Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

DIPLOMOVÁ PRÁCE

Doležel Pavel

Dopravní problém, jeho zobecnění a aplikace v pravděpodobnosti a statistice

Katedra teorie pravděpodobnosti a matematické statistiky
Vedoucí diplomové práce: ***Prof. RNDr. Jitka Dupačová, DrSc.***
Studijní program: ***Matematika, Ekonometrie***

Rád bych poděkoval vedoucí diplomové práce, paní profesorce Jitce Dupačové, za poskytnutí literatury, vedení práce, její kontrolu a motivující přístup.

Prohlašuji, že jsem svou diplomovou práci napsal samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce.

V Praze dne 5. srpna 2009 Pavel Doležel

Obsah

Název práce: Dopravní problém, jeho zobecnění a aplikace v pravděpodobnosti a statistice
Autor: Pavel Doležel
Katedra (ústav): Katedra teorie pravděpodobnosti a matematické statistiky
Vedoucí diplomové práce: Prof. RNDr. Jitka Dupačová, DrSc.
e-mail vedoucího: Jitka.Dupacova@mff.cuni.cz
Abstrakt:

V práci se autor zabývá specifickou optimalizační úlohou, tzv. dopravním problémem a jeho řešením. Uvádí různé metody řešení dopravního problému a jeho aplikace v teorii pravděpodobnosti a matematické statistice, zejména na statistické třídění v L1 normě a rekonstrukci kontingenčních tabulek. Zvláštní místo je věnováno různým modifikacím klasického dopravního problému, především vícerozměrnému dopravnímu problému. Podstatnou částí práce jsou vybrané aplikace dopravního problému na řešení konkrétních úloh a uvedení některých algoritmů, které se k řešení používají.
Klíčová slova:

Dopravní problém, Vícerozměrný dopravní problém, Mongeovy matice, kontingenční tabulky

Title: Transportation problem, it's generalisations and aplications in probability and statistics
Author: Pavel Doležel
Department: Department of the Theory of Probability and Mathematical Statistics
Supervisor: Prof. RNDr. Jitka Dupačová, DrSc.
Supervisor's e-mail address: Jitka.Dupacova@mff.cuni.cz
Abstract: Author describes a specific optimization problem-the transportation problem and analyzes relevant solution methods. Several methods of solving the transportation problem are listed, applied or introduced and applications of the transportation poblem in the theory of probability and mathematical statistics are presented, namely the statistical sorting in L1 norm and re-construction of contingency tables. Special interest is devoted to several modifications of ordinary transportation problem, mainly the multi-index transportation problem. The crucial part of the work are selected applications of the transportation problem to particular problems and showing some algorithms used for finding solutions.
Keywords:

Transportation problem, Multi-index transportation problem, Monge matrices, contingency tables

Kapitola 1

Úvod

Dopravní problém je název pro úlohu, která byla motivována minimalizací nákladů na dopravu zboží od dodavatelů k odběratelům. Může jít jak o zboží libovolně dělitelné, tak o zboží, které je možné převážet, poptávat nebo dodávat pouze v celých kusech, přičemž tyto dvě úlohy jsou dosti odlišné. První je předmětem studia lineárního programování, druhá programování celočíselného. Celočíselným programováním se v práci zabývat nebudu. Zobecněním klasického dopravního problému je Monge-Kantorovičův dopravní problém. Jde v něm o nalezení infima reálného funkcionálu na množině všech pravděpodobnostních měr (definovaných na kartézském součinu $\mathcal{S} \times \mathcal{S}$, kde $\mathcal{S}$ je separabilní metrický prostor) s danými fixními marginály, viz. definice na straně 1 v [35]. Pokud jsou tyto míry diskrétní (atomické) a nabývají kladných hodnot pouze na konečné množině, jde o diskrétní dopravní problém. Monge-Kantorovičův dopravní problém je velmi podrobně popsán v knihách [35], [36] a [47] a zde jej zmiňuji pouze aby bylo zřejmé, proč v definicích dopravních problémů explicitně uvádím, že jde o problémy diskrétní.

V práci se budu detailně zabývat pouze klasickým dvourozměrným diskrétním dopravním problémem, vícerozměrným diskrétním dopravním problémem a některými jejich modifikacemi, jako je kupříkladu diskrétní dopravní problém s kapacitními omezeními (aplikace na statistické třídění v L1 normě), či diskrétní dopravní problém s nelineární separovatelnou účelovou funkcí (aplikace na rekonstrukce kontingenčních tabulek). Zmíním i některé další typy problémů a modifikací. Těmi se ale z důvodu rozsahu práce nemohu zabývat hlouběji a tak u nich pouze odkáži na relevantní literaturu. Soustředím se na aplikace v teorii pravděpodobnosti a matematické statistice, byť jsem si vědom, že aplikace dopravního problému sahají mnohem dále do mnoha jiných oblastí matematiky či teoretické informatiky. Jak je uvedeno v různé literatuře, zejména v knihách [37], [35] a [36], Monge-Kantorovičův dopravní problém má aplikace v řadě oblastí matematiky od funkcionální analýzy, přes ekonometrii, teorii pravděpodobnosti až po matematickou ekonomii a operační výzkum. Budu v práci předpokládat základní znalosti čtenáře v teorii pravděpodobnosti, v matematické statistice a lineárním programování.

Rozsah známé teorie v oblasti řešení dopravního problému a jeho zobecnění je značný a proto se nebudu pokoušet o podání úplného přehledu a soustředím se na způsoby řešení a aplikace podle mého uvážení úzce související s teorií pravděpodobnosti a statistikou. V práci budu vycházet zejména z knih [1], [35], [36], [32], [21], [37] a [24]. Aplikace ve statistice jsou k nalezení v knihách [21], [1], [35], [36] a [47] a v mnoha desítkách, možná i stovkách článků, z nichž některé jsou v práci citovány. V knihách [26] a [32] je zohledněn přístup kombinatorické optimalizace, teorie grafů, toků v sítích a teorie matroidů[1]. Jak se později ukáže, je vhodné uvažovat zvlášť přípustné a zvlášť optimální řešení dopravního problému, přičemž za jistých okolností je konkrétní metodou nalezené přípustné řešení zároveň řešením optimálním (viz. například [4],[14]). Řešení dvourozměrného Monge-Kantorovičova dopravního problému pro některé speciální typy nákladových funkcí (ryze konkávní či ryze konvexní funkce vzdálenosti) jsou explicitně uvedena v článku [19]. V knize [35] je sepsána velká část teorie dvourozměrného Monge-Kantorovičova dopravního problému, včetně teorie duality. Klasickému dvourozměrnému diskrétnímu dopravnímu problému, který je znám ze základních přednášek lineárního programování (viz. například [24]) se budu věnovat poměrně důkladně, byť jde o problém, jehož řešení je velmi dobře prozkoumáno. Nebude to však pouze z didaktických důvodů, ale hlavně z důvodu rozšíření teorie Mongeových matic. Teorie optimalizace a lineárního programování je nejvíce zohledněna v knihách [24] a [37]. V článcích [48], [28], [29], [44] a [42] jsou uvedeny různé nutné podmínky pro existenci přípustného řešení planárního trojrozměrného diskrétního dopravního problému (viz. definice 1). V článku [43] je uveden postup, který vede k nalezení nutných a postačujících podmínek pro existenci přípustného řešení planárního vícerozměrného diskrétního dopravního problému. Článek [13] se zabývá mohutností množiny všech matic s danými řádkovými a sloupcovými součty. V článcích [45], [46], [41], [30] a dalších a v knize [35] lze najít základy teorie vícerozměrného dopravního problému. V článku [22] autor ukazuje, že lze převést libovolný symetrický diskrétní dopravní problém (viz. definice 1) na problém axiální (viz. definice 1). V článku [25] autoři studují vlastnosti množiny všech přípustných řešení axiálního a planárního troj-

[1]Ta má aplikace v oblasti tzv. hladových algoritmů, které vedou k nalezení přípustného řešení některých typů diskrétních dopravních problémů.

rozměrného diskrétního dopravního problému a mimojiné dokazují triviální tvrzení o existenci přípustného řešení axiálního trojrozměrného diskrétního dopravního problému a zavádějí dolní meze pro maximální počet vrcholů (resp. celočíselných vrcholů) polytopu všech přípustných řešení planárního trojrozměrného diskrétního dopravního problému. Zajímavou aplikaci v podobě rekonstrukce kontingenčních tabulek lze najít v článku [18]. Tato aplikace je zajímavá i z teoretického hlediska, neboť zavádí alternativní postup k řešení dvourozměrného diskrétního dopravního problému s nelineární separovatelnou účelovou funkcí a navíc při zobecnění do více rozměrů vede k algoritmu, kterým lze (jak je uvedeno na straně 266 (věta 3.2) v článku [8]) ověřit existenci přípustného řešení planárního vícerozměrného diskrétního dopravního problému.

Práce je strukturována do čtyř kapitol.

Kapitola 1 obsahuje stručný úvod do problematiky teorie a aplikací dopravního problému, jeho modifikací a různých zobecnění a obsahuje základní přehled o nejpoužívanější literatuře.

V kapitole 2 definuji jednotlivé úlohy dopravního problému, jeho modifikací a zobecnění, kterými se v práci zabývám. Je zde uveden i krátký úvod do teorie dvourozměrného diskrétního dopravního problému, krátce popsán vícerozměrný diskrétní dopravní problém a duální úloha k dvourozměrnému dopravnímu problému. Různá další zobecnění a modifikace dopravního problému je možné nalézt zejména v knihách [35], [36] a [47][2].

V kapitole 3 popisuji nejpoužívanější algoritmy pro nalezení přípustného a optimálního řešení diskrétního dopravního problému (Vogelův algoritmus, Dantzigův simplexový algoritmus (SIMPLEX), metodu řádkových a sloupcových čísel (MODI) vycházející z duální úlohy k dopravnímu problému, metodu severozápadního rohu, Berkovu metodu, Ford-Fulkersonův algoritmus a maďarskou metodu). Nesoustředím se však pouze na samotné algoritmy, nýbrž v některých případech i na jejich teoretické aspekty[3].

V kapitole 4 popisuji různé aplikace dopravního problému. K aplikacím patří kupříkladu notoricky známá minimalizace nákladů na dopravu zboží od dodavatelů k odběratelům (odkud získal dopravní problém svůj název), dále pak přidělovací či přiřazovací úlohy, v nichž jde o přiřazení úkolů jednotlivým procesorům (lidem, strojům, apod.) a další úlohy, které lze na přiřazovací převést (plánování rezerv a oprav, plánování výroby a zásob, apod.), nicméně protože se v těchto aplikacích nejedná o přímou aplikaci v teorii pravděpodobnosti či matematické statistice, budu se jimi zabývat pouze

[2]V práci se jimi z důvodu jejího rozsahu nemohu hlouběji zabývat.

[3]Kupříkladu u metody severozápadního rohu zkoumám, za jakých dodatečných podmínek vede k optimálnímu řešení.

okrajově. Jsou probrány detailněji v knihách [24], [37], [32], [49] a [26]. V této kapitole probereme možnost rekonstrukce kontingenční tabulky při znalosti součtů jejích řádků a sloupců a dodatečných informací o její struktuře pomocí řešení dopravního problému, viz. [18]. Dále uvedeme použití v důkazu Strassenovy věty pro omezenou Lipschitzovskou metriku (metrizace slabé topologie), viz. [21], navržení a analýzu statistického experimentu (dvojné třídění), viz. [1] a aplikaci na řízené zaokrouhlování ve statistice. Aplikacím Monge-Kantorovičova dopravního problému v teorii pravděpodobnosti je věnována celá kniha [36].

K práci je přiloženo CD, na nějž se v práci odkazuji. Jsou na něm uloženy popisy vybraných algoritmů ve formě obecně přijímaného způsobu zápisu algoritmů[4] a u většiny z těchto algoritmů je přiložen i spustitelný kód v jazyku VBA. Popis uživatelského prostředí a způsobu použití je uveden u každého algoritmu zvlášť v podobě nápovědy. Kódy je možné spustit přímo v tabulkovém procesoru Excel od firmy Microsoft, pro který jsem se rozhodl ze dvou důvodů. Jednak z důvodu široké kompatibility a jednak proto, že Excel již má zabudováno uživatelské rozhraní, které je vhodné k zapisování vstupů ve formě matic a vektorů a to velmi efektivně a bez znalostí konkrétního programovacího či skriptovacího jazyka.

Důkazy tvrzení jsou v práci uvedeny buď v plném rozsahu, nebo je uveden pouze odkaz na literaturu v níž je možné je nalézt. Do textu jsem zařadil v plném rozsahu pouze krátké důkazy. Důkazy které jsou delší a náročnější, jsem vložil do adresáře ***Důkazy*** na přiložené CD. Všechny v textu i na CD uvedené důkazy jsem vytvářel sám. Žádný z uvedených důkazů tedy není přímo opsán z citované literatury. Pokud byl důkaz v literatuře srozumitelně proveden, nepovažoval jsem za nutné jej do práce opisovat.

Úmluvy

1. Množinu přirozených čísel budeme značit $\mathbb{N}$, množinu celých čísel $\mathbb{Z}$, množinu racionálních čísel $\mathbb{Q}$, množinu reálných čísel $\mathbb{R}$, množinu všech nezáporných reálných čísel $\mathbb{R}_0^+$ a množinu všech kladných reálných čísel $\mathbb{R}^+$. Rozšířenou reálnou osu $(\mathbb{R} \cup \{-\infty\} \cup \{+\infty\})$ s obvyklou topologií budeme značit $\overline{\mathbb{R}}$.

2. Klasický d-rozměrný reálný prostor budeme značit $\mathbb{R}^d$ a d-rozměrný racionální prostor budeme značit $\mathbb{Q}^d$.[5]

[4] Důvod pro tuto formu zápisu je zpřístupnění algoritmů čtenářům bez znalostí konkrétních programovacích jazyků.

[5] Racionálním prostorem zde míníme množinu všech d-rozměrných vektorů racionálních čísel na níž je zavedena euklidova metrika.

3. Binární operací dělení beze zbytku (označme ji $|$) budeme rozumět binární operaci definovanou na množině $\mathbb{Z}$ tak, že pro všechna $a, b \in \mathbb{Z}$, $b \neq 0$ $a|b$ je rovno největšímu celému číslu z takovému, že $b \cdot z \leq a$.

4. Matici typu $m \times n$ budeme značit buď $\mathbf{C} = [c_{ij}]_{i=1,\ldots,m}^{j=1,\ldots,n}$, nebo, bude-li uvedeno, že jde o matici typu $m \times n$, prostě jen $\mathbf{C} = (c_{ij})$, nebo jen tučně $\mathbf{C}$. Obdobné označení budeme někdy používat i pro vícerozměrné tabulky, čili kupříkladu trojrozměrnou tabulku s rozměry m, n a o budeme někdy značit $\mathbf{C}$, přičemž v takovém případě budeme uvádět, že jde o trojrozměrnou tabulku typu $m \times n \times o$. Později v textu budeme obecně n-rozměrné tabulky značit $\mathcal{T}_n$. Tučně budeme značit i vektory, ale ty budeme narozdíl od matic značit malými písmeny.

5. Někdy budeme užívat označení $I = \{1, \ldots, m\}$ a $J = \{1, \ldots, n\}$, kde $m, n \in \mathbb{N}$, aniž bychom explicitně uvedli definice těchto množin.

6. Operaci transponování nějaké matice $\mathbf{A}$ budu značit $\mathbf{A}^T$.

7. Skutečnost, že všechny prvky matice, resp. vícerozměrné tabulky jsou nezáporné, budeme pro matici, resp. tabulku $\mathbf{X}$ značit $\mathbf{X} \geq \mathbf{0}$. Skutečnost, že všechny prvky matice, resp. vícerozměrné tabulky jsou kladné, budeme pro matici, resp. tabulku $\mathbf{X}$ značit $\mathbf{X} > \mathbf{0}$.[6]

8. Minimalizaci přes všechny nezáporné reálné matice $\mathbf{X} = [x_{ij}]_{i=1,\ldots,m}^{j=1,\ldots,n}$ typu $m \times n$ kde $m, n \in \mathbb{N}$ budeme značit $\min_{\mathbf{X} \in \mathbb{R}_{0,+}^{m \times n}}$, nebo $\min_{x_{ij} \geq 0}$, případně $\min_{\mathbf{X} \geq \mathbf{0}}$.

9. Koeficienty v účelové funkci diskrétního dopravního problému budeme nazývat náklady. V případě dvourozměrného diskrétního dopravního problému tyto koeficienty tvoří matici, kterou budeme nazývat nákladovou maticí a v případě vícerozměrného diskrétního dopravního problému tvoří tyto koeficienty vícerozměrnou tabulku. V takovém případě budeme hovořit o tabulce nákladů.

[6]Podobné značení se užívá pro matice pozitivně semidefinitní, či pozitivně definitní. Tyto vlastnosti matic v práci nedefinuji ani nepoužívám.

Kapitola 2

Dopravní problém a jeho modifikace

V této části formálně zavedeme dopravní problém. Nebudu definovat modifikace a zobecnění dopravního problému, kterými se v práci nezabývám, nicméně existuje jich celá řada (dopravní problémy s náhodnými požadavky, dopravní problémy s částečnou znalostí marginálů, dopravní problémy s uvolněnými okrajovými podmínkami, dopravní problémy s lokálními omezeními, spojité dopravní problémy, dopravní problémy s nelineární neseparovatelnou účelovou funkcí, atd.). Uvedu i duální úlohu k dvourozměrnému diskrétnímu dopravnímu problému a to jednak proto, že jedna z metod vedoucích k nalezení optimálního řešení dvourozměrného diskrétního dopravního problému využívá duality a jednak proto, že duální úloha se využívá v aplikaci na třídění v L1 normě.

Definice 1. Nechť $V = \{1, \ldots, n\}$, kde $n \in \mathbb{N}$. Nechť $\mathcal{C}$ je rodina navzájem různých podmnožin V takových, že $\bigcup_{E \in \mathcal{C}} E = V$. Pro každé $v \in V$ nechť $M_v = \{1, \ldots, m_v\}$, $m_v \in \mathbb{N}$ a $m_v \geq 2$. Nechť dále $b_{i_1 \ldots, i_n} \in \mathbb{R}^+$ pro každé $i_1 \in M_1, \ldots, i_n \in M_n$. $V, \mathcal{C}$ a $M_1, \ldots, M_n$ reprezentují **n-rozměrný diskrétní dopravní problém**:

$$\min_{x_{i_1 \ldots i_n} \geq 0} \sum_{i_1=1}^{m_1} \cdots \sum_{i_n=1}^{m_n} c_{i_1 \ldots i_n} x_{i_1 \ldots i_n}, \tag{2.1}$$

kde $i_1 \in M_1, \ldots, i_n \in M_n$ vzhledem k podmínkám

$$\sum_{i_v \in M_v; v \in E} x_{i_1 \ldots i_n} = b_{i(E^C)}, \text{ kde } i\left(E^C\right) = \left\{i_w \in M_w : w \in E^C\right\} \tag{2.2}$$

pro každé $E \in \mathcal{C}$, kde $E^C = V \setminus E$.

Formálně je n-rozměrný diskrétní dopravní problém dán nejen svým rozměrem n, ale i rodinou $\mathcal{C}$ podmnožin V a přirozenými čísly $M_1, \ldots, M_n$. Protože ale na konkrétních hodnotách $M_1, \ldots, M_n$, (jsou-li vesměs rovny alespoň 2) nezáleží a v práci se budu téměř vždy zabývat pouze problémy se speciálními rodinami $\mathcal{C}$, nebudu při označování n-rozměrného diskrétního dopravního problému používat ani parametr $\mathcal{C}$, ani parametry $M_1, \ldots, M_n$. Pokud je $\mathcal{C}$ množinou právě všech k-prvkových podmnožin množiny V, kde $k \in \{1, \ldots, n-1\}$ je pevně dané, pak n-rozměrný diskrétní dopravní problém nazveme **symetrickým** a značíme jej DMKP$_{n,k}$. Problém DMKP$_{n,1}$ nazýváme **planárním n-rozměrným diskrétním dopravním problémem** a problém DMKP$_{n,n-1}$ nazýváme **axiálním n-rozměrným diskrétním dopravním problémem**.

V práci se zabývám téměř výhradně symetrickými diskrétními dopravními problémy. Pokud nebude podstatné, jaký rozměr má daný diskrétní vícerozměrný dopravní problém, budeme jej značit zkráceně (DMKP).

Požadavek na to, aby v definici 1 platilo $\bigcup_{E \in \mathcal{C}} E = V$ není samozřejmě nutný pro nadefinování vícerozměrného diskrétního dopravního problému. I když není splněn, jedná se o vícerozměrný diskrétní dopravní problém, nicméně v takovém případě lze problém redukovat na problém nižší dimenze a tudíž již není možné jej označit za n-rozměrný, viz. Lemma 1 v [22]. Proto tuto podmínku v definici n-rozměrného diskrétního problému uvádím.

Nezáporné řešení soustavy rovnic dané podmínkami (2.2) pro axiální n-rozměrný dopravní problém existuje právě tehdy, když platí

$$\sum b_{i(\{1\})} = \ldots = \sum b_{i(\{n\})}. \tag{2.3}$$

Takovým řešením je kupříkladu řešení dané

$$x_{i_1 \ldots i_n} = \frac{\prod_{k=1}^{n} b_{i(\{k\})}}{\left(\sum b_{i(\{1\})}\right)^{n-1}}. \tag{2.4}$$

Toto řešení je samozřejmě nezáporné, neboť $b_{i_1 \ldots i_n} \in \mathbb{R}^+$ pro všechna $i_1 \in M_1, \ldots, i_n \in M_n$.

Definice 2. Nechť $m, n \in \mathbb{N}$ a nechť $i \in I \stackrel{\text{def}}{=} \{1, \ldots, m\}$ a $j \in J \stackrel{\text{def}}{=} \{1, \ldots, n\}$. Nechť dále $\forall i \in I : a_i \in \mathbb{R}^+$ a $\forall j \in J : b_j \in \mathbb{R}^+$ a $\forall\,(i,j) \in I \times J : c_{ij} \in \mathbb{R}$. **Dvourozměrným diskrétním dopravním problémem** budeme rozumět problém DMKP$_{2,1}$, tj. problém:

$$\min_{\mathbf{x}} \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} \tag{2.5}$$

na množině dané následujícími omezeními:

$$\sum_{j=1}^{n} x_{ij} = a_i \quad \forall i \in I \tag{2.6}$$

$$\sum_{i=1}^{m} x_{ij} = b_j \quad \forall j \in J \tag{2.7}$$

$$x_{ij} \geq 0 \quad \forall (i,j) \in I \times J. \tag{2.8}$$

Diskrétní dvourozměrný dopravní problém budeme značit zkráceně (DTP). Soustava rovnic daná omezeními (2.6) a (2.7) má řešení právě tehdy, když

$$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j. \tag{2.9}$$

Takovým řešením je kupříkladu

$$x_{ij} = \frac{a_i b_j}{\sum_{i=1}^{m} a_i}, \tag{2.10}$$

které je nezáporné.

(DTP) má optimální řešení, právě tehdy, když má řešení přípustné. Množina všech přípustných řešení je totiž vždy omezená. Žádný prvek libovolného přípustného řešení nemůže být větší, než $\sum_{i=1}^{m} a_i < \infty$ a menší než 0.

Úloha dopravního problému je jakožto úloha minimalizace součtu (2.5) vzhledem k omezením daným (2.6), (2.7) a (2.8), úlohou lineárního programování. Podmínka (2.9) je nutná a postačující pro existenci přípustného řešení problému (DTP). Platí-li (2.9), nazýváme dopravní problém vyrovnaným (případně vyváženým)[1].

Obecně je možné uvažovat i problémy, kdy rovnice (2.9) neplatí (nejde o vyrovnaný dopravní problém). V takovém případě je buď

$$\sum_{i=1}^{m} a_i > \sum_{j=1}^{n} b_j,$$

z čehož plyne, že podmínka (2.6) neplatí jako rovnost, ale jako nerovnost

$$\sum_{j=1}^{m} x_{ij} \leq a_i \quad \forall i \in I,$$

[1] My jej budeme nazývat vyrovnaným, neboť pojem vyváženosti použijeme později pro označení jiné vlastnosti modelu dvojného třídění.

nebo

$$\sum_{i=1}^{m} a_i < \sum_{j=1}^{n} b_j,$$

a podmínka (2.7) neplatí jako rovnost, ale jako nerovnost

$$\sum_{i=1}^{n} x_{ij} \leq b_j \quad \forall j \in J.$$

Obě tyto možnosti však lze bez újmy na obecnosti převést na problém vyrovnaný (viz. třeba kapitola 4.9 v [24]) přidáním umělých proměnných. Protože nevyrovnaný dopravní problém je možné bez újmy na obecnosti převést na problém vyrovnaný vždy, nebudeme se nevyrovnaným problémem dále zabývat.

Speciálním typem dopravního problému je tzv. **problém přiřazovací (přidělovací)**. Jde v něm o nalezení v jistém smyslu optimálního přiřazení. Například může jít o přiřazení pracovníků k jednotlivým činnostem podle jejich výkonnosti v těchto činnostech tak, aby celkový výkon byl co nejvyšší, nebo doba potřebná ke zpracování nějaké úlohy byla co nejmenší. Jde o problém minimalizace

$$\min_{\mathbf{X}} \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} \tag{2.11}$$

za podmínek

$$\sum_{i=1}^{m} x_{ij} = 1 \text{ pro všechna } j = 1, \ldots, n, \tag{2.12}$$

$$\sum_{j=1}^{n} x_{ij} = 1 \text{ pro všechna } i = 1, \ldots, m, \tag{2.13}$$

$$x_{ij} \in \{0, 1\} \text{ pro všechna } i = 1, \ldots, m \text{ a } j = 1, \ldots, n. \tag{2.14}$$

Duální úlohou k problému (DTP) je problém

$$\max_{\alpha,\beta} \sum_{i=1}^{m} a_i \alpha_i + \sum_{j=1}^{n} b_j \beta_j \tag{2.15}$$

za podmínek

$$\alpha_i + \beta_j \leq c_{ij} \text{ pro všechna } = 1, \ldots, m \text{ a } j = 1, \ldots, n, \tag{2.16}$$

$$\alpha \in \mathbb{R}^m \text{ a } \beta \in \mathbb{R}^n. \tag{2.17}$$

Definice 3. Dvourozměrným diskrétním dopravním problémem s kapacitními omezeními budeme nazývat takový problém (DTP), že existuje matice $\mathbf{H} \in \mathbb{R}_{+,0}^{m \times n}$ taková, že pro všechna $i \in I$ a $j \in J$ platí dodatečná podmínka pro $\mathbf{X}$:

$$x_{ij} \leq h_{ij}. \tag{2.18}$$

Dvourozměrný diskrétní dopravní problém s kapacitními omezeními je možné snadno zobecnit na vícerozměrný případ, ale toto zobecnění v práci nepoužiji a tak jej explicitně neuvádím.

Definice 4. Nechť $f : \mathbb{R}^{m \times n} \to \mathbb{R}$ je nějaká nelineární funkce. Pak **dvourozměrným diskrétním dopravním problémem s nelineární separovatelnou účelovou funkcí** budeme rozumět problém (DTP), v němž (2.5) nahradíme:

$$\min_{\mathbf{X}} \sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij}\left(x_{ij}\right) \tag{2.19}$$

Kapitola 3

Řešení dopravního problému

Nejprve se v této kapitole budeme zabývat problémem (DTP). Na řešení tohoto problému si ukážeme základní algoritmy, které vedou buď k nalezení přípustného řešení (viz. definice 5), nebo přímo k nalezení optimálního řešení (viz. definice 6), případně vedou k řešení, které je v nějakém smyslu přibližně rovno optimálnímu. V této kapitole se budu rovněž detailněji zabývat Mongeovými maticemi a jejich využitím a provedu rozšíření teorie na Mongeův rozklad. Později se zaměříme na problémy (DMKP).

3.1 Řešení diskrétního dvourozměrného dopravního problému

Uvažujme problém (DTP).

Definice 5. Řekneme, že matice $X = [x_{ij}]$, $i \in I$ a $j \in J$ je **přípustným řešením (DTP)**, pokud platí (2.6), (2.7) a (2.8).

Definice 6. Řekneme, že matice $X = [x_{ij}]$, $i \in I$ a $j \in J$ je **optimálním řešením (DTP)** pokud je přípustným řešením (DTP) a zároveň je dosaženo minima v (2.5).

Poznámka 1. Jde o problém lineárního programování o mn proměnných a $m + n$ omezeních. Omezení jsou soustavou $m + n$ lineárních rovnic, které ale mají vesměs koeficienty nulové, nebo jednotkové. Navíc omezení nejsou lineárně nezávislá a tedy jde fakticky pouze o $m + n - 1$ omezení. Pokud bychom označili m omezení (2.6) za řádková omezení a n omezení (2.7) za sloupcová omezení, pak snadno nahlédneme, že všechna tato omezení nejsou lineárně nezávislá například tím, že od součtu všech řádkových omezení odečteme součet všech sloupcových omezení vyjma jednoho. Toto vynechané omezení je potom rovno výsledku této lineární kombinace ostatních omezení.

3.1.1 Přípustné řešení

Poznámka 2. Ukažme, že problém (DTP) má přípustné řešení právě tehdy, když platí rovnost (2.9). Platí-li (2.9), je přípustným řešením například (2.10), které je skutečně přípustným řešením problému (DTP), neboť je nezáporné. Označme $\mathbf{A}$ matici soustavy dané rovnostmi (2.6) a (2.7) a $\mathbf{b} = (a_1, \ldots, a_m, b_1, \ldots, b_n)^T$. Z definice (DTP) máme, že $a_i > 0$ pro všechna $i = 1, \ldots, m$ a $b_j > 0$ pro všechna $j = 1, \ldots, n$. Soustava rovnic $\mathbf{Ax} = \mathbf{b}$ má obecně řešení právě tehdy, když se hodnost matice $\mathbf{A}$ rovná hodnosti rozšířené matice soustavy $(\mathbf{A}|\mathbf{b})$. Matice $\mathbf{A}$ je typu $(m+n) \times mn$ a jak již víme, její hodnost je $m+n-1$. Neplatí-li tedy (2.9), je

$$\sum_{i=1}^{m} a_i - \sum_{\substack{j=1 \\ j \neq k}}^{n} b_j \neq b_k$$

pro všechna $k = 1, \ldots, n$ a tudíž hodnost rozšířené matice soustavy $(\mathbf{A}|\mathbf{b})$ má hodnost $m+n$ a nikoliv $m+n-1$. Proto soustava $\mathbf{Ax} = \mathbf{b}$ v tomto případě nemá řešení. Základním řešením soustavy rovnic $\mathbf{Ax} = \mathbf{b}$ nazýváme takové řešení $\mathbf{x}$, které obsahuje nejvýše $r(\mathbf{A})$ nenulových členů. Základní přípustné řešení problému (DTP) má nejvýše $m+n-1$ kladných kořenů.

Definice 7. Má-li základní přípustné řešení (DTP) méně než $m+n-1$ kladných kořenů, říkáme, že je problém (DTP) **degenerovaný**.

Metoda severozápadního rohu vede k nalezení přípustného řešení a je nejjednodušší z používaných metod. Na druhou stranu nevede v případě hledání optimálního řešení obecně k příliš dobrým vstupům do simplexové metody, která obecně řeší problém nalezení optimálního řešení.

Metoda severozápadního rohu spočívá v tom, že postupně obsazujeme pole tabulky[1] od levého horního rohu. Přitom vkládáme maximální množství, které je vzhledem k omezením možné. Po každém vložení pokračujeme směrem vpravo, pokud ještě není řádkový součet roven součtu prvků daného řádku anebo směrem dolů, pokud není sloupcový součet roven součtu prvků daného sloupce. Tímto jednoduchým postupem se dostaneme k přípustnému řešení (DTP). Algoritmický zápis metody:

Nechť $a_1, \ldots, a_m$ a $b_1, \ldots, b_n$ jsou kladná reálná čísla a $\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$. Označme dále $\mathbf{X} = (x_{ij})$, $i \in I$ a $j \in J$ matici nezáporných čísel a položme $x_{ij} := 0$ pro všechna $i \in I$ a všechna $j \in J$.

1. $i := 1$ a $j := 1$

[1] reprezentující matici řešení $\mathbf{X}$, která je na počátku algoritmu nulová

2. $x_{ij} := min\left\{a_i - \sum_{k=1}^{j-1} x_{ik}, b_j - \sum_{k=1}^{i-1} x_{kj}\right\}$

3. Pokud $a_i = \sum_{k=1}^{n} x_{ik}$, pak $i := i + 1$. Pokud je $b_j = \sum_{k=1}^{m} x_{kj}$, pak $j := j + 1$.

4. Pokud $i > m$ nebo $j > n$, skončíme s maticí $\mathbf{X}$, která je přípustným řešením (DTP), jinak se vrátíme na krok 2.

Ukažme nyní, že uvedený algoritmus skutečně vede k přípustnému řešení. Vezměmě si i-tý řádek matice $\mathbf{X}$. Vzhledem k tomu, že v kroku 3 se dostaneme k $i + 1$-nímu řádku pouze tehdy, když je již $a_i = \sum_{k=1}^{n} x_{ik}$, což znamená splnění i-tého řádkového omezení, máme za předpokladu, že se vůbec k i-tému řádku dostaneme, splněna všechna řádková a vzhledem k symetrii problému i všechna sloupcová omezení. Ukažme nyní, že k i-tému řádku se dostaneme vždy. K tomu potřebujeme ukázat, že v kroku 3 nastane buď $a_i = \sum_{k=1}^{n} x_{ik}$, nebo $b_j = \sum_{k=1}^{m} x_{kj}$, nebo obojí. To plyne z kroku 2, neboť v tomto kroku vždy doplňujeme buď řádkový součet na hodnotu a_i, nebo sloupcový součet na hodnotu b_j, případně obojí. Jediným způsobem, který by mohl vést k tomu, že se algoritmus k i-tému řádku nedostane, je tedy možnost, že se dříve dostane index j na hodnotu $n + 1$. To by ale znamenalo, že součet všech sloupcových součtů není roven součtu všech řádkových součtů a to je ve sporu s vyrovnaností problému (DTP).

Nyní zavedeme zajímavou a pro další zkoumání užitečnou vlastnost dopravního problému, respektive matice nákladů $\mathbf{C} = [c_{ij}], c_{ij} \geq 0$ pro všechna $1 \leq i \leq m, 1 \leq j \leq n$, která úzce souvisí s metodou severozápadního rohu. Budu dále uvažovat pouze nezápornou matici nákladů, ačkoliv je zřejmé, že uvedené metody řešení dopravního problému nezápornost matice nákladů nevyžadují. Je však možné každý problém, který má záporné sazby nákladů převést na problém s nezápornou maticí nákladů se stejným řešením. Budeme vycházet z článku [4], z něhož vybereme pro nás podstatné informace a doplníme je o některé další, případně odvodíme platnost v tomto článku uvedených tvrzení.

Definice 8. Řekneme, že matice $\mathbf{C} = [c_{ij}]$ splňuje **Mongeovu podmínku**, pokud platí

$$c_{ij} + c_{rs} \leq c_{is} + c_{rj} \quad \forall 1 \leq i < r \leq m, \forall 1 \leq j < s \leq n. \tag{3.1}$$

Řekneme, že matice $\mathbf{C} = [c_{ij}]$ splňuje **inverzní Mongeovu podmínku**, pokud platí

$$c_{ij} + c_{rs} \geq c_{is} + c_{rj} \quad \forall 1 \leq i < r \leq m, \forall 1 \leq j < s \leq n. \tag{3.2}$$

Poznámka 3. Místo abychom říkali, že matice splňuje Mongeovu podmínku, budeme říkat, že matice je Mongeova a místo splnění inverzní Mongeovy podmínky budeme psát o inverzní Mongeově matici.

Lemma 1. "O nejmenším prvku v řádku"
Nechť $\mathbf{C} = [c_{ij}]$ je Mongeova matice typu $m \times n$. Označme $j(i)$ pořadí takového sloupce, že c_{ij} je minimální prvek i-tého řádku matice $\mathbf{C}$ (v případě, kdy je minimálních prvků v i-tém řádku více, označme je $c_{ij_1}, \ldots, c_{ij_k}$, položíme $j(i) := \min\{j_1, \ldots, j_k\}$) pro všechna $1 \leq i \leq m$. Pak platí $j(1) \leq j(2) \leq \ldots \leq j(m)$.

Důkaz lemmatu 1. : Důkaz provedu sporem. Nechť existuje přirozené $1 \leq \bar{i} \leq m$ takové, že $j(\bar{i}) > j(\bar{i}+1)$. Protože $c_{\bar{i},j(\bar{i})} < c_{\bar{i},j(\bar{i}+1)}$ a $c_{\bar{i}+1,j(\bar{i}+1)} < c_{\bar{i}+1,j(\bar{i})}$, platí $c_{\bar{i}+1,j(\bar{i})} + c_{\bar{i},j(\bar{i}+1)} > c_{\bar{i},j(\bar{i})} + c_{\bar{i}+1,j(\bar{i}+1)}$ a matice $\mathbf{C}$ tedy nemůže být Mongeova.

□

Věta 1. "Test Mongeovy podmínky"
Reálná matice $\mathbf{C}$ s vesměs konečnými prvky je Mongeova právě tehdy, když platí

$$c_{ij} + c_{i+1,j+1} \leq c_{i,j+1} + c_{i+1,j} \qquad \forall 1 \leq i < m, \forall 1 \leq j < n.$$

Důkaz věty 1. : Je-li matice $\mathbf{C} \in \mathbb{R}^{m \times n}$ Mongeova, pak tvrzení věty platí triviálně, neboť stačí položit $r := i + 1$ a $s := j + 1$. Dokažme nyní opačnou implikaci. Chceme ověřit, že pro každou čtveřici indexů $1 \leq i < r \leq m$, $1 \leq j < s \leq n$, platí $c_{ij} + c_{rs} \leq c_{is} + c_{rj}$. Pokud platí tvrzení věty pro všechna $i \in \{1, \ldots, m-1\}$ a všechna $j \in \{1, \ldots, n-1\}$, pak platí pro všechna $1 \leq p < r \leq m$ a $1 \leq q < s \leq n$

$$\sum_{u=0}^{r-p-1} \sum_{v=0}^{s-q-1} c_{p+u,q+v} + \sum_{u=1}^{r-p} \sum_{v=1}^{s-q} c_{p+u,q+v} \leq \sum_{u=1}^{r-p} \sum_{v=0}^{s-q-1} c_{p+u,q+v} + \sum_{u=0}^{r-p-1} \sum_{v=1}^{s-q} c_{p+u,q+v},$$

což je ekvivalentní s

$$c_{pq} + c_{rs} \leq c_{ps} + c_{rq}.$$

To je ekvivalentní s definicí Mongeovy matice.

□

Poznámka 4. K tomu, abychom otestovali, zda matice typu $m \times n$ je Mongeova nám tedy podle předchozí věty stačí projít $(m-1)(n-1)$ podmínek a je tedy možné tento test provést v čase $O(mn)$.

Věta 2. "Vlastnosti Mongeovy matice"
Nechť $\mathbf{C}$ a $\mathbf{D}$ jsou dvě Mongeovy matice typu $m \times n$, které mají prvky z množiny $\overline{\mathbb{R}}$ a nechť $\mathbf{u} \in \overline{\mathbb{R}}^m$ a $\mathbf{v} \in \overline{\mathbb{R}}^n$. Potom následující matice jsou rovněž Mongeovy:

- $\mathbf{C}^T$
- $\lambda\mathbf{C}$ pro libovolné $\lambda \in \mathbb{R}_0^+$
- $\mathbf{C} + \mathbf{D}$
- matice $\mathbf{A} = (a_{ij})$ definovaná $a_{ij} := c_{ij} + u_i + v_j$

Důkaz věty 2. : Důkaz je triviální a plyne přímo z definice Mongeovy matice.

Definice 9. Reálnou matici $\mathbf{D} = [d_{ij}]$ typu $m \times n$ nazveme **distribuční maticí**, pokud existuje reálná matice $\mathbf{Q} = [q_{ij}]$ typu $m \times n$ s vesměs nezápornými prvky taková, že pro všechna $1 \leq i \leq m, 1 \leq j \leq n$ platí

$$d_{ij} = \sum_{k=i}^{m} \sum_{l=1}^{j} q_{kl}. \tag{3.3}$$

Poznámka 5. Každá distribuční matice je Mongeova, neboť $d_{i,j} + d_{i,j} + q_{i-1,j} + q_{i,j+1} + q_{i-1,j+1} \geq d_{i,j} + q_{i-1,j} + d_{i,j} + q_{i,j+1}$ což platí proto, že $q_{i-1,j+1} \geq 0$ a to je podle věty 1 ekvivalentní Mongeově podmínce.

Lemma 2. "O distribuční matici"
Matice $\mathbf{C} = [c_{ij}]$ typu $m \times n$ je Mongeova právě tehdy, když existuje distribuční matice $\mathbf{D} = [d_{ij}]$ typu $m \times n$ a dva vektory $\mathbf{u} \in \mathbb{R}^m$ a $\mathbf{v} \in \mathbb{R}^n$ tak, že pro všechna $1 \leq i \leq m, 1 \leq j \leq n$ platí

$$c_{ij} = d_{ij} + u_i + v_j. \tag{3.4}$$

Důkaz lemmatu 2. : Důkaz je na přiloženém CD v adresáři ***Důkazy*** pod názvem **"O distribuční matici".**

Následující věta je vysvětlením důvodu uvedení Mongeovy podmínky v pasáži o metodě severozápadního rohu.

Věta 3. "O Mongeově matici a metodě severozápadního rohu"
Metoda severozápadního rohu vede k optimálnímu řešení vyrovnaného (DTP) pro všechny kladné sloupcové a řádkové součty právě tehdy, když matice nákladů $\mathbf{C} = [c_{ij}]$ splňuje Mongeovu podmínku.

Důkaz věty 3. : Důkaz je na přiloženém CD v adresáři ***Důkazy*** pod názvem **"O Mongeově matici a metodě severozápadního rohu".**

Poznámka 6. Snadno nahlédneme, že Mongeovu matici převedeme na inverzní Mongeovu matici vynásobením všech jejích prvků číslem -1, nebo seřazením jejích sloupců v opačném pořadí.

Poznámka 7. Z věty 2 je patrné, že Mongeovy matice typu $m \times n$ tvoří v prostoru všech matic typu $m \times n$ s prvky z $\mathbb{R}$ konvexní kužel. Označme $\mathbf{H}^{(i)} = \left(h_{pq}^{(i)}\right)$ pro $1 \leq i \leq m$ matice definované $h_{pq}^{(i)} = 1$ pro $p = i$, $h_{pq}^{(i)} = 0$ jinak, $\mathbf{V}^{(j)} = \left(v_{pq}^{(j)}\right)$ pro $1 \leq j \leq n$ matice definované $v_{pq}^{(j)} = 1$ pro $q = j$, $v_{pq}^{(j)} = 0$ jinak, $\mathbf{R}^{(ij)} = \left(r_{pq}^{(ij)}\right)$ pro $1 \leq i \leq m-1, 2 \leq j \leq n$ matice definované $r_{pq}^{(ij)} = 1$ pro $p \leq i$, $q \geq j$ $r_{pq}^{(ij)} = 0$ jinak a $\mathbf{L}^{(ij)} = \left(l_{pq}^{(ij)}\right)$ pro $2 \leq i \leq m, 1 \leq j \leq n-1$ matice definované $l_{pq}^{(ij)} = 1$ pro $p \geq i$, $q \leq j$ $l_{pq}^{(ij)} = 0$ jinak. Z lemmatu 2 víme, že každou Mongeovu matici typu $m \times n$ lze zapsat jako lineární kombinaci matic $\mathbf{H}^{(i)}$, $\mathbf{V}^{(j)}$ a $\mathbf{L}^{(kl)}$, pro $i = 1, \ldots, m$, $j = 1, \ldots, n$, $k = 1, \ldots, m$ a $l = 1, \ldots, n$, přičemž koeficienty této lineární kombinace u matic typu $\mathbf{L}$ jsou vesměs nezáporné. Přidáme-li podmínku nezápornosti, zjistíme, že každou nezápornou Mongeovu matici $\mathbf{C}$ typu $m \times n$ lze zapsat jako nezápornou lineární kombinaci matic $\mathbf{V}^j$, $\mathbf{H}^i$, $\mathbf{R}^{kl}$ a $\mathbf{L}^{uv}$, pro $i = 1, \ldots, m$, $j = 1, \ldots, n$, $k = 1, \ldots, m-1$, $l = 2, \ldots, n$, $u = 2, \ldots, m$ a $v = 1, \ldots, n-1$. Přesněji, existují nezáporná

$$\kappa_1, \ldots, \kappa_m, \mu_1, \ldots, \mu_n,$$

a nezáporná

$$\lambda_{2,1}, \ldots, \lambda_{m,1}, \ldots, \lambda_{2,n-1}, \ldots, \lambda_{m,n-1},$$

$$\nu_{1,2}, \ldots, \nu_{m-1,2}, \ldots, \nu_{1,n}, \ldots, \nu_{m-1,n}$$

taková, že

$$\mathbf{C} = \sum_{i=1}^{m} \kappa_i \mathbf{H}^{(i)} + \sum_{j=1}^{n} \mu_j \mathbf{V}^{(j)} + \sum_{i=2}^{m} \sum_{j=1}^{n-1} \lambda_{ij} \mathbf{R}^{(ij)} + \sum_{i=1}^{m-1} \sum_{j=2}^{n} \nu_{ij} \mathbf{L}^{(ij)}. \tag{3.5}$$

Získat vyjádření (3.5) z lemmatu 2 není nijak obtížné. Stačí si uvědomit několik skutečností:

- Platí

$$\mathbf{R}^{(ij)} = \frac{1}{2} \sum_{k=1}^{i} \mathbf{H}^{(k)} + \frac{1}{2} \sum_{l=j}^{n} \mathbf{V}^{(l)} - \frac{1}{2} \sum_{l=1}^{j-1} \mathbf{V}^{(l)} - \frac{1}{2} \sum_{k=i+1}^{m} \mathbf{H}^{(k)} + \mathbf{L}^{(i+1,j-1)}$$

a tudíž každou matici, která je zapsána ve tvaru (3.5) s nezápornými koeficienty lze zapsat prostřednictvím (3.4)[2].

- Stačí uvažovat matice $\mathbf{R}^{(ij)}$ pouze pro $i = 1, \ldots, m-1$ a $j = 2, \ldots, n$ a matice $\mathbf{L}^{(ij)}$ pouze pro $i = 2, \ldots, m$ a $j = 1, \ldots, n-1$, neboť ve všech ostatních případech je lze vyjádřit prostřednictvím vhodně voleného součtu matic typu $\mathbf{H}$ či vhodně voleného součtu matic typu $\mathbf{V}$.

- Z předchozích bodů je zřejmé, že matice zapsaná ve tvaru (3.5) s nezápornými koeficienty je nezáporná Mongeova. Důkaz, že je možné každou nezápornou Mongeovu matici vyjádřit ve tvaru (3.5) s nezápornými koeficienty je k nahlédnutí v adresáři ***Důkazy*** na přiloženém CD pod názvem ***Mongeova baze***.

Matice typu $m \times n$ je nezáporná Mongeova právě tehdy, když ji lze zapsat ve tvaru (3.5) s vesměs nezápornými koeficienty. Popisem (3.5) evidentně není nezáporná Mongeova matice popsána jednoznačně, neboť kupříkladu matici samých jedniček, která je jistě Mongeova, můžeme dostat jako součet $\sum_{i=1}^{m} \mathbf{H}^{(i)}$, nebo jako součet $\sum_{j=1}^{n} \mathbf{V}^{(j)}$. Množina všech nezáporných Mongeových matic tvoří v prostoru matic typu $m \times n$ konvexní kužel, neboť součet dvou nezáporných Mongeových matic je nezáporná matice (Mongeova je dle věty 2) a nezáporný násobek nezáporné matice je nezápornou maticí (Mongeovou maticí je dle věty 2). Tento kužel je polyedrický, neboť jak jsme uvedli, každý jeho prvek je možné získat jako nezápornou linerání kombinaci nějaké pevně dané konečné množiny matic. Tuto pevně danou množinu matic budu nazývat bazí nezáporných Mongeových matic (viz. [39]) typu $m \times n$.

Definice 10. Označme $\overline{\mathbf{H}}^{(i)} = \left(\overline{h}_{pq}^{(i)}\right)$ pro $1 \leq i \leq m$ matice definované $\overline{h}_{pq}^{(i)} = -1$ pro $p = i$, $\overline{h}_{pq}^{(i)} = 0$ jinak, $\overline{\mathbf{V}}^{(j)} = \left(\overline{v}_{pq}^{(j)}\right)$ pro $1 \leq j \leq n$ matice definované $\overline{v}_{pq}^{(j)} = -1$ pro $q = j$, $\overline{v}_{pq}^{(j)} = 0$ jinak, $\mathbf{W}^{(ij)} = \left(w_{pq}^{(ij)}\right)$ pro $1 \leq i \leq m-1, 1 \leq j \leq n-1$ matice definované $w_{pq}^{(ij)} = 1$ pro $p \leq i$, $q \leq j$ a $w_{pq}^{(ij)} = 0$ jinak a $\mathbf{E}^{(ij)} = \left(e_{pq}^{(ij)}\right)$ pro $2 \leq i \leq m, 2 \leq j \leq n$ matice definované $e_{pq}^{(ij)} = 1$ pro $p \geq i$, $q \geq j$ a $e_{pq}^{(ij)} = 0$ jinak. Každou matici $\mathbf{Z}$ typu $m \times n$, kterou lze zapsat jako jistou lineární kombinaci těchto matic, přesněji matici pro níž existují reálná $\omega_1, \ldots, \omega_m, \tau_1, \ldots, \tau_n$ a nezáporná $\xi_{2,1}, \ldots, \xi_{m,1}, \ldots,$

[2] Z toho plyne, že matice zapsatelné ve tvaru (3.5) s vesměs nezápornými koeficienty κ, μ, λ a ν jsou podmnožinou matic zapsatelných ve tvaru (3.4). Navíc je zřejmé, že matice zapsaná ve tvaru (3.5) s nezápornými koeficienty je nezáporná.

$\xi_{2,n-1}, \dots, \xi_{m,n-1}, \chi_{1,2}, \dots, \chi_{m-1,2}, \dots, \chi_{1,n}, \dots, \chi_{m-1,n}$ tak, že

$$\mathbf{Z} = \sum_{i=1}^{m} \omega_i \overline{\mathbf{H}}^{(i)} + \sum_{j=1}^{n} \tau_j \overline{\mathbf{V}}^{(j)} + \sum_{i=1}^{m-1} \sum_{j=1}^{n-1} \xi_{ij} \mathbf{W}^{(ij)} + \sum_{i=2}^{m} \sum_{j=2}^{n} \chi_{ij} \mathbf{E}^{(ij)}. \tag{3.6}$$

budeme nazývat **zbytkovou maticí**.

Věta 4. "O existenci Mongeova rozkladu"

Každou nezápornou matici typu $m \times n$ lze zapsat jako součet nezáporné Mongeovy matice a zbytkové matice.

Důkaz věty 4. : Důkaz je na přiloženém CD v adresáři ***Důkazy*** pod názvem **"O existenci Mongeova rozkladu".**

Lemma 3. "Vlastnosti zbytkových matic"
Nechť $\mathbf{C}$ a $\mathbf{D}$ jsou dvě zbytkové matice typu $m \times n$, které mají prvky z množiny $\overline{\mathbb{R}}$ a nechť $\mathbf{u} \in \overline{\mathbb{R}}^m$ a $\mathbf{v} \in \overline{\mathbb{R}}^n$. Potom následující matice jsou rovněž zbytkové:

- $\mathbf{C}^T$
- $\lambda\mathbf{C}$ pro libovolné $\lambda \in \mathbb{R}_0^+$
- $\mathbf{C} + \mathbf{D}$
- matice $\mathbf{A} = (a_{ij})$ definovaná $a_{ij} := c_{ij} + u_i + v_j$

Důkaz lemmatu 3. : Důkaz je na přiloženém CD v adresáři ***Důkazy*** pod názvem **"Vlastnosti zbytkových matic".**

Zbytková matice nemusí mít všechny prvky nezáporné, nicméně pokud ji použijeme jako nákladovou matici v (DTP), vede ke stejnému optimálnímu řešení, jako vhodně vybraná matice nákladů s vesměs nezápornými členy. Tuto matici zvolíme velmi snadno a to využitím následujícího lemmatu, v němž položíme $\lambda := \max\{0, -\min_{i\in I, j\in J} c_{ij}\} + 1$.

Lemma 4. "O nezápornosti"
(DTP) daný v (2.5) s maticí nákladů $C \in \mathbb{R}^{m\times n}$ má stejné optimální řešení, jako (DTP) s maticí nákladů $C + \lambda \cdot \mathbb{E}$, kde $\lambda \in \mathbb{R}$ a $\mathbb{E} = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{m\times n}$.

Důkaz lemmatu 4. : Důkaz je triviální.

Lemma 5. "1. lemma o inverzní Mongeově matici"
Každá zbytková matice $\mathbf{Z}$ je inverzní Mongeovou maticí.

Důkaz lemmatu 5. : Důkaz je na přiloženém CD v adresáři ***Důkazy*** pod názvem **"1. lemma o inverzní Mongeově matici".**

Lemma 6. "2. lemma o inverzní Mongeově matici"
Nechť $\mathbf{C}$ je inverzní Mongeova matice. Pak $\mathbf{C}$ je možné zapsat ve tvaru (3.6), tj. $\mathbf{C}$ je zbytková.

Důkaz lemmatu 6. : Důkaz je na přiloženém CD v adresáři ***Důkazy*** pod názvem **"2. lemma o inverzní Mongeově matici".**

Lemma 7. "O konvexní kombinaci"
Nechť $\mathbf{A}$ je matice typu $m \times n$ a nechť existují různé Mongeovy matice $\mathbf{C}$ a $\mathbf{C}'$ a dvě zbytkové matice $\mathbf{Z}$ a $\mathbf{Z}'$ takové, že platí $\mathbf{A} = \mathbf{C} + \mathbf{Z}$ a zároveň $\mathbf{A} = \mathbf{C}' + \mathbf{Z}'$. Pak pro každé $\lambda \in [0, 1]$ platí

$$\mathbf{A} = \mathbf{C}\,(1 - \lambda) + \mathbf{C}'\lambda + \mathbf{Z}\,(1 - \lambda) + \mathbf{Z}'\lambda,$$

kde $\mathbf{Z}\,(1 - \lambda) + \mathbf{Z}'\lambda$ je zbytková matice a $\mathbf{C}\,(1 - \lambda) + \mathbf{C}'\lambda$ je nezáporná Mongeova matice.

Důkaz lemmatu 7. : Matice $\mathbf{C}$ a $\mathbf{C}'$ jsou Mongeovy a tudíž jejich nezáporné násobky jsou rovněž Mongeovy. Jelikož součet dvou Mongeových matic je opět Mongeova matice, je $\mathbf{C}\,(1 - \lambda) + \mathbf{C}'\lambda$ Mongeova matice pro všechna $\lambda \in [0, 1]$. Matice $\mathbf{Z}$ a $\mathbf{Z}'$ jsou zbytkové a tudíž jejich nezáporné násobky jsou rovněž zbytkové matice. Jelikož součet dvou zbytkových matic je opět zbytková matice, je $\mathbf{Z}\,(1 - \lambda) + \mathbf{Z}'\lambda$ zbytková matice.

□

Poznámka 8. Nechť $\mathbf{A}$ je matice typu $m \times n$ a nechť $\mathbf{A} = \mathbf{C} + \mathbf{Z}$, kde $\mathbf{C}$ je Mongeova matice a $\mathbf{Z}$ je zbytková matice. Pak pro každou Mongeovu matici $\mathbf{Q}$ platí, že $\mathbf{C} + \mathbf{Q}$ je Mongeova a $\mathbf{Z} - \mathbf{Q}$ je zbytková. Přitom samozřejmě $\mathbf{A} = \mathbf{C} + \mathbf{Q} + \mathbf{Z} - \mathbf{Q}$.

Poznámka 9. Jak víme z důkazu věty 4, každou matici $\mathbf{A}$ typu $m \times n$ je možné zapsat jako součet nzáporné Mongeovy matice a zbytkové matice. To znamená, že pro každou matici $\mathbf{A}$ existuje nezáporná Mongeova matice $\mathbf{C}$ a existuje zbytková matice $\mathbf{Z}$ takové, že $\mathbf{A} = \mathbf{C} + \mathbf{Z}$. Takových rozkladů na součet Mongeovy matice a zbytkové matice existuje více, kdykoliv $m \geq 2$ a $n \geq 2$.

Definice 11. Libovolný rozklad matice $\mathbf{A}$ na součet Mongeovy matice a zbytkové matice budu nazývat **Mongeovým rozkladem matice A**. Mongeovu matici v Mongeově rozkladu matice $\mathbf{A}$ budu nazývat **Mongeovou částí matice A** a množinu všech Mongeových částí matice $\mathbf{A}$ budu značit $\pi^M(\mathbf{A})$. K ní příslušnou zbytkovou matici budu nazývat **zbytkovou částí matice A** a množinu všech zbytkových částí matice $\mathbf{A}$ budu značit $\pi^Z(\mathbf{A})$.

Z věty 4 víme, že každou matici $\mathbf{A}$ typu $m \times n$ lze zapsat ve tvaru

$$\forall 1 \leq p \leq m,\ 1 \leq q \leq n: \quad a_{pq} = \mu_q + \kappa_p + \sum_{k=2}^{p} \sum_{l=q}^{n-1} \nu_{kl} + \sum_{k=p}^{m-1} \sum_{l=2}^{q} \lambda_{kl}, \tag{3.7}$$

kde $\mu_j \in \mathbb{R}$ pro všechna $j = 1, \ldots, n$, $\kappa_i \in \mathbb{R}$ pro všechna $i = 1, \ldots, m$, $\mu_{ij} \in \mathbb{R}$ pro všechna $i = 2, \ldots, m$ a $j = 1, \ldots, n-1$ a $\lambda_{ij} \in \mathbb{R}$ pro všechna $i = 1, \ldots, m-1$ a $j = 2, \ldots, n$. Dále víme, že ke každé matici existuje nejen rozklad na Mongeovu a zbytkovou matici, ale dokonce na nezápornou Mongeovu matici a zbytkovou matici.

K praktickému použití Mongeova rozkladu musíme být schopni k dané matici $\mathbf{A}$ najít nějakou její Mongeovu a zbytkovou část.

Mějme matici nákladů $\mathbf{C}$ typu $m \times n$. Tu lze zapsat prostřednictvím (3.7). Mongeovu část matice $\mathbf{C}$ z množiny $\pi^M(\mathbf{C})$ najdeme následujícím postupem: Nejprve vyřešíme minimalizační úlohu kvadratického programování

$$\min_{\substack{\kappa \in \mathbb{R}^m,\, \mu \in \mathbb{R}^n, \\ \lambda \in \mathbb{R}^{(m-1)\times(n-1)}, \\ \nu \in \mathbb{R}^{(m-1)\times(n-1)}}} \sum_{k=1}^{m} \sum_{l=1}^{n} \left(c_{kl} - \kappa_k - \mu_l - \sum_{i=2}^{k} \sum_{j=l}^{n-1} \nu_{ij} - \sum_{i=k}^{m-1} \sum_{j=2}^{l} \lambda_{ij} \right)^2, \tag{3.8}$$

Řešení této úlohy najdeme vyřešením následující soustavy lineárních rovnic pro vektory parametrů $\kappa^*, \mu^*, \lambda^*, \nu^*$, jejichž pravá strana vznikla derivováním účelové funkce v (3.8) podle těchto parametrů.

$$0=\sum_{i=1}^{m}\left(c_{iq}-\mu_q^*-\kappa_i^*-\sum_{k=2}^{i}\sum_{j=q}^{n-1}\nu_{kj}^*-\sum_{k=i}^{m-1}\sum_{j=2}^{q}\lambda_{kj}^*\right)$$
$$\text{pro}\quad q=1,\ldots,n,$$
$$0=\sum_{j=1}^{n}\left(c_{pj}-\mu_j^*-\kappa_p^*-\sum_{k=j}^{n-1}\sum_{i=2}^{p}\nu_{ik}^*-\sum_{k=2}^{j}\sum_{i=p}^{m-1}\lambda_{ik}^*\right)$$
$$\text{pro}\quad p=1,\ldots,m,$$
$$0=\sum_{i=p}^{m}\sum_{j=1}^{q}\left(c_{ij}-\mu_j^*-\kappa_i^*-\sum_{k=2}^{j}\sum_{l=i}^{m-1}\lambda_{lk}^*-\sum_{k=2}^{i}\sum_{l=j}^{n-1}\nu_{kl}^*\right)$$
$$\text{pro}\quad p=2,\ldots,m,\ q=1,\ldots,n-1,$$
$$0=\sum_{i=1}^{p}\sum_{j=q}^{n}\left(c_{ij}-\mu_j^*-\kappa_i^*-\sum_{k=2}^{j}\sum_{l=i}^{m-1}\lambda_{lk}^*-\sum_{k=2}^{i}\sum_{l=j}^{n-1}\nu_{kl}^*\right)$$
$$\text{pro}\quad p=1,\ldots,m-1,\ q=2,\ldots,n.$$

Jde o soustavu $2mn-n-m+2$ rovnic s $2mn-n-m+2$ proměnných. Označme $\bar{n}:=2mn-m-n+2$. Soustavu lze řešit kupříkladu metodou LUP rozkladu[3]. Soustava vede k nalezení koeficientů, které po dosazení do (3.7) dávají matici $\mathbf{C}$. Pokud $m\geq 2$ a $n\geq 2$, není matice této soustavy regulární, protože zápis (3.7) není jednoznačným zápisem matice, ale nám k nalezení Mongeova rozkladu stačí jedno z možných řešení.

Nalezení Mongeova rozkladu matice typu $m\times n$ je díky LUP rozkladu proveditelné v čase $O\left(\bar{n}^3\right)$ a použitím Coppersmithova-Winogradova algoritmu (viz. [6]) pro násobení matic dokonce v čase $O\left(\bar{n}^{2.376}\right)$. Nějaký Mongeův rozklad k libovolné nezáporné matici tedy jsme schopni nalézt v tomto čase.

Mongeovu část matice $\mathbf{C}$ najdeme po vyřešení soustavy rovnic jednoduše tak, že položíme v (3.5):

- $\kappa_i=\max\left\{\kappa_i^*,0\right\}$, pro $i=1,\ldots,m$,
- $\mu_j=\max\left\{\mu_j^*,0\right\}$, pro $j=1,\ldots,n$,
- $\nu_{ij}=\max\left\{\nu_{ij}^*,0\right\}$, pro $i=1,\ldots,m-1$ a $j=2,\ldots,n$,
- $\lambda_{ij}=\max\left\{\lambda_{ij}^*,0\right\}$, pro $i=2,\ldots,m$ a $j=1,\ldots,n-1$.

[3]Jde o zobecnění klasického LU rozkladu, které je detailně popsáno v knize [7] na stranách 750-756. LUP rozklad má složitost $O\left(\bar{n}^3\right)$.

Zbytkovou část pak najdeme odečtením nalezené Mongeovy části od matice **C**. Algoritmus nalezení Mongeova rozkladu je implementován v excelovském souboru uloženém v adresáři ***Mongeův rozklad*** na přiloženém CD.

Poznámka 10. Optimální řešení (DTP) s maticí nákladů, která je zbytkovou maticí, najdeme pomocí metody jihozápadního rohu, která je jednoduchou modifikací metody severozápadního rohu. Zbytková matice je lineární kombinací matic $\mathbf{H}^{(1)}, \ldots, \mathbf{H}^{(m)}$, $\mathbf{V}^{(1)}, \ldots, \mathbf{V}^{(n)}$,

$$\mathbf{W}^{(1,1)}, \ldots, \mathbf{W}^{(m-1,1)}, \ldots, \mathbf{W}^{(1,n)}, \ldots, \mathbf{W}^{(m-1,n)},$$

$$\mathbf{E}^{(2,2)}, \ldots, \mathbf{E}^{(m,2)}, \ldots, \mathbf{E}^{(2,n)}, \ldots, \mathbf{E}^{(m,n)},$$

což je matice, která vznikne z Mongeovy matice jejím otočením o 90° po směru hodinových ručiček, tedy transformací $\overline{c}_{j,m+1-i} := c_{i,j}$. Pro tuto matici tedy platí $\overline{c}_{j,m+1-i} + \overline{c}_{s,m+1-r} \leq \overline{c}_{s,m+1-i} + \overline{c}_{j,m+1-r} \quad \forall 1 \leq i < r \leq m, \forall 1 \leq j < s \leq n$. Metoda severozápadního rohu se tedy po otočení matice stává metodou jihozápadního rohu.

V čase $O\left(\bar{n}^{2.376}\right)$ jsme díky Mongeovu rozkladu schopni získat dolní odhad pro minimální dosažitelné náklady v dopravním problému.

Definice 12. Nechť $\mathbf{X} \in \mathbb{R}^{m \times n}$ je reálná matice. **ε-změnou na místech** (u,v) a (s,t) nazveme funkci $f^{\varepsilon}_{(u,v),(s,t)} : \mathbb{R}^{m\times n} \longrightarrow \mathbb{R}^{m \times n}$ takovou, že platí:

$$f^{\varepsilon}_{(u,v),(s,t)}\left(\mathbf{X}\right) = \begin{pmatrix} x_{11} & \ldots & x_{1s} & \ldots & x_{1t} & \ldots & x_{1n} \\ \vdots & \ldots & \vdots & \ldots & \vdots & \ldots & \vdots \\ x_{u1} & \ldots & x_{us}+\varepsilon & \ldots & x_{ut}-\varepsilon & \ldots & x_{un} \\ \vdots & \ldots & \vdots & \ldots & \vdots & \ldots & \vdots \\ x_{v1} & \ldots & x_{vs}-\varepsilon & \ldots & x_{vt}+\varepsilon & \ldots & x_{vn} \\ \vdots & \ldots & \vdots & \ldots & \vdots & \ldots & \vdots \\ x_{m1} & \ldots & x_{ms} & \ldots & x_{mt} & \ldots & x_{mn} \end{pmatrix}, \text{ kde}$$

$$\varepsilon \in \mathbb{R} \setminus \{0\}.$$

Neboli matici **X** se přiřazuje matice, která je s touto maticí totožná na všech místech vyjma míst $(u,s), (u,t), (v,s), (v,t)$, na nichž se na dvou přičítá reálné nenulové číslo ε a na dvou naopak odčítá.

Lemma 8. "O maximální ε-změně"
Nechť matice **C** je nezáporná a Mongeova. Pak matice $\mathbf{D} := f^{\varepsilon}_{(i,r),(j,s)}\left(\mathbf{C}\right)$ je nezáporná Mongeova matice, jestliže

$$\varepsilon \in \left(\max\left\{-c_{ij}, -c_{rs}\right\}, 0\right]. \tag{3.9}$$

Důkaz lemmatu 8. :

Označme místa na nichž provádíme ε-změnu $(\bar{i}, \bar{r})$ a $(\bar{j}, \bar{s})$. Platí

$$c_{\bar{i}\bar{j}} + c_{\bar{r}\bar{s}} \leq c_{\bar{i}\bar{s}} + c_{\bar{r}\bar{j}}$$

a protože ε je nekladné platí triviálně také

$$c_{\bar{i}\bar{j}} + \varepsilon + c_{\bar{r}\bar{s}} + \varepsilon \leq c_{\bar{i}\bar{s}} - \varepsilon + c_{\bar{r}\bar{j}} - \varepsilon$$

čili

$$d_{\bar{i}\bar{j}} + d_{\bar{r}\bar{s}} \leq d_{\bar{i}\bar{s}} + d_{\bar{r}\bar{j}}$$

a přitom $d_{\bar{r}\bar{j}} \geq 0$, $d_{\bar{i}\bar{j}} \geq 0$, $d_{\bar{i}\bar{s}} \geq 0$ a $d_{\bar{r}\bar{s}} \geq 0$. Pro všechny ostatní přípustné dvojice indexů zůstává v platnosti Mongeova podmínka, neboť předpokládáme, že matice $\mathbf{C}$ je Mongeova. Z toho plyne, že matice $\mathbf{D}$ je nezáporná a Mongeova.

$\square$

Lemma 9. "O stejně velkých maticích"
Nechť $\mathbf{C} \in \mathbb{R}^{m \times n}$ a $\mathbf{D} \in \mathbb{R}^{m \times n}$ jsou dvě různé nezáporné matice a nechť existuje posloupnost ε-změn taková, že

$$\mathbf{D} = f^{\varepsilon_1}_{(u_1,v_1),(s_1,t_1)} \left(\ldots \left(f^{\varepsilon_k}_{(u_k,v_k),(s_k,t_k)} (\mathbf{C}) \right) \ldots \right). \tag{3.10}$$

Nechť $\overline{\delta}\,(C) = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij}^2}$ je Euklidova norma na prostoru všech matic typu $m \times n$. Pak

$$\overline{\delta}\,(\mathbf{C}) = \overline{\delta}\,(\mathbf{D}),$$

právě tehdy, když platí

$$\sum_{g=1}^{k} \left(c_{u_g s_g} + c_{v_g t_g} \right) = \sum_{g=1}^{k} \left(c_{u_g t_g} + c_{v_g s_g} \right) - 2 \sum_{g=1}^{k} \varepsilon_g.$$

Důkaz lemmatu 9. : Důkaz je na přiloženém CD v adresáři ***Důkazy*** pod názvem **"O stejně velkých maticích".**

Lemma 10. "O čtyřech nulách"
Pro každý problém (DTP) existuje takové optimální řešení $\mathbf{X} \in \mathbb{R}^{m \times n}$, že

$$\forall 1 \leq u \neq v \leq m, 1 \leq s \neq t \leq n : x_{u,s} = 0 \vee x_{u,t} = 0 \vee x_{v,s} = 0 \vee x_{v,t} = 0. \tag{3.11}$$

Důkaz lemmatu 10. : Důkaz je na přiloženém CD v adresáři ***Důkazy*** pod názvem **"O čtyřech nulách".**

Lemma 11. "O ε-změně"
Nechť $\mathbf{X} \in \mathbb{R}^{m\times n}$ a $\mathbf{Y} \in \mathbb{R}^{m\times n}$. Nechť vektor řádkových součtů matice $\mathbf{X}$ je roven vektoru řádkových součtů matice $\mathbf{Y}$ a vektor sloupcových součtů matice $\mathbf{X}$ je roven vektoru sloupcových součtů matice $\mathbf{Y}$. Pak existuje posloupnost ε-změn $f^{\varepsilon_1}_{(u_1,v_1),(s_1,t_1)}, \ldots, f^{\varepsilon_k}_{(u_k,v_k),(s_k,t_k)}$ taková, že

$$f^{\varepsilon_1}_{(u_1,v_1),(s_1,t_1)}\left(\ldots f^{\varepsilon_k}_{(u_k,v_k),(s_k,t_k)}(\mathbf{X})\ldots\right) = \mathbf{Y}.$$

Důkaz lemmatu 11. : Důkaz je na přiloženém CD v adresáři ***Důkazy*** pod názvem **"O epsilon změně".**

Lemma 12. "O kuželu optimálních řešení"
Nechť $\mathbf{C} \geq \mathbf{0}$ a $\mathbf{X} \in \mathbb{R}$ je optimálním řešením problému (DTP):

$$\min_{z_{ij}\geq 0} \sum_{i=1}^{m}\sum_{j=1}^{n} z_{ij}c_{ij} \tag{3.12}$$

vzhledem k podmínkám

$$\sum_{i=1}^{m} z_{ij} = b_j \text{ pro } j = 1,\ldots,n, \tag{3.13}$$

$$\sum_{j=1}^{n} z_{ij} = a_i \text{ pro } i = 1,\ldots,m. \tag{3.14}$$

Pak platí

$(\forall\varepsilon > 0, \varepsilon \in \mathbb{R})$ ($\mathbf{X}$ je rovněž optimálním řešením (3.12) vzhledem k (3.13) a (3.14) s maticí nákladů $\varepsilon\mathbf{C}$)

Je-li navíc $\mathbf{X}$ optimálním řešením (3.12) vzhledem k (3.13) a (3.14) s maticí nákladů $\mathbf{0} \leq \mathbf{D} \neq \mathbf{C}$, pak $\mathbf{X}$ je optimálním řešením problému

$$\min_{z_{ij}\geq 0} \sum_{i=1}^{m}\sum_{j=1}^{n} z_{ij}\left(c_{ij} + d_{ij}\right)$$

vzhledem k podmínkám (3.13) a (3.14).

Důkaz lemmatu 12. : Důkaz je triviální.

Poznámka 11. Z lemmatu 12 plyne, že množina všech nákladových matic v dopravním problému (DTP), které vedou ke stejnému optimálnímu řešení, je konvexní kužel.

Nyní jsme v situaci, kdy umíme najít vždy přípustné řešení (DTP), existuje-li (tj. platí-li (2.9)) a víme, kdy toto řešení, nalezené metodou severozápadního rohu, je zároveň řešením optimálním. Metoda severozápadního rohu nebere v potaz nákladovou matici **C** a nemůže tedy obecně vést k nalezení optimálního řešení. Navíc ale v obecném případě nevede ani k dobrému výchozímu řešení pro simplexovou metodu, která se používá k nalezení optimálního řešení diskrétních dopravních problémů. Za účelem nalezení přípustného řešení, které by bylo lepším výchozím řešením, tj. výchozím řešením, které vede na nižší počet kroků simplexové metody, bylo vyvinuto několik metod. Všechny jsou nějakou modifikací pravidla pro volbu pořadí obsazování volných polí tabulky, které používá metoda severozápadního rohu v kroku 3. Zmíním zde tři takové metody: metodu minimálního nákladu, Vogelovu metodu a Berkovu metodu.

Metoda minimálního nákladu je velmi podobná metodě severozápadního rohu, jediný rozdíl je v tom, že algoritmus nepostupuje ze "severozápadního" rohu k rohu "jihovýchodnímu", ale postupuje od nejnižších hodnot matice **C** k vyšším, čili v této metodě již hodnoty matice nákladů roli hrají. Je známo, že metoda minimálního nákladu nepřináší obecně téměř žádné zlepšení výchozího řešení pro simplexovou metodu oproti metodě severozápadního rohu.

Vogelova metoda je o něco složitější než metoda minimálního nákladu, ale je také efektivnější (vede většinou k přípustnému řešení s nižší hodnotou součtu v (2.5)). Metoda je postavena na výpočtu penalizací ke každému sloupci i řádku, který je v daném kroku algoritmu neobsazen. Postup je následující:

1. Ke každému řádku matice nákladů , který nebyl dosud eliminován, spočítáme rozdíl mezi nejnižším a druhým nejnižším nákladem v daném řádku a totéž provedeme pro každý dosud neeliminovaný sloupec.

2. Vybereme ze všech řádků a sloupců matice ten, u kterého je spočtený rozdíl nejvyšší[4]. V tomto řádku, či sloupci, alokujeme minimum z příslušného aktuálního řádkového a sloupcového omezení na takové místo matice řešení, které odpovídá nejnižšímu nákladu ve vybraném

[4]Je-li jich více, vytvoříme z nich submatici a obsazujeme místo s nejnižším nákladem v této submatici.

řádku či sloupci. Od příslušného řádkového a sloupcového omezení odečteme takto alokovanou hodnotu.

3. Řádek, či sloupec, který má nulové omezení, eliminujeme a pokračujeme dále s maticí obsahující pouze prvky z neeliminovaných řádků a sloupců. Pokud jsou všechna omezení již nulová, algoritmus končí s přípustným řešením. Pokud je nějaké omezení stále kladné, vrátíme se na první krok.

Přesný algoritmus je detailně popsán v adresáři ***Vogelova metoda*** na přiloženém CD.

Berkova metoda byla publikována v roce 1986 a vychází z Littlova algoritmu, který se používá v teorii grafů k nalezení minimální Hamiltonovské[5] kružnice grafu. Metoda je mnohdy lepší z hlediska dosažené hodnoty účelové funkce přípustného řešení, než metoda Vogelova. Není to však pravidlem. Tato metoda není tak známa jako metody předchozí. Algoritmus je velmi jednoduchý a je určen následujícím postupem:

1. V každém řádku matice nákladů najdeme minimální náklad a ten odečteme od všech prvků daného řádku matice.

2. V každém sloupci matice nákladů najdeme minimální náklad a ten odečteme od všech prvků daného sloupce matice nákladů upravené v prvním kroku.

3. Každému nulovému prvku matice nákladů upravené v předchozích krocích přiřadíme součet minima všech dosud neeliminovaných prvků řádku, v němž se nachází daná nula a minima všech dosud neeliminovaných prvků sloupce, v němž se nachází daná nula.

4. Nulovému prvku s nejvyšším přiřazeným součtem alokujeme maximální hodnotu, kterou umožňují řádková a sloupcová omezení. Je-li tato hodnota rovna aktuálnímu řádkovému omezení, pak z matice eliminujeme daný řádek a odečteme tuto hodnotu od řádkového omezení, je-li rovna aktuálnímu sloupcovému omezení, pak z matice eliminujeme daný sloupec a odečteme tuto hodnotu od sloupcového omezení. Pokud jsou eliminovány všechny sloupce a všechny řádky, pak algoritmus končí s nalezeným přípustným řešením. Pokud ne, vrátíme se na první krok.

[5]Hamiltonovská kružnice je taková kružnice, která obsahuje všechny vrcholy grafu. Tento pojem má význam zejména pro řešení problému obchodního cestujícího.

Přesný algoritmus je detailně popsán v adresáři ***Berkova metoda*** na přiloženém CD a při jeho vytváření jsem čerpal z webových stránek autora (http://home.eunet.cz/berka/o/matempro.htm), kde je metoda stručně popsána.

3.1.2 Optimální řešení

Maďarská metoda je metoda, která se používá k řešení přidělovacího problému a existuje i její zobecnění, které lze použít pro nalezení optimálního řešení problému (DTP). Je založena na jednoduché úvaze, která vychází z následujícího lemmatu.

Lemma 13. Nechť $\mathbf{C}$ je matice nákladů v problému (DTP). Pak přičtením libovolného reálného čísla ke všem prvkům libovolného řádku matice $\mathbf{C}$ a přičtením libovolného reálného čísla ke všem prvkům libovolného sloupce matice $\mathbf{C}$ se optimální řešení problému (DTP) nezmění, pouze účelová funkce v optimu může nabývat jiné hodnoty.

Důkaz lemmatu 13. : Předpokládejme, že matice $\mathbf{C}$ je typu $m \times n$, kde $m, n \in \mathbb{N}$. Definujme novou matici nákladů $\mathbf{D}$ takto: $d_{ij} = c_{ij} + u_i + v_j$, pro všechna $i = 1, \ldots, m$ a $j = 1, \ldots, n$. Účelová funkce v (DTP) pro matici nákladů $\mathbf{D}$ je rovna

$$\sum_{i=1}^{m}\sum_{j=1}^{n} d_{ij}x_{ij} = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} + \sum_{i=1}^{m} u_i \sum_{j=1}^{n} x_{ij} + \sum_{j=1}^{n} v_j \sum_{i=1}^{m} x_{ij}.$$

Protože

$$\sum_{i=1}^{m} x_{ij} = b_j \text{ a } \sum_{j=1}^{n} x_{ij} = a_i \text{ pro všechna } i = 1, \ldots, m \text{ a } j = 1, \ldots, n,$$

dostáváme hodnotu účelové funkce v (DTP) s maticí nákladů $\mathbf{D}$:

$$\sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} + \sum_{i=1}^{m} a_i u_i + \sum_{j=1}^{n} b_j v_j.$$

Protože poslední dva sčítance v tomto vyjádření jsou konstanty, které volbou matice řešení $\mathbf{X}$ neovlivníme a minima je tedy dosaženo ve stejném bodě. Proto je optimální řešení problému (DTP) s maticí nákladů $\mathbf{D}$ totožné s optimálním řešením problému (DTP) s maticí nákladů $\mathbf{C}$, pouze minimum účelové funkce se liší o hodnotu těchto dvou sčítanců.

□

1. Upravíme matici nákladů původního problému (DTP) na matici, která má v každém sloupci i každém řádku minimálně jednu nulu. Toho lze dosáhnout postupnou aplikací lemmatu 13 na matici nákladů $\mathbf{C}$ tak, že nejprve od všech prvků každého řádku odečteme minimální prvek daného řádku, čímž získáme v každém řádku alespoň jednu nulu a poté od všech prvků každého sloupce odečteme minimální prvek daného sloupce, čímž získáme v každém sloupci alespoň jednu nulu.

2. Vybereme maximální možný počet nul v aktuální matici nákladů a to takových, že žádné dvě nejsou ve stejném sloupci ani ve stejném řádku. K určení maximálního možného počtu nul, které je možné vybrat, se užívá tzv. Königova věta, která "říká", že maximální počet nulových prvků s danými vlastnostmi v dané matici je roven minimálnímu počtu krycích čar[6], kterými je možné pokrýt všechny nulové prvky matice.

3. Existuje-li matice s prvky z množiny $\{0, 1\}$, která má v každém sloupci a každém řádku právě jednu jedničku a přitom každá jednička je na místě vybrané nuly z kroku 2, je tato matice optimálním řešením přiřazovacího problému a algoritmus končí. Pokud taková matice neexistuje, pokračujeme krokem 4.

4. Proškrtneme všechny řádky a sloupce v nichž je alespoň jedna nula vybraná v kroku 2. Najdeme minimum z prvků matice, které nejsou proškrtnuté a toto minimum odečteme od všech prvků matice a poté přičteme ke všem proškrtnutým prvkům a to jednou pro prvky proškrtnuté jednou a dvakrát pro prvky proškrtnuté dvakrát (horizontální i vertikální čarou). Vrátíme se zpět na krok 2.

Maďarská metoda po konečném počtu kroků vede k optimálnímu řešení přiřazovacího problému.

Maďarská metoda se původně používala pouze k řešení přidělovacího problému, ale lze ji, kupříkladu za předpokladu celočíselnosti (ale stačí i racionalita, jen je úloha o něco technicky složitější, neboť se nejprve musí provést vynásobení všech omezení jejich nejmenším společným násobkem) všech řádkových a sloupcových omezení, zobecnit i na řešení problému (DTP) tak, že problém (DTP) převedeme na přidělovací problém. Tento převod spočívá v tom, že místo množiny n odběratelů s celkovými požadavky $b = \sum_{j=1}^{n} b_j$ a množiny m dodavetelů s celkovými kapacitami $a = \sum_{i=1}^{m} a_i$,

[6] Krycí čarou je míněno proškrtnutí řádku nebo sloupce.

budeme uvažovat b odběratelů s jednotkovým požadavkem a a dodavatelů s jednotkovou kapacitou. Matice nákladů problému (DTP) je typu $m \times n$. Matice nákladů přidělovacího problému, na který problém (DTP) převádíme, bude typu $a \times b$ a bude mít na místě (i, j), kde $i = 1, \ldots, a$ a $j = 1, \ldots, b$ prvek c_{kl}, kde $k := 1 + \max\left\{h : h \in \{0, \ldots, m-1\} \wedge \sum_{j=1}^{h} a_j < i\right\}$ a $l := 1+\max\left\{h : h \in \{0, \ldots, n-1\} \wedge \sum_{i=1}^{h} b_i < j\right\}$. Označme takto vzniklou matici $\mathbf{D}$. Nyní můžeme řešit původní problém (DTP) jako přiřazovací problém s nákladovou maticí $\mathbf{D}$. Problémem této metody je skutečnost, že ačkoliv původní matice nákladů může být velmi malá, matice nákladů vytvořeného přiřazovacího problému může být obrovská. Tuto nepříjemnost však lze s mírnými technickými obtížemi zdolat.

Simplexová metoda je metoda, která byla vyvinuta americkým matematikem Georgem Dantzigem na konci čtyřicátých let 20. století. Je založena na procházení vrcholů mnohorozměrného polyedru (jenž je dán množinou omezení) tak, že hodnota účelové funkce v žádném kroku nevzroste a řešení zůstávají nezáporná. Simplexová metoda má celou řadu modifikací a je o ní známo, že v průměru řeší diskrétní dopravní problém velmi rychle, tj. v polynomiálním čase. V roce 1972 však přišli v [23] američané V. Klee a G. J. Minti s úlohou lineárního programování jejíž složitost je $O(\exp(2n-1))$. Algoritmus simplexové metody (včetně důkazů tvrzení, na nichž je postaven) je detailně popsán mimojiné ve skriptech [38] a knihách [32], [26] a [37].

Simplexový algoritmus je použitelný jak na problém (DTP), tak na problém (DMKP) i na jejich různé modifikace, jako jsou například dodatečná kapacitní omezení.

Metoda MODI (metoda řádkových a sloupcových čísel) je používána často pro svoji efektivitu a vychází z řešení duální úlohy k (DTP) prostřednictvím simplexové metody (jde tedy v podstatě o simplexovou metodu). Využívá ε-změn. Spočívá v následujícím iteračním postupu:

- Nejprve najdeme přípustné řešení pomocí jedné ze zmíněných metod, které vedou k jeho nalezení.

- Dále pro každé obsazené pole matice (tj. pro takovou dvojici indexů (i, j), pro níž nalezené přípustné řešení je kladné $x_{ij} > 0$) vytvoříme rovnici $R_i + K_j = c_{ij}$, kde R_i je hodnota příslušná i-tému řádku, K_j hodnota příslušná j-tému sloupci a c_{ij} je hodnota nákladové matice na místě (i, j).

- Položme $R_1 := 0$ a vyřešme soustavu vzniklých rovnic pro všechna R a K.

- Spočítáme redukovaný náklad $I_{kl} := c_{ij} - R_i - K_j$ pro všechna pole matice řešení, která aktuálně nejsou obsazena.
- Najdeme nejmenší záporný redukovaný náklad (označme jej bez újmy na obecnosti I_{i_1,j_1}) a provedeme ε-změnu na místech (i_1, i_2) a (j_1, j_2), kde pole v matici na místech (i_1, j_2) a (i_2, j_1) jsou obsazena, přičemž zvolíme $\varepsilon = \min\{x_{i_1,j_2}, x_{i_2,j_1}\}$ tak, že hodnotu ε přičítáme k hodnotám matice x_{i_1,j_1} a x_{i_2,j_2} a odčítáme od hodnot x_{i_2,j_1} a x_{i_1,j_2}, přičemž víme, že $x_{i_1,j_1} = 0$. Pokud neexistuje žádný záporný redukovaný náklad, pak jsme dosáhli optimálního řešení a algoritmus končí. Jinak pokračujeme v algoritmu dalším krokem.
- Vrátíme se na krok následující po nalezení přípustného řešení, nyní však již s přípustným řešením po v předchozím kroku provedené ε-změně.

Přesný algoritmus je uveden například v knihách [15] (kapitola 3, str. 39-71) a [24] (str. 98-102).

Ford-Fulkersonův algoritmus vychází z teorie grafů, přesněji z teorie toků v sítích. Problém (DTP) lze totiž převést s použitím duální úlohy na problém typu max-flow (maximálního toku v síti). Mějme tedy dán problém (DTP) a převeďme jej na problém k němu duální. Dostaneme následující maximalizační úlohu:

$$\max_{\alpha,\beta} \sum_{i=1}^{m} a_i\alpha_i + \sum_{j=1}^{n} b_j\beta_j \tag{3.15}$$

za podmínek

$$\alpha_i + \beta_j \leq c_{ij} \text{ pro všechna } i = 1, \ldots, m \text{ a všechna } j = 1, \ldots, n.$$

Triviální přípustné řešení této úlohy je $\alpha_i = 0$ a $\beta_j = \min_{1\leq i\leq m} c_{ij}$. Definujme množinu G takto: $G := \{(i,j) : \alpha_i + \beta_j = c_{ij}\}$. Nyní se podívejme na následující minimalizační úlohu:

$$\min \sum_{i=1}^{m+n} y_i \tag{3.16}$$

za podmínek

$$\begin{aligned} \sum_{j=1}^{n} x_{ij} + y_i = a_i \text{ pro všechna} \qquad & i = 1, \ldots, m, \\ \sum_{i=1}^{m} x_{ij} + y_{m+j} = b_j \text{ pro všechna} \qquad & j = 1, \ldots, n, \\ y_i \geq 0 \text{ pro všechna} \qquad & i = 1, \ldots, m+n, \\ x_{ij} \geq 0 \text{ pokud} \qquad & (i,j) \in G \text{ a} \\ x_{ij} = 0 \text{ pokud} \qquad & (i,j) \notin G. \end{aligned}$$

Z těchto podmínek snadno plyne, že

$$\sum_{i=1}^{m+n} y_i = \sum_{i=1}^{m} a_i + \sum_{j=1}^{n} b_j - 2 \sum_{(i,j)\in G} x_{ij}.$$

Proto je tato úloha ekvivalentní maximalizační úloze:

$$\max \sum_{(i,j)\in G} x_{ij} \tag{3.17}$$

za podmínek

$$\begin{aligned} \sum_{j=1}^{n} x_{ij} \leq a_i \text{ pro všechna} \qquad & i = 1, \ldots, m, \\ \sum_{i=1}^{m} x_{ij} \leq b_j \text{ pro všechna} \qquad & j = 1, \ldots, n, \\ x_{ij} \geq 0 \text{ pokud} \qquad & (i,j) \in G \text{ a} \\ x_{ij} = 0 \text{ pokud} \qquad & (i,j) \notin G. \end{aligned}$$

Toto je problém maximalizace toku v sítích, v němž jde o nalezení maximálního průtoku sítí.

Definice 13. Sítí rozumíme orientovaný graf s celočíselnými kladnými kapacitními omezeními každé jeho hrany, který obsahuje zdrojový vrchol (bez hrany orientované k němu) a terminální vrchol (bez hrany orientované z něj).

Definice 14. Tokem v síti t rozumíme vektor z $\mathbb{R}^k$, kde k je počet všech orientovaných hran sítě, takový, že každá jeho složka je přiřazena právě jedné orientované hraně sítě a platí následující dvě podmínky:

1. Pro všechny vrcholy u a v takové, že existuje v síti hrana jdoucí z vrcholu u do vrcholu v, je složka vektoru toku odpovídající hraně z u do v nezáporná a přitom nejvýše rovna kapacitnímu omezení této hrany.

2. Pro každý vrchol v vyjma terminálního a zdrojového je součet všech složek vektoru toku odpovídajících hranám, které jdou do vrcholu v roven součtu všech složek vektoru toku odpovídajících hranám, které jdou z vrcholu v.

Úloha (3.17) tedy odpovídá maximalizaci toku v síti, v níž ze zdrojového vrcholu vychází m hran s kapacitami a_i pro i-tou z těchto hran. Každá z těchto m hran vede do jednoho z m vrcholů (nazvěme je dodavatelskými vrcholy) a z každého z těchto dodavatelských vrcholů vychází n hran s nekonečnými kapacitami, každá do jednoho z n jiných vrcholů (nazvěme je odběratelskými). Z odběratelských vrcholů vychází nakonec do terminálního vrcholu n hran s kapacitami b_i pro i-tou z těchto hran. Řešení x_{ij} problému (3.17) je dáno tokem v této síti a to jako složka toku odpovídající hraně, která jde z i-tého dodavatelského vrcholu do j-tého odběratelského vrcholu.

Řešení je aplikace Ford-Fulkersonova algoritmu na problém maximálního toku v sítích (3.17). Tento algoritmus je uveden v knize [32], str. 123. Jeho výhodou je možnost zavedení kapacitních omezení a lze jím tedy řešit i dopravní problém s kapacitními omezeními, viz. definice 3. Nevýhodou algoritmu je, že nemusí skončit, pokud nějaké z omezení a_i, $i = 1, \ldots, m$ a b_j, $j = 1, \ldots, n$ není racionální. V takovém případě nejen, že nemusí skončit, ale může dokonce konvergovat k řešení, které je striktně nižší, než optimální, viz. [17].

Metoda Lagrangeových multiplikátorů je obecnou optimalizační metodou, která se používá spíše v nelineárním programování (tj. při vázané optimalizaci s nelineární diferencovatelnou účelovou funkcí). Lze ji však použít i na dopravní problém a zejména pak na dopravní problém s nelineární účelovou funkcí. Tato metoda je velmi známá a proto se jí zde nebudu věnovat, nicméně je v práci použita při aplikaci dopravního problému na rekonstrukce kontingenčních tabulek a je základem důkazu Strassenovy věty (str. 82). Detaily o této metodě je možné najít například v knize [2].

3.1.3 Přibližná řešení (heuristické algoritmy)

Heuristické algoritmy, které vedou pouze k přibližnému řešení využívají různých teoretických konceptů, od neuronových sítí (viz. [31]), přes genetické algoritmy (viz. [20]), až po metody typu "Local Search" a evoluční algoritmy[7].

3.1.4 Algoritmy pro speciální případy

Existuje celá řada postupů, které řeší nějakou specifickou podmnožinu problémů typu (DTP) (například matice nákladů má jisté vlastnosti[8]–je Mongeova, náklady jsou ryze konvexní funkcí vzdálenosti, jsou celočíselné, apod.). Přidáme-li například podmínku $n = 2$, pak v článku [27] najdeme algoritmus, který tento problém řeší v lineárním čase převedením na známý problém batohu[9].

Další specifický typ dopravního problému, který je řešitelný v lineárním čase, je problém s maticí nákladů, kde každý prvek je součinem dvou hodnot, z nichž jedna je dána řádkovým indexem a druhá indexem sloupcovým, tedy pro každé $i \in \{1, \ldots, m\}$ a každé $j \in \{1, \ldots, n\}$ platí $c_{ij} = u_i v_j$, kde $u_i > 0$ a $v_j > 0$. V tomto případě je možné problém řešit metodou severozápadního rohu, ale až poté, co se řádky uspořádají podle jim příslušných hodnot u_i a sloupce podle jim příslušných hodnot v_j, ale v opačném pořadí. Označme největší hodnotu u_p příslušející po uspořádání řádků prvnímu řádku u_1^{usp}, druhou nejvyšší u_2^{usp}, atd. a stejně tak nejvyšší hodnotu v_r po uspořádní sloupců jako v_n^{usp}, druhou v_{n-1}^{usp}, atd. Jak víme, metoda severozápadního rohu dává optimální řešení právě tehdy, když je matice nákladů Mongeova. Pro $k < l$ a $p < o$ dostáváme $0 \geq \left(u_k^{usp} - u_l^{usp}\right)\left(v_p^{usp} - v_o^{usp}\right) = u_k^{usp} v_p^{usp} + u_l^{usp} v_o^{usp} - u_k^{usp} v_o^{usp} - u_l^{usp} v_p^{usp} = c_{kp} + c_{lo} - c_{ko} - c_{lp}$, což je definice Mongeovy matice. Tímto problémem se zabývá článek [14]. Obdobně je-li každý prvek matice nákladů $c_{ij} = u_i + v_j$, pak platí $c_{ij} + c_{kl} - c_{il} - c_{kj} = u_i + v_j + u_k + v_l - u_i - u_k - v_j - v_l = 0$. Jde tedy rovněž o Mongeovu matici a problém je řešitelný metodou severozápadního rohu v lineárním čase.

3.2 Řešení diskrétního vícerozměrného dopravního problému

Uvažujme vyrovnaný vícerozměrný diskrétní dopravní problém (DMKP).

Definice 15. Nechť $n \in \mathbb{N}$, $N = \{1, \ldots, n\}$ a $m_1, \ldots, m_n \in \mathbb{N}$. Nechť $A_i = \{1, \ldots, m_i\}$, pro všechna $i = 1, \ldots, n$ a $A = \otimes_{i \in N} A_i$. n**-rozměrnou reálnou**

[7]Disertační práce (Hasan Timucin Ozdemir): "Graph based evolutionary algorithms for transportation problems".

[8]Jak jsme ukázali, když je matice nákladů Mongeova, pak lze (DTP) řešit v lineárním čase metodou severozápadního rohu.

[9]Problém batohu je popsán kupříkladu v článku [33].

tabulkou typu $m_1 \times \ldots \times m_n$ budu nazývat uspořádanou dvojici (T, f), kde $T = \{r_1, \ldots, r_{m_1 \cdots m_n}\}$, $r_j \in \mathbb{R}$, pro všechna $j = 1, \ldots, m_1 \cdots m_n$ je multimnožina reálných čísel mohutnosti $m_1 \cdots m_n$ a f je prosté zobrazení z T na množinu A.

n-rozměrnou tabulku budu značit $\mathcal{T}_n$. Množinu všech n-rozměrných tabulek budu značit $\{\mathcal{T}_n\}$. Prvek t z množiny T, pro nějž platí $f(t) = (i_1, \ldots, i_n)$ budeme značit $t_{i_1,\ldots,i_n}$. Matice nákladů v (DTP) je tedy zřejmě dvourozměrnou tabulkou, tj. $\mathbf{C} \in \{\mathcal{T}_2\}$ takovou, že $t_{ij} := c_{ij}$. (DTP) obsahuje vyjma omezení na nezápornost omezení na součty řádků a součty sloupců. Tato omezení je možné zapsat ve formě 1-rozměrných tabulek. Chceme-li popsat omezení v případě trojrozměrného problému (DMKP), pak jsou dvě možnosti. Buď je toto omezení dvourozměrnou tabulkou, nebo tabulkou jednorozměrnou. Obecně můžeme mít v (DMKP) kombinaci těchto dvou typů omezení. Pokud jsou všechna omezení 1-rozměrné tabulky, pak jde o trojrozměrný diskrétní axiální dopravní problém, pokud jsou všechna omezení 2-rozměrné tabulky, jde o trojrozměrný diskrétní planární dopravní problém. V dalším textu se budu zabývat především těmito dvěma trojrozměrnými problémy. Metody řešení axiálního a metody řešení planárního problému jsou značně odlišné. Zatímco na řešení axiálního dopravního problému lze s úspěchem použit jednoduchá zobecnění metod používaných při řešení problému (DTP), řešení planárního dopravního problému je značně obtížnější. Již samotné zjištění existence přípustného řešení planárního dopravního problému je velmi obtížnou úlohou. Abych mohl pokračovat v popisu vícerozměrných dopravních problémů a definovat přípustné a optimální řešení, musím nejprve definovat součty n-rozměrné tabulky v různých směrech.

Definice 16. Nechť $n \in \mathbb{N}$, $N = \{1, \ldots, n\}$ a $m_1, \ldots, m_n \in \mathbb{N}$. Nechť $\mathcal{T}_n$ je n-rozměrná reálná tabulka typu $m_1 \times \ldots \times m_n$. Nechť $h \in \{1, \ldots, n-1\}$. Označme $(j_1, \ldots, j_h)$ uspořádanou h-tici takovou, že $j_i \in N$, pro všechna $i = 1, \ldots, h$ a $j_1 < \ldots < j_h$. Označme dále

$$\mathcal{F}_n^{j_1,\ldots,j_h} = \left\{ \sum_{k_1=1}^{m_{j_1}} \cdots \sum_{k_h=1}^{m_{j_h}} t_{i_1,\ldots,i_{j_1-1},k_1,i_{j_1+1},\ldots,i_{j_2-1},k_2,i_{j_2+1},\ldots,i_{j_h-1},k_h,i_{j_h+1},\ldots,i_n} : \right.$$
$$i_s \in \{1, \ldots, m_s\}, s \in \{1, \ldots, j_1 - 1, j_1 + 1, \ldots, j_2 - 1, j_2 + 1, \ldots,$$
$$\left. j_h - 1, j_h + 1, \ldots, n\} \right\}.$$

$(n-h)$-**tým marginálem** $j_1, \ldots, j_h$-**tých směrů** tabulky $\mathcal{T}_n$ budu nazývat uspořádanou dvojici $\left(\mathcal{F}_n^{j_1,\ldots,j_h}, f_n^{j_1,\ldots,j_h}\right)$, kde $f_n^{j_1,\ldots,j_h}$ je prosté zobrazení z $\mathcal{F}_n^{j_1,\dots,j_h}$ na množinu všech uspořádaných $(n-h)$-tic přirozených čísel takových, že l-tý prvek této $(n-h)$-tice je z množiny

$$\begin{array}{l} \{1,\dots,m_l\}\,,\ \text{pokud}\ l < j_1 \\ \{1,\dots,m_{l+1}\}\,,\ \text{pokud}\ j_1 \le l < j_2 \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ \{1,\dots,m_{l+h-1}\}\,,\ \text{pokud}\ j_{h-1} \le l < j_h \\ \{1,\dots,m_{l+h}\}\,,\ \text{pokud}\ j_h < l \end{array}$$

$$f_n^{j_1,\dots,j_h}\left(\left\{\sum_{k_1=1}^{m_{j_1}}\cdots\sum_{k_h=1}^{m_{j_h}} t_{i_1,\dots,i_{j_1-1},k_1,i_{j_1+1},\dots,i_{j_2-1},k_2,i_{j_2+1},\dots,i_{j_h-1},k_h,i_{j_h+1},\dots,i_n}\right\}\right)$$
$$= (i_1,\dots,i_{j_1-1},i_{j_1+1},\dots,i_{j_2-1},i_{j_2+1},\dots,i_{j_h-1},i_{j_h+1},\dots,i_n)$$

a to pro všechny uspořádané $(n-h)$-tice

$$(i_1,\dots,i_{j_1-1},i_{j_1+1},\dots,i_{j_2-1},i_{j_2+1},\dots,i_{j_h-1},i_{j_h+1},\dots,i_n)$$

takové, že l-tý prvek této $(n-h)$-tice je z množiny

$$\begin{array}{l} \{1,\dots,m_l\}\,,\ \text{pokud}\ l < j_1 \\ \{1,\dots,m_{l+1}\}\,,\ \text{pokud}\ j_1 \le l < j_2 \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ \{1,\dots,m_{l+h-1}\}\,,\ \text{pokud}\ j_{h-1} \le l < j_h \\ \{1,\dots,m_{l+h}\}\,,\ \text{pokud}\ j_h < l. \end{array}$$

$(n-h)$-tý marginál $j_1,\dots,j_h$-tých směrů tabulky $\mathcal{T}_n$ budu značit

$$\mathcal{E}_n^{n-h}(j_1,\dots,j_h).$$

Místo k-tý marginál, budu často zkráceně psát o $k-$-marginálu.

Pro tabulku $\mathcal{T}_n$ existuje n různých $(n-1)$-rozměrných tabulek, které jsou jejími $(n-1)$-marginály, každý v jiném směru. Indexy j_i ve výrazu $(n-1)$-marginál j_i-tého směru značí "směr" sčítání prvků tabulky. Pro n-rozměrnou tabulku existuje n různých směrů v nichž je možné tabulku sčítat. Každý n-rozměrný marginál nějaké tabulky vyššího rozměru je sám o sobě n-rozměrnou tabulkou a má tudíž marginály všech rozměrů nižších než n. n-rozměrné tabulky a jejich marginály jsme zavedli obecně, nicméně pro dopravní problémy, které jsou vždy dány nějakou n-rozměrnou tabulkou nákladů a nějakou množinou marginálů, budeme uvažovat pouze takové marginály, jejichž všechny 1-marginály mají všechny složky kladné. Pokud by totiž 1-rozměrný marginál k-rozměrného marginálu nějaké n-rozměrné tabulky typu $m_1 \times \ldots \times m_n$, kde $n > k \geq 1$ obsahoval nulu, bylo by možné původní n-rozměrný problém redukovat na problém některého z typů $m_1 - 1 \times \ldots \times m_n, \ldots, m_1 \times \ldots \times m_n - 1$.

Axiální 3-rozměrný diskrétní dopravní problém značíme $\mathrm{DMKP}_{3,2}$ a planární 3-rozměrný diskrétní dopravní problém $\mathrm{DMKP}_{3,1}$. Při řešení $\mathrm{DMKP}_{3,2}$ hledáme nezápornou reálnou tabulku $\mathcal{T}_3$ při daných 1-marginálech $\mathcal{E}_3^1(1,2)$, $\mathcal{E}_3^1(2,3)$ a $\mathcal{E}_3^1(1,3)$ a při řešení $\mathrm{DMKP}_{3,1}$ hledáme nezápornou reálnou tabulku $\mathcal{T}_3$ při daných 2-marginálech $\mathcal{E}_3^2(1)$, $\mathcal{E}_3^2(2)$ a $\mathcal{E}_3^2(3)$. Problémy $\mathrm{DMKP}_{3,2}$ a $\mathrm{DMKP}_{3,1}$ jsou jediné dva symetrické diskrétní 3-rozměrné dopravní problémy.

Definice 17. Řekneme, že n-rozměrná tabulka $\mathbf{X} = [x_{i_1 \ldots i_n}]$, $i_r \in \{1, \ldots, m_r\}$, kde $r = 1, \ldots, n$ je **přípustným řešením (DMKP)**, pokud je nezáporná a jsou splněny všechny rovnosti (2.2) v definici (DMKP).

Definice 18. Řekneme, že n-rozměrná tabulka $\mathbf{X} = [x_{i_1 \ldots i_n}]$, $i_r \in I_r$, kde $r = 1, \ldots, n$ je **optimálním řešením (DMKP)**, pokud je přípustným řešením (DMKP) a zároveň neexistuje jiné přípustné řešení, na němž by účelová funkce daná v definici (DMKP) nabývala nižší hodnoty.

Poznámka 12. n-rozměrný diskrétní dopravní problém definovaný v definici 1 je problémem lineárního programování o $m_1 \ldots m_n$ proměnných a $\sum_{i=1}^{n} m_1 \ldots m_{i-1} m_{i+1} \ldots m_n$ omezeních, kde $m_r = card\,(I_r)$ a některá omezení jsou nadbytečná. Jako takový je samozřejmě řešitelný pomocí simplexové metody.

3.2.1 Přípustné řešení

Jak ukážeme, je při hledání přípustného řešení vícerozměrného diskrétního dopravního problému důležité oddělit problémy axiální od problémů planárních a problémů, které nejsou ani planární, ani axiální.

Axiální dopravní problém

Víme, že nutnou a postačující podmínkou pro existenci přípustného řešení axiálního problému $\mathrm{DMKP}_{n,n-1}$ je (2.3) a přípustným řešením je kupříkladu řešení dané (2.4). Existuje-li, lze přípustné řešení nalézt pomocí zobecněné metody severozápadního rohu, případně nějaké jednoduché modifikace této metody, jako je kupříkladu zobecnění metody minimálního nákladu. Zobecněná metoda severozápadního rohu je jednoduchým hladovým algoritmem a je popsána kupříkladu v [16] na stranách 453-454.

Zobecnění metody severozápadního rohu spočívá v tom, že obdobně jako v dvourozměrném případě postupně obsazujeme pole tabulky $\mathcal{T}_n$ od pole daného uspořádanou n-ticí indexů $(1,\ldots,1)$ k poli danému uspořádanou n-ticí indexů $(m_1,\ldots,m_n)$. Přitom vkládáme v každém kroku algoritmu maximální množství, které je vzhledem k omezením možné. Máme-li tedy v nějakém kroku obsadit pole dané n-ticí indexů $(i_1,\ldots,i_n)$, vložíme do něj hodnotu $\gamma := \min\left\{(\mathcal{E}_n^1(2,\ldots,n))_{i_1},\ldots,(\mathcal{E}_n^1(1,\ldots,n-1))_{i_n}\right\}$. Předpokládejme bez újmy na obecnosti, že $\gamma = (\mathcal{E}_n^1(2,\ldots,n))_{i_1}$. Pak pro všechna $k=1,\ldots,n$ položíme

$$\left(\mathcal{E}_n^1(1,\ldots,k-1,k+1,\ldots,n)\right)_{i_k} := \left(\mathcal{E}_n^1(1,\ldots,k-1,k+1,\ldots,n)\right)_{i_k} - \gamma$$

a přejdeme na obsazování pole určeného n-ticí indexů $(i_1+1,i_2,\ldots,i_n)$. Takto postupujeme dokud neexistuje žádný index, který by nabýval vyšší hodnoty, než je přípustné, tj. skončíme, jakmile existuje $j\in\{1,\ldots,n\}$, takové, že $i_j > m_j$. Tímto jednoduchým postupem se dostaneme k přípustnému řešení $\mathrm{DMKP}_{n,n-1}$. Přesný algoritmus je detailně popsán v adresáři ***Zobecněná metoda severozápadního rohu*** na přiloženém CD. Dále je v témže adresáři uložen spustitelný soubor, který umožňuje nalézt řešení zobecněnou metodou severozápadního rohu pro axiální dopravní problémy $\mathrm{DMKP}_{2,1}$, $\mathrm{DMKP}_{3,2}$, až $\mathrm{DMKP}_{10,9}$, s rozměry až do velikosti 50 splňující omezení na součin všech rozměrů, který musí být z technických důvodů menší než 65 530 (nesmí přesáhnout počet řádků listu v Excelu). Jak je vidět, tento soubor zahrnuje i řešení problému (DTP) pomocí klasické metody severozápadního rohu a proto jsem ji samostatně neimplementoval.

Zobecněná metoda minimálního nákladu je obdobou metody minimálního nákladu z dvourozměrného případu a je jednoduchou modifikací zobecněné metody severozápadního rohu. Postupuje se v ní obsazováním dosud neobsazených polí n-rozměrné tabulky, přičemž na počátku nejsou obsazena žádná pole. Narozdíl od zobecněné metody severozápadního rohu se ale nepostupuje od pole daného uspořádanou n-ticí indexů $(1,\ldots,1)$ k poli daného uspořádanou n-ticí indexů $(m_1,\ldots,m_n)$, nýbrž od pole, kterému odpovídá nejnižší hodnota v nákladové tabulce k poli, kterému odpovídá hodnota nejvyšší.

Definujme nyní Mongeovu tabulku, která je zobecněním Mongeovy matice z teorie (DTP). Lze totiž ukázat, že je-li n-rozměrná tabulka nákladů $\mathcal{T}_n$ Mongeova, je řešení axiálního problému $\mathrm{DMKP}_{n,n-1}$ nalezené pomocí zobecněné metody severozápadního rohu řešením optimálním.

Definice 19. Nechť $n\in\mathbb{N}$, $N=\{1,\ldots,n\}$ a $m_1,\ldots,m_n\in\mathbb{N}$. Nechť $A_i=\{1,\ldots,m_i\}$, pro všechna $i=1,\ldots,n$ a $A=\otimes_{i\in N}A_i$. Řekneme, že

n-rozměrná reálná tabulka typu $m_1 \times \ldots \times m_n$ splňuje Mongeovu podmínku, nebo že je Mongeova, pokud

$$\forall \mathbf{a}, \mathbf{b} \in A: \quad t_{(\max\{a_1,b_1\},\ldots,\max\{a_n,b_n\})} + t_{(\min\{a_1,b_1\},\ldots,\min\{a_n,b_n\})} \leq t_{\mathbf{a}} + t_{\mathbf{b}}, \tag{3.18}$$

kde t_a, $a \in A$ značí prvek tabulky $\mathcal{T}_n$ na místě daném uspořádanou n-ticí indexů $(a_1, \ldots, a_n)$.

Věta 5. Řešení DMKP$_{n,n-1}$ zobecněnou metodou severozápadního rohu je optimální pro všechny kladné 1-marginály splňující (2.3) právě tehdy, když tabulka nákladů je Mongeova.

Důkaz věty 5. : Důkaz je uveden v článku [3].

Planární dopravní problém

Ačkoliv by se mohlo zdát, že nalezení přípustného řešení planárního problému DMKP$_{n,1}$ je pouze analogií k nalezení přípustného řešení problému DMKP$_{n,n-1}$, situace je zde znatelně komplikovanější. Zabývejme se nejjednodušším planárním problémem DMKP$_{3,1}$. Je zřejmé, že k tomu, aby existovalo přípustné řešení DMKP$_{3,1}$ dopravního problému s trojrozměrnou tabulkou nákladů C typu $2 \times 2 \times 2$ a s maticí pravo-levých sloupcových součtů

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix},$$

maticí dolno-horních sloupcových součtů

$$\mathbf{B} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$$

a maticí předo-zadních sloupcových součtů

$$\mathbf{D} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix},$$

musí platit určitě alespoň $a_{11}+a_{12} = c_{11}+c_{12}$, $a_{21}+a_{22} = c_{21}+c_{22}$, $a_{11}+a_{21} = b_{11} + b_{21}$, $a_{12} + a_{22} = b_{12} + b_{22}$, $b_{11} + b_{12} = c_{11} + c_{21}$ a $b_{21} + b_{22} = c_{12} + c_{22}$. Neboli součet zadní čtveřice musí být stejný při sčítání po sloupcích jako při sčítání po řádcích a totéž musí platit pro přední, levou, pravou, horní i dolní čtveřici prvků matice řešení typu $2 \times 2 \times 2$. Tyto podmínky jsou sice nutné pro existenci přípustného řešení, ale nejsou postačující. To nahlédneme na elementárním příkladu.

Mějme DMKP$_{3,1}$ typu $2 \times 2 \times 2$ s následujícími 2-marginály maticí pravo-levých sloupcových součtů

$$\mathcal{E}_3^2(2) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

maticí dolno-horních sloupcových součtů

$$\mathcal{E}_3^2(1) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

a maticí předo-zadních sloupcových součtů

$$\mathcal{E}_3^2(3) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Tato omezení sice splňují všechny předchozí rovnosti, ale přesto neexistuje žádné přípustné řešení. Například prvek $x_{1,1,2}$ je dán omezením $(\mathcal{E}_3^2(2))_{1,2}$ jedničkou, omezením $(\mathcal{E}_3^2(3))_{1,1}$ rovněž jedničkou, ale omezením $(\mathcal{E}_3^2(1))_{2,1}$ nulou. Když ale na toto místo alokujeme nulu, už nemůžeme naplnit omezení $(\mathcal{E}_3^2(2))_{1,2}$, neboť prvek $x_{1,2,2}$ je dán omezením $(\mathcal{E}_3^2(3))_{1,2}$ nulou a je tedy nutné na toto místo alokovat také nulu. Tím je prokázána neexistence přípustného řešení. Vzniká tedy přirozená otázka, jaké vlastnosti musejí mít omezení v DMKP$_{3,1}$, aby existovalo přípustné řešení. Problémem existence přípustného řešení planárního problému DMKP$_{3,1}$ při daných omezeních se zabývala celá řada autorů, ale jako jedni z prvních se jím začali zabývat M. Vlach a J. Morávek a to již v padesátých letech dvacátého století. Dále se jím zabývali K. B. Haley a G. Smith. Jde zejména o články [28], [29], [44], [42], [43], [49], [9], [10] a [8].

V článku [49] je uveden seznam různých nutných podmínek existence přípustného řešení problému DMKP$_{3,1}$. Explicitně dané podmínky na 2-marginály, které by byly nutné a zároveň postačující pro existenci přípustného řešení problému DMKP$_{3,1}$ typu $m \times n \times o$, kde $m \geq 3$, $n \geq 3$ a $o \geq 3$ nejsou známy, nicméně v článku [43] je uveden obecný postup pro jejich nalezení. Tento postup využívá duality a je třeba jej aplikovat na každý konkrétní problém zvlášť. Nejedná se o explicitně obecně vyjádřené podmínky, nýbrž o testování všech bazí duálního problému, kterých může být velmi mnoho. Domnívám se, že k testování existence přípustného řešení lze využít i dříve zavedený Mongeův rozklad, viz. definice 11 a algoritmus v adresáři ***Mongeův rozklad*** na přiloženém CD. Pokud totiž provedeme Mongeův rozklad tří daných 2-marginálů, můžeme z těchto rozkladů zrekonstruovat trojrozměrnou tabulku, pokud existuje. Je totiž zřejmé, že máme-li Mongeův rozklad prvního 2-marginálu, dostáváme tím omezení na rozklad druhého 2-marginálu a máme-li dány Mongeovy rozklady dvou 2-marginálů trojrozměrné tabulky, pak je Mongeův rozklad třetího 2-marginálu již dán. Existenci a rychlost algoritmu, který by testoval existenci přípustné tabulky planárního trojrozměrného diskrétního dopravního problému pomocí Mongeova rozkladu ponechávám v rámci této práce jako hypotézu a budu se jím dále zabývat.

Analyticky vyjádřených podmínek na 2-marginály nutných pro existenci přípustného řešení $\mathrm{DMKP}_{3,1}$ byla vytvořena celá řada, ale pro žádné nebylo dokázáno, že jsou postačující. Některé z nich jsou poměrně komplikované a navíc téměř u všech je dokázáno, že postačující nejsou. Uveďme tedy alespoň elementární podmínky nutné pro existenci přípustného řešení $\mathrm{DMKP}_{3,1}$.

Uvažujme dvourozměrné tabulky $\mathbf{A}$ typu $m_1 \times m_2$, $\mathbf{B}$ typu $m_1 \times m_3$ a $\mathbf{C}$ typu $m_2 \times m_3$.

$$\text{Předpokládáme } a_{ij} \geq 0,\ b_{ik} \geq 0,\ c_{jk} \geq 0 \text{ pro všechna}$$
$$i \in \{1, \ldots, m_1\},\ j \in \{1, \ldots, m_2\},\ k \in \{1, \ldots, m_3\}.$$
$$\sum_{j=1}^{m_2} a_{ij} = \sum_{k=1}^{m_3} b_{ik} \text{ pro každé } i \in \{1, \ldots, m_1\} \tag{3.19}$$
$$\sum_{i=1}^{m_1} a_{ij} = \sum_{k=1}^{m_3} c_{jk} \text{ pro každé } j \in \{1, \ldots, m_2\} \tag{3.20}$$
$$\sum_{i=1}^{m_1} b_{ik} = \sum_{j=1}^{m_2} c_{jk} \text{ pro každé } k \in \{1, \ldots, m_3\} \tag{3.21}$$

Jak jsme ovšem ukázali, splnění těchto omezení nestačí k tomu, aby existovala trojrozměrná tabulka s 2-marginály $\mathbf{A}$, $\mathbf{B}$ a $\mathbf{C}$. Pro obecný planární problém $\mathrm{DMKP}_{n,1}$ můžeme tyto nutné podmínky vyjádřit pomocí následujícího lemmatu.

Lemma 14. "O existenci řešení"
Nutnou podmínkou pro existenci přípustného řešení n-rozměrného diskrétního dopravního problému je, že pro libovolné dvě množiny indexů $E \in \mathcal{C}$ a $F \in \mathcal{C}$ platí

$$\sum_{i_v \in M_v; v \in E^C \cap F} b_{i(E^C)} = \sum_{i_v \in M_v; v \in F^C \cap E} b_{i(F^C)}. \tag{3.22}$$

Pokud $E^C \cap F^C \neq \emptyset$, pak podmínka (3.22) musí platit pro všechny indexy $i_w \in M_w$, kde $w \in E^C \cap F^C$.

Důkaz lemmatu 14. : Důkaz je uveden v článku [22].

V článku [8] se autor zabývá nutnými a postačujícími podmínkami pro existenci vícerozměrné tabulky a uvádí, že algoritmus uvedený v adresáři ***Rekonstrukce 3d tabulky*** na přiloženém CD, který jako vstup použije tabulku obsahující samé jedničky, konverguje k přípustnému řešení, pokud toto existuje. Autor v článku dokazuje následující tvrzení.

Věta 6. "Test existence řešení"
Pro dané 2-marginály $\mathcal{E}_3^2(1)$, $\mathcal{E}_3^2(2)$ a $\mathcal{E}_3^2(3)$ existuje přípustná tabulka $\mathcal{T}^3$ typu $m \times n \times o$ právě tehdy, když algoritmus z adresáře ***Rekonstrukce 3d tabulky*** z přiloženého CD aplikovaný na tabulku typu $m \times n \times o$ obsahující samé jedničky a na dané 2-marginály konverguje pro každý prvek tabulky.

Důkaz věty 6. : Důkaz je uveden na straně 266 v článku [8].

Zabývejme se nyní nalezením přípustného řešení planárního dopravního problému $\mathrm{DMKP}_{3,1}$. Zobecnění metoda severozápadního rohu obecně nefunguje, viz. následující příklad.

Uvažujme příklad kdy máme tabulku typu $4 \times 4 \times 4$. Nechť tato tabulka má následující 2-marginály:

$$\begin{pmatrix} 13 & 8 & 4 & 6 \\ 8 & 3 & 16 & 14 \\ 9 & 4 & 14 & 7 \\ 6 & 6 & 17 & 12 \end{pmatrix} \begin{pmatrix} 8 & 14 & 14 & 10 \\ 11 & 8 & 7 & 8 \\ 8 & 15 & 9 & 11 \\ 4 & 4 & 4 & 12 \end{pmatrix} \begin{pmatrix} 16 & 4 & 15 & 11 \\ 9 & 5 & 13 & 7 \\ 6 & 6 & 18 & 13 \\ 5 & 6 & 5 & 8 \end{pmatrix}$$

Taková tabulka jistě existuje. Kupříkladu je to tabulka:

$$\begin{pmatrix} 5 & 2 & 0 & 1 \\ 5 & 3 & 1 & 2 \\ 3 & 0 & 2 & 3 \\ 0 & 3 & 1 & 0 \end{pmatrix} \begin{pmatrix} 2 & 1 & 7 & 4 \\ 2 & 0 & 4 & 2 \\ 2 & 1 & 4 & 8 \\ 2 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 6 & 1 & 3 & 4 \\ 0 & 0 & 5 & 2 \\ 1 & 2 & 5 & 1 \\ 2 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 3 & 0 & 5 & 2 \\ 2 & 2 & 3 & 1 \\ 0 & 3 & 7 & 1 \\ 1 & 1 & 2 & 8 \end{pmatrix}.$$

Existuje tedy přípustné řešení, ale zobecněná metoda severozápadního rohu vede k následujícímu řešení, které přípustné není:

$$\begin{pmatrix} 8 & 0 & 0 & 0 \\ 5 & 5 & 1 & 0 \\ 0 & 3 & 3 & 2 \\ 0 & 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} 8 & 3 & 3 & 0 \\ 0 & 0 & 8 & 0 \\ 0 & 0 & 5 & 10 \\ 0 & 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} 0 & 1 & 12 & 1 \\ 4 & 0 & 2 & 1 \\ 5 & 3 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 10 \\ 0 & 0 & 2 & 2 \\ 1 & 0 & 10 & 0 \\ 5 & 6 & 1 & 0 \end{pmatrix}.$$

Že řešení není přípustné je vidět například z posledního sloupce třetí matice, který má dávat součet 7, ale dává součet 3.

Pokusím se zde navrhnout elementární algoritmus, který vede k nalezení přípustného řešení trojrozměrného planárního dopravního problému, pokud toto existuje. Pokud by přípustné řešení neexistovalo, nebo pokud by došlo k zacyklení algoritmu, skončí algorimus ohlášením, že přípustné řešení nenalezl. Zacyklení algoritmu v situaci, kdy přípustné řešení existuje, je velmi nepravděpodobné, nicméně nedokázal jsem formálně, že k němu nemůže dojít. K nalezení optimálního řešení můžeme použít simplexovou metodu, kterou spustíme s přípustným výchozím řešením nalezeným prostřednictvím tohoto algoritmu. Algoritmus je detailně popsán v adresáři ***Přípustné řešení planárního DP*** na přiloženém CD.

Mějme tedy dán problém nalezení přípustného řešení s rozměry $m \times n \times o$, $m, n, o \in \mathbb{N}$, $m, n, o > 1$. Mějme dány tři nezáporné 2-marginály, které označíme $M^{n \times o}$, $N^{m \times o}$ a $O^{m \times n}$. Algoritmus sestává ze tří fází. V první fázi "rozřežeme" tabulku na dvourozměrné "plátky" v libovolném směru a na každý z těchto "plátků" spustíme metodu severozápadního rohu. Dostaneme tedy přípustná řešení pro každý z "plátků" a tedy po tomto kroku budou jistě alespoň 2 z uvedených 2-marginálů tabulky odpovídat nalezenému řešení. Třetí 2-marginál nalezenému řešení obecně odpovídat nemusí. Po první fázi algoritmu tedy máme matice, tj. dvourozměrné tabulky, které můžeme poskládat do tabulky trojrozměrné, která splňuje alespoň dva ze tří daných 2-marginálů. Budeme dále bez újmy na obecnosti uvažovat, že jsme tabulku "rozřezali" tak, že nám vzniklo o "plátků" (dalšími možnostmi je vznik m či vznik n "plátků").

Ve druhé fázi algoritmu postupně upravujeme hodnoty matic nalezených v první fázi pomocí ε-změn tak, aby se jejich součet vyrovnal s příslušnou hodnotou nesplněného 2-marginálu. Začneme na místě $(1, 1)$ první matice a budeme se postupně posouvat po řádcích až k prvku $(m-1, n-1)$. V kroku, který bude odpovídat prvkům na místech (i, j) matic z první fáze nejprve položíme $k := 1$, poté provedeme požadovanou ε-změnu a zvýšíme k o jedničku a opět posoudíme použití požadované ε-změny a takto postupujeme až do okamžiku, kdy $k = o$. Poté se posuneme na prvky $(i, j+1)$, případně $(i+1, 1)$, pokud $j = n-1$. Požadovanou ε-změnu určíme podle toho, která z následujících tří možností nastane.

1. Je-li součet prvků na místech (i, j) z k-té až o-té matice (označme jej $S_{i,j}^k$) větší, než prvek na místě (i, j) z nesplněného 2-marginálu (označme jej $o_{i,j}$) a zároveň součet prvků na místech (i, j) $k+1$-ní až o-té matice (označme jej $S_{i,j}^{k+1}$) není větší, než prvek $o_{i,j}$, pak provedeme $\left(S_{i,j}^k - o_{i,j}\right)$-změnu na místech $(i, i+1)$ a $(j, j+1)$.

2. Je-li součet $S_{i,j}^k$ meší, než prvek $o_{i,j}$, pak provedeme $\left(o_{1i,j} - S_{i,j}^k\right)$-změnu na místech $(i, i+1)$ a $(j, j+1)$.

3. Je-li součet $S_{i,j}^{k+1}$ větší, než prvek $o_{i,j}$, pak provedeme $\left(S_{i,j}^k - S_{i,j}^{k+1}\right)$-

změnu na místech $(i, i+1)$ a $(j, j+1)$.

Po ukončení druhé fáze algoritmu sice jsou splněny všechny tři 2-marginály, ale obecně nemusí být všechny prvky nalezené tabulky nezáporné. Obsahuje-li nalezená tabulka pouze nezáporné prvky, jsme hotovi. Neobsahuje-li, pustíme se do třetí fáze algoritmu v níž aplikujeme trojrozměrnou ε-změnu. Tu jsme ale zatím nedefinovali, čili tak učiňme.

Definice 20. Nechť $\mathbf{X} \in \mathbb{R}^{m\times n\times o}$ je reálná trojrozměrná tabulka. **Trojrozměrnou (či třídimenzionální)** ε**-změnou** na místech (u, r), (v, s) a (w, t) nazveme funkci $f^{\varepsilon}_{(u,r),(v,s),(w,t)} : \mathbb{R}^{m\times n\times o} \longrightarrow \mathbb{R}^{m\times n\times o}$ takovou, že platí: $f^{\varepsilon}_{(u,r),(v,s),(w,t)}(\mathbf{X}) = \mathbf{Z}_{ijk}$, kde

$$z_{ijk} = x_{ijk} + \varepsilon \text{ pro } i = u, j = v, k = w,$$

$$z_{ijk} = x_{ijk} + \varepsilon \text{ pro } i = u, j = s, k = t,$$

$$z_{ijk} = x_{ijk} + \varepsilon \text{ pro } i = r, j = s, k = w,$$

$$z_{ijk} = x_{ijk} + \varepsilon \text{ pro } i = r, j = v, k = t,$$

$$z_{ijk} = x_{ijk} - \varepsilon \text{ pro } i = u, j = v, k = t,$$

$$z_{ijk} = x_{ijk} - \varepsilon \text{ pro } i = u, j = s, k = w,$$

$$z_{ijk} = x_{ijk} - \varepsilon \text{ pro } i = r, j = v, k = w,$$

$$z_{ijk} = x_{ijk} - \varepsilon \text{ pro } i = r, j = s, k = t,$$

a jinak pro všechny jiné indexy je $z_{ijk} = x_{ijk}$, kde $\varepsilon \in \mathbb{R} \setminus \{0\}$.

Je zřejmé, že aplikace trojrozměrné ε-změny na trojrozměrnou tabulku nezmění její 2-marginály a tedy pochopitelně ani 1-marginály.

Algoritmus pokračuje třetí fází, která spočívá ve vyhledávání a aplikování "vhodných" trojrozměrných ε-změn pro všechny záporné hodnoty v tabulce, která je výstupem druhé fáze algoritmu. Druhá fáze je konstruována tak, že záporné prvky mohou být v maticích pouze v posledním sloupci, nebo v posledním řádku. Najdeme tedy všechny záporné prvky tabulky a aplikujeme na ně vhodně zvolené ε-změny[10].

[10] Algoritmus postupuje tak, že najde nejvyšší možné ε ze všech trojrozměrných ε-změn takové, že žádný záporný prvek nepřibyde. Pokud je toto maximální ε ostře větší než nula, pak provede ε-změnu, která buď vede k úplné, nebo alespoň k částečné eliminaci záporného prvku. Pokud je $\varepsilon \leq 0$, pak není možné eliminovat záporný prvek přímo a tudíž se provede trojrozměrná ε-změna tak, aby na místě záporného prvku zůstala nula, čímž se záporný prvek přesune na jiné místo, či místa a na ně se posléze aplikuje stejný postup.

Postup tohoto algoritmu na konkrétním příkladu je uveden v adresáři ***Přípustné řešení planárního DP*** na přiloženém CD pod názvem ***Příklad*** .

Tento algoritmus zároveň dokazuje, že k libovolným třem nezáporným 2-rozměrným tabulkám splňujícím všechny podmínky (3.22), existuje vždy nějaká reálná 3-rozměrná tabulka, taková, že tyto tři tabulky jsou jejími 2-marginály. Nemusí však existovat tabulka nezáporná.

K dalšímu zkoumání třetí fáze algoritmu uveďme následující lemma.

Lemma 15. "O trojrozměrné ε-změně"
Nechť $\mathbf{X}$ a $\mathbf{Y}$ jsou reálné tabulky typu $m \times n \times o$ které mají stejné 2-marginály. Pak existuje taková posloupnost trojrozměrných ε-změn, že jejich postupnou aplikací na tabulku $\mathbf{X}$ získáme tabulku $\mathbf{Y}$.

Důkaz lemmatu 15. : Důkaz je na přiloženém CD v adresáři ***Důkazy*** pod názvem **"O trojrozměrné epsilon změně".**

Víme tedy, že existuje-li přípustné řešení, pak existuje posloupnost trojrozměrných ε-změn taková, že tabulku, která je výstupem druhé fáze algoritmu, lze převést na takovou tabulku, která je přípustným řešením, tj. která je nezáporná.

3.2.2 Optimální řešení

Problémem řešení vícerozměrného diskrétního dopravního problému se zabývá mimojiné článek [22]. Optimální řešení je za předpokladu existence řešení přípustného jistě možné nalézt pomocí simplexové metody, či metod podobných (zobecněné MODI metody, apod.), neboť jde o problém lineárního programování, nicméně jak jsme ukázali, máme jednak problém s nalezením přípustného řešení a tudíž první fáze simplexové metody, která hledá přípustné řešení, nemusí skončit a jednak jsou rozsah zápisu úlohy i jeho výpočetní složitost zbytečně veliké. Skutečnost, že trojrozměrný axiální dopravní problém je možné řešit zobecněním metody MODI (metody řádkových a sloupcových čísel) je uveden v článku [41].

Lemma 16. "O osmici nul"
Pro každý problém (DMKP) existuje takové optimální řešení $\mathbf{X} \in \mathbb{R}^{m \times n \times o}$, že

$$\begin{aligned} &\forall 1 \leq u \neq v \leq m, 1 \leq s \neq t \leq n, 1 \leq k \neq l \leq o : \\ &\quad x_{u,s,k} = 0 \vee x_{u,t,k} = 0 \vee x_{u,t,l} = 0 \vee x_{u,s,l} = 0 \vee \\ &\quad\quad x_{v,s,k} = 0 \vee x_{v,s,l} = 0 \vee x_{v,t,k} = 0 \vee x_{v,t,l} = 0. \end{aligned} \tag{3.23}$$

Důkaz lemmatu 16. : Důkaz je na přiloženém CD v adresáři ***Důkazy*** pod názvem **"O osmici nul".**

Aspekty řešení obecného n-rozměrného dopravního problému s danými 1-marginály se zabývají články [46] a [45]. V článku [22] autor ukazuje, že libovolný symetrický vícerozměrný dopravní problém může být vyřešen použitím vícerozměrného planárního dopravního problému. Z tohoto poznatku plyne, že pokud umíme řešit planární n-rozměrný dopravní problém, pak umíme řešit i libovolný jiný symetrický n-rozměrný dopravní problém a to pouze převedením na planární problém. Tento převod je uveden v obecném tvaru pro problémy $\mathrm{DMKP}_{n,q}$, kde $q = 1, \ldots, n-1$ v důkazu věty 2 v článku [22].

Aplikací vícerozměrného dopravního problému je kupříkladu nalezení optimálního školního rozvrhu, je-li dána množina T učitelů, množina C předmětů, množina R místností a množina H hodin v rozvrhu. Úlohou je nalézt takovou matici řešení $\mathbf{X}$ s prvky, které jsou z množiny $\{0, 1\}$, která minimalizuje

$$\sum_{t\in T}\sum_{c\in C}\sum_{r\in R}\sum_{h\in H} d_{tchr}x_{tchr}$$

za podmínek

$$\sum_{c\in C}\sum_{r\in R} x_{tchr} = v_{th},\ \ t\in T\ h\in H,$$

$$\sum_{t\in T}\sum_{r\in R} x_{tchr} = w_{ch},\ \ c\in C\ h\in H,$$

$$\sum_{c\in C}\sum_{t\in T} x_{tchr} = u_{rh},\ \ r\in R\ h\in H,$$

$$\sum_{h\in H}\sum_{r\in R} x_{tchr} = g_{tc},\ \ t\in T\ c\in C,$$

kde $\mathbf{V}$ je matice obsahující prvky $v_{th} = 1$, pokud učitel t je volný v hodině h a $v_{th} = 0$, pokud volný v hodině h není, $\mathbf{W}$ je matice obsahující prvky $w_{ch} = 1$, pokud předmět c je možné zařadit na hodinu h a $w_{ch} = 0$, pokud to možné není, $\mathbf{G}$ je matice přirozených čísel obsahující prvky g_{tc} reprezentující počet hodin, které má učitel t odučit v předmětu c a $\mathbf{U}$ je matice obsahující prvky $u_{rh} = 1$, pokud třída r je volná v hodině h a $u_{rh} = 0$, pokud volná není. Nejde sice o aplikaci v teorii pravděpodobnosti, ani ve statistice, ale poslouží nám k objasnění následující definice v níž definujeme obecnější vícerozměrný diskrétní dopravní problém, v němž omezení mohou být dána nejen $(n-1)$-marginály.

Lemma 17. "O převodu dopravního problému"
Nechť $\mathrm{DMKP}_{n,k}$, kde $n > 2$ a $1 < k < n$ je symetrický n-rozměrný diskrétní

dopravní problém. Jeho optimální řešení lze získat řešením vhodného problému typu $\mathrm{DMKP}_{n,1}$.

Důkaz lemmatu 17. : Důkaz je uveden v článku [22].

V článcích [50] a [51] autoři navrhují algoritmus pro řešení čtyřrozměrného diskrétního dopravního problému s kapacitními omezeními, provádějí jeho numerickou analýzu a testují rychlost výpočtu. V článku [34] autoři aplikují aproximační algoritmy k nalezení přibližného optimálního řešení při separovatelných nákladech.

Kapitola 4

Aplikace dopravního problému

V této kapitole se budu zabývat statistickým tříděním při minimalizaci absolutních odchylek, metrizací slabé topologie, rekonstrukcí kontingenční tabulky (dvou i vícerozměrné), a řízeným zaokrouhlováním ve statistice.

4.1 Třídění bez interakcí v L1 normě

V knize [1] je uvedena zajímavá aplikace lineárního programování ve statistice, kterou lze prostřednictvím teorie duality převést na řešení dopravního problému s kapacitními omezeními. Na diskrétní dvourozměrný dopravní problém s kapacitními omezeními vede v knize uvedený problém dvojného třídění bez interakcí při odhadu parametrů modelu pomocí minimalizace absolutních odchylek. V knize je detailně probrán model s dvěma faktory bez interakcí. Zde ukážeme, že je možné problém rozšířit na problém vícečetných třídění bez interakcí. Autorům zmíněné aplikace jde v zásadě o co největší zrychlení a zjednodušení algoritmu vedoucího k řešení problému dvojného třídění při minimalizaci součtu absolutních odchylek a ukazují, že jako nejvhodnější se jeví řešení dopravního problému s kapacitními omezeními.

Nejprve krátce uveďme, o co nám v analýze rozptylu dvojného třídění jde. Cílem analýzy rozptylu dvojného třídění bez interakcí je testování hypotéz o vlivu dvou znaků na vysvětlovanou proměnnou a odhad parametrů modelu daného rovnicí (4.1).

Nechť máme naměřeny hodnoty vysvětlované proměnné a ke každému pozorování máme dány dva znaky, podle nichž jsme schopni jednotlivá pozorování třídit (například máme pacienty, u nichž nás zajímá doba dožití od data diagnózy a jsme schopni je k datu určení diagnózy roztřídit do skupin podle pohlaví a podle stáří v letech). Zajímá nás, zda je možné prohlásit, že doba dožití závisí v nějakém smyslu na hodnotě prvního znaku, nebo na hodnotě druhého znaku.

Vysvětlovanou proměnnou označme Y. Situaci modelujeme rovnicí

$$Y_{ijp} = \mu + \alpha_i + \beta_j + \varepsilon_{ijp}, \tag{4.1}$$

kde $\varepsilon \sim N(0, \sigma^2 > 0)$ je náhodná složka, $i = 1, \ldots, I$, $j = 1, \ldots, J$ a $p = 1, \ldots, P_{ij}$ pro každou přípustnou dvojici (i, j). Pro jednoduchost zde budu uvažovat vyvážené třídění, tedy situaci, kdy $P_{ij} = P \in \mathbb{N}$ pro každou přípustnou dvojici (i, j). I značí počet stavů prvního faktoru (u pohlaví by bylo $I = 2$), J značí počet stavů druhého faktoru a P_{ij} značí počet pozorování s hodnotou prvního faktoru i a hodnotou druhého faktoru j. Označme pro zjednodušení $n = IJP$. Protože takto definovaný model obsahuje nadbytečné parametry, zavádějí se reparametrizační rovnice

$$\sum_{i=1}^{I} \alpha_i = 0, \tag{4.2}$$

$$\sum_{j=1}^{J} \beta j = 0. \tag{4.3}$$

Označme

$$\mathbf{X} = \begin{pmatrix} \mathbf{1} & \mathbf{X}_1 & \mathbf{U}_1 \\ \mathbf{1} & \mathbf{X}_1 & \mathbf{U}_2 \\ \vdots & \vdots & \vdots \\ \mathbf{1} & \mathbf{X}_1 & \mathbf{U}_J \\ \mathbf{1} & \mathbf{X}_2 & \mathbf{U}_1 \\ \mathbf{1} & \mathbf{X}_2 & \mathbf{U}_2 \\ \vdots & \vdots & \vdots \\ \mathbf{1} & \mathbf{X}_2 & \mathbf{U}_J \\ \ldots & \ldots & \ldots \\ \mathbf{1} & \mathbf{X}_I & \mathbf{U}_1 \\ \mathbf{1} & \mathbf{X}_I & \mathbf{U}_2 \\ \vdots & \vdots & \vdots \\ \mathbf{1} & \mathbf{X}_I & \mathbf{U}_J \end{pmatrix} \text{ a } \mathbf{Y} = \begin{pmatrix} \mathbf{V}_{11} \\ \mathbf{V}_{12} \\ \vdots \\ \mathbf{V}_{1J} \\ \mathbf{V}_{21} \\ \mathbf{V}_{22} \\ \vdots \\ \mathbf{V}_{2J} \\ \ldots \\ \mathbf{V}_{I1} \\ \mathbf{V}_{I2} \\ \vdots \\ \mathbf{V}_{IJ} \end{pmatrix}, \tag{4.4}$$

kde $\mathbf{1}$ je vektor jedniček typu $P \times 1$,

$$\mathbf{X}_i = \begin{pmatrix} 0 & \ldots & 0 & 1 & 0 & \ldots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & 0 & 1 & 0 & \ldots & 0 \end{pmatrix}, \tag{4.5}$$

je typu $P \times I$ a jedničky jsou pouze v i-tém sloupci,

$$\mathbf{U}_j = \begin{pmatrix} 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 1 & 0 & \dots & 0 \end{pmatrix}, \tag{4.6}$$

je typu $P \times J$ a jedničky jsou pouze v j-tém sloupci a

$$\mathbf{V}_{ij} = \begin{pmatrix} Y_{ij1} \\ \vdots \\ Y_{ijP} \end{pmatrix},$$

je typu $P \times 1$, pro všechna $i = 1, \dots, I$ a $j = 1, \dots, J$. Označme $\theta = (\mu, \alpha_1, \dots, \alpha_m, \beta_1, \dots, \beta_n)$. Maticový zápis modelu dvojného třídění bez interakcí (4.1) je nyní dán rovnicí $\mathbf{X}\theta^T = \mathbf{Y}$. Před reparametrizací je počet parametrů roven $I + J + 1$, po reparametrizaci $I + J - 1$. Matice $\mathbf{X}$ před reparametrizací je typu $n \times (I + J + 1)$ a nemá úplnou sloupcovou hodnost, neboť kupříkladu první sloupec matice $\mathbf{X}$ lze získat také součtem druhého až $I+1$-ního sloupce, nebo součtem $I+2$-hého až posledního sloupce matice $\mathbf{X}$. Matice $\mathbf{X}$ po reparametrizaci je typu $n \times (I + J - 1)$ a získáme ji kupříkladu vynecháním prvního a $I + 2$-hého sloupce. Matice po reparametrizaci již má plnou sloupcovou hodnost. Zajímá nás odhad parametrů modelu. Většinou je odhadujeme pomocí metody nejmenších čtverců, nicméně ta má několik podstatných nedostatků. Zejména je velmi citlivá na odlehlá pozorování[1]. Proto se občas užívá odhadu pomocí minimalizace absolutních odchylek.

Zabývejme se nyní odhadem parametrů μ, $\alpha = (\alpha_1, \dots, \alpha_m)$ a $\beta = (\beta_1, \dots, \beta_n)$ modelu (4.1) pomocí minimalizace absolutních odchylek

$$\min_{\mu,\alpha,\beta} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{p=1}^{P} |Y_{ijp} - \mu - \alpha_i - \beta_j| \,. \tag{4.7}$$

Označme $\tau_i = \mu + \alpha_i$. Tím provedeme první reparametrizaci a dostáváme pouze $I + J$ parametrů modelu[2]. Ekvivalentní zápis problému ve tvaru problému lineárního programování je

$$\min_{\mathbf{d}_1,\mathbf{d}_2} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{p=1}^{p} (d_{1ijp} + d_{2ijp}) \tag{4.8}$$

[1] V tomto případě pouze typu "outlier", nikoliv typu "laverage point".

[2] Hodnoty μ a α dopočítáme na závěr pomocí reparametrizační rovnice (4.2).

vzhledem k

$$\mathbf{X}' \begin{pmatrix} \tau_1 \\ \tau_2 \\ \vdots \\ \tau_I \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_J \end{pmatrix} + \mathbf{I}\mathbf{d}_1 - \mathbf{I}\mathbf{d}_2 = \mathbf{Y} \tag{4.9}$$

a $\mathbf{d}_1 \geq \mathbf{0}$ a $\mathbf{d}_2 \geq \mathbf{0}$, kde $\mathbf{d}_1 = (d_{1111}, \ldots, d_{1IJP})^T$ je vektor typu $n \times 1$, $\mathbf{d}_2 = (d_{2111}, \ldots, d_{2IJP})^T$ je rovněž vektor typu $n \times 1$ a $\mathbf{I}$ je jednotková matice typu $n \times n$ a

$$\mathbf{X}' = \begin{pmatrix} \mathbf{X}_1 & \mathbf{U}_1 \\ \mathbf{X}_1 & \mathbf{U}_2 \\ \vdots & \vdots \\ \mathbf{X}_1 & \mathbf{U}_J \\ \mathbf{X}_2 & \mathbf{U}_1 \\ \mathbf{X}_2 & \mathbf{U}_2 \\ \vdots & \vdots \\ \mathbf{X}_2 & \mathbf{U}_J \\ \cdots & \cdots \\ \mathbf{X}_I & \mathbf{U}_1 \\ \mathbf{X}_I & \mathbf{U}_2 \\ \vdots & \vdots \\ \mathbf{X}_I & \mathbf{U}_J \end{pmatrix}. \tag{4.10}$$

Označíme-li dále

$$\tilde{\mathbf{X}} = \left(\mathbf{X}' \mid \mathbf{I} \mid -\mathbf{I} \right) \in \mathbb{R}^{n \times (I+J+2n)}, \tag{4.11}$$

$$\lambda = (\tau_1, \ldots, \tau_I, \beta_1, \ldots, \beta_J, \mathbf{d}_1, \mathbf{d}_2), \ \lambda \in \mathbb{R}^{(I+J+2n) \times 1} \tag{4.12}$$

a

$$\mathbf{c} = \Bigg(\underbrace{0, \ldots, 0}_{I+J}, \underbrace{1, \ldots, 1}_{2n} \Bigg), \tag{4.13}$$

pak lze problém daný v (4.8) a (4.9) zapsat maticově

$$\min_{\lambda} \lambda \mathbf{c}^T \tag{4.14}$$

vzhledem k podmínkám
$$\tilde{\mathbf{X}}\lambda^T = \mathbf{Y} \tag{4.15}$$
a $\mathbf{d}_1 \geq \mathbf{0}$ a $\mathbf{d}_2 \geq \mathbf{0}$, kde $\mathbf{0}$ je nulový vektor typu $n \times 1$.

Budeme nyní procházet jednotlivé kroky simplexové metody. Uvažujme nějakou bazi $\mathbf{B}$. Nejvýše $I + J - 1$ proměnných z množiny $\{\tau_1, \ldots, \tau_I, \beta_1, \ldots, \beta_J\}$ může být bazických. Máme-li bazi, můžeme rozdělit řádky matice $\mathbf{X}'$ do tří skupin takto:

$$J_0 := \{(i,j,k) : d_{1ijk}, d_{2ijk} \text{ jsou nebazické}\} \tag{4.16}$$
$$J_1 := \{(i,j,k) : d_{1ijk} \text{ je bazické}\} \tag{4.17}$$
$$J_2 := \{(i,j,k) : d_{2ijk} \text{ je bazické}\}, \tag{4.18}$$

kde (i,j,k) je uspořádaná trojice indexů odpovídající řádku $\tau_i + \beta j + d_{1ijk} - d_{2ijk} = Y_{ijk}$, tj. $(i-1)JP + (j-1)P + k$-tému řádku v soustavě rovnic (4.9). Přeházením řádků matice $\mathbf{X}'$ tak, aby byly všechny řádky, které odpovídají trojici indexů zařazené do skupiny J_0 nad všemi řádky, které odpovídají trojici indexů, která do J_0 nenáleží a proházením sloupců tak, aby všechny, které odpovídají bazickým proměnným, byly vlevo od každého, který odpovídá nebazické proměnné, získáme matici $\bar{\mathbf{X}}$. Matici $\mathbf{X}_{(1)}$ dostáváme nyní jako podmatici matice $\bar{\mathbf{X}}$ s levým horním rohem na místě $(1,1)$ a pravým dolním rohem na místě $(I+J-1, I+J-1)$ a matici $\mathbf{X}_{(2)}$ dostáváme jako podmatici matice $\bar{\mathbf{X}}$ s levým horním rohem na místě $(I+J, 1)$ a pravým dolním rohem na místě $(n, I+J-1)$. Bazi $\mathbf{B}$ můžeme zapsat takto

$$\mathbf{B} = \begin{pmatrix} \mathbf{X}_{(1)} & \mathbf{0} \\ \mathbf{X}_{(2)} & \mathbf{D} \end{pmatrix}, \tag{4.19}$$

kde $\mathbf{X}_{(1)}$ je regulární matice typu $(I+J-1) \times (I+J-1)$, $\mathbf{0}$ je nulová matice typu $(I+J-1) \times (n-I-J+1)$, $\mathbf{D}$ je diagonální matice typu $(n-I-J+1) \times (n-I-J+1)$, která má na diagonále pouze prvky z množiny $\{-1, 1\}$. Matice $\mathbf{X}_{(1)}$ je navíc trojúhelníková. Proto matici $\mathbf{B}^{-1}$ můžeme zapsat jako

$$\mathbf{B}^{-1} = \begin{pmatrix} \mathbf{X}_{(1)}^{-1} & \mathbf{0} \\ -\mathbf{D}\mathbf{X}_{(2)}\mathbf{X}_{(1)}^{-1} & \mathbf{D} \end{pmatrix}. \tag{4.20}$$

Invertibilita matice $\mathbf{B}$ je zřejmá, neboť matice $\mathbf{X}_{(1)}$ je regulární trojúhelníková a matice $\mathbf{D}$ je diagonální (determinant matice $\mathbf{B}$ je tudíž nenulový). Všimněme si, že matice $\mathbf{X}^{(1)}$ i matice $\mathbf{X}^{(2)}$ mají o jeden sloupec méně, než matice $\bar{\mathbf{X}}$. Tím je provedena druhá reparametrizace. Hodnoty λ_i příslušné bazi získáme Gaussovou eliminací. Navíc platí, že $d_{1ijk}d_{2ijk} = 0$ pro všechny přípustné trojice (i, j, k) a $d_{1ijk} \geq 0$ a $d_{2ijk} \geq 0$. Pro všechny řádky s indexy (i, j, k) z množiny J_0 jsou hodnoty příslušných d_{1ijk} i d_{2ijk} rovny nule. Podmínku optimality (podmínka optimality simplexového algoritmu, viz. [38], věta 173, str. 168) je možné zapsat ve tvaru

$$-\tilde{\mathbf{e}}^T \leq \mathbf{e}^T \mathbf{D} \mathbf{X}_{(2)} \mathbf{X}_{(1)}^{-1} \leq \tilde{\mathbf{e}}^T, \tag{4.21}$$

kde $\mathbf{e} = (1, \ldots, 1)^T$ je typu $1 \times (n - I - J + 1)$ a $\tilde{\mathbf{e}} = (1, \ldots, 1)^T$ je typu $1 \times (I + J - 1)$. K řešení se ukazuje být nejvhodnější použití duální úlohy k úloze (4.8). Úloha duální k této úloze je dopravním problémem s kapacitními omezeními. Maticový zápis duální úlohy je

$$\max_{\pi} \mathbf{Y}^T \pi \tag{4.22}$$

$$\text{vzhledem k } \mathbf{X}'^T \pi = \mathbf{0} \tag{4.23}$$

$$\text{a } -\mathbf{e}'^T \leq \pi^T \leq \mathbf{e}'^T, \tag{4.24}$$

kde $\pi = (\pi_{111}, \ldots, \pi_{IJP})$ a $\mathbf{e}'^T = (1, \ldots, 1)$ je typu $1 \times n$.

Z teorie duality víme, že $\mathbf{c}_B \mathbf{B}^{-1}$, kde $\mathbf{c}_B$ jsou prvky $\mathbf{c}$ odpovídající bazickým proměnným, dává duální proměnné příslušné bazi $\mathbf{B}$ původního problému. $\mathbf{c}_B \mathbf{B}^{-1}$ splňuje podmínku (4.23), ale nemusí obecně splňovat podmínku (4.24). Pokud $\mathbf{B}$ není optimální baze původního problému, pak dostáváme řešení, které podmínku (4.21) nesplňuje a nejde tedy o přípustné řešení duální úlohy a není tedy splněna podmínka (4.24). Z podmínky optimality (4.21) plyne snadno nezápornost redukovaných nákladů RN_{1ijk} a RN_{2ijk} definovaných v prvním kroku simplexového algoritmu takto:

$$RN_{1ijk} := 1 + \mathbf{e}^T \mathbf{D} \mathbf{X}_{(2)} \mathbf{X}_{(1)}^{-1} \mathbf{e}_p \tag{4.25}$$

pro každé $d_{1ijk}, (i, j, k) \in J_0$ a

$$RN_{2ijk} := 1 - \mathbf{e}^T \mathbf{D} \mathbf{X}_{(2)} \mathbf{X}_{(1)}^{-1} \mathbf{e}_p \tag{4.26}$$

pro každé $d_{2ijk}, (i, j, k) \in J_0$, kde $\mathbf{e}_p$ je p-tý sloupec jednotkové matice hodnosti $I + J - 1$ a p je index řádku matice $\mathbf{X}_{(1)}$ odpovídajícího trojici indexů (i, j, k). Protože

$$\mathbf{e}^T \mathbf{D} \mathbf{X}_{(2)} \mathbf{X}_{(1)}^{-1}$$

je řádkový sloupec duálních proměnných π_k, které odpovídají řádkovým indexům v J_0, mohou být hodnoty těchto proměnných snadno dopočítány pomocí Gaussovy eliminace, která se navíc pro trojúhelníkovou matici $X_{(1)}$ zjednoduší na pouhé dosazování do rovnic. Jakmile jsou hodnoty těchto duálních proměnných spočteny, použijí se na výpočet (4.25) a (4.26), přičemž součet (4.25) a (4.26) je vždy roven 2 a stačí tedy vždy spočítat pouze jednu z těchto hodnot. Pokud je $RN_{1ijk} \geq 0$ pro všechna $d_{1ijk}, (i,j,k) \in J_0$ a $RN_{2ijk} \geq 0$ pro všechna $d_{2ijk}, (i,j,k) \in J_0$, máme optimální bazi.

Umíme tedy najít přípustnou bazi a umíme snadno ověřit její optimalitu. Ke spuštění simplexové metody potřebujeme umět v každém kroku vybrat nebazickou proměnnou, kterou vložíme do nové baze, pokud ta nalezená ještě není optimální a potřebujeme vědět, kterou proměnnou z baze vypustit. V každém kroku zařadíme tu nebazickou proměnnou, jejíž redukovaný náklad je nejnižší ze všech redukovaných nákladů odpovídajících nebazickým proměnným (zápornou, jinak by stávající baze byla optimální). Předpokládejme tedy, že máme bazi $\mathbf{B}$, podmínka optimality (4.21) není splněna a určili jsme, že do baze vložíme proměnnou odpovídající p-tému řádku matice $\mathbf{X}^{(1)}$. Nyní už musíme pouze rozhodnout, který vektor z baze vyloučíme. K tomu použijeme následující postup. Spočítáme

$$\arg\min_s \frac{|r(s)|}{\pm d_{ss}\mathbf{X}_{(2)(s)}\mathbf{X}^{-1}_{(1)(p)}}, \tag{4.27}$$

kde minimalizace jde přes všechna s taková, že jmenovatel je kladný, $r(s)$ je reziduum odpovídající s-tému řádku matice $\mathbf{X}_{(2)}$, d_{ss} je diagonální prvek matice $\mathbf{D}$, $\mathbf{X}_{(2)(s)}$ je s-tý řádek matice $\mathbf{X}_{(2)}$ a $\mathbf{X}_{(1)(p)}$ je p-tý sloupec matice $X^{-1}_{(1)}$ odpovídající do baze vstupující proměnné. Pokud je to d_{1ijk} bereme ve jmenovateli (4.27) mínus, je-li do baze vstupující proměnná d_{2ijk} bereme ve jmenovateli (4.27) plus. Pokud minimum redukovaného nákladu je dosaženo pro vstupní proměnnou d_{1tuv}, resp. d_{2tuv}, pak změníme redukované náklady RN_{1utv}, resp. RN_{2utv} odpovídající této proměnné přičtením $2\left|\mathbf{X}_{(2)(s)}\mathbf{X}^{-1}_{(1)(p)}\right|$. Pokud je po přičtení této hodnoty k aktuálnímu redukovanému nákladu výsledná hodnota stále záporná, vezmeme v (4.27) místo minimální hodnoty, druhou nejmenší a postupujeme takto tak dlouho, až je nový redukovaný náklad kladný. Tím je postup simplexové metody zcela dán. Alternativou je potom řešení zmíněného dopravního problému s kapacitními omezeními, které získáme kupříkladu použitím Ford-Fulkersonova algoritmu.

Ukažme popsaný algoritmus na konkrétním příkladu. Mějme 30 hypotečních úvěrů, které byly ukončeny v důsledku nesplácení. Nechť 10 z těchto hypoték mělo počáteční LTV[3] do 70% nevčetně, 10 s počátečním LTV z intervalu [70%, 100%) a 10 s počátečním LTV nejméně 100%. Nechť 5 z každé z těchto skupin jsou hypotéky s jedním žadatelem a pět s více

[3]"Loan to Value" je nominální hodnota úvěru k datu podpisu smlouvy o úvěru vydělená oceněnou hodnotou nemovitostního zajištění platnou k témuž datu.

než jedním žadatelem. Protože hypoteční úvěry byly již ukončeny, máme ke každé napočtenu hodnotu LGD[4], viz. tabulka 4.1.

žadatel	$LTV < 70\%$	$70\% <= LTV < 100\%$	$LTV >= 100\%$
1	0.00	0.25	0.81
	0.04	0.11	0.85
	0.02	0.54	1.00
	0.12	0.25	0.95
	0.08	0.31	0.62
> 1	0.01	0.24	0.78
	0.00	0.21	0.68
	0.05	0.19	0.97
	0.02	0.09	0.81
	0.09	0.14	0.78

Tabulka 4.1: Data k příkladu o hypotékách

Máme minimalizovat

$$\sum_{i=1}^{3}\sum_{j=1}^{2}\sum_{k=1}^{5}(d_{1ijk} + d_{2ijk})$$

[4]"Loss Given Default" je ztráta podmíněná defaultem, která je definována v pravidlech obezřetného podnikání bank daných v ČR vyhláškou ČNB 123/2007 Sb. Tato pravidla jsou odvozena na základě pravidel Basel II vydaných Basilejskou komisí pro bankovní dohled.

za podmínek

$$\begin{aligned}
\tau_1 + \beta_1 + d_{1111} - d_{2111} &= 0.00 \\
\tau_1 + \beta_1 + d_{1112} - d_{2112} &= 0.04 \\
\tau_1 + \beta_1 + d_{1113} - d_{2113} &= 0.02 \\
\tau_1 + \beta_1 + d_{1114} - d_{2114} &= 0.12 \\
\tau_1 + \beta_1 + d_{1115} - d_{2115} &= 0.08 \\
\tau_1 + \beta_2 + d_{1121} - d_{2121} &= 0.01 \\
\tau_1 + \beta_2 + d_{1122} - d_{2122} &= 0.00 \\
\tau_1 + \beta_2 + d_{1123} - d_{2123} &= 0.05 \\
\tau_1 + \beta_2 + d_{1124} - d_{2124} &= 0.02 \\
\tau_1 + \beta_2 + d_{1125} - d_{2125} &= 0.09 \\
\tau_2 + \beta_1 + d_{1211} - d_{2211} &= 0.25 \\
\tau_2 + \beta_1 + d_{1212} - d_{2212} &= 0.11 \\
\tau_2 + \beta_1 + d_{1213} - d_{2213} &= 0.54 \\
\tau_2 + \beta_1 + d_{1214} - d_{2214} &= 0.25 \\
\tau_2 + \beta_1 + d_{1215} - d_{2215} &= 0.31 \\
\tau_2 + \beta_2 + d_{1221} - d_{2221} &= 0.24 \\
\tau_2 + \beta_2 + d_{1222} - d_{2222} &= 0.21 \\
\tau_2 + \beta_2 + d_{1223} - d_{2223} &= 0.19 \\
\tau_2 + \beta_2 + d_{1224} - d_{2224} &= 0.09 \\
\tau_2 + \beta_2 + d_{1225} - d_{2225} &= 0.14 \\
\tau_3 + \beta_1 + d_{1311} - d_{2311} &= 0.81 \\
\tau_3 + \beta_1 + d_{1312} - d_{2312} &= 0.85 \\
\tau_3 + \beta_1 + d_{1313} - d_{2313} &= 1.00 \\
\tau_3 + \beta_1 + d_{1314} - d_{2314} &= 0.95 \\
\tau_3 + \beta_1 + d_{1315} - d_{2315} &= 0.62 \\
\tau_3 + \beta_2 + d_{1321} - d_{2321} &= 0.78 \\
\tau_3 + \beta_2 + d_{1322} - d_{2322} &= 0.68 \\
\tau_3 + \beta_2 + d_{1323} - d_{2323} &= 0.97 \\
\tau_3 + \beta_2 + d_{1324} - d_{2324} &= 0.81 \\
\tau_3 + \beta_2 + d_{1325} - d_{2325} &= 0.78
\end{aligned} \tag{4.28}$$

a podmínek $d_{1ijk} \geq 0$ a $d_{2ijk} \geq 0$ pro všechna přípustná (i, j, k). Maticově lze zapsat tyto podmínky s užitím (4.11), kde

$$\mathbf{X}' = \begin{pmatrix} \mathbf{X}_1 & \mathbf{U}_1 \\ \mathbf{X}_1 & \mathbf{U}_2 \\ \mathbf{X}_2 & \mathbf{U}_1 \\ \mathbf{X}_2 & \mathbf{U}_2 \\ \mathbf{X}_3 & \mathbf{U}_1 \\ \mathbf{X}_3 & \mathbf{U}_2 \end{pmatrix},$$

$\mathbf{X}_i$, pro $i = 1,2,3$ jsou typu 5×3 dané předpisem (4.5) a $\mathbf{U}_j$, pro $j = 1,2$ jsou typu 5×2 dané předpisem (4.6) a $\mathbf{I}$ i $-\mathbf{I}$ v (4.11) jsou typu 30×30. Označíme-li dále

$$\lambda = (\tau_1, \tau_2, \tau_3, \beta_1, \beta_2, \mathbf{d}_1, \mathbf{d}_2)\,,$$

kde

$$\mathbf{d}_1 = (d_{1111}, d_{1112}, \ldots, d_{1324}, d_{1325})\,,$$

$$\mathbf{d}_2 = (d_{2111}, d_{2112}, \ldots, d_{2324}, d_{2325})$$

a $\mathbf{Y} = (0.00, 0.04, 0.02, 0.12, \ldots, 0.97, 0.81, 0.78)^T$, můžeme zapsat omezení (4.28) maticově

$$\tilde{\mathbf{X}}\lambda^T = \mathbf{Y}.$$

Najdeme optimální řešení použitím popsaného postupu.

Do množiny J_0 zahrneme kupříkladu řádky $(1,1,1)$, $(2,2,5)$, $(3,1,5)$ a $(2,1,5)$. Tomu odpovídá matice

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

jenž má hodnost 4. Poslední sloupec je lineárně závislý na předchozích čtyřech a tudíž jej z matice vypustíme (tím je provedena reparametrizace). Dostáváme matici

$$\mathbf{X}_{(1)} = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix},$$

matici

$$\mathbf{X}_{(2)} = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

a diagonální matici $\mathbf{D}$ typu 26×26 s prvky z množiny $\{-1, 1\}$. Dostáváme soustavu následujících rovnic

$$\begin{array}{ccccccccc} \lambda_1 & + & & & & & \lambda_4 & = & 0 \\ & & \lambda_2 & & & & & = & 0.14 \\ & & & & \lambda_3 & + & \lambda_4 & = & 0.62 \\ & & \lambda_2 & + & & + & \lambda_4 & = & 0.31, \end{array}$$

kterou vyřešíme bez aplikace Gaussovy eliminace prostým postupným dosazováním. Dostáváme $\lambda_2 = 0.14$, $\lambda_4 = 0.17$, $\lambda_1 = -0.17$ a $\lambda_3 = 0.45$. Nyní máme přípustnou bazi a zjišťujeme platnost podmínky optimality (4.21).

Spočítáme $\mathbf{X}_{(1)}^{-1}$. Ta je rovna

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

Matice $\mathbf{D}$ má na diagonále po řadě prvky

$$(1,1,1,1,1,1,1,1,1,1,1,1,1,-1,1,1,1,1,1,1,1,1,-1,-1,1,-1)$$

a tudíž

$$\mathbf{e}^T\mathbf{D}\mathbf{X}_{(2)}\mathbf{X}_{(1)}^{-1} = (9,12,9,-12)\,.$$

Podmínka optimality tedy není splněna a musíme provést výměnu baze. Hodnota účelové funkce, neboli součet absolutních hodnot všech reziduí je pro tuto bazi rovna 5. Duální úloha má tvar

$$\max_{\pi} \mathbf{Y}^T\pi$$

$$\text{vzhledem k } \mathbf{X}'^T\pi = 0$$

$$-\mathbf{e}^T \leq \pi^T \leq \mathbf{e}^T,$$

kde $\mathbf{e} = (1,\ldots,1)$ je vektor jedniček typu 1×30. Duální úloha má přípustné řešení.

$$(9,12,9,-12) = (\pi_{111},\pi_{225},\pi_{315},\pi_{215}) \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix}.$$

Z toho ihned plyne $\pi_{111} = 9$, $\pi_{225} = 12$, $\pi_{315} = 9$ a $\pi_{215} = -12$. Spočítáme redukované náklady

$$C_{1,111} = 1 + 9 = 10 \text{ a } C_{2,111} = 1 - 9 = -8,$$

$$C_{1,225} = 1 + 12 = 13 \text{ a } C_{2,225} = 1 - 12 = -11,$$

$$C_{1,315} = 1 + 9 = 10 \text{ a } C_{2,315} = 1 - 9 = -8,$$

$$C_{1,215} = 1 - 12 = -11 \text{ a } C_{2,215} = 1 + 12 = 13.$$

Nyní najděme proměnnou, kterou do baze zařadíme a proměnnou, kterou vyřadíme. Proměnnou, kterou do baze zařadíme, bude proměnná odpovídající minimu z redukovaných nákladů, což je -11 a tudíž budeme zařazovat proměnnou d_{2225}. Abychom zjistili, kterou proměnnou z baze vyřadíme, spočteme dle (4.27)

$$\arg\min_s \frac{|r(s)|}{\pm d_{ss}\mathbf{X}_{(2)(s)}\mathbf{X}_{(1)(p)}^{-1}}$$

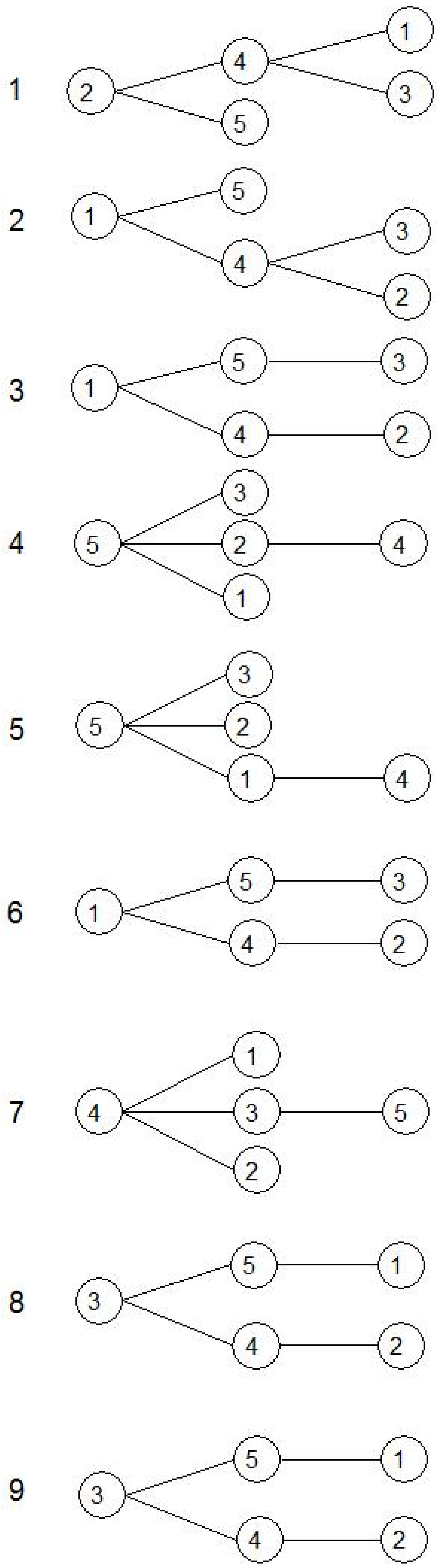
pro jednotlivá s. Nechť tedy $s = 1$. Pak dostáváme $d_{11} = 1$, $\mathbf{X}_{(2)(1)}\mathbf{X}_{(1)}^{-1} = (0, 0, 1, 1)$ a $r(1) = -0.19$. Zařazujeme proměnnou d_{2225} a tudíž $p = 2$. Hodnotu podílového kritéria označíme NO, neboť jmenovatel je nulový. Podobně pro $s = 2, \ldots, 26$, dostáváme hodnoty podílového kritéria postupně $NO, NO, NO, NO, 0.18, 0.17, 0.22, 0.19, 0.26, 0.1, 0.07, 0.05, NO, NO, 0.33, 0.23, 0.52, 0.36, 0.33, NO, NO, NO, NO, NO, NO$, kde NO značí, že pro dané s je jmenovatel v podílovém kritériu roven nule a tudíž proměnnou odpovídající tomuto řádku nevyřazujeme. V prvním kroku algoritmu vyměníme proměnnou, která přísluší třináctému řádku matice $X_{(2)}$, kde hodnoty podílového kritéria nabývají minimální kladné hodnoty.

Spustíme-li tento algoritmus, pak posloupností výměn bazických proměnných za nebazické dospějeme k řešení původní úlohy dvojného třídění. Hodnota účelové funkce přitom postupně klesá: 5, 3.67, 2.69, 2.16, 2.03, 1.99, 1.95, 1.93, 1.92, a 1.91, což již je globální minimum. Algoritmus je velmi rychlý. Odhadnuté parametry modelu dvojného třídění jsou pro náš příklad: $\mu = 0.36$, $\alpha_1 = -0.32$, $\alpha_2 = -0.14$, $\alpha_3 = 0.46$, $\beta_1 = 0.03$ a $\beta_2 = -0.03$.

V knize [1] autoři uvádějí zajímavou vlastnost množiny indexů J_0. Přiřadíme-li každé hodnotě prvního faktoru nabývajícího n_1 hodnot indexy z množiny $N_1 = \{1, \ldots, n_1\}$ tak, že každé hodnotě faktoru přísluší právě jeden index z této množiny a každé hodnotě druhého faktoru nabývajícího n_2 hodnot indexy z množiny $N_2 = \{n_1 + 1, \ldots, n_1 + n_2\}$ rovněž jednoznačně, pak stačí uvažovat pouze některé množiny indexů J_0. Vytvoříme-li graf tak, že každému z indexů množiny N_1 a množiny N_2 přiřadíme vrchol a každému řádku matice $\mathbf{X}_{(1)}$ přiřadíme hranu mezi těmi vrcholy, které odpovídají tomuto řádku, pak v algoritmu stačí uvažovat pouze takové množiny indexů J_0, které odpovídají maticím $\mathbf{X}_{(1)}$, pro něž takto vytvořený graf je stromem.

Posloupnost stromů, které byly vygenerovány algoritmem při řešení našeho problému s hypotékami je zobrazena na obrázku 4.1.

Na závěr tohoto příkladu ještě uveďme, že robustní odhady parametrů regrese se za přítomnosti odlehlých pozorování ukazují jako mnohem lepší, než odhady pomocí metody nejmenších čtverců, viz. např. [40]. Značný počet článků rozvíjí metody používané k řešení úlohy minimalizace absolutních odchylek pomocí algoritmů neuronových sítí a evolučních algoritmů. Jedná se pochopitelně o heuristické algoritmy, které trpí klasickými neduhy použitých metod, ale při problémech s větším rozsahem jsou často dostatečně dobrou



Obrázek 4.1: Grafy k problému s hypotékami

alternativou.

Aplikace v knize [1] se omezuje pouze na dvojné třídění bez interakcí, nicméně problém může být rozšířen i na vícečetná třídění bez interakcí. Jsou dvě možnosti jak k problému vícečetného třídění přistoupit. Buď odvodíme celý postup minimalizace, včetně odvození dopravního problému s kapacitními omezeními pro model (4.29), nebo prostřednictvím (4.32) převedeme model vícečetného třídění na model dvojného třídění. Uvažujme pětičetné třídění. Vysvětlovanou proměnnou označme opět Y. Situaci pětičetného třídění modelujeme rovnicí

$$Y_{ijklmp} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + \kappa_m + \varepsilon_{ijklmp}, \tag{4.29}$$

kde $\varepsilon \sim N(0, \sigma^2 > 0)$ je náhodná složka, $i = 1, \ldots, I$, $j = 1, \ldots, J$, $k = 1, \ldots, K$, $l = 1, \ldots, L$, $m = 1, \ldots, M$ a $p = 1, \ldots, P$.[5] Označme $n = IJKLMP$.

Problém lineárního programování příslušný problému (4.29) má tvar

$$\min_{d_{1qrstuv}, d_{2qrstuv}} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{l=1}^{L} \sum_{m=1}^{M} \sum_{p=1}^{P} (d_{1ijklmp} + d_{2ijklmp}) \tag{4.30}$$

vzhledem k

$$\tilde{\mathbf{X}} \lambda^T = \mathbf{Y}, \tag{4.31}$$

kde

$$\tilde{\mathbf{X}} = (\mathbf{X}' \mid \mathbf{I} \mid -\mathbf{I}),$$

[5] Model je tedy vyvážený.

$$\mathbf{X}' = \begin{pmatrix} \mathbf{X}_1 & \mathbf{U}_1 & \mathbf{Z}_1 & \mathbf{W}_1 & \mathbf{V}_1 \\ \mathbf{X}_1 & \mathbf{U}_1 & \mathbf{Z}_1 & \mathbf{W}_1 & \mathbf{V}_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{X}_1 & \mathbf{U}_1 & \mathbf{Z}_1 & \mathbf{W}_1 & \mathbf{V}_M \\ \mathbf{X}_1 & \mathbf{U}_1 & \mathbf{Z}_1 & \mathbf{W}_2 & \mathbf{V}_1 \\ \mathbf{X}_1 & \mathbf{U}_1 & \mathbf{Z}_1 & \mathbf{W}_2 & \mathbf{V}_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{X}_1 & \mathbf{U}_1 & \mathbf{Z}_1 & \mathbf{W}_2 & \mathbf{V}_M \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \mathbf{X}_1 & \mathbf{U}_1 & \mathbf{Z}_1 & \mathbf{W}_L & \mathbf{V}_1 \\ \mathbf{X}_1 & \mathbf{U}_1 & \mathbf{Z}_1 & \mathbf{W}_L & \mathbf{V}_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{X}_1 & \mathbf{U}_1 & \mathbf{Z}_1 & \mathbf{W}_L & \mathbf{V}_M \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \mathbf{X}_I & \mathbf{U}_J & \mathbf{Z}_K & \mathbf{W}_L & \mathbf{V}_1 \\ \mathbf{X}_I & \mathbf{U}_J & \mathbf{Z}_K & \mathbf{W}_L & \mathbf{V}_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{X}_I & \mathbf{U}_J & \mathbf{Z}_K & \mathbf{W}_L & \mathbf{V}_M \end{pmatrix},$$

$\mathbf{I}$ je jednotková matice typu $n \times n$,

$$\lambda = (\tau_1, \dots, \tau_I, \beta_1, \dots, \beta_J, \gamma_1, \dots, \gamma_K, \delta_1, \dots, \delta_L, \kappa_1, \dots, \kappa_M, \mathbf{d}_1, \mathbf{d}_2)$$

$$\mathbf{Y} = (\mathbf{Y}_{11111}, \dots, \mathbf{Y}_{1111M}, \mathbf{Y}_{11121}, \dots, \mathbf{Y}_{1112M}, , \mathbf{Y}_{IJKL1}, \dots, \mathbf{Y}_{IJKLM})^T,$$

kde

$$\mathbf{X}_i = \begin{pmatrix} 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 1 & 0 & \dots & 0 \end{pmatrix},$$

je typu $P \times I$ a jedničky jsou pouze v i-tém sloupci,

$$\mathbf{U}_j = \begin{pmatrix} 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 1 & 0 & \dots & 0 \end{pmatrix},$$

je typu $P \times J$ a jedničky jsou pouze v j-tém sloupci,

$$\mathbf{Z}_k = \begin{pmatrix} 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 1 & 0 & \dots & 0 \end{pmatrix},$$

je typu $P \times K$ a jedničky jsou pouze v k-tém sloupci,

$$\mathbf{W}_l = \begin{pmatrix} 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 1 & 0 & \dots & 0 \end{pmatrix},$$

je typu $P \times L$ a jedničky jsou pouze v l-tém sloupci,

$$\mathbf{V}_m = \begin{pmatrix} 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 1 & 0 & \dots & 0 \end{pmatrix},$$

je typu $P \times M$ a jedničky jsou pouze v m-tém sloupci a

$$\mathbf{Y}_{ijklm} = \begin{pmatrix} Y_{ijklm1} \\ \vdots \\ Y_{ijklmP} \end{pmatrix},$$

je typu $P \times 1$ a $\mathbf{d}_1 = (d_{1111111}, \dots, d_{1IJKLMP})$ je vektor typu $n \times 1$ a $\mathbf{d}_2 = (d_{2111111}, \dots, d_{2IJKLMP})$ je rovněž vektor typu $n \times 1$ a vzhledem k $\mathbf{d}_1 \geq \mathbf{0}$ a $\mathbf{d}_2 \geq \mathbf{0}$, kde $\mathbf{0}$ je nulový vektor typu $1 \times n$.

Zcela analogicky jako v případě dvojného třídění bez interakcí můžeme z matice $\mathbf{X}'$ vyjmout čtyři sloupce tak, že výsledná matice má plnou sloupcovou hodnost a určit množiny pětic indexů J_0, J_1 a J_2 a jim příslušné podmatice $\mathbf{X}_{(1)}$ a $\mathbf{X}_{(2)}$ přeuspořádané matice $\mathbf{X}'$. K danému problému pak přísluší duální úloha ve tvaru

$$\max_{\pi} \mathbf{Y}^T \pi$$

$$\text{vzhledem k } \mathbf{X}'^T \pi = 0$$

$$-\mathbf{e}^T \leq \pi^T \leq \mathbf{e}^T,$$

kde $\mathbf{e}^T$ je vektor jedniček typu $1 \times n$. Nebudu zde již detailně procházet algoritmus a uvedu raději příklad. Ten je dán tabulkou 4.2.

Jde o data pacientů u nichž se zjišťuje hodnota pěti faktorových proměnných. Prvním faktorem je pohlaví pacienta, druhým věková skupina, třetím odpověď na otázku, zda bydlí ve městě nad 100 000 obyvatel, čtvrtým kuřáctví a pátým konzumace alkoholu. Vysvětlovaná proměnná má celočíselnou hodnotu[6]. Vyberme do množiny J_0 následující šestice indexů $(1, 1, 1, 1, 1, 3)$,

[6] Nejde mi zde o konkrétní data a odhadování kauzálních závislostí, nýbrž pouze o teoretickou aplikaci na situaci vícečetného třídění. K tomuto účelu jsem data vytvořil uměle a nemají tedy žádnou konkrétní interpretaci.

M	old	yes	smoker	alcohol	96	M	old	yes	nonsmoker	no alcohol	25
M	old	yes	smoker	no alcohol	76	M	old	yes	nonsmoker	alcohol	43
M	old	no	nonsmoker	alcohol	2	M	old	no	smoker	no alcohol	20
M	old	no	nonsmoker	no alcohol	64	M	old	no	nonsmoker	alcohol	27
M	old	no	smoker	alcohol	65	M	middle	yes	smoker	alcohol	97
M	middle	yes	nonsmoker	no alcohol	45	M	middle	yes	smoker	no alcohol	58
M	middle	yes	nonsmoker	alcohol	7	M	middle	no	smoker	alcohol	18
M	middle	no	nonsmoker	no alcohol	25	M	middle	no	nonsmoker	alcohol	66
M	middle	no	smoker	no alcohol	94	M	young	yes	smoker	no alcohol	67
M	young	yes	nonsmoker	no alcohol	98	M	young	yes	smoker	alcohol	38
M	young	yes	nonsmoker	alcohol	10	M	young	yes	nonsmoker	alcohol	53
M	young	no	nonsmoker	no alcohol	39	M	young	no	smoker	no alcohol	94
M	young	no	smoker	alcohol	60	M	young	no	nonsmoker	alcohol	32
M	young	no	nonsmoker	no alcohol	36	F	young	yes	nonsmoker	alcohol	18
F	young	yes	nonsmoker	no alcohol	11	F	young	yes	smoker	alcohol	15
F	young	yes	smoker	alcohol	58	F	young	yes	smoker	no alcohol	51
F	young	no	nonsmoker	alcohol	79	F	young	no	smoker	alcohol	47
F	young	no	smoker	alcohol	76	F	young	no	nonsmoker	alcohol	1
F	young	no	nonsmoker	alcohol	85	F	middle	yes	nonsmoker	no alcohol	46
F	middle	yes	nonsmoker	alcohol	86	F	middle	yes	nonsmoker	no alcohol	79
F	middle	yes	smoker	alcohol	74	F	middle	yes	smoker	alcohol	1
F	middle	yes	smoker	alcohol	68	F	middle	no	smoker	alcohol	33
F	middle	no	smoker	alcohol	6	F	middle	no	nonsmoker	no alcohol	37
F	middle	no	nonsmoker	no alcohol	79	F	middle	no	nonsmoker	alcohol	17
F	middle	no	smoker	no alcohol	93	F	middle	no	smoker	alcohol	9
F	old	yes	smoker	alcohol	48	F	old	yes	smoker	alcohol	80
F	old	yes	smoker	no alcohol	85	F	old	yes	nonsmoker	no alcohol	16
F	old	yes	nonsmoker	no alcohol	48	F	old	yes	nonsmoker	alcohol	65
F	old	no	nonsmoker	no alcohol	51	F	old	no	nonsmoker	no alcohol	71
F	old	no	nonsmoker	alcohol	95	F	old	no	smoker	no alcohol	92
F	old	no	smoker	alcohol	23	F	old	no	smoker	alcohol	61
F	old	no	smoker	alcohol	88	F	old	no	nonsmoker	no alcohol	35
F	old	no	nonsmoker	no alcohol	26						

Tabulka 4.2: Data k ukázce aplikace na pětičetné třídění

$(2,1,2,2,2,1)$, $(2,2,2,1,1,2)$, $(2,3,2,2,1,1)$, $(1,1,2,2,1,1)$, $(1,1,2,1,2,1)$ a $(1,1,2,2,2,3)$. Algoritmus v tomto příkladě končí po dvaceti krocích (přičemž jeden krok znamená jednu změnu baze) v nichž postupně snižujeme hodnotu, kterou nabývá účelová funkce (neboli součet absolutních odchylek modelem predikovaných hodnot od hodnot skutečných). Posloupnost hodnot, kterých nabývá účelová funkce je následující: 6125, 3056, 2458, 2382, 2169, 1960, 1817, 1733, 1697, 1653, 1617, 1584, 1524, 1511, 1493, 1488, 1484, 1483, 1476 a 1475, která je již minimální hodnotou, které je možné dosáhnout. Odhadnuté parametry jsou $\mu = 52$, $\alpha_1 = 1.5$, $\alpha_2 = -1.5$, $\beta_1 = 5$, $\beta_2 = -1$, $\beta_3 = -4$, $\gamma_1 = 1.5$, $\gamma_2 = -1.5$, $\delta_1 = 12.5$, $\delta_2 = -12.5$ $\kappa_1 = -4$, a $\kappa_2 = 4$. Stejně jako v případě dvojného třídění je možné při řešení testovat, zda indexům v množině J_0 odpovídá struktura stromu. V tomto případě ale musíme nejprve informaci o hodnotách všech faktorů, vyjma prvního, soustředit do jednoho faktoru, tj. musíme přemapovat všechny kombinace, kterých může nabývat druhý až poslední faktor na přirozená čísla $I+1, \dots, I+JKLM$. To učiníme obecně velmi snadno například tak, že nabývá-li i-tý faktor z n faktorů I_i různých hodnot, pak zkonstruujeme zobrazení $\Lambda : (\lambda_2, \dots, \lambda_n) \to \mathbb{N} \setminus \{1, \dots, I_1\}$, které přiřazuje každé kombinaci hodnot druhého až n-tého faktoru přirozené číslo takto:

$$\Lambda(\lambda_2, \ldots, \lambda_n) = 1 + I_1 + \sum_{i=2}^{n} (\lambda_i - 1) \left(\prod_{k=1}^{i-1} I^{(k)} \right), \tag{4.32}$$

kde $I^{(1)} = 1$ a $I^{(k)} = I_k$ pro $k = 2, \ldots, n$. Grafy vygenerované při řešení úlohy minimalizace absolutních odchylek aplikované na data pacientů jsou na obrázku 4.1, z kterého je patrné, že pouze 9 z 20 vygenerovaných grafů má strukturu stromu.

4.2 Rekonstrukce kontingenční tabulky

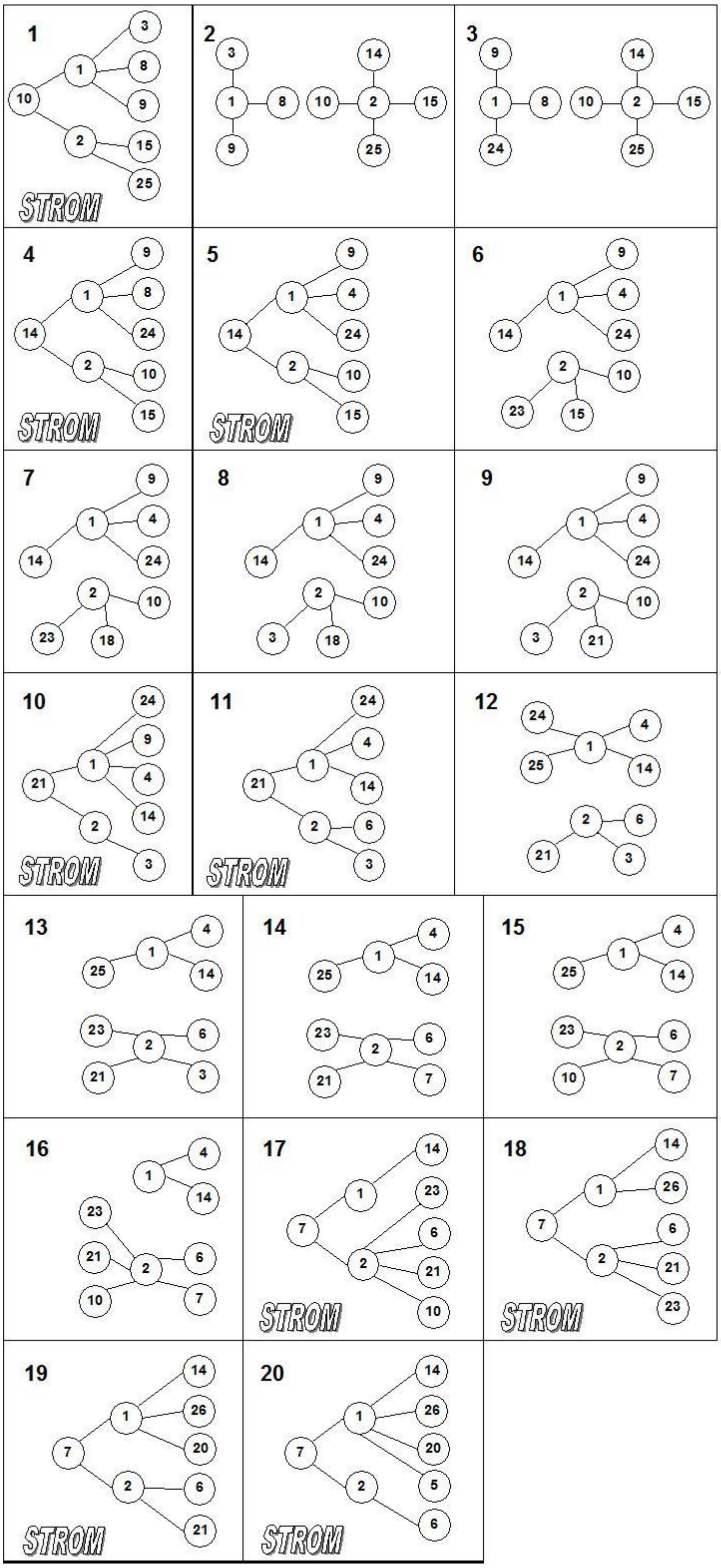
V článku [18] je zmíněna aplikace (DTP) na rekonstrukci kontingenční tabulky při znalosti řádkových a sloupcových součtů a dalších informací o vnitřní struktuře tabulky. Autor popisuje dvě techniky. První je technika používaná GROEW[7], druhá je technika navržená v článku [11]. Obě techniky konvergují ke stejnému řešení a to velmi rychle. Spočívají v nalezení kontingenční tabulky při daných sloupcových a řádkových součtech, která je v nějakém smyslu nejblíže jiné kontingenční tabulce o níž předpokládáme, že má stejnou strukturu jako námi hledaná tabulka.

Mějme dvourozměrnou kontingenční tabulku v níž jsou dány všechny hodnoty. Může se jednat kupříkladu o demografická data (počet lidí v různých věkových kategoriích a rodinných stavech) z nějakého roku. Představme si nyní, že jsme v situaci, kdy jsme naměřili počty lidí ve stejně definovaných věkových kategoriích i počty lidí se všemi možnými rodinnými stavy, pouze v jiném roce, ale nemáme k dispozici počty lidí v jednotlivých kombinacích věkových kategorií a rodinných stavů. Pak zmíněnými technikami lze jednoduše odhadnout skutečný počet lidí v těchto kombinacích věkových kategorií a rodinných stavů a tím odhadout jednotlivé prvky kontingenční tabulky.

Obě techniky jsou velmi podobné. Popíšeme tedy detailně pouze jednu z nich a to techniku používanou GROEW. Tato technika je iterační a spočívá v postupném přičítání přírůstků k hodnotám ze známé kontingenční tabulky střídavě podle požadovaných řádkových a požadovaných sloupcových součtů. Popišme nyní tuto techniku a dokažme, že skutečně konverguje k řešení dopravního problému se ztrátovou funkcí (4.33).

Pro všechna $i = 1, \ldots, m$ a $j = 1, \ldots, n$ označme $X_{ij} > 0$ prvek známé

[7] General Register Office of England and Wales



Obrázek 4.2: Grafy k ukázce zobecnění na pětičetné třídění

kontingenční tabulky v i-tém řádku a j-tém sloupci[8]. Označme dále

$$\sum_{j=1}^{n} X_{ij} = X_{i\bullet},$$

$$\sum_{i=1}^{m} X_{ij} = X_{\bullet j},$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n} X_{ij} = X_{\bullet\bullet}.$$

Nechť $L_i > 0$ je požadovaný součet prvků i-tého řádku hledané kontingenční tabulky a $M_j > 0$ je požadovaný součet prvků j-tého sloupce hledané kontingenční tabulky pro $i = 1, \ldots, m$ a $j = 1, \ldots, n$. Pochopitelně platí $\sum_{i=1}^{m} L_i = \sum_{j=1}^{n} M_j$. Označme nakonec x_{ij} odhadovanou neznámou hodnotu prvku hledané kontingenční tabulky v i-tém řádku a j-tém sloupci a zaveďme ztrátovou funkci

$$\sum_{i=1}^{m}\sum_{j=1}^{n} \frac{(x_{ij} - X_{ij})^2}{X_{ij}}. \tag{4.33}$$

Minimalizací (4.33) vzhledem k podmínkám

$$\sum_{i=1}^{m} x_{ij} = M_j \text{ a } \sum_{j=1}^{n} x_{ij} = L_i \text{ pro } i = 1, \ldots, m \text{ a } j = 1, \ldots, n$$

řešíme problém rekonstrukce kontingenční tabulky. Jde tedy o dopravní problém s nelineární účelovou funkcí bez omezení na nezáporné proměnné, který je možné řešit kupříkladu metodou Lagrangeových multiplikátorů (viz. str. 35). Námi zkoumaný iterační postup je alternativou k řešení tohoto dopravního problému, byť iterační postup k optimálnímu řešení pouze konverguje. Konvergence je však velmi rychlá. K témuž řešení konverguje i zmíněný, tzv. Stephanův-Demingův algoritmus, odvozený v [11]. Protože nemáme omezení na nezápornost prvků hledané tabulky, může se stát, že nalezená tabulka bude obsahovat záporné prvky. Pokud tedy požadujeme nezápornost prvků hledané tabulky (což se u kontingenčních tabulek dá předpokládat), pak je-li nějaký prvek tabulky získané prostřednictvím jedné z popisovaných technik záporný, nelze tyto techniky k rekonstrukci tabulky použít. Algoritmus používaný GROEW, založený na rozdělování přírůstků

[8] Neuvažujeme $X_{ij} = 0$ pro žádnou uspořádanou dvojici indexů $(i, j) \in I \times J$, protože funkce (4.33) by pak nebyla definována.

k prvkům tabulky, je detailně popsán v adresáři ***Algoritmus Friedlander (přírůstky)*** na přiloženém CD. V témže adresáři je spustitelný soubor obsahující implementaci tohoto algoritmu (excelovský soubor). Stephanův-Demingův algoritmus z článku [11] je detailně popsán v adresáři ***Algoritmus Friedlander (bez přírůstků)*** na přiloženém CD. V tomto adresáři je také spustitelný soubor obsahující implementaci Stephanova-Demingova algoritmu (excelovský soubor).

Nyní dokážeme, že algoritmus skutečně konverguje k řešení zmíněného dopravního problému s účelovou funkcí (4.33). Lagrangeova funkce pro (4.33) má tvar

$$L(\mathbf{x}, \lambda', \mu') = \tag{4.34}$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{(x_{ij}-X_{ij})^2}{X_{ij}} + \sum_{i=1}^{m}\lambda'_i\left(L_i - \sum_{j=1}^{n}x_{ij}\right) + \sum_{j=1}^{n}\mu'_j\left(M_j - \sum_{i=1}^{m}x_{ij}\right).$$

Dle metody Lagrangeových multiplikátorů (viz. Lagrangeova metoda, str. 35) derivujeme Lagrangeovu funkci podle x_{ij} a položíme každou takto získanou funkci rovnu nule. Tím dostáváme pro každé $i \in \{1, \ldots, m\}$ a $j \in \{1, \ldots, n\}$ rovnici:

$$x_{ij} = X_{ij} + \frac{\lambda'_i}{2}X_{ij} + \frac{\mu'_j}{2}X_{ij}.$$

Sečtením všech těchto rovností přes všechna $j = 1, \ldots, n$, resp. všechna $i = 1, \ldots, m$ a přeznačením $\lambda := \frac{\lambda'}{2}$ a $\mu := \frac{\mu'}{2}$ dostáváme

$$\left(\frac{L_i}{X_{i\bullet}} - 1\right) - \lambda_i - \sum_{j=1}^{n}\mu_j\frac{X_{ij}}{X_{i\bullet}} = 0, \text{ pro všechna } i = 1, \ldots, m, \tag{4.35}$$

$$\left(\frac{M_j}{X_{\bullet j}} - 1\right) - \sum_{i=1}^{m}\lambda_i\frac{X_{ij}}{X_{\bullet j}} - \mu_j = 0, \text{ pro všechna } j = 1, \ldots, n. \tag{4.36}$$

Označme nyní pro všechna $i = 1, \ldots, m$ a všechna $j = 1, \ldots, n$

$$P_{ij} = \left(\frac{X_{ij}}{X_{i\bullet}}\right) \text{ a } Q_{ij} = \left(\frac{X_{ij}}{X_{\bullet j}}\right)$$

a matice typu $m \times n$ $\mathbf{P} = (P_{ij})$ a $\mathbf{Q} = Q_{ij}$. Dále označme vektory typu $1 \times m$

$$\lambda = (\lambda_1, \ldots, \lambda_m) \text{ a } \mathbf{L}^T = \begin{pmatrix} \frac{L_1}{X_{1\bullet}} - 1 \\ \vdots \\ \frac{L_m}{X_{m\bullet}} - 1 \end{pmatrix}$$

a vektory typu $1 \times n$

$$\mathbf{M}^T = \begin{pmatrix} \frac{M_1}{X_{\bullet 1}} - 1 \\ \vdots \\ \frac{M_n}{X_{\bullet n}} - 1 \end{pmatrix} \text{ a } \mu = (\mu_1, \dots, \mu_n) .$$

Rovnice (4.35) a (4.36) mají nyní maticový zápis

$$\mathbf{L} - \lambda - \mu \mathbf{P}^T = \mathbf{0} \tag{4.37}$$

$$\mathbf{M} - \lambda \mathbf{Q} - \mu = \mathbf{0}. \tag{4.38}$$

Odečtením $\mathbf{P}^T$ násobku (násobeno zprava) rovnice (4.38) od rovnice (4.37) (případně vyjádřením μ z rovnice (4.38) a jeho dosazením do rovnice (4.37)) dostáváme

$$\left(\mathbf{L} - \mathbf{M}\mathbf{P}^T\right) - \lambda \left(\mathbf{I} - \mathbf{Q}\mathbf{P}^T\right) = \mathbf{0}, \tag{4.39}$$

kde $\mathbf{I}$ je jednotková matice typu $m \times m$. Ukážeme, že matice $\mathbf{QP}^T$ je stochastická, resp. že je sloupcově stochastická, tj. součet prvků v libovolném sloupci této matice je roven jedné. Prvek na místě (i, j), kde $i \in \{1, \dots, m\}$ a $j \in \{1, \dots, m\}$ matice $\mathbf{QP}^T$ označíme $\left(\mathbf{QP}^T\right)_{ij}$. Součet prvků j-tého sloupce matice $\mathbf{QP}^T$ je

$$\begin{aligned} \sum_{i=1}^{m} \left(\mathbf{QP}^T\right)_{ij} &= \sum_{i=1}^{m} \sum_{k=1}^{n} \left[\frac{X_{ik} X_{jk}}{X_{\bullet k} X_{j\bullet}}\right] \\ &= \sum_{k=1}^{n} \frac{X_{jk}}{X_{j\bullet}} \underbrace{\left[\sum_{i=1}^{m} \frac{X_{ik}}{X_{\bullet k}}\right]}_{=1} = \sum_{k=1}^{n} \frac{X_{jk}}{x_{j\bullet}} = 1. \end{aligned}$$

Tato rovnost platí pro každé $j = 1, \dots, m$. Protože $\mathbf{QP}^T$ je stochastická, obsahuje její spektrum jedničku. Dosazením jedničky do charakteristického polynomu dostáváme $\det\left(\mathbf{I} - \mathbf{QP}^T\right) = 0$, z čehož je zřejmé, že matice $\left(\mathbf{I} - \mathbf{QP}^T\right)$ je singulární. Zapišme nyní (4.39) ve tvaru

$$\left(\mathbf{I} - \mathbf{QP}^T\right)^T \lambda^T = \left(\mathbf{L} - \mathbf{MP}^T\right)^T . \tag{4.40}$$

Protože matice $\mathbf{QP}^T$ je singulární, nemá obecně soustava rovnic (4.40) vzhledem k λ jednoznačné řešení. Předpokládejme, že existuje k lineárně nezávislých řešení λ a μ. Pak libovolné řešení vzhledem k nejednoznačnosti můžeme zapsat ve tvaru $\lambda = \mathbf{R} + \mathbf{AS}$, kde $\mathbf{R}$ je matice typu $1 \times m$, $\mathbf{A}$ je matice typu $1 \times k$ a $\mathbf{S}$ je matice typu $k \times m$, kde $\mathbf{R}$ je nějaké řešení soustavy (4.40), $\mathbf{S}$ je k lineárně nezávislých netriviálních řešení homogenní soustavy

$$\left(\mathbf{I} - \mathbf{QP}^T\right)^T \lambda^T = \mathbf{0}$$

a $\mathbf{A}$ je matice k parametrů, které mohou nabývat libovolných hodnot a pro libovolnou takovou matici $\mathbf{A}$ je λ řešením (4.39). Obdobně můžeme řešení μ zapsat ve tvaru $\mu = \mathbf{T} + \mathbf{BU}$, kde $\mathbf{T}$ je matice typu $1 \times n$, $\mathbf{U}$ je matice typu $k \times n$ a $\mathbf{B}$ je matice typu $1 \times k$ libovolných hodnot, kde $\mathbf{T}$ je nějaké řešení soustavy

$$\left(\mathbf{I} - \mathbf{P}^T\mathbf{Q}\right)^T \mu^T = (\mathbf{M} - \mathbf{LQ})^T\,, \tag{4.41}$$

$\mathbf{U}$ je k lineárně nezávislých netriviálních řešení homogenní soustavy

$$\left(\mathbf{I} - \mathbf{P}^T\mathbf{Q}\right)^T \mu^T = \mathbf{0}$$

a $\mathbf{B}$ je matice parametrů, které mohou nabývat libovolných hodnot a pro každé $\mathbf{B}$ je μ řešením (4.41). Je-li $\mathbf{A} = \mathbf{B}$, řekneme, že řešení μ a řešení λ si odpovídají. Rovnice (4.37) musí platit pro libovolná řešení λ a μ nezávisle na hodnotách matic $\mathbf{A}$ a $\mathbf{B}$. Vezměme tedy $\mathbf{A} = \mathbf{B} = \mathbf{0}$ a dosazením těchto parametrických vyjádření λ a μ do (4.37) dostáváme $\mathbf{L} - \mathbf{R} - \mathbf{TP}^T = \mathbf{0}$, což musí platit pro všechna řešení λ a μ a proto i $\mathbf{S} + \mathbf{UP}^T = \mathbf{0}$. Pokud nyní vezmeme obecné řešení $\lambda = \mathbf{R} + \mathbf{AS}$ a $\mu = \mathbf{T} + \mathbf{BU}$, dostaneme z (4.37), $\mathbf{L} = \mathbf{R} + \mathbf{TP}^T$ a $\mathbf{S} = -\mathbf{UP}^T$, rovnici $(\mathbf{A} - \mathbf{B})\,\mathbf{S} = \mathbf{0}$ a z toho plyne buď $\mathbf{A} = \mathbf{B}$, nebo je možné redukovat rozměr matice $\mathbf{A}$. To ale vede ke sporu, neboť indukcí bychom došli k tomu, že λ je určeno jednoznačně, což je ovšem ve sporu se singularitou matice $\left(\mathbf{I} - \mathbf{QP}^T\right)$. Proto musí být $\mathbf{A} = \mathbf{B}$ a kdykoliv řešení λ a μ splňují (4.37), pak si odpovídají.

Popišme iterační algoritmus, který vede v každém kroku k řešení, které s rostoucím počtem kroků konverguje k výše popsanému řešení dopravního problému s účelovou funkcí (4.33). Vstupem pro algoritmus je původní matice $\mathbf{X}$ s hodnotami X_{ij} v i-tém řádku a j-tém sloupci. Problém spočívá v tom, že řádkové i sloupcové součty nemusí souhlasit s těmi, které máme dány pro naši hledanou matici, resp. kontingenční tabulku, tj. obecně existuje nějaký řádek, že $X_{i\bullet} - L_i \neq 0$ nebo sloupec, že $X_{\bullet j} - M_j \neq 0$. V každém kroku algoritmu budeme upravovat hodnoty matice obdržené v předešlém kroku tím, že k nim budeme přičítat přírůstky získané dále popsaným způsobem.

Označme přírůstek, který budeme v N-tém kroku přičítat k prvku matice v i-tém řádku a j-tém sloupci ${}_N E_{ij}$. V prvním kroku máme v i-tém řádku a j-tém sloupci hodnotu X_{ij} a přírůstek, který budeme k této hodnotě přičítat je

$${}_1 E_{ij} := X_{ij}\mathbf{L}^{(i)},$$

kde $\mathbf{L}^{(i)}$ značí i-tý prvek vektoru $\mathbf{L}$, který je maticí typu $1 \times m$. Pro každé $i = 1, \ldots, m$ a $j = 1, \ldots, n$ položíme

$$X_{ij} := X_{ij} + X_{ij}\left(\frac{L_i}{X_{i\bullet}} - 1\right).$$

Po přičtení přírůstků ke všem prvkům matice dostáváme rozdíl mezi j-tým sloupcovým součtem v prvním kroku vzniklé matice a požadovaným sloupcovým součtem M_j

$$M_j - X_{\bullet j} - \sum_{i=1}^{m} \frac{X_{ij}}{X_{i\bullet}} \left(L_i - X_{i\bullet}\right).$$

V druhém kroku algoritmu dostáváme na místě (i, j), $i = 1, \ldots, m$ a $j = 1, \ldots, n$

$${}_2E_{ij} = X_{ij} \left(\left(\frac{M_j}{X_{\bullet j}} - 1 \right) - \sum_{i=1}^{m} \left(\frac{L_i}{X_{i\bullet}} - 1 \right) \frac{X_{ij}}{X_{\bullet j}} \right) = X_{ij} \left(\mathbf{M} - \mathbf{LQ}\right)^{(j)},$$

kde $(\mathbf{M} - \mathbf{LQ})^{(j)}$ značí j-tý prvek vektoru $(\mathbf{M} - \mathbf{LQ})$, který je maticí typu $1 \times n$. Podobně pak

$${}_3E_{ij} = X_{ij} \left(-\mathbf{MP}^T + \mathbf{LQP}^T\right)^{(i)},$$

$${}_4E_{ij} = X_{ij} \left(\mathbf{MP}^T\mathbf{Q} - \mathbf{LQP}^T\mathbf{Q}\right)^{(j)},$$

atd.

Hodnota prvku v i-tém řádku a j-tém sloupci matice po dokončení kroku $2N$ tohoto algoritmu je rovna

$$\begin{aligned}
X_{ij} +_1 E_{ij} +_2 E_{ij} + \ldots +_{2N} E_{ij} = \\
&X_{ij}\Big(1 + \Big(\mathbf{L} - \mathbf{MP}^T + \mathbf{LQP}^T - \mathbf{MP}^T\mathbf{QP}^T \ldots - \mathbf{MP}^T \left(\mathbf{QP}^T\right)^{N-2} \\
&+ \mathbf{L}\left(\mathbf{QP}^T\right)^{N-1}\Big)^{(i)} + \Big(\mathbf{M} - \mathbf{LQ} + \mathbf{MP}^T\mathbf{Q} - \mathbf{LQP}^T\mathbf{Q} \ldots + \mathbf{M}\left(\mathbf{P}^T\mathbf{Q}\right)^{N-1} \\
&- \mathbf{LQ}\left(\mathbf{P}^T\mathbf{Q}\right)^{N-1}\Big)^{(j)}\Big) \\
= X_{ij} &\left(1 + \left(\left(\mathbf{L} - \mathbf{MP}^T\right)\left(\mathbf{I} + \mathbf{QP}^T + \ldots + \left(\mathbf{QP}^T\right)^{N-2}\right) + \mathbf{L}\left(\mathbf{QP}^T\right)^{N-1}\right)^{(i)}\right. \\
&\left. + \left(\left(\mathbf{M} - \mathbf{LQ}\right)\left(\mathbf{I} + \mathbf{P}^T\mathbf{Q} + \ldots + \left(\mathbf{P}^T\mathbf{Q}\right)^{N-1}\right)\right)^{(j)}\right).
\end{aligned}$$

Ukažme, že pro $N \to \infty$,

$$\left(\left(\mathbf{L} - \mathbf{MP}^T\right)\left(\mathbf{I} + \mathbf{QP}^T + \ldots + \left(\mathbf{QP}^T\right)^{N-2}\right) + \mathbf{L}\left(\mathbf{QP}^T\right)^{N-1}\right) \tag{4.42}$$

konverguje pro $N \to \infty$ k řešení (4.39) pro λ a

$$\left(\left(\mathbf{M} - \mathbf{LQ}\right)\left(\mathbf{I} + \mathbf{P}^T\mathbf{Q} + \ldots + \left(\mathbf{P}^T\mathbf{Q}\right)^{N-1}\right)\right) \tag{4.43}$$

konverguje k řešení (4.41) pro μ. Vynásobením levé strany (4.42) členem $\left(\mathbf{I}-\mathbf{QP}^T\right)$ dostáváme snadno

$$\left(\mathbf{L}-\mathbf{MP}^T\right)+\mathbf{MP}^T\left(\mathbf{QP}^T\right)^{N-1}-\mathbf{L}\left(\mathbf{QP}^T\right)^N.$$

Protože je $\left(\mathbf{QP}^T\right)$ sloupcově stochastická matice, platí (viz. [12], str. 173) $\left(\mathbf{QP}^T\right)^N \overset{N\to\infty}{\to} \left(\frac{X_{i\bullet}}{X_{\bullet\bullet}}\right)_{i=1,\dots,m}^{j=1,\dots,n}$, což je matice typu $m\times m$, jejíž prvky nezávisí na sloupcovém indexu. Dále odvoďme

$$\mathbf{MP}^T\left(\mathbf{QP}^T\right)^{N-1} \overset{N\to\infty}{\to} \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{X_{ij}X_{i\bullet}}{X_{i\bullet}X_{\bullet\bullet}}\left(\frac{M_j}{X_{\bullet j}}-1\right)=$$

$$\frac{\left(\sum_{j=1}^{n}M_j\right)}{X_{\bullet\bullet}}-1$$

a analogicky $\mathbf{L}\left(\mathbf{QP}^T\right)^N \overset{N\to\infty}{\to} \frac{\left(\sum_{i=1}^{m}L_i\right)}{X_{\bullet\bullet}}-1$. Tedy

$$\mathbf{MP}^T\left(\mathbf{QP}^T\right)^{N-1}-\mathbf{L}\left(\mathbf{QP}^T\right)^N \overset{N\to\infty}{\to} 0.$$

Celkově tedy dostáváme, že (4.42) konverguje pro $N\to\infty$ k řešení λ splňující (4.39). Vynásobením levé strany (4.43) členem $\left(\mathbf{I}-\mathbf{P}^T\mathbf{Q}\right)$ dostáváme

$$(\mathbf{M}-\mathbf{LQ})+\mathbf{LQ}\left(\mathbf{P}^T\mathbf{Q}\right)^N-\mathbf{M}\left(\mathbf{P}^T\mathbf{Q}\right)^N,$$

a obdobně jako v případě λ toto konverguje pro $N\to\infty$ k řešení μ splňující (4.41). Tudíž iterační algoritmus konverguje k řešení dopravního problému se ztrátovou funkcí (4.33) a zbývá pouze dokázat, že řešení λ a μ si odpovídají. To je ale zřejmé, neboť v každém kroku je buď splněna podmínka $\sum_{i=1}^{m}x_{ij}=M_j$ pro všechna $j=1,\dots,n$, nebo podmínka $\sum_{j=1}^{n}x_{ij}=L_i$, pro všechna $i=1,\dots,m$, kde x_{ij} značí matici získanou v daném kroku iteračního postupu. V limitě potom jsou splněny obě tyto rovnosti a tudíž je v limitě splněna i rovnice (4.37) a jak jsme ukázali výše, řešení si tedy odpovídají.

Technika používaná GROEW je použitelná i pro trojrozměrné kontingenční tabulky. Řešení konverguje k optimálnímu řešení planárního trojrozměrného diskrétního dopravního problému typu $m\times n\times o$ s danými 2-marginály $\mathbf{K}>\mathbf{0}$ typu $m\times n$ $\mathbf{L}>\mathbf{0}$ typu $n\times o$ a $\mathbf{M}>\mathbf{0}$ typu $m\times o$ s účelovou funkcí

$$\sum_{k=1}^{o}\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{\left(x_{ijk}-X_{ijk}\right)^2}{X_{ijk}} \tag{4.44}$$

za podmínek $K_{i,j} = \sum_{k=1}^{o} x_{ijk}$ pro všechna $i = 1, \ldots, m$ a $j - 1, \ldots, n$, $L_{j,k} = \sum_{i=1}^{m} x_{ijk}$ pro všechna $j = 1, \ldots, n$ a $k = 1, \ldots, o$ a $M_{i,k} = \sum_{j=1}^{n} x_{ijk}$ pro všechna $i = 1, \ldots, m$ a $k = 1, \ldots, o$. Detailně je algoritmus, který konverguje k řešení tohoto trojrozměrného dopravního problému, popsán v adresáři ***Rekonstrukce 3d tabulky*** na přiloženém CD. Ve stejném adresáři je i spustitelný soubor v němž je algoritmus implementován (excelovský soubor).

Podívejme se nyní na minimalizaci účelové funkce

$$\sum_{i=1}^{m} \sum_{j=1}^{n} \frac{|x_{ij} - X_{ij}|}{X_{ij}} \tag{4.45}$$

za podmínek $L_i = \sum_{j=1}^{n} x_{ij}$ pro všechna $i = 1, \ldots, m$ a $M_j = \sum_{i=1}^{m} x_{ij}$ pro všechna $j = 1, \ldots, n$. Tentokrát jde o dopravní problém s po částech lineární účelovou funkcí, bez omezení na nezápornost proměnných. Podobně jako v kapitole o statistickém třídění vede její minimalizace na řešení následujícího problému lineárního programování:

$$\min_{\mathbf{d} \geq \mathbf{0}} \mathbf{F}\mathbf{d}^T, \tag{4.46}$$

kde $\mathbf{F} = (\mathbf{F}' \mid \mathbf{F}')$ a $\mathbf{F}' = \left(\frac{1}{X_{11}}, \ldots, \frac{1}{X_{1n}}, \ldots, \frac{1}{X_{m1}}, \ldots, \frac{1}{X_{mn}}\right)$ je řádkový vektor typu $1 \times mn$ a

$$\mathbf{d} = \left(d_{11}^+, \ldots, d_{1n}^+, \ldots, d_{m1}^+, \ldots, d_{mn}^+, d_{11}^-, \ldots, d_{1n}^-, \ldots, d_{m1}^-, \ldots, d_{mn}^-\right)$$

je nezáporný řádkový vektor, vzhledem k podmínkám

$$\mathbf{A}_1\mathbf{d}^T = (X_{1\bullet}, \ldots, X_{m\bullet})^T - \mathbf{L}^T \tag{4.47}$$

$$\mathbf{A}_2\mathbf{d}^T = (X_{\bullet 1}, \ldots, X_{\bullet n})^T - \mathbf{M}^T, \tag{4.48}$$

kde

$$\begin{aligned}
\mathbf{L} &:= (L_1, \ldots, L_m) \\
\mathbf{M} &:= (M_1, \ldots, M_n) \\
\mathbf{A}_1 &:= [\mathbf{E}_1 \mid -\mathbf{E}_1] \\
\mathbf{A}_2 &:= [\mathbf{E}_2 \mid -\mathbf{E}_2] \\
\mathbf{E}_1 &:= \begin{bmatrix} \mathbf{e}_n & \mathbf{0}_n & \cdots & \mathbf{0}_n \\ \mathbf{0}_n & \mathbf{e}_n & \cdots & \mathbf{0}_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_n & \mathbf{0}_n & \cdots & \mathbf{e}_n \end{bmatrix} \\
\mathbf{E}_2 &:= \begin{bmatrix} \mathbf{I}_n & \mathbf{I}_n & \cdots & \mathbf{I}_n \end{bmatrix},
\end{aligned}$$

kde $\mathbf{e}_n$ je řádkový vektor typu $1 \times n$ obsahující samé jedničky ,matice $\mathbf{E}_1$ je typu $m \times mn$, matice $\mathbf{E}_2$ je typu $n \times mn$ a $\mathbf{I}_n$ je jednotková matice typu $n \times n$. Tuto úlohu je možné řešit klasickými metodami lineárního programování. Podívejme se nyní ještě na úlohu minimalizace funkce

$$\sum_{i=1}^{m} \sum_{j=1}^{n} |x_{ij} - X_{ij}| \tag{4.49}$$

za podmínek

$$\sum_{i=1}^{m} x_{ij} = M_j \text{ a } \sum_{j=1}^{n} x_{ij} = L_i,$$

pro všechna $i = 1, \ldots, m$ a $j = 1, \ldots, n$. Není již nutné předpokládat $X_{ij} > 0$ pro každé $i = 1, \ldots, m$ a $j = 1, \ldots, n$ a dokonce ani $X_{ij} \geq 0$. Tuto úlohu je jistě možné řešit jako problém lineárního programování obdobně jako právě popsanou minimalizaci funkce (4.45), nicméně je možné použít také následující algoritmus, kterým najdeme řešení v lineárním čase a který je detailně popsán v adresáři ***Rekonstrukce tabulky (absolutní odchylky)*** na přiloženém CD, kde je uložen i spustitelný soubor s implementovaným algoritmem (excelovský soubor).

1. Pro každé i spočteme $S_i^R := L_i - \sum_{j=1}^{n} X_{ij}$ a pro každé j spočteme $S_j^C := M_j - \sum_{i=1}^{m} X_{ij}$.

2. Hodnoty S_i^R i hodnoty S_j^C rozdělíme na hodnoty nezáporné a hodnoty záporné.

3. Aplikujeme metodu severozápadního rohu při řádkových omezeních S_i^R, v nichž **záporné** hodnoty nahradíme nulami a sloupcových omezeních S_j^C, v nichž **záporné** hodnoty rovněž nahradíme nulami. V každém kroku průběžně upravujeme řádkové i sloupcové součty. Jakmile se algoritmus dostane do posledního sloupce, nebo řádku, vyplní se do zbývajících polí matice zbývající hodnoty řádkových, resp. sloupcových součtů a do pravého dolního rohu matice se vyplní maximum ze zbývajícího řádkového a sloupcového součtu[9].

4. Aplikujeme metodu severozápadního rohu při řádkových omezeních S_i^R, v nichž **kladné** hodnoty nahradíme nulami a sloupcových omezeních S_j^C, v nichž **kladné** hodnoty rovněž nahradíme nulami. V každém kroku průběžně upravujeme řádkové i sloupcové součty. Jakmile se algoritmus

[9] Tím je zajištěno, že všechny sloupcové i řádkové součty odpovídají vytvořené matici až na jeden, který skutečný součet převyšuje.

dostane do posledního sloupce, nebo řádku, vyplní se do zbývajících polí matice zbývající hodnoty řádkových, resp. sloupcových součtů a do pravého dolního rohu matice se vyplní maximum ze zbývajícího řádkového a sloupcového součtu[10].

5. Hledanou matici $\mathbf{x}$ získáme jako součet matice $\mathbf{X}$ a matic získaných v předchozích dvou krocích.

Dokažme, že algoritmus skutečně vede k optimálnímu řešení. Nejprve ukážeme, že v (4.49) není možné dosáhnout hodnoty nižší, než

$$\tau := \max\left\{\sum_{i=1}^{m}\left|S_i^R\right|, \sum_{j=1}^{n}\left|S_j^C\right|\right\}.$$

Předpokládejme, že $\min_{x_{ij}} \sum_{i=1}^{m}\sum_{j=1}^{n}|X_{ij}-x_{ij}| < \sum_{i=1}^{m}\left|S_i^R\right|$. Jelikož nyní pro libovolnou matici řešení x_{ij} platí

$$\sum_{i=1}^{m}\sum_{j=1}^{n}|X_{ij}-x_{ij}| < \sum_{i=1}^{m}\left|S_i^R\right| = \sum_{i=1}^{m}\left|L_i - \sum_{j=1}^{n}X_{ij}\right| = \sum_{i=1}^{m}\left|\sum_{j=1}^{n}x_{ij} - \sum_{j=1}^{n}X_{ij}\right| =$$

$$\sum_{i=1}^{m}\left|\sum_{j=1}^{n}(X_{ij}-x_{ij})\right| \le \sum_{i=1}^{m}\sum_{j=1}^{n}|X_{ij}-x_{ij}|,$$

dostáváme spor, neboť tato ostrá nerovnost musí platit i pro řešení minimalizující ztrátovou funkci a přitom evidentně platit nemůže. Předpokládejme tedy, že $\min_{x_{ij}} \sum_{i=1}^{m}\sum_{j=1}^{n}|X_{ij}-x_{ij}| < \sum_{j=1}^{n}\left|S_j^C\right|$. Jelikož nyní pro libovolnou matici řešení x_{ij} platí

$$\sum_{i=1}^{m}\sum_{j=1}^{n}|X_{ij}-x_{ij}| < \sum_{j=1}^{n}\left|S_j^C\right| = \sum_{j=1}^{n}\left|M_j - \sum_{i=1}^{m}X_{ij}\right| = \sum_{j=1}^{n}\left|\sum_{i=1}^{m}x_{ij} - \sum_{i=1}^{m}X_{ij}\right| =$$

$$\sum_{j=1}^{n}\left|\sum_{i=1}^{m}(X_{ij}-x_{ij})\right| \le \sum_{j=1}^{n}\sum_{i=1}^{m}|X_{ij}-x_{ij}|,$$

dostáváme opět spor, neboť tato ostrá nerovnost musí platit i pro řešení minimalizující ztrátovou funkci a přitom evidentně platit nemůže. Z těchto dvou sporných tvrzení vyplývá, že žádná matice řešení $\mathbf{x}$ nemůže vést k nižší

[10] Tím je zajištěno, že všechny sloupcové i řádkové součty odpovídají vytvořené matici až na jeden, který skutečný součet převyšuje.

hodnotě ztrátové funkce, než τ. Nyní ukážeme, že navržený algoritmus vede k takové matici řešení $\mathbf{x}$, že platí

$$\sum_{i=1}^{m}\sum_{j=1}^{n}|X_{ij}-x_{ij}|=\tau.$$

Mějme matici $\mathbf{X}$ a požadované sloupcové součty M_j a řádkové součty L_i, pro všechna $i=1,\ldots,m$ a $j=1,\ldots,n$. V prvním kroku algoritmu spočteme S_i^R a S_j^C pro každý řádek a každý sloupec. Nechť $I=\{1,\ldots,m\}$ a $J=\{1,\ldots,n\}$. Označme

$$I^+=\left\{k:k\in I\wedge S_k^R\geq 0\right\},\ I^-=\left\{k:k\in I\wedge S_k^R<0\right\},$$

$$J^+=\left\{k:k\in J\wedge S_k^C\geq 0\right\}\ \text{ a }\ J^-=\left\{k:k\in J\wedge S_k^C<0\right\}.$$

Uvažujme nulovou matici $\mathbf{Y}$ typu $m\times n$ a požadované řádkové součty $L_i^+:=S_i^R$ pokud $i\in I^+$ a $L_i^+:=0$ pokud $i\notin I^+$ a sloupcové součty $M_j^+:=S_j^C$ pokud $j\in J^+$ a $M_j^+:=0$ pokud $j\notin J^+$. Aplikujeme metodu severozápadního rohu a obsazujeme jednotlivá místa matice od severozápadního rohu k rohu jihovýchodnímu. To je jistě možné dělat tradičním způsobem až do okamžiku, kdy dosáhneme posledního sloupce, či řádku. Předpokládejme bez újmy na obecnosti, že jsme dosáhli posledního řádku. Mohou nastat dvě možnosti, buď v okamžiku dosažení posledního řádku je hodnota sloupcového indexu nižší než n, nebo je rovna n. Je-li rovna n, pak prvku na místě (m,n) přiřadíme maximum z hodnoty m-tého aktuálního řádkového součtu a hodnoty n-tého aktuálního sloupcového součtu. Tím dojde k tomu, že poslední řádkový součet, nebo poslední sloupcový součet je nekladný. Je-li hodnota sloupcového indexu nižší než n, pak přiřadíme všem prvkům posledního řádku, které ještě nebyly obsazeny, vyjma prvku v posledním sloupci zbývající hodnoty sloupcových součtů. Průběžně upravujeme řádkový součet a při dosažení prvku na místě (m,n) vezmeme maximum z aktuálního řádkového a sloupcového součtu.

Matice $\mathbf{Y}$ má nyní sloupcové součty i řádkové součty totožné s požadovanými sloupcovými součty M_j^+ a požadovanými řádkovými součty L_i^+ až na nejvýše jeden, který je roven $-\left|\sum_{i=1}^m L_i^+-\sum_{j=1}^n M_j^+\right|$. Označme $L_i^-:=S_i^R$ pokud $i\notin I^+$ a $L_i^-:=0$ pokud $i\in I^+$ a sloupcové součty $M_j^-:=S_j^C$ pokud $j\notin J^+$ a $M_j^+:=0$ pokud $j\in J^+$. Je tedy $S_i^R=L_i^++L_i^-$ pro všechna $i=1,\ldots,m$ a $S_j^C=M_j^++M_j^-$ pro všechna $j=1,\ldots,n$. Protože $\sum_{i=1}^m S_i^R=\sum_{j=1}^n S_j^C$, je $\sum_{i=1}^m L_i^+-\sum_{j=1}^n M_j^+=\sum_{j=1}^n M_j^--\sum_{i=1}^m L_i^-$.

Předpokládejme bez újmy na obecnosti, že $\sum_{i=1}^m L_i^+\leq\sum_{j=1}^n M_j^+$. Pak $-\left|\sum_{i=1}^m L_i^+-\sum_{j=1}^n M_j^+\right|=\sum_{i=1}^m L_i^+-\sum_{j=1}^n M_j^+=\sum_{j=1}^n M_j^--\sum_{i=1}^m L_i^-$.

Nyní máme matici $\mathbf{Y}$, řádkové součty $K_i := L_i^-$ pro $i = 1, \ldots, m-1$ a $K_m := L_m^- + \left(\sum_{j=1}^n M_j^- - \sum_{i=1}^m L_i^-\right)$ a sloupcové součty $N_j := M_j^-$ pro $j = 1, \ldots, n$. Dostáváme součet řádkových součtů $\sum_{i=1}^m K_i = \sum_{i=1}^{m-1} L_i^- + L_m^- + \sum_{j=1}^n M_j^- - \sum_{i=1}^m L_i^- = \sum_{j=1}^n M_j^- = \sum_{j=1}^n N_j$ a jedná se tedy již o vyrovnaný dopravní problém s nekladnými řádkovými a sloupcovými součty. Přípustné řešení najdeme tedy pomocí metody severozápadního rohu s drobnou modifikací–místo přičítání minima z aktuálního řádkového a sloupcového součtu přičítáme maximum (to je záporné). Dostáváme řešení $\mathbf{Y} = (y_{ij})$, pro nějž platí $\sum_{i=1}^m \sum_{j=1}^n |y_{ij}| = \sum_{j=1}^n M_j^+ - \sum_{j=1}^n M_j^- = \sum_{j=1}^n \left|S_j^C\right| = \tau$. Matici řešení $\mathbf{x}$ získáme jako součet $\mathbf{x} = \mathbf{Y} + \mathbf{X}$.

Příklad použití tohoto algoritmu je uveden v adresáři ***Rekonstrukce tabulky (absolutní odchylky)*** na přiloženém CD pod názvem ***Příklad*** .

4.3 Jiné aplikace

V této kapitole krátce zmiňuji několik dalších aplikací dopravního problému ve statistice a teorii pravděpodobnosti a odkazuji na relevantní literaturu.

Řízené zaokrouhlování

V článku [5] autoři uvádějí aplikaci řízeného zaokrouhlování ve statistice.

Definice 21. Nechť $a \in \mathbb{R}$ a $0 < B \in \mathbb{N}$. Zobrazení $R : \mathbb{R} \to \mathbb{Z}$, které každému reálnému číslu a přiřadí $B\left\lceil \frac{a}{B} \right\rceil$ nebo $B\left\lfloor \frac{a}{B} \right\rfloor$ budeme nazývat zaokrouhlením s bazí B. Pokud pro R platí, že pro libovolné $k \in \mathbb{Z}$ je $R(k \cdot B) = k \cdot B$, pak říkáme, že zaokrouhlení R je nula-omezující.

Nechť $\mathbf{A}$ je matice typu $m \times n$. Nechť dále $A_{\bullet j}$ značí pro každé $j = 1, \ldots, n$ j-tý sloupcový součet matice $\mathbf{A}$ a $A_{i\bullet}$ značí pro každé $i = 1, \ldots, m$ i-tý sloupcový součet matice $\mathbf{A}$ a $A_{\bullet\bullet}$ značí součet všech prvků matice $\mathbf{A}$. Problém optimálního řízeného zaokrouhlování spočívá v nalezení zobrazení $R : \mathbb{R}^{m\times n} \to \mathbb{Z}^{m\times n}$, takového, že $R(a_{ij}) = R_{ij}(\mathbf{A})$, kde a_{ij} je prvek matice $\mathbf{A}$ na místě (i, j), kde $i = 1, \ldots, m$ a $j = 1, \ldots, n$ splňujícího následující podmínky

$$\forall i \in \{1, \ldots, m\}\, \forall j \in \{1, \ldots, n\}\; R_{ij}(\mathbf{A}) \text{ je zaokrouhlením prvku } a_{ij} \tag{4.50}$$

$$\forall j \in \{1, \ldots, n\} \sum_{i=1}^m R_{ij}(\mathbf{A}) = R(A_{\bullet j}) \tag{4.51}$$

$$\forall i \in \{1, \ldots, m\} \sum_{j=1}^n R_{ij}(\mathbf{A}) = R(A_{i\bullet}) \tag{4.52}$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n} R_{ij}\left(\mathbf{A}\right) = R\left(A_{\bullet\bullet}\right) \tag{4.53}$$

$$R\left(\mathbf{A}\right) \text{ minimalizuje } \left(\sum_{i=1}^{m}\sum_{j=1}^{n} |R(a_{ij}) - a_{ij}|^p\right)^{\frac{1}{p}} \tag{4.54}$$

pro nějaké pevně dané $1 \leq p < \infty$, nebo pro $p = \infty$

$$R\left(\mathbf{A}\right) \text{ minimalizuje } \max\left\{|R(a_{ij}) - a_{ij}| : i = 1, \ldots, m \text{ a } j = 1, \ldots, n\right\}.$$

Běžně používané zaokrouhlení $R(a) = \lfloor a + 0.5 \rfloor$ minimalizuje (4.54) pro $1 \leq p \leq \infty$, ale obecně nesplňuje (4.51), (4.52) a (4.53). Cox a Ernst formulovali problém řízeného zaokrouhlování jako problém celočíselného programování. Tento problém celočíselného programování je pro $1 \leq p < \infty$ dán následovně

$$\begin{aligned}
x_{ij} &:= R\left(a_{ij}\right) \\
x_{i\bullet} &:= R\left(a_{i\bullet}\right) - \lfloor a_{i\bullet} \rfloor \\
x_{\bullet j} &:= R\left(a_{\bullet j}\right) - \lfloor a_{\bullet j} \rfloor \\
x_{\bullet\bullet} &:= R\left(a_{\bullet\bullet}\right) - \lfloor a_{\bullet\bullet} \rfloor,
\end{aligned}$$

čili $x_{ij} \in \{0, 1\}$ pro každé $i = 1, \ldots, m$ a každé $j = 1, \ldots, n$.

$$\min_{x_{ij}} \left(\sum_{i=1}^{m}\sum_{j=1}^{n} |R\left(a_{ij}\right) - a_{ij}|^p\right)^{\frac{1}{p}} \tag{4.55}$$

$$\begin{aligned}
x_{ij} + (1 - x_{i\bullet}) &= \lfloor a_{i\bullet} + 1 \rfloor \text{ pro všechna } i = 1, \ldots, m \text{ a } j = 1, \ldots, n \\
(1 - x_{\bullet j}) + x_{\bullet\bullet} &= \sum_{j=1}^{n} \lfloor a_{\bullet j} + 1 \rfloor - \lfloor a_{\bullet\bullet} \rfloor \text{ pro všechna } j = 1, \ldots, n \quad (4.56) \\
x_{ij} + (1 - x_{\bullet j}) &= \lfloor a_{\bullet j} + 1 \rfloor \text{ pro všechna } i = 1, \ldots, m \text{ a } j = 1, \ldots, n \\
(1 - x_{i\bullet}) + x_{\bullet\bullet} &= \sum_{i=1}^{m} \lfloor a_{i\bullet} + 1 \rfloor - \lfloor a_{\bullet\bullet} \rfloor \text{ pro všechna } i = 1, \ldots, m
\end{aligned}$$

Pro dané $1 \leq p < \infty$ je množina všech řešení tohoto problému ekvivalentní množině řešení tohoto problému s lineární účelovou funkcí

$$\sum_{i=1}^{m}\sum_{j=1}^{n} \left((1 - a_{ij})^p - (a_{ij})^p\right) x_{ij}. \tag{4.57}$$

Jde tedy o dopravní problém daný minimalizací (4.57) za podmínek (4.56) s celočíselnými řádkovými i sloupcovými součty. Řešení tohoto problému je celočíselné, nicméně aby bylo zajištěno, že všechny prvky řešení x_{ij} jsou z množiny $\{0,1\}$, musíme přidat další omezení $0 \leq x_{ij} \leq 1$. Tím dostáváme dopravní problém s kapacitními omezeními. Navíc lze dokázat, že optimální řešení tohoto problému vždy existuje, neboť existuje vždy řešení přípustné. Dopravní problém s kapacitními omezeními je řešitelný jak simplexovým algoritmem, tak kupříkladu pomocí Ford-Fulkersonova algoritmu. Aplikace řízeného zaokrouhlování ve statistice, zejména v teorii výběrů z konečných populací, jsou uvedeny v článku [5].

Důkaz Strassenovy věty

Další aplikací dopravního problému je důkaz Strassenovy věty pro omezenou Lipschitzovu metriku. Budu vycházet z knihy [21].

Definice 22. Nechť X je normovaný lineární prostor. Množina $G \subset X$ je **slabě otevřená**, jestliže ke každému $x \in G$ existují funkcionály $\psi_1, \ldots, \psi_n \in X^*$ tak, že

$$V_x(\psi_1, \ldots, \psi_n) := \{t \in X : |\psi_1(x-t)| < 1, \ldots, |\psi_n(x-t)| < 1\} \subset G \tag{4.58}$$

Definice 23. Nechť X je normovaný lineární prostor a w je soustava všech slabě otevřených podmnožin X. Tuto soustavu podmnožin nazveme **slabou topologií** na X.

Poznámka 13. Předpokládejme, že množina stavů Ω je n-rozměrný Euklidův prostor. Na něm máme definovánu klasickou Euklidovu vzdálenost $\delta(\mathbf{x}, \mathbf{y}) := \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$, kde $\mathbf{x}, \mathbf{y} \in \Omega$. Označme d funkci vzdálenosti δ omezenou jedničkou, tj. např. $d(\mathbf{x}, \mathbf{y}) := \frac{\delta(\mathbf{x},\mathbf{y})}{1+\delta(\mathbf{x},\mathbf{y})}$ pro všechny prvky $\mathbf{x}, \mathbf{y} \in \Omega$.

Definice 24. Nechť F a G jsou pravděpodobnostní míry. Pak **omezenou Lipschitzovou metrikou** nazveme funkci

$$d_{BL}(F, G) = \sup \left| \int \psi dF - \int \psi dG \right|, \tag{4.59}$$

kde supremum se bere přes všechny funkce ψ, které splňují Lipschitzovu podmínku, tj.

$$|\psi(x) - \psi(y)| \leq d(x, y).$$

d_{BL} je metrika.

Věta 7. "Strassenova věta"
Nechť F a G jsou pravděpodobnostní míry. Pak následující tvrzení jsou ekvivalentní:

1. $d_{BL}(F, G) \leq \varepsilon$
2. Existují náhodné veličiny X a Y s $\mathcal{L}(X) = F$ a $\mathcal{L}(Y) = G$ takové, že $Ed(X, Y) \leq \varepsilon$.

Důkaz věty 7. Důkaz je uveden v knize [21] a využívá řešení dopravního problému se speciální maticí nákladů $\mathbf{D}$ pro jejíž prvky platí $d_{ij} \geq 0$, $d_{ii} = 0$, $d_{ij} = d_{ji}$ a $d_{ik} \leq d_{ij} + d_{jk}$ pro všechna relevantní i, j, k. Jde tedy o matici vzdáleností a ta umožňuje velmi výhodně řešit dopravní problém pomocí Lagrangeovy metody, přestože jde o problém lineárního programování.

Literatura

[1] T. S. Arthanari and Y. Dodge. *Mathematical Programming in Statistics*. John Wiley and Sons, New York, 1981.

[2] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty. *Nonlinear programming. Theory and algorithms, 2. ed.* John Wiley and Sons, New York, 1993.

[3] W. W. Bein, P. Brucker, J. K. Park, and P. K. Pathak. A Monge property for the d-dimensional transportation problem. *Discrete Applied Mathematics*, 58(2):97–109, 1995.

[4] R. E. Burkard, B. Klinz, and R. Rudolf. Perspectives of Monge Properties in Optimization. *Discrete Applied Mathematics*, 70(2):95–161, 1996.

[5] B. D. Causey, L. H. Cox, and L. R. Ernst. Applications of transportation theory to statistical problems. *Journal of the American Statistical Association*, 80(392):903–909, 1985.

[6] D. Coppersmith and S. Winograd. Matrix multiplication via arithmetic progressions. *Journal of Symbolic Computation*, 9(3):251–280, 1990.

[7] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms, Second Edition.* The MIT Press, Cambridge, 2001.

[8] L. H. Cox. On properties of multi-dimensional statistical tables. *Journal of Statistical Planning and Inference*, 117(2):251–273, 2003.

[9] J. A. De Loera and S. Onn. The complexity of three-way statistical tables. *SIAM Journal on Computing*, 33(4):819–836, 2004.

[10] J. A. De Loera and S. Onn. Markov bases of three-way tables are arbitrarily complicated. *Journal of Symbolic Computation*, 41(2):173–181, 2006.

[11] W. E. Deming and F. F. Stephan. On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *The Annals of Mathematical Statistics*, 11(4):427–444, 1940.

[12] J. Doob. *Stochastic Processes.* John Wiley and Sons, New York, 1953.

[13] M. Dyer, R. Kannan, and J. Mount. Sampling contingency tables. *Random Structures and Algorithms*, 10:487–506, 1997.

[14] J. R. Evans. The factored transportation problem. *Management Science*, 30(8):1021–1024, 1984.

[15] R. O. Ferguson and L. F. Sargent. *Linear Programming: Fundamentals and Applications.* McGraw-Hill, New York, 1958.

[16] C. A. Floudas and P. M. Pardalos, editors. *Encyclopedia of Optimization, Volume III, Interior - M.* Kluwer Academic Publisher, 2001.

[17] L. R. Ford Jr. and D. R. Fulkerson. *Flows in Networks.* Princeton University Press, Princeton, 1962.

[18] D. Friedlander. A technique for estimating a contingency table, given the marginal totals and some supplementary data. *Journal of the Royal Statistical Society, Series A*, 124(3):412–420, 1961.

[19] W. Gangbo and R. J. McCann. The geometry of optimal transportation. *Acta Mathematica*, 177:113–161, 1996.

[20] W. Ho and P. Ji. A genetic algorithm for the generalised transportation problem. *International Journal of Computer Applications in Technology*, 22(4):190–197, 2005.

[21] P. J. Huber. *Robust Statistics.* John Wiley and Sons, Hoboken, 2004.

[22] W. Juninger. On representatives of multi-index transportation problems. *European Journal of Operational Research*, 66(3):353–371, 1993.

[23] V. Klee and G. J. Minty. *'How Good is the Simplex Algorithm?', Inequalities, III.* Academic Press, New York, 1972.

[24] B. Korda. *Učebnice lineárního programování.* SNTL, Praha, 1962.

[25] M. K. Kravtsov and A. P. Krachkovskij. On some properties of three-index transportation polytopes. *Discrete Applied Mathematics*, 9(5):545–562, 1999.

[26] E. Lawler. *Combinatorial Optimization, Networks and Matroids.* Dover Publications, New York, 2001.

[27] T. Matsui. A Linear Time Algorithm for the Hitchcock Transportation Problem with Fixed Number of Supply Points. In *Cooperative Research Report 35, The Institute of Statistical Mathematics, Minami-Azabu, Minato-ku*, pages 128–138, Tokyo, Japan, 1992.

[28] J. Morávek and M. Vlach. On the necessary conditions for the existence of a solution to the multi-index problem. *Operations Research*, 15:542–545, 1967.

[29] J. Morávek and M. Vlach. On necessary conditions for a class of systems of linear inequalities. *Aplikace matematiky*, 13:299–303, 1968.

[30] T. S. Motzkin. The multi-index transportation problem. *Bulletin of The American Mathematical Society*, 58:494, 1952.

[31] V. C. Mouli, J. Srinivas, and K. V. Subbaiah. An unsupervised neural network to solve transportation problems. *The Institution of Engineers*, 86, September 2005.

[32] C. H. Papadimitriou and K. Steiglitz. *Combinatorial Optimization, Algorithms and Complexity.* Dover Publications, Mineola, 1998.

[33] D. Pisinger. Where are the hard knapsack problems? *Computers and Operations Research*, 32(9):2271–2284, 2005.

[34] M. Queyranne and F. C. R. Spieksma. Approximation algorithms for multi-index transportation problems with decomposable costs. *Discrete Applied Mathematics*, 76(1-3):239–253, 1997.

[35] S. T. Rachev and L. Rüschendorf. *Mass Transportation Problems, Volume I: Theory.* Springer, New York, 1998.

[36] S. T. Rachev and L. Rüschendorf. *Mass Transportation Problems, Volume II: Applications.* Springer, New York, 1998.

[37] R. L. Rardin. *Optimization in Operations Research.* Prentice Hall, Upper Saddle River, 2000.

[38] J. Rohn. *Lineární algebra a optimalizace.* Karolinum, Praha, 2004.

[39] R. Rudolf and G. J. Woeginger. The Cone of Monge Matrices: Extremal Rays and Applications. *Zeitschrift für Operations-Research*, 42(22):161–168, 1995.

[40] K. Sanli and A. Apaydin. Mathematical programming for estimation of parameters in random blocks model (review). *G.U. Journal of Science, 19(1)*, 19(1):41–48, 2006.

[41] E. D. Schell. Distribution of a product by several properties. *Proceedings of the Second Symposium in Linear Programming*, pages 615–642, 1955.

[42] G. Smith. Further necessary conditions for the existence of a solution to the multi-index problem. *Operations Research*, 21(1):380–386, 1973.

[43] G. Smith. A procedure for determining necessary and sufficient conditions for the existence of a solution to the multi-index problem. *Applications of Mathematics*, 19(3):177–183, 1974.

[44] G. Smith. On the Morávek and Vlach conditions for the existence of a solution to the multi-index problem. *Applications of Mathematics*, 20(6):432–435, 1975.

[45] M. Švábová and E. Marunová. A multidimensional transportation problem. *Ekonomicko-matematický obzor*, 3:308–323, 1967.

[46] B. S. Verkhovskij. Symmetric multidimensional transportation problems. *Moskovskij gosudarstvennij universitet*, pages 482–498, 1963.

[47] C. Villani. *Topics in Optimal Transportation.* American Mathematical Society, Providence, 2003.

[48] M. Vlach. *Deterministické modely rozvrhování výroby.* SNTL, Praha, 1983.

[49] M. Vlach. Conditions for the existence of solutions of the three-dimensional planar transportation problem. *Discrete Applied Mathematics*, 13(1):61–78, January 1986.

[50] R. Zitouni and A. Keraghel. Resolution of a capacitated transportation problem with four subscripts. *Kybernetes*, 32(9/10):1450–1463, 2003.

[51] R. Zitouni, A. Keraghel, and D. Benterki. Elaboration and implantation of an algorithm solving a capacitated four-index transportation problem. *Applied Mathematical Sciences*, 1(53):2643–2657, 2007.