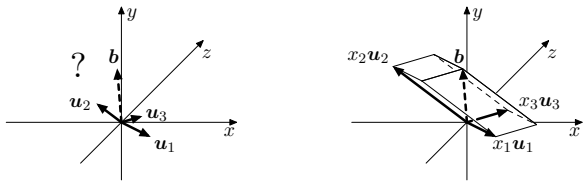
Povídání o lineární algebře

Pavel Klavík

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



verze 8. ledna 2015

Obsah

Předmluva

Lineární algebru považuji za neuvěřitelně krásnou oblast matematiky. Pamatuji si, že jako student prvního ročníku jsem z ní nebyl příliš nadšený. Působila na mě jenom jako seznam prázdných definic a tvrzení, jejichž důkazy byly spíše triviální. Kouzlo lineární algebry se však skrývá v dobře zvolených základních pojmech, které jsou velice propojené a kolem kterých je následně vybudována. Je až neuvěřitelné, kolik různých pohledů lze objevit na některé klíčové definice.

Samotný úvod lineární algebry lze vybudovat více způsoby. Lze na ni pohlížet jako na (abstraktní) teorii vektorových prostorů a lineárních zobrazení. Lze ji také budovat jako zobecnění geometrie. My však zvolíme praktičtější směr (podobně jako v přednášce na matfyzu) a budeme se pokoušet řešit soustavy lineárních rovnic. Čtenář se však nemusí obávat, že by byl o teorii vektorových prostorů nebo geometrii ochuzen, záhy se k ní také dostaneme.

Tento text nemůže vzhledem ke svému rozsahu složit jako kompletní učebnice lineární algebry (a ani nemá takové ambice). Snaha je popsat lineární algebru spíše shora, vysvětlit klíčové definice a pojmy. Jak už to bývá, pokud člověk chce něco pořádně pochopit, měl by se zabývat základními pojmy a vztahy. Bez pochopení definic nemá smysl pokračovat a třeba dokazovat složité věty. Budeme rádi, pokud tento text povede čtenáře k zamyšlení nad lineární algebrou; ostatně jenom tak lze dosáhnout hlubšího pochopení.

V textu naleznete celou řadu poznámek pod čarou. V těch se snažíme o vysvětlení z jiného úhlu, například často budeme srovnávat pojmy lineární algebry s jinými pojmy v jiných oblastech matematiky, například analýze. Důležité je, že poznámky pod čarou lze z textu vynechat. Pokud čtenář některé z nich nebude rozumět, může bez obav pokračovat ve čtení. Některé obtížnější bonusové pasáže jsou označeny hvězdičkami a čtenář je může bez obav přeskočit. Pokud je u odkazu na některou kapitolu uveden otazník, znamená to, že dosud nebyla napsána a bude uvedena v budoucí verzi textu.

Klíčové poznatky, které by si čtenář měl z dané kapitoly odnést, jsou uvedeny v rámečku, například:

> Protože tento text teprve vzniká, provádím v něm řadu úprav. Nezapomeňte si před čtením stáhnout aktuální verzi:
> `http://pavel.klavik.cz/vyuka/texty/povidani_o_la.html`.

Jako žádný text ani tento není bez chyb. Pokud nějakou chybu objevíte, napište mi prosím na `klavik@kam.mff.cuni.cz`. Za řadu poznámek k textu bych rád poděkoval Aleně Bušákové, Milanu Hladíkovi, Jiřímu Matouškovi, Anetě Šťastné, Tomáši Vyskočilovi a Petrovi Zemanovi. Tento text byl podpořen grantem CE-ITI — P202/12/G061 GAČR.

Kapitola 1

Soustavy lineárních rovnic

Soustavy lineárních rovnic se zkoumají v matematice od dob dávných a postup zvaný Gaussova eliminace určený na jejich řešení je znám již od staré Číny.[1] Cílem této kapitoly je soustavy lineárních rovnic popsat a vysvětlit Gaussovu eliminaci.

1.1 Úvod

Začněme příkladem soustavy tří rovnic o třech neznámých x, y a z:

$$\begin{aligned} 2x &- y &&= -1,\\ -x &+ 2y &- z &= 3,\\ &- y &+ 2z &= 1. \end{aligned} \tag{1.1}$$

Hledáme všechny možné hodnoty reálných čísel x, y a z, které splňují všechny tři rovnice současně. Každé přiřazení hodnot proměnným x, y a z se nazývá *ohodnocení* a zapíše se jako uspořádaná trojice (α, β, γ). Ohodnocení $(1, 3, 0)$ (tedy $x = 1$, $y = 3$ a $z = 0$) není řešením soustavy, protože nesplňuje druhou a třetí rovnici: $-1+6-0 \neq 3$ a $-3+0 \neq 1$. Můžete ověřit, že ohodnocení $(1, 3, 2)$ řeší soustavu. *Množina všech řešení soustavy* je množina všech ohodnocení, která splňují všechny rovnice současně. V případě této soustavy je $(1, 3, 2)$ jediné řešení, což brzo ukážeme.

Obecně soustava m lineárních rovnic o n neznámých vypadá takto:

$$\begin{aligned} a_{1,1}x_1 &+ a_{1,2}x_2 &+ a_{1,3}x_3 &+ \cdots + a_{1,n}x_n &= b_1\\ a_{2,1}x_1 &+ a_{2,2}x_2 &+ a_{2,3}x_3 &+ \cdots + a_{2,n}x_n &= b_2\\ a_{3,1}x_1 &+ a_{3,2}x_2 &+ a_{3,3}x_3 &+ \cdots + a_{3,n}x_n &= b_3\\ \vdots & & \vdots & \ddots & \vdots\\ a_{m,1}x_1 &+ a_{m,2}x_2 &+ a_{m,3}x_3 &+ \cdots + a_{m,n}x_n &= b_m \end{aligned} \tag{1.2}$$

[1] Matematik Gauss pochopitelně nebyl Číňan. Toto je příklad jednoho z řady pojmů, které se jmenují po někom, kdo není jejich původním objevitelem. Konkrétní soustavy lineárních rovnic se v matematice objevují již před více než pěti tisíci lety. V čínské knize *Devět kapitol* napsané zhruba dvě stě let před naším letopočtem se objevuje postup, jak vyřešit konkrétní soustavu tří lineárních rovnic o třech neznámých. Jakékoliv formální zdůvodnění chybí, ale při zobecnění dostaneme Gaussovu eliminaci.

V západní matematice Gaussovu eliminaci poprvé popsal Newton ve své algebraické knize, i když se jednalo o postup mezi matematiky běžně známý. Gauss se zabýval *metodou nejmenších čtverců* (v souvislosti s geodézií), pro níž popsal algoritmus podobný Gaussově eliminaci. Jméno Gaussova eliminace se začalo používat až v padesátých letech díky implementaci metody nejmenších čtverců v počítačích. Historie řady pojmů lineární algebry je značně složitá!

Reálná čísla a_{ij} se nazývají koeficienty, reálná čísla b_i jsou pravé strany, pro konkrétní soustavu všechny a_{ij} a b_i známe. Symboly $x_1, \ldots, x_n$ označují neznámé, jejichž ohodnocení chceme nalézt.[2] Množina všech řešení soustavy je množina všech n-tic reálných čísel $(x_1, \ldots, x_n)$, které splňují současně všechny rovnice.

Ukažme si, jak vypadá obecný zápis v případě soustavy (1.1). To je soustava tří lineárních rovnic o třech neznámých, tedy $n = 3$ a $m = 3$. Pro neznámé platí

$$x_1 = x, \qquad x_2 = y \quad \text{a} \quad x_3 = z.$$

Koeficienty mají následující hodnoty:

$$\begin{array}{llll} a_{1,1} = 2, & a_{1,2} = -1, & a_{1,3} = 0, & b_1 = -1, \\ a_{2,1} = -1, & a_{2,2} = 2, & a_{2,3} = -1, & b_2 = 3, \\ a_{3,1} = 0, & a_{3,2} = -1, & a_{3,3} = 2, & b_3 = 1. \end{array}$$

Všimněte si, jak jsou koeficienty číslovány: $a_{i,j}$ znamená koeficient na i-tém řádku a v j-tém sloupci. Toto pořadí je zavedená konvence, kterou budeme dodržovat.

Co to znamená, že rovnice jsou *lineární*? Levá strana je tvořena součtem proměnných vynásobených nějakými (pevnými) koeficienty, pravá strana je nějaké pevné číslo. Rovnici splňují ta ohodnocení proměnných, pro které po dosazení dostaneme na levé straně stejné číslo jako na pravé straně. Lineární rovnice (a jejich soustavy) patří k tomu nejjednoduššímu v matematice. V tomto textu si ukážeme, že jsou dostatečně zajímavé a mají celou řadu aplikací.

Soustava m lineárních rovnic o n neznámých je zadána koeficienty $a_{i,j}$ a b_i pro $i = 1, \ldots, m$ a $j = 1, \ldots, n$. Hledáme všechna ohodnocení neznámých $(x_1, \ldots, x_n)$ splňující všechny rovnice soustavy.

Dosazovací metoda. Ukážeme si, jak vyřešit soustavu (1.1) pomocí dosazování. Běžně se dosazování neprovádí a Gaussova eliminace je založena na jiném principu. Avšak dosazování má jednu velkou výhodu: Je snadno vidět, že funguje správně.

Začneme vyjádřením x z první rovnice:

$$x = \tfrac{1}{2}y - \tfrac{1}{2}. \tag{1.3}$$

Pro každé řešení víme, že hodnota x je totožná s hodnotou $\frac{1}{2}y - \frac{1}{2}$. Proto můžeme ve zbývajících dvou rovnicích nahradit výskyty x za $\frac{1}{2}y - \frac{1}{2}$. Dostáváme:

$$\begin{array}{rcrcrcl} -\frac{1}{2}y + \frac{1}{2} & + & 2y & - & z & = & 3, \\ & & -y & + & 2z & = & 1, \end{array} \quad \text{po úpravách} \quad \begin{array}{rcrcrcl} \frac{3}{2}y & - & z & = & \frac{5}{2}, \\ -y & + & 2z & = & 1. \end{array}$$

Získali jsme menší systém pouze dvou rovnic o dvou neznámých, jednu neznámou se nám podařilo vyloučit neboli *eliminovat* (s použitím jedné rovnice). Pokud bychom uměli vyřešit tuto menší soustavu, můžeme dopočítat hodnotu x pomocí (1.3). Jak ale vyřešit tuto menší soustavu? Budeme postupovat dál stejnou metodou, eliminujeme y a soustavu opět zmenšíme:

$$y = \tfrac{2}{3}z + \tfrac{5}{3}, \tag{1.4}$$

Po dosazení do poslední rovnice dostaneme $-\frac{2}{3}z - \frac{5}{3} + 2z = 1$, po úpravách dostaneme $z = 2$. Nyní využijeme vztahy (1.4) a (1.3), abychom dopočítali $y = 3$ a $x = 1$. Tedy dostáváme, že trojice $(1, 3, 2)$ je řešením soustavy. Toto dopočítání řešení se nazývá *zpětná substituce*.

[2] Pokud je neznámých málo, označují se většinou různými písmeny z konce abecedy, třeba výše uvedené x, y a z. Pokud je neznámých více, musíme je číslovat.

Dosazením můžeme zkontrolovat, že $(1, 3, 2)$ je skutečně řešením. To však nepomůže k dokázání, že je to jediné řešení soustavy. Jak dokázat jednoznačnost? Stačí si všimnout, že z původní soustavy jsme dosazováním vyvodili $z = 2$. Hodnota z je v každém řešení jednoznačně určena. Ale podle (1.4) je z hodnoty z jednoznačně určena hodnota y. Tedy i hodnota y je jednoznačně určena. Nakonec z (1.3) je i hodnota x jednoznačně určena z hodnot y a z. Existuje tedy jediné řešení $(1, 3, 2)$.

1.2 Úpravy a počty řešení

Dosazování lze zobecnit na řešení libovolných soustav. Neprovádí se však pohodlně, proto bychom chtěli použít jiný postup, který se snadněji aplikuje.

Odvozování dalších rovnic. Nejprve si ukažme, že ze soustavy můžeme odvodit další rovnice. Budeme požadovat, aby každé řešení původní soustavy splňovalo i tyto nové odvozené rovnice. Pokud tedy odvozené rovnice přidáme do soustavy (čímž zvětšíme počet rovnic), množina řešení se *nezmění*.

V dalším textu budeme označovat strany rovnic velkými písmeny. Například pro rovnici $x - 3y = 5$ označme levou stranu $x - 3y$ jako A a pravou stranu 5 jako B. Tuto rovnici můžeme zapsat jako $A = B$.

Budeme uvažovat dvě základní odvození:

1. *Vynásobení libovolné rovnice číslem:* Z rovnice $A = B$ můžeme odvodit rovnici $\alpha \cdot A = \alpha \cdot B$ pro libovolné $\alpha \in \mathbb{R}$. Všimněte si, že pokud ohodnocení splňuje $A = B$, potom se levá strana A rovná pravé straně B. Tato rovnost platí i pro vynásobené strany $\alpha \cdot A$ a $\alpha \cdot B$. Koeficient α může být i nulový, i když odvozená rovnice nebude zajímavá. Například z rovnice $x - 3y = 5$ můžeme odvodit rovnice $2x - 6y = 10$ (pro $\alpha = 2$), $-\frac{1}{3}x + y = -\frac{5}{3}$ (pro $\alpha = -\frac{1}{3}$) a $0x + 0y = 0$ (pro $\alpha = 0$).
2. *Sečtení dvou libovolných rovnic:* Pokud soustava obsahuje rovnice $A = B$ a $C = D$, můžeme odvodit rovnici $A + C = B + D$. Podobně jako předtím, pro každé řešení soustavy je A rovné B a C rovné D, tedy i součet $A + C$ je rovný $B + D$. Například z rovnic $x + y = 3$ a $x - y = 5$ můžeme odvodit $2x = 8$.

Můžeme vždy odvodit jednu rovnici a přidat ji do soustavy. I odvozené rovnice lze použít k dalšímu odvozování, tedy je možné aplikovat libovolnou sérii těchto operací. Důležité je, že přidáním těchto odvozených rovnic do soustavy žádné řešení neztratíme. A poslední příklad s odvozením $2x = 8$ ukazuje, že odvozené rovnice mohou být jednodušší než ty původní; dozvěděli jsme se, že $x = 4$.

Regulární úpravy. Odvozování má nevýhodu, že soustavu nezjednodušujeme, počet rovnic v ní naopak roste. Budeme postupovat tedy jinak. Chtěli bychom vždy odvodit nějakou novou rovnici a nahradit s ní nějakou jinou rovnici původní soustavy. Počet rovnic tedy zůstane stejný, ale pokud budeme odvozovat správně, bude novou soustavu jednodušší vyřešit. Této operaci budeme říkat *úprava soustavy*.

Musíme však dát pozor, abychom *nezměnili* množinu řešení. Víme, že libovolné řešení původní soustavy bude splňovat i upravenou soustavu. Pokud ale budeme upravovat neopatrně, mohli bychom množinu řešení zvětšit. Úpravy, které nemění množinu řešení, nazýváme *regulární*.[3]

Například pro rovnici $x - 3y = 5$ existuje pro každou volbu y jen jediná volba x, která ji řeší. Její násobek $0x + 0y = 0$ však ztratil tuto informaci a řeší ho libovolné ohodnocení. Úprava, ve které nahradíme $x - 3y = 5$ pomocí $0x + 0y = 0$ není regulární (alespoň ne obecně, pro každou soustavu). Na druhou stranu rovnice $2x - 6y = 10$ obsahuje úplně stejnou informaci a úprava $x - 3y = 5$ na $2x - 6y = 10$ je regulární.

Tvrzení 1.1 (Elementární úpravy)**.** *Uvažme elementární úpravy vynásobení nenulovým číslem a přičtení jedné rovnice k druhé:*

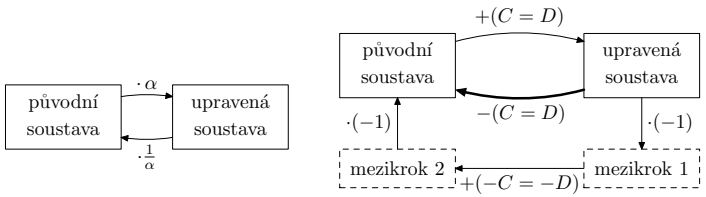
[3] Používá se pro ně také označení *ekvivalentní*.

- *Úprava $A = B$ na $\alpha \cdot A = \alpha \cdot B$ pro $\alpha \neq 0$.*
- *Úprava $A = B$ na $A + C = B + D$, pokud soustava také obsahuje rovnici $C = D$.*

Tyto úpravy jsou regulární.

Důkaz. Stačí ukázat, že jsme schopni upravenou soustavu upravit zpět do původní podoby, tedy že můžeme obě elementární úpravy vrátit (invertovat). Proč? Pokud ohodnocení řeší soustavu, řeší i libovolně upravenou soustavu; množina řešení může s každou úpravou pouze vzrůst. Pokud můžeme aplikovat úpravu oběma směry, musí být množina řešení identická. Kdyby se zvětšila, musela by se po vrácení opět zmenšit, což není možné.

Důkaz je schématicky naznačen na obrázku 1.1. Vynásobení rovnice α můžeme vrátit inverzním vynásobením $\frac{1}{\alpha}$ (proto musí být $\alpha \neq 0$, jinak by $\frac{1}{\alpha}$ neexistovalo!). Přičtení rovnice $C = D$ můžeme vrátit odečtením $C = D$. Že odečtení nemáme jako úpravu? Lze ho snadno získat složením tří úprav: vynásobením $C = D$ pomocí $\alpha = -1$, přičtením $-C = -D$ k $A = B$ a vynásobením $-C = -D$ pomocí $\alpha = -1$ zpět na $C = D$. □



Obrázek 1.1: Schéma důkazu: Nalevo inverze vynásobení, napravo přičtení jedné rovnice k druhé.

Těmto základním úpravám se říká *elementární řádkové úpravy*. Všechny úpravy, které budeme uvažovat, jsou řádkové. Proto budeme zkráceně používat *elementární úpravy*. Pro pohodlí budeme za elementární úpravu také považovat přičtení násobku jedné rovnice k druhé (kterou lze složit pomocí tří úprav, viz důkaz). S rovnicemi navíc budeme pracovat v pevném pořadí. Další elementární úprava je prohození dvou rovnic $A = B$ a $C = D$; čtenář může zkusit odvodit tuto úpravu zřetězením výše uvedených úprav.

Elementární úpravy můžeme na soustavu libovolně aplikovat a máme zaručeno, že se množina řešení nezmění. Úpravu soustavy budeme značit pomocí $\sim$.

Elementární úpravy nemění množinu řešení soustavy. Jejich aplikováním upravíme soustavu do tvaru, v kterém bude snadné nalézt všechna řešení.

Příklad. Ukažme si na příkladu soustavy (1.1), jak lze úpravami zjednodušovat. Nejprve vynásobíme první rovnici $\frac{1}{2}$ a přičteme k druhé rovnici:

$$\begin{array}{rcrcrcrcl} & & x & - & \frac{1}{2}y & & & = & -\frac{1}{2}, \\ \cdots & \sim & -x & + & 2y & - & z & = & 3, \\ & & & & -y & + & 2z & = & 1, \end{array} \qquad \begin{array}{rcrcrcrcl} & & x & - & \frac{1}{2}y & & & = & -\frac{1}{2}, \\ & \sim & & & \frac{3}{2}y & - & z & = & \frac{5}{2}, \\ & & & & -y & + & 2z & = & 1. \end{array}$$

Nyní druhou rovnici vynásobíme $\frac{2}{3}$ a přičteme ke třetí:

$$\begin{array}{rcrcrcrcl} & & x & - & \frac{1}{2}y & & & = & -\frac{1}{2}, \\ \cdots & \sim & & & y & - & \frac{2}{3}z & = & \frac{5}{3}, \\ & & & & -y & + & 2z & = & 1, \end{array} \qquad \begin{array}{rcrcrcrcl} & & x & - & \frac{1}{2}y & & & = & -\frac{1}{2}, \\ & \sim & & & y & - & \frac{2}{3}z & = & \frac{5}{3}, \\ & & & & & & \frac{4}{3}z & = & \frac{8}{3}. \end{array}$$

Upravenou soustavu vyřešíme zpětnou substitucí stejně jako předtím a dostaneme řešení $(1, 3, 2)$. Pokud tento postup zobecníme, dostaneme Gaussovu eliminaci, kterou popíšeme na závěr kapitoly.

Poznámka: Překvapuje vás, že po úpravách vyšly přesně stejné koeficienty jako při předchozím vyjadřování a dosazování? To není žádná náhoda a můžete si to zkusit jako malé cvičení rozmyslet. Mějme dvě rovnice

$$\begin{array}{rcrcrcrcl} a_1x_1 & + & a_2x_2 & + & \cdots & + & a_nx_n & = & b, \\ a'_1x_1 & + & a'_2x_2 & + & \cdots & + & a'_nx_n & = & b'. \end{array}$$

a předpokládejme, že koeficienty a_1 i a'_1 jsou nenulové. Pokud chceme neznámou x_1 eliminovat z druhé rovnice, musíme vynulovat koeficient a'_1. To můžeme udělat přičtením $(-a'_1/a_1)$-násobku první rovnice k druhé, nebo vyjádřením x_1 z první rovnice a dosazením do druhé. Dokažte, že při obou postupech dostaneme u druhé rovnice po úpravách přesně stejné koeficienty.

Maticový zápis. Na předchozím výpočtu si můžeme všimnout, že použitý zápis není úplně ideální. Neustále opisujeme názvy proměnných a přitom nás zajímají pouze koeficienty. Pokud s proměnnými budeme pracovat v předem určeném pořadí (typicky x_1 až x_n), můžeme zapisovat pouze koeficienty uspořádané do tabulky spolu s pravou stranou (oddělenou čarou). Tomuto zápisu se říká *maticový*.

V dalším textu budeme vynechávat nulové koeficienty, pokud je nebudeme chtít explicitně zdůraznit. Budeme používat označení *řádek matice* pro koeficienty příslušící k jedné rovnici a *sloupec matice* pro koeficienty příslušící k jedné neznámé.

Například výše uvedené úpravy můžeme zapsat kompaktněji takto:

$$\left(\begin{array}{ccc|c} 2 & -1 & & -1 \\ -1 & 2 & -1 & 3 \\ & -1 & 2 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & -\frac{1}{2} & & -\frac{1}{2} \\ -1 & 2 & -1 & 3 \\ & -1 & 2 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & -\frac{1}{2} & & -\frac{1}{2} \\ & \frac{3}{2} & -1 & \frac{5}{2} \\ & -1 & 2 & 1 \end{array}\right) \sim$$

$$\left(\begin{array}{ccc|c} 1 & -\frac{1}{2} & & -\frac{1}{2} \\ & 1 & -\frac{2}{3} & \frac{5}{3} \\ & -1 & 2 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & -\frac{1}{2} & & -\frac{1}{2} \\ & 1 & -\frac{2}{3} & \frac{5}{3} \\ & & \frac{4}{3} & \frac{8}{3} \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & -\frac{1}{2} & & -\frac{1}{2} \\ & 1 & -\frac{2}{3} & \frac{5}{3} \\ & & 1 & 2 \end{array}\right).$$

Počet řešení. Pro soustavu (1.1) existuje právě jedno řešení $(1, 3, 2)$, a to jsme nalezli. V obecném případě může existovat řešení spousta, nebo dokonce nemusí existovat žádné. Ukážeme si dva ilustrativní příklady:

$$\begin{array}{rcrcrcl} x & + & 2y & = & 3, \\ 2x & + & 4y & = & 6, \end{array} \qquad \begin{array}{rcrcrcl} x & + & 2y & = & 3, \\ 2x & + & 4y & = & 8, \end{array}$$

v maticovém zápisu s úpravou:

$$\left(\begin{array}{cc|c} 1 & 2 & 3 \\ 2 & 4 & 6 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 2 & 3 \\ 0 & 0 & 0 \end{array}\right), \qquad \left(\begin{array}{cc|c} 1 & 2 & 3 \\ 2 & 4 & 8 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 2 & 3 \\ 0 & 0 & 2 \end{array}\right). \tag{1.5}$$

Rozeberme nejprve soustavu nalevo. Druhou rovnici můžeme odvodit z první (vynásobením), je tedy zcela nadbytečná. Přičtením (-2)-násobku první rovnice k druhé dostaneme upravenou soustavu s rovnicí $0x + 0y = 0$. Tato rovnice je splněna pro každou volbu x a y. V rovnici $x + 2y = 3$ můžeme zvolit neznámou y zcela libovolně. Pro každou hodnotu y je hodnota x určena jednoznačně, platí $x = 3 - 2y$. Soustava má tedy nekonečně mnoho řešení ve tvaru $(3 - 2y, y)$, množinu všech řešení můžeme zapsat následovně:

$$\big\{(3 - 2y, y) : y \in \mathbb{R}\big\}.$$

Dvojtečka v zápisu množiny kvantifikuje přes vlastnost uvedenou napravo.[4] Tento zápis tedy znamená: „Množina všech dvojic $(3 - 2y, y)$, kde y je libovolné reálné číslo.“

[4] Někdy se též místo dvojtečky „:“ používá svislítko „|“.

Řekneme, že rovnice je *nadbytečná*, pokud ji lze odvodit z ostatních rovnic. Nadbytečných rovnic může být víc. Rádi bychom soustavu zredukovali tak, aby žádnou nadbytečnou rovnici neobsahovala. Dokonce může být nadbytečná každá z rovnic (libovolnou rovnici můžeme odebrat a nezměnit řešení), jako v případě soustavy (1.5) nalevo. Nadbytečné rovnice však musíme odebírat po jedné. Po každém odebrání se totiž nadbytečnost rovnic může změnit. Například v případě soustavy (1.5) nalevo po odebrání jedné rovnice již ta druhá není nadbytečná. Gaussova eliminace bude v průběhu výpočtu nadbytečné rovnice vynulovávat.

V případě soustavy napravo jsou první a druhá rovnice v rozporu. Neexistují x a y, které splní obě současně. Pokud přičteme (-2)-násobek první rovnice k druhé, obsahuje upravená soustava rovnici $0x+0y=2$, pro kterou neexistuje žádné řešení. Soustavě, která nemá řešení, budeme říkat *nekonzistentní*. Pokud aplikujeme Gaussovu eliminaci na nekonzistentní soustavu, vyprodukuje neřešitelnou rovnici, která má levou stranu nulovou a pravou stranu nenulovou.

Poznamenejme, že pokud je pravá strana nulová (u všech rovnic), nemůže tato situace nastat. Soustava s nulovou pravou stranou je vždy konzistentní:

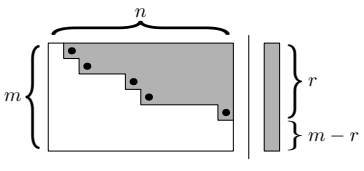
Pozorování 1.2. *Soustava s nulovou pravou stranou má vždy alespoň jedno řešení tvořené samými nulami.*

1.3 Gaussova eliminace

Popíšeme si, jak funguje obecný algoritmus (postup) na řešení soustav lineárních rovnic zvaný Gaussova eliminace. Máme soustavu m rovnic o n neznámých. Gaussova eliminace se provádí ve dvou fázích. První fáze se nazývá *dopředná eliminace* a upraví soustavu do *odstupňovaného tvaru*. Druhá fáze, *zpětná substituce*, dopočítá všechna řešení soustavy z odstupňovaného tvaru.

Nejprve vysvětlíme odstupňovaný tvar. Poté popíšeme obě fáze a dokážeme, že skutečně fungují.

Odstupňovaný tvar. Zaveďme nejprve značení. Budeme ignorovat pravou stranu. Řádek je *nulový*, pokud jsou všechny jeho koeficienty nulové. Řádek je *nenulový*, pokud má alespoň jeden koeficient nenulový. Nenulový koeficient v řádku, který je nejvíc nalevo, budeme označovat jako *vedoucí koeficient*. Během upravování se může pozice vedoucích koeficientů měnit. Nulové řádky nemají vedoucí koeficienty.



Obrázek 1.2: Odstupňovaný tvar soustavy. Bílé pozice jsou nuly a šedé pozice mohou obsahovat jak nuly, tak nenuly. Pivoty vyznačené černými body jsou vždy nenulové.

Odstupňovaný tvar je naznačen na obrázku 1.2, kde neformálně vedoucí koeficienty tvoří schody. Formálně je soustava v odstupňovaném tvaru, pokud jsou splněny dvě podmínky:

- Všechny nenulové řádky se nachází nad nulovými řádky. Označme r počet nenulových řádku a $m-r$ nulových.
- Pozice (souřadnice sloupce) vedoucích koeficientů v r nenulových řádcích jsou ostře rostoucí. Jinými slovy, vedoucí koeficient se v každém řádku nachází víc napravo než všechny vedoucí koeficienty v řádcích nad ním.

V odstupňovaném tvaru se vedoucí koeficienty nazývají *pivoty*, na obrázku jsou vyznačené černě. Pivoty navíc očíslujeme, i-tý pivot je pivot v i-tém řádku. Nulové řádky pochopitelně neobsahují pivot.

První fáze: Dopředná eliminace. Nejprve si připomeňme, které elementární úpravy můžeme používat:

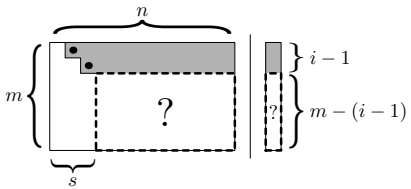
(Ú1) Můžeme libovolný řádek vynásobit *nenulovým* reálným číslem.
(Ú2) Můžeme jeden řádek přičíst k druhému.
(Ú3) Můžeme přičíst libovolný násobek jednoho řádku k druhému.
(Ú4) Můžeme dva řádky matice prohodit.

Poznamenejme, že bychom vystačili pouze s úpravami 1 a 2, neboť ty další dvě se dají složit pomocí několika úprav 1 a 2. Pro úpravu 3 jsme to popsali v důkazu tvrzení 1.1, pro úpravu 4 si to zkuste rozmyslet sami.

Odstupňovaný tvar budeme konstruovat v několika krocích. Na konci $(i-1)$-ního kroku budou platit dvě podmínky:

(A) Prvních $i-1$ řádků je převedeno do výsledného odstupňovaného tvaru, tedy pozice již nalezených pivotů jsou ostře rostoucí.
(B) Nechť se $(i-1)$-ní pivot nachází v s-tém sloupečku. Prvních s koeficientů v řádcích i až m je vynulovaných.

Podmínka říká jinými slovy, že prvních s sloupečků je také ve výsledném odstupňovaném tvaru. Dopředná eliminace již nikdy nebude modifikovat koeficienty v prvních $i-1$ řádcích a prvních s sloupečcích. Stav soustavy na konci $(i-1)$-ního kroku je naznačen na obrázku 1.3.

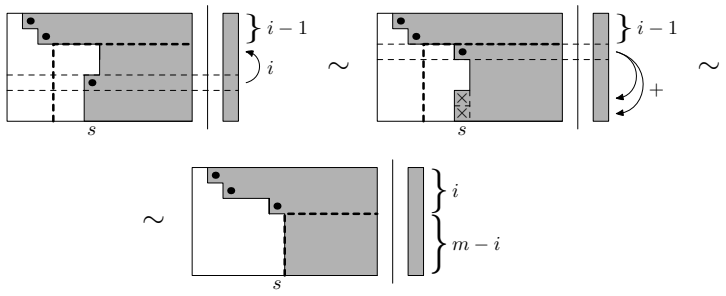


Obrázek 1.3: Matice po $i-1$ krocích dopředné eliminace. Další kroky budou upravovat pouze koeficienty v oblastech vyznačených otazníky.

Popišme, jak probíhá i-tý krok, postup je naznačen na obrázku 1.4. Krok i začíná tam, kde $(i-1)$-ní krok skončil. Pokud jsou zbývající řádky i až m nulové, je soustava v odstupňovaném tvaru (rozmyslete si, že platí obě podmínky) a dopředná eliminace je dokončena. V opačném případě vybereme z řádků i až m ten, jehož vedoucí koeficient je nejvíc vlevo (pokud je víc takových řádků, zvolíme libovolný z nich). Tento řádek bude i-tým řádkem v odstupňovaném tvaru, a proto ho prohodíme se současným řádkem i.

Tvrdíme, že je nyní splněna pro krok i podmínka (A). Tedy tvrdíme, že prvních i řádků je v odstupňovaném tvaru a pozice jejich pivotů jsou rostoucí. Pro pivoty jedna až $i-1$ to platilo, neboť na konci $(i-1)$-ního kroku byla splněna podmínka A. Nově nalezený pivot i (vedoucí koeficient i-tého řádku) je napravo všech z nich, neboť na konci kroku $i-1$ platila podmínka (B), že všechny nenulové koeficienty řádků i až m se nachází napravo od $(i-1)$-ního pivotu.[5]

[5] Povšimněte si, že výše uvedené funguje i v případě, že provádíme první krok dopředné eliminace, je $i=1$. Pivot i je první pivot, který jsme zkonstruovali. Podmínka platí zcela triviálně: Totiž i-tý pivot je napravo od všech předchozích $i-1$ pivotů, protože žádný předchozí pivot není! Tomuto obratu, který činí studentům značné problémy, se říká *kvantifikace*



Obrázek 1.4: Krok i dopředné eliminace. Tlustou čarou je oddělena část matice, která již není modifikována. Nejprve prohodíme řádek s vedoucím koeficientem nejvíc vlevo s i-tým řádkem. Poté vynulujeme koeficienty ostatních řádků ve sloupečku s přičtením vhodných násobků i-tého řádku.

Abychom splnili pro krok i podmínku (B), musíme vynulovat koeficienty řádků $i+1$ až m až po sloupeček s, v kterém se nachází i-tý pivot. Protože jsme volili jako i-tý řádek ten, jehož vedoucí koeficient byl nejvíc vlevo, potřebujeme pouze vynulovat koeficienty ve sloupečku s, ostatní jsou již nulové. Koeficienty ve sloupečku s vynulujeme tak, že přičteme vhodné násobky i-tého řádku. Jaké přesně? K řádku j přičteme $(-a_{j,s}/a_{i,s})$-násobek i-tého řádku. Pokud totiž přičteme tento násobek, vynulujeme s-tý koeficient v j-tém řádku $a_{j,s}$:

$$a_{j,s} + \underbrace{\frac{-a_{j,s}}{a_{i,s}}}_{\text{násobek}} a_{i,s} = 0.$$

Hodnotu násobku lze také vysvětlit takto: Pokud by byl koeficient $a_{i,s} = 1$, přičtením $(-a_{j,s})$-násobku řádku i bychom koeficient $a_{j,s}$ vynulovali. Protože koeficient $a_{i,s}$ může být libovolný, vyrobíme z něj vynásobením $1/a_{i,s}$ onu jedničku. Pokud je koeficient $a_{j,s}$ nulový, není potřeba nic přičítat, ale výše uvedený násobek je stejně nulový a nic nemění.

Všimněte si, že se může stát, že vynulujeme několik dalších sloupečků napravo od s, například v následujícím příkladu. V takovém případě dostaneme v odstupňovaném tvaru "dlouhý schod", kdy v některém sloupečku chybí pivot.

Ukažme si získání odstupňovaného tvaru na konkrétní soustavě. V prvním kroku si vybereme řádek s vedoucím koeficientem nejvíce vlevo, například ten první. Přičteme $(-\frac{1}{2})$-násobek první rovnice k druhé rovnici, $(-\frac{2}{2})$-násobek (tedy (-1)-násobek) ke třetí rovnici a $(-\frac{3}{2})$-násobek ke čtvrté. Tím jsme vynulovali dokonce první dva sloupečky řádků dva až čtyři.

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

V druhém kroku nejprve prohodíme třetí rovnici s druhou, neboť třetí má vedoucí koeficient nejvíc vlevo. Nyní přičteme (-1)-násobek druhé rovnice ke čtvrté rovnici, čímž vynulujeme zbývající dva sloupečky

přes prázdnou množinu a používá se v matematice velice často. Pokud X je prázdná množina a $\mathcal{P}$ je podmínka, tvrzení „Pro každý prvek X platí podmínka $\mathcal{P}$" (velký kvantifikátor) je vždy pravdivé a tvrzení „Existuje prvek X, pro který platí podmínka $\mathcal{P}$" (malý kvantifikátor) je vždy nepravdivé, *nezávisle na tom*, co je podmínka $\mathcal{P}$. Zamyslete se nad tím, proč to dává smysl.

řádků tři až čtyři.

$$\cdots \sim \left(\begin{array}{cccc|c} 2 & 2 & -2 & 0 & -2 \\ & & 1 & 2 & 3 \\ & & & & 0 \\ & & 1 & 2 & 3 \end{array}\right) \sim \left(\begin{array}{cccc|c} 2 & 2 & -2 & 0 & -2 \\ & & 1 & 2 & 3 \\ & & & & 0 \\ & & & & 0 \end{array}\right). \tag{1.6}$$

Tím eliminace končí, protože zbývající rovnice jsou nulové. Dosáhli jsme odstupňovaného tvaru.

Tvrzení 1.3. *Dopředná eliminace vždy upraví soustavu do odstupňovaného tvaru.*

Důkaz. Povšimněte si, že dopředná eliminace se vždy zastaví po nejvýše m krocích. Potřebujeme dokázat, že matice, kterou dostaneme po provedení dopředné eliminace, splňuje dvě podmínky, které pro odstupňovaný tvar požaduje. Před provedením prvního kroku jsou splněny podmínky A a B (pro $i = 0$). Již jsme odargumentovali: Pokud jsou podmínky A a B splněny po provedení $(i-1)$-ního kroku, potom platí i po provedení i-tého kroku. Tedy použitím indukce (podle počtu kroků) dostaneme, že podmínky A a B platí i na konci dopředné eliminace.

Nechť se eliminace zastavila po r krocích. Potom prvních r řádků je nenulových (obsahují pivot). Řádky $r+1$ až m jsou nulové, protože jsme se zastavili. Proto je splněna první podmínka odstupňovaného tvaru (nenulové řádky jsou nad nulovými). Podmínka A přímo říká, že pozice pivotů v prvních r řádcích jsou rostoucí, tedy je splněna i druhá podmínka odstupňovaného tvaru. □

Druhá fáze: zpětná substituce. Dopředná eliminace upravila soustavu do odstupňovaného tvaru. Použila k tomu elementární úpravy, které nemění množinu řešení. Tedy množina řešení původní soustavy a odstupňovaného tvaru je totožná. Zpětná substituce dopočte všechna řešení odstupňovaného tvaru.

Nejprve ověříme, zda je některá z pravých stran odpovídající vynulovaným $m - r$ rovnicím je nenulová. Pokud ano, podařilo se odvodit rovnici, která je nesplnitelná. Soustava je nekonzistentní a nemá žádné řešení. Pokud jsou naopak všechny tyto pravé strany nulové, je soustava konzistentní a řešení existují.

Neznámým, jejichž sloupce neobsahují pivot, říkejme *volné*. Neznámým ve sloupcích s pivotem říkejme *určené*. Volným neznámým můžeme přiřadit libovolnou hodnotu, podle čehož dostaneme různá řešení. Hodnoty určených neznámých jsou již určeny jednoznačně a stačí je dopočítat. Proč? Hodnoty určených neznámých budeme dopočítávat zpětně zprava doleva, vždy s použitím řádku, v kterém je pivot (tedy budeme používat odstupňované řádky odzdola vzhůru). Ve chvíli, kdy dopočítáváme neznámou x_i, známe již hodnoty všech neznámých napravo. Hodnotu x_i dopočítáváme pomocí řádku j, kde již známe hodnoty všech ostatních neznámých s nenulovým koeficientem. Proto je i hodnota x_i jednoznačně určena.

Ukážeme si toto na příkladu odstupňovaného tvaru (1.6):

$$\left(\begin{array}{cccc|c} \underline{2} & 2 & -2 & 0 & -2 \\ & & \underline{1} & 2 & 3 \\ & & & & 0 \\ & & & & 0 \end{array}\right).$$

Dva pivoty jsou vyznačené podtržením. Protože poslední dvě pravé strany jsou nulové, existuje řešení. Pro zvolené hodnoty volných neznámých x_2 a x_4 dopočítáme jednoznačně hodnoty určených neznámých x_1 a x_3:

$$x_3 = 3 - 2x_4, \quad 2x_1 = -2 - 2x_2 + 2x_3.$$

Každou určenou neznámou můžeme vyjádřit pouze pomocí hodnot volných neznámých. Proč? Hodnota určené neznámé x_i záleží pouze na hodnotách (volných i určených) neznámých napravo. Za

hodnoty určených neznámých můžeme dosadit získané odvození. Nejpravější určená neznámá záleží pouze na volných neznámých. Každá další určená neznámá záleží na volných neznámých a na určených neznámých, pro které už známe vyjádření pomocí volných neznámých. Tedy stačí tato vyjádření dosadit a upravit.

Ve výše uvedeném příkladu závisí hodnota x_3 pouze na volných neznámých. Pro hodnotu x_1 to neplatí, a proto dosadíme hodnotu x_3:

$$2x_1 = -2 - 2x_2 + 2x_3 = -2 - 2x_2 + 6 - 4x_4 = 4 - 2x_2 - 4x_4, \text{ tedy } x_1 = 2 - x_2 - 2x_4.$$

Tím hodnoty určených neznámých x_1 a x_3 závisí pouze na volných neznámých x_2 a x_4. Množina všech řešení soustavy je

$$\big\{(2 - x_2 - 2x_4, x_2, 3 - 2x_4, x_4) : x_2, x_4 \in \mathbb{R}\big\},$$

neboť všechna řešení dostaneme z výše uvedených vztahů pro různé volby x_2 a x_4.

Tvrzení 1.4. *Zpětná substituce správně určí množinu všech řešení soustavy.*

Důkaz. Musíme dokázat dvě inkluze: Za prvé každé nalezené přiřazení neznámým je skutečně řešením soustavy. Za druhé žádné řešení nechybí, zpětná substituce nalezla všechna.

Nechť $(x_1, x_2, \ldots, x_n)$ je ohodnocení neznámých, které zpětná substituce nalezla. Musíme ověřit, že splňuje všech r prvních rovnic. Vynulované rovnice mají nulovou pravou stranu a jsou splněny triviálně. Volné neznámé jsme volili libovolně. Pro každou z r prvních rovnic máme jednu určenou neznámou, jejíž hodnotu volíme tak, aby byla splněna.

Nechť $(x_1, x_2, \ldots, x_n)$ je řešení soustavy. Musíme ukázat, že ho zpětná substituce nalezne. Hodnoty volných neznámých volíme libovolně, proto se jedna z voleb bude shodovat s tímto řešení. Hodnoty určených neznámých jsou ale jednoznačné, proto se nalezené ohodnocení musí shodovat s řešením $(x_1, x_2, \ldots, x_n)$ (jinak by řešení nebylo řešením). □

Můžeme si všimnout, že proměnných, jejich hodnotu můžeme libovolně zvolit, je $n - r$. Proto pokud $r = n$, soustava má právě jedno řešení.

> Gaussova eliminace se skládá ze dvou fází. Dopředná eliminace upraví soustavu do odstupňovaného tvaru. Zpětná substituce dopočítá všechna řešení soustavy.

Gaussův-Jordanův tvar matice. Výše popsaný odstupňovaný tvar matice se někdy nazývá Gaussův tvar matice. Gaussův tvar matice lze ještě dalšími úpravami zjednodušit a získat *Gaussův-Jordanův tvar* (někdy se mu také říká RREF, *reduced row echelon form* neboli v překladu *redukovaný (řádkově) odstupňovaný tvar*).

Nejprve vynásobíme každý z prvních r řádků tak, aby vzniklý pivot byl jednička (tedy pokud je pivot $a_{i,j}$, vynásobíme i-tý řádek $1/a_{i,j}$). Nyní vynulujeme nenulové koeficienty nad všemi pivoty. Sloupečky vynulováváme v pořadí zprava doleva, vždy přičtením vhodných násobků řádku s pivotem. Detaily necháme čtenáři na rozmyšlení.

Ukažme si Jordanův tvar výše uvedeného odstupňovaného tvaru. Nejprve vynásobíme první řádek $\frac{1}{2}$. Poté přičteme druhý řadek k prvnímu. Dostaneme:

$$\left(\begin{array}{cccc|c} \underline{2} & 2 & -2 & 0 & -2 \\ & & \underline{1} & 2 & 3 \\ & & & & 0 \\ & & & & 0 \end{array}\right) \sim \left(\begin{array}{cccc|c} \underline{1} & 1 & -1 & 0 & -1 \\ & & \underline{1} & 2 & 3 \\ & & & & 0 \\ & & & & 0 \end{array}\right) \sim \left(\begin{array}{cccc|c} \underline{1} & 1 & 0 & 2 & 2 \\ & & \underline{1} & 2 & 3 \\ & & & & 0 \\ & & & & 0 \end{array}\right).$$

Všimněte si, že v Gaussově-Jordanově tvaru dostaneme přesně stejné koeficienty jako ve vyjádření řešení zpětnou substitucí (pochopitelně s obráceným znaménkem):

$$\big\{(2 - x_2 - 2x_4, x_2, 3 - 2x_4, x_4) : x_2, x_4 \in \mathbb{R}\big\}.$$

To není náhoda! Zkuste si to jako cvičení rozmyslet.

Shrnutí

V této kapitole jsme si ukázali soustavy lineárních rovnic a popsali Gaussovu eliminaci.

Nejprve jsme vysvětlili, jak vypadají soustavy lineárních rovnic. Popsali jsme, které úpravy můžeme provádět, aniž bychom měnili množinu řešení. Také jsme ukázali, že soustava může mít jediné řešení, může mít řešení nekonečně mnoho nebo nemusí mít žádné. Soustava s nulovou pravou stranou má vždy alespoň jedno řešení tvořené samými nulami.

Nakonec jsme popsali Gaussovu eliminaci. Ta se skládá ze dvou fází. Dopředná eliminace převede soustavu do odstupňovaného tvaru. V tomto tvaru je snadné dopočítat množinu všech řešení, což dělá druhá fáze zvaná zpětná substituce. Dokázali jsme, že dopředná eliminace vždy uspěje a zpětná substituce správně zkonstruuje množinu všech řešení. Popsali jsme také Jordanův tvar matice, jako další zjednodušení odstupňovaného tvaru.

Cvičení

⇒ 1.1 Vyřešte následující soustavy rovnic:

$$\left(\begin{array}{c|c} 5 & 4 \end{array}\right), \quad \left(\begin{array}{cc|c} 2 & 1 & 4 \\ 3 & 1 & 5 \end{array}\right), \quad \left(\begin{array}{ccc|c} 1 & 0 & 1 & 2 \\ 2 & 1 & 0 & 5 \end{array}\right), \quad \left(\begin{array}{ccc|c} 6 & 1 & 3 & 2 \\ -1 & 1 & 2 & 5 \\ 4 & 1 & 3 & 4 \end{array}\right).$$

1.2 Vyřešte následující soustavy rovnic $n \times n$:

$$\left(\begin{array}{ccccc|c} 1 & 1 & 1 & \cdots & 1 & b_1 \\ 1 & 2 & 2 & \cdots & 2 & b_2 \\ 1 & 2 & 3 & \cdots & 3 & b_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 2 & 3 & \cdots & n & b_n \end{array}\right), \quad \left(\begin{array}{ccccc|c} x & y & y & \cdots & y & b_1 \\ y & x & y & \cdots & y & b_2 \\ y & y & x & \cdots & y & b_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ y & y & y & \cdots & x & b_n \end{array}\right), \text{ kde } x \text{ a } y \text{ jsou libovolné konstanty.}$$

⋆ 1.3 Aplikujte Gaussovu eliminaci na následující matice. Z úprav provádějte pouze přičítání násobků výše položených řádků k těm níže položeným, tedy neprohazujte řádky a nenásobte je konstantou. Jaké hodnoty mají pivoti v odstupňovaném tvaru? Dokažte správnost pro obecnou velikost matice $n \times n$.

$$\begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & \ddots & \ddots & \\ & & \ddots & 2 & -1 \\ & & & -1 & 2 \end{pmatrix}, \quad \begin{pmatrix} 1 & -1 & & & \\ 1 & 1 & -1 & & \\ & 1 & \ddots & \ddots & \\ & & \ddots & 1 & -1 \\ & & & 1 & 1 \end{pmatrix}.$$

1.4 Jak vypadá soustava, pro kterou dopředná eliminace neprovede žádný krok (zastaví se na začátku prvního)? Jak vypadá soustava, pro kterou se provede pouze jediný krok? Co se stane, pokud dopřednou eliminaci použijeme na matici, která již je v odstupňovaném tvaru.

1.5 Nalezněte soustavu třech rovnic o dvou neznámých, která nemá žádné řešení, ale levá strana žádné z rovnic není násobkem jiné levé strany. Podle kapitoly 2 nakreslete řádkovou geometrickou interpretaci soustavy.

⋆ 1.6 Mějme polynom

$$P(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

s neznámými koeficienty $a_0, \ldots, a_n$. Známe jeho hodnoty v $n+1$ různých bodech $P(x_1), \ldots, P(x_n)$. Nalezněte koeficenty tohoto polynomu.

1.7 Spočítejte počet operací, které provede Gaussova eliminace pro soustavu n lineárních rovnic o n neznámých. Předpokládejte, že daným řádkem vždy eliminujeme všechny ostatní řádky a vždy vynulujeme pouze jeden sloupeček. Chceme odhadnout pouze dominatní člen u počtu operací, ostatní členy zanedbejte.

1.8 (a) Naprogramujte Gaussovu eliminace, třeba v Pascalu nebo C. Výpočty počítejte ve floatech a naprogramujte i zkoušku.
(b) Vyzkoušejte naprogramovaný algoritmus na slavné Hilbertově matici 20×20 s jednotkovou pravou stranou, a porovnejte velikost chyby v provedené zkoušce, tedy porovnejte $\boldsymbol{b}$ a $A\boldsymbol{x}$. Koeficient na pozici (i, j) v Hilbertově matici je $\frac{1}{i+j}$.

Kapitola 2

Vektory a vektorové prostory

V kapitole 1 jsme popsali soustavy lineárních rovnic a algoritmus na jejich řešení zvaný Gaussova eliminace. Ukázali jsme také, že soustavy mohou mít různý počet řešení (klidně žádné).

Cílem této kapitoly je ukázat, že všechna řešení libovolné soustavy mají velice jednoduchou strukturu. Ve zkratce by se dalo říct, že všechna řešení tvoří *afinní podprostor*. Navíc si více přiblížíme geometrickou stránku lineární algebry.

Aby se nám lépe popisovalo, zavedeme nejprve klíčovou definici vektorů a jejich operací.

2.1 Vektory a jejich operace

Začněme motivací, mějme body v rovině. Každému bodu můžeme přiřadit dvě *souřadnice* x a y vůči osám, které budeme zapisovat jako dvojici (x, y). Klíčové je, že každé dva různé body mají různé souřadnice a každá dvojice souřadnic určuje nějaký bod roviny. Máme body spárované se souřadnicemi, jsou v korespondenci jedna ku jedné. To znamená, že body v rovině můžeme reprezentovat pomocí uspořádaných dvojic reálných čísel.

Podobně body v trojrozměrném prostoru jsou popsané třemi souřadnicemi x, y a z, a odpovídají tedy trojicím (x, y, z). Zobecněme to nyní na n-rozměrné prostory.

Vektory. Zaveďme n-rozměrný prostor $\mathbb{R}^n$ jako množinu všech n-tic reálných čísel.[1] Každá n-tice odpovídá souřadnicím jednoho bodu. Zde vidíme jednu z krásných vlastností lineární algebry, umožňuje snadno zobecňovat do více dimenzí. Body v n-rozměrném prostoru se špatně představují a ještě hůře vizualizují. Naproti tomu s uspořádanými n-ticemi reálných čísel se pracuje pohodlně a představí si je každý. Ostatně známý vtip praví, že pokud si matematik chce představit čtyřrozměrný prostor, uváží n-rozměrný prostor a za n dosadí čtyřku.

Bodům budeme říkat *vektory* a n-rozměrnému prostoru se říká *vektorový prostor*. Vektory budeme značit tučnými písmeny, například $\boldsymbol{u}$. Pro

$$\boldsymbol{u} = (u_1, u_2, \ldots, u_n)$$

se číslům $u_1, \ldots, u_n$ říká *složky*. Někdy budeme označovat i-tou složku u_i vektoru $\boldsymbol{u}$ jako $(\boldsymbol{u})_i$. Vektoru $\mathbf{0}$, jehož všechny souřadnice jsou nulové, se říká *počátek*.

[1] $\mathbb{R}^n$ značí n-tou *kartézskou mocninu* reálných čísel, množinu všech n-tic reálných čísel. Obecně X^n značí množinu všech n-tic prvků z X, tedy:

$$X^n = \{(x_1, \ldots, x_n) : x_1, \ldots, x_n \in X\}.$$

Geometricky se vektory často značí jako šipka vycházející z počátku do souřadnic vektoru.[2] Fyzikálně jsou vektory síly, které působí nějakou velikostí nějakým směrem.

Uveďme si příklady n-rozměrných prostorů pro malé hodnoty n. Patologický případ $\mathbb{R}^0$ obsahuje pouze jediný vektor, kterým je počátek $\mathbf{0}$. Proč? Existuje právě jedna „nulatice“ reálných čísel. Pro ni platí, že má všechny složky nulové. Vektorový prostor $\mathbb{R}^1$ je přímka, prostor $\mathbb{R}^2$ je rovina, $\mathbb{R}^3$ je trojrozměrný prostor, a tak dál. Na názorná vysvětlení většiny tvrzení lineární algebry vystačíme s $\mathbb{R}^2$ a $\mathbb{R}^3$, které se dobře vizualizují.

Vektorové operace. Na vektorech můžeme provádět dvě základní operace. První z nich je *násobení skalárem* (reálným číslem),[3] které vynásobí každou ze složek vektoru tímto skalárem:

$$\alpha\boldsymbol{u} = (\alpha u_1, \alpha u_2, \ldots, \alpha u_n).$$

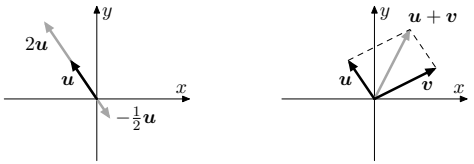
Geometrický význam násobení skalárem je naznačen na obrázku 2.1 vlevo. Násobení skalárem α natáhne vektor α-krát. Násobení záporným skalárem navíc obrátí vektor. Vektor $(-1)\boldsymbol{u}$ je *opačný vektor* k $\boldsymbol{u}$ a budeme ho značit $-\boldsymbol{u}$. Pro libovolný vektor $\boldsymbol{u}$ platí, že $0\boldsymbol{u} = \mathbf{0}$.

Druhá operace je *součet* dvou vektorů, který se opět provede po složkách:

$$\boldsymbol{u} + \boldsymbol{v} = (u_1 + v_1, u_2 + v_2, \ldots, u_n + v_n).$$

Pozor, aby bylo vektory možné sčítat, musí mít *stejný* počet složek, tedy musí patřit do stejného vektorového prostoru $\mathbb{R}^n$! Kdykoliv budeme sčítání aplikovat, budeme stejnou velikost předpokládat. Ostatně typicky pracujeme s vektory jedné velikosti.

Fyzikální význam je sečtení (složení) dvou sil. Geometricky tato operace odpovídá doplnění na rovnoběžník, jak je naznačeno na obrázku 2.1 vpravo. Sčítání funguje jako složení dvou posunutí, z počátku o vektor $\boldsymbol{u}$ a poté o vektor $\boldsymbol{v}$. Za $\boldsymbol{u} + (-\boldsymbol{v})$ zavedeme zkratku $\boldsymbol{u} - \boldsymbol{v}$. Toto značení je velice vhodné, neboť $\boldsymbol{u} - \boldsymbol{v} = (u_1 - v_1, u_2 - v_2, \ldots, u_n - v_n)$. Všimněte si, že pro libovolný vektor $\boldsymbol{u}$ platí $\boldsymbol{u} - \boldsymbol{u} = \mathbf{0}$.



Obrázek 2.1: Ukázka vektorových operací v rovině. Nalevo násobení skalárem, napravo sčítání.

Vlastnosti operací. Tyto operace mají řadu hezkých vlastností, díky kterým se pohodlně používají. Důvodem je, že po složkách aplikujeme operace na reálná čísla. Proto se velká část pěkných vlastností reálných čísel přenese i na vektory. Které to jsou? Jedná se například o komutativitu, asociativitu a distributivitu.[4] Věříme, že čtenář alespoň zhruba tyto vlastnosti zná. Proto si je jen ve zkratce popíšeme, což by mělo pro pochopení dalšího textu stačit. Přesně si všechny vlastnosti popíšeme na konci kapitoly, kde je použijeme k alternativní definici vektorového prostoru.

Komutativita říká, že můžeme prohodit pořadí sčítanců, tedy $\boldsymbol{u}+\boldsymbol{v} = \boldsymbol{v}+\boldsymbol{u}$. Asociativita říká, že posloupnost sčítání $\boldsymbol{u}_1 + \cdots + \boldsymbol{u}_n$ můžeme uzávorkovat v libovolném pořadí a nezměníme výsledek, tedy například $(\boldsymbol{u}_1+\boldsymbol{u}_2)+\boldsymbol{u}_3 = \boldsymbol{u}_1+(\boldsymbol{u}_2+\boldsymbol{u}_3)$. Distributivita říká: $\alpha(\boldsymbol{u}+\boldsymbol{v}) = \alpha\boldsymbol{u}+\alpha\boldsymbol{v}$ a $(\alpha+\beta)\boldsymbol{u} = \alpha\boldsymbol{u}+\beta\boldsymbol{u}$.

Zkusme alespoň jednu z vlastností dokázat. Jako cvičení můžete dokázat i ostatní.

Tvrzení 2.1. *Sčítání vektorů je komutativní, tedy pro libovolné vektory $\boldsymbol{u}$ a $\boldsymbol{v}$ platí $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$.*

Důkaz. Dva vektory jsou stejné, pokud se shodují ve všech složkách, tedy jsou to stejné n-tice. Abychom ukázali, že $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$, musíme ukázat, že $(\boldsymbol{u} + \boldsymbol{v})_i = (\boldsymbol{v} + \boldsymbol{u})_i$ pro každou ze složek $i = 1, \ldots, n$. Protože $(\boldsymbol{u} + \boldsymbol{v})_i = u_i + v_i$ a $(\boldsymbol{v} + \boldsymbol{u})_i = v_i + u_i$ jsou součty reálných čísel, můžeme použít komutativitu $\mathbb{R}$:

$$(\boldsymbol{u} + \boldsymbol{v})_i = u_i + v_i \overset{\text{kom. } \mathbb{R}}{=} v_i + u_i = (\boldsymbol{v} + \boldsymbol{u})_i, \qquad \forall i = 1, \ldots, n.$$

Tedy vektory $\boldsymbol{u} + \boldsymbol{v}$ a $\boldsymbol{v} + \boldsymbol{u}$ mají všechny složky shodné, a proto jsou stejné. □

Vektorový prostor $\mathbb{R}^n$ je množina všech n-tic reálných čísel, kterým říkáme vektory a na kterých máme definované dvě operace: Násobení skalárem a sčítání. Tyto operace se provádí po složkách.

2.2 Geometrické interpretace soustavy

Vraťme se zpět k soustavám lineárních rovnic. Ukážeme si dvě geometrické interpretace, pomocí řádků soustavy a pomocí sloupců soustavy.

Řádková interpretace. Zkusme nahlédnout na soustavy po řádcích. Každé řešení je n-složkový vektor $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ z $\mathbb{R}^n$. Množina všech řešení je tedy nějaká množina vektorů. Pokusíme se zjistit, jak vypadá tato množina geometricky.

Položme si otázku, jak vypadá množina všech řešení pouze pro jeden řádek

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b.$$

Pokud budeme totiž umět vyřešit každý řádek samostatně, budeme umět vyřešit celou soustavu. Množina všech řešení soustavy je totiž průnik řešení jednotlivých řádků. Budeme předpokládat, že alespoň jeden z koeficientů $a_1, \ldots, a_n$ je nenulový. Pokud by totiž všechny byly nulové, je triviální určit množinu všech řešení: buď je prázdná (pro $b \neq 0$), nebo je to naopak celý prostor $\mathbb{R}^n$ (pro $b = 0$).

Podívejme se nejprve na případ $n = 2$, tedy na $a_1x_1 + a_2x_2 = b$. Každé řešení je dvousložkový vektor $\boldsymbol{x} = (x_1, x_2)$. Geometricky řešení této rovnice tvoří přímku. Proč? Pokud je jeden z koeficientů nulový, dostaneme pro $a_1x_1 = b$ vertikální přímku $\{(b/a_1, x_2) : x_2 \in \mathbb{R}\}$ a pro $a_2x_2 = b$ horizontální přímku $\{(x_1, b/a_2) : x_1 \in \mathbb{R}\}$. Pokud jsou oba koeficienty nenulové, můžeme vyjádřit x_1:

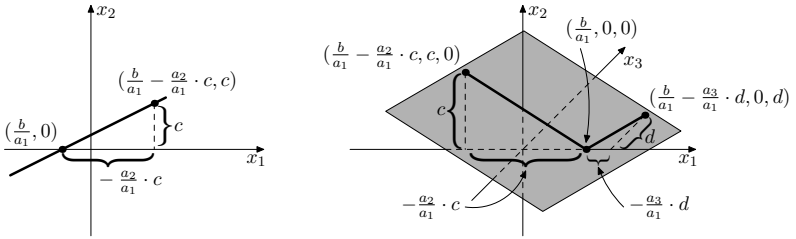
$$x_1 = -\frac{a_2}{a_1}x_2 + \frac{b}{a_1}.$$

Pro každou hodnotu x_2 je jednoznačně určena hodnota x_1. Situace je naznačena na obrázku 2.2 vlevo. Předně pro $x_2 = 0$ dostaneme řešení $(\frac{b}{a_1}, 0)$. Pokud změníme hodnotu x_2 o c, změní se hodnota x_1 *lineárně* o $-\frac{a_2}{a_1} \cdot c$. Rozmyslete si, že z lineárního vztahu mezi x_1 a x_2 plyne, že všechna řešení tvoří přímku.

[2] Všimněte si, že pro bod v n-rozměrném prostoru používáme dvě označení: vektor a bod. Důvod pro to je historický, bodům se říkalo body a vektory reprezentovaly posunutí určitým směrem o určitou vzdálenost (proto šipka z počátku). Z našeho pohledu body a vektory splývají a není potřeba je rozlišovat. Někdy se bude hodit interpretace jako bod, někdy jako směr.

[3] Proč se používá název násobení skalárem místo přirozenějšího násobení reálným číslem? Důvodem je, že později budeme uvažovat obecnější vektorové prostory, jejichž vektory budou tvořeny obecněji definovanými čísly. Skalár je označení pro toto obecné číslo. Obecnější čísla jsou prvky *algebraických těles*, která si popíšeme v kapitole 6.

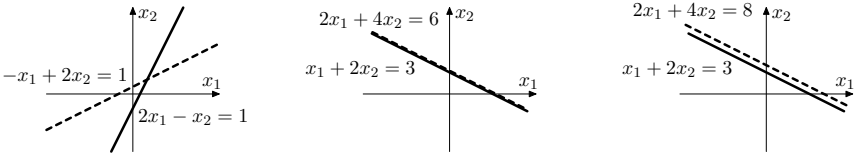
[4] Že jsou tyto vlastnosti skutečně hezké a důležité docení čtenář až ve chvíli, kdy bude pracovat s operacemi, které některou z těchto vlastností postrádají, je to mnohem těžší.



Obrázek 2.2: Množina všech řešení jedné rovnice pro $n = 2$ (nalevo) a $n = 3$ (napravo).

Pro $n = 2$ dostaneme z každé rovnice jednu přímku a řešení soustavy je průnik těchto přímek. Na obrázku 2.3 jsou naznačeny tři typické příklady. Pro dvě rovnice se nejčastěji setkáme s případem nalevo, kde mají dvě přímky jediný společný bod a existuje jediné řešení (v tomto případě $\boldsymbol{x} = (1, 1)$). Na obrázku 2.3 uprostřed a vpravo naleznete soustavy (1.5) z kapitoly 1. Pro soustavu uprostřed určují obě rovnice stejnou přímku, tedy množina řešení je celá přímka. Pro soustavu napravo dostáváme dvě různé rovnoběžné přímky. Ty mají prázdný průnik, a tedy žádné řešení neexistuje.

Tyto příklady ilustrují všechny typy řešení, která můžeme pro $n = 2$ dostat: prázdný průnik, jediný bod (dimenze 0), přímka (dimenze 1) nebo celá rovina (dimenze 2; triviální případ, kdy jsou všechny koeficienty soustavy nulové). Význam dimenzí lze zatím chápat intuitivně, formálně je popíšeme v kapitole 4. Pochopitelně vše, co o dimenzích tvrdíme, je neformální, bez přesné definice totiž není možné o dimenzích hovořit formálně.



Obrázek 2.3: Množina všech řešení soustav o dvou neznámých jako průnik přímek v rovině. Jedna přímka je vyznačena plnou čarou, druhá přerušovaně.

V případě $n = 3$ tvoří řešení každé rovnice jednu rovinu v $\mathbb{R}^3$, jak je naznačeno na obrázku 2.2 vpravo. Předpokládáme, že alespoň jeden koeficient je nenulový (jinak je rovnice triviální). Bez újmy na obecnosti nechť je a_1 nenulový, vyjádříme z rovnice neznámou x_1. Pokud by koeficient a_1 byl nulový, vyjádříme neznámou x_2 nebo x_3 s nenulovým koeficientem, vyjádření se bude pouze lišit jiným očíslováním proměnných. Platí

$$x_1 = -\frac{a_2}{a_1}x_2 - \frac{a_3}{a_1}x_3 + \frac{b}{a_1}$$

a pro každou volbu x_2 a x_3 je jednoznačně určena hodnota x_1.

Pro $x_2 = x_3 = 0$ dostaneme řešení $(\frac{b}{a_1}, 0, 0)$. Pokud změníme hodnotu x_2 o c (a nezměníme hodnotu x_3), změní se hodnota x_1 o $-\frac{a_2}{a_1} \cdot c$. Pokud změníme hodnotu x_3 o d (a nezměníme hodnotu x_2), změní se hodnota x_1 o $-\frac{a_3}{a_1} \cdot d$. To znamená, že změna x_1 je lineárně závislá na změnách x_2 a x_3 a při změně pouze jedné z hodnot x_2 nebo x_3 řešení tvoří přímku. Pokud změníme hodnoty obou x_2 a x_3, změny x_1 se sečtou a přímky se zkombinují. Dostaneme jako řešení rovinu v $\mathbb{R}^3$.

Vše platí i v obecném případě, neboť hodnota jedné proměnné (s nenulovým koeficientem) záleží lineárně na hodnotách ostatních proměnných. Při změně hodnoty pouze jediné proměnné všechna řešení

leží na přímce, při obecné změně se tyto přímky zkombinují. Protože můžeme zvolit hodnoty $n - 1$ proměnných, je řešení každé rovnice množina dimenze $n - 1$, které se říká *nadrovina*.

Řešením soustavy je průnik několika nadrovin. Dimenze tohoto průniku je počet volných proměnných $n - r$ z odstupňovaného tvaru, tedy počet proměnných, jejichž hodnotu můžeme libovolně zvolit. Za každou volnou proměnnou dimenze řešení vzroste o jedničku. Pokud má soustava nulovou pravou stranu, pozorování 1.2 říká, že počátek je řešením. Pro nulovou pravou stranu prochází každá z nadrovin počátkem, a proto počátek leží i v jejich průniku.

Toto pro ilustraci řádkové interpretace stačí, formálně si vše zdůvodníme později v kapitole 4.

Množina všech řešení každé rovnice tvoří nadrovinu a množina všech řešení soustavy je průnik několika nadrovin. Řešení mají hezkou geometrickou strukturu, jsou to vícerozměrná zobecnění objektů jako přímka nebo rovina.

Sloupcová interpretace. Na soustavu lze nahlédnout geometricky úplně jinak, tentokrát po sloupcích. Každý sloupec dává jeden m-složkový vektor, jehož složky jsou koeficienty v tomto sloupci. Abychom zdůraznili, že se jedná o sloupcové vektory, zapíšeme vektory do sloupečku. Označme vektor i-tého sloupce $\boldsymbol{u}_i$ a vektor pravé strany $\boldsymbol{b}$, tedy

$$\boldsymbol{u}_i = \begin{pmatrix} a_{1,i} \\ a_{2,i} \\ \vdots \\ a_{m,i} \end{pmatrix}, \qquad i = 1, 2, \ldots, n \qquad \text{a} \qquad \boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix},$$

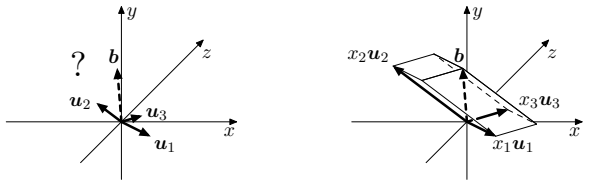
připomeňte si zápis koeficientů z (1.2).

Nyní učiňme klíčové pozorování: V soustavě násobí proměnná x_i pouze koeficienty z i-tého sloupečku, tedy z vektoru $\boldsymbol{u}_i$. Toto násobení není nic jiného než skalární násobek $\boldsymbol{u}_i$ skalárem x_i. Tedy spolu s proměnnou x_i máme v i-tém sloupečku soustavy vektor $x_i\boldsymbol{u}_i$. Soustava všechny tyto vektory sčítá a pokládá rovné pravé straně $\boldsymbol{b}$. Dostaneme vektorovou rovnici

$$x_1\boldsymbol{u}_1 + x_2\boldsymbol{u}_2 + \cdots + x_n\boldsymbol{u}_n = \boldsymbol{b}, \tag{2.1}$$

což není nic jiného než přepsání soustavy do vektorové řeči.

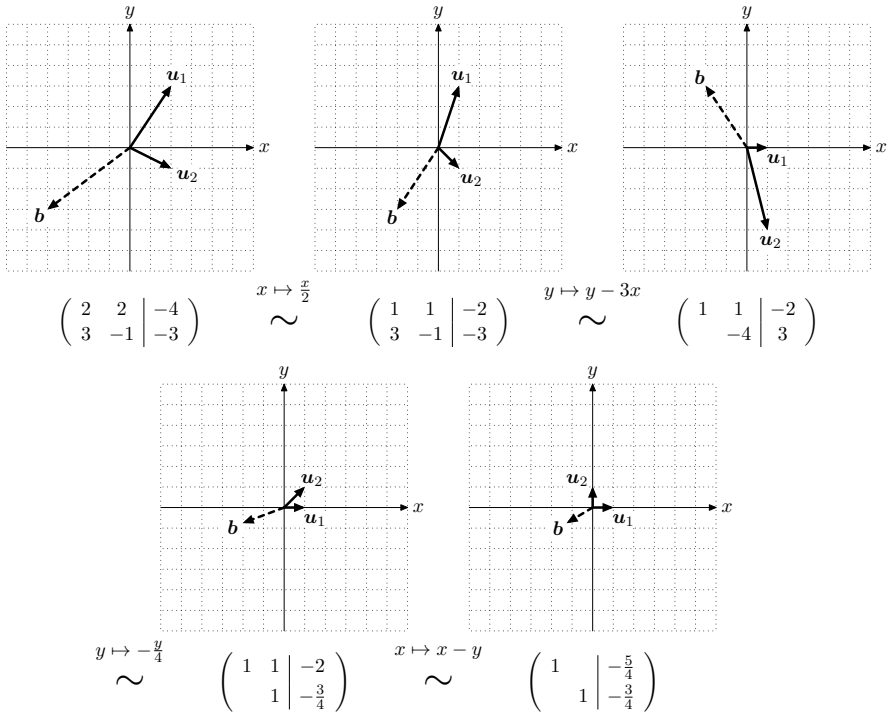
Nyní na tuto vektorovou rovnici použijeme geometrickou interpretaci vektorových operací. Hledáme reálná čísla $x_1, \ldots, x_n$ taková, že tato rovnice bude platit. Tedy hledáme natažení vektorů $\boldsymbol{u}_1, \ldots \boldsymbol{u}_n$ taková, že po jejich sečtení dostaneme pravou stranu $\boldsymbol{b}$. Sloupcovou interpretaci soustavy (1.1) naleznete na obrázku 2.4.



Obrázek 2.4: Sloupcová interpretace soustavy (1.1). Nalevo máme vektory $\boldsymbol{u}_1$, $\boldsymbol{u}_2$, $\boldsymbol{u}_3$ a pravou stranu $\boldsymbol{b}$. Napravo jsou vektory vynásobené řešením $x_1 = 1$, $x_2 = 3$ a $x_3 = 2$. Platí $x_1\boldsymbol{u}_1 + x_2\boldsymbol{u}_2 + x_3\boldsymbol{u}_3 = \boldsymbol{b}$, graficky sčítáme doplněním na rovnoběžnostěn.

Sloupcová interpretace soustavy říká: Hledáme skaláry $x_1, \dots, x_n$ takové, aby se násobky $x_1\boldsymbol{u}_1, \dots, x_n\boldsymbol{u}_n$ sloupcových vektorů sečetly na pravou stranu $\boldsymbol{b}$.

Gaussova eliminace geometricky. Zkusíme geometricky ilustrovat fungování Gaussovy eliminace. Proč fungují regulární řádkové úpravy a naopak sloupcové úpravy typicky mění množinu řešení? Z pohledu sloupcové interpretace řádkové úpravy geometricky transformují prostor, ve kterém se nachází vektory $\boldsymbol{u}_1, \dots, \boldsymbol{u}_n$ a $\boldsymbol{b}$. Regularita říká, že tyto geometrické transformace nikde prostor "nesplácnou", a tedy nezmění vzájemné vztahy mezi $\boldsymbol{u}_1, \dots, \boldsymbol{u}_n$ a $\boldsymbol{b}$. Na obrázku 2.5 se nachází příklad sekvence transformace prostoru sloupcových vektorů, který provádí Gaussova eliminace. Gaussova eliminace je tedy strategie, jak aplikovat na prostor se sloupcovými vektory sérii transformací tak, aby se vzajemné vztahy mezi vektory $\boldsymbol{u}_1, \dots, \boldsymbol{u}_n$ co nejvíce zjednodušili.



Obrázek 2.5: Jednotlivé kroky Gaussovy eliminace, vyobrazené po sloupcích. Matice je upravena až do Gauss-Jordanova tvaru, v kterém je jednoduché najít řešení $(-\frac{5}{4}, -\frac{3}{4})$. Vždy platí, že $-\frac{5}{4}\boldsymbol{u}_1 - \frac{3}{4}\boldsymbol{u}_2 = \boldsymbol{b}$. Toto řešení lze také vypočíst zpětnou substitucí z odstupňovaného tvaru. Jediný vektor $\boldsymbol{u}_2$ má nenulovou složku ve směru osy y, tedy natažení x_2 je jednoznačně určené souřadnicí $\boldsymbol{b}$ ve směru osy y. Zbývá určit x_1, které je jednoduché ze znalosti x_2 dopočítat.

Naproti tomu sloupcové úpravy by měnili jednotlivé sloupcové vektory. Tím by se však měnili koeficienty natažení $x_1, \dots, x_n$, aby platilo $x_1\boldsymbol{u}_1 + \cdots + x_n\boldsymbol{u}_n = \boldsymbol{b}$. Protože však tyto koeficienty neznáme, neumíme určit vliv úpravy a příslušně změnit pravou stravu $\boldsymbol{b}$, aby $x_1, \dots, x_n$ byly zachovány.[5]

Konkrétně elementární řádkové úpravy uvažované v Gaussově eliminaci jsou dvě: vynásobení řádku nenulovým koeficientem α a přičtení jednoho řádku k druhému. Geometricky vynásobení α-krát natahuje prostor ve směru jedné souřadné osy. S tím se natahují i vektory $\boldsymbol{u}_1, \dots, \boldsymbol{u}_n$ a $\boldsymbol{b}$, všechny však stejně, a proto se jejich vzájemné vztahy nemění. Operace přičtení jednoho řádku k druhému odpovídá zkosení jedné dimenze do druhé. Tato operace umožňuje v Gaussově eliminaci narovnávat sloupcové vektory ve směru souřadných os.

Sloupcová interpretace vysvětluje, proč je tak jednoduché dopočítat množinu všech řešení z odstupňovaného tvaru matice, kdy máme sloupcové vektory $\boldsymbol{u}_1, \dots, \boldsymbol{u}_n$ mnohem lépe uspořádané.

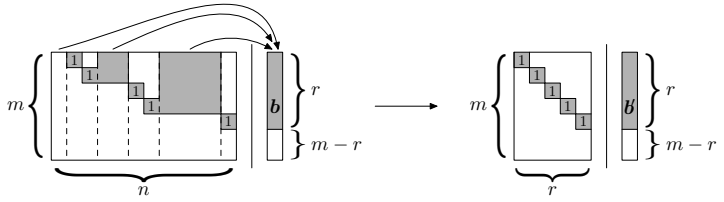
Aby řešení existovalo, musí mít nulové řádky i nulové pravé strany. Totiž pokud by existoval nulový řádek s nenulovou pravou stranou, obsahoval by vektor $\boldsymbol{b}$ složku, která ve všech vektorech $\boldsymbol{u}_i$ chybí. Proto libovolným vynásobením a sečtením vektorů $\boldsymbol{u}_i$ nikdy nedostaneme $\boldsymbol{b}$ a rovnice (2.1) nemá řešení. Geometricky vektor $\boldsymbol{b}$ ukazuje „jiným směrem" než vektory $\boldsymbol{u}_i$, což přesně vysvětlíme v kapitole 4.

Nyní jak nalézt řešení. Nejprve zvolíme hodnoty volných proměnných zcela libovolně. Dostaneme tedy konkrétní násobky $x_i\boldsymbol{u}_i$ sloupcových vektorů $\boldsymbol{u}_i$ volných proměnných. Tyto konkrétní násobky převedeme na pravou stranu soustavy a odečteme od $\boldsymbol{b}$. Dostaneme novou pravou stranu $\boldsymbol{b}'$:

$$\boldsymbol{b}' = \boldsymbol{b} - \sum_{x_i \text{ volná}} x_i\boldsymbol{u}_i$$

a nová soustavu bude obsahovat pouze určené proměnné, které přečíslujme $x'_1, \dots, x'_r$, a jejich sloupcové vektory $\boldsymbol{u}'_1, \dots, \boldsymbol{u}'_r$. Nová soustava je naznačena na obrázku 2.6.

Potom postupně dopočítáváme koeficienty $x'_r, \dots, x'_1$, jejichž hodnoty již jsou jednoznačně určeny. Důvod je, že pouze nejpravější sloupcový vektor $\boldsymbol{u}'_r$ ukazuje ve směru r-té souřadné osy. Proto je jeho koeficient natažení x'_r jednoznačně určen hodnotou b'_r. Toto postupně platí i pro ostatní koeficienty x'_i, když jsou všechny proměnné $x'_{i+1}, \dots, x'_r$ určeny. Speciálně pro Gaussův-Jordanův tvar, který jsme popsali na konci kapitoly 1, je triviální určit hodnoty jednotlivých proměnných x'_i, neboť platí $x'_i = b'_i$ pro každou proměnnou.



Obrázek 2.6: Dopočet řešení z Gaussova-Jordanova tvaru.

Gaussova eliminace je strategie, jak transformovat prostor sloupcových vektorů, aby bylo snadné určit vyjádření pravé strany.

[5] V určitém smyslu lze provádět sloupcové úpravy, vypočítat koeficienty $x_1, \dots, x_n$ a aplikovat inverzní sloupcové úpravy na tyto koeficienty. Tato možnost je složitější a budeme ji prozatím ignorovat.

2.3 Vektorové podprostory

Často dostaneme určitou množinu vektorů, například množinu všech řešení soustavy. Budeme uvažovat množiny, které mají určitou strukturu a chovají se hezky vůči vektorovým operacím. Zaveďme si klíčovou definici *vektorových podprostorů*.

Vektorový podprostor. Užitečná vlastnost množiny je, aby byla *uzavřená na operace*. Co to znamená? Množina vektorů W musí obsahovat všechny natažení a součty vektorů z W. Formálně:

- *Uzavřenost na násobení skalárem:* Pokud $\boldsymbol{u} \in W$, také $\alpha\boldsymbol{u} \in W$ pro každé $\alpha \in \mathbb{R}$.
- *Uzavřenost na sčítání:* Pokud $\boldsymbol{u}, \boldsymbol{v} \in W$, také $\boldsymbol{u} + \boldsymbol{v} \in W$.

Neprázdnou množinu $W \subseteq \mathbb{R}^k$ uzavřenou na operace nazveme *vektorový podprostor* prostoru $\mathbb{R}^k$ (nebo jen zkráceně *podprostor*). Povšimněme si, že počátek $\mathbf{0}$ leží v každém vektorovém podprostoru. Proč? Podprostor je množina neprázdná, proto obsahuje alespoň jeden vektor $\boldsymbol{u}$. Pro ten platí $0\boldsymbol{u} = \mathbf{0}$, a tedy i $\mathbf{0}$ leží v tomto podprostoru.

Jako příklad si ukažme, jak vypadají vektorové podprostory prostoru $\mathbb{R}^3$. Množina obsahující pouze počátek je triviální vektorový podprostor. Pokud všechny vektory množiny ukazují stejným směrem, dostaneme přímku procházející počátkem (podprostor obsahuje všechny násobky). Pokud vektory ukazují dvěma směry, dostaneme rovinu procházející počátkem. Pokud vektory ukazují třemi směry, je to celý prostor $\mathbb{R}^3$. Podprostory $\mathbb{R}^3$ tedy jsou: triviální (dimenze 0), přímky procházející počátkem (dimenze 1), roviny procházející počátkem (dimenze 2) a celý prostor $\mathbb{R}^3$ (dimenze 3).

Podprostory mají řadu hezkých vlastností, ukažme si jednu z nich:

Tvrzení 2.2. *Nechť $W_1, \ldots, W_n$ jsou vektorové podprostory prostoru $\mathbb{R}^k$. Potom jejich průnik*

$$W = W_1 \cap W_2 \cap \cdots \cap W_n = \bigcap_{i \in \{1,\ldots,n\}} W_i$$

je také vektorový podprostor.

Důkaz. Potřebujeme ukázat, že W je uzavřený na operace násobení skalárem a sčítání. Nechť $\boldsymbol{u} \in W$, chceme ukázat, že také $\alpha\boldsymbol{u} \in W$ pro každé $\alpha \in \mathbb{R}$. Protože $\boldsymbol{u} \in W$, také $\boldsymbol{u} \in W_i$ pro každé W_i. Ale protože W_i jsou vektorové podprostory, obsahují též $\alpha\boldsymbol{u}$. Proto $\alpha\boldsymbol{u}$ leží v průniku W. Podobně dokážeme uzavřenost W na sčítání, zkuste jako cvičení. □

Tvrzení platí i pro průnik nekonečně mnoha podprostorů, rozmyslete si, že úplně stejný důkaz bude fungovat. Tato poznámka je důležitá, protože v matematice existuje řada tvrzení, které platí pro konečné průniky a pro nekonečné průniky obecně neplatí.⟨6⟩

Motivace. Čtenář se možná ptá, proč jsou vektorové podprostory tolik zajímavé. Proč chceme množiny vektorů uzavřené na operace? Vektory většinou nepoužíváme samostatně, ale aplikujeme na ně operace. Vždyť i samotnou definici vektorů hned doplňujeme zavedením vektorových operací. Pokud máme množinu vektoru, na kterou chceme aplikovat operace, je dobré mít jistotu, že i výsledek bude ležet opět uvnitř množiny. Uzavřenost je jedna ze základních vlastností operací, která se v algebře požaduje.

Uveďme si jiný příklad, který dobře znáte; přirozená čísla. Ty velice dobře fungují s operacemi jako sčítání a násobení, neboť jsou na ně uzavřená. Pokud ale začneme pracovat i s odčítáním, rychle se dostaneme do problému, protože výsledek rozdílu nemusí být přirozené číslo. Aby se lépe pracovalo, rozšíříme přirozená čísla na celá čísla. Pokud začneme uvažovat i dělení, hodí se zavést racionální čísla.⟨7⟩

Řešení soustavy s nulovou pravou stranou. Existuje další motivace pro vektorové podprostory. Vektorové podprostory se přirozeně objevuji v lineární algebře.

Tvrzení 2.3. *Mějme soustavu s nulovou pravou stranou. Množina všech řešení soustavy tvoří vektorový podprostor prostoru $\mathbb{R}^n$.*

Důkaz. Jak to dokážeme? Každé řešení soustavy je n-složkový vektor z $\mathbb{R}^n$. Chceme ukázat, že množina všech řešení tvoří vektorový podprostor. Je neprázdná, protože $\mathbf{0}$ je řešením. Zbývá ukázat, že je uzavřená na násobení skalárem a na sčítání.

1. *Uzavřenost na násobení skalárem:* Mějme vektor $\boldsymbol{x}$, který je řešením soustavy. Chceme ukázat, že také $\alpha\boldsymbol{x}$ pro každé α reálné je řešením. Důkaz provedeme pro každou rovnici zvlášť. Rozepišme si i-tou rovnici:
$$a_{i,1}x_1 + a_{i,2}x_2 + \cdots + a_{i,n}x_n = 0. \tag{2.2}$$
Z vlastností reálných čísel (komutativita, asociativita a distributivita) plyne:
$$a_{i,1}(\alpha x_1) + a_{i,2}(\alpha x_2) + \cdots + a_{i,n}(\alpha x_n) = \alpha(a_{i,1}x_1 + a_{i,2}x_2 + \cdots + a_{i,n}x_n) \overset{(2.2)}{=} \alpha \cdot 0 = 0.$$
Tedy i $\alpha\boldsymbol{x}$ řeší i-tou rovnici. Toto platí pro každou rovnici, tedy $\alpha\boldsymbol{x}$ je řešením soustavy.
2. *Uzavřenost na sčítání:* Chceme ukázat, že pokud vektory $\boldsymbol{x}$ a $\boldsymbol{y}$ jsou řešeními, je i $\boldsymbol{x} + \boldsymbol{y}$ řešením. To se ukáže zcela obdobně, zvlášť pro každou rovnici. Pro i-tou rovnici platí:
$$a_{i,1}(x_1 + y_1) + a_{i,2}(x_2 + y_2) + \cdots + a_{i,n}(x_n + y_n) = \underbrace{a_{i,1}x_1 + a_{i,2}x_2 + \cdots a_{i,n}x_n}_{=0 \text{ podle (2.2)}} + \underbrace{a_{i,1}y_1 + a_{i,2}y_2 + \cdots a_{i,n}y_n}_{=0 \text{ podle (2.2) s } \boldsymbol{y} \text{ místo } \boldsymbol{x}} = 0 + 0 = 0.$$
Tedy $\boldsymbol{x} + \boldsymbol{y}$ splňuje každou z rovnic a je řešením soustavy. □

> Podmnožina vektorového prostoru se nazývá vektorový podprostor, pokud je uzavřená na operace násobení skalárem a sčítání. Podprostory mají řadu hezkých vlastností a často se v lineární algebře vyskytují; například množina řešení každé soustavy s nulovou pravou stranou tvoří vektorový podprostor $\mathbb{R}^n$.

2.4 Afinní podprostory

Podařilo se nám ukázat, že pro soustavu s nulovou pravou stranou všechna řešení tvoří vektorový podprostor. Jak je to v případě obecné pravé strany? Množiny řešení odpovídají *afinním podprostorům*, které nyní zavedeme.⟨8⟩ Vektorové podprostory jsou geometricky přímky, roviny a vícerozměrná zobecnění procházející počátkem. Afinní podprostory jsou vektorové podprostory posunuté z počátku.

⟨6⟩ Například uvažme intervaly na reálné ose. Průnik konečně mnoha otevřených intervalů je otevřený interval. Avšak průnik nekonečně mnoha již otevřený být nemusí. Například pro intervaly $(0, 1 + \frac{1}{n})$, pro všechna přirozená n, platí $\bigcap_{n=1}^{\infty}(0, 1 + \frac{1}{n}) = (0, 1]$. Důvodem je, že intervaly postupně obsahují menší a menší část osy napravo od 1 a ta se zmenšuje k nule. V jejich průniku proto nezůstane nic většího než 1.

⟨7⟩ Pro srovnání s analýzou: Neexistuje limita každé posloupnosti, pouze některé konvergují. Všimněte si, o kolik komplikovanější je práce s limitami než třeba s reálnými čísly. To je jeden z důvodů, proč je analýza mnohem méně intuitivní než lineární algebra, pracuje se slabšími definicemi.

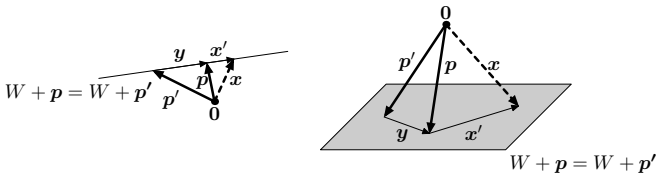
⟨8⟩ Překvapivě v angličtině se používá slovo *affine*, které se na rozdíl od českého afinní píše se dvěma písmeny 'f' a jedním písmenem 'n'.

Afinní podprostor. Nechť W je libovolný vektorový podprostor prostoru $\mathbb{R}^k$ a $\boldsymbol{p}$ je libovolný vektor z $\mathbb{R}^k$. Afinní podprostor $W + \boldsymbol{p}$ je vektorový podprostor W posunutý o $\boldsymbol{p}$:

$$W + \boldsymbol{p} = \{\boldsymbol{x} + \boldsymbol{p} \mid \boldsymbol{x} \in W\}.$$

Jako triviální případ připouštíme i prázdný afinní podprostor. Poznamenejme, že platí $W = W + \boldsymbol{0}$, tedy každý vektorový podprostor je také afinní podprostor. Protože $\boldsymbol{0} \in W$, vektor posunutí $\boldsymbol{p} \in W + \boldsymbol{p}$. Také si můžeme všimnout, že rozdíl libovolných dvou vektorů z $W + \boldsymbol{p}$ leží ve W, protože vektor posunutí se rozdílem vyruší.

Víme, že vektor posunutí leží v afinním podprostoru. Platí však i naopak, že libovolný vektor z afinního podprostoru můžeme zvolit jako vektor posunutí. Tedy formálně pro každý vektor $\boldsymbol{p'} \in W + \boldsymbol{p}$ platí $W + \boldsymbol{p'} = W + \boldsymbol{p}$. Jak toto dokážeme? Důkaz je naznačen na obrázku 2.7. Klíčové je, že $\boldsymbol{y} = \boldsymbol{p} - \boldsymbol{p'}$, rozdíl těchto dvou posunutí, leží ve W. Ukažme inkluzi $W + \boldsymbol{p} \subseteq W + \boldsymbol{p'}$. Pokud $\boldsymbol{x} \in W + \boldsymbol{p}$, potom $\boldsymbol{x} = \boldsymbol{x'} + \boldsymbol{p}$. Ale z toho plyne, že $\boldsymbol{x} = \boldsymbol{x'} + \boldsymbol{y} + \boldsymbol{p'}$, kde $\boldsymbol{x'} + \boldsymbol{y} \in W$. Tedy $\boldsymbol{x} \in W + \boldsymbol{p'}$. Druhá inkluze se dokáže odečtením $\boldsymbol{y}$, důkaz necháme na čtenáři.



Obrázek 2.7: Geometricky naznačený důkaz ve dvou dimenzích (nalevo) a ve třech dimenzích (napravo).

Řešení soustavy s obecnou pravou stranou. V případě obecné pravé strany $\boldsymbol{b}$ je řešením afinní podprostor. Pokud žádné řešení neexistuje, jedná se o prázdný afinní podprostor. Pokud nějaké řešení existuje, platí:

Tvrzení 2.4. *Množina všech řešení soustavy s pravou stranou $\boldsymbol{b}$ je $W + \boldsymbol{p}$, kde:*

- *Vektor $\boldsymbol{p}$ je libovolné řešení s pravou stranou $\boldsymbol{b}$ a*
- *vektorový podprostor W je množina všech řešení této soustavy s vynulovanou pravou stranou.*

Důkaz. Připomeňme, že podle tvrzení 2.3 je W skutečně vektorový podprostor. V důkazu budeme používat: $\boldsymbol{x} \in W + \boldsymbol{p}$, právě když $\boldsymbol{x} - \boldsymbol{p} \in W$. Musíme dokázat dvě implikace:

- $\boldsymbol{x}$ je řešení $\implies$ $\boldsymbol{x} \in W + \boldsymbol{p}$: Stačí ukázat, že $\boldsymbol{x} - \boldsymbol{p} \in W$, tedy že $\boldsymbol{x} - \boldsymbol{p}$ řeší soustavu s vynulovanou pravou stranou. Rozepišme si i-tou rovnici a upravme ji s použitím distributivity a komutativity:

$$a_{i,1}(x_1 - p_1) + a_{i,2}(x_2 - p_2) + \cdots + a_{i,n}(x_n - p_n) = \underbrace{\big(a_{i,1}x_1 + a_{i,2}x_2 + \cdots + a_{i,n}x_n\big)}_{=b_i,\text{ neboť } \boldsymbol{x} \text{ je řešení}} - \underbrace{\big(a_{i,1}p_1 + a_{i,2}p_2 + \cdots + a_{i,n}p_n\big)}_{=b_i,\text{ neboť } \boldsymbol{p} \text{ je řešení}} = b_i - b_i = 0.$$

  Tedy $\boldsymbol{x} - \boldsymbol{p}$ splňuje i-tou rovnici. To platí pro každou rovnici, a proto $\boldsymbol{x} - \boldsymbol{p}$ řeší soustavu s vynulovanou pravou stranou. Platí $\boldsymbol{x} - \boldsymbol{p} \in W$, neboli $\boldsymbol{x} \in W + \boldsymbol{p}$.
- $\boldsymbol{x} \in W + \boldsymbol{p}$ $\implies$ $\boldsymbol{x}$ je řešení: Pokud $\boldsymbol{x} \in W + \boldsymbol{p}$, platí $\boldsymbol{x} - \boldsymbol{p} \in W$, tedy $\boldsymbol{x} - \boldsymbol{p}$ řeší soustavu s vynulovanou pravou stranou. Rozepišme si opět i-tou rovnici soustavy s vynulovanou pravou stranou:

$$0 = a_{i,1}(x_1 - p_1) + a_{i,2}(x_2 - p_2) + \cdots + a_{i,n}(x_n - p_n) = \underbrace{\big(a_{i,1}x_1 + a_{i,2}x_2 + \cdots + a_{i,n}x_n\big)}_{=?,\text{ chceme } b_i} - \underbrace{\big(a_{i,1}p_1 + a_{i,2}p_2 + \cdots + a_{i,n}p_n\big)}_{=b_i,\text{ neboť } \boldsymbol{p} \text{ je řešení}}.$$

  Hodnota levé závorky musí být také b_i, neboť rozdíl je nulový. Tedy $\boldsymbol{x}$ řeší každou z rovnic původní soustavy (s nevynulovanou pravou stranou). □

Množina všech řešení libovolné soustavy tvoří afinní podprostor, což je vektorový podprostor posunutý z počátku.

2.5 Abstraktní definice vektorového prostoru

Na začátku kapitoly jsme definovali vektorový prostor, vektory a jejich operace. Tato definice byla konkrétní, popsali jsme přesně, jak vypadají. V moderní matematice se typicky volí jiný abstraktní přístup, který si nyní ukážeme.

Struktury. Základními objekty matematiky jsou množiny. Často však nepracujeme pouze s množinami, ale uvažujeme na jejich prvcích i nějaké *operace* či *relace*. Například pokud pracujeme s množinou reálných čísel, používáme většinou i jejich operace sčítání, násobení nebo třeba relaci uspořádání $<$. Dává proto smysl na některé množiny nazírat společně s jejich operacemi a relacemi. *Struktura* je objekt, který obsahuje množinu a nějaké operace či nějaké relace.

Nadefinujme si vše formálně. Struktura $\mathbb{S}$ je tvořena:

- *Množinou prvků S.*[9]
- *Nějakými operacemi $\circ_1, \ldots, \circ_k$.* Uvažujme operaci $\circ_i$, která má nějakou aritu r. Operace přiřazuje *každé* r-tici prvků z S *nějaký* prvek S. Výsledek operace je definovaný pro každou r-tici. Formálně je operace $\circ_i$ zobrazení $\circ_i : S^r \to S$. Protože se s operacemi arity jedna a dva setkáváme nejčastěji, mají speciální jména *unární* a *binární*. Uvažme například reálná čísla. Sčítání $+ : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ a násobení $\cdot : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ jsou binární operace. Druhá mocnina $^2 : \mathbb{R} \to \mathbb{R}$ je unární operace. Druhá odmocnina však unární operace není, neboť není definována pro záporná čísla.
- *Nějakými relacemi $R_1, \ldots, R_\ell$.* Každá relace R_i má opět nějakou aritu r. Relace R_i popisuje nějaký vztah na množině S; říká, které r-tice prvků z S jsou v tomto vztahu. Formálně je $R_i \subseteq S^r$. Nejtypičtější relace jsou binární, relace arity dva. Například struktura reálných čísel obsahuje binární relaci uspořádání $< \subseteq \mathbb{R}^2$ takovou, že $(a, b) \in <$, právě když a je menší než b. Příklady struktur s relacemi si ukážeme na závěr této kapitoly, prozatím budou klíčové operace.

Pochopitelně může být $k = 0$ (struktura neobsahuje žádné operace) nebo $\ell = 0$ (neobsahuje žádné relace). Struktura $\mathbb{S}$ se spolu se svými operacemi a relacemi zapisuje do závorek (jako uspořádaná $(k+\ell+1)$-tice):

$$\mathbb{S} = (S, \circ_1, \ldots, \circ_k, R_1, \ldots, R_\ell).$$

Podmnožina $S' \subseteq S$ prvků struktury $\mathbb{S}$ určuje *podstrukturu* $\mathbb{S}'$, pokud je S' uzavřená na operace. Tedy výsledek libovolné z operací $\mathbb{S}$ na prvky z S' je opět prvek S'. Příkladem podstruktury jsou vektorové podprostory vektorového prostoru.

[9] Abychom struktury odlišili od množin, používáme pro struktury druh písma $\mathbb{S}$. Čtenář pravděpodobně zná tento druh značení pro reálná čísla $\mathbb{R}$ nebo třeba komplexní čísla $\mathbb{C}$. To říká, že reálná čísla a komplexní čísla již uvažujeme jako strukturu. Často se také zaměňuje S a $\mathbb{S}$, například se říká $x \in \mathbb{S}$, i když formálně správné by bylo $x \in S$.

Příklady struktur. Ukažme si několik příkladů struktur a podstruktur, které čtenář dobře zná:

- Přirozená čísla spolu s binárními operacemi sčítání a násobení tvoří strukturu $\mathbb{N} = (N, +, \cdot)$. Odčítání není na přirozených číslech korektní operace, rozdíl není vždy definovaný.
- Celá čísla spolu s binárními operacemi sčítání a násobení, tedy $\mathbb{Z} = (Z, +, \cdot)$. Všimněte si, že přirozená čísla $\mathbb{N}$ tvoří podstrukturu $(N, +, \cdot)$ celých čísel. Další příklady podstruktur celých čísel jsou určené jednoprvkovou množinou $\{0\}$ nebo třeba množinou všech sudých čísel.
- Grafy z diskrétní matematiky jsou struktury $\mathbb{G} = (V, E)$, kde V jsou vrcholy a E je symetrická relace „být spojený hranou“.

Abstraktní definice. Připomeňme si, jak jsme na začátku kapitoly popsali vektorové prostory. Vektorový prostor je množina všech n-tic reálných čísel $\mathbb{R}^n$ spolu s konkrétně definovanými operacemi sčítání a násobení. Popsali jsme tedy přesně, jak vektory a operace vypadají, a dokázali jsme řadu vlastností. Vytvořili jsme tedy konkrétní strukturu $\mathbb{R}^n = (\mathbb{R}^n, \cdot, +)$.

V moderní matematice se typicky volí jiný abstraktní přístup. Všimněme si, že při práci s vektory moc nezáleží na tom, jak přesně vektory a jejich operace vypadají. Stačí vědět, že splňují vlastnosti, které potřebujeme. Abstraktní přístup definuje vektorový prostor jako strukturu tvořenou libovolnou množinou prvků (těm budeme říkat vektory) spolu s dvěma operacemi, které splňují krátký seznam vlastností. Při práci s takovou abstraktní strukturou můžeme používat pouze těchto vlastností (a jejich důsledků, které odvodíme), nemůžeme činit žádné další předpoklady (třeba jak operace nebo prvky vypadají).

Operace na vektorech budeme značit do konce kapitoly v kroužku. Důvod je, abychom je odlišili od operací na reálných číslech. Navíc tímto odlišíme konkrétní a abstraktní definici. V dalším textu od tohoto upustíme. Protože vektory značíme tučně, vždy by mělo být jasné, o jakou operaci se jedná.

Abstraktní definice říká, že *vektorový prostor* $\mathbb{V}$ je libovolná struktura $(V, \odot, \oplus)$, která splňuje seznam vlastností uvedený níže. Platí, že V je *množina vektorů*, $\odot$ je binární operace *násobení skalárem* $\odot : \mathbb{R} \times V \to V$ a $\oplus$ je binární operace sčítání $\oplus : V \times V \to V$.[10] Čtenář si může ověřit, že konkrétní definice vektorového prostoru všechny tyto vlastnosti splňuje.

- *Operace sčítání:*
  - *Je komutativní:* Pro každé $\boldsymbol{u}, \boldsymbol{v} \in V$ platí $\boldsymbol{u} \oplus \boldsymbol{v} = \boldsymbol{v} \oplus \boldsymbol{u}$. Vektory můžeme sčítat v libovolném pořadí.
  - *Je asociativní:* Pro každé $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$ platí $(\boldsymbol{u} \oplus \boldsymbol{v}) \oplus \boldsymbol{w} = \boldsymbol{u} \oplus (\boldsymbol{v} \oplus \boldsymbol{w})$. Díky této vlastnosti můžeme vypustit závorky a zapisovat součet prostě jako $\boldsymbol{u} \oplus \boldsymbol{v} \oplus \boldsymbol{w}$, čehož běžně používáme. Pokud by sčítání nebylo asociativní a chtěli bychom přesto závorky vynechávat, museli bychom nadefinovat pořadí, v jakém se operace aplikují. Toho se běžně používá v programovacích jazycích, kde například sčítání čísel s pohyblivou desetinou čárkou není asociativní, rozmyslete si. Pokud vynecháme závorky, aplikuje se součet zleva, tedy $x+y+z$ se interpretuje jako $(x + y) + z$.
  - *Existuje nulový prvek:* Existuje vektor $\mathbf{0} \in V$, že pro každý vektor $\boldsymbol{u} \in V$ platí $\boldsymbol{u} \oplus \mathbf{0} = \boldsymbol{u}$. V konkrétní definici jsme věděli, jak tento neutrální prvek vypadá; vektor samých nul. V abstraktní definici jeho podobu neznáme. Definice pouze zaručuje, že takový prvek existuje. Dále platí, že existuje pouze jediný neutrální prvek. Proč? Kdyby existovaly dva, označme je $\mathbf{0}$ a $\bar{\mathbf{0}}$, platilo by $\mathbf{0} = \mathbf{0} \oplus \bar{\mathbf{0}} = \bar{\mathbf{0}}$, tedy $\mathbf{0} = \bar{\mathbf{0}}$. To ale znamená, že $\mathbf{0}$ a $\bar{\mathbf{0}}$ jsou stejné vektory.

[10] Algebraikům se moc nelíbí operace jako $\odot$, které míchají prvky z různých struktur; prostě občas není dobré míchat jablka s hruškami. Lze však vektorový prostor zadefinovat tak, že každý skalár α bude mít svoji unární operaci $\odot_\alpha : V \to V$, která přiřazuje vektoru $\boldsymbol{u}$ jeho násobek $\alpha \odot \boldsymbol{u}$. Jako malé cvičení zkuste upravit seznam vlastností tak, aby fungoval pro takto pozměněné skalární násobení pomocí unárních operací.

  - *Existují inverzní prvky:* Pro každý vektor $\boldsymbol{u}$ existuje inverzní prvek $-\boldsymbol{u}$ takový, že platí $\boldsymbol{u} \oplus (-\boldsymbol{u}) = \mathbf{0}$. V konkrétní definici se jednalo o opačný vektor. Zkuste rozmyslet, že pro každý vektor je jeho inverzní prvek určený jednoznačně, podobně jako platí jednoznačnost nulového prvku.
- *Násobení skalárem:*
  - *Je „asociativní“:* Pro každé $\alpha, \beta \in \mathbb{R}$ a $\boldsymbol{u} \in V$ platí $\alpha \odot (\beta \odot \boldsymbol{u}) = (\alpha \cdot \beta) \odot \boldsymbol{u}$. Jinými slovy nezáleží, jestli nejprve vektor vynásobíme skalárem α a poté skalárem β, nebo jestli ho rovnou vynásobíme skalárem $\alpha \cdot \beta$. Všimněte si, že také nezáleží na pořadí; vynásobení β a poté α dá stejný výsledek (neboť násobení reálných čísel je komutativní).
  - *Násobení jedničkou:* Pro každý vektor $\boldsymbol{u} \in V$ platí $1 \odot \boldsymbol{u} = \boldsymbol{u}$, natažení jedničkou vektory vůbec nemění.
- *Distributivita:* Ukázali jsme, jak fungují jednotlivé operace samostatně. O jejich společném chování nevíme skoro nic. Tuto chybějící vazbu doplní distributivita.
  - *Distributivita násobení skalárem:* Pro každé $\alpha, \beta \in \mathbb{R}$ a $\boldsymbol{u} \in V$ platí $(\alpha + \beta) \odot \boldsymbol{u} = (\alpha \odot \boldsymbol{u}) \oplus (\beta \odot \boldsymbol{u})$ (závorky budeme vypouštět, násobení skalárem bude mít přednost, často se pro zdůraznění píše $\alpha\boldsymbol{u} \oplus \beta\boldsymbol{u}$). Všimněte si, že z distributivity vyplývá $(-1) \odot \boldsymbol{u} = -\boldsymbol{u}$ a $0 \odot \boldsymbol{u} = \mathbf{0}$. Jako cvičení si to zkuste dokázat.
  - *Distributivita součtu vektorů:* Pro každé $\alpha \in \mathbb{R}$ a pro každé vektory $\boldsymbol{u}, \boldsymbol{v} \in V$ platí $\alpha \odot (\boldsymbol{u} \oplus \boldsymbol{v}) = (\alpha \odot \boldsymbol{u}) \oplus (\alpha \odot \boldsymbol{v})$.

Konkrétní definice, kterou jsme si popsali na začátku kapitoly, všechny tyto vlastnosti splňuje. Čtenář se jistě ptá, jestli je abstraktní definice obecnější než ta konkrétní, tedy jestli ji splňují i nějaké jiné struktury. Tak tomu v případě vektorových prostorů není, od každé velikosti existuje pouze jediný vektorový prostor $\mathbb{R}^n$ a ostatní se liší pouze přejmenováním prvků; to si dokážeme v kapitole 4. Poznamenejme, že aby toto byla přesně pravda, museli bychom konkrétní definici maličko zesílit i na vektory tvořené nekonečně mnoha složkami.

Proč zavádíme abstraktní definici vektorových prostorů, která nic nového nepřinese? Jsou pro to alespoň dva dobré důvody. Často je lepší přemýšlet o vektorech jinak než jen o uspořádaných n-ticích. Například, jak si hned ukážeme, polynomy tvoří vektorový prostor. Pokud tyto vektory budeme uvažovat jako polynomy, má řada operací jako násobení nebo derivování mnohem lepší smysl. Jinými slovy v každé situaci je nejlepší uvážit nejpřirozenější definici, neboť to usnadní uvažování. Druhý důvod je následující. I když se abstraktně definovaný vektorový prostor liší od konkrétního prostoru pouze přejmenováním vektorů, toto přejmenování může být v případě nekonečně složek naprosto obludné. Dokonce se může stát, že sice budeme vědět, že přejmenování existuje, ale budeme schopni dokázat, že se ho nikdy nepodaří zkonstruovat.

> Abstraktní definice říká: Vektorový prostor je libovolná struktura $(V, \odot, \oplus)$ splňující několik základních vlastností. Pokud pracujeme s takto definovanými vektory, můžeme použít pouze těchto vlastností (a jejich důsledků, které vyvodíme).

Exotické příklady. Ukážeme si na závěr kapitoly několik exotických příkladů vektorových prostorů. Čtenář si může jako cvičení dokázat, že to skutečně jsou vektorové prostory.

- Triviální vektorový prostor, který obsahuje pouze počátek $\mathbf{0}$.
- Vektorový prostor všech posloupností reálných čísel $(a_0, a_1, a_2, \dots)$, které se sčítají a násobí po složkách.

- Vektorový prostor všech polynomů $P[x]$ spolu s operacemi sčítání a násobení konstantou. Polynomy stupně nejvýše k jsou jeho podprostor. Každý polynom lze reprezentovat jako vektor, který má nekonečně mnoho složek:

  $$a_0 + a_1 x + a_2 x^2 + \cdot + a_n x^n \text{ lze reprezentovat jako vektor } (a_0, a_1, a_2, \ldots, a_n, 0, 0, \ldots),$$

  pouze konečně mnoho složek je nenulových. Prostor všech polynomů je podprostorem prostoru všech posloupností.
- Vektorový prostor všech funkcí na intervalu $[0, 1]$, spolu s jejich sčítáním a násobením konstantou. Podprostory jsou třeba množina všech spojitých funkcí nebo množina všech funkcí nulových na intervalu $[0, \frac{1}{2}]$. Opět každou funkci lze reprezentovat jako vektor, který má nekonečně mnoho reprezentujících složek, odpovídajících hodnotám v jednotlivých číslech intervalu $[0, 1]$. Poznamenejme, že tentokrát těch složek bude mnohem víc.⟨11⟩
- Komplexní čísla spolu se sčítáním a násobením reálnou konstantou jsou vektorový prostor, vektory jsou dvojice $(x, y) = x + i \cdot y$. Komplexní čísla získají mnohem bohatší strukturu, pokud zavedeme operaci násobení (komplexním číslem). Jeden z významných podprostorů komplexních čísel jsou reálná čísla. To, že reálná čísla jsou vektorový prostor (jehož skaláry jsou zase reálná čísla) může působit trochu zvláštně.

Srovnání s objektovým programováním. Objektové programování je založené na vytváření objektů, které jsou tvořeny daty a funkcemi, kterým se říká metody. Metody pracují s daty a popisují tak chování objektu. Objekt je tedy jakýsi uzavřený samostatný celek. Struktury jsou založené na podobném principu. Struktura je tvořena množinou (což jsou data) a nějakými operacemi a relacemi (což jsou metody), které popisují chování a vlastnosti struktury.

V objektovém programování se také používá princip zvaný dědičnost. Ten funguje tak, že vytvoříme základní objekt, který je často abstraktní, bez konkrétní implementace. Základní objekt popisuje rozhraní, tedy slibuje určitá data a určité metody. Od tohoto základního objektu odvodíme odvozené objekty, které již popisují konkrétní implementace těchto dat a metod. Každý odvozený objekt může mít jinou vnitřní implementaci, ale vnější rozhraní má podle základního objektu. To je přímá paralela s abstraktní definicí vektorového prostoru. Abstraktní definice funguje jako onen základní objekt, popisuje vektorový prostor jako nějakou množinu vektorů spolu s dvě operacemi, které mají splňovat několik základních vlastností. Konkrétní podoba (implementace) vektorů a operací předepsána není. Odvozeným objektem je konkrétní vektorový prostor, který je tvořen konkrétní množinou vektorů a dvěma konkrétními operacemi.

Poznamenejme, že algebraická abstrakce jde podstatně dále než objektové programování. Předně v každé struktuře nalezneme hierarchii podstruktur. Další silný nástroj je faktorizace. Co to znamená? Faktorizace umožňuje ze složitých struktur vyrábět struktury mnohem jednodušší tak, že neztratíme žádné klíčové vlastnosti původní složité struktury. Existují i opačné techniky, které vezmou jednoduchou strukturu a vyrobí z ní strukturu složitější, která ale navíc splňuje nějaké nové vlastnosti. Tyto nástroje jsou základem *univerzální algebry*.

2.6 Svaz vektorových podprostorů a lineární obaly

Uvažme pro daný vektorový prostor množinu $\mathcal{P}$ všech jeho vektorových podprostorů. Na závěr kapitoly ukážeme, že $\mathcal{P}$ tvoří velice pravidelnou strukturu, která se nazývá *úplný svaz*. Nejprve začneme maličkou

⟨11⟩ Kolik? Přesně tolik, kolik je reálných čísel. Slavný Cantorův výsledek říká, že reálných čísel je mnohem víc než přirozených čísel. Cantor to dokázal geniálním trikem, kterému se dnes říká *Cantorova diagonální metoda*. Že nezní toto tvrzení vůbec překvapivě? Pak je třeba poznamenat, že množina racionálních čísel je naopak stejně velká jako množina přirozených čísel. Ale na první pohled vypadají velikosti racionálních a reálných čísel stejně.

odbočkou ze světa lineární algebry.

Částečně uspořádané množiny. Uvažme nějakou množinu X. Jak už jsme zmínili, v matematice se typicky množiny neuvažují samostatně, nýbrž se studují jako struktura spolu s nějakými operacemi nebo relacemi. Například pokud by X byla množina vektorů, můžeme ji studovat spolu s operacemi sčítání a násobení skalárem jako vektorový prostor. Tentokrát však opatříme X jednou binární relací $\leq$. Tato relace se jmenuje *částečné uspořádání* a splňuje následující tři přirozené podmínky:

- *Reflexivita:* Částečné uspořádání $\leq$ je neostré. Pro každé $a \in X$ platí $a \leq a$. Pochopitelně existuje i varianta, které se říká *ostré částečné uspořádání* $<$, pro kterou se naopak vyžaduje antireflexivita; tedy pro žádné $a \in X$ neplatí $a < a$.
- *Antisymetrie:* Neexistují dva různé prvky $a, b \in X$, že současně platí $a \leq b$ a $b \leq a$. Jinak by mezi a a b nebylo žádné uspořádání.
- *Tranzitivita:* Pro každé $a, b, c \in X$ platí, že $a \leq b$ a $b \leq c$ implikuje $a \leq c$.

Vzniklá struktura $(X, \leq)$ splňující tři uvedené podmínky se nazývá *částečně uspořádaná množina*. Zavedeme také užitečnou zkratku $x < y$ za „$x \leq y$ a zároveň $x \neq y$“.

Čtenář určitě zná příklady částečně uspořádaných množin, například $(\mathbb{N}, \leq)$, $(\mathbb{Q}, \leq)$ nebo $(\mathbb{R}, \leq)$. Tato částečná uspořádání jsou značně specifická a nazývají se *lineární uspořádání*. Lineární uspořádání je částečné uspořádání, pro které platí, že libovolné dva různé prvky x a y jsou *porovnatelné*, tedy platí buď $x \leq y$, nebo $y \leq x$. Obecné částečné uspořádání může mít *neporovnatelné* dvojice prvků (x, y), pro které $x \not\leq y$ a $y \not\leq x$.

Hasseho diagramy. Definujeme, že y je *přímý následník* x (a naopak x je *přímý předchůdce* y), pokud platí $x < y$ a zároveň neexistuje z, aby $x < z < y$. Pro uspořádání $(\mathbb{N}, \leq)$ je $k - 1$ přímým předchůdcem čísla k a $k + 1$ jeho přímým následníkem. Pro některá uspořádání jako $(\mathbb{Q}, \leq)$ nebo $(\mathbb{R}, \leq)$ neexistují přímí předchůdci a následníci. Pro prvky $x \leq y$ existuje *řetěz přímých následníků* $a_0, \ldots, a_k$, pokud

$$x = a_0 < a_1 < \cdots < a_{k-1} < a_k = y$$

a a_i je přímý předchůdce a_{i+1}. Pokud je množina X konečná, takový řetěz existuje pro libovolné dva prvky x a y, pro které platí $x \leq y$; důkaz si může čtenář rozmyslet.

Hasseho diagram umožňuje nakreslit některé částečně uspořádané množiny. Podmínkou je, že pro libovolné $x \leq y$ existuje řetěz přímých následníků z x do y. Tedy například pro $(\mathbb{R}, \leq)$ Hasseho diagramem neexistuje. Hasseho diagram reprezentuje prvky množiny X jako body v rovině s podmínkou, že pro $x < y$ umístíme x níže do roviny než y. V Hasseho diagramu spojíme bod reprezentující x úsečkou s každým bodem reprezentujícím přímého následníka x. Myšlenka je, že nekreslíme nadbytečné úsečky pro dvojice $x < y$, které vyplynou z tranzitivity $(X, \leq)$. Příklady Hasseho diagramů jsou na obrázku 2.8.

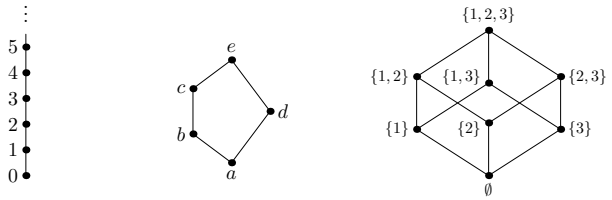
Množinové systémy. Množina $\mathcal{X}$ se nazývá *množinový systém*, pokud prvky této množiny jsou podmnožiny jedné pevné množiny E. Pokud 2^E značí množinu všech podmnožin E, platí $\mathcal{X} \subseteq 2^E$.⟨12⟩

Pro každý množinový systém existuje jedno zcela přirozené částečné uspořádání, a to uspořádání inkluzí $\subseteq$ (relací „být podmnožinou“). Ostatně čtenář si může všimnout nápadné podobnosti mezi symboly $\leq$ a $\subseteq$. Příklad takového uspořádání je na obrázku 2.8 vpravo.

Tvrzení 2.5. *Pro libovolný množinový systém $\mathcal{X}$ je struktura $(\mathcal{X}, \subseteq)$ částečně uspořádaná množina.*

Důkaz. Abychom dokázali, že $\subseteq$ je částečné uspořádání, stačí ověřit tři podmínky z definice. Tento důkaz je cvičením na definici podmnožiny. Ta říká, že $A \subseteq B$, právě když pro každé $x \in A$ platí $x \in B$.

⟨12⟩ Pro zpřehlednění budeme značit množinové systémy kaligrafickým písmem jako $\mathcal{X}, \mathcal{Y}, \ldots$, abychom je odlišili od množin obsahující prvky. Poznamenejme, že obecně v teorii množin je každá množina složena z množin, tedy formálně toto odlišení nedává smysl.



$(\mathbb{N}, \leq)$ $(\{a, b, c, d, e\}, \leq)$ $(2^{\{1,2,3\}}, \subseteq)$

Obrázek 2.8: Tři příklady Hasseho diagramů. Příklad vlevo je lineární uspořádání přirozených čísel. Příklad usprostřed je částečné uspořádání, které není lineární uspořádání, protože dvojice (b, d) a (c, d) jsou neporovnatelné. Příklad vpravo je uspořádání množinového systému všech podmnožin $\{1, 2, 3\}$ inkluzí, vysvětleno níže.

- Reflexivita je splněna, protože pro libovolnou množinu A platí $A \subseteq A$.
- Pokud jsou A a B dvě různé množiny, potom buď A obsahuje prvek, který není v B, nebo B obsahuje prvek, který není v A; jinak by A a B nebyly různé. Proto není možné, aby platilo zároveň $A \subseteq B$ a $B \subseteq A$, tedy dostáváme antisymetrii.
- Tranzitivita také zjevně platí. Když $A \subseteq B \subseteq C$, množina B obsahuje všechny prvky, které leží v A, a C obsahuje všechny prvky, které leží v B. Proto C obsahuje všechny prvky, co leží v A; tedy $A \subseteq C$. □

Svazy. Částečně uspořádaná množina $(X, \leq)$ se nazývá *úplný svaz*, pokud existují infima a suprema pro každou podmnožinu množiny X. Nejprve definujme pro $Y \subseteq X$ minimum a maximum:

$$\min(Y) = x \in Y, \text{ že } \forall y \in Y \text{ platí } x \leq y, \qquad \text{a} \qquad \max(Y) = x \in Y, \text{ že } \forall y \in Y \text{ platí } x \geq y.$$

Pochopitelně minimum či maximum nemusí pro množinu Y existovat a často neexistuje.

Infima a suprema jsou zobecněním minim a maxim na některé další množiny, pro která minima a maxima neexistují. Nechť $S \subseteq X$, potom $\inf(S)$ je největší dolní závora množiny S a $\sup(S)$ je nejmenší horní závora. Formálně:

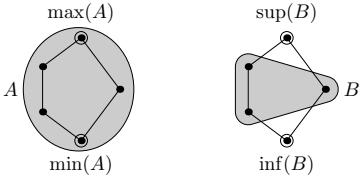
$$\inf(S) = \max\{x \in X : \forall y \in S \; x \leq y\} \qquad \text{a} \qquad \sup(S) = \min\{x \in X : \forall y \in S \; x \geq y\}.$$

Pro obecnou částečně uspořádanou množinu opět nemusí infimum či supremum existovat. V případě úplného svazu $(X, \leq)$ však existují infima a suprema pro každou podmnožinu $S \subseteq X$.

Na obrázku 2.9 je srovnání minim/maxim a infim/suprem pro částečně uspořádanou množinu z obrázku 2.8 uprostřed. Čtenář může zkusit dokázat, že tato částečně uspořádaná množina je úplný svaz.

Uveďme příklad úplného svazu z matematické analýzy. Lineárně uspořádaná množina racionálních čísel $(\mathbb{Q}, \leq)$ netvoří úplný svaz, protože nemusí existovat suprema a infima nekonečně velkých podmnožin. Například supremum množiny $\{x : x^2 \leq 2\}$ neexistuje, neboť $\sqrt{2}$ není racionální číslo. Lineárně uspořádaná množina reálných čísel $(\mathbb{R}, \leq)$ je zúplněním $(\mathbb{Q}, \leq)$, které přidává existenci suprem a infim. Avšak podle výše uvedené definice ani $(\mathbb{R}, \leq)$ netvoří úplný svaz. Axiom o supremu totiž zaručuje existenci suprem pouze neprázdných shora omezených množin. Pokud bychom chtěli získat úplný svaz, můžeme reálná čísla rozšířit o $+\infty$ a $-\infty$ tak, že $-\infty < x < \infty$ pro libovolné $x \in \mathbb{R}$; označme $\mathbb{R}^* = \mathbb{R} \cup \{-\infty, +\infty\}$. Čtenář může dokázat jako cvičení, že rozšířená reálná čísla $(\mathbb{R}^*, \leq)$ tvoří úplný svaz.

Následující pozorování ukazuje, že infima a suprema jsou skutečně zobecněním minim a maxim.



Obrázek 2.9: Vlevo je vyznačena množina A a její minimum a maximum. Vpravo je vyznačena množina B spolu se svým infimem a suprem. Minimum a maximum pro množinu B neexistuje.

Pozorování 2.6. *Pokud pro $S \subseteq X$ existuje minimum, existuje i infimum a platí* $\inf(S) = \min(S)$. *Podobně pokud existuje maximum, existuje i supremum a platí* $\sup(S) = \max(S)$.

Důkaz. Dokážeme pouze první část pro infima, pro suprema je důkaz obdobný. Nejprve si všimněme, že $\min(S)$ je dolní závora S, protože platí $\min(S) \leq y$ pro každé $y \in S$. Protože $\min(S) \in S$ a pro každou dolní závoru x množiny S platí, že je menší či rovna libovolnému prvku z S, platí také $x \leq \min(S)$. Proto je $\min(S)$ největší dolní závora neboli $\inf(S)$. □

Tedy například pro množinu A z obrázku 2.9 platí $\inf(A) = \min(A)$ a $\sup(A) = \max(A)$. Čtenář si může rozmyslet, že pokud naopak $\inf(S) \in S$, potom existuje i minimum a platí $\min(S) = \inf(S)$, a podobně platí pro suprema a maxima.

Svaz vektorových podprostorů. Vraťme se zpátky k lineární algebře. Nechť $\mathcal{P}$ značí množinu všech vektorových podprostorů daného vektorového prostoru. Pokud podprostory uspořádáme inkluzí, dostáváme částečně uspořádanou množinu $(\mathcal{P}, \subseteq)$. Naším hlavním výsledkem je následující tvrzení:

Tvrzení 2.7. *Částečně uspořádaná množina $(\mathcal{P}, \subseteq)$ je úplný svaz.*

Nejprve zkonstruujeme infima. Pro prázdnou množinu je infimum celý prostor. Dále pro $\emptyset \neq \mathcal{S} \subseteq \mathcal{P}$ a pro každé $X \in \mathcal{S}$ zjevně platí $\inf(\mathcal{S}) \subseteq X$. Jako dobrý kandidát na infimum se jeví být průnik všech podprostorů v $\mathcal{S}$:

Pozorování 2.8. *Pro $\emptyset \neq \mathcal{S} \subseteq \mathcal{P}$ je* $\inf(\mathcal{S}) = \bigcap_{X \in \mathcal{S}} X$.

Důkaz. Označme $U = \bigcap_{X \in \mathcal{S}} X$. Podle tvrzení 2.2 je U podprostor, tedy $U \in \mathcal{P}$. Zbývá ověřit, že U splňuje podmínky z definice infima. Protože $U \subseteq X$ pro každé $X \in \mathcal{S}$, je U dolní závora $\mathcal{S}$. Navíc pokud je W dolní závora $\mathcal{S}$, platí $W \subseteq X$ pro každé $X \in \mathcal{S}$, tedy $W \subseteq U$. Tedy U je největší dolní závora $\mathcal{S}$, jinými slovy $\inf(\mathcal{S}) = U$. □

Zkonstruovat supremum je trochu složitější, nechť $\mathcal{S} \subseteq \mathcal{P}$. Pro supremum $\sup(\mathcal{S})$ platí, že $X \subseteq \sup(\mathcal{S})$ pro každé $X \in \mathcal{S}$. Označme pro podprostor W vlastnost „$X \subseteq W$ pro každé $X \in \mathcal{S}$“ pomocí $\mathcal{S} \subseteq W$. Z podprostorů W splňujících $\mathcal{S} \subseteq W$ potřebujeme vybrat ten nejmenší, k čemuž lze využít průnik všech takových podprostorů.

Pozorování 2.9. *Pro $\mathcal{S} \subseteq \mathcal{P}$ je* $\sup(\mathcal{S}) = \bigcap_{W \in \mathcal{P}, \mathcal{S} \subseteq W} W$.

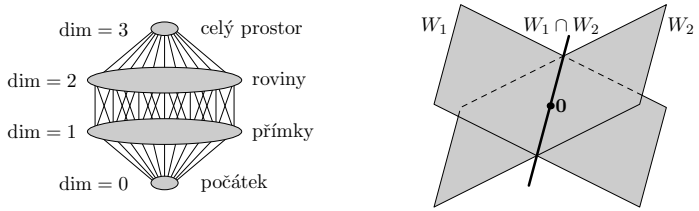
Důkaz. Opět je potřeba ověřit, že popsaná množina je supremum. Označme $U = \bigcap_{W \in \mathcal{P}, \mathcal{S} \subseteq W} W$. Jedná se o průnik podprostorů, tedy U je podle tvrzení 2.2 je podprostor. Protože pro každý podprostor W v průniku platí $\mathcal{S} \subseteq W$, platí i $\mathcal{S} \subseteq U$. Podprostor U je horní závora $\mathcal{S}$ a pro každou horní závoru W platí $U \subseteq W$. Tedy U je supremum $\sup(\mathcal{S})$. □

Možná tato konstrukce suprema působí zvláštně a není moc jasné, jak jsme na ni přišli. Na druhou stranu, když už známe tu správnou konstrukci, je velice jednoduché dokázat její správnost. Pro lepší pochopení si čtenář může rozmyslet následující cvičení: Mějme částečně uspořádanou množinu $(X, \leq)$ s vlastností, že pro každé $S \subseteq X$ existuje infimum $\inf(S)$. Potom dokažte, že existuje pro každé $S' \subseteq X$ supremum $\sup(S')$, a určete, jak vypadají suprema v závislosti na infimech. Pochopitelně i obráceně plyne z existence suprem existence infim. Tedy pokud chceme pro danou částečně uspořádanou množinu $(X, \leq)$ ověřit, zda je to úplný svaz, stačí ověřit pouze existenci infim nebo pouze existenci suprem.

Důkaz tvrzení 2.7. Infimum prázdné množiny je celý prostor, existence infim neprázdných množin plyne z pozorování 2.8. Pozorování 2.9 dokazuje existence suprem. Proto je $(\mathcal{P}, \subseteq)$ úplný svaz. □

Pro daný vektorový prostor, množina všech jeho vektorových podprostorů $\mathcal{P}$ uspořádaná inkluzí tvoří úplný svaz: Existuje infimum a supremum pro každou podmnožinu $\mathcal{S} \subseteq \mathcal{P}$.

Příklad. Jako příklad uvažme vektorový prostor $\mathbb{R}^3$. Jak už jsme zmínili (zatím bez důkazu), podprostory $\mathbb{R}^3$ jsou počátek, přímky procházející počátkem, roviny procházející počátkem a celý prostor. Na obrázku 2.10 vlevo je schématicky naznačen Hasseho diagram $(\mathcal{P}, \subseteq)$. Diagram se skládá ze čtyř vrstev, a jak uvidíme později, tyto vrstvy odpovídají podprostorům dimenze nula, jedna, dva a tři. Na obrázku 2.10 vpravo je zobrazeno infimum pro dvě roviny v $\mathbb{R}^3$.



Obrázek 2.10: Vlevo je Hasseho diagram $(\mathcal{P}, \subseteq)$ pro $\mathbb{R}^3$, ve kterém jsou úsečkou spojeny počátek s každou přímkou, celý prostor s každou rovinou a rovina s přímkou, pokud přímka leží v této rovině. Napravo jsou dvě roviny W_1 a W_2 v $\mathbb{R}^3$. Jejich infimum je tučně vyznačená přímka $W_1 \cap W_2$. Supremum W_1 a W_2 je celý prostor $\mathbb{R}^3$, neboť žádný jiný podprostor neobsahuje současně obě roviny.

Lineární obaly. Nyní uvedeme jednu z klíčových definic lineární algebry. Mějme množinu vektorů X, která není vektorový podprostor. To může být občas problém, pokud potřebujeme u X některou z užitečných vlastností vektorových podprostorů. Proto bychom chtěli X maličko pozměnit a vytvořit z něj X', které bude vektorový podprostor. Jednou z možností je odebrat prvky z X, avšak nemusí existovat žádný vektorový podprostor X' s vlastností $X' \subseteq X$. Druhou možností je do X přidat prvky a vytvořit tak vektorový podprostor. Můžeme za X' pochopitelně zvolit celý vektorový prostor, ale rádi bychom zvolili X' co nejmenší, aby se od X co nejméně lišilo. Takovému X' říkejme *lineární obal* X.

Zadefinujme vše přesně. Nedává smysl požadovat, aby X' byl nejmenší podprostor do velikosti, protože každý vektorový podprostor (až na triviální počátek) obsahuje nekonečně mnoho prvků. Proto budeme požadovat minimalitu do inkluze, se kterou jsem již pracovali při konstrukci infim a suprem úplného svazu $(\mathcal{P}, \subseteq)$. Pro množinu vektorů X definujeme její lineární obal $\mathcal{L}(X)$ jako do inkluze nejmenší vektorový podprostor W s vlastností $X \subseteq W$.[13]

[13] Občas se používají i jiná označení, například $\operatorname{span}(X)$ nebo $\langle X \rangle$.

Z tvrzení 2.7 vyplývá, že lineární obal je pro libovolnou množinu X dobře definovaný, tedy že pro libovolnou množinu X do inkluze nejmenší podprostor existuje:

Důsledek 2.10. *Platí*

$$\mathcal{L}(X) = \inf\{W : W \in \mathcal{P}, X \subseteq W\} = \bigcap_{\substack{W \in \mathcal{P} \\ X \subseteq W}} W.$$

Důkaz. Druhá rovnost platí z konstrukce infim popsané v pozorování 2.8, zbývá odargumentovat první rovnost. Označme $\mathcal{U} = \{W : W \in \mathcal{P}, X \subseteq W\}$. Podle tvrzení 2.7 infimum množiny $\mathcal{U}$ existuje, je to vektorový podprostor a platí $\inf(\mathcal{U}) \subseteq W$ pro každý podprostor W splňující $X \subseteq W$. Protože platí $X \subseteq W$ pro každý podprostor z $\mathcal{U}$, platí z konstrukce infima také, že $X \subseteq \inf(\mathcal{U})$. Tedy $\inf(\mathcal{U})$ je do inkluze nejmenší podprostor obsahující X, což je lineární obal $\mathcal{L}(X)$. □

Podobnost s výše uvedenou konstrukcí suprema není náhodná; čtenář může ověřit, že platí rovnost $\sup(\mathcal{S}) = \mathcal{L}(\bigcup_{X \in \mathcal{S}} X)$.

Klíčová definice lineární algebry: Pro množinu vektorů X je lineární obal $\mathcal{L}(X)$ do inkluze nejmenší vektorový podprostor obsahující všechny vektory z X. Lineární obal X je roven průniku všech vektorových podprostorů obsahujících X.

Lineární obaly mají další důležitou aplikaci, kterou si teď naznačíme a která je centrální pro kapitolu 4. Mějme vektorový podprostor W, který chceme co nejjednodušeji popsat. Přestože W může obsahovat spoustu vektorů, geometrická struktura podprostoru umožní následující jednoduchou charakterizaci. Chtěli bychom nalézt množinu X, pro kterou platí $\mathcal{L}(X) = W$. Zjevně pro každý vektorový podprostor W taková množina existuje, protože platí $\mathcal{L}(W) = W$. Cílem kapitoly 4 bude zkonstruovat množinu X, která bude co nejmenší.

Velikost nejmenší množiny X splňující $\mathcal{L}(X) = W$ se nazývá *dimenze* vektorového podprostoru W. Například pro rovinu procházející počátkem existuje dvouprvková množina X taková, že lineární obal X je tato rovina. Navíc neexistuje jednoprvková množina X s touto vlastností, a tedy dimenze roviny je dva. V tomto textu převážně studujeme vektorové prostory konečné dimenze, pro které každý podprostor má konečnou dimenzi; pro každý podprostor W existuje konečná množina X, pro kterou platí $\mathcal{L}(X) = W$.

Shrnutí

V této kapitole jsme popsali dvě definice vektorového prostoru. Vektorový prostor je struktura tvořená množinou vektorů a dvěma operacemi násobení skalárem a sčítání. Konkrétní definice přesně popisuje tyto vektory a operace. Výhoda je, že konkrétní definice je intuitivní a dobře se geometricky představuje; vektory odpovídají bodům v n-rozměrném prostoru, násobení skalárem natahuje vektory, sčítání funguje jako skládání sil. V algebře se spíše používá abstraktní definice. Ta říká, že vektorový prostor je množina vektorů spolu s dvěma operacemi, které splňují několik základních vlastností. Tedy abstraktní definice nepopisuje, jak přesně vektory a operace vypadají, pouze zaručuje určité vlastnosti.

Také jsme ukázali dvě geometrické interpretace soustavy lineárních rovnic, s použitím řádků a s použitím sloupců. Množina všech řešení každého řádku je nadrovina a řešení celé soustavy je průnik těchto nadrovin. Sloupcová interpretace říká, že při řešení soustavy hledáme koeficienty natažení sloupcových vektorů tak, aby se tyto vektory sečetly na vektor pravé strany.

Zavedli jsme vektorové podprostory jako množiny vektorů uzavřené na operace. Také jsme zavedli afinní podprostory jako vektorové podprostory posunuté z počátku o nějaký vektor. Dokázali jsme, že

množina všech řešení soustavy s pravou stranou nulovou tvoří vektorový podprostor $\mathbb{R}^n$ a množina všech řešení obecné soustavy tvoří afinní podprostor $\mathbb{R}^n$.

Nakonec jsme dokázali, že množina všech vektorových podprostorů uspořádaná inkluzí tvoří úplný svaz, tedy existují infima a suprema pro libovolnou podmnožinu podprostorů. Pomocí toho jsme zadefinovali klíčovou definici lineárního obalu, která danou množinu vektorů X rozšíří na do inkluze nejmenší vektorový podprostor, který obsahuje X.

Cvičení

2.1 Uvažte množinu $\mathbb{P}$ všech polynomů stupně nejvýše n, spolu s operací sčítání a násobení konstantou. Dokažte, že $\mathbb{P}$ tvoří vektorový prostor, tedy ověřte, že jsou splněny všechny axiomy.

Kapitola 3

Matice

V předchozí kapitole jsme zavedli definici vektorových prostorů s vektory a jejich operacemi. Cílem této kapitoly je zavést další klíčovou definici: matice spolu s jejich operacemi. Matice a vektory tvoří základní jazyk lineární algebry. Ostatně matice se objevily již v kapitole 1, takže musí být důležité.

Vektory geometricky odpovídají bodům v n-rozměrném prostoru. Pro matice zatím nepopíšeme žádnou geometrickou reprezentaci, alespoň ne pořádně. Geometrické reprezentaci se budeme věnovat až v kapitole 5, protože musíme nejprve v kapitole 4 vybudovat teorii bází. Geometrická reprezentace matic je složitější; matice odpovídají *lineárním zobrazením* mezi vektorovými prostory.

V této kapitole si také představíme klíčový koncept maticové inverze a regulárních matic. Ukážeme si, že soustavy lineárních rovnic a elementární úpravy lze vyjádřit velice pohodlně v řeči matic. Nakonec si představíme první z řady slavných maticových dekompozic, LU dekompozici.

3.1 Matice a jejich operace

Nejprve matice zadefinujeme a popíšeme si jejich operace.

Matice. Matice je tabulka reálných čísel $m \times n$, kde m je počet řádků a n je počet sloupců. Matice se značí velkými písmeny, například A. Koeficienty matice se značí malými písmeny, například $a_{i,j}$ značí koeficient na pozici (i, j) (v i-tém řádku a j-tém sloupci). Celá matice se zapíše do závorek, jak jsme už jsme viděli v kapitole 1:

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & a_{2,3} & \cdots & a_{2,n} \\ a_{3,1} & a_{3,2} & a_{3,3} & \cdots & a_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & a_{m,3} & \cdots & a_{m,n} \end{pmatrix}, \qquad \text{například} \qquad A = \begin{pmatrix} 2 & 1 \\ 0 & -1 \end{pmatrix}.$$

Koeficient na pozici (i, j) matice A se někdy také značí $(A)_{i,j}$. Množina všech matic velikosti $m \times n$ se značí $\mathbb{R}^{m\times n}$. Maticím, které mají stejný počet řádků a sloupců, se říká *čtvercové*.

Na matice lze syntakticky nahlížet jako na dvojrozměrné vektory, ostatně značení je velice podobné. S vektory se často pracuje jako kdyby to byly matice, které mají pouze jediný řádek nebo pouze jediný sloupec (typicky se uvažují v lineární algebře sloupcové vektory). Tedy operace, které pro matice zavedeme, fungují i pro vektory, čehož budeme využívat. Zde je však důležité zmínit, že matice a vektory mají v lineární algebře jiný význam. Geometricky n-složkové vektory odpovídají bodům v $\mathbb{R}^n$ a matice $m\times n$ odpovídají *lineárním zobrazením* z $\mathbb{R}^n$ do $\mathbb{R}^m$. Co je to lineární zobrazení přesně vysvětlíme později, čtenář si zatím může představovat zobrazení hezkých vlastností. Interpretaci matic pomocí lineárních zobrazení vysvětlíme podrobně v kapitole 5.

Na jednotlivé řádky matice budeme pohlížet jako na *řádkové vektory* z $\mathbb{R}^n$, jejich složky jsou koeficienty v jednotlivých řádcích. Podobně z koeficientů ve sloupcích dostaneme *sloupcové vektory* z $\mathbb{R}^m$.

Maticové operace. Na matice lze stejně jako na vektory aplikovat operace násobení skalárem a sčítání, které se provádí opět po složkách. Matice A vynásobená skalárem α tedy vypadá:

$$\alpha A = \begin{pmatrix} \alpha \cdot a_{1,1} & \alpha \cdot a_{1,2} & \alpha \cdot a_{1,3} & \cdots & \alpha \cdot a_{1,n} \\ \alpha \cdot a_{2,1} & \alpha \cdot a_{2,2} & \alpha \cdot a_{2,3} & \cdots & \alpha \cdot a_{2,n} \\ \alpha \cdot a_{3,1} & \alpha \cdot a_{3,2} & \alpha \cdot a_{3,3} & \cdots & \alpha \cdot a_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha \cdot a_{m,1} & \alpha \cdot a_{m,2} & \alpha \cdot a_{m,3} & \cdots & \alpha \cdot a_{m,n} \end{pmatrix}.$$

Sčítat můžeme pouze matice stejné velikosti. Pro dvě matice A a B z $\mathbb{R}^{m\times n}$ je součet $A+B$ matice z $\mathbb{R}^{m\times n}$:

$$A + B = \begin{pmatrix} a_{1,1}+b_{1,1} & a_{1,2}+b_{1,2} & a_{1,3}+b_{1,3} & \cdots & a_{1,n}+b_{1,n} \\ a_{2,1}+b_{2,1} & a_{2,2}+b_{2,2} & a_{2,3}+b_{2,3} & \cdots & a_{2,n}+b_{1,n} \\ a_{3,1}+b_{3,1} & a_{3,2}+b_{3,2} & a_{3,3}+b_{3,3} & \cdots & a_{3,n}+b_{1,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m,1}+b_{m,1} & a_{m,2}+b_{m,2} & a_{m,3}+b_{m,3} & \cdots & a_{m,n}+b_{m,n} \end{pmatrix}.$$

Podobně jako pro vektory zavedeme zkratku $(-A)$ za $(-1)A$ a zkratku $A-B$ za $A+(-B)$.

Všimněme si, že množina všech matic $\mathbb{R}^{m\times n}$ tvoří spolu operacemi násobení skalárem a sčítání vektorový prostor; stačí ověřit definici abstraktního vektorového prostoru. Na tom není nic překvapivého, protože můžeme vzít řádky (či sloupce) matice a uspořádat je do jednoho dlouhého vektoru, který bude mít mn složek: Sčítání matic přesně odpovídá sčítání těchto dlouhých vektorů a násobení skalárem odpovídá násobení skalárem pro dlouhé vektory. Tento vektorový prostor matic není nic jiného než $\mathbb{R}^{mn}$.

Pro matici $A \in \mathbb{R}^{m\times n}$ zavedeme *transponovanou matici* $A^\mathsf{T} \in \mathbb{R}^{n\times m}$, jejíž koeficienty jsou zrcadlový obraz koeficientů A podle diagonály:

$$(A^\mathsf{T})_{i,j} = (A)_{j,i}, \qquad \forall i = 1,\dots,n,\ j = 1,\dots,m.$$

Na T lze nahlížet jako na unární operaci *transpozice*, která maticím z $\mathbb{R}^{m\times n}$ přiřazuje jejich transponované matice z $\mathbb{R}^{n\times m}$. Čtvercová matice A, pro kterou platí $A = A^\mathsf{T}$, se nazývá *symetrická*. Všimněte si, že pokud je $\boldsymbol{x}$ sloupcový vektor, je vektor $\boldsymbol{x}^\mathsf{T}$ řádkový vektor. Pokud budeme vektory používat v kombinaci s maticemi, budou všechny vektory sloupcové. Pokud budeme chtít řádkový vektor, použijeme explicitně transpozici.

> Matice $m \times n$ je tabulka reálných čísel, která má m řádků a n sloupců. Množina všech matic $m\times n$ se značí $\mathbb{R}^{m\times n}$. Matice z $\mathbb{R}^{m\times n}$ spolu s operacemi sčítání a násobením skalárem tvoří vektorový podprostor. Transpozice zrcadlí koeficienty matice podle diagonály. Čtvercová matice se nazývá symetrická, pokud je rovná své transpozici.

Maticové násobení. Na maticích se však také definuje *maticové násobení*, což je trochu složitější operace. Maticové násobení je definováno pouze pro kompatibilní velikosti matic. Nejprve si popišme, jak násobení *nevypadá*. Nejpřirozenější definice, která každého určitě hned napadne, je násobit pouze matice stejné velikosti a provádět násobení stejně jako sčítání po složkách. Tak tomu není, definice násobení je zcela odlišná![1]

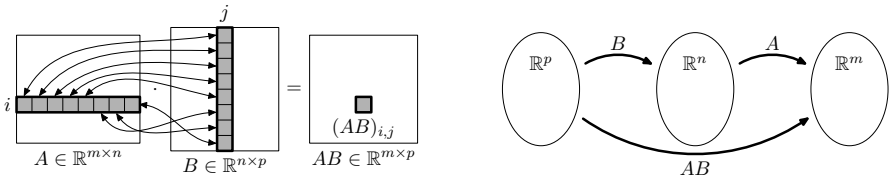
[1] Na druhou stranu i tento druh násobení má svoje uplatnění a v určitých oblastech teorie lineární algebry se používá. Nazývá se *Hadamardův součin* nebo *součin po složkách*.

Násobit můžeme pouze matici $m \times n$ s maticí $n \times p$ a výsledkem je matice $m \times p$, pro libovolné rozměry m, n a p. Jak si toto dobře zapamatovat? Pokud zapíšeme tyto rozměry za sebe, dostaneme čtyři čísla m, n, n a p. Vnitřní dva rozměry, tedy počet sloupců první matice a počet řádků druhé matice, musí být stejné. Výsledkem je matice, jejíž velikost odpovídá dvěma vnějším rozměrům, tedy počtu řádků první matice a počtu sloupců druhé matice.

Tedy nechť $A \in \mathbb{R}^{m\times n}$ a $B \in \mathbb{R}^{n\times p}$, výsledek násobení je matice $AB \in \mathbb{R}^{m\times p}$, pro kterou platí:

$$(AB)_{i,j} = a_{i,1}b_{1,j} + a_{i,2}b_{2,j} + \cdots + a_{i,n}b_{n,j} = \sum_{k=1}^{n} a_{i,k}b_{k,j}, \qquad \forall i = 1,\dots,m,\ j = 1,\dots,p.$$

Rozmysleme si, co tato formule říká. Chceme spočítat koeficient AB na pozici (i,j), tedy koeficient v i-tém řádku a j-tém sloupci. Uvážíme vektor i-tého řádku matice A a vektor j-tého sloupce matice B. Tyto dva vektory mají stejný počet složek, roven n (jinak by násobení AB nebylo definované). Definice říká, že složky vynásobíme po dvojicích a sečteme. Pro lepší představu se podívejte na obrázek 3.1 vlevo.

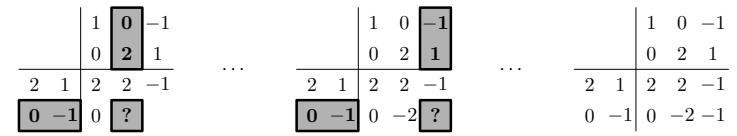


Obrázek 3.1: Nalevo je naznačeno, které složky sloupcového vektoru A a řádkového vektoru B se spolu násobí. Po sečtení těchto součinů dostaneme hodnotu koeficientu $(AB)_{i,j}$. Napravo je interpretace násobení matic jako skládání lineárních zobrazení.

Příklad. Uveďme příklad maticového součinu. Mějme matice $A \in \mathbb{R}^{2\times 2}$ a $B \in \mathbb{R}^{2\times 3}$, potom součin $AB \in \mathbb{R}^{2\times 3}$ a součin BA není definován. Pro konkrétní čísla:

$$\text{Pro} \quad A = \begin{pmatrix} 2 & 1 \\ 0 & -1 \end{pmatrix} \quad \text{a} \quad B = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 2 & 1 \end{pmatrix} \quad \text{je} \quad AB = \begin{pmatrix} 2 & 2 & -1 \\ 0 & -2 & -1 \end{pmatrix}. \tag{3.1}$$

Pokud chceme násobit matice na papíře, můžeme si A a B zapsat do tabulky jako na obrázku 3.2; minimalizujeme tak šanci, že vynásobíme políčka špatných souřadnic. Hodnoty AB dopočítáváme postupně, vždy vezmeme jeden řádek A a jeden sloupec B.



Obrázek 3.2: Součin zapíšeme do tabulky a dopočítáváme hodnoty AB jednu po druhé.

> Maticové násobení se definuje pro matice $A \in \mathbb{R}^{m\times n}$ a $B \in \mathbb{R}^{n\times p}$. Výsledek AB je matice z $\mathbb{R}^{m\times p}$. Koeficienty AB jsou sumy koeficientů A a B:
>
> $$(AB)_{i,j} = \sum_{k=1}^{n} a_{i,k}b_{k,j}, \qquad \forall i = 1,\dots,m,\ j = 1,\dots,p.$$

Význam maticového násobení. Proč se násobí matice tak zvláštně? Důvod je ten, že toto násobení má výhodné vlastnosti. Již jsme zmínili, že matice z $\mathbb{R}^{m\times n}$ odpovídá lineárnímu zobrazení z $\mathbb{R}^n$ do $\mathbb{R}^m$. Násobení odpovídá skládání těchto zobrazení, což je velice důležitá vlastnost. Konkrétně násobek AB matice $A \in \mathbb{R}^{m\times n}$ a $B \in \mathbb{R}^{n\times p}$ odpovídá $A \circ B$, kde se nejprve aplikuje zobrazení B a poté zobrazení A; zobrazení se skládají zprava doleva. Všimněme si, že

$$B : \mathbb{R}^p \to \mathbb{R}^n, \qquad A : \mathbb{R}^n \to \mathbb{R}^m, \qquad \text{tedy} \qquad A \circ B : \mathbb{R}^p \to \mathbb{R}^m.$$

Proto dává smysl, že výsledek AB je matice z $\mathbb{R}^{m\times p}$; složení je lineární zobrazení $\mathbb{R}^p \to \mathbb{R}^m$. Aby složení bylo možné, musí na sebe zobrazení B a A pasovat (a proto musí být vnitřní rozměry při násobení stejné). Skládání lineárních zobrazení je naznačeno na obrázku 3.1 vpravo.

A jaké zobrazení tedy určuje matice A z $\mathbb{R}^{m\times n}$? Nechť $\boldsymbol{x} \in \mathbb{R}^n$. Součin $A\boldsymbol{x}$ je nějaký vektor z $\mathbb{R}^m$. Matice A tedy přiřazuje každému vektoru $\boldsymbol{x}$ z $\mathbb{R}^n$ nějaký vektor $A\boldsymbol{x}$ z $\mathbb{R}^m$, což je slibované zobrazení:

$$A : \boldsymbol{x} \mapsto A\boldsymbol{x}.$$

Tato zobrazení určená maticí mají řadu pěkných vlastností a budeme se jim podrobně věnovat v kapitole 5.

Soustavy v řeči matic. Uvažme součin nějaké matice $A \in \mathbb{R}^{m\times n}$ a nějakého vektoru $\boldsymbol{x} \in \mathbb{R}^n$ (matice budou vždy násobit vektory zleva). Výsledek tohoto součinu je nějaký jiný vektor $\boldsymbol{b} \in \mathbb{R}^m$. Rozepišme z definice násobení, jak vypadají jeho složky:

$$b_i = \sum_{j=1}^{n} a_{i,j}x_j = a_{i,1}x_1 + a_{i,2}x_2 + a_{i,3}x_3 + \cdots + a_{i,n}x_n, \qquad \forall i = 1,\dots,m.$$

Nepřipomíná vám tento součin jeden řádek soustavy lineárních rovnic? V řeči matic je soustava lineárních rovnic prostě $A\boldsymbol{x} = \boldsymbol{b}$. V řeči lineárních zobrazení jsou řešením soustavy všechny vzory $\boldsymbol{b}$ v lineárním zobrazení definovaným maticí A.

Soustavu lineárních rovnic lze zapsat v řeči matic jako

$$A\boldsymbol{x} = \boldsymbol{b},$$

hledáme pro pevnou matici A a pevný vektor $\boldsymbol{b}$ všechny vektory $\boldsymbol{x}$ splňující tuto rovnost.

Vlastnosti operací. Maticové operace splňují řadu hezkých vlastností, uveďme si alespoň některé. Protože matice spolu s operacemi sčítání a násobení skalárem tvoří vektorový prostor, platí pro ně všechny vlastnosti vektorového prostoru. Popíšeme tedy pouze nové vlastnosti, v kterých se vyskytuje maticové násobení či transpozice. Následující rovnosti mají smysl pouze tehdy, pokud všechny matice mají kompatibilní rozměry, což budeme předpokládat.

- *Násobení je asociativní*, tedy pro libovolné matice A, B a C platí $(AB)C = A(BC)$. Proto v maticových součinech můžeme vynechávat závorky. Poznamenejme, že obecně skládání libovolných zobrazení je asociativní.
- *Násobení obecně není komutativní.* Ostatně pro součin AB nemusí součin BA být vůbec definován, viz příklad (3.1) výše. I když jsou oba součiny AB a BA definovány (potom jsou nutně obě matice čtvercové $n \times n$), mohou být různé:

$$A = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}, \qquad B = \begin{pmatrix}1 & 2\\ 3 & 4\end{pmatrix}, \qquad AB = \begin{pmatrix}3 & 4\\ 1 & 2\end{pmatrix}, \qquad BA = \begin{pmatrix}2 & 1\\ 4 & 3\end{pmatrix}.$$

  Tento protipříklad není jakkoliv speciální, součin komutuje pouze výjimečně. Skládání zobrazení také typicky není komutativní.

- *Násobení je distributivní.* Pro libovolné matice A, B a C platí $(A+B)C = AC + BC$, podobně z druhé strany $C(A+B) = CA + CB$.
- *Součin skalárních násobků.* Pro libovolné matice A a B a pro libovolná reálná čísla α a β platí $(\alpha A)(\beta B) = (\alpha\cdot\beta)(AB)$. Opět to ospravedlňuje opomíjení závorek a zápis $\alpha\beta AB$.
- *Transpozice součinu.* Pro libovolné matice A a B platí $(AB)^\mathsf{T} = B^\mathsf{T}A^\mathsf{T}$. Povšimněme si, že se transponované matice objevují v obráceném pořadí. Uvědomme si, že rovnost bez obráceného pořadí $(AB)^\mathsf{T} = A^\mathsf{T}B^\mathsf{T}$ nemůže obecně platit, neboť pro $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times p}$ a pro $m \neq p$ není součin $A^\mathsf{T}B^\mathsf{T}$ na rozdíl od $(AB)^\mathsf{T}$ vůbec definován!
- *Transpozice součtu.* Podobně pro libovolné matice A a B platí $(A+B)^\mathsf{T} = A^\mathsf{T} + B^\mathsf{T}$.

Dokažme si alespoň dvě vlastnosti, zbývající může čtenář dokázat jako cvičení.

Tvrzení 3.1. *Pro libovolné matice $A \in \mathbb{R}^{m\times n}$ a $B \in \mathbb{R}^{n\times p}$ platí $(AB)^\mathsf{T} = B^\mathsf{T}A^\mathsf{T}$.*

Důkaz. Abychom ukázali, že matice $(AB)^\mathsf{T}$ a $B^\mathsf{T}A^\mathsf{T}$ jsou stejné, musíme ukázat, že se shodují ve všech koeficientech, tedy pro každé i a j platí $(AB)^\mathsf{T}_{i,j} = (B^\mathsf{T}A^\mathsf{T})_{i,j}$. Všimněte si, že jsme se od matic dostali k reálným číslem, pro která můžeme použít jejich vlastnosti. Rozepíšeme tedy koeficient $(AB)^\mathsf{T}_{i,j}$ podle definice násobení, prohodíme pořadí v sumě a přepíšeme sumu jako $(B^\mathsf{T}A^\mathsf{T})_{i,j}$:

$$(AB)^\mathsf{T}_{i,j} = (AB)_{j,i} = \sum_{k=1}^{p}(A)_{j,k}(B)_{k,i} = \sum_{k=1}^{p}(B^\mathsf{T})_{i,k}(A^\mathsf{T})_{k,j} = (B^\mathsf{T}A^\mathsf{T})_{i,j}, \qquad \forall i = 1,\dots,p,\ j = 1,\dots,m.$$

□

Tvrzení 3.2. *Násobení matic je asociativní, tedy pro libovolné matice $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times p}$ a $C \in \mathbb{R}^{p\times q}$ platí $(AB)C = A(BC)$.*

Důkaz. Potřebujeme dokázat, že se matice $(AB)C$ a $A(BC)$ shodují v každém koeficientu. Myšlenka důkazu je úplně stejná, nejprve rozepíšeme součin $(AB)C$ podle definice, vzniklé sumy přeuspořádáme a nakonec přepíšeme na součin $A(BC)$. Úprava je tentokrát trochu složitější:

$$\big((AB)C\big)_{i,j} \overset{(1)}{=} \sum_{k=1}^{p}(AB)_{i,k}c_{k,j} \overset{(2)}{=} \sum_{k=1}^{p}\left(\sum_{\ell=1}^{n}a_{i,\ell}b_{\ell,k}\right)c_{k,j} \overset{(3)}{=} \sum_{k=1}^{p}\left(\sum_{\ell=1}^{n}a_{i,\ell}b_{\ell,k}c_{k,j}\right) \overset{(4)}{=}$$
$$\overset{(4)}{=} \sum_{\ell=1}^{n}\left(\sum_{k=1}^{p}a_{i,\ell}b_{\ell,k}c_{k,j}\right) \overset{(5)}{=} \sum_{\ell=1}^{n}a_{i,\ell}\left(\sum_{k=1}^{p}b_{\ell,k}c_{k,j}\right) \overset{(6)}{=} \sum_{\ell=1}^{n}a_{i,\ell}(BC)_{\ell,j} \overset{(7)}{=} \big(A(BC)\big)_{i,j}.$$

Vysvětleme si pořádně, proč platí jednotlivé rovnosti:

(1–2) Rozepíšeme maticové součiny do sum podle definice.

(3) Přesuneme člen $c_{k,j}$ do vnitřní sumy podle ℓ. To umožňuje distributivita násobení reálných čísel, která říká $\beta\sum_i\alpha_i = \sum_i\beta\alpha_i$. Můžeme tedy β přesunout do sumy. Pokud každý člen sumy vynásobíme $c_{k,j}$, dostaneme stejnou hodnotu jako když celou sumu vynásobíme $c_{k,j}$.

(4) Prohodíme pořadí sum, což má následující význam. Představme si, že potřebujeme sečíst všechna čísla v tabulce. Pokud nejprve sečteme čísla v řádcích a poté sečteme součty řádků, dostaneme stejný výsledek jako když nejprve sečteme čísla ve sloupečcích a poté sečteme součty sloupců. Toto platí díky komutativitě sčítání.

(5) Pomocí distributivity vytkneme člen $a_{i,\ell}$ ze sumy podle k. Proč to můžeme udělat? Pro celou sumu podle k je $a_{i,\ell}$ jedno pevné reálné číslo (které nezávisí na hodnotě k), hodnota $a_{i,\ell}$ se mění pouze podle vnější sumy podle ℓ. Proto každý člen $b_{\ell,k}c_{k,j}$ sumy podle k je vynásobený stejným $a_{i,\ell}$, které můžeme vytknout.

(6–7) Přepíšeme sumy zpět na součiny matic.

Tato úprava je značně technická, i když ve své podstatě jednoduchá. Nelíbí se vám tento důkaz plný indexů? Autorovi tohoto textu také ne. V kapitole 5 dokážeme, že násobení matic odpovídá skládání lineárních zobrazení. Z toho dostaneme asociativitu násobení matic zadarmo, skládání libovolných zobrazení je asociativní, což si můžete rozmyslet. □

> Maticové násobení splňuje řadu užitečných vlastností jako asociativita či distributivita, ale většinou nekomutuje. Ostatně maticové násobení odpovídá skládání lineárních zobrazení a skládá zobrazení je vždy asociativní, ale komutuje pouze výjimečně!

Umocňování matic. Čtvercovou matici A můžeme násobit samu sebou, a proto dává smysl zavést *mocninu* A^k:

$$A^k = \underbrace{A \cdot A \cdot A \cdots A \cdot A}_{k\text{-krát}}.$$

Navíc s použitím asociativity můžeme spočítat k-tou mocninu pouze s pomocí $2\log_2 k$ součinů (to platí obecně pro libovolnou operaci, která je asociativní). Jak na to? Všimněte si, že pro výpočet A^k stačí spočítat $A^{k/2}$, které umocníme na druhou. Tedy přesněji platí:

$$A^k = \begin{cases} A^{k/2} \cdot A^{k/2} & \text{pro } k \text{ sudé a} \\ A^{\lfloor k/2 \rfloor} \cdot A^{\lfloor k/2 \rfloor} \cdot A & \text{pro } k \text{ liché.} \end{cases}$$

Tedy pokud chceme spočítat třeba A^9, stačí udělat 4 součiny:

$$A^9 = A^4 \cdot A^4 \cdot A, \qquad A^4 = A^2 \cdot A^2, \qquad A^2 = A \cdot A.$$

To sice nevypadá jako veliká úspora, ale v případě $k = 1000000$ stačí udělat pouze 40 součinů. Logaritmické funkce rostou zatraceně pomalu!

Ukažme si to na příkladu:

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \quad A^2 = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \quad A^4 = \begin{pmatrix} 5 & 3 \\ 3 & 2 \end{pmatrix} \quad \text{a} \quad A^9 = A^4 \cdot A^4 \cdot A = \begin{pmatrix} 55 & 34 \\ 34 & 21 \end{pmatrix}. \tag{3.2}$$

⋆**Fibonacciho čísla.** Posloupnost Fibonacciho čísel se definuje lineární rekurencí:

$$f_0 = 0, \qquad f_1 = 1, \qquad f_{n+2} = f_{n+1} + f_n.$$

Tato posloupnost roste velice rychle a splňuje řadu hezkých vlastností (například souvislosti se zlatým řezem). Několik prvních členů posloupnosti je:

$$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, \ldots$$

Pokud chceme vypočítat n-té Fibonacciho číslo, nepoužijeme k tomu přímo definici (rozmyslete si, že bychom potřebovali exponenciálně mnoho operací). Místo toho postupně spočteme celý počáteční úsek posloupnosti až po n-té číslo. Ze znalosti Fibonacciho čísel 1 až k můžeme pomocí jediného součtu spočítat $(k+1)$-ní Fibonacciho číslo. Potřebujeme tedy pouze lineárně mnoho operací na výpočet n-tého čísla.

Proč zde popisujeme Fibonacciho čísla, která na první pohled nemají s lineární algebrou a maticemi nic společného? Protože na druhý pohled už mají, Fibonacciho čísla jsou totiž definována *lineární* rekurencí. Navíc pokud se podíváme na výše uvedenou matici A z (3.2) a její mocniny, Fibonacciho čísla se v nich objevují překvapivě často.

Uvažme vektor dvou po sobě jdoucích Fibonacciho čísel f_{n+1} a f_n a vynásobme ho zleva maticí A. Dostaneme:

$$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} f_{n+1} \\ f_n \end{pmatrix} = \begin{pmatrix} f_{n+1} + f_n \\ f_{n+1} \end{pmatrix} = \begin{pmatrix} f_{n+2} \\ f_{n+1} \end{pmatrix},$$

tedy vektor obsahující následující dvojici Fibonacciho čísel f_{n+2} a f_{n+1}; každé vynásobení A udělá posun o jedna v posloupnosti Fibonacciho čísel. Pokud tedy chceme nalézt n-té Fibonacciho číslo, stačí vzít vektor (f_0, f_1) a vynásobit ho zleva n-krát maticí A. Protože maticové násobení je asociativní, můžeme nejprve spočítat A^n a poté výslednou mocninou vynásobit vektor (f_0, f_1):

$$A^n \cdot \begin{pmatrix} f_1 \\ f_0 \end{pmatrix} = \begin{pmatrix} f_{n+1} & f_n \\ f_n & f_{n-1} \end{pmatrix} \cdot \begin{pmatrix} f_1 \\ f_0 \end{pmatrix} = \begin{pmatrix} f_{n+1} \\ f_n \end{pmatrix}.$$

Navíc si můžeme všimnout, že A^n je tvořené třemi Fibonacciho čísly f_{n-1}, f_n a f_{n+1}. Tedy například z výše uvedeného výpočtu A^9 jsme zjistili, že $f_8 = 21$, $f_9 = 34$ a $f_{10} = 55$.

Výše jsme popsali, jak můžeme spočítat mocninu A^n pomocí logaritmicky mnoho součinů, jeden součin potřebuje konstantě mnoho operací. Proto umíme spočítat n-té Fibonacciho číslo pomocí logaritmicky mnoha operací (místo původních lineárně mnoha).

Poznamenejme, že z matice A lze zjistit o Fibonacciho číslech mnohem víc. Například lze pomocí *vlastních čísel* a *diagonalizace* ukázat pozoruhodný vzoreček pro obecné Fibonacciho číslo:

$$f_n = \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2} \right)^n - \frac{1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2} \right)^n.$$

Pro tento vzoreček ani není zřejmé, že výsledkem je vždy přirozené číslo, natož Fibonacciho! Zkuste dokázat jeho správnost indukcí.

3.2 Speciální matice

Popišme si některé matice, které se objevují tak často, že mají speciální jména. Mají často jednoduchou strukturu a budou se vyskytovat ve zbytku textu. Také platí, že řada problémů je pro ně mnohem jednodušší. V zápisu matic budeme opět nulové koeficienty vynechávat, pokud je nebudeme chtít zdůraznit.

Nulová matice. Matice, jejíž všechny koeficienty jsou nulové, se nazývá *nulová* a značí se 0. Součin nulové matice s libovolnou jinou maticí (z libovolné strany) je opět nulová matice.

Diagonální matice. Koeficienty $a_{i,i}$ se nazývají *diagonála matice*. Matice je *diagonální*, pokud má všechny koeficienty mimo diagonálu nulové, tedy pro každé i a j, $i \neq j$, platí $a_{i,j} = 0$. Diagonální matice budeme uvažovat typicky čtvercové, pokud explicitně neuvedeme jinak, a značí se například písmenem D. Příklady:

$$D = \begin{pmatrix} 1 & & \\ & 2 & \\ & & 3 \end{pmatrix}, \qquad D' = \begin{pmatrix} 0 & & \\ & 0 & \\ & & 0 \end{pmatrix}, \qquad D'' = \begin{pmatrix} 3 & & \\ & 0 & \\ & & -4 \end{pmatrix}.$$

Jednotková matice. *Jednotková matice* I_n je diagonální matice $n \times n$, která má diagonálu tvořenu jedničkami. Tedy například:

$$I_3 = \begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix}, \qquad I_4 = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix}.$$

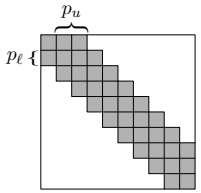
Maticím se říká jednotkové, protože jsou neutrálními prvky pro maticové násobení. Tedy pro libovolnou matici $A \in \mathbb{R}^{m \times n}$ platí $I_m A = A I_n = A$.

Trojúhelníková matice. Trojúhelníková matice neobsahuje nenulové hodnoty nad či pod diagonálou. Pokud neobsahuje nenulové hodnoty pod diagonálou, formálně $a_{i,j} = 0$ pro $i > j$, nazývá se matice *horní trojúhelníková* a typicky se značí U (z anglického *upper*). Pokud neobsahuje nenulové hodnoty nad diagonálou, formálně $a_{i,j} = 0$ pro $i < j$, nazývá se *dolní trojúhelníková* a značí typicky písmenem L (z anglického *lower*). Speciálně každá diagonální matice je současně horní a dolní trojúhelníková matice. Trojúhelníkové matice budeme také typicky uvažovat čtvercové, pokud neuvedeme jinak. Příklady:

$$L = \begin{pmatrix} 1 & & \\ 2 & 1 & \\ 3 & 2 & 1 \end{pmatrix}, \quad L' = \begin{pmatrix} 4 & & & \\ 0 & 0 & & \\ 0 & 1 & 0 & \\ 1 & 0 & 0 & \frac{1}{2} \end{pmatrix}, \quad U = \begin{pmatrix} 1 & 2 & 3 \\ & 2 & 3 \\ & & 3 \end{pmatrix}, \quad U' = \begin{pmatrix} -1 & 1 & -1 & 1 \\ & -1 & 1 & -1 \\ & & -1 & 1 \\ & & & -1 \end{pmatrix}.$$

> Nejdůležitější matice lineární algebry, které se objevují velice často, jsou matice jednotkové, diagonální a trojúhelníkové. Mají jednoduchou strukturu a splňují celou řadu užitečných vlastností.

Pásová matice. Matice se nazývá pásová, pokud se všechny její nenulové koeficienty vyskytují *blízko diagonály*. Co to znamená přesně? Pásová matice je charakterizována dvěma čísly p_ℓ (vzdálenost pod diagonálou) a p_u (vzdálenost nad diagonálou) a platí $a_{i,j} = 0$, pokud $i-j > p_\ell$ či $j-i > p_u$. Pásová matice je naznačena na obrázku 3.3. Například každá diagonální matice je pásová matice pro $p_\ell = p_u = 0$. Pásové matice se často objevují v různých aplikacích (třeba v souvislosti s řešením diferenciálních rovnic) a řada algoritmů pro ně funguje mnohem efektivněji. Například Gaussova eliminace pracuje pouze s koeficienty uvnitř pásu. Rozmyslete si, že během eliminace zůstávají koeficienty mimo pás nulové.



$$\begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & -1 & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{pmatrix}$$

Obrázek 3.3: Nalevo je pásová matice pro parametry p_ℓ a p_u. Napravo je příklad konkrétní pásové matice pro $p_\ell = p_u = 1$.

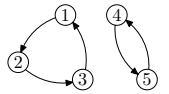
Permutační matice. Permutace množiny M je bijektivní zobrazení $M \to M$. Permutace π množiny $\{1, \ldots, n\}$ prohazuje pořadí čísel $1, \ldots, n$. Permutace π určuje permutační matici P, která je definována takto: V každém sloupečku j je právě jedna jednička na pozici $a_{\pi(j),j}$ a zbytek sloupečku je nulový. Příklad pro $n = 5$ je na obrázku 3.4.

Permutační matice mají řadu hezkých vlastností. Součin dvou permutačních matic je zase permutační matice, která odpovídá permutaci vzniklé složením těchto dvou permutací. Pokud násobíme matici A permutační maticí zleva, prohazujeme řádky matice A (podle permutace), pokud násobíme A zprava, prohazujeme sloupce matice A. Rozmyslete si detaily jako cvičení.

Bloková matice. *Blokové matice* jsou tvořené menšími maticemi, které se nazývají bloky. S blokovými maticemi lze provádět pohodlně většina operací a mají celou řadu hezkých vlastností.

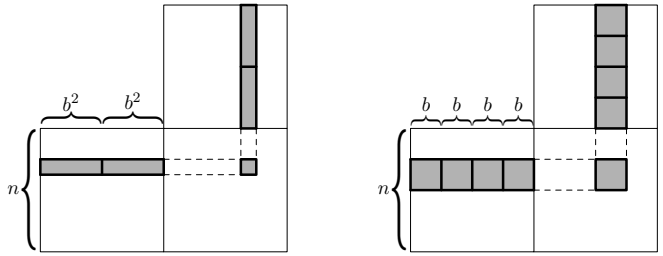
Například uvažme dvě matice A a B tvořené čtyřmi bloky velikosti $n \times n$, tedy matice A a B jsou z $\mathbb{R}^{2n \times 2n}$. Součin AB je následující bloková matice:

$$A = \begin{pmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{pmatrix}, \quad B = \begin{pmatrix} B_{1,1} & B_{1,2} \\ B_{2,1} & B_{2,2} \end{pmatrix}, \quad AB = \begin{pmatrix} A_{1,1}B_{1,1} + A_{1,2}B_{2,1} & A_{1,1}B_{1,2} + A_{1,2}B_{2,2} \\ A_{2,1}B_{1,1} + A_{2,2}B_{2,1} & A_{2,1}B_{1,2} + A_{2,2}B_{2,2} \end{pmatrix}. \quad (3.3)$$



$$P = \begin{pmatrix} & & 1 & & \\ 1 & & & & \\ & 1 & & & \\ & & & & 1 \\ & & & 1 & \end{pmatrix}.$$

Obrázek 3.4: Nalevo je grafické znázornění permutace π, s šipkami $(i, \pi(i))$. Napravo je příslušná permutační matice.



Obrázek 3.5: Nalevo je násobení podle definice. Napravo je blokové násobení, které potřebuje b-krát méně přístupů do paměti RAM. Při blokovém násobení na vypočtení jedné hodnoty AB sice potřebuje b-krát více čtení, ale při každém čtení se počítá b^2 hodnot současně.

Nepřipomínají vám koeficienty klasické maticové násobení?

Podobně mějme blokovou soustavu

$$\begin{pmatrix} A & B \\ & C \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \end{pmatrix},$$

kde všechny matice jsou $n \times n$ a všechny vektory mají n složek. Můžeme ji vyřešit následujícím způsobem. Nejprve vyřešíme spodní polovinu

$$C\boldsymbol{x}_2 = \boldsymbol{b}_2, \quad (3.4)$$

nalezneme všechna řešení $\boldsymbol{x}_2$. Při řešení horní poloviny už známe hodnoty proměnných z $\boldsymbol{x}_2$, které násobí blok B. Můžeme tedy $B\boldsymbol{x}_2$ odečíst od pravé strany a stačí vyřešit soustavu

$$A\boldsymbol{x}_1 = \boldsymbol{b}_1 - B\boldsymbol{x}_2.$$

Tento postup může být výhodný, pokud existuje málo řešení soustavy (3.4).

⋆**Blokové násobení v počítačích.** Blokové matice nejsou užitečné pouze v teorii lineární algebry, používají se i v počítačích. Zkusme si alespoň ve stručnosti naznačit, jak se implementuje násobení matic po blocích. Dnešní architektury počítačů mají *paměťovou hierarchii*, od malých rychlých pamětí až po velké pomalé. Zaměřme se pouze na paměti *cache* a *RAM*. Cache je rychlá paměť blízko procesoru, která má velikost v řádu kB až MB. Čtení z RAM trvá mnohonásobně déle než z cache (můžete si představit třeba stokrát), ale velikost RAM je v řádech GB.

Předpokládejme, že máme obě matice A a B uložené v RAM a chceme spočítat součin AB. Matice se nám celé nevejdou do cache, proto chceme vždy načíst do cache pouze malý kus, spočítat část součinu a zapsat výsledek do RAM. Chtěli bychom minimalizovat počet přístupů do RAM. Pro zjednodušení budeme předpokládat, že obě matice jsou $n \times n$ a že se do cache vždy vejdou buď tři bloky $b \times b$ nebo část

řádku a část sloupce délky b^2. Abychom se nemuseli zabývat s horními celými částmi, předpokládejme, že b^2 dělí n. Situace je naznačena na obrázku 3.5.

Pokud bychom postupovali přesně podle definice násobení, musíme pro každý koeficient matice AB načíst jeden řádek matice A a jeden sloupec matice B, vypočítat hodnotu sumy a zapsat výsledek. To dokážeme udělat na n/b^2 čtení. Tedy celkem budeme potřebovat n^3/b^2 čtení z paměti RAM.

Matice A, B a AB však můžeme rozdělit na bloky $b \times b$ a počítat součin po blocích jako v (3.3). Pokud chceme spočítat hodnotu jednoho bloku AB, musíme postupně vynásobit bloky ve daném řádku A s příslušnými bloky v daném sloupci B. Na vypočítání jednoho bloku AB potřebujeme číst z paměti RAM (n/b)-krát. Protože matice AB obsahuje n^2/b^2 bloků, musí celkově číst (n^3/b^3)-krát. To je b-krát méně než v případě násobení podle definice.

3.3 Inverzní matice a regularita

Jednotková matice I_n funguje jako neutrální prvek na násobení. Položme otázku, zda existují i inverzní prvky. Tedy zda pro matici A existuje nějaká jiná matice A^{-1}, pro kterou by platilo $AA^{-1} = I_n$. Taková matice A^{-1} se nazývá *(pravá) inverze* matice A.

Inverze existují pouze někdy. Například pro nulovou matici inverze neexistuje. Její součin s libovolnou jinou maticí je vždy nulová matice, a tedy nulovou matici není možné invertovat. Existuje celá řada matic, které nemají inverze, jak si hned ukážeme. Například pro libovolnou matici z $\mathbb{R}^{m\times n}$, kde $m > n$, neexistuje inverze. Na druhou stranu když inverze existuje, nemusí být určena jednoznačně.

Výpočet inverze. Označme sloupcové vektory I_n jako $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$. Tedy $\boldsymbol{e}_i$ je n-složkový vektor, jehož všechny složky jsou nulové až na i-tou složku rovnou jedné. Pro matici A hledáme matici A^{-1}, aby platilo $AA^{-1} = I_n$. Uvědomme si, že součin AB lze popsat tak, že uvážíme sloupcové vektory B, každý z nich vynásobíme maticí A a tyto násobky poskládáme vedle sebe do matice AB. Označme $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ sloupcové vektory hledané matice A^{-1}. Pokud má rovnost $AA^{-1} = I_n$ platit, musí platit

$$A\boldsymbol{u}_i = \boldsymbol{e}_i, \qquad \forall i = 1, \ldots, n.$$

Dostáváme tedy n soustav lineárních rovnic, které se liší pouze v pravé straně. Pokud některá z těchto rovnic nemá řešení, inverze neexistuje.

Připomeňme si, jak funguje Gaussova eliminace. Dopředná eliminace převádí matici do odstupňovaného tvaru přičítáním vhodných násobků řádků matice. Povšimněme si, že dopředná eliminace se provádí pouze podle koeficientů matice, nezávisle na hodnotách pravé strany. Pravá strana se v průběhu dopředné eliminace modifikuje a nakonec se využívá ve zpětné substituci k dopočtení množiny všech řešení.

V případě inverze chceme vyřešit soustavy s maticí A pro n různých pravých stran. Na každou z nich můžeme aplikovat úplně stejnou dopřednou eliminaci, která je promění v

$$C\boldsymbol{u}_i = \boldsymbol{b}_i, \qquad \forall i = 1, \ldots, n,$$

kde C je matice v odstupňovaném tvaru a $\boldsymbol{b}_i$ je pravá strana, která vznikla z $\boldsymbol{e}_i$ během dopředné eliminace. Protože se všechny pravé strany upravují v průběhu dopředné eliminace stejným způsobem, zapisují se běžně všechny pravé strany vedle sebe a provádí se eliminace naráz.

Ukažme jako příklad vypočtení inverze matice 2×2:

$$A = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}, \qquad \left(\begin{array}{cc|cc} 2 & -1 & 1 & 0 \\ -1 & 2 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{cc|cc} 1 & -\frac{1}{2} & \frac{1}{2} & 0 \\ -1 & 2 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{cc|cc} 1 & -\frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{3}{2} & \frac{1}{2} & 1 \end{array}\right) \sim \cdots$$

$$\cdots \sim \left(\begin{array}{cc|cc} 1 & -\frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 1 & \frac{1}{3} & \frac{2}{3} \end{array}\right) \sim \left(\begin{array}{cc|cc} 1 & 0 & \frac{2}{3} & \frac{1}{3} \\ 0 & 1 & \frac{1}{3} & \frac{2}{3} \end{array}\right), \qquad A^{-1} = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}.$$

Soustavu jsme upravili dokonce do Gaussova-Jordanova tvaru a tím jsme na pravé straně přímo dostali hledanou inverzní matici. V tomto případě je inverze jednoznačně určena.

Existence a jednoznačnost. Pokusme se zjistit, za jakých podmínek obecně existuje inverzní matice. Nejprve učiňme jednoduché pozorování:

Pozorování 3.3. *Soustava $A\boldsymbol{x} = \boldsymbol{b}$ má řešení pro libovolnou pravou stranu $\boldsymbol{b}$, právě když má řešení pro všechny pravé strany $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$.*

Důkaz. Jedna implikace je triviální: Pokud má soustava řešení pro každé $\boldsymbol{b}$, má řešení i pro $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$.

Při druhé implikaci existují pro pravé strany $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ nějaká řešení $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$. Uvažme vektor

$$\boldsymbol{x} = b_1\boldsymbol{x}_1 + b_2\boldsymbol{x}_2 + \cdots + b_n\boldsymbol{x}_n.$$

Ten je řešením soustavy $A\boldsymbol{x} = \boldsymbol{b}$, protože

$$A\boldsymbol{x} = A(b_1\boldsymbol{x}_1 + b_2\boldsymbol{x}_2 + \cdots + b_n\boldsymbol{x}_n) = b_1A\boldsymbol{x}_1 + b_2A\boldsymbol{x}_2 + \cdots + b_nA\boldsymbol{x}_n = b_1\boldsymbol{e}_1 + b_2\boldsymbol{e}_2 + \cdots + b_n\boldsymbol{e}_n = \boldsymbol{b}. \quad \square$$

Toto pozorování mimo jiné říká:

Tvrzení 3.4. *Pro matici A existuje pravá inverze právě tehdy, když má soustava $A\boldsymbol{x} = \boldsymbol{b}$ řešení pro libovolnou pravou stranu $\boldsymbol{b}$.* $\square$

A to lze snadno ověřit v odstupňovaném tvaru matice A:

Tvrzení 3.5. *Soustava $A\boldsymbol{x} = \boldsymbol{b}$ má řešení pro libovolnou pravou stranu $\boldsymbol{b}$, právě když v odstupňovaném tvaru neobsahuje nulový řádek.*

Důkaz. Nechť U je odstupňovaný tvar A. Pokud U nulový řádek neobsahuje, nemůže zpětná substituce selhat a vždy zkonstruuje řešení.

Naopak pokud U nulový řádek obsahuje, potom neexistuje řešení $U\boldsymbol{x} = \boldsymbol{b}'$ pro libovolný vektor $\boldsymbol{b}'$ s nenulovou m-tou složkou (třeba $\boldsymbol{b} = \boldsymbol{e}_m$). Pokud invertujeme kroky dopředné eliminace (což lze udělat, viz důkaz tvrzení 1.1), upravíme pravou stranu $\boldsymbol{b}'$ na pravou stranu $\boldsymbol{b}$. Protože množina řešení $U\boldsymbol{x} = \boldsymbol{b}'$ a $A\boldsymbol{x} = \boldsymbol{b}$ je stejná, nemá ani soustava $A\boldsymbol{x} = \boldsymbol{b}$ žádné řešení. $\square$

Pokud má matice A více řádků než sloupců, obsahuje její odstupňovaný tvar vždy nulový řádek. Tedy pro matici $A \in \mathbb{R}^{m\times n}$, kde $m > n$, neexistuje pravá inverze.

Uvažme nyní čtvercovou matici. Pokud nemá v odstupňovaném tvaru nulový řádek, je odstupňovaný tvar horní trojúhelníková matice s *nenulovou* diagonálou. Odstupňovaný tvar neobsahuje žádnou volnou proměnnou a řešení je pro každou pravou stranu určené jednoznačně. Tedy:

Tvrzení 3.6. *Pokud pro čtvercovou matici existuje pravá inverze, je určená jednoznačně.*

Čtenář si může rozmyslet, že naopak pokud pro nečtvercovou matici existuje inverze, není nikdy určena jednoznačně. Dokonce není ani libovolný sloupec inverze určen jednoznačně.

Matice A má pravou inverzi A^{-1} právě tehdy, pokud soustava $A\boldsymbol{x} = \boldsymbol{b}$ má řešení pro libovolnou pravou stranu $\boldsymbol{b}$. A to platí tehdy, když matice A nemá v odstupňovaném tvaru žádný nulový řádek. Jednotlivé sloupečky inverze jsou řešení $A\boldsymbol{u}_i = \boldsymbol{e}_i$; typicky se počítá eliminací pro n pravých stran současně. Pokud čtvercová matice má inverzi, je určena jednoznačně.

Levá a oboustranná inverze. Čtenář si možná všiml, že inverzi uvažujeme zprava a explicitně ji nazýváme pravá inverze. Podobně můžeme zadefinovat *levou inverzi* A^{-1} matice A jako matici, pro kterou platí $A^{-1}A = I_n$. Mezi levou a pravou inverzí platí následující vztah:

$$A^{-1}A = I_n = I_n^\mathsf{T} = (A^{-1}A)^\mathsf{T} = A^\mathsf{T}(A^{-1})^T, \tag{3.5}$$

tedy pravá inverze matice A^T je transpozice levé inverze matice A.

Pochopitelně aby levá inverze vůbec mohla existovat, musí pro $A \in \mathbb{R}^{m\times n}$ platit $m \leq n$. Tedy pokud není matice čtvercová, nemůže mít současně levou i pravou inverzi. Jak je to pro čtvercové matice? Platí následující překvapivý fakt, který je jeden z divů lineární algebry a který si dokážeme na závěr kapitoly:

Věta 3.7. *Nechť čtvercová matice A má nějakou inverzi A^{-1} (levou či pravou). Potom je A^{-1} současně levá i pravá inverze, tedy*

$$A^{-1}A = AA^{-1} = I_n.$$

Navíc je inverze A určena jednoznačně jako A^{-1}.

Pro čtvercovou matice A tedy nemusíme rozlišovat mezi levou a pravou inverzí, matice A^{-1} bude prostě *inverzní matice* nebo *inverze*.

Invertovatelné čtvercové matice (ty, které mají inverzi) jsou natolik důležité, že dostaly speciální jméno *regulární*. Slovo regulární jsme používali již v kapitole 1 pro regulární úpravy, které nemění množinu řešení soustavy. Toto není náhoda, jak si brzo ukážeme, každá regulární úprava odpovídá nějaké regulární matici. Čtvercová matice, která nemá inverzi (a tedy není regulární), se nazývá *singulární*. Pokusme se vysvětlit, proč je výše uvedená věta tolik překvapivá. Podle (3.5) lze znění věty přepsat pomocí transpozice. Věta říká, že:

1. Čtvercová matice A je zprava invertovatelná, právě když je matice A^T zprava invertovatelná.
2. Navíc inverze a transpozice pro čtvercové matice komutují:
$$(A^\mathsf{T})^{-1} = (A^{-1})^\mathsf{T}.$$
3. Pro čtvercové matice A a A^T jsou inverze určené jednoznačně, což už jsme dokázali výše.

Pokud je matice A symetrická, je věta triviální, což si můžete rozmyslet. Pokud však není symetrická, jsou $A\boldsymbol{x} = \boldsymbol{b}$ a $A^\mathsf{T}\boldsymbol{x} = \boldsymbol{b}$ naprosto odlišné soustavy!

První část říká, že soustava $A\boldsymbol{x} = \boldsymbol{b}$ má řešení pro každou pravou stranu $\boldsymbol{b}$, právě když má soustava $A^\mathsf{T}\boldsymbol{x} = \boldsymbol{b}$ má řešení pro každou pravou stranu $\boldsymbol{b}$. To je překvapivé, ale jak uvidíme v kapitole 4, platí zobecnění, že soustavy $A\boldsymbol{x} = \boldsymbol{b}$ a $A^\mathsf{T}\boldsymbol{x} = \boldsymbol{b}$ mají řešení pro „stejně mnoho“ pravých stran $\boldsymbol{b}$ (i v případě libovolných obdélníkových matic).

Pokud první část není dostatečně překvapivá, druhá rozhodně je. Uvědomme si, že řešení soustav $A\boldsymbol{x} = \boldsymbol{b}$ a $A^\mathsf{T}\boldsymbol{x} = \boldsymbol{b}$ pro jedno pevné $\boldsymbol{b}$ jsou zcela odlišná. Dokažme si alespoň tuto druhou část krasným algebraickým trikem:

Lemma 3.8. *Nechť X je levá inverze A a Y je pravá inverze A, tedy $XA = I_n$ a $AY = I_n$. Potom $X = Y$.*

Důkaz. Důkaz provedeme trikovou úpravou využívající asociativitu násobení matic:

$$X = XI_n = X(AY) = (XA)Y = I_nY = Y. \qquad \square$$

První část věty dokážeme na konci této kapitoly s využitím LU dekompozice.

Pokud čtvercová matice má libovolnou inverzi, má současně levou i pravou inverzi. Navíc jsou tyto inverze shodné, tedy pro čtvercovou matici existuje jednoznačně určená oboustranná inverze. Invertovatelné čtvercové matice se nazývají regulární, neinvertovatelné se nazývají singulární.

Vlastnosti inverze. Inverze matic mají řadu dalších pěkných vlastností, z nichž některé si zmíníme. Jejich správnost spíše naznačíme a detaily si může čtenář sám ověřit. Uvažujme nyní pouze čtvercové matice, i když podobná tvrzení platí i pro nečtvercové matice s jednostrannými inverzemi; klidně je zkuste zformulovat a dokázat.

- *Inverze a součin.* Nechť A a B jsou regulární matice. Potom i matice AB a BA jsou regulární, navíc platí $(AB)^{-1} = B^{-1}A^{-1}$ a $(BA)^{-1} = A^{-1}B^{-1}$. Pamatujete si na prohození pořadí u transpozice součinu, pro inverze to funguje úplně stejně.
- *Regularita inverze.* Pokud je A regulární matice, je také A^{-1} regulární matice, navíc její inverze je A. Tedy A a A^{-1} jsou svoje vzájemné inverze. To vyplyne z důkazu věty 3.7.
- *Inverze a transpozice.* Pro připomenutí, pokud je matice A regulární, je i A^T regulární a navíc inverze a transpozice komutují: $(A^T)^{-1} = (A^{-1})^T$. Proto se často inverze transpozice značí $A^{-\mathsf{T}}$.
- *Inverze a mocnina.* Nechť A je regulární matice. Potom také A^n je regulární matice pro libovolné n. Navíc platí $(A^n)^{-1} = (A^{-1})^n$. Tedy inverze a mocnina opět komutuje a často se značí jako A^{-n}. Pro toto značení například platí, že $A^iA^j = A^{i+j}$ pro libovolná celá čísla i a j.
- *Inverze trojúhelníkových matic.* Inverze regulární horní trojúhelníkové matice je horní trojúhelníková matice. Podobně inverze regulární dolní trojúhelníkové matice je dolní trojúhelníková matice. Rozmyslete si také, jak vypadá inverze diagonální matice.
- *Inverze a součet.* Mezi inverzí a součtem není *obecně žádný vztah*, součet se chová k inverzi hezky pouze výjimečně! Rozhodně obecně neplatí $(A+B)^{-1} = A^{-1}+B^{-1}$. Jako protipříklad stačí uvážit $B = -A$, neboť pro $A + B = 0$ žádná inverze neexistuje. Obecně to ale neplatí ani tehdy, kdyby inverze $A + B$ existovala! Taková vlastnost totiž neplatí ani pro reálná čísla (na která lze nahlížet jako na matice 1×1). Například $(1 + 2)^{-1} = \frac{1}{3}$, ale $1^{-1} + 2^{-1} = 1 + \frac{1}{2} = \frac{3}{2}$.

3.4 Soustavy v řeči matic

Již jsme ukázali, že soustava má v řeči matic elegantní zápis

$$A\boldsymbol{x} = \boldsymbol{b}$$

a viděli jsme řadu výhod tohoto zápisu. Naším cílem bude popsat elementární úpravy soustavy v řeči násobení matic.

Množina všech řešení soustavy. V kapitole 2 jsme dokázali, že množina všech řešení soustavy s nulovou pravou stranou tvoří vektorový podprostor (tvrzení 2.3) a že množina všech řešení soustavy s obecnou pravou stranou tvoří afinní podprostor (tvrzení 2.4). Ukažme si tyto důkazy znovu v elegantnějším maticovém zápisu.

Tvrzení 3.9. *Množina všech řešení soustavy $A\boldsymbol{x} = \boldsymbol{0}$ tvoří vektorový podprostor prostoru $\mathbb{R}^n$.*

Důkaz. Potřebujeme ukázat, že je množina všech řešení uzavřená na násobení skalárem a sčítání. Nechť $\boldsymbol{x}$ a $\boldsymbol{y}$ jsou libovolná dvě řešení (tedy $A\boldsymbol{x} = \boldsymbol{0}$ a $A\boldsymbol{y} = \boldsymbol{0}$) a nechť α je libovolné reálné číslo. Budeme využívat vlastností maticového násobení. Platí

$$A(\alpha\boldsymbol{x}) = \alpha(A\boldsymbol{x}) = \alpha\boldsymbol{0} = \boldsymbol{0} \qquad \text{a} \qquad A(\boldsymbol{x}+\boldsymbol{y}) = A\boldsymbol{x} + A\boldsymbol{y} = \boldsymbol{0}+\boldsymbol{0} = \boldsymbol{0}. \qquad \square$$

Tento vektorový podprostor je natolik důležitý, že získal speciální jméno *jádro matice* A a značí se $\mathrm{Ker}(A)$. Jádro je jeden ze čtyř fundamentální podprostorů definovaných maticí, které si popíšeme v kapitole 4.

Tvrzení 3.10. *Nechť $\boldsymbol{p}$ je libovolné řešení soustavy $A\boldsymbol{x} = \boldsymbol{b}$. Množina všech řešení této soustavy tvoří afinní podprostor* $\mathrm{Ker}(A) + \boldsymbol{p}$.

Důkaz. Potřebujeme dokázat dvě implikace:

- $\boldsymbol{x} \in \mathrm{Ker}(A) + \boldsymbol{p} \implies A\boldsymbol{x} = \boldsymbol{b}$: Platí $\boldsymbol{x} = \boldsymbol{x}_0 + \boldsymbol{p}$, kde $\boldsymbol{x}_0 \in \mathrm{Ker}(A)$. Z toho vyplývá:
$$A\boldsymbol{x} = A(\boldsymbol{x}_0 + \boldsymbol{p}) = A\boldsymbol{x}_0 + A\boldsymbol{p} = \boldsymbol{0} + \boldsymbol{b} = \boldsymbol{b}.$$
- $A\boldsymbol{x} = \boldsymbol{b} \implies \boldsymbol{x} \in \mathrm{Ker}(A) + \boldsymbol{p}$: Využijeme, že $\boldsymbol{x} \in \mathrm{Ker}(A) + \boldsymbol{p}$, právě když $\boldsymbol{x} - \boldsymbol{p} \in \mathrm{Ker}(A)$. Což platí, protože
$$A(\boldsymbol{x} - \boldsymbol{p}) = A\boldsymbol{x} - A\boldsymbol{p} = \boldsymbol{b} - \boldsymbol{b} = \boldsymbol{0}. \qquad \square$$

Maticové rovnice. S příklady maticových rovnic jsme se už setkali, například soustava lineárních rovic v maticovém zápisu, $A\boldsymbol{x} = \boldsymbol{b}$, je maticová rovnice. Budeme však uvažovat úplně obecné maticové rovnice, například

$$X + I_n = YZ, \tag{3.6}$$

kde X, Y a Z jsou neznámé matice. Řešením jsou všechny trojice matic (X, Y, Z), které splňují tuto rovnici; například jedna taková trojice je $(I_n, I_n, 2I_n)$.

Nyní vynásobme zleva rovnici libovolnou (kompatibilní) maticí A, dostaneme rovnici

$$A(X + I_n) = AYZ. \tag{3.7}$$

Pokud nějaká trojice (X, Y, Z) řeší rovnici (3.6), dostaneme po dosazení shodnou levou a pravou stranu. Proto trojice (X, Y, Z) řeší i vynásobenou rovnici (3.7), po vynásobení A opět platí rovnost levé a pravé strany. Na druhou stranu rozhodně neplatí, že by se množina všech řešení nemohla vynásobením zvětšit. Například pokud by A byla nulová matice, platila by rovnost (3.7) pro libovolnou trojici (X, Y, Z). Množina řešení však může po vynásobení A pouze vzrůst.

Budeme tedy zkoumat, co musí platit pro matici A, aby se množina řešení obecně (nezávisle na podobě maticové rovnice) nezměnila. Například co kdyby matice A byla zleva invertovatelná? Potom můžeme rovnici (3.7) vynásobit zleva A^{-1} a s využitím asociativity dostaneme zpět rovnici (3.6). S každým vynásobením může množina řešení pouze vzrůst a nakonec je zpět stejná jako na začátku; tedy nemohla se vynásobením změnit. Násobení zleva maticí, která je zleva invertovatelná, nemění množinu řešení.

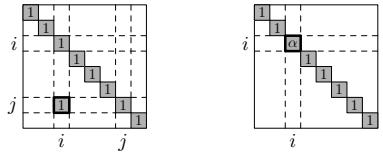
Podobně to platí při násobení zprava maticí, která je zprava invertovatelná, případně pro regulární matice z obou stran. Poznamenejme, že i když matice A není invertovatelná, můžeme vynásobením nezměnit množinu řešení, v *obecnosti* však nic takového neplatí.

Regulární a elementární úpravy. Nepřipomíná vám předchozí popis maticových rovnic úpravy soustav z kapitoly 1? Přesně stejně platilo, že úpravy mohou množinu řešení pouze zvětšit a regulární úpravy ji nemění. Navíc jsme to dokazovali tak, že jsme regulární úpravy invertovali. Nyní si ukážeme, že podobnost není náhodná, regulární úpravy odpovídají regulárním maticím.

Budeme uvažovat úpravy, které fungují v obecnosti pro libovolnou soustavu a budeme uvažovat pouze čtvercové regulární matice. Každá úprava bude odpovídat nějaké matici R, kdy upravíme soustavu $A\boldsymbol{x} = \boldsymbol{b}$ v soustavu $RA\boldsymbol{x} = R\boldsymbol{b}$. Již jsme si dokázali, že pro regulární matice se množina řešení soustavy nezmění.

Každá regulární úprava je tvořena konečně mnoha aplikacemi elementárních úprav, které jsou přičtení jednoho řádku k druhému a vynásobení jednoho řádku.[2] Pokud ukážeme, že tyto úpravy lze reprezentovat regulární maticí, máme ekvivalenci regulárních úprav a regulárních matic dokázánu. A nalézt tyto matice není nic obtížného.

Pro přičtení i-tého řádku k j-tému stačí uvážit jednotkovou matici I_n s pozměněným koeficientem $a_{j,i} = 1$. Pro vynásobení i-tého řádku koeficientem α stačí opět uvážit jednotkovou matici, kde pozměníme koeficient $a_{i,i} = \alpha$. Graficky jsou tyto matice naznačeny na obrázku 3.6. Čtenář může sám ověřit, že tyto matice fungují a nalézt po vzoru důkazu tvrzení 1.1 jejich inverze. Také může zkusit nalézt matice reprezentující rozšiřující elementární úpravy přičtení násobku jednoho řádku k druhému a prohození dvou řádků.



Obrázek 3.6: Nalevo je matice elementární úpravy přičtení i-tého řádku k j-tému řádku, napravo matice úpravy vynásobení i-tého řádku koeficientem α.

> Regulární úpravy odpovídají regulárním maticím, jimiž násobíme soustavu zleva. Pokud je totiž matice zleva invertovatelná, nemění násobením množinu řešení. Pokud matice není invertovatelná, může se množina řešení pouze zvětšit.

Úprava inverzní maticí. Nechť A je regulární matice. Přirozená otázka je, jaká matice R zjednoduší soustavu $A\boldsymbol{x} = \boldsymbol{b}$ co nejvíce, ideálně ji převede do Gaussova-Jordanova tvaru. Stačí si uvědomit, že Gaussův-Jordanův tvar regulární matice je jednotková matice. Hledáme tedy regulární matici R, která splňuje $RA = I_n$. To platí pro inverzní matice A^{-1}.

Pokud tedy známe pro matici A její inverzní matici A^{-1}, stačí soustavu $A\boldsymbol{x} = \boldsymbol{b}$ vynásobit A^{-1} a dostaneme

$$A^{-1}A\boldsymbol{x} = A^{-1}\boldsymbol{b}, \qquad \text{neboli} \qquad \boldsymbol{x} = A^{-1}\boldsymbol{b}. \tag{3.8}$$

Místo řešení soustavy stačí vynásobit vektor $\boldsymbol{b}$ maticí A^{-1}, což je mnohem jednodušší a efektivnější. Samozřejmě to má háček, že inverzní matici A^{-1} většinou neznáme a její nalezení je ještě obtížnější než vyřešení samotné soustavy. Mělo by smysl spočítat inverzi pouze tehdy, pokud bychom potřebovali vyřešit stejnou soustavu pro hodně pravých stran. V praxi se spíše používá LU dekompozice, která je vedlejším produktem Gaussovy eliminace a kterou si nyní popíšeme. Na druhou stranu vztah (3.8) je velice užitečný teoreticky.

[2] To jsme nedokázali a budeme tomu věřit. Abychom to mohli dokázat, museli bychom přesně zadefinovat, co považujeme za regulární úpravu a co ne. Definice z kapitoly 1 nebyla příliš formální.

3.5 LU dekompozice

Nyní si popíšeme jednu ze základních maticových dekompozic zvanou LU. Tato dekompozice má řadu hezkých teoretických důsledků, jak brzo uvidíme. Navíc se i prakticky používá, při řešení soustavy lineárních rovnic se často nejprve spočte LU dekompozice a pomocí LU dekompozice se soustava snadno vyřeší. Ostatně žebříček nejrychlejších superpočítačů planety Top500 srovnává stroje pomocí toho, jak rychle umějí spočítat LU dekompozici obrovské matice.

LU je první z několika slavných dekompozic, které v tomto textu popíšeme. Další slavné dekompozice jsou QR dekompozice, spektrální dekompozice a SVD dekompozice. Dekompozice tvoří jeden z fundamentálních kamenů numerické lineární algebry, mají spoustu úžasných využití a je známa celá řada algoritmů na jejich nalezení.

Maticové dekompozice. Pokud řešíme nějaký problém s maticí A (třeba soustavu $A\boldsymbol{x} = \boldsymbol{b}$), může nám pomoci vhodná *maticová dekompozice*. Idea je rozložit složitou matici A na součet či součin jednodušších matic. Pro tyto jednoduché matice bude snadné problém vyřešit a při troše štěstí se z těchto řešení podaří složit řešení problému pro původní matici A.

Nejjednodušší dekompozice jsou tvořeny součtem několika matic. Například $A = D + E$, kde D je diagonální část matice A a E je část mimo diagonálu, formálně:

$$d_{i,j} = \begin{cases} a_{i,j} & \text{pro } i = j, \\ 0 & \text{jinak,} \end{cases} \qquad \text{a} \qquad e_{i,j} = \begin{cases} a_{i,j} & \text{pro } i \neq j, \\ 0 & \text{jinak.} \end{cases}$$

Tyto součtové dekompozice však nejsou moc vhodné pro vyřešení soustavy $A\boldsymbol{x} = \boldsymbol{b}$, i kdybychom uměli rychle vyřešit soustavy $D\boldsymbol{y} = \boldsymbol{b}$ a $E\boldsymbol{z} = \boldsymbol{b}$. Problém je, že $\boldsymbol{x}$ není obecně v žádném vztahu k vektorům $\boldsymbol{y}$ a $\boldsymbol{z}$, například rozhodně neplatí $\boldsymbol{x} = \boldsymbol{y} + \boldsymbol{z}$. Jinými slovy s využitím (3.8) můžeme říct, že matice $A^{-1} = (D + E)^{-1}$ a $D^{-1} + E^{-1}$ nejsou obecně v žádném vztahu, mohou se libovolně lišit. Na druhou stranu, součtové dekompozice mají využití v jiných oblastech lineární algebry.[3]

Naproti tomu dekompozice A na součin dvou (či více) jednoduchých matic MN jsou pro vyřešení soustavy $A\boldsymbol{x} = \boldsymbol{b}$ mnohem užitečnější. Pokud by totiž byly soustavy s maticemi M a N snadno vyřešitelné, můžeme $A\boldsymbol{x} = (MN)\boldsymbol{x} = \boldsymbol{b}$ vyřešit ve dvou krocích

$$M\boldsymbol{y} = \boldsymbol{b} \qquad \text{a} \qquad N\boldsymbol{x} = \boldsymbol{y}.$$

LU je příklad takové dekompozice.

LU dekompozice. Budeme uvažovat matici A, která je regulární. Navíc předpokládejme, že lze matici A převést do odstupňovaného tvaru bez prohazování řádků; pokud přičítáme i-tý řádek k j-tému, vždy platí $i < j$. Tato podmínka není jakkoliv zásadní a vzápětí LU dekompozici zobecníme na obecné matice.

LU dekompozice matice A je, jak už název napovídá, rozklad A na součin dvou matic L a U, tedy $A = LU$. Matice L je dolní trojúhelníková s jednotkovou diagonálou a matice U je horní trojúhelníková s obecnou diagonálou.

LU dekompozici A získáme aplikováním dopředné eliminace na A; matice U je odstupňovaný (Gaussův) tvar A. Matice L popisuje průběh dopředné eliminace. Dopředná eliminace je aplikace série elementárních úprav na matici A a každá elementární úprava odpovídá nějaké regulární matici R. Tedy platí

$$R_k R_{k-1} \cdots R_2 R_1 A = U,$$

[3] Například na této konkrétní dekompozici $A = D + E$ je založena Jacobiho metoda, numerická metoda pro řešení soustavy $A\boldsymbol{x} = \boldsymbol{b}$, která funguje dobře pro matice s dominantními hodnotami na diagonále. Více si o numerických metodách povíme v kapitole **??**.

kde k je počet elementárních úprav provedených dopřednou eliminací. Proto platí $A = LU$ pro

$$L = (R_k R_{k-1} \cdots R_2 R_1)^{-1} = R_1^{-1} R_2^{-1} \cdots R_{k-1}^{-1} R_k^{-1}. \tag{3.9}$$

Zbývá dokázat, že matice L bude opravdu dolní trojúhelníková.

Vlastnosti trojúhelníkových matic. Odvodíme si tři jednoduchá lemmata, která popíšou základní vlastnosti trojúhelníkových matic. Učiňme užitečné pozorování, že transpozice převádí dolní trojúhelníkové matice na horní trojúhelníkové matice a naopak.

Lemma 3.11. *Součin horních (resp. dolních) trojúhelníkových matic je horní (resp. dolní) trojúhelníková matice, tedy*

$$U_1 U_2 = U \qquad a \qquad L_1 L_2 = L.$$

Důkaz. Dokážeme pouze pro horní trojúhelníkové matice, pro dolní se to dokáže transponováním na horní (čtenář si může rozmyslet detaily). Potřebujeme ukázat, že $(U)_{i,j} = 0$, kdykoliv $i > j$. Rozepišme tedy koeficient $(U)_{i,j}$ podle definice součinu matic:

$$(U)_{i,j} = \sum_{k=1}^{n} (U_1)_{i,k} (U_2)_{k,j}.$$

Ukážeme, že každý z členů sumy je nulový, neboť vždy $(U_1)_{i,k} = 0$ nebo $(U_2)_{k,j} = 0$. Protože U_1 a U_2 jsou horní trojúhelníkové matice, platí $(U_1)_{i,k} = 0$ pro $i > k$ a $(U_2)_{k,j} = 0$ pro $k > j$. Aby byl k-tý člen sumy nenulový, muselo by platit $i \leq k$ a zároveň $k \leq j$. Takové k ale neexistuje, protože $(U)_{i,j}$ je pod diagonálou, a tedy $i > j$. □

Ukázali jsme si řadů nástrojů, jak pracovat s pravými inverzemi; například tvrzení 3.4 dává nutnou a postačující podmínku existence. S levými inverzemi, o kterých toho zatím moc nevíme, se vypořádáme pro trojúhelníkové matice pomocí transpozice.

Například uvažme dolní trojúhelníkovou matici L a ptáme se, jestli má levou inverzi L^{-1}. Pokud uvážíme horní trojúhelníkovou matici $L^\intercal = U$ a její pravou inverzi U^{-1}, platí

$$(U^{-1})^\intercal = L^{-1}, \qquad \text{neboť} \qquad I_n = I_n^\intercal = (UU^{-1})^\intercal = (U^{-1})^\intercal L.$$

Speciálně levá inverze L^{-1} existuje, právě když existuje pravá inverze U^{-1}.

Lemma 3.12. *Trojúhelníková matice má inverzi (levou i pravou) právě tehdy, když má nenulovou diagonálu.*

Důkaz. Pokud má trojúhelníková matice nenulovou diagonálu, má v odstupňovaném tvaru všechny řádky nenulové (v případě U se přímo nachází v odstupňovaném tvaru, v případě L vynuluje Gaussova eliminace hodnoty pod diagonálou). Proto podle tvrzení 3.5 a tvrzení 3.4 existuje pravá inverze. S využitím transpozice získáme i existenci levé inverze.

Naopak pokud má trojúhelníková matice diagonálu někde nulovou, bude v odstupňovaném tvaru obsahovat sloupeček bez pivota, a tedy bude mít nulový řádek. Proto dostaneme neexistenci pravé inverze a s pomocí transpozice i neexistenci levé inverze. □

Poznamenejme, že výše uvedené lemma je vlastně speciální případ věty 3.7 pro trojúhelníkové matice. Navíc podle lemmatu 3.8 víme, že se levé a pravé inverze shodují.

Lemma 3.13. *Inverze horní (resp. dolní) trojúhelníkové matice je horní (resp. dolní) trojúhelníková matice.*

Důkaz. Dokažme to opět pouze pro horní trojúhelníkovou matici. Inverze se počítá tak, že soustavu $(U|I_n)$ převedeme úpravami v $(I_n|U^{-1})$. Ale protože upravujeme koeficienty U pouze v horním trojúhelníku, všechny tyto úpravy modifikují I_n také pouze v horním trojúhelníku, tedy U^{-1} je horní trojúhelníková matice.

Pro dolní trojúhelníkovou matici se to dokáže transpozicí na horní trojúhelníkovou matici. □

Existence a jednoznačnost LU. Dokažme nyní existenci a jednoznačnost LU dekompozice:

Tvrzení 3.14. *Pokud A je regulární matice, kterou lze odstupňovat bez prohazování řádků, potom existuje LU dekompozice a je určena jednoznačně.*

Důkaz. Protože matici lze odstupňovat bez prohazování řádků, jsou všechny matice elementárních úprav dolní trojúhelníkové matice, viz obrázek 3.6. Podle lemmat 3.11 a 3.13 je tedy i $L = R_1^{-1}R_2^{-1}\cdots R_{k-1}^{-1}R_k^{-1}$ dolní trojúhelníková matice. Spolu s odstupňovaným tvarem U dostáváme LU dekompozici $A = LU$.

Kdyby existovali dvě různé LU dekompozice, platilo by $A = L_1U_1 = L_2U_2$. Podle definice LU dekompozice mají všechny tyto matice nenulové diagonály; tedy podle lemmatu 3.12 existují jejich inverzní matice a platí

$$L_2^{-1}L_1 = U_2U_1^{-1}.$$

S využitím lemmat 3.11 a 3.13 víme, že $L_2^{-1}L_1$ je dolní trojúhelníková matice s jednotkovou diagonálou a $U_2U_1^{-1}$ je horní trojúhelníková matice.

Pokud platí rovnost mezi dolní a horní trojúhelníkovou maticí, musí to být matice diagonální. Navíc musí mít jednotkovou diagonálu, tedy je to jednotková matice I_n. Avšak potom $L_2^{-1}L_1 = I_n$ a L_2^{-1} je levá inverze L_1; L_2^{-1} je inverzí jak L_1, tak L_2. Naopak však platí, že L_1 i L_2 jsou obě inverzí matice L_2^{-1}. Protože všechny matice jsou čtvercové, platí podle tvrzení 3.6 o jednoznačnosti inverze, že $L_1 = L_2$. Podobně platí $U_1 = U_2$. □

Často se pro regulární matice (spíše pro přehlednost) uvažuje LDU dekompozice, kde matice D je diagonální matice obsahující pivoty a obě trojúhelníkové matice L a U mají jednotkovou diagonálu.

> LU dekompozice matice je rozklad $A = LU$, kde L je dolní trojúhelníková matice s jednotkovou diagonálou a U je horní trojúhelníková matice. Matice U je odstupňovaný tvar matice A a matice L vznikne složením všech úprav potřebných k odstupňování A. Aby LU dekompozice existovala, musí být možné odstupňovat A bez prohazování řádků. Pokud je A regulární, je LU dekompozice určena jednoznačně. LDU dekompozice je rozklad $A = LDU$, kde obě matice L a U mají jednotkovou diagonálu a pivoty jsou umístěny na diagonálu D.

Příklad. Ukážeme si příklad, jak vypadá LU dekompozice matice

$$A = \begin{pmatrix} 2 & -1 & \\ -1 & 2 & -1 \\ & -1 & 2 \end{pmatrix}.$$

První úprava R_1 je přičtení $\frac{1}{2}$-násobku prvního řádku k druhému. Dostaneme:

$$R_1 = \begin{pmatrix} 1 & & \\ \frac{1}{2} & 1 & \\ & & 1 \end{pmatrix}, \qquad \begin{pmatrix} 1 & & \\ \frac{1}{2} & 1 & \\ & & 1 \end{pmatrix}\begin{pmatrix} 2 & -1 & \\ -1 & 2 & -1 \\ & -1 & 2 \end{pmatrix} = \begin{pmatrix} 2 & -1 & \\ 0 & \frac{3}{2} & -1 \\ & -1 & 2 \end{pmatrix}$$

Druhá úprava R_2 je přičtení $\frac{2}{3}$-násobku druhého řádku k třetímu:

$$R_2 = \begin{pmatrix} 1 & & \\ & 1 & \\ & \frac{2}{3} & 1 \end{pmatrix}, \qquad \begin{pmatrix} 1 & & \\ & 1 & \\ & \frac{2}{3} & 1 \end{pmatrix}\begin{pmatrix} 1 & & \\ \frac{1}{2} & 1 & \\ & & 1 \end{pmatrix}\begin{pmatrix} 2 & -1 & \\ -1 & 2 & -1 \\ & -1 & 2 \end{pmatrix} = \begin{pmatrix} 2 & -1 & \\ 0 & \frac{3}{2} & -1 \\ & 0 & \frac{4}{3} \end{pmatrix}$$

Matice je v odstupňovaném tvaru U. Zbývá dopočítat $L = R_1^{-1}R_2^{-1}$:

$$R_1^{-1} = \begin{pmatrix} 1 & & \\ -\frac{1}{2} & 1 & \\ & & 1 \end{pmatrix}, \qquad R_2^{-1} = \begin{pmatrix} 1 & & \\ & 1 & \\ & -\frac{2}{3} & 1 \end{pmatrix}, \qquad L = R_1^{-1}R_2^{-1} = \begin{pmatrix} 1 & & \\ -\frac{1}{2} & 1 & \\ & -\frac{2}{3} & 1 \end{pmatrix}.$$

Dostáváme tedy následující LU (resp. LDU) dekompozici matice A:

$$LU = \begin{pmatrix} 1 & & \\ -\frac{1}{2} & 1 & \\ & -\frac{2}{3} & 1 \end{pmatrix}\begin{pmatrix} 2 & -1 & \\ & \frac{3}{2} & -1 \\ & & \frac{4}{3} \end{pmatrix}, \qquad LDU = \begin{pmatrix} 1 & & \\ -\frac{1}{2} & 1 & \\ & -\frac{2}{3} & 1 \end{pmatrix}\begin{pmatrix} 2 & & \\ & \frac{3}{2} & \\ & & \frac{4}{3} \end{pmatrix}\begin{pmatrix} 1 & -\frac{1}{2} & \\ & 1 & -\frac{2}{3} \\ & & 1 \end{pmatrix}. \tag{3.10}$$

LU dekompozice obecně. LU dekompozici lze provést i pro matice, které nelze odstupňovat bez prohazování řádků. Pokud bychom však postupovali jako uvedeno výše, nebyla by matice L dolní trojúhelníková. Platí však, že na konci v odstupňovaném tvaru A se řádky objeví v nějakém pořadí. Můžeme tedy řádky do tohoto pořadí přeházet již na začátku a místo A uvažovat pozměněnou matici PA, kde P je vhodná permutační matice. Tuto matici již lze odstupňovat bez prohazování řádků a platí

$$PA = LU.$$

Pochopitelně pokud LU dekompozici počítáme, neznáme tu správnou permutační matici. Proto provádíme dopřednou eliminaci běžným způsobem a pouze si stranou zapisujeme, jakou permutační matici P prohazováním řádků vytváříme.

Výše uvedené tvrzení 3.14 o jednoznačnosti LU dekompozice lze rozšířit i pro prohazování řádků. Platí, že pro každou permutační matici P buď LU dekompozice *neexistuje* (pokud soustavu PA nelze odstupňovat bez prohazování řádků), nebo je LU dekompozice $PA = LU$ *určena jednoznačně.*

LU dekompozici lze provést i pro neregulární a nečtvercové matice. Pokud je $A \in \mathbb{R}^{m\times n}$, potom existují $P \in \mathbb{R}^{m\times m}$, $L \in \mathbb{R}^{m\times m}$ a $U \in \mathbb{R}^{m\times n}$, že platí $PA = LU$. Matice P je permutační, L je dolní trojúhelníková a U je odstupňovaný (Gaussův tvar) matice A. Všimněte si, že obě matice L a P jsou čtvercové a regulární. Důvod je, že matice L a P popisují posloupnost kroků aplikovaných v průběhu dopředné eliminace mezi řádky, proto jejich rozměr nezávisí na počtu sloupečků matice A. Čtenář si může sám rozmyslet detaily. Poznamenejme, že již neplatí jednoznačnost LU dekompozice.

> Obecně je LU dekompozice rozklad $PA = LU$, kde P je vhodná permutační matice, aby bylo možné odstupňovat PA bez prohazování řádků. V případě regulární matice A je pro každou permutační matici LU dekompozice určena jednoznačně, pokud existuje.

LU dekompozice a transpozice. Předpokládejme, že máme regulární matici A, kterou lze odstupňovat bez prohazování řádků, tedy platí $A = LDU$. Potom platí

$$A^\mathsf{T} = (LDU)^\mathsf{T} = U^\mathsf{T}D^\mathsf{T}L^\mathsf{T},$$

kde U^T je dolní trojúhelníková matice, D^T je diagonální matice a L^T je horní trojúhelníková matice. Dostali jsme tedy, že LDU dekompozice A a A^T jsou v pěkném vztahu.

Speciálně pokud A je symetrická matice, platí $LDU = U^\mathsf{T}D^\mathsf{T}L^\mathsf{T}$ a z jednoznačnosti LDU dekompozice plyne $L = U^\mathsf{T}$, tedy pro symetrickou matici si liší L a U pouze transpozicí. Všimněte si, že LDU dekompozice z příkladu (3.10) je skutečně symetrická.

Inverze regulárních matic. Nyní si dokončíme důkaz věty 3.7. Již víme podle lemmatu 3.8, že pokud existují inverze z obou stran, musí se rovnat. Zbývá dokázat, že regulární matice má inverzi i z levé strany, tedy pokud A je regulární, také A^T je regulární. Nejprve si ukažme jedno užitečné lemma:

Lemma 3.15. *Nechť pro regulární matici R je R^{-1} oboustranná inverze.*[4] *Soustava $A\boldsymbol{x} = \boldsymbol{b}$ má řešení, právě když soustava $AR\boldsymbol{y} = \boldsymbol{b}$ má řešení.*

Důkaz. Protože matice R je regulární, má podle tvrzení 3.4 soustava $R\boldsymbol{y} = \boldsymbol{z}$ řešení pro každé $\boldsymbol{z}$, speciálně pro pravou stranu $\boldsymbol{x}$. Pokud tedy soustava $A\boldsymbol{x} = \boldsymbol{b}$ má řešení, můžeme soustavu $AR\boldsymbol{y} = \boldsymbol{b}$ vyřešit ve dvou krocích:

$$A\boldsymbol{x} = \boldsymbol{b} \qquad \text{a následně} \qquad R\boldsymbol{y} = \boldsymbol{x}.$$

Naopak pokud má soustava $AR\boldsymbol{y} = \boldsymbol{b}$ řešení, má podle výše uvedeného i soustava $ARR^{-1}\boldsymbol{x} = A\boldsymbol{x} = \boldsymbol{b}$ řešení, což dokazuje druhou implikaci.

Navíc platí mezi řešeními $\boldsymbol{x}$ a $\boldsymbol{y}$ vztah $\boldsymbol{x} = R\boldsymbol{y}$, tedy násobení regulární maticí R zprava transformuje množinu řešení zobrazením R^{-1}. □

Důkaz věty 3.7. Pokud A je regulární matice, uvážíme její LU dekompozici $PA = LU$, pro kterou mají matice L a U nenulovou diagonálu. Pokud rozklad transponujeme, dostaneme

$$A^\mathsf{T}P^\mathsf{T} = U^\mathsf{T}L^\mathsf{T}.$$

Protože soustavy $L^\mathsf{T}\boldsymbol{x} = \boldsymbol{b}$ a $U^\mathsf{T}\boldsymbol{x} = \boldsymbol{b}$ mají řešení pro libovolnou pravou stranu $\boldsymbol{b}$, má i soustava $A^\mathsf{T}P^\mathsf{T}\boldsymbol{x} = \boldsymbol{b}$ řešení pro libovolnou pravou stranu.

Čtenář si může rozmyslet, že P^T je také permutační matice a navíc obecně má permutační matice P oboustrannou inverzi $P^{-1} = P^\mathsf{T}$. Proto má podle lemmatu 3.15 i soustava $A^\mathsf{T}\boldsymbol{x} = \boldsymbol{b}$ řešení pro každou pravou stranu. Matice A^T je regulární a má pravou inverzi $(A^\mathsf{T})^{-1}$, která je levou inverzí A. S pomocí lemmatu 3.8 je věta dokázána. □

Shrnutí

V této kapitole jsme představili matice, které spolu s vektory tvoří základní jazyk lineární algebry. Například ústřední problém tohoto textu, soustava lineárních rovnic, má v řeči matic elegantní zápis $A\boldsymbol{x} = \boldsymbol{b}$. Na maticích se definují operace sčítání, násobení skalárem, transpozice a maticové násobení. Tyto operace mají řadu hezkých vlastností. Například maticové násobení je asociativní a distributivní. Maticové násobení je komplikovanější operace, která například obecně nekomutuje. Příklady speciálních matic, které mají jednoduchou strukturu a objevují se často, jsou matice jednotkové I_n, diagonální a trojúhelníkové.

Pro matici A existuje pravá inverze pouze někdy; právě když má soustava $A\boldsymbol{x} = \boldsymbol{b}$ řešení pro každou pravou stranu $\boldsymbol{b}$. A to platí právě tehdy, když odstupňovaný tvar A neobsahuje nulový řádek. Levá inverze A existuje právě tehdy, když existuje pravá inverze A^T. Pokud pro čtvercovou matici existuje libovolná inverze, invertuje tuto matici z obou stran a navíc je určena jednoznačně. Čtvercové invertovatelné matice jsou natolik důležité, že mají speciální název regulární.

Popsali jsme také, že soustavy lineárních rovnic mají elegantní maticový zápis $A\boldsymbol{x} = \boldsymbol{b}$. Navíc regulární úpravy odpovídají regulárním maticím, které násobí soustavu zleva. Pro regulární matici R má soustava $RA\boldsymbol{x} = R\boldsymbol{b}$ stejnou množinu řešení jako $A\boldsymbol{x} = \boldsymbol{b}$. Speciálně pokud je A regulární, dostáváme pro $R = A^{-1}$ vztah $\boldsymbol{x} = R^{-1}\boldsymbol{b}$, tedy se znalostí inverze je snadné soustavu vyřešit.

[4] To platí obecně pro každou regulární matici R. Ale to vyplyne až z věty 3.7, kterou zatím nemáme dokázanou. Ve znění lemmatu budeme proto předpokládat oboustrannost inverze a v důkazu 3.7 níže použijeme lemma ve speciálním případě, pro který to bude snadné ověřit.

V závěru kapitoly jsme si ukázali LU dekompozici matice. To je rozklad $PA = LU$, kde L je dolní trojúhelníková matice s jednotkovou diagonálou, U je odstupňovaný tvar matice A a P je nějaká permutační matice. Navíc pokud A je regulární, je pro každou matici P určena LU dekompozice jednoznačně (pokud existuje). Aby LU dekompozice existovala, musí jít PA odstupňovat bez prohazování řádků. S využitím LU dekompozice jsme dokázali, že matice A je regulární, právě když A^T je regulární; tedy čtvercová regulární matice má oboustrannou inverzi.

Cvičení

⇒ **3.1** Určete inverzi $n \times n$ matice

$$\begin{pmatrix} & & & & 1 \\ & & & 1 & \\ & & \iddots & & \\ & 1 & & & \\ 1 & & & & \end{pmatrix}.$$

⋆ **3.2** Na konci podkapitoly 3.1 jsme popsali způsob, jak počítat Fibonacciho čísla pomocí umocňování matice $\left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right)$. Zobecněte tento postup pro libovolnou lineární rekurentní posloupnost $(x_0, x_1, \dots)$. Taková posloupnost má prvních k členů libovolných a pro každý další člen platí

$$x_{n+k} = \alpha_0 x_n + \alpha_1 x_{n+1} + \cdots + \alpha_{k-1} x_{n+k-1},$$

kde $\alpha_0, \dots, \alpha_{k-1}$ jsou libovolná pevná čísla. Nalezněte matici, jejímž umocňováním lze určit n-tý člen x_n.

⋆ **3.3** Uvažme třídu $\mathcal{T}$ všech matic, které mají na diagonále jednu hodnotu x a mimo diagonálu hodnotu y. Ukažte, že $\mathcal{T}$ je uzavřená na sčítání, násobení a inverze (pokud existují). Tedy pokud A a B jsou dvě matice z $\mathcal{T}$, potom i $A + B$, AB a A^{-1} patří do $\mathcal{T}$. Navíc určete, za jakých podmínek je matice z $\mathcal{T}$ invertovatelná.

⋆ **3.4** Zmínili jsme, že násobení čtvercových matic není typicky komutativní. Charakterizujte všechny matice $K \in \mathbb{R}^{n\times n}$, které komutují s libovolnou jinou maticí $A \in \mathbb{R}^{n\times n}$. Pochopitelně dokažte, že jste nalezli všechny takové matice K.

Kapitola 4

Lineární kombinace, nezávislost a báze

V kapitole 2 jsme zadefinovali vektorové podprostory jako množiny vektorů, které jsou uzavřené na operace sčítání a násobení skalárem. Tato definice popisuje vlastnosti, které musí podmnožina vektorového prostoru splňovat, abychom ji nazývali podprostorem. Přesto z ní není vůbec patrná struktura vektorových podprostorů, tedy jak tyto speciální podmnožiny vypadají. V této kapitole se pokusíme nalézt pro podprostory co nejjednodušší popis, k čemuž pomohou *lineární kombinace* a *lineární obaly*.

Také si ukážeme, jak zavést různé systémy souřadnic nad vektorovým prostorem či uvnitř jeho podprostorů. Těmto souřadným systémům budeme říkat *báze*. Každý souřadný systém musí splňovat dvě přirozené podmínky. Za prvé žádná souřadnice nesmí být nadbytečná, systém souřadnic musí být *lineárně nezávislý*. Za druhé každý vektor musí jít pomocí souřadnic vyjádřit, souřadnice musí *vygenerovat* celý prostor. Počet těchto souřadnic vyjadřuje velikost prostoru a nazývá se *dimenze*.

Pojem vektorového podprostoru je centrální, protože se v lineární algebře objevuje tak často. Ukázali jsme v kapitole 2, že podprostory jsou úzce propojené se soustavami lineárních rovnic. Pro soustavu s pravou stranu nulovou tvoří množina řešení vektorový podprostor a pro obecnou pravou stranu afinní podprostor, což je vektorový podprostor posunutý z počátku. Strukturální výsledky této kapitoly umožní lépe nahlédnout, jak všechna řešení soustavy vypadají.

V kapitole 3 jsme v souvislosti s množinou všech řešení soustavy popsali jádro matice $\mathrm{Ker}(A)$. Jádro je jeden ze čtyř fundamentálních podprostorů, kterými se budeme na konci kapitoly zabývat. Tím výrazně prohloubíme znalosti o maticích získané v kapitole 3. Také zavedeme důležitou definici hodnosti matice, která říká, jak moc je daná matice blízká regulární. Vybudované pochopení matic bude velice užitečné v následující kapitole 5.

4.1 Posloupnosti operací a lineární kombinace

Mějme nějaký vektorový prostor $\mathbb{V} = (V, +, \cdot)$. Pro něj máme definované dvě základní operace: unární operaci násobení skalárem a binární operace sčítání. Nyní budeme uvažovat další n-nární operace, kterým říkáme *posloupnosti operací*. Každá posloupnost operací je zobrazení $o : V^n \to V$, které vznikne zřetězením/složením konečně mnoha základních operací.[1] Uveďme dva příklady posloupnosti operací

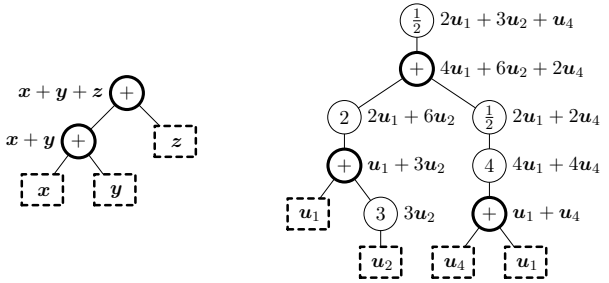
[1] Proč požadujeme konečnost? Protože je to jednodušší a pro prostory konečné dimenze, které uvažujeme v tomto textu, nejsou nekonečné posloupnosti potřeba. Pochopitelně nekonečné posloupnosti operací lze uvažovat, ale člověk se rychle dostane do problémů, kterými se zabývá matematická analýza. Výsledek takové operace nemusí vůbec být definován, například $\boldsymbol{u} - \boldsymbol{u} + \boldsymbol{u} - \boldsymbol{u} + \boldsymbol{u} - \boldsymbol{u} + \cdots$ pro $\boldsymbol{u} \neq \boldsymbol{0}$ vůbec nedává smysl, neboť částečné součty střídavě mění svoji hodnotu mezi $\boldsymbol{u}$ a $\boldsymbol{0}$. Také může být výsledek nekonečný, třeba $\boldsymbol{u}+2\boldsymbol{u}+3\boldsymbol{u}+4\boldsymbol{u}+\cdots$, což nepatří do vektorového prostoru. Také při práci s nekonečnými posloupnostmi operací řada základních vlastností přestane platit; třeba komutativita a asociativita sčítání, rozmyslete si.

arity tři a pět:

$$o(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) = (\boldsymbol{x} + \boldsymbol{y}) + \boldsymbol{z}, \qquad o'(\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3, \boldsymbol{u}_4, \boldsymbol{u}_5) = \frac{1}{2}\Big(2(\boldsymbol{u}_1 + 3\boldsymbol{u}_2) + \frac{1}{2}\big(4(\boldsymbol{u}_4 + \boldsymbol{u}_1)\big)\Big). \tag{4.1}$$

Mějme nějaký vektorový podprostor W. Uzavřenost na základní operace implikuje uzavřenost také na libovolnou konečnou posloupnost operací; což je snadné dokázat indukcí dle délky posloupnosti. Naším cílem bude nalézt pro každou posloupnost operací co nejjednodušší formu.

Aritmetické stromy. Pozastavme se na chvíli nad tím, co je to přesně posloupnost operací. Místo formální definice stačí vědět, že každou posloupnost operací lze reprezentovat *aritmetickým stromem*. To je zakořeněný strom, který obsahuje ve *vnitřních uzlech* jednotlivé operace a v každém *listu* jeden z vektorů $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$. Strom nemusí mít přesně n listů, vektory $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ se mohou v listech libovolně opakovat a některé mohou chybět. Formální definici aritmetického stromu vynecháme, čtenář si ji může zkusit sám zkonstruovat. Příklad aritmetického stromu pro posloupnosti operací (4.1) je na obrázku 4.1.



Obrázek 4.1: Nalevo je aritmetický strom pro posloupnost operací $(\boldsymbol{x}+\boldsymbol{y})+\boldsymbol{z}$, napravo pro posloupnost $\frac{1}{2}\big(2(\boldsymbol{u}_1 + 3\boldsymbol{u}_2) + \frac{1}{2}(4(\boldsymbol{u}_4 + \boldsymbol{u}_1))\big)$. Vedle vnitřních uzlů jsou připsána vyhodnocení.

Nejjednodušší tvar. Dvě posloupnosti operací arity n jsou *ekvivalentní*, pokud pro libovolnou n-tici vektorů $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ dávají shodný výsledek. Mějme jednu posloupnost operací. Rádi bychom ji co nejvíc zjednodušili, tedy chceme nalézt co nejjednodušší posloupnost operací, která je s původní posloupností ekvivalentní.

Ukážeme si to nejprve na příkladu posloupnosti operací o' z (4.1). Nejprve využijeme distributivitu na vnitřní závorky:

$$2(\boldsymbol{u}_1 + 3\boldsymbol{u}_2) = 2\boldsymbol{u}_1 + 6\boldsymbol{u}_2 \qquad \text{a} \qquad \frac{1}{2}\big(4(\boldsymbol{u}_4 + \boldsymbol{u}_1)\big) = \frac{1}{2}\big(4\boldsymbol{u}_4 + 4\boldsymbol{u}_1\big) = 2\boldsymbol{u}_4 + 2\boldsymbol{u}_1,$$

což nám zredukovalo posloupnost na jednodušší ekvivalentní tvar

$$o'(\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3, \boldsymbol{u}_4, \boldsymbol{u}_5) = \frac{1}{2}\Big(2\boldsymbol{u}_1 + 6\boldsymbol{u}_2 + 2\boldsymbol{u}_4 + 2\boldsymbol{u}_1\Big).$$

Nyní výraz uvnitř závorky přeuspořádáme s využitím komutativity a sečteme koeficienty u $\boldsymbol{u}_1$:

$$2\boldsymbol{u}_1 + 6\boldsymbol{u}_2 + 2\boldsymbol{u}_4 + 2\boldsymbol{u}_1 = (2+2)\boldsymbol{u}_1 + 6\boldsymbol{u}_2 + 2\boldsymbol{u}_4 = 4\boldsymbol{u}_1 + 6\boldsymbol{u}_2 + 2\boldsymbol{u}_4.$$

Nakonec rozdistribuujeme vnější skalární násobek $\frac{1}{2}$ a dostaneme jednoduchý ekvivalentní tvar:

$$o'(\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3, \boldsymbol{u}_4, \boldsymbol{u}_5) = \frac{1}{2}\Big(4\boldsymbol{u}_1 + 6\boldsymbol{u}_2 + 2\boldsymbol{u}_4\Big) = 2\boldsymbol{u}_1 + 3\boldsymbol{u}_2 + \boldsymbol{u}_4.$$

Výše uvedené zjednodušení je ilustrováno na aritmetických stromech na obrázku 4.1. V řeči stromů postupujeme odzdola vzhůru a zjednodušujeme mezivýsledky v jednotlivých vnitřních vrcholech.

Dokázali jsme složitou posloupnost vektorových operací (4.1) převést na jednoduché vyjádření. Stačí vzít vektory $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_5$, natáhnout je vhodnými skaláry (vektory $\boldsymbol{u}_3$ a $\boldsymbol{u}_5$ násobíme nulovým skalárem) a sečíst tyto natažené vektory dohromady. Takovému vyjádření budeme říkat *lineární kombinace*. Byla to náhoda, že se podařilo tuto posloupnost operací zjednodušit až na lineární kombinaci? Nikoliv, za okamžik ukážeme, že libovolnou posloupnost operací lze převést na nějakou lineární kombinaci.

Lineární kombinace. Zaveďme si nejprve lineární kombinace formálně. Mějme pevná reálná čísla $\alpha_1, \ldots, \alpha_n$. Lineární kombinace je n-ární posloupnost operací ℓ následujícího tvaru:

$$\ell(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n) = \sum_{i=1}^{n} \alpha_i \boldsymbol{u}_i = \alpha_1 \boldsymbol{u}_1 + \alpha_2 \boldsymbol{u}_2 + \cdots + \alpha_n \boldsymbol{u}_n. \tag{4.2}$$

Reálná čísla $\alpha_1, \ldots, \alpha_n$ se nazývají *koeficienty lineární kombinace*.

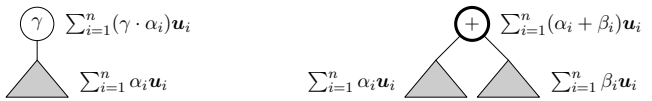
Často nás zajímají pouze lineární kombinace nějaké množiny vektorů X, což znamená, že volíme $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ pouze z této množiny X. Pochopitelně je možné, aby X byla nekonečná. V každém případě má však lineární kombinace pouze *konečně mnoho* koeficientů. Proč? Vzpomeňte si, že povolujeme jenom konečné posloupnosti operací.

Samotné spojení „lineární kombinace“ má v lineární algebře různé významy. Je to označení pro výraz (4.2); kde $\alpha_1, \ldots, \alpha_n$ a $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ mohou a nemusí být konkrétní hodnoty. Například pro tři konkrétní vektory $\boldsymbol{x}$, $\boldsymbol{y}$ a $\boldsymbol{z}$ je výraz $\boldsymbol{x} + 2\boldsymbol{y} + 3\boldsymbol{z}$ lineární kombinace. Také může lineární kombinace označovat výsledný vektor $\boldsymbol{v}$, který vznikne z nějaké lineární kombinace. Například vektor $(a, b) \in \mathbb{R}^2$ je lineární kombinace vektorů $(1, 0)$ a $(0, 1)$, protože platí $(a, b) = a(1, 0) + b(0, 1)$. Budeme se snažit, aby z kontextu vždy bylo jasné, který význam máme na mysli.

Tvrzení 4.1. *Pro každou konečnou posloupnost operací existuje lineární kombinace, která je s ní ekvivalentní.*

Důkaz. Důkaz provedeme indukcí podle počtu operací v posloupnosti, neboli podle velikosti aritmetického stromu. Pokud strom obsahuje pouze jediný vrchol, je tento vrchol list $\boldsymbol{u}_i$ a posloupnost lze triviálně zapsat jako lineární kombinaci.

Mějme aritmetický strom velikosti n. Chceme dokázat, že ho lze zjednodušit na nějakou lineární kombinaci. Klíčové je, že tento strom má nějakou operaci v kořeni (sčítání nebo násobení skalárem), za který jsou zavěšené menší podstromy (jeden nebo dva). O těchto menších podstromech víme z indukčního předpokladu, že jsou ekvivalentní nějakým lineárním kombinacím. Proto stačí na tyto lineární kombinace aplikovat operaci z kořene. Indukční krok důkazu je naznačen na obrázku 4.2.



Obrázek 4.2: Nalevo indukční krok, pokud je v kořeni násobení skalárem, napravo pro sčítání.

- *V kořeni je skalární násobení:* Máme kořen, který reprezentuje skalární násobení γ, a jeho podstrom je ekvivalentní lineární kombinaci $\sum_{i=1}^n \alpha_i \boldsymbol{u}_i$. Stačí γ rozdistribuovat[2] dovnitř lineární kombinace:

$$\gamma \cdot \left(\sum_{i=1}^{n} \alpha_i \boldsymbol{u}_i \right) = \sum_{i=1}^{n} (\gamma \cdot \alpha_i) \boldsymbol{u}_i.$$

[2] Když jsme v kapitole 2 popisovali vlastnosti vektorových prostorů, požadovali jsme pouze distributivitu jednoho sčítání, tedy $\alpha(\boldsymbol{u} + \boldsymbol{v}) = \alpha\boldsymbol{u} + \alpha\boldsymbol{v}$. Zkuste si rozmyslet a dokázat indukcí, že lze distributivitu zobecnit i na více sčítanců, tedy že platí:

$$\alpha(\boldsymbol{u}_1 + \boldsymbol{u}_2 + \cdots + \boldsymbol{u}_n) = \alpha\boldsymbol{u}_1 + \alpha\boldsymbol{u}_2 + \cdots + \alpha\boldsymbol{u}_n.$$

Podobně je třeba dokázat pro n vektorů správnost přeuspořádání pomocí komutativity a správnost asociativity, která umožňuje pro součet $\boldsymbol{u}_1 + \boldsymbol{u}_2 + \cdots + \boldsymbol{u}_n$ libovolně přeuspořádat pořadí závorek; obojí se používá v druhém případě, kdy je v kořeni operace sčítání.

Vzniklá lineární kombinace má každý z koeficientů vynásobený γ.

- *V kořeni je sčítání:* Podobně jako předtím, podstromy jsou ekvivalentní lineárním kombinacím $\sum_{i=1}^n \alpha_i \boldsymbol{u}_i$ a $\sum_{i=1}^n \beta_i \boldsymbol{u}_i$. Využijeme komutativity, asociativity a distributivity:

$$\sum_{i=1}^{n} \alpha_i \boldsymbol{u}_i + \sum_{i=1}^{n} \beta_i \boldsymbol{u}_i = \sum_{i=1}^{n} (\alpha_i \boldsymbol{u}_i + \beta_i \boldsymbol{u}_i) = \sum_{i=1}^{n} (\alpha_i + \beta_i) \boldsymbol{u}_i,$$

čtenář si může rozmyslet detaily. Tedy operace součet prostě sečte koeficienty lineárních kombinací levého a pravého podstromu.

Tím jsme ukázali, že umíme přepsat i strom velikosti n a důkaz indukcí je hotov. □

Tedy pokud chceme zkoumat posloupnosti operací na vektorech, stačí uvažovat pouze lineární kombinace. Ty mají sice jednoduchou strukturu, ale umožňují vyjádřit libovolnou konečnou posloupnost operací na vektorech. To je jeden z důvodů, proč jsou lineární kombinace tak centrálním pojmem lineární algebry; cokoliv, co lze z množiny vektorů X vytvořit, odpovídá nějaké lineární kombinaci množiny X.

> Lineární kombinace je výraz $\sum_{i=1}^n \alpha_i \boldsymbol{u}_i$, zároveň i výsledky těchto výrazů označujeme jako lineární kombinace. Pro každou konečnou posloupnost operací platí, že je ekvivalentní nějaké lineární kombinaci.

Lineární obal. V kapitole 2 jsme zadefinovali lineární obal $\mathcal{L}(X)$ množiny X jako do inkluze nejmenší vektorový podprostor obsahující X. Protože vektorové podprostory tvoří úplný svaz, má tato definice pro libovolnou množinu X smysl, lineární obal $\mathcal{L}(X)$ vždy existuje a je roven průniku všech podprostorů obsahujících X. Nyní si ukážeme alternativní definici pomocí lineárních kombinací.

Tvrzení 4.2. *Množina všech lineárních kombinací množiny vektorů X tvoří vektorový podprostor. Tento vektorový podprostor je roven $\mathcal{L}(X)$.*

Důkaz. Označme množinu všech lineárních kombinací množiny X pomocí U. Z důkazu indukčního kroku v tvrzení 4.1 vyplývá, že U je uzavřená na sčítání a násobení skalárem. Také platí, že U je neprázdná; i v případě $X = \emptyset$, protože vždy $\boldsymbol{0} \in U$. Tedy U je vektorový podprostor.

Protože platí $X \subseteq U$, je U jeden z podprostorů v průniku definujícím $\mathcal{L}(X)$ a platí $\mathcal{L}(X) \subseteq U$. Na druhou stranu pokud nějaký podprostor W obsahuje množinu X, obsahuje i libovolný vektor, který lze vytvořit konečnou posloupností operací z množiny X. Speciálně W obsahuje každou z lineárních kombinací X, a tedy platí $U \subseteq W$. Proto platí $U \subseteq \mathcal{L}(X)$ a dostáváme požadovanou rovnost. Důkaz je naznačen na obrázku 4.3. □

Ostatně správnost výše uvedeného tvrzení dává smysl. Pokud totiž vektorový podprostor obsahuje množinu X, obsahuje také všechno, co lze z X vytvořit konečným aplikováním operací vektorového prostoru; což je podle tvrzení 4.1 ekvivalentní s tím, že obsahuje libovolnou lineární kombinaci množiny X. Na druhou stranu nic dalšího nemusí vektorový podprostor podle definice obsahovat. Tedy platí uvedená rovnost, že $\mathcal{L}(X)$ je množina všech lineárních kombinací X.

Říkáme, že vektor $\boldsymbol{x}$ lze *vygenerovat* z množiny vektorů X, pokud $\boldsymbol{x} \in \mathcal{L}(X)$. To je podle tvrzení 4.2 ekvivalentní s definicí, že $\boldsymbol{x}$ lze zapsat jako lineární kombinaci množiny X. Pro zjednodušení zapisujeme lineární obal konečné množiny $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ jako $\mathcal{L}(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n)$ místo formálního $\mathcal{L}\big(\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}\big)$.

Hledání koeficientů lineární kombinace. Mějme ve vektorovém prostoru $\mathbb{R}^m$ nějaký vektor $\boldsymbol{b}$ a konečnou množinu vektorů $X = \{\boldsymbol{u}_1, \dots, \boldsymbol{u}_n\}$. Chceme vymyslet způsob, jak zjistit, zda $\boldsymbol{b} \in \mathcal{L}(X)$. To je ekvivalentní s otázkou, zda existují reálné koeficienty $\alpha_1, \dots, \alpha_n$, pro které platí rovnost

$$\alpha_1\boldsymbol{u}_1 + \alpha_2\boldsymbol{u}_2 + \cdots + \alpha_n\boldsymbol{u}_n = \boldsymbol{b}.$$

Přesně se stejnou vektorovou rovnicí (2.1) jsme se setkali v kapitole 2 (pouze hledané koeficienty byly x_i místo α_i), když jsme popisovali sloupcovou interpretaci soustavy lineárních rovnic. Tehdy jsme zmínili, že soustava $A\boldsymbol{x} = \boldsymbol{b}$ má řešení, právě když lze pravou stranu $\boldsymbol{b}$ vygenerovat ze sloupcových vektorů $\boldsymbol{u}_1, \dots, \boldsymbol{u}_n$ matice A.

Tento vztah můžeme pochopitelně použít obráceně k testování, zda $\boldsymbol{b} \in \mathcal{L}(X)$. Pro to zkonstruujeme soustavu v následujícím tvaru. Matice A má n sloupečků tvořených vektory $\boldsymbol{u}_1, \dots, \boldsymbol{u}_n$ z množiny X. Vektor pravé strany je zadaný vektor $\boldsymbol{b}$, pro který testujeme náležení do $\mathcal{L}(X)$. Chceme nalézt vektor koeficientů $\boldsymbol{\alpha}$ splňující

$$A\boldsymbol{\alpha} = \boldsymbol{b}. \tag{4.3}$$

Pochopitelně $\boldsymbol{b}$ nemusí vůbec být vyjádřitelné jako lineární kombinace množiny X, což odpovídá tomu, že soustava $A\boldsymbol{\alpha} = \boldsymbol{b}$ nemusí mít řešení. Také nemusí být koeficienty lineární kombinace určeny jednoznačně, pokud soustava $A\boldsymbol{\alpha} = \boldsymbol{b}$ má víc různých řešení.

Proto jsou soustavy tak centrálním pojmem lineární algebry. I kdybychom zvolili více algebraický přístup, že lineární algebra je studium vektorových prostorů, narazíme na soustavy lineárních rovnic přirozeně při práci s lineárními kombinacemi.

Obraz matice. Mějme libovolnou matici $A \in \mathbb{R}^{m\times n}$. V kapitole 2 jsme představili jádro matice $\mathrm{Ker}(A)$ jako jeden ze čtyř fundamentálních podprostorů definovaných maticí. Další fundamentální podprostor se nazývá *obraz matice* a definuje se jako

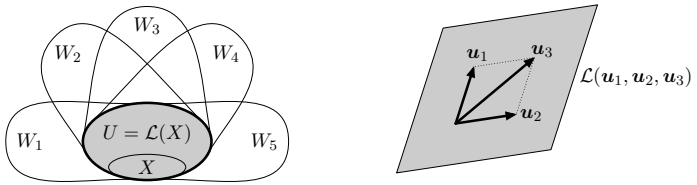
$$\mathrm{Im}(A) = \{\boldsymbol{b} \in \mathbb{R}^m : \text{soustava } A\boldsymbol{x} = \boldsymbol{b} \text{ má řešení}\}.$$

Název obraz je založen na tom, že tento podprostor odpovídá množině všech obrazů lineárního zobrazení $\boldsymbol{x} \mapsto A\boldsymbol{x}$; detailněji popsáno v kapitole 5.

Je snadné nahlédnout, že obraz je skutečně vektorový podprostor. Pro dokázání stačí ověřit uzavřenost na sčítání a násobení skalárem. Mějme $\boldsymbol{b}, \boldsymbol{c} \in \mathrm{Im}(A)$, potom podle definice obraz existují $\boldsymbol{x}$ a $\boldsymbol{y}$ splňující $A\boldsymbol{x} = \boldsymbol{b}$ a $A\boldsymbol{y} = \boldsymbol{c}$. Pak také platí $A(\boldsymbol{x}+\boldsymbol{y}) = \boldsymbol{b}+\boldsymbol{c}$, a tedy $\boldsymbol{b}+\boldsymbol{c} \in \mathrm{Im}(A)$. Tím je ověřena uzavřenost na sčítání a podobně se dokáže uzavřenost na násobení skalárem.

Výše popsaný vztah mezi koeficienty lineárních kombinací a soustavami lineárních rovnic dává elegantní popis toho, jak obraz matice A přesně vypadá. Je roven lineárnímu obalu sloupcových vektorů $\boldsymbol{u}_1, \dots, \boldsymbol{u}_n$, tedy

$$\mathrm{Im}(A) = \mathcal{L}\big(\boldsymbol{u}_1, \dots, \boldsymbol{u}_n\big).$$



Obrázek 4.3: Nalevo je naznačen důkaz tvrzení 4.2, kde podprostory W_1 až W_5 patří do průniku z definice $\mathcal{L}(X)$. Každý z nich obsahuje množinu U všech lineárních kombinací. Napravo je příklad lineárního obalu tří vektorů $\boldsymbol{u}_1$, $\boldsymbol{u}_2$ a $\boldsymbol{u}_3$. Protože platí $\boldsymbol{u}_3 = \boldsymbol{u}_1 + \boldsymbol{u}_2$, je tento obal geometricky rovina.

To je důvod, proč se někdy $\mathrm{Im}(A)$ nazývá *sloupcový prostor* matice A. Obsahuje totiž přesně ty vektory, které se dají vygenerovat jako lineární kombinace sloupcových vektorů matice A.

4.2 Lineární nezávislost a báze

Mějme konečnou množinu $X = \{\boldsymbol{u}_1, \dots, \boldsymbol{u}_n\}$. Víme, že každý vektor z lineárního obalu $\mathcal{L}(X)$ lze zapsat jako nějakou lineární kombinaci $\sum_{i=1}^n \alpha_i\boldsymbol{u}_i$. Proto množina X zavádí jakýsi souřadnicový systém nad $\mathcal{L}(X)$, protože vektory můžeme popisovat pouze pomocí n-tic koeficientů $(\alpha_1, \dots, \alpha_n)$, neboli pomocí vektorů $\boldsymbol{\alpha}$ z $\mathbb{R}^n$.

Mohlo by se zdát, že můžeme vektory $\mathcal{L}(X)$ identifikovat jedna ku jedné s vektory $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)$. Toto identifikování jedna ku jedné však není obecně možné; pro vektor $\boldsymbol{x} \in \mathcal{L}(X)$ mohou existovat různé vektory koeficientů $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^n$, pro které platí $\boldsymbol{x} = \sum_{i=1}^n \alpha_i\boldsymbol{u}_i = \sum_{i=1}^n \beta_i\boldsymbol{u}_i$. Například pro $X = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$ z obrázku 4.3 napravo lze vektor $\boldsymbol{u}_3 \in \mathcal{L}(X)$ vyjádřit pomocí různých lineárních kombinací $\boldsymbol{u}_3$ a $\boldsymbol{u}_1 + \boldsymbol{u}_2$, odpovídajících různým vektorům koeficientů $(0,0,1)$ a $(1,1,0)$. Naším cílem bude ukázat, že pokud toto identifikovaní není možné, X obsahuje nadbytečné vektory, které je možné odebrat a nezměnit přitom lineární obal.

Nadbytečné vektory. Ukážeme si nejprve situaci na příkladu $X = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$ z obrázku 4.3 napravo. Všimněme si, že vektor $\boldsymbol{u}_3$ je v množině X zcela nadbytečný. Umíme totiž $\boldsymbol{u}_3$ vygenerovat z vektorů $\boldsymbol{u}_1$ a $\boldsymbol{u}_2$, a tedy libovolnou lineární kombinaci můžeme výskyt $\boldsymbol{u}_3$ nahradit pomocí $\boldsymbol{u}_1$ a $\boldsymbol{u}_2$:

$$\alpha_1\boldsymbol{u}_1 + \alpha_2\boldsymbol{u}_2 + \alpha_3\boldsymbol{u}_3 = \alpha_1\boldsymbol{u}_1 + \alpha_2\boldsymbol{u}_2 + \alpha_3(\boldsymbol{u}_1 + \boldsymbol{u}_2) = (\alpha_1+\alpha_3)\boldsymbol{u}_1 + (\alpha_2+\alpha_3)\boldsymbol{u}_2.$$

To znamená, že můžeme místo množiny $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$ uvažovat menší množinu $\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$, z které vygenerujeme přesně to samé: $\mathcal{L}(\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3) = \mathcal{L}(\boldsymbol{u}_1, \boldsymbol{u}_2)$.

Stejně tak jsme mohli ponechat $\boldsymbol{u}_3$ a odebrat jeden z vektorů $\boldsymbol{u}_1$ a $\boldsymbol{u}_2$. Každý z nich se dá totiž vyjádřit z ostatních dvou vektorů, platí $\boldsymbol{u}_1 = \boldsymbol{u}_3 - \boldsymbol{u}_2$ a $\boldsymbol{u}_2 = \boldsymbol{u}_3 - \boldsymbol{u}_1$. Tedy platí

$$\alpha_1\boldsymbol{u}_1 + \alpha_2\boldsymbol{u}_2 + \alpha_3\boldsymbol{u}_3 = \underbrace{(\alpha_2-\alpha_1)\boldsymbol{u}_2 + (\alpha_1+\alpha_3)\boldsymbol{u}_3}_{\text{lineární kombinace bez vektoru } \boldsymbol{u}_1} = \underbrace{(\alpha_1-\alpha_2)\boldsymbol{u}_1 + (\alpha_2+\alpha_3)\boldsymbol{u}_3}_{\text{lineární kombinace bez vektoru } \boldsymbol{u}_2},$$

tedy $\mathcal{L}(\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3) = \mathcal{L}(\boldsymbol{u}_2, \boldsymbol{u}_3) = \mathcal{L}(\boldsymbol{u}_1, \boldsymbol{u}_3)$. V našem případě jsou tedy všechny tři vektory $\boldsymbol{u}_1$, $\boldsymbol{u}_2$ a $\boldsymbol{u}_3$ nadbytečné.

Uveďme nyní definici nadbytečného vektoru přesně. Vektor $\boldsymbol{u}$ je v množině X *nadbytečný*, pokud se jeho odebráním nezmění lineární obal; tedy $\mathcal{L}(X) = \mathcal{L}(X \setminus \{\boldsymbol{u}\})$. Ekvivalentně lze říct, že $\boldsymbol{u}$ nadbytečný, pokud ho lze vyjádřit jako lineární kombinaci ostatních vektorů z X (zkuste si rozmyslet, proč).

Je důležité zmínit, že v každém kroku můžeme odebrat pouze jeden nadbytečný vektor. Po každém odebrání se může nadbytečnost ostatních vektorů změnit. Pokud odebereme v našem příkladě libovolný vektor, ani jeden ze zbývajících dvou vektorů není nadbytečný; potřebujeme oba, abychom vygenerovali rovinu $\mathcal{L}(X)$.

Lineární závislost a nezávislost. Množina vektorů X je *lineárně závislá*, pokud obsahuje nadbytečný vektor. Naopak pokud žádný nadbytečný vektor neobsahuje, je množina X *lineárně nezávislá*. Pro lepší pochopení pojmů nejprve ukažme několik ekvivalentních definic lineární závislosti. Všechny tyto definice fungují i v případě nekonečně velké množiny X, i když obecně je to spíše technický detail. Lineární kombinace se nazývá *netriviální*, pokud má alespoň jeden koeficient nenulový.

Tvrzení 4.3. *Následující definice lineární závislosti množiny X jsou ekvivalentní:*

(i) Množina X obsahuje nadbytečný vektor $\boldsymbol{u}$.
(ii) Existuje v X netriviální lineární kombinace nuly.

(iii) Pro nějaký vektor $\boldsymbol{x} \in \mathcal{L}(X)$ existují dvě různé lineární kombinace množiny X, které ho vyjadřují.

Důkaz. Nejprve dokažme, že (i) je ekvivalentní s (ii). Jak už jsme zmínili, pokud $\boldsymbol{u}$ je nadbytečný vektor, lze ho vyjádřit jako lineární kombinaci vektorů $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n \in X \setminus \{\boldsymbol{u}\}$; nechť $\boldsymbol{u} = \sum_{i=1}^{n} \alpha_i \boldsymbol{u}_i$. Potom převedením $\boldsymbol{u}$ na druhou stranu získáme netriviální vyjádření nuly (neboť alespoň pro $\boldsymbol{u}$ je koeficient roven -1):

$$\boldsymbol{0} = \sum_{i=1}^{n} \alpha_i \boldsymbol{u}_i + (-1)\boldsymbol{u}.$$

Naopak mějme netriviální lineární kombinaci nuly $\boldsymbol{0} = \sum_{i=1}^{n} \alpha_i \boldsymbol{u}_i$ a předpokládejme, že $\alpha_1 \neq 0$, jinak vektory přeuspořádáme a přejmenujeme. Potom můžeme $\boldsymbol{u}_1$ vyjádřit pomocí ostatních vektorů:

$$\boldsymbol{u}_1 = \sum_{i=2}^{n} \left(-\frac{\alpha_i}{\alpha_1}\right) \boldsymbol{u}_i,$$

a tedy $\boldsymbol{u}_1$ je nadbytečný v X.

Nyní dokažme, že (ii) je ekvivalentní s (iii). Nejprve trochu upřesněme, co přesně znamená (iii). Existence dvou různých vyjádření $\boldsymbol{x}$ znamená, že existují $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n \in X$ a dva různé vektory koeficientů $\boldsymbol{\alpha}$ a $\boldsymbol{\beta}$, pro které platí $\boldsymbol{x} = \sum_{i=1}^{n} \alpha_i \boldsymbol{u}_i = \sum_{i=1}^{n} \beta_i \boldsymbol{u}_i$.[3]

Pokud máme netriviální vyjádření nuly, dostáváme pro $\boldsymbol{x} = \boldsymbol{0}$ dvě různá vyjádření: To netriviální a to triviální. Naopak pokud má $\boldsymbol{x}$ dvě různá vyjádření, je jejich rozdíl nula:

$$\boldsymbol{0} = \boldsymbol{x} - \boldsymbol{x} = \sum_{i=1}^{n} \alpha_i \boldsymbol{u}_i - \sum_{i=1}^{n} \beta_i \boldsymbol{u}_i = \sum_{i=1}^{n} (\alpha_i - \beta_i) \boldsymbol{u}_i.$$

Tato lineární kombinace nuly je netriviální, neboť vyjádření $\sum_{i=1}^{n} \alpha_i \boldsymbol{u}_i$ a $\sum_{i=1}^{n} \beta_i \boldsymbol{u}_i$ jsou různá. □

Pochopitelně negací těchto ekvivalentních definic dostaneme alternativní definice lineární nezávislosti. Množina X je lineárně nezávislá, pokud podle (ii) neobsahuje netriviální lineární kombinaci nuly nebo podle (iii) není možné vyjádřit libovolný vektory $\boldsymbol{x}$ dvěma různými způsoby.

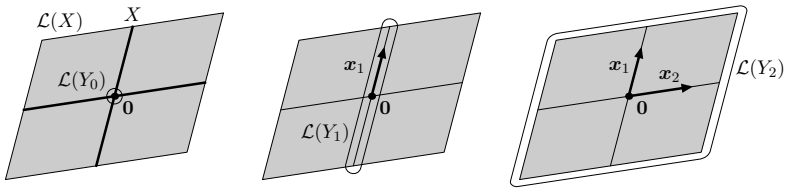
> Množina vektorů je lineárně závislá, pokud obsahuje nadbytečný vektor, jehož odebráním se nezmění lineární obal. Naopak je lineárně nezávislá, pokud žádný nadbytečný vektor neobsahuje. Ekvivalentní definice lineární závislosti jsou, že existuje více různých lineárních kombinacích vyjadřujících ten samý vektor, nebo že existuje netriviální lineární kombinace nuly.

Konstrukce lineárně nezávislé podmnožiny. Mějme nějakou lineární závislou množinu X, z které chceme odstranit nadbytečné vektory a vyrobit její lineárně nezávislou podmnožinu Y s vlastností $\mathcal{L}(X) = \mathcal{L}(Y)$. První nápad je odebírat nadbytečné vektory jeden za druhým, dokud nebude vzniklá množina lineárně nezávislá. Avšak tento postup selže, pokud je množina X nekonečná. Předpokládejme však, že libovolná lineárně nezávislá množina má konečnou velikost nejvýše d, kde d je dimenze vektorového prostoru; že takové d existuje, je vlastnost vektorových prostorů konečné dimenze, se kterými pracujeme ve většině textu. Potom můžeme vybudovat Y následujícím způsobem zdola.

Začneme s prázdnou množinou a v každém kroku přidáme jeden vektor. Budeme tedy konstruovat posloupnost množin $\emptyset = Y_0 \subsetneq Y_1 \subsetneq \cdots \subsetneq Y_k = Y$, kde v i-tém kroku vyrobíme z Y_{i-1} přidáním jednoho vektoru Y_i. Konstrukci Y_i z Y_{i-1} provedeme přidáním libovolného vektoru, který v $\mathcal{L}(Y_{i-1})$ chybí; tedy

[3] Pokud pracujeme s nekonečnou množinou X, mohla by každá z lineárních kombinací obsahovat v úplné obecnosti jiné vektory $\boldsymbol{u}_i$ z množiny X. Protože však obě lineární kombinace obsahují pouze konečně těchto vektorů, můžeme množiny použitých vektorů sloučit (a přidané vektory budou mít nulové koeficienty).

pro libovolný vektor $\boldsymbol{x}_i$ z $\mathcal{L}(X) \setminus \mathcal{L}(Y_{i-1})$ bude $Y_i = Y_{i-1} \cup \{\boldsymbol{x}_i\}$. Všimněme si, že po celou dobu udržujeme vlastnost, že Y_i je lineárně nezávislá; to lze dokázat indukcí, zjevně lineární nezávislost platí na začátku pro Y_0, a pokud je Y_{i-1} lineárně nezávislá, je i Y_i lineárně nezávislá, neboť platí $\mathcal{L}(Y_i) \subsetneq \mathcal{L}(Y_{i-1})$. Pochopitelně přidávání vektorů skončí v k-tém kroku, kdy platí $\mathcal{L}(Y_k) = \mathcal{L}(X)$, podařilo se zkonstruovat Y. Víme, že proces skončí po nejvýše d krocích, tedy $k \leq d$. Příklad této konstrukce je naznačen na obrázku 4.4.



Obrázek 4.4: Ukázka konstrukce lineárně nezávislé podmnožiny pro množinu X tvořenou dvěma přímkami procházejícími počátkem. Začneme s $Y_0 = \emptyset$ s $\mathcal{L}(Y_0) = \{\boldsymbol{0}\}$. Ve dvou krocích zkonstruujeme $Y_2 = \{\boldsymbol{x}_1, \boldsymbol{x}_2\}$, která je lineárně nezávislá a pro kterou platí $\mathcal{L}(Y_2) = \mathcal{L}(X)$.

Báze. Definujme, že *báze* je lineárně nezávislá množina vektorů, která generuje celý vektorový prostor. Tedy báze B neobsahuje nadbytečné vektory a $\mathcal{L}(B)$ je celý vektorový prostor. Podobně pro libovolný vektorový podprostor W definujeme *bázi B podprostoru W* jako lineárně nezávislou množinu s vlastností $\mathcal{L}(B) = W$. Existují další dvě ekvivalentní definice báze:

- *Báze je do inkluze nejmenší generátor.* Pokud je generátor G do inkluze nejmenší, potom neobsahuje žádný nadbytečný vektor $\boldsymbol{x}$ s vlastností $\mathcal{L}(G) = \mathcal{L}\big(G \setminus \{\boldsymbol{x}\}\big)$. To je přímo definice lineární nezávislosti.
- *Báze je do inkluze největší lineárně nezávislá množina.* Pokud by do inkluze maximální lineárně nezávislá množina M negenerovala celý vektorový prostor, existoval by nějaký vektor $\boldsymbol{x} \notin \mathcal{L}(M)$. To by ale byl spor s maximalitou, protože by množina šla zvětšit na lineárně nezávislou množinu $M \cup \{\boldsymbol{x}\}$. Proto je M generátor vektorového prostoru.

> Báze je do inkluze nejmenší generátor a zároveň do inkluze největší lineárně nezávislá množina. Celý vektorový prostor lze popsat pomocí lineárních kombinací báze.

Věta o isomorfismu. Předpokládejme, že máme konečnou bázi B vektorového prostoru $\mathbb{V}$ tvořenou vektory $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$. S využitím těchto bazických vektorů ukážeme, že prostor $\mathbb{V}$ je algebraicky zcela totožný s vektorovým prostorem $\mathbb{R}^n$.

Věta 4.4. *Zobrazení $f : \sum_{i=1}^{n} \alpha_i \boldsymbol{b}_i \mapsto (\alpha_1, \ldots, \alpha_n)$ má následující vlastnosti:*

(i) Je to bijektivní zobrazení mezi $\mathbb{V}$ a $\mathbb{R}^n$.
(ii) Pro libovolné dva vektory $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{V}$ platí $f(\boldsymbol{x} + \boldsymbol{y}) = f(\boldsymbol{x}) + f(\boldsymbol{y})$.
(iii) Pro libovolný vektor $\boldsymbol{x} \in \mathbb{V}$ a libovolné $\gamma \in \mathbb{R}$ platí $f(\gamma \cdot \boldsymbol{x}) = \gamma \cdot f(\boldsymbol{x})$.

Důkaz. Nejprve dokažme (i). Protože je množina B lineárně nezávislá, má každý vektor $\boldsymbol{x}$ podle tvrzení 4.3 jednoznačně určené vyjádření jako $\sum_{i=1}^{n} \alpha_i \boldsymbol{b}_i$, tedy zobrazení f je prosté. Na druhou stranu pro libovolně zvolené koeficienty $\alpha_1, \ldots, \alpha_n$ definuje $\sum_{i=1}^{n} \alpha_i \boldsymbol{b}_i$ nějaký vektor z $\mathbb{V}$, a tedy zobrazení f je na. Dohromady dostáváme, že f je bijektivní zobrazení.

Pro dokázání (ii) a (iii) stačí využít vlastností vektorového prostoru, ukažme pouze část (ii) a část (iii) si může čtenář dokázat jako cvičení. Mějme $\boldsymbol{x} = \sum_{i=1}^{n} \alpha_i \boldsymbol{b}_i$ a $\boldsymbol{y} = \sum_{i=1}^{n} \beta_i \boldsymbol{b}_i$. Potom s využitím komutativity, asociativity a distributivity platí

$$\boldsymbol{x} + \boldsymbol{y} = \sum_{i=1}^{n} \alpha_i \boldsymbol{b}_i + \sum_{i=1}^{n} \beta_i \boldsymbol{b}_i = \sum_{i=1}^{n} (\alpha_i + \beta_i) \boldsymbol{b}_i,$$

a tedy $f(\boldsymbol{x} + \boldsymbol{y}) = (\alpha_1 + \beta_1, \ldots, \alpha_n + \beta_n)$. □

Pozastavme se nad chvilku nad tím, proč je předchozí tvrzení natolik významné, že dostalo označení věta. Popsané zobrazení f vlastností (i) až (iii) se v algebře nazývá *izomorfismus*, a tedy jsme dokázali, že vektorové prostory $\mathbb{V}$ a $\mathbb{R}^n$ jsou *izomorfní*. Tyto tři vlastnosti jsou zcela klíčové, protože umožňují algebraicky ztotožnit $\mathbb{V}$ a $\mathbb{R}^n$. Vlastnost (i) říká, že toto ztotožnění je jedna ku jedné; každému prvku $\mathbb{V}$ přiřazujeme právě jeden prvek z $\mathbb{R}^n$. Vlastnosti (ii) a (iii) říkají, že zobrazení f zachovává algebraickou strukturu operací.

Tedy pokud máme nějakou množinu vektorů ve $\mathbb{V}$, na kterou chceme aplikovat posloupnost operací a zobrazit výsledek do $\mathbb{R}^n$, je zcela jedno, zda vektory nejprve aplikujeme posloupnost operací, nebo je nejprve zobrazíme do $\mathbb{R}^n$. Také platnost vektorových rovnic, například $\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{z}$, se přenáší mezi prostory $\mathbb{V}$ a $\mathbb{R}^n$. Prakticky všechny věci zatím uvažované v tomto textu využívají pouze algebraických vlastností prostorů, a tedy fungují ve $\mathbb{V}$ a $\mathbb{R}^n$ totožně. Například soustavy lineárních rovnic můžeme řesit ve $\mathbb{V}$ tak, že pomocí zobrazení f převedeme abstraktní vektory v konkrétní vektory $\mathbb{R}^n$, vyřešíme v konkrétních koeficientech, a inverzním zobrazením f^{-1} interpretujeme výsledek v rámci $\mathbb{V}$.

To, že jsou vektorové prostory $\mathbb{V}$ a $\mathbb{R}^n$ algebraicky totožné, ještě neznamená, že jsou to stejné prostory; pouze se algebraicky chovají stejně. Uvažme například vektorový prostor $\mathbb{V}$ všech polynomů $a_k x^k + a_{k-1} x^{k-1} + \cdots + a_1 x^1 + a_0 x^0$ stupně k, na nichž definujeme sčítání a násobení skalárem po složkách. Tento prostor má $(k+1)$-prvkovou bázi $B = \{x^0, x^1, x^2, \ldots, x^k\}$ a proto ho ho lze podle věty 4.4 identifikovat s vektorovým prostorem $\mathbb{R}^{k+1}$. I když se tyto dva prostory chovají algebraicky totožně (vzhledem k operacím sčítání a násobení skalárem), jsou polynomy stupně k matematicky zcela jiné objekty než uspořádané $(k+1)$-tice reálných čísel.

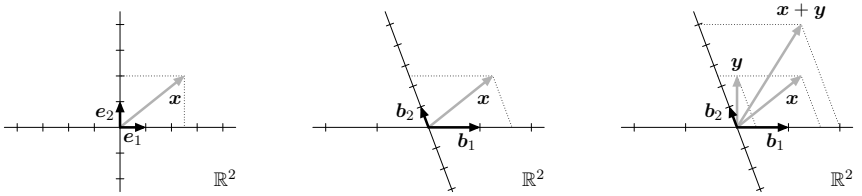
Věta o izomorfismu říká, že vektorový prostor s n-prvkovou bází je algebraicky totožný s vektorovým prostorem $\mathbb{R}^n$.

Báze jako systémy souřadnic. Na výše popsaný izomorfismus f lze nahlédnout také tak, že báze B zavádí nad vektorovým prostorem $\mathbb{V}$ souřadný systém. Pro vektor $\boldsymbol{x} = \sum_{i=1}^{n} \alpha_i \boldsymbol{b}_i$ definujeme $\alpha_1, \ldots, \alpha_n$ jeho souřadnice vůči bázi B. Souřadnice říkají, jak moc vektor $\boldsymbol{x}$ ukazuje ve směru bazických vektorů $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$. Bazické vektory určují směry souřadných os a jednotkové délky na nich.

Připomeňme si, jak jsme zaváděli na začátku kapitoly 2 vektorový prostor $\mathbb{R}^n$. Každému bodu v n-rozměrném prostoru jsme přiřadili n-tici souřadnic $(x_1, \ldots, x_n)$. Co jsou přesně tyto souřadnice? Tyto souřadnice přesně odpovídají souřadnicím vůči kanonické bázi tvořené vektory $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$, kde (jak jsme již definovali v kapitole 3) vektor $\boldsymbol{e}_i$ je vektor s jediným koeficientem nenulovým, který je na pozici i a je roven jedné. Tedy

$$(\boldsymbol{e}_i)_j = \begin{cases} 1, & \text{pro } i = j, \text{ a} \\ 0 & \text{jinak.} \end{cases}$$

Této bázi $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$ se říká *kanonická báze* a čtenář si může jako cvičení dokázat, že je to skutečně báze $\mathbb{R}^n$. Protože $\boldsymbol{x} = (x_1, \ldots, x_n) = \sum_{i=1}^{n} x_i \boldsymbol{e}_i$, jsou souřadnice vektoru $\boldsymbol{x}$ vůči kanonické bázi přesně jednotlivé koeficienty vektoru $\boldsymbol{x}$. Na obrázku 4.5 vlevo a uprostřed je příklad vyjádření jednoho vektoru vůči dvěma různým bázím.



Obrázek 4.5: Nalevo je vektor $\boldsymbol{x}$ vyjádřen vůči kanonické bázi $\{\boldsymbol{e}_1, \boldsymbol{e}_2\}$, uprostřed vůči jiné bázi $\{\boldsymbol{b}_1, \boldsymbol{b}_2\}$, tečkovaně jsou naznačeny hodnoty jednotlivých souřadnic $\boldsymbol{x}$. Napravo je součet dvou vektorů $\boldsymbol{x} + \boldsymbol{y}$, jehož souřadnice jsou rovny součtu souřadnic $\boldsymbol{x}$ a $\boldsymbol{y}$. Povšimněme si, že pro různé báze mají vektory různé souřadnice, například $\boldsymbol{x} + \boldsymbol{y}$ má vůči bázi $\{\boldsymbol{b}_1, \boldsymbol{b}_2\}$ souřadnice $(2, 5)$, zatímco vůči kanonické bázi $(\frac{5}{2}, 4)$.

Předchozí věta o izomorfismu f má následující význam v řeči souřadnic. Pokud chceme dva vektory $\boldsymbol{x}$ a $\boldsymbol{y}$ sečíst, potom stačí po složkách sečíst jejich souřadnice. A pokud chceme vektor $\boldsymbol{x}$ vynásobit skalárem α, potom stačí vynásobit jeho souřadnice skalárem α. To přesně odpovídá definici $\mathbb{R}^n$ z kapitoly 2, kdy jsme vektorové operace definovali po složkách. Na obrázku 4.5 vpravo je ukázku sečtení dvou vektorů v řeči souřadnic.

Každá n-prvková báze zavádí nad prostorem systém souřadných os a přiřazuje každému vektoru n-tici reálných čísel. Tato reálná čísla jsou koeficienty lineární kombinace vektorů báze.

Steinitzova věta o výměně. Mějme nějakou množinu generující nějaký vektorový podprostor. Naším cílem bude ukázat, že na této množině není nic speciální a že za určitých podmínek můžeme její vektory nahradit za jiné vektory, aniž bychom změnili lineární obal. Pro začátek dokažme, že můžeme nahradit alespoň jeden vektor.

Lemma 4.5 (o výměně)**.** *Nechť X je libovolná množina vektorů a $\boldsymbol{y} \neq \boldsymbol{0}$ leží v $\mathcal{L}(X)$. Potom existuje $\boldsymbol{z} \in X$, že*

$$\mathcal{L}(X) = \mathcal{L}\big(X \setminus \{\boldsymbol{z}\} \cup \{\boldsymbol{y}\}\big).$$

Důkaz. Protože $\boldsymbol{y} \in \mathcal{L}(X)$, je $\boldsymbol{y}$ lineární kombinace konečně mnoha vektorů $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in X$:

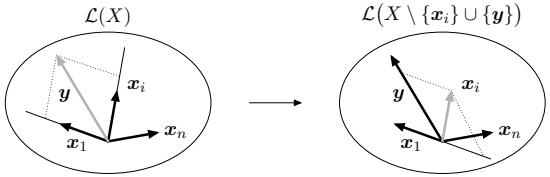
$$\boldsymbol{y} = \alpha_1 \boldsymbol{x}_1 + \alpha_2 \boldsymbol{x}_2 + \cdots + \alpha_n \boldsymbol{x}_n.$$

Navíc protože $\boldsymbol{y} \neq \boldsymbol{0}$, platí, že alespoň jeden z těchto koeficientů α_i je nenulový. Zvolíme tedy $\boldsymbol{z} = \boldsymbol{x}_i$, a zbývá pouze dokázat, že záměnou $\boldsymbol{y}$ za $\boldsymbol{x}_i$ se nezmění lineární obal.

Abychom toto dokázali, stačí ukázat, že $\boldsymbol{x}_i \in \mathcal{L}\big(X \setminus \{\boldsymbol{x}_i\} \cup \{\boldsymbol{y}\}\big)$. Pro to stačí z výše uvedené lineární kombinace vyjádřit $\boldsymbol{x}_i$:

$$\boldsymbol{x}_i = \frac{\boldsymbol{y} - \alpha_1 \boldsymbol{x}_1 - \cdots - \alpha_{i-1} \boldsymbol{x}_{i-1} - \alpha_{i+1} \boldsymbol{x}_{i+1} - \cdots - \alpha_n \boldsymbol{x}_n}{\alpha_i} =$$
$$= \frac{1}{\alpha_i} \boldsymbol{y} - \frac{\alpha_1}{\alpha_i} \boldsymbol{x}_1 - \cdots - \frac{\alpha_{i-1}}{\alpha_i} \boldsymbol{x}_{i-1} - \frac{\alpha_{i+1}}{\alpha_i} \boldsymbol{x}_{i+1} - \cdots - \frac{\alpha_n}{\alpha_i} \boldsymbol{x}_n.$$

Důkaz je ilustrován na obrázku 4.6. □



Obrázek 4.6: Vyměníme vektor $\boldsymbol{y}$ za libovolný vektor $\boldsymbol{x}_i$, který má nenulovový koeficient v lineární kombinaci vyjadřující $\boldsymbol{y}$. Protože můžeme z ostatních vektorů X spolu s $\boldsymbol{y}$ vygenerovat $\boldsymbol{x}_i$, nezměnil se lineární obal.

Co když ale chceme v množině X zaměnit místo jednoho vektoru $\boldsymbol{y}$ několik vektorů $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k$? První nápad je aplikovat lemma 4.5 o výměně na množinu X několikrát a postupně vložit vektory $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k$ jeden za druhým. S tímto postupem se snadno můžeme dostat do problémů, protože při vkládání $\boldsymbol{y}_i$ může lemma 4.5 o výměně odebrat předtím vložený $\boldsymbol{y}_j$. S tímto problémem se můžeme vypořádat pouze někdy.

Obecně nemusí být možné zaměnit v množině X vektory $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k$ současně. Může se totiž stát, že málo vektorů z množiny X ukazuje směrem $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k$, a tedy nemáme jich dost na výměnu. Pokud jsou však vektory $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k$ lineárně nezávislé, tedy žádný z nich není nadbytečný, lze výměnu vždy provést. To dokazuje Steinitzova věta, jejíž znění je ilustrováno na obrázku 4.7:

Věta 4.6 (Steinitzova o výměně)**.** *Nechť X je libovolná množina vektorů a $Y = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k\}$ je libovolná lineárně nezávislá množina vektorů z $\mathcal{L}(X)$.*

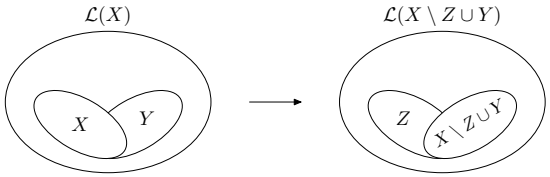
(i) Vždy existuje $Z \subseteq X$, že $|Z| = k$ a platí

$$\mathcal{L}(X) = \mathcal{L}\big(X \setminus Z \cup Y\big).$$

(ii) Pokud je X navíc lineárně nezávislá, je i množina $X \setminus Z \cup Y$ lineárně nezávislá.

Důkaz. (i) Aplikujeme výše uvedený nápad s opakovaným použitím lemmatu 4.5 o výměně, ale musíme si dát pozor, abychom neodebrali předtím vložené vektory $\boldsymbol{y}_j$. Klíčové z důkazu lemmatu o výměně je, že za $\boldsymbol{z}$ můžeme zvolit libovolný vektor, který má nenulový koeficient v lineární kombinaci rovné $\boldsymbol{y}$. Tedy předpokládejme, že jsme v množině X už zaměnili vektory $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{i-1}$, a žádný z těchto vektorů jsme neodebrali. Označme tuto pozměněnou množinu X'. Protože $\mathcal{L}(X) = \mathcal{L}(X')$, lze $\boldsymbol{y}_i$ vyjádřit jako nějakou lineární kombinaci:

$$\boldsymbol{y}_i = \alpha_1 \boldsymbol{x}_1 + \cdots + \alpha_n \boldsymbol{x}_n + \beta_1 \boldsymbol{y}_1 + \cdots \beta_{i-1} \boldsymbol{y}_{i-1}.$$



Obrázek 4.7: Pokud je Y lineárně nezávislá množina vektorů z $\mathcal{L}(X)$, je možné vyměnit k vektorů v množině X za vektory $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k$ tak, že výsledná množina $X \setminus Z \cup Y$ má stejný lineární obal.

Podle výše uvedeného důkazu lze vektorem $\boldsymbol{y}_i$ nahradit libovolný vektor, který má nenulový koeficient v této lineární kombinaci. Není možné, že by se všechny nenulové koeficienty byly pouze $\beta_1, \ldots, \beta_{i-1}$, protože jinak by platilo

$$\boldsymbol{y}_i = \beta_1 \boldsymbol{y}_1 + \cdots + \beta_{i-1} \boldsymbol{y}_{i-1},$$

a množina Y by byla lineárně závislá. Proto existuje alespoň jeden koeficient α_j, který je nenulový. Proto nahradíme vektorem $\boldsymbol{y}_i$ příslušný vektor $\boldsymbol{x}_j$, který vložíme do konstruované množiny Z.

(ii) Pro dokázaní druhé části stačí ukázat, že toto platí pro lemma 4.5 o výměně, a zbytek plyne indukcí. Předpokládejme, že X je lineárně nezávislá. Pokud by $X \setminus \{\boldsymbol{z}\} \cup \{\boldsymbol{y}\}$ byla lineárně závislá, existovala by netriviální lineární kombinace nuly. Protože ale X byla lineárně nezávislá, musí v této lineární kombinací být nenulový koeficient u $\boldsymbol{y}$, a tedy $\boldsymbol{y}$ umíme vyjádřit pomocí vektorů z $X \setminus \{\boldsymbol{z}\}$. Potom však umíme vyjádřit i $\boldsymbol{z}$ pomocí těchto vektorů, a tedy dostáváme spor s lineární nezávislostí X. □

Možná je překvapivé, že jsme výše uvedené tvrzení nazvali větou, když je jeho důkaz relativně snadný. Avšak síla Steinitzovy věty spočívá v tom, že má řadu netriviálních a důležitých důsledků v souvislosti se strukturou podprostorů a lineárně nezávislých množin. Speciálně víme, že pro daný prostor existuje spousta bází a na jejich vektorech není nic speciálního, můžeme je celkem snadno vyměnit za jiné. Dostáváme například následující snadný fakt:

Důsledek 4.7. *Libovolnou lineárně nezávislou množinu lze rozšířit na bázi.* □

Steinitzova věta o výměně říká, že v libovolné množině X můžeme zaměnit k jejích vektorů za jinou k-prvkovou lineárně nezávislou množinu z $\mathcal{L}(X)$, aniž bychom změnili lineární obal. Na vektorech množiny X není oproti vektorům v $\mathcal{L}(X)$ nic speciálního.

Dimenze prostoru. Klíčový důsledek však je, že Steinitzova věta umožňuje zavést pojem *dimenze prostoru* (či podprostoru), který říká, jak je daný prostor velký. Dimenze prostoru se definuje jako velikost libovolné báze, a budeme nyní předpokládat, že každá báze obsahuje pouze konečně mnoho vektorů. Problém s touto definicí je ten, že není zřejmé, proč by měla mít každá báze stejnou velikost. Pojďme si to tedy dokázat.

Důsledek 4.8. *Nechť $\mathbb{V}$ je vektorový prostor a nechť $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ a $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m$ jsou dvě jeho báze. Potom platí $n = m$.*

Důkaz. Označme první bázi X a druhou Y, obě množiny jsou zjevně lineárně nezávislé. Lze na ně tedy aplikovat obě části Steinitzovy věty 4.6, a to oběma směry. Protože je možné povyměňovat vektory Y za vektory X, dostáváme $n \geq m$. Podobně je možné povyměňovat vektory X za vektory Y, a tedy dostáváme $n \leq m$. Proto $n = m$. □

Označme v dalším textu dimenzi (pod)prostoru $\mathbb{V}$ jako $\dim \mathbb{V}$. Podle věty 4.4 víme, že vektorový prostor $\mathbb{V}$ dimenze n je izomorfní $\mathbb{R}^n$. Přitom platí, že pro různé hodnoty n jsou prostory $\mathbb{R}^n$ navzájem neizomorfní. Liší se totiž svojí velikostí bází, ale izomorfní prostory musejí mít stejně velké báze. Tedy $\mathbb{V}$ je izomorfní právě $\mathbb{R}^n$ a žádnému jinému $\mathbb{R}^m$ pro $m \neq n$.

Možná čtenáři nepřijde existence dimenze příliš překvapivá. Dává přece geometricky smysl, že v $\mathbb{R}^n$ potřebujeme n souřadných os v libovolném souřadném systému. Tedy velikost každé báze musí být n. Avšak něco takového rozhodně není automatické a ukazuje to, jak neuvěřitelné silná je struktura vektorových prostorů. Ukažme si, že něco takového je v matematice spíše výjimka.

V kapitole 6 si popíšeme jinou slavnou matematickou strukturu zvanou *grupa*. Podobně pro grupy lze uvažovat generátory, avšak do inkluze minimální generátory nemusí být stejné velikosti. (Zde do

inkluze minimální generátory jsou obdoba bází, i když se tak pro grupy nenazývají.) Například uvažme grupu $\mathbb{Z}_6$ tvořenou přirozenými čísly $\{0,1,2,3,4,5\}$, spolu s operací sčítaní modulo šest. (Tedy například $3+4=1$.) Jeden minimální generátor je $\{1\}$, protože opakovaným přičítáním jedničky lze vygenerovat celou grupu $\mathbb{Z}_6$. Ale jiný do inkluze minimální generátor je $\{2,3\}$, například protože umíme vygenerovat jedničku jako $2+2+3$. A tento generátor je do inkluze minimální, protože odebráním libovolného z těchto dvou prvků celou grupu nevygenerujeme.

Jinými slovy tohle ukazuje, že struktura grup je výrazně složitější než struktura vektorových prostorů. Ostatně po stovkách let zkoumání nejsou grupy pořád dostatečně pochopené. Jeden z největších výsledků nedávné matematiky se týká klasifikace konečných jednoduchých grup, jejíž důkaz je dlouhý tisíce stránek.

> Pro každý vektorový (pod)prostor existuje číslo zvané dimenze, které udává jeho velikost. Toto číslo je velikost libovolné báze, podle Steinitzovy věty jsou všechny báze stejně velké.

Podprostory geometricky. Na konci kapitoly 2 na obrázku 2.10 jsme bez důkazu zmínili, jak vypadají geometricky podprostory prostoru $\mathbb{R}^3$. Nyní si to odvodíme pomocí vybudované teorie bází. Dimenze $\mathbb{R}^3$ je tři, neboť vektory $(1,0,0)$, $(0,1,0)$ a $(0,0,1)$ tvoří jeho bázi. Dimenze každého podprostoru je určitě menší než dimenze celého prostoru, tedy patří do množiny $\{0,1,2,3\}$. Podprostory potom můžeme rozčlenit právě podle této dimenze:

- *Dimenze 0.* Lineární obal prázdné množiny vektorů obsahuje pouze počátek, který dostaneme z prázdné lineární kombinace.[4] Proto jediný podprostor dimenze nula je podprostor $\{\mathbf{0}\}$ obsahující pouze počátek. Tento podprostor je infimum svazu všech podprostorů.
- *Dimenze 1.* Lineární obal jediného nenulového vektoru $\mathcal{L}(\boldsymbol{u})$ je přímka procházející počátkem ve směru $\boldsymbol{u}$. Lineární obal spolu s $\boldsymbol{u}$ obsahuje i jeho libovolné natažení. Všechna natažení vytvoří přímku. Podprostorů dimenze 1 je nekonečně mnoho.
- *Dimenze 2.* Lineární obal dvou lineárně nezávislých vektorů je rovina. Rovina bude obsahovat přímky ve směru těchto dvou vektorů a všechny body, které můžeme z těchto dvou přímek zkombinovat. Můžete si to třeba představit tak, že vezmeme jednu přímku a posouváme ji ve směru druhé přímky. Podprostorů dimenze 2 je také nekonečně mnoho.
- *Dimenze 3.* Tři lineárně nezávislé vektory v $\mathbb{R}^3$ už generují celý prostor. Tedy i když existuje mnoho bází velikosti tři, všechny generují jeden a ten samý vektorový podprostor, a to $\mathbb{R}^3$. Tento podprostor je supremum svazu všech podprostorů.

Poznamenejme, že vektorové podprostory dimenze 0 a plné dimenze se nazývají *triviální*.

Proč uvažovat různé báze? Zadefinovali jsme si pojem báze a ukázali jsme, že báze odpovídají různým souřadnicovým systémům nad vektorovým prostorem. Proč jsou však báze natolik klíčovým pojmem lineární algebry? A proč si nevystačíme s klasickou kanonickou bází a potřebujeme uvažovat i jiné souřadné systémy? Důvodů je hned několik.

Například můžeme dostat (třeba z experimentu) množinu vektorů v $\mathbb{R}^n$ jako data, u kterých chceme pochopit strukturu. Ta nemusí být vůbec patrná ze souřadnic těchto vektorů vůči kanonické bázi. Avšak při zvolení vhodné báze můžeme snadno objevit vlastnosti skryté v kanonické bázi. Takové aplikace jsou klíčové v matematické statistice.

[4] Obvykle se v matematice používá následující konvence. Pokud sčítáme přes prázdnou množinu, výsledek je nula (v našem případě reprezentovaná počátkem). Pokud násobíme přes prázdnou množinu (což v případě lineární algebry neděláme moc často, protože násobení není lineární), výsledek je jednička. Zkuste si rozmyslet, proč dává smysl mít takovou konvenci.

S tím úzce souvisí druhý důvod. V lineární algebře zkoumáme vektorové prostory a jejich transformace, kterými si budeme zabývat v kapitole 5. Ukážeme si, že pro každou bázi odpovídá daná transformace jedné matici. Naším cílem je pochopit strukturu této transformace, a proto chceme zvolit bázi (souřadný systém), vůči které se transformace chová co nejjednodušeji. Toto je přesně motivace pro studium *vlastních vektorů*, kterými se budeme časem zabývat.

Ostatně příkladem výše uvedeného zvolení správné báze je slavná *Fourierova transformace*, což je transformace, která mění jednu funkci v jinou. Z pohledu lineární algebry však nejde o nic jiného než zvolení vhodné báze na prostoru funkcí a transformace přepočítává souřadnice mezi kanonickou a touto bází. Ta je tvořena funkcemi sinus a cosinus a má celou řadu úžasných vlastností a aplikacích. Fourierova transformace se právem řadí k jednomu z nejdůležitějších matematických objevů.

⋆**Vzorec pro Fibonacciho čísla.** Ukažme si na jednom příkladu, že zvolením správné báze se mohou problémy zjednodušit. V kapitole 3 jsme popsali, jak vypočítat n-té Fibonacciho číslo efektivně pomocí umocňování jedné matice. Připomeňme si, že Fibonacciho čísla se definují lineární rekurencí:

$$f_0 = 0, \qquad f_1 = 1, \qquad f_{n+2} = f_{n+1} + f_n.$$

Také jsme si uvedli vzorec pro n-té Fibonacciho číslo:

$$f_n = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^n. \tag{4.4}$$

Pojďme si tento vzorec odvodit zvolením vhodné báze.

Uvažme vektorový prostor všech posloupností $(a_0, a_1, a_2, \dots)$ spolu se sčítáním po složkách a násobení skalárem. Tento prostor má nekonečnou dimenzi, a proto se zaměříme na podprostor všech *fibonacciovských posloupností*. Posloupnost $\{a_n\}$ se nazývá fibonacciovská, pokud splňuje rekurenci $a_{n+2} = a_{n+1} + a_n$; tedy na rozdíl od Fibonacciho posloupnosti může mít libovolné první dva členy. Čtenář může snadno ověřit, že fibonacciovské posloupnosti tvoří podprostor.

Zjevně dimenze tohoto podprostoru je dva, protože stačí zvolit první dva členy a zbytek posloupnosti je jednoznačně určený, tedy tento prostor je izomorfní $\mathbb{R}^2$. Posloupnosti $(1,0,1,1,2,3,5,8,\dots)$ a $(0,1,1,2,3,5,8,13,\dots)$ tvoří kanonickou bázi; tyto dvě slavné posloupnosti jsou posunutá Fibonacciho čísla a Fibonacciho čísla. Zkusíme však nalézt jednodušší bázi tvořenou posloupnostmi, pro něž budeme schopni určit n-tý člen.

Příkladem takové jednoduché fibonacciovské posloupnosti by byla posloupnost $a_n = x^n$ pro nějakou hodnotu x. Ukážeme si, že existují dvě takové posloupnosti, které tvoří bázi. Protože a_n splňuje rekurenci, musí x splňovat rovnost $x^{n+2} = x^{n+1} + x^n$, neboli $x^2 - x - 1 = 0$. Snadným výpočtem zjistíme, že tato rovnost platí pro $x = \frac{1\pm\sqrt{5}}{2}$, a tedy dostáváme dvě fibonacciovské posloupnosti:

$$a_n = \left(\frac{1+\sqrt{5}}{2}\right)^n \qquad \text{a} \qquad b_n = \left(\frac{1-\sqrt{5}}{2}\right)^n.$$

Protože jsou tyto dvě posloupnosti lineárně nezávislé, tvoří bázi. To, že se tyto dva výrazy objevují v (4.4) pochopitelně není náhoda.

Označme tyto dvě posloupnosti jako vektory $\boldsymbol{a}$ a $\boldsymbol{b}$. Protože tvoří bázi, existují pro libovolnou fibonacciovskou posloupnost koeficienty α a β, že lze tuto posloupnost zapsat jako $\alpha\boldsymbol{a} + \beta\boldsymbol{b}$. Označme Fibonacciho posloupnost $\boldsymbol{f}$. Zbývá dopočítat koeficienty α a β, pro které platí $\boldsymbol{f} = \alpha\boldsymbol{a} + \beta\boldsymbol{b}$. Vyjádříme si tyto vektory vůči kanonické bázi, což vede na vektorovou rovnici $(0,1) = \alpha(1, \frac{1+\sqrt{5}}{2}) + \beta(1, \frac{1-\sqrt{5}}{2})$, když zapíšeme pouze první dva členy. Tedy dostáváme soustavu lineárních rovnic:

$$\begin{array}{rcrcl} \alpha & + & \beta & = & 0, \\ \frac{1+\sqrt{5}}{2}\alpha & + & \frac{1-\sqrt{5}}{2}\beta & = & 1. \end{array}$$

Vyřešením této soustavy dostaneme $\alpha = \frac{1}{\sqrt{5}}$ a $\beta = -\frac{1}{\sqrt{5}}$, čímž dostáváme vzorec (4.4). Dokonce popsaná metoda umožňuje určit vzorec pro libovolnou fibonacciovskou posloupnost, například pro Lucasova čísla, která jsou fibonacciovská posloupnost $(2, 1, 3, 4, 7, 11, 18, \dots)$; zkuste si určit vzorec jako cvičení.

Výše uvedený postup pravděpodobně působí překvapivě. Vytáhli jsme králíka z klobouku v podobě báze $a_n = x^n$, ono to náhodou vyšlo a nalezli jsme vzorec pro Fibonacciho čísla. Není však vůbec zřejmé, jak takový postup objevit, když ho člověk nezná. V kapitole **??** si ukážeme, že na tomto postupu není nic překvapivého. Protože rekurzivní vztah pro Fibonacciho čísla je lineární, musí podobná báze vždy existovat. Tato báze souvisí s maticí

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

zmíněnou v kapitole 3 a jejími vlastními vektory. Abychom tuto souvislost mohli popsat, potřebujeme znát podstatně víc teorie lineární algebry.

4.3 Fundamentální podprostory a hodnost matice

Pokusíme se aplikovat vybudovanou teorii bází na matice. S každou maticí jsou spojeny čtyři fundamentální podprostory, o nichž jsme si již ukázali částečné výsledky. Ukážeme se další vlastnosti těchto podprostorů, které nám prozradí mnoho užitečné o maticích, například jaké mají dimenze. Tím zobecníme a lépe pochopíme řadu výsledků získaných v kapitole 3. Navíc fundamentální podprostory jsou skvělými příklady podprostorů, takže si lépe osvojíme strukturální vlastnosti podprostorů. Také zavedeme důležitou definici hodnosti matice, který udává, jak moc je daná matice regulární.

Fundamentální podprostory. Nechť $A \in \mathbb{R}^{m\times n}$. Definujeme pro ní následující čtyři fundamentální podprostory, první dva už jsme v textu zmínili:

- *Jádro* $\mathrm{Ker}(A)$*:* Množina všech řešení $A\boldsymbol{x} = \boldsymbol{0}$.
- *Obraz* $\mathrm{Im}(A)$*:* Množina všech pravých stran $\boldsymbol{b}$, pro které existuje řešení $A\boldsymbol{x} = \boldsymbol{b}$. Na začátku této kapitoly jsme si dokázali, že tento prostor je roven lineárnímu obalu sloupcových vektorů matice, a tedy se také nazývá *sloupcový prostor*.
- *Řádkový prostor* $\mathcal{R}(A)$*:* Lineární obal řádkových vektorů matice. Tento podprostor se také občas nazývá levý obraz, neboť je roven $\mathrm{Im}(A^\mathsf{T})$.
- *Levé jádro* $\mathrm{Ker}(A^\mathsf{T})$*:* Množina všech řešení $A^\mathsf{T}\boldsymbol{x} = \boldsymbol{0}$. Název levé jádro dostal proto, že odpovídá násobení $\boldsymbol{x}$ zleva, tedy množině všech řešení $\boldsymbol{x}^\mathsf{T}A = \boldsymbol{0}^\mathsf{T}$.

První dva podprostory se v textu objevily, protože nám pomáhaly vysvětlit probíraná témata. Například víme, že množina všech řešení $A\boldsymbol{x} = \boldsymbol{b}$ je rovna afinnímu podprostoru tvořenému posunutí $\mathrm{Ker}(A)$ do libovolného řešení. (Samozřejmě za předpokladu, že alespoň jednou řešení existuje.) Obraz matice neboli sloupcový prostor se zase přirozeně objevil při zkoumání lineárního obalu dané množiny vektorů.

Zatím však není zřejmé, proč zavádíme druhé dva podprostory. Ty jsou totiž rovny obrazu a jádru A^T. V kapitole 3 zavedli transpozici matice tak, že prohodíme v matici pořadí indexů; tedy koeficienty se „zrcadlí“ podle diagonály. Jeden z hlavních cílů bude ukázat, že je mezi maticemi A a A^T výrazně hlubší souvislost, které jsme se již dotkli při důkazu věty 3.7, že regularita implikuje existenci inverze z obou stran. Klíčové je, že pokud chceme geometricky pochopit, co lineární zobrazení reprezentované maticí A dělá, musíme uvažovat také $\mathcal{R}(A) = \mathrm{Im}(A^\mathsf{T})$ a $\mathrm{Ker}(A^\mathsf{T})$.

Protože matice A je velikosti $m \times n$, uvažme dva vektorové prostory $\mathbb{R}^m$ a $\mathbb{R}^n$. Čtyři fundamentální podprostory jsou podprostory těchto prostorů po dvou: Platí, že $\mathcal{R}(A)$ a $\mathrm{Ker}(A)$ jsou podprostory $\mathbb{R}^n$, a že $\mathrm{Im}(A)$ a $\mathrm{Ker}(A^T)$ jsou podprostory $\mathbb{R}^m$. Fundamentální podprostory jsou společně zachyceny na obrázku 4.8, spolu s lineárním zobrazením $A : \boldsymbol{x} \mapsto A\boldsymbol{x}$. Platí, že obraz celého $\mathbb{R}^n$ je $\mathrm{Im}(A)$. Tedy

alternativní definice obrazu matice je

$$\mathrm{Im}(A) = \{A\boldsymbol{x} : \boldsymbol{x} \in \mathbb{R}^n\}.$$

Definice hodnosti. Jak už jsme zmínili, hodnost matice udává, jak moc je daná matice regulární. Existuje řada ekvivalentních definic hodnosti, a cílem této kapitoly bude ukázat jejich ekvivalenci a vztahy. Definujme *hodnost* $\mathrm{rank}(A)$ matice A jako dimenzi $\mathrm{Im}(A)$, což geometricky odpovídá tomu, jak moc lineární zobrazení $A : \boldsymbol{x} \mapsto A\boldsymbol{x}$ zužuje $\mathbb{R}^n$. Protože $\mathrm{Im}(A)$ je podprostor $\mathbb{R}^m$, zjevně platí $\mathrm{rank}(A) \le m$. Na druhou stranu platí, že $\mathrm{Im}(A)$ je lineární obal n sloupcových vektorů $\boldsymbol{u}_1, \dots, \boldsymbol{u}_n$, a tedy $\mathrm{rank}(A) \le n$. Dohromady tedy platí pro libovolnou matici $A \in \mathbb{R}^{m\times n}$, že $\mathrm{rank}(A) \le \min\{m, n\}$. Matice nabývající toho minima mají *plnou hodnost*, například regulární matice.

Přirozenou otázkou je, jak se hodnost chová vůči maticovým operacím.

> Typicky $\mathrm{rank}(A + B) \ne \mathrm{rank}(A) + \mathrm{rank}(B)$ a $\mathrm{rank}(AB) \ne \mathrm{rank}(A) \cdot \mathrm{rank}(B)$. Žádná přesná formule pro výslednou hodnost obecně neexistuje, protože hodnost závisí na tom, jak se sejdou koeficienty v jednotlivých maticích.

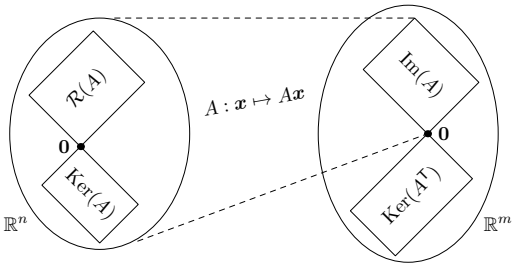
Čtenář může jako cvičení zkusit zkonstruovat dvě matice nenulové hodnosti, že jejich součet, respektive součin má nulovou hodnost. Pokud tedy nejsme schopni určit přesně hodnotu, spokojíme se alespoň s následujícími horními odhady:

Lemma 4.9. *Pro libovolné dvě matice* $A, B \in \mathbb{R}^{m\times n}$ *platí*

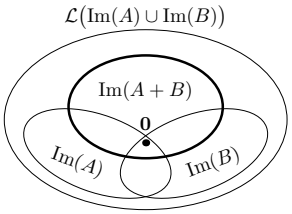
$$\mathrm{rank}(A + B) \le \mathrm{rank}(A) + \mathrm{rank}(B).$$

Důkaz. Mějme $\boldsymbol{c} \in \mathrm{Im}(A + B)$, potom existuje $\boldsymbol{x}$ splňující $(A + B)\boldsymbol{x} = \boldsymbol{c}$. Označme $A\boldsymbol{x} = \boldsymbol{a}$ a $B\boldsymbol{x} = \boldsymbol{b}$, tedy $\boldsymbol{a} \in \mathrm{Im}(A)$ a $\boldsymbol{b} \in \mathrm{Im}(B)$. Klíčové je, že $\boldsymbol{c} = \boldsymbol{a} + \boldsymbol{b}$, a tedy $\mathrm{Im}(A + B)$ lze vygenerovat společně pomocí $\mathrm{Im}(A)$ a $\mathrm{Im}(B)$. Pokud si zvolíme bázi $\boldsymbol{a}_1, \dots, \boldsymbol{a}_k$ pro $\mathrm{Im}(A)$ a bázi $\boldsymbol{b}_1, \dots, \boldsymbol{b}_\ell$ pro $\mathrm{Im}(B)$, je obraz $\mathrm{Im}(A+B)$ podprostor $\mathcal{L}(\boldsymbol{a}_1, \dots, \boldsymbol{a}_k, \boldsymbol{b}_1, \dots, \boldsymbol{b}_\ell) = \mathcal{L}\big(\mathrm{Im}(A) \cup \mathrm{Im}(B)\big)$. Povšimněme si, že množina vektorů $\boldsymbol{a}_1, \dots, \boldsymbol{a}_k, \boldsymbol{b}_1, \dots, \boldsymbol{b}_\ell$ nemusí být lineárně nezávislá a všechno, co vygenerujeme, nemusí být v $\mathrm{Im}(A+B)$. Dostáváme však $\mathrm{rank}(A + B) \le k + \ell = \mathrm{rank}(A) + \mathrm{rank}(B)$. Důkaz je naznačen na obrázku 4.9. □

Poznamenejme, že pokud bychom znali dimenzi $\mathrm{Im}(A) \cap \mathrm{Im}(B)$, šlo by podle cvičení 4.1 horní odhad ještě zpřesnit.



Obrázek 4.8: Matice A reprezentuje lineární zobrazení z $\mathbb{R}^n$ do $\mathbb{R}^m$. Je vyobrazena silná geometrická struktura fundamentálních podprostorů, jejíž část si dokážeme v této kapitole a zbytek v následujících kapitolách.



Obrázek 4.9: Protože $\mathrm{Im}(A+B)$ je podprostor $\mathcal{L}\big(\mathrm{Im}(A) \cup \mathrm{Im}(B)\big)$, platí horní odhad na hodnost.

Lemma 4.10. *Pro libovolné dvě matice $A \in \mathbb{R}^{m\times n}$ a $B \in \mathbb{R}^{n\times p}$ platí*

$$\mathrm{rank}(AB) \leq \min\big\{\mathrm{rank}(A), \mathrm{rank}(B)\big\}.$$

Důkaz. Platí, že $\mathrm{rank}(AB)$ je dimenze $\mathrm{Im}(AB)$. Mějme $\boldsymbol{b} \in \mathrm{Im}(AB)$, potom existuje $\boldsymbol{x}$, že platí $(AB)\boldsymbol{x} = \boldsymbol{b}$. S využitím asociativity však dostáváme, že také existuje $\boldsymbol{y}$ splňující $A(B\boldsymbol{x}) = A\boldsymbol{y} = \boldsymbol{b}$. Speciálně je tedy $\mathrm{Im}(AB)$ podprostor $\mathrm{Im}(A)$, a proto platí první nerovnost $\mathrm{rank}(AB) \leq \mathrm{rank}(A)$.

Pro druhou nerovnost uvažme bázi $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k$ podprostoru $\mathrm{Im}(AB)$. Jak uvedeno výše, podle definice existují $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k$, že $A\boldsymbol{y}_i = \boldsymbol{b}_i$. Klíčové je, že tyto vektory také musí být lineárně nezávislé, a čtenář to může zkusit dokázat. Protože $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k \in \mathrm{Im}(B)$, platí, že $\mathrm{rank}(AB) = k \leq \mathrm{rank}(B)$. □

Poznamenejme, že existují i dolní odhady, které však nebudeme přímo potřebovat a čtenář si je může dokázat jako cvičení 4.7.

Hodnost a regularita. V kapitole 3 jsme se zabývali podmínkami, za kterých existuje inverzní matice. Pravá inverze existuje podle definice, právě když n soustav $A\boldsymbol{x}_i = \boldsymbol{e}_i$ lze vyřešit zároveň. V řeči fundamentálních podprostorů to není nic jiného než $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n \in \mathrm{Im}(A)$. Pokud však $\mathrm{Im}(A)$ obsahuje tyto vektory, musí také obsahovat jejich lineární obal. Protože se jedná o kanonickou bázi $\mathbb{R}^n$, je lineární obal roven celému $\mathbb{R}^n$. Pravá inverze tedy existuje, právě když $\mathrm{Im}(A) = \mathbb{R}^n$. V řeči hodnosti má matice $A \in \mathbb{R}^{m\times n}$ pravou inverzi, právě když je její hodnost rovna počtu sloupců n. Čtvercová matice je regulární, právě když má plnou hodnost.

To, že $\mathrm{Im}(A)$ musí být celé $\mathbb{R}^n$, jsme už ostatně dokázali v tvrzení 3.4. To říká, že matice A má pravou inverzi právě tehdy, když soustava $A\boldsymbol{x} = \boldsymbol{b}$ má řešení pro každou pravou stranu $\boldsymbol{b}$. Důkaz jsme v kapitole 3 provedli bez použití lineárních obalů a bází, protože jsme tyto pojmy ještě neměli zavedené. Toto není jediný důkaz, který by šel s nově zavedenými pojmy zjednodušit.

Podobně levá inverze existuje, právě když $\mathrm{rank}(A) = m$; to ale zatím neumíme dokázat. Víme pouze, že matice A má levou inverzi, právě když matice A^T má pravou inverzi. Ta existuje, právě když $\mathrm{rank}(A^\mathsf{T}) = m$. K dokončení důkazu budeme muset ukázat následující větu:

Věta 4.11. *Pro libovolnou matici $A \in \mathbb{R}^{m\times n}$ platí*

$$\mathrm{rank}(A) = \mathrm{rank}(A^\mathsf{T}).$$

Věta 4.11 zobecňuje větu 3.7, že pro čtvercovou matici existuje levá inverze, právě když existuje inverze pravá. Pokud čtvercová matice $A \in \mathbb{R}^{n\times n}$ má pravou inverzi, potom $\mathrm{rank}(A) = n$. Tedy podle věty 4.11 je $\mathrm{rank}(A^\mathsf{T}) = n$ a matice A^T má pravou inverzi, což je levá inverze matice A. Druhá implikace se dokáže identicky. Připomeňme si, že hodnost matice udává, jak moc je daná matice blízká regulární. Věta tedy říká, že matice A a A^T jsou stejně blízko regularitě. V textu ukážeme několik důkazů, již v této kapitole dva různé.

Pro důkaz nejprve ukažme klíčovou vlastnost regulárních matic. Podle lemmatu 4.10 platí nerovnost $\mathrm{rank}(AB) \leq \min\big\{\mathrm{rank}(A), \mathrm{rank}(B)\big\}$. Zajímá nás, za jakých podmínek nastane rovnost. Obecně něco takového charakterizovat není snadné; teorie lineární algebry k tomu sice dává nástroje, ale to je pro naše účely zbytečně komplikované. Pokud však jedna z matic je regulární, lépe řečeno postačuje existence inverze ze správné strany, vždy nastane rovnost. Tedy násobení regulární maticí nemění hodnost.

Lemma 4.12. *Nechť $A \in \mathbb{R}^{m\times n}$ je libovolná matice. Nechť $X \in \mathbb{R}^{p\times m}$ má levou inverzi, a nechť $Y \in \mathbb{R}^{n\times q}$ má pravou inverzi. Potom platí*

$$\mathrm{rank}(XA) = \mathrm{rank}(A) = \mathrm{rank}(AY).$$

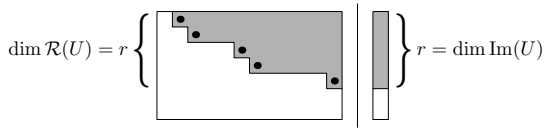
Důkaz. Ukážeme pouze v případě matice X. Pokud vynásobíme A zleva, podle lemmatu 4.10 nemůže hodnost vzrůst. Protože však vynásobení X zleva můžeme invertovat, nemůže se hodnost ani snížit:

$$\mathrm{rank}(A) \geq \mathrm{rank}(XA) \geq \mathrm{rank}(X^{-1}XA) = \mathrm{rank}(A).$$ □

> Násobení regulární maticí nemění hodnost. Platí i pro násobení obdélníkovou maticí, která má inverzi ze správné strany.

Poznamenejme, že znění lemmatu lze i otočit. Dostáváme tak alternativní trochu zvláštní definici regulární matice, více ve cvičení 4.2. Následuje první důkaz věty.

Důkaz věty 4.11. Důkaz provedeme podobně jako v případě vety 3.7, s využitím LU dekompozice. Ta říká, že pro každou matici A existuje rozklad $PA = LU$, kde P a L jsou regulární matice a U je horní trojúhelníková matice v odstupňovaném tvaru. Podle lemmatu 4.12 platí, že $\mathrm{rank}(A) = \mathrm{rank}(U)$. Podobně platí $A^\mathsf{T}P^\mathsf{T} = U^\mathsf{T}L^\mathsf{T}$, a tedy $\mathrm{rank}(A^\mathsf{T}) = \mathrm{rank}(U^\mathsf{T})$. K dokončení stačí ukázat, že $\mathrm{rank}(U) = \mathrm{rank}(U^\mathsf{T})$. Situace je naznačena na obrázku 4.10.



Obrázek 4.10: Obraz $\mathrm{Im}(U)$ obsahuje všechny vektory, které mají nulové koeficienty $b_{r+1}, \ldots, b_m$. Prvních r nenulových řádků tvoří bázi řádkového podprostoru $\mathcal{R}(U)$.

Mějme matici U v odstupňovaném tvaru, kde má r nenulových řádků a $m - r$ nulových řádků. Obraz $\mathrm{Im}(U)$ je množina všech pravých stran $\boldsymbol{b}$, pro které má soustava $U\boldsymbol{x} = \boldsymbol{b}$ řešení. Soustava má řešení právě tehdy, když jsou koeficienty $b_{r+1}, \ldots, b_m$ nulové. Ostatní koeficienty můžeme zvolit libovolně, a tedy $\dim \mathrm{Im}(U) = \mathrm{rank}(U) = r$.

Pro $\mathrm{rank}(U^\mathsf{T})$ platí, že je roven dimenzi řádkového prostoru $\mathcal{R}(U)$. Protože každý z prvních r nenulových řádků obsahuje pivot na jiném místě, jsou tyto řádkové vektory lineárně nezávislé. A zjevně generují $\mathcal{R}(U)$, tedy $\dim \mathcal{R}(U) = \mathrm{rank}(U^\mathsf{T})$ je také rovna r. Tedy $\mathrm{rank}(U) = \mathrm{rank}(U^\mathsf{T})$, jak jsme potřebovali. □

Jako důsledek dostáváme, že levá inverze existuje pro matici $A \in \mathbb{R}^{m\times n}$, právě když $\mathrm{rank}(A) = m$. Hodnost matice lépe osvětluje výsledky získané v kapitole 3.

Dimenze fundamentálních podprostorů. Již jsme se zabývali hodností, která v řeči fundamentálních podprostorů říká, jak velký je obraz matice $\mathrm{Im}(A)$. Nyní se budeme zabývat i ostatními fundamentálními podprostory a jejich velikostmi. Také si ukážeme, jak se fundamentální podprostory mění při aplikování maticového násobení.

Uvažme například větu 4.11 přeformulovanou v řeči fundamentálních podprostorů. Ta říká klíčovou vlastnost, že dimenze řádkového prostoru $\mathcal{R}(A)$ a dimenze sloupcového podprostoru/obrazu $\mathrm{Im}(A)$ je stejná. Důvod je ten, že $\mathcal{R}(A) = \mathrm{Im}(A^\mathsf{T})$ a dimenze $\mathrm{Im}(A)$ a $\mathrm{Im}(A^\mathsf{T})$, což jsou hodnosti A a A^T, se podle věty rovnají. Tuto vlastnost jsme záměrně vyobrazili na obrázku 4.8 pomocí stejně velkých obdélníků reprezentujících $\mathcal{R}(A)$ a $\mathrm{Im}(A)$.

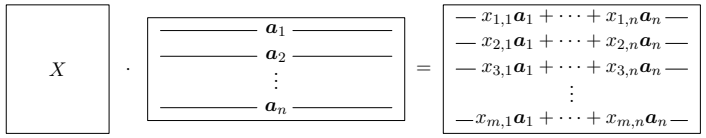
Pro libovolnou matici $A \in \mathbb{R}^{m\times n}$ platí

$$\dim \mathcal{R}(A) = \mathrm{rank}(A^T) = \mathrm{rank}(A) = \dim \mathrm{Im}(A).$$

Mějme libovolnou matici A a vynásobme ji nějakou maticí X zleva. Uvažme, jak se změní řádkový podprostor a jádro. Jádro $\mathrm{Ker}(XA)$ je množina všech řešení $XA\boldsymbol{x} = 0$. Tedy pokud $\boldsymbol{x} \in \mathrm{Ker}(A)$, také $\boldsymbol{x} \in \mathrm{Ker}(XA)$. Ohledně řádkového podprostoru, podle definice maticového násobení je z každý řádků XA lineární kombinaci řádků A, jak je naznačeno na obrázku 4.11. Tedy naopak $\mathcal{R}(XA)$ je podprostor $\mathcal{R}(A)$. Dostáváme tedy:

$$\mathrm{Ker}(A) \subseteq \mathrm{Ker}(XA) \qquad \text{a} \qquad \mathcal{R}(A) \supseteq \mathcal{R}(XA). \tag{4.5}$$

Při násobení X zprava něco podobného neplatí. Předně $\mathrm{Ker}(AX)$ a $\mathcal{R}(AX)$ mohou být podprostory jiného prostoru než $\mathrm{Ker}(A)$ a $\mathcal{R}(A)$. I když matice X je čtvercová, mohou být tyto podprostory zcela odlišné.



Obrázek 4.11: Řádkové vektory XA jsou lineární kombinace řádkových vektorů A, kde koeficienty těchto kombinací jsou řádkové vektory X.

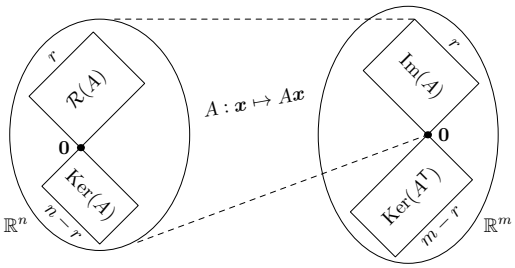
Obdobné vztahy však platí pro $\mathrm{Ker}(A^\mathsf{T})$ a $\mathrm{Im}(A)$ při násobení X zprava. Tedy $\mathrm{Ker}(A^\mathsf{T})$ se může zvětšit a $\mathrm{Im}(A)$ se může pouze zmenšit. Násobení X zleva může fundamentální podprostory $\mathrm{Ker}(A^\mathsf{T})$ a $\mathrm{Im}(A)$ libovolně změnit.

Podobně jako v důkazu lemmatu 4.12, pokud matice X má levou inverzi, dostáváme v (4.5) rovnosti mezi podprostory. Pokud by se totiž například jádro $\mathrm{Ker}(XA)$ zvětšilo, muselo by se jádro $\mathrm{Ker}(X^{-1}XA)$ opět zmenšit na $\mathrm{Ker}(A)$, což podle (4.5) není možné.

Nyní už víme dost na to, abychom určili dimenzi jádra.

Tvrzení 4.13. *Nechť $A \in \mathbb{R}^{m\times n}$ je libovolná matice. Potom* $\dim \mathcal{R}(A) + \dim \mathrm{Ker}(A) = n$.

Důkaz. Nejprve ukažme, že to platí pro matici U v odstupňovaném tvaru. Zde je $\dim \mathcal{R}(U)$ rovna počtu nenulových řádků, což je také počet pivotů a $\mathrm{rank}(U)$. Dimenze jádra je naopak počet sloupců, v kterých se nevyskytuje pivot, protože hodnoty těchto proměnných můžeme libovolně zvolit. Pro libovolnou z těchto voleb už je hodnota proměnných ve sloupcích s pivoty jednoznačně určena, a tedy $\dim \mathrm{Ker}(U) = n - \dim \mathcal{R}(U)$. Pro libovolnou matici v odstupňovaném tvaru tvrzení platí.



Obrázek 4.12: Fundamentální podprostory z obrázku 4.8 s doplněnými dimenzemi podprostorů.

Nechť A je libovolná matice a uvažme její LU dekompozici $PA = LU$. Platí, že $A = (P^\mathsf{T}L)U$, tedy A vznikne z U vynásobením zleva maticí, která má levou inverzi. Tedy podle výše uvedeného platí $\mathcal{R}(A) = \mathcal{R}(U)$ a $\mathrm{Ker}(A) = \mathrm{Ker}(U)$, speciálně platí výše dokázané velikosti jejich dimenzí. □

Poznamenejme, že tento výsledek dává smysl vzhledem k tomu, jak funguje Gaussova eliminace. Ta provádí regulární úpravy, které odpovídají násobení matice A regulárními maticemi úprav R zleva. Tím neměníme množinu řešení, která je afinní podprostor vzniklý posunutím $\mathrm{Ker}(A)$. Také neměníme to, co lze vygenerovat z řádkových vektorů matice, tedy řádkový podprostor $\mathcal{R}(A)$.

Aplikováním tvrzení 4.13 na A^T dostáváme, že součet dimenzí obrazu a levého jádra je vždy roven m. Násobení zleva regulární maticí typicky mění tyto dva fundamentální podprostory, jejich velikosti se však podle věty 4.11 nemění. Dostáváme tedy následující fundamentální fakt lineární algebry:

Nechť $A \in \mathbb{R}^{m\times n}$ je libovolná matice a nechť $r = \mathrm{rank}(A)$. Potom:

$$\dim \mathcal{R}(A) = \dim \mathrm{Im}(A) = r, \qquad \dim \mathrm{Ker}(A) = n - r, \qquad \dim \mathrm{Ker}(A^\mathsf{T}) = m - r.$$

Nezávislost fundamentálních podprostorů. Zaměřme se pouze na řádkový podprostor a jádro, protože pro obraz a levý kernel dostaneme analogický výsledek aplikováním na A^T. Víme, že $\mathcal{R}(A)$ a $\mathrm{Ker}(A)$ jsou podprostory $\mathbb{R}^n$ dimenze r a $n - r$. Dokážeme si, že není náhoda, že se tyto dimenze přesně sečtou na n, což je $\dim \mathbb{R}^n$; řádkový prostor a jádro jsou totiž lineárně nezávislé a společně generují $\mathbb{R}^n$.

Tvrzení 4.14. *Pro libovolnou matici A platí* $\mathcal{R}(A) \cap \mathrm{Ker}(A) = \{\mathbf{0}\}$.

Důkaz. Nechť $\boldsymbol{x} \in \mathcal{R}(A) \cap \mathrm{Ker}(A)$, chceme ukázat, že $\boldsymbol{x} = \mathbf{0}$. Protože $\boldsymbol{x}$ leží v jádru, platí $A\boldsymbol{x} = \mathbf{0}$. A protože náleží do řádkového podprostoru, což je $\mathrm{Im}(A^\mathsf{T})$, existuje $\boldsymbol{y}$, pro které $A^\mathsf{T}\boldsymbol{y} = \boldsymbol{x}$. Složením těchto dvou rovností dostáváme $AA^\mathsf{T}\boldsymbol{y} = \mathbf{0}$, tedy $\boldsymbol{y} \in \mathrm{Ker}(AA^\mathsf{T})$. Ukážeme, že $\boldsymbol{y} \in \mathrm{Ker}(A^\mathsf{T})$, a tedy vyjde $A^\mathsf{T}\boldsymbol{y} = \boldsymbol{x} = \mathbf{0}$.

Pro libovolnou matici A totiž platí $\mathrm{Ker}(AA^\mathsf{T}) = \mathrm{Ker}(A^\mathsf{T})$, což nyní dokážeme. Inkluzi $\mathrm{Ker}(AA^\mathsf{T}) \supseteq \mathrm{Ker}(A^\mathsf{T})$ jsme už argumentovali výše, neboť platí při libovolném násobení maticí zleva. Pro druhou inkluzi však nepotřebujeme existenci levé inverze matice A, platí v případě AA^T obecně. Uvažme $AA^\mathsf{T}\boldsymbol{y} = \mathbf{0}$. Vynásobením zleva $\boldsymbol{y}^\mathsf{T}$ zachováme rovnost a uděláme výraz více symetrický:

$$\boldsymbol{y}^\mathsf{T}\mathbf{0} = \boldsymbol{y}^\mathsf{T}AA^\mathsf{T}\boldsymbol{y} = (A^\mathsf{T}\boldsymbol{y})^\mathsf{T}(A^\mathsf{T}\boldsymbol{y}) = \boldsymbol{z}^\mathsf{T}\boldsymbol{z}.$$

Oba součiny $\boldsymbol{y}^\mathsf{T}\mathbf{0}$ a $\boldsymbol{z}^\mathsf{T}\boldsymbol{z}$ jsou matice 1×1 a budeme na ně nahlížet jako na reálná čísla.

Zjevně $\boldsymbol{y}^\mathsf{T}\boldsymbol{0} = 0$. Pro vektor $\boldsymbol{z} = (z_1, \ldots, z_n)$ dostáváme po rozepsání podle definice maticového násobení

$$\boldsymbol{z}^\mathsf{T}\boldsymbol{z} = \sum_{i=1}^{n} z_i^2 = z_1^2 + \cdots + z_n^2. \tag{4.6}$$

Pro reálná čísla platí $z_i^2 \geq 0$ a rovnost nastává pouze pro nulu. Tedy součet (4.6) může být roven nule pouze tehdy, když $A^\mathsf{T}\boldsymbol{y} = \boldsymbol{z} = \boldsymbol{0}$. Proto $\boldsymbol{y} \in \mathrm{Ker}(A^\mathsf{T})$ a dostáváme druhou inkluzi $\mathrm{Ker}(AA^\mathsf{T}) \subseteq \mathrm{Ker}(A^\mathsf{T})$. □

Aplikováním na A^T dostaneme, že také $\mathrm{Im}(A) \cap \mathrm{Ker}(A^\mathsf{T}) = \{\boldsymbol{0}\}$. Proto se fundamentální podprostory protínají přesně tak, jak je to naznačeno na obrázku 4.12. Doplňme si tvrzení několika důležitými poznámkami.

Předně důkaz může působit trikově, neboť není jasné, jak jsme přišli na to výraz $AA^\mathsf{T}\boldsymbol{y} = \boldsymbol{0}$ zleva vynásobit $\boldsymbol{y}^\mathsf{T}$. Jedná se však o hlubší geometrickou souvislost, kterou lépe osvětlíme v kapitole **??** pomocí skalárního součinu. Důvod je, že toto tvrzení přirozeně nepatří do této kapitoly a mělo by spíš být obsaženo později v textu. Prozatím se tedy spokojíme s tím, že jsme důkaz takto „vyhaluzili“.

Mimochodem objevená rovnost mezi jádry A^T a AA^T není jediná společná vlastnost těchto matic. Čtenář může zkusit dokázat, že se shodují i jejich řádkové podprostory. Aplikováním na A^T dostáváme také $\mathrm{Ker}(A^\mathsf{T}A) = \mathrm{Ker}(A)$. Obecně matice $A^\mathsf{T}A$ a AA^T mají řadu společných vlastností s A a A^T vzhledem k fundamentálním podprostorům. Navíc obě matice $A^\mathsf{T}A$ a AA^T jsou symetrické a mají další silné vlastnosti. Tedy v řadě situací je možné uvažovat tyto matice místo původních A a A^T a zjednodušit si řešené problémy. Poznamenejme, že druhý způsob zesymetričtění čtvercové matice je $\frac{1}{2}(A + A^\mathsf{T})$.

Důkaz tvrzení 4.14 je naprosto závislý na vlastnosti reálných čísel, že $x^2 \geq 0$ a rovnost nastává přesně pro $x = 0$. V kapitole 6 zavedeme zobecnění reálných čísel zvané *algebraické těleso*, nad kterými lze podobně vybudovat vektorové prostory a celou lineární algebru. V některých tělesech vlastnost $x^2 \geq 0$ neplatí, například protože symboly $\geq$ a $>$ vůbec nemají smysl. Potom $\mathrm{Ker}(AA^\mathsf{T})$ může být striktně větší než $\mathrm{Ker}(A^\mathsf{T})$ a může existovat nenulový vektor $\boldsymbol{x}$ v průniku $\mathcal{R}(A)$ a $\mathrm{Ker}(A)$.

Přímý součet. Dva podprostory U a V nazýváme *lineárně nezávislé* nebo zkráceně *nezávislé*, pokud $U \cap V = \{\boldsymbol{0}\}$. Lineární obal sjednocení dvou nezávislých podprostorů se nazývá *přímý součet* U a V a používá se následující značení $U \oplus V = \mathcal{L}(U \cup V)$. Čtenář si může zkusit ve cvičení 4.3 dokázat, že $\dim U + \dim V = \dim(U \oplus V)$. Navíc pro libovolný vektor $\boldsymbol{x} \in U \oplus V$ existují jednoznačně určené vektory $\boldsymbol{x}_u \in U$ a $\boldsymbol{x}_v \in V$, že $\boldsymbol{x}_u + \boldsymbol{x}_v = \boldsymbol{x}$. Libovolný vektor $\boldsymbol{x} \in U \oplus V$ lze zapsat jako dvojici $(\boldsymbol{x}_u, \boldsymbol{x}_v)$, což vede na alternativní definici přímého součet, více ve cvičení 4.4.

Tvrzení 4.14 má následující alternativní interpretaci. Pro libovolnou matici jsou podprostory $\mathcal{R}(A)$ a $\mathrm{Ker}(A)$ lineárně nezávislé. Protože $\dim \mathcal{R}(A) + \dim \mathrm{Ker}(A) = n$, generují tyto dva nezávislé podprostory dohromady celé $\mathbb{R}^n$, tedy

$$\mathcal{R}(A) \oplus \mathrm{Ker}(A) = \mathbb{R}^n.$$

Libovolné $\boldsymbol{x} \in \mathbb{R}^n$ lze rozložit na složku $\boldsymbol{x}_r$ v řádkovém podprostoru a $\boldsymbol{x}_k$ v jádru, že $\boldsymbol{x} = \boldsymbol{x}_r + \boldsymbol{x}_k$. V příští kapitole si ukážeme, že lineární zobrazení $\boldsymbol{x} \mapsto A\boldsymbol{x}$ se chová velice pěkně vůči těmto složkám. Spolu s dimenzemi fundamentálních podprostorů jsou toto dvě části fundamentální věty lineární algebry, kterou si postupně budeme odhalovat.

Pro libovolnou matici $A \in \mathbb{R}^{m\times n}$ platí, že $\mathcal{R}(A)$ a $\mathrm{Ker}(A)$ jsou lineárně nezávislé a podobně $\mathrm{Im}(A)$ a $\mathrm{Ker}(A^\mathsf{T})$ jsou lineárně nezávislé. Navíc

$$\mathcal{R}(A) \oplus \mathrm{Ker}(A) = \mathbb{R}^n \qquad \text{a} \qquad \mathrm{Im}(A) \oplus \mathrm{Ker}(A^\mathsf{T}) = \mathbb{R}^m.$$

Důvodem je, že $\mathrm{Ker}(AA^\mathsf{T}) = \mathrm{Ker}(A^\mathsf{T})$ a $\mathrm{Ker}(A^\mathsf{T}A) = \mathrm{Ker}(A)$.

⋆**Matice hodnosti jedna.** Uvažme matici $A \in \mathbb{R}^{m\times n}$, která má hodnost jedna. Ukážeme si, že všechny tyto matice mají velice jednoduchou algebraickou strukturu. Může to působit jako zábavná hříčka, ale nakonec si ukážeme, že z toho vyplývají zajímavé vlastnosti. Připomeňme, že hodnost matice je dimenze řádkového podprostoru a současně dimenze obrazu. V následujícím textu budeme hodně pracovat se sloupci matice, a proto budeme obraz alternativně nazývat jako sloupcový podprostor.

Ukažme si nejprve příklad matice hodnosti jedna:

$$A = \begin{pmatrix} 1 & -2 & 3 \\ 2 & -4 & 6 \\ -3 & 6 & -9 \end{pmatrix}.$$

Můžeme si všimnout, že řádky a sloupce se hodně opakují. Důvodem je, že jak řádkový, tak sloupcový podprostor mají dimenzi jedna. Dají se tedy vygenerovat pomocí jediného vektoru.

Zaměřme se na sloupcový podprostor. Zvolme libovolný vektor $\boldsymbol{x} \in \mathrm{Im}(A)$, všechny sloupce jsou jeho lineární kombinace, což jsou násobky $\alpha\boldsymbol{x}$. Pokud $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ jsou sloupce matice A, existují reálná čísla $y_1, \ldots, y_m$, že $\boldsymbol{x}_i = y_i\boldsymbol{x}$. Pokud uspořádáme tyto reálná čísla do sloupcového vektoru $\boldsymbol{y} = (y_1, \ldots, y_m)$, dostáváme v řeči maticového násobení elegantní zápis $A = \boldsymbol{x}\boldsymbol{y}^\mathsf{T}$. Povšimněme si také, že $\boldsymbol{y} \in \mathcal{R}(A)$, tedy že jsme mohli postupovat obráceně, zvolit vektor $\boldsymbol{y}$ z řádkového podprostoru a získat vektor $\boldsymbol{x}$ jako vektor koeficientů lineární kombinace.

V případě výše uvedeného příkladu můžeme například jako $\boldsymbol{y}$ použít první sloupec a dostáváme

$$A = \begin{pmatrix} 1 & -2 & 3 \\ 2 & -4 & 6 \\ -3 & 6 & -9 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix} \cdot \begin{pmatrix} 1 & -2 & 3 \end{pmatrix} = \boldsymbol{x}\boldsymbol{y}^\mathsf{T}.$$

Následující lemma elegantně charakterizuje matice hodnosti jedna.

Lemma 4.15. *Matice $A \in \mathbb{R}^{m\times n}$ má hodnost jedna, právě když existují nenulové vektory $\boldsymbol{x} \in \mathbb{R}^m$ a $\boldsymbol{y} \in \mathbb{R}^n$, pro které platí*

$$A = \boldsymbol{x}\boldsymbol{y}^\mathsf{T}. \tag{4.7}$$

Navíc vždy platí $\boldsymbol{x} \in \mathrm{Im}(A)$ a $\boldsymbol{y} \in \mathcal{R}(A)$.

Důkaz. Již jsme ukázali, že pro matici A hodnosti jedna lze zvolit řekněme $\boldsymbol{x} \in \mathrm{Im}(A)$ a $\boldsymbol{y} \in \mathbb{R}^n$, že platí $A = \boldsymbol{x}\boldsymbol{y}^\mathsf{T}$. Protože je matice A nenulová, určitě jsou oba vektory nenulové. Zbývá ukázat obrácenou implikaci.

Předpokládejme, že existují nenulové vektory $\boldsymbol{x} = (x_1, \ldots, x_m)$ a $\boldsymbol{y} = (y_1, \ldots, y_n)$, pro které platí $A = \boldsymbol{x}\boldsymbol{y}^\mathsf{T}$. Matice A je nenulová, tedy její hodnost je alespoň jedna. Podle definice maticového násobení je i-tý řádek A roven $x_i\boldsymbol{y}^\mathsf{T}$, tedy dimenze řádkového podprostoru je jedna. Podobně je j-tý sloupec A roven $y_j\boldsymbol{x}$, tedy i dimenze sloupcového podprostoru je rovna jedné. Proto je hodnost A rovna jedné a platí $\boldsymbol{x} \in \mathrm{Im}(A)$ a $\boldsymbol{y} \in \mathcal{R}(A)$.

Alternativně jsme mohli k dokázání jedničkové hodnosti použít lemma 4.10 o hodnosti součinu matic. Přímo by z toho však nevyplynulo, že $\boldsymbol{x} \in \mathrm{Im}(A)$ a $\boldsymbol{y} \in \mathcal{R}(A)$. □

Poznamenejme, že součin $\boldsymbol{x}\boldsymbol{y}^\mathsf{T}$ je zcela rozdílný od součinu $\boldsymbol{x}^\mathsf{T}\boldsymbol{y}$ z důkazu tvrzení 4.14. Ten první se nazývá *tenzorový součin* nebo také *vnější součin* a jeho výsledkem je matice $m \times n$. Pro druhý součin, zvaný *skalární* nebo *vnitřní*, musí být oba vektory stejné délky a výsledkem je matice 1×1, alternativně reálné číslo. Oba součiny mají klíčové místo v lineární algebře a budeme se jimi později zabývat.

Přirozená otázka je, jestli by lemma 4.15 nešlo zobecnit na vyšší hodnost, třeba dva nebo obecně k. Následující tvrzení je takové zobecnění:

Tvrzení 4.16. *Nechť $A \in \mathbb{R}^{m\times n}$ je libovolná matice hodnosti k. Potom existují vektory $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \in \mathrm{Im}(A)$ a $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k \in \mathcal{R}(A)$, že*

$$A = \boldsymbol{x}_1\boldsymbol{y}_1^\mathsf{T} + \cdots \boldsymbol{x}_k\boldsymbol{y}_k^\mathsf{T}. \tag{4.8}$$

Důkaz. V důkazu můžeme postupovat úplně stejně jako v důkazu lemmatu 4.15. Pokud má matice hodnost k, má sloupcový prostor dimenzi přesně k a můžeme zvolit libovolně jeho bázi $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \in \mathrm{Im}(A)$. Platí, že i-tý sloupec matice A je lineární kombinace $y_{1,i}\boldsymbol{x}_1 + \cdots + y_{k,i}\boldsymbol{x}_k$. Zapsáním koeficientů $y_{1,i}$ do vektoru $\boldsymbol{y}_1$, koeficientů $y_{2,i}$ do $\boldsymbol{y}_2$, a tak dále až $y_{k,i}$ do $\boldsymbol{y}_k$ dostáváme slibovanou rovnost (4.8). Je snadné nahlédnout, že $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k \in \mathcal{R}(A)$. Dokonce generují řádkový podprostor, neboť libovolný řádek matice A z nich lze vygenerovat. □

V důkazu jsme pochopitelně mohli postupovat obráceně, tedy zvolit si k-prvkovou bázi $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k$ řádkového podprostoru a zkonstruovat příslušné vektory $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ generující sloupcový podprostoru. Z toho dostáváme alternativní důkaz věty 4.11, že dimenze řádkového a sloupcového podprostoru je pro libovolnou matici stejná. Z důkazu tvrzení totiž vyplývá, že $\dim \mathrm{Im}(A) \geq \dim \mathcal{R}(A)$, a z alternativního postupu vyplývá obrácená nerovnost $\dim \mathrm{Im}(A) \leq \dim \mathcal{R}(A)$. Obešli jsme se tedy úplně bez odstupňovaného tvaru a LU dekompozice.

Tvrzení 4.16 má alternativní znění. Matici hodnosti k lze zapsat jako součet k matic A_i hodnosti jedna, kde $A_i = \boldsymbol{x}_i\boldsymbol{y}_i^\mathsf{T}$. Podle lemmatu 4.9 víme, že tento počet je nejlepší možný, neboť

$$\mathrm{rank}(A_1 + \cdots + A_k) \leq \mathrm{rank}(A_1) + \cdots + \mathrm{rank}(A_k) = k.$$

Cvičení 4.5 je zesílení tvrzení ve stylu lemmatu 4.15. Cvičení 4.6 se zabývá tím, jak vektory $\boldsymbol{x}_i$ a $\boldsymbol{y}_i$ nalézt algoritmicky.

Každá matice A hodnosti k je rovna součtu k matic hodnosti jedna:

$$A = A_1 + \cdots + A_k = \boldsymbol{x}_1\boldsymbol{y}_1^\mathsf{T} + \cdots + \boldsymbol{x}_k\boldsymbol{y}_k^\mathsf{T}.$$

⋆**Komprese obrázků.** Možná to může působit překvapivě, ale tvrzení 4.16 je výrazně hlubší než se může zdát a má praktické aplikace. Lze ho totiž použít ke kompresi dat uložených ve formě matice malé hodnosti k. Při standardním uložení matice $A \in \mathbb{R}^{m\times n}$ potřebujeme uložit mn čísel. Místo toho však můžeme pouze uložit k dvojic vektorů $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ a matici zrekonstruovat podle (4.8). Pro uložení těchto vektorů si stačí pamatovat $k(m+n)$ čísel, což může být v případě malé hodnosti k výrazná úspora.

Tento postup lze například použít při ukládání obrázků. Při rozlišení $m \times n$ můžeme obrázek reprezentovat jako matici $A \in \mathbb{R}^{m\times n}$, kde koeficient $a_{i,j}$ kóduje barvu pixelu na pozici (i, j), řekněme jako číslo v intervalu $[0, 1]$ reprezentující odstín šedi. Pokud má matice A malou hodnost k, dokážeme obrázek uložit v mnohem menší paměti než mn.

Potíž s tímto postupem je, že typická matice A bude mít hodnost blízkou plné hodnosti $\min\{m, n\}$, a tedy dokonce $k(n+m) > nm$ a žádné úspory nedosáhneme. Lze však matice $A_1, \ldots, A_k$ zvolit tak, že A_1 obsahuje dominantní hodnoty z A, A_2 druhé nejdůležitější, až A_k ty nejméně důležité. Potom můžeme použít kompresi, při které zapíšeme pouze prvních ℓ nejdůležitějších matic $A_1, \ldots, A_\ell$. Pochopitelně $A_1 + \cdots + A_\ell$ není přesně rovno matici A, ale můžeme se s výsledkem dostat hodně blízko a významně ušetřit paměť.

Obrázek 4.13 obsahuje příklad použití této komprese. Samozřejmě reálně používané postupy v kompresi obrázků jsou výrazně složitější a dosahují mnohem lepších výsledků. Například nejpoužívanější metoda komprese fotek JPEG je založená na jiném principu a je výrazně efektivnější. Tato komprese je tedy spíše ilustrací tvrzení 4.16. Přesto můžeme dosáhnout zejména na větších obrázcích slušné úspory.

Všimněte si, že jsme dosud neřekli, co to znamená, že matice A_i obsahuje i-té nejdůležitější hodnoty z matice A. Také nevíme, jak tyto speciální matice $A_1, \ldots, A_k$ najít. Oběma nedostatkům se zatím nebudeme věnovat, protože vyžadují výrazně pokročilejší znalost lineární algebry. Souvisí totiž se singulárním rozkladem matice (SVD), což je jeden z poměrně nedávných objevů lineární algebry. Tento rozklad poví o matici A prakticky všechno a má celou řadu úžasných aplikací například v analýze dat, signálu, statistice. Na druhou stranu není úplně levné a snadné ho spočítat.

Shrnutí

V této kapitole jsme popsali řadu klíčových pojmů lineární algebry a jedná se o dosud nejhutnější část textu. Získali jsem výrazně lepší náhled do struktury vektorových podprostorů. Ukázali jsme, že lineární kombinace jsou ekvivalentní libovolné posloupnosti vektorových operací, a tedy pro uzavřenost na vektorové operace stačí studovat uzavřenost na lineární kombinace. Jako elegantní popis vektorového podprostoru můžeme zvolit ideálně malou množinu vektorů tak, že podprostor je roven množině všech jejich lineárních kombinací. Tím dostáváme alternativní popis lineárního obalu jako nejmenšího vektorového podprostoru obsahující danou množinu vektorů.

Chceme, aby množina vektorů popisující vektorový podprostor byla co nejmenší. Přirozeně jsme zavedli lineární nezávislost, která říká, že množina neobsahuje žádné nadbytečné vektory. Lineárně nezávislá množina generující celý prostor se nazývá báze, lze uvažovat i báze pro podprostory. Protože se každý vektor dá vyjádřit jednoznačně jako lineární kombinace bazických vektorů, zavádí báze souřadný systém nad vektorový prostorem. Steinitzova věta umožňuje rozšiřovat nezávislé množiny na báze a popisuje velice bohatou strukturu všech bází. Klíčový důsledek Steinitzovy věty je, že každá báze vektorového podprostoru má stejnou velikost a tato velikost se nazývá dimenze. Navíc věta o izomorfismu říká, že vektorový prostor dimenze n má stejnou algebraickou strukturu jako $\mathbb{R}^n$.

Získané poznatky jsme dále aplikovali na čtyři fundamentální podprostory definované maticí: obraz (neboli sloupcový podprostor), jádro, řádkový podprostor a levý kernel. Navíc jsme zavedli klíčovou definici hodnosti matice, která udává, jak moc je matice blízká regulární. Hodnost matice $\mathrm{rank}(A)$ má několik možných ekvivalentních definic, jak jsme v textu ukázali:

- Hodnost je dimenze obrazu, neboli počet lineárně nezávislých sloupců matice.
- Hodnost je dimenze řádkového podprostoru, neboli počet lineárně nezávislých řádků.
- Hodnost je počet nenulových řádků v odstupňovaném tvaru matice.
- Hodnost je nejmenší počet matic jedničkové hodnosti, které se sečtou na matici A. (Matici hodnosti jedna lze zadefinovat i bez hodnosti jako $\boldsymbol{x}\boldsymbol{y}^\mathsf{T}$ pro nenulové vektory $\boldsymbol{x}$ a $\boldsymbol{y}$.)

Ohledně fundamentálních podprostorů jsme ukázali první dvě části fundamentální věty lineární algebry. Zatím víme, že pro matici $A \in \mathbb{R}^{m\times n}$ platí:

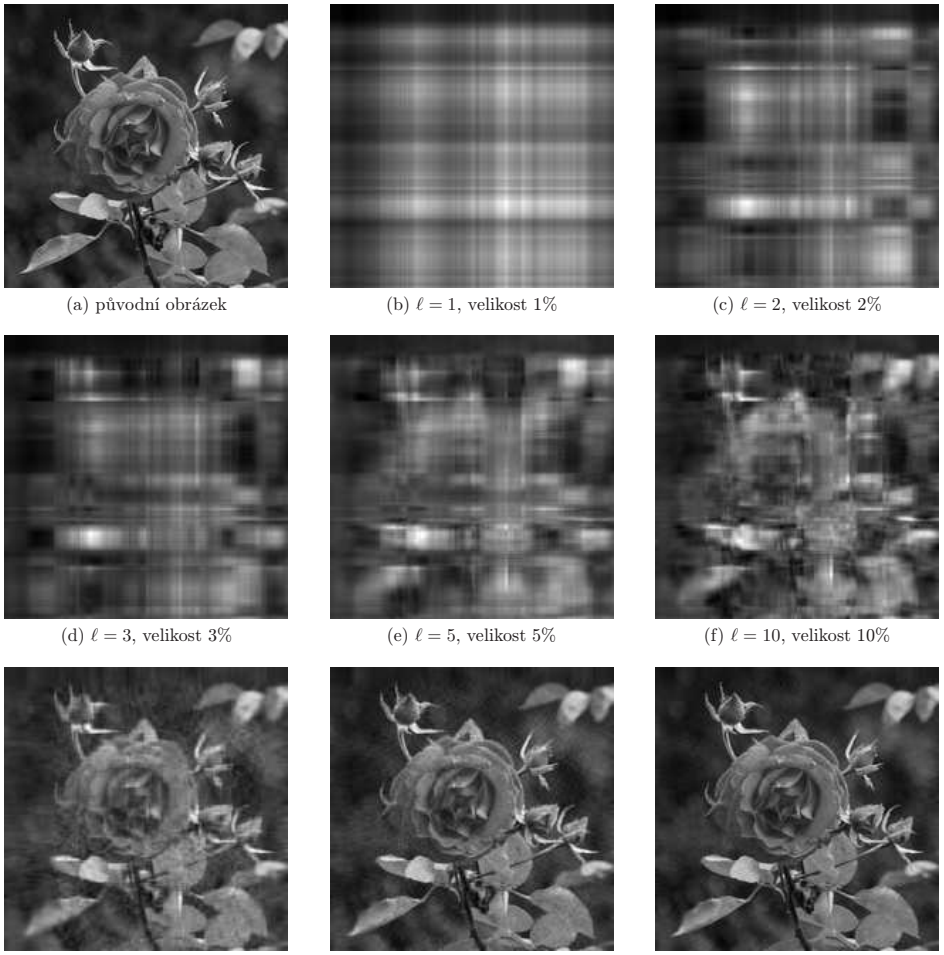
- Nechť $r = \mathrm{rank}(A)$. Potom

$$\dim \mathcal{R}(A) = \dim \mathrm{Im}(A) = r, \qquad \dim \mathrm{Ker}(A) = n - r, \qquad \dim \mathrm{Ker}(A^\mathsf{T}) = m - r.$$

- Řádkový podprostor je lineárně nezávislý s jádrem, obraz je lineárně nezávislý s levým jádrem a tyto podprostory generují příslušné vektorové prostory:

$$\mathcal{R}(A) \oplus \mathrm{Ker}(A) = \mathbb{R}^n \qquad \text{a} \qquad \mathrm{Im}(A) \oplus \mathrm{Ker}(A^\mathsf{T}) = \mathbb{R}^m.$$

Jeden možný způsob, jak si tyto klíčové vlastnosti zapamatovat, je obrázek 4.12, který nás bude provázet v různých variantách po zbytek textu.



(a) původní obrázek (b) $\ell = 1$, velikost 1% (c) $\ell = 2$, velikost 2%

(d) $\ell = 3$, velikost 3% (e) $\ell = 5$, velikost 5% (f) $\ell = 10$, velikost 10%

(g) $\ell = 25$, velikost 25% (h) $\ell = 50$, velikost 50% (i) $\ell = 75$, velikost 75%

Obrázek 4.13: Autorova fotografie růže velikosti 200×200 pixelů. Původní matice (a) má plnou hodnost 200. Zkomprimované obrázky (b) až (i) jsou ukázány pro osm různých voleb ℓ, spolu s procentuální velikostí vůči původnímu obrázku. Pro $\ell = 1$ je v obrázku (b) dobře vidět struktura $\boldsymbol{x}\boldsymbol{y}^\mathsf{T}$. V případě posledních dvou voleb (h) a (i) je už výsledný obrázek velice podobný původnímu.

Cvičení

4.1 Nechť U a V jsou dva libovolné podprostory. Dokažte důležitou rovnost, které se říká *modularita*:

$$\dim U + \dim V = \dim (U \cap V) + \dim \mathcal{L}(U \cup V).$$

4.2 Dokažte následující alternativní definici regulární matice. Čtvercová matice $R \in \mathbb{R}^{n\times n}$ je regulární, právě když při násobení libovolné matice $A \in \mathbb{R}^{n\times n}$ nemění hodnost:

$$\operatorname{rank}(RA) = \operatorname{rank}(A) = \operatorname{rank}(AR).$$

4.3 Nechť U a V jsou dva lineárně nezávislé podprostory. Dokažte, že $\dim U + \dim V = \dim U \oplus V$. Také dokažte, že pro libovolný vektor $\boldsymbol{x} \in U \oplus V$ existují jednoznačně určené vektory $\boldsymbol{x}_u$ a $\boldsymbol{x}_v$, že $\boldsymbol{x}_u + \boldsymbol{x}_v = \boldsymbol{x}$.

4.4 Způsob, jakým jsme nadefinovali přímý součet v této kapitole, se někdy nazývá *vnitřní přímý součet*. Ze dvou nezávislých podprostorů U a V jsme uvnitř nějakého vektorového prostoru $\mathbb{X}$ vybudovali větší podprostor $U \oplus V$ uvnitř $\mathbb{X}$.

Pochopitelně existuje také *vnější přímý součet* dvou vektorových prostorů $\mathbb{X}$ a $\mathbb{Y}$, který budeme značit $\mathbb{X} \times \mathbb{Y}$. Prvky $\mathbb{X} \times \mathbb{Y}$ jsou všechny dvojice $(\boldsymbol{x}, \boldsymbol{y})$, kde $\boldsymbol{x} \in \mathbb{X}$ a $\boldsymbol{y} \in \mathbb{Y}$. Vektorové operace sčítání a násobení skalárem se aplikují na dvojicích po složkách:

$$(\boldsymbol{x}_1, \boldsymbol{y}_1) + (\boldsymbol{x}_2, \boldsymbol{y}_2) = (\boldsymbol{x}_1 + \boldsymbol{x}_2, \boldsymbol{y}_1 + \boldsymbol{y}_2) \qquad \text{a} \qquad \alpha \cdot (\boldsymbol{x}, \boldsymbol{y}) = (\alpha \cdot \boldsymbol{x}, \alpha \cdot \boldsymbol{y}).$$

Dokažte, že výsledný objekt $\mathbb{X} \times \mathbb{Y}$ je vždy vektorový prostor.

Také ukažte, že v případě dvou lineárně nezávislých podprostorů U a V jsou $U \oplus V$ a $U \times V$ izomorfní. (Uvědomme si, že to jsou matematicky jiné objekty: $U \oplus V$ je podprostor $\mathbb{X}$, zatímco $U \times V$ je vektorový prostor dvojic $(\boldsymbol{u}, \boldsymbol{v})$. Algebraicky však mají stejnou strukturu.)

4.5 Zesílíme tvrzení 4.16 ve stylu lemmatu 4.15. Dokažte, že vektory $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ a $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k$ jsou lineárně nezávislé. Navíc pokud pro libovolné lineárně nezávislé vektory platí

$$A = \boldsymbol{x}_1\boldsymbol{y}_1^\mathsf{T} + \cdots + \boldsymbol{x}_k\boldsymbol{y}_k^\mathsf{T},$$

potom $\operatorname{rank}(A) = k$, $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ je báze $\operatorname{Im}(A)$ a $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k$ je báze $\mathcal{R}(A)$. Navíc pro libovolnou volbu jedné báze existuje jednoznačně určená druhá báze.

4.6 Navrhněte algoritmus, který pro zadanou matici A v koeficientech nalezne nejmenší počet vektorů $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ a $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k$, aby platilo

$$A = \boldsymbol{x}_1\boldsymbol{y}_1^\mathsf{T} + \cdots + \boldsymbol{x}_k\boldsymbol{y}_k^\mathsf{T}.$$

⋆ 4.7 Jak už jsme zmínili, obecně může být hodnost $A + B$ klidně nulová, i když obě matice A a B mají nenulovou hodnost. Pokud však má řekněme A velkou hodnost a B malou, dává smysl, že i $A + B$ musí mít velkou hodnost. Nalezněte co nejlepší dolní odhady pro hodnosti $\operatorname{rank}(A + B)$ a $\operatorname{rank}(AB)$ a ukažte na příkladech, že jsou optimální.

Kapitola 5

Lineární zobrazení

V předchozích kapitolách jsme se seznámili se strukturou vektorových prostorů a jejich podprostorů. Také jsme popsali základní vlastnosti matic. Zatím však chybí propojení mezi maticemi a vektorovými prostory, vyjma toho, že je můžeme použít například k elegantnějšímu zápisu soustavy lineárních rovnic jako $A\boldsymbol{x} = \boldsymbol{b}$. V této kapitole vytvoříme tento chybějící most a ukážeme, proč jsou matice klíčovým pojmem lineární algebry.

Na matici lze nahlížet dvěma rozdílnými způsoby. Můžeme s ní pracovat jako s tabulkou čísel, která reprezentují určitá data. Tento pohled je zejména u studentů informatiky nejrozšířenější. Neméně důležitý je však i druhý pohled. Každá matice popisuje geometrickou transformaci vektorového prostoru. Velká část lineární algebry se zabývá tím, jak s těmito transformacemi pracovat a pochopit jejich vlastnosti. Matice často popisuje vývoj nějakého uzavřeného systému a naším úkolem je určit, co se se systémem stane.

Představme si například most. Ten se skládá z řady komponent, které na sebe působí. Přirozeně můžeme tyto působení reprezentovat maticí, kde do složky (i, j) zapíšeme, jak i-tá komponenta působí na j-tou. Všimněte si, že zde nahlížíme na matici jako na data. Avšak matice také popisuje proces nad komponentami mostu, tedy to, jak se působením sil most transformuje z jednoho stavu do jiného. Je zcela přirozené zabývat se následujícím problémem. Most se nachází v počátečním stavu a necháme jeho komponenty na sebe působit po nějaký čas. Tím, jak na sebe komponenty působí, se dostane most do jiného stavu. Nás pochopitelně zajímá, jaký ten stav bude a jestli například most nespadne.

Dnešní stavitelé mostu využívají software založený na algoritmech lineární algebry, který jim umožňuje analyzovat strukturu mostů a dalších staveb. V dávnějších dobách podobné analýzy nebyly možné, což například vedlo v roce 1940 k pádu špatně navrženého mostu přes úžinu Tacoma vlivem relativně slabého působení větru.[1] I drobné vlivy mohou mít velké následky, pokud se nevhodně složí dohromady.

5.1 Matice jako reprezentace

Z pohledu abstraktní algebry je lineární algebra studium vektorových prostorů a lineárních zobrazení. Vektorovým prostorům jsme se už věnovali a je na čase se podívat na lineární zobrazení.

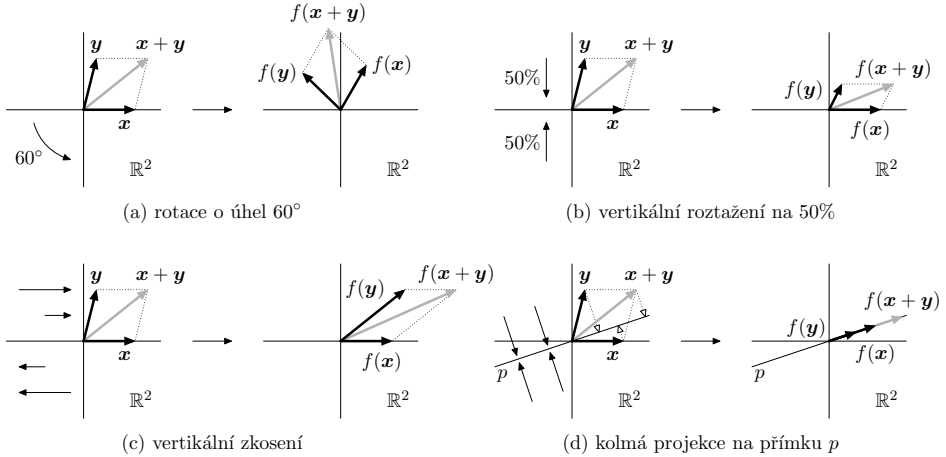
Lineární zobrazení. Mějme dva vektorové prostory $\mathbb{U}$ a $\mathbb{V}$. *Lineární zobrazení* je zobrazení $f : \mathbb{U} \to \mathbb{V}$ splňující následující dvě podmínky, souhrnně nazývané *linearita*:

$$\begin{aligned} f(\boldsymbol{x}+\boldsymbol{y}) &= f(\boldsymbol{x})+f(\boldsymbol{y}) && \text{pro libovolné } \boldsymbol{x},\boldsymbol{y} \in \mathbb{U}, \\ f(\alpha\boldsymbol{x}) &= \alpha f(\boldsymbol{x}) && \text{pro libovolné } \boldsymbol{x} \in \mathbb{U} \text{ a } \alpha \in \mathbb{R}. \end{aligned}$$

[1] Doporučuji vyhledat třeba na YouTube video „Tacoma Narrows Bridge“, které ukazuje most houpající se a vlnící se pod vlivem větru.

Povšimněme si, že na levé straně rovností jsou použité vektorové operace $\mathbb{U}$, zatímco na pravé straně operace $\mathbb{V}$. Mohli bychom linearitu zapsat jednou rovností jako $f(\alpha\boldsymbol{x}+\beta\boldsymbol{y}) = \alpha f(\boldsymbol{x})+\beta f(\boldsymbol{y})$, nicméně rozepsaní do dvou podmínek považuji za přehlednější.

Dobrými příklady lineárních zobrazení jsou geometrické transformace, které čtenář určitě zná. Zůstaňme pro jednoduchost v rovině, tedy $\mathbb{U} = \mathbb{V} = \mathbb{R}^2$. Na obrázku 5.1 jsou příklady čtyř geometrických transformací. Čtenář si může rozmyslet, proč jsou tato zobrazení lineární. Abychom to však mohli dokázat, potřebovali bychom je geometricky definovat, což dělat nebudeme. Čtenář si také může rozmyslet, že posunutí o nenulový vektor není lineární zobrazení.



(a) rotace o úhel 60°

(b) vertikální roztažení na 50%

(c) vertikální zkosení

(d) kolmá projekce na přímku p

Obrázek 5.1: Ukázka čtyř geometrický transformací, které jsou lineární, spolu s obrazy vektorů $f(\boldsymbol{x})$, $f(\boldsymbol{y})$ a $f(\boldsymbol{x}+\boldsymbol{y})$. Obrázek ilustruje, že pro tyto vektory platí linearita $f(\boldsymbol{x}+\boldsymbol{y}) = f(\boldsymbol{x})+f(\boldsymbol{y})$.

Motivace. Proč je zrovna linearita tak klíčová vlastnost zobrazení, že jí chceme věnovat celou kapitolu? Jeden důvod je, že řada zobrazení linearitu přirozeně splňuje. Druhý důležitější důvod je, že tyto vlastnosti přenáší strukturu $\mathbb{U}$ dovnitř struktury $\mathbb{V}$. Každé lineární zobrazení je *homomorfismus*, což je v algebře obecný název pro zobrazení chovající se hezky vůči algebraickým operacím.

Platí, že velká část struktury je zachována homomorfismem. Například mějme rovnost $\boldsymbol{u}+\boldsymbol{v} = \boldsymbol{w}$. Zobrazením obou stran lineárním zobrazením $f : \mathbb{U} \to \mathbb{V}$ a rozepsáním levé strany podle linearity dostáváme:

$$f(\boldsymbol{u})+f(\boldsymbol{v}) = f(\boldsymbol{w}).$$

Rovnost z $\mathbb{U}$ je tedy zobrazením do $\mathbb{V}$ zachována.

Historická motivace pro studium homomorfismů vychází ze snahy o řešení *diofantických rovnic*. Diofantické rovnice jsou rovnice pro celá čísla. Například rovnice $x^n + y^n = z^n$, pro kterou hledáme řešení $x, y, z \in \mathbb{Z}$. Slavná Velká Fermatova věta, dokázaná v roce 1995 Andrew Wilesem, říká, že pro $n \geq 3$ neexistuje žádné řešení. Obecně je velice obtížné vyřešit diofantické rovnice, ostatně v roce 1970 Yuri Matiyasevich dokázal, že problém řešení diofantických rovnic je algoritmicky neřešitelný; neexistuje obecný algoritmus, který by pro zadanou diofantickou rovnici rozhodl, zda má řešení nebo ne. Pochopitelně může algoritmus zkoušet dosazovat všechna možná celočíselná ohodnocení a zastavit se v případě, že nalezne libovolné řešení. Pokud však řešení neexistuje, algoritmus se nikdy nezastaví. Algoritmická neřešitelnost říká, že žádný esenciálně lepší postup neexistuje.

A jak tohle souvisí s homomorfismy? Pokud řešíme obtížný problém, je často užitečné řešit jednodušší variantu, ideálně pokud tato jednodušší varianta řekne něco o řešení původního problému. Užitečná technika pro řešení diofantických rovnic je řešit modulo nějaké pevně zvolené číslo k, což odpovídá řešení v $\mathbb{Z}_k = \{0, 1, \ldots, k-1\}$. Například lze ukázat, že pro libovolné přirozené číslo n má n^2 zbytek 0, 1 nebo 4 modulo 8. Tedy rovnice obsahující druhé mocniny může být výrazně jednodušší v $\mathbb{Z}_8$ než v $\mathbb{Z}$. Pochopitelně není pravda, že každé nalezené řešení v $\mathbb{Z}_k$ odpovídá řešení v $\mathbb{Z}$, přesto se však můžeme ledacos o původní diofantické rovnici dozvědět, uvažujeme-li například modulo různá k. A kde se nachází onen homomorfismus? Počítání modulo k není nic jiného než použití homomorfismu $\phi : \mathbb{Z} \to \mathbb{Z}_k$. Lidé se snažili tyto techniky pochopit a zdokonalit, a tím došlo k rozvoji celé abstraktní algebry.

Obrazy lineárních kombinací. Mějme lineární zobrazení $f : \mathbb{U} \to \mathbb{V}$. Můžeme si nejprve všimnout, že linearita implikuje, že obrazem nuly musí nutně být nula:

$$f(\mathbf{0}) = f(0 \cdot \mathbf{0}) = 0 \cdot f(\mathbf{0}) = \mathbf{0}, \qquad \text{alternativně} \qquad f(\mathbf{0}) = f(\mathbf{0} - \mathbf{0}) = f(\mathbf{0}) - f(\mathbf{0}) = \mathbf{0}.$$

Dokonce jsme ukázali dva různé důkazy, každý využívající jinou z vlastností linearity.

Podobně obrazem libovolné lineární kombinace je lineární kombinace obrazů:

$$f\left(\sum_{i=1}^{n} \alpha_i \boldsymbol{x}_i\right) = \sum_{i=1}^{n} f(\alpha_i \boldsymbol{x}_i) = \sum_{i=1}^{n} \alpha_i f(\boldsymbol{x}_i), \tag{5.1}$$

kde první rovnost platí opakovaným použitím linearity na součty, a druhá rovnost platí aplikováním linearity na součin v každém členu sumy.

Pokud známe obrazy $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, známe podle (5.1) také obrazy lineárního obalu $\mathcal{L}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$. Speciálně pokud tyto vektory tvoří bázi $\mathbb{U}$, je celé lineární zobrazení jednoznačně určené. Pokud chceme popsat libovolné lineární zobrazení, stačí uvážit libovolnou bázi a popsat pouze její hodnoty. Libovolný další vektor má jednoznačně určenou lineární kombinaci vůči této bázi, a tedy i jednoznačně určenou hodnotu lineárního zobrazení.

> Pokud známe hodnoty lineárního zobrazení pro libovolnou bázi, je celé lineární zobrazení jednoznačně určené.

Maticová reprezentace. Představme si, že chceme lineární zobrazení $f : \mathbb{U} \to \mathbb{V}$ uložit třeba do počítače a chceme najít co nejelegantnější způsob, jak ho reprezentovat. Podle (5.1) víme, že si stačí pamatovat jeho hodnoty pro libovolnou bázi $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$. Pokud chceme určit hodnotu $f(\boldsymbol{x})$, nejprve vyjádříme $\boldsymbol{x}$ vůči této bázi: $\boldsymbol{x} = \alpha_1 \boldsymbol{x}_1 + \cdots + \alpha_n \boldsymbol{x}_n$, což vede na řešení soustavy lineárních rovnic. Pomocí tohoto vyjádření můžeme spočítat

$$f(\boldsymbol{x}) = \alpha_1 f(\boldsymbol{x}_1) + \cdots + \alpha_n f(\boldsymbol{x}_n). \tag{5.2}$$

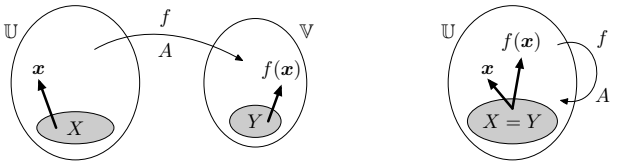
Musíme však ještě vyřešit, jak reprezentovat obrazy $f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_n)$. To jsou vektory z vektorového prostoru $\mathbb{V}$. Proto můžeme zvolit libovolnou bázi $Y = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m\}$ prostoru $\mathbb{V}$. Libovolný vektory $\boldsymbol{y} \in \mathbb{V}$ má jednoznačně určené vyjádření $\boldsymbol{y} = \beta_1 \boldsymbol{y}_1 + \cdots + \beta_m \boldsymbol{y}_m$. Stačí si tedy pamatovat tato vyjádření pro vektory $f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_n)$, což je n vyjádření, každé má m složek. Pomocí (5.2) a těchto vyjádření můžeme určit obraz libovolného vektoru $f(\boldsymbol{x})$.

Tedy pro popis lineárního zobrazení musíme zvolit dvě báze X a Y a uložit mn reálných čísel. Je velice přirozené tyto reálná číslo uspořádat do tabulky $m \times n$, kde vyjádření $f(\boldsymbol{x}_i)$ zapíšeme do i-tého sloupce. Nechť $f(\boldsymbol{x}_i) = a_{1,i}\boldsymbol{y}_1 + \cdots a_{m,i}\boldsymbol{y}_m$ je vyjádření vektoru $f(\boldsymbol{x}_i)$. Potom vzniklá tabulka vypadá

takto a nazývá se *matice*:

$$A = \begin{pmatrix} | & | & | & & | \\ f(\boldsymbol{x}_1) & f(\boldsymbol{x}_2) & f(\boldsymbol{x}_3) & \cdots & f(\boldsymbol{x}_n) \\ | & | & | & & | \end{pmatrix} = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & a_{2,3} & \cdots & a_{2,n} \\ a_{3,1} & a_{3,2} & a_{3,3} & \cdots & a_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & a_{m,3} & \cdots & a_{m,n} \end{pmatrix}.$$

Schématicky je lineární zobrazení naznačeno na obrázku 5.2 vlevo.



Obrázek 5.2: Nalevo schéma lineárního zobrazení $f : \mathbb{U} \to \mathbb{V}$ a jeho reprezentace A pro volby bází X a Y. Napravo situace, když f je endomorfismus $\mathbb{U}$ a $X = Y$.

Možná může být divné, proč znovu zavádíme pojem matice, když už jsme to udělali v kapitole 3. Jedna z neuspokojivých vlastností původního přístupu je, že se definice matic a jejich operací objevily z ničeho nic, bez žádného odvození nebo vysvětlení. Speciálně pro maticové násobení není vůbec zřejmé, proč jsme ho definovali tímto způsobem. Nyní se pokusíme s tímto nedostatkem vypořádat a zkusíme odvodit definici maticového násobení. Povšimněte si, že jediný předpoklad, který jsme museli udělat, je následující: matice je tabulka, která po sloupcích obsahuje vyjádření obrazů vektorů báze X vůči bázi Y.

Podle Steinitzovy věty má libovolná báze vektorového prostoru stejnou velikost. Pokud $\dim \mathbb{U} = n$ a $\dim \mathbb{V} = m$, platí podle věty 4.4 o izomorfismu $\mathbb{U} = \mathbb{R}^n$ a $\mathbb{V} = \mathbb{R}^m$ a toto lineární zobrazení ekvivalentně $f : \mathbb{R}^n \to \mathbb{R}^m$. Libovolná matice reprezentující $f : \mathbb{U} \to \mathbb{V}$ je velikosti $m \times n$. Pokud mají oba prostory stejnou dimenzi, dostáváme čtvercovou matici. V případě stejných dimenzí $\mathbb{U}$ a $\mathbb{V}$ se často volí pouze jedna báze $X = Y$. Protože platí $\mathbb{U} = \mathbb{V}$, lineární zobrazení $f : \mathbb{U} \to \mathbb{U}$ transformuje vektory uvnitř vektorového podprostoru. Takovému zobrazení se často místo homomorfismus říká *endomorfismus*, schéma na obrázku 5.2 vpravo.

Pro různé báze X a Y dostáváme různé matice reprezentující jedno lineární zobrazení. Pochopitelně pokud na prostory $\mathbb{U}$ a $\mathbb{V}$ nahlížíme jako na $\mathbb{R}^n$ a $\mathbb{R}^m$, můžeme za X a Y zvolit kanonické báze a často je to rozumná volba. Avšak existují situace, kdy můžeme výrazně lépe pochopit lineární zobrazení volbou bází X a Y vhodných pro toto lineární zobrazení. Ostatně na této myšlence je postavena celá teorie vlastních čísel a vlastních vektorů. Pokud existuje báze tvořená z vlastních vektorů, příslušná matice reprezentující lineární zobrazení je diagonální. Poznamenejme, že dvě různá lineární zobrazení f a g mohou být pro vhodné volby bází reprezentována stejnou maticí A.

> Pro volbu dvou bází X a Y dostáváme matici A reprezentující lineární zobrazení tak, že zapíšeme do sloupců souřadnice obrazů bazický vektorů báze X vůči bázi Y. Pro volby různých bází dostáváme různé matice reprezentující stejné lineární zobrazení.

Již jsme popsali, jak lze pomocí matic pro zvolené báze reprezentovat lineární zobrazení. Ale ještě nevíme, jakým způsobem lze s takovými reprezentacemi pracovat. Ideální reprezentace totiž umožňuje provést ty úkony, které by člověk mohl dělat přímo se znalostí lineárního zobrazení. Matice jsou vskutku vynikajícími reprezentacemi a všechny běžné úkony umožňují snadno provádět.

Zobrazování vektorů. Například přirozený úkon je pro daný vektor $\boldsymbol{x}$ nalézt jeho obraz $f(\boldsymbol{x})$. Předpokládejme, že máme $\boldsymbol{x}$ vyjádřený vůči bázi X, tedy $\boldsymbol{x} = x_1\boldsymbol{x}_1 + \cdots + x_n\boldsymbol{x}_n$. (Zde x_i je koeficient a $\boldsymbol{x}_i$ je vektor báze X.) Chceme nalézt vyjádření $f(\boldsymbol{x})$ vůči bázi Y.

Pochopitelně bychom se mohli spokojit s odpovědí (5.2), ale chceme zjistit, jak přesně budou vypadat koeficienty ve vyjádření $f(\boldsymbol{x})$. Můžeme tedy do (5.2) dosadit vyjádření $f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_n)$, což jsou sloupečky A, a dostáváme:

$$x_1\begin{pmatrix} a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{m,1} \end{pmatrix} + x_2\begin{pmatrix} a_{1,2} \\ a_{2,2} \\ \vdots \\ a_{m,2} \end{pmatrix} + \cdots + x_n\begin{pmatrix} a_{1,n} \\ a_{2,n} \\ \vdots \\ a_{m,n} \end{pmatrix} = \begin{pmatrix} a_{1,1}x_1 + \cdots + a_{1,n}x_n \\ a_{2,1}x_1 + \cdots + a_{2,n}x_n \\ \vdots \\ a_{m,1}x_1 + \cdots + a_{m,n}x_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n a_{1,i}x_i \\ \sum_{i=1}^n a_{2,i}x_i \\ \vdots \\ \sum_{i=1}^n a_{m,i}x_i \end{pmatrix}.$$

Možná čtenáři přijdou tyto výrazy povědomé, podobné těm z kapitoly 3.

Pokud $\boldsymbol{x}$ uvážíme jako sloupcový vektor $(x_1, \ldots, x_n)$, potom to není nic jiného vynásobení vektoru $\boldsymbol{x}$ maticí zleva:

$$f(\boldsymbol{x}) = A\boldsymbol{x}. \tag{5.3}$$

Z definice matice A jako reprezentace lineárního zobrazení f vychází přirozeně potřeba definovat součin matice s vektorem (nebo jinak nazvanou operaci), protože to odpovídá zobrazování vektorů lineárním zobrazením f. V řeči reprezentace máme tedy lineární zobrazení $f : \boldsymbol{x} \mapsto A\boldsymbol{x}$.

Matice A s volbou bází X a Y reprezentuje lineární zobrazení f jako

$$f : \boldsymbol{x} \mapsto A\boldsymbol{x},$$

kde vektory $\boldsymbol{x}$ jsou vyjádřené vůči bázi X a jejich obrazy $A\boldsymbol{x}$ vůči bázi Y.

Geometrické příklady. Pro lepší seznámení s definicí si ukažme, jak vypadají maticové reprezentace základních geometrických transformací z obrázku 5.1. Další geometrické transformace může čtenář prozkoumat ve cvičeních 5.1 a 5.2. Každá transformace je endomorfismus a uvažujeme ji vůči kanonické bázi.

Nejjednodušší transformací je α-násobné roztažení celého vektorového prostoru, tedy natažení všech vektorů koeficientem α. Protože $f(\boldsymbol{e}_i) = \alpha\boldsymbol{e}_i$, dostáváme v i-tém sloupci jediný nenulový koeficient na pozici (i, i). Tedy matice reprezentující transformaci vůči kanonické bázi je αI_n; ostatně vůči libovolné bázi. Maličko obecnější je transformace, která v směru každé osy natahuje s jiným koeficientem α_i. Ta je reprezentována diagonální maticí

$$\begin{pmatrix} \alpha_1 & & & \\ & \alpha_2 & & \\ & & \ddots & \\ & & & \alpha_n \end{pmatrix}.$$

S diagonálními maticemi se pracuje tak dobře, protože reprezentují ty nejjednodušší transformace. Každý vektor báze je transformován nezávisle na ostatních.

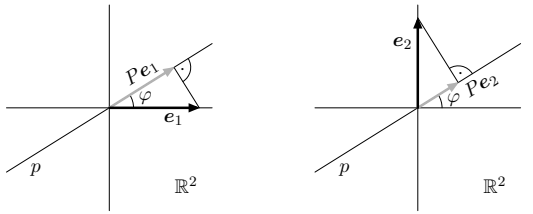
Pro ostatní zobrazení se zaměřme na rovinu $\mathbb{R}^2$. Popsat je v $\mathbb{R}^n$ je totiž složitější, i když k tomu lineární algebra dává prostředky. Zkosení zobrazuje $(x, y) \mapsto (x + \alpha y, y)$, tedy je reprezentováno maticí

$$\begin{pmatrix} 1 & \alpha \\ & 1 \end{pmatrix}.$$

Co se týká rotace kolem počátku o úhel φ, k určení matice stačí zjistit, jak vypadají obrazy $\boldsymbol{e}_1$ a $\boldsymbol{e}_2$ vůči kanonické bázi. S využitím základních znalostí geometrie a goniometrických funkcí dostáváme, že matice rotace o úhel φ je

$$\begin{pmatrix} \cos\varphi & \cos(\varphi + 90°) \\ \sin\varphi & \sin(\varphi + 90°) \end{pmatrix} = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}.$$

Dále uvažme kolmou projekci na přímku p procházející počátkem pod úhlem φ. Pro určení matice projekce P stačí nalézt projekce vektorů kanonické báze. Protože libovolný obraz leží na p, lze ho vyjádřit jako $(\alpha\cos\varphi, \alpha\sin\varphi)$ vůči kanonické bázi pro nějaký koeficient α. Zbývá určit tyto koeficienty pro $\boldsymbol{e}_1$ a $\boldsymbol{e}_2$, k čemuž nám postačí základní znalosti z geometrie; situace je naznačena na obrázku 5.3.



Obrázek 5.3: Nalevo vektor $\boldsymbol{e}_1$ a jeho obraz, napravo $\boldsymbol{e}_2$ a jeho obraz. Délka projektovaného vektoru je délka odvěsny v pravoúhlém trojúhelníku.

Protože $\cos\varphi$ je v pravoúhlém trojúhelníku délka přilehlé strany vůči délce přepony rovné jedna, dostáváme

$$P\boldsymbol{e}_1 = (\cos\varphi \cdot \cos\varphi, \cos\varphi \cdot \sin\varphi) \qquad \text{a} \qquad P\boldsymbol{e}_2 = (\cos(\varphi - 90°) \cdot \cos\varphi, \cos(\varphi - 90°) \cdot \sin\varphi).$$

Tedy výsledná matice projekce vůči kanonické bázi je rovna

$$P = \begin{pmatrix} \cos\varphi \cdot \cos\varphi & \cos(\varphi - 90°) \cdot \cos\varphi \\ \cos\varphi \cdot \sin\varphi & \cos(\varphi - 90°) \cdot \sin\varphi \end{pmatrix} = \begin{pmatrix} \cos^2\varphi & \cos\varphi\sin\varphi \\ \cos\varphi\sin\varphi & \sin^2\varphi \end{pmatrix}.$$

Povšimněme si, že matice P je symetrická, což není náhoda. Mimochodem z posledního zápisu vůbec není patrné, že má matice hodnost jedna.

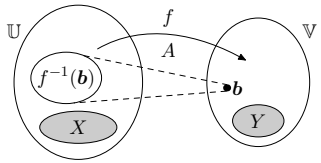
Hledání vzorů. Lze si však položit i obrácenou otázku. Pro daný vektor $\boldsymbol{b} \in \mathbb{V}$ chceme nalézt množinu jeho vzorů

$$f^{-1}(\boldsymbol{b}) = \big\{\boldsymbol{x} \mid \boldsymbol{x} \in \mathbb{U} \text{ a platí } f(\boldsymbol{x}) = \boldsymbol{b}\big\}.$$

Mějme vektor $\boldsymbol{b}$ vyjádřený vůči bázi Y. Podle (5.3) dostáváme soustavu lineárních rovnic $A\boldsymbol{x} = \boldsymbol{b}$. Situace je nakreslena na obrázku 5.4. I kdybychom lineární algebru budovali abstraktně přes vektorové prostory a lineární zobrazení, musíme umět řešit soustavy kvůli hledání vzorů lineárních zobrazení.

Pochopitelně žádný vzor $\boldsymbol{b}$ nemusí existovat, a také může vzorů existovat nekonečně mnoho. To odpovídá tomu, že soustava $A\boldsymbol{x} = \boldsymbol{b}$ nemusí mít žádné řešení, respektive může mít nekonečně mnoho řešení. Z výsledků popsaných v kapitolách 2 a 3 víme, že množina vzorů $f^{-1}(\boldsymbol{b})$ vždy tvoří afinní podprostor vektorového prostoru $\mathbb{U}$, který nalezneme vyřešením soustavy vyjádřený vůči bázi X.

Množina vzorů $f^{-1}(\boldsymbol{b})$ je množina všech řešení soustavy $A\boldsymbol{x} = \boldsymbol{b}$.



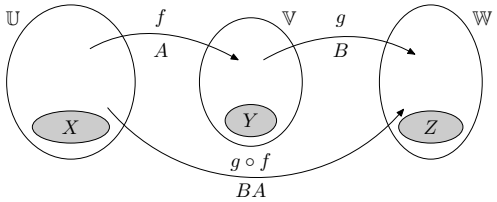
Obrázek 5.4: Množina vzorů $f^{-1}(\boldsymbol{b})$, které f zobrazuje na $\boldsymbol{b}$. Ukázali jsme, že množina $f^{-1}(\boldsymbol{b})$ je rovna množině řešení soustavy $A\boldsymbol{x} = \boldsymbol{b}$.

Skládání lineárních zobrazení. Není obtížné dokázat, že složení dvou lineárních zobrazení $g \circ f$ je lineární zobrazení.[2] Ukažme alespoň, že složení dvou lineárních zobrazení splňuje linearitu pro součet:

$$g\big(f(\boldsymbol{x}+\boldsymbol{y})\big) = g\big(f(\boldsymbol{x})+f(\boldsymbol{y})\big) = g\big(f(\boldsymbol{x})\big) + g\big(f(\boldsymbol{y})\big),$$

kde první rovnost platí kvůli linearitě f a druhá díky linearitě g.

Přirozená otázka je, jak vypadá maticová reprezentace složení dvou lineárních zobrazení. Mějme tři vektorové prostory $\mathbb{U}$, $\mathbb{V}$ a $\mathbb{W}$ s bázemi X, Y a Z a dimenzemi n, m a p. Dále máme lineární zobrazení $f : \mathbb{U} \to \mathbb{V}$ a $g : \mathbb{V} \to \mathbb{W}$ reprezentovaná maticemi A a B vůči bázím X a Y, a vůči Y a Z; tedy prostřední báze je společná, aby na sebe matice A a B navazovaly. Zajímá nás, jak vypadá maticová reprezentace $g \circ f$ vůči bázím X a Z, a tvrdíme, že je to BA. Pro lepší orientaci se podívejte na obrázek 5.5.



Obrázek 5.5: Matice reprezentující $g \circ f$ vůči bázím X a Z není nic jiného než maticový součin BA.

Tvrzení 5.1. *Nechť matice A reprezentuje lineární zobrazení f vůči X a Y a B reprezentuje g vůči Y a Z. Potom matice BA reprezentuje $g \circ f$ vůči bázím X a Z*

Důkaz. Můžeme postupovat přímo podle definice maticové reprezentace. Matice reprezentující $g \circ f$ má mít v i-tém sloupci vyjádření $(g \circ f)(\boldsymbol{x}_i)$ vůči bázi Z. Matice A obsahuje informace o tom, jak vypadají vyjádření $f(\boldsymbol{x}_i)$ vůči bázi Y. Podobně matice B obsahuje vyjádření $g(\boldsymbol{y}_i)$ vůči bázi Z. Víme, že $f(\boldsymbol{x}_i) = a_{1,i}\boldsymbol{y}_1 + \cdots + a_{m,i}\boldsymbol{y}_m$. Pro určení $g\big(f(\boldsymbol{x}_i)\big)$ stačí dosadit za $\boldsymbol{y}_i$ vyjádření $g(\boldsymbol{y}_i)$ vůči bázi Z.

Zbytek důkazu je maličko pracné hraní s koeficienty. Dostáváme

$$g\big(f(\boldsymbol{x}_i)\big) = a_{1,i}(b_{1,1}\boldsymbol{z}_1 + \cdots + b_{p,1}\boldsymbol{z}_p) + \cdots + a_{m,i}(b_{1,m}\boldsymbol{z}_1 + \cdots + b_{p,m}\boldsymbol{z}_p).$$

Přeuspořádejme násobky jednotlivých $\boldsymbol{z}_i$ do společných závorek, a prohodíme pořadí koeficientů matic A a B:

$$g\big(f(\boldsymbol{x}_i)\big) = (b_{1,1}a_{1,i} + \cdots b_{1,m}a_{m,i})\boldsymbol{z}_1 + \cdots + (b_{p,1}a_{1,i} + \cdots + b_{p,m}a_{m,i})\boldsymbol{z}_p.$$

[2] Při skládání zapisujeme zobrazení zprava doleva podle toho, jak se aplikují. To může působit na první pohled podivně, ale umožňuje to zapsat $(g \circ f)(x)$ také jako $g\big(f(x)\big)$ bez změny pořadí.

Tím dostáváme vyjádření $g\big(f(\boldsymbol{x}_i)\big)$ vůči jednotlivým vektorům báze Z, a tedy jednotlivé koeficienty ve výsledné matici. Na pozici (j,i) budeme mít koeficient $\sum_{k=1}^{n} b_{j,k}a_{k,i}$, což není nic jiného než maticový součin. Tedy matice reprezentující $g \circ f$ vůči bázím X a Z je BA. □

Máme tedy vysvětlenou definici maticového násobení. Maticové násobení se musí definovat na první pohled tak zvláštně, aby součin matic odpovídal skládání zobrazení. I kdybychom budovali lineární algebru přes lineární zobrazení a jejich reprezentace, odvodili bychom definici maticového násobení. Opět připojme k tvrzení několik poznámek.

Povšimněme si, že pořadí součinu matic je stejné jako pořadí zobrazení při skládáni. To je způsobené naší volbou, že souřadnice $f(\boldsymbol{x}_i)$ zapíšeme do i-tého sloupce. Mohli bychom pochopitelně zvolit i transponovanou definici a umístit souřadnicemi do i-tého řádku. Pak by skládání vypadalo jako $A^\mathsf{T}B^\mathsf{T}$ a zobrazovalo by se násobením zprava: $f : \boldsymbol{x}^\mathsf{T} \mapsto \boldsymbol{x}^\mathsf{T}A^\mathsf{T}$. Museli bychom pracovat s řádkovými vektory místo sloupcovými vektory. Ale dostali bychom formálně úplně stejný svět lineární algebry. Jeden z důvodu, proč v tomto textu upřednostňujeme zápis po sloupcích je, že hledání vzorů vede na řešení soustavy $A\boldsymbol{x} = \boldsymbol{b}$. Pokud bychom zapisovali po řádcích, dostali bychom soustavu $\boldsymbol{x}^\mathsf{T}A^\mathsf{T} = \boldsymbol{b}^\mathsf{T}$, tedy jednotlivé rovnice by byly zapsané ve sloupcích; to není příliš typické. Při čtení jiných knih, obzvláště o aplikacích lineární algebry, by si čtenář měl dát pozor, že může být použita řádková reprezentace a zobrazení se mohou skládat z druhé strany.

Připomeňme si, že součin BA dává smysl pouze tehdy, když $B \in \mathbb{R}^{p\times m}$ a $A \in \mathbb{R}^{m\times n}$. Prostřední velikost se musí shodovat a výsledná matice $BA \in \mathbb{R}^{p\times n}$. V řeči zobrazení $A : \mathbb{R}^n \to \mathbb{R}^m$ a $B : \mathbb{R}^m \to \mathbb{R}^p$. Abychom tyto dvě zobrazení mohli složit, musí být prostřední prostor stejně velký. Složením dostáváme $BA : \mathbb{R}^n \to \mathbb{R}^p$, které podle definice dává $BA \in \mathbb{R}^{p\times n}$.

Násobení matic odpovídá skládání zobrazení. Mějme dvě lineární zobrazení:

$f : \mathbb{U} \to \mathbb{V}$ reprezentované A vůči bázím X a Y,
$g : \mathbb{V} \to \mathbb{W}$ reprezentované B vůči bázím Y a Z.

Jejich složení $g \circ f$ je reprezentované maticí BA vůči X a Z.

Vztah s asociativitou. V kapitole 3 jsme slibovali jednodušší důkaz tvrzení 3.2. Chceme ukázat, že $(AB)C = A(BC)$. Nechť A reprezentuje lineární zobrazení f, B reprezentuje g a C reprezentuje h (třeba vůči kanonickým bázím). Potom se rovnost převede na $(f \circ g) \circ h = f \circ (g \circ h)$. Je snadné ukázat, že skládání lineárních zobrazení je asociativní, protože dokonce skládání libovolných zobrazení je asociativní. Čtenář si může rozmyslet proč.

Předchozí důkaz má jednu vadu na kráse. Víme, že zobrazení jsou stejná, ale proč by z toho mělo vyplývat, že i matice jsou stejné. Musíme ukázat, že pro dané lineární zobrazení a danou volbu bází X a Y je matice A určená jednoznačně. Předpokládejme, že by byly dvě různé matice A a B, které by reprezentovaly f vůči X a Y. Potom existuje koeficient (i,j), že $a_{i,j} \neq b_{i,j}$. Avšak obrazy $A\boldsymbol{x}_i$ a $B\boldsymbol{x}_i$ mají jiný koeficient ve vyjádření vůči $\boldsymbol{y}_j$, a tedy jsou různé. To je spor.

Naopak nelíbí-li se vám důkaz tvrzení 5.1, lze toto tvrzení alternativně dokázat pomocí asociativity maticového násobení z tvrzení 3.2. Platí totiž $(BA)\boldsymbol{x} = B(A\boldsymbol{x})$. To vpravo podle definice odpovídá obrazu $(g \circ f)(\boldsymbol{x})$, protože nejprve zobrazíme A, a poté zobrazíme B. To je podle asociativity ekvivalentní zobrazením BA. Tedy $g \circ f$ musí být reprezentováno BA, jak jsme chtěli. Můžete si rozmyslet detaily.

Izomorfismy a automorfismy. Připomeňme si standardní značení. Zobrazení $f : \mathbb{U} \to \mathbb{V}$ se nazývá *prosté*, pokud zobrazuje každý vektor $\mathbb{U}$ jinam, tedy $f(\boldsymbol{x}) = f(\boldsymbol{y})$ implikuje $\boldsymbol{x} = \boldsymbol{y}$. Zobrazení f se nazývá *na*, pokud se na každý vektor ve $\mathbb{V}$ něco zobrazuje, tedy pro každý vektor $\boldsymbol{b} \in \mathbb{V}$ existuje vzor $\boldsymbol{x} \in \mathbb{U}$, že

$f(\boldsymbol{x}) = \boldsymbol{b}$. Zobrazení se nazývá *bijektivní*, pokud splňuje obě podmínky současně. Například pro čtyři geometrická zobrazení z obrázku 5.1 platí, že první tři jsou bijekce a čtvrté není ani prosté, ani na.

Pojem izomorfismu jsme zmínili v kapitole 4, kde jsme dokázali, že libovolný vektorový prostor dimenze n má algebraicky totožnou strukturu s $\mathbb{R}^n$. Zadefinujme si ho pořádně. *Izomorfismus* je bijektivní lineární zobrazení $f : \mathbb{U} \to \mathbb{V}$. Vektorové prostory $\mathbb{U}$ a $\mathbb{V}$ jsou *izomorfní*, pokud mezi nimi existuje izomorfismus $f : \mathbb{U} \to \mathbb{V}$. Izomorfní prostory se značí $\mathbb{U} \cong \mathbb{V}$. Bijektivnost f tedy páruje jednotlivé prvky $\mathbb{U}$ a $\mathbb{V}$, a linearita zaručuje, že toto párování respektuje algebraickou strukturu.[3]

Existence izomorfismu f mezi $\mathbb{U}$ a $\mathbb{V}$ zaručuje totožnou algebraickou strukturu, například rovnost $\boldsymbol{x} = \boldsymbol{y} + \boldsymbol{z}$ se přenese do $\mathbb{V}$ jako $f(\boldsymbol{x}) = f(\boldsymbol{y}) + f(\boldsymbol{z})$, a naopak. Pokud ukážeme libovolnou algebraickou vlastnost o jednom z prostorů, platí tato vlastnost i pro druhý. Tedy z určitého pohledu je můžeme považovat za totožné objekty. Avšak z jiného pohledu totožné být nemusí, jejich vektory mohou být velice odlišné objekty. Například vektorový prostor všech polynomů stupně nejvýše n je izomorfní s $\mathbb{R}^{n+1}$, avšak polynomu jsou zcela jiné matematické objekty než uspořádané $(n+1)$-tice.

Už jsme uvedli, že lineární zobrazení (homomorfismus) $f : \mathbb{U} \to \mathbb{U}$ v rámci jednoho prostoru se nazývá endomorfismus. Endomorfismy, které jsou současně izomorfismy, se nazývají *automorfismy*. Každý automorfismus je reprezentovaný regulární maticí a naopak každá regulární matice reprezentuje nějaký automorfismus. Automorfismy jsou tedy důležitá lineární zobrazení, s kterými jsme se setkali na řadě míst v tomto textu. Složení dvou automorfismů je opět automorfismus, a množina všech automorfismů vektorového prostoru má velice hezké matematické vlastnosti.

> Izomorfismus je bijektivní lineární zobrazení, které páruje vektory dvou podprostorů. Existence izomorfismu implikuje totožnou algebraickou strukturu. Bijektivní endomorfismy se nazývají automorfismy a jsou reprezentovány regulárními maticemi.

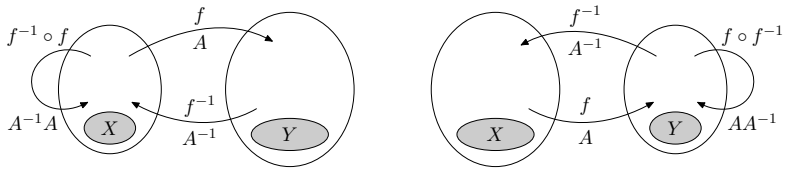
Inverzní zobrazení. Často chceme pro lineární zobrazení nalézt jeho inverzi. Obecně pro zobrazení f je potřeba rozlišovat jeho levou inverzi $f^{-1} \circ f = \mathrm{id}$ a jeho pravou inverzi $f \circ f^{-1} = \mathrm{id}$. Rozdíl mezi nimi je ten, že levá inverze invertuje zobrazení provedené f, zatímco pravá inverze f^{-1} je invertována zobrazením f. Zobrazení může mít pouze jednu z nich, klidně neurčenou jednoznačně. Pokud však má obě, jsou tyto inverze určené jednoznačně a rovnají se; důkaz je stejný jako důkaz lemmatu 3.8, neboť skládání zobrazení je asociativní. Ostatně to je důvod, proč levou a pravou inverzi označujeme stejným symbolem f^{-1}. Jako cvičení 5.3 si čtenář může rozmyslet, že inverzní zobrazení k lineárnímu je zase lineární. Jak čtenář asi očekává, bude inverzní zobrazení reprezentováno inverzní maticí.

Zaměřme se nejprve na levou inverzi. Máme matici A reprezentující f vůči bázím X a Y. Hledáme neznámou matici R, aby RA reprezentovalo identitu vůči jedné bázi X. Povšimněme si, že identita vůči jedné bázi je vždy reprezentována jednotkovou maticí I_n, neboť $\mathrm{id}(\boldsymbol{x}_i) = \boldsymbol{x}_i$. Tedy hledaná matice R musí splňovat rovnost $RA = I_n$, což je přesně definice levé inverze. V případě pravé inverze dostáváme identitu reprezentovanou I_m vůči jedné bázi Y, tedy $AS = I_m$ a S je přesně pravá inverze matice. Dvě inverze jsou nakresleny na obrázku 5.6.

> Nechť f je lineární zobrazení reprezentované A vůči X a Y. Inverzní zobrazení f^{-1} je reprezentované vůči bázím Y a X inverzní maticí A^{-1}.

Následující fakt platí obecně pro libovolné zobrazení v matematice:

[3] Obecně libovolné bijektivní zobrazení z U do V páruje prvky množiny U a množiny V; ostatně pomocí existence bijekce se v teorii množin definuje totožná velikost. Oproti tomu izomorfismus respektuje dodatečnou algebraickou strukturu $\mathbb{U}$ a zobrazuje ji do struktury $\mathbb{V}$. Například $\mathbb{R}$ a $\mathbb{R}^2$ jsou neizomorfní vektorové prostory, ale existuje mezi nimi bijektivní zobrazení; čtenář si může rozmyslet ve cvičení 5.5.



Obrázek 5.6: Nalevo je naznačena levá inverze, napravo pravá inverze. V obrázcích se nejprve aplikuje horní šipka, poté dolní šipka. Oba obrázky obsahují také šipku pro složené zobrazení.

- Levá inverze existuje, právě když je zobrazení prosté.
- Pravá inverze existuje, právě když je zobrazení na.
- Oboustranná inverze tedy existuje, právě když je zobrazení bijekce.

V případě lineárních zobrazení to koresponduje k našim zjištěním z kapitoly 3. Tvrzení 3.4 říká, že pravá inverze existuje, právě když soustava $A\boldsymbol{x} = \boldsymbol{b}$ má řešení pro libovolnou pravou stranu. To ale v řeči lineárních zobrazení přesně říká, že $A : \boldsymbol{x} \mapsto A\boldsymbol{x}$ má pravou inverzi, právě když je na.

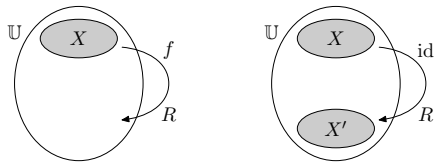
Připomeňme, že čtvercové matice odpovídají endomorfismům $f : \mathbb{U} \to \mathbb{U}$. Věta 3.7 říká, že lineární zobrazení, které je endomorfismus, je buď bijektivní (tedy má oboustrannou inverzi a reprezentující matice je regulární), nebo není ani prosté, ani na (a nemá ani jednu z inverzí). Toto dává smysl: Pokud $f(\boldsymbol{x}) = f(\boldsymbol{y})$ pro $\boldsymbol{x} = \boldsymbol{y}$, potom nebudeme mít dost vektorů, aby endomorfismus byl na. Toto ale pochopitelně není důkaz, neboť je potřeba využít linearitu! Čtenář si může rozmyslet, že může existovat nelineární zobrazení $g : \mathbb{U} \to \mathbb{U}$, které je na a není prosté, a naopak. S linearitou to však není možné.

Matice přechodu. Uvažme regulární matici R. Ta může reprezentovat automorfismus vektorového prostoru vůči jedné bázi X. Existuje však i jiná interpretace naznačená na obrázku 5.7, která je užitečná a často se používá. Uvažovaný automorfismus je identita, kterou ale uvažujeme vůči dvěma bázím X a X'. V takovém případě se matice R nazývá *matice přechodu*. Matice přepočítává souřadnice $\boldsymbol{\alpha}$ vůči bázi X na souřadnice $\boldsymbol{\beta}$ vůči bázi X':

$$R\boldsymbol{\alpha} = \boldsymbol{\beta}.$$

V této interpretaci obsahuje i-tý sloupec vyjádření $\boldsymbol{\beta}_i$ vektoru $\boldsymbol{x}_i$ vůči bázi X'. Maticové násobení pak provede lineární kombinaci těchto vyjádření podle koeficientů $\alpha_1, \ldots, \alpha_n$, tedy

$$\boldsymbol{\beta} = \alpha_1\boldsymbol{\beta}_1 + \cdots + \alpha_n\boldsymbol{\beta}_n.$$



Obrázek 5.7: Nalevo je matice R interpretována jako automorfismus vůči jedné bázi X. Napravo je interpretována jako matice přechodu od báze X a X'. Uvažovaný automorfismus je identita vůči bázím X a X'.

Dává smysl, že matice přepočítávající souřadnice musí reprezentovat automorfismus, což je bijektivní lineární zobrazení. Pokud by totiž zobrazení nebylo prosté, sdílely by dva vektory souřadnice vůči

bázi X', což není možné. A pokud by nebylo na, existoval by vektor vyjádřitelný v souřadnicích X' a nevyjádřitelný vůči bázi X.

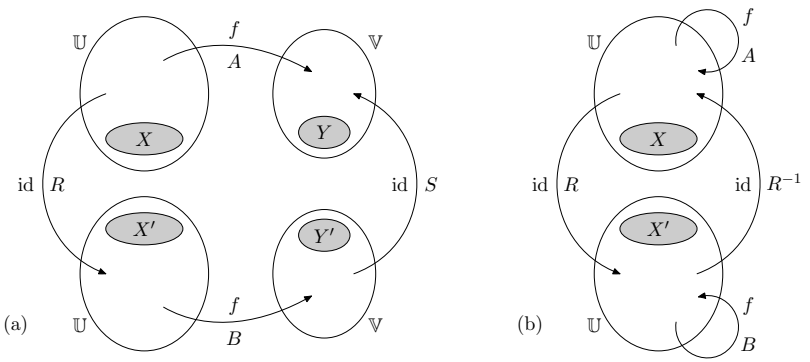
> Regulární matice mohou reprezentovat identický automorfismus, který přepočítává souřadnice vůči bázi X na souřadnice vůči bázi Y.

Ekvivalence a podobnost. Zmínili jsme, že jedno lineární zobrazení může mít pro různé volby bází různé maticové reprezentace. Maticové reprezentace však nemohou být úplně libovolné, protože jsou omezené strukturou lineárního zobrazení. Například pro libovolný automorfismus je každá reprezentující matice regulární, i když můžeme dostat různé regulární matice pro různé volby X a Y. Dvě matice nazveme *ekvivalentní*, pokud reprezentují stejné zobrazení vůči jiné volbě bází. Zkusíme nyní blíže pochopit tuto maticovou ekvivalenci.

Mějme zobrazení f reprezentované maticí A vůči X a Y a maticí B vůči X' a Y'. Můžeme pochopitelně uvažovat matice přechodu mezi těmito bázemi. Nechť R je matice přechodu od báze X k bázi X' a nechť S je matice přechodu od Y' k Y. Potom složením těchto zobrazení dostáváme rovnost

$$A = SBR, \qquad \text{neboli} \qquad B = S^{-1}AR^{-1}. \tag{5.4}$$

Pro lepší představu se podívejte na obrázek 5.8(a).



Obrázek 5.8: (a) Dvě ekvivalentní matice A a B spolu s dvěma maticemi přechodu. (b) Dvě podobné matice A a B reprezentující stejný endomorfismus.

Alternativní definice maticové ekvivalence A a B je, že existují regulární matice R a S splňující rovnost (5.4). Z této alternativní definice je snadné nahlédnout, že maticová ekvivalence je skutečně ekvivalence,[4] která rozděluje matice do tříd ekvivalencí. Každé dvě matice ze stejné třídy jsou ekvivalentní a žádné dvě matice z různých tříd nejsou ekvivalentní. Platí následující charakterizace ekvivalentních matic:

Tvrzení 5.2. *Dvě matice $A, B \in \mathbb{R}^{m\times n}$ jsou ekvivalentní, právě když mají stejnou hodnost.*

[4] V matematice je relace na množině prvků X ekvivalencí, pokud je tato relace reflexivní, symetrická a tranzitivní. Čtenář může zkusit dokázat, že tyto vlastnosti maticová ekvivalence splňuje. Také může zkusit dokázat, že prvky jsou v každé ekvivalenci rozděleny na třídy navzájem ekvivalentních prvků.

Důkaz. Protože násobení regulární maticí zachovává hodnost, musejí mít ekvivalentní matice stejnou hodnost. Druhou implikaci ukážeme tak, že každá matice A hodnosti k je matice ekvivalentní obdélníkové diagonální matici K, která má na diagonále nejprve k jedniček a zbytek diagonály jsou nuly. Z tranzitivity ekvivalence pak vyplývá, že A je ekvivalentní B.

Nejprve matici převedeme řádkovými úpravami do odstupňovaného tvaru, tedy získáme její LU dekompozici $A = P^{\mathsf{T}}LDU$. Matice U má všechny pivoty jedničkové. Dále sloupcovými úpravami převedeme U na výše popsanou diagonální matici. Přeházíme sloupečky s pivoty vlevo podle pořadí, poté vyeliminujeme všechny ostatní nenulové prvky vyjma diagonály. Platí $U = K(P'U')$. Tedy A je ekvivalentní s K, neboť

$$A = (P^{\mathsf{T}}LD)U = (P^{\mathsf{T}}LD)K(P'U') = SKR. \qquad \square$$

Protože endomorfismy se často reprezentují maticí pouze vůči jedné bázi, tedy $X = Y$, zavádí se podobná definice *maticové podobnosti*. Dvě čtvercové matice A a B jsou podobné, pokud reprezentují stejný endomorfismus vůči různým bázím X a X'. Schéma podobnosti je naznačeno na obrázku 5.8(b). Pokud chceme zobrazit maticí A, můžeme alternativně přejít maticí R od báze X k bázi X', zobrazit maticí B a poté přejít zpátky od báze X' do X' maticí R^{-1}. Je snadné si všimnout, že inverzní matice přechodu přechází mezi bázemi v obráceném směru. Proto dostáváme, že dvě matice A a B jsou si podobné, pokud

$$A = R^{-1}BR, \qquad \text{neboli} \qquad B = RAR^{-1}.$$

Protože maticová podobnost je také ekvivalentní relace, rozděluje čtvercové matice na třídy ekvivalence reprezentující jednotlivá lineární zobrazení vůči různým bázím. Tentokrát však není tak snadné charakterizovat jednotlivé třídy jako v případě maticové ekvivalence. Tříd ekvivalence je tentokrát mnohem víc a jejich porozumění se bude zaobírat velká část tohoto textu.

> Dvě matice jsou ekvivalentní, pokud reprezentují stejné lineární zobrazení vůči různým volbám bází. Dvě čtvercové matice jsou podobné, pokud reprezentují stejný endomorfismus (při volbě pouze jedné báze). Pro regulární matice R a S,
>
> *ekvivalence:* $A = SBR$,
> *podobnost:* $A = R^{-1}BR$.

5.2 Revize fundamentálních podprostorů

V kapitole 4 jsme popsali fundamentální podprostory a hodnost matice. Tyto pojmy jsou mnohem přirozenější v kontextu lineárních zobrazení. Zkusíme je přeformulovat, ukázat jiné důkazy a geometrický náhled. Pokusíme se získat lepší vhled do struktury a vlastností lineárních zobrazení.

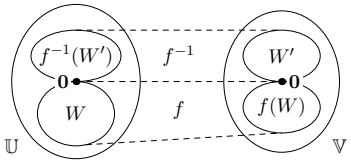
Lineární zobrazení a podprostory. Zmínili jsme, že podmínka linearity byla zvolena tak, aby lineární zobrazení zachovávalo strukturu vektorového prostoru. V kapitole 2 jsme se zabývali strukturou vektorových prostorů a jeden z hlavních pojmů byl podprostor. Pojďme ukázat, že se lineární zobrazení chovají pěkně vůči vektorovým podprostorům.

Nechť $f : \mathbb{U} \to \mathbb{V}$ je lineární zobrazení a nechť W je vektorový podprostor $\mathbb{U}$. Jeho obraz $f(W)$ je opět vektorový podprostor, což nyní dokážeme. Řekněme, že $\boldsymbol{b}, \boldsymbol{c} \in f(W)$. Podle definice existují $\boldsymbol{x}, \boldsymbol{y} \in W$, pro které $f(\boldsymbol{x}) = \boldsymbol{b}$ a $f(\boldsymbol{y}) = \boldsymbol{c}$. Protože W je podprostor, platí $\boldsymbol{x} + \boldsymbol{y} \in W$. Z linearity dostáváme, že

$$f(\boldsymbol{x} + \boldsymbol{y}) = f(\boldsymbol{x}) + f(\boldsymbol{y}) = \boldsymbol{b} + \boldsymbol{c},$$

tedy $\boldsymbol{b}+\boldsymbol{c}$ také leží v $f(W)$. Podobně lze ukázat, že $f(W)$ je uzavřené na násobení skalárem. Navíc obrazem generátoru W je generátor $f(W)$. Obrazem báze W však nemusí být báze $f(W)$, neboť nezávislost se může zobrazením báze ztratit.

Podobný vztah však platí i obráceným směrem; vzorem vektorového podprostoru je podprostor. Nechť W' je podprostor $\mathbb{V}$. Potom $f^{-1}(W')$ je vektorový podprostor $\mathbb{U}$. Důkaz je velice podobný, stačí opět ukázat uzavřenost $f^{-1}(W')$ na vektorové operace. Čtenář si může rozmyslet detaily jako cvičení. Zachovávání podprostorů je ilustrováno na obrázku 5.9.



Obrázek 5.9: Důležitá vlastnost lineárních zobrazení je, že obrazem a vzorem libovolného podprostoru je podprostor.

Obraz $f(W)$ libovolného podprostoru $W \subseteq \mathbb{U}$ je podprostor $\mathbb{V}$ a vzor $f^{-1}(W')$ libovolného podprostoru $W' \subseteq \mathbb{V}$ je podprostor $\mathbb{U}$.

Na konci kapitoly 2 jsme uvažovali pro jeden vektorový prostor všechny jeho podprostory uspořádané inkluzí. Ukázali jsme, že tvoří svaz, tedy že existují suprema a infima libovolné množiny podprostorů. Libovolné lineární zobrazení $f : \mathbb{U} \to \mathbb{V}$ vnořuje svaz podprostorů $\mathbb{U}$ do svazu podprostorů $\mathbb{V}$. Důvodem je, že lineární zobrazení zachovává inkluzi, neboť pro $W \subseteq W'$ platí $f(W) \subseteq f(W')$. Čtenář si může rozmyslet, že linearita zobrazení implikuje, že struktura suprem/infim je přenesena; obraz suprema je roven supremu obrazů a podobně pro infima. Tedy zobrazení f *indukuje* homomorfismus svazů.[5]

Poznamenejme, že podobné vlastnosti neplatí pouze pro vektorové prostory a lineární zobrazení, ale obecně pro libovolné algebraické struktury. Homomorfismy zachovávají velkou řadu vlastností těchto struktury.

Jádro a obraz. Pro pochopení homomorfismů matematických struktur je velice užitečné uvažovat jádro a obraz. *Jádro* je množina vzorů nuly (pokud dává ve struktuře smysl) a *obraz* je obraz celé struktury po aplikování homomorfismu. Pro vektorové prostory $\mathbb{U}$ a $\mathbb{V}$ a lineární zobrazení $f : \mathbb{U} \to \mathbb{V}$ definujeme

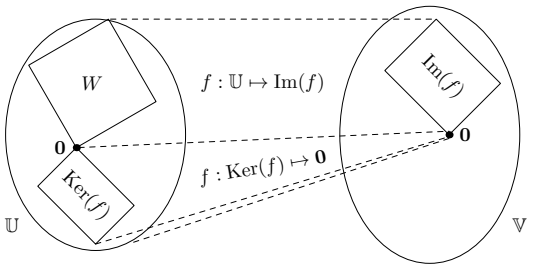
$$\mathrm{Ker}(f) = f^{-1}(\mathbf{0}) \qquad \text{a} \qquad \mathrm{Im}(f) = f(\mathbb{U}).$$

Tyto dva podprostory jsou naznačeny na obrázku 5.10.

V kapitole 3 jsme definovali $\mathrm{Ker}(A)$ jako množinu všech řešení soustavy $A\boldsymbol{x} = \mathbf{0}$. To jsou v řeči lineárních zobrazení přesně vzory nuly, tedy abstraktní $\mathrm{Ker}(f)$ se shoduje s definicí $\mathrm{Ker}(A)$. Navíc dostáváme zdarma tvrzení 3.9, že $\mathrm{Ker}(A)$ je podprostor, neboť $\{\mathbf{0}\}$ je podprostor a vzor libovolného podprostoru je podprostor. Podobně definice $\mathrm{Im}(A)$ z kapitoly 4 jako množina pravých stran $\boldsymbol{b}$, pro kterou má soustava $A\boldsymbol{x} = \boldsymbol{b}$ řešení, je totožná s abstraktní definicí $\mathrm{Im}(f)$. Protože obrazem libovolného podprostoru je podprostor, dostáváme zdarma, že $\mathrm{Im}(A)$ je podprostor.

Jádro je tvořeno přesně množinou vektorů, které se zobrazí na nulu. Každý vektor, který je mimo jádro, se zobrazí mimo nulu. Intuitivně čím menší je jádro, tím větší počet vektorů se zobrazí mimo nulu, a tedy tím větší je obraz. Tato intuice je zcela správná, neboť v řeči dimenzí dostáváme:

[5] Co znamená slovo indukuje? Zobrazení f není přímo homomorfismus svazů, protože je to homomorfismus z $\mathbb{U}$ do $\mathbb{V}$. Z jeho existence však přímo vyplývá jiný homomorfismus f' ze svazu $\mathbb{U}$ do svazu $\mathbb{V}$. Toto indukované zobrazení f' je definováno jako $f'(W) = f(W)$; rozdíl je, že f zobrazuje vektory, kdežto f' zobrazuje podprostory.



Obrázek 5.10: Dva podprostory $\mathrm{Ker}(f)$ a $\mathrm{Im}(f)$ klíčové pro libovolné lineárního zobrazení. Podprostor W jako doplněk $\mathrm{Ker}(f)$ z důkazu tvrzení 5.3.

Tvrzení 5.3. *Pro libovolné lineární zobrazení $f : \mathbb{U} \to \mathbb{V}$ platí, že*

$$\dim \mathrm{Im}(f) = \dim \mathbb{U} - \dim \mathrm{Ker}(f).$$

Důkaz. Zvolme libovolnou bázi $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ v jádře $\mathrm{Ker}(f)$ a doplňme ji vektory $\boldsymbol{x}_{k+1}, \ldots, \boldsymbol{x}_n$ na bázi $\mathbb{U}$. Označme $W = \mathcal{L}(\boldsymbol{x}_{k+1}, \ldots, \boldsymbol{x}_n)$. Tedy W je jakýsi doplněk $\mathrm{Ker}(f)$, neboť $W \oplus \mathrm{Ker}(f) = \mathbb{U}$. Dimenze W je rovna $\dim \mathbb{U} - \dim \mathrm{Ker}(f)$ a naším cílem je ukázat, že je totožná s dimenzí $\mathrm{Im}(f)$. Situace je naznačena na obrázku 5.10.

Zaměříme se na to, kam f zobrazuje vektory W. Zjevně $f(W)$ je podprostor $\mathrm{Im}(f)$. Platí však, že $f(W) = \mathrm{Im}(f)$, což nyní dokážeme. Mějme libovolný vektor $\boldsymbol{b} \in \mathrm{Im}(f)$. Pro něj existuje vzor $\boldsymbol{x} \in \mathbb{U}$ splňující $f(\boldsymbol{x}) = \boldsymbol{b}$. Vyjádříme $\boldsymbol{x}$ vůči bázi jako $\alpha_1\boldsymbol{x}_1 + \cdots + \alpha_n\boldsymbol{x}_n$. Vynulováním prvních k složek vektorů z $\mathrm{Ker}(f)$ dostáváme $\boldsymbol{y} = \alpha_{k+1}\boldsymbol{x}_{k+1} + \cdots + \alpha_n\boldsymbol{x}_n$, jehož obraz je totožný: $f(\boldsymbol{y}) = \boldsymbol{b}$. Tedy $f(W) = \mathrm{Im}(f)$.

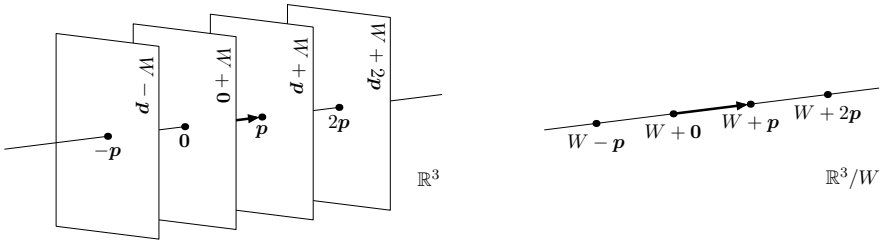
Nyní ukážeme, že f zobrazuje W prostě. Nechť $\boldsymbol{x}, \boldsymbol{y} \in W$ a $f(\boldsymbol{x}) = f(\boldsymbol{y})$. Potom $f(\boldsymbol{x} - \boldsymbol{y}) = \mathbf{0}$ z linearity, a tedy $\boldsymbol{x} - \boldsymbol{y} \in \mathrm{Ker}(f)$. Avšak $\boldsymbol{x} - \boldsymbol{y} \in W$, a protože $\mathrm{Ker}(f)$ a W jsou lineárně nezávislé, musí platit $\boldsymbol{x} = \boldsymbol{y}$. Dohromady získáváme, že f zobrazuje W izomorfně na $\mathrm{Im}(f)$. Tedy báze W se zobrazí na bázi $f(W)$, z čehož vyplývá, že $\dim W = \dim \mathrm{Im}(f)$. □

V kapitole 4 jsme získali výše uvedený vztah o velikosti dimenzí jiným způsobem. Věta 4.11 říká, že dimenze řádkového podprostoru a obrazu je stejná. Podle tvrzení 4.13 je součet dimenzí řádkového podprostoru a kernelu roven $\dim \mathbb{U}$. Složením získáme výše uvedené tvrzení. Zde popsaný důkaz je však z určitého pohledu elegantnější, protože vychází přímo z pojmů spojených s lineárním zobrazením a vůbec nepotřebuje pracovat s koeficienty reprezentace. Ostatně pořád nevíme, co to znamená řádkový podprostor lineárního zobrazení f, protože tento pojem je spojen s maticovou reprezentací.

Mimochodem v důkazu můžeme rozšířit bázi $\mathrm{Ker}(f)$ pomocí $\boldsymbol{x}_{k+1}, \ldots, \boldsymbol{x}_n$ libovolně. Mějme maticovou reprezentaci A lineárního zobrazení f. Podle tvrzení 4.14 je $\mathcal{R}(A)$ nezávislý s $\mathrm{Ker}(A)$. Proto můžeme zvolit vektory $\boldsymbol{x}_{k+1}, \ldots, \boldsymbol{x}_n$ z $\mathcal{R}(A)$. Dostáváme hezký vztah, že lineární zobrazení $A : \boldsymbol{x} \mapsto A\boldsymbol{x}$ zobrazuje $\mathcal{R}(A)$ bijektivně na $\mathrm{Im}(A)$. Libovolné bijektivní zobrazení lze invertovat, což umožňuje zavést *pseudoinverzi* A^+. Pokud je $\mathrm{Ker}(A)$ netriviální, levá inverze neexistuje a zobrazením $\boldsymbol{x} \to A\boldsymbol{x}$ jsou souřadnice $\boldsymbol{x}$ vůči $\mathrm{Ker}(A)$ ztraceny. Avšak můžeme alespoň invertovat souřadnice vůči $\mathcal{R}(A)$, což přesně dělá A^+. Přesný popis konstrukce A^+ vyžaduje další teorii a úzce souvisí s *metodou nejmenších čtverců*.

Lineární zobrazení $f : \mathbb{U} \to \mathbb{V}$ zobrazuje izomorfně doplněk $\mathrm{Ker}(f)$ na $\mathrm{Im}(f)$. Pro maticovou reprezentaci A můžeme jako doplněk zvolit řádkový prostor $\mathcal{R}(A)$. Zobrazení $\boldsymbol{x} \mapsto A\boldsymbol{x}$ je izomorfismus mezi $\mathcal{R}(A)$ a $\mathrm{Im}(A)$.

Faktorprostor a věta o homomorfismu. Abstraktní algebra popisuje jiný náhled do důkazu tvrzení 5.3. Uvažme libovolný podprostor $W \subseteq \mathbb{U}$. V kapitole 2 jsme popsali afinní podprostory $W + \boldsymbol{p}$ jako posunutí W o vektor $\boldsymbol{p}$. Navíc jsme ukázali, že $W + \boldsymbol{p} = W + \boldsymbol{q}$, právě když $\boldsymbol{p} - \boldsymbol{q} \in W$. Existuje matematická operace zvaná *faktorizace*, která umožňuje z $\mathbb{U}$ vyrobit pomocí W menší vektorový prostor. Tento prostor zvaný *faktorprostor* se značí $\mathbb{U}/W$ a jeho prvky jsou všechny afinní podprostory vzniklé posunutím W. Pro tyto afinní podprostory definujeme sčítání $(W + \boldsymbol{p}) + (W + \boldsymbol{q}) = W + (\boldsymbol{p} + \boldsymbol{q})$ a násobení skalárem $\alpha(W + \boldsymbol{p}) = W + (\alpha\boldsymbol{p})$. Čtenář může jako cvičení 5.4 dokázat, že vzniklý objekt je skutečně vektorový prostor a jeho dimenze je $\dim \mathbb{U} - \dim W$. Příklad je na obrázku 5.11.



Obrázek 5.11: Struktura afinních podprostorů $\mathbb{R}^3$ určených posunutím roviny W. Faktorprostor $\mathbb{R}^3/W$ má dimenzi jedna a je izomorfní $\mathbb{R}^1$. Můžeme si představit, že jednotlivé vektory $\mathbb{R}^3/W$ vzniknou geometricky sloučením jednotlivých afinních podprostorů do bodů.

Jak toto souvisí s výše uvedeným důkazem tvrzení 5.3? Vektorový prostor $\mathbb{U}/\mathrm{Ker}(f)$ má dimenzi $\dim \mathbb{U} - \dim \mathrm{Ker}(f)$, což je přesně dimenze $\mathrm{Im}(f)$. To pochopitelně není náhoda. Vektorové prostory $\mathbb{U}/\mathrm{Ker}(f)$ a $\mathrm{Im}(f)$ jsou totiž izomorfní, a proto musí mít stejnou dimenzi. Tyto podprostory mají stejnou algebraickou strukturu, i když jsou tvořené jinými prvky. Prvky $\mathbb{U}/\mathrm{Ker}(f)$ jsou afinní podprostory $\mathbb{U}$ vzniklé posunutím $\mathrm{Ker}(f)$. Na druhou stranu obraz $\mathrm{Im}(f)$ je tvořený prvky podprostoru $\mathbb{V}$.

V důkazu tvrzení 5.3 jsme ukázali, že f zobrazuje W izomorfně na obraz $\mathrm{Im}(f)$; tedy že toto zobrazení je prosté a na. Z pohledu faktorizace, vektory $\boldsymbol{x}_{k+1}, \ldots, \boldsymbol{x}_n$ generující W popisují všechna možná posunutí $\mathrm{Ker}(f)$, která vytvoří různé afinní podprostory. Každý afinní podprostor $\mathrm{Ker}(f) + \boldsymbol{x}$ je zobrazen f na jiný prvek $\mathrm{Im}(f)$. Toto je zmiňovaný izomorfismus mezi $\mathbb{U}/\mathrm{Ker}(f)$ a $\mathrm{Im}(f)$, který je obsažen v lineárním zobrazení f.

Připomeňme si tvrzení 2.4 z kapitoly 2. To říká, že pro libovolný vektor $\boldsymbol{b} \in \mathrm{Im}(f)$ je množina jeho vzorů $f^{-1}(\boldsymbol{b})$ afinní podprostor $\mathrm{Ker}(f) + \boldsymbol{x}$, kde $\boldsymbol{x}$ je libovolný vzor $\boldsymbol{b}$. Pochopitelně pro různé vektory $\boldsymbol{b}$ dostáváme různé množiny vzorů, tedy různé afinní podprostory. Na druhou stranu každý vektor $\boldsymbol{x} \in \mathbb{U}$ je vzorem pro nějaký vektor $\boldsymbol{b} \in \mathrm{Im}(f)$. Afinní podprostory vzniklé posunutím $\mathrm{Ker}(f)$ jsou spárované s vektory $\mathrm{Im}(f)$ podle izomorfismu.

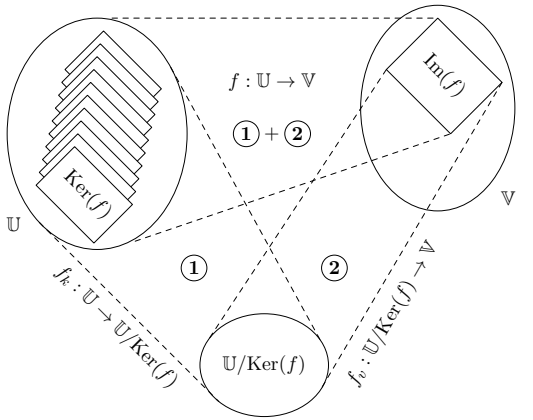
Tato vlastnost je natolik důležitá, že se jí říká věta o homomorfismu. Navíc podobné tvrzení platí i pro homomorfismy obecných algebraických struktur.

Věta 5.4 (o homomorfismu)**.** *Pro libovolné lineární zobrazení $f : \mathbb{U} \to \mathbb{V}$ platí*

$$\mathbb{U}/\mathrm{Ker}(f) \cong \mathrm{Im}(f).$$

Věta říká, že každé lineární zobrazení $f : \mathbb{U} \to \mathbb{V}$ lze přiloženě rozdělit na dvě části: *kvocientovou* $f_k : \mathbb{U} \to \mathbb{U}/\mathrm{Ker}(f)$ a *vnořovací* $f_v : \mathbb{U}/\mathrm{Ker}(f) \to \mathbb{V}$; ilustrováno na obrázku 5.12. Každé lineární zobrazení f vznikne složením těchto dvou lineárních zobrazení:

$$f = f_v \circ f_k.$$



Obrázek 5.12: Libovolné lineární zobrazení $f : \mathbb{U} \to \mathbb{V}$ je složení dvou lineárních zobrazení f_k a f_v. Vlevo nahoře je prostor $\mathbb{U}$ s vyznačenými afinními podprostory tvořenými posunutím $\mathrm{Ker}(f)$. Kvocientové zobrazení f_k slučuje tyto afinní podprostory do jednotlivých vektorů prostoru $\mathbb{U}/\mathrm{Ker}(f)$. Vnořovací zobrazení f_v zobrazuje jednotlivé prvky $\mathbb{U}/\mathrm{Ker}(f)$ prostě do $\mathrm{Im}(f)$, tedy funguje jako přejmenování.

Zaměřme se nejprve na kvocientové zobrazení f_k. Jeho název vychází z toho, že f_k zobrazuje do faktorprostoru $\mathbb{U}/\mathrm{Ker}(f)$, kterému se také někdy říká *kvocient*. Tento název dává smysl, ostatně operaci faktorizace značíme matematickým symbolem po dělení. Zobrazení f_k shlukuje jednotlivé afinní podprostory vzniklé posunutím $\mathrm{Ker}(f)$ do jednotlivých vektorů $\mathbb{U}/\mathrm{Ker}(f)$. Je jednoduché ukázat, že f_k je zobrazení na, neboť každý vektor $\mathbb{U}/\mathrm{Ker}(f)$ má jako vzor jeden afinní podprostor.

Vnořovací zobrazení f_v dostalo svůj název podle toho, že vnořuje prostor $\mathbb{U}/\mathrm{Ker}(f)$ do $\mathbb{V}$. Toto zobrazení je prosté, a proto je jeho obraz $\mathrm{Im}(f_v)$ izomorfní $\mathbb{U}/\mathrm{Ker}(f)$. Protože je f_k zobrazení na, platí $\mathrm{Im}(f) = \mathrm{Im}(f_v)$. Zobrazení f_v přiřadí různým vektorům z $\mathbb{U}/\mathrm{Ker}(f)$ různé obrazy z $\mathrm{Im}(f)$. Je to izomorfismus mezi těmito dvěma podprostory, který můžeme chápat jako přejmenování.

> Věta o homomorfismu říká, že každé lineární zobrazení $f : \mathbb{U} \to \mathbb{V}$ je složení $f_v \circ f_k$ dvou jednodušších lineárních zobrazení:
>
> - kvocientového zobrazení $f_k : \mathbb{U} \to \mathbb{U}/\mathrm{Ker}(f)$, které je na, a
> - vnořovacího zobrazení $f_v : \mathbb{U}/\mathrm{Ker}(f) \to \mathbb{V}$, které je prosté.
>
> Zobrazení f_v definuje izomorfismus mezi $U/\mathrm{Ker}(f)$ a $\mathrm{Im}(f)$.

Lineární zobrazení a dimenze. Klíčová vlastnost lineárního zobrazení je, že nikde „nerozšiřuje“ obraz. Co tím myslíme? Pro libovolný podprostor $W \subseteq \mathbb{U}$ platí $\dim W \geq \dim f(W)$. Obrazy podprostorů jsou do dimenze nejvýše tak velké jako podprostory samotné. Důvodem je, že obraz generátoru W generuje $f(W)$. Proto je báze $f(W)$ nejvýše tak velká jako báze W. Podobně pro libovolný podprostor $W' \subseteq \mathbb{V}$ platí, že $\dim W' \leq \dim f^{-1}(W')$.

Pokud f zobrazuje W prostě, platí mezi dimenzemi rovnost: $\dim W = \dim f(W)$. Pokud je celé zobrazení f prosté, platí tato rovnost pro libovolný podprostor. Ostatně alternativní definice prostého lineárního zobrazení je, že zachovává dimenzi libovolného podprostoru. Pokud zobrazení není prosté, pro některé podprostory W je $\dim W > \dim f(W)$.

Jaký je ale přesný vztah mezi dimenzemi W a $f(W)$? Ukážeme nejprve dolní odhad. Jak lze asi očekávat, odpověď bude souviset s velikostí jádra $\mathrm{Ker}(f)$.

Lemma 5.5. *Nechť $f : \mathbb{U} \to \mathbb{V}$ je lineární zobrazení. Pro libovolný podprostor $W \subseteq \mathbb{U}$ platí:*

$$\dim W - \dim \mathrm{Ker}(f) \leq \dim f(W) \leq \dim W.$$

Důkaz. Nechť $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ tvoří bázi W. Jejich obrazy $f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_k)$ generují $f(W)$, avšak nemusí být nezávislé. Z toho plyne horní odhad. Pro dolní odhad budeme odebírat nadbytečné vektory, dokud nezískáme bázi $f(W)$. Bez újmy na obecnosti až na přejmenování můžeme předpokládat, že jsme odebrali $f(\boldsymbol{x}_{\ell+1}), \ldots, f(\boldsymbol{x}_k)$ a zbývající vektory $f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_\ell)$ jsou lineárně nezávislé. Tedy tvoří bázi $f(W)$ a $\dim f(W) = \ell$.

Pro každé $j > \ell$ lze vektor $f(\boldsymbol{x}_j)$ vyjádřit jako lineární kombinaci $\sum_{i=1}^{\ell} \alpha_i f(\boldsymbol{x}_i)$. Tedy

$$\boldsymbol{0} = f(\boldsymbol{x}_j) - \sum_{i=1}^{\ell} \alpha_i f(\boldsymbol{x}_i) = f\left(\boldsymbol{x}_j - \sum_{i=1}^{\ell} \alpha_i \boldsymbol{x}_i\right).$$

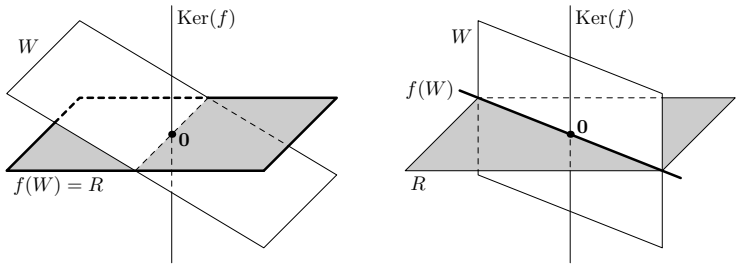
Z toho vyplývá, že $\boldsymbol{y}_j = \boldsymbol{x}_j - \sum_{i=1}^{\ell} \alpha_i \boldsymbol{x}_i$ leží v $\mathrm{Ker}(f)$. Čtenář si může rozmyslet, že lineární nezávislost vektorů $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ implikuje nezávislost $\boldsymbol{y}_{\ell+1}, \ldots, \boldsymbol{y}_k$. Proto $\dim \mathrm{Ker}(f) \geq k - \ell = \dim W - \dim f(W)$, z čehož úpravou získáváme dolní odhad. □

Z důkazu lze vyvodit přesnou hodnotu $\dim f(W)$. Ta je tím menší, čím více vektorů W leží v jádru $\mathrm{Ker}(f)$. Pokud W bude s $\mathrm{Ker}(f)$ zcela nezávislé, nastane rovnost dimenzí W a $f(W)$; přestože f nemusí vůbec být prosté, pouze lokálně vůči W. Následující důsledek zobecňuje tvrzení 5.3.

Důsledek 5.6. *Pro libovolné lineární zobrazení $f : \mathbb{U} \to \mathbb{V}$ a libovolný podprostor $W \subseteq \mathbb{U}$ platí, že*

$$\dim f(W) = \dim W - \dim\big(W \cap \mathrm{Ker}(f)\big).$$

Uvažme například kolmou projekci na rovinu R, což je geometrické zobrazení $f : \mathbb{R}^3 \to \mathbb{R}^3$. Jádro $\mathrm{Ker}(f)$ má dimenzi jedna a je to kolmá přímka k rovině R procházející počátkem. Z lemmatu 5.5 víme, že pro libovolný podprostor W platí, že $\dim W - 1 \leq \dim f(W) \leq \dim W$. Pokud W neobsahuje přímku tvořící jádro $\mathrm{Ker}(f)$, je nezávislý s $\mathrm{Ker}(f)$ a platí rovnost $\dim f(W) = \dim W$. Pokud ji W obsahuje, platí $\dim f(W) = \dim W - 1$. Obrázek 5.13 ukazuje obě situace.



Obrázek 5.13: Rovina R kolmé projekce je vyznačena šedě, podprostor W bíle a tučně jeho obraz $f(W)$. Nalevo je $W \cap \mathrm{Ker}(f)$ triviální a dimenze W je zobrazením zachována. Napravo W obsahuje $\mathrm{Ker}(f)$ a dimenze $f(W)$ je o jedna menší.

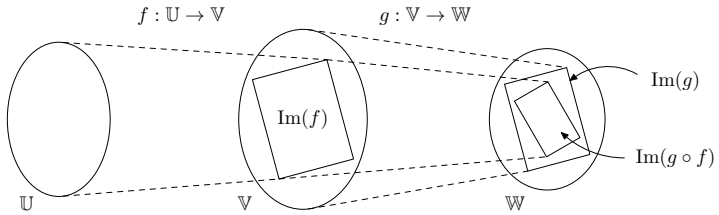
Hodnost zobrazení. Hodnost lineárního zobrazení $\mathrm{rank}(f)$ se definuje úplně stejně jako pro matice, tedy jako $\dim \mathrm{Im}(f)$. S využitím geometrie lineárního zobrazení můžeme lépe pochopit výsledky získané v kapitole 4. Předně $\dim \mathrm{Im}(f)$ je menší či rovna $\dim \mathbb{U}$ (podle lemmatu 5.5) a $\dim \mathbb{V}$ (protože $\mathrm{Im}(f) \subseteq \mathbb{V}$). Toto odpovídá nerovnosti

$$\mathrm{rank}(A) \leq \min\{n, m\}.$$

Horní odhad pro hodnost $\mathrm{rank}(A + B)$ bohužel nemá pěknou geometrickou interpretaci. Avšak nerovnost $\mathrm{rank}(AB) \leq \min\{\mathrm{rank}(A), \mathrm{rank}(B)\}$ má velice přirozené geometrické vysvětlení. Podle tvrzení 5.1 to odpovídá v řeči zobrazení nerovnosti

$$\mathrm{rank}(g \circ f) \leq \min\big\{\mathrm{rank}(f), \mathrm{rank}(g)\big\}.$$

Tato nerovnost je ilustrována na obrázku 5.14. Protože $\mathrm{Im}(g \circ f) \subseteq \mathrm{Im}(g)$, dostáváme, že $\mathrm{rank}(g \circ f) \leq \mathrm{rank}(g)$. Druhá nerovnost $\mathrm{rank}(g \circ f) \leq \mathrm{rank}(f)$ platí podle lemmatu 5.5, neboť $\mathrm{Im}(g \circ f) = g(\mathrm{Im}(f))$ a $\dim g(\mathrm{Im}(f)) \leq \dim \mathrm{Im}(f)$.



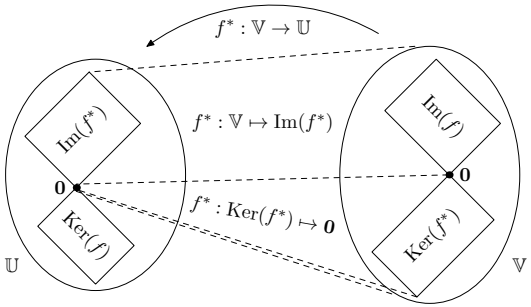
Obrázek 5.14: Geometrická ilustrace ukazuje, že složené zobrazení $g \circ f$ menší hodnost $\mathrm{rank}(g \circ f)$ než obě zobrazení f a g.

Pokud je f prosté, dostáváme rovnost $\mathrm{rank}(g \circ f) = \mathrm{rank}(g)$. Podobně pokud je g na, nastane rovnost $\mathrm{rank}(g \circ f) = \mathrm{rank}(f)$. Regulární matice odpovídají izomorfismům, které jsou prosté a zároveň na. Proto získáváme, že násobení regulární maticí zleva/zprava nemění hodnost, což je výsledek odvozený v kapitole 4.

Duální zobrazení. Podle kapitoly 4 jsou zbývající dva fundamentální podprostory řádkový podprostor a levé jádro. Tyto pojmy jsme vybudovali pomocí transponované matice $A^\intercal$, jejíž definice je závislá na maticové reprezentaci pomocí koeficientů. Abychom mohli tyto podprostory zavést v řeči abstraktních lineárních zobrazení, je potřeba zavést pojem *duálního zobrazení*. Existuje matematicky elegantní definice ve formě duálního vektorového prostoru a skalárního součinu. Tu si ale prozatím odpustíme, protože tyto pojmy nemáme vybudované. Místo toho můžeme duální zobrazení definovat s využitím maticové reprezentace přes transpozici.

Mějme libovolné zobrazení $f : \mathbb{U} \to \mathbb{V}$ a zvolme libovolně X bázi $\mathbb{U}$ a Y bázi $\mathbb{V}$. Duální zobrazení je lineární zobrazení $f^* : \mathbb{V} \to \mathbb{U}$, tedy zobrazuje v opačném směru než f. Již jsme se setkali s inverzní zobrazením $f^{-1} : \mathbb{V} \to \mathbb{U}$, avšak duální zobrazení je něco jiného. Například jeden rozdíl je, že duální zobrazení vždy existuje. Nechť je $A \in \mathbb{R}^{m \times n}$ je matice reprezentující f vůči X a Y. Duální zobrazení $f^* : \mathbb{V} \to \mathbb{U}$ je lineární zobrazení reprezentované maticí $A^\intercal$ vůči duálním bázím Y^* a X^*. Vektory těchto duálních bází jsou sloupcové vektory $Y^\intercal$ a $X^\intercal$, tedy řádkové vektory Y a X. Duální zobrazení f^* je určené jednoznačně nezávisle na volbě bází X a Y, což si čtenář může rozmyslet ve cvičení 5.6. Čtyři fundamentální podprostory lineárního zobrazení f jsou $\mathrm{Ker}(f)$, $\mathrm{Im}(f)$, $\mathrm{Ker}(f^*)$ a $\mathrm{Im}(f^*)$, nakresleny na obrázku 5.15.

Transpozici matice, definovanou v koeficientech matice, lze zavést i pro abstraktní lineární zobrazení. Reprezentuje duální zobrazení.



Obrázek 5.15: Duální zobrazeni f^* zobrazuje zprava doleva. Jsou vyobrazeny jeho dva fundamentální podprostory $\mathrm{Ker}(f^*)$ a $\mathrm{Im}(f^*)$.

5.3 ⋆Maticové reprezentace grafů

Ukážeme si aplikace získaných poznatků lineární algebry v teorii grafů. Pro to máme dvě motivace. Předně tyto aplikace jsou velice elegantní, dávají lepší vhled to některých aspektů grafů a mají přesah i do reálných aplikací. Druhá motivace je čistě pragmatická z pohledu lineární algebry. Zatím jsme si ukazovali lineární transformace v řeči geometrických transformací (rotace, roztažení, projekce, ...). Matice reprezentující grafy jsou elegantní příklady zajímavých lineárních zobrazení, na kterých můžeme ilustrovat vybudovanou teorii z jiného úhlu pohledu.

Graf je jednou ze základních matematických struktur a jejich zkoumáním se zabývá velká část diskrétní matematiky. Každý graf G sestává z množiny *vrcholů* $V(G)$ a množiny *hran* $E(G)$, kde každá hrana propojuje dvojici vrcholů. Například vrcholy mohou reprezentovat města na mapě a hrany jsou silnice mezi nimi. V chemii můžeme každou molekulu reprezentovat grafem, kde vrcholy jsou atomy a hrany reprezentují chemické vazby. Není proto překvapivé, že se grafy vyskytují na řadě míst v matematice a různých vědeckých disciplínách.

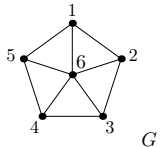
Mějme graf G s n vrcholy $\{1, \ldots, n\}$ a m hranami. Pokud je graf malý, přirozeně se reprezentuje obrázkem. Pro větší grafy si například můžeme pamatovat seznam všech hran, případně seznam sousedů pro každý vrchol zvlášť. V lineární algebře se graf typicky reprezentuje maticí, což lze udělat hned několika způsoby.

Matice sousednosti a Laplaceova matice. Často se používají následující dvě čtvercové matice $n \times n$. *Matice sousednosti* A_G je velice přirozená, naproti tomu *Laplaceova matice* L_G je trochu zvláštní. Nechť $d(i)$ označuje *stupeň* vrcholu i, což je počet jeho sousedů. Koeficienty těchto matic jsou definovány jako

$$(A_G)_{i,j} = \begin{cases} 1, & \text{pokud } ij \in E(G), \\ 0 & \text{jinak,} \end{cases} \qquad \text{a} \qquad (L_G)_{i,j} = \begin{cases} d(i), & \text{pokud } i = j, \\ -1, & \text{pokud } ij \in E(G), \\ 0 & \text{jinak.} \end{cases}$$

Alternativně $L_G = D_G - A_G$, kde D_G je diagonální matice se stupni jednotlivých vrcholů na diagonále. Příklady jsou na obrázku 5.16. Pro lepší seznámení s těmito maticemi může čtenář vyzkoušet následující dvě cvičení. Cvičení 5.7 ukazuje souvislost mocnin A_G^k se strukturou grafu, a cvičení 5.8 řeší vztah mezi Laplaceovou maticí a řezy v grafu.

Možná čtenáři přijde Laplaceova matice jako podivný způsob reprezentace grafu. Vždyť přeci neobsahuje žádnou další informaci, kterou by nešlo vyčíst z matice sousednosti. To je zcela pravda, avšak pouze z určitého pohledu. V úvodu kapitoly jsme zmínili, že každou matici lze interpretovat dvěma



G

$$A_G = \begin{pmatrix} & 1 & & & 1 & 1 \\ 1 & & 1 & & & 1 \\ & 1 & & 1 & & 1 \\ & & 1 & & 1 & 1 \\ 1 & & & 1 & & 1 \\ 1 & 1 & 1 & 1 & 1 & \end{pmatrix}, \qquad L_G = \begin{pmatrix} 3 & -1 & & & -1 & -1 \\ -1 & 3 & -1 & & & -1 \\ & -1 & 3 & -1 & & -1 \\ & & -1 & 3 & -1 & -1 \\ -1 & & & -1 & 3 & -1 \\ -1 & -1 & -1 & -1 & -1 & 5 \end{pmatrix}.$$

Obrázek 5.16: Graf G spolu s maticemi A_G a L_G. V maticích vynecháváme nulové koeficienty.

způsoby. Náš dosavadní náhled byl na tyto matice jako na tabulky čísel, popisující data k danému grafu. Z tohohle pohledu je veškerá informace L_G obsažena v A_G. Druhý pohled je však přes lineární zobrazení, kdy každá z matic popisuje jiný proces nad grafem. Obě matice popisují lineární zobrazení z $\mathbb{R}^n$ do $\mathbb{R}^n$. Koeficienty vektoru $\boldsymbol{x} \in \mathbb{R}^n$ můžeme chápat jako ohodnocení jednotlivých vrcholů. Každá z matic tato ohodnocení transformuje jiným způsobem.

Matice sousednosti jako transformace. Zaměřme se nejprve na matici sousednosti. Ta definuje lineární zobrazení $A_G : \mathbb{R}^n \to \mathbb{R}^n$. Nechť $\boldsymbol{x}$ je libovolný vektor a nechť $\boldsymbol{y} = A_G\boldsymbol{x}$. Z definice A_G a maticového násobení platí, že

$$y_i = \sum_{ij \in E(G)} x_j.$$

Transformace sčítá pro každý vrchol ohodnocení všech jeho sousedů. Alternativně každý vrchol rozešle svoje ohodnocení všem svým sousedů a zároveň přijme jejich ohodnocení. Zkoumáním algebraických vlastností matice A_G se můžeme dozvědět spoustu globálních vlastností grafu G.

Ukažme ještě souvislost matice A_G s modelováním *pravděpodobnosti*. Pomocí pravděpodobnosti chceme předpovídat reálný svět. Například uvažme hrací kostku. Výsledek hodu ovlivňuje celá řada faktorů: poloha kostky; směr hodu; místo, kam kostka dopadne; síla hodu; přesné rozmístění atomů kostky a povrchu; .... Problém je, že těchto faktorů je hrozně moc a jsou neuvěřitelně složité. Ani nejmodernější fyzika není schopná predikovat výsledek hodu. Myšlenka pravděpodobnosti je, že vytvoříme jednoduchý model kostky třeba tak, že si řekneme, že každá z šesti možností padá stejně často, tedy v jedné šestině hodů dostaneme každé ze šesti čísel. Tento model samozřejmě neumožňuje předpovědět, jak konkrétní hod dopadne. Pokud však hodů provedeme hodně, popisuje model velice dobře očekávané chování. Tedy zkoumáním i takto jednoduchého matematického modelu se můžeme ledacos o reálném světě dozvědět a na tomto principu například fungují kasína.

Jak tohle souvisí s maticí sousednosti? Tuto matici lze snadno modifikovat na matici P_G, která popisuje *náhodnou procházku* nad grafem. To je matematický model částice, která se pohybuje v grafu a v každém vrcholu si rovnoměrně vybere další směr cesty ze všech jeho sousedů. Tento model má řadu aplikací například ve fyzice. Matice P_G se získá z matice A_G vydělením každého sloupce i číslem $d(i)$. Například pro matici A_G z obrázku 5.16 získáme P_G vynásobením prvních pěti sloupců $\frac{1}{3}$ a posledního sloupce $\frac{1}{5}$. Povšimněte si, že součet v každém sloupci P_G je přesně jedna, ale součty v jednotlivých řádcích mohou být odlišné.

Uvažme například situaci, kdy je částice na začátku umístěna ve vrcholu i. Potom po jednom kroku náhodné procházky dostáváme *distribuci* pravděpodobnosti jednotlivých výskytů jako $P_G\boldsymbol{e}_i$, obecně po k krocích jako $P_G^k\boldsymbol{e}_i$. Důvod je následující. Pro $\boldsymbol{y} = P_G\boldsymbol{x}$ platí, že

$$y_i = \sum_{ij \in E(G)} \frac{1}{d(j)} x_j.$$

Částice stojí ve vrcholu j s pravděpodobností x_j. Hodnota $1/d(j)$ odpovídá pravděpodobnosti, že si v j-tém vrcholu vybere hranu vedoucí do i. Dohromady získáváme, že se částice pohne z j do i s pravděpodobností $\frac{1}{d(j)}x_j$. Tedy y_i je správně pravděpodobnost, že částice po jednom kroku stojí v i.

Přirozená otázka je, jak se tato distribuce vyvine, pokud uděláme velké množství kroků náhodné procházky. Tedy například jestli existuje

$$\lim_{k\to\infty} P_G^k \boldsymbol{e}_i$$

a jak případně tato limita vypadá. Tyto a další podobné otázky lze odpovědět pomocí metod lineární algebry. Příklad distribucí pro náhodnou procházku je tabulce 5.1.

$\boldsymbol{x}$	$P_G\boldsymbol{x}$	$P_G^2\boldsymbol{x}$	$P_G^3\boldsymbol{x}$	$P_G^4\boldsymbol{x}$	$P_G^5\boldsymbol{x}$	$P_G^6\boldsymbol{x}$	$P_G^7\boldsymbol{x}$	$P_G^8\boldsymbol{x}$	$P_G^9\boldsymbol{x}$	$\lim P_G^k\boldsymbol{x}$
1.0	0.0	0.2889	0.0889	0.1852	0.1317	0.1599	0.1447	0.1529	0.1485	0.15
0.0	0.3333	0.0667	0.2000	0.1235	0.1646	0.1421	0.1543	0.1477	0.1512	0.15
0.0	0.0	0.1778	0.1259	0.1605	0.1440	0.1531	0.1483	0.1509	0.1495	0.15
0.0	0.0	0.1778	0.1259	0.1605	0.1440	0.1531	0.1483	0.1509	0.1495	0.15
0.0	0.3333	0.0667	0.2000	0.1235	0.1646	0.1421	0.1543	0.1477	0.1512	0.15
0.0	0.3333	0.2222	0.2593	0.2469	0.2510	0.2497	0.2501	0.2500	0.2500	0.25

Tabulka 5.1: Vývoj pravděpodobnostní distribuce náhodné procházky a limitní distribuce pro graf z obrázku 5.16 s počátečním vrcholem 1. Všimněte si, že je výrazně pravděpodobnější, že se částice bude nacházet v centrálním vrcholu, což dává smysl.

Poznamenejme, že pokud se čtenáři nelíbí pravděpodobnost, popisuje P_G i zcela reálný proces. Představme si, že na začátku je nějaké množství vody rozmístěné mezi vrcholy grafu, v poměrech podle vektoru $\boldsymbol{x}$. Potom P_G popisuje proces, při kterém každý vrchol rovnoměrně rozdělí svoje množství vody a pošle ho sousedům, a zároveň od svých sousedů přijme nějaké množství vody. Vektor $P_G^k\boldsymbol{x}$ udává distribuci vody po k krocích. Přirozená otázka je, jestli takový systém konverguje k *equilibriu*, což je ustálený stav, kdy se distribuce vody v jednotlivých krocích nemění. Model s výměnou vody odpovídá situaci, kdy ve výše uvedeném modelu náhodné procházky sledujeme spoustu částic, u kterých očekáváme, že se typicky rovnoměrně rozdělí mezi sousedy daného vrcholu. Takové situace například nastane při pohybu spousty elektronů v elektrické síti.

Laplaceova matice jako transformace. Nyní uvažme tajemnější Laplaceovu matici jako transformaci $L_G : \mathbb{R}^n \to \mathbb{R}^n$. Nechť $\boldsymbol{y} = L_G\boldsymbol{x}$, potom rozepsáním a přeuspořádáním maticového násobení dostáváme:

$$y_i = d(i)\cdot x_i - \sum_{ij\in E(G)} x_j = d(i)\cdot\left(x_i - \frac{1}{d(i)}\sum_{ij\in E(G)} x_j\right).$$

Nemá smysl iterovat vektor Laplaceovou maticí. Představme si však, že jednotlivé vrcholy umístíme do různých výšek, kde i-tý vrchol je ve výšce $\boldsymbol{x}_i$. Potom $\boldsymbol{y}_i$ udává, o kolik je výška vrcholu i vyšší než průměrná výška jeho sousedů. (Výsledek je vynásobený faktorem $d(i)$, který pro jednoduchost můžeme ignorovat.)

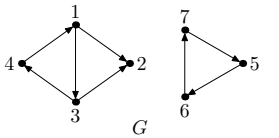
Laplaceova matice popisuje *Laplaceův operátor* Δ, který hraje klíčovou roli v řadě oblastí fyziky souvisejících s difuzí. Například v termodynamice popisuje druhý zákon termodynamiky, že se teploty jednotlivých míst průměrují. Pokud je nějaké místo teplejší než jeho okolí, bude se postupně ochlazovat, a podobně naopak. Uvažujeme-li situaci diskrétně nad grafem, dostaneme Laplaceovu matici. Pokud x_i jsou teploty jednotlivých vrcholů, hodnota y_i udává, jak moc se teplota daného vrcholu liší od průměru na jeho okolí, a tedy jak moc rychle se bude měnit. Laplaceova matice a Laplaceův operátor se objevují v celé řadě dalších oblastí fyziky, například při studiu elektřiny a magnetismu nebo při difuzy chemických látek.

Zkusme vyřešit podobnou otázku jako předtím, jak bude vypadat equlibrium s ustálenou distribucí teplot. Libovolné equilibrium $\boldsymbol{x}$ splňuje *Laplaceovu rovnici* $L_G\boldsymbol{x} = \boldsymbol{0}$. To platí pouze řešení se stejnou teplotou v celém grafu a důvod je následující. Pokud by existoval vrchol, jehož ohodnocení by bylo vyšší,

než ohodnocení některého z jeho sousedů, existoval by další jeho soused s ještě vyšším ohodnocením. Tento argument můžeme opakovat a získáme tak nekonečnou posloupnost vrcholů se zvyšujícím se ohodnocením. To není možné, protože předpokládáme, že graf je konečný. (V argumentu předpokládáme, že graf je *souvislý*, tedy neobsahuje oddělené části, mezi kterými by nevedly žádné hrany. V opačném případě může mít equilibrium v každé *komponentě souvislosti* jiné ohodnocení.)

V praxi jsou mnohem zajímavější situace s *okrajovými podmínkami*. Řekněme, že pro nějakou podmnožinu *okrajových vrcholů* dostaneme předepsané teploty. Představme si například, že máme nějaký systém, který je v těchto vrcholech zahřívaný. Zajímá nás, jak bude vypadat equlibrium za okrajových podmínek. Tedy hledáme ohodnocení teplot pro zbývající *vnitřní vrcholy*, které bude splňovat $L_G\boldsymbol{x} = \boldsymbol{0}$ pro všechny vnitřní vrcholy. Protože se jedná o soustavu lineárních rovnic, umíme tento problém efektivně vyřešit pomocí počítače. Ve spoustě aplikací se také objevuje řešení *Poissonovy rovnice* $L_G\boldsymbol{x} = \boldsymbol{b}$, které odpovídá předepsaným tokům teplot.

Matice incidence. Zkusme se zaměřit na fundamentální podprostory. Ty nejsou pro předcházející dvě čtvercové matice příliš zajímavé. Existuje však ještě další obdélníková matice reprezentující graf, jejíž fundamentální podprostory si teď popíšeme. Této matici $n \times m$ se říká *matice incidence* a značí se I_G. Nejprve si zvolme libovolnou orientaci hran grafu, na konkrétní volbě nezáleží, ale nějakou potřebujeme. Každý řádek odpovídá jednomu vrcholu a každý sloupeček odpovídá jedné hraně. Nechť k-tý sloupeček odpovídá hraně zorientované z i do j. Potom tento sloupeček obsahuje přesně dva nenulové koeficienty, a to -1 v i-tém řádku a 1 v j-tém řádku. Příklad matice incidence je na obrázku 5.17. Čtenář si může rozmyslet, že nezávisle na zvolené orientaci platí $I_G I_G^{\mathsf{T}} = L_G$.



$$I_G = \begin{pmatrix} -1 & -1 & & 1 & & & & \\ 1 & & 1 & & & & & \\ & 1 & -1 & & -1 & & & \\ & & & -1 & 1 & & & \\ & & & & & -1 & & 1 \\ & & & & & 1 & -1 & \\ & & & & & & 1 & -1 \end{pmatrix}.$$

Obrázek 5.17: Graf G se zorientovanými hranami spolu s maticí incidence I_G.

Abychom pochopili, jak vypadají fundamentální podprostory I_G, budeme uvažovat zobrazení $I_G : \mathbb{R}^m \to \mathbb{R}^n$ a duální zobrazení $I_G^{\mathsf{T}} : \mathbb{R}^n \to \mathbb{R}^m$. Každý vektor $\boldsymbol{x} \in \mathbb{R}^n$ je ohodnocení vrcholů a každý vektor $\boldsymbol{y} \in \mathbb{R}^m$ popisuje ohodnocení hran. Ohodnocení vrcholů budeme uvažovat fyzikálně jako *potenciály*.[6] Ohodnocení hran odpovídá *toku*, který teče po směru hran. Proto jsme hrany zorientovali a toku proti směru hrany odpovídá záporné ohodnocení.

Zkusme tyto transformace popsat přesněji. Obrázek 5.18 ukazuje příklady. Lineární zobrazení $I_G : \mathbb{R}^m \to \mathbb{R}^n$ transformuje ohodnocení hran $\boldsymbol{y}$ v ohodnocení vrcholů $\boldsymbol{x}$. Označme E_i^+ množinu hran směřujících do i-tého vrcholu a E_i^- množinu hran směřujících z i-tého vrcholu. Potom pro $\boldsymbol{x} = I_G\boldsymbol{y}$ platí:

$$x_i = \sum_{E_i^+} y_j - \sum_{E_i^-} y_k.$$

Výsledné ohodnocení vrcholu x_i odpovídá *přebytku* v i-tém vrcholu, tedy rozdílu toho, co do vrcholu přiteče a co odteče. Studium toků hraje klíčovou roli například v elektromagnetismu a je důležité v

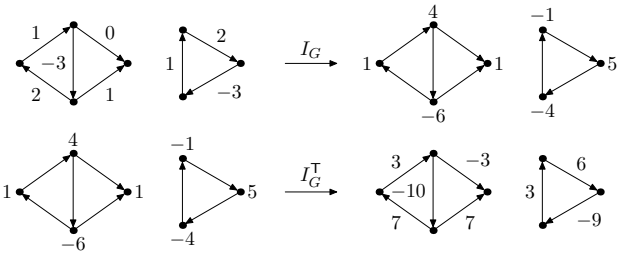
[6] Co jsou to potenciály? Ty mohou třeba odpovídat výškám, ve kterých jsou jednotlivé vrcholy umístěny. Představme si, že se nějaký objekt přesouvá po grafu. Pokud se přesune do výše položeného vrcholů, musí vykonat práci úměrnou rozdílu výšek. Tato práce se uloží ve formě (gravitačního) potenciálu, a při opětovném sestupu do níže položeného vrcholu je část potenciálu spotřebována. Samozřejmě potenciál může souviset i například s abstraktnějším elektomagnetismem, kde funguje podobně.

aplikacích, například při síťové komunikaci nebo optimalizaci dopravních spojení. Matice I_G a nástroje lineární algebry umožňují získat vzhled do problematiky toků.

Duální zobrazení $I_G^\mathsf{T} : \mathbb{R}^n \to \mathbb{R}^m$ naopak transformuje ohodnocení vrcholů $\boldsymbol{x}$ v ohodnocení hran $\boldsymbol{y}$. Nechť $\boldsymbol{y} = I_G^\mathsf{T}\boldsymbol{x}$. Potom

$$y_k = x_j - x_i, \quad \text{kde } k\text{-tá hrana vede z } i \text{ do } j.$$

Tedy duální zobrazení přiřazuje hranám rozdíly potenciálů.



Obrázek 5.18: Lineární zobrazení I_G transformuje tok $\boldsymbol{y}$ v potenciály $\boldsymbol{x}$. Duální lineární zobrazení I_G^T transformuje potenciály $\boldsymbol{x}$ v rozdíly potenciálů $\boldsymbol{y}$.

Jádro duálního zobrazení. Je snadné určit první fundamentální podprostor $\mathrm{Ker}(I_G^\mathsf{T})$. Každý vektor $\boldsymbol{x}$ z jádra splňuje $I_G^\mathsf{T}\boldsymbol{x} = \boldsymbol{0}$, tedy rozdíl potenciálů na libovolné hraně je nulový. Proto všechny vrcholy v celé komponentě musí stejný potenciál. Dimenze $\mathrm{Ker}(I_G^\mathsf{T})$ je rovna c, kde c je počet komponent grafu. Přirozená bázi podprostoru tvoří c vektorů, kde každý má hodnotu jedna v rámci jedné komponenty a nulovou hodnotu pro ostatní komponenty. Tento podprostor nazveme *podprostor komponent*. Poznamenejme, že $\mathrm{Ker}(L_G)$ je stejný jako podprostor komponent; v důkazu tvrzení 4.14 jsme ukázali, že $\mathrm{Ker}(AA^\mathsf{T}) = \mathrm{Ker}(A^\mathsf{T})$ pro libovolnou matici A.

Protože jsme určili dimenzi jednoho fundamentálního prostoru, umíme snadno dopočítat dimenze ostatních fundamentálních podprostorů. Matice I_G má hodnost $\mathrm{rank}(I_G) = n - c$. Proto je

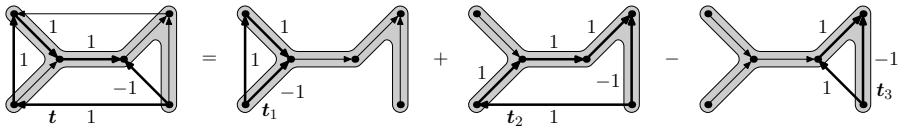
$$\dim \mathcal{R}(I_G) = \dim \mathrm{Im}(I_G) = n - c \qquad \text{a} \qquad \dim \mathrm{Ker}(I_G) = m - n + c.$$

Zbývá určit, jak kombinatoricky vypadají tři zbývající fundamentální podprostory a vysvětlit zvláštní hodnoty dimenzí.

Jádro. Zkusme pochopit kombinatorickou strukturu $\mathrm{Ker}(I_G)$. Tok $\boldsymbol{y}$ leží v $\mathrm{Ker}(I_G)$, pokud platí $I_G\boldsymbol{y} = \boldsymbol{0}$, tedy tok vytváří ve všech vrcholech nulové přebytky. Tato vlastnost se ve fyzice nazývá *Kirchhofův zákon*. Ten říká, že se v elektrické síti nevytváří přebytky proudu v žádném uzlu. Podprostoru $\mathrm{Ker}(I_G)$ se typicky říká *podprostor cyklů*, neboť je generovaný všemi cykly grafu. Pokud zvolíme libovolný cyklus a pošleme po něm v nějakém směru tok nějaké velikosti, nezmění se přebytky. Přesněji řečeno přebytky se nezmění, pokud přičteme následující tok $\boldsymbol{t}$. Ten má na hranách mimo cyklus nulovou hodnotu, na hranách po směru cyklu hodnotu 1 a na hranách proti směru cyklu zápornou hodnotu toku -1. Příklady takových toků $\boldsymbol{t}$ jsou na obrázku 5.19 spolu s tučně vyznačenými cykly.

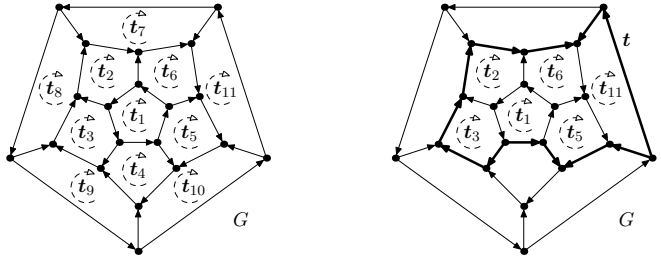
Proč má však jádro $\mathrm{Ker}(I_G)$ dimenzi právě $m - n + c$? Pro toto zvláštní číslo získáme vysvětlení, když sestrojíme nějakou přirozenou bázi. Již víme, že podprostor je generovaný cykly, kterých je však v typickém grafu hrozně moc. Ukážeme, že stačí zvolit malou část těchto cyklů. Uvažme libovolnou *kostru*, což je maximální acyklický podgraf. Kostra souvislého grafu obsahuje $n - 1$ hran, v případě c komponent obsahuje $n - c$ hran. Počet hran kostry je totožný s dimenzí $\mathcal{R}(A)$ a $\mathrm{Im}(A)$, a v případě $\mathrm{Ker}(I_G)$ je dimenze $m - n + c$ počet zbývajících hran mimo kostru. To pochopitelně není náhoda.

Každá hrana e mimo kostru definuje jeden specifický cyklus C_e, který se nazývá *fundamentální cyklus*. Tento cyklus je jediný cyklus, který vznikne přidáním hrany e do kostry, tedy všechny zbývající hrany C_e patří do kostry. Tomuto cyklu odpovídá tok $\boldsymbol{t}_e$ definovaný jako výše. Libovolný cyklus C s libovolným tokem $\boldsymbol{t}$ lze vyjádřit jako lineární kombinaci toků $\boldsymbol{t}_e$ přes fundamentální cykly C_e pro všechny hrany e, které patří do C a nejsou z kostry. Detaily si může čtenář rozmyslet ve cvičení 5.9.



Obrázek 5.19: Nalevo je tučně vyznačený cyklus spolu s šedě vyznačenou kostrou. Tok $\boldsymbol{t}$ má po směru cyklu hodnotu 1 a proti směru cyklu -1. Protože tento cyklus obsahuje tři hrany mimo kostru, lze vyjádřit $\boldsymbol{t}$ jako lineární kombinaci tří toků $\boldsymbol{t}_1$, $\boldsymbol{t}_2$ a $\boldsymbol{t}_3$ pro fundamentální cyklů nakreslené vpravo.

Eulerova věta. Udělejme malou odbočku a ukažme velice zajímavou aplikaci získaných výsledků. Naši znalost dimenze $\mathrm{Ker}(I_G)$ využijeme k důkazu Eulerovy věty. To je slavný vztah mezi počtem vrcholů, hran a stěn v rovinných grafech. Graf se nazývá *rovinný*, pokud ho lze nakreslit do roviny bez křížení hran, na obrázku 5.20 je příklad rovinného grafu. V každém rovinném nakreslení rozdělují hrany grafu rovinu na několik oblastí, které se nazývají *stěny*. Nakreslení obsahuje právě jednu neomezenou stěnu, která se nazývá *vnější*, a ostatní stěny se nazývají *vnitřní*. Označme počet stěn rovinného grafu pomocí f; pro libovolné nakreslení je stejný. Mimochodem rovinné grafy úzce souvisí s mnohostěny, ostatně proto mají vrcholy, hrany a stěny svoje geometrická jména.



Obrázek 5.20: Nalevo je příklad rovinného grafu, což je slavný dvanáctistěn, jedno z pěti Platonských těles. Pro tento graf je $n = 20$, $m = 30$ a $f = 12$. Jedenáct cyklů odpovídajících vnitřním stěnám, které tvoří bázi $\mathrm{Ker}(I_G)$, je vyznačeno. Napravo je zvýrazněn cyklus s odpovídajícím tokem $\boldsymbol{t}$, který lze vyjádřit jako $\boldsymbol{t}_1 + \boldsymbol{t}_2 + \boldsymbol{t}_3 + \boldsymbol{t}_6 + \boldsymbol{t}_5 + \boldsymbol{t}_{11}$.

Zaměřme se pro jednoduchost na souvislé grafy, i když obecné znění věty je velmi podobné. (Čtenář si může rozmyslet obecné znění a jeho důkaz.) Eulerova věta říká:

Věta 5.7 (Euler, 1750)**.** *Pro libovolný souvislý rovinný graf platí*

$$m - n + 1 = f - 1.$$

Důkaz. Existuje celá řada důkazů. Například standardní důkaz je indukcí podle počtu hran. Pro strom zjevně rovnost platí. Přidáním jedné hrany se jedna stěna rozdělí na dvě, což zvětší se počet stěn o jedna. Obě strany rovnosti se tedy zvětší o jedničku a rovnost zůstane zachována.

Ukážeme si alternativní důkaz pomocí lineární algebry. Levá strana rovnosti by měla být čtenáři povědomá, je to přesně dimenze $\mathrm{Ker}(I_G)$ pro souvislý graf G. Jak už často v matematice bývá, podobné věci nejsou náhody. Stačí ukázat, že $\mathrm{Ker}(I_G)$ má jinou bázi velikosti $f-1$, a rovnost vyplyne ze Steinitzovy věty 4.6 o výměně.

Uvažme libovolné rovinné nakreslení grafu. Pro každou vnitřní stěnu (kterých je $f-1$) zvolíme tok $\boldsymbol{t}_i$ ve směru hodinových ručiček podél hran této stěny. Tyto toky jsou vyznačeny na obrázku 5.20 vlevo. Tvrdíme, že toky $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_{f-1}$ tvoří bázi prostoru cyklů $\mathrm{Ker}(I_G)$.

Množina $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_{f-1}$ generuje všechny cykly v grafu z následujícího důvodu. Uvažme libovolný cyklus s příslušným tokem $\boldsymbol{t}$, řekneme velikosti jedna po směru hodinových ručiček v rovinném nakreslení. Potom $\boldsymbol{t}$ lze vyjádřit jako součet toků $\boldsymbol{t}_i$ pro všechny stěny umístěné uvnitř cyklu. Hrany uvnitř cyklu se navzájem odečtou. Zůstanou pouze hrany na hranici se správným znaménkem. Toto vyjádření je naznačeno na obrázku 5.20 vpravo.

Podobně lze dokázat, že množina $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_{f-1}$ je lineárně nezávislá. Argument je, že v lineární kombinaci libovolné množiny stěn určitě zůstane nenulová hranice těchto stěn. Detaily si může čtenář rozmyslet. Dokázali jsme, že $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_{f-1}$ tvoří bázi, z čehož vyplývá rovnost. □

Tato věta má celou řadu zajímavých důsledků a zobecnění. Například z ní vyplývá, že v libovolném rovinném grafu platí $m \leq 3n-6$. Tedy, na rozdíl od obecných grafů, rovinné grafy obsahují pouze lineárně mnoho hran vzhledem k počtu vrcholů. Poznamenejme, že typičtější zápis rovnosti je $n+f=m+2$. Z našeho zápisu je však jasnější důkaz pomocí lineární algebry.

Obraz. Jeden ze dvou zbývajících fundamentálních podprostorů, obraz $\mathrm{Im}(I_G)$, se někdy nazývá *podprostor potenciálů*. Z definice je generován jednotlivými sloupci matice I_G, kterým odpovídají jednotlivé hrany. Pokud po hraně z u do v pošleme nějaký tok velikosti t, zvýšíme přebytek ve vrcholu v o hodnotu t a snížíme přebytek v u o hodnotu t. Na přebytek vrcholu můžeme nahlížet jako *potenciál*, který může poslat svým sousedům. Potenciály některých vrcholů mohou být záporné, což matematicky nevytváří jakýkoliv problém. Z pohledu reálného světa si můžeme představit, že všechny vrcholy začínají s nějakým výchozím množstvím, a potenciál popisuje rozdíl vůči počátečnímu stavu. Jediná podmínka pro potenciály je, že jejich součet je v libovolné komponentě nulový. To například odpovídá fyzikálním principům zachování hmoty nebo energie. (Podle toho, co tok vyjadřuje.)

Víme, že dimenze $\mathrm{Im}(I_G)$ je $n-c$, což je počet hran v kostře. Pokud chceme poslat tok z u do v, musejí se nacházet ve stejné komponentě. Tento tok můžeme poslat podél kostry, pouze s využitím hran kostry. Tedy hrany kostry tvoří bázi $\mathrm{Im}(I_G)$. Čtenář si může rozmyslet, že sloupce odpovídající podmnožině hran jsou nezávislá, právě když hrany neobsahují cyklus.

Řádkový podprostor. Podstatně zajímavější je řádkový podprostor, který se nazývá *podprostor řezů*. Ten je generován řádky matice. Uvědomme si, že i-tý řádek odpovídá množině hran, které vedou z a do i-tého vrcholu. Nechť $\boldsymbol{x}$ je libovolný vektor potenciálů vrcholů. Potom $\boldsymbol{y} = I_G^{\mathsf{T}}\boldsymbol{x}$ je vektor, který popisuje pro jednotlivé hrany rozdíly potenciálů koncových vrcholů. Řádkový prostor je tedy generovaný změnou potenciálů jednotlivých vrcholů. Přičtení násobku i-tého řádku odpovídá změně potenciálu i-tého vrcholu o tento násobek, což způsobí stejnou změnu rozdílu potenciálu pro všechny hrany vedoucí z a do vrcholu i.

Víme, že dimenze řádkového podprostoru je $n-c$. Rozdíly potenciálů hrají roli pouze v rámci komponent, tedy jednotlivé komponenty lze uvažovat samostatně. Klíčové je, že pouze rozdíly potenciálů ovlivňují hodnotu $\boldsymbol{y}$, nikoliv jejich absolutní hodnoty. Pokud změníme potenciály v celé komponentě stejně, rozdíly se nezmění. Proto můžeme v každé komponentě zafixovat potenciál jednoho vrcholu a stačí měnit zbývající potenciály k vygenerování celého řádkového podprostoru. Čtenář může zkusit odvodit jinou bázi $\mathcal{R}(I_G)$ ve cvičení 5.10.

Proč se tomuto podprostoru říká podprostor řezů? Protože popisuje *řezy* v grafu. Rozdělme vrcholy grafu na dvě části, které označme A a B. Řez mezi A a B je množina všech hran, které vedou mezi

vrcholem z A a vrcholem z B. Rozdělení vrcholů mezi A a B lze popsat následujícím vektorem $\boldsymbol{x}$:

$$x_i = \begin{cases} 0, & \text{pokud } i \in A, \text{ a} \\ 1, & \text{pokud } i \in B. \end{cases}$$

Potom $\boldsymbol{y} = I_G^{\mathsf{T}}\boldsymbol{x}$ je vektor, který přiřazuje hranám z A do B hodnotu jedna, hranám z B do A hodnotu mínus jedna a ostatním hranám hodnotu nula. Tedy $\boldsymbol{y}$ popisuje řez mezi A a B. Obecně vektory řádkové podprostoru popisují zobecnění řezů. Mimochodem z rovnosti $L_G = I_G I_G^{\mathsf{T}}$ lze vyvodit alternativní důkaz cvičení 5.8.

Shrnutí

Cvičení

⇒ 5.1 Určete, jak vypadá matice transformace zrcadlení v $\mathbb{R}^2$ podél přímky procházející počátkem pod úhlem φ vůči kanonické bázi.

⇒ 5.2 Určete pro následující matice uvažované vůči kanonické bázi v $\mathbb{R}^2$, jakým geometrickým transformacím odpovídají. Vynechané koeficienty jsou nulové.

$$A = \begin{pmatrix} & -1 \\ 1 & \end{pmatrix}, \qquad B = \begin{pmatrix} & 1 \\ 1 & \end{pmatrix}, \qquad C = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \qquad D = \begin{pmatrix} 1 & \\ & \end{pmatrix}, \qquad E = \begin{pmatrix} 1 & \\ & -1 \end{pmatrix}.$$

5.3 Dokažte, že inverze lineárního zobrazení je lineární. Budete muset zvlášť dokázat pro levou inverzi $f^{-1} \circ f = \mathrm{id}$ a pro pravou inverzi $f \circ f^{-1} = \mathrm{id}$.

⋆ 5.4 Nechť W je podprostor $\mathbb{U}$. Definujme *faktorprostor* $\mathbb{U}/W$ jako množinu všech afinních podprostorů $W + \boldsymbol{p}$ spolu s operacemi sčítání a násobení skalárem:

$$(W+\boldsymbol{p}) + (W+\boldsymbol{q}) = W + (\boldsymbol{p}+\boldsymbol{q}) \qquad \text{a} \qquad \alpha(W+\boldsymbol{p}) = W + (\alpha\boldsymbol{p}).$$

Dokažte, že vzniklá struktura je vektorový prostor. Určete, jak vypadá nulový vektor. Dokažte, že $\dim \mathbb{U}/W = \dim \mathbb{U} - \dim W$.

⋆ 5.5 Dokažte, že neexistuje izomorfismus mezi $\mathbb{R}$ a $\mathbb{R}^2$. Na druhou stranu sestojte bijektivní zobrazení $f : \mathbb{R} \to \mathbb{R}^2$. K tomu využijte desetinný rozvoj reálného čísla x a rozdělte ho vhodným způsobem mezi dvě reálná čísla (y, z).

5.6 Dokažte, že duální zobrazení f^* je určené jednoznačně. Tedy ukažte, že pro libovolnou volbu bází X a Y určuje matice A^{T} stejné lineární zobrazení vůči bázím Y a X.

⋆ 5.7 Uvažme matici sousednosti A_G grafu G. Objevte, co popisují koeficienty A_G^2 a obecně A_G^k pro libovolné k. Výsledek pochopitelně dokažte.

⋆ 5.8 Nechť L_G je Laplaceova matice grafu G a nechť $\boldsymbol{x}$ je libovolný vektor, jehož koeficienty jsou pouze nuly a jedničky. Ukažte, že výraz $\boldsymbol{x}^{\mathsf{T}} L_G \boldsymbol{x}$ počítá velikost určitého řezu v grafu. Řez je množina hran, které vedou mezi dvěma skupinami vrcholů.

⋆ 5.9 Dokažte, že jádro $\mathrm{Ker}(I_G)$ je generované všemi cykly grafu a že fundamentální cykly C_e generují všechny cykly grafu.

⋆ **5.10** Pro jednoduchost uvažme souvislý graf a zvolme v něm kostru. Prostor řezů $\mathcal{R}(I_G)$ má následující bázi tvořenou *fundamentálními řezy*, každý odpovídající jedné hraně kostry. Každá hrana rozděluje vrcholy kostry na dvě části. Vektor zvětšuje potenciál všech vrcholů na jedné straně o jedna, a na druhé straně potenciály nemění. Dokažte, že zvolené vektory tvoří bázi $\mathcal{R}(I_G)$ a určete, které hrany tyto fundamentální řezy obsahují.

Kapitola 6

Grupy a tělesa

V této kapitole uděláme malou odbočku ze světa lineární algebry a popíšeme si dvě základní struktury abstraktní algebry. První strukturou jsou *grupy*, které slouží k popisu symetrií objektů. Také úzce souvisí se strukturou transformací aplikovaných na objekty. Druhou strukturou jsou *(algebraická) tělesa*. Název je maličko matoucí, neboť naznačuje souvislost s geometrickými tělesy (například krychle). Algebraická tělesa jsou zobecnění čísel, například čísel reálných nebo racionálních. Myšlenka je, že řada úkonů, které můžeme provádět s reálnými čísly, lze stejně dobře provádět v tomto zobecnění. Tato zobecnění lze aplikovat na řadu situací, které není možné popsat reálnými čísly.

Tyto struktury jsou základními matematickými objekty, se kterými se čtenář setká na každém rohu. Přirozená otázka je však následující. Jak tyto dvě struktury souvisí s lineární algebrou? Proč jim věnujeme celou kapitolu? Grupy úzce souvisí se strukturou transformací, což jsou například lineární zobrazení popsaná v kapitole 5. Například množina všech automorfismů (bijektivních lineárních zobrazení $f : \mathbb{U} \to \mathbb{U}$) tvoří grupu. V řeči matic je to grupa všech regulárních matic.

Motivace pro zkoumání těles je ještě přímočařejší. V celém textu jsme uvažovali lineární algebru pouze nad reálnými čísly: vektory byly n-tice reálných čísel, matice byly tabulky reálných čísel, typické vektorové prostory byly $\mathbb{R}^n$. Čtenář si však může povšimnout, že reálná čísla mají celou řadu vlastností, které jsme nikdy nepoužili. Například pro libovolné nezáporné reálné číslo x existuje jeho odmocnina $\sqrt{x}$. V lineární algebře však uvažujeme lineární pojmy, a proto jsme existenci odmocniny (alespoň zatím) nikde nepotřebovali.

Myšlenka je, že můžeme lineární algebru provádět nad zobecněnými čísly. Potřebujeme, že tato čísla splňují vlastnosti jako komutativita, asociativita a distributivita. Prozkoumáním všech vlastností z tohoto textu dostaneme definici algebraického tělesa. Algebraická tělesa jsou tedy přesně ta čísla, nad kterými funguje lineární algebra. Vysvětlíme si, které vlastnosti se v tomto zobecnění maličko liší a které jsou naopak zcela stejné. A ukážeme si aplikace těles.

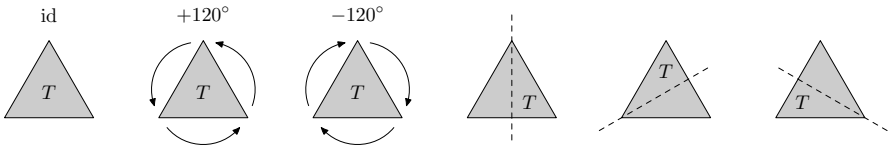
6.1 Grupy

V této sekci zavedeme matematickou strukturu zvanou grupa, ukážeme její motivaci a základní vlastnosti. Také popíšeme katalog grup, které se často vyskytují v matematice.

Struktura symetrií. Uvažme například rovnostranný trojúhelník T v rovině. Budeme uvažovat všechny jeho *symetrické transformace* (neboli *symetrie*). To jsou geometrické transformace roviny, které zobrazují T přesně na T. Budeme uvažovat shodné transformace roviny, což jsou všechny rotace, posunutí a zrcadlení.

Identita je vždy symetrická transformace, neboť rozhodně zachovává T (a libovolný jiný útvar). Dále můžeme uvažovat rotace roviny ve středu trojúhelníka. Pokud ji otočíme o 120° nebo o 240°, bude T

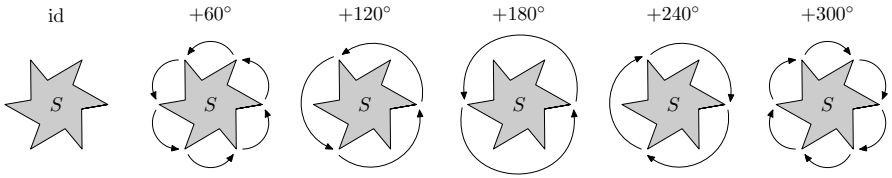
zachován. Nakonec jsou symetrické transformace tři zrcadlení podle přímek, které prochází jednotlivými vrcholy trojúhelníka a středy protějších stran. Tyto symetrické transformace jsou na obrázku 6.1.



Obrázek 6.1: Šest symetrických transformací rovnostranného trojúhelníka T.

Celkem jsme získali šest symetrických transformací a tvrdíme, že žádné další nejsou. Jak toto můžeme ukázat? Klíčové pozorování je, že symetrická transformace zobrazuje vrcholy T na vrcholy T. Budeme předpokládat následující geometrický fakt, který nemůžeme dokázat, neboť nemáme shodné transformace roviny formálně definovány. Platí, že pokud dvě shodné transformace roviny zobrazují vrcholy T stejně, musí být totožné. Proto existuje pro T nejvýše šest symetrických transformací, které jsme již objevili.

Počet symetrií T však není všechno. Přirozená otázka je, jakou mají strukturu. Co tím myslíme? Například ozubené kolo S na obrázku 6.2 má také šest symetrických transformací, které však vypadají velice odlišně než symetrie T. Strukturou zde myslíme to, jak se symetrické transformace skládají. Pokud máme totiž dvě symetrické transformace a složíme je, dostaneme opět symetrickou transformaci; první i druhá transformace zachovávají objekt, proto i jejich složení zachovává objekt a je symetrickou transformací. Například složením dvou rotací T o 120° je rotace o 240° neboli −120°.



Obrázek 6.2: Šest symetrií ozubeného kola S, všechny jsou rotace roviny o násobek šedesáti stupňů.

Další strukturální vlastnost symetrií je existence *inverzních prvků*. Libovolná symetrická transformace f má inverzní symetrickou transformací f^{-1}, která splňuje

$$f \circ f^{-1} = f^{-1} \circ f = \mathrm{id}.$$

Pro libovolné f je inverze f^{-1} jednoznačně určená. Například inverze rotace T o 120° je obrácená rotace o −120°. Poznamenejme, že může platit $f^{-1} = f$, tedy symetrická transformace může být sama k sobě inverzní; v takovém případě se f nazývá *involuce*. Identita je vždy involuce a další příklady involucí jsou zrcadlení T a rotace S o 180°. Struktury symetrií S a T jsou odlišné už jen proto, že S obsahuje přesně dvě involuce a T obsahuje přesně čtyři involuce.

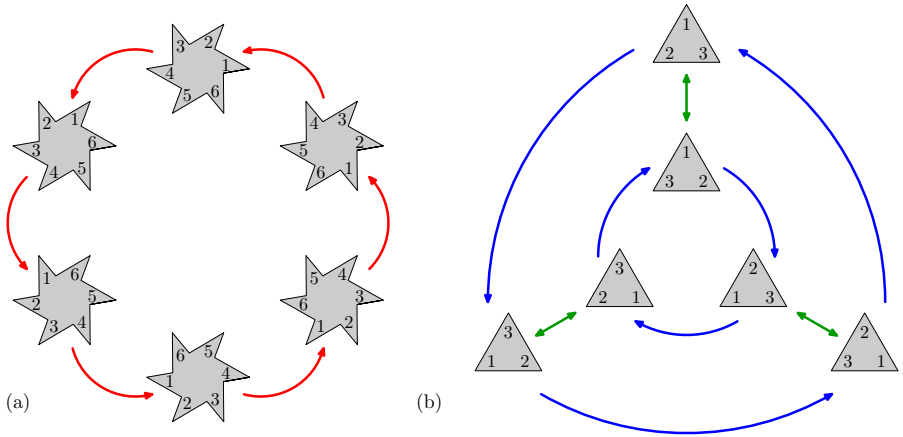
Cayleyho grafy. Abychom mohli ukázat, jak moc je odlišná struktura symetrií S a T, můžeme použít *Cayleyho grafy*. Cayleyho graf je diagram, který umožňuje vizualizovat strukturu symetrií. Právě za tímto účelem vymyslel v roce 1878 Cayley tyto diagramy.

Uvažme nějaký objekt O a jeho symetrické transformace $\mathrm{id} = f_1, \ldots, f_n$. Vrcholy Cayleyho grafu pro objekt O jsou jednotlivé obrazy $O = O_1, \ldots, O_n$ objektu po aplikování symetrických transformací,

kde $O_i = f_i(O)$. Tedy obrazy lze identifikovat se symetrickými transformacemi objektu. Tyto obrazy jsou totožné objekty, které se liší pouze přejmenováním bodů. Například v případě rovnostranného trojúhelníku T můžeme jeho vrcholy očíslovat 1, 2 a 3 a získáváme obrazy $T_1, \ldots, T_6$ s různě očíslovanými vrcholy.

Hrany popisují vztahy jednotlivých symetrií. Pro každou symetrickou transformaci f_i zvolíme nějakou barvu. Z každého obrazu O_j nakreslíme šipku této barvy směřující do obrazu $f_i(O_j)$. Protože tímto způsobem by Cayleyho graf obsahoval příliš mnoho hran[1] a nebyl by přehledný, zobrazuje se pouze část symetrických transformací.

Zvolíme malou množinu symetrických transformací zvanou *generující množina*. Důvodem pro tento název je, že *generuje* všechny symetrické transformace. Prvky generující množiny se nazývají *generátory*. Každá symetrická transformace se dá vyjádřit jako složení několika generátorů. Obrázek 6.3 ukazuje Cayleyho grafy pro symetrie objektů S a T. Pokud by zvolené symetrie negenerovaly všechny symetrie, Cayleyho graf by nevyjadřoval celou strukturu symetrií.



Obrázek 6.3: (a) Struktura symetrií ozubeného kola S. Stačí jednoprvkový generátor tvořený rotací o 60°. Jednotlivé symetrie jsou uspořádané do cyklu. (b) Struktura symetrií rovnostranného trojúhelníku T. Žádný prvek sám negeneruje všechny symetrie, a proto zvolíme dvouprvkový generátor tvořený rotací o 120° (vyobrazena modře) a zrcadlením podle vertikální přímky (vyobrazeno zeleně). Povšimněte si, že modrá rotace posouvá symetrie „uvnitř“ opačným směrem než „venku“. Jako cvičení zkuste zkonstruovat jiný Cayleyho graf pro T, kde zvolený generátor je tvořený dvěma zrcadleními.

Typicky se volí do inkluze minimální generující množinu, protože se chceme vyhnout zobrazování nadbytečné informace, generátory již jednoznačně určují strukturu symetrií. Do inkluze minimální generující množiny neobsahují žádné nadbytečné prvky, tedy zde je přirozená paralela s nezávislými množinami vektorů z kapitole 4. Pro symetrie však neplatí obdoba Steinitzovy věty 4.6, že by každý do inkluze minimální generující množina obsahovala stejně prvků. Například pro množinu symetrií objektu S tvoří rotace o 120° a 180° do inkluze minimální generující množinu.

Můžeme si všimnout, že struktura těchto Cayleyho grafů je velice pravidelná. To není náhoda, podobně pravidelně vypadá každý Cayleyho graf. Této vlastnosti se v teorii grup říká *regularita*. To je

[1] Přesněji z libovolného O_i by do libovolného O_j vedla přesně jedna hrana nějaké barvy. Čtenář si může rozmyslet, proč tomu tak je.

mnohem silnější než regularita v teorii grafů, která říká, že každý vrchol má stejně sousedů. Grupařská regularita Cayleyho grafu přibližně znamená, že z pohledu každého vrcholu vypadá zcela totožně. Například pro libovolný vrchol Cayleyho grafu pro T z obrázku 6.3b platí, že aplikování modré, zelené a modré šipky z libovolného vrcholu vede do stejného vrcholu jako samotná zelená šipka. Podrobněji vysvětlíme na konci kapitoly 6.1

Definice grupy. Výše popsané struktury symetrií S a T jsou konkrétní příklady matematických struktur zvaných *grupy*, a Cayleyho grafy jsou nástroje určené pro vizualizaci grup. Z popsaných vlastností symetrií vyvodíme abstraktní definici grupy jako matematické struktury. Připomeňte si jejich definici popsanou v kapitole 2.5.

Grupa $\mathbb{G}$ je matematická struktura tvořena množinou prvků G a operací $\circ : G \to G$. V našich konkrétních příkladech grup byla G množina symetrií a $\circ$ operace skládání, která musí splňovat následující tři vlastnosti:

- *Asociativita:* Pro libovolné tři prvky $f, g, h \in G$ platí
$$(f \circ g) \circ h = f \circ (g \circ h).$$
Tato vlastnost dává pro symetrie smysl, protože skládání libovolných zobrazení je asociativní. Neasociativní matematické struktury jsou velice divné, tedy typicky je to rozumná vlastnost.
- *Existence neutrálního prvku:* Existuje *neutrální prvek* $e \in G$, který pro libovolný prvek $f \in G$ splňuje
$$f \circ e = e \circ f = f.$$
V případě symetrii tuto roli plní identické zobrazení id. Čtenář si může rozmyslet, že neutrální prvek je jednoznačně určený.
- *Existence inverzních prvků:* Pro libovolný prvek $f \in G$ existuje *oboustranná inverze* f^{-1}, která splňuje
$$f \circ f^{-1} = f^{-1} \circ f = e.$$
Protože symetrické transformace zachovávají objekt, dají se vždy invertovat. Poznamenejme, že z asociativity a existence levé a pravé inverze vyplývá, že tyto inverze jsou jednoznačně určené a totožné; tohle jsme dokázali konkrétně pro čtvercové matice v lemmatu 3.8.

Operaci $\circ$ budeme v textu také podle situace značit jako $\cdot$, $+$ nebo přímo symbol vynechávat; například asociativita říká, že $f(gh) = (fg)h$. Počet prvků grupy se nazývá *řád* a značí se $|\mathbb{G}|$. V textu se budeme převážně zabývat konečnými grupami (s konečným řádem), avšak ukážeme také důležité příklady nekonečných grup.

Na rozdíl od vektorových prostorů nemusí být operace $\circ$ komutativní. Obvykle ani nebude, neboť skládání zobrazení typicky komutativní není. Například v diagramu na obrázku 6.3b dostáváme jiný obraz aplikováním nejprve zelené a poté modré symetrie, než aplikováním nejprve modré a poté zelené. Pokud je $\circ$ komutativní, nazývá se grupa *komutativní* nebo *Abelovská* (podle jednoho z velkých matematiků spojených se vznikem teorie grup). Později ukážeme, že Abelovské grupy mají velice omezenou strukturu, a tedy typické grupy nejsou komutativní.

Povšimněme si, že každý vektorový prostor definuje Abelovskou grupu vůči operaci sčítání. Struktura vektorových prostorů je velice limitovaná operací násobení skalárem, strukturou reálných čísel a dalších vlastností. Na druhou stranu obecně o grupách vlastnosti poví velice málo, a proto je jejich struktura mnohem složitější. Při práci s grupami totiž máme mnohem méně vodítek, kterých se můžeme chopit. Nástroje teorie grup vznikaly několik stovek let za účelem lépe pochopit strukturu složitých grup. V tomto textu si ukážeme pouze střípky z této rozsáhlé teorie.

Historická motivace. Různé výsledky z teorie grup se objevovaly v matematice už od pradávna, v souvislosti se studiem symetrií geometrických objektů. Například Platonská tělesa a jejich symetrie fascinovaly již antické matematiky, kteří jim přisuzovali hluboký filozofický význam. Techniky využívání symetrií se také budovaly v teorii čísel při řešení rovnic, kdy se uvažovaly různé transformace množiny řešení. Samotný pojem grupy zavedl Galois ve své práci, když se snažil pochopit neřešitelnost rovnic s polynomem pátého a vyššího stupně.

O co se přesně jedná? Již několik tisíc let je známý vzorec pro řešení kvadratické rovnice $ax^2 + bx + c = 0$. Ten říká, že řešením je množina

$$\left\{ \frac{-b + \sqrt{b^2 - 4ac}}{2a}, \frac{-b - \sqrt{b^2 - 4ac}}{2a} \right\}.$$

V průběhu šestnáctého století byly objeveny podobné, ale mnohem komplikovanější vzorce pro řešení kubických rovnic $ax^3 + bx^2 + cx + d = 0$ a quartických rovnic $ax^4 + bx^3 + cx^2 + dx + e = 0$, obsahující třetí a čtvrté odmocniny. Žádné další vzorce pro polynomy vyššího stupně však nebyly nalezeny a jejich existence byla záhadou následujících 200 let. V roce 1823 dokázal matematik Abel, že žádný podobný vzorec pro řešení rovnic s polynomy pátého a vyššího stupně neexistuje; konkrétně neexistuje například pro rovnici $x^5 - x + 1 = 0$.

Co však myslíme podobným vzorcem? Základní věta algebry říká, že každý polynom s komplexními koeficienty má komplexní kořen. Kořenem můžeme polynom vydělit a pokračovat dál. Tedy pro každý polynom existuje rozklad:

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 = a_n (x - x_1)(x - x_2) \cdots (x - x_n),$$

kde $x_1, \ldots, x_n$ jsou nějaká komplexní čísla. Abelova věta však říká, že obecně tato komplexní čísla nemůžeme popsat vzorečky, které obsahují pouze koeficienty $a_0, \ldots, a_n$, další racionální konstanty a jejich součty, rozdíly, součiny, podíly a n-té odmocniny (které se nazývají *radikály*), a to ani, když je povolíme libovolně poskládat a zanořit do sebe. Proto se tomuto výsledků říká "neřešitelnost polynomiálních rovnic stupně pět v radikálech".

Jak toto souvisí s grupami? Povšimněme si, že Abelova věta neříká, že pro každou polynomiální rovnici stupně alespoň pět nelze množinu řešení popsat vzorečkem. Například pro rovnici $x^n = 1$ umíme množinu řešení popsat v radikálech jako $\sqrt[n]{1}$ (což je množina n komplexních čísel). Pouze říká, že existují špatné rovnice, například $x^5 - x + 1 = 0$, pro které takový popis není možný. Přirozené otázky jsou:

- Pro které polynomiální rovnice popis vzorečkem existuje?
- Co je tolik speciální na čísle pět, tedy proč podobné vzorce existují pro všechny polynomiální rovnice do čtvrtého stupně?

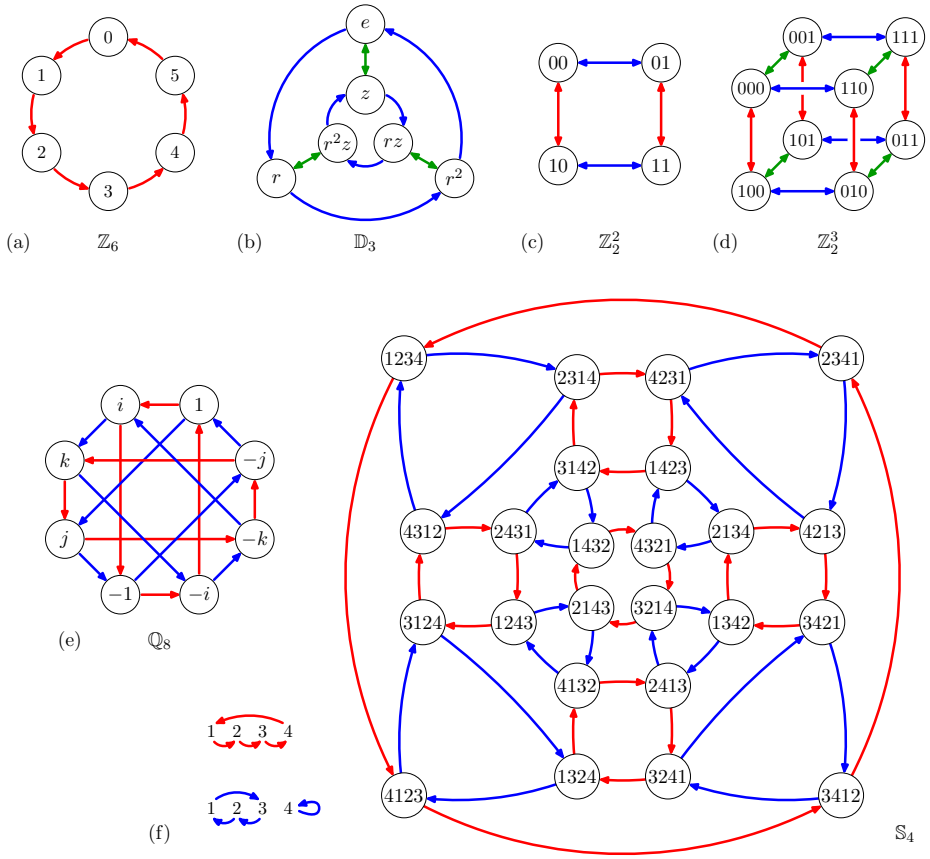
Na tyto otázky odpověděl Galois a využil k tomu strukturu symetrií množiny řešení dané polynomiální rovnice. V jeho článku se poprvé objevil pojem grupa a ukázal, že existence vzorce závisí na struktuře této grupy. Grupy, pro které lze popsat řešení polynomiální rovnice vzorcem, nazval *řešitelné*. Galoisova teorie je propojení mezi grupami a tělesy, a podobné úvahy lze použít i v důkazech neřešitelnosti řady dalších matematických problémů.

Grupy hrají klíčovou roli v řadě oblastí matematiky. Obecná myšlenka je, že spousta matematických objektů má velice bohatou pravidelnou strukturu. Teorie grup dává silné obecné nástroje pro zkoumání této struktury a umožňuje se tak vyznat ve složitých objektech a zjednodušit je. Díky tomu je možné tyto objekty efektivně konstruovat, popisovat a analyzovat. Grupy mají také celou řadu aplikací mimo matematiku a mají přesah například do fyziky a chemie. V krystalografii se zkoumají krystaly, což jsou pravidelně uspořádané struktury atomů. Ze znalosti symetrií krystalů se můžeme dozvědět řadu jejich vlastností.

Revize Cayleyho grafů. Nejprve definujme formálně generující množinu. Podmnožina prvků $H \subseteq G$ se nazývá *generující množina*, pokud *generuje* G. Te znamená, že libovolný prvek $g \in G$ lze získat aplikováním konečně mnoha operací na prvky z H a jejich inverze $H^{-1} = \{h^{-1} : h \in H\}$, formálně:

$$\forall g \in G\ \exists h_1, \ldots, h_k \in H \cup H^{-1} : \qquad g = h_k \circ h_{k-1} \circ \cdots \circ h_2 \circ h_1.$$

Prvky generující množiny se nazývají *generátory*.



Obrázek 6.4: Cayleyho grafy malých grup, s typickým označením používaným v teorii grup. (a) Symetrie ozubeného kola S jako grupa sčítání celých čísel modulo 6. (b) Dihedrální grupa $\mathbb{D}_3$ symetrii rovnostranného trojúhelníku T. (c) Kleinova čtyřgrupa, neboli grupa binárních dvojsložkových vektorů. (d) Grupa binárních vektorů se třemi složkami tvoří krychli. (e) Grupa imaginárních jednotek kvaternionů, což jsou zobecnění komplexních čísel ve tvaru $a + bi + cj + dk$. (f) Grupa všech čtyřprvkových permutací $\mathbb{S}_4$, generovaná dvěma vyznačenými permutacemi.

Cayleyho graf pro grupu $\mathbb{G}$ je zkonstruován následovně. Jeho vrcholy odpovídají prvkům G. Zvolíme si libovolnou, typicky do inkluze minimální, generující množinu H. Každému prvku $h \in H$ přiřadíme jednu barvu a v grafu definuje množinu orientovaných hran

$$\big\{(g, h \circ g) : g \in G\big\}$$

této barvy. Příklady Cayleyho grafů pro několik významných grup jsou na obrázku 6.4. Pokud chceme zjistit, jaká je hodnota prvku $g \circ f$, stačí nalézt f v Cayleyho grafu a postupně se z něj posouvat podle šipek $h_1, \ldots, h_k$, případně proti směru pro $h_i \in H^{-1}$.

Multiplikativní tabulka. Nechť $\mathbb{G}$ je grupa řádu n s prvky $e = g_1, \ldots, g_n$. Operaci $\circ$ lze popsat tabulkou $n \times n$, kde na pozici (i, j) je umístěn prvek $g_j \circ g_i$. Tím je operace jednoznačně popsaná. Příklady těchto tabulek jsou na obrázku 6.5. Poznamenejme, že toto není moc praktický způsob popisu grupy, protože i pro malé grupy moc neprozradí o jejich struktuře. Proto ho vesměs nebudeme používat. Přesto si však můžeme všimnout několika vlastností. Grupa $\mathbb{G}$ je komutativní, právě když její multiplikativní je symetrická podle hlavní diagonály; podobně jako symetrie matic $A = A^\mathsf{T}$. Také si můžeme všimnout, že v každém řádku a sloupci se vyskytuje každý prvek grupy právě jednou. Jak brzo uvidíme, to pochopitelně není náhoda.

+	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	2	3	4	5	0
2	2	3	4	5	0	1
3	3	4	5	0	1	2
4	4	5	0	1	2	3
5	5	0	1	2	3	4

ℤ6

∘	e	r	r^2	z	rz	r^2z
e	e	r	r^2	z	rz	r^2z
r	r	r^2	e	r^2z	z	rz
r^2	r^2	e	r	rz	r^2z	z
z	z	rz	r^2z	e	r^2	r
rz	rz	r^2z	z	r^2	r	e
r^2z	r^2z	z	rz	r	e	r^2

𝔻3

Obrázek 6.5: Multiplikativní tabulky pro $\mathbb{Z}_6$ a $\mathbb{D}_3$ z obrázků 6.4a a b.

Nyní popíšeme několik základních typů grup, se kterými se čtenář často setká v matematice. Tyto grupy jsou natolik důležité, že dostaly speciální jména.

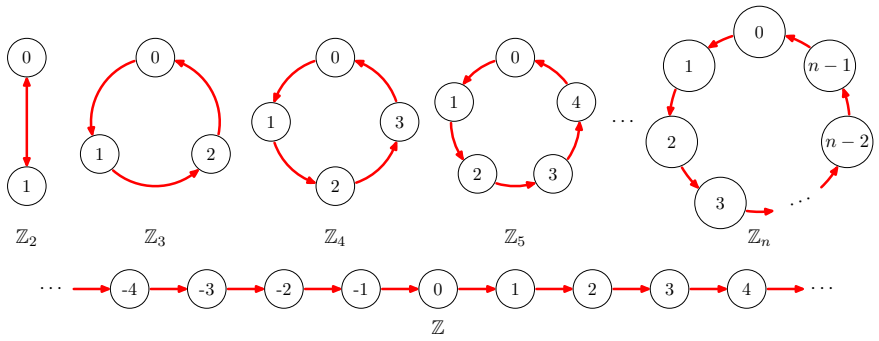
Cyklické grupy. Nejjednodušší třídou grup jsou *cyklické grupy* $\mathbb{Z}_n$ (někdy též $\mathbb{C}_n$). Ty jsou tvořené čísly $0, 1, \ldots, n-1$ a operací $+$, která se provádí modulo n. Tyto grupy jsou generované číslem 1, protože libovolné číslo lze získat opakovaným přičítáním jedničky. Cayleyho grafy pro několik prvních velikostí jsou vyobrazeny na obrázku 6.6. Nazývají se cyklické, protože jsou tvořené jediným cyklem. Popisují symetrie objektů, které lze pouze otáčet, ale nelze je překlápět; například výše uvedené ozubené kolo S má grupu symetrií $\mathbb{Z}_6$.

Značení $\mathbb{Z}_n$ je podle celých čísel $\mathbb{Z}$. Ta také tvoří grupu vzhledem k operaci $+$, která je tentokrát nekonečná. Je opět generovaná číslem 1. Cayleyho graf grupy $\mathbb{Z}$ je na obrázku 6.6 dole.

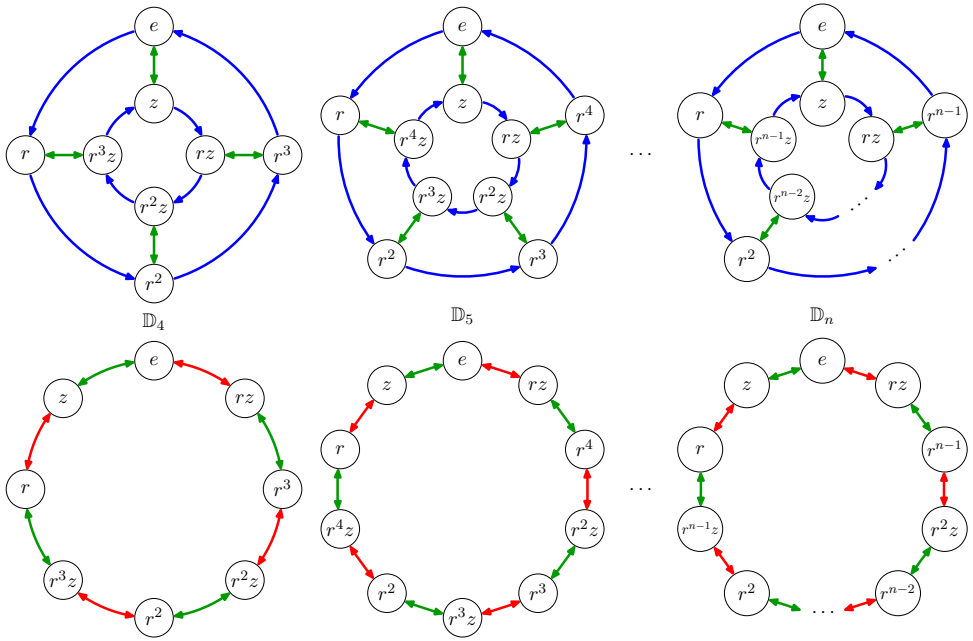
Dihedrální grupy. Třída dihedrálních grup $\mathbb{D}_n$ je tvořena grupami symetrií pravidelného n-úhelníka. Jak už jsme popsali, $\mathbb{D}_n$ obsahuje $2n$ prvků, což je n otočení a n zrcadlení. Označme r rotaci o úhel $360°/n$ a z jedno symetrické zrcadlení. Platí, že $\{r, z\}$ generuje $\mathbb{D}_n$ a na obrázcích 6.7 nahoře a obrázku 6.4b jsou takto generované Cayleyho grafy. Alternativně lze zvolit jako generující množinu dvě zrcadlení, například $\{rz, z\}$, a získáme jiné Cayleyho grafy ukázané na obrázku 6.7 dole.

Dihedrální grupy jsou nejjednodušší nekomutativní grupy. Důvodem je, že rotace a zrcadlení spolu nekomutují, tedy $rz \neq zr$. Z Cayleyho grafů lze nahlédnout, že obecně platí

$$zr^k = r^{n-k}z$$



Obrázek 6.6: Cayleyho grafy pro několik základních cyklických grup, obecně $\mathbb{Z}_n$ a $\mathbb{Z}$.



Obrázek 6.7: Cayleyho grafy malých dihedrálních grup a obecně $\mathbb{D}_n$. Nahoře generované $\{r, z\}$, dole generované $\{rz, z\}$. Rozmyslete si, kde by se v horních grafech vyskytovali červené šipky.

a také $rzrz = e$. Poznamenejme, že se někdy $\mathbb{D}_n$ značí $\mathbb{D}_{2n}$ podle počtu prvků prvků; toto však nebudeme nikde v textu využívat.

Permutace. Permutace, se kterými jsme se již zabývali v kapitole 3 v souvislosti s permutačními maticemi, hrají klíčovou roli v teorii grup. Nechť $X = \{1, 2, \ldots, n\}$. Permutace π na X je bijektivní zobrazení

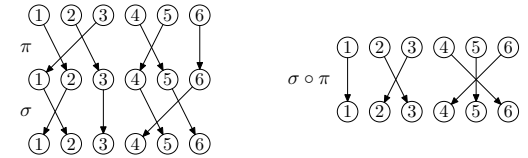
$\pi : X \to X$. Tedy π přiřazuje každému prvku $i \in X$ nějaký prvek $\pi(i) \in X$, a to každému prvku X jiný prvek. Permutaci je transformace, která nějak přehází prvky X.

Permutaci lze reprezentovat několik způsoby. Je možný *maticový zápis*, kde do prvního řádku zapíšeme prvky X a do druhého řádku jejich obrazy po aplikování permutace. Tedy například dvě permutace π a σ můžeme reprezentovat jako

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 3 & 1 & 5 & 4 & 6 \end{pmatrix} \qquad \text{a} \qquad \sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 1 & 3 & 5 & 6 & 4 \end{pmatrix}.$$

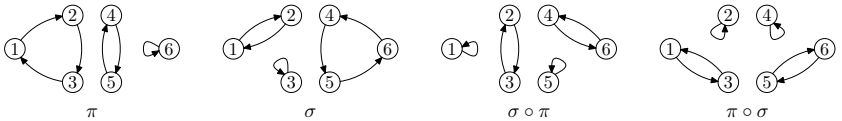
V řadě situací je navíc možné vynechat první řádek a zapsat $\pi = (2, 3, 1, 5, 4, 6)$ a $\sigma = (2, 1, 3, 5, 6, 4)$.

Permutace na množině X lze skládat a čtenář může ověřit, že výsledkem jsou opět permutace na množině X. Skládání je dobře vidět v reprezentaci grafem, kde umístíme prvky X ve dvou kopiích, do dvou řádků. Z prvku i v horním řádku vedeme šipku do prvku $\pi(i)$ v dolním řádku. Složení $\sigma \circ \pi$ pak vypadá tak, že jejich reprezentace dáme pod sebe a z každého prvku i uděláme dva kroky podél šipek; viz obrázek 6.8.



Obrázek 6.8: Složení permutací π a σ a výsledná permutace $\sigma \circ \pi$ napravo.

Druhý způsob zápisu permutace je *cyklová reprezentace*. To je graf, kde vrcholy jsou jednotlivé prvky X a z každého prvku i vede právě jedna šipka do $\pi(i)$. Tato reprezentace se nazývá cyklová, protože se graf skládá z kolekce cyklů. Každý cyklus popisuje skupinu prvků, které permutace točí dokola. Cyklus může být tvořený i jediným prvkem, kterému se říká *pevný bod* permutace. Obrázek 6.9 obsahuje cyklové reprezentace výše uvedených permutací.



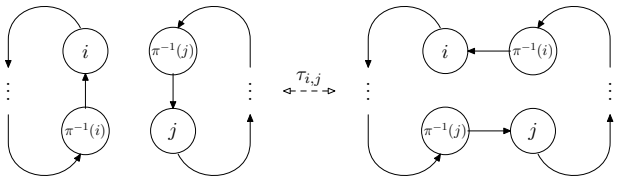
Obrázek 6.9: Permutace π a σ a jejich dvě složení. Jak je ilustrováno, skládání permutací je typicky nekomutativní.

Cyklová reprezentace ukazuje hodně ze struktury permutace. Je obzvlášť vhodná pro určování *mocnin*, neboť π^k je permutace, ve které zobrazujeme každý prvek podél cyklu o k šipek. Pokud délka tohoto cyklu je přesně k, stane se z něj množina pevných bodů. Čtenář si může rozmyslet, jaké je nejmenší k takové, že $\pi^k = \mathrm{id}$, v závislosti na struktuře permutace π. Permutace je involuce, pokud je tvořena pouze pevnými body a cykly délky dva. Pro každou permutaci π existuje také oboustranná inverzní permutace π^{-1}, která vznikne z π obrácením směru šipek.

Symetrické grupy. Symetrická grupa $\mathbb{S}_n$ je grupa všech n-prvkových permutací s operací skládání. Tyto grupy jsou velmi důležité v teorii grup kvůli svojí univerzalitě, což vysvětlíme později. Není moc vhodné je vizualizovat pomocí Cayleyho diagramů, protože obsahují příliš mnoho prvků; řád $\mathbb{S}_n$ je $n!$ a faktoriály rostou velice rychle. Grupy $\mathbb{S}_0$ a $\mathbb{S}_1$ jsou triviální jednoprvkové grupy a $\mathbb{S}_2$ je stejná jako $\mathbb{Z}_2$. Grupa $\mathbb{S}_3$ má

totožnou strukturu jako $\mathbb{D}_3$ z obrázku 6.4b, čtenář si může rozmyslet, které dvě permutace odpovídají r a z. Grupa $\mathbb{S}_4$ je vyobrazena na obrázku 6.4f.

Speciální druh permutací jsou *transpozice*. Označme permutaci složenou z pevných bodů a jednoho cyklu délky dva tvořeného i a j jako transpozici $\tau_{i,j}$. Transpozice je tedy permutace, která prohodí i a j, a nechá ostatní prvky na místě. Klíčovou vlastností je, že libovolnou permutaci π lze vyjádřit jako složení transpozic. Obrázek 6.10 popisuje vliv transpozice $\tau_{i,j}$ na permutaci π. Libovolnou permutaci π tedy můžeme vybudovat postupně z id vytvářením jednotlivých cyklů. Pokud chceme vložit pevný bod j mezi $\pi^{-1}(i)$ a i, stačí složit dosud vybudovanou permutaci s $\tau_{i,j}$.



Obrázek 6.10: Vliv složení s $\tau_{i,j}$ na permutace π. Pokud jsou i a j obsaženy v různých cyklech, postupuje se zleva doprava. Pokud jsou obsaženy v jednom cyklu, postupuje se zprava doleva. Tedy pokud jsou obsaženy v jednom cyklu před složením, jsou v rozdílných cyklech po složení, a naopak.

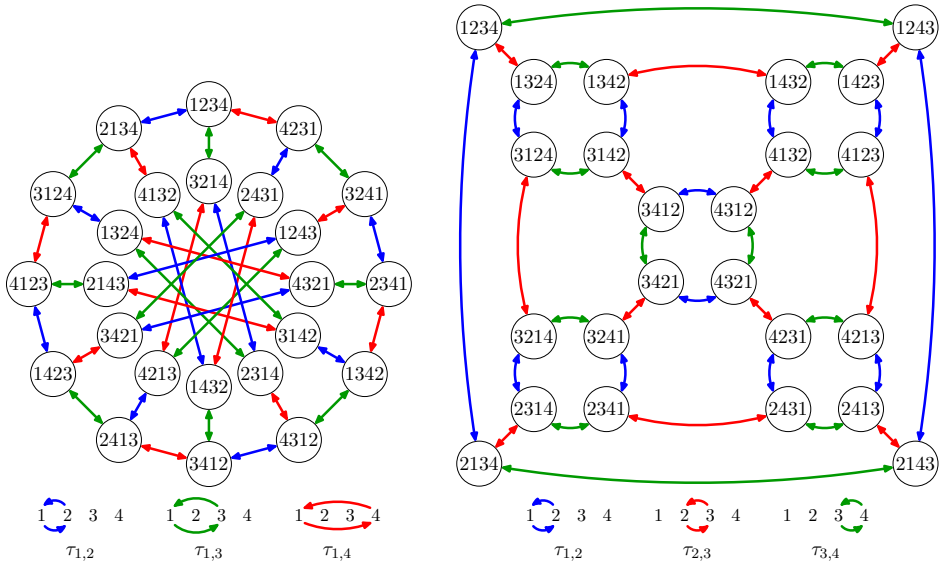
Tedy množina všech transpozic generuje grupu $\mathbb{S}_n$, avšak tato generující množina není do inkluze minimální. Existuje řada rozumných voleb do inkluze minimální generujících množin. Například je možné zvolit množinu $n-1$ transpozic $\{\tau_{1,2}, \tau_{1,3}, \ldots, \tau_{1,n}\}$. Podobně je možné zvolit množinu $n-1$ transpozic $\{\tau_{1,2}, \tau_{2,3}, \ldots, \tau_{n-1,n}\}$. Obrázek 6.11 ukazuje takto generované Cayleyho grafy pro $\mathbb{S}_4$. Pokud netrváme na transpozicích, je možné zvolit dvouprvkovou generující množinu $\{\pi = (2, 3, 4, \ldots, n, 1), \tau_{i,i+1}\}$ pro libovolný index i. Čtenář si může ve cvičení 6.1, že jsou tyto množiny vskutku do inkluze minimální generující množin $\mathbb{S}_n$.

Podgrupy. Budeme uvažovat podstruktury v rámci grupy, podobně jako jsme v kapitole 2 uvažovali vektorové podprostory uvnitř vektorového prostoru. Definice je téměř totožná, jako obecně pro libovolnou algebraickou podstrukturu. Nechť $\mathbb{G}$ je grupa. Neprázdná podmnožina $H \subseteq G$ spolu s operací $\circ$ tvoří *podgrupu*, pokud je uzavřená na operaci $\circ$ a na inverze. Podgrupu H můžeme také uvažovat jako strukturu $\mathbb{H}$, neboť je sama grupou. Vztah býti podgrupou budeme zapisovat jako $\mathbb{H} \leq \mathbb{G}$, a symbol $<$ se používá při ostré inkluzi $H \subsetneq G$. Z vlastností podgrupy automaticky vyplývá, že H obsahuje neutrální prvek e. Protože je H neprázdná, existují $g, g^{-1} \in H$ a z uzavřenosti na součin také $gg^{-1} = e$ leží v H. Pokud je $H = \{e\}$ nebo $H = G$, nazývá se podgrupa *triviální*. Obrázek 6.12 ukazuje všechny podgrupy dihedrální grupy $\mathbb{D}_3$.

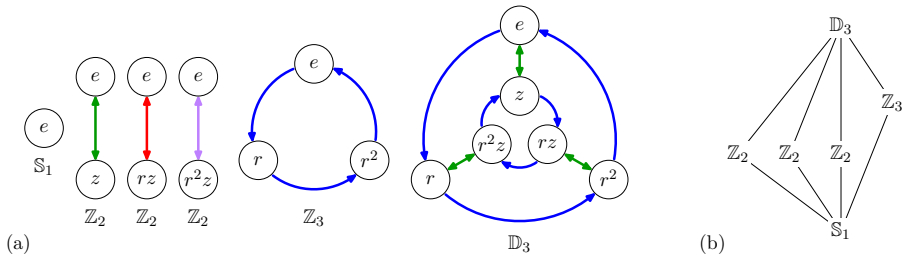
V kapitole 2.6 jsme popsali, že podprostory vektorového prostoru tvoří uspořádanou strukturu zvanou úplný svaz; tedy že existují infima a suprema. Podobná věc platí i pro grupy a jejich podgrupy. Infimum množiny podgrup je opět jejich průnik. Pro supremum nejprve definujme *algebraický uzávěr* $\langle H \rangle$ libovolné podmnožiny $H \subseteq G$. Podobně jako v definici generující množiny, nechť $H^{-1} = \{h^1 : h \in H\}$.

$$\langle H \rangle = \{h_k \circ h_{k-1} \circ \cdots \circ h_2 \circ h_1 : k \in \mathbb{N} \text{ a } h_1, \ldots, h_k \in H \cup H^{-1}\}.$$

Tedy H je generující množina, právě když $\langle H \rangle = G$. Algebraický uzávěr je obdoba lineárního obalu, pouze název lineární obal je typicky rezervovaný pro vektorové prostory. Supremum množiny podgrup je uzávěr jejich sjednocení. Čtenář si může zkusit dokázat ve cvičení 6.3, že jsme skutečně sestrojili úplný svaz. Tento svaz poví ledacos o struktuře grupy, avšak rozhodně nemá tak silné vlastnosti jako svaz vektorových podprostorů.



Obrázek 6.11: Dva Cayleyho grafy grupy $\mathbb{S}_4$ generované transpozicemi.



Obrázek 6.12: Všechny podgrupy dihedrální grupy $\mathbb{D}_3$ a Hasseho diagram jejich inkluzí.

Parita permutace. Zmínili jsme, že každá permutace π se dá zapsat jako složení transpozic. Položme si otázku, kolik transpozic v tomto zápisu π je. Rychle si můžeme všimnout, že toto číslo není jednoznačně určené. Pokud totiž lze π vyjádřit jako složení k transpozic, lze výsledný zápis dvakrát složit s třeba $\tau_{1,2}$, čímž se výsledek nezmění, a tedy získat zápis π pomocí $k+2$ transpozic.

Položme tedy jinou otázku: Pro která čísla k lze π vyjádřit jako složení k transpozic? Čtenář může jako cvičení nalézt pro danou permutací π nejmenší takové k. V předchozím odstavci jsme ukázali, že to splňuje i každé větší číslo se stejnou paritu. Ukážeme si, že tato parita je jedna z vlastností permutace. Tedy že není možné najít permutaci π, která by byla složením sudého počtu transpozic a zároveň složeným lichého počtu transpozic. Proto se podle parity k nazývá permutace buď *lichá* nebo *sudá*.

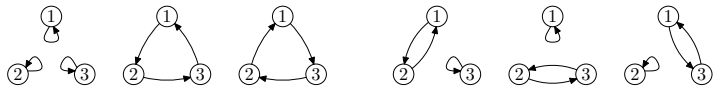
Lemma 6.1. *Každá permutace je buď lichá, nebo sudá.*

Důkaz. Mějme libovolnou permutaci a složme ji s transpozicí $\tau_{i,j}$. Již jsme popsali následující: Pokud i

a j patří do různých cyklů v π, budou jejich cykly po aplikování $\tau_{i,j}$ sloučeny. Pokud naopak i a j patří do jednoho cyklu, transpozice $\tau_{i,j}$ rozdělí tento cyklus na dva. V obou příkladech se počet cyklů změní o jedna; buď se zvýší, nebo se sníží.

Proto je parita permutace určena paritou počtu jejích cyklů. Identita id je sudá permutace a má n cyklů. Parita libovolné permutace π s k cykly je rovná paritě čísla $n-k$. Pokud by totiž bylo možné vyjádřit π jako složení počtu transpozic s opačnou paritou, potom by π nemohla obsahovat k cyklů. □

Čtenář si může rozmyslet, že alternativní definice parity permutace je parita počtu jejích sudých cyklů. Obrázek 6.13 ukazuje paritu všech tříprvkových permutací. Vždy platí, že polovina permutací je sudá a polovina je lichá. Důvod je, že sudé a liché permutace můžeme spárovat složením například s $\tau_{1,2}$; toto složení mění paritu a vytváří bijekci mezi sudými a lichými permutacemi. Například v případě obrázku 6.13 je levá sudá permutace spárována s levou lichou permutací, prostřední s prostřední a pravá s pravou. Tedy sudých a lichých n-prvkových permutací je shodně $n!/2$.



Obrázek 6.13: Všechny tříprvkové permutace rozdělené podle parity.

Typicky se parita permutace označuje *znaménkem* $\operatorname{sgn}(\pi) \in \{-1, 1\}$, kde 1 je pro sudou permutaci a -1 pro lichou. Důvod pro toto značení je následující. Dvě permutace můžeme složit tak, že složíme jejich zápisy pomocí transpozic, a tedy délka tohoto zápisu se sečte. Proto složení dvou sudých permutací je sudá permutace, složení dvou lichých je také sudá permutace a složení liché a sudé permutace (v libovolném pořadí) je lichá permutace. Z pohledu znaménka říká objevený vztah, že

$$\operatorname{sgn}(\sigma \circ \pi) = \operatorname{sgn}(\sigma) \cdot \operatorname{sgn}(\pi).$$

Alternující grupy. Jak už jsme zmínili, sudé permutace jsou uzavřené na skládání, a podobně jsou i uzavřené na inverze. Proto množina všech sudých permutací tvoří podgrupu $\mathbb{S}_n$, které se nazývá *alternující grupa* a značí se $\mathbb{A}_n$. Tyto grupy jsou strukturálně velice zajímavé a mají mnoho důsledků v teorii grup. Grupy $\mathbb{A}_2$ a $\mathbb{A}_3$ mají totožnou strukturu s $\mathbb{S}_1$ a $\mathbb{Z}_3$. Cayleyho grafy dalších dvou alternujících grup jsou na obrázku 6.14.

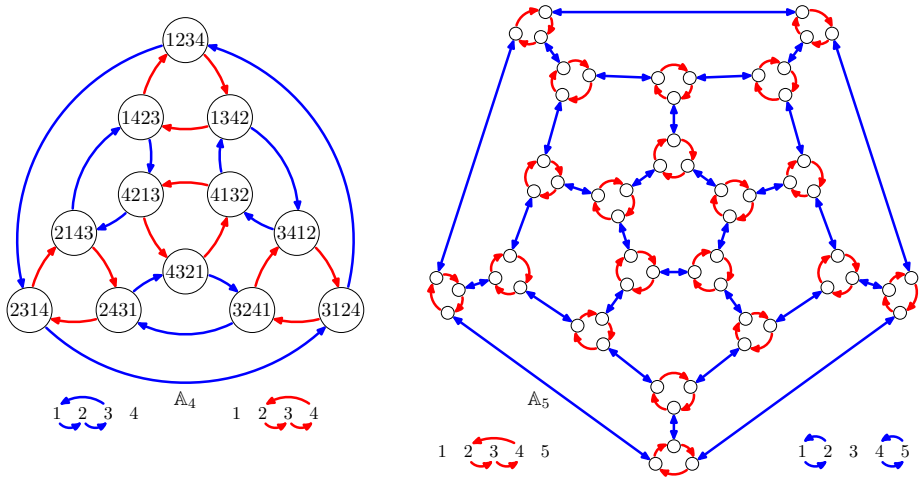
Zobrazené Cayleyho grafy pro malé symetrické a alternující grupy jsou zajímavé objekty související s Platonskými tělesy. Například Cayleyho graf pro $\mathbb{A}_5$ vznikl z dvanáctistěnu (modré hrany) nahrazením každého vrcholu červeným cyklem délky tři zorientovaným po směru hodinových ručiček. Pokud uvažujeme pouze rotace, grupa symetrií pravidelného dvanáctistěnu je přesně $\mathbb{A}_5$. Proto tato souvislost není náhodná.

Grupové homomorfismy. Podobně jako pro vektorové prostory v kapitole 5, definujeme důležité zobrazení mezi grupami zvané homomorfismy. Nechť $\mathbb{G}$ a $\mathbb{H}$ jsou dvě grupy. Zobrazení $f : G \to H$ se nazývá *homomorfismus*, pokud splňuje dvě následující vlastnosti:

(i) Zobrazení f zachovává grupovou operaci. Pro libovolné prvky $x, y \in G$ platí, že

$$f(xy) = f(x) \circ f(y).$$

(ii) Obrazem neutrálního prvku je neutrální prvek: $f(e) = e$.



Obrázek 6.14: Cayleyho grafy alternujících grup $\mathbb{A}_4$ a $\mathbb{A}_5$. Jejich generující množiny jsou dvě sudé permutace, vzniklé složení sudě mnoha transpozic. V případě Cayleyho grafu $\mathbb{A}_5$ jsme vynechali z důvodu nedostatku místa popisky, čtenář si je může zkusit doplnit.

Čtenář může dokázat, že z těchto dvou vlastností také vyplývá, že pro libovolný prvek $x \in G$ platí $f(x^{-1}) = f(x)^{-1}$; tedy i inverzní prvky jsou zachovány. Podobně jako v případě vektorových prostorů, homomorfismus f vnořuje strukturu grupy $\mathbb{G}$ do grupy $\mathbb{H}$. Podrobněji se důsledky této vlastnosti budeme zabývat v kapitole **??**. Ve cvičeni 6.5 si může čtenář rozmyslet definici homomorfismu obecné matematické struktury.

Podobně jako předtím, homomorfismus f se nazývá *izomorfismus*, pokud je f bijektivní zobrazení. Dvě grupy $\mathbb{G}$ a $\mathbb{H}$ jsou *izomorfní*, pokud mezi nimi existuje izomorfismus, což se značí $G \cong H$. Existence izomorfismu implikuje, že mají stejnou algebraickou strukturu. Homomorfismus z $\mathbb{G}$ do $\mathbb{G}$ se nazývá *endomorfismus*, a pokud je navíc bijektivní, je to *automorfismus*.

Cayleyho věta. Již několik stránek se zabýváme strukturou permutací a symetrickými grupami. Proto tyto pojmy musí být velice užitečné v teorii grup. Nyní si dokážeme Cayleyho větu, která říká, že symetrické grupy jsou univerzální.

Věta 6.2 (Cayley). *Libovolná grupa $\mathbb{G}$ řádu n je izomorfní nějaké podgrupě $\mathbb{S}_n$.*

Proč tato věta dává smysl zjistíme po prozkoumání Cayleyho grafů ve výše uvedených obrázcích. Hrany každé barvy odpovídající jednomu z generátorů popisují permutaci prvků grupy. Totiž z každého prvku vychází právě jedna šipka této barvy a do každého prvku vchází právě jedna šipka této barvy. Cayleyho graf celý, a tedy i struktura grupy, vznikne zkombinováním několika takových permutací. Podobně v multiplikativních tabulkách na obrázku 6.5 se v každém sloupci a každém řádku nachází každý prvek grupy právě jednou. Tedy opět každý sloupec a řádek definuje permutaci ve stylu maticového zápisu. Dokažme toto obecně pro každou grupu:

Lemma 6.3. *Násobení prvkem f zleva definuje permutaci na množině prvků grupy.*

Důkaz. Stačí si uvědomit, že v grupách můžeme krátit. Tedy jestliže platí $fg = fh$, potom také platí

$g = h$. Důvodem je existence inverzí, stačí totiž vynásobit rovnici zleva f^{-1}. Proto musíme pro různé prvky g dostat různé hodnoty fg, a násobení zleva je skutečně permutace. □

Výše uvedená permutace se nazývá *levá translace* prvku f. Podobně násobení zprava definuje permutaci zvanou *pravá translace* prvku f. Poznamenejme, že pro nekomutativní grupy jsou tyto translace typicky různé permutace. V souvislosti s tímto si čtenář může rozmyslet cvičení 6.4.

Důkaz Cayleyho věty 6.2. Podle lemma 6.3 víme, že každý prvek $f \in \mathbb{G}$ definuje permutaci na množině prvků zvanou levá translace, kterou označme $\ell(f)$. Prvky grupy chceme identifikovat s jejich levými translacemi. Tvrdíme, že toto zobrazení ℓ je hledaný izomorfismus, tedy že grupa $\mathbb{G}$ má stejnou strukturu jako levé translace s operací skládání.

Pro dokázání izomorfismu musíme ukázat tři věci:

1. Každé dva prvky v $\mathbb{G}$ mají rozdílné levé translace, tedy přiřazení $\ell : \mathbb{G} \to \mathbb{S}_n$ je prosté.
2. Levá translace neutrálního prvku $\ell(e)$ je identita.
3. Skládání levých translací má stejnou algebraickou strukturu jako násobení v $\mathbb{G}$. Tedy kdykoliv platí $f = gh$, potom
$$\ell(f) = \ell(g) \circ \ell(h).$$

První bod platí mnohem silněji, pro libovolné dva různé prvky grupy $\mathbb{G}$ se jejich translace neshodují v žádném bodě. Pokud by totiž platilo, že $\ell(f)(x) = \ell(g)(x)$, potom v řeči $\mathbb{G}$ dostáváme $fx = gx$. Protože můžeme krátit zprava (vynásobením x^{-1} zprava), dostáváme $f = g$. V Cayleyho grafu to odpovídá vlastnosti, že dva vrcholy mohou být spojené pouze jednou šipkou. Tedy zobrazení ℓ je prosté.

Druhý bod vyplývá triviálně, neboť násobení neutrálním prvkem e zleva žádný prvek grupy $\mathbb{G}$ nemění. Tedy $\ell(e) = \mathrm{id}$.

Třetí část platí z asociativity $\mathbb{G}$. Totiž pro libovolný prvek $x \in \mathbb{G}$ platí $g(hx) = (gh)x = fx$. Tedy levá translace $\ell(f)(x)$ má stejnou hodnotu jako nejprve aplikování levé translace $\ell(h)(x)$, a na výsledek aplikování levé translace $\ell(g)(\ell(h)(x))$. Což je přesně skládání permutací. Proto je přiřazení ℓ mezi grupou $\mathbb{G}$ a jejími levými translacemi izomorfismus. □

Maticové grupy. Zaměřme se na čtvercové matice $n \times n$. Jako hlavní výsledek jsme v kapitole 3 dokázali, že pokud má čtvercová matice inverzi z jedné strany, má i inverzi z druhé strany. Tedy regulární matice splňují vlastnost o existenci oboustranné inverze. Protože je snadné nahlédnout, že i ostatní vlastnosti z definice grupy jsou splněny, dokázali jsme, že regulární matice tvoří grupu. Tato velice důležitá nekonečná grupa regulárních matic $n \times n$ se nazývá *lineární grupa* stupně n a značí se GL_n.

Opět platí, že maticové grupy jsou univerzální. Tedy každá grupa $\mathbb{G}$ řádu n je podgrupa GL_n. To snadno vyplývá z Cayleyho věty, která umožňuje identifikovat prvky $\mathbb{G}$ izomorfně s podgrupou $\mathbb{S}_n$. Klíčové je, že $\mathbb{S}_n$ je izomorfní s grupou všech permutačních matic, což jsou matice, které jsme definovali v kapitole 3. Proto můžeme jednotlivým prvkům $\mathbb{G}$ místo permutací přiřadit příslušné permutační matice, a výsledkem je podgrupa GL_n izomorfní $\mathbb{G}$. Výhoda toho přístupu je, že často můžeme i velice složité (i nekonečné) grupy vygenerovat jako podgrupy GL_n pro malé n. Uveďme si několik zajímavých příkladů maticových podgrup GL_2.

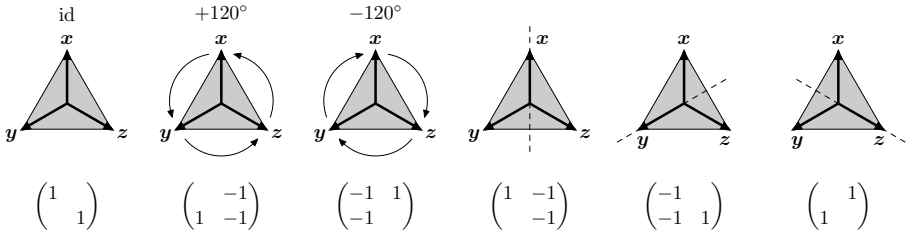
Uvažme množinu komplexních čísel $\mathbb{C}^* = \mathbb{C} \setminus \{0\}$ a operaci násobení. Tato množina tvoří grupu a lze ji reprezentovat pomocí reálných matic 2×2. Pro pochopení tohoto výsledku je potřeba porozumět, jak funguje násobení komplexním číslem. Na násobení x zleva můžeme nahlédnout jako na operaci $\mathbb{C}^* \to \mathbb{C}^*$ definovanou jako zobrazení $y \mapsto x \cdot y$. Tristam Niedham využívá ve své knize Visual Complex Analysis pro popis této operaci krásnou anglickou složeninu *amplitwist*. Násobení komplexním číslem x roztahuje (amplifikuje) a rotuje (kroutí) komplexní rovinu. Konkrétně roztažení je podle koeficientu $|x|$ a otočení je o úhel, který svírá x s reálnou osou. Pokud uvážíme komplexní rovinu bez nuly, je toto

násobení permutace komplexní roviny, což je levá translace grupy. Tento geometrický pohled je mnohem užitečnější než vzoreček
$$(a + bi)(c + di) = ac - bd + (ad + bc)i.$$
Existuje velice elegantní zápis pomocí komplexní exponenciely. Označme $r = |x|$ a φ úhel s reálnou osou. Potom $x = re^{i\varphi}$.

A jaká je tedy maticová reprezentace $\mathbb{C}^*$? Stačí popsat příslušnou transformaci komplexní roviny jako lineární zobrazení z $\mathbb{R}^2$ do $\mathbb{R}^2$, které můžeme uvažovat vůči kanonické bázi tvořené 1 a i. Protože potřebujeme, aby každá matice byla regulární, nemůžeme uvažovat násobení nulou, které by odpovídalo nulové matici. Výsledná matice reprezentující $a + bi = re^{i\varphi}$ je kombinace roztažení a rotace:
$$r \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix},$$
kde levá část je odvozena v kapitole 3. Násobení matic odpovídá skládání zobrazení, což je přesně násobení komplexních čísel.

Jako další příklad uveďme maticovou reprezentaci grupy $\mathbb{D}_n$, což jsou symetrie pravidelného n-úhelníka. Ten vnoříme do $\mathbb{R}^2$ tak, že jeho střed umístíme do počátku a jeden vrchol řekneme na souřadnici $(0, 1)$. Budeme uvažovat lineární zobrazení vůči kanonické bázi. Rotace n-úhelníka o úhel $\varphi = 360°/n$ odpovídají rotacím roviny a překlopení n-úhelníka odpovída zrcadlení podle osy y. Dostáváme, že
$$\mathbb{D}_n \cong \left\langle \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}, \begin{pmatrix} -1 & \\ & 1 \end{pmatrix} \right\rangle.$$

Ještě elegantnější reprezentaci lze vytvořit pro grupu $\mathbb{D}_3$ zvolením jiné báze. Místo kanonické báze budeme uvažovat bázi tvořenou vektory $\boldsymbol{x}$ a $\boldsymbol{y}$ ve vrcholech trojúhelníka. Třetí vrchol roven $-\boldsymbol{x} - \boldsymbol{y}$ označme $\boldsymbol{z}$. Potom dostáváme na obrázku 6.15 jednoduché matice odpovídající jednotlivým symetrickým transformacím. Tedy platí, že
$$\mathbb{D}_3 \cong \left\langle \begin{pmatrix} & -1 \\ 1 & -1 \end{pmatrix}, \begin{pmatrix} & 1 \\ 1 & \end{pmatrix} \right\rangle.$$



Obrázek 6.15: Elegantní maticová reprezentace grupy $\mathbb{D}_3$ pro volbu vhodné báze $\boldsymbol{x}$ a $\boldsymbol{y}$.

Řád prvku. Jak už jsme zmínili, každý prvek definuje permutace zvané levé a pravá translace. Při prohlédnutí ilustrovaných Cayleyho grafů si můžeme všimnout, že tyto permutace jsou velice speciální. Vypadají zcela odlišně například od permutací vyobrazených na obrázku 6.9. Konkrétně každá levá/pravá translace prvku f je permutace tvořená cykly pouze jedné délky. Tato délka se nazývá *řád prvku* f, i když se typicky definuje jiným způsobem.

Nejprve se podívejme na mocniny f^k, tedy uvažme posloupnost $e = f^0, f^1, f^2, \ldots$. Protože je grupa konečná, musí se tyto od určité hodnoty k opakovat. Označme k nejmenší mocninu takovou, že existuje $0 \leq \ell < k$ splňující $f^k = f^\ell$. Tvrdíme, že $\ell = 0$, tedy $f^k = \mathrm{id}$. Pokud by totiž $\ell > 0$, potom by levá translace prvku f zobrazovala dva různé prvky f^{k-1} a $f^{\ell-1}$ na stejný prvek, což není možné.

Nyní každý cyklus levé translace prvku f musí mít délku určitě dělitelnou k, neboť $f^k x = ex = x$. Naopak pokud platí $f^\ell x = x$, potom vynásobením x^{-1} zprava dostáváme $f^\ell = e$, a tedy ℓ je násobek k. Poznamenejme, že k musí dělit řád grupy.

Akce grupy na množině. Na začátku této kapitoly jsme motivovali grupy přes struktury symetrií geometrických objektů. Udělejme revizi této motivace, ve které si vysvětlíme souvislosti grup se strukturou transformací objektů.

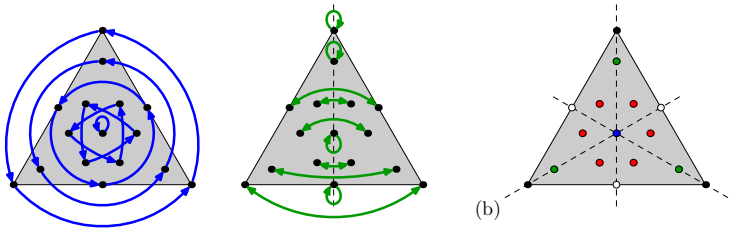
Nechť X je libovolná množina. Akce grupy $\mathbb{G}$ na množině X přiřazuje každému prvku $f \in G$ jednu permutaci množiny X. Tedy akce grupy na množině je operace $\circ : G \times X \to X$. Tato operace musí splňovat několik základních vlastností, aby struktura $\mathbb{G}$ byla aplikována na transformace X:

- Pro libovolný prvek $f \in \mathbb{G}$ je $\circ$ permutace, tedy pokud $f \circ x = f \circ y$, potom nutně $x = y$.
- Neutrální prvek $e \in \mathbb{G}$ odpovídá identitě. Tedy pro libovolný prvek $x \in X$ platí $e \circ x = x$.
- Akce je kompatibilní se strukturou grupy. Pro libovolné $f, g \in \mathbb{G}$ a libovolný prvek $x \in X$ platí

$$g \circ (f \circ x) = (gf) \circ x.$$

Alternativně se akce grupy na množině definuje jako homomorfismus z $\mathbb{G}$ do $\mathbb{S}_X$, kde $\mathbb{S}_X$ je grupa všech permutací prvků X.

Například uvažme rovnostranný trojúhelník T z úvodu kapitoly. Jeho grupa symetrií je $\mathbb{D}_3$, která definuje akci na množině T. V úvodu kapitoly jsme uvažovali pouze to, jak symetrie prohazují vrcholy trojúhelníku. Avšak můžeme i uvažovat, jak symetrie permutují všechny body T, kterých je nekonečně mnoho. Obrázek 6.16a ukazuje, jaké permutace jsou generátory $\mathbb{D}_3$. Pro popsání akce pochopitelně stačí přiřadit permutace libovolné generující množině grupy, zbytek je již jednoznačně určený skládáním.



(a) (b)

Obrázek 6.16: (a) Akce generátorů $\mathbb{D}_3$ na množinu bodů rovnostranného trojúhelníka, vyobrazeny jsou některé body. Rotace je permutace složená z cyklů délky tři a přesně jednoho pevného bodu. Zrcadlení má pevné body na těžnici a cykly délky dva všude jinde. (b) Různými barvami vyobrazeno několik orbit této akce. Střed trojúhelníka tvoří jedinou jednoprvkovou orbitu. Body ležící právě na jedné těžnici patří do tříprvkových orbit, ostatní body trojúhelníka patří do orbit velikosti šest.

Zaveďme užitečnou terminologii pro popisování akcí.

- *Orbita* $[x]$ - pro libovolný prvek $x \in X$ je to množina

$$\{g \circ x : g \in G\}.$$

Tedy orbita je přesně podmnožina prvků X, na které lze zobrazit prvek x zobrazit akcí grupy. Čtenář si může rozmyslet, že z vlastností akce mají libovolné dva prvky ze stejné orbity stejnou orbitu, tedy $[x] = [g \circ x]$. Proto orbity rozdělují prvky X na třídy ekvivalence. Na obrázku 6.16b jsou naznačeny některé orbity akce symetrií na rovnostranném trojúhelníku T. Můžeme si všimnout, že orbity mohou mít různou velikost, ale jejich velikost dělí řád $\mathbb{G}$; to není náhoda a platí to obecně.

- *Stabilizátor* $\mathrm{Stab}(x)$ - pro libovolný prvek x je to množina

$$\{g : g \in G \text{ a } g \circ x = x\}.$$

Tedy stabilizátor x je množina všech prvků grupy, pro které je x pevným bodem. Čtenář si může rozmyslet, že stabilizátor libovolného prvku je vždy podgrupa $\mathbb{G}$. Intuitivně platí, že čím větší je orbita $[x]$, tím méně prvků grupy může fixovat x, a tedy tím menší je $\mathrm{Stab}(x)$. Toto přesně vyjadřuje věta o orbitě a stabilizátoru, která říká, že vždy platí

$$\big|[x]\big| \cdot \big|\mathrm{Stab}(x)\big| = |\mathbb{G}|.$$

- Akce se nazývá *tranzitivní*, pokud obsahuje pouze jednu orbitu $[x] = X$. Tedy libovolný prvek lze zobrazit na libovolný vhodným prvkem grupy.
- Akce je *semiregulární*, pokud žádný prvek grupy výjma identity neobsahuje pevný bod. Tedy alternativně jsou všechny stabilizátory triviální podgrupy. Můžeme si povšimnout, že semiregularita implikuje následující: Kdykoliv $g \circ x = h \circ x$, potom $g = h$.
- Akce se označuje *regulární*, pokud je současně tranzitivní a semiregulární.

Ukažme si další příklady akcí grup na množině, kvůli kterým jsme zaváděli tuto terminologii. Každá grupa $\mathbb{G}$ definuje akci sama na sobě pomocí levých translací. Tato akce je velice speciální, je totiž vždy regulární. Semiregularitu jsme dokázali v důkazu Cayleyho věty 6.2. Regularita platí, neboť pro libovolné prvky $g, h \in \mathbb{G}$ zobrazuje levá translace prvku hg^{-1} prvek g na prvek h. Cayleyho grafy jsou tedy vizualizace generátorů této akce, podobně jako obrázek 6.16 vizualizuje akci symetrií na rovnostranném trojúhelníku T.

Grupy automorfismů. Automorfismy jsme v kapitole 5 viděli definované pro vektorové prostory, avšak mohou se uvažovat pro libovolnou matematickou strukturu. Nechť X je libovolná množina se strukturou, tedy nějakými operacemi a relacemi, jak jsme popsali v kapitole 2.5. Automorfismus $\pi : X \to X$ je bijektivní zobrazení, které zachovává strukturu X. Množina všech automorfismů X v matematice vždy tvoří grupu, která se značí $\mathrm{Aut}(X)$.

Zkoumáním grupy automorfismů se dozvíme celou řadu vlastností matematické struktury X. Například grupa $\mathrm{Aut}(\mathbb{R}^n)$ je tvořena regulárních maticemi $n \times n$ a tvoří již popsanou lineární grupu GL_n. V teorii grup hrají významnou roli grupy automorfismů grup, tedy $\mathrm{Aut}(\mathbb{G})$, kde $\mathbb{G}$ je nějaká grupa. Například platí, že $\mathrm{Aut}(\mathbb{Z}_2^2) \cong \mathbb{S}_3$.

Grupa automorfismů $\mathrm{Aut}(X)$ popisuje akci na množině X. Každý automorfismus definuje nějakou permutaci na množině X, a jejich skládání je skládání těchto permutací. Libovolná orbita $[x]$ je tvořena prvky X, které hrají ve struktuře totožnou roli. Například akce $\mathrm{Aut}(\mathbb{R}^n)$ má dvě orbity, jedna je tvořena pouze nulovým vektorem $\mathbf{0}$ a druhá všemi ostatními vektory. Je jednoduché sestrojit automorfismus, který zobrazuje libovolný nenulový vektor $\boldsymbol{x}$ na libovolný jiný nenulový vektor $\boldsymbol{y}$. Toto rozlišení prvků $\mathbb{R}^n$ dává smysl, neboť $\mathbf{0}$ se zásadně odlišuje od libovolného nenulového vektoru.

Pochopitelně můžeme uvažovat libovolnou podgrupu $\mathrm{Aut}(X)$ a akci tvořenou pouze prvky této podgrupy. Například zvolme jako podgrupu $\mathrm{Aut}(\mathbb{R}^n)$ množinu všech permutačních matic. Potom do jedné orbity patří vektory s danými koeficienty, pouze s proházeným pořádím. Například pro $\mathbb{R}^3$ získáváme orbity

$$\begin{aligned}\big[(1,2,2)\big] &= \big\{(1,2,2),(2,1,2),(2,2,1)\big\},\\ \big[(1,2,3)\big] &= \big\{(1,2,3),(1,3,2),(2,1,3),(2,3,1),(3,1,2),(3,2,1)\big\}.\end{aligned}$$

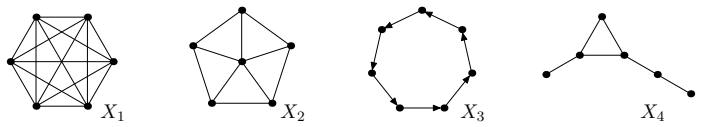
Čtenář si může rozmyslet, jak velká je orbita $[\boldsymbol{x}]$ v závislosti na tom, jaké koeficienty vektor $\boldsymbol{x}$ obsahuje.

Automorfismy grafů. Zaměřme se však nyní na matematickou strukturu grafu. Každý graf X je tvořen množinou prvků $V(X)$ zvaných vrcholy a binární relací $E(X)$, která popisuje orientované hrany. Tedy $(a,b) \in E(X)$ znamená, že graf obsahuje orientovanou hranu z a do b. V případě neorientovaného grafu je relace $E(X)$ symetrická.

Pro takto definovanou matematickou strukturu je přirozené uvažovat její homomorfismy. Pro dva grafy X a Y se zobrazení $\pi : V(X) \to V(Y)$ nazývá *homomorfismus*, pokud splňuje

$$(a,b) \in E(X) \implies \big(\pi(a), \pi(b)\big) \in E(Y).$$

Tato definice je zcela přirozená a totožná s definicí homomorfismu ostatních matematických struktur uvažovaných v tomto textu. *Automorfismus* grafu X je bijektivní homomorfismus $\pi : V(X) \to V(X)$. Množina všech automorfismů grafu X tvoří grupu $\mathrm{Aut}(X)$, která popisuje strukturu symetrií grafu X; příklady jsou na obrázku 6.17.

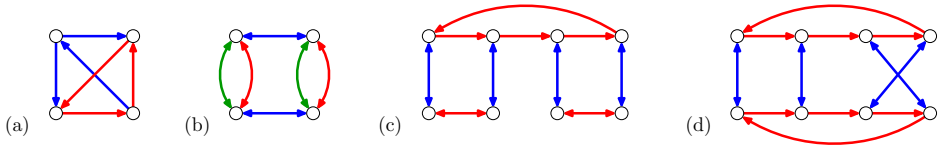


Obrázek 6.17: Několik grafů a jejich grup automorfismů.

Struktura Cayleyho grafu. V textu jsme grupy popisovali pomocí Cayleyho grafů, které jsou zjevně velice symetrické objekty. Co je však na Cayleyho grafech strukturálně tak speciálního? Řekneme, že máme graf s obarvenými orientovanými hranami. Za jakých podmínek je to Cayleyho graf, a tedy popisuje strukturu nějaké grupy?

Ukažme nejprve několik příkladů grafů, které nejsou Cayleyho grafy. Předně přesně jedna šipka každé barvy musí z každého vrcholu vycházet a vcházet; proto graf na obrázku 6.18a. Dále v Cayleyho grafu nemohou být paralelní hrany vedoucí mezi stejnými dvojicemi vrcholů jako na obrázku 6.18b. Důvodem je, že Cayleyho graf je vytvořen podle množiny generátorů grupy.

Zbývající dva příklady jsou složitější. Jak jsme již zmínili, každá levá translace musí být složena z cyklů jedné délky rovné řádu daného generátoru, což neplatí na obrázku 6.18c. Avšak ani toto není postačující, jak je ilustrováno na obrázku 6.18d. To není Cayleyho graf, protože v něm musí být lokální struktura šipek všude stejná.



Obrázek 6.18: Toto nejsou Cayleyho grafy z následujících důvodů: (a) S levým horním vrcholem nesousedí žádná červená hrana a vedou z něj dvě modré hrany. (b) Modré a zelené hrany jsou paralelní, proto je modrý a zelený generátor. (c) Červený generátor není levá translace, neboť obsahuje cyklus délky čtyři a cykly délky dva. (d) Z levé části vyplývá, že $c \circ m = m \circ c$, což neplatí pravé části grafu.

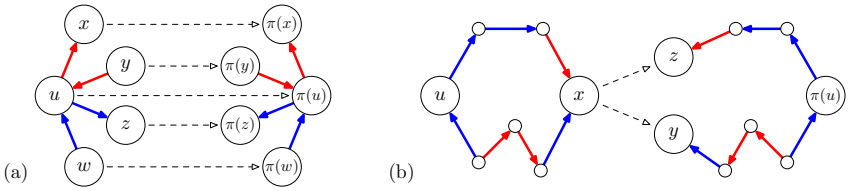
Co je tedy správná charakteristika Cayleyho grafu? Ta je přes jejich grupu automorfismů, která, jak uvidíme, je velice speciální. Protože Cayleyho graf obsahuje hrany různých barev, budeme uvažovat

pouze automorfismy π, které zachovávají barvy hran, tedy hrany (u,v) a $(\pi(u), \pi(v))$ mají vždy stejnou barvu. Nejprve si charakterizujme, jak vypadají grupy automorfismů Cayleyho grafů:

Tvrzení 6.4. *Nechť X je Cayleyho graf reprezentující grupu $\mathbb{G}$. Akce $\mathrm{Aut}(X)$ je regulární a platí, že*

$$\mathrm{Aut}(X) \cong \mathbb{G}.$$

Důkaz. Nejprve pochopíme, jak strukturálně vypadá $\mathrm{Aut}(X)$. Ukážeme, že akce této grupy na Cayleyho grafu X je regulární. To znamená, že pro každé dva vrcholy $u, v \in X$ existuje právě jeden automorfismus π, pro který platí $\pi(u) = v$. Tento automorfismus musí zobrazit sousedy u na sousedy v. Protože π zachovává barvy a směry hran, je pro každého souseda u jednoznačně určen jeho obraz, jak je naznačeno na obrázku 6.19a. Podobně obrazy sousedů sousedů u jsou jednoznačně určeny, a tak dál. Protože je graf X souvislý, je π jednoznačně určené na celém grafu.



Obrázek 6.19: (a) Automorfismus π, který zobrazuje u na v, je jednoznačně určený na sousedech u, podle barev a směru šipek vycházejících z u. (b) Jednoznačnost procesu, který definuje π. Pokud dvě posloupnosti šipek vedou z u do x, potom musejí odpovídat stejným prvkům v grupě. Označíme-li m modrý a c červený prvek, potom v grupě platí $m \circ c^2 \circ m^{-1} = c \circ m^2$.

Zatím víme, že existuje nejvýše jeden automorfismus zobrazující u na v, který musí být určený výše popsaným způsobem. Musíme však ukázat, že automorfismus π skutečně existuje, tedy že tento proces nepřiřadí jednomu vrcholu $x \in X$ dva různé obrazy $y \in X$ a $z \in X$. Připomeňme, jak proces konstrukce π funguje. Kdykoliv určuje obraz $\pi(w)$ nějakého vrcholu $w \in X$, je tento obraz určen podle posloupnosti barevných šipek vycházející v u a končící ve w. Jako $\pi(w)$ je zvolen vrchol, který je na konci stejné posloupnosti šipek vycházejících z $\pi(u)$.

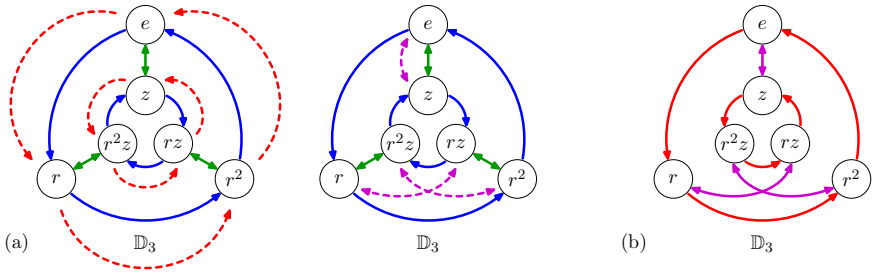
Předpokládejme, že dvě posloupnosti barevných šipek vedou z u do x, jak na obrázku 6.19b, a tyto posloupnosti definují $\pi(x) = y$ a $\pi(x) = z$. Chceme ukázat, že definice π je korektní, tedy že $y = z$. Nechť H je generující množina, pomocí kterého je Cayleyho graf X vytvořen. Tyto dvě posloupnosti šipek odpovídají násobení následujícími prvky:

$$\begin{aligned} &\text{Posloupnost definující } \pi(x) = y: && h = h_k \circ h_{k-1} \circ \cdots \circ h_1, \quad \text{kde } h_i \in H \cup H^{-1}. \\ &\text{Posloupnost definující } \pi(x) = z: && h' = h'_\ell \circ h'_{\ell-1} \circ \cdots \circ h'_1, \quad \text{kde } h'_i \in H \cup H^{-1}. \end{aligned}$$

Protože obě posloupnosti vedou z u do x, platí $h \circ u = h' \circ u$. V grupě funguje krácení, tudíž $h = h'$. Protože však $y = h \circ \pi(u)$ a $z = h' \circ \pi(u)$, musí platit, že $y = z$. Tedy automorfismus π skutečně existuje a akce $\mathrm{Aut}(X)$ je regulární.

Zbývá ukázat, že i struktura $\mathrm{Aut}(X)$ je totožná se strukturou $\mathbb{G}$. První nápad je, že každý z těchto automorfismů bude odpovídá vynásobení nějakým prvkem h zleva. Pokud π zobrazuje u na v, potom by muselo platit $h = vu^{-1}$. Toto však nefunguje pro libovolnou nekomutativní grupu, příklad je na obrázku 6.20a.

Mějme automorfismus π zobrazující u na v. Již víme, že $\pi(x)$ je prvek, který dostaneme posunutím z prvku v o libovolnou posloupnost šipek, která vede z u do x. Každá taková posloupnost šipek odpovídá



Obrázek 6.20: (a) Akce dvou automorfismů Cayleyho grafu grupy $\mathbb{D}_3$. Červený automorfismus zobrazuje e na r, fialový e na z. Tyto automorfismy neodpovídají násobení prvkem grupy zleva. (b) Cayleyho graf $\mathbb{D}_3$ s pravými translacemi prvky r a z. Automorfismy Cayleyho grafu jsou tvořené pravými translacemi.

xu^{-1}, a tedy platí $\pi(x) = (xu^{-1}) \circ v$. Z asociativity dostaneme automorfismus π jako pravou translaci prvkem $u^{-1}v$, tedy $\pi : x \mapsto x \circ (u^{-1}v)$. Pravé translace $\mathbb{D}_3$ jsou na obrázku 6.20b. Ty nejsou narozdíl od levých translací generátorů vyobrazeny v Cayleyho grafech.

Ukázali jsme, že jednotlivé automorfismy $\mathrm{Aut}(X)$ Cayleyho grafu X můžeme identifikovat s pravými translacemi v grupě $\mathbb{G}$. Abychom ukázali, že je to izomorfismus, musíme ukázat, že skládání automorfismů odpovídá násobení v grupě G. Mějme automorfismy π odpovídající pravé translaci $u^{-1}v$ a σ odpovídající $v^{-1}w$. Složení těchto pravých translací $(u^{-1}v) \circ (v^{-1}w) = u^{-1}w$ přesně odpovídá automorfismu $\sigma \circ \pi$ zobrazujícímu u na w. Protože identitě přiřazujeme identický automorfismus, je nalezené přiřazení skutečně izomorfismus mezi $\mathrm{Aut}(X)$ a $\mathbb{G}$. □

Právě regularita grupy automorfismů je charakterizující vlastnost.

Tvrzení 6.5. *Graf X je Cayleyho graf, právě když je souvislý, bez multihran, z každého vrcholu vychází a do každého vrcholu vchází přesně jedna hrana každé barvy a $\mathrm{Aut}(X)$ je regulární.*

Důkaz. Pokud je X Cayleyho graf, všechny tyto vlastnosti platí, poslední podle tvrzení 6.4. Zbývá dokázat obrácenou implikaci. Mějme $\mathrm{Aut}(X)$ regulární a tvrdíme, že X je Cayleyho graf reprezentující grupu $\mathrm{Aut}(X)$. Pro libovolné dva vrcholy u a v existuje automorfismus π, který zobrazuje u na v. Tento automorfismus musí být ze souvislosti grafu přesně popsaný jako v tvrzení 6.4. Odpovída tomu, že když z vrcholu u do jiného vrcholu x vede libovolná posloupnost šipek, potom z vrcholu $v = \pi(u)$ do vrcholu $\pi(x)$ vede stejná posloupnost šipek.

Nyní chceme přiřadit jednotlivým vrcholům grafu X prvky grupy $\mathrm{Aut}(X)$. Protože je graf symetrický, zvolíme libovolný vrchol jako neutrální prvek id. Každém vrcholu u přiřadíme, z regularity, jednoznačný automorfismus π_u, který zobrazuje id na u. Podobně jednotlivým barevným hranám můžeme přiřadit prvky grupy, jestliže vede hrana z u do v, odpovídá tato barva prvku $\pi_v \circ \pi_u^{-1}$. Z regularity vyplývá, že takto definované přiřazení je skutečně jednoznačné, tedy nedojdeme k rozporu, že by jedna barevná hrana odpovídala dvěma generátorům. Je jednoduché ověřit, že graf X je skutečně Cayleyho graf pro grupu $\mathrm{Aut}(X)$, protože složení hrany z u do v a hrany z v do w odpovídá násobení generátorů, neboť

$$\pi_w \circ \pi_v^{-1} \circ \pi_v \circ \pi_u^{-1} = \pi_w \circ \pi_u^{-1}.$$

□

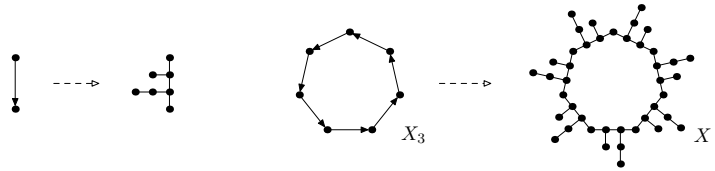
Poznamejme, že předpoklad souvislosti je nadbytečný, neboť vyplývá z regularity $\mathrm{Aut}(X)$. Čtenář si může rozmyslet, proč tomu tak je.

Fruchtova věta. Ukázali jsme, že grupy automorfismů barevných orientovaných grafů umožňují realizovat libovolnou abstraktní konečnou grupu. Předchozí konstrukci lze snadno upravit i na nekonečné grupy pomocí nekonečných grafů. Zaměřme se však na neorientované grafy bez barevných hran. Fruchtova věta říká, že i grupy automorfismů těchto grafů jsou univerzální, tedy libovolná konečná abstraktní grupa je izomorfní grupě automorfismů nějakého neorientovaného grafu.

Věta 6.6 (Frucht)**.** *Pro libovolnou abstraktní konečnou grupu $\mathbb{G}$ existuje neorientovaný graf X, že*

$$\mathrm{Aut}(X) \cong \mathbb{G}.$$

Důkaz. Ukažme nejprve konstrukci neorientovaného grafu X, pro který je $\mathrm{Aut}(X) \cong \mathbb{Z}_7$. Tento graf sestrojíme z grafu X_3 z obrázku 6.17. Každou orientovanou hranu nahradíme malým *udělátkem*, které je asymetrický neorientovaný podgrafem. Konstrukce je vysvětlena na obrázku 6.21. Tento postup můžeme aplikovat na libovolný orientovaný graf X a zachovat jeho grupu automorfismů, avšak v udělátkách prodloužíme další navěšenou cestu na dostatečnou délku, aby se v grafu X takové udělátko předtím nevyskytovalo.



Obrázek 6.21: Každou orientovanou hranu nahradíme udělátkem. To obsahuje dvě navěšený cesty, délky jedna a dva. Proto neexistuje žádná symetrie, která by udělátko zrcadlila. Každý automorfismus prohazuje udělátka nějakým způsobem. V získaném grafu X je možné je rotovat pouze jedním směrem, a proto je $\mathrm{Aut}(X) \cong \mathbb{Z}_7$.

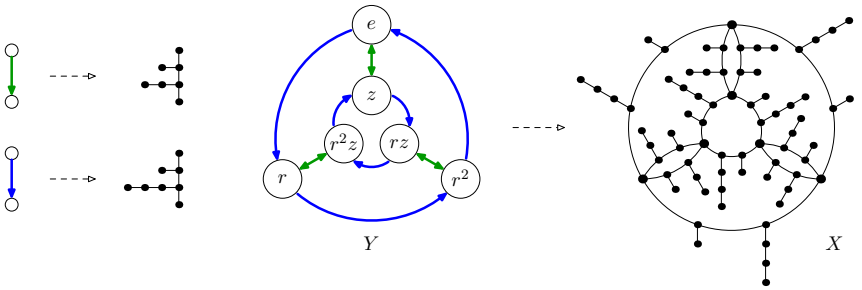
K důkazu věty pro libovolnou grupu $\mathbb{G}$ použijeme Cayleyho graf Y této grupy. Podle tvrzení 6.4 víme, že $\mathrm{Aut}(Y) \cong \mathbb{G}$. Modifikací Y vyrobíme neorientovaný jednobarevný graf X se stejnou grupou automorfismů. Myšlenka je, že každou orientovanou hranu nahradíme udělátkem, podobně jako předtím. Avšak tímto způsobem bychom ztratili informace o barvách, a tím pravděpodobně zvětšili grupu automorfismů. Proto hrany jednotlivých barev nahradíme různými udělátkami, s různě dlouhými délkami dalších cest. Každý automorfismus prohazuje pouze udělátka stejného typu, a proto $\mathrm{Aut}(X) \cong \mathrm{Aut}(Y)$. Obrázek 6.22 ukazuje příklady této konstrukce. □

Cvičení

6.1 Dokažte, že následující množiny permutací jsou do inkluze minimální generující množiny grupy $\mathbb{S}_n$:

(a) $\{\tau_{1,2}, \tau_{1,3}, \ldots, \tau_{1,n}\}$.
(b) $\{\tau_{1,2}, \tau_{2,3}, \ldots, \tau_{n-1,n}\}$.
(c) $\{\pi = (2, 3, 4, \ldots, n, 1), \tau_{i,i+1}\}$.

⋆ 6.2 Zesílíme předcházející cvičení 6.1, části a až b. Nechť T je množina transpozic $\mathbb{S}_n$. Sestrojme graf X, kde $ij \in E(X)$, právě když $\tau_{i,j} \in T$. Jaké grafy X dostáváme pro množiny T ze cvičení 6.1 v části a až b? Dokažte, že množina T generuje $\mathbb{S}_n$, právě když X je souvislý. Dokažte, že T je do inkluze minimální generující množina $\langle T \rangle$, právě když X neobsahuje žádný cyklus. Jak obecně vypadá $\langle T \rangle$?



Obrázek 6.22: Konstrukce neorientovaného grafu X, jehož grupa automorfismů je stejná jako Cayleyho grafu Y, tedy $\mathrm{Aut}(X) \cong \mathbb{D}_3$. Původní vrcholy grafu Y jsou zvýrazněny. Pochopitelně existují mnohem jednodušší grafy X s touto grupou automorfismů (třeba trojúhelník), ale tato konstrukce funguje obecně pro libovolnou grupu $\mathbb{G}$.

6.3 Dokažte, že všechny podgrupy libovolné grupy tvoří vzhledem k inkluzi úplný svaz. Nechť $\mathcal{H}$ je množina podgrup, potom operace infimum a supremum jsou definované následovně:

$$\inf(\mathcal{H}) = \bigcap_{H\in\mathcal{H}} H \qquad \text{a} \qquad \sup(\mathcal{H}) = \Big\langle \bigcup_{H\in\mathcal{H}} H \Big\rangle.$$

6.4 Definujme *centrum grupy* jako množinu prvků grupy, které komutují s každým dalším prvkem. Tedy jsou to přesně ty prvky, které mají stejné levé a pravé translace. Dokažte, že pro libovolnou grupu tvoří její centrum podgrupu. V jakém vztahu jsou centrum a Abelovské grupy? Jaký je vztah se cvičením 3.4?

6.5 V tomto cvičení zkusíme vymyslet definici homomorfismu obecné matematické struktury. V tomto textu jsme například uvažovali tyto struktury:

- Vektorový prostor je množina spolu s binární operací $+$ a unárními operacemi násobení skalárem α, pro každý skalár jedna operace.
- Grupa je množina spolu s binární operací $\circ$ a nulární operací neutrálního prvku e; takto lze definovat konstantu.
- Graf je množina vrcholů spolu s binární relací E, což je množina orientovaných hran.
- Barevný graf je množina vrcholů spolu s binárními relacemi $E_1, \ldots, E_k$, což jsou množiny orientovaných hran jednotlivých barev.

Jaké podmínky musel homomorfismus mezi těmito strukturami splňovat.

Nechť X a Y jsou dvě množiny opatřené operacemi a relacemi stejného typu: operacemi $\circ_1, \ldots, \circ_k$ a relacemi $R_1, \ldots, R_\ell$. Zobecněte výše uvedené definice pro obecné matematické struktury. Jaké vlastnosti musí splňovat zobrazení $f : X \to Y$, aby se nazývalo homomorfismus?

⋆ 6.6 Nechť H je do inkluze minimální generující množina grupy $\mathbb{G}$ řádu n. Dokažte, že H je nejvýše velikosti $\lceil \log_2 n \rceil$. Jak velký graf X dostaneme pomocí konstrukce z důkazu Fruchtovy věty 6.6, v závislosti na řádu n grupy $\mathbb{G}$? Zkuste navrhnout efektivnější konstrukci, vytvářející menší graf X. Zkuste také konstrukci modifikovat, aby každý vrchol měl nejvýše tři sousedy.

Příloha A

Nápovědy ke cvičení

Následují nápovědy k vybraným cvičením.

Kapitola 4: Lineární nezávislost, kombinace a báze

4.7 Pro hodnost platí obecně následující dolní odhady. Ten první není příliš známý, ten druhý dokázal poprvé Sylvester:

$$\begin{aligned} \mathrm{rank}(A+B) &\geq \big|\mathrm{rank}(A) - \mathrm{rank}(B)\big|, \\ \mathrm{rank}(AB) &\geq \mathrm{rank}(A) + \mathrm{rank}(B) - n, \qquad \text{kde } A \in \mathbb{R}^{m\times n} \text{ a } B \in \mathbb{R}^{n\times p}. \end{aligned}$$