak a (3×3)-as mátrixnak, amelynek a sorait épp $\underline{u}$, $\underline{v}$ és $\underline{w}$ koordinátái alkotják. A térvektorok vegyesszorzatáról mondottak szerint (lásd az 1.4.20. és az 1.4.21. Tételeket) viszont ez a determináns $\underline{u}\ \underline{v}\ \underline{w}$-vel, vagyis az $\underline{u}$, $\underline{v}$ és $\underline{w}$ által kifeszített paralelepipedon térfogatával egyenlő. A térszemlélet alapján jól látszik, hogy ez a térfogat pontosan akkor nem nulla, ha a három térvektor nem esik egy origón átmenő síkba – és valóban, korábban már kiderült, hogy éppen ez $\underline{u}$, $\underline{v}$ és $\underline{w}$ lineáris függetlenségének a szükséges és elégséges feltétele.

1.5.17. Feladat. Legyen $n \geq 1$ egész és minden $1 \leq i \leq n$ esetén legyen $\underline{v}_i$ az az $\mathbb{R}^n$-beli vektor, amelynek az i-edik koordinátája p, az összes többi koordinátája 1. A $p \in \mathbb{R}$ paraméter mely értékeire lesz $\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_n$ lineárisan független?

Megoldás: Legyen A a $\underline{v}_i$ vektorok egyesítésével keletkező $(n \times n)$-es mátrix. Az 1.4.10. Feladat eredménye szerint $\det A = (p+n-1) \cdot (p-1)^{n-1}$. Látható, hogy $\det A = 0$ a $p = 1$ és a $p = 1-n$ értékekre teljesül. Az 1.5.16. Tétel szerint tehát $\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_n$ a $p \neq 1$, $p \neq 1-n$ értékekre lineárisan független. □

1.6. Az inverz mátrix

Az 1.5.14. Tételből kiderült, hogy a lineáris egyenletrendszerek $A \cdot \underline{x} = \underline{b}$ alakba írhatók – amiből némi bátorsággal felvethető a kérdés, hogy nem lehetne-e az egyenletrendszert „mindkét oldalt A-val osztva” megoldani? A mátrixok körében az osztás művelete nem értelmezett, így a kérdésre a közvetlen válasz az, hogy nem. Van azonban egy olyan fogalom, aminek segítségével (bizonyos esetekben) helyettesíthető az osztás a mátrixok körében is: ez az inverz mátrix. Az alábbi fogalom motivációja az, hogy a valós számok körében mindegy, hogy a-val osztunk vagy $1/a$-val szorzunk valamit; az inverz mátrix pedig a reciprok fogalmának a mátrixokra való alkalmazása – hiszen az 1 számnak az E egységmátrix felel meg.

1.6.1. Definíció. *Egy $(n \times n)$-es A mátrix* inverzének *nevezzük az $(n \times n)$-es X mátrixot, ha $A \cdot X = E = X \cdot A$ teljesül. Ennek a jele: $X = A^{-1}$.*

A definícióból kiemelendő egyrészt az, hogy inverze csak $(n \times n)$-es mátrixnak lehet (de – amint azt látni fogjuk – nem mindnek van) másrészt az, hogy A-t A^{-1}-zel bármelyik oldalról szorozva E-t kell kapnunk. Az inverz mátrix jelölése nyilván a reciprokkal való kapcsolatra utal (hiszen a számok körében $a^{-1} = 1/a$), de mátrixok esetében A^{-1}-et nem szokás „A a mínusz egyedikenként" kiolvasni (hanem egyszerűen „A inverzként").

A fogalommal való ismerkedésként próbáljuk meg eldönteni, hogy van-e inverze az $A = \begin{pmatrix} 3 & 2 \\ 5 & 4 \end{pmatrix}$ mátrixnak. Legyen $X = \begin{pmatrix} x_1 & x_3 \\ x_2 & x_4 \end{pmatrix}$ a keresett inverz (ahol a kettős indexek használatát most kényelmi okokból mellőztük). Ekkor az $A \cdot X = E$ egyenletből a következő feltételek adódnak X elemeire:

$$\begin{array}{ccc} & \begin{pmatrix} x_1 & x_3 \\ x_2 & x_4 \end{pmatrix} & \\ \begin{pmatrix} 3 & 2 \\ 5 & 4 \end{pmatrix} & \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} & \longrightarrow \quad \begin{array}{ll} 3x_1 + 2x_2 = 1 & 3x_3 + 2x_4 = 0 \\ 5x_1 + 4x_2 = 0 & 5x_3 + 4x_4 = 1 \end{array} \end{array}$$

Első pillantásra látszik, hogy egy (4×4)-es lineáris egyenletrendszert kaptunk. De mielőtt elkezdenénk ezt megoldani, érdemes egy második pillantást is vetni rá: mivel az első két egyenletben csak x_1 és x_2, a második kettőben csak x_3 és x_4 szerepel, ezért jobban járunk, ha a feltételeket két különálló (2×2)-es lineáris egyenletrendszerként kezeljük. Így két Gauss-eliminációval kereshetjük a megoldást: először x_1 és x_2 értékeit az

$$\left(\begin{array}{cc|c} 3 & 2 & 1 \\ 5 & 4 & 0 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 2/3 & 1/3 \\ 0 & 2/3 & -5/3 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 2/3 & 1/3 \\ 0 & 1 & -5/2 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 0 & 2 \\ 0 & 1 & -5/2 \end{array}\right)$$

eliminációból kapjuk: $x_1 = 2$, $x_2 = -5/2$; utána x_3 és x_4 értékét az

$$\left(\begin{array}{cc|c} 3 & 2 & 0 \\ 5 & 4 & 1 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 2/3 & 0 \\ 0 & 2/3 & 1 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 2/3 & 0 \\ 0 & 1 & 3/2 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 0 & -1 \\ 0 & 1 & 3/2 \end{array}\right)$$

eliminációból: $x_3 = -1$, $x_4 = 3/2$. Mindezekből tehát az $X = \begin{pmatrix} 2 & -1 \\ -5/2 & 3/2 \end{pmatrix}$ mátrix adódik. Erre tehát $A \cdot X = E$ teljesül (ami közvetlenül is ellenőrizhető) – ebből azonban még nem következik, hogy $X = A^{-1}$, mert ehhez $X \cdot A = E$ is szükséges. Ezt azonban könnyen kipróbálhatjuk:

$$\begin{array}{ccc} & & \begin{pmatrix} 3 & 2 \\ 5 & 4 \end{pmatrix} = A \\ X = \begin{pmatrix} 2 & -1 \\ -5/2 & 3/2 \end{pmatrix} & & \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = X \cdot A = E \end{array}$$

Látszik, hogy „véletlenül" $X \cdot A = E$ is igaz, így a kapott X az A (egyetlen) inverze. (Alább látni fogjuk, hogy $X \cdot A = E$ teljesülése valójában egyáltalán nem véletlen.)

A fenti számolás persze tetszőleges $(n \times n)$-es A-ra megismételhető, de az persze már nem igaz, hogy A^{-1} mindig létezik is: például ha $A = 0$ (vagyis A a csupa nulla

mátrix), akkor a mátrixszorzás definíciójából rögtön látszik, hogy $A \cdot X = 0$ is igaz minden $X \in \mathbb{R}^{n\times n}$-re, vagyis $A \cdot X = E$ sosem teljesülhet. Az alábbi tétel ad választ arra a kérdésre, hogy mely négyzetes mátrixoknak létezik inverze.

1.6.2. Tétel. *Az $(n \times n)$-es A mátrixnak akkor és csak akkor létezik inverze, ha $\det A \neq 0$. Ha A^{-1} létezik, akkor az egyértelmű.*

Bizonyítás: Először tegyük fel, hogy $X = A^{-1}$ létezik; megmutatjuk, hogy $\det A \neq 0$. A (definíció szerint teljesülő) $A \cdot X = E$ egyenlet mindkét oldalának determinánsát véve: $\det(A \cdot X) = \det E$. Itt $\det E = 1$ nyilván igaz (lásd az 1.4.6. Tételt), a bal oldalon pedig alkalmazhatjuk a determinánsok szorzástételét (1.5.13. Tétel): $\det A \cdot \det X = 1$. Ebből $\det A \neq 0$ rögtön következik.

A fordított irány bizonyításához az alábbi lemmát fogjuk használni.

1.6.3. Lemma. *Ha $A \in \mathbb{R}^{n\times n}$ és $\det A \neq 0$, akkor egyértelműen létezik egy olyan $X \in \mathbb{R}^{n\times n}$ mátrix, amelyre $A \cdot X = E$.*

A Lemma bizonyítása: A fenti példában látott számolást általánosítjuk. Legyenek $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ a keresett X mátrix oszlopai (mégpedig sorban). Mivel az $A \cdot X$ szorzat i-edik oszlopa a mátrixszorzás definíciója szerint azonos $A \cdot \underline{x}_i$-vel, ezért $A \cdot X = E$ ekvivalens az $A \cdot \underline{x}_1 = \underline{e}_1$, $A \cdot \underline{x}_2 = \underline{e}_2$, …, $A \cdot \underline{x}_n = \underline{e}_n$ egyenletek teljesülésével (ahol $\underline{e}_i$ az egységmátrix i-edik oszlopát jelöli, összhangban az 1.2.22. Definícióval). Az 1.5.14. Tételben láttuk, hogy $A \cdot \underline{x}_i = \underline{e}_i$ egy lineáris egyenletrendszert jelöl (amelynek a kibővített együtthatómátrixa $(A|\underline{e}_i)$). Mivel $\det A \neq 0$, ezért az 1.4.11. Tételből következik, hogy az $A \cdot \underline{x}_i = \underline{e}_i$ lineáris egyenletrendszer egyértelműen megoldható. Ezzel tehát a lemmát beláttuk: a keresett X i-edik oszlopa az $A \cdot \underline{x}_i = \underline{e}_i$ rendszer egyértelmű megoldása minden $1 \leq i \leq n$ esetén. ◇

A lemmából rögtön következik a tételnek az az állítása, hogy ha A^{-1} létezik, akkor az egyértelmű: valóban, már az $A \cdot X = E$ feltételnek is csak egyetlen X tehet eleget.

Azt azonban még be kell látnunk, hogy erre az X-re $X \cdot A = E$ is teljesül (vagyis hogy ami a fenti példában „véletlenül” teljesült, az általában is igaz). Ehhez a következő egyszerű, de ravasz ötlet vezet: a fenti lemmát alkalmazzuk A helyett az (ugyanebből a lemmából nyert) X mátrixra! Ezt megtehetjük, mert az $A \cdot X = E$ egyenletből a bizonyítás elején látott módon következik, hogy $\det A \cdot \det X = 1$, így $\det X \neq 0$. A lemmából tehát azt kapjuk, hogy létezik (mégpedig egyértelműen) egy olyan $Y \in \mathbb{R}^{n\times n}$ mátrix, amelyre $X \cdot Y = E$. Így készen leszünk a bizonyítással, ha sikerül megmutatnunk, hogy $Y = A$.

Ehhez az 1.5.8. Tételből a mátrixszorzás asszociativitását használjuk fel: $(A \cdot X) \cdot Y = A \cdot (X \cdot Y)$. A bal oldalt az $A \cdot X = E$ egyenlet (és az 1.5.10. Állítás) felhasználásával átalakítva: $(A \cdot X) \cdot Y = E \cdot Y = Y$. Hasonlóan, a jobb oldalt az $X \cdot Y = E$ egyenlet felhasználásával alakíthatjuk át: $A \cdot (X \cdot Y) = A \cdot E = A$. Mindezek összevetéséből $Y = A$ valóban következik. □

1.6.1. Az inverz kiszámítása

Alkalmazásokban felmerülő, fontos feladat egy adott $(n \times n)$-es A mátrix inverzének a kiszámítása. A módszer alapgondolatát már a 78. oldalon, illetve az 1.6.2. Tétel bizonyításában láttuk: ha megoldjuk az $A \cdot \underline{x}_1 = \underline{e}_1, A \cdot \underline{x}_2 = \underline{e}_2, \ldots, A \cdot \underline{x}_n = \underline{e}_n$ lineáris egyenletrendszereket, akkor a megoldások egyesítéséből kapott X mátrixra $A \cdot X = E$ teljesül; az pedig az 1.6.2. Tétel bizonyítása közben kiderült, hogy ekkor $X \cdot A = E$ is automatikusan igaz, így $X = A^{-1}$.

Ezt a gondolatot érdemes kiegészíteni azzal, hogy ennek az n lineáris egyenletrendszernek a megoldásához valójában nem kell n-szer lefuttatni a Gauss-eliminációt, ezek párhuzamosan is végezhetők. Ugyanis az n egyenletrendszer csak a jobb oldalakban különbözik, az együtthatómátrix mind az n esetben A. Márpedig a Gauss-elimináció futása során a megteendő lépések mindig csak az együtthatóktól függenek, a jobb oldalak csak követik a változásokat. Már a 78. oldalon látott példában is végezhettük volna párhuzamosan a számításokat, így:

$$\left(\begin{array}{cc|cc} 3 & 2 & 1 & 0 \\ 5 & 4 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{cc|cc} 1 & 2/3 & 1/3 & 0 \\ 0 & 2/3 & -5/3 & 1 \end{array}\right) \sim$$

$$\sim \left(\begin{array}{cc|cc} 1 & 2/3 & 1/3 & 0 \\ 0 & 1 & -5/2 & 3/2 \end{array}\right) \sim \left(\begin{array}{cc|cc} 1 & 0 & 2 & -1 \\ 0 & 1 & -5/2 & 3/2 \end{array}\right)$$

Egy nemnulla determinánsú A mátrix inverzének meghatározása tehát lényegében azonos n lineáris egyenletrendszer párhuzamos megoldásával – ismét a Gauss-elimináció egy változatáról van szó. Az elimináció elindítása előtt A mellé másoljuk (a hagyományokat – legalábbis írásban számoláskor – követve: vonallal elválasztva) az E egységmátrixot; ez felel meg az n egyenletrendszer jobb oldalainak. Majd $(A|E)$-re az elemi sorekvivalens lépéseket alkalmazzuk addig, amíg az A helyén az E egységmátrix jelenik meg (ez felel meg a redukált lépcsős alaknak – legalábbis egyértelmű megoldhatóság esetén, amit az 1.4.11. Tétel szerint $\det A \neq 0$ miatt feltételezhetünk). A módszert különösen kényelmessé teszi, hogy ekkor a vonaltól jobbra épp A^{-1} jelenik meg; valóban, az n egyenletrendszert megoldottuk és (ahogyan az a 78. oldalon látott számolás fenti megismétlésénél is látszott) épp az $\underline{x}_i$ megoldások jelennek sorban a vonaltól jobbra – márpedig a keresett X épp ezek egyesítése. Így az eljárás az $(E|A^{-1})$ alakkal ér véget.

A módszer további előnye, hogy nem szükséges előre meggyőződni afelől, hogy a bemenetként kapott A mátrixra $\det A \neq 0$. Valóban: az elemi sorekvivalens lépések nem változtatják meg azt, hogy $\det A$ értéke 0 vagy sem (miközben persze $\det A$ maga változhat). Ha tehát $\det A$ értéke mégis 0 volna, az az elimináció (annak is az első fázisa) közben úgyis kiderül: az algoritmusnak a determináns kiszámítására vonatkozó változatánál (lásd az 53. oldalt) láttuk, hogy ebben az esetben a felsőháromszög-mátrix felé vezető úton egy ponton a főátlóbeli elem 0 és alatta is mindenhol 0 áll. Ha tehát ez bekövetkezik, akkor az eljárás leállhat (és kiírhatja, hogy „A^{-1} nem létezik”).

1.6.4. Feladat. a) Döntsük el, hogy létezik-e inverze az alábbi A mátrixnak; ha igen, akkor számítsuk ki A inverzét.

$$A = \begin{pmatrix} 1 & -3 & 7 \\ -1 & 3 & -6 \\ 2 & -5 & 12 \end{pmatrix}$$

b) Oldjuk meg az alábbi lineáris egyenletrendszert a $p, q, r \in \mathbb{R}$ paraméterek minden értékére.

$$\begin{aligned} x_1 - 3x_2 + 7x_3 &= p \\ -x_1 + 3x_2 - 6x_3 &= q \\ 2x_1 - 5x_2 + 12x_3 &= r \end{aligned}$$

Megoldás: a) A Gauss-elimináció fenti változatát használjuk. Indítás előtt A mellé másoljuk a (3×3)-as egységmátrixot (alább, balra). Mivel a bal felső sarokban rögtön 1-es áll, az 1. sorhoz nem kell nyúlnunk, az 1. oszlop másik két elemét változtatjuk 0-vá az 1. sor megfelelő többszörösének hozzáadásával (jobbra):

$$\left(\begin{array}{ccc|ccc} 1 & -3 & 7 & 1 & 0 & 0 \\ -1 & 3 & -6 & 0 & 1 & 0 \\ 2 & -5 & 12 & 0 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & -3 & 7 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & -2 & -2 & 0 & 1 \end{array}\right) \sim$$

Mivel a 2. sor 2. helyén 0 áll, a 2. sort felcseréljük a 3.-kal. Ezzel a lépcsős alakot el is értük. Azt már most állíthatjuk, hogy A^{-1} létezik (hiszen a vonaltól balra álló mátrix determinánsa 1, így kezdetben sem lehetett 0).

$$\sim \left(\begin{array}{ccc|ccc} 1 & -3 & 7 & 1 & 0 & 0 \\ 0 & 1 & -2 & -2 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{array}\right) \sim$$

Áttérve a Gauss-elimináció második fázisára, a vezéregyesek fölötti nemnulla elemeket változtatjuk 0-vá. Először a 3. sor 2-szeresét, illetve (-7)-szeresét adjuk a 2., illetve 1. sorhoz (balra). Végül a 2. sor 3-szorosát az 1.-höz.

$$\sim \left(\begin{array}{ccc|ccc} 1 & -3 & 0 & -6 & -7 & 0 \\ 0 & 1 & 0 & 0 & 2 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -6 & -1 & 3 \\ 0 & 1 & 0 & 0 & 2 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{array}\right)$$

Ezzel az elimináció véget ért, A^{-1} a fenti, a vonaltól jobbra eső mátrix. A számolási hibákat kiszűrendő érdemes meggyőződni arról, hogy a kapott X mátrixra $A \cdot X = E$ valóban teljesül (vagy akár arról, hogy $X \cdot A = E$ igaz – fentebb láttuk, hogy bármelyikből következik a másik).

b) A lineáris egyenletrendszer $A\underline{x} = \underline{b}$ alakba írható, ahol A az a) feladatbeli mátrix, $\underline{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$ és $\underline{b} = \begin{pmatrix} p \\ q \\ r \end{pmatrix}$. Az $A\underline{x} = \underline{b}$ egyenlet mindkét oldalát balról A^{-1}-zel

szorozva: $A^{-1}(A\underline{x}) = A^{-1}\underline{b}$. Felhasználva a mátrixszorzás asszociativitását és az inverz definícióját, a kapott egyenlet bal oldala: $A^{-1}(A\underline{x}) = (A^{-1}A)\underline{x} = E\underline{x} = \underline{x}$. Következik, hogy $\underline{x} = A^{-1}\underline{b}$. Ezért a feladat megoldásához egyszerűen $\underline{b}$-t megszorozzuk balról az a) feladatban kiszámolt A^{-1}-zel:

$$\begin{pmatrix} p \\ q \\ r \end{pmatrix} = \underline{b}$$

$$A^{-1} = \begin{pmatrix} -6 & -1 & 3 \\ 0 & 2 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} -6p-q+3r \\ 2q+r \\ p+q \end{pmatrix} = A^{-1}\underline{b} = \underline{x}$$

Így a rendszer egyértelmű megoldása: $x_1 = -6p-q+3r$, $x_2 = 2q+r$, $x_3 = p+q$. □

1.7. Mátrix rangja

Egy $(k \times n)$-es mátrix kétféle értelemben is értelmezhető vektorrendszerként: az oszlopait n darab $\mathbb{R}^k$-beli vektornak, a sorait k darab n hosszú sorvektornak tekinthetjük. Ennek a két vektorrendszernek látszólag nem sok köze van egymáshoz. Mégis, azt már tudjuk, hogy az $(n \times n)$-es esetben fontos kapcsolat van köztük: az 1.5.16. Tétel szerint az egyik pontosan akkor lineárisan független, ha a másik; ráadásul mindkettő pontosan akkor, ha a mátrix determinánsa nem 0. A mátrixrang fogalmának segítségével ezeket a kapcsolatokat fogjuk általánosítani tetszőleges mátrixra is.

Egy $(k \times n)$-es mátrix *rangja* alatt egy természetes számot fogunk érteni. A fogalom ereje abban rejlik, hogy ezt az értéket többféleképpen is definiálhatjuk – az 1.7.3. Tétel pedig épp azt fogja állítani, hogy a különböző definíciók mindig azonos értéket adnak. Először szükségünk lesz a következő fogalomra:

1.7.1. Definíció. *Legyen A $(k \times n)$-es mátrix és $r \leq k,n$ egész. Válasszunk ki tetszőlegesen A sorai és oszlopai közül r-r darabot. Ekkor a kiválasztott sorok és oszlopok kereszteződéseiben kialakuló $(r \times r)$-es mátrixot A egy* négyzetes részmátrixának *nevezzük.*

A négyzetes részmátrix fogalma tehát elképzelhető úgy is, hogy A-ból elhagyunk néhány (tetszőlegesen választott) oszlopot, majd a maradék mátrixból elhagyunk néhány sort úgy, hogy végül négyzetes mátrixot kapjunk. Ha például A (5×10)-es, akkor

$$A = \left(\begin{array}{cccccccccc} \square & \square & \blacksquare & \square & \square & \square & \blacksquare & \square & \blacksquare & \square \\ \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare \\ \square & \square & \blacksquare & \square & \square & \square & \blacksquare & \square & \blacksquare & \square \\ \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare \end{array}\right) \quad \longrightarrow \quad M = \begin{pmatrix} a_{2,3} & a_{2,7} & a_{2,9} \\ a_{3,3} & a_{3,7} & a_{3,9} \\ a_{5,3} & a_{5,7} & a_{5,9} \end{pmatrix}$$

A-nak M (3×3)-as részmátrixa, amely a 2., 3. és 5. sorok és a 3., 7. és 9. oszlopok kiválasztásával keletkezett. A négyzetes részmátrix fogalma kicsit emlékeztet az 1.4.13. Tételben használt előjeles aldeterminánsfogalmára, de több lényeges

ponton különbözik attól: egyrészt négyzetes részmátrixai minden mátrixnak vannak (és nem csak a négyzeteseknek), másrészt az A-ból elhagyott sorok és oszlopok száma nem feltétlen 1 (de általában még csak nem is egyenlő), harmadrészt a négyzetes részmátrix értelemszerűen egy mátrix és nem egy szám (szemben az előjeles aldeterminánssal), negyedrészt az előjeles aldeterminánsfogalmában kulcsszerepet játszó sakktáblaszabály a négyzetes részmátrix fogalmából hiányzik.

1.7.2. Definíció. *Legyen A tetszőleges mátrix. Azt mondjuk, hogy*

(i) A oszloprangja *r, ha A oszlopai közül kiválasztható r darab úgy, hogy a kiválasztott oszlopok (mint $\mathbb{R}^k$-beli vektorok) lineárisan függetlenek, de $r+1$ már nem választható ki így;*

(ii) A sorrangja *r, ha A sorai közül kiválasztható r darab úgy, hogy a kiválasztott sorok (mint n hosszú sorvektorok) lineárisan függetlenek, de $r+1$ már nem választható ki így;*

(iii) A determinánsrangja *r, ha A-nak van nemnulla determinánsú $(r\times r)$-es részmátrixa, de $(r+1)\times(r+1)$-es nemnulla determinánsú már nincs.*

Ha A a nullmátrix, akkor az oszloprangja és a sorrangja (az 1.2.11. Definíció után írtakkal összhangban) 0. Mivel a determinánsrang fenti definíciója ebben az esetben közvetlenül nem értelmezhető, ezért megállapodunk abban, hogy a nullmátrix determinánsrangja is 0.

Azonnal látszik, hogy minden A mátrixra mindhárom definíció egy-egy egyértelmű értéket határoz meg: a legnagyobb méretű A-beli lineárisan független oszloprendszer, sorrendszer, illetve nemnulla determinánsú négyzetes részmátrix méretét. (Valóban: ha $r+1$ lineárisan független oszlop vagy sor már nem létezik, akkor $(r+1)$-nél több sem létezhet, mert lineárisan független rendszer minden része is lineárisan független. Hasonlóan, ha $(r+1)\times(r+1)$-es nemnulla determinánsú részmátrix már nincs, akkor – az 1.4.13. Kifejtési Tételből könnyen láthatóan – ennél nagyobb sem lehet.)

A oszloprangjának, sorrangjának, illetve determinánsrangjának értékét – ideiglenesen – $o(A)$, $s(A)$, illetve $d(A)$ fogja jelölni (de amint az egyenlőségüket bebizonyítjuk, közös jelölést vezetünk be rájuk).

Tekintsük például az alábbi A mátrixot:

$$A=\begin{pmatrix}1&3&5&7\\9&11&13&15\\17&19&21&23\end{pmatrix}$$

Jelölje A oszlopait sorban $\underline{o}_1,\underline{o}_2,\underline{o}_3,\underline{o}_4$, a sorait $\underline{s}_1,\underline{s}_2,\underline{s}_3$. A oszlopai közül például $\underline{o}_1$ és $\underline{o}_2$ lineárisan független (mert egyik sem skalárszorosa a másiknak), de bármelyik hármat választjuk, lineárisan összefüggő rendszert kapunk: ezt mutatják az $\underline{o}_1-2\underline{o}_2+\underline{o}_3=\underline{0}$, $2\underline{o}_1-3\underline{o}_2+\underline{o}_4=\underline{0}$, $\underline{o}_1-3\underline{o}_3+2\underline{o}_4=\underline{0}$ és az $\underline{o}_2-2\underline{o}_3+\underline{o}_4=\underline{0}$ nemtriviális lineáris kombinációk. Következik, hogy $o(A)=2$. Hasonlóan, például $\underline{s}_1$ és $\underline{s}_2$ lineárisan független, de $\underline{s}_1-2\underline{s}_2+\underline{s}_3=\underline{0}$ mutatja, hogy $\underline{s}_1,\underline{s}_2,\underline{s}_3$ (az egyetlen lehetőség három sor kiválasztására) már lineárisan összefüggő. Így $s(A)=2$.

Végül A-ból például a négy sarkot (vagyis az $\underline{s}_1, \underline{s}_3$ sorok és az $\underline{o}_1, \underline{o}_4$ oszlopok által meghatározott részmátrixot) választva $\begin{vmatrix} 1 & 7 \\ 17 & 23 \end{vmatrix} = 1 \cdot 23 - 17 \cdot 7 = -96 \neq 0$, de bárhogyan választanánk A-ból (3×3)-as részmátrixot, annak a determinánsa mindig 0 volna; így $d(A) = 2$ is igaz. (Valóban, egy (3×3)-as részmátrixhoz mindhárom sort ki kell választanunk és az oszlopok közül hármat. Mivel mind a négy szóba jövő esetben az imént láttuk, hogy a keletkező (3×3)-as részmátrixok oszlopai lineárisan összefüggők, ezért az 1.5.16. Tétel szerint a determinánsuk is 0.) Így tehát a fenti A mátrixra $o(A) = s(A) = d(A)$ valóban teljesül – de persze nem csak erre.

1.7.3. Tétel. *Minden A mátrixra $o(A) = s(A) = d(A)$.*

Bizonyítás: Elég lesz belátni, hogy $o(A) = d(A)$ igaz minden A mátrixra, ebből már a tétel teljes állítása következni fog. Valóban, mivel A sorai (lényegében) azonosak A^T oszlopaival, ezért $s(A) = o(A^T)$. Az 1.5.4. Tételből következik, hogy $d(A) = d(A^T)$, mert az A^T-ból választható négyzetes részmátrixok épp az A-ból választhatók transzponáltjai; mivel ezek determinánsa az 1.5.4. Tétel szerint egyenlő, a legnagyobb nemnulla determinánsúnak a mérete is azonos a két esetben. Ha az $o(A) = d(A)$ állítást minden mátrixra – így A^T-ra is – igaznak feltételezzük, akkor mindezeket összevetve az $s(A) = o(A^T) = d(A^T) = d(A) = o(A)$ egyenlőségeket kapjuk, amiből a tétel állítása már következik.

Azt kell tehát csak bizonyítanunk, hogy $o(A) = d(A)$. Ezt két lépésben tesszük: először megmutatjuk, hogy $o(A) \geq d(A)$, utána azt, hogy $o(A) \leq d(A)$.

Tegyük fel tehát először, hogy $d(A) = r$. Meg kell mutatnunk, hogy $o(A) \geq r$, vagyis hogy A oszlopai közül kiválasztható r darab lineárisan független. A-ból $d(A) = r$ miatt kiválasztható egy $(r \times r)$-es, nemnulla determinánsú M részmátrix. Jelölje A_M az A-nak abból az r oszlopából álló mátrixot, amelyeket az M készítésekor kiválasztottunk; ekkor tehát M sorai az A_M sorainak részhalmaza. Állítjuk, hogy A_M oszlopai lineárisan függetlenek. Ha nem így volna, akkor az 1.5.15. Következmény szerint az $A_M \cdot \underline{x} = \underline{0}$ lineáris egyenletrendszernek volna egy $\underline{x}^* \neq \underline{0}$ megoldása. Ekkor azonban $\underline{x}^*$ megoldása volna az $M \cdot \underline{x} = \underline{0}$ lineáris egyenletrendszernek is, hiszen az utóbbi rendszert az előbbiből bizonyos egyenletek (mégpedig az M-hez nem tartozó A_M-beli soroknak megfelelők) elhagyásával kapjuk. Következésképp M oszlopai lineárisan összefüggők volnának, ami az 1.5.16. Tétel szerint ellentmondana annak, hogy $\det M \neq 0$. Így $o(A) \geq d(A)$ valóban igaz.

Az $o(A) \leq d(A)$ állítás bizonyításához az alábbi lemmát fogjuk használni.

1.7.4. Lemma. *Legyen C egy $(k \times n)$-es mátrix, amelynek az oszlopai (mint $\mathbb{R}^k$-beli vektorok) lineárisan függetlenek. Ha $k > n$, akkor C sorai közül kiválasztható egy úgy, hogy ezt a sort elhagyva a kapott $(k-1) \times n$-es C' mátrix oszlopai szintén lineárisan függetlenek.*

A Lemma bizonyítása: Legyenek C oszlopai $\underline{c}_1, \underline{c}_2, \ldots, \underline{c}_n$, az ezek által generált $\mathbb{R}^k$-beli altér $W = \langle \underline{c}_1, \underline{c}_2, \ldots, \underline{c}_n \rangle$. Mivel W-ben van n elemű generátorrendszer

(mégpedig $\underline{c}_1, \underline{c}_2, \ldots, \underline{c}_n$) és $k > n$, ezért az 1.2.15. F-G egyenlőtlenség miatt nem lehet benne k elemű lineárisan független rendszer. Ebből következik, hogy az $\mathbb{R}^k$-beli standard bázis vektorai (más szóval: az E_k egységmátrix oszlopai) között van olyan, amelyik nem tartozik W-hez. Legyen $\underline{e}_j$ ilyen (amelyben tehát az 1-es a j-edik helyen áll). Állítjuk, hogy a C j-edik sora teljesíti a lemma feltételét: az elhagyásával kapott C' mátrix oszlopai lineárisan függetlenek. Tegyük fel indirekt, hogy nem így van: a $C' \cdot \underline{x} = \underline{0}$ lineáris egyenletrendszernek van egy $\underline{x}^* \neq \underline{0}$ megoldása. Ekkor $C \cdot \underline{x}^* \neq \underline{0}$, mert C oszlopai lineárisan függetlenek. Mivel a $C \cdot \underline{x}^*$ szorzat csak abban különbözik a $C' \cdot \underline{x}^*$ szorzattól, hogy az utóbbiba a j-edik helyre „beszúródik" a C j-edik sorának és $\underline{x}^*$-nak a skaláris szorzata, ezért a $C \cdot \underline{x}^*$ oszlopvektor j-edik koordinátája egy $\alpha \neq 0$ szám, a többi koordinátája 0. Következik, hogy $C \cdot \left(\frac{1}{\alpha} \cdot \underline{x}^*\right) = \frac{1}{\alpha} \cdot (C \cdot \underline{x}^*) = \underline{e}_j$. Ez ellentmond annak, hogy $\underline{e}_j \notin W$: az 1.5.14. Tétel szerint a C oszlopainak az $\left(\frac{1}{\alpha} \cdot \underline{x}^*\right)$ vektor koordinátáival, mint együtthatókkal képzett lineáris kombinációja épp $\underline{e}_j$-t adja. Ez az ellentmondás bizonyítja a lemmát. ◇

Legyen most $o(A) = r$ és válasszunk A oszlopai közül r lineárisan függetlent, alkossák ezek a C mátrixot. Célunk tehát megmutatni, hogy $d(A) \geq r$. C (és A) sorainak számát k-val jelölve C oszlopai $\mathbb{R}^k$-beli vektorok, így az 1.2.15. F-G egyenlőtlenség miatt $k \geq r$ (hiszen $\mathbb{R}^k$-ban van k elemű generátorrendszer: bármely bázis ilyen). Ha $k > r$, akkor a fenti lemmát C-re alkalmazva kapjuk a $(k-1) \times r$-es C' mátrixot, amelynek az oszlopai továbbra is lineárisan függetlenek. Ha $k - 1 > r$, akkor ismét alkalmazhatjuk a lemmát C'-re és ezt folytathatjuk egészen addig, amíg ($k - r$ ilyen lépés után) egy $(r \times r)$-es C^* mátrixot nem kapunk. Ekkor az 1.5.16. Tétel szerint $\det C^* \neq 0$ (mert C^* oszlopai lineárisan függetlenek). Mivel C^* az A-nak $(r \times r)$-es részmátrixa, ezért ez bizonyítja a $d(A) \geq r$ állítást – és ezáltal a tételt is. □

A fenti tétel tehát létjogosultságot ad az alábbi definíciónak.

1.7.5. Definíció. *Az A mátrix* rangjának *nevezzük és $r(A)$-val jelöljük $o(A)$, $s(A)$ és $d(A)$ közös értékét.*

A mátrix rangja számos, eddig megismert fogalommal szoros kapcsolatban áll. Az alábbi tételek ilyen kapcsolatokat mutatnak be: először az alterek dimenzióját, utána a lineáris egyenletrendszerek (egyértelmű) megoldhatóságának kérdését tárgyaljuk a rang segítségével.

1.7.6. Tétel. *Legyen A $(k \times n)$-es mátrix, az oszlopai legyenek $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n$. Ekkor $r(A) = \dim\langle \underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n \rangle$.*

Bizonyítás: Válasszunk ki A oszlopai közül a lehető legtöbbet úgy, hogy ezek lineárisan függetlenek legyenek. A kiválasztott oszlopok száma ekkor az oszloprang definíciója szerint $r = r(A)$. A jelöléseinket egyszerűsítendő tegyük föl, hogy A-nak épp az első r oszlopát választottuk ki (hiszen az oszlopok sorrendje érdektelen). Állítjuk, hogy $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_r$ bázist alkot a $W = \langle \underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n \rangle$ altérben. Ha ezt belátjuk, abból a tétel állítása (a dimenzió 1.2.20. Definíciója szerint) következni fog.

Világos, hogy $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_r$ lineárisan független, azt kell tehát csak belátnunk, hogy generátorrendszer W-ben. Legyen ezért $U = \langle \underline{a}_1, \underline{a}_2, \ldots, \underline{a}_r \rangle$, célunk belátni, hogy $U = W$. Ebből annyi nyilvánvaló, hogy $U \subseteq W$, a $W \subseteq U$ tartalmazást kell megmutatnunk. Tetszőleges $r < i \leq n$ esetén az $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_r, \underline{a}_i$ rendszer lineárisan összefüggő, hiszen A-ból $r+1$ lineárisan független oszlop nem választható ki. Az „újonnan érkező vektor" 1.2.14. Lemmája szerint ekkor $\underline{a}_i \in \langle \underline{a}_1, \underline{a}_2, \ldots, \underline{a}_r \rangle = U$. Azt kaptuk tehát, hogy az $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n$ vektorok mindegyike U-beli (hiszen ez $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_r$ esetében magától értetődő). Mivel azonban U altér, zárt az összeadásra és a skalárral szorzásra, így minden W-beli, vagyis az $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n$ vektorokból lineáris kombinációval kifejezhető vektor is U-beli kell legyen. Ezzel a $W \subseteq U$ tartalmazást, és így a tételt is beláttuk. □

1.7.7. Tétel. *Tetszőleges $(k \times n)$-es A mátrix és $\underline{b} \in \mathbb{R}^k$ esetén az $A \cdot \underline{x} = \underline{b}$ lineáris egyenletrendszer akkor és csak akkor megoldható, ha $r(A) = r\big((A|\underline{b})\big)$. (Itt $(A|\underline{b})$ a lineáris egyenletrendszer kibővített együtthatómátrixát jelöli.)*

Bizonyítás: Jelölje az A oszlopait $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n$ és legyen $U = \langle \underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n \rangle$, valamint $W = \langle \underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n, \underline{b} \rangle$. Nyilvánvaló, hogy $U \subseteq W$. Az 1.7.6. Tételt alkalmazva az A, illetve az $(A|\underline{b})$ mátrixra: $r(A) = \dim U$, illetve $r\big((A|\underline{b})\big) = \dim W$.

Kezdjük a feltétel szükségességének a bizonyításával: ha az $A \cdot \underline{x} = \underline{b}$ rendszer megoldható, akkor $r(A) = r\big((A|\underline{b})\big)$, vagyis $\dim U = \dim W$. $A \cdot \underline{x} = \underline{b}$ megoldhatósága miatt az 1.5.14. Tétel szerint $\underline{b} \in U$. Ebből pedig $W \subseteq U$ következik: mivel U (lévén altér) zárt az összeadásra és a skalárral szorzásra és $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n, \underline{b} \in U$, ezért az ezekből a vektorokból lineáris kombinációval előálló minden vektor – vagyis W minden eleme – U-beli. Így ($U \subseteq W$ és $W \subseteq U$ miatt) $U = W$, amiből persze $\dim U = \dim W$ valóban következik.

Most belátjuk a feltétel elégségességét: ha $r(A) = r\big((A|\underline{b})\big)$, akkor $A \cdot \underline{x} = \underline{b}$ megoldható. Az $r(A) = r\big((A|\underline{b})\big)$ feltételből $\dim U = \dim W$. Belátjuk, hogy $U = W$. Vegyük ugyanis U egy tetszőleges bázisát; ez $U \subseteq W$, $\dim U = \dim W$ és az 1.2.28. Következmény miatt W-nek is bázisa. Mivel van közös bázisuk, ezért $U = W$ valóban igaz (mindkettő a közös bázis által generált altér). Így $\underline{b} \in W$-ből $\underline{b} \in U$ is következik, ami az 1.5.14. Tétel szerint ekvivalens az $A \cdot \underline{x} = \underline{b}$ megoldhatóságával. □

1.7.8. Tétel. *Tetszőleges $(k \times n)$-es A mátrix és $\underline{b} \in \mathbb{R}^k$ esetén az $A \cdot \underline{x} = \underline{b}$ lineáris egyenletrendszer akkor és csak akkor egyértelműen megoldható, ha $r(A) = r\big((A|\underline{b})\big) = n$.*

Bizonyítás: Először a feltétel szükségességét látjuk be: ha $A \cdot \underline{x} = \underline{b}$ egyértelműen megoldható, akkor $r(A) = r\big((A|\underline{b})\big) = n$. Mivel az egyértelmű megoldhatóság magában foglalja a megoldhatóságot, $r(A) = r\big((A|\underline{b})\big)$ következik a fenti tételből. Mivel A-nak n oszlopa van, világos az oszloprang definíciójából, hogy $r(A) \leq n$. Tegyük fel ezért indirekt, hogy $r(A) < n$; ekkor A oszlopai lineárisan összefüggők (mert n-nél kevesebb lineárisan független választható csak ki közülük). Így az 1.5.15. Következmény szerint az $A \cdot \underline{x} = \underline{0}$ lineáris egyenletrendszernek létezik egy $\underline{z} \neq \underline{0}$ megoldása. Ha most az $A \cdot \underline{x} = \underline{b}$ rendszernek $\underline{x}^*$ megoldása, akkor megoldása lesz $\underline{x}^* + \underline{z}$ is, mert $A \cdot (\underline{x}^* + \underline{z}) = A \cdot \underline{x}^* + A \cdot \underline{z} = \underline{b} + \underline{0} = \underline{b}$. Mivel $\underline{z} \neq \underline{0}$ miatt $\underline{x}^* \neq \underline{x}^* + \underline{z}$, ez ellentmond annak, hogy $A \cdot \underline{x} = \underline{b}$ egyértelműen megoldható.

Következik a feltétel elégségessége: ha $r(A) = r\big((A|\underline{b})\big) = n$, akkor $A \cdot \underline{x} = \underline{b}$ egyértelműen megoldható. Mivel $r(A) = r\big((A|\underline{b})\big)$, annyit a fenti tételből már tudunk, hogy $A \cdot \underline{x} = \underline{b}$ megoldható. Tegyük fel ezért indirekt, hogy van két különböző megoldása: $\underline{x}_1^*$ és $\underline{x}_2^*$. Legyen $\underline{z} = \underline{x}_1^* - \underline{x}_2^*$. Ekkor $\underline{x}_1^* \neq \underline{x}_2^*$ miatt $\underline{z} \neq \underline{0}$ és

$$A \cdot \underline{z} = A \cdot (\underline{x}_1^* - \underline{x}_2^*) = A \cdot \underline{x}_1^* - A \cdot \underline{x}_2^* = \underline{b} - \underline{b} = \underline{0}.$$

Így az $A \cdot \underline{x} = \underline{0}$ rendszernek van a $\underline{0}$-tól különböző megoldása, ami az 1.5.15. Következmény szerint azt jelenti, hogy A oszlopai lineárisan összefüggők. Ez az oszloprang definíciója szerint azzal ekvivalens, hogy $r(A) < n$. Ez ellentmond az $r(A) = n$ feltevésnek és ezáltal a tételt bizonyítja. □

A rang kiszámítása

Egy adott mátrix rangjának kiszámítása alkalmazásokban felmerülő, fontos algoritmikus feladat. Az alábbi állítás következménye, hogy ennek a feladatnak a megoldására is alkalmas lesz a Gauss-elimináció (egy változata). Az állításban elemi sorekvivalens lépés, illetve lépcsős alak alatt pontosan azt értjük, amit az 1.3.1., illetve az 1.3.3. Definíciók is annak neveztek – az egyetlen különbség az, hogy most egy tetszőleges mátrixra alkalmazzuk ezeket, az utolsó oszlop nem bír speciális szereppel, mint az $(A|\underline{b})$ kibővített együtthatómátrix esetében.

1.7.9. Állítás.
(i) Az elemi sorekvivalens lépések a mátrix rangját nem változtatják meg.
(ii) Lépcsős alakú mátrix rangja egyenlő a sorainak a számával.

Bizonyítás: Válasszunk ki A oszlopai közül tetszőlegesen néhányat, alkossák ezek együtt az A' mátrixot. A' oszlopai az 1.5.15. Következmény szerint akkor és csak akkor lineárisan függetlenek, ha az $A' \cdot \underline{x} = \underline{0}$ lineáris egyenletrendszernek az egyetlen megoldása $\underline{x} = \underline{0}$. Amikor A-ra alkalmazzuk valamelyik elemi sorekvivalens lépést, ezáltal ugyanezt a lépést alkalmazzuk az $(A'|\underline{0})$ kibővített együtthatómátrixra is: valóban, egyrészt A' sorai az A sorainak részei (így az A teljes sorain végzett lépés A'-re is azonos hatással van), másrészt ha a jobb oldalakon csupa 0 áll, akkor ezt a tulajdonságot mindegyik elemi sorekvivalens lépés fenntartja. Azonban az $(A'|\underline{0})$-n végzett lépések az $A' \cdot \underline{x} = \underline{0}$ lineáris egyenletrendszer megoldásait nem változtatják meg – ezt mondja ki az 1.3.2. Állítás (és éppen ezért lettek ezek a Gauss-elimináció megengedett lépései). Ebből következik, hogy az A-n (és ezáltal A'-n) végzett elemi sorekvivalens lépések nem változtatnak azon, hogy A' oszlopai lineárisan függetlenek-e. Így természetesen az A oszlopai közül kiválasztható legnagyobb lineárisan független rendszer mérete, vagyis az (oszlop)rang sem változik.

Ezzel *(i)*-et beláttuk, *(ii)*-t legegyszerűbb a determinánsrang felől megközelíteni. Ha a lépcsős alakú A mátrix sorainak száma k, akkor A-ból az összes sor és a vezéregyeseket tartalmazó oszlopok kiválasztásával keletkező M négyzetes részmátrix egy felsőháromszög-mátrix, amelynek a főátlójában álló minden elem 1-es

(ezek épp a vezéregyesek). Így $\det M = 1 \neq 0$, vagyis A-nak van $(k \times k)$-as, nemnulla determinánsú négyzetes részmátrixa. Ennél nagyobb pedig nyilván nem lehet, hiszen A-nak csak k sora van. Következik, hogy A (determináns)rangja valóban k. □

A fenti állításból kiolvasható egy, a rang meghatározására szolgáló hatékony algoritmus. Tudjuk, hogy minden A mátrix elemi sorekvivalens lépésekkel lépcsős alakra hozható: ezt a Gauss-elimináció (annak is a lineáris egyenletrendszerekre vonatkozó, a 39. oldalon leírt változatának) vizsgálata során be is láttuk. (Folytathatnánk persze az eliminációt a redukált lépcsős alak eléréséig is, de a fenti állítás szerint ez fölösleges: a rang már a lépcsős alakból is kiolvasható.) Az ott leírt algoritmust (annak az első fázisát) alkalmazzuk tehát most is – azzal az egyetlen apró változtatással, hogy a 3. lépéséből töröljük a „ha van olyan $i < t \leq k$, amelyre $b_t \neq 0$, akkor…” lépést (hiszen most a jobb oldalakról, b_t-ről beszélni értelmetlen). Ha pedig elértük a lépcsős alakot, akkor a fenti állítás szerint a rang egyszerűen a (megmaradt) sorok száma.

A fenti állítást fontos kiegészíteni a következővel: ha a Gauss-elimináció lépéseit A-nak nem a soraira, hanem az oszlopaira alkalmazzuk, a rangot az sem változtatja meg. Ez következik a fenti állításból, ha azt A^T-ra alkalmazzuk: $r(A) = r(A^T)$ (hiszen A oszloprangja azonos A^T sorrangjával), az A oszlopain végzett lépés pedig egyenértékű az A^T sorain végzett lépéssel. Ez a megfigyelés számos feladat megoldását egyszerűsíti: időnként az oszlopokon végzett lépéseket (is) alkalmazva hamarabb elérhetjük a lépcsős alakot (vagy egy olyan mátrixot, amelynek a rangját könnyen megállapíthatjuk).

Előfordulhat azonban (sőt: gyakori) olyan algoritmikus feladat is, ahol nem csak $r(A)$-t magát, hanem (az oszloprang definíciója szerint) egy, az A oszlopai közül kiválasztható legnagyobb lineárisan független oszloprendszert is keresünk. A fentiekből ennek a feladatnak a megoldása is következik: az 1.7.9. Állítás bizonyításából kiderült, hogy az elemi sorekvivalens lépések nem csak $r(A)$-t nem változtatják, hanem az A oszlopaiból kiválasztható rendszerek lineáris függetlenségét sem. Ha pedig Gauss-eliminációval (a fent leírt módon) elérjük a lépcsős alakot, akkor ott a vezéregyeseket tartalmazó oszlopok $r(A)$ méretű lineárisan független rendszert adnak (ez következik az 1.5.16. Tételből és abból, hogy az ezek által az oszlopok által alkotott mátrix determinánsa 1). Így az ezeknek megfelelő oszlopok az eredeti mátrixban is $r(A)$ méretű lineárisan független rendszert alkotnak. Fontos azonban hozzátenni, hogy az ebben a bekezdésben leírt eljárás csak akkor működik, ha az elemi sorekvivalens lépéseket csak a sorokra alkalmazzuk: az oszlopokon végzett lépések $r(A)$-t ugyan nem, de az A-ból választott egyes oszloprendszerek lineáris függetlenségét már megváltoztathatják.

1.7.10. Feladat. Adjunk meg egy bázist a $W = \langle \underline{a}_1, \underline{a}_2, \underline{a}_3, \underline{a}_4, \underline{a}_5 \rangle$ generált altérben a p paraméter minden értékére, ahol

$$\underline{a}_1 = \begin{pmatrix} 2 \\ 1 \\ -1 \\ 5 \end{pmatrix}, \underline{a}_2 = \begin{pmatrix} 8 \\ 2 \\ -1 \\ 22 \end{pmatrix}, \underline{a}_3 = \begin{pmatrix} 6 \\ -1 \\ 3 \\ 19 \end{pmatrix}, \underline{a}_4 = \begin{pmatrix} 4 \\ 12 \\ -12 \\ p \end{pmatrix} \text{ és } \underline{a}_5 = \begin{pmatrix} 2 \\ 7 \\ 0 \\ p+3 \end{pmatrix}.$$

Megoldás: Az 1.7.6. Tétel szerint $\dim W = r(A)$, ahol A az $\underline{a}_1, \ldots, \underline{a}_5$ vektorok egyesítésével keletkező mátrix (alább, balra). Így ha mutatunk W-ben $r(A)$ darab lineárisan független vektort, akkor az 1.2.28. Következmény szerint ezek bázist is alkotnak W-ben. Mivel A oszlopai mind W-beliek, ezért ezt az $r(A)$ darab lineárisan független oszlopot A oszlopai közül is választhatjuk. Az oszloprang definíciója szerint ez lehetséges – és a fentiekből kiderült, hogy a Gauss-eliminációval (annak a lépéseit csak az A soraira alkalmazva) meg is tehető.

A Gauss-eliminációt A-val indítva a következőket kapjuk:

$$\begin{pmatrix} 2 & 8 & 6 & 4 & 2 \\ 1 & 2 & -1 & 12 & 7 \\ -1 & -1 & 3 & -12 & 0 \\ 5 & 22 & 19 & p & p+3 \end{pmatrix} \sim \begin{pmatrix} 1 & 4 & 3 & 2 & 1 \\ 0 & -2 & -4 & 10 & 6 \\ 0 & 3 & 6 & -10 & 1 \\ 0 & 2 & 4 & p-10 & p-2 \end{pmatrix} \sim$$

$$\sim \begin{pmatrix} 1 & 4 & 3 & 2 & 1 \\ 0 & 1 & 2 & -5 & -3 \\ 0 & 0 & 0 & 5 & 10 \\ 0 & 0 & 0 & p & p+4 \end{pmatrix} \sim \begin{pmatrix} 1 & 4 & 3 & 2 & 1 \\ 0 & 1 & 2 & -5 & -3 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 4-p \end{pmatrix}$$

(Az utolsó mátrix 4. sorát úgy kaptuk, hogy az előző mátrix 4. sorából az 5-tel elosztott 3. sor p-szeresét vontuk ki.)

Ha $p = 4$, akkor az utolsó sor csupa 0 sor, így elhagyható. A kapott 3 sorú mátrix lépcsős alakú, így ekkor $r(A) = \dim W = 3$. Vezéregyest a lépcsős alakban az 1., 2., és 4. oszlopok tartalmaznak, így az ezeknek megfelelő A-beli oszlopok lineárisan függetlenek. Következik, hogy a $p = 4$ esetben $\underline{a}_1, \underline{a}_2, \underline{a}_4$ bázis W-ben.

Ha $p \neq 4$, akkor az utolsó sor $(4-p)$-vel való osztása után kapjuk a lépcsős alakot. Ekkor tehát $r(A) = \dim W = 4$ és W-nek bázisa $\underline{a}_1, \underline{a}_2, \underline{a}_4, \underline{a}_5$. Ebből persze következik, hogy valójában $W = \mathbb{R}^4$, hiszen $\underline{a}_1, \underline{a}_2, \underline{a}_4, \underline{a}_5$ az 1.2.28. Következmény szerint (és $\dim W = 4$ miatt) $\mathbb{R}^4$-ben is bázis, így W minden $\mathbb{R}^4$-beli vektort tartalmaz. Így a $p \neq 4$ esetben W-nek bázisa bármely más $\mathbb{R}^4$-beli bázis is (például a standard bázis). □

1.8. Lineáris leképezések

A matematika számos alkalmazásában fordulnak elő vektor-vektor függvények (más néven: leképezések), amelyek tehát szám-n-esekhez szám-k-asokat rendelnek, vagyis $f : \mathbb{R}^n \to \mathbb{R}^k$ típusúak. Ezek közül is különösen fontosak a lineárisak – ami azt jelenti, hogy a függvény kimenetének mindegyik koordinátája a bemenet koordinátáinak valamilyen lineáris kifejezése (alább ezt pontosan definiáljuk). Számos geometriai transzformáció ilyen, de többváltozós differenciálszámítást használva sokkal összetettebb függvények is közelíthetők lokálisan lineáris leképezések segítségével – hasonlóan ahhoz, ahogy az $f : \mathbb{R} \to \mathbb{R}$ függvényeket a grafikon egy pontjába húzott érintő közelíti (ha az létezik). A lineáris leképezések vizsgálata egyike a lineáris algebra gyakorlati alkalmazások által leginkább használt területeinek.

1.8.1. A lineáris leképezés fogalma

1.8.1. Definíció. *Az $f:\mathbb{R}^n\to\mathbb{R}^k$ függvényt* lineáris leképezésnek *hívjuk, ha létezik egy olyan $(k\times n)$-es A mátrix, amelyre $f(\underline{x})=A\cdot\underline{x}$ teljesül minden $\underline{x}\in\mathbb{R}^n$ esetén. Az $n=k$ esetben f-et* lineáris transzformációnak *is nevezzük. Ha $f:\mathbb{R}^n\to\mathbb{R}^k$ lineáris leképezés és $f(\underline{x})=A\cdot\underline{x}$ minden $\underline{x}\in\mathbb{R}^n$-re, akkor azt mondjuk, hogy f-*nek a mátrixa *A és ezt a tényt így jelöljük: $A=[f]$.*

A definíció szerint tehát az oszlopvektorok rögzített mátrixszal való szorzását hívjuk lineáris leképezésnek. Ha például A az alábbi mátrix, akkor az ebből keletkező $f:\mathbb{R}^4\to\mathbb{R}^3$ lineáris leképezés hozzárendelési szabálya látható A-tól jobbra.

$$A=\begin{pmatrix}2&-3&4&-5\\1&0&0&1\\0&-6&7&8\end{pmatrix}\quad\longrightarrow\quad f:\begin{pmatrix}x_1\\x_2\\x_3\\x_4\end{pmatrix}\mapsto\begin{pmatrix}2x_1-3x_2+4x_3-5x_4\\x_1+x_4\\-6x_2+7x_3+8x_4\end{pmatrix}$$

A lineáris leképezés fogalma az egyik olyan ok, ami miatt $\mathbb{R}^n$ elemeit inkább szokás oszlopvektorként és nem sorvektorként felfogni – hiszen A-t (jobbról) csak egy oszlopvektorral tudjuk megszorozni. Ennek ellenére, azokban a példákban és alkalmazásokban, ahol sík- vagy térvektorokról van szó, használni fogjuk a sorvektoros jelölést is (a korábbiakkal összhangban); ilyenkor értelemszerűen a szóban forgó sorvektor transzponáltjával kell A-t jobbról szorozni és az így kapott oszlopvektor lesz a függvény kimenetének a transzponáltja.

1.8.2. Feladat. Lineáris leképezések-e az alábbi függvények?
a) $f_1:\mathbb{R}^2\to\mathbb{R}^2$, $f_1:(x,y)\mapsto(x+2,y+3)$;
b) $f_2:\mathbb{R}^2\to\mathbb{R}^2$, $f_2:(x,y)\mapsto(x^2,y^2)$;
c) $f_3:\mathbb{R}^2\to\mathbb{R}^2$, $f_3:(x,y)\mapsto\begin{cases}(2x,0), & \text{ha } y=0\\(0,2x), & \text{ha } y\neq 0\end{cases}$;
d) $f_4:\mathbb{R}^n\to\mathbb{R}^n$, minden $x\in\mathbb{R}^n$ esetén $f_4(\underline{x})=3\cdot\underline{x}$;
e) $f_5:\mathbb{R}^2\to\mathbb{R}^2$, minden $\underline{v}\in\mathbb{R}^2$-re $f_5(\underline{v})$ a $\underline{v}$ origó körüli $+90^\circ$-os elforgatottja;
f) $f_6:\mathbb{R}^3\to\mathbb{R}^2$, $f_6:(x,y,z)\mapsto(x-y,x+z)$.

Megoldás: a) Mivel $A\cdot\underline{0}=\underline{0}$ minden A mátrixra, ezért $f(\underline{0})=\underline{0}$ teljesül minden f lineáris leképezésre. Azonban f_1 a $(0;0)$-hoz a $(2;3)$ vektort rendeli, így nem lehet lineáris leképezés.

b) Az 1.5.8. Tétel *(i)* állítása szerint $A\cdot(\lambda\cdot\underline{v})=\lambda\cdot(A\cdot\underline{v})$ minden A, λ és $\underline{v}$ esetén; így minden f lineáris leképezésnek teljesíteni kell, hogy a $\lambda\underline{v}$-n felvett függvényérték λ-szorosa a $\underline{v}$-n felvett függvényértéknek. f_2 azonban ezt nem teljesíti: $\lambda\underline{v}$-hez λ^2-szeresét rendeli a $\underline{v}$-n felvett értékének, így nem lehet lineáris leképezés. Például: $f_2\big((1;1)\big)=(1;1)$; ha f_2 lineáris leképezés volna és $[f_2]=A$, akkor

$$f_2\big((2;2)\big)=f_2\big(2\cdot(1;1)\big)=A\cdot\left(2\cdot\begin{pmatrix}1\\1\end{pmatrix}\right)=2\cdot\left(A\cdot\begin{pmatrix}1\\1\end{pmatrix}\right)=2\cdot f_2\big((1;1)\big)=(2;2).$$

Azonban f_2 képlete szerint $f_2\big((2;2)\big) = (4;4)$, így f_2 nem lineáris leképezés.

c) f_3 már betartja az $f_3(\lambda \cdot \underline{v}) = \lambda \cdot f_3(\underline{v})$ szabályt, így a b) feladat megoldása itt nem alkalmazható. Azonban az 1.5.8. Tétel *(ii)* állításából $A \cdot (\underline{u} + \underline{v}) = A \cdot \underline{u} + A \cdot \underline{v}$ következik minden A, $\underline{u}$ és $\underline{v}$ esetén. Így ha f_3 lineáris leképezés, akkor bármely két vektor összegéhez a külön-külön vett képek összegét kellene rendelje. Ezt viszont f_3 megsérti, így nem lehet lineáris leképezés. Például: f_3 képletéből $f_3\big((1;0)\big) = (2;0)$ és $f_3\big((0;1)\big) = (0;0)$, így ha f_3 lineáris leképezés volna és $[f_3] = A$, akkor

$$f_3\big((1;1)\big) = f_3\big((1;0)+(0;1)\big) = A \cdot \left(\begin{pmatrix}1\\0\end{pmatrix} + \begin{pmatrix}0\\1\end{pmatrix}\right) =$$

$$= A \cdot \begin{pmatrix}1\\0\end{pmatrix} + A \cdot \begin{pmatrix}0\\1\end{pmatrix} = f_3\big((1;0)\big) + f_3\big((0;1)\big) = (2;0)+(0;0) = (2;0).$$

Azonban f_3 képlete szerint $f_3\big((1;1)\big) = (0;2)$, így f_3 nem lineáris leképezés.

d) Legyen $A = 3E$ (ahol E az $(n \times n)$-es egységmátrix). Mivel $E \cdot \underline{x} = \underline{x}$ minden $\underline{x} \in \mathbb{R}^n$-re, ezért (az 1.5.8. Tétel *(i)* állításából) $A \cdot \underline{x} = (3E) \cdot \underline{x} = 3 \cdot (E \cdot \underline{x}) = 3 \cdot \underline{x}$. Ezért f_4 azonos az A-val való szorzással, így lineáris leképezés. (Az $A \cdot \underline{x} = 3 \cdot \underline{x}$ összefüggés közvetlenül a mátrixszorzás definíciójából is látszik – hiszen A főátlójának minden eleme 3, az összes többi eleme pedig 0.)

e) Tudjuk, hogy a $\underline{v} = (a,b)$ síkvektor 90°-os elforgatottja a $(-b,a)$ vektor – tehát ez az f_5 hozzárendelési szabálya. Könnyű találni olyan (2×2)-es A mátrixot, amellyel végzett szorzás ugyanezt a szabályt valósítja meg:

$$\begin{pmatrix}a\\b\end{pmatrix} = \underline{v}$$

$$A = \begin{pmatrix}0 & -1\\1 & 0\end{pmatrix}\begin{pmatrix}-b\\a\end{pmatrix} = A \cdot \underline{v}$$

Ezért f_5 lineáris leképezés és $[f_5] = A$, ahol A a fenti mátrix.

f) Legyen $A = \begin{pmatrix}a_1 & a_3 & a_5\\a_2 & a_4 & a_6\end{pmatrix}$ és próbáljuk megválasztani az $a_1,\dots,a_6$ értékeket úgy, hogy $f(\underline{v}) = A \cdot \underline{v}$ teljesüljön minden $\underline{v} \in \mathbb{R}^3$-ra. Legyenek $\underline{e}_1, \underline{e}_2, \underline{e}_3$ a (3×3)-as egységmátrix oszlopai (avagy az $\mathbb{R}^3$-beli standard bázis vektorai). Ekkor $A \cdot \underline{e}_1 = \begin{pmatrix}a_1\\a_2\end{pmatrix}$, $A \cdot \underline{e}_2 = \begin{pmatrix}a_3\\a_4\end{pmatrix}$ és $A \cdot \underline{e}_3 = \begin{pmatrix}a_5\\a_6\end{pmatrix}$. Mivel az f_6 képlete szerint $f_6(\underline{e}_1) = (1;1)$, $f_6(\underline{e}_2) = (-1;0)$ és $f_6(\underline{e}_3) = (0;1)$, ezért az $f(\underline{v}) = A \cdot \underline{v}$ összefüggést $\underline{v} = \underline{e}_1$-re, $\underline{v} = \underline{e}_2$-re és $\underline{v} = \underline{e}_3$-ra alkalmazva sorra megkapjuk az A oszlopait: $\begin{pmatrix}a_1\\a_2\end{pmatrix} = \begin{pmatrix}1\\1\end{pmatrix}$, $\begin{pmatrix}a_3\\a_4\end{pmatrix} = \begin{pmatrix}-1\\0\end{pmatrix}$ és $\begin{pmatrix}a_5\\a_6\end{pmatrix} = \begin{pmatrix}0\\1\end{pmatrix}$. Ezért $A = \begin{pmatrix}1 & -1 & 0\\1 & 0 & 1\end{pmatrix}$ az egyetlen lehetséges A, amelyre $[f_6] = A$ teljesülhet. Azt viszont könnyen kipróbálhatjuk, hogy a kapott A-ra $f(\underline{v}) = A \cdot \underline{v}$ fennáll-e minden $\underline{v} \in \mathbb{R}^3$ esetén:

$$\begin{pmatrix}x\\y\\z\end{pmatrix} = \underline{v}$$

$$A = \begin{pmatrix}1 & -1 & 0\\1 & 0 & 1\end{pmatrix}\begin{pmatrix}x-y\\x+z\end{pmatrix} = A \cdot \underline{v}$$

Mivel $f(\underline{v}) = A \cdot \underline{v}$ igaz, ezért f_6 lineáris leképezés és $[f_6]$ a fenti A mátrix. □

A fenti megoldásban látott gondolatok sok általános tanulsággal is szolgálnak a lineáris leképezéseket illetően. A b) és c) feladat függvényei megsértettek egy-egy olyan tulajdonságot, amelyet minden lineáris leképezésnek be kell tartani; az alábbi tételből kiderül, hogy ezek nem véletlenszerűek voltak: ezek a tulajdonságok jellemzik is a lineáris leképezéseket. Az f) feladat megoldása pedig arra mutat példát, hogy f ismeretében hogyan kapható meg annak az $[f]$ mátrixa. (Érdemes megfigyelni, hogy az f) feladat megoldásának mintájára az 1.8.2. Feladat összes korábbi részfeladata is megoldható: az $f(\underline{e}_i)$ vektorokból összeálló A mátrix az egyetlen lehetséges jelölt $[f]$-re, így ha $f(\underline{v}) = A \cdot \underline{v}$ minden $\underline{v}$-re teljesül, akkor f lineáris leképezés, ha pedig nem, akkor nem.)

1.8.3. Tétel. *Az $f : \mathbb{R}^n \to \mathbb{R}^k$ függvény akkor és csak akkor lineáris leképezés, ha teljesül rá az alábbi két tulajdonság:*
(i) $f(\underline{x}+\underline{y}) = f(\underline{x}) + f(\underline{y})$ igaz minden $x, y \in \mathbb{R}^n$ esetén;
(ii) $f(\lambda \cdot \underline{x}) = \lambda \cdot f(\underline{x})$ igaz minden $x \in \mathbb{R}^n$ és $\lambda \in \mathbb{R}$ esetén.
Ha pedig f teljesíti ezt a két tulajdonságot (és így lineáris leképezés), akkor f-nek az $[f]$ mátrixa egyértelmű és azonos azzal a $(k \times n)$-es mátrixszal, amelynek minden $1 \leq i \leq n$ esetén az i-edik oszlopa $f(\underline{e}_i)$. (Itt $\underline{e}_i$ az $\mathbb{R}^n$-beli standard bázis vektora).

Bizonyítás: Tegyük fel, hogy f lineáris leképezés és $[f] = A$. Ekkor az 1.5.8. Tétel *(ii)* állítása szerint $f(\underline{x}+\underline{y}) = A \cdot (\underline{x}+\underline{y}) = A \cdot \underline{x} + A \cdot \underline{y} = f(\underline{x}) + f(\underline{y})$ és az 1.5.8. Tétel *(i)* állítása miatt $f(\lambda \cdot \underline{x}) = A \cdot (\lambda \cdot \underline{x}) = \lambda \cdot (A \cdot \underline{x}) = \lambda \cdot f(\underline{x})$. Ezzel a tételbeli feltétel szükségességét beláttuk: ha f lineáris leképezés, akkor betartja az *(i)* és *(ii)* tulajdonságokat. Most azt látjuk be, hogy ha f lineáris leképezés, akkor $[f]$ egyértelmű. Legyen f-nek A (egy lehetséges) mátrixa, jelölje A-nak az i-edik oszlopát $\underline{a}_i$ minden i-re. A mátrixszorzás definíciójából $A \cdot \underline{e}_i = \underline{a}_i$ adódik (ez kiolvasható az 1.13. ábrából is – hiszen $\underline{e}_i$ az egységmátrix i-edik oszlopa). Ebből $[f] = A$ miatt $f(\underline{e}_i) = A \cdot \underline{e}_i = \underline{a}_i$ következik, ami bizonyítja $[f]$ egyértelműségét: $[f]$ csak az a mátrix lehet, amelynek az i-edik oszlopa $f(\underline{e}_i)$ – vagyis csak A.

Végül belátjuk a tétel feltételének az elégségességét: ha az *(i)* és *(ii)* tulajdonságok teljesülnek, akkor f lineáris leképezés. Kell tehát mutatnunk egy olyan A mátrixot, amelyre $f(\underline{x}) = A \cdot \underline{x}$ teljesül minden $\underline{x} \in \mathbb{R}^n$ esetén. Ebben azonban a fenti bekezdés gondolatmenete segít: azt már tudjuk, hogy az egyetlen szóba jövő A mátrix az, amelynek az i-edik oszlopa $f(\underline{e}_i)$ minden i-re (hasonlóan az 1.8.2. Feladat f) részének megoldásához). Legyen tehát A az így definiált mátrix, az i-edik oszlopát (amely tehát $f(\underline{e}_i)$-vel egyenlő) jelölje $\underline{a}_i$. Ekkor az $f(\underline{x}) = A \cdot \underline{x}$ feltétel teljesül az $\underline{x} = \underline{e}_i$ vektorokra; azt kell belátni, hogy minden más $\underline{x}$-re is.

Ehhez először azt mutatjuk meg, hogy a tételbeli *(i)* feltétel (illetve az azzal analóg állítás) n tagú összegekre is teljesül, ha kéttagúakra igaz:

$$f(\underline{v}_1 + \underline{v}_2 + \underline{v}_3 + \ldots + \underline{v}_n) = f(\underline{v}_1) + f(\underline{v}_2 + \underline{v}_3 + \ldots + \underline{v}_n) =$$
$$f(\underline{v}_1) + f(\underline{v}_2) + f(\underline{v}_3 + \ldots + \underline{v}_n) = \ldots = f(\underline{v}_1) + f(\underline{v}_2) + f(\underline{v}_3) + \ldots + f(\underline{v}_n),$$

vagyis $(n-1)$-szer egymás után alkalmazva az *(i)*-et kapjuk, hogy az n vektor összegére is érvényes.

Legyen most $\underline{x} \in \mathbb{R}^n$ tetszőleges, az i-edik koordinátáját jelölje x_i. Ekkor $\underline{x} = x_1 \cdot \underline{e}_1 + x_2 \cdot \underline{e}_2 + \ldots + x_n \cdot \underline{e}_n$ (lásd az 1.2.21. Állítás bizonyítását). Ebből

$$f(\underline{x}) = f(x_1 \cdot \underline{e}_1 + x_2 \cdot \underline{e}_2 + \ldots + x_n \cdot \underline{e}_n) = f(x_1 \cdot \underline{e}_1) + f(x_2 \cdot \underline{e}_2) + \ldots + f(x_n \cdot \underline{e}_n) =$$
$$= x_1 \cdot f(\underline{e}_1) + x_2 \cdot f(\underline{e}_2) + \ldots + x_n \cdot f(\underline{e}_n) = x_1 \cdot \underline{a}_1 + x_2 \cdot \underline{a}_2 + \ldots + x_n \cdot \underline{a}_n = A \cdot \underline{x}.$$

Itt először azt használtuk fel, hogy f teljesíti a tételbeli *(i)*-et (n tagú összegre), utána azt, hogy a *(ii)*-t. Az utolsó egyenlőség pedig a mátrixszorzás definíciójából következik: az $A \cdot \underline{x}$ szorzat nem más, mint az A oszlopaiból az $\underline{x}$ koordinátáival, mint együtthatókkal képzett lineáris kombináció (lásd az 1.5.14. Tételt, illetve az utána írt megjegyzést).

Beláttuk tehát, hogy $f(\underline{x}) = A \cdot \underline{x}$ teljesül minden $\underline{x} \in \mathbb{R}^n$ esetén, amivel a tétel bizonyítása teljes. □

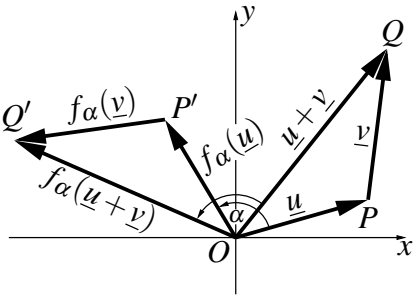
A fenti tétel segítségével számos fontos geometriai transzformációról lehet belátni, hogy lineáris leképezés. Erre mutat példát az alábbi állítás.

1.8.4. Állítás. *Legyen $f_\alpha : \mathbb{R}^2 \to \mathbb{R}^2$ az a függvény, amely minden $\underline{v} \in \mathbb{R}^2$ síkvektorhoz annak az origó körüli α szöggel való elforgatottját rendeli. Ekkor f_α lineáris transzformáció, amelynek a mátrixa*

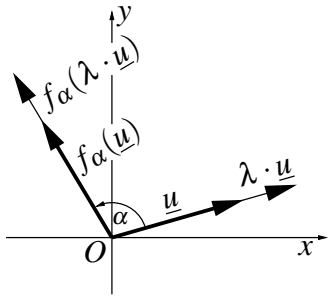
$$[f_\alpha] = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$$

Bizonyítás: Az 1.8.3. Tétel értelmében az $f_\alpha(\underline{u} + \underline{v}) = f_\alpha(\underline{u}) + f_\alpha(\underline{v})$ és az $f_\alpha(\lambda \cdot \underline{u}) = \lambda \cdot f_\alpha(\underline{u})$ összefüggéseket kell megmutatnunk minden $\underline{u}$ és $\underline{v}$ síkvektorra, illetve λ skalárra. Mindkét összefüggés kiolvasható az 1.14. ábrából. Az első esetben $\underline{u} = \overrightarrow{OP}$ és $\underline{v} = \overrightarrow{PQ}$, ekkor $\underline{u} + \underline{v} = \overrightarrow{OQ}$. Az OPQ háromszöget O körül α szöggel elforgatva kapjuk az $OP'Q'$ háromszöget, így $f_\alpha(\underline{u}) = \overrightarrow{OP'}$, $f_\alpha(\underline{v}) = \overrightarrow{P'Q'}$ és $f_\alpha(\underline{u}+\underline{v}) = \overrightarrow{OQ'}$. A síkvektorok összeadásának definíciójából következik, hogy $\overrightarrow{OQ'} = \overrightarrow{OP'} + \overrightarrow{P'Q'}$, így $f_\alpha(\underline{u}+\underline{v}) = f_\alpha(\underline{u}) + f_\alpha(\underline{v})$ valóban igaz. A másik összefüggés indoklása hasonló: $f_\alpha(\underline{u})$-ból $f_\alpha(\lambda \cdot \underline{u})$ szintén λ-val való szorzással kapható meg, így $f_\alpha(\lambda \cdot \underline{u}) = \lambda \cdot f_\alpha(\underline{u})$. Tehát f_α teljesíti az 1.8.3. Tételben szereplő két feltételt, ezért valóban lineáris leképezés.

Szintén az 1.8.3. Tételből tudjuk, hogy $[f_\alpha]$ első oszlopa $f_\alpha\big((1;0)\big)$, a második oszlopa $f_\alpha\big((0;1)\big)$ (illetve mindkét esetben nyilván a képvektorok transzponáltja). A $\cos\alpha$ és a $\sin\alpha$ definíciója szerint $f_\alpha\big((1;0)\big) = (\cos\alpha, \sin\alpha)$. Mivel $(0;1)$ az $(1;0)$-nak $90°$-kal való elforgatottja, ezért ugyanez elmondható $f_\alpha\big((1;0)\big)$ és $f_\alpha\big((0;1)\big)$ viszonyában is. Így $f_\alpha\big((0;1)\big) = (-\sin\alpha, \cos\alpha)$ (felhasználva, hogy (a,b) $90°$-os elforgatottja $(-b,a)$). Ezekből tehát valóban következik, hogy $[f_\alpha]$ az állításban megadott mátrix. □



1.14a ábra



1.14b ábra

A fentihez hasonlóan mutatható meg, hogy a síkban az origón átmenő egyenesre való tükrözés vagy vetítés, térben az origón átmenő tengely körüli forgatás, az origón átmenő síkra való tükrözés és vetítés és még néhány további, az alkalmazásokban gyakran előkerülő geometriai transzformáció lineáris leképezés. Ugyanez viszont már nem mondható el az eltolásról: az 1.8.2. Feladat a) részében látott f_1 a $(2;3)$ vektorral való eltolás és f_1-ről kiderült, hogy nem lineáris leképezés; hasonló okból nem az egyetlen nemnulla vektorral való eltolás sem. (Érdemes azonban a részletek mellőzésével azt is megemlíteni, hogy a számítógépes grafikában a sík és a tér pontjait általában nem a Descartes-féle koordinátarendszerben megszokott módon, hanem úgynevezett *homogén koordinátákkal* adják meg. Ezt az alakot használva a fent felsorolt transzformációk mellett már az eltolás is lineáris leképezés.)

1.8.2. Lineáris leképezések szorzata

A középiskolai tanulmányokból is ismert a függvények kompozíciójának, vagyis egymás után alkalmazásának fogalma: ha $f : A \to B$ és $g : B \to C$ tetszőleges függvények, akkor ezek *kompozíciója* a $h : A \to C$ függvény, ha minden $x \in A$-ra $h(x) = g\big(f(x)\big)$. Ennek a jele: $h = g \circ f$. (Vigyázni kell tehát arra, hogy először a jobbra írt függvényt alkalmazzuk.) Ha f és g lineáris leképezések, akkor a kompozíciójukat inkább *szorzatnak* szokás nevezni; ennek oka az alábbi, egyszerűsége dacára is igen fontos tételben rejlik.

1.8.5. Tétel. *Legyenek $f : \mathbb{R}^n \to \mathbb{R}^k$ és $g : \mathbb{R}^k \to \mathbb{R}^m$ lineáris leképezések. Ekkor ezeknek a $g \circ f$ szorzata is lineáris leképezés, amelyre $[g \circ f] = [g] \cdot [f]$. (Szavakban: a szorzatleképezés mátrixa a tagok mátrixának a – megfelelő sorrendben vett – szorzata.)*

Bizonyítás: Legyen $[f] = A$ és $[g] = B$. Ekkor tehát minden $\underline{x} \in \mathbb{R}^n$-re $f(\underline{x}) = A \cdot \underline{x}$ és minden $\underline{y} \in \mathbb{R}^k$-ra $g(\underline{y}) = B \cdot \underline{y}$. Alkalmazzuk a $g \circ f$ függvényt egy tetszőleges $\underline{x} \in \mathbb{R}^n$-re:

$$(g \circ f)(\underline{x}) = g\big(f(\underline{x})\big) = g\big(A \cdot \underline{x}\big) = B \cdot (A \cdot \underline{x}) = (B \cdot A) \cdot \underline{x},$$

ahol az utolsó lépésben az 1.5.8. Tétel *(iii)* állítását, a mátrixszorzás asszociativitását alkalmaztuk. A kapott összefüggés a tételnek mindkét állítását bizonyítja: látjuk, hogy $g \circ f$ azonos a $B \cdot A$ mátrixszal való szorzással, így valóban lineáris leképezés és a mátrixa $B \cdot A = [g] \cdot [f]$. □

A fenti bizonyítás kapcsán érdemes megjegyezni a következőket. A $g \circ f$ szorzat lineáris leképezés voltát definíció szerint igazoltuk, a bizonyítás egyetlen lényeges eszköze a mátrixszorzás asszociativitása volt. Megtehettük volna azt is, hogy az 1.8.3. Tétel alapján bizonyítunk: a szükséges és elégséges feltétel teljesülését jelentő két tulajdonság $(g \circ f)$-re könnyen levezethető lett volna abból, hogy ugyanezt a két tulajdonságot f és g külön-külön teljesíti – ami viszont az 1.5.8. Tétel *(i)* és *(ii)* állításából következett. Ez a bizonyítás a fentinél valamivel bonyolultabb lett volna, de mégis van egy elvi jelentőségű előnye: lehetőséget adott volna arra, hogy a mátrixszorzás asszociativitását bebizonyítsuk a fenti és az 1.8.3. Tételekből. (Így végül is az 1.5.8. Tétel *(iii)* állítása áttételesen következett volna ugyanannak a tételnek az *(i)* és *(ii)* állításából.) Mindez azonban csak elvi jelentőséggel bír, a részletek végiggondolását az érdeklődő olvasóra bízzuk.

A fenti tétel alkalmazásaképpen bebizonyítjuk az alábbi, a matematika számos területén alapvető fontosságú tételt. A bizonyítás jól mutatja, hogy a lineáris leképezések elmélete miért annyira hasznos számos alkalmazás (például a számítógépes grafika) számára: a fenti tételt még két igen egyszerű síkbeli transzformációra alkalmazva is cseppet sem triviális összefüggésekre juthatunk.

1.8.6. Tétel. (Addíciós képletek a sin és cos függvényekre)
Tetszőleges α és β szögekre teljesülnek az alábbi összefüggések:
(i) $\sin(\alpha+\beta) = \sin\alpha \cdot \cos\beta + \cos\alpha \cdot \sin\beta$
(ii) $\cos(\alpha+\beta) = \cos\alpha \cdot \cos\beta - \sin\alpha \cdot \sin\beta$

Bizonyítás: Legyen $f_\alpha : \mathbb{R}^2 \to \mathbb{R}^2$, illetve $f_\beta : \mathbb{R}^2 \to \mathbb{R}^2$ a síkban az origó körüli α, illetve β szöggel való elforgatás. Az 1.8.4. Állításból tudjuk, hogy ezek lineáris leképezések. Alkalmazzuk ezekre az 1.8.5. Tételt! Nyilván igaz, hogy $f_\alpha \circ f_\beta$ azonos $f_{\alpha+\beta}$-val, az origó körüli $\alpha+\beta$ szögű elforgatással (hiszen egy tetszőleges $\underline{v}$-t először β, majd α szöggel elforgatva ugyanazt kapjuk, mintha $\alpha+\beta$ szöggel forgattuk volna el). Az f_α, f_β és $f_{\alpha+\beta}$ lineáris transzformációk mátrixa kiolvasható az 1.8.4. Állításból, ezekre az 1.8.5. Tétel szerint fennáll az $[f_{\alpha+\beta}] = [f_\alpha] \cdot [f_\beta]$ összefüggés:

$$\begin{pmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{pmatrix} = [f_\beta]$$

$$[f_\alpha] = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} \cos(\alpha+\beta) & -\sin(\alpha+\beta) \\ \sin(\alpha+\beta) & \cos(\alpha+\beta) \end{pmatrix} = [f_{\alpha+\beta}]$$

Az $[f_{\alpha+\beta}]$ és $[f_\alpha] \cdot [f_\beta]$ mátrixok egyenlősége a tétel mindkét állítását bizonyítja: a mátrixszorzás definíciója szerint kiszámítva a szorzat bal alsó, illetve bal fölső sarkában álló elemet éppen az *(i)*, illetve *(ii)* addíciós képleteket kapjuk. □

1.8.3. Magtér, képtér

A lineáris leképezésekkel kapcsolatos két alapvető fogalmat definiálunk.

1.8.7. Definíció. *Legyen $f:\mathbb{R}^n\to\mathbb{R}^k$ lineáris leképezés. f* magterének *nevezzük és* $\operatorname{Ker} f$*-fel jelöljük azon $\mathbb{R}^n$-beli vektorok halmazát, amelyeknek a képe az $\mathbb{R}^k$-beli nullvektor:*

$$\operatorname{Ker} f=\left\{\underline{x}\in\mathbb{R}^n : f(\underline{x})=\underline{0}\right\}.$$

f képterének *nevezzük és* $\operatorname{Im} f$*-fel jelöljük azon $\mathbb{R}^k$-beli vektorok halmazát, amelyek megkaphatók (legalább) egy alkalmas $\mathbb{R}^n$-beli vektor f-fel vett képeként:*

$$\operatorname{Im} f=\left\{\underline{y}\in\mathbb{R}^k : \exists\underline{x}\in\mathbb{R}^n, f(\underline{x})=\underline{y}\right\}.$$

$\operatorname{Im} f$ tehát valójában nem más, mint az f függvény értékkészlete. Érdemes a két definíciót az f mátrixán keresztül is megfogalmazni: ha $[f]=A$ (és így $f(\underline{x})=A\cdot\underline{x}$ minden $\underline{x}\in\mathbb{R}^n$-re), akkor definíció szerint $\operatorname{Ker} f$ az $A\cdot\underline{x}=\underline{0}$ lineáris egyenletrendszer megoldásaiból áll, $\operatorname{Im} f$ pedig az $A\cdot\underline{x}=\underline{y}$ alakban előálló $\underline{y}$ vektorokból. Legyen például $f:\mathbb{R}^4\to\mathbb{R}^3$ a 90. oldalon már látott lineáris leképezés:

$$A=[f]=\begin{pmatrix}2&-3&4&-5\\1&0&0&1\\0&-6&7&8\end{pmatrix}\longrightarrow f:\begin{pmatrix}x_1\\x_2\\x_3\\x_4\end{pmatrix}\mapsto\begin{pmatrix}2x_1-3x_2+4x_3-5x_4\\x_1+x_4\\-6x_2+7x_3+8x_4\end{pmatrix}$$

Ekkor $\operatorname{Ker} f$ a $2x_1-3x_2+4x_3-5x_4=0, x_1+x_4=0, -6x_2+7x_3+8x_4=0$ lineáris egyenletrendszer $\mathbb{R}^4$-beli megoldásaiból áll, $\operatorname{Im} f$ pedig azokból az $(y_1,y_2,y_3)\in\mathbb{R}^3$ vektorokból, amelyekre megoldható a $2x_1-3x_2+4x_3-5x_4=y_1,\ x_1+x_4=y_2,$ $-6x_2+7x_3+8x_4=y_3$ lineáris egyenletrendszer.

Fontos hangsúlyozni, hogy ha $f:\mathbb{R}^n\to\mathbb{R}^k$ lineáris leképezés, akkor míg $\operatorname{Ker} f$ $\mathbb{R}^n$-nek, addig $\operatorname{Im} f$ $\mathbb{R}^k$-nak a részhalmaza. Mégpedig nem is tetszőleges részhalmaza – ezt mondja ki az alábbi állítás (amely egyben a két fogalom nevében szereplő „tér” utótagra is magyarázatul szolgál).

1.8.8. Állítás. *Ha $f:\mathbb{R}^n\to\mathbb{R}^k$ lineáris leképezés, akkor*
(i) $\operatorname{Ker} f\le\mathbb{R}^n$*, vagyis* $\operatorname{Ker} f$ *altér $\mathbb{R}^n$-ben;*
(ii) $\operatorname{Im} f\le\mathbb{R}^k$*, vagyis* $\operatorname{Im} f$ *altér $\mathbb{R}^k$-ban.*

Bizonyítás: Az *(i)* belátásához az 1.2.4. Definíció szerint azt kell megmutatnunk, hogy bármely $\underline{x}_1,\underline{x}_2\in\operatorname{Ker} f$ és $\lambda\in\mathbb{R}$ esetén $\underline{x}_1+\underline{x}_2\in\operatorname{Ker} f$ és $\lambda\cdot\underline{x}_1\in\operatorname{Ker} f$ teljesülnek. Ha $\underline{x}_1,\underline{x}_2\in\operatorname{Ker} f$, akkor $f(\underline{x}_1)=\underline{0}$ és $f(\underline{x}_2)=\underline{0}$; ebből az 1.8.3. Tételbeli *(i)* tulajdonságot felhasználva $f(\underline{x}_1+\underline{x}_2)=f(\underline{x}_1)+f(\underline{x}_2)=\underline{0}+\underline{0}=\underline{0}$ következik, így $\underline{x}_1+\underline{x}_2\in\operatorname{Ker} f$. Hasonlóan, az 1.8.3. Tételbeli *(ii)* tulajdonságból $f(\lambda\cdot\underline{x}_1)=\lambda\cdot f(\underline{x}_1)=\lambda\cdot\underline{0}=\underline{0}$, így $\lambda\cdot\underline{x}_1\in\operatorname{Ker} f$ is igaz. Kiegészítve ezt azzal,

hogy $\operatorname{Ker} f$ nem lehet üres, hiszen $\underline{0} \in \operatorname{Ker} f$ definíció szerint mindig igaz, az *(i)* bizonyítása teljes.

A *(ii)* állítást a fentihez hasonlóan is meg lehetne mutatni, de valójában erre nincs szükség. Említettük, hogy ha $[f] = A$, akkor $\operatorname{Im} f$ definíció szerint azokból az $\underline{y} \in \mathbb{R}^k$ vektorokból áll, amelyek kifejezhetők $A \cdot \underline{x} = \underline{y}$ alakban – vagyis amelyekre az $A \cdot \underline{x} = \underline{y}$ lineáris egyenletrendszer megoldható. Az 1.5.14. Tétel szerint ez azzal ekvivalens, hogy $\underline{y} \in \langle \underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n \rangle$, ahol $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n$ az A oszlopait jelölik. Más szóval: $\operatorname{Im} f$ nem más, mint az $\langle \underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n \rangle$ generált altér, amelyről az 1.2.8. Tételben már beláttuk, hogy altér. □

Érdemes külön nyomatékkal kiemelni a fenti bizonyítás második felének azt a megfigyelését, hogy $\operatorname{Im} f$ azonos az $[f]$ oszlopai által generált altérrel. Akár azt is mondhatjuk, hogy a képtér fogalma a generált altér korábbról már ismert fogalmának ad új nevet, de attól érdemben nem különbözik. Éppen ezért egyes tankönyvek $\operatorname{Im} f$-et az A mátrix *oszlopterének* nevezik, ha $[f] = A$; hasonlóan, $\operatorname{Ker} f$-et az A *nullterének* is szokás nevezni, mert az $A \cdot \underline{x} = \underline{0}$ lineáris egyenletrendszer megoldásaiból áll. (Mi a továbbiakban is maradunk a képtér és magtér elnevezéseknél.) Akárhogy is hívjuk őket, ez a két fogalom rendkívül fontos – elsősorban az alábbi, őket összekötő tétel miatt. Lényeges kiemelni, hogy az alábbi tételnek a fenti állítás ad létjogosultságot – hiszen $\operatorname{Ker} f$ és $\operatorname{Im} f$ dimenziójáról beszélni csak akkor lehet, ha tudjuk, hogy ezek alterek.

1.8.9. Tétel. (Dimenziótétel)
Ha $f : \mathbb{R}^n \to \mathbb{R}^k$ *lineáris leképezés, akkor* $\dim \operatorname{Ker} f + \dim \operatorname{Im} f = n$.

Bizonyítás: Legyen $\dim \operatorname{Ker} f = m$ és válasszunk egy tetszőleges bázist $\operatorname{Ker} f$-ben, (az 1.2.27. Következmény szerint ezt megtehetjük); legyen ez a bázis $\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_m$. Mivel $\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_m$ lineárisan független, ezért az 1.2.26. Tétel szerint ez a rendszer kiegészíthető $\mathbb{R}^n$ egy bázisává. Mivel $\dim \mathbb{R}^n = n$, ezért ehhez további $n-m$ vektor szükséges, legyen $\underline{c}_1, \underline{c}_2, \ldots, \underline{c}_{n-m}$ egy ilyen rendszer. (Tehát $\underline{b}_1, \ldots, \underline{b}_m, \underline{c}_1, \ldots, \underline{c}_{n-m}$ bázis $\mathbb{R}^n$-ben.) Megmutatjuk, hogy a (definíció szerint $\operatorname{Im} f$-beli vektorokból álló) $f(\underline{c}_1), f(\underline{c}_2), \ldots, f(\underline{c}_{n-m})$ rendszer bázis $\operatorname{Im} f$-ben. Ebből következni fog, hogy $\dim \operatorname{Im} f = n-m$ és ezáltal a tétel állítása is (hiszen $\dim \operatorname{Ker} f = m$).

Először belátjuk, hogy $f(\underline{c}_1), f(\underline{c}_2), \ldots, f(\underline{c}_{n-m})$ generátorrendszer $\operatorname{Im} f$-ben. Legyen ugyanis $\underline{y} \in \operatorname{Im} f$ tetszőleges, ekkor $\underline{y} = f(\underline{x})$ valamely $\underline{x} \in \mathbb{R}^n$-re. Mivel $\underline{b}_1, \ldots, \underline{b}_m, \underline{c}_1, \ldots, \underline{c}_{n-m}$ generátorrendszer $\mathbb{R}^n$-ben (hiszen bázis), ezért $\underline{x}$ kifejezhető a lineáris kombinációjukként: $\underline{x} = \beta_1 \underline{b}_1 + \ldots + \beta_m \underline{b}_m + \gamma_1 \underline{c}_1 + \ldots + \gamma_{n-m} \underline{c}_{n-m}$. Alkalmazzuk itt mindkét oldalra f-et, majd használjuk ki f-nek az 1.8.3. Tétel szerinti tulajdonságait (amelyek közül az első az 1.8.3. Tétel bizonyításában írtak szerint többtagú összegre is érvényes):

$$\begin{aligned}\underline{y} = f(\underline{x}) = f(\beta_1\underline{b}_1 + \beta_2\underline{b}_2 + \ldots + \beta_m\underline{b}_m + \gamma_1\underline{c}_1 + \gamma_2\underline{c}_2 + \ldots + \gamma_{n-m}\underline{c}_{n-m}) = \\ = f(\beta_1\underline{b}_1) + f(\beta_2\underline{b}_2) + \ldots + f(\beta_m\underline{b}_m) + f(\gamma_1\underline{c}_1) + f(\gamma_2\underline{c}_2) + \ldots + f(\gamma_{n-m}\underline{c}_{n-m}) = \\ = \beta_1 f(\underline{b}_1) + \beta_2 f(\underline{b}_2) + \ldots + \beta_m f(\underline{b}_m) + \gamma_1 f(\underline{c}_1) + \gamma_2 f(\underline{c}_2) + \ldots + \gamma_{n-m} f(\underline{c}_{n-m}) = \\ = \beta_1\underline{0} + \beta_2\underline{0} + \ldots + \beta_m\underline{0} + \gamma_1 f(\underline{c}_1) + \gamma_2 f(\underline{c}_2) + \ldots + \gamma_{n-m} f(\underline{c}_{n-m}) = \\ = \gamma_1 f(\underline{c}_1) + \gamma_2 f(\underline{c}_2) + \ldots + \gamma_{n-m} f(\underline{c}_{n-m})\end{aligned}$$

Itt az utolsó előtti lépésben felhasználtuk, hogy $f(\underline{b}_1) = f(\underline{b}_2) = \ldots = f(\underline{b}_m) = \underline{0}$, ami $\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_m \in \operatorname{Ker} f$ miatt igaz. Azt kaptuk, hogy a tetszőlegesen választott $\underline{y} \in \operatorname{Im} f$ kifejezhető $f(\underline{c}_1), f(\underline{c}_2), \ldots, f(\underline{c}_{n-m})$ lineáris kombinációjaként, így ezek valóban generátorrendszert alkotnak $\operatorname{Im} f$-ben.

Most belátjuk, hogy $f(\underline{c}_1), f(\underline{c}_2), \ldots, f(\underline{c}_{n-m})$ lineárisan független. Tegyük fel ehhez, hogy $\gamma_1 f(\underline{c}_1) + \gamma_2 f(\underline{c}_2) + \ldots + \gamma_{n-m} f(\underline{c}_{n-m}) = \underline{0}$; azt kell megmutatnunk (az 1.2.12. Tétel szerint), hogy ekkor $\gamma_1 = \gamma_2 = \ldots = \gamma_{n-m} = 0$. Ismét alkalmazzuk az 1.8.3. Tétel szerinti tulajdonságokat:

$$\begin{aligned}\underline{0} = \gamma_1 f(\underline{c}_1) + \gamma_2 f(\underline{c}_2) + \ldots + \gamma_{n-m} f(\underline{c}_{n-m}) = \\ = f(\gamma_1\underline{c}_1) + f(\gamma_2\underline{c}_2) + \ldots + f(\gamma_{n-m}\underline{c}_{n-m}) = f(\gamma_1\underline{c}_1 + \gamma_2\underline{c}_2 + \ldots + \gamma_{n-m}\underline{c}_{n-m}).\end{aligned}$$

Ebből $\operatorname{Ker} f$ definíciója szerint $\gamma_1\underline{c}_1 + \gamma_2\underline{c}_2 + \ldots + \gamma_{n-m}\underline{c}_{n-m} \in \operatorname{Ker} f$. Így ez a vektor kifejezhető $\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_m$ lineáris kombinációjaként (hiszen $\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_m$ bázis $\operatorname{Ker} f$-ben):

$$\gamma_1\underline{c}_1 + \gamma_2\underline{c}_2 + \ldots + \gamma_{n-m}\underline{c}_{n-m} = \beta_1\underline{b}_1 + \beta_2\underline{b}_2 + \ldots + \beta_m\underline{b}_m.$$

Átrendezve:

$$-\beta_1\underline{b}_1 - \beta_2\underline{b}_2 - \ldots - \beta_m\underline{b}_m + \gamma_1\underline{c}_1 + \gamma_2\underline{c}_2 + \ldots + \gamma_{n-m}\underline{c}_{n-m} = \underline{0}.$$

Azonban $\underline{b}_1, \underline{b}_2 \ldots, \underline{b}_m, \underline{c}_1, \underline{c}_2, \ldots, \underline{c}_{n-m}$ lineárisan független (hiszen bázis $\mathbb{R}^n$-ben), ezért csak a triviális lineáris kombinációjuk adhatja a nullvektort; következik, hogy $\gamma_1 = \gamma_2 = \ldots = \gamma_{n-m} = 0$ (és $\beta_1 = \beta_2 = \ldots = \beta_m = 0$ is igaz).

Tehát megmutattuk, hogy $f(\underline{c}_1), f(\underline{c}_2), \ldots, f(\underline{c}_{n-m})$ lineárisan független. Így bázis is (mert azt már beláttuk, hogy generátorrendszer). Ezzel a bizonyítás teljes. □

Fentebb említettük, hogy $\operatorname{Im} f$ definíció szerint azonos az $[f]$ mátrix oszlopai által generált altérrel, így az 1.7.6. Tétel miatt $\dim \operatorname{Im} f = r([f])$. $\operatorname{Ker} f$ viszont az $[f] \cdot \underline{x} = \underline{0}$ feltételt kielégítő $\underline{x}$ vektorokból áll, ennek a dimenzióját szokás az $[f]$ mátrix *nullitásának* is nevezni. Így bizonyos tankönyvek a dimenziótételt *rang-nullitás-tétel* néven említik (hiszen az más megfogalmazásban azt mondja ki, hogy bármely mátrix rangjának és nullitásának összege egyenlő az oszlopainak számával).

1.8.10. Feladat. Legyen $f:\mathbb{R}^5\to\mathbb{R}^4$ az a lineáris leképezés, amelynek a mátrixa az alábbi A mátrix. Határozzuk meg $\dim \operatorname{Ker} f$ és $\dim \operatorname{Im} f$ értékét és adjunk meg egy-egy bázist $\operatorname{Ker} f$-ben és $\operatorname{Im} f$-ben.

$$[f]=A=\begin{pmatrix} 2 & 8 & 6 & 4 & 2\\ 1 & 2 & -1 & 12 & 7\\ -1 & -1 & 3 & -12 & 0\\ 5 & 22 & 19 & 4 & 7 \end{pmatrix}$$

Megoldás: $\operatorname{Im} f$-ről fentebb már mondtuk, hogy az azonos az A oszlopai által generált altérrel. Ezt pedig az 1.7.10. Feladatban már meghatároztuk: ott a $p=4$ értékre az $\underline{a}_1,\underline{a}_2,\dots,\underline{a}_5$ vektorok épp A oszlopaival azonosak. Akkor megmutattuk, hogy az $\langle \underline{a}_1,\underline{a}_2,\dots,\underline{a}_5\rangle$ generált altérben – tehát $\operatorname{Im} f$-ben – az $\underline{a}_1,\underline{a}_2,\underline{a}_4$ vektorok bázist alkotnak (a $p=4$ esetben). Így $\dim \operatorname{Im} f=3$.

Az 1.7.10. Feladat megoldásához a Gauss-eliminációt használtuk. Szerencsére $\operatorname{Ker} f$ meghatározásában is az A sorain végzett Gauss-elimináció segít: valóban, $\operatorname{Ker} f$-et definíció szerint az $A\cdot \underline{x}=\underline{0}$ lineáris egyenletrendszer megoldásai alkotják, ezeket pedig nyilván Gauss-eliminációval kereshetjük meg. Az $(A|\underline{0})$ kibővített együtthatómátrixon futtatva az eliminációt a vonaltól jobbra végig megmarad a nullvektor, ezért mindegy, hogy az eliminációt A-ra, vagy $(A|\underline{0})$-ra végezzük. Az 1.7.10. Feladatban csak a lépcsős alakig kellett eljutnunk a Gauss-eliminációval: ezt az ott leírt számolásban az utolsó mátrixból kapjuk az utolsó sor törlésével (a $p=4$ értékválasztásnak megfelelően). Most az $A\cdot\underline{x}=\underline{0}$ lineáris egyenletrendszer megoldásához folytatjuk az ottani számolást a redukált lépcsős alakig:

$$\sim\left(\begin{array}{ccccc|c} 1 & 4 & 3 & 2 & 1 & 0\\ 0 & 1 & 2 & -5 & -3 & 0\\ 0 & 0 & 0 & 1 & 2 & 0 \end{array}\right)\sim\left(\begin{array}{ccccc|c} 1 & 4 & 3 & 0 & -3 & 0\\ 0 & 1 & 2 & 0 & 7 & 0\\ 0 & 0 & 0 & 1 & 2 & 0 \end{array}\right)\sim$$

$$\sim\left(\begin{array}{ccccc|c} 1 & 0 & -5 & 0 & -31 & 0\\ 0 & 1 & 2 & 0 & 7 & 0\\ 0 & 0 & 0 & 1 & 2 & 0 \end{array}\right)$$

A redukált lépcsős alakból kiolvasva a lineáris egyenletrendszer megoldásait: $x_3=\alpha\in\mathbb{R}$ és $x_5=\beta\in\mathbb{R}$ szabad paraméterek, ezekből pedig a másik három változó így fejezhető ki: $x_1=5\alpha+31\beta$, $x_2=-2\alpha-7\beta$, $x_4=-2\beta$. Másképpen fogalmazva: az $A\cdot\underline{x}=\underline{0}$ megoldásai, vagyis $\operatorname{Ker} f$ elemei

$$\underline{x}=\begin{pmatrix} 5\alpha+31\beta\\ -2\alpha-7\beta\\ \alpha\\ -2\beta\\ \beta \end{pmatrix}=\alpha\begin{pmatrix} 5\\ -2\\ 1\\ 0\\ 0 \end{pmatrix}+\beta\begin{pmatrix} 31\\ -7\\ 0\\ -2\\ 1 \end{pmatrix}$$

alakúak. Ebből következik, hogy az itt jobbra látható (α-val, illetve β-val szorzott) két oszlopvektor bázist alkot $\operatorname{Ker} f$-ben: eddig azt mutattuk meg, hogy generátorrendszert alkotnak (hiszen minden $\operatorname{Ker} f$-beli vektor a fent látható módon kifejezhető a lineáris kombinációjukként), az pedig ránézésre látható, hogy lineárisan függetlenek (hiszen egyik sem skalárszorosa a másiknak). Így $\dim \operatorname{Ker} f = 2$ – összhangban a dimenziótétel állításával. □

A fenti megoldásból kiolvasható az 1.8.9. Dimenziótétel egy alternatív bizonyításának az alapgondolata is: tetszőleges f lineáris leképezés esetében is $[f]$-re lefuttatva a Gauss-eliminációt a redukált lépcsős alakban a vezéregyest tartalmazó, illetve nem tartalmazó oszlopok száma $\dim \operatorname{Im} f$-fel, illetve $\dim \operatorname{Ker} f$-fel azonos – így a kettő összege valóban az $[f]$ oszlopainak száma.

1.8.4. Lineáris transzformációk inverze

Ismert, hogy egy tetszőleges $f : A \to B$ függvény akkor *invertálható*, ha bármely $x_1, x_2 \in A$, $x_1 \neq x_2$ esetén $f(x_1) \neq f(x_2)$. Ha pedig ez teljesül, akkor f *inverze* az az f^{-1}-zel jelölt függvény, amely f értékkészletének elemeihez rendeli A elemeit, mégpedig az $y \in B$, $x \in A$ elemekre $f^{-1}(y) = x$ akkor igaz, ha $f(x) = y$. Most az invertálhatóság kérdését lineáris transzformációkra (vagyis $\mathbb{R}^n$-ről $\mathbb{R}^n$-be menő lineáris leképezésekre) vizsgáljuk. (A kérdés általános lineáris leképezésekre is feltehető, de csak a lineáris transzformációkra vonatkozó esetet fogjuk használni.) Mivel egy $f : \mathbb{R}^n \to \mathbb{R}^n$ lineáris transzformáció egy $(n \times n)$-es $[f] = A$ mátrixszal való szorzás, nem meglepő, hogy f^{-1}-nek az A^{-1} inverz mátrixhoz van köze.

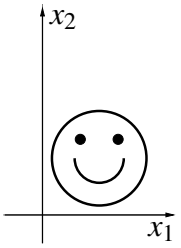
1.8.11. Tétel. *Egy $f : \mathbb{R}^n \to \mathbb{R}^n$ lineáris transzformáció akkor és csak akkor invertálható, ha $\det[f] \neq 0$. Ha pedig ez a feltétel fennáll, akkor $[f^{-1}] = [f]^{-1}$ – vagyis az f^{-1} inverz transzformáció mátrixa az f mátrixának az inverze.*

Bizonyítás: Legyen $[f] = A$, vagyis $f(\underline{x}) = A \cdot \underline{x}$ minden $\underline{x} \in \mathbb{R}^n$-re. Először a feltétel szükségességét látjuk be: ha f invertálható, akkor $\det A \neq 0$. Tegyük fel indirekt, hogy $\det A = 0$. Ekkor az 1.5.16. Tétel szerint A oszlopai lineárisan összefüggők, ami az 1.5.15. Következmény szerint úgy is fogalmazható, hogy az $A \cdot \underline{x} = \underline{0}$ lineáris egyenletrendszernek van egy $\underline{x}^* \neq \underline{0}$ megoldása. Ekkor tehát $A \cdot \underline{x}^* = \underline{0}$ és nyilván $A \cdot \underline{0} = \underline{0}$ is igaz. Ez ellentmond annak, hogy f invertálható: $\underline{x}^* \neq \underline{0}$, de $f(\underline{x}^*) = f(\underline{0})$. Ez az ellentmondás bizonyítja a feltétel szükségességét.

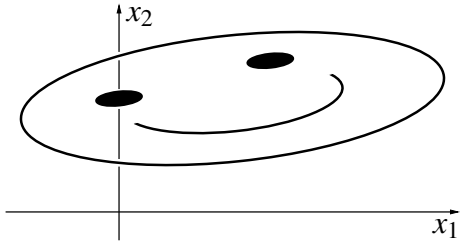
Most a feltétel elégségességét látjuk be: ha $\det A \neq 0$, akkor f invertálható. Mivel $\det A \neq 0$, ezért az 1.6.2. Tétel szerint létezik az A^{-1} inverz mátrix. Tetszőleges $\underline{x} \in \mathbb{R}^n$ esetén $f(\underline{x}) = \underline{y}$ azt jelenti, hogy $\underline{y} = A \cdot \underline{x}$. Mindkét oldalt balról A^{-1}-zel szorozva és felhasználva a mátrixszorzás asszociativitását (1.5.8. Tétel, *(iii)* állítás): $A^{-1} \cdot \underline{y} = A^{-1} \cdot (A \cdot \underline{x}) = (A^{-1} \cdot A) \cdot \underline{x} = E \cdot \underline{x} = \underline{x}$ (ahol E az $(n \times n)$-es egységmátrix). Ez éppen azt mutatja, hogy az $\underline{y} \mapsto A^{-1} \cdot \underline{y}$ függvény azonos az f inverzével – amivel beláttuk egyrészt azt, hogy f^{-1} létezik, másrészt azt, hogy $[f^{-1}] = A^{-1} = [f]^{-1}$. □

1.8.5. Bázistranszformáció

Alkalmazzuk az 1.15a ábrán látható szmájlira (annak minden $\underline{x} = (x_1, x_2)$ pontjára) az $f : \mathbb{R}^2 \to \mathbb{R}^2$, $f : (x_1, x_2) \mapsto (4x_1 - 2x_2, x_1 + x_2)$ lineáris transzformációt. Az eredmény (vagyis a kapott képpontok összessége) az 1.15b ábrán látható.



1.15a ábra



1.15b ábra

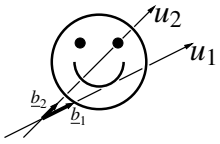
Hogyan lehetne a látott jelenséget jobban megérteni – vagyis az f működéséről a fenti képleténél jobban értelmezhető információhoz jutni? Alább ki fog derülni, hogy f hozzárendelési szabálya sokkal áttekinthetőbbé válhat, ha áttérünk egy (szerencsésen választott) új koordinátarendszerre.

Ennek az ötletnek a pontosabb megvalósításához a koordinátavektor 1.2.24. Definícióját hívjuk segítségül. Ha f $\underline{x}$-hez az $\underline{y} = f(\underline{x})$ vektort rendeli, akkor $\underline{x}$-et és $\underline{y}$-t is helyettesítjük egy $B = \{\underline{b}_1, \underline{b}_2\}$ bázis szerinti koordinátavektorával, majd megvizsgáljuk, hogy $[\underline{x}]_B$-ből milyen függvény állítja elő $[\underline{y}]_B$-t. Szemléletesen ez valóban úgy fogalmazható, hogy a pontokat abban a koordinátarendszerben írjuk fel, amelyben a tengely irányú egységvektoroknak $\underline{b}_1$ és $\underline{b}_2$ felelnek meg és f működését is ebben a koordinátarendszerben írjuk le.

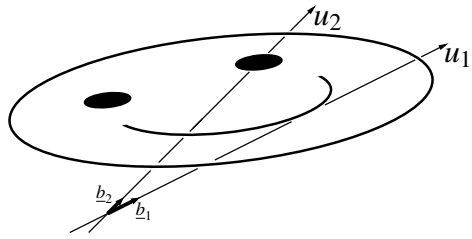
A fenti f esetében kísérletezzünk például a $\underline{b}_1 = (2;1)$, $\underline{b}_2 = (1;1)$ vektorok alkotta B bázissal. ($\underline{b}_1$ és $\underline{b}_2$ nem párhuzamosak, így valóban bázist alkotnak a síkban. Hogy miért épp ezt a bázist választottuk, arra később visszatérünk.) Ha $\underline{x} = (x_1, x_2)$ síkvektor és $[\underline{x}]_B = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$, akkor a koordinátavektor definíciójából $(x_1, x_2) = \underline{x} = u_1 \cdot \underline{b}_1 + u_2 \cdot \underline{b}_2 = u_1 \cdot (2;1) + u_2 \cdot (1;1) = (2u_1 + u_2, u_1 + u_2)$. Megoldva a $2u_1 + u_2 = x_1$, $u_1 + u_2 = x_2$ egyenletrendszert: $u_1 = x_1 - x_2$, $u_2 = -x_1 + 2x_2$. Így $[\underline{x}]_B = \begin{pmatrix} x_1 - x_2 \\ -x_1 + 2x_2 \end{pmatrix}$ igaz minden $\underline{x} = (x_1, x_2)$ síkvektorra. Alkalmazva ugyanezt az összefüggést az $f(\underline{x}) = (4x_1 - 2x_2, x_1 + x_2)$ képvektorra is: $[f(\underline{x})]_B = \begin{pmatrix} (4x_1 - 2x_2) - (x_1 + x_2) \\ -(4x_1 - 2x_2) + 2(x_1 + x_2) \end{pmatrix} = \begin{pmatrix} 3x_1 - 3x_2 \\ -2x_1 + 4x_2 \end{pmatrix}$. Összevetve az $[\underline{x}]_B$-re és az $[f(\underline{x})]_B$-re kapott képleteket rögtön látszik, hogy az utóbbi nagyon egyszerű szabály szerint kapható az előbbiből: $[\underline{x}]_B$ első koordinátáját 3-mal, a másodikat 2-vel szorozva épp $[f(\underline{x})]_B$-t kapjuk. Más szóval a $g : [\underline{x}]_B \mapsto [f(\underline{x})]_B$ függvény hozzárendelési szabálya: $g : \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \mapsto \begin{pmatrix} 3u_1 \\ 2u_2 \end{pmatrix}$.

Ez az eredmény szemléletesen úgy is megfogalmazható, hogy áttérve arra az

új koordinátarendszerre, amelyben a két tengelyirányú egységvektor $\underline{b}_1 = (2;1)$ és $\underline{b}_2 = (1;1)$, f az első tengely irányában 3-szorosára, a második irányában 2-szeresére nyújtja a vektorokat. Ez f eredeti hozzárendelési szabályánál még akkor is sokkal többet mond f működéséről, ha a $\underline{b}_1$ és $\underline{b}_2$ által alkotott koordinátarendszer nem derékszögű. Az 1.16. ábrán újra látható az eredeti és az f-fel transzformált szmájli az új koordinátarendszerben is.



1.16a ábra



1.16b ábra

Egy tetszőleges $f : \mathbb{R}^n \to \mathbb{R}^n$ lineáris transzformációról is sokat elmondhat, ha választunk egy $B = \{\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_n\}$ bázist $\mathbb{R}^n$-ben és f helyett a $g : [\underline{x}]_B \mapsto [f(\underline{x})]_B$ hozzárendelést vizsgáljuk meg. Ezt *bázistranszformációnak* hívjuk, az alábbi tétel pedig ennek a megvalósítását írja le. A tétel kimondásához szükségünk lesz a B bázis vektorainak egyesítésével keletkező $(n \times n)$-es mátrixra is. A továbbiakban a B jelölést fogjuk használni egyrészt a $B = \{\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_n\}$ bázisra, másrészt annak az $(n \times n)$-es mátrixnak a jelölésére is, amelynek az oszlopai sorban $\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_n$. Ez a kétértelműség nem fog félreértést okozni.

1.8.12. Tétel. *Legyen $f : \mathbb{R}^n \to \mathbb{R}^n$ lineáris transzformáció és B egy $(n \times n)$-es mátrix, amelynek az oszlopai bázist alkotnak $\mathbb{R}^n$-ben. Jelölje $g : \mathbb{R}^n \to \mathbb{R}^n$ azt a függvényt, amely minden $\underline{x} \in \mathbb{R}^n$ esetén $[\underline{x}]_B$-hez $[f(\underline{x})]_B$-t rendeli. Ekkor g is lineáris transzformáció, amelynek a mátrixa $[g] = B^{-1} \cdot [f] \cdot B$.*

Bizonyítás: Először is vegyük észre, hogy a B oszlopai akkor és csak akkor alkotnak bázist, ha $\det B \neq 0$. Valóban, az 1.2.28. Következmény szerint $\mathbb{R}^n$ bázisai az n tagú lineárisan független rendszerek. Azt pedig az 1.5.16. Tételből tudjuk, hogy B oszlopainak lineáris függetlensége ekvivalens azzal, hogy $\det B \neq 0$. Ebből az 1.6.2. Tétel szerint következik, hogy a B^{-1} inverz mátrix valóban létezik.

A bizonyításhoz használni fogjuk az alábbi, önmagában is fontos lemmát.

1.8.13. Lemma. *Legyen $h : \mathbb{R}^n \to \mathbb{R}^n$ az a függvény, amely minden $\underline{x} \in \mathbb{R}^n$ esetén $[\underline{x}]_B$-hez $\underline{x}$-et rendeli. Ekkor h lineáris transzformáció, amelynek a mátrixa $[h] = B$.*

A Lemma bizonyítása: Jelölje valamely $\underline{x} \in \mathbb{R}^n$-re az $[\underline{x}]_B$ koordinátavektor i-edik koordinátáját α_i minden $1 \leq i \leq n$ esetén. Ekkor tehát $\underline{x} = \alpha_1 \underline{b}_1 + \alpha_2 \underline{b}_2 + \ldots + \alpha_n \underline{b}_n$.

Azonban a mátrixszorzás definíciója szerint a $B \cdot [\underline{x}]_B$ szorzat is azonos a B oszlopaiból (vagyis a $\underline{b}_i$-kből) az $[\underline{x}]_B$ koordinátáival (vagyis az α_i-kkel), mint együtthatókkal képzett lineáris kombinációval (lásd az 1.5.14. Tételt, illetve az utána írt megjegyzést). Így $\underline{x} = B \cdot [\underline{x}]_B$, ami mutatja, hogy a $h : [\underline{x}]_B \mapsto \underline{x}$ függvény lineáris transzformáció, amelynek a mátrixa B. ◇

Mivel $\det B \neq 0$, ezért az 1.8.11. Tétel szerint a h^{-1} inverz transzformáció is létezik és a mátrixa $[h^{-1}] = [h]^{-1} = B^{-1}$. Ez nyilván $\underline{x}$-hez rendeli $[\underline{x}]_B$-t minden $\underline{x} \in \mathbb{R}^n$ esetén.

Rátérve a tétel bizonyítására, annak a fő gondolata az a megfigyelés, hogy a tétel szövegében definiált $g : [\underline{x}]_B \mapsto [f(\underline{x})]_B$ függvény azonos a $h^{-1} \circ f \circ h$ függvénnyel (ahol h a fenti lemma által bevezett függvény, a $\circ$ pedig a kompozíciót jelöli). Valóban, ha $[\underline{x}]_B$-re először alkalmazzuk h-t, akkor $\underline{x}$-et kapjuk; erre f-et alkalmazva kapjuk $f(\underline{x})$-et; végül erre h^{-1}-et alkalmazva valóban $[f(\underline{x})]_B$ az eredmény. Így az 1.8.5. Tétel szerint a $g = h^{-1} \circ f \circ h$ függvény valóban lineáris transzformáció és a mátrixa $[g] = [h^{-1}] \cdot [f] \cdot [h] = B^{-1} \cdot [f] \cdot B$. □

A tételben bevezetett g lineáris transzformáció mátrixának ad nevet az alábbi definíció.

1.8.14. Definíció. *Legyen $f : \mathbb{R}^n \to \mathbb{R}^n$ lineáris transzformáció és B bázis $\mathbb{R}^n$-ben. Ekkor a $g : [\underline{x}]_B \mapsto [f(\underline{x})]_B$ lineáris transzformáció mátrixát az f* transzformáció *B* bázis szerinti mátrixának nevezzük. *Ennek a jele: $[f]_B$.*

Fontos kiemelni a különbséget az $[f]$ és az $[f]_B$ jelölések között: $[f]$ magának az f transzformációnak a mátrixa, vagyis $f(\underline{x}) = [f] \cdot \underline{x}$ minden $\underline{x} \in \mathbb{R}^n$-re; $[f]_B$ viszont már nem csak magától f-től, hanem egy B bázistól is függ és egy másik lineáris transzformáció (mégpedig a $g : [\underline{x}]_B \mapsto [f(\underline{x})]_B$) mátrixát jelöli.

A fogalmat a fentebb már vizsgált $f : (x_1, x_2) \mapsto (4x_1 - 2x_2, x_1 + x_2)$ lineáris transzformációval és a $B = \{(2;1), (1;1)\}$ bázissal illusztráljuk. Ekkor f mátrixa: $[f] = \begin{pmatrix} 4 & -2 \\ 1 & 1 \end{pmatrix}$, mert $\begin{pmatrix} 4 & -2 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 4x_1 - 2x_2 \\ x_1 + x_2 \end{pmatrix}$. A B-nek megfelelő mátrix $B = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$, ennek az inverze $B^{-1} = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$ (ez Gauss-eliminációval kiszámítható, vagy közvetlenül is ellenőrizhető). Ezért az 1.8.12. Tétel szerint

$$[f]_B = B^{-1} \cdot [f] \cdot B = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} \cdot \begin{pmatrix} 4 & -2 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}.$$

Ez tehát azt jelenti, hogy a $g : [\underline{x}]_B \mapsto [f(\underline{x})]_B$ függvény azonos a kapott $[f]_B$-vel való szorzással, vagyis az $\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \mapsto \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix} \cdot \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} 3u_1 \\ 2u_2 \end{pmatrix}$ lineáris transzformációval. Ez megfelel annak, amit fentebb kiszámítottunk.

Az alábbi tétel összefoglalja az $[f]_B$ mátrixszal kapcsolatban eddig mondottakat és egy további, alkalmazásokban nagyon hasznos állítást is kimond.

1.8.15. Tétel. *Legyen $f : \mathbb{R}^n \to \mathbb{R}^n$ lineáris transzformáció és B egy $(n \times n)$-es mátrix, amelynek az oszlopai bázist alkotnak $\mathbb{R}^n$-ben. Ekkor az $[f]_B$ mátrixra az alábbiak teljesülnek:*

(i) $[f(\underline{x})]_B = [f]_B \cdot [\underline{x}]_B$ *minden* $\underline{x} \in \mathbb{R}^n$*-re;*

(ii) $[f]_B = B^{-1} \cdot [f] \cdot B$*;*

(iii) az $[f]_B$ *i-edik oszlopa egyenlő az* $[f(\underline{b}_i)]_B$ *koordinátavektorral minden* $1 \leq i \leq n$ *esetén.*

Bizonyítás: A *(ii)* állítást már beláttuk az 1.8.12. Tételben, az *(i)* pedig közvetlenül következik az $[f]_B$ 1.8.14. Definíciójából (és az annak létjogosultságot adó 1.8.12. Tételből): mivel $[f]_B$ annak a g lineáris transzformációnak a mátrixa, amely minden $\underline{x} \in \mathbb{R}^n$-re $[\underline{x}]_B$-hez $[f(\underline{x})]_B$-t rendeli, ezért (a lineáris leképezés 1.8.1. Definíciója szerint) $[f(\underline{x})]_B = [f]_B \cdot [\underline{x}]_B$ valóban igaz.

A *(iii)* állítás pedig az 1.8.3. Tétel közvetlen következménye: mivel $[f]_B$ a $g : [\underline{x}]_B \mapsto [f(\underline{x})]_B$ lineáris transzformáció mátrixa, ezért az i-edik oszlopa $g(\underline{e}_i)$-vel egyenlő minden i-re. Mivel a koordinátavektor definíciója szerint $\underline{e}_i$ éppen a $\underline{b}_i$ koordinátavektora (vagyis $\underline{e}_i = [\underline{b}_i]_B$), ezért $g(\underline{e}_i) = g\big([\underline{b}_i]_B\big) = [f(\underline{b}_i)]_B$. □

A fenti tétel állításait érdemes abban a speciális esetben végiggondolni, amikor B a standard bázis (vagyis mátrixként $B = E$, az egységmátrix). Ekkor $[f] = [f]_B$, hiszen a $B = E$ esetben $\underline{x} = [\underline{x}]_B$ minden $\underline{x} \in \mathbb{R}^n$-re, így a g transzformáció azonos f-fel. Ugyanez természetesen az $[f]_B = B^{-1} \cdot [f] \cdot B$ összefüggésből is következik (hiszen $E^{-1} = E$ és az E-vel végzett szorzás $[f]$-et nem változtatja meg). A *(iii)* pedig a $B = E$ esetben azt állítja, hogy $[f]_E = [f]$ i-edik oszlopa $[f(\underline{e}_i)]_E = f(\underline{e}_i)$, amit az 1.8.3. Tételből már tudunk.

1.8.16. Feladat. Az $f : \mathbb{R}^3 \to \mathbb{R}^3$ lineáris transzformáció rendelje minden térvektorhoz annak az $x + 3y + z = 0$ egyenletű S síkra való tükörképét. Adjuk meg f hozzárendelési szabályát.

Megoldás: Először is figyeljük meg, hogy f valóban lineáris transzformáció; ez az 1.8.3. Tételből az 1.8.4. Állításban látott módszerrel könnyen következik.

A feladatot koordinátageometriai eszközökkel is megoldhatnánk, de ehelyett most a bázistranszformáció módszerét használjuk. Az alapgondolat az, hogy egy olyan bázist választunk $\mathbb{R}^3$-ben, amelyben $[f]_B$ nagyon könnyen felírható, majd ebből határozzuk meg $[f]$-et.

A $B = \{\underline{b}_1, \underline{b}_2, \underline{b}_3\}$ bázis például megfelel a célnak, ha $\underline{b}_1$ és $\underline{b}_2$ S-re illeszkedő vektorok, $\underline{b}_3$ pedig merőleges rá. Ekkor ugyanis nyilván $f(\underline{b}_1) = \underline{b}_1$, $f(\underline{b}_2) = \underline{b}_2$ és $f(\underline{b}_3) = -\underline{b}_3$, így az 1.8.15. Tétel *(iii)* állításából $[f]_B$ könnyen kiolvasható lesz.

Legyen ezért például $\underline{b}_1 = (1,-1,2)$, $\underline{b}_2 = (2,-1,1)$ és $\underline{b}_3 = (1,3,1)$. Ekkor $\underline{b}_1$ és $\underline{b}_2$ kielégítik S egyenletét, ezért valóban illeszkednek rá, a $\underline{b}_3$ pedig az S egyenletéből kiolvasható normálvektor, így merőleges S-re. Mivel $\underline{b}_1$, $\underline{b}_2$ és $\underline{b}_3$ nem illeszkednek közös (origón átmenő) síkra, ezért bázist alkotnak $\mathbb{R}^3$-ben. Továbbá

az 1.8.15. Tétel *(iii)* állításából következik, hogy

$$[f]_B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

(Valóban, például $f(\underline{b}_3) = -\underline{b}_3 = 0 \cdot \underline{b}_1 + 0 \cdot \underline{b}_2 + (-1) \cdot \underline{b}_3$, amiből $[f]_B$ harmadik oszlopa adódik az 1.8.15. Tétel *(iii)* állítása szerint.)

Az 1.8.12. Tétel szerint $[f]_B = B^{-1} \cdot [f] \cdot B$, ahol a B oszlopai sorban $\underline{b}_1$, $\underline{b}_2$ és $\underline{b}_3$ (persze oszlopvektorként). B^{-1}-et az 1.6.1. szakaszban látott módszerrel határozhatjuk meg (a számítás részleteit mellőzzük):

$$B^{-1} = \begin{pmatrix} 1 & 2 & 1 \\ -1 & -1 & 3 \\ 2 & 1 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} -\frac{4}{11} & -\frac{1}{11} & \frac{7}{11} \\ \frac{7}{11} & -\frac{1}{11} & -\frac{4}{11} \\ \frac{1}{11} & \frac{3}{11} & \frac{1}{11} \end{pmatrix}.$$

Az $[f]_B = B^{-1} \cdot [f] \cdot B$ egyenletet balról B-vel, jobbról B^{-1}-zel szorozva (valamint felhasználva a mátrixszorzás asszociativitását és az inverz definícióját) kapjuk, hogy $B \cdot [f]_B \cdot B^{-1} = [f]$. Így $[f]$-et megkaphatjuk az alábbi szorzás elvégzésével:

$$[f] = \begin{pmatrix} 1 & 2 & 1 \\ -1 & -1 & 3 \\ 2 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \cdot \begin{pmatrix} -\frac{4}{11} & -\frac{1}{11} & \frac{7}{11} \\ \frac{7}{11} & -\frac{1}{11} & -\frac{4}{11} \\ \frac{1}{11} & \frac{3}{11} & \frac{1}{11} \end{pmatrix} = \begin{pmatrix} \frac{9}{11} & -\frac{6}{11} & -\frac{2}{11} \\ -\frac{6}{11} & -\frac{7}{11} & -\frac{6}{11} \\ -\frac{2}{11} & -\frac{6}{11} & \frac{9}{11} \end{pmatrix}.$$

$[f]$-ből pedig f hozzárendelési szabálya $f(\underline{x}) = [f] \cdot \underline{x}$ miatt már kiolvasható:
$f\big((x_1,x_2,x_3)\big) = (\frac{9}{11}x_1 - \frac{6}{11}x_2 - \frac{2}{11}x_3, -\frac{6}{11}x_1 - \frac{7}{11}x_2 - \frac{6}{11}x_3, -\frac{2}{11}x_1 - \frac{6}{11}x_2 + \frac{9}{11}x_3)$. □

A fenti megoldás elején említettük, hogy ez a feladat bázistranszformáció nélkül is megoldható lett volna. A most látott módszerrel azonban olyan transzformációk mátrixát is meghatározhatjuk, amelyeket egyszerű koordinátageometriai eszközökkel már nagyon körülményes volna. Például egy origón átmenő egyenes, mint tengely körüli elforgatás esetében olyan B-t volna érdemes választani, amelynek a vektorai páronként merőlegesek egymásra, az egyikük az elforgatás tengelyével párhuzamos, a másik kettő pedig egyenlő hosszúságú; ekkor $[f]_B$ az 1.8.15. Tétel *(iii)* állítását használva könnyen kiolvasható (az 1.8.4. Állításhoz hasonlóan), amiből $[f]$ a fenti megoldáshoz hasonlóan megkapható.

1.9. Sajátérték, sajátvektor

A 101. oldalon vizsgált $f : (x_1,x_2) \mapsto (4x_1 - 2x_2, x_1 + x_2)$ lineáris transzformáció esetében szerencsés választás volt a $B = \{(2;1),(1;1)\}$ bázis, mert az $[f]_B$ mátrixnak csak a főátlójában állnak 0-tól különböző elemek, így a $g : [\underline{x}]_B \mapsto [f(\underline{x})]_B$ függvény hozzárendelési szabálya nagyon egyszerű: mindkét koordinátát egy-egy

rögzített konstanssal szorozza. Természetesen a sík bármely más B bázisa alapján is elvégezhettük volna a bázistranszformációt, de az eredmény általában semmivel nem adott volna több információt f-ről, mint az f eleve ismert hozzárendelési szabálya. Most azt a kérdést fogjuk megvizsgálni, hogy mely $f:\mathbb{R}^n\to\mathbb{R}^n$ lineáris transzformációkra és hogyan található olyan B bázis, amely a fenti példához hasonlóan szerencsés – vagyis amelyre $[f]_B$-nek minden főátlón kívüli eleme 0. Az ilyen tulajdonságú négyzetes mátrixokat *diagonális mátrixnak* nevezzük.

A 101. oldalon látott példában a $B=\{\underline{b}_1,\underline{b}_2\}$ bázis azért volt szerencsés választás, mert $f(\underline{b}_1)=3\cdot\underline{b}_1$ és $f(\underline{b}_2)=2\cdot\underline{b}_2$ teljesült. Ezért volt ugyanis igaz, hogy f a $\underline{b}_1$ és $\underline{b}_2$ meghatározta koordinátarendszerben az első tengely mentén 3-szorosukra, a második mentén 2-szeresükre nyújtotta a vektorokat és ezen múlt az is, hogy $[f]_B$ diagonális mátrix lett. Később látni fogjuk, hogy az ebben a példában látott jelenség általánosítható is (lásd az 1.9.5. Állítást).

Az f lineáris transzformáció szempontjából tehát különös fontossággal bírnak azok az $\underline{x}$ vektorok, amelyekre valamely λ skalárral $f(\underline{x})=\lambda\underline{x}$, vagyis $[f]\cdot\underline{x}=\lambda\underline{x}$ teljesül. Az ilyen tulajdonságú $\underline{x}$ vektorok és λ értékek nem csak az $[f]_B$ diagonalizálásában játszanak alapvető szerepet, hanem a lineáris algebra számtalan más alkalmazásában is. Ezeknek ad nevet az alábbi definíció.

1.9.1. Definíció. *Legyen A egy $(n\times n)$-es mátrix.*

(i) A sajátértékének *nevezzük a $\lambda\in\mathbb{R}$ skalárt, ha létezik olyan $\underline{x}\in\mathbb{R}^n$, $\underline{x}\neq\underline{0}$ vektor, amelyre $A\cdot\underline{x}=\lambda\cdot\underline{x}$ teljesül.*

(ii) A sajátvektorának *nevezzük az $\underline{x}\in\mathbb{R}^n$ vektort, ha $\underline{x}\neq\underline{0}$ és létezik olyan $\lambda\in\mathbb{R}$ skalár, amelyre $A\cdot\underline{x}=\lambda\cdot\underline{x}$ teljesül.*

A sajátérték és a sajátvektor definícióját csak a világosabb érthetőség kedvéért választottuk szét, de persze szorosan összetartozó fogalmakról van szó: ha $A\cdot\underline{x}=\lambda\underline{x}$ teljesül valamely $\underline{x}\neq\underline{0}$-ra és λ-ra, akkor λ sajátértéke, $\underline{x}$ sajátvektora A-nak; ilyenkor azt mondjuk, hogy *$\underline{x}$ a λ-hoz tartozó sajátvektora* A-nak.

Ezeket a fogalmakat használva elmondhatjuk, hogy a 101. oldalon vizsgált f lineáris transzformáció $[f]=\begin{pmatrix}4&-2\\1&1\end{pmatrix}$ mátrixának a $\underline{b}_1=\begin{pmatrix}2\\1\end{pmatrix}$ és a $\underline{b}_2=\begin{pmatrix}1\\1\end{pmatrix}$ vektorok sajátvektorai, a 3 és a 2 számok pedig sajátértékei – ez következik az $[f]\cdot\underline{b}_1=3\cdot\underline{b}_1$ és az $[f]\cdot\underline{b}_2=2\cdot\underline{b}_2$ összefüggésekből.

A sajátérték és a sajátvektor definíciójában fontos kikötés az $\underline{x}\neq\underline{0}$ feltétel: ha ezt elhagynánk, akkor minden négyzetes mátrixnak sajátvektora lenne a $\underline{0}$ és sajátértéke volna minden valós szám (hiszen $A\cdot\underline{0}=\lambda\cdot\underline{0}$ minden A-ra és λ-ra fennáll). Nem szabad azonban ezt a feltételt összetéveszteni a sajátérték definíciójával: $\lambda=0$ lehet sajátértéke egy négyzetes mátrixnak.

A fogalmakkal való ismerkedésként meghatározzuk az $A=\begin{pmatrix}1&3\\3&9\end{pmatrix}$ mátrix sajátértékeit és sajátvektorait. Az $A\cdot\underline{x}=\lambda\cdot\underline{x}$ egyenlet rendezés után az alábbi egyenletrendszerre vezet:

$$\begin{pmatrix}1&3\\3&9\end{pmatrix}\cdot\begin{pmatrix}x_1\\x_2\end{pmatrix}=\begin{pmatrix}\lambda x_1\\\lambda x_2\end{pmatrix}\longrightarrow\begin{array}{l}x_1+3x_2=\lambda x_1\\3x_1+9x_2=\lambda x_2\end{array}\longrightarrow\begin{array}{l}(1-\lambda)x_1+3x_2=0\\3x_1+(9-\lambda)x_2=0\end{array}$$

Ha λ értékét rögzítettnek tekintjük, akkor lineáris egyenletrendszert kaptunk az x_1 és x_2 változókkal. A probléma azonban éppen az, hogy λ értékét sem ismerjük – márpedig ha azt is változónak tekintenénk, akkor az egyenletrendszer nem volna többé lineáris. Konkrét λ értékekkel viszont minden további nélkül kísérletezhetünk: ha például $\lambda = 1$, akkor az első egyenletből $x_2 = 0$ adódik, amit a másodikba helyettesítve $x_1 = 0$; ez tehát azt jelenti, hogy $\lambda = 1$ nem sajátértéke A-nak, mert az $A \cdot \underline{x} = 1 \cdot \underline{x}$ egyetlen megoldása $\underline{x} = \underline{0}$. Több sikerrel járunk, ha a $\lambda = 0$ értékkel próbálkozunk: ekkor az $x_1 + 3x_2 = 0$, $3x_1 + 9x_2 = 0$ egyenletrendszert kapjuk, ahol a második egyenlet nyilván elhagyható, mert az 3-szorosa az elsőnek. Vagyis a $\lambda = 0$ esetben van $\underline{0}$-tól különböző megoldás: minden, az $x_1 + 3x_2 = 0$ egyenletnek eleget tevő $\underline{x}$. Ez tehát azt jelenti, hogy $\lambda = 0$ sajátértéke és például $\underline{x} = \begin{pmatrix} 3 \\ -1 \end{pmatrix}$ sajátvektora A-nak (mert erre az $\underline{x}$-re $A \cdot \underline{x} = 0 \cdot \underline{x}$).

A két kísérletből szerzett tapasztalatainkat általánosíthatjuk is: egy konkrét λ akkor lesz sajátérték, ha a fenti lineáris egyenletrendszernek van a $\underline{0}$-tól különböző megoldása. Ez az 1.5.15. Következmény szerint úgy is fogalmazható, hogy az $\begin{pmatrix} 1-\lambda & 3 \\ 3 & 9-\lambda \end{pmatrix}$ mátrix oszlopai lineárisan összefüggőek – ami viszont az 1.5.16. Tétel szerint azzal ekvivalens, hogy a mátrix determinánsa 0. Kiszámítva a determinánst az $(1-\lambda)(9-\lambda) - 3 \cdot 3 = \lambda^2 - 10\lambda$ értéket kapjuk. Következik, hogy a sajátértékek a $\lambda^2 - 10\lambda = 0$ egyenlet megoldásai: $\lambda = 0$ és $\lambda = 10$. Ezek közül a 0-t már megtaláltuk, a $\lambda = 10$ értéket visszahelyettesítve a $-9x_1 + 3x_2 = 0$, $3x_1 - x_2 = 0$ egyenletrendszert kapjuk. Látszik, hogy ennek valóban végtelen sok megoldása van (mert az első egyenlet (-3)-szorosa a másodiknak). Azt kaptuk tehát, hogy a $\lambda = 10$ is sajátértéke A-nak és ehhez tartozó sajátvektor minden olyan $\underline{x}$, amelyre $x_2 = 3x_1$. Több sajátértéke pedig nincs A-nak.

Az alábbi tétel bizonyítása a fenti gondolatmenetet általánosítja.

1.9.2. Tétel. *A négyzetes A mátrixnak a $\lambda \in \mathbb{R}$ skalár akkor és csak akkor sajátértéke, ha* $\det(A - \lambda \cdot E) = 0$ *(ahol E az egységmátrixot jelöli).*

Bizonyítás: λ definíció szerint akkor sajátérték, ha $A \cdot \underline{x} = \lambda \cdot \underline{x}$-nek van egy $\underline{x} \neq \underline{0}$ megoldása. Az egyenlet jobb oldalán álló $\lambda\underline{x}$ helyett $(\lambda \cdot E) \cdot \underline{x}$-et is írhatunk: az 1.5.8. Tétel *(i)* állítása szerint $(\lambda \cdot E) \cdot \underline{x} = \lambda \cdot (E \cdot \underline{x}) = \lambda \cdot \underline{x}$. (De $(\lambda \cdot E) \cdot \underline{x} = \lambda \cdot \underline{x}$ a mátrixszorzás definíciójából közvetlenül is látszik.) Az $A \cdot \underline{x} = (\lambda \cdot E) \cdot \underline{x}$ egyenletet átrendezve, majd $\underline{x}$-et (az 1.5.8. Tétel *(ii)* állítása szerint) kiemelve:

$$A \cdot \underline{x} - (\lambda \cdot E) \cdot \underline{x} = \underline{0}$$
$$\big(A - \lambda \cdot E\big) \cdot \underline{x} = \underline{0}$$

Tehát λ akkor és csak akkor sajátértéke A-nak, ha az $\big(A - \lambda \cdot E\big) \cdot \underline{x} = \underline{0}$ lineáris egyenletrendszernek (amelynek tehát az együtthatómátrixa $A - \lambda \cdot E$ és a jobb oldalakon mindenhol 0 áll) van egy $\underline{x} \neq \underline{0}$ megoldása. Ez az 1.5.15. Következmény szerint azzal ekvivalens, hogy az $A - \lambda \cdot E$ mátrix oszlopai lineárisan összefüggőek – ami viszont az 1.5.16. Tétel szerint valóban azzal, hogy $\det(A - \lambda \cdot E) = 0$. □

A fenti tétel jelentősége abban rejlik, hogy a segítségével (legalábbis elvileg) meghatározhatók egy tetszőleges A mátrix sajátértékei: ehhez ki kell számítani a $\det(A-\lambda\cdot E)$ determináns értékét a λ paraméter függvényében, majd megkeresni azokat a λ-kat, amelyekre ez 0. Azt pedig láttuk, hogy a sajátértékek ismeretében a sajátvektorok meghatározása már csak (sajátértékenként) egy lineáris egyenletrendszer megoldásából áll: a λ sajátértékhez tartozó sajátvektorok az $A\cdot\underline{x}=\lambda\cdot\underline{x}$ egyenletet kielégítő $\underline{x}\neq\underline{0}$ vektorok, vagyis (a fenti bizonyításban látott átrendezés után) az $\left(A-\lambda\cdot E\right)\cdot\underline{x}=\underline{0}$ lineáris egyenletrendszer $\underline{0}$-tól különböző megoldásai.

Ezt a módszert alkalmaztuk már a fenti tétel előtt látott példában is:

$$A=\begin{pmatrix}1&3\\3&9\end{pmatrix}\longrightarrow A-\lambda E=\begin{pmatrix}1-\lambda&3\\3&9-\lambda\end{pmatrix}\longrightarrow\det(A-\lambda E)=\lambda^2-10\lambda$$

A $\lambda^2-10\lambda=0$ egyenlet megoldásával kaptuk A sajátértékeit: $\lambda=0$ és $\lambda=10$.

A $\det(A-\lambda\cdot E)$ kifejezés értéke a fenti példában λ-nak másodfokú polinomja volt, de $(n\times n)$-es A mátrixra általában is igaz, hogy $\det(A-\lambda\cdot E)$ λ-nak n-edfokú polinomja. Valóban, a determináns definíciója szerint $\det(A-\lambda\cdot E)$ meghatározásakor az $A-\lambda\cdot E$ mátrix elemeiből választott n tényezős szorzatokat előjelezünk és adunk össze. Ezeknek a szorzatoknak minden tényezője vagy az A mátrix egy $a_{i,j}$ eleme, vagy egy $(a_{i,i}-\lambda)$ alakú kifejezés. Az n tényezős szorzatokban található λ-t tartalmazó zárójeleket képzeletben felbontva összevonás után valóban $c_0+c_1\cdot\lambda+c_2\cdot\lambda^2+\ldots+c_n\lambda^n$ alakú kifejezést, vagyis λ egy n-edfokú polinomját kell kapjuk. (Mivel a legtöbb λ-t tartalmazó tényező nyilván a főátlónak megfelelő szorzatban található, a zárójeleket felbontva itt egy $\pm\lambda^n$ alakú tag is keletkezik. Ez mutatja egyrészt, hogy a $\det(A-\lambda\cdot E)$ kiszámításakor keletkező polinom valóban n-edfokú, másrészt hogy a λ^n tag előjele n paritásától függően ±1.) Az alábbi definíció ad nevet ennek a polinomnak.

1.9.3. Definíció. *Az $(n\times n)$-es A mátrix* karakterisztikus polinomjának *nevezzük a $\det(A-\lambda\cdot E)$ determináns értékét, ahol λ változó. Ennek a jele: $k_A(\lambda)$.*

Az 1.9.2. Tétel állítása tehát úgy is fogalmazható, hogy A sajátértékei a $k_A(\lambda)$ karakterisztikus polinom gyökei (vagyis a $k_A(\lambda)=0$ egyenlet megoldásai). Az algebra egyik sokat alkalmazott (és könnyen bizonyítható) tétele szerint bármely n-edfokú polinomnak legföljebb n gyöke lehet, amiből következik, hogy minden $(n\times n)$-es mátrixnak legföljebb n sajátértéke van. Ezeknek a meghatározásához azonban egy n-edfokú polinom gyökeit (vagyis egy n-edfokú egyenlet megoldásait) kell megkeresni, ami nagy n-ekre általában csak közelítő módszerekkel lehetséges.

1.9.4. Feladat. Határozzuk meg az alábbi A mátrix minden sajátértékét és sajátvektorát.

$$A=\begin{pmatrix}3&1&1\\0&4&0\\4&2&3\end{pmatrix}$$

Megoldás: A sajátértékek meghatározásával kezdjük: az 1.9.2. Tétel szerint kiszámítjuk az $A-\lambda E$ determinánsát. Ehhez az 1.4.13. Kifejtési tételt alkalmazzuk a második sorra (mert abban két nulla is van):

$$\begin{vmatrix} 3-\lambda & 1 & 1 \\ 0 & 4-\lambda & 0 \\ 4 & 2 & 3-\lambda \end{vmatrix} = (4-\lambda)\cdot \begin{vmatrix} 3-\lambda & 1 \\ 4 & 3-\lambda \end{vmatrix} =$$

$$= (4-\lambda)\big((3-\lambda)^2-4\cdot 1\big) = (4-\lambda)(\lambda^2-6\lambda+5) = (4-\lambda)(\lambda-1)(\lambda-5)$$

Így a sajátértékek $\lambda=4$, $\lambda=1$ és $\lambda=5$, mert $\det(A-\lambda\cdot E)=0$ ezekre az értékekre teljesül. (A karakterisztikus polinomot a fenti kifejezésből a zárójelek felbontásával kaphatjuk: $k_A(\lambda)=(4-\lambda)(\lambda^2-6\lambda+5)=-\lambda^3+10\lambda^2-29\lambda+20$. De ennek a gyökeit – vagyis a sajátértékeket – könnyebb volt a szorzat alakból kiszámítani.)

A sajátvektorokat a három sajátértékhez külön-külön az $(A-\lambda E)\underline{x}=\underline{0}$ lineáris egyenletrendszer megoldásával nyerjük. A $\lambda=4$ esetben a $-x_1+x_2+x_3=0$, $0x_1+0x_2+0x_3=0$, $4x_1+2x_2-x_3=0$ egyenletrendszert kapjuk. A középső egyenlet nyilván elhagyható, a másik kettőre alkalmazhatjuk a Gauss-eliminációt:

$$\left(\begin{array}{ccc|c} -1 & 1 & 1 & 0 \\ 4 & 2 & -1 & 0 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & -1 & -1 & 0 \\ 0 & 6 & 3 & 0 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 0 & -1/2 & 0 \\ 0 & 1 & 1/2 & 0 \end{array}\right)$$

Ebből az $x_3=\alpha\in\mathbb{R}$ szabad paraméter, $x_1=\frac{1}{2}\alpha$, $x_2=-\frac{1}{2}\alpha$ megoldásokat kapjuk. Így a $\lambda=4$ sajátértékhez tartozó sajátvektorok: $\underline{x}=\begin{pmatrix} 1/2\alpha \\ -1/2\alpha \\ \alpha \end{pmatrix}$, ahol $\alpha\neq 0$ tetszőleges. A $\lambda=1$ és $\lambda=5$ sajátértékekhez tartozó sajátvektorok kiszámításának módszere azonos: az előbbi esetben az $\underline{x}=\begin{pmatrix} -1/2\alpha \\ 0 \\ \alpha \end{pmatrix}$, az utóbbiban az $\underline{x}=\begin{pmatrix} 1/2\alpha \\ 0 \\ \alpha \end{pmatrix}$ sajátvektorokat kapjuk, ahol $\alpha\neq 0$ tetszőleges (a számításokat nem részletezzük). □

Érdemes megfigyelni, hogy a fenti feladatban mindhárom sajátérték esetén az ahhoz tartozó sajátvektorok egy konkrét $\underline{v}$ vektor nemnulla skalárszorosai voltak. Ebből annyi általában is igaz, hogy ha $\underline{v}$ az f lineáris transzformáció λ sajátértékhez tartozó sajátvektora és $\alpha\neq 0$, akkor $\alpha\cdot\underline{v}$ is a λ-hoz tartozó sajátvektor. Valóban: ha valamely $\underline{v}\neq\underline{0}$-ra $f(\underline{v})=\lambda\cdot\underline{v}$, akkor $f(\alpha\underline{v})=\alpha\cdot f(\underline{v})=\alpha\cdot(\lambda\underline{v})=\lambda\cdot(\alpha\cdot\underline{v})$.

Végül visszatérünk a szakasz elején felvetett kérdésre: az f lineáris transzformációhoz milyen B bázist választva teljesül, hogy $[f]_B$ diagonális mátrix (vagyis a főátlóján kívül minden elem 0)?

1.9.5. Állítás. *Legyen $f:\mathbb{R}^n\to\mathbb{R}^n$ lineáris transzformáció és $B=\{\underline{b}_1,\underline{b}_2,\ldots,\underline{b}_n\}$ bázis $\mathbb{R}^n$-ben. Ekkor $[f]_B$ akkor és csak akkor diagonális mátrix, ha B minden eleme sajátvektora $[f]$-nek.*

Bizonyítás: $[f]_B$ pontosan akkor diagonális, ha minden $1\leq i\leq n$ esetén az i-edik oszlopa $\lambda_i\cdot\underline{e}_i$-vel egyenlő valamilyen $\lambda_i\in\mathbb{R}$ skalárra (ahol $\underline{e}_i$ a standard bázis

i-edik vektora). Ez az 1.8.15. Tétel *(iii)* állításának értelmében azzal ekvivalens, hogy $[f(\underline{b}_i)]_B = \lambda_i \cdot \underline{e}_i$. Ez viszont (a koordinátavektor 1.2.24. Definíciója szerint) azt jelenti, hogy $f(\underline{b}_i) = 0 \cdot \underline{b}_1 + 0 \cdot \underline{b}_2 + \ldots + \lambda_i \cdot \underline{b}_i + \ldots + 0 \cdot \underline{b}_n$, vagyis $[f] \cdot \underline{b}_i = \lambda_i \cdot \underline{b}_i$. Más szóval: $\underline{b}_i$ (a λ_i sajátértékhez tartozó) sajátvektora $[f]$-nek. □

1.9.6. Feladat. Van-e az alábbi $f : \mathbb{R}^2 \to \mathbb{R}^2$ lineáris transzformációkhoz olyan B bázis, amelyben $[f]_B$ diagonális mátrix? Ha van, akkor adjunk meg egy ilyen B-t és írjuk fel $[f]_B$-t.

a) Minden $\underline{x} \in \mathbb{R}^2$ esetén $f(\underline{x})$ az $\underline{x}$ tükörképe az $x_2 = 2x_1$ egyenletű egyenesre;
b) Minden $\underline{x} \in \mathbb{R}^2$ esetén $f(\underline{x})$ az $\underline{x}$ origó körüli 20°-os elforgatottja;
c) $f : (x_1, x_2) \mapsto (x_1 + 2x_2, -5x_1 + 8x_2)$.

Megoldás: Az 1.9.5. Állítás szerint a válasz mindig attól függ, hogy létezik-e az $[f]$ sajátvektoraiból álló bázis $\mathbb{R}^2$-ben. Mindhárom esetben meg kell tehát keresnünk $[f]$ sajátvektorait.

Az a) feladat esetében ehhez nincs szükség az 1.9.2. Tétel alkalmazására. Ugyanis $\underline{x} \neq \underline{0}$ akkor sajátvektora $[f]$-nek, ha $[f] \cdot \underline{x} = \lambda \cdot \underline{x}$, vagyis ha $f(\underline{x}) = \lambda \cdot \underline{x}$ valamilyen λ-ra; ez geometriailag annyit jelent, hogy $f(\underline{x})$ párhuzamos $\underline{x}$-szel (beleértve ebbe, hogy $f(\underline{x})$ a nullvektor is lehet). A síkban egy tetszőleges origón átmenő t tengelyre való tükrözés esetén $\underline{x}$ tükörképe nyilván akkor lesz párhuzamos $\underline{x}$-szel, ha vagy $\underline{x}$ maga párhuzamos t-vel (és ekkor a tükörkép azonos $\underline{x}$-szel) vagy $\underline{x}$ merőleges t-re (és ekkor a tükörkép ellentettje $\underline{x}$-nek). Ha tehát a t tengely az $x_2 = 2x_1$ egyenletű egyenes, akkor a $\underline{b}_1 = (1;2)$ és a $\underline{b}_2 = (2;-1)$ vektorok sajátvektorai $[f]$-nek: ezekre $f(\underline{b}_1) = 1 \cdot \underline{b}_1$, illetve $f(\underline{b}_2) = (-1) \cdot \underline{b}_2$. Ez a két vektor bázist is alkot a síkban (mert nem párhuzamosak), így $B = \{\underline{b}_1, \underline{b}_2\}$ jó választás f-hez. $[f]_B$-t pedig legkönnyebben az 1.8.15. Tétel *(iii)* állítása alapján írhatjuk fel: ehhez az $f(\underline{b}_1)$ és $f(\underline{b}_2)$ vektorok B szerinti koordinátavektorára van szükségünk. Azonban ezeket épp az előbb határoztuk meg: $f(\underline{b}_1) = 1 \cdot \underline{b}_1 + 0 \cdot \underline{b}_2$ és $f(\underline{b}_2) = 0 \cdot \underline{b}_1 - 1 \cdot \underline{b}_2$, így $[f]_B = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

A b) feladat esetében a gondolatmenet hasonló: olyan $\underline{x} \neq \underline{0}$ síkvektorokat keresünk, amelyekre $f(\underline{x})$ párhuzamos $\underline{x}$-szel. Rögtön látszik, hogy ilyen $\underline{x}$ most nincs: az $\underline{x}$-et az origó körül 20°-kal elforgatva nem kaphatunk $\underline{x}$-szel párhuzamos vektort. Így $[f]$-nek nincs sajátvektora (és így nyilván sajátértéke sem), $[f]_B$ semmilyen B bázis esetén sem lesz diagonális.

A c) feladat megoldásában a geometria már nem segít, így az algebrai utat járjuk: felírjuk $[f]$ mátrixát és megkeressük annak a sajátértékeit és a sajátvektorait. $[f]$ oszlopai az 1.8.3. Tétel szerint $f(\underline{e}_1)$ és $f(\underline{e}_2)$, így $[f] = \begin{pmatrix} 1 & 2 \\ -5 & 8 \end{pmatrix}$. Ebből

$$k_{[f]}(\lambda) = \begin{vmatrix} 1-\lambda & 2 \\ -5 & 8-\lambda \end{vmatrix} = (1-\lambda)(8-\lambda) - (-5) \cdot 2 = \lambda^2 - 9\lambda + 18.$$

Így a sajátértékek a $\lambda^2 - 9\lambda + 18 = 0$ másodfokú egyenlet megoldásai: $\lambda = 3$ és $\lambda = 6$. A sajátvektorok meghatározásához mindkét esetben az $(A - \lambda E)\underline{x} = \underline{0}$ li-

neáris egyenletrendszert kell megoldanunk. A $\lambda = 3$ esetben ez a $-2x_1 + 2x_2 = 0$, $-5x_1 + 5x_2 = 0$ egyenletrendszert, a $\lambda = 6$ esetben a $-5x_1 + 2x_2 = 0$, $-5x_1 + 2x_2 = 0$ egyenletrendszert jelenti. Az első esetben tehát azok az $\underline{x} \neq \underline{0}$ vektorok a sajátvektorok, amelyekre $x_1 = x_2$, a másodikban az $5x_1 = 2x_2$ feltételt kielégítők. Így 3-hoz, illetve 6-hoz tartozó sajátvektor például a $\underline{b}_1 = (1;1)$, illetve a $\underline{b}_2 = (2;5)$. Valóban, gyorsan ellenőrizhető, hogy $f(\underline{b}_1) = (3;3) = 3 \cdot \underline{b}_1$ és $f(\underline{b}_2) = (12;30) = 6 \cdot \underline{b}_2$. Mivel $\underline{b}_1$ és $\underline{b}_2$ nem párhuzamosak, ezért $B = \{\underline{b}_1, \underline{b}_2\}$ bázis, amelyben tehát $[f]_B$ diagonális lesz. Mégpedig az 1.8.15. Tétel *(iii)* állítását alkalmazva: $[f]_B = \begin{pmatrix} 3 & 0 \\ 0 & 6 \end{pmatrix}$. □

1.10. Kitekintés, ajánlott irodalom

A lineáris algebra hatalmas terület, a fentiekben csupán néhány alapvető fogalom és eredmény rövid ismertetésére szorítkozhattunk. A fejezet zárásaként a teljesség igénye nélkül megemlítünk néhány további irányt, általánosítási lehetőséget, amelyek a lineáris algebrát még sokkal hasznosabb eszköztárrá teszik.

A lineáris algebra alapfogalmait eddig a valós számok $\mathbb{R}$ halmazára építettük: $\mathbb{R}^n$-beli oszlopvektorokat, $\mathbb{R}^{k\times n}$-beli mátrixokat vizsgáltunk, stb. Vannak azonban más olyan „számkörök”, amelyekben a lineáris algebra ugyanilyen jól működik – például a racionális vagy a komplex számok $\mathbb{Q}$, illetve $\mathbb{C}$ halmaza. Valójában minden olyan számkör megfelel, ahol a négy alapművelet (a nullával való osztást leszámítva) akadálytalanul elvégezhető. Az ilyen „számköröket” az algebrában *testnek* hívják (ennek a fogalomnak a pontos definíciója számos tankönyvben megtalálható, például az alább ajánlott [1] és [2] könyvekben is). Így $\mathbb{Q}$, $\mathbb{R}$ és $\mathbb{C}$ is test – de vannak ezektől lényegesen különböző, például véges sok elemet tartalmazó testek is. A lineáris algebra pedig tetszőleges testet választva működőképes, a fentiekben megismert szinte minden tétel (és persze az ebben a jegyzetben nem említett rengeteg további eredmény is) érvényben marad. (A kevés kivétel is csak annak tulajdonítható, hogy ha az alaptest véges, akkor érvényét veszíti néhány olyan állítás, amely kihasználja a valós számok halmazának végtelenségét. Például ha az alaptest véges, akkor minden lineáris egyenletrendszernek véges sok megoldása van.)

Van azonban a lineáris algebra módszereinek egy, még a fentieknél is messzebb menő általánosítási lehetősége. A legtöbb, a témával foglalkozó tankönyv egy, az absztrakt algebra világába tartozó fogalmat, a *vektorteret* választja a lineáris algebra alapfogalmául. Ennek az alapgondolata az, hogy nem szükséges előre rögzítenünk, hogy a „vektor” kifejezés alatt mit is értsünk: lehet az bármi, csak teljesüljön a vektorokra néhány alapkövetelmény, amelyekből kiindulva a lineáris algebra (fentebb megismertekkel analóg) fogalmai és tételei felépíthetők. Pontosabban: válasszunk egy tetszőleges nemüres alaphalmazt, ennek az elemeit tekintjük majd vektoroknak; továbbá mondjuk meg előre, hogy két tetszőleges vektort összeadva melyik vektort (vagyis az alaphalmaz melyik elemét) kapjuk, illetve hogy egy tetszőleges vektort egy skalárral megszorozva melyik vektort kapjuk. (Itt a „skalár” szó jelenthet valós számot, de a fentiek szerint akár más testet is választhatunk.) Ha az így definiált műveletek betartanak néhány, egészen pontosan lefektetett alapkövetelményt – más

néven: *axiómát* –, akkor a lineáris algebra fogalmai és tételei a fentebb látottakhoz hasonló módon bevezethetők, illetve bizonyíthatók lesznek. (Ezeknek az axiómáknak egy jó részét mi az 1.2.2. Tételben soroltuk fel – az ott leírt tulajdonságok mellett még meg kell követelni az $1 \cdot \underline{v} = \underline{v}$ azonosság teljesülését, illetve külön axióma rendelkezik a nullvektor és az ellentett vektor létezéséről.) Ennek az absztrakt megközelítésnek komoly előnye, hogy ezáltal a lineáris algebra alkalmazási körét a lehető legtágabbra vonjuk – egy konkrét szituációban csak a vektortér axiómáit kell ellenőrizni ahhoz, hogy a lineáris algebra teljes eszköztára rendelkezésre álljon.

Mindezekkel az általánosítási lehetőségekkel ebben a jegyzetben nem foglalkozunk, az érdeklődő olvasóknak az alábbi tankönyveket ajánljuk:

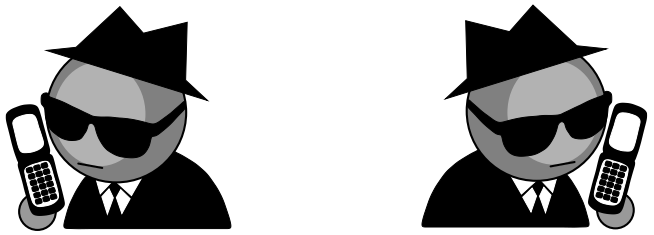
[1] Freud Róbert: *Lineáris algebra*, ELTE Eötvös Kiadó, Budapest, 2006.

[2] Wettl Ferenc: *Lineáris algebra: azoknak, akik érteni is szeretnék*, TypoTeX Kiadó, Budapest, 2011.
http://oszkdk.oszk.hu/storage/00/00/58/66/dd/1/BME_TTK_Wettl_Linearis_algebra_OSZK-nak.pdf

2. fejezet

Számelmélet

Két titkosügynöknek telefonon kell megbeszélni egy – a szakma természetéből fakadóan – bizalmas témát. Biztosak benne, hogy az ellenség lehallgatja a telefonbeszélgetést, de nincs lehetőségük más csatornán kommunikálni. Korábban soha nem találkoztak egymással, nincs előre egyeztetett titkos kódjuk. A telefonbeszélgetést kezdhetik ugyan egy kód egyeztetésével, de persze az ellenség ezt is le fogja hallgatni. Hogyan tudnak akkor mégis biztonságosan kommunikálni – tehát úgy, hogy az ellenség a lehallgatott beszélgetést biztosan ne tudja megfejteni?



A feladat első hallásra valószínűleg megoldhatatlannak tűnik – és talán ugyanennyire életszerűtlennek is. Pedig mindkettő tévedés: a probléma hatékonyan megoldható és számtalan gyakorlati alkalmazásban felmerül. Aki például vásárolt vagy bankolt már az interneten, az használta a *https* protokollt is, ami a világhálón való adattovábbításra széles körben használt *http* protokoll titkosított változata. Könnyű belegondolni, hogy például egy internetes fizetőfelület használatakor a szolgáltató és az ügyfél közötti biztonságos kommunikáció megvalósítása a két titkosügynökével analóg problémát jelent – hiszen az interneten zajló adatforgalom még könnyebben is „lehallgatható”, mint a telefonbeszélgetés.

Mi zajlik tehát, amikor a böngésző átvált biztonságos kapcsolatra, hogyan titkosít a *https* protokoll? A választ ennek a fejezetnek a végén megtudjuk, ehhez előbb a számelmélet körébe tartozó néhány alapvető ismeretet kell megszereznünk.

A számelmélet a matematika egyik legősibb ága, az egész számok körében felmerülő kérdéseket tanulmányozza. Közismert, hogy a mai értelemben vett (tehát

axiomatikus felépítésen, definíciókon, tételeken és bizonyításokon alapuló) matematika az ókori görögökkel kezdődött. Ők a matematikát két nagy területre osztották: a geometriára és az *aritmetikára*, vagyis a számelméletre. A görögök számfogalma a (pozitív) egészekre korlátozódott, a törteket nem tekintették valódi számnak (helyette az egészek arányáról beszéltek), az irracionális számokat pedig még kevésbé (bár tudták például, hogy a négyzet oldala és átlója „nem összemérhető" – amit ma úgy fejezünk ki, hogy $\sqrt{2}$ irracionális). A görögökig visszanyúló hagyomány okán értjük számelmélet alatt ma is csupán az egész számok vizsgálatát.

Számos más örökséget is őriz a számelmélet a görögöktől: több alapfogalom nevét és definícióját (mint például a prímszámét), rengeteg alapvető tételt és azok bizonyítását (mint például hogy végtelen sok prím létezik) – és sok olyan nyitott kérdést is, amelyeket az utóbbi 2500 év matematikusainak sem sikerült megválaszolni (mint például hogy létezik-e páratlan tökéletes szám – vagyis olyan pozitív, páratlan szám, ami egyenlő a saját magánál kisebb, pozitív osztóinak összegével).

A számelméletet egészen az 1970-es évek végéig a matematika „belügyének" tekintették – senki nem gondolta, hogy a prímszámokkal kapcsolatos, mégoly izgalmas és patinás eredmények az elméleti matematika rajongóin kívül bárkit érdekelhetnének. 1977-ben fedezte fel *Ronald Rivest*, *Adi Shamir* és *Len Adleman* a róluk elnevezett *RSA algoritmust* – azt a kriptográfiai eljárást, amelyet többek között a fent említett *https* protokoll is használ. Ez a felfedezés állította reflektorfénybe a számelmélet évezredes kutatásának eredményeit – és emiatt váltak mára ennek a területnek az alapjai az informatikus alapképzettség elengedhetetlen részévé is.

2.1. Alapismeretek

Ebben a szakaszban a középiskolai tanulmányokból már nagyrészt ismert alapokat foglaljuk össze és egészítjük ki. Egy később nagyon hasznosnak bizonyuló részletben viszont el fogunk térni a középiskolában megszokottaktól: a számelméletet az egész számok $\mathbb{Z}$ halmazára dolgozzuk ki – beleértve tehát a negatív egészeket is. A teljes fejezetre vonatkozóan elfogadjuk azt a megállapodást, hogy minden változó egész számot jelöl (akkor is, ha ezt külön nem említjük). Először a számelmélet néhány alapfogalmát vezetjük be.

2.1.1. Definíció. *Azt mondjuk, hogy az $a \in \mathbb{Z}$ egész* osztója *a $b \in \mathbb{Z}$ egésznek, ha létezik olyan $c \in \mathbb{Z}$, amelyre $a \cdot c = b$. Ugyanezt fejezzük ki, ha b-t az a* többszörösének *mondjuk. Ennek a jele: $a \mid b$; ha pedig a nem osztója b-nek, azt így jelöljük: $a \nmid b$. Az a* valódi osztója *b-nek, ha $a \mid b$ fennáll és $1 < |a| < |b|$.*

Így például igazak a $13|91$, a $-7|63$, a $2|0$ és a $-8 \nmid -36$ állítások. (Talán meglepő, de $0|0$ is igaz – hiszen $0 \cdot c = 0$ bármely c-re teljesül. Ennek ellenére, a 0-val való osztást természetesen a továbbiakban sem definiáljuk.) A 10 valódi osztói definíció szerint 2, 5, -2 és -5, a nem valódi osztói 1, 10, -1 és -10.

2.1.2. Definíció. *A $p \in \mathbb{Z}$ egészt* prímszámnak *(röviden:* prímnek*) nevezzük, ha $|p| > 1$ és p-nek nincs valódi osztója. Más szóval: $p = a \cdot b$ csak akkor lehetséges, ha $a = \pm 1$ vagy $b = \pm 1$. Ha $|p| > 1$ és p nem prím, akkor* összetett számnak *mondjuk.*

Így például a 3, 103, -7, -89 számok prímek. Általában: a negatív prímek nyilván épp a pozitívak ellentettjei. A prímszám definíciójában a $|p| > 1$ feltétel azért szükséges, mert a -1, 0 és 1 számoknak sincs valódi osztója, de ezeket nem definiáljuk prímnek (mert ezzel elrontanánk a számelmélet 2.1.3. alaptételének az állítását). A -1, 0 és 1 számok se nem prímek, se nem összetettek.

Megjegyezzük, hogy számos tankönyv a fenti definícióban bevezetett fogalmat *felbonthatatlan számnak* nevezi és prímszám alatt mást ért: olyan p egészt, amelyre $p|a \cdot b$-ből $p|a$ vagy $p|b$ következik. Az ezt a terminológiát használó tankönyvek ezután tételként mondják ki, hogy a felbonthatatlan számok azonosak a prímekkel. (Ez az állítás egyébként messze nem magától értetődő, egyenértékű a számelmélet 2.1.3. alaptételével.) Mivel a két fogalom között végül is nincs tartalmi különbség, ezért mi megmaradunk a középiskolában megszokott (egyébként az ókori görögökével is azonos) szóhasználatnál.

Az alábbi tétel nem véletlenül kapta a nevét: ez a számelmélet egész felépítését megalapozó struktúratétel, amely egyben a prímek meghatározó szerepét is mutatja az egész számok vizsgálatában.

2.1.3. Tétel. (A számelmélet alaptétele)
Minden 1-től, 0-tól és (-1)-től különböző egész szám felbontható prímek szorzatára és ez a felbontás a tényezők sorrendjétől és előjelétől eltekintve egyértelmű.

Például a 100 esetében a tétel által garantált felbontás lehet $2 \cdot 2 \cdot 5 \cdot 5$, de lehet akár $(-5) \cdot 2 \cdot (-2) \cdot 5$ is. Ez a példa mutatja, miért kellett a tétel egyértelműségre vonatkozó részében a sorrendtől és az előjelektől eltekinteni: a 100-nak ezt a két felbontását azonosnak szeretnénk tekinteni. A tétel állításában az egytényezős szorzatokat is megengedjük, így a tétel prímekre is igaz (a p prím a saját maga által alkotott egytényezős szorzattal egyenlő). A 0-t és ± 1-et viszont valóban nem lehet prímek szorzatára bontani, ezért kellett ezeket kivételként felsorolni.

A felbonthatóság bizonyítása a 2.1.3. Tételben: Megadunk egy egyszerű eljárást, amely tetszőleges $n \in \mathbb{Z}$, $|n| > 1$ egészt prímtényezők szorzatára bont. Az eljárás végig fenntartja az n egy (± 1-től különböző) egészek szorzatára való bontását, kezdetben ez lehet az n egytényezős szorzat. Ha egy ponton az $n = a_1 \cdot a_2 \cdot \ldots \cdot a_k$ szorzatnál tart és az a_i tényezők mind prímek, akkor az eljárás megáll. Ha a tényezők között van összetett és például a_i ilyen, akkor a_i-nek van valódi osztója, így felírható $a_i = b \cdot c$ alakban, ahol $|b|, |c| > 1$ egészek. Ekkor az eljárás az n felbontásában a_i-t helyettesíti $b \cdot c$-vel, majd ugyanígy folytatódik tovább. Mivel az n felbontásában a tényezők száma minden lépésben növekszik 1-gyel és minden tényező abszolút értéke legalább 2, ezért az eljárás véges sok lépésben (egy legföljebb $\log_2 |n|$ tényezős szorzattal) megáll és szolgáltatja n egy prímtényezőkre bontását. □

Később visszatérünk arra a kérdésre, hogy a fenti bizonyításban leírt eljárás mennyire hatékony a gyakorlatban. Azt mindenképp jól mutatja, hogy a számelmélet alaptételéből a felbonthatóság bizonyítása egyszerű és rövid. Fontos azonban rámutatni arra, hogy az alaptétel valódi erejét nem önmagában a felbonthatóság ténye hordozza, hanem annak az egyértelműsége. Ha egy számnak több különböző (nem csak sorrendben és előjelekben eltérő) prímtényezős felbontása lehetne, akkor a tétel elveszítené a valódi erejét, keveset mondana az egész számok struktúrájáról; például az n egy lehetséges prímtényezős felbontásából nem lehetne kiolvasni n összes osztóját.

Az egyértelműség fontosságát illusztrálandó érdemes egy példát végiggondolni: mi történne, ha „betiltanánk" a páratlan számokat, csak a párosakat tartanánk meg? Első látásra ez talán nem tűnne akkora veszteségnek, legalábbis nem nagyobbnak egy „mértékegységváltásnál" – hiszen a páros számok szorzata és összege is páros. Pedig a páratlan számok elvesztése komoly bajt okozna, ugyanis a prímtényezőkre bontás egyértelműsége többé nem volna igaz: például a 36-nak a $36 = 2 \cdot 18$ és a $36 = 6 \cdot 6$ is olyan felbontása volna, amelyben egyik tényező sem bontható tovább. (A teljes igazsághoz tartozik, hogy ez a példa kicsit sántít. Ugyanis a páros számok világában a prímszám 2.1.2 Definícója eleve értelmezhetetlenné válik: mivel az 1 páratlan, ezért a 4-gyel osztva 2 maradékot adó számokat semmilyen módon nem lehet szorzattá alakítani, még nem valódi osztójuk sincs. Sokkal több leleményességgel azonban mutathatók olyan „számkörök" is, amelyek az egészek műveleti tulajdonságait minden tekintetben imitálják, a prímfelbontás egyértelműsége mégis sérül; ilyen példa az $a + b \cdot \sqrt{5} \cdot i$ alakú komplex számok halmaza, ahol $a, b \in \mathbb{Z}$.)

A prímfelbontás egyértelműsége tehát messze nem magától értetődő, bizonyításra szorul. Ennek az egyetlen valódi nehézsége az, hogy az állítás igazságát általában annyira természetesnek tekintjük, hogy könnyű szem elől téveszteni a célt és magát a bizonyítandó állítást (vagy annak egy következményét) is felhasználni a bizonyításban.

Az egyértelműség bizonyítása a 2.1.3. Tételben: Elég lesz megmutatni, hogy az állítás igaz a pozitív egészek körében – vagyis hogy minden $n > 1$ egésznek sorrendtől eltekintve egyértelmű a felbontása pozitív prímek szorzatára. Valóban, ha $n < -1$ és n-nek volna két lényegesen (vagyis nem csak sorrendben és előjelekben) különböző prímfelbontása, akkor mindkét felbontásban az összes prímtényező abszolút értékét véve $(-n)$ két lényegesen különböző felbontását kapnánk pozitív prímek szorzatára.

A pozitív egészekre vonatkozó állítást pedig n-re vonatkozó teljes indukcióval látjuk be. Ha $n = 2$, akkor az állítás magától értetődő (mert 2 prím). Legyen most $n > 2$ és tegyük fel, hogy az állítás már minden $1 < n' < n$ pozitív egészre igaz. (A teljes indukció tehát most nem egyesével lépked, hanem a $2, 3, \ldots, n-1$ számokra vonatkozó állítás igazságából látjuk be az n-re vonatkozót.)

Tegyük fel indirekt, hogy n-nek két lényegesen különböző (pozitív prímekre való) felbontása létezik:

$$n = p_1 \cdot p_2 \cdot \ldots \cdot p_r = q_1 \cdot q_2 \cdot \ldots \cdot q_s.$$

Nyilván $r, s \geq 2$ (mert n maga nem lehet prím). Ha $p_i = q_j$ teljesülne valamely $1 \leq i \leq r$, $1 \leq j \leq s$ párra, akkor mindkét oldalt leosztva ezzel a közös prímmel azt

kapnánk, hogy a hányadosnak is két különböző prímfelbontása volna; ez ellentmond az indukciós feltevésnek, így lehetetlen. Azt is feltehetjük az általánosság megszorítása nélkül, hogy a két felbontásban szereplő összes prím között nincs p_1-nél kisebb (de persze a p_i-k között lehet még p_1-gyel egyenlő).

Legyen most $n' = (q_1 - p_1) \cdot q_2 \cdot q_3 \cdot \ldots \cdot q_s$. Ekkor ($p_1 < q_1$ miatt) $n' > 1$. Megmutatjuk, hogy n'-nek létezik egy olyan prímtényezős felbontása, amelyben p_1 nem szerepel és egy olyan is, amelyben igen; mivel ez két lényegesen különböző prímfelbontást jelent és $n' < n$, ezért ez ellentmond az indukciós feltevésnek és így bizonyítja a tételt.

Ehhez először is vegyük észre, hogy az n'-nek megkaphatjuk egy prímfelbontását, ha $(q_1 - p_1)$-et prímtényezőkre bontjuk (már beláttuk, hogy ez lehetséges) és ezt az n' fenti definíciójába helyettesítjük. Mivel $p_1 \nmid q_1$ (hiszen q_1 prím), ezért $p_1 \nmid (q_1 - p_1)$ is igaz. Ezért az n' így kapott felbontásában a p_1 biztosan nem fog szerepelni (ugyanis azt már tudjuk, hogy $p_1 \neq q_2, q_3, \ldots, q_s$).

Másrészt viszont

$$n' = q_1 \cdot q_2 \cdot \ldots \cdot q_s - p_1 \cdot q_2 \cdot \ldots \cdot q_s = n - p_1 \cdot q_2 \cdot \ldots \cdot q_s = \\ = p_1 \cdot p_2 \cdot \ldots \cdot p_r - p_1 \cdot q_2 \cdot \ldots \cdot q_s = p_1(p_2 \cdot \ldots \cdot p_r - q_2 \cdot \ldots \cdot q_s).$$

Ha most itt $p_2 \cdot \ldots \cdot p_r - q_2 \cdot \ldots \cdot q_s$ helyére annak egy tetszőleges prímfelbontását helyettesítjük, akkor n'-nek valóban egy olyan prímfelbontását kapjuk, amelyben p_1 szerepel. Ezzel tehát a tételt beláttuk. □

A számelmélet alaptételét a szakasz elején írtakkal összhangban egész számokra mondtuk ki és láttuk be, de persze ebből következik a pozitív egészekre vonatkozó változat is: minden $n > 1$ egész a sorrendtől eltekintve egyértelműen bomlik pozitív prímek szorzatára. Ez lehetőséget ad a pozitív egészek *kanonikus alakjának* a bevezetésére: ez nem mást jelent, mint hogy az $n > 1$ egész (egyértelmű) pozitív prímekre bontásában a prímek ismétlődő példányait egy hatványban fogjuk össze, így az $n = p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdot \ldots \cdot p_k^{\alpha_k}$ alakot kapjuk (ahol tehát $p_1, p_2, \ldots, p_k > 1$ prímek és $\alpha_1, \alpha_2, \ldots, \alpha_k \geq 1$ egészek). Például a 600 kanonikus alakja: $600 = 2^3 \cdot 3^1 \cdot 5^2$. A kanonikus alak birtokában számos, elsősorban osztókra és többszörösökre vonatkozó kérdés könnyen megválaszolhatóvá válik – hiszen két egész szorzásakor a prímfelbontásaik nyilván egyesítődnek. Így az alábbi állítás közvetlen következménye a számelmélet alaptételének.

2.1.4. Állítás. *Legyen az $n > 1$ egész kanonikus alakja $n = p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdot \ldots \cdot p_k^{\alpha_k}$. Ekkor egy pozitív egész m-re $m|n$ akkor és csak akkor igaz, ha $m = p_1^{\beta_1} \cdot p_2^{\beta_2} \cdot \ldots \cdot p_k^{\beta_k}$ valamely $0 \leq \beta_1 \leq \alpha_1$, $0 \leq \beta_2 \leq \alpha_2$, …, $0 \leq \beta_k \leq \alpha_k$ kitevőkre.*

Ennek az állításnak a következménye (és középiskolából is ismert) az a módszer, amellyel az $n, m > 1$ egészek (n, m) *legnagyobb közös osztóját* és $[n, m]$ *legkisebb közös többszörösét* megkaphatjuk. (Ezt a két fogalmat a nevük egyben definiálja is – az utóbbi esetben azzal a magától értetődő kiegészítéssel, hogy csak pozitív többszörösökről van szó.) Ehhez n-et és m-et nem pontosan a fenti kanonikus alakjukban érdemes felírni: mindkettejük prímfelbontásában feltüntetjük

az összes olyan prímet, amellyel bármelyikük osztható – cserébe viszont a kitevőben a 0-t is megengedjük. Például $n = 600$ és $m = 84$ esetén ezek kanonikus alakja $600 = 2^3 \cdot 3^1 \cdot 5^2$ és $84 = 2^2 \cdot 3^1 \cdot 7^1$, de (n,m) és $[n,m]$ meghatározása előtt ezeket így írjuk: $600 = 2^3 \cdot 3^1 \cdot 5^2 \cdot 7^0$, $84 = 2^2 \cdot 3^1 \cdot 5^0 \cdot 7^1$; ezek után $(600,84) = 2^2 \cdot 3^1 \cdot 5^0 \cdot 7^0 = 12$ és $[600,84] = 2^3 \cdot 3^1 \cdot 5^2 \cdot 7^1 = 4200$ közvetlenül adódik (n,m) és $[n,m]$ definíciójából, illetve a 2.1.4. Állításból. Ugyanezt általában a következő tétel mondja ki.

2.1.5. Tétel. *Legyenek $p_1, p_2, \ldots, p_k > 1$ prímek és $n = p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdot \ldots \cdot p_k^{\alpha_k}$, valamint $m = p_1^{\beta_1} \cdot p_2^{\beta_2} \cdot \ldots \cdot p_k^{\beta_k}$ egészek, ahol $\alpha_1, \alpha_2, \ldots, \alpha_k, \beta_1, \beta_2, \ldots, \beta_k \geq 0$. Ekkor*

$$(n,m) = p_1^{\min\{\alpha_1,\beta_1\}} \cdot p_2^{\min\{\alpha_2,\beta_2\}} \cdot \ldots \cdot p_k^{\min\{\alpha_k,\beta_k\}} \text{ és}$$
$$[n,m] = p_1^{\max\{\alpha_1,\beta_1\}} \cdot p_2^{\max\{\alpha_2,\beta_2\}} \cdot \ldots \cdot p_k^{\max\{\alpha_k,\beta_k\}}.$$

2.1.6. Feladat. Tetszőleges n és m pozitív egészek esetén jelölje $\langle n,m \rangle$ a legkisebb olyan x pozitív egészt, amelyre $n|m \cdot x$ és $m|n \cdot x$ teljesül. Adjunk $\langle n,m \rangle$ meghatározására a 2.1.5. Tételbelihez hasonló képletet és számítsuk ki $\langle 10080, 99000 \rangle$ értékét!

Megoldás: A 2.1.5. Tételben látotthoz hasonlóan írjuk fel n-et és m-et a közös $p_1, p_2, \ldots, p_k > 1$ prímek hatványainak szorzataként (0 kitevőket is megengedve): $n = p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdot \ldots \cdot p_k^{\alpha_k}$ és $m = p_1^{\beta_1} \cdot p_2^{\beta_2} \cdot \ldots \cdot p_k^{\beta_k}$. Ekkor a keresett x is felírható $x = p_1^{\gamma_1} \cdot p_2^{\gamma_2} \cdot \ldots \cdot p_k^{\gamma_k}$ alakban; valóban, ha x prímtényezős felbontásában szerepelne a p_i-től különböző prím pozitív hatványa, akkor ezt elhagyva a kapott x'-re $n|m \cdot x'$ és $m|n \cdot x'$ továbbra is fennállna, ami $x' < x$ miatt ellentmondás volna.

Mivel $m \cdot x = p_1^{\beta_1+\gamma_1} \cdot p_2^{\beta_2+\gamma_2} \cdot \ldots \cdot p_k^{\beta_k+\gamma_k}$, ezért $n|m \cdot x$ (a 2.1.4. Állítás szerint) ekvivalens az $\alpha_1 \leq \beta_1 + \gamma_1$, $\alpha_2 \leq \beta_2 + \gamma_2$, $\ldots$, $\alpha_k \leq \beta_k + \gamma_k$ feltételek teljesülésével. Átrendezve: $\gamma_i \geq \alpha_i - \beta_i$ minden $1 \leq i \leq k$-ra. Hasonlóan, az $m|n \cdot x$ feltétel ekvivalens azzal, hogy $\gamma_i \geq \beta_i - \alpha_i$ fennáll minden $1 \leq i \leq k$-ra. Összefoglalva: a $\gamma_i \geq |\alpha_i - \beta_i|$ feltételek adódtak, így a legkisebb ilyen x-et akkor kapjuk, ha $\gamma_i = |\alpha_i - \beta_i|$ minden $1 \leq i \leq k$-ra. Tehát:

$$\langle n,m \rangle = p_1^{|\alpha_1-\beta_1|} \cdot p_2^{|\alpha_2-\beta_2|} \cdot \ldots \cdot p_k^{|\alpha_k-\beta_k|}.$$

Ezt $\langle 10080, 99000 \rangle$ kiszámítására alkalmazva: $10080 = 2^5 \cdot 3^2 \cdot 5^1 \cdot 7^1 \cdot 11^0$ és $99000 = 2^3 \cdot 3^2 \cdot 5^3 \cdot 7^0 \cdot 11^1$, így $\langle 10080, 99000 \rangle = 2^2 \cdot 3^0 \cdot 5^2 \cdot 7^1 \cdot 11^1 = 7700$.

Érdemes megfigyelni, hogy a 2.1.5. Tételből és az $\langle n,m \rangle$-re adott fenti képletünkből következik az $\langle n,m \rangle = \dfrac{[n,m]}{(n,m)}$ összefüggés is. □

2.2. Prímszámok

A prímekkel kapcsolatos kérdések az ókori görögök óta izgatják a matematikusokat, köztük máig is sok a megoldatlan. Az első alapvető eredmény ebben a témában azt állítja, hogy a prímek száma végtelen. Ez nem magától értetődő, hiszen akár egyetlen prímből is végtelen sok egész szám áll össze – miért is ne lehetne, hogy a számítógépekkel már nem belátható magasságokban egy ponton megszakad a prímek sorozata? Az alábbi tétel bizonyítása már Euklidesznek az i. e. 300 körül megjelent *Elemek* című művében is szerepel.

2.2.1. Tétel. *A prímek száma végtelen.*

Bizonyítás: Indirekt tegyük fel, hogy a prímek száma véges: $p_1, p_2, \ldots, p_k$ az összes (pozitív) prím. Legyen $N = p_1 \cdot p_2 \cdot \ldots \cdot p_k + 1$. Ekkor N (a 2.1.3. Tétel állítása szerint – itt az egyértelműséget nem is kell kihasználni) prímtényezők szorzatára bomlik (vagy maga is prím). Azonban N nem osztható a $p_1, p_2, \ldots, p_k$ prímek egyikével sem (ugyanis mindegyikkel osztva 1 maradékot ad), így N minden prímtényezője hiányzik a $p_1, p_2, \ldots, p_k$ felsorolásból. Ez az ellentmondás bizonyítja a tételt. □

A prímek (a fentiek szerint végtelen) sorozata bizonyos értelemben nagyon kiszámíthatatlan, más értelemben viszont meglepően szabályos. Alább megemlítünk néhány ide tartozó eredményt, de bebizonyítani csak a (messze) legkönnyebbet fogjuk – ebben a témában ugyanis az egyszerűnek hangzó kérdések is könnyen bizonyulhatnak rendkívül nehéznek. Ilyen például a híres ikerprímsejtés: eszerint végtelen sok, egymástól 2 távolságra lévő prímpár létezik (mint például az 5 és a 7, a 101 és a 103 vagy a 4127 és a 4129) – ez tehát máig is nyitott probléma. (Az ikerprímsejtés vizsgálatában áttörést ért el 2013-ban *Yitang Zhang*: belátta, hogy végtelen sok, egymástól legföljebb 70 millió távolságra lévő prímpár van. Bár ez még nagyon messze van az ikerprímsejtés által állított 2 távolságtól, korábban 70 millió helyett semmilyen fix távolsággal sem volt ismert hasonló eredmény. Azóta rengeteg matematikus összefogása révén a 70 milliós távolságot sikerült 246-ra javítani.)

Az ikerprímsejtés egyfajta ellentétét fogalmazza meg az alábbi tétel.

2.2.2. Tétel. *Minden $N > 1$ egészhez találhatók olyan $p < q$ prímek, hogy p és q között nincs további prím és $q - p > N$.*

Bizonyítás: Legyen $N > 1$ rögzített. Azt kell belátnunk, hogy létezik N darab szomszédos összetett szám; valóban, ekkor az ezeknél kisebb prímek közül a legnagyobbat p-nek és az ezeknél nagyobb prímek közül a legkisebbet q-nak választva a tétel állítása teljesül.

Legyen $a_i = (N+1)! + i$ minden $i = 2, 3, \ldots, (N+1)$ esetén. Ekkor az $a_2, a_3, \ldots, a_{N+1}$ egészek szomszédosak és a darabszámuk N. Továbbá mindegyikük összetett, mert minden $2 \leq i \leq N+1$ esetén a_i-nek valódi osztója i; valóban, $(N+1)!$ nyilván osztható i-vel (hiszen $(N+1)!$ az $1, 2, \ldots, N+1$ egészek szorzata és ezek között i is szerepel), így ehhez i-t adva ismét i-vel oszható számot kapunk. Az a_i egészek tehát bizonyítják a tétel állítását. □

A fentiekből tehát az derül ki, hogy a prímek sorozata a szomszédos tagok különbségét vizsgálva szeszélyesen viselkedik: nagyon kicsi (az ikerprímsejtés szerint akár 2) és tetszőlegesen nagy távolságok is végtelen sokszor előfordulnak. Ehhez képest meglepő, hogy ha „statisztikailag” vizsgáljuk a prímeket, akkor szabályosságot mutatnak. Ennek a pontos megfogalmazásához vezessük be az n-nél nem nagyobb (pozitív) prímek számára a $\pi(n)$ jelölést (így például $\pi(5)=3$ és $\pi(10)=4$). Az alábbi, 1896-ban *Jacques Hadamard* és *Charles Jean de la Vallée-Poussin* által, mély komplex függvénytani eszközökkel bizonyított tétel alapvető fontosságú a prímszámok elméletében – ezt tükrözi az a név is, ami már a huszadik század matematikusai révén ragadt rá.

2.2.3. Tétel. (A Nagy Prímszámtétel)
$\pi(n) \approx \dfrac{n}{\ln n}$, *vagyis* $\lim\limits_{n\to\infty} \dfrac{\pi(n)}{\frac{n}{\ln n}} = 1$.

Az itt használt „$\approx$” jel a számsorozatok *aszimptotikus egyenlőségét* jelöli; ez azt jelenti, amit a tétel kimondása is sugall: az a_n sorozat akkor aszimptotikusan egyenlő b_n-nel (jelölésben: $a_n \approx b_n$), ha a két sorozat hányadosa 1-hez konvergál. A Nagy Prímszámtétel állítása tehát úgy értelmezhető, hogy $\pi(n)$ értékére jó becslés $\frac{n}{\ln n}$ – abban az értelemben, hogy a becslés relatív hibája n növekedtével 0-hoz konvergál.

Megemlítünk még egy tételt, amely a prímek egy egészen másfajta értelemben vett szabályosságát mondja ki. Mivel a prímek a 2 kivételével páratlanok és a páratlan számok tovább oszthatók 4-gyel osztva 1, illetve 3 maradékot adó számokra, felvetődik a kérdés: vajon a $4k+1$, illetve a $4k+3$ alakú prímek száma is végtelen, vagy az egyik típusból csak véges sok van? Hasonló kérdés a 4 helyett más szám szerinti maradékokkal is felvethető. Például a 12-es maradékokat vizsgálva világos, hogy csak a $12k+1$, a $12k+5$, a $12k+7$ és a $12k+11$ alakú számok esetében van esély arra, hogy végtelen sok prímet tartalmazzanak, vagyis a 12-höz relatív prím maradékok jönnek szóba. Valóban, ha $(12,b)=d>1$, akkor a $12k+b$ alakú számok mindegyike osztható d-vel, így legföljebb 2 prím lehet köztük (mégpedig $\pm d$, ha d prím). Így az alábbi tétel a legáltalánosabb esetben válaszolja meg ezt a kérdést.

2.2.4. Tétel. (Johann Dirichlet, 1837)
Ha $(a,b)=1$, *akkor végtelen sok* $a\cdot k+b$ *alakú prím van.*

Végül felsorolunk néhány híres sejtést a prímekkel kapcsolatban, amelyek – a fentebb már említett ikerprímsejtéshez hasonlóan – egy általános iskolás szintjén is megérthetők, mégis régóta ellenállnak minden bizonyítási kísérletnek.

1. Minden $n>2$ páros szám felírható két prím összegeként. *(Goldbach-sejtés)*
2. n^2 és $(n+1)^2$ között minden n-re van prím.
3. Végtelen sok n^2+1 alakú prím van.
4. Végtelen sok olyan p prím létezik, amelyre $2p+1$ szintén prím.

2.3. Kongruencia

Az egész számok világában az osztás a négy alapművelet közül az egyetlen, amely nem végezhető el bármely két elem között (a 0-val való osztástól eltekintve sem). Ezért különösen fontos a maradékos osztás: valamely $a,b,k,r \in \mathbb{Z}$ egészekre azt mondjuk, hogy *a-ban k-szor van meg a b és a maradék r*, ha $a = k \cdot b + r$ és $0 \leq r \leq |b| - 1$. Fontos tehát, hogy a maradék mindig egy $|b|$-nél kisebb, nemnegatív egész. Így például (-30)-at 9-cel maradékosan osztva azt kapjuk, hogy az (-4)-szer van meg benne és a maradék 6 (hiszen $-30 = (-4) \cdot 9 + 6$). A maradékos osztás bármely $a,b \in \mathbb{Z}$, $b \neq 0$ egészek esetén elvégezhető (vagyis léteznek hozzájuk a megfelelő k,r értékek), hiszen az a-tól az egész számok sorozatában lefelé lépegetve legföljebb $|b| - 1$ távolságban nyilván találunk b-vel osztható számot.

A maradékos osztáson alapul az alábbi egyszerű, de a számelméleti vizsgálatok szempontjából alapvető fontosságú fogalom.

2.3.1. Definíció. *Legyenek $a,b,m \in \mathbb{Z}$, $m \neq 0$ tetszőleges egészek. Azt mondjuk, hogy a* kongruens *b*-vel modulo *m, ha a-t és b-t m-mel maradékosan osztva azonos maradékokat kapunk. Ennek a jele: $a \equiv b \pmod{m}$, vagy rövidítve $a \equiv b \ (m)$. Az m számot az $a \equiv b \pmod{m}$ kongruencia* modulusának *nevezzük.*

Így például: $17 \equiv 52 \pmod{7}$, mert 17 és 52 7-es maradéka is 3. Igaz a $33 \equiv -30 \pmod{9}$ kongruencia is, mert 33 és -30 9-es maradéka egyaránt 6 (ugyanis a -30 is 6-tal nagyobb egy 9-cel osztható számnál). A kongruencia jele nem véletlen emlékeztet az egyenlőségjelre: a fogalom mögött meghúzódó szemlélet az, hogy az a és b egészeket azonosnak tekintjük az m szerinti maradékuk szempontjából. Kevésbé szemléletes, de bizonyításokban kényelmesebben kezelhető a kongruencia definíciójának alábbi átfogalmazása.

2.3.2. Állítás. *Tetszőleges $a,b,m \in \mathbb{Z}$, $m \neq 0$ egészekre $a \equiv b \pmod{m}$ akkor és csak akkor igaz, ha $m|a-b$.*

Bizonyítás: Jelölje r_1, illetve r_2 az a, illetve a b maradékát m-mel osztva. Eszerint $a = k_1 \cdot m + r_1$ és $b = k_2 \cdot m + r_2$ valamely k_1,k_2 és $0 \leq r_1,r_2 \leq m-1$ egészekre. Mivel a és b szerepe szimmetrikus, ezért feltehető, hogy $r_1 \geq r_2$. Ezekből $a - b = (k_1 - k_2) \cdot m + (r_1 - r_2)$, vagyis $(a-b)$-t m-mel osztva a maradék $r_1 - r_2$. Így $m|a-b$ akkor és csak akkor teljesül, ha $r_1 = r_2$, ami definíció szerint valóban ekvivalens azzal, hogy $a \equiv b \pmod{m}$ fennáll. □

A kongruencia fogalmát az teszi igazán hasznossá, hogy az alapműveletek – az osztás kivételével – hasonlóan végezhetők vele, mint az egyenletekkel. Az összeadásra, kivonásra és szorzásra vonatkozó azonosságokat foglalja össze az alábbi tétel.

2.3.3. Tétel. *Tegyük fel, hogy $a \equiv b \pmod m$ és $c \equiv d \pmod m$ fennállnak az a,b,c,d,m egészekre és $k \geq 1$ tetszőleges. Ekkor igazak az alábbiak:*
(i) $a+c \equiv b+d \pmod m$
(ii) $a-c \equiv b-d \pmod m$
(iii) $a \cdot c \equiv b \cdot d \pmod m$
(iv) $a^k \equiv b^k \pmod m$

Bizonyítás: A tétel feltételei a 2.3.2. Állítás szerint úgy fogalmazhatók, hogy $m|a-b$ és $m|c-d$. Mivel m-mel osztható számok összege és különbsége is nyilván m-mel osztható, ezért ezekből következik, hogy $m|(a-b)+(c-d) = (a+c)-(b+d)$ és $m|(a-b)-(c-d) = (a-c)-(b-d)$; vagyis (ismét a 2.3.2. Állítás miatt) a tétel *(i)* és *(ii)* állításai igazak.

Mivel egy m-mel osztható szám bármely többszöröse is nyilván m-mel osztható, ezért $m|a-b$-ből $m|c(a-b) = ac-bc$ következik és hasonlóan, $m|c-d$ miatt $m|b(c-d) = bc-bd$. Ismét kihasználva, hogy m-mel oszható számok összege m-mel osztható kapjuk, hogy $m|(ac-bc)+(bc-bd) = ac-bd$. Ez épp a tétel *(iii)* állítása (a 2.3.2. Állítást használva).

Végül a *(iv)* állítás következik a *(iii)*-ból: ha azt a $c = a$ és $d = b$ szereposztásban alkalmazzuk, akkor az $a^2 \equiv b^2 \pmod m$ kongruenciát kapjuk; erre és az $a \equiv b \pmod m$ kongruenciára újra *(iii)*-at alkalmazva kapjuk az $a^3 \equiv b^3 \pmod m$ kongruenciát, stb., $(k-3)$ további ilyen lépés után jutunk az $a^k \equiv b^k \pmod m$ eredményre. □

A fenti tétel első három állítását gyakran használjuk már a $c = d$ speciális esetben is: mivel $c \equiv c \pmod m$ nyilván igaz, ezért $a \equiv b \pmod m$-ből következik, hogy $a+c \equiv b+c \pmod m$, $a-c \equiv b-c \pmod m$, illetve $a \cdot c \equiv b \cdot c \pmod m$. Hasonló állítás az osztás esetében nyilván nem érvényes – már csak azért sem, mert a kongruencia mindkét oldalán egész számnak kell állnia. De még ha ez a feltétel nem is sérülne, az osztás akkor sem feltétlen működik: például $40 \equiv 64 \pmod{12}$ igaz, de mindkét oldalt 8-cal elosztva a hamis $5 \equiv 8 \pmod{12}$ állítást kapnánk. A kongruenciák osztására vonatkozó szabályt az alábbi tétel mondja ki.

2.3.4. Tétel. *Legyenek a,b,c,m tetszőlegesek és $d = (c,m)$ (ahol a gömbölyű zárójel a legnagyobb közös osztót jelöli). Ekkor $a \cdot c \equiv b \cdot c \pmod m$ akkor és csak akkor igaz, ha $a \equiv b \pmod{\frac{m}{d}}$.*

Bizonyítás: Legyen $c' = \frac{c}{d}$ és $m' = \frac{m}{d}$. Nyilván c' és m' egészek (mert d közös osztója c-nek és m-nek). Továbbá $(c',m') = 1$, ugyanis ellenkező esetben $(c',m') \cdot d$ egy d-nél nagyobb közös osztója volna c-nek és m-nek.

$a \cdot c \equiv b \cdot c \pmod m$ a 2.3.2. Állítás szerint azt jelenti, hogy $m|ac-bc = c(a-b)$. Ez ekvivalens azzal, hogy $m'|c'(a-b)$ (mert az $m \cdot k = c(a-b)$ egyenlet is ekvivalens az $m' \cdot k = c'(a-b)$ egyenlettel). Megmutatjuk, hogy ez tovább ekvivalens az $m'|a-b$ oszthatósággal – amivel (ismét a 2.3.2. Állítás miatt) a tétel bizonyítása teljes lesz.

Valóban, egyrészt $m'|a-b$-ből nyilván következik $m'|c'(a-b)$ (hiszen m'-vel osztható szám bármely többszöröse is az). Megfordítva, ha $m'|c'(a-b)$, akkor c' és $a-b$ prímtényezős felbontásának egyesítése már tartalmazza az m' prímtényezős felbontását. Mivel azonban $(c',m')=1$, ezért c' és m' prímtényezős felbontásában nincs közös prím, így m' prímtényezős felbontását teljes egészében az $a-b$ prímtényezős felbontásának kell tartalmaznia. Vagyis $m'|a-b$ valóban igaz. □

A fenti tételnek gyakran használjuk azt a következményét, hogy $(c,m)=1$ esetén $a\cdot c\equiv b\cdot c \pmod m$ ekvivalens az $a\equiv b \pmod m$ kongruenciával.

2.3.5. Feladat. Milyen maradékot ad
a) 100^{100} 11-gyel osztva; b) 654^{321} 655-tel osztva; c) 111^{41} 35-tel osztva?

Megoldás: A megoldásban végig a 2.3.3. Tétel állításait fogjuk használni.

a) Tudjuk, hogy $100\equiv 1 \pmod{11}$ (hiszen 99 osztható 11-gyel). A kongruencia mindkét oldalát 100-adik hatványra emelve: $100^{100}\equiv 1^{100}=1 \pmod{11}$. Így 100^{100} 1 maradékot ad 11-gyel osztva.

b) Itt a $654\equiv -1 \pmod{655}$ kongruenciából érdemes kiindulni. Ezt a 321-edikre emelve: $654^{321}\equiv(-1)^{321}=-1 \pmod{655}$. Így 654^{321} 654 maradékot ad 655-tel osztva.

c) A $111\equiv 6 \pmod{35}$ kongruenciát 41-edikre emelve: $111^{41}\equiv 6^{41} \pmod{35}$. Most azonban ezzel a feladat még nincs megoldva. A továbblépéshez a kulcs az lesz, hogy $6^2\equiv 1 \pmod{35}$. Ezt 20-adikra emelve: $6^{40}=(6^2)^{20}\equiv 1^{20}=1 \pmod{35}$. Mindkét oldalt 6-tal szorozva: $6^{41}=6^{40}\cdot 6\equiv 6 \pmod{35}$. Mindezekből tehát $111^{41}\equiv 6 \pmod{35}$, így a keresett maradék a 6. □

2.4. Lineáris kongruenciák

A lineáris kongruencia feladata úgy viszonylik a kongruencia fogalmához, mint ahogyan a lineáris egyenlet az egyenlőséghez: adott a,b,m egészek esetén azokat az x egészeket keressük (ha léteznek), amelyekre $a\cdot x\equiv b \pmod m$ teljesül. Például a $4x\equiv 3 \pmod{11}$ lineáris kongruenciának megoldása az $x=9$ (mert $36\equiv 3 \pmod{11}$). A $8x\equiv 3 \pmod{14}$ lineáris kongruencia viszont nyilván nem megoldható, mert $8x$ páros szám, így nem adhat 14-gyel osztva 3 maradékot.

Azonnal megfigyelhetjük, hogy ha az $a\cdot x\equiv b \pmod m$ lineáris kongruenciának x_0 egy megoldása, akkor ez rögtön végtelen sok további megoldást is garantál: ha $x_1\equiv x_0 \pmod m$, akkor a 2.3.3. Tétel szerint $a\cdot x_1\equiv a\cdot x_0\equiv b \pmod m$ is teljesül, így x_1 is megoldás. Például az imént látott $4x\equiv 3 \pmod{11}$ lineáris kongruenciának megoldása minden olyan x, amelyre $x\equiv 9 \pmod{11}$. Nem nehéz végiggondolni, hogy ezzel ennek a feladatnak az összes megoldását megkaptuk: valóban, ha az x_1 megoldás 11-gyel osztva x_0 maradékot ad, akkor $0\le x_0\le 10$ és $x_1\equiv x_0 \pmod{11}$ miatt x_0 is megoldás, márpedig az $x=0,1,\dots,10$ egészeket behelyettesítve látható, hogy ezek közül csak az $x=9$ megoldás. De például a $6x\equiv 4 \pmod{10}$ lineáris

kongruenciának már 2 megoldása is van a $0,1,\dots,9$ számok között: a 4 és a 9; ezért ennek a feladatnak az összes megoldását így adhatjuk meg: $x \equiv 4$ vagy $9 \pmod{10}$.

A továbbiakban elfogadjuk azt a megállapodást, hogy az $a \cdot x \equiv b \pmod{m}$ lineáris kongruencia megoldásait mindig „modulo m keressük" – vagyis arra vagyunk kíváncsiak, hogy x milyen maradékot adhat m-mel osztva. Az $a \cdot x \equiv b \pmod{m}$ feladat megoldásainak a számáról is fogunk beszélni – de ezeket is „modulo m számoljuk", vagyis arra keressük a választ, hogy hány különböző m szerinti maradékot adhat x. Például az előbb látott $6x \equiv 4 \pmod{10}$ lineáris kongruenciának 2 megoldása van modulo 10.

A fenti példákból egyben az $a \cdot x \equiv b \pmod{m}$ egy elvi megoldási módszere is kiolvasható: a $0,1,\dots,m-1$ számokat egyesével behelyettesítve az összes megoldást adó maradékot megtaláljuk. Persze ez a módszer nagyobb m-ekre rendkívül lassú. Az alábbi feladatban látni fogjuk, hogy a 2.3.3. és a 2.3.4. Tételeket használva hogyan lehet egyes konkrét lineáris kongruenciákat ennél sokkal gyorsabban megoldani, később pedig (a 2.6.5. szakaszban) egy hatékony algoritmust is megadunk.

2.4.1. Feladat. Oldjuk meg az alábbi lineáris kongruenciákat!

a) $68x \equiv 12 \pmod{98}$ b) $39x \equiv 24 \pmod{198}$

c) $59x \equiv 4 \pmod{222}$ d) $51x \equiv 100 \pmod{170}$

Megoldás: a) Első lépésként elosztjuk mindkét oldalt 4-gyel. Mivel $(4,98) = 2$, ezért a 2.3.4. Tétel szerint a modulus $\frac{98}{2} = 49$-re változik:

$$17x \equiv 3 \pmod{49}.$$

Most mindkét oldalt megszorozzuk 3-mal:

$$51x \equiv 9 \pmod{49}.$$

Ez a lépés azért volt hasznos, mert $51 \equiv 2 \pmod{49}$ miatt $51x \equiv 2x \pmod{49}$ is igaz minden x-re, így a feladat a következőre egyszerűsödik:

$$2x \equiv 9 \pmod{49}.$$

Ez viszont $9 \equiv 58 \pmod{49}$ miatt így is írható:

$$2x \equiv 58 \pmod{49}.$$

Így ismét eloszthatjuk mindkét oldalt 2-vel:

$$x \equiv 29 \pmod{49}$$

(ahol a modulus most $(49,2) = 1$ miatt nem változott). Ezzel ugyan már az összes megoldást megkaptuk, de még hátra van, hogy a fentebb írtak szerint az eredményt modulo 98 is megadjuk. Ha $x \equiv 29 \pmod{49}$, akkor $x = 49k + 29$ valamilyen k-ra; ha most $k = 2\ell$ páros, akkor $x = 98\ell + 29$, ha viszont $k = 2\ell + 1$ páratlan, akkor

$x = 49(2\ell+1)+29 = 98\ell+78$. (Vagyis egy 49-cel osztva 29 maradékot adó szám 98-cal osztva nyilván 29 vagy $29+49 = 78$ maradékot ad.) Így a végeredmény:

$$x \equiv 29 \text{ vagy } 78 \pmod{98}.$$

Ellenőrizve az eredményeket: $68\cdot 29 \equiv 12 \pmod{98}$ és $68\cdot 78 \equiv 12 \pmod{98}$ valóban igazak.

b) Mindkét oldalt 3-mal osztva: $13x \equiv 8 \pmod{66}$.
(Itt az új modulus: $\frac{198}{(198,3)} = \frac{198}{3} = 66$.) Most 5-tel szorozva: $65x \equiv 40 \pmod{66}$.
Ez $65 \equiv -1 \pmod{66}$ miatt így is írható: $-x \equiv 40 \pmod{66}$.
(-1)-gyel szorozva: $x \equiv -40 \equiv 26 \pmod{66}$.
Tehát a feladat megoldása: $x \equiv 26$ vagy 92 vagy $158 \pmod{198}$.
(Valóban, ha $x = 66k+26$, akkor a $k = 3\ell$, $k = 3\ell+1$ és $k = 3\ell+2$ alakokat sorban behelyettesítve $x = 198\ell+26$, $x = 198\ell+92$, illetve $x = 198\ell+158$ adódik.)
Ellenőrizve kiderül, hogy $39\cdot 26 \equiv 24 \pmod{198}$, $39\cdot 92 \equiv 24 \pmod{198}$ és $39\cdot 158 \equiv 24 \pmod{198}$ egyaránt igazak.

c) 4-gyel szorozva: $236x \equiv 16 \pmod{222}$.
Ez $236 \equiv 14 \pmod{222}$ miatt így írható: $14x \equiv 16 \pmod{222}$.
2-vel osztva: $7x \equiv 8 \pmod{111}$
(ahol az új modulus $\frac{222}{(222,2)} = \frac{222}{2} = 111$). 16-tal szorozva: $112x \equiv 128 \pmod{111}$.
Ez $112 \equiv 1 \pmod{111}$ és $128 \equiv 17 \pmod{111}$ miatt így írható:$x \equiv 17 \pmod{111}$.
Ebből $x \equiv 17$ vagy $128 \pmod{222}$.
Azonban az ellenőrzés most azzal az első látásra meglepő eredménnyel zárul, hogy $59\cdot 128 \equiv 4 \pmod{222}$ ugyan igaz, de $59\cdot 17 \equiv 4 \pmod{222}$ nem: valóban, $59\cdot 17 = 1003$, ez pedig 115 maradékot ad 222-vel osztva. Hol követtük el a „hibát”, hogy jöhetett be hamis gyök? A titok nyitja a legelső lépésben, a 4-gyel való szorzásban rejlik: ez ugyan „helyes lépés” volt abban az értelemben, hogy ha $59x \equiv 4 \pmod{222}$ igaz, akkor $236x \equiv 16 \pmod{222}$ is (ez következik a 2.3.3. Tétel *(iii)* állításából). A fordított következtetés azonban már hamis: ha $236x \equiv 16 \pmod{222}$, akkor 4-gyel osztva (a 2.3.4. Tételnek megfelelően, $(222,4) = 2$ miatt) az $59x \equiv 4 \pmod{111}$ kongruenciát kapjuk. Ez valóban kevesebb az $59x \equiv 4 \pmod{222}$ kongruencia állításánál, hiszen $(59x-4)$-nek csak a 111-gyel való oszthatóságát állítja a 222-vel való oszthatóság helyett. Vagyis a fenti megoldásban a 4-gyel szorzás nem ekvivalens lépés volt – hasonlóan ahhoz, ahogyan például a négyzetre emelés sem ekvivalens lépés az egyenletek megoldásakor.

A tanulság is ugyanaz, mint amit az egyenletek négyzetre emelésével kapcsolatban megszoktunk: ha az $a\cdot x \equiv b \pmod{m}$ lineáris kongruencia mindkét oldalát egy m-hez nem relatív prím számmal szorozzuk, akkor ez a lépés nem ekvivalens, hamis gyökök megjelenésével jár. Ha tehát ilyen lépést teszünk (ami időnként – mint például a fenti megoldásban is – nagyon kényelmes), akkor az eredmények ellenőrzésével ki kell szűrni a hamis gyököket. (Érdemes megfigyelni, hogy az a) és a b) feladatban egyszer sem szoroztunk a modulushoz nem relatív prím számmal, csak ekvivalens lépéseket tettünk. Így ezeknél az ellenőrzés szerepe csak a számolási hibák kiszűrésére korlátozódott, a megoldások elvileg – szemben a c) feladatra adott fenti megoldásunkkal – enélkül is teljes értékűek voltak.)

Végül is tehát a c) feladat végeredménye: $x \equiv 128 \pmod{222}$.

d) Ha x_0 megoldása a feladatnak, akkor $170 | 51x_0 - 100$. Ebből $17 | 170$ miatt $17 | 51x_0 - 100$. Azonban ez lehetetlen: $17 | 51$ miatt $17 | 51x_0$, így (mivel 17-tel osztható számok különbsége is 17-tel osztható) a $17 | 100$ ellentmondást kapnánk. Tehát ennek a lineáris kongruenciának nincs megoldása. □

A fenti feladat d) részében a megoldhatatlanság hátterében az volt, hogy 51-nek és 170-nek van egy olyan közös osztója – nevezetesen a 17 –, ami 100-nak nem osztója. Az alábbi tételből az derül ki, hogy egy tetszőleges $a \cdot x \equiv b \pmod{m}$ lineáris kongruencia is csak akkor lehet megoldhatatlan, ha ezt hasonlóan meg lehet mutatni a-nak és m-nek egy olyan közös d osztójával, ami b-nek nem osztója. Persze felmerül a kérdés: hogyan lehet egy ilyen d létezését eldönteni? A válasz egyszerű: ha van ilyen d, akkor a és m legnagyobb közös osztója is rendelkezik ugyanezzel a tulajdonsággal (mert az a számelmélet alaptételéből következően d-nek többszöröse).

2.4.2. Tétel. *Az $a \cdot x \equiv b \pmod{m}$ lineáris kongruencia akkor és csak akkor megoldható, ha $(a,m) | b$. Ha pedig ez a feltétel teljesül, akkor $a \cdot x \equiv b \pmod{m}$ megoldásainak száma modulo m (a,m)-val egyenlő.*

Bizonyítás: Legyen $d = (a,m)$. Először a feltétel szükségességét látjuk be. Tegyük fel tehát, hogy $a \cdot x \equiv b \pmod{m}$ megoldható és legyen x_0 egy megoldás. Ekkor tehát $a \cdot x_0 \equiv b \pmod{m}$, így $m | a \cdot x_0 - b$. Ebből $d | m$ miatt $d | a \cdot x_0 - b$ is következik. Emellett $d | a$ miatt $d | a \cdot x_0$ is igaz. Mivel d-vel osztható számok különbsége is d-vel osztható, ezekből $(a,m) = d | a \cdot x_0 - (a \cdot x_0 - b) = b$ valóban következik.

A feltétel elégségességét először a $(a,m) = 1$ esetben látjuk be. Mivel ekkor $(a,m) | b$ mindenképp igaz, meg kell mutatnunk, hogy $a \cdot x \equiv b \pmod{m}$ megoldható. Helyettesítsük be x helyére a $0, 1, \ldots, m-1$ számok mindegyikét – vagyis mindet szorozzuk meg a-val. Állítjuk, hogy a kapott eredmények közül semelyik kettő nem ad azonos maradékot m-mel osztva. Valóban, ha valamely $0 \leq x_1 < x_2 \leq m-1$ értékekre $a \cdot x_1 \equiv a \cdot x_2 \pmod{m}$ teljesülne, akkor a 2.3.4. Tétel szerint mindkét oldalt a-val osztva az $x_1 \equiv x_2 \pmod{m}$ kongruenciát kapnánk (ugyanis az osztásnál $(a,m) = 1$ miatt a modulus nem változik); ez azonban $0 \leq x_1 < x_2 \leq m-1$ miatt lehetetlen. Ezzel a feltétel elégségességét az $(a,m) = 1$ esetben valóban beláttuk: mivel az $a \cdot 0, a \cdot 1, a \cdot 2, \ldots, a \cdot (m-1)$ számok közül bármely kettő különböző maradékot ad m-mel osztva, így ezek között a maradékok között minden m szerinti maradék pontosan egyszer fordul elő, vagyis pontosan az egyikük egyenlő a b maradékával. Ezzel tehát a megoldhatóság ténye mellett azt is beláttuk, hogy az $(a,m) = 1$ esetben a megoldások száma 1-gyel egyenlő modulo m – összhangban a tétel állításával.

Most belátjuk a feltétel elégségességét az $(a,m) > 1$ esetben is. Legyen továbbra is $d = (a,m)$ és tegyük fel, hogy $d | b$. Vezessük be az $a' = \frac{a}{d}$, $m' = \frac{m}{d}$, $b' = \frac{b}{d}$ jelöléseket. Ekkor tehát a', m' és b' is egészek, továbbá $(a',m') = 1$. (Az utóbbi állítás indoklását a 2.3.4. Tétel bizonyításában már láttuk: ha $(a',m') > 1$, akkor $(a',m') \cdot d$ egy d-nél nagyobb közös osztója volna a-nak és m-nek.) A 2.3.4. Tétel szerint az $a \cdot x \equiv b \pmod{m}$ kongruencia ekvivalens a d-vel való osztással kapott

$a' \cdot x \equiv b' \pmod{m'}$ kongruenciával. (Itt az új modulus $\frac{m}{(a,m)} = \frac{m}{d} = m'$.) Erre azonban $(a',m') = 1$ miatt már alkalmazható az előző bekezdésben már bebizonyított speciális eset: az $a' \cdot x \equiv b' \pmod{m'}$ (és így az $a \cdot x \equiv b \pmod{m}$) lineáris kongruencia megoldható és egyetlen megoldása van modulo m'.

A tétel bizonyításából tehát mindössze annak a megmutatása van hátra, hogy ha az $a \cdot x \equiv b \pmod{m}$ lineáris kongruenciának x_0 az egyetlen megoldása modulo m' (ahol $0 \leq x_0 \leq m'-1$), akkor ez $d = (a,m)$ darab megoldást jelent modulo m. Ezt a 2.4.1. Feladatban látotthoz hasonlóan látjuk be. Az összes megoldás $x = k \cdot m' + x_0$ alakú, ahol k tetszőleges egész. Azt kell tehát megvizsgálnunk, hogy ha itt k helyére két különböző egészt – jelölje ezeket k_1 és k_2 – helyettesítünk, akkor az ezekből adódó $k_1 \cdot m' + x_0$ és $k_2 \cdot m' + x_0$ megoldások azonos vagy különböző maradékot adnak m-mel osztva. A $k_1 \cdot m' + x_0 \equiv k_2 \cdot m' + x_0 \pmod{m}$ kongruencia mindkét oldalából először x_0-t levonva, majd m'-vel osztva a $k_1 \equiv k_2 \pmod{d}$ kongruenciát kapjuk, ahol az új modulus $\frac{m}{(m,m')} = \frac{m}{m'} = d$. Ez tehát azt jelenti, hogy k helyére a $0,1,\ldots,d-1$ értékeket helyettesítve megkapjuk az összes lehetséges megoldást modulo m; vagyis az $a \cdot x \equiv b \pmod{m}$ összes megoldása

$$x \equiv x_0 \text{ vagy } x_0 + m' \text{ vagy } x_0 + 2m' \text{ vagy } \ldots \text{ vagy } x_0 + (d-1)m' \pmod{m}.$$

Így a megoldások száma modulo m valóban $d = (a,m)$. □

Ez a tétel tehát megválaszolja egy tetszőleges lineáris kongruencia megoldhatóságának a kérdését, de a fenti bizonyítás a megoldások megtalálására nem ad hatékony módszert. Valóban, az $(a,m) = 1$ esetben a bizonyítás alapgondolata egyszerűen az összes eset végigpróbálgatása volt, ez pedig nagy m-ekre használhatatlanul lassú. A bizonyítás mégis szolgál egy, a gyakorlat szempontjából is fontos tanulsággal: az $a \cdot x \equiv b \pmod{m}$ lineáris kongruencia megoldása az $(a,m) > 1$ esetben egyszerűen egy (a,m)-val való osztás révén visszavezethető az $a' \cdot x \equiv b' \pmod{m'}$ feladatra, amelyre már $(a',m') = 1$. Így később (a 2.6.5. szakaszban) elég lesz csak az ebbe a speciális esetbe tartozó lineáris kongruenciák megoldására szolgáló algoritmust megadni, ezzel az általános esetet is megoldjuk.

Emellett a fenti tétel akkor is jól használható, ha egy lineáris kongruenciát a 2.4.1. Feladatban látotthoz hasonló módszerekkel oldunk meg. Például a 2.4.1. Feladat d) részében $(51,170) = 17 \not| 100$ mutatja, hogy nincs megoldás. De (a,m) meghatározása akkor is hasznos, ha $(a,m) \big| b$ teljesül (és így a lineáris kongruencia megoldható), mert előre tudható, hogy a megoldások száma (a,m) lesz modulo m, ami segítség lehet az esetleges számolási hibák felismerésében és a hamis gyökök kiszűrésében.

2.4.1. Kétváltozós, lineáris diofantikus egyenletek

Az i. sz. 3. században élt görög matematikus, *Diophantosz* munkássága nyomán *diofantikus egyenletnek* nevezzük azokat a feladatokat, ahol egy (több változós) egyenletnek az egész megoldásait keressük. Ezek között a feladatok között nagyon nehezek is vannak, ilyen például a híres *Nagy Fermat-sejtés*, amit a 17. században élt *Pierre de Fermat* éppen Diophantosz *Arithmetika* című könyvének olvasása közben,

annak a margójára jegyzett fel: eszerint az $x^n+y^n=z^n$ egyenletnek az $n\geq 3$ esetben nincs megoldása, ha x, y és z nullától különböző egészek. Ezt csak több, mint 350 évvel később, 1994-ben sikerült bebizonyítania *Andrew Wilesnak*.

A diofantikus egyenletek közül a legegyszerűbbek a kétváltozós, lineáris egyenletek: adott a,b,c egészekre az $a\cdot x+b\cdot y=c$ egyenlet egész x,y megoldásait keressük. Látni fogjuk, hogy ezek a feladatok könnyen visszavezethetők egy lineáris kongruencia megoldására.

2.4.3. Feladat. Adjuk meg a $31x+153y=928$ egyenlet összes egész megoldását!

Megoldás: Átrendezés után a $153y=928-31x$ egyenlet egész megoldásait keressük – vagyis azokat az x egészeket, amelyekre $153\big|928-31x$. Ez pedig a 2.3.2. Állítás szerint ekvivalens a $31x\equiv 928 \pmod{153}$ lineáris kongruenciával.

Mivel $(31,153)=1$, ezért a 2.4.2. Tétel szerint ez a lineáris kongruencia megoldható és egyetlen megoldása van modulo 153. Ezt a 2.4.1. Feladatban látott módszerekkel keressük meg.

$928\equiv 10 \pmod{153}$ miatt a feladat így írható: $31x\equiv 10 \pmod{153}$.
5-tel szorozva: $155x\equiv 50 \pmod{153}$.
Ez $155\equiv 2 \pmod{153}$ miatt így írható: $2x\equiv 50 \pmod{153}$.
2-vel osztva: $x\equiv 25 \pmod{153}$,
ahol $(2,153)=1$ miatt az osztás a kongruencia modulusát nem változtatta meg.

Ezzel tehát a lineáris kongruenciát megoldottuk. Egyetlen megoldást kaptunk modulo 153, ami összhangban van a 2.4.2. Tétel állításából nyert fenti várakozásunkkal, így ez az eredmény biztosan helyes. De hivatkozhatunk arra is, hogy a lépéseink ekvivalensek voltak – igaz ez még az 5-tel való szorzásra is $(153,5)=1$ miatt (lásd a 2.4.1. Feladat c) részének megoldása kapcsán írtakat).

A diofantikus egyenletben x helyére tehát $x=153k+25$ alakú számot írhatunk. Ezt az egyenletbe visszahelyettesítve, majd y-t kifejezve: $y=1-31k$. Így a diofantikus egyenlet megoldásai az $(x,y)=(153k+25,1-31k)$ alakú számpárok. □

A fenti megoldásban látott módszer tetszőleges $a\cdot x+b\cdot y=c$ alakú diofantikus egyenletre működik: $b\cdot y=c-a\cdot x$ miatt x-re a $b\big|c-a\cdot x$ feltétel adódik, ami ekvivalens az $a\cdot x\equiv c \pmod{b}$ lineáris kongruenciával. Ha ennek nincs megoldása, akkor a diofantikus egyenletnek sincs. Ha van, akkor a 2.4.2. Tétel szerint (a,b) darab megoldás létezik modulo b – ami úgy is fogalmazható, hogy egyetlen megoldás van modulo $b'=\frac{b}{(a,b)}$ (lásd a 2.4.2. Tétel bizonyításának utolsó bekezdését). Ha ezt a modulo b' vett egyértelmű megoldást x_0 jelöli, akkor az egyenletben x helyére $x=x_0+k\cdot b'$ alakú szám írható; ezt az egyenletbe helyettesítve és y-t kifejezve megkapjuk az összes (x,y) megoldást.

A most leírt módszer segítségével könnyen nyerhető általános érvényű tétel is a kétváltozós, lineáris diofantikus egyenletek megoldhatóságáról.

2.4.4. Tétel. *Az $a\cdot x+b\cdot y=c$ egyenletnek akkor és csak akkor létezik egész x,y megoldása, ha $(a,b)\big|c$.*

Bizonyítás: Az imént megmutattuk, hogy $a\cdot x+b\cdot y=c$ egészekkel való megoldhatósága ekvivalens az $a\cdot x\equiv c \pmod{b}$ lineáris kongruencia megoldhatóságával. Ez pedig a 2.4.2. Tétel szerint valóban ekvivalens az $(a,b)\big|c$ feltétellel. □

2.4.2. Szimultán kongruenciarendszerek

A számelmélet számos alkalmazásában felmerül olyan feladat, amelyben az egyszerre több kongruenciának is eleget tevő x számokat keressük: $x \equiv a_1 \pmod{m_1}$, $x \equiv a_2 \pmod{m_2}$, $\ldots$, $x \equiv a_k \pmod{m_k}$. Ezt a feladatot *szimultán kongruenciarendszernek* nevezzük és a megoldását gyakran használják egyes számítási feladatok gyorsítására, illetve tárkapacitásának csökkentésére (például egész elemű mátrixok determinánsának kiszámításánál, vagy egész együtthatós lineáris egyenletrendszerek megoldásánál). Szerencsére ezek a feladatok is könnyen visszavezethetők lineáris kongruenciákra.

2.4.5. Feladat. Oldjuk meg az alábbi kongruenciarendszereket!
a) $x \equiv 11 \pmod{42}$, $x \equiv 10 \pmod{199}$
b) $x \equiv 16 \pmod{78}$, $x \equiv 56 \pmod{364}$
c) $x \equiv 1 \pmod{15}$, $x \equiv 4 \pmod{21}$, $x \equiv 6 \pmod{50}$

Megoldás: a) Az első kongruencia szerint $x = 42k + 11$ valamely k egészre. Ezt a második kongruenciába helyettesítve: $42k + 11 \equiv 10 \pmod{199}$.
11-et levonva lineáris kongruencia feladatra jutunk: $42k \equiv -1 \pmod{199}$.
5-tel szorozva: $210k \equiv -5 \pmod{199}$,
ami $210 \equiv 11 \pmod{199}$ miatt így is írható: $11k \equiv -5 \pmod{199}$.
18-cal szorozva: $198k \equiv -90 \pmod{199}$,
vagyis $-k \equiv -90 \pmod{199}$.
(-1)-gyel szorozva: $k \equiv 90 \pmod{199}$.
Minden lépésünk ekvivalens volt (mert $(5,199) = 1$ és $(18,199) = 1$), ezért ezzel valóban a lineáris kongruencia összes megoldását kaptuk meg. Így tehát $k = 199\ell + 90$ valamely ℓ egészre. Ezt visszahelyettesítve az $x = 42k + 11$ egyenletbe: $x = 42(199\ell + 90) + 11 = 8358\ell + 3791$. Vagyis a kongruenciarendszer megoldásai az $x \equiv 3791 \pmod{8358}$ kongruenciát kielégítő x egészek.

b) Követve az a) feladat megoldását: $x = 78k + 16$, amit a második kongruenciába helyettesítve, majd 16-ot levonva a $78k \equiv 40 \pmod{364}$ lineáris kongruenciára jutunk. Itt $(78,364) = 26$, így $26 \nmid 40$ miatt (a 2.4.2. Tétel szerint) a lineáris kongruenciának nincs megoldása – így a kongruenciarendszernek sincs.

c) Először az első két kongruencia közös megoldásait keressük meg. Az a) és b) feladatban már látott módszert követjük: $x = 15k + 1$, ezt a második kongruenciába helyettesítve és 1-et levonva: $15k \equiv 3 \pmod{21}$.
3-mal osztva ($\frac{21}{(3,21)} = \frac{21}{3} = 7$ miatt): $5k \equiv 1 \pmod{7}$.
Ez $5 \equiv -2 \pmod{7}$ és $1 \equiv -6 \pmod{7}$ miatt így is írható: $-2k \equiv -6 \pmod{7}$.
(-2)-vel osztva: $k \equiv 3 \pmod{7}$
(és csak ekvivalens lépéseket tettünk). Így $k = 7\ell + 3$, amiből megkapjuk az első két kongruencia közös megoldásait: $x = 15(7\ell + 3) + 1 = 105\ell + 46$. Ezt behelyettesíthetjük a harmadik kongruenciába: $105\ell + 46 \equiv 6 \pmod{50}$.
46-ot levonva: $105\ell \equiv -40 \pmod{50}$.
Felhasználva, hogy $105 \equiv 5 \pmod{50}$ és $-40 \equiv 10 \pmod{50}$: $5\ell \equiv 10 \pmod{50}$.
5-tel osztva: $\ell \equiv 2 \pmod{10}$.

Vagyis $\ell = 10t+2$ alkalmas t-re, amit az $x = 105\ell + 46$ egyenletbe visszahelyettesítve: $x = 1050t + 256$. Vagyis a három kongruenciából álló rendszer megoldásait az $x \equiv 256 \pmod{1050}$ kongruenciát kielégítő x egészek adják. □

A fenti feladatban látott módszer nyilván tetszőleges kongruenciarendszer megoldhatóságának eldöntésére és a megoldások megtalálására alkalmazható. A két kongruenciából álló $x \equiv a_1 \pmod{m_1}$, $x \equiv a_2 \pmod{m_2}$ rendszer megoldásához az első kongruenciából kapott $x = k \cdot m_1 + a_1$ alakot a másodikba helyettesítjük, amiből rendezés után az $m_1 \cdot k \equiv a_2 - a_1 \pmod{m_2}$ lineáris kongruenciát kapjuk k-ra. Ha ennek nincs megoldása, akkor a kongruenciarendszernek sincs. Ha van, akkor (a 2.4.2. Tétel szerint) (m_1, m_2) darab megoldás létezik modulo m_2 – vagyis egyetlen megoldás van modulo $m' = \frac{m_2}{(m_1,m_2)}$. Ha ezt a modulo m' vett egyértelmű megoldást k_0 jelöli, akkor a $k = m' \cdot \ell + k_0$ alakot visszahelyettesítve az $x = k \cdot m_1 + a_1$ egyenletbe az $x = m_1 \cdot (m' \cdot \ell + k_0) + a_1 = m_1 \cdot m' \cdot \ell + m_1 \cdot k_0 + a_1$ megoldásokat kapjuk. Vagyis a kongruenciarendszer megoldásai az $x \equiv m_1 \cdot k_0 + a_1 \pmod{m_1 \cdot m'}$ kongruenciát kielégítő x egészek.

Mivel a két kongruenciából álló rendszerek megoldásai (ha léteznek) a fentiek szerint egyetlen kongruencia megoldáshalmazaként írhatók le, ez (a fenti c) feladat megoldásához hasonlóan) a kettőnél több kongruenciából álló rendszerek megoldását is lehetővé teszi: ehhez a fenti bekezdésben leírt módszerrel mindig eggyel csökkentjük a rendszert alkotó kongruenciák számát, míg végül egyetlen kongruenciából álló rendszert (vagyis a megoldásokat) kapjuk.

A fenti leírt módszer segítségével könnyen nyerhető általános érvényű tétel is a kongruenciarendszerek megoldhatóságáról.

2.4.6. Tétel. *Az $x \equiv a_1 \pmod{m_1}$, $x \equiv a_2 \pmod{m_2}$ kongruenciarendszer akkor és csak akkor megoldható, ha $(m_1, m_2) \mid a_1 - a_2$. Ha ez a feltétel teljesül, akkor a megoldáshalmaz $x \equiv x_0 \pmod{[m_1, m_2]}$ alakú valamilyen x_0 egészre (ahol $[m_1, m_2]$ az m_1 és m_2 legkisebb közös többszörösét jelöli).*

Bizonyítás: Fentebb megmutattuk, hogy a rendszer megoldhatósága ekvivalens az $m_1 \cdot k \equiv a_2 - a_1 \pmod{m_2}$ lineáris kongruencia megoldhatóságával. Ez pedig a 2.4.2. Tétel szerint valóban ekvivalens az $(m_1, m_2) \mid a_1 - a_2$ feltétellel.

A fentiekből az is kiderül, hogy ha ez a feltétel teljesül, akkor a rendszer összes megoldását az $x \equiv x_0 \pmod{m_1 \cdot m'}$ kongruencia adja egy alkalmas x_0 egészre, ahol $m' = \frac{m_2}{(m_1,m_2)}$. Így a tétel állítása következik, az $\frac{m_1 \cdot m_2}{(m_1,m_2)} = [m_1, m_2]$ azonosságból, ami viszont a 2.1.5. Tétel következménye. □

Ennek a tételnek közvetlen következménye az alábbi, *Kínai maradéktétel* néven közismert tétel. Ez a valamikor az i. sz. 3. és 5. század között élt *Szun Cu* kínai természettudós és matematikus *Zhou Bi Suan Jing* című művében már szerepel, innen kapta a nevét.

2.4.7. Tétel. (Kínai maradéktétel)
Ha az $m_1, m_2, \ldots, m_k$ egészek közül bármelyik kettő relatív prím, akkor az $x \equiv a_1 \pmod{m_1}$, $x \equiv a_2 \pmod{m_2}$, $\ldots$, $x \equiv a_k \pmod{m_k}$ kongruenciarendszer megoldható és a megoldáshalmaza $x \equiv x_0 \pmod{m_1 \cdot m_2 \cdot \ldots \cdot m_k}$ alakú valamilyen x_0 egészre.

Bizonyítás: Az $x \equiv a_1 \pmod{m_1}$, $x \equiv a_2 \pmod{m_2}$ rendszer a fenti tétel szerint megoldható (mert $(m_1, m_2) = 1$) és a megoldáshalmaza $x \equiv x_1 \pmod{m_1 \cdot m_2}$ alakú (mert $[m_1, m_2] = m_1 \cdot m_2$). Ezt párba állítva az $x \equiv a_3 \pmod{m_3}$ kongruenciával és ismét alkalmazva a fenti tételt kapjuk, hogy a rendszer ismét megoldható és a megoldáshalmaza $x \equiv x_2 \pmod{m_1 \cdot m_2 \cdot m_3}$ alakú. Ezt értelemszerűen folytatva kapjuk a tétel bizonyítását. □

2.5. Az Euler-Fermat tétel

A (fejezet elején említett) RSA algoritmusban a dekódolás kulcsa az lesz, hogy adott a és m egészekre az a-nak egy m-mel osztva 1 maradékot adó hatványát megtaláljuk – vagyis egy olyan k-t, amelyre $a^k \equiv 1 \pmod{m}$. Ilyen k nem mindig létezik: ha $(a,m) = d > 1$, akkor $a^k - 1$ nemhogy m-mel, de még d-vel sem osztható (mert a^k osztható d-vel). Az Euler-Fermat tételből következik, hogy az $(a,m) = 1$ esetben viszont mindig van ilyen k kitevő – még olyan is, ami a-tól nem függ, csak m-től.

2.5.1. Az Euler-féle φ függvény

A kongruencia fentebb megismert fogalma az m-mel vett maradékuk szerint „azonosítja" az egészeket. Láttuk, hogy a modulo m egymással kongruens egészek valóban sok szempontból azonosan viselkednek – az alábbi állítás szerint pedig még az m-mel vett legnagyobb közös osztójuk szempontjából is.

2.5.1. Állítás. *Ha $a \equiv b \pmod{m}$ teljesül, akkor $(a,m) = (b,m)$.*

Bizonyítás: $a \equiv b \pmod{m}$ miatt $m \mid a - b$, így $b = a + k \cdot m$ alkalmas k-ra. Legyen $d = (a,m)$. Mivel egy d-vel osztható szám bármely többszöröse, illetve ilyenek összege is d-vel osztható, ezért a $d \mid a$ és $d \mid m$ oszthatóságokból $d \mid k \cdot m$ és így $d \mid a + k \cdot m = b$ is következik. Ezek szerint d közös osztója b-nek és m-nek, így ezek legnagyobb közös osztójánál nem lehet nagyobb, vagyis $d = (a,m) \leq (b,m)$. Mivel a és b szerepe az állításban szimmetrikus, hasonlóan belátható a $(b,m) \leq (a,m)$ állítás is, a kettőből együtt pedig $(a,m) = (b,m)$ valóban következik. □

A fenti állításnak fontos következménye, hogy $a \equiv b \pmod{m}$ esetén $(a,m) = 1$ akkor és csak akkor igaz, ha $(b,m) = 1$ – vagyis az m-mel osztva azonos maradékot adó számok vagy mind relatív prímek m-hez, vagy egyikük sem az. Az alábbi, alapvető fontosságú függvény az előbbi eset előfordulásait számolja meg.

2.5.2. Definíció. *Ha $n \geq 2$ egész, akkor az $1, 2, \ldots, n-1$ számok között az n-hez relatív prímek számát $\varphi(n)$-nel jelöljük. Az $n \mapsto \varphi(n)$ függvényt* Euler-féle φ függvénynek *nevezzük. (A φ görög betű, kiolvasáskor „fí"-nek ejtjük.)*

Például $\varphi(10)$ meghatározásához az $1,2,\ldots,9$ számok között kell a 10-hez relatív prímeket megszámolnunk. A páros számok nyilván kiesnek és a páratlanok közül is az 5 (mert ezeknek van 1-nél nagyobb közös osztójuk a 10-zel). A megmaradt 1,3,7 és 9 számok viszont már relatív prímek 10-hez, így $\varphi(10)=4$. Ha $n=p$ prímszám, akkor a definícióból rögtön adódik, hogy $\varphi(p)=p-1$; valóban az $1,2,\ldots,p-1$ számok mind relatív prímek p-hez (mert p-nek nincs 1-nél nagyobb és p-nél kisebb osztója). Persze nagyobb, összetett n-ekre $\varphi(n)$ kiszámítása a definíció alapján nagyon fáradságos volna. Az alábbi tétel ezt nagyban megkönnyíti: kiderül belőle, hogy n prímtényezős felbontásából $\varphi(n)$ könnyen megkapható.

2.5.3. Tétel. *Legyen az $n>1$ egész kanonikus alakja $n=p_1^{\alpha_1}\cdot p_2^{\alpha_2}\cdot\ldots\cdot p_k^{\alpha_k}$. Ekkor*

$$\varphi(n)=\left(p_1^{\alpha_1}-p_1^{\alpha_1-1}\right)\cdot\left(p_2^{\alpha_2}-p_2^{\alpha_2-1}\right)\cdot\ldots\cdot\left(p_k^{\alpha_k}-p_k^{\alpha_k-1}\right).$$

Bizonyítás: Tegyük fel először, hogy n prímtényezős felbontásában csak egyetlen prím van, vagyis $n=p^\alpha$ valamilyen p prímre és $\alpha\geq 1$ egészre. Ekkor $(n,a)>1$ akkor és csak akkor igaz, ha $p|a$ (különben a és n prímtényezős felbontásában nem lehetne közös prím). Ez tehát azt jelenti, hogy az $1,2,\ldots,n$ számok közül $\frac{n}{p}=p^{\alpha-1}$ darab nem relatív prím n-hez (ugyanis nyilván ennyi a p-vel oszthatók száma). Így definíció szerint $\varphi(n)=n-p^{\alpha-1}=p^\alpha-p^{\alpha-1}$. Következésképp a tétel igaz minden prímhatványra – amiből általános n-re is könnyen következni fog az alábbi lemmát felhasználva.

2.5.4. Lemma. *Ha az a,b egészekre $(a,b)=1$, akkor $\varphi(a\cdot b)=\varphi(a)\cdot\varphi(b)$.*

A Lemma bizonyítása: Először is figyeljük meg, hogy $(x,a\cdot b)=1$ akkor és csak akkor igaz, ha $(x,a)=1$ és $(x,b)=1$ egyaránt teljesülnek. Valóban, mivel az $a\cdot b$ prímtényezős felbontása az a és a b prímtényezős felbontásának az egyesítése, ezért $a\cdot b$ prímfelbontásában pontosan akkor nincs az x prímfelbontásában is szereplő prím, ha sem a-ében, sem b-ében nincs. Így azt kell megmutatnunk, hogy az $1,2,\ldots,a\cdot b$ egészek között $\varphi(a)\cdot\varphi(b)$ darab olyan van, ami a-hoz és b-hez is relatív prím.

Írjuk be az $1,2,\ldots,a\cdot b$ egészeket sorfolytonosan egy $(a\times b)$-s M mátrixba, vagyis legyen $m_{i,j}=(i-1)b+j$ minden $1\leq i\leq a$ és $1\leq j\leq b$ esetén. Majd keretezzük be a mátrixban az a-hoz és b-hez (és így $a\cdot b$-hez) is relatív prím elemeket. Például az $a=3$, $b=8$ értékekre készült mátrix az alábbi:

$$\begin{pmatrix} \boxed{1} & 2 & 3 & 4 & \boxed{5} & 6 & \boxed{7} & 8 \\ 9 & 10 & \boxed{11} & 12 & \boxed{13} & 14 & 15 & 16 \\ \boxed{17} & 18 & \boxed{19} & 20 & 21 & 22 & \boxed{23} & 24 \end{pmatrix}$$

Oszloponként fogjuk megszámolni M-ben a bekeretezett elemeket.

Először is vegyük észre, hogy a 2.5.1. Állításból $m_{i,j}=(i-1)b+j\equiv j \pmod b$ miatt $(m_{i,j},b)=(j,b)$ következik. Így $(m_{i,j},b)=1$ akkor és csak akkor igaz, ha $(j,b)=1$, vagyis M-ben a b-hez relatív prím elemek a b-hez relatív prím sorszámú

oszlopokban vannak (és bekeretezett elemek is csak ezekben az oszlopokban lehetnek). Azt kell tehát megszámolnunk, hogy $(j,b) = 1$ esetén a j-edik oszlop hány a-hoz is relatív prím elemet tartalmaz.

Ehhez megmutatjuk, hogy $(j,b) = 1$ esetén a j-edik oszlop bármely két tagja különböző maradékot ad a-val osztva. Ha ugyanis az oszlop $m_{i,j} = (i-1)b + j$ és $m_{k,j} = (k-1)b + j$ elemeire $(i-1)b + j \equiv (k-1)b + j \pmod{a}$, akkor j-t levonva majd b-vel osztva és 1-et hozzáadva az $i \equiv k \pmod{a}$ kongruenciát kapjuk – ahol az osztásnál a modulus $(a,b) = 1$ miatt nem változott. Mivel $1 \leq i,k \leq a$, ez valóban azt jelenti, hogy $m_{i,j} \equiv m_{k,j} \pmod{a}$ csak az $i = k$ esetben fordulhatna elő.

Mivel az M j-edik oszlopában a elem van és nincs köztük két a-val osztva azonos maradékot adó, ezért ezeknek az a szerinti maradékait véve a $0,1,\ldots,a-1$ maradékok mindegyikét pontosan egyszer kapjuk meg. Ebből és a 2.5.1. Állításból következik, hogy a j-edik oszlopban pontosan $\varphi(a)$ darab a-hoz relatív prím van; valóban, $(m_{i,j},a) = 1$ pontosan akkor igaz, ha $m_{i,j}$-nek az a szerinti maradéka relatív prím a-hoz, márpedig a $0,1,\ldots,a-1$ számok között épp $\varphi(a)$ ilyen van.

Összefoglalva az eddigieket: M-ben a b-hez relatív prím elemek $\varphi(b)$ darab oszlopot töltenek meg (hiszen ennyi darab j-re teljesül $(j,b) = 1$) és minden ilyen oszlopban $\varphi(a)$ darab a-hoz relatív prím elem van. Ebből pedig valóban következik, hogy összesen $\varphi(a) \cdot \varphi(b)$ darab a-hoz és b-hez is relatív prím elem van M-ben. ◇

A fenti lemmával analóg állítás nyilván a kettőnél több tagú $a_1 \cdot a_2 \cdot \ldots \cdot a_k$ szorzatra is érvényes, ha az a_i-k közül bármelyik kettő relatív prím. Valóban, először a_1-re és a_2-re alkalmazva a fenti lemmát kapjuk, hogy $\varphi(a_1 \cdot a_2) = \varphi(a_1) \cdot \varphi(a_2)$. Mivel $(a_1 \cdot a_2, a_3) = 1$ (ugyanis sem a_1, sem a_2 prímtényezős felbontásában nem szerepel az a_3 prímtényezős felbontásában előforduló prím, így $a_1 \cdot a_2$-ében sem szerepelhet), ezért $\varphi(a_1 \cdot a_2 \cdot a_3) = \varphi(a_1) \cdot \varphi(a_2) \cdot \varphi(a_3)$ következik ismét a fenti lemmából. Hasonlóan folytatva kapjuk az $a_1 \cdot a_2 \cdot \ldots \cdot a_k$-ra vonatkozó állítást.

Ha pedig a kapott összefüggést az $a_1 = p_1^{\alpha_1}, a_2 = p_2^{\alpha_2}, \ldots, a_k = p_k^{\alpha_k}$ számokra alkalmazzuk (amelyek közül nyilván bármely kettő relatív prím), akkor a már belátott $\varphi(p^\alpha) = p^\alpha - p^{\alpha-1}$ állítást felhasználva valóban a tétel állítását kapjuk. □

2.5.2. Redukált maradékrendszer

A 2.5.1. Állítás következménye, hogy bármely $0 \leq c \leq m-1$ esetén az m-vel osztva c maradékot adó számok közül vagy mind relatív prím m-hez, vagy egyikük sem az. Ha tehát a lehető legtöbb számból álló R halmazt akarjuk megalkotni úgy, hogy R minden tagja relatív prím legyen m-hez, de semelyik kettő ne adjon azonos maradékot m-mel osztva, akkor minden m-hez relatív prím c maradékra egyetlen m-mel osztva c maradékot adó számot vehetünk R-be – és így nyilván $\varphi(m)$ méretű R halmazt kapunk. Az ilyen tulajdonságú halmazoknak ad nevet az alábbi definíció.

2.5.5. Definíció. *Az $R = \{c_1, c_2, \ldots, c_k\}$ számhalmaz* redukált maradékrendszer modulo *m, ha a következő feltételeknek eleget tesz:*

(i) $(c_i, m) = 1$ *minden $i = 1,2,\ldots,k$ esetén;*

(ii) $c_i \not\equiv c_j \pmod{m}$ *bármely $i \neq j$, $1 \leq i,j \leq k$ esetén;*

(iii) $k = \varphi(m)$.

Így például modulo 10 redukált maradékrendszer az $\{1,3,7,9\}$ halmaz, de ugyanez elmondható a $\{21,43,67,89\}$ vagy az $\{1,-1,3,-3\}$ halmazról is. Az alábbi állítás lesz a szakasz fő eredményét jelentő Euler-Fermat tétel bizonyításának a legfontosabb segédeszköze.

2.5.6. Állítás. *Legyen $R=\{c_1,c_2,\ldots,c_k\}$ redukált maradékrendszer modulo m és legyen a a tetszőleges egész, amelyre $(a,m)=1$. Ekkor az $R'=\{a\cdot c_1,a\cdot c_2,\ldots,a\cdot c_k\}$ halmaz szintén redukált maradékrendszer modulo m.*

Bizonyítás: Azt kell megmutatnunk, hogy a 2.5.5. Definíció feltételei teljesülnek R'-re, ha R-re igazak. Ezek közül *(i)* a számelmélet alaptételéből következik, hiszen $a\cdot c_i$ és m prímtényezős felbontásaiban nem lehet közös prím, ha sem a, sem c_i prímfelbontásában nincs az m felbontásában is szereplő prím. A *(ii)* feltétel bizonyításához tegyük fel, hogy $a\cdot c_i\equiv a\cdot c_j \pmod m$ valamely $1\le i,j\le k$ esetben; ekkor a kongruencia mindkét oldalát a-val osztva a $c_i\equiv c_j \pmod m$ kongruenciát kapjuk, hiszen $(a,m)=1$ miatt (a 2.3.4. Tétel szerint) osztáskor a modulus nem változik. Mivel R-re teljesül a *(ii)* feltétel, ez valóban csak az $i=j$ esetben fordulhat elő. Végül R és R' elemszáma nyilván azonos, így *(iii)* is teljesül R'-re. □

2.5.3. Az Euler-Fermat tétel

Ennyi előkészítés után végül kimondhatjuk a szakasz legfontosabb tételét.

2.5.7. Tétel. (Euler-Fermat tétel)
Ha az a és $m\ge 2$ egészekre $(a,m)=1$, akkor $a^{\varphi(m)}\equiv 1 \pmod m$.

Bizonyítás: Legyen $R=\{c_1,c_2,\ldots,c_k\}$ tetszőleges redukált maradékrendszer modulo m. Mivel $(a,m)=1$, a 2.5.6. Állítás szerint $R'=\{a\cdot c_1,a\cdot c_2,\ldots,a\cdot c_k\}$ szintén redukált maradékrendszer modulo m. Ebből következik, hogy R és R' elemei párba állíthatók úgy, hogy a párok tagjai kongruensek legyenek modulo m (hiszen minden, az m-hez relatív prím c maradékra R és R' is pontosan egy m-mel osztva c maradékot adó számot tartalmaz). Ebből (a 2.3.3. Tétel *(iii)* tulajdonságát használva) következik, hogy R és R' elemeit összeszorozva egymással modulo m kongruens eredményeket kapunk:

$$c_1\cdot c_2\cdot\ldots\cdot c_k\equiv(a\cdot c_1)\cdot(a\cdot c_2)\cdot\ldots\cdot(a\cdot c_k) \pmod m.$$

A jobb oldalt átrendezve és felhasználva, hogy $k=\varphi(m)$:

$$c_1\cdot c_2\cdot\ldots\cdot c_k\equiv a^{\varphi(m)}\cdot c_1\cdot c_2\cdot\ldots\cdot c_k \pmod m.$$

Mivel $(c_i,m)=1$, ezért (a számelmélet alaptételéből következően, hasonlóan a 2.5.6. Állítás bizonyításában írtakhoz) $(c_1\cdot c_2\cdot\ldots\cdot c_k,m)=1$ is igaz.Ezért a fenti kongruencia mindkét oldalát $(c_1\cdot c_2\cdot\ldots\cdot c_k)$-val osztva a modulus nem változik, így éppen a tétel állítását kapjuk. □

Érdemes megfigyelni, hogy a fenti tételt használva új bizonyítást kaphatunk a 2.4.2. Tételből a megoldhatóságra vonatkozó feltétel elégségességére, legalábbis az $(a,m)=1$ esetben: valóban, ekkor $x=a^{\varphi(m)-1}\cdot b$ megoldása az $a\cdot x\equiv b \pmod m$ lineáris kongruenciának.

A 2.5.7. Tételt valójában a 18. században élt *Leonhard Euler* bizonyította be. A nála egy évszázaddal korábban élt *Pierre de Fermat-tól* a tétel alábbi, önmagában is fontos speciális esete származik.

2.5.8. Következmény. („Kis" Fermat-tétel)
Ha p prím és a tetszőleges egész, akkor $a^p\equiv a \pmod p$.

Bizonyítás: A tétel állítása magától értetődő, ha $p|a$: valóban, ekkor $p|a^p$ is nyilván igaz, így $a^p\equiv 0\equiv a \pmod p$. Ha viszont $p\nmid a$, akkor $(a,p)=1$ is igaz (mert p prím), így alkalmazhatjuk a 2.5.7. Euler-Fermat tételt a-ra és p-re. Ez $\varphi(p)=p-1$ miatt a következőt állítja: $a^{p-1}\equiv 1 \pmod p$. A kapott kongruencia mindkét oldalát a-val szorozva épp a tétel állítását kapjuk. □

2.5.9. Feladat. Milyen maradékot ad
a) 11^{111} 63-mal osztva; b) $51^{41^{32}}$ 140-nel osztva?

Megoldás: a) $(11,63)=1$, így alkalmazható az Euler-Fermat tétel az $a=11$, $m=63$ választással. $\varphi(63)=\varphi(3^2\cdot 7)=(3^2-3^1)\cdot(7^1-7^0)=6\cdot 6=36$ a 2.5.3. Tétel szerint, így az Euler-Fermat tételből $11^{36}\equiv 1 \pmod{63}$. Mindkét oldalt a harmadik hatványra emelve (2.3.3. Tétel, *(iv)* állítás): $11^{108}=\left(11^{36}\right)^3\equiv 1^3=1 \pmod{63}$. Így a feladat megoldásához már csak 11^3 63-as maradékát kell meghatároznunk: $11^2=121\equiv -5 \pmod{63}$, így $11^3=11^2\cdot 11\equiv(-5)\cdot 11=-55\equiv 8 \pmod{63}$. Tehát $11^{111}=11^{108}\cdot 11^3\equiv 1\cdot 8=8 \pmod{63}$ vagyis a keresett maradék a 8.

b) Most az $a=51$, $m=140$ választással alkalmazzuk az Euler-Fermat tételt. $51=3\cdot 17$ és $140=2^2\cdot 5\cdot 7$, így $\varphi(140)=(2^2-2^1)(5^1-5^0)(7^1-7^0)=2\cdot 4\cdot 6=48$ és $(51,140)=1$. Az Euler-Fermat tételből tehát $51^{48}\equiv 1 \pmod{140}$ adódik. Első látásra nem világos, hogy ez hogyan segít a megoldásban, de felhasználhatjuk az a) feladat megoldásának ötletét: tetszőleges $k\geq 1$ egészre mindkét oldalt a k-adikra emelve az $51^{48k}\equiv 1 \pmod{140}$ kongruenciát kapjuk. Bár a feladatban szereplő 41^{32} nyilván nem $48k$ alakú, de (ismét az a) feladathoz hasonlóan) közel eshet egy 48-cal osztható számhoz. Ezért a továbblépéshez azt a kérdést kell feltennünk, hogy 41^{32} milyen maradékot ad 48-cal osztva? Ezt ismét az Euler-Fermat tétellel válaszoljuk meg: $\varphi(48)=\varphi(2^4\cdot 3)=(2^4-2^3)(3-1)=16$ és $(41,48)=1$, így $41^{16}\equiv 1 \pmod{48}$ adódik az Euler-Fermat tételből. Ezt négyzetre emelve: $41^{32}\equiv 1 \pmod{48}$, vagyis $41^{32}=48k+1$ egy alkalmas k-ra. Ezt felhasználva:

$$51^{41^{32}}=51^{48k+1}=51^{48k}\cdot 51\equiv 1\cdot 51=51 \pmod{140},$$

vagyis a keresett maradék az 51. □

2.6. Számelméleti algoritmusok

Mi köze vajon az eddig megismert számelméleti ismereteknek a biztonságos internetes kommunikációhoz? A *kriptográfia*, vagyis a titkosírás, a rejtjelezés tudománya közel egyidős az emberi kultúrával, az írással. Különböző katonai, hírszerzési, diplomáciai alkalmazásokban már az ókortól kezdve alapvető igény volt, hogy egy üzenetet csak a címzett tudjon elolvasni. Manapság, az internetes kommunikáció korában pedig ez a terület már sokkal komplexebb kérdéseket vet föl és az idevágó módszerek széles körben, egészen hétköznapi helyzetekben is alkalmazást nyernek.

A kódolás és a dekódolás valójában nem más, mint két függvény alkalmazása, amelyek egymásnak az inverzei: az x üzenetre a feladó alkalmazza a C kódoló függvényt és a $C(x)$ kódolt üzenetet továbbítja; ha ez eljutott a címzetthez, akkor az alkalmazza rá a $D = C^{-1}$ dekódoló függvényt és így visszakapja a $D(C(x)) = x$ üzenetet. A kriptográfia feladata a megfelelő C és D inverz függvénypárok megtalálása.

A rejtjelező eljárások működéséhez természetesen hozzátartozik, hogy a C és D függvényeket titokban kell tartani – legalábbis így volt ez egészen az 1970-es évek végéig. Ekkor forradalmi változás állt be a kriptográfiában: felfedezték a *nyilvános kulcsú titkosítást*. Ennek a lényege az, hogy a C kódoló függvényt közzéteszik, csak a D dekódoló függvényt tartják titokban. Ha például Aliz és Bonifác leveleznek, akkor Aliz megkeresi a Bonifác által előzetesen nyilvánosságra hozott C_B kódolófüggvényt és ezzel kódolja a neki szánt üzeneteket. Bonifác viszont Aliz nyilvános C_A kódoló függvényét alkalmazza a neki szánt üzenetekre. A kapott üzeneteket mindketten a saját, titkosan kezelt D_A, illetve D_B dekódoló függvényükkel fejtik meg.

A nyilvános kulcsú titkosítás gondolata első hallásra valószínűleg képtelenségnek tűnik: hogyan is lehetne a $D = C^{-1}$ függvényt titokban tartani, ha a C nyilvános? Például egy magyar-navajo szótárból is készíthető kellő türelemmel navajo-magyar szótár. Miközben ez persze elvileg igaz, a gyakorlatban mégsem szívesen vállalkoznánk egy navajo nyelvű szöveg magyarra fordítására egy magyar-navajo szótár birtokában. Hasonló a helyzet a nyilvános kulcsú titkosítás esetében is: az ezekben alkalmazott C kódoló függvények értelmezési tartománya olyan hatalmas, hogy még a legmodernebb szuperszámítógépekkel is évmilliárdokig tartana minden egyes elemét végigpróbálgatni. Ezért bátran közzétehető a C függvény hozzárendelési szabálya – feltéve, hogy ebből illetéktelenek nem tudják kiszámítani a D inverz függvény hozzárendelési szabályát.

A modern kriptográfiai eljárások első lépése az, hogy a kódolandó üzenetet számok – mégpedig a fentieknek megfelelően hatalmas számok – sorozatává alakítják. Persze a számítógépes alkalmazásokban ez eleve adott, hiszen bármi legyen is az „üzenet”, az egy bitsorozatként tárolódik. Ma a legbiztonságosabbnak tartott módszerek 2048 bites titkosítást használnak – ami azt jelenti, hogy a C és D függvények értelmezési tartománya egy, a kettes számrendszerben 2048 jegyű N számnál kisebb egészekből áll. N nagyságrendje tehát durván 10^{616} – vagyis a 10-es számrendszerben 600-nál több jegyű számokról van szó. (Összehasonlításképpen: a világegyetemben körülbelül 10^{80} proton van és az ősrobbanás óta eltelt idő „csak” $4,354 \cdot 10^{17}$ másodperc.)

A nyilvános kulcsú titkosítást az teszi működőképessé, hogy ezek a felfogha-

tatlanul hatalmas számok is (a számítógépes alkalmazások mértéke szerint) nevetségesen kis helyen tárolhatók. Persze nem elég tárolni ezeket a számokat: számos alapvető műveletet kell rajtuk végezni, mégpedig nagyon gyorsan.

2.6.1. Számelméleti algoritmusok hatékonysága

Mielőtt egy adott számítási feladatra (legyen az számelméleti vagy akár más jellegű) algoritmust terveznénk, fel kell tennünk a kérdést: milyen futásidejű megoldást tartunk elfogadhatónak? Ez pedig messze nem könnyen megválaszolható kérdés, hiszen egyes feladatoknál egy töredékmásodperc is túl hosszú lehet, másoknál akár hónapokig is futhat a program. Ráadásul ugyanaz a programkód egy nagyobb teljesítményű számítógépen sokkal gyorsabban lefut. De ezeknél is fontosabb szempont, hogy bármilyen ügyes programot is írunk, az nagyobb bemenetre nyilván tovább fog futni – vagyis a futásidő (még a hardver rögzítése után is) nem egy számnak, hanem egy függvénynek tekintendő, amely a bemenet méretétől függ.

Vizsgáljuk meg például a következő feladatot: határozzuk meg egy N szám prímtényezős felbontását. (Ezt a feladatot szokás *prímfaktorizációnak* is nevezni.) Erre a problémára már az iskolai tanulányainkból ismerünk egy egyszerű algoritmust (amelyet néha „akasztófa módszer" néven emlegetnek): 2-től egyesével fölfelé lépkedve minden egésszel elosztjuk N-et és ha megtaláljuk egy p osztóját, akkor az eljárást megismételjük $\frac{N}{p}$-re (de a lépkedést elég p-től folytatni). Ha pedig egészen $\sqrt{N}$-ig lépkedve nem találtuk meg N egy osztóját sem, akkor N prím (hiszen ha összetett, akkor $N = a \cdot b$ valamely $a \leq b$ valódi osztóira, amiből $a \leq \sqrt{N}$). Ez az eljárás kétségtelenül helyesen működik és viszonylag kis számokra akár kézzel is elvégezhető, számítógéppel pedig akár 10-20 jegyű számokra is. De ha N például egy 70 jegyű prím, akkor az eljárás durván 10^{35} osztást végez, mire ezt megállapítja – ez pedig a jelenlegi legmodernebb szuperszámítógépeknek is tovább tartana az ősrobbanás óta eltelt időnél. Így ezt az algoritmust a többszáz jegyű számokat használó gyakorlati alkalmazások szempontjából használhatatlannak kell minősítenünk. Persze ettől még létezhetne ugyanerre a feladatra egy (esetleg nagyságrendekkel komplikáltabb) másik algoritmus is, amellyel az ilyen hatalmas számok is elfogadható időn belül prímtényezőkre bonthatók – ilyen módszer azonban nem ismert. Ha viszont csak azt kell eldöntenünk egy N számról, hogy az prím-e, akkor (mint azt később látni fogjuk) erre a feladatra már vannak a többszáz jegyű számok körében is jól alkalmazható, a fentinél sokkal ravaszabb algoritmusok.

Mikor tartsunk tehát egy **A** algoritmust hatékonynak? Erre a kérdésre a számítástudomány szűk évszázados története alatt kialakult egy olyan válasz, amely bár sok kérdést nyitva hagy, mégis mind a terület elméleti kutatói, mind a gyakorlati szakemberek számára többé-kevésbé elfogadható: akkor, ha *polinomiális futásidejű*. Ez a következőt jelenti: jelölje $f(n)$ azt, hogy egy n méretű bemeneten az **A** legföljebb hány lépés megtétele után áll le. **A** akkor polinomiális futásidejű, ha léteznek olyan c és k rögzített konstansok, hogy $f(n) \leq c \cdot n^k$ minden n-re fennáll. Ha tehát például az **A** algoritmus minden n méretű bemeneten legföljebb $c \cdot n^2$, vagy $c \cdot n^3$, vagy akár csak $c \cdot n^{10}$ lépés után biztosan megáll (ahol c valamilyen fix kons-

tans), akkor polinomiális (és így hatékonynak tekinthető). Ha azonban előfordulhat, hogy **A** egy n méretű bemeneten például 2^n, vagy akár csak $1,01^n$ lépést tesz, akkor nem polinomiális (mert elegendően nagy n-re $1,01^n > c \cdot n^k$ teljesül, bárhogyan is választjuk c-t és k-t). Közismert például a tetszőleges számhalmaz nagyság szerinti rendezésére szolgáló *buborékrendezés* módszere; könnyű végiggondolni, hogy ez n szám rendezéséhez legfeljebb $\frac{1}{2}n(n-1)$ összehasonlítást végez, így (a szükséges elemcserékkel és egyéb elemi lépésekkel együtt is) legfeljebb $c \cdot n^2$ lépésszámú valamilyen c konstansra, így polinomiális (bár számos, ennél hatékonyabb módszer is létezik ugyanerre a feladatra). Polinomiális algoritmus (megfelelő implementáció esetén) az 1. Fejezetben megismert Gauss-elimináció is.

Feltétlen meg kell jegyeznünk, hogy a polinomiális futásidejű algoritmus fogalmának fenti leírása semmiképp nem tekintendő precíz matematikai definíciónak. Nem tisztáztuk ugyanis, pontosan hogyan is mérjük a bemenet méretét, hogy mit is értünk egy algoritmus egy lépése alatt – sőt, még azt sem, mit nevezünk algoritmusnak. Ezek a kérdések a legszigorúbb matematikai elvárásokat is kielégítő módon megválaszolhatók (ezekkel az informatikus képzés későbbi tárgyai foglalkoznak). A mostani céljainknak megfelel, ha algoritmus alatt egyszerűen egy (például *C* nyelven írt) programkódot értünk, a bemenet mérete helyett az annak tárolásához szükséges memória méretére, a lépésszám helyett pedig (egy tetszőleges, rögzített hardveren mért) futásidőre gondolunk.

Térjünk most vissza a számelméleti problémákhoz: hogyan értendő a polinomialitás az ide tartozó algoritmusok esetében? Vizsgáljuk meg újra a fent már felidézett akasztófa módszert az N prímfaktorizációjára. Itt a bemenet mérete az N számjegyeinek száma – jelölje ezt n –, hiszen ennyi helyet foglalunk a memóriában az N tárolásakor. Ekkor tehát $10^{n-1} \leq N < 10^n$, így $n-1 \leq \log_{10} N < n$; vagyis $n = \lfloor \log_{10} N \rfloor + 1$. Az algoritmus viszont akár $\sqrt{N}$ osztást is végezhet, ha N prím. Így a lépésszám exponenciális függvénye lehet a bemenet méretének (hiszen $\sqrt{N} \approx (\sqrt{10})^n$), ezért az akasztófa módszer nem polinomiális – összhangban a fenti megállapításunkkal, hogy nagy N-ekre használhatatlan.

Figyeljük meg, hogy bár a fenti számítás feltételezte, hogy N a tízes számrendszerben van felírva, ez elvi szempontból érdektelen. Ha N-et például binárisan írjuk fel, akkor $\lfloor \log_2 N \rfloor + 1$ jegyű. De $\log_2 N = \log_2 10 \cdot \log_{10} N \approx 3,32 \cdot \log_{10} N$ miatt a decimális és a bináris jegyek száma csak egy konstans szorzóban különbözik – ami a lépésszám polinomialitásának szempontjából érdektelen. A továbbiakban ezért a szemléletesség kedvéért mindig feltesszük, hogy a bemenetet alkotó számok decimálisan vannak megadva – bár a számítógépes alkalmazásoknál nyilván nem az a helyzet.

Összefoglalva a fentieket: a számelméleti algoritmusoknál a bemenet méretét mindig a bemenetet adó számok összes számjegyeinek számával mérjük, ami lényegében azonosítható a számok (például 10-es alapú) logaritmusával. Egy algoritmust pedig akkor tekintünk (elméletben és többnyire a gyakorlatban is) hatékonynak, ha n jegyű számokon legfeljebb $c \cdot n$, vagy legfeljebb $c \cdot n^2$, vagy általában legfeljebb $c \cdot n^k$ lépést tesz (valamilyen fix k-ra).

A fejezet hátralévő részében bemutatandó algoritmusok kapcsán mindig érde-

mes szem előtt tartani, hogy ezeket a gyakorlatban többszáz jegyű számokra alkalmazzák – noha a módszereket természetesen csak 2-3 jegyű számokon illusztráljuk.

2.6.2. Alapműveletek

A számelmélet körébe tartozó algoritmikus feladatok közül a legegyszerűbbek és legalapvetőbbek az alapműveletek. Például az összeadás feladata így írható le:

Bemenet: a és b egészek;
Kimenet: $a+b$.

Ezzel analóg a kivonás és a szorzás feladata, osztás alatt pedig maradékos osztást értünk: itt $\frac{a}{b}$ egészrészét és a-nak a b szerinti osztási maradékát kell kiszámítani; az előbbit a továbbiakban $\lfloor \frac{a}{b} \rfloor$, az utóbbit $(a \mod b)$ fogja jelölni.

Erre a négy feladatra szerencsére ismerünk hatékony algoritmusokat, mégpedig már alsó tagozatból: az „írásbeli" összeadás, kivonás, stb. ilyenek. (Az írásbeli osztás esetében ugyan az alsó tagozatban tanult módszer a következő jegy „megtippelését" írja elő, de ez ismételt visszaszorzások segítségével könnyen kiváltható a keresett jegy módszeres meghatározásával.) Ha a és b jegyeinek száma k, illetve ℓ, akkor az írásbeli összeadás és kivonás $c \cdot (k+\ell)$, a szorzás és az osztás pedig $c \cdot k \cdot \ell$ lépésszámú algoritmusok (valamilyen c konstansra); így ha $n = k+\ell$ jelöli a bemenet méretét, akkor az összeadás és kivonás futásidejére a $c \cdot n$, a szorzáséra és az osztáséra pedig a $c \cdot n^2$ felső becslés adódik. Így ezek nem csak polinomiális, de még ezen belül is igen hatékony algoritmusok – és természetesen a tízes mellett tetszőleges alapú számrendszerben működnek. (Érdemes megjegyezni, hogy a szorzás és az osztás esetében léteznek a $c \cdot n^2$ futásidőnél is gyorsabb algoritmusok. Bár ezek jóval komplikáltabbak és a gyakorlatban csak óriási számok esetében lehet velük futásidőt megtakarítani, épp a kriptográfiai alkalmazások számára már hasznosak.)

Vizsgáljuk meg a hatványozás feladatát is:

Bemenet: a és b egészek;
Kimenet: a^b.

Szemben a négy alapművelettel, a hatványozásra már nem adható hatékony (polinomiális futásidejű) algoritmus – mégpedig abból az egyszerű okból, hogy még a kimenet kiírása is túl sok ideig tartana. Valóban, ha b jegyeinek száma n (a tízes számrendszerben), akkor még az $a = 2$ esetben is 2^b jegyeinek száma (lényegében) $\log_{10} 2^b = b \cdot \log_{10} 2 \geq \log_{10} 2 \cdot 10^{n-1} > 0{,}03 \cdot 10^n$, vagyis 2^b jegyeinek száma exponenciális függvénye b jegyei számának. (Ha b például 100 jegyű, akkor 2^b jegyeinek száma $3 \cdot 10^{98}$-nál több, így 2^b kiírása még akkor is lehetetlen volna, ha a világegyetemben található minden protonra ráírhatnánk a kimenet egy számjegyét.)

2.6.3. Hatványozás modulo m

A nyilvános kulcsú titkosításhoz alapvető lesz, hogy bár a^b-t kiszámítani a fentiek szerint reménytelen, annak egy adott m szerinti osztási maradékát mégis meg tudjuk határozni. A következő feladatot fogjuk tehát vizsgálni:

Bemenet: a, b és m egészek;
Kimenet: $a^b \mod m$ (vagyis a^b osztási maradéka m szerint).

Ennél a feladatnál már nem okoz problémát, hogy a kimenet túl nagy volna, hiszen az m-nél kisebb. Másrészt viszont nem járható út a^b kiszámítása, majd annak m-mel való maradékos osztása, hiszen láttuk, hogy ez a terv már az első lépésénél meghiúsul. Az exponenciális tárigény problémáján segít, ha az $a, a^2, \ldots, a^b$ hatványok m szerinti maradékát sorra kiszámítjuk (úgy, hogy mindig az előző maradék a-szorosának m szerinti maradékát vesszük), de ez az eljárás is használhatatlanul lassú volna: $b-1$ darab ilyen lépést kellene tennünk, ami exponenciális lépésszámú algoritmust jelentene.

Szerencsére létezik hatékony algoritmus is a modulo m hatványozás feladatára: az *ismételt négyzetre emelések módszere*. Ezt először egy példán illusztráljuk: meghatározzuk 13^{53} maradékát 97-tel osztva. Az alábbi számolásban mindegyik sor az előző négyzetre emelésével keletkezik – amint azt a módszer neve is mutatja:

$$\begin{aligned}
13^1 &\equiv 13 && \pmod{97}\\
13^2 = 169 &\equiv 72 && \pmod{97}\\
13^4 = (13^2)^2 \equiv 72^2 = 5184 &\equiv 43 && \pmod{97}\\
13^8 = (13^4)^2 \equiv 43^2 = 1849 &\equiv 6 && \pmod{97}\\
13^{16} = (13^8)^2 \equiv 6^2 = 36 && && \pmod{97}\\
13^{32} = (13^{16})^2 \equiv 36^2 = 1296 &\equiv 35 && \pmod{97}
\end{aligned}$$

Látszik, hogy ezzel a módszerrel a 13-nak az $1, 2, 4, 8, \ldots$, vagyis a 2-hatvány kitevőjű hatványait tudjuk közvetlenül meghatározni. A sort 13^{32} után tovább folytatni értelmetlen volna, hiszen 13^{64} már nagyobb a kiszámítandó 13^{53}-nál. De még így is kérdés, hogy a fenti számítások hogyan segítenek 13^{53} 97-es maradékának a meghatározásában? A válasz egyszerű: $13^{53} = 13^{1+4+16+32} = 13^1 \cdot 13^4 \cdot 13^{16} \cdot 13^{32}$, így ennek a négy tagnak a 97-es maradékait összeszorozva 13^{53}-nal kongruens számot kapunk modulo 97. Persze ezeket a szorzásokat is lépésenként végezzük és az eredményeknek – a fentiekhez hasonlóan – mindig csak a 97-es maradékát tartjuk meg:

$$\begin{aligned}
13^5 = 13^1 \cdot 13^4 &\equiv 13 \cdot 43 = 559 \equiv 74 && \pmod{97}\\
13^{21} = 13^5 \cdot 13^{16} &\equiv 74 \cdot 36 = 2664 \equiv 45 && \pmod{97}\\
13^{53} = 13^{21} \cdot 13^{32} &\equiv 45 \cdot 35 = 1575 \equiv 23 && \pmod{97}
\end{aligned}$$

Ezzel tehát megkaptuk a végeredményt: $13^{53} \equiv 23 \pmod{97}$. Persze felvetődik a kérdés: hogyan működik ez a módszer $a^b \mod m$ meghatározására tetszőleges a, b és m esetén? Hiszen a fenti példában „szerencsénk volt”: $53 = 1+4+16+32$, vagyis a $b = 53$ kitevő előállt nála kisebb 2-hatványok összegeként. Azonban ebben a „szerencsében” természetesen minden b kitevő részesül: ez következik b-nek a 2-es számrendszerben való felírhatóságából. Például a $b = 53$ bináris alakja 110101, ami épp azt jelenti, hogy $53 = 2^0 + 2^2 + 2^4 + 2^5$.

Az algoritmus tehát az általános esetben ismételt négyzetre emelésekkel meghatározza a^t maradékát m szerint minden $t \le b$ 2-hatványra; vagyis a $t = 2^k$ kitevőkre, ahol $k = 0, 1, \ldots, \lceil \log_2 b \rceil$. Majd az így kapott maradékokból állítja elő a^b maradékát: ez a b bináris alakjában az 1-es jegyeknek megfelelő 2-hatványoknak, mint

kitevőknek megfelelő maradékok szorzatának m szerinti maradéka. Valójában az utóbbi számítást érdemes párhuzamosan végezni a négyzetre emelésekkel, hogy ne kelljen az azok során meghatározott maradékokat tárolni. (A fent bemutatott példa ezzel tehát annyiban módosul, hogy 13^5 97-es maradékát nem utólag számítjuk ki, hanem közvetlenül a 13^4 maradékának meghatározása után és hasonlóan, 13^{21} és 13^{53} maradékát közvetlenül a 13^{16}, illetve a 13^{32} maradéka után kapjuk meg).

Az algoritmus általános leírását elegendő a $0 < a < m$ esetre megadni (mert ha nem ez a helyzet, akkor a-t helyettesíthetjük az m szerinti maradékával). Nem feltételezzük, hogy b bináris alakja eleve adott, ezt is párhuzamosan határozzuk meg a négyzetre emelésekkel. Ez a következő egyszerű észrevételen alapul: b bináris alakjának utolsó jegye b 2-es maradékával egyenlő, a többi jegy pedig megegyezik $\lfloor \frac{b}{2} \rfloor$ bináris alakjával. Ezek alapján végül is az eljárás a következőképpen írható le.

AZ ISMÉTELT NÉGYZETRE EMELÉSEK MÓDSZERE ($a^b \bmod m$ KISZÁMÍTÁSÁRA)

Bemenet: a, b és m (amelyekre $0 < a < m$ és $b \geq 1$ teljesül)

0. lépés. $c \leftarrow 1$

1. lépés. Ha b páratlan, akkor: $c \leftarrow c \cdot a \mod m$

2. lépés. $b \leftarrow \lfloor \frac{b}{2} \rfloor$

3. lépés. Ha $b = 0$, akkor: PRINT „$a^b \mod m =''$, c; STOP.

4. lépés. $a \leftarrow a^2 \mod m$

Folytassuk az **1. lépésnél**.

Érdemes a $13^{53} \bmod 97$ meghatározását a fenti leírásnak megfelelően megismételni. Az alábbi táblázatban a, b és c értékének változását követhetjük nyomon (tehát az $a = 13$, $b = 53$, $m = 97$ bemenő adatok esetén), továbbá k számlálja, hogy a ciklust hányadszorra hajtjuk végre.

k	a	b	c
0	13	53	1
1	72	26	13
2	43	13	13
3	6	6	74
4	36	3	74
5	35	1	45
6	–	0	23

Látható, hogy a táblázat a és c oszlopában sorra ugyanazok az értékek keletkeztek (és azonos számolások eredményeképpen), mint amiket a korábbi számításban kaptunk. A b oszlopban kapott értékek megfelelnek a bináris alak kiszámításának a bemenetként kapott b-hez (az 1-es jegyek a páratlan elemeknek felelnek meg).

Az algoritmus működésének a helyességét a következőképpen is megindokolhatjuk. Jelölje a_k, b_k és c_k az a, b és c változók aktuális értékét a ciklus k-adik végrehajtása után (illetve a_0 és b_0 a bemenetként kapott a és b értékeket és $c_0 = 1$). Továbbá jelölje r a kiszámítandó $(a^b \mod m)$ maradékot. Ekkor k-ra vonatkozó teljes indukcióval könnyen megmutatható, hogy az algoritmus futása során minden ciklus végén (vagyis a 4. lépés végrehajtása után) igaz, hogy $r \equiv a_k^{b_k} \cdot c_k \pmod m$. Valóban, ez a $k = 0$ esetben magától értetődő. Tegyük most fel, hogy valamely k-ra az állítás már igaz. Ha b_k páros, akkor $r \equiv a_k^{b_k} \cdot c_k = (a_k^2)^{\frac{b_k}{2}} \cdot c_k \pmod m$ miatt az $a_{k+1} \equiv a_k^2 \pmod m$, $b_{k+1} = \frac{b_k}{2}$ és $c_{k+1} = c_k$ választással valóban igaz lesz az $r \equiv a_{k+1}^{b_{k+1}} \cdot c_{k+1} \pmod m$ állítás, ami összhangban van az algoritmus működésével. Hasonlóan, ha b_k páratlan, akkor $r \equiv a_k^{b_k} \cdot c_k = (a_k^2)^{\frac{b_k-1}{2}} \cdot a_k \cdot c_k \pmod m$ miatt az $a_{k+1} \equiv a_k^2 \pmod m$, $b_{k+1} = \frac{b_k-1}{2}$ és $c_{k+1} \equiv a_k \cdot c_k \pmod m$ választással lesz igaz az $r \equiv a_{k+1}^{b_{k+1}} \cdot c_{k+1} \pmod m$ állítás, ismét összhangban az algoritmus működésével. Ezzel tehát az $r \equiv a_k^{b_k} \cdot c_k \pmod m$ állítást minden k-ra beláttuk. Mivel az eljárás leállásakor $b_k = 0$, ebből $r \equiv a_k^0 \cdot c_k = c_k \pmod m$. Így az algoritmus valóban a helyes r értéket adja meg.

Végül megmutatjuk, hogy az ismételt négyzetre emelések módszere hatékony algoritmus. Ciklusonként legfőljebb két szorzást, két maradékos osztást és egy felezést kell elvégeznünk, a szorzásokat mindig m-nél kisebb, az osztásokat pedig m^2-nél kisebb számokon. A 2.6.2. szakaszban már tisztáztuk, hogy ezek hatékonyan elvégezhető műveletek. Ezen kívül már csak azt kell észrevennünk, hogy a ciklust $(\lfloor \log_2 b \rfloor + 1)$-szer hajtjuk végre, hiszen ennyi a b bináris felírásában a jegyek száma. Azt pedig már láttuk, hogy $\log_2 b$ csak egy konstans szorzóban tér el b jegyeinek a számától (függetlenül attól, hogy b milyen alapú számrendszerben van felírva). Mindezekből következik, hogy az algoritmus valóban polinomiális lépésszámú. (Érdemes azonban megjegyezni, hogy ha a szorzásokat és osztásokat egyszerűen a 2.6.2. szakaszban említett „írásbeli” műveletekkel végezzük, akkor az ismételt négyzetre emelések módszerének lépésszámára a $c \cdot n^3$ felső becslés adódik, ahol n a bemenet mérete – vagyis a, b és m összes jegyeinek száma – és c egy konstans. Ez a lépésszám ugyan valóban polinomiális, de a kriptográfiai alkalmazások számára már használhatatlanul lassú volna. Szerencsére a módszert néhány ügyes algebrai trükkel jelentősen fel lehet gyorsítani, de ezekre itt nem térünk ki.)

2.6.4. A legnagyobb közös osztó kiszámítása

Két adott szám legnagyobb közös osztója kiszámítható a prímtényezős felbontásukból (a 2.1.5 Tétel alapján), de – amint azt már a 2.6.1. szakaszban említettük – a prímfaktorizációra nem ismert hatékony (vagyis polinomiális futásidejű) algoritmus. Szerencsére létezik a legnagyobb közös osztó kiszámítására egy ennél sokkal hatékonyabb módszer is: az *Euklideszi algoritmus*. Ez a matematikatörténet egyik első algoritmusa, szerepel már Euklidesznek az i. e. 300 körül megjelent *Elemek* című művében is. A következő feladatot vizsgáljuk tehát:

Bemenet: a és m egészek (amelyekre feltesszük, hogy $0 < a < m$);
Kimenet: (a, m) (vagyis a és m legnagyobb közös osztója).

Az Euklideszi algoritmus ismételt maradékos osztásokon alapul: az első lépésben m-et osztjuk a-val, a másodikban a-t a kapott maradékkal, stb., az i-edik lépésben mindig az $(i-2)$-edik lépésben kapott maradékot osztjuk az $(i-1)$-edikben kapottal. Az eljárást először egy példán illusztráljuk: meghatározzuk $(567, 1238)$ értékét.

$$\begin{array}{lrl}
(1) & [1238/567]: & 1238 = 2\cdot 567 + 104 \\
(2) & [567/104]: & 567 = 5\cdot 104 + 47 \\
(3) & [104/47]: & 104 = 2\cdot 47 + 10 \\
(4) & [47/10]: & 47 = 4\cdot 10 + 7 \\
(5) & [10/7]: & 10 = 1\cdot 7 + 3 \\
(6) & [7/3]: & 7 = 2\cdot 3 + 1 \\
(7) & [3/1]: & 3 = 3\cdot 1 + 0
\end{array}$$

Az Euklideszi algoritmus kimenete mindig az utolsó nemnulla maradék; a fenti példában tehát $(1238, 567) = 1$. Általános a és m esetén az eljárás által végzett maradékos osztások sorozatát a következőképpen írhatjuk le:

$$\begin{array}{lrll}
(1) & [m/a]: & m = t_1\cdot a + r_1 & (0 \le r_1 < a) \\
(2) & [a/r_1]: & a = t_2\cdot r_1 + r_2 & (0 \le r_2 < r_1) \\
(3) & [r_1/r_2]: & r_1 = t_3\cdot r_2 + r_3 & (0 \le r_3 < r_2) \\
(4) & [r_2/r_3]: & r_2 = t_4\cdot r_3 + r_4 & (0 \le r_4 < r_3) \\
\vdots & \vdots & \vdots & \vdots \\
(k) & [r_{k-2}/r_{k-1}]: & r_{k-2} = t_k\cdot r_{k-1} + r_k & (0 \le r_k < r_{k-1}) \\
(k+1) & [r_{k-1}/r_k]: & r_{k-1} = t_{k+1}\cdot r_k + 0 &
\end{array}$$

Itt az utolsó nemnulla maradék r_k, ez tehát az algoritmus kimenete; persze azt még be kell látnunk, hogy ez valóban megegyezik a és m legnagyobb közös osztójával.

2.6.1. Állítás. *Az Euklideszi algoritmus fenti végrehajtásakor $r_k = (a, m)$.*

Bizonyítás: Az eljárás (1) lépéséből következik, hogy $m \equiv r_1 \pmod{a}$. Ebből a 2.5.1. Állítás szerint $(m, a) = (a, r_1)$ adódik. Hasonlóan, a (2) lépés miatt $a \equiv r_2 \pmod{r_1}$, amiből $(a, r_1) = (r_1, r_2)$. Folytatva a sort kapjuk, hogy $(m, a) = (a, r_1) = (r_1, r_2) = \ldots = (r_{k-1}, r_k)$. Azonban a legutolsó, $(k+1)$ lépés szerint $r_k \mid r_{k-1}$, így $(r_{k-1}, r_k) = r_k$. Ezzel tehát az állítást beláttuk. □

Természetesen be kell még látnunk, hogy az eljárás hatékony, vagyis polinomiális idejű. Ehhez csak azt kell megmutatnunk, hogy a maradékos osztások száma kellően kevés (hiszen egy ilyen lépés a 2.6.2. szakaszban írtak szerint hatékonyan végrehajtható). Az nyilvánvaló, hogy az eljárás véges sok lépésben véget ér, hiszen $a > r_1 > r_2 > \ldots$ miatt előbb-utóbb el kell érnie a 0 maradékot. Ez azonban még kevés: ha az r_i maradékok mindig csak 1-gyel csökkennének az előzőhöz képest, akkor a maradékos osztásra volna szükség, ami pedig a bemenet méretének (vagyis a és m jegyei számának) exponenciális függvénye lehet. A valóság ennél szerencsére sokkal kedvezőbb: az alábbi állítás bizonyításából kiderül, hogy az r_i maradékok kétlépésenként legalábbis megfeleződnek.

2.6.2. Állítás. *Az Euklideszi algoritmus legföljebb* $2 \cdot \lceil \log_2 a \rceil$ *maradékos osztás után megáll.*

Bizonyítás: Vizsgáljuk meg az eljárás egy tetszőleges lépését: $r_{i-2} = t_i \cdot r_{i-1} + r_i$, ahol a fentiek szerint $r_{i-2} > r_{i-1} > r_i$. Itt tehát $t_i \geq 1$ ($r_{i-2} > r_{i-1}$ miatt), amiből $r_{i-2} \geq r_{i-1} + r_i$ következik. Ebből viszont $r_{i-1} > r_i$ miatt $r_{i-2} > 2r_i$ adódik. Így az eljárás páros sorszámú soraiból az $a = r_0 > 2r_2 > 4r_4 > \ldots > 2^k \cdot r_{2k}$ becsléssort kapjuk. A $k = \lceil \log_2 a \rceil$ választással $2^k \geq a$, így indirekt feltételezve, hogy az eljárás az r_{2k} maradékkal még nem ért véget, a $0 < r_{2k} < \frac{a}{2^k} \leq 1$ ellentmondást kapnánk. □

Ebből tehát valóban következik, hogy az Euklideszi algoritmus polinomiális futásidejű (de még ezen belül is nagyon hatékony), hiszen $\log_2 a$ az a jegyei számának konstansszorosa. (Az eljárás hatékonysága tovább javítható, ha a maradékos osztás fogalmát kicsit módosítjuk: megengedjük a negatív maradékokat is és mindig a legkisebb abszolút értékű maradékot választjuk. Mivel az a-val osztható számok egymástól a távolságra követik egymást, ezért minden m-től legföljebb $\frac{a}{2}$ távolságra található a-val osztható szám. Más szóval: minden a és m esetén egyértelműen létezik egy olyan t egész és egy $-\frac{a}{2} < r \leq \frac{a}{2}$ egész, amelyre $m = t \cdot a + r$. Az Euklideszi algoritmusban a maradékos osztásokat így végezve tehát nem csak kétlépésenként feleződnek meg a maradékok, hanem minden lépésben. Így az eljáráshoz szükséges maradékos osztások száma ezzel a módosítással $\lceil \log_2 a \rceil$-ra csökkenthető.)

Az algoritmusnak további komoly előnye, hogy a futtatásakor nem szükséges tárolni az összes keletkező adatot: r_i kiszámításához elegendő r_{i-1} és r_{i-2} értékét ismerni. Így háromnál több változó értékének tárolására sosincs szükség.

EUKLIDESZI ALGORITMUS ((a,m) KISZÁMÍTÁSÁRA)

Bemenet: a és m (amelyekre $0 < a < m$ teljesül)

1. lépés. $r \leftarrow m \mod a$

2. lépés. Ha $r = 0$, akkor: PRINT „$(a,m) =''$, a; STOP

3. lépés. $m \leftarrow a$, $a \leftarrow r$

Folytassuk az **1. lépésnél**.

Az eljárás során az 1. lépés minden végrehajtásakor m a kettővel, a az eggyel korábbi maradékot jelöli (illetve kezdetben a a bemenetként kapott adatokat). Az aktuálisan kiszámolt maradék mindig r. Így ha $r = 0$ adódik, akkor a 2. lépésnek megfelelően valóban a, vagyis az utolsó nemnulla maradék az algoritmus kimenete.

2.6.5. Lineáris kongruenciák megoldása

A lineáris kongruenciák feladatával már a 2.4. szakaszban megismerkedtünk és a 2.4.2. Tételből tudjuk, hogy $a \cdot x \equiv b \pmod{m}$ akkor és csak akkor megoldható, ha $(a,m) \mid b$ és ebben az esetben a megoldások száma modulo m egyenlő (a,m)-val. Így

a lineáris kongruenciák megoldhatóságának, illetve a megoldások számának kérdését már hatékonyan meg tudjuk válaszolni az Euklideszi algoritmussal. Az alábbiakból ki fog derülni, hogy ennél jóval több is igaz: az Euklideszi algoritmus kisebb módosításával hatékony módszert kaphatunk a lineáris kongruenciák megoldására. A következő feladatot vizsgáljuk tehát:

Bemenet: a, b és m egészek;

Kimenet: A c és m' egészek, amelyekre az $a \cdot x \equiv b \pmod{m}$ lineáris kongruencia megoldáshalmaza az $x \equiv c \pmod{m'}$ feltételt kielégítő x-ekből áll; vagy „nincs megoldás”, ha az $a \cdot x \equiv b \pmod{m}$ lineáris kongruencia nem megoldható.

Ha egy tetszőleges bemenetre (a,m) értékét már (az Euklideszi algoritmussal) meghatároztuk, akkor ennek az $(a,m) \mid b$ feltétel ellenőrizhetőségén kívül további haszna is van: a 2.4.2. Tétel után írtak szerint az $a \cdot x \equiv b \pmod{m}$ lineáris kongruencia ekvivalens az (a,m)-val való osztás után kapott $a' \cdot x \equiv b' \pmod{m'}$ feladattal, amelyre már $(a',m') = 1$. A továbbiakban tehát feltételezhetjük, hogy a bemenetként kapott a, b és m egészekre $(a,m) = 1$ teljesül.

Az Euklideszi algoritmusnak a lineáris kongruenciák megoldására szolgáló módosított változatát alább először az $567x \equiv 123 \pmod{1238}$ feladaton illusztráljuk. Ehhez először felírjuk az $1238x \equiv 0 \pmod{1238}$ kongruenciát (ami nyilván minden x-re igaz, ezt $(*)$ jelöli), majd a megoldandó, bemenetként kapott ((B)-vel jelölt) lineáris kongruenciát. Ezután megismételjük az $(567,1238)$ meghatározására szolgáló számítást (amelyet a 2.6.4. szakaszban már láttunk), de minden lépést kiegészítünk azzal, hogy a kettővel korábbi kongruenciából kivonjuk az eggyel korábbinak egy alkalmas többszörösét úgy, hogy a kapott kongruenciában x együtthatója az Euklideszi algoritmus által éppen kiszámolt maradékra változzon. Eközben a keletkező lineáris kongruenciákban a jobb oldalakon álló konstansokat mindig helyettesítjük az 1238-cal vett maradékukkal (ezáltal elkerülve, hogy az eljárásban túl nagy számok keletkezzenek). Végül a (6) lépésben az $x \equiv 819 \pmod{1238}$ kongruenciát kapjuk – ami tehát $(567,1238) = 1$ miatt a feladat egyetlen megoldását adja.

$$
\begin{array}{rrl|l}
(*) & 1238x \equiv 0 & \pmod{1238} & \\
(B) & 567x \equiv 123 & \pmod{1238} & \\
(*)-2\cdot(B):(1) & 104x \equiv -246 \equiv 992 & \pmod{1238} & 1238 = 2\cdot 567 + 104 \\
(B)-5\cdot(1):(2) & 47x \equiv -4837 \equiv 115 & \pmod{1238} & 567 = 5\cdot 104 + 47 \\
(1)-2\cdot(2):(3) & 10x \equiv 762 & \pmod{1238} & 104 = 2\cdot 47 + 10 \\
(2)-4\cdot(3):(4) & 7x \equiv -2933 \equiv 781 & \pmod{1238} & 47 = 4\cdot 10 + 7 \\
(3)-1\cdot(4):(5) & 3x \equiv -19 \equiv 1219 & \pmod{1238} & 10 = 1\cdot 7 + 3 \\
(4)-2\cdot(5):(6) & x \equiv -1657 \equiv 819 & \pmod{1238} & 7 = 2\cdot 3 + 1 \\
 & & & 3 = 3\cdot 1 + 0
\end{array}
$$

Ugyanezt a módszert tetszőleges $a \cdot x \equiv b \pmod{m}$ lineáris kongruenciára alkalmazhatjuk. Az (a,m) kiszámítására a 2.6.4. szakaszban látott eljárás az alábbiak szerint módosul. A fenti példához hasonlóan a lépések során számolt c_i konstansokról feltesszük, hogy $0 \le c_i < m-1$. Az eljárás végén $r_k = 1$, hiszen a 2.6.1. Állítás-

ban beláttuk, hogy $r_k = (a,m)$, másrészt fentebb feltettük, hogy $(a,m) = 1$. Ezért a (k) lépésben kapott kongruencia a feladat megoldását adja: $x \equiv c_k \pmod{m}$.

$$
\begin{array}{rll|l}
 & (*)\ m\cdot x \equiv 0 & \pmod m & \\
 & (B)\ \ a\cdot x \equiv b & \pmod m & \\
(*) - t_1\cdot(B): & (1)\ r_1\cdot x \equiv -t_1\cdot b \equiv c_1 & \pmod m & m = t_1\cdot a + r_1 \\
(B) - t_2\cdot(1): & (2)\ r_2\cdot x \equiv b - t_2\cdot c_1 \equiv c_2 & \pmod m & a = t_2\cdot r_1 + r_2 \\
(1) - t_3\cdot(2): & (3)\ r_3\cdot x \equiv c_1 - t_3\cdot c_2 \equiv c_3 & \pmod m & r_1 = t_3\cdot r_2 + r_3 \\
(2) - t_4\cdot(3): & (4)\ r_4\cdot x \equiv c_2 - t_4\cdot c_3 \equiv c_4 & \pmod m & r_2 = t_4\cdot r_3 + r_4 \\
 & \vdots \qquad \vdots & & \vdots \\
 & (k)\ r_k\cdot x \equiv c_{k-2} - t_k\cdot c_{k-1} \equiv c_k & \pmod m & r_{k-2} = t_k\cdot r_{k-1} + r_k \\
 & & & r_{k-1} = t_{k+1}\cdot r_k + 0
\end{array}
$$

Az algoritmus működésének helyessége szinte magától értetődő: a kongruenciákon végzett alapműveletekre vonatkozó szabályok miatt (lásd a 2.3.3. Tételt) minden lépésben olyan kongruenciát kapunk, amelyet a bemenetként kapott feladat minden x megoldása kielégít. Mivel ez az utolsóként kapott $x \equiv c_k \pmod{m}$ kongruenciára is igaz és $(a,m) = 1$ miatt tudjuk, hogy a feladatnak egyetlen megoldása van modulo m, ezért ez a megoldás valóban c_k kell legyen. (Érdemes azonban megfigyelni, hogy az eljárás közben kapott kongruenciák önmagukban nem feltétlen ekvivalensek a bemenetként kapottal, a megoldáshalmazuk lehet bővebb is. Így például a fenti példában az (1) és (3) lépésekben keletkező lineáris kongruenciáknak két megoldása is van modulo 1238.) Emellett nyilván továbbra is igaz, hogy az eljárás során legföljebb $2\cdot\lceil\log_2 a\rceil$ maradékos osztást végzünk (lásd a 2.6.2. Állítást), így az eljárás polinomiális futásidejű.

EUKLIDESZI ALGORITMUS (AZ $a\cdot x \equiv b \pmod{m}$ LINEÁRIS KONGRUENCIA MEGOLDÁSÁRA AZ $(a,m) = 1$ ESETBEN)

Bemenet: a, b és m (amelyekre $0 < a,b < m$ és $(a,m) = 1$ teljesül)

0. lépés. $M \leftarrow m$, $p \leftarrow 0$, $q \leftarrow b$

1. lépés. $t \leftarrow \lfloor \frac{m}{a} \rfloor$, $r \leftarrow m \mod a$.

2. lépés. Ha $r = 0$, akkor
PRINT „A lineáris kongruencia megoldáshalmaza: $x \equiv$”, $q \pmod{M}$; STOP

3. lépés. $c \leftarrow p - t\cdot q \mod M$

4. lépés. $m \leftarrow a$, $a \leftarrow r$, $p \leftarrow q$, $q \leftarrow c$

Folytassuk az **1. lépésnél**.

Az m és a változók továbbra is a kettővel, illetve az eggyel korábbi maradékokat tárolják. Most azonban a keletkező lineáris kongruenciák jobb oldalaiból is el kell tárolni a kettővel, illetve eggyel korábbit; ezeket p, illetve q jelöli. Így a 0. lépésben p értékét 0-nak, q értékét b-nek inicializáljuk összhangban a fenti leírásbeli $(*)$

és (B) jobb oldalaival. A 0. lépésben ugyancsak eltároljuk a bemenetként kapott m értékét, mert arra az eljárás 3. lépésében végig szükségünk lesz. Az 1. lépésben továbbra is csak egyetlen maradékos osztást hajtunk végre, de most $m \mod a$ mellett a t változóban eltároljuk $\lfloor \frac{m}{a} \rfloor$ értékét is – erre szintén a 3. lépésben lesz szükség. Ezektől a kiegészítésektől eltekintve azonban a módosított eljárás azonos az Euklideszi algoritmus eredeti változatával.

Ha egy tetszőleges $a \cdot x \equiv b \pmod{m}$ lineáris kongruenciát a fentiek szerint oldunk meg, akkor ehhez kétszer kell futtatni az Euklideszi algoritmust: először kiszámítjuk (a,m) értékét és (az $(a,m) \mid b$ feltétel ellenőrzése után) leosztunk vele, majd a kapott $a' \cdot x \equiv b' \pmod{m'}$ feladatra futtatjuk az algoritmus most megismert, módosított változatát. Valójában azonban egyszer is elég futtatni az algoritmust, a két futás összevonható. Ha ugyanis a fenti eljárást közvetlenül a bemenetként kapott $a \cdot x \equiv b \pmod{m}$ feladatra futtatjuk, akkor az a $d \cdot x \equiv c_k \pmod{m}$ kongruenciával ér véget, ahol $d = r_k = (a,m)$. Így ezen a ponton a $d \mid b$ feltétel ellenőrzésével eldönthető, hogy a lineáris kongruencia megoldható-e. Ha igen, akkor pedig könnyű megmutatni, hogy a megoldáshalmaza azonos a kapott $d \cdot x \equiv c_k \pmod{m}$ kongruenciáéval. Valóban, egyrészt ennek a kongruenciának a megoldáshalmaza tartalmazza a bemenetként kapott feladat összes megoldását (hiszen ez az eljárás közben keletkező mindegyik kongruenciáról elmondható), másrészt hamis gyököket nem tartalmazhat, mert a megoldásainak a száma modulo m egyenlő $(d,m) = d$-vel, ami a 2.4.2. Tétel szerint azonos az $a \cdot x \equiv b \pmod{m}$ megoldásainak számával. Ebből persze következik az is, hogy egyrészt c_k osztható $(d,m) = d$-vel (hiszen $d \cdot x \equiv c_k \pmod{m}$ megoldható), másrészt hogy az $a \cdot x \equiv b \pmod{m}$ megoldáshalmaza azonos a d-vel való osztás után előálló $x \equiv \frac{c_k}{d} \pmod{\frac{m}{d}}$ kongruenciáéval. Mindezek alapján végül is az algoritmus a következő formában írható le.

EUKLIDESZI ALGORITMUS (AZ $a \cdot x \equiv b \pmod{m}$ LINEÁRIS KONGRUENCIA MEGOLDÁSÁRA)

Bemenet: a, b és m (amelyekre $0 < a,b < m$ teljesül)

0. lépés. $M \leftarrow m$, $p \leftarrow 0$, $q \leftarrow b$

1. lépés. $t \leftarrow \lfloor \frac{m}{a} \rfloor$, $r \leftarrow m \mod a$.

2. lépés. Ha $r = 0$, akkor:
- Ha $a \nmid b$, akkor
  PRINT „A lineáris kongruenciának nincs megoldása.”; STOP
- Ha $a \mid b$, akkor
  PRINT „A lineáris kongruencia megoldáshalmaza: $x \equiv$”, $\frac{q}{a} \pmod{\frac{M}{a}}$; STOP

3. lépés. $c \leftarrow p - t \cdot q \mod M$

4. lépés. $m \leftarrow a$, $a \leftarrow r$, $p \leftarrow q$, $q \leftarrow c$

Folytassuk az **1. lépésnél**.

2.6.6. Prímtesztelés

Többször említettük már, hogy a prímfaktorizálás feladatára nem ismert hatékony algoritmus; később látni fogjuk, hogy ez nem feltétlen „rossz hír”, ezen alapszik a nyilvános kulcsú titkosítás működése. Ehhez képest meglepő, hogy egy adott számról hatékonyan el lehet dönteni, hogy prím-e vagy sem – ezt a feladatot hívják prímtesztelésnek:

Bemenet: m egész;
Kimenet: „IGEN”, ha m prím és „NEM”, ha m nem prím.

A prímtesztelő algoritmusoknak tehát az m összetettségét úgy kell tudniuk kimutatni, hogy közben m egyetlen valódi osztóját sem találják meg. (Valóban, ha minden összetett számnak hatékonyan kiszámítható volna egy valódi osztója, akkor ennek az ismételt alkalmazásával a prímtényezős felbontáshoz is hatékonyan eljuthatnánk.)

Az egyik legegyszerűbb prímtesztelő módszer a 2.5.7. Euler-Fermat tételen alapszik: ha m prím és $1 \leq a \leq m-1$ tetszőleges, akkor $\varphi(m) = m-1$ és $(a,m) = 1$, így az $a^{m-1} \equiv 1 \pmod{m}$ kongruenciának teljesülnie kell. Ha tehát sikerül találnunk egy olyan $1 \leq a \leq m-1$ egészt, amelyre $a^{m-1} \not\equiv 1 \pmod{m}$, akkor m biztosan nem prím. Az alábbi, *Fermat-teszt* néven közismert eljárás a lehető legegyszerűbb módon próbál ilyen a-t keresni: egymás után generál véletlen a számokat és minden a-ra ellenőrzi az $a^{m-1} \equiv 1 \pmod{m}$ feltételt; ha ez nem teljesül, akkor az algoritmus megáll (és közli, hogy m nem prím), ha viszont teljesül, akkor új a-t generál.

Az algoritmust ehhez az igen egyszerű alapötlethez képest csak két ponton egészítjük ki. Egyrészt az $a^{m-1} \equiv 1 \pmod{m}$ feltétel ellenőrzése előtt kiszámítjuk (a,m) értékét; ha olyan szerencsénk van, hogy $(a,m) > 1$, akkor m-nek nem csak az összetettségét tudtuk meg, hanem még egy valódi osztóját is megkaptuk. Másrészt beiktatunk egy (k-val jelölt) számlálót, ami egy előre rögzített számú sikertelen kísérlet után (alább ezt 100-nak választottuk) megállítja az eljárást.

FERMAT-TESZT

Bemenet: m egész

0. lépés. $k \leftarrow 1$

1. lépés. Generáljunk egy a véletlen számot 1 és $m-1$ között.

2. lépés. Számítsuk ki (a,m) értékét az EUKLIDESZI ALGORITMUSSAL. Ha ez nem 1, akkor: PRINT „m NEM prím”; STOP

3. lépés. Számítsuk ki $(a^{m-1} \bmod m)$ értékét az ISMÉTELT NÉGYZETRE EMELÉSEK MÓDSZERÉVEL. Ha ez nem 1, akkor: PRINT „m NEM prím”; STOP

4. lépés. Ha $k = 100$, akkor: PRINT „IGEN, m (valószínűleg) prím”; STOP

5. lépés. $k \leftarrow k+1$

Folytassuk az **1. lépésnél**.

Szokás a fenti eljárást a következő, a krimik nyelvezetét idéző módon is elmondani. Az a véletlen számokat sorban a tanúk padjára idézzük, az a vallomása (az $(a,m)=1$ esetben) az $(a^{m-1} \mod m)$ értéke. Ha ez 1, akkor a nem közöl információt m prímségét illetően; ilyenkor a-t az m *cinkosának* mondjuk. Ha viszont $a^{m-1} \not\equiv 1 \pmod m$, akkor a leleplezi m összetettségét; ilyenkor a-t az m *árulójának* nevezzük. (Nem szokás árulónak nevezni a-t az $(a,m)>1$ esetben annak ellenére sem, hogy természetesen ilyenkor is kiderül m összetettsége. Így a cinkos és az áruló elnevezések egyaránt csak az 1 és m közötti, m-hez relatív prím a-kra vonatkoznak.)

Az algoritmus 1. lépése véletlen szám generálását írja elő, de ennek a megvalósításával itt nem foglalkozunk annak ellenére sem, hogy ez a feladat messze nem magától értetődő. (A kriptográfiai alkalmazások megbízhatósága szempontjából alapvető, hogy jó minőségű – tehát a „valódi véletlent" nagyon jól imitáló – véletlen szám generátor álljon rendelkezésre. A legtöbb programnyelv tartalmaz beépített véletlen szám generátort, de ezek jó része nem megfelelő a kriptográfiai alkalmazások számára.) Nem foglalkozunk itt emellett azzal sem, hogy a véletlen használata a gyakorlati szempontok mellett alapvető algoritmuselméleti kérdéseket is felvet, mert az algoritmus fogalmának elméleti megalapozása elsősorban a *determinisztikus algoritmusokra* vonatkozik. (Ezek azok az algoritmusok, amelyek azonos bemenet esetén mindig ugyanazt a futást és kimenetet produkálják. Véletlen számok használata esetén ez a feltétel nyilván sérül.)

A Fermat-teszt működésével kapcsolatos legégetőbb kérdés azonban jól láthatóan a 4. lépés értelmezése: mit értsünk azalatt, hogy „m valószínűleg prím"? Más szóval: mire lehet következtetni 100 semmitmondó vallomásból, vagyis 100 sikertelen kísérletből m összetettségének a bizonyítására? Ha például egy m összetett számra előfordulhatna, hogy a hozzá relatív prím a-knak a 0,1%-a áruló, akkor 90%-nál is nagyobb a valószínűsége annak, hogy a 100 kísérlet során egy árulót sem találunk és m-et tévesen prímnek nyilvánítjuk. Az alábbi tétel szerint ez nem fordulhat elő.

2.6.3. Tétel. *Ha $m>1$ összetett szám és m-nek van árulója, akkor az 1 és m közötti, m-hez relatív prím számoknak legalább a fele áruló.*

Bizonyítás: Legyen a egy tetszőleges árulója m-nek és legyen $c_1, c_2, \ldots, c_k$ az m összes cinkosa. Megmutatjuk, hogy az $a_i = (a \cdot c_i \mod m)$, $i = 1,2,\ldots,k$ számok páronként különböző árulói m-nek. Ebből következni fog, hogy az árulók száma legalább akkora, mint a cinkosok száma, ami ekvivalens a tétel állításával.

Először is vegyük észre, hogy $(a,m)=1$ és $(c_i,m)=1$ (valamint a számelmélet alaptétele) miatt $(a \cdot c_i, m)=1$, így a 2.5.1. Állítás szerint $(a_i,m)=1$ is igaz, mert $a_i \equiv a \cdot c_i \pmod m$. Továbbá az $a_i \equiv a \cdot c_i \pmod m$ kongruenciát az $(m-1)$-edik hatványra emelve kapjuk, hogy

$$a_i^{m-1} \equiv (a \cdot c_i)^{m-1} = a^{m-1} \cdot c_i^{m-1} \equiv a^{m-1} \cdot 1 = a^{m-1} \not\equiv 1 \pmod m.$$

(Itt felhasználtuk, hogy $c_i^{m-1} \equiv 1 \pmod m$ és $a^{m-1} \not\equiv 1 \pmod m$, mert c_i cinkos és a áruló.) Így ebből valóban következik, hogy a_i is áruló.

Végül megmutatjuk, hogy az $a_1, a_2, \ldots, a_k$ árulók páronként különbözők. Ugyanis indirekt feltéve, hogy $a_i = a_j$ valamely $1 \leq i, j \leq k$, $i \neq j$ esetén, ebből $a \cdot c_i \equiv a \cdot c_j \pmod{m}$ következik. Mindkét oldalt a-val osztva a $c_i \equiv c_j \pmod{m}$ kongruenciát kapjuk (hiszen $(a,m) = 1$ miatt a modulus nem változik). Ez azonban $1 \leq c_i, c_j \leq m-1$, $c_i \neq c_j$ miatt ellentmondás, amivel a tételt beláttuk. □

A fenti tételnek alapvető fontosságú következménye van a Fermat-teszt használhatóságával kapcsolatban: ha m összetett és van árulója, akkor a teszt legföljebb $\frac{1}{2^{100}}$ valószínűséggel nyilvánítja m-et tévesen prímnek. Bár ez a valószínűség kétségtelenül pozitív, olyan felfoghatatlanul kicsi, hogy az minden gyakorlati szempontból elhanyagolható. (Ha az ősrobbanás pillanatától kezdve másodpercenként egymilliárdszor futtattuk volna a tesztet ezzel a hibavalószínűséggel, akkor 99,96% valószínűséggel még sohasem hibázott volna. De aki ezt is túl rizikósnak találja, felemelheti az algoritmus 4. lépésében szereplő 100-as határt például 1000-re.)

Még mindig nyitva maradt azonban egy olyan kérdés, amely a Fermat-teszt gyakorlati alkalmazhatóságát kétségessé teszi. A fenti tétel ugyanis feltételezi, hogy m-nek van legalább egy árulója. Előfordulhat-e, hogy m ugyan összetett, de nincs egyetlen árulója sem? A válasz sajnos igen, példa erre az $561 = 3 \cdot 11 \cdot 17$.

2.6.4. Definíció. *Az $m > 1$ összetett számot* univerzális álprímnek*, vagy más néven* Carmichael-számnak *nevezzük, ha nincs árulója; vagyis ha minden $1 < a < m$, $(a,m) = 1$ esetén $a^{m-1} \equiv 1 \pmod{m}$.*

Ha tehát a Fermat-tesztet egy m Carmichael-számra futtatjuk, akkor az nagy valószínűséggel tévesen prímnek fogja nyilvánítani azt (m egyedül a 2. lépésben, az $(a,m) = 1$ feltétel sérülésével „bukhatna le”, de ilyen a-ba botlani csak nagyon kis valószínűséggel fogunk). A Carmichael-számok viszonylag ritkák (az 561 a legkisebb, az utána következő az $1105 = 5 \cdot 13 \cdot 17$, egymillióig pedig összesen csak 43 darab van), de 1994-ben bebizonyították, hogy a számuk végtelen.

A Fermat-teszt tehát ugyan nagyon hatékony algoritmus, de – összefoglalva a fentieket – az alábbi három hiányosság merül fel vele kapcsolatban:

(i) Az algoritmus a Carmichael-számokat nagy valószínűséggel tévesen prímnek nyilvánítja.

(ii) Az eljárás az m prímnek nyilvánításakor egy elhanyagolhatóan kicsi, de pozitív valószínűséggel akkor is tévedhet, ha m tetszőleges összetett szám.

(iii) Az algoritmus nem determinisztikus, véletlent használ.

A gyakorlat szempontjából a fenti három hiányosság közül csak az *(i)* érdemel figyelmet, a másik kettő csupán elvi jelentőségű. Szerencsére a Fermat-tesztnek ismertek olyan módosításai, amelyek éppen ezt a problémát oldják meg; ezek közül a *Miller-Rabin-teszt* a legtöbbet használt, ezt alább röviden ismertetjük. Ez azon alapszik, hogy a tanúk „alaposabb kikérdezésével” már kiszűrhetők a Carmichael-számok is, miközben az algoritmus futásideje ettől alig romlik.

A Miller-Rabin-teszt azonban nem érinti a fenti *(ii)* és *(iii)* szempontot és ezáltal nyitva hagyja a kérdést: létezik-e olyan polinomiális futásidejű algoritmus, amely

minden m szám prímségét determinisztikusan és teljes biztonsággal eldönti? A válasz erre a kérdésre 2002 óta ismert: ekkor publikált *Agrawal*, *Kayal* és *Saxena* egy ilyen algoritmust. Bár ez az eredmény áttörés volt a prímszámok elméletében, a gyakorlati alkalmazások a nagyságrendekkel jobb lépésszáma miatt továbbra is a Miller-Rabin-tesztet (vagy más, véletlent használó módszert) használnak.

A Miller-Rabin-teszt

A Miller-Rabin-teszt a struktúráját tekintve azonos a Fermat-teszttel, attól érdemben csak a 3. lépésében különbözik. Az alapötlet a következő: az $a^{m-1} \equiv 1 \pmod{m}$ feltételt helyettesítjük egy annál szigorúbb olyan feltétellel, amelyet minden m prímszámnak teljesítenie kell, de az összetett számoknak nem feltétlenül. Ehhez a következő, rendkívül egyszerű megfigyelést használjuk.

2.6.5. Állítás. *Ha m prím és $x^2 \equiv 1 \pmod{m}$ teljesül valamely x egészre, akkor $x \equiv 1 \pmod{m}$ vagy $x \equiv -1 \pmod{m}$.*

Bizonyítás: $x^2 \equiv 1 \pmod{m}$ miatt $m \mid x^2 - 1 = (x-1)(x+1)$. Ha m prím, akkor ebből (a számelmélet alaptétele miatt) $m \mid x-1$ vagy $m \mid x+1$ következik. Az első esetben $x \equiv 1 \pmod{m}$, a másodikban $x \equiv -1 \pmod{m}$. □

Megjegyezzük, hogy a fenti állítás alkalmazásakor az alábbiakban sokszor jutnak szerephez az $x \equiv -1 \pmod{m}$ feltételt kielégítő x-ek. Ezeket az egyszerűség kedvéért (-1) maradékot adónak fogjuk mondani annak ellenére, hogy egy ilyen x-nek az m-mel vett osztási maradéka helyesen nyilván $m-1$.

A tesztelendő m számról feltehetjük, hogy páratlan (ha $m > 2$ páros, akkor azonnal összetettnek nyilváníthatjuk), így $m-1$ páros. Ezért ha egy $1 \le a \le m-1$, $(a,m) = 1$ egészre $a^{m-1} \equiv 1 \pmod{m}$ adódik a Fermat-teszt 3. lépésében, akkor (mielőtt az a tanút elbocsátjuk a tanúk padjáról) érdemes megvizsgálni $(a^{\frac{m-1}{2}} \bmod m)$ értékét is: ha ez ± 1-től különböző, akkor a fenti állítás értelmében m összetettsége mégis kiderült. Ha viszont $a^{\frac{m-1}{2}} \equiv \pm 1 \pmod{m}$, akkor – noha ebből nem tudtunk meg semmit – még mindig érdemes lehet tovább vizsgálódni: ha $a^{\frac{m-1}{2}} \equiv 1 \pmod{m}$ és $\frac{m-1}{2}$ is páros, akkor a fenti állítás szerint $(a^{\frac{m-1}{4}} \bmod m)$ értéke is ± 1 kell legyen. Ehhez hasonlóan léphetünk tovább az $\frac{m-1}{8}, \frac{m-1}{16}, \ldots$ kitevőkre is egészen addig, amíg páratlan kitevőig nem jutunk vagy a megfelelő hatványa 1-től különböző maradékot nem ad. Ha ez a maradék (-1), akkor hívjuk a következő tanút (vagyis új véletlen a-t generálunk), ellenkező esetben m biztosan összetetett.

A valóságban ezeket a számításokat nem a kitevők folyamatos felezésével érdemes végezni, mert így mindig újra kellene futtatni az ismételt négyzetre emelések módszerét a feleakkora kitevőjű hatványok kiszámításához. Ehelyett rögtön az algoritmus futásának az elején meghatározzuk a 2-nek azt a legnagyobb hatványát, amellyel $m-1$ még osztható – vagyis felírjuk azt $m-1 = 2^t \cdot c$ alakban, ahol c már páratlan. Ezután a tesztelés közben az ismételt négyzetre emelések módszerével először mindig $(a^c \bmod m)$ értékét számítjuk ki, ebből további négyzetre emelésekkel jutunk el $(a^{2c} \bmod m)$, $(a^{4c} \bmod m)$, … értékéhez.

MILLER-RABIN-TESZT

Bemenet: $m > 1$ páratlan egész

0. lépés.
- $k \leftarrow 1$
- $m-1$ ismételt felezésével határozzuk meg azt a c páratlan egészt és $t \geq 1$ egészt, amelyekre $m-1 = 2^t \cdot c$

1. lépés. Generáljunk egy a véletlen számot 1 és $m-1$ között.

2. lépés. Számítsuk ki (a,m) értékét az EUKLIDESZI ALGORITMUSSAL. Ha ez nem 1, akkor: PRINT „m NEM prím"; STOP

3. lépés. Számítsuk ki $(a^c \mod m)$, $(a^{2c} \mod m)$, $(a^{4c} \mod m)$, …, $(a^{2^{t-1}c} \mod m)$ értékeit az ISMÉTELT NÉGYZETRE EMELÉSEK MÓDSZERÉVEL. Ha ezek közül egyik sem (-1) és $(a^c \mod m) \neq 1$, akkor: PRINT „m NEM prím"; STOP

4. lépés. Ha $k = 100$, akkor: PRINT „IGEN, m (valószínűleg) prím"; STOP

5. lépés. $k \leftarrow k+1$

Folytassuk az **1. lépésnél**.

Az eljárás 3. lépésében számolt maradékokra nyilván igaz, hogy ha valamelyikük (± 1), akkor az összes további 1-gyel egyenlő. Így a 3. lépésben meghozott döntés valóban megfelel a fentebb mondottaknak: az m-hez relatív prím a egész akkor tanúsítja m összetettségét, ha az $(a^c \mod m), \ldots, (a^{2^{t-1}c} \mod m)$, $(a^{m-1} \mod m)$ maradékok között vagy nincs 1 (és így sérül a 2.5.7. Euler-Fermat tételből következő feltétel) vagy az első 1-est közvetlenül megelőzi egy (-1)-től különböző maradék (és ekkor sérül a 2.6.5. Állítás feltétele). Érdemes megfigyelni, hogy az algoritmus 3. lépésében az utolsó kiszámított maradék $(a^{\frac{m-1}{2}} \mod m)$, az eljárás $(a^{m-1} \mod m)$ kiszámításáig nem jut el. De erre nincs is szükség: ha $a^{\frac{m-1}{2}} \equiv \pm 1 \pmod m$, akkor ebből $a^{m-1} \equiv 1 \pmod m$ úgyis következik, ha pedig $a^{\frac{m-1}{2}} \not\equiv \pm 1 \pmod m$, akkor $(a^{m-1} \mod m)$ értékétől függetlenül m mindenképp összetett.

Az eljárás működését $m = 561$-re illusztráljuk. Említettük, hogy $561 = 3 \cdot 11 \cdot 17$ Carmichael-szám, vagyis a Fermat-teszt csak abban az esetben leplezheti le az összetettségét, ha 561-hez nem relatív prím a véletlen számot generál. (Mivel $\varphi(561) = 2 \cdot 10 \cdot 16 = 320$, ezért egy a véletlen számra ennek az esélye $\frac{560-320}{560} = \frac{3}{7}$. De ez a valószínűség csak azért ilyen nagy, mert 561 nagyon kicsi; egy sokszáz jegyű m Carmichael-szám esetében már csak elenyésző valószínűséggel ütköznénk m-hez nem relatív prím a-ba.) A Miller-Rabin-teszt viszont például az $a = 2$ esetben is kimutatja $m = 561$ összetettségét: $m-1 = 560 = 2^4 \cdot 35$ és az ismételt négyzetre emelések módszerével könnyen kiszámítható, hogy $2^{35} \equiv 263 \pmod{561}$, $2^{70} \equiv 166 \pmod{561}$, $2^{140} \equiv 67 \pmod{561}$ és $2^{280} \equiv 1 \pmod{561}$. Mivel ezek között a maradékok között nincs (-1) (vagyis 560) és köztük az első nem 1, ezért a Miller-Rabin teszt az 561-et az $a = 2$ „vallomása" alapján helyesen összetettnek nyilvánítja (annak ellenére is, hogy $(2,561) = 1$).

A Miller-Rabin-teszt gyakorlati alkalmazhatóságával kapcsolatban ugyanazt a kérdést kell feltennünk, mint a Fermat-teszt kapcsán: milyen valószínűséggel derül ki egy véletlen a vizsgálatakor m összetettsége? Nevezzünk egy $1 \leq a \leq m-1$, $(a,m) = 1$ egészt *Miller-Rabin-árulónak*, ha teljesíti a fenti eljárás 3. lépésében

írt feltételt (és így a Miller-Rabin-teszt az a vizsgálata alapján az „m nem prím” következtetésre jutna). Egyszerű megfigyelés, hogy ha a árulója m-nek (a kifejezésnek a Fermat-teszt kapcsán használt értelmében) akkor Miller-Rabin-áruló is. (Valóban, $a^{2^s c} \equiv \pm 1 \pmod{m}$ nem teljesülhet semmilyen $0 \leq s < t$-re, ha $a^{m-1} = a^{2^t c} \not\equiv 1 \pmod{m}$.) Ebből rögtön következik, hogy ha egy m összetett szám nem Carmichael-szám, akkor a a 2.6.3. Tétellel analóg állítás érvényes rá: az 1 és m közötti, m-hez relatív prím a-k legalább fele Miller-Rabin-áruló. Az alábbi, bizonyítás nélkül közölt tétel szerint ugyanez még a Carmichael-számokra is igaz.

2.6.6. Tétel. (Gary Miller, 1975)
Ha $m > 1$ páratlan, összetett szám, akkor az 1 és m közötti, m-hez relatív prím számoknak legalább a fele Miller-Rabin-áruló.

A Miller-Rabin-teszt tehát a Fermat-teszt korábban felsorolt hiányosságai közül éppen a leglényegesebbet javítja ki: *minden* m összetett számot (az eljárásnak a fent leírt formájában) legföljebb $\frac{1}{2^{100}}$ valószínűséggel nyilvánít prímnek. Bár ez a hibavalószínűség minden gyakorlati szempontból elhanyagolható és a 100-as küszöbérték növelésével tetszőlegesen kicsire csökkenthető, egy szám prímségét a Miller-Rabin-teszt sem tudja minden elméleti kétséget kizáróan bizonyítani. (Erre polinomiális futásidőben jelenleg csak a már említett Agrawal-Kayal-Saxena-teszt alkalmas.) A gyakorlati alkalmazások szempontjaiból azonban ez érdektelen, a Miller-Rabin-tesztet használják többek között a legelterjedtebb kriptográfiai módszerek is.

Prímgenerálás

A prímtesztelő algoritmusokat arra is lehet használni, hogy hatalmas prímszámokat állítsunk elő velük (ezt a feladatot szokták prímgenerálásnak nevezni). Ha például egy 300 jegyű prímszámra van szükségünk, akkor egymás után generálunk 300 jegyű véletlen számokat és mindegyikre lefuttatjuk (például) a Miller-Rabin-tesztet.

Persze felvetődik a kérdés: nem tart-e reménytelenül soká, mire az első prímet így megtaláljuk? Ennek a megválaszolására segítségül hívhatjuk a 2.2.3. Nagy Prímszámtételt: ha $\pi(n)$-re, vagyis az n-nél nem nagyobb prímek számára a $\pi(n) \approx \frac{n}{\ln n}$ közelítést használjuk, akkor azt kapjuk, hogy az n-nél nem nagyobb számok közül durván minden $(\ln n)$-edik prím. Ezt $n = 10^{300}$-ra alkalmazva az derül ki, hogy a legföljebb 300 jegyű számok közül minden $\ln 10^{300} = 300 \cdot \ln 10 = 690,78$-adik prím. (De ha rögtön kiszűrjük a 2-vel, 3-mal vagy 5-tel oszthatókat, a maradékból már minden 184-edik prím.) Következésképp már néhány száz véletlen szám tesztelése után a 300 jegyű számok között is nagyon nagy valószínűséggel találunk prímet – ez pedig bőven a megvalósíthatóság határain belül van.

(Megjegyezzük, hogy a fenti bekezdésben valójában helytelenül használtuk a 2.2.3. Nagy Prímszámtételt: önmagában abból, hogy a $\pi(n)$ és az $\frac{n}{\ln n}$ sorozatok hányadosa 1-hez tart, konkrét n-ekre semmi nem következik a két érték egymáshoz való viszonyáról. Szerencsére ismertek a $\pi(n)$ értékéről a 2.2.3. Tételnél többet is mondó eredmények, ezeknek a részleteiben azonban itt nem mélyedünk el. Megelégszünk azzal, hogy valójában $\frac{n}{\ln n}$ „megnyugtatóan jól” – és $n \geq 10$ esetén ráadásul alulról – becsli $\pi(n)$-et. Így például valóban igaz, hogy a legföljebb – vagy akár a pontosan – 300 jegyű számoknak legalább a 690-ed része prím.)

2.6.7. A nyilvános kulcsú titkosítás

Ha a fentiek szerint generálunk két (például) 300 jegyű prímszámot – jelölje ezeket p és q –, akkor az $N = p \cdot q$ szorzatukat bátran közzétehetjük, a jelenlegi tudásunk szerint ebből p és q értékét senki sem fogja tudni kiszámítani. Ennek a ténynek már önmagában is vannak alkalmazásai. Képzeljük el például, hogy egy nagyobb összeget szeretnénk egy bankban elhelyezni, de – biztosan jó okunk van erre – a kilétünket szeretnénk homályban hagyni. Ezért a bankbetétet jelszóval kívánjuk védeni: azt az utasítást adjuk a banknak, hogy a pénzünket kiadhatják bárkinek, aki az általunk megadott jelszót ismeri. Ez azonban kockázatos eljárás: ha a banknak egy kellően korrupt alkalmazottja hozzáfér a jelszavunkhoz és azt elárulja egy bűntársának, akkor búcsút mondhatunk a pénzünknek. A kérdés tehát ez: hogyan lehetne a banknak olyan információt adni, ami a jelszó ellenőrzésére képessé teszi őket, de magát a jelszót mégsem ismerik meg? Egy megoldás erre a következő: megadjuk a banknak $N = p \cdot q$ értékét, de p-t és q-t titokban tartjuk; N mellé azt az utasítást adjuk, hogy a pénzünket kiadhatják bárkinek, aki N egy valódi osztóját megadja. Ezt a feltételt N ismeretében a bank könnyen tudja ellenőrizni, de hiába van tele a bank korrupt alkalmazottakkal, N ismerete kevés egy valódi osztójának a kiszámításához.

Az a két tény tehát, hogy a prímtesztelés (és így a prímgenerálás) feladata hatékonyan megoldható, de a prímfaktorizálás a jelenlegi tudásunk szerint reménytelenül nehéz, együtt lehetőséget teremtenek „megfejthetetlen titkok” nagy tételben való előállítására. Ez az alapja számos, a nyilvános kulcsú titkosítás alapfeladatát megoldó módszernek is. A feladatot a 2.6. szakasz elején már leírtuk: olyan $C, D : \{0,1,\ldots,N-1\} \to \{0,1,\ldots,N-1\}$ kölcsönösen egyértelmű függvényeket keresünk, amelyek a következő feltételeknek eleget tesznek:

(i) minden $x \in \{0,1,\ldots,N-1\}$ esetén $D(C(x)) = x$, vagyis C és D egymás inverzei;

(ii) a kód „tulajdonosa” $C(x)$ és $D(x)$ értékét is hatékonyan ki tudja számítani;

(iii) a $C(x)$ kiszámítására vonatkozó eljárás nyilvánosságra hozható, kívülállók számára ebből $D(x)$ nem lesz hatékonyan kiszámítható.

A feltételekből természetesen következik, hogy N hatalmas, például (decimálisan) 600 jegyű szám. (Valóban, ha N túl kicsi, akkor egyszerűen az $x = 0,1,\ldots,N-1$ értékek kipróbálásával $C(x)$-ből x megkapható.)

Egy ilyen függvénypár birtokában a kód tulajdonosa biztonságosan tud üzenetet fogadni bárkitől anélkül, hogy az illetővel előtte kódot kellene egyeztetnie: elküldi (vagy nyilvánosan elérhetővé teszi) a C függvényt kiszámító eljárást, a partner pedig az x üzenet helyett mindig annak az $y = C(x)$ kódját küldi el. Ezt a kódolt üzenetet a kód tulajdonosa D-vel meg tudja fejteni ($D(y) = D(C(x)) = x$ miatt), de kívülállók nem. Ha pedig a kommunikációban részt vevő mindkét fél rendelkezik egy-egy ilyen C, D függvénypárral, akkor a köztük zajló teljes kommunikáció biztonságosan lebonyolítható. Ezen alapszik például a fejezet elején már említett *https* protokoll is (de az ott szereplő két titkosügynök problémájára is lehet ez a megoldás).

Messze nem magától értetődő a fenti három feltételt kielégítő C, D függvénypárok konstrukciója, de mostanra rendelkezésünkre állnak az ehhez szükséges eszközök. A feladat többféleképpen is megoldható, alább a legszélesebb körben elterjedt

(a Rivest-Shamir-Adleman szerzőhármas neveinek kezdőbetűi alapján) RSA-nak nevezett algoritmust ismertetjük. Ehhez szükségünk lesz az alábbi állításra.

2.6.7. Állítás. *Legyenek p és q különböző prímek és $N = p \cdot q$. Ekkor tetszőleges x és $k \geq 1$ egészekre $x^{k\cdot\varphi(N)+1} \equiv x \pmod{N}$.*

Bizonyítás: Ha $(x,N) = 1$, akkor az állítás közvetlen következménye a 2.5.7. Euler-Fermat tételnek: az $x^{\varphi(N)} \equiv 1 \pmod{N}$ kongruenciát először a k-adik hatványra emelve, majd mindkét oldalt x-szel szorozva épp a bizonyítandó állítást kapjuk.

Ha $(x,N) \neq 1$, akkor $p|x$ vagy $q|x$. Ha mindkettő teljesül, akkor $N|x$, így a bizonyítandó állítás $0 \equiv 0 \pmod{N}$ miatt magától értetődő. Tegyük fel ezért, hogy $p \not| x$, de $q|x$ (a fordított, $p|x$, $q \not| x$ eset bizonyítása ezzel analóg). Mivel p prím és $p \not| x$, ezért $(x,p) = 1$ és $\varphi(p) = p-1$, így az Euler-Fermat tétel miatt $x^{p-1} \equiv 1 \pmod{p}$. Ezt a $k\cdot(q-1)$-edik hatványra emelve, majd mindkét oldalt x-szel szorozva $\varphi(N) = (p-1)(q-1)$ miatt a $x^{k\cdot\varphi(N)+1} \equiv x \pmod{p}$ kongruenciát kapjuk. De $q|x$ miatt ugyanez a kongruencia nyilván modulo q is fennáll. Ebből azonban következik, hogy modulo N is teljesül: a $p\left|x^{k\cdot\varphi(N)+1} - x\right.$ és $q\left|x^{k\cdot\varphi(N)+1} - x\right.$ oszthatóságokból együtt a $p\cdot q\left|x^{k\cdot\varphi(N)+1} - x\right.$ oszthatóság következik (mert p és q különböző prímek), ez pedig ekvivalens a bizonyítandó állítással. □

(Megjegyezzük, hogy a fenti állítás hasonló bizonyítással érvényes akkor is, ha N kettő helyett tetszőlegesen sok, de csupa különböző prím szorzata.)

Az RSA algoritmus

Generáljunk az előző szakaszban írtak szerint két (például) 300 jegyű prímet, legyenek ezek p és q. Legyen továbbá $N = p\cdot q$ és válasszunk még egy olyan c egészt is, amelyre $(c,\varphi(N)) = 1$ (ennek a feltételnek a szerepére később visszatérünk). Majd tegyük közzé, hogy a nyilvános C kódoló függvényünk a következő:

$$C : x \mapsto x^c \mod N.$$

Kétségtelen, hogy ez a függvény (az ismételt négyzetre emelések módszerével) hatékonyan kiszámítható – de hogyan találjunk hozzá egy D dekódoló függvényt? Keressük D-t is a fentihez hasonló alakban:

$$D : y \mapsto y^d \mod N.$$

A d értékét úgy szeretnénk megválasztani, hogy D ezáltal C inverze legyen. Ez tehát akkor teljesül, ha $D(C(x)) = x$ minden $0 \leq x \leq N-1$ esetén, ami $C(x) \equiv x^c \pmod{N}$ és ebből $D(C(x)) \equiv x^{c\cdot d} \pmod{N}$ miatt ekvivalens az $x^{c\cdot d} \equiv x \pmod{N}$ feltétellel.

Itt jut szerephez a 2.6.7. Állítás: eszerint D inverze lesz C-nek, ha a d értékét sikerül úgy megválasztanunk, hogy $c\cdot d = k\cdot\varphi(N)+1$ teljesül valamilyen $k \geq 1$ egészre. Más szóval: célunk a $\varphi(N)\left|c\cdot d - 1\right.$ oszthatóság, vagy ismét másképp a

$$c\cdot d \equiv 1 \pmod{\varphi(N)}$$

kongruencia kielégítése. Mivel itt c és $\varphi(N)$ adottak, ez d-re egy lineáris kongruencia feladat, amely (a c választásakor előrelátóan teljesített) $(c,\varphi(N))=1$ feltétel miatt megoldható (a 2.4.2. Tétel szerint). Ráadásul egy d megoldás hatékonyan ki is számítható az Euklideszi algoritmussal a 2.6.5. szakaszban látott módon.

Összefoglalva a fentieket: a nyilvános $C: x \mapsto (x^c \mod N)$ kódoló függvényhez a $D: y \mapsto (y^d \mod N)$ jó dekódoló függvény lesz, ha d a $c\cdot d \equiv 1 \pmod{\varphi(N)}$ lineáris kongruencia megoldása. Az Euklideszi algoritmust a d kiszámítására elég egyszer, a kód generálásakor lefuttatni, a kapott d ezután minden $C(x)=y$ kódolt üzenet dekódolására alkalmazható. Az RSA algoritmust az teszi biztonságossá, hogy a d értékét csak a $\varphi(N)=(p-1)(q-1)$ ismeretében lehet kiszámítani, ehhez pedig szükség van az $N=p\cdot q$ felbontásra. Ha tehát p és q (valamint d és $\varphi(N)$) értékét titokban tartjuk, akkor a kódoló függvény megadásához szükséges c és N bátran nyilvánosságra hozható, ebből a dekódoló függvény nem kiszámítható – legalábbis a fenti módszerrel biztosan nem.

Elvileg nem kizárt, hogy a D dekódoló függvényt valaki egy más alakban, pusztán c és N ismeretében előállítsa (vagyis a „c-edik gyökvonást modulo N” megvalósítsa N prímfelbontásának a hiányában is) – mint ahogyan az sem, hogy a prímfaktorizáció feladatának megoldására hatékony algoritmus születik. Mindez azonban a jelenlegi tudásunk és az RSA algoritmussal kapcsolatos több évtizedes tapasztalatok szerint nagyon valószínűtlen, a módszer rendkívül biztonságosnak tűnik.

Érdemes azonban megjegyezni, hogy az RSA algoritmus valóban biztonságos implementálása számos további, messze nem nyilvánvaló kérdést vet föl. Ezek egy részének a matematikai háttérhez kevés köze van – például kellő körültekintés híján eredményesen támadható a rendszer pusztán a kódolás és dekódolás idejének vagy az azt végző hardver energiafelhasználásának a mérésével. De az N és c paraméterek nem megfelelő választása is okozhat problémát – például egy ügyetlenül választott N prímfaktorizációja esetleg gyorsan meghatározható lehet. Mindezekkel a szempontokkal itt nem foglalkozunk.

Digitális aláírás

Az RSA algoritmussal a fent leírt módon megvalósított nyilvános kulcsú titkosításnak egyelőre van egy komoly hátránya a „hagyományos” (vagyis kódegyeztetésen alapuló) módszerekkel szemben: az üzenetek feladója könnyen hamisítható. Ha például Aliz és Bonifác levelezik és ehhez mindketten a másik által nyilvánosságra hozott C_B, illetve C_A kódoló kulcsot használják, akkor egy rosszindulatú kívülállót semmi nem akadályoz meg abban, hogy Bonifác nevében félrevezető üzenetet hamisítson Aliznak. Ugyanez a hiányosság még egy problémát felvet: még ha Bonifác is volt az üzenet valódi feladója, ezt ő később (ha az érdekei úgy kívánják) nyugodtan letagadhatja; sem Aliz, sem más nem tudja majd hitelt érdemlően bizonyítani a feladó kilétét. A korábbi, hagyományos módszerek használatakor mindez nem fordulhatott elő: mivel a használt kulcsot Alizon és Bonifácon kívül más nem ismerte, egy ezzel titkosított üzenet egyben arra is biztosíték volt, hogy az valóban a másiktól érkezett.

Szerencsére a nyilvános kulcsú titkosításnak a következő módosításával ezek a problémák is kiküszöbölhetők. Ha Bonifác az x üzenetetet szánja Aliznak, ak-

kor erre először a saját, titkos D_B dekódoló kulcsát alkalmazza, majd a kapott $y = D_B(x)$-re Aliz nyilvános C_A függvényét; végül $z = C_A(y) = C_A(D_B(x))$-et küldi el Aliznak. A dekódolást Aliz könnyen el tudja végezni: a kapott z-re először a saját titkos D_A kulcsát alkalmazva megkapja $y = D_A(z) = D_B(x)$-et, majd erre Bonifác nyilvánosan elérhető C_B függvényét alkalmazva nyeri vissza az $x = C_B(D_A(z))$ üzenetet. Alizon kívül más nem tudja dekódolni z-t, mert ehhez D_A ismerete szükséges. De most már a feladó kilétében is biztos lehet Aliz, hiszen Bonifácon kívül senki más nem alkalmazhatta volna x-re D_B-t.

Ha pedig Bonifácnak később kedve támadna letagadni, hogy a z üzenet tőle származik, akkor Aliz egy független bíróság előtt még úgy is tudja hitelt érdemlően cáfolni Bonifácot, hogy ehhez a saját titkos kulcsát nem kell felfednie: bemutatja a kapott $z = C_A(D_B(x))$ üzenetet és az általa félig dekódolt $y = D_A(z) = D_B(x)$-et. A bíróság ekkor Aliz és Bonifác nyilvános kulcsaival ellenőrizheti, hogy $C_A(y) = z$, valamint $C_B(y) = x$ fennállnak és ezekből meggyőződhet arról, hogy Aliz valóban igazat mond.

A számos további részlet és implementációs nehézség részletezésének mellőzésével megemlítjük, hogy a fenti módszeren alapulnak azok az elektronikus aláírási sémák is, amelyeket mára a világ számos országában (köztük a vonatkozó EU-s irányelv nyomán 2001-ben elfogadott törvény alapján Magyarországon is) a hagyományos aláírással egyenrangúnak tekintenek.

2.6.8. Feladat. Elfogtunk egy gyanús üzenetet: `159,111,5,140,39,68,6`. Tudjuk, hogy a feladó az üzenet karaktereit egyszerűen az ASCII kódjukkal helyettesítette, majd ezekre az $x \to (x^{29} \mod 161)$ kódoló függvényt alkalmazta. (Ezzel a függvénnyel tehát a $0, 1, \ldots, 160$ számokat lehet kódolni, de csak az első 128-nak van valódi jelentése.) Használjuk ki, hogy a feladó túl kicsi számokat választott a kód generálásakor: készítsünk dekódoló függvényt és fejtsük meg vele az üzenetet!

Megoldás: A kódolófüggvény az $N = 161$ modulust alkalmazza. Mivel a 161 prímfaktorizációja könnyű feladat, a dekódolófüggvényt ugyanúgy el tudjuk készíteni, ahogyan a kód generálója tette: $N = 161 = 7 \cdot 23$, amiből $\varphi(N) = 6 \cdot 22 = 132$. Így a $c = 29$ kitevőhöz a $29d \equiv 1 \pmod{132}$ lineáris kongruencia megoldásával találunk megfelelő d-t. A megoldást az Euklideszi algoritmussal keressük meg (bár a feladat paraméterei olyan kicsik, hogy a 2.4.1. Feladatban látott módszerek is gyorsan célhoz vezetnének):

$$\begin{array}{rll}
(*) & 132x \equiv 0 & \pmod{132} \\
(B) & 29x \equiv 1 & \pmod{132} \\
(*) - 4 \cdot (B) : (1) & 16x \equiv -4 \equiv 128 & \pmod{132} \\
(B) - 1 \cdot (1) : (2) & 13x \equiv -127 \equiv 5 & \pmod{132} \\
(1) - 1 \cdot (2) : (3) & 3x \equiv 123 & \pmod{132} \\
(2) - 4 \cdot (3) : (4) & x \equiv -487 \equiv 41 & \pmod{132}
\end{array}$$

Mivel $(29, 132) = 1$ (ha ez nem így volna, a kód generálója hibázott volna), ezért a 41 az egyetlen megoldás modulo 132. Ezzel elkészült a dekódolófüggvényünk:

$y \to (y^{41} \mod 161)$. Ezt alkalmazzuk az elfogott üzenetet alkotó számokra, természetesen az ismételt négyzetre emelések módszerét használva. Az üzenet első tagjára, 159-re a 2.6.3. szakaszban látotthoz hasonlóan részletezzük $(159^{41} \mod 161)$ kiszámítását:

k	a	b	c
0	159	41	1
1	4	20	159
2	16	10	159
3	95	5	159
4	9	2	132
5	81	1	132
6	–	0	66

Így $159^{41} \equiv 66 \pmod{161}$, a 66 pedig a B ASCII kódja. Az üzenet többi karakterének a dekódolását a kíváncsi olvasókra hagyjuk. □

2.7. Ajánlott irodalom

Az alábbi, kiváló tankönyv a tárgyalt anyag megértését segítő magyarázatokkal és feladatokkal bőségesen körített bevezetést nyújt a számelmélet világába, az ebben a jegyzetben is érintett kezdetektől egészen a jóval mélyebb eredményekig.

[1] Freud Róbert, Gyarmati Edit: *Számelmélet*, Nemzeti Tankönyvkiadó, Budapest, 2000, 2006.

A nyilvános kulcsú titkosítás elmélete iránt érdeklődőknek pedig az alábbi tankönyvet ajánljuk:

[2] Györfi László, Győri Sándor, Vajda István: *Információ- és kódelmélet*, Typotex Kiadó, Budapest, 2005, 2010.

3. fejezet

Végtelen halmazok számossága

Miből van több: természetes számból, vagy páros természetes számból? – Micsoda kérdés? Nyilván természetes számból, hiszen ezek közül csak minden második páros – mondhatná valaki. – Micsoda kérdés? Nyilván mindkettőből végtelen sok van, vagyis ugyanannyi – mondhatná valaki más ugyanolyan meggyőződéssel. Melyiküknek van igaza? Egyelőre nem mondhatunk mást, mint hogy mindkét vélemény egyformán megalapozatlan – ugyanis a halmazok számosságának (vagyis elemszámának) a fogalma egyelőre csak véges halmazokra bír értelemmel.

A matematika fejlődésében forradalmat hozott az a *Georg Cantortól*, az 1870-es évekből származó gondolat, hogy a végtelen halmazok számosságának az egyenlőségét (vagy épp nem egyenlőségét) is lehet definiálni. Az általa alkotott fogalom – megdöbbentő egyszerűsége ellenére – érdekes kérdések és meglepő válaszok áradatát indította el és egyben a matematika *halmazelmélet* nevű ágának a születését is jelentette.

Mára Cantor felfedezései a matematikai alapműveltség körébe tartoznak. Az informatikus számára is hasznos, ha ennek az elméletnek legalább az alapvetéseit megismeri: az algoritmikus megoldhatóság kérdéskörének vizsgálatában számos negatív (vagyis valaminek a lehetetlenségét kimondó) eredmény bizonyításában játszanak kulcsszerepet (lásd a 3.5. szakaszt).

3.1. Halmazok számosságának egyenlősége

Ha egy téli napon végignézünk a zsúfolt jégpályán, reménytelennek tűnő feladat volna megszámlálni, hogy hány korcsolya siklik egyszerre a jégen. Egyben azonban biztosak lehetünk: ugyanannyi a bal korcsolya, mint a jobb; ez természetes, hiszen ezek párban állnak. Ezen a szinte nevetségesen egyszerű gondolaton alapszik a modern halmazelmélet fent említett alapfogalma: két halmaz számossága akkor egyenlő, ha az egyik elemei párba állíthatók a másik elemeivel. Ezt a párba állítást egy függvénnyel fogjuk tudni pontosan megadni; persze akármilyen függvény nem felel meg a célnak, ezért van szükségünk az alábbi definíciókra.

3.1.1. Definíció. *Legyenek A és B tetszőleges halmazok és $f : A \to B$ egy függvény. Az f függvényt*

(i) injektívnek *(vagy más néven* invertálhatónak*) nevezzük, ha bármely $x_1, x_2 \in A$, $x_1 \neq x_2$ esetén $f(x_1) \neq f(x_2)$;*

(ii) szürjektívnek *(vagy más néven* ráképezésnek*) nevezzük, ha minden $y \in B$ esetén létezik olyan $x \in A$, amelyre $f(x) = y$;*

(iii) bijektívnek *(vagy más néven* kölcsönösen egyértelműnek*) nevezzük, ha egyszerre injektív és szürjektív.*

Injektív függvény, szürjektív függvény, illetve bijektív függvény helyett a rövidebb injekció*,* szürjekció*, illetve* bijekció *elnevezéseket is használjuk.*

A definíciót úgy is fogalmazhatjuk, hogy $f : A \to B$ akkor injektív, szürjektív, illetve bijektív, ha minden $y \in B$ elemnek legföljebb egy, legalább egy, illetve pontosan egy *ősképe* van (ahol az ősép olyan $x \in A$ elemet jelent, amelyre $f(x) = y$).

Fontos megjegyezni, hogy ha egy f függvényről kijelentjük, hogy az *egyértelmű*, azzal semmit nem mondunk róla. Ugyanis minden függvény egyértelmű: az $f : A \to B$ függvény definíciójában benne foglaltatik, hogy az minden A-beli elemhez egyetlen B-belit rendel hozzá. Ennél sokkal többet mondunk f-ről, ha kölcsönösen egyértelműnek (vagyis bijektívnek) nevezzük, mert egy tetszőleges f függvény sem az injektivitás, sem a szürjektivitás követelményét nem feltétlen teljesíti. Ha például A a (valaha élt) férfiak, B pedig a nők halmaza és az $f : A \to B$ függvény minden férfihoz az édesanyját rendeli, akkor f valóban függvény (hiszen minden férfinak egyetlen édesanyja van), de se nem injektív (mert egy nőnek több fia is lehet), se nem szürjektív (mert nem minden nőnek van fia). Ha viszont A és B a jelenleg házasságban élő (magyar állampolgár) férfiak, illetve nők halmazát jelöli és az $f : A \to B$ függvény minden férfihoz a feleségét rendeli, akkor f most is függvény (mert Magyarországon nincs többnejűség) és ráadásul injektív is (mert többférjűség sem lehetséges) és szürjektív is (mert minden feleségnek van férje). Amit tehát szemléletesen úgy fejezünk ki, hogy A és B elemeit párba állítjuk, azt precízen egy $f : A \to B$ bijekcióval adhatjuk meg (hasonlóan ahhoz, ahogyan az iménti f párba állította a házaspárok tagjait). Ez a szemlélet áll az alábbi definíció mögött.

3.1.2. Definíció. *Azt mondjuk, hogy a (tetszőleges) A és B halmazok* számossága egyenlő*, ha létezik egy $f : A \to B$ bijekció. Ezt a tényt így jelöljük: $|A| = |B|$.*

Ennek a definíciónak a birtokában már visszatérhetünk a fejezetet nyitó kérdéshez: ha $\mathbb{N}$ jelöli a természetes számok, P pedig páros természetes számok halmazát, akkor igaz-e, hogy $|\mathbb{N}| = |P|$? A válasz igenlő: például az $f : \mathbb{N} \to P$, $f(n) = 2n$ függvény bijektív és így bizonyítja az $|\mathbb{N}| = |P|$ állítást. (El kell tehát fogadnunk, hogy egy végtelen halmaz lehet egyenlő számosságú a saját valódi részhalmazával – ez következik a fenti definícióból.) Ennek ellenére, a fejezet elején másodjára véleményt nyilvánító képzeletbeli beszélőnek annak dacára sem volt igaza, hogy

véletlenül helyeset állított: látni fogjuk, hogy két végtelen halmaz számossága nem feltétlen egyenlő.

Érdemes megfigyelni, hogy a 3.1.2. Definíció által elvárt f bijekció messze nem egyértelmű. Akár az $|\mathbb{N}| = |P|$ állítást is bizonyíthattuk volna egy másik függvénnyel (például ha a $g : \mathbb{N} \to P$ függvényre $g(0) = 2$, $g(1) = 0$ és minden $n \geq 2$ esetén $g(n) = 2n$, akkor g szintén bijekció). Általában, ha az $|A| = |B|$ állítást akarjuk belátni, akkor ezt bármilyen $f : A \to B$ bijekció megadásával (vagy létezésének bizonyításával) megtehetjük, a lehetőségek széles köre (és a kreativitás tág tere) nyílik meg. Az sem elvárás, hogy az f bijekció valamilyen tömören leírható „képlettel" legyen megadva, bármilyen világosan definiált függvény megfelel.

A halmazok számossága közti egyenlőség fenti definíciója természetesen véges halmazokra is működik és ebben az esetben $|A| = |B|$ nyilván ekvivalens azzal, hogy A és B elemszáma azonos. Ez az eset azonban érdektelennek tekinthető, a 3.1.2. Definíció valódi újdonságot a végtelen halmazok világában hoz. (Ennek megfelelően, végtelen halmazok „elemszámáról" a továbbiakban sem fogunk beszélni – a „számosság" elnevezés szándékosan különbözik ettől.)

Fontos megjegyezni, hogy az egyenlő számosságú halmaz jelölésében (és nevében) az egyenlőség valóban úgy viselkedik, ahogyan azt elvárjuk. Így $|A| = |A|$ minden A halmazra igaz (ezt bizonyítja az $f(x) = x$ identitás függvény), $|A| = |B|$ esetén $|B| = |A|$ is fennáll (mert ha $f : A \to B$ bijekció, akkor az $f^{-1} : B \to A$ inverz függvény is az), valamint az $|A| = |B|$ és $|B| = |C|$ állításokból $|A| = |C|$ is következik (mert ha $f : A \to B$ és $g : B \to C$ bijekciók, akkor a $g \circ f : A \to C$ kompozíciójuk is az). (Ezeket a tulajdonságokat fejezzük ki sorban azzal, ha az $|A| = |B|$ fogalmát *reflexív*, *szimmetrikus* és *tranzitív* relációnak nevezzük.)

3.1.3. Feladat. Igaz-e az $|A| = |B|$ állítás az alábbi A és B halmazokra?
a) $A = \mathbb{N}$ és B a (pozitív) prímszámok halmaza;
b) $A = (0,1)$ és $B = (2,100)$ (ahol (a,b) a valós számok a és b végpontú nyílt intervallumát, vagyis az $\{x : a < x < b, x \in \mathbb{R}\}$ halmazt jelöli);
c) $A = \mathbb{R}$ és $B = (0,\infty)$ (a pozitív valós számok halmaza);
d) $A = (0,1)$ és $B = \mathbb{R}$;
e) $A = [0,\infty)$ (a nemnegatív valós számok halmaza) és $B = (0,\infty)$;
f) $A = [0,1]$ és $B = (0,1)$ (ahol a [,] zárt intervallumot jelöl).

Megoldás: a) Számozzuk meg a prímszámokat növekvő sorrendben, de kezdjük a sorszámokat a nullával: $p_0 = 2$, $p_1 = 3$, $p_2 = 5$, stb. (Más szóval: minden $n \in \mathbb{N}$ esetén p_n az $(n+1)$-edik prímet jelöli.) Mivel a prímek száma végtelen, ezért ezzel végtelen sorozatot definiáltunk. Ez egyben igazolja az $|A| = |B|$ állítást is, hiszen $n \mapsto p_n$ bijekció A-ról B-re.

b) Könnyű olyan függvényt találni, amely A és B között kölcsönösen egyértelmű megfeleltetést teremt: az $f : x \mapsto 98x+2$ függvény bijekció A-ról B-re, így bizonyítja az $|A| = |B|$ állítást.

c) A válasz ismét igen: például az $x \mapsto 2^x$ függvény igazolja ezt.

d) Kicsit több leleményességgel itt is találunk a középiskolából ismert valós függvények közt olyat, ami bizonyítja az $|A| = |B|$ állítást. Az $x \mapsto \operatorname{tg} x$ függvényt

megszorítva a $(-\frac{\pi}{2},\frac{\pi}{2})$ intervallumra bijekciót kapunk $(-\frac{\pi}{2},\frac{\pi}{2})$ és $\mathbb{R}$ között. A $(0,1)$ és a $(-\frac{\pi}{2},\frac{\pi}{2})$ intervallumok között pedig a b) feladat ötletét felhasználva könnyű bijekciót mutatni: $x\mapsto \pi\cdot x-\frac{\pi}{2}$. Ezzel megmutattuk az $|A|=|(-\frac{\pi}{2},\frac{\pi}{2})|$ és a $|(-\frac{\pi}{2},\frac{\pi}{2})|=|B|$ állításokat, amiből a fenti megjegyzés szerint következik $|A|=|B|$ is. (Valóban, az $f:x\mapsto \operatorname{tg}\left(\pi\cdot x-\frac{\pi}{2}\right)$ függvény bijekció A-ról B-re.)

e) Egy $f:A\to B$ bijekció megalkotásához egyetlen elemet, a 0-t kell „eltüntetnünk". Ennél már jóval nagyobbnak tűnő feladat sem okozott problémát: a természetes számokon értelmezett $n\mapsto 2n$ függvény minden páratlan számot eltüntetett. Ha csak a 0-t kellett volna eltüntetni, az $n\mapsto n+1$ függvény is megfelelt volna – ez tehát bizonyítja az $|\mathbb{N}|=|\mathbb{N}^+|$ állítást (ahol $\mathbb{N}^+$ a pozitív egészek halmazát jelöli). Nem nehéz rájönni, hogy ugyanez a gondolat az $|A|=|B|$ állítást is bizonyítja – annyival kell csak kiegészíteni, hogy a $[0,\infty)\setminus\mathbb{N}$ halmazon a függvény legyen az identitás. Összefoglalva: az $f(x)=\begin{cases} x+1, & \text{ha } x\in\mathbb{N}\\ x, & \text{ha } x\notin\mathbb{N}.\end{cases}$ függvény egy $f:A\to B$ bijekció, így $|A|=|B|$.

f) Próbáljuk meg az e) feladat megoldását alkalmasan módosítani. Az nem tűnik valódi problémának, hogy most két elemet (a 0-t és az 1-et) kell „eltüntetni", az $n\mapsto n+2$ függvény ezt megtenné. Nagyobb különbségnek látszik, hogy most A-nak és B-nek nem részhalmaza $\mathbb{N}$, pedig az előbb épp ez nyelte el a 0-t. Valójában azonban már az e) feladatban is csak kényelmes volt $\mathbb{N}$-et használni, helyette bármilyen végtelen számsorozat megfelelt volna. Most például használhatjuk az 1-nél nagyobb egészek reciprokait, vagyis az $\frac{1}{2},\frac{1}{3},\frac{1}{4},\dots$ sorozatot: ha ennek a tagjait „eltoljuk" két pozícióval, akkor a felszabadul két „hely" a 0 és az 1 számára. Összefoglalva: az

$$f(x)=\begin{cases} \frac{1}{2}, & \text{ha } x=0,\\ \frac{1}{n+2}, & \text{ha } x=\frac{1}{n} \text{ valamely } n\geq 1 \text{ egészre},\\ x, & \text{egyébként.}\end{cases}$$

függvény egy $f:A\to B$ bijekció, így $|A|=|B|$ ismét igaz. □

A fenti feladat a) részének megoldásából fontos következtetést vonhatunk le: a gondolatmenet szempontjából érdektelen, hogy a kérdés épp a prímszámokról szólt, a természetes számok bármely végtelen részhalmazáról ugyanígy elmondhatjuk, hogy az egyenlő számosságú $\mathbb{N}$-nel. Az ilyen halmazoknak különös fontosságuk miatt saját nevük van, ezt vezeti be az alábbi definíció.

3.1.4. Definíció. *Az A halmazt* megszámlálhatóan végtelennek *nevezzük, ha a számossága egyenlő a természetes számok halmazáéval (vagyis $|A|=|\mathbb{N}|$). Ennek a jele: $|A|=\aleph_0$.*

Egyes tankönyvek a *megszámlálható halmaz* elnevezést is bevezetik a véges vagy megszámlálhatóan végtelen halmazokra; bár ezt mi a továbbiakban nem használjuk, hasznos rögzíteni, hogy ez a két kifejezés különböző tartalommal bír. Később látni fogjuk, hogy nem minden végtelen halmaz megszámlálhatóan végtelen –

mégis, a „megszámlálhatatlanul végtelen” kifejezés nem létezik a szakirodalomban (helyette „nem megszámlálhatóan végtelen” használható). A definícióban bevezetett jelölés a héber „alef” ($\aleph$) betűt használja; ez a jelölés még Cantortól származik és széles körben elterjedt. (Az $\aleph_0$ jelölésből sejthető, hogy az indexben a 0 helyett más szimbólumok is előfordulhatnak; ez ugyan igaz, de itt nem foglalkozunk vele.)

A fentiekből kiderült, hogy a természetes számok bármely végtelen részhalmaza megszámlálhatóan végtelen; az alábbiakban viszont azt vizsgáljuk meg, hogy mi mondható az $\mathbb{N}$-nél bővebb számhalmazokról.

3.1.5. Tétel. *Az egész számok $\mathbb{Z}$ halmaza és a racionális számok $\mathbb{Q}$ halmaza egyaránt megszámlálhatóan végtelen.*

Bizonyítás: Mindkét esetben egy $(a_n) = (a_0, a_1, a_2, \ldots)$ számsorozat megadása lesz a célunk úgy, hogy a sorozat bármely két tagja különböző legyen és az értékkészlete (vagyis az $\{a_0, a_1, a_2, \ldots\}$ számhalmaz) $\mathbb{Z}$, illetve $\mathbb{Q}$ legyen. Ezzel csak átfogalmazzuk (és valamivel szemléletesebbé tesszük) a bizonyítandó állítást, hiszen így (hasonlóan a 3.1.3. Feladat a) részének a megoldásához) valójában az $f : n \mapsto a_n$ bijekciót adjuk meg $\mathbb{N}$ és $\mathbb{Z}$, illetve $\mathbb{N}$ és $\mathbb{Q}$ között (ezáltal definíció szerint bizonyítva a $|\mathbb{Z}| = \aleph_0$, illetve a $|\mathbb{Q}| = \aleph_0$ állítást).

$\mathbb{Z}$ esetében viszonylag egyszerűen készíthető ilyen sorozat, például így:

a_0	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	$\ldots$
0	1	-1	2	-2	3	-3	4	-4	5	$\ldots$

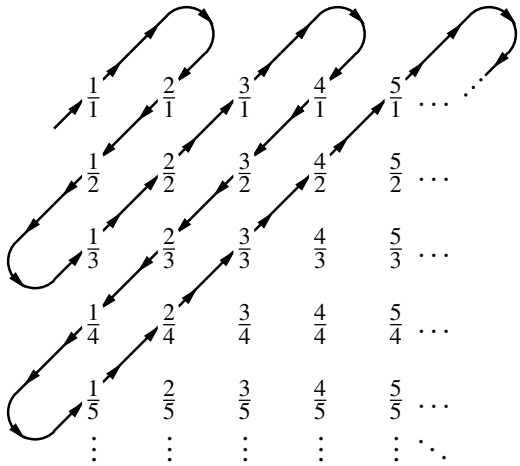
Bár nincs erre szükség, de ez a sorozat elég egyszerű ahhoz, hogy akár képlettel is megadhassuk: $a_n = \begin{cases} -k, & \text{ha } n \text{ páros és } n = 2k, \\ k, & \text{ha } n \text{ páratlan és } n = 2k-1. \end{cases}$ Látszik, hogy az (a_n) sorozat valóban megfelel a feltételeknek: a tagjai kimerítik $\mathbb{Z}$-t és nincs köztük két azonos. Így (a_n) (és az $f : n \mapsto a_n$ bijekció) bizonyítja az $|\mathbb{N}| = |\mathbb{Z}|$ állítást.

$\mathbb{Q}$ esetében egy megfelelő (a_n) sorozat készítésének az alapötletét a 3.1. ábra mutatja. Ezen a pozitív egész számlálójú és nevezőjű törteket egy jobbra és lefelé is végtelen „táblázatban” rendeztük el: a j-edik sorban az i-edik helyen áll $\frac{i}{j}$ (ahol $i, j \geq 1$ egészek). Az ábrán látható bejárás szerint ezek a törtek elrendezhetők egyetlen sorozatban:

$$\frac{1}{1}, \frac{2}{1}, \frac{1}{2}, \frac{1}{3}, \frac{2}{2}, \frac{3}{1}, \frac{4}{1}, \frac{3}{2}, \frac{2}{3}, \frac{1}{4}, \frac{1}{5}, \frac{2}{4}, \frac{3}{3}, \frac{4}{2}, \frac{5}{1}, \frac{6}{1}, \frac{5}{2}, \ldots$$

Persze ezzel a sorozattal egyelőre két probléma is van: egyrészt csak a pozitív racionális számokat tartalmazza, másrészt ezek mindegyikét többször (valójában végtelen sokszor), hiszen ugyanannak a racionális számnak minden bővített alakját külön felsorolja (például $\frac{1}{2}$, $\frac{2}{4}$, $\frac{3}{6}$, … mind külön szerepel a sorozatban). Mindkét problémán könnyű segíteni. Először is hagyjuk el a sorozat minden olyan tagját, amellyel egyenlő értékű tört korábban más alakban már szerepelt:

$$\frac{1}{1}, \frac{2}{1}, \frac{1}{2}, \frac{1}{3}, \frac{3}{1}, \frac{4}{1}, \frac{3}{2}, \frac{2}{3}, \frac{1}{4}, \frac{1}{5}, \frac{5}{1}, \frac{6}{1}, \frac{5}{2}, \ldots$$



3.1. ábra

Ezzel egy olyan $r_1, r_2, r_3, \ldots$ sorozatot kaptunk, amely minden pozitív racionális számot pontosan egyszer sorol fel. Ebből pedig a $\mathbb{Z}$-re vonatkozó fenti bizonyítás gondolatát használva már könnyen készíthető olyan sorozat, amely minden racionális számot pontosan egyszer érint: $0, r_1, -r_1, r_2, -r_2, r_3, -r_3, \ldots$ Ezzel tehát (minden $n \in \mathbb{N}$-hez az így készült sorozat $(n+1)$-edik tagját rendelve) valóban egy $\mathbb{N} \to \mathbb{Q}$ bijekciót adtunk meg és így a tételt beláttuk. □

Érdemes külön kiemelni a fenti (két) bizonyításnak azt a gondolatát, hogy ha egy A halmaz megszámlálhatóan végtelen voltát akarjuk belátni, akkor ezt egy, az A minden elemét pontosan egyszer felsoroló $(a_n) = (a_0, a_1, a_2, \ldots)$ számsorozat megadásával tehetjük meg. Valóban, ez csak a megfogalmazását tekintve különbözik attól, hogy az $|\mathbb{N}| = |A|$ állítást az $f : n \mapsto a_n$ bijekcióval mutatjuk meg. (Amint azt a fenti bizonyításban láttuk, az (a_n) sorozat akár többször is tartalmazhatja az A egyes elemeit, a lényeges követelmény csak az, hogy A minden elemét tartalmazza. Valóban, a korábban már szerepelt A-beli elemek ugyanúgy elhagyhatók, mint ahogyan azt az $|\mathbb{N}| = |\mathbb{Q}|$ állítás bizonyításában tettük.) Ezt szokás úgy kifejezni, hogy az A halmaz elemeit *sorozatba rendezzük* – és így a megszámlálhatóan végtelen halmaz fogalma azonos a *sorozatba rendezhető* halmazéval.

$\mathbb{N}$, $\mathbb{Z}$ és $\mathbb{Q}$ után a következő természetes kérdés a valós számok $\mathbb{R}$ halmazára vonatkozik: vajon még ez is egyenlő számosságú $\mathbb{N}$-nel? Cantor egyik első, alapvető fontosságú felfedezése éppen az volt, hogy a válasz itt már nemleges. Ez egyben az első példa arra is, hogy két végtelen halmaz számossága nem feltétlen egyenlő.

3.1.6. Tétel. (Cantor)
A valós számok $\mathbb{R}$ halmaza nem megszámlálhatóan végtelen, vagyis $|\mathbb{N}| \neq |\mathbb{R}|$.

Bizonyítás: Indirekt bizonyítunk: tegyük fel, hogy $|\mathbb{N}| = |\mathbb{R}|$, vagyis (a fentiek sze-

rint) $\mathbb{R}$ sorozatba rendezhető. Ebből következik, hogy a $(0,1)$ nyílt intervallum elemei is sorozatba rendezhetők: valóban, az $\mathbb{R}$ elemeit felsoroló sorozatból hagyjuk ki az ezen az intervallumon kívül eső elemeket. Létezik tehát egy $(a_n) = (a_1, a_2, a_3, \ldots)$ sorozat, ami minden 0-nál nagyobb és 1-nél kisebb valós számot felsorol.

Ellentmondásra úgy fogunk jutni, hogy konstruálunk egy olyan $b \in (0,1)$ valós számot, ami hiányzik az (a_n) sorozatból. Ehhez az (a_n) tagjainak a tizedestört alakját fogjuk használni: ez a sorozat minden tagjára a „0," jelekkel kezdődik (hiszen 0 és 1 közötti számokról van szó), ezután pedig számjegyek (vagyis 0 és 9 közötti egészek) egy végtelen sorozata következik. (Úgy tekintjük, hogy a véges tizedestörteknek is végtelen sok tizedesjegye van: egy ponttól kezdve csupa 0.) Jelölje minden $n, i \geq 1$ esetén az a_n tizedestört alakjában (a tizedesvessző után) az i-edik tizedesjegyet $x_{n,i}$ (lásd a 3.2. ábrát).

$$
\begin{aligned}
a_1 &= 0,\boxed{x_{1,1}}\, x_{1,2}\, x_{1,3}\, x_{1,4}\, x_{1,5}\, \ldots \\
a_2 &= 0,x_{2,1}\, \boxed{x_{2,2}}\, x_{2,3}\, x_{2,4}\, x_{2,5}\, \ldots \\
a_3 &= 0,x_{3,1}\, x_{3,2}\, \boxed{x_{3,3}}\, x_{3,4}\, x_{3,5}\, \ldots \\
a_4 &= 0,x_{4,1}\, x_{4,2}\, x_{4,3}\, \boxed{x_{4,4}}\, x_{4,5}\, \ldots \\
a_5 &= 0,x_{5,1}\, x_{5,2}\, x_{5,3}\, x_{5,4}\, \boxed{x_{5,5}}\, \ldots \\
&\vdots \qquad\qquad\quad \vdots
\end{aligned}
$$

3.2. ábra

A keresett $b \in (0,1)$ számot is tizedestört alakban fogjuk megadni, az n-edik tizedesjegyét (a tizedesvessző után) y_n fogja jelölni: $b = 0,y_1\ y_2\ y_3\ y_4 \ldots$ A b elkészítéséhez az $x_{n,n}$ tizedesjegyeket (vagyis a 3.2. ábrán látható végtelen táblázat „főátlóját") fogjuk használni. Az y_1 megválasztásánál csak arra figyeljünk, hogy az $x_{1,1}$-től különbözzön: ha $y_1 \neq x_{1,1}$, akkor már biztosak lehetünk benne, hogy $b \neq a_1$. Hasonlóan, ha y_2-t az $y_2 \neq x_{2,2}$ feltételnek megfelelően választjuk, akkor $b \neq a_2$. Általában: y_n lehet bármilyen számjegy (legalábbis majdnem – lásd a következő bekezdést), csak $y_n \neq x_{n,n}$ teljesüljön. Az így konstruált $b \in (0,1)$ valós szám valóban különbözik az (a_n) sorozat minden tagjától: a_n-től eltér az n-edik tizedesjegyében, így $b \neq a_n$. Ez ellentmond annak, hogy (a_n) a $(0,1)$ minden elemét felsorolja, amivel a tételt beláttuk.

Ehhez a gondolatmenethez egy apró, technikai jellegű kiegészítést kell fűzni ahhoz, hogy valóban tökéletes legyen. Ismert ugyanis, hogy a valós számok tizedestört alakja nem mindig egyértelmű: ha egy véges tizedestört utolsó (nem nulla) jegyét 1-gyel csökkentjük és onnantól csupa 9-eseket írunk, akkor a kapott tizedestört alak más, de ugyanazt a számot jelöli. (Például: $0,5432 = 0,54319999\ldots$) Ez a fenti bizonyításban problémát okozhatna: előfordulhatna, hogy az elkészült b tizedestört alakja ugyan különbözik az (a_n) sorozat minden tagjáétól, de az értéke mégis megegyezik valamelyikkel. Emellett ha b minden tizedesjegyét óvatlanul 0-nak vagy

mindet 9-nek választjuk, akkor a kapott b nem is volna a $(0,1)$ intervallumban – az utóbbi esetben $0,9999\ldots = 1$ miatt. Persze ezeken a problémákon könnyű segíteni: tiltsuk be a 9-est és a 0-t az y_n-ek megválasztásánál – így is marad legalább hét választási lehetőség még $x_{n,n}$ kizárása után is. (Nem feltétlen szükséges, de akár teljesen pontossá is tehetjük y_n megválasztását, például így: $y_n = \begin{cases} 2, & \text{ha } x_{n,n} = 1 \\ 1, & \text{ha } x_{n,n} \neq 1 \end{cases}$.) Ezzel a kiegészítéssel pedig a bizonyítás már hézagmentes. □

A fenti bizonyítás alapgondolatát – vagyis a 3.2. ábrán látható táblázat főátlójától való tagonkénti eltérést – *átlós módszernek* szokás nevezni. Ezt Cantor fedezte fel épp a fenti bizonyításhoz, de azóta sok más helyzetben is alkalmazták.

Hasonlóan ahhoz, ahogyan $\mathbb{N}$ „prototípusként” szolgált a megszámlálhatóan végtelen halmaz fogalmához, $\mathbb{R}$ révén is egy hasonló, alapvető fontosságú fogalmat definiálhatunk.

3.1.7. Definíció. *Az A halmazt* kontinuum számosságúnak *nevezzük, ha a számossága egyenlő a valós számok halmazáéval (vagyis $|A| = |\mathbb{R}|$). Ennek a jele: $|A| = c$.*

Például a 3.1.3. Feladat b)–f) részeinek megoldásából levonható közös tanulság, hogy a valós számegyenes bármely nem elfajuló (vagyis legalább két pontot tartalmazó) részintervalluma kontinuum számosságú. (Cantor eredetileg a gót „c” betűvel jelölte a kontinuum számosságot, sok tankönyv ezt máig is megőrizte; mi a mára szintén elterjedt latin betűs jelölést használjuk.)

3.2. Nagysági reláció a halmazok számossága között

A 3.1.6. Tételben beláttuk, hogy $|\mathbb{N}| \neq |\mathbb{R}|$; talán „szívesen” mondanánk, hogy ebből $|\mathbb{N}| < |\mathbb{R}|$ következik, de ennek az állításnak egyelőre nem adtunk értelmet, csak az egyenlő (és ezáltal a nem egyenlő) fogalmát vezettük be. Az alábbi definíció ezt pótolja.

3.2.1. Definíció. *Legyenek A és B (tetszőleges) halmazok.*

(i) Azt mondjuk, hogy A számossága kisebb vagy egyenlő *B számosságánál, ha létezik egy $f : A \to B$ injektív függvény. Ezt a tényt így jelöljük: $|A| \leq |B|$ (vagy $|B| \geq |A|$).*

(ii) Azt mondjuk, hogy A számossága kisebb *B számosságánál, ha $|A| \leq |B|$, de $|A| \neq |B|$. Ezt a tényt így jelöljük: $|A| < |B|$ (vagy $|B| > |A|$).*

A definíció birtokában már valóban mondhatjuk például, hogy $|\mathbb{N}| < |\mathbb{R}|$: az $|\mathbb{N}| \leq |\mathbb{R}|$ állítást bizonyítja az $f : \mathbb{N} \to \mathbb{R}$ identitás függvény (vagyis $f(n) = n$ minden $n \in \mathbb{N}$-re), ami nyilván injekció, az $|\mathbb{N}| \neq |\mathbb{R}|$ állítást pedig a 3.1.6. Tételben igazoltuk.

Általában is elmondható, hogy ha az A és B halmazokra $A \subseteq B$ fennáll, akkor az $f : A \to B$, $f(a) = a$ identitásfüggvény injekció, így bizonyítja az $|A| \leq |B|$ állítást. Más szóval: ugyan el kellett fogadnunk azt az első hallásra szokatlan gondolatot, hogy egy (végtelen) halmaz lehet egyenlő számosságú a saját valódi részhalmazával, de az az abszurditás azért szerencsére már nem fordulhat elő, hogy egy halmaz kisebb számosságú is lehessen egy részhalmazánál.

Könnyű megérezni a fenti két definíció mögött rejlő intuíciót: egy $f : A \to B$ injekció révén bijekciót adunk meg A és B egy részhalmaza (f értékkészlete) között és ezzel megmutatjuk, hogy A egyenlő számosságú B egy részhalmazával; természetes gondolat ezt megtenni az $|A| \leq |B|$ definíciójának. Ha pedig az $|A| \leq |B|$ és az $|A| = |B|$ fogalmát már definiáltuk, akkor ezekből $|A| < |B|$ jelentése már „magától" adódik. A definíciókat értelmezve tehát $|A| < |B|$ akkor igaz, ha $f : A \to B$ injekció létezik, de $f : A \to B$ bijekció már nem.

Hasznos lehet felhívni a figyelmet a következő, gyakori félreértésre: $|A| < |B|$ *nem azt jelenti*, hogy létezik egy $f : A \to B$ injekció, ami nem bijekció. Így például az $f : \mathbb{N} \to \mathbb{N}$, $f(n) = 2n$ függvény injekció és nem bijekció, de természetesen nem bizonyítja a (hamis) $|\mathbb{N}| < |\mathbb{N}|$ állítást. Az $|A| < |B|$ definíciója ennél sokkal többet mond: létezik egy $f : A \to B$ injekció, de nem csak ez az f nem bijekció, hanem egyáltalán nem létezik bijekció A és B között.

Hasonlóan az $|A| = |B|$ fogalmával kapcsolatban mondottakhoz, az $|A| \leq |B|$ reláció is úgy működik, ahogyan azt az intuíciónk alapján elvárnánk. Például: ha $|A| \leq |B|$ és $|B| \leq |C|$, akkor $|A| \leq |C|$ is igaz – bizonyítja ezt az $f : A \to B$ és a $g : B \to C$ injekciók $g \circ f : A \to C$ kompozíciója. Azonban itt már fölvetődik egy alapvető fontosságú és messze nem magától értetődő kérdés is: vajon az $|A| \leq |B|$ és $|B| \leq |A|$ állításokból együtt következik-e, hogy $|A| = |B|$? Nyilván szívesen válaszolnánk erre is igennel, de a definíciókból kiindulva ez nem megy könnyen: egy $f : A \to B$ és egy $g : B \to A$ injekció létezéséből miért következne egy $h : A \to B$ bijekció létezése? Ezt a kérdést még Cantor sem tudta hiánytalanul megválaszolni – végül egy akkor 19 éves tanítványa, *Felix Bernstein* adta az első teljes bizonyítást.

3.2.2. Tétel. (Cantor, Bernstein)
Az A és B halmazokra $|A| = |B|$ akkor és csak akkor igaz, ha $|A| \leq |B|$ és $|B| \leq |A|$ egyaránt fennállnak.

Bizonyítás: Az állítás „csak akkor" része magától értetődő: ha $|A| = |B|$, akkor létezik egy $h : A \to B$ bijekció; ekkor h, illetve h^{-1} bizonyítják az $|A| \leq |B|$, illetve a $|B| \leq |A|$ állításokat (mert $h : A \to B$ és $h^{-1} : B \to A$ injekciók).

Az „akkor" irány bizonyítása már jóval több munkát igényel. Tegyük fel, hogy $|A| \leq |B|$ és $|B| \leq |A|$ fennállnak, vagyis léteznek az $f : A \to B$ és a $g : B \to A$ injekciók. Azt kell megmutatnunk, hogy ekkor létezik egy $h : A \to B$ bijekció is.

Először is az általánosság megszorítása nélkül feltehetjük, hogy A és B diszjunktak; ha nem így volna, akkor minden $A \cap B$-beli elemet B-ben „lecserélhetnénk" egy új, A-n kívüli elemre anélkül, hogy a bizonyítandó állításon érdemben változtatnánk.

f és g injektivitása azt jelenti, hogy léteznek az f^{-1} és a g^{-1} inverz függvények. Itt például f^{-1} általában nem a teljes B-n, hanem annak csak egy részhalmazán (mégpedig f értékkészletén) van értelmezve; így egy tetszőleges $b \in B$ esetén $f^{-1}(b) \in A$ lehet értelmezett vagy nem az. Hasonlóan, minden $a \in A$ elemhez a $g^{-1}(a) \in B$ elem vagy értelmezett vagy nem.

A bizonyítás alapgondolata az, hogy egy tetszőleges $a \in A$ elemből indulva és az f és g függvényeket felváltva alkalmazva egy végtelen sorozatot kapunk:

$$a \mapsto f(a) \mapsto g\big(f(a)\big) \mapsto f\Big(g\big(f(a)\big)\Big) \mapsto \dots$$

Sőt: ezt a sorozatot a-tól a másik irányba is elindíthatjuk a g^{-1} és az f^{-1} függvények váltott alkalmazásával:

$$\dots \mapsto g^{-1}\Big(f^{-1}\big(g^{-1}(a)\big)\Big) \mapsto f^{-1}\big(g^{-1}(a)\big) \mapsto g^{-1}(a) \mapsto a \mapsto f(a) \mapsto g\big(f(a)\big) \mapsto \dots$$

Jelölje minden $a \in A$ esetén $S(a)$ az a-hoz a fenti módon rendelt (a-tól jobbra és balra is haladó) sorozatot. Persze figyelembe kell vennünk, hogy míg a-tól jobbra az $S(a)$ biztosan a végtelenségig folytatható, addig a-tól balra egy ponton (vagy akár rögtön a-nál) elakadhat – hiszen (a fentiek szerint) f^{-1} és g^{-1} nem minden B, illetve A-beli elemre értelmezett. Így négy esetet különböztethetünk meg:

1. *A-ban elakadó sorozat.* $S(a)$-ban a-tól balra haladva egy A-beli elemnél elakadunk (vagyis vagy $g^{-1}(a)$ eleve nem értelmezett, vagy utoljára az f^{-1} függvényt tudjuk alkalmazni, de a kapott függvényértékre g^{-1}-et már nem).
2. *B-ben elakadó sorozat.* Ezzel analóg módon $S(a)$ az a-tól balra haladva egy B-beli elemnél is elakadhat (ha utoljára g^{-1}-et lehet alkalmazni).
3. *Mindkét irányban végtelen sorozat.* $S(a)$ az a-tól balra indulva is a végtelenségig folytatható, sosem akad el és csupa különböző elemet érint.
4. *Ciklizáló sorozat.* $S(a)$-ban az a-tól jobbra indulva véges sok lépés után újra a-hoz érkezünk. (Ekkor persze a-tól balra indulva is visszajutunk a-hoz, ugyanennyi lépésben.) Nyilván igaz, hogy a újbóli eléréséig ekkor páros sok lépést tettünk meg (mert páratlan sok lépés után B-beli elemnél kell tartsunk).

A bizonyítás kulcsát az a megfigyelés jelenti, hogy az $S(a)$ sorozatok páronként diszjunkt halmazokra vágják $A \cup B$-t. Valóban, az $A \cup B$ minden eleme szerepel egy ilyen sorozatban (egy $a \in A$ például $S(a)$-ban, egy $b \in B$ pedig $S\big(g(b)\big)$-ben) és mindegyik nyilván csak egyben (mert ha egy $x \in A \cup B$-re x az $S(a_1)$-ben és az $S(a_2)$-ben is szerepel, akkor x az a_1-ből és az a_2-ből is elérhető az f, g, f^{-1} és g^{-1} függvények véges sokszori alkalmazásával, így a_1 és a_2 egymásból is elérhetők, következésképp $S(a_1)$ és $S(a_2)$ elemhalmaza azonos).

Ez viszont lehetőséget ad arra, hogy a keresett $h : A \to B$ bijekciót az $A \cup B$ így kapott részein külön-külön adjuk meg. (Ezek a részek a fent felsorolt első három esetben nyilván végtelenek, a ciklizáló sorozatok esetében pedig végesek.) Ha $S(a)$ egy A-ban elakadó sorozat, akkor az elemein f egy bijekciót definiál. Hasonlóan, ha $S(a)$ egy B-ben elakadó sorozat, akkor az elemein g definiál egy bijekciót – és ekkor persze ugyanez g^{-1}-ről is elmondható (amely már $S(a)$-nak az A-beli elemeihez rendeli a B-belieket). Ha pedig $S(a)$ mindkét irányban végtelen vagy ciklizáló, akkor az elemein akár f, akár g^{-1} bijekciót definiál. Az így részenként megadott bijekciókból együtt pedig valóban összeáll a keresett $h : A \to B$ bijekció. □

A fenti tétel gyakran nélkülözhetetlen eszköz a végtelen halmazok számosságával kapcsolatos gondolatmenetekben. Ugyanis az $|A| = |B|$ definíció szerinti igazolásához szükséges $h : A \to B$ bijekció közvetlen megadása helyett gyakran nagyságrendekkel könnyebb egy $f : A \to B$ és egy $g : B \to A$ injekció megadása. Erre több példát is fogunk látni, de első illusztrációként újra bebizonyítjuk az $|\mathbb{N}| = |\mathbb{Q}|$ állítást (amit a 3.1.5. Tételben már megtettünk). Az alábbi bizonyításban szereplő ötlet – a számelmélet alaptételének alkalmazása – egyben számos feladat megoldásában is jól alkalmazható.

3.2.3. Állítás. $|\mathbb{N}| = |\mathbb{Q}|$ *(vagyis* $\mathbb{Q}$ *megszámlálhatóan végtelen).*

Bizonyítás: Először az $|\mathbb{N}| \leq |\mathbb{Q}|$, utána a $|\mathbb{Q}| \leq |\mathbb{N}|$ egyenlőtlenséget látjuk be. Ezekből a 3.2.2. Tétel szerint valóban következni fog $|\mathbb{N}| = |\mathbb{Q}|$. Ezek közül az $|\mathbb{N}| \leq |\mathbb{Q}|$ állítás magától értetődő (bizonyítja ezt az identitásfüggvény, hiszen $\mathbb{N} \subseteq \mathbb{Q}$).

A $|\mathbb{Q}| \leq |\mathbb{N}|$ egyenlőtlenség megmutatásához egy $g : \mathbb{Q} \to \mathbb{N}$ injekciót kell megadnunk. Legyen $r \in \mathbb{Q}$ tetszőleges. Ekkor r felírható $r = (-1)^s \cdot \frac{a}{b}$ alakban, ahol $s \in \{0,1\}$, $a \geq 0$ egész és $b \geq 1$ egész. (Persze r többféleképp is felírható így – ezek közül egy tetszőlegeset válasszunk.) A g függvény r-en felvett értékét definiáljuk így: $g(r) = 2^s \cdot 3^a \cdot 5^b$. Ekkor g valóban injektív, hiszen a számelmélet alaptétele szerint a $g(r)$ függvényérték (egyértelmű) prímtényezős felbontásából s, a és b (és ezáltal r) visszakövetkeztethető. (Érdemes megfigyelni, hogy g nem bijekció: csak olyan természetes számokat vesz föl értékként – ezek közül sem mindet –, amelyeknek a prímtényezős felbontásában nincs 5-nél nagyobb prím.) □

3.2.4. Feladat. Mi a számossága az alábbi halmazoknak?
a) azon térvektorok halmaza, amelyeknek mindhárom koordinátája racionális;
b) azon (tetszőleges, de véges magasságú) oszlopvektorok halmaza, amelyeknek minden koordinátája racionális;
c) az irracionális számok $\mathbb{Q}^*$ halmaza.

Megoldás: a) Jelölje a szóban forgó halmazt A. Ekkor $|\mathbb{N}| \leq |A|$ nyilván igaz: például az $f : \mathbb{N} \to A$, $f(n) = (n,0,0)$ függvény injekció. Megadunk még egy $g : A \to \mathbb{N}$ injekciót is; ebből $|A| \leq |\mathbb{N}|$ és ezáltal (a 3.2.2. Tétel szerint) $|A| = |\mathbb{N}|$ következni fog. Legyen $\underline{v} = (r_1, r_2, r_3) \in A$, ahol tehát $r_1, r_2, r_3 \in \mathbb{Q}$. Követve a 3.2.3. Állítás bizonyításának gondolatát, legyen $r_i = (-1)^{s_i} \cdot \frac{a_i}{b_i}$ minden $1 \leq i \leq 3$ esetén, ahol $s_i \in \{0,1\}$, $a_i \geq 0$ és $b_i \geq 1$ egészek. Értelmezzük $g(\underline{v})$-t így:

$$g(\underline{v}) = 2^{s_1} \cdot 3^{a_1} \cdot 5^{b_1} \cdot 7^{s_2} \cdot 11^{a_2} \cdot 13^{b_2} \cdot 17^{s_3} \cdot 19^{a_3} \cdot 23^{b_3}$$

A számelmélet alaptételéből ismét következik, hogy g injekció. Így tehát $|A| = \aleph_0$.

b) A feladatbeli halmazt B-vel jelölve $|\mathbb{N}| \leq |B|$ megint nyilvánvaló: az $f : \mathbb{N} \to B$, $f(n) = (n)$ (ahol a jobb oldalon 1 magasságú oszlopvektor áll) injekció. Az a) feladathoz hasonlóan ismét megadunk egy $g : B \to \mathbb{N}$ injekciót, ami az

iméntivel megegyezően bizonyítani fogja, hogy $|B| = \aleph_0$ is igaz. Ehhez egyszerűen módosíthatnánk is az a) feladat megoldását (több prímet használva az oszlopvektor „elkódolásához"). Ehelyett inkább egy másik megoldást keresünk – de valóban csak a nagyobb változatosság kedvéért. Legyen tehát $\underline{v} \in B$ oszlopvektor, az i-edik koordinátáját jelölje r_i (minden értelmes $i \geq 1$ egészre), továbbá legyen $|r_i| = \frac{a_i}{b_i}$ az $|r_i|$ egy lehetséges tört alakja, ahol $a_i \geq 0$ és $b_i \geq 1$. Kezdetnek írjunk le egy 1-est vagy egy 2-est: 1-est, ha $r_1 < 0$ és 2-est, ha $r_1 \geq 0$; most írjunk le a_1 darab 3-ast és utána b_1 darab 4-est. Járjunk el ugyanígy r_2-vel: az eddig leírtakat folytatva írjunk le egy 1-est vagy egy 2-est r_2 előjelétől függően, majd a_2 darab 3-ast és utána b_2 darab 4-est. Ezt folytassuk egészen addig, amíg $\underline{v}$ koordinátái el nem fogynak. Az így keletkezett (csak 1,2,3 és 4 számjegyekből álló) természetes szám legyen $g(\underline{v})$. Rögtön látszik, hogy $g : B \to \mathbb{N}$ valóban injekció, hiszen a $\underline{v}$-t valóban „elkódoltuk", $g(\underline{v})$ ismeretében $\underline{v}$ egyszerűen rekonstruálható.

c) Láttuk, hogy $\mathbb{Q}$ megszámlálhatóan végtelen, de $\mathbb{R}$ már kontinuum számosságú. Ebből érezhető, hogy az $\mathbb{R} \setminus \mathbb{Q} = \mathbb{Q}^*$ halmaz nem lehet megszámlálhatóan végtelen. Valóban: ha indirekt feltesszük, hogy létezik a $\mathbb{Q}^*$-nak egy $i_1, i_2, i_3, \ldots$ sorozatba rendezése, $r_1, r_2, r_3, \ldots$ pedig a $\mathbb{Q}$ egy sorozatba rendezése (amiről láttuk, hogy létezik), akkor a két sorozat összefésülésével keletkező $r_1, i_1, r_2, i_2, r_3, i_3, \ldots$ sorozat minden racionális és irracionális – vagyis minden valós számot felsorolna. Így azt kapnánk, hogy $\mathbb{R}$ is sorozatba rendezhető, ami ellentmond a 3.1.6. Tételnek.

Fontos hangsúlyozni, hogy ezzel csak azt mutattuk meg, hogy $\mathbb{Q}^*$ nem megszámlálhatóan végtelen, de önmagában *ebből nem következik*, hogy kontinuum számosságú. (Erre a kérdésre a 3.4. szakaszban visszatérünk.) Ettől még a $|\mathbb{Q}^*| = c$ állítás igaz, de teljesen más bizonyítást kíván.

A $|\mathbb{Q}^*| \leq |\mathbb{R}|$ állítás magától értetődő, hiszen $\mathbb{Q}^* \subseteq \mathbb{R}$. Így (ismét a a 3.2.2. Tétel szerint) elég lesz az ellenkező irányú relációt belátnunk. Nagyban egyszerűsíti ezt, ha $\mathbb{R}$ helyett a $(0,1)$ nyílt intervallumot használjuk: egy $g : (0,1) \to \mathbb{Q}^*$ injekciót fogunk megadni. Ezzel közvetlenül a $|(0,1)| \leq |\mathbb{Q}^*|$ állítást látjuk be, de mivel korábban (a 3.1.3. Feladat d) részében) a $|(0,1)| = |\mathbb{R}|$ egyenlőséget már beláttuk, ebből $|\mathbb{R}| \leq |\mathbb{Q}^*|$ is következik.

A $(0,1)$-en értelmezett g függvény hozzárendelési szabálya legyen a következő:

$$g(x) = \begin{cases} x, & \text{ha } x \in \mathbb{Q}^*, \\ x + \sqrt{2}, & \text{ha } x \in \mathbb{Q}. \end{cases}$$

Az nyilvánvaló, hogy g injekció: két különböző $x \in (0,1)$ képe nem lehet azonos, mert a racionális számok képe ($\sqrt{2} > 1$ miatt) 1-nél nagyobb és páronként különböző, az irracionális számokon pedig a függvény azonos az identitással. Azt kell megmutatnunk, hogy minden $x \in (0,1)$-re $g(x) \in \mathbb{Q}^*$. Ez nyilvánvaló akkor, ha $x \in \mathbb{Q}^*$. Legyen most $x \in \mathbb{Q} \cap (0,1)$ és $y = g(x) = x + \sqrt{2}$. Ha $y \in \mathbb{Q}$ teljesülne, akkor ebből $\sqrt{2} \in \mathbb{Q}$ is következne, mert $\sqrt{2} = y - x$ és racionális számok különbsége nyilván racionális. Mivel ismert, hogy $\sqrt{2} \notin \mathbb{Q}$, ebből $y = g(x) \in \mathbb{Q}^*$ valóban következik. Beláttuk tehát, hogy $g : (0,1) \to \mathbb{Q}^*$ injekció, így $|\mathbb{Q}^*| = c$. □

3.3. A kontinuumon túl

Eddig a végtelen halmazok között csak kétféle számosságúval találkoztunk: megszámlálhatóan végtelen és kontinuum számosságú halmazokkal. Felmerül a kérdés, hogy van-e még a kontinuumnál is nagyobb számosságú halmaz? A választ az alábbi, ismét alapvető fontosságú és megint Cantortól származó tétel adja meg: minden halmaznál van nagyobb számosságú halmaz.

A tétel kimondásához elevenítsük fel a következő fogalmat: egy tetszőleges A halmaz *hatványhalmazának* nevezzük és $P(A)$-val jelöljük az A összes (véges és végtelen) részhalmaza által alkotott halmazt. Így például $\emptyset \in P(A)$ és $A \in P(A)$ minden A-ra teljesül – hiszen az üres halmaz és A részhalmaza A-nak; általában $H \in P(A)$ egyenértékű a $H \subseteq A$ állítással. Egy példa: ha $A = \{\text{uk}, \text{muk}, \text{fuk}\}$, akkor $P(A) = \big\{\emptyset, \{\text{uk}\}, \{\text{muk}\}, \{\text{fuk}\}, \{\text{uk}, \text{muk}\}, \{\text{uk}, \text{fuk}\}, \{\text{muk}, \text{fuk}\}, \{\text{uk}, \text{muk}, \text{fuk}\}\big\}$.

3.3.1. Tétel. (Cantor)
Minden A halmazra $|A| < |P(A)|$.

Bizonyítás: A 3.2.1. Definíció szerint az $|A| \leq |P(A)|$ és az $|A| \neq |P(A)|$ állításokat kell belátnunk. Ebből az első szinte magától értetődő: az $f : A \to P(A)$, $f(a) = \{a\}$ függvény (ami tehát minden $a \in A$-hoz az egyedül a-t tartalmazó, 1 elemű részhalmazt rendeli) injektív.

Az $|A| \neq |P(A)|$ állítást indirekt úton bizonyítjuk: tegyük fel, hogy $|A| = |P(A)|$, vagyis létezik egy $f : A \to P(A)$ bijekció. Célunk ebből ellentmondásra jutni.

Az f függvény A elemeihez A részhalmazait rendeli. Így egy tetszőleges $a \in A$ elemre két eset lehetséges: $a \in f(a)$ vagy $a \notin f(a)$ – vagyis a benne van vagy nincs benne az f által sajátmagához rendelt részhalmazban. Az első esetben nevezzük a-t *kedvesnek*, a második esetben *undoknak*. Az undok elemek halmazát pedig jelölje U, vagyis definíció szerint $U = \{a \in A : a \notin f(a)\}$.

Mivel $U \subseteq A$, ezért $U \in P(A)$. Így létezik egy olyan $u \in A$ elem, amire $f(u) = U$, hiszen f bijekció (és így szürjektív is). Két eset lehetséges: u lehet kedves vagy undok.

Először tegyük fel, hogy u kedves. Ez azt jelenti, hogy $u \in f(u)$, vagyis $u \in U$. Így egy kedves elem (nevezetesen u) került az undok elemek U halmazába – ez ellentmondás.

Tegyük fel ezért, hogy u undok. Ez azt jelenti, hogy $u \notin f(u)$, vagyis $u \notin U$. Így egy undok elem (nevezetesen u) kimaradt az undok elemek U halmazából – ez is ellentmondás.

Mindkét lehetőségből ellentmondásra jutottunk, ezzel tehát a tételt beláttuk. □

A fenti tétel segítségével már tudunk olyan végtelen halmazt mutatni, ami se nem megszámlálhatóan végtelen, se nem kontinuum számosságú: $P(\mathbb{R})$ biztosan ilyen (mert $|P(\mathbb{R})| > |\mathbb{R}|$). Sőt: végtelen sok, egymástól páronként különböző számosságú halmazt is tudunk készíteni: az $\mathbb{R}$, $P(\mathbb{R})$, $P\big(P(\mathbb{R})\big)$, ... sorozat tagjai egyre (szigorúan) növekvő számosságúak. De a lehetőségek még ezzel sem merültek ki: az $\mathbb{R}$, $P(\mathbb{R})$, $P\big(P(\mathbb{R})\big)$, ... halmazok mindegyikénél nagyobb számosságú az ezek

uniójaként előálló H halmaz – hiszen H-nak van a sorozat bármelyik tagjánál nagyobb számosságú részhalmaza (például a sorozat következő tagja), így egyikükkel sem lehet azonos számosságú. Ezek után H-ból újra lehetne indítani a H, $P(H)$, $P\big(P(H)\big)$ sorozatot, stb.

Ezzel tehát konstruáltunk végtelen sok, páronként különböző számosságú végtelen halmazt – de végiggondolható, hogy a fenti bekezdésben „csak" megszámlálhatóan végtelen sokat. Valójában létezik $\aleph_0$-nál több számosság is és talán logikusnak hangzik a kérdés, hogy végül is „milyen számosságú a számosságok halmaza" – de ennek a kérdésnek még a pontos feltétele is olyan mély halmazelméleti eszközöket igényelne, amelyek messze túlmutatnak ennek a jegyzetnek a keretein. Anélkül, hogy ennek az állításnak akár a pontos jelentésébe belemennénk, megemlítjük, hogy valójában a kérdés is rossz: számosságokból „olyan sok van", hogy nincs is olyan halmaz, amelyik mindet tartalmazná.

Ehelyett egy sokkal könnyebben megközelíthető kérdéssel zárjuk a szakaszt: a 3.3.1. Tételből következik, hogy $|\mathbb{N}| < |P(\mathbb{N})|$ és korábban az $|\mathbb{N}| < |\mathbb{R}|$ állítást is beláttuk – de mi mondható $\mathbb{R}$ és $P(\mathbb{N})$ egymáshoz való viszonyáról?

3.3.2. Tétel. $|P(\mathbb{N})| = |\mathbb{R}|$

Bizonyítás: A bizonyítást egy hasznos lemmával kezdjük. Ehhez jelölje $\mathbb{B}$ azoknak a (végtelen) számsorozatoknak a halmazát, amelyeknek minden tagja 0 vagy 1 – ezeket röviden bitsorozatoknak fogjuk hívni.

3.3.3. Lemma. $|P(\mathbb{N})| = |\mathbb{B}|$

A Lemma bizonyítása: Az állítást a 3.1.2. Definíció alapján fogjuk belátni: megadunk egy $f : P(\mathbb{N}) \to \mathbb{B}$ bijekciót. Az alapötlet nagyon egyszerű: $\mathbb{N}$ egy tetszőleges H részhalmaza (vagyis $P(\mathbb{N})$ egy eleme) megadható azáltal, ha minden $n \in \mathbb{N}$-ről megmondjuk, hogy H-beli vagy sem. Ezeket az „igen–nem döntéseket" pedig 0-kkal és 1-esekkel kódolhatjuk – más szóval egy bitsorozattal. Pontosabban: a $H \in P(\mathbb{N})$ részhalmaz esetén $f(H)$ az a $(b_n) = (b_0, b_1, b_2, \ldots) \in \mathbb{B}$ sorozat lesz, amelyre minden $n \in \mathbb{N}$ esetén $b_n = \begin{cases} 1, & \text{ha } n \in H, \\ 0, & \text{ha } n \notin H. \end{cases}$ Az alábbi táblázat példaként három H részhalmazt és a hozzájuk rendelt $f(H) \in \mathbb{B}$ sorozatokat mutatja.

H		b_0	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	…
{páratlan számok}	$\mapsto$	0	1	0	1	0	1	0	1	0	1	…
{prímszámok}	$\mapsto$	0	0	1	1	0	1	0	1	0	0	…
$\{0,4,6\}$	$\mapsto$	1	0	0	0	1	0	1	0	0	0	…

Az így definiált $f : P(\mathbb{N}) \to \mathbb{B}$ pedig valóban bijektív, mert minden $(b_n) \in \mathbb{B}$ bitsorozatnak létezik és egyértelműen rekonstruálható az őskép: ez az azok által az n-ek által alkotott $H \subseteq \mathbb{N}$ részhalmaz, amelyekre $b_n = 1$. ◇

A tétel bizonyításához, a fenti lemma és a korábban (a 3.1.3. Feladat d) részében) már igazolt $|(0,1)| = |\mathbb{R}|$ állítás alapján elég lesz megmutatnunk, hogy $|\mathbb{B}| = |(0,1)|$. Ezt a 3.2.2. Tétel alapján tesszük: megmutatjuk, hogy $|\mathbb{B}| \leq |(0,1)|$ és $|(0,1)| \leq |\mathbb{B}|$.

A $|\mathbb{B}| \leq |(0,1)|$ állításhoz egy $f : \mathbb{B} \to (0,1)$ injekciót kell mutatnunk. Ez nagyon egyszerűen megtehető: egy tetszőleges $(b_n) \in \mathbb{B}$ képe legyen az a $(0,1)$-beli valós szám, amelynek a tizedestört alakja $0,b_0\ b_1\ b_2\ b_3\ \ldots$ Mivel azonban a csupa nullákból álló $\mathbb{B}$-beli sorozat képe így a 0 volna, ami nincs a $(0,1)$-ben, ezért ehhez a sorozathoz f rendelje mondjuk a 0,5-öt. Ekkor f valóban injekció, hiszen $f\big((b_n)\big)$ ismeretében, annak tizedestört alakjából (ami a csupa nullához rendelt 0,5 kivételével csak 0-kat és 1-eseket tartalmazhat) (b_n) egyértelműen visszaállítható.

A $|(0,1)| \leq |\mathbb{B}|$ megmutatásához szükséges $g : (0,1) \to \mathbb{B}$ injekció konstrukciójának ötlete nagyon hasonló: egy tetszőleges $x \in (0,1)$ tizedestört alakjának minden jegye elkódolható 4 biten (ehhez egyszerűen kettes számrendszerben írjuk fel a tizedesjegyeket és szükség esetén az elejére annyi 0-t írunk, hogy éppen 4 bitet kapjunk). Ezeknek a 4 tagú bitsorozatoknak az egymás után fűzéséből álljon definíció szerint $g(x)$. Például az $x = \frac{1}{\pi} = 0,3183\ldots$ esetében a $g(x) \in \mathbb{B}$ sorozat így kezdődik: $0,0,1,1,0,0,0,1,1,0,0,0,0,0,1,1,\ldots$ Látszik, hogy g valóban injektív: $g(x)$-ből x tizedestört alakját egyértelműen „dekódolhatjuk”, ha a sorozatot 4 bitnyi darabokra tördeljük és mindegyiket visszaírjuk 10-es számrendszerbe.

Végül érdemes megemlíteni, hogy bár ebben a bizonyításban is a valós számok tizedestört alakjával dolgoztunk, most nem okozott problémát, hogy a tizedestört alak nem mindig egyértelmű (lásd a 3.1.6. Tétel bizonyítását): a fenti f esetében csak csupa 0 és 1 tizedesjegyeket tartalmazó tizedestört alakokra szorítkoztunk, márpedig ilyenből minden $(0,1)$-beli valós számnak legföljebb egy van; a g esetében pedig x bármelyik tizedestört alakját választhatjuk $g(x)$ definiálásához. □

Érdemes megfigyelni, hogy a fenti tételből és a 3.3.1. Tételből együtt új bizonyítást kapunk a 3.1.6. Tételre: valóban, a $|P(\mathbb{N})| = |\mathbb{R}|$ és a $|P(\mathbb{N})| > |\mathbb{N}|$ állításokból $|\mathbb{R}| > |\mathbb{N}|$ következik. Azonban a fenti tétel ennél sokkal hasznosabb: ha egy halmazról ki akarjuk mutatni, hogy az kontinuum számosságú, akkor ezt gyakran megkönnyíti, ha „referenciának” nem $\mathbb{R}$-et, hanem $P(\mathbb{N})$-et, vagy még inkább $\mathbb{B}$-t használjuk. Az alábbi feladatban erre is látunk példát.

3.3.4. Feladat. Mi a számossága az alábbi halmazoknak?
a) $\mathbb{R}^2$, vagyis a sík pontjainak halmaza;
b) a valós számsorozatok halmaza (jelölje ezt $\mathbb{S}$);
c) az $f : \mathbb{R} \to \mathbb{R}$ függvények halmaza (jelölje ezt $\mathbb{F}$).

Megoldás: a) A 3.3.2. Tételből (illetve annak bizonyításából) tudjuk, hogy $|\mathbb{R}| = |\mathbb{B}|$ (ahol tehát $\mathbb{B}$ továbbra is a bitsorozatok halmazát jelöli). Rögzítsünk ezért egy $g : \mathbb{R} \to \mathbb{B}$ bijekciót – tudjuk, hogy ilyen létezik. Definiálni fogunk egy $f : \mathbb{R}^2 \to \mathbb{B}$ bijekciót g segítségével – ezzel tehát igazolni fogjuk, hogy a sík pontjainak számossága is kontinuum (hiszen $\mathbb{B}$-ről ezt már beláttuk). Legyen $(x,y) \in \mathbb{R}^2$ tetszőleges és g rendelje x-hez, illetve y-hoz a $g(x) = (x_1,x_2,x_3,\ldots) \in \mathbb{B}$, illetve a $g(y) = (y_1,y_2,y_3,\ldots) \in \mathbb{B}$ sorozatokat. Ezek után $f\big((x,y)\big)$ legyen a $g(x)$ és a $g(y)$

összefésülésével keletkező sorozat: $f\big((x,y)\big) = (x_1, y_1, x_2, y_2, x_3, y_3, \ldots) \in \mathbb{B}$. Az így definiált $f : \mathbb{R}^2 \to \mathbb{B}$ függvény valóban bijektív, mert minden $(b_n) \in \mathbb{B}$ bitsorozathoz egyértelműen rekonstruálható egy olyan $(x,y) \in \mathbb{R}^2$ pont, amelyre $f\big((x,y)\big) = (b_n)$: x-et, illetve y-t úgy kapjuk, hogy a (b_n) páratlan, illetve páros sorszámú tagjai által alkotott bitsorozatokra alkalmazzuk g^{-1}-et.

b) A sík pontjai a 2 hosszúságú valós számsorozatoknak felelnek meg, ezeknek a halmazáról láttuk be, hogy kontinuum számosságú. Ehhez képest most a végtelen hosszú valós számsorozatok halmazát vizsgáljuk. Így talán meglepő, hogy $\mathbb{S}$ számossága még mindig „csak" kontinuum. Ezt ismét egy $f : \mathbb{S} \to \mathbb{B}$ bijekció elkészítésével igazoljuk, amihez megint az a) feladat megoldásában rögzített $g : \mathbb{R} \to \mathbb{B}$ bijekciót hívjuk segítségül. Legyen tehát $(s_n) = (s_1, s_2, s_3, \ldots) \in \mathbb{S}$ tetszőleges számsorozat. Készítsünk el egy, a 3.1. ábrán láthatóhoz hasonló, jobbra és lefelé is végtelen táblázatot: ennek az i-edik sorában a $g(s_i)$ bitsorozat álljon minden $i \geq 1$-re (vagyis az i-edik sorban a j-edik helyen a $g(s_i)$ bitsorozat j-edik tagja található minden $i, j \geq 1$ esetén). Majd a táblázat elemeit a 3.1. ábrán látható kígyóvonal mentén fűzzük fel és az így kapott (egyetlen) bitsorozat legyen definíció szerint $f\big((s_n)\big)$. Ezzel valóban egy $f : \mathbb{S} \to \mathbb{B}$ bijekciót adtunk meg: tetszőleges $(b_n) \in \mathbb{B}$ bitsorozat tagjait a 3.1. ábra kígyóvonala mentén leírva, majd a kapott (végtelen) táblázat soraira a g^{-1} függvényt alkalmazva egyértelműen rekonstruálhatók annak az $(s_n) \in \mathbb{S}$ sorozatnak a tagjai, amelyre $f\big((s_n)\big) = (b_n)$.

c) A 3.3.3. Lemma bizonyításának alapgondolatát átvihetjük $P(\mathbb{N})$-ről $P(\mathbb{R})$-re is: egy tetszőleges $H \in P(\mathbb{R})$ (vagyis $H \subseteq \mathbb{R}$) részhalmaznak megfeleltethetjük azt az $f : \mathbb{R} \to \{0,1\}$ függvényt, amelyre minden $x \in \mathbb{R}$ esetén $f(x) = \begin{cases} 1, & \text{ha } x \in H \\ 0, & \text{ha } x \notin H \end{cases}$. Ezzel egy $P(\mathbb{R}) \to \mathbb{F}$ injekciót adtunk meg (ami mellesleg bijekció $P(\mathbb{R})$ és az $f : \mathbb{R} \to \{0,1\}$ függvények halmaza között). Ezzel beláttuk, hogy $|P(\mathbb{R})| \leq |\mathbb{F}|$. Megmutatjuk az $|\mathbb{F}| \leq |P(\mathbb{R})|$ relációt is, vagyis megadunk egy $h : \mathbb{F} \to P(\mathbb{R})$ injekciót. Ehhez felhasználjuk a már megoldott a) feladatot: legyen $g : \mathbb{R}^2 \to \mathbb{R}$ egy rögzített bijekció. Tetszőleges $f : \mathbb{R} \to \mathbb{R}$ függvény *grafikonja* alatt a síknak az azon (x,y) pontjai által alkotott részhalmazát értjük, amelyekre $f(x) = y$. Alkalmazzuk most egy tetszőleges $f : \mathbb{R} \to \mathbb{R}$ függvény grafikonjának minden pontjára a g függvényt, így $\mathbb{R}$ egy részhalmazát kapjuk; legyen ez a részhalmaz definíció szerint a h függvénynek az $f \in \mathbb{F}$-en felvett értéke. Ekkor $h : \mathbb{F} \to P(\mathbb{R})$ valóban injekció, hiszen egy tetszőleges $H \in P(\mathbb{R})$ részhalmazhoz egyértelműen visszakereshető az az $f \in \mathbb{F}$ függvény, amelyre $h(f) = H$: ehhez a H minden pontjára a g^{-1} függvényt alkalmazva megkapjuk f grafikonját és így f-et is.

Beláttuk tehát a $|P(\mathbb{R})| \leq |\mathbb{F}|$ és az $|\mathbb{F}| \leq |P(\mathbb{R})|$ relációkat, így a 3.2.2. Tétel szerint $|\mathbb{F}| = |P(\mathbb{R})|$. Ez a legtöbb, amit az $\mathbb{F}$ számosságáról mondani tudunk (hiszen a $P(\mathbb{R})$-rel azonos számosságú halmazokra nem vezettünk be külön elnevezést). □

3.4. A kontinuumhipotézis

A végtelen halmazok között eddig láttunk megszámlálhatóan végtelen és kontinuum számosságúakat és megmutattuk, hogy léteznek $\mathbb{R}$-nél is nagyobb számosságú halmazok – sőt, még ezen belül is végtelen sok további lehetőség van. Egy fontos kérdés azonban nyitva maradt: van-e számosság $\aleph_0$ és c között? Pontosabban: létezik-e olyan H halmaz, amelyre $|\mathbb{N}| < |H| < |\mathbb{R}|$? Ezt a kérdést Cantor is felvetette és azt sejtette, hogy a válasz nemleges: *kontinuumhipotézisnek* nevezte el azt az állítást, amely egy ilyen H létezésének a lehetetlenségét mondja ki. Cantor hosszan próbálkozott a kontinuumhipotézis bizonyításával, sikertelenül.

A huszadik század elejére ez a kérdés a kor matematikájának egyik legfontosabb nyitott problémájává nőtte ki magát. David Hilbert, a kor egyik legjelentősebb és legnagyobb hatású matematikusa 1900-ban megfogalmazott 23 megoldatlan problémát, amelyekről azt gondolta, hogy ezek a matematika jövőjét jelentős mértékben meghatározzák; ezek közül is az első helyre tette a kontinuumhipotézist. Ekkorról származik Hilbert azóta híressé vált mondata is: „Senki sem űzhet ki bennünket abból a paradicsomból, melyet Cantor teremtett nekünk."

Ma már ismerjük a valóságot a kontinuumhipotézissel kapcsolatban, de a válasz mind Cantort, mind Hilbertet jócskán meglepné. A helyzet ugyanis a következő: nem létezik olyan bizonyítás, amely a kontinuumhipotézist bebizonyítaná és olyan sem, amely megcáfolná azt. Nem abban az értelemben nem léteznek ezek a bizonyítások, hogy még senki sem talált ilyeneket: mindkét bizonyítás létezésének a lehetetlensége bizonyított tétel – az előbbit *Paul Cohen* látta be 1963-ban, az utóbbit *Kurt Gödel* 1940-ben. (Kicsit pontosabban ez a következőt jelenti: ha feltételezzük, hogy a halmazelméletet – és ezáltal a matematikát – megalapozó, a szakirodalomban ZFC-vel jelölt axiómarendszerből nem lehet ellentmondást levezetni, akkor ezt az axiómarendszert akár a kontinuumhipotézissel, akár annak a tagadásával kiegészítve ismét olyan axiómarendszereket kapunk, amelyekben nem lehet ellentmondást levezetni.)

A huszadik század elejéig a matematikusok körében elfogadott volt az a nézet, hogy a matematikában minden „értelmesen megfogalmazott" kérdés megválaszolható, legalábbis elvileg; lehet, hogy egy állítás bizonyítása vagy cáfolata olyan nehéz vagy hosszú, hogy soha nem leszünk képesek megtalálni azt, de elvileg legalábbis létezik. Ehhez képest az 1930-as évek elején megrázkódtatásként érték a matematikát (és a filozófiát) Gödel híres nemteljességi tételei: ezek azt állítják, hogy bárhogyan is választunk egy olyan axiómarendszert, ami egyrészt nem használhatatlanul semmitmondó, másrészt nem lehet benne ellentmondást levezetni, mindenképp megfogalmazható olyan állítás, amit sem bebizonyítani, sem megcáfolni nem lehet ebben az axiómarendszerben. (Ez persze csak egy durván leegyszerűsített megfogalmazása Gödel eredményeinek, eredeti formájukban ezek tökéletesen precízen kimondott, erősen formalizált állítások.)

A kontinuumhipotézis pedig az egyik első példa volt arra, hogy az ilyen, a fenti értelemben eldönthetetlen állítások létezése nem csak egy életidegen elvi lehetőség: egészen konkrét, a matematika fejlődése során természetesen felvetődő kérdések is bizonyulhatnak eldönthetetlennek.

3.5. Egy alkalmazás az informatika világából

Létezik-e olyan, például C nyelven írt program, amely egy tetszőleges $n \geq 1$ egész bemenetre kiszámítja a π szám n-edik tizedesjegyét? Elsőre kétségtelenül nem tűnik könnyűnek ez a programozási feladat, de megoldható – ehhez az analízis közelítő módszereit kellene segítségül hívni (de végül maga a kód nem is feltétlen volna rendkívül bonyolult). A π minél pontosabb kiszámítása régóta egyfajta sokakat lázba hozó (bár minden praktikus szempontból meglehetősen értelmetlen) versengéssé nőtte ki magát, a mai rekord már tíztrillió (10^{13}) tizedesjegy fölött van. Bár ma még a legmodernebb szuperszámítógépek tár- és időkapacitása sem elegendő ahhoz, hogy a π-nek például a 10^{50}-edik tizedesjegyét kiszámítsuk (és biztosra vehető, hogy az emberiség – még ha ennél égetőbb problémái is bőven akadnak – sosem fogja megismerni a π-nek a 10^{1000}-edik tizedesjegyét), ez nem változtat azon, hogy létezik olyan (nem is túlságosan bonyolult) programkód, ami ezt *elvileg* megtenné.

Nevezzünk egy $f : \mathbb{N} \to \mathbb{N}$ függvényt *kiszámíthatónak*, ha létezik olyan (mondjuk C nyelven írt) programkód, ami minden $n \in \mathbb{N}$ bemenetre elvileg helyesen kiszámítja $f(n)$ értékét. Nyilván minden programozással hivatásszerűen foglalkozó szakember számára megnyugtató volna a tudat, hogy minden $f : \mathbb{N} \to \mathbb{N}$ függvény kiszámítható, de a valóság sajnos ennek az ellenkezője.

3.5.1. Tétel. *Létezik olyan $f : \mathbb{N} \to \mathbb{N}$ függvény, amely nem kiszámítható.*

Bizonyítás: Jelölje F az $f : \mathbb{N} \to \mathbb{N}$ függvények halmazát és jelölje K ezek közül a kiszámíthatóak halmazát. A bizonyítás azon alapszik, hogy meghatározzuk mindkét halmaz számosságát.

F számosságát valójában már ismerjük: ehhez csak azt kell (újra) észrevennünk, hogy az $f : \mathbb{N} \to \mathbb{R}$ függvények azonosíthatók a valós számsorozatokkal. Valóban: az $f : \mathbb{N} \to \mathbb{R}$ függvény helyett beszélhetünk az $(f(0), f(1), f(2), \ldots)$ számsorozatról is, ez csak szóhasználatbeli különbséget jelent. Így F azokból a számsorozatokból áll, amelyeknek minden tagja természetes szám. Korábban (a 3.3.2. Tételben, illetve a 3.3.4. Feladat b) részében) már beláttuk, hogy a bitsorozatok $\mathbb{B}$ halmaza és az összes valós számsorozat $\mathbb{S}$ halmaza egyaránt kontinuum számosságú. Mivel (az F elemeinek természetes számokból álló sorozatokkal való azonosítása után) $\mathbb{B} \subseteq F \subseteq \mathbb{S}$ fennáll, ezért (a 3.2.2. Tétel szerint) F is kontinuum számosságú.

A K halmazról viszont belátjuk, hogy megszámlálhatóan végtelen. $|\mathbb{N}| \leq |K|$ szinte magától értetődő: minden $m \in \mathbb{N}$-re a konstans m függvény (vagyis amelyre $f(n) = m$ minden n-re) nyilván kiszámítható, amivel tehát megadtunk egy $g : \mathbb{N} \to K$ injekciót. A $|K| \leq |\mathbb{N}|$ bizonyításához egy $h : K \to \mathbb{N}$ injekciót mutatunk. Legyen tehát $f : \mathbb{N} \to \mathbb{N}$ kiszámítható függvény és legyen P egy olyan programkód, ami $f(n)$-et minden $n \in \mathbb{N}$ bemenetre kiszámítja. P természetesen – mint minden fájl – felfogható egy véges hosszúságú bitsorozatként. (Bár a P kódra nyilván inkább egy szövegfájlként gondolunk, ez karakterenként megfeleltethető 8 bitnek, P pedig ezek egymás után fűzéséből áll.) Írjunk ez elé a bitsorozat elé (például) egy 2-est és az így kapott (az első 2-estől eltekintve csupa 0 és 1 számjegyeket tartalmazó) természetes

számot rendelje a h függvény az f-hez. Ekkor $h : K \to \mathbb{N}$ valóban injekció, hiszen $h(f)$-ből az első 2-est levágva egyértelműen visszanyerhető a P programkód és ezáltal az f kiszámítható függvény. (A $h(f)$ elejére azért kellett egy 2-est írni, mert a P-nek megfelelő bitsorozat 0-kkal is kezdődhet.) A fentiekből tehát a 3.2.2. Tétel szerint valóban következik, hogy $|K| = \aleph_0$.

Beláttuk tehát, hogy F kontinuum számosságú, míg K csak megszámlálhatóan végtelen. Így valóban kell léteznie K-ba nem tartozó F-beli függvénynek. □

A fenti bizonyításból még sokkal sötétebb kép rajzolódik ki, mint a tétel állításából: nem csak hogy létezik nem kiszámítható függvény, hanem valójában az $f : \mathbb{N} \to \mathbb{N}$ függvények túlnyomó többsége ilyen. Ennek ellenére, a bizonyításból csak (rengeteg) ilyen függvény létezése következik, ez alapján egyetlen konkrét ilyen függvényt sem tudunk mutatni. Hasonlóan az előző szakasz végén írtakhoz, a nem kiszámítható függvények létezése megint nem csak egy életidegen, elvi lehetőség: konkrét gyakorlati alkalmazások is felvethetnek olyan feladatokat, amelyekre egyszerűen nem készíthető program. Mindezekről bővebben az informatikus képzés későbbi tárgyaiban lesz szó.

3.6. Ajánlott irodalom

Bár hasonlókat a korábbi fejezetek végén is mondtunk, ennek a témájára még inkább érvényes, hogy csak a felszínt borzoltuk meg a fenti rövid ismertetőben. A terület iránt mélyebben érdeklődő olvasóknak az alábbi tankönyveket ajánljuk:

[1] Hajnal András, Hamburger Péter: *Halmazelmélet*, Nemzeti Tankönyvkiadó, Budapest, 1983, 1994.

[2] Komjáth Péter: *Halmazelmélet* (digitális egyetemi jegyzet), Budapest, 2007. http://www.cs.elte.hu/~kope/oktatas/ma1.pdf

Az algoritmikus megoldhatóság kérdésében elmélyedni kívánóknak pedig az alábbi könyvet ajánljuk:

[3] Lovász László: *Algoritmusok bonyolultsága*, Nemzeti Tankönyvkiadó, Budapest, 2001. Digitális változat:
http://www.cs.elte.hu/~kiraly/alg.pdf

Köszönetnyilvánítás

Hálásan köszönöm azoknak a hallgatóknak, akik a jegyzetben talált kisebb-nagyobb hibákat jelezték és ezzel hozzájárultak annak fejlődéséhez, javulásához.

A 2014. őszi félév hallgatói közül: **Czövek Mártonnak**, Harcsa-Pintér Bálintnak, Kemenes Balázsnak, László Balázsnak, Tóth Kristófnak és Vincze Ádámnak.

A 2015. őszi félév hallgatói közül: Balogh Norbertnek, **Juhos Attilának**, Király Péternek, Nagy Marcellnek, Neubrandt Dórának, **Péterfalvi Ferencnek**, Sgánetz Bencének, Szakács Béla Benedeknek és Varga Rudolfnak.

A 2016. őszi félév hallgatói közül: Baranyai Gergelynek, Hübner Krisztiánnak, Kertész Gergőnek, Pelles Kingának és Varga Flóriánnak.

A 2017. őszi félév hallgatói közül: Balog Istvánnak és Szőke Tibornak.