ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

CZECH TECHNICAL UNIVERSITY IN PRAGUE

RNDr. Ivana Pultarová, Ph.D.

Víceúrovňové metody pro stochastické matice

Multilevel methods for stochastic matrices

Summary

Multilevel numerical methods for the solution of stationary probability distribution vectors of stochastic matrices can accelerate the convergence of stationary and Krylov iterative methods. A present development of these methods is caused by needs of many branches of technical and social sciences, biology and medicine. These methods called iterative aggregation-disaggregation (IAD) methods are an analogy of algebraic multigrid methods which are used mainly for discretized differential equations. However prevailing nonsymmetry of stochastic matrices gives rise to a different convergence behavior. In this lecture, the main results in the convergence theory of the IAD methods are summarized and several open questions are mentioned, the solution of which could contribute also to the classical methods of algebraic multigrid for nonsymmetric matrices.

Souhrn

Víceúrovňové numerické metody pro nalezení vektoru stacionárního rozdělení pravděpodobnosti stochastických matic umožňují urychlení konvergence stacionárních nebo krylovovských iteračních metod. Současný vývoj těchto metod je vyvolán potřebou mnoha oborů technických a sociálních věd, biologie a medicíny. Tyto metody, nazývané iterační agregační-desagregační (IAD) metody, jsou obdobou metod algebraického multigridu, které se používají zejména pro řešení diskretizovaných diferenciálních rovnic. Obecná nesymetrie stochastických matic je však příčinou odlišných konvergenčních vlastností. V přednášce jsou shrnuty hlavní dosud získané výsledky v teorii konvergence IAD metod a je zmíněna řada otevřených otázek, jejichž vyřešení by bylo přínosem také pro klasické metody algebraického multigridu pro nesymetrické úlohy.

Klíčová slova

Stochastické matice, Markovovy řetězce, stacionární rozdělení pravděpodobnosti Markovova řetězce, numerické řešení Markovova řetězce, víceúrovňové metody, algebraický multigrid, iterační agregační-desagregační metody.

Keywords

Stochastic matrices, Markov chains, stationary probability distribution of Markov chains, numerical solution of Markov chains, multilevel methods, algebraic multigrid, iterative aggregation-disaggregation methods.

Obsah

1 Úvod **6**

2 Stochastické matice a Markovovy řetězce **6**
2.1 Definice a charakteristiky 7
2.2 Numerické řešení Perronova vektoru 8
2.3 Víceúrovňové metody: iterační agregační-desagregační metody 8

3 Konvergence iteračních agregačních-desagregačních metod **11**
3.1 Dvouúrovňové IAD metody 11
3.2 Víceúrovňové IAD metody 14
3.3 Diskuse a otevřené otázky 16

Literatura **17**

RNDr. Ivana Pultarová, Ph.D. **19**

1 Úvod

V mnoha oblastech vědeckých a technických výpočtů se setkáme s úkolem určit vektor stacionárního rozdělení pravděpodobnosti stochastické matice. Mezi takové obory patří technické, informatické a ekonomické vědy (teorie obsluhy, zpracování dat, výpočet charakteristik elektronických sítí, výpočty rovnovážných stavů reakcí, odhadování spolehlivosti zabezpečovacích zařízení, ohodnocení uzlů sítě podle struktury odkazů, tzv. PageRank) a nové medicínské a biologické obory (určení metastabilních stavů velkých molekul, zkoumání vlastností genů).

Náročnější úlohy tohoto typu často nelze řešit přímými metodami. Lze použít řadu přibližných iteračních metod. Podobně jako při numerickém řešení parciálních diferenciálních rovnic získávají i zde popularitu metody víceúrovňové. Hlavním principem takových přístupů je konstrukce úloh menších rozměrů, jejichž řešení je použito pro urychlení konvergence základních iteračních metod. Tyto postupy používané pro stochastické matice se nazývají iterační agregační-desagregační (IAD) metody.

V přednášce shrnujeme známé konvergenční vlastnosti IAD metod. Obecná nesymetrie stochastických matic způsobuje, že důkazové prostředky se liší od postupů používaných u klasického algebraického multigridu. Přestože IAD metody získavají na popularitě a roste šíře jejich uplatnění, jsou jejich obecné konvergenční vlastnosti překvapivě málo prostudovány. Je k dispozici pouze jediná relevantní monografie. Příspěvky v literatuře se opírají většinou o heuristická zdůvodnění. Existuje mnoho základních otevřených otázek.

V kapitole 2 je stručně popsána řešená úloha a používané metody. V kapitole 3 jsou představeny dosud známé konvergenční vlastosti IAD metod a v závěru několik nezodpovězených otázek.

2 Stochastické matice a Markovovy řetězce

Nechť M^T je matice transponovaná k M. Prvek matice M v r-tém řádku a s-tém sloupci označme $[M]_{rs}$. Nezápornou maticí M míníme matici, jejíž prvky jsou nezáporná reálná čísla, a značíme $M \geq 0$. Obdobně kladnou matici značíme $M > 0$. Spektrum matice M značíme $\sigma(M)$ a spektrální poloměr $\rho(M)$. Jednotková matice je označena I a vektor jedniček se značí e.

Čtvercová matice M se nazývá nerozložitelná, jestliže neexistuje matice permutace P taková, že

$$PBP^T = \begin{pmatrix} \tilde{B}_{11} & \tilde{B}_{12} \\ 0 & \tilde{B}_{22} \end{pmatrix},$$

kde diagonální bloky $\tilde{B}_{11}$ a $\tilde{B}_{22}$ jsou čtvercové.

2.1 Definice a charakteristiky

Uvažujme nezápornou matici B typu $N \times N$ takovou, že její sloupcové součty jsou rovny jedné. Tedy $e^T B = e^T$. Taková matice se nazývá stochastická matice.

Uvažujme náhodný proces s diskrétní konečnou množinou stavů očíslovaných $1, 2, \ldots, N$ a s diskrétními hodnotami času $t_1, t_2, \ldots$. Nechť pravděpodobnost nabytí každého stavu v čase t_k závisí pouze na stavu procesu v čase t_{k-1}, ale nazávisí na k. Takový náhodný proces se nazývá diskrétní konečný Markovův řetězec. Jelikož se v tomto textu nebudeme zabývat spojitými Markovovými řetězcy ani nekonečnou množinou stavů, budeme používat pouze název Markovovy řetězce.

Nechť $[B]_{rs}$, $r, s = 1, 2, \ldots, N$, jsou pravděpodobnosti přechodu ze stavu s do stavu r během jednoho libovolného časového úseku. Matice B se pak nazývá matice pravděpodobností přechodů příslušného Markovova řetězce. Jelikož $e^T B = e^T$, je B stochastická.

Označme $x^k \in \mathcal{R}^N$ vektor pravděpodobností nabývání jednotlivých stavů v čase t_k. Potom vektor pravděpodobností nabývání jednotlivých stavů v čase t_{k+1} je $x^{k+1} = Bx^k$. Jednou ze základních charakteristik Markovova řetězce je jeho vektor stacionárního rozdělení pravděpodobnosti, jehož složky jsou pravděpodobnosti nabytí jednotlivých stavů v nekonečném čase.

Podle Perronovy-Frobeniovy věty [13, 15] vektor stacionárního rozdělení pravděpodobnosti Markovova řetězce s nerozložitelnou maticí pravděpodobností přechodů B nezávisí na počátečním stavu, je kladný a je jednoznačně určen. Označme jej $\hat{x}$. Je roven vlastnímu vektoru matice B příslušnému vlastnímu číslu 1, které je jednonásobné. Současně platí $\rho(B) = 1$. Vektor $\hat{x}$ tedy splňuje

$$B\hat{x} = \hat{x}, \qquad e^T \hat{x} = 1. \tag{1}$$

Vektor $\hat{x}$ se také nazývá Perronův vektor matice B.

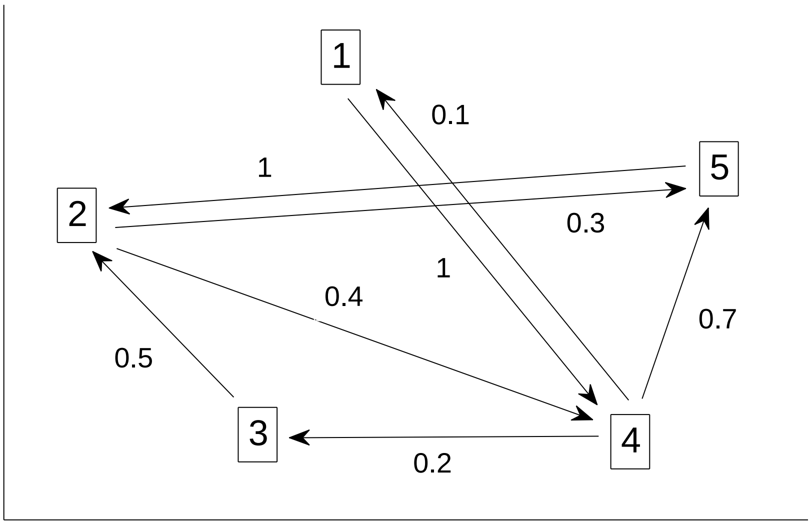
Některým metodám numerického řešení úlohy (1), kde B je nerozložitelná stochastická matice, a podmínkám jejich konvergence je věnována tato přednáška.

Příklad 2.1 Na Obrázku 1 je příklad pravděpodobností přechodů mezi stavy 1 - 5 Markovova řetězce odpovídajícího stochastické matici

$$B = \begin{pmatrix} 0 & 0 & 0 & 0.1 & 0 \\ 0 & 0.3 & 0.5 & 0 & 1 \\ 0 & 0 & 0.5 & 0.2 & 0 \\ 1 & 0.4 & 0 & 0 & 0 \\ 0 & 0.3 & 0 & 0.7 & 0 \end{pmatrix}. \tag{2}$$

Snadno se ověří, že Perronův vektor matice B je

$$\hat{x} = \frac{1}{205} (4, 90, 16, 40, 55)^T .$$



Obrázek 1: Grafické schema pravděpodobností přechodů mezi stavy 1 - 5 Markovova řetězce odpovídajícího stochastické matici (2) z Příkladu 2.1.

2.2 Numerické řešení Perronova vektoru

Předpokládejme nadále, že B je nerozložitelná stochastická matice. Označme projekci $P = \hat{x}e^T$ a matici $Z = B - P$. Platí $\sigma(Z) = \sigma(B) \cup \{0\} \setminus \{1\}$. Označme

$$\mu(B) = \max\{|\lambda|;\ \lambda \in \sigma(B),\ \lambda \neq 1\}.$$

Řekneme, že stochastická matice B je primitivní, jestliže je nerozložitelná a $\mu(B) < 1$. Je-li $\mu(B) = 1$, je B cyklická. Stochastická matice B je primitivní, právě když existuje celé $m > 0$ takové, že $B^m > 0$.

Úlohu (1) lze řešit jako soustavu lineárních rovnic přímými metodami nebo iteračními metodami, např. mocninnou, Richardsonovou, Jacobiovou nebo Gaussovou-Seidelovou metodou. Iterační matice T jsou obecně odvozeny z regulárního rozkladu [15] matice $I - B$. Tedy $T = M^{-1}W$, kde $M - W = I - B$, M je regulární, $M^{-1} \geq 0$ a $W \geq 0$. Redukce chyby v každém kroku je pak dána faktorem $\mu(T)$. Různé varianty blokových iteračních metod, jako jsou bloková Jacobiova a Gaussova-Seidelova metoda nebo Schwarzovy metody, umožňují urychlení konvergence i paralelizaci výpočtu Perronova vektoru. V případech, kdy $\mu(T) \approx 1$, lze zlepšit konvergenci použitím víceúrovňového schematu.

2.3 Víceúrovňové metody: iterační agregační-desagregační metody

Základní myšlenkou víceúrovňových metod je konstrukce vhodných snadno řešitelných úloh a využití jejich řešení pro přiblížení aktuální aproximace přesnému řešení. V kontextu úloh se stochastickými maticemi se takové metody nazývají iterační agregační-desagregační (IAD) metody. Jde o obdobu algebraických multigridních metod [1, 2, 3, 14]. Algebraické multigridní metody se používají pro řešení úloh se symetrickými pozitivně definitvními maticemi vzniklými např. diskretizací diferenciálních rovnic. Obecná nesymetrie stochastických matic je však příčinou odlišnosti těchto dvou teorií.

Uvažujme Markovův proces s N stavy a maticí pravděpodobností přechodů B. Označme L počet úrovní. Původní (největší) úloha přísluší úrovni 1 a nejhrubší (nejmenší) úloha urovni L. V každé úrovni kromě L-té zvolíme agregační skupiny stavů G_{mj}, kde m je číslo úrovně, $m = 1, 2, \ldots, L-1$, a j je číslo skupiny, $j = 1, 2, \ldots, N_{m+1}$. Tedy N_{m+1} je současně počet agregačních skupin v úrovni m a počet stavů v úrovni $m+1$ pro $m = 1, 2, \ldots, L-1$. Současně

$$\cup_{j=1}^{N_{m+1}} G_{mj} = \{1, 2, \ldots, N_m\}, \qquad G_{mj} \cap G_{mk} = \emptyset, \quad \text{pro} \quad j \neq k.$$

Velikost původní matice B je $N_1 = N$ a velikost nehrubší matice je N_L. Bez újmy na obecnosti v tomto textu předpokládejme, že v každé úrovni $m = 1, 2, \ldots, L-1$ jsou agregační skupiny vždy zvoleny tak, že jestliže $i_1 \in G_{m,k_1}$, $i_2 \in G_{m,k_2}$ a $i_1 < i_2$, potom $k_1 < k_2$. Symetrické přerovnání matice tedy obecně vede k jinému rozdělení stavů do skupin.

Přechody mezi úrovněmi jsou zprostředkovány zobrazeními (maticemi) redukce a prolongace. Matice redukce R_m zobrazuje $\mathcal{R}^{N_m}$ do $\mathcal{R}^{N_{m+1}}$,

$$\begin{aligned} [R_m]_{ij} &= 1 \quad \text{pro} \quad j \in G_{mi}, \\ &= 0 \quad \text{jinak.} \end{aligned}$$

Matice prolongace $S(x)_m$ zobrazuje $\mathcal{R}^{N_{m+1}}$ do $\mathcal{R}^{N_m}$ v závislosti na kladném vektoru $x \in \mathcal{R}^{N_m}$,

$$\begin{aligned} [S(x)_m]_{ij} &= \frac{x_i}{\sum_{k \in G_{mj}} x_k} \quad \text{pro} \quad i \in G_{mj}, \\ &= 0 \quad \text{jinak.} \end{aligned}$$

Matice $B_m = R_{m-1} \ldots R_1 B S(y^1)_1 \ldots S(y^{m-1})_{m-1}$ je agregovaná matice v m-té úrovni příslušející matici B pro vektory $y^1 \in \mathcal{R}^{N_1}, \ldots, y^{m-1} \in \mathcal{R}^{N_{m-1}}$.

Pro každou úroveň označme projekci $\mathcal{R}^{N_1}$ do $\mathcal{R}^{N_1}$

$$P(y^1, \ldots, y^{m-1})_m = S(y^1)_1 \ldots S(y^{m-1})_{m-1} R_{m-1} \ldots R_1.$$

Příklad 2.2 Nechť $N_1 = N = 5$, $L = 2$ a agregační skupiny $G_{1,1} = \{1, 2\}$ a $G_{1,2} = \{3, 4, 5\}$. Potom

$$R_1 = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix}, \qquad S(x^0)_1 = \begin{pmatrix} 1/3 & 0 \\ 2/3 & 0 \\ 0 & 2/6 \\ 0 & 3/6 \\ 0 & 1/6 \end{pmatrix} \qquad \text{pro} \quad x^0 = \frac{1}{12} \begin{pmatrix} 2 \\ 4 \\ 2 \\ 3 \\ 1 \end{pmatrix}.$$

Pro matici B z Příkladu 2.1 danou (2) je agregovaná matice $B_2 = R_1BS(x^0)_1 =$

$$= \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 0.1 & 0 \\ 0 & 0.3 & 0.5 & 0 & 1 \\ 0 & 0 & 0.5 & 0.2 & 0 \\ 1 & 0.4 & 0 & 0 & 0 \\ 0 & 0.3 & 0 & 0.7 & 0 \end{pmatrix} \begin{pmatrix} 1/3 & 0 \\ 2/3 & 0 \\ 0 & 2/6 \\ 0 & 3/6 \\ 0 & 1/6 \end{pmatrix} = \begin{pmatrix} 1/5 & 23/60 \\ 4/5 & 37/60 \end{pmatrix}.$$

Projekce $P(x^0)_2$ je

$$P(x^0)_2 = S(x^0)_1R_1 = \begin{pmatrix} 1/3 & 1/3 & 0 & 0 & 0 \\ 2/3 & 2/3 & 0 & 0 & 0 \\ 0 & 0 & 2/6 & 2/6 & 2/6 \\ 0 & 0 & 3/6 & 3/6 & 3/6 \\ 0 & 0 & 1/6 & 1/6 & 1/6 \end{pmatrix}.$$

Popišme jeden cyklus obecného algoritmu víceúrovňové IAD metody. Iterační matice T_m odpovídají nějaké stacionární iterační metodě uvedené v části 2.2, např. $T_m = B_m$ pro mocninnou metodu. Násobením maticí T_m říkáme základní iterace IAD metody nebo podle analogie s algebraickými multigridními metodami hladící kroky. Parametry μ_m a ν_m značí počet základních iterací v úrovni m před hrubým krokem a po něm. Hrubým krokem v úrovni m míníme (přibližné) řešení úlohy v úrovni $m+1$.

Jeden cyklus IAD metody: vstup $x^n > 0$, $m := 1$

1. $x := T_m^{\mu_m} x^n$ (základní iterace před hrubým krokem)
2. $y := R_m x$ (redukce)
3. $B_{m+1} := R_m B_m S(x)_m$; je-li $m = L-1$, vyřeš přesně $B_{m+1}z = z$, jinak přibližně použitím kroků 1. - 5. pro m větší o 1 s počátečním vektorem y (hrubý krok)
4. $x := S(x)_m z$ (prolongace)
5. $x^{n+1} := T_m^{\nu_m} x$ (základní iterace po hrubém kroku)

Toto základní schema IAD metody lze použít ve formě V-cyklu nebo W-cyklu nebo může být součástí předpodmínění gradientních metod. Lze snadno ukázat, že Perronův vektor matice B je pevným bodem tohoto procesu.

V literatuře se první IAD metody se dvěma úrovněmi objevily přibližně před padesáti lety [12]. Některé varianty získaly pojmenování [13]. Bloková Gaussova-Seidelova metoda s agregačním krokem se nazývá metoda Kouryova-McAllisterova-Stewartova. V našem značení je to tedy dvouúrovňová IAD metoda s $\mu_1 + \nu_1 = 1$ a s iterační maticí T_1 odpovídající blokové Gaussově-Seidelově metodě. Takahashiova metoda je dvouúrovňová

IAD metoda s $\mu_1 + \nu_1 = 1$ a se základní iterací odpovídající upravené blokové Gaussově-Seidelově metodě. V blokové Gaussově-Seidelově metodě se po výpočtu j-té části (odpovídající j-tému diagonálnímu bloku) vektoru normuje tato část tak, aby její $||\cdot||_1$-norma byla rovna $[z]_k$, kde z je právě získané řešení hrubé úlohy. Vantilborghova metoda je dvouúrovňová IAD metoda s $\mu_1 + \nu_1 = 1$. Základní iterace odpovídá výpočtu vektoru, jehož jednotlivé j-té části jsou Perronovy vektory stochastických doplňků matic B_{jj} v maticích $\tilde{B}_{jj}$. Matice $\tilde{B}_{jj}$ vzniknou z matice B agregováním všech bloků (skupin) kromě j-tého do jediného stavu. (Matice $\tilde{B}_{jj}$ tedy mají o 1 větší rozměr než B_{jj}.) Základní iterace Vantilborghovy metody tedy odpovídá jednomu kroku upravené blokové Jacobiovy metody. V posledních několika letech nastal pravděpodobně v důsledku potřeby aplikací nový rozvoj IAD metod, bylo publikováno mnoho článků a výsledků experimentů. Je zdůrazňována zejména podobnost IAD metod s algebraickým multigridem [2, 3].

3 Konvergence iteračních agregačních-desagregačních metod

Řekneme, že numerická metoda pro řešení úlohy (1) konverguje globálně k $\hat{x}$, jestliže pro každé počáteční přiblížení x^0, pro které $x^0 > 0$ a $e^T x^0 = 1$, konverguje získaná posloupnost k $\hat{x}$. Řekneme, že numerická metoda pro řešení (1) konverguje lokálně k $\hat{x}$, jestliže existuje okolí $U(\hat{x})$ vektoru $\hat{x}$ takové, že pro každé počáteční přiblížení $x^0 \in U(\hat{x})$, pro které $x^0 > 0$ a $e^T x^0 = 1$, konverguje získaná posloupnost k $\hat{x}$. Pouze pro některé speciální typy IAD metod a pro některé typy stochastických matic máme k dispozici důkazy lokální nebo globální konvergence.

3.1 Dvouúrovňové IAD metody

Věnujme se nejprve dvouúrovňovým metodám. Důkladně je zpracovaná teorie konvergence téměř zcela rozdělitelných stochastických matic [13]. Tyto matice mají blokovou strukturu a jejich diagonální bloky jsou blízké primitivním stochastickým maticím, normy mimodiagonálních bloků jsou velikosti $O(\varepsilon)$. Je dokázáno, že pro matici základní iterace T_1 odpovídající blokové Gaussově-Seidelově metodě konverguje IAD metoda globálně s faktorem konvergence $O(\varepsilon)$ [13]. Tento povzbudivý výsledek má dvě nevýhody. Ověření potřebných podmínek je obecně náročnější než samotný výpočet. V praktických případech není snadné identifikovat diagonální bloky s uvedenými vlastnostmi.

Další pokusy ověřit konvergenci IAD metod pro obecné stochastické matice jsou založeny na znalosti vzorce pro vývoj chyby [6]. Pro dvouúrovňovou IAD metodu platí

$$x^{n+1} - x = J(x^n)(x^n - x),$$

kde matice přenosu chyby je

$$J(x^n) = T_1^{\mu_1+\nu_1} \left(I - P(x^n)_1 Z\right)^{-1} \left(I - P(x^n)_1\right). \tag{3}$$

Spektrální poloměry matic $J(x^n)$ přesahují hodnotu 1 i v případech, kdy posloupnost přiblížení x^n konverguje k $\hat{x}$. Uvedený vzorec je proto významnou pomůckou zejména pro dokazování lokální konvergence. Dále v této části vynecháme index u T_1.

Bylo dokázáno, že IAD metoda s $T = \alpha B + (1 - \alpha)I$, $\alpha \in (0, 1)$, $\mu_1 + \nu_1 = 1$ vždy konverguje lokálně pro jakoukoliv nerozložitelnou matici B [7, 9]. Bylo dokázáno, že IAD metoda s $T = B$ a s $\mu_1 + \nu_1 = 1$ konverguje lokálně pro každou nerozložitelnou matici B, která má alespoň jeden kladný řádek nebo kladný sloupec nebo kladnou diagonálu [5, 7, 9].

Nutná a postačující podmínka pro globální konvergenci IAD metody se speciální volbou agregačních skupin byla publikována v roce 2006 [4]. Jestliže $T = B$, $\nu_1 = 1$ a nejvýše jedna agregační skupina obsahuje více než jeden stav, je globální konvergence IAD metody dána strukturou nezáporných prvků diagonálního bloku odpovídajícího skupině s více stavy. Přesněji řečeno, metoda globálně konverguje, právě když je stochastický doplněk tohoto bloku primitivní matice. Současně lze shora odhadnout asymptotický faktor konvergence na základě vztahu $\mu(B) \leq \frac{1}{2} \max_{i,j} ||Be_i - Be_j||_1$, kde e_i je i-tý sloupec jednotkové matice.

Nutná a postačující podmínka pro lokální konvergenci IAD metody pro $T = B$ a $\mu_1 + \nu_1 = 1$ avšak pro libovolnou volbu agregačních skupin a odhad asymptotického faktoru konvergence byly získány v roce 2008 [10]. Podmínka opět závisí pouze na struktuře nenulových prvků matice B.

Pro speciální struktury matice B, vhodnou volbu agregačních skupin a iterační matice T poskytne IAD metoda přesné řešení po konečném počtu cyklů [6]. Např. mají-li všechny mimodiagonální blokové řádky hodnost jedna a je-li základní iterace bloková Jacobiova metoda, dostaneme $x^2 = \hat{x}$.

O podobnost IAD metod a algebraických multigridních metod se opírají nedávné publikace, např. [2, 3], demonstrující možnost využití zhlazených agregací pro IAD metody. Jelikož je známo, že metody algebraického multigridu konvergují pro jakoukoliv volbu agregačních skupin [14], je vhodné se ptát, zda je tomu tak i pro IAD metody. Ukazuje se že nikoliv. Obecná absence symetrie a pozitivní definitnosti u úloh řešených IAD metodami neumožňuje využít obdobné důkazové prostředky jako u multigridních metod. Pouze v případech, kdy stochastická matice B v (1) je symetrická nebo $B = DSD^{-1}$, kde D je diagonální a S je symetrická, lze využít analogie IAD metod a multigridních metod. V těchto speciálních případech je tedy také např. zaručena lokální konvergence IAD metod.

Obecně je tedy žádoucí získat výsledky pro matice nesymetrické. Jednoduchým příkladem nesymetrické matice typu $N \times N$ je matice B cyklická např. s prvky $[B]_{rc} = 1$ pro $(r - c) \bmod n = 1$ a $[B]_{rc} = 0$ jinak. Schema nenulových pravděpodobností přechodů mezi N stavy je tedy

$$1 \to 2 \to 3 \to 4 \to \ldots \to N - 1 \to N \to 1. \tag{4}$$

Perronův vektor této matice je $\hat{x} = (1, \ldots, 1)^T / N$, není tedy potřeba jej numericky počítat. Nicméně cyklické matice jsou vhodné pro demonstrování odlišností IAD metod a algebraického multigridu. Ze spojitosti pak plynou podobné vlastnosti i pro cyklické matice s poruchami.

Pro sudé N uvažujme symetrickou permutaci uvedené cyklické matice B odpovídající schematu přechodů

$$1 \to 3 \to 5 \to \ldots \to N-1 \to N \to N-2 \to \ldots \to 4 \to 2 \to 1. \qquad (5)$$

Následující příklad ukazuje, jak důležitá je volba přerovnání matice B nebo (při daném přerovnání) volba skupin. Zdůrazněme, že při přerovnání (4) odpovídá volba skupin často používanému agregování "podle síly vazby" [2, 3]: stavy r a s jsou ve stejné agregační skupině, jestliže $[B]_{rs}+[B]_{sr} \gg 0$. Dosud však není k dispozici žádné teoretické zdůvodnění této volby. Tato volba dokonce není prospěšná vždy, jak ukazuje Příklad 3.1.

Příklad 3.1 Nechť $N = 200$. Uvažujme cyklickou stochastickou matici typu $N \times N$ a její dvě různá symetrická přerovnání B_a a B_b, kde B_a odpovídá struktuře pravděpodobností přechodů (4) a B_b odpovídá (5). Obě matice porušíme maticí $ee^T/N/1000$, přesněji $B_* := (1000B_* + ee^T/N)/1001$, $* = a, b$. Perronův vektor pro matice B_a i B_b je zřejmě $\hat{x} = (1, \ldots, 1)^T/N$.

Porovnejme nejprve hodnoty $\mu(T)$ iteračních matic některých stacionárních iteračních metod. Hodnoty $\mu(T)$ jsou současně faktory redukce chyby. V Richardsonově metodě volíme iterační matici $T_* = 0.7B_* + 0.3I$, $* = a, b$. U blokových metod volíme velikosti bloků 2.

$\mu(T)$	mocninná m.	Richardsonova m.	blok. Jacobiova m.	blok. Gaussova-Seidelova m.
B_a	0.9990	0.9992	0.9980	0.0038
B_b	0.9990	0.9992	0.9990	0.9985

Pouze bloková Gaussova-Seidelova metoda pro přerovnání typu (4) (matice B_a) konverguje podstatně rychleji než ostatní metody.

Nyní použijeme IAD metody. Porovnáme spektrální poloměry r_a a r_b asymptotických matic vývoje chyby $J_a(\hat{x})$ a $J_b(\hat{x})$ pro matice B_a a B_b pro IAD metodu s $\nu_1 = 0$ a se čtyřmi různými typy základní iterace: mocninnou, Richardsonovou, blokovou Jacobiovou a blokovou Gaussovou-Seidelovou metodou. Hodnoty r_* jsou horními odhady faktorů asymptotické redukce chyby (pro $r_* < 1$) nebo znamenají divergenci metody v lokálním smyslu (pro $r_* \geq 1$). Hodnoty spektrálních poloměrů r_a a r_b nalezneme v následujících tabulkách. Číslo $\#G_{1,i}$ je počet prvků v každé skupině.

r_a $\qquad T \approx$	mocninná	Richardsonova	blok. Jacobiova	blok. Gaussova-Seidelova
$\#G_{1,i} = 2$, $\mu_1 = 1$	0.9990	0.3993	0.0005	8E-5
$\#G_{1,i} = 2$, $\mu_1 = 3$	0.9970	0.2601	2E-5	3E-9
$\#G_{1,i} = 4$, $\mu_1 = 1$	0.9990	0.7609	0.0015	0.0002
$\#G_{1,i} = 4$, $\mu_1 = 3$	1.8311	0.7589	5E-5	2E-7

r_b $\quad T \approx$	mocninná	Richardsonova	blok. Jacobiova	blok. Gaussova-Seidelova
$\#G_{1,i} = 2,\ \mu_1 = 1$	0.9990	0.5370	0.9990	9.9830
$\#G_{1,i} = 2,\ \mu_1 = 3$	4.5266	2.9379	4.5228	10.5734
$\#G_{1,i} = 4,\ \mu_1 = 1$	0.9990	0.7609	0.9980	6.1335
$\#G_{1,i} = 4,\ \mu_1 = 3$	2.0108	0.8428	4.3293	6.9810

Je zřejmé, že lokální konvergence je dosaženo pro téměř všechny použité typy IAD metody pro přerovnání cyklické matice typu (4) (matice B_a). Výjimkou je IAD metoda se čtyřprvkovými skupinami, která používá jako základní iteraci tři kroky mocninné metody. Přerovnání typu (5) (matice B_b) vede k uspokojivě konvergentní IAD metodě pouze pro Richardsonovu metodu až na výjimku, kdy se používají dvouprvkové skupiny a tři kroky této metody. Je zajímavé, že bloková varianta Gaussovy-Seidelovy metody zde nevede ke konvergenci.

Následující věta specifikuje důsledky nevhodné volby přerovnání matice.

Věta 3.2 [11] *Nechť $N \geq 4$ je dělitelné čtyřmi. Uvažujme dvouúrovňovou IAD metodu, kde základní iterace odpovídá mocninné metodě nebo blokové Jacobiově metodě s počtem kroků $\mu_1 + \nu_1 = N/4$. Nechť každá agregační skupina obsahuje právě dva prvky. Nechť B je matice typu $N \times N$ odpovídající struktuře (5). Potom $\rho(J(\hat{x})) \geq N/4$.*

3.2 Víceúrovňové IAD metody

Teoretické výsledky pro konvergenci víceúrovňových metod jsou v literatuře dosud vzácné. Obecnou víceúrovňovou IAD metodu uvažujme nyní ve formě V-cyklu odpovídajícího popisu v části 2.3. Nechť L je počet úrovní.

Nechť $T = M^{-1}W$, kde $M - W = I - B$ je regulární rozklad. Tudíž $T\hat{x} = \hat{x}$. Budeme předpokládat $TB = BT$. Iterační matice v úrovních $m = 1, 2, \ldots, L-1$ nechť jsou

$$T_m = R_{m-1} \ldots R_1 T S(u_1)_1 \ldots S(u_{m-1})_{m-1}.$$

Pro nejhrubší úrovneň L budeme vždy uvažovat $T = B$, tedy

$$B_L = R_{L-1} \ldots R_1 B S(u_1)_1 \ldots S(u_{L-1})_{L-1}.$$

Označme vektory $u_k, v_k \in \mathcal{R}^{N_k}$ počítané během jednoho cyklu IAD metody

$$\begin{aligned} u_1 &= T^{\mu_1} x^n, \\ u_m &= (R_{m-1} \ldots R_1 T S(u_1)_1 \ldots S(u_{m-1})_{m-1})^{\mu_m} R_{m-1} u_{m-1}, \\ v_m &= (R_{m-1} \ldots R_1 T S(u_1)_1 \ldots S(u_{m-1})_{m-1})^{\nu_m} S(u_m)_m v_{m+1} \end{aligned}$$

pro $m = 2, 3, \ldots, L-1$. Nechť v_L je přesné řešení nejhrubší úlohy, $B_L v_L = v_L$. Vektory u_k se počítají během procesu zhrubování, zatímco vektory v_k se počítají během návratu zpět na úroveň 1.

Označme projekce

$$P(u_k, u_{k+1}, \dots, u_{m-1})_{km} = S(u_k)_k \dots S(u_{m-1})_{m-1} R_{m-1} \dots R_k.$$

Ve Větách 3.3 a 3.4 budeme používat zkratku $P_m \equiv P(u_1, u_2, \dots, u_{m-1})_{1m}$ a $P_1 = I$.

V následujících větách jsou uvedeny vzorce pro chybu x^{n+1} ve víceúrovňových IAD metodách, kde buď $L = 3$ nebo $\mu_m = \nu_m = 1$. Principy důkazů lze snadno upravit pro odvození vzorců pro $L > 2$ a libovolná $\mu_m \geq 0$, $\nu_m \geq 0$. Používáme konvenci

$$\sum_{k=m}^{n} X_k = 0, \qquad \prod_{k=m}^{n} X_k = 1,$$

jestliže $m > n$.

Věta 3.3 [8] *Uvažujme tříúrovňovou IAD metodu, $L = 3$, $\mu_m + \nu_m \geq 1$, $m = 1, 2$. Pro chybu v $n+1$-ním cyklu platí*

$$x^{n+1} - \hat{x} = J(x^n)(x^n - \hat{x}),$$

kde

$$\begin{aligned} J(x^n) &= T^{\nu_1}\Big((P_2 T)^{\nu_2}(I - P_3 Z)^{-1}\Big((P_2 - P_3)\sum_{k=0}^{\mu_2 - 1}(TP_2)^k (T - I) + I - P_3\Big) \\ &\quad + \sum_{k=0}^{\nu_2 - 1}(P_2 T)^k (I - P_2)\Big) T^{\mu_1}. \end{aligned}$$

Věta 3.4 [8] *Uvažujme IAD metodu s libovolným počtem úrovní $L \geq 2$ a s jedním krokem základní iterace před hrubým krokem a s jedním krokem základní iterace po hrubém kroku, tedy $\mu_m = \nu_m = 1$ pro $m = 1, 2, \dots, L-1$. Pro chybu v $n+1$-ním cyklu platí*

$$x^{n+1} - \hat{x} = J(x^n)(x^n - \hat{x}),$$

kde

$$\begin{aligned} J(x^n) &= T \prod_{k=2}^{L-1}(P_k T)(I - P_L Z)^{-1} \sum_{k=1}^{L-1}(P_k - P_{k+1}) M_{k-1} \\ &\quad + T \sum_{m=1}^{L-2} \prod_{k=2}^{m}(P_k T) \sum_{k=1}^{m}(P_k - P_{k+1}) M_{k-1}, \end{aligned}$$

kde $M_0 = T$ a

$$M_k = (T + \sum_{j=2}^{k} TP_j (T - I))T,$$

pro $k = 1, 2, \dots, L-2$.

Příklad 3.5 Uvažujme nesymetrickou nerozložitelnou primitivní nezápornou matici B typu 8×8,

$$B = \begin{bmatrix} \beta & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \alpha+\gamma & 1 & 0 & \alpha & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ \gamma & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \alpha & 0 & 0 & 0 & 0 & 0 & 1 & \beta \end{bmatrix}. \tag{6}$$

Uvažujme tři jednoduché varianty IAD metody s $T_m = B_m$ a jim příslušné matice přenosu chyby. Ve všech úrovních všech metod je zvoleno $\mu_m = \nu_m = 1$:
a) $L = 2$, $\#G_{1,i} = 2$, tedy počet prvků v každé agregační skupině je 2;
b) $L = 2$, $\#G_{1,i} = 4$, tedy počet prvků v každé agregační skupině je 4;
c) $L = 3$, $\#G_{1,i} = 2$, $\#G_{2,i} = 2$, tedy počet prvků v každé agregační skupině v každé úrovni je 2.
Pro několik různých hodnot parametrů α, β a γ spočítáme spektrální poloměry asymptotických matic přenosu chyby, tedy $\rho(J(\hat{x}))$, kde vektory $\hat{x}$, jsou Perronovy vlastní vektory matic B pro danou volbu parametrů. Při každé volbě trojice parametrů α, β a γ jsou sloupce normovány tak, aby výsledná matice byla stochastická. V následující tabulce nalezneme přibližné hodnoty spektrálních poloměrů asymptotických matic přenosu chyby pro uvedené tři metody a tři volby parametrů.

$(\alpha, \beta, \gamma) =$	$(10, 0, 0)$	$(0, 10, 0)$	$(0, 0, 10)$
a) $L = 2$, $\#G_{1,i} = 2$	0.0789	2.0463	0.1289
b) $L = 2$, $\#G_{1,i} = 4$	0.9090	0.4092	1.5757
c) $L = 3$, $\#G_{1,i} = 2$, $\#G_{2,i} = 2$	1.3390	0.3448	0.6861

Pro volbu $(\alpha, \beta, \gamma) = (10, 0, 0)$ obě dvouúrovňové metody konvergují lokálně, zatímco tříúrovňová metoda nekonverguje ani v lokálním smyslu. Pro volbu $(\alpha, \beta, \gamma) = (0, 10, 0)$ lokálně konvergují pouze metody variant b) a c). Pro volbu $(\alpha, \beta, \gamma) = (0, 0, 10)$ lokálně konvergují pouze metody variant a) a c). Tento příklad dokazuje, že z lokální konvergence IAD metody s L úrovněmi obecně neplyne lokální konvergence IAD metody se stejným typem a počtem základních iterací pro $L - 1$ ani pro $L + 1$ úrovní.

3.3 Diskuse a otevřené otázky

IAD metody pro řešení Perronova vektoru stochastické matice lze vnímat jako modifikaci algebraických multigridních metod pro úlohy typu $Ax = b$, kde A je symetrická pozitivně definitní matice. Pro důkazy konvergence IAD metod však nelze použít postupy z teorie multigridních metod. Další odlišností je skutečnost, že vhodná volba IAD metody významně urychlí výpočet oproti výpočtu stacionární iterační metodou, zatímco nevhodně zvolená IAD metoda může divergovat, jak ilustruje Příklad 3.1.

Novými výsledky v teorii konvergence IAD metod jsou podmímky pro lokální konvergenci některých typů IAD metod, důkaz neomezenosti asymptotického spektrálního poloměru matice šířeni chyby a odvození přesného vzorce pro chybu IAD metody s libovolným počtem úrovní a s libovolnými počty kroků základní iterace v jednotlivých úrovních. Poslední jmenovaný výsledek poskytuje základ pro další obecné zkoumání konvergence víceúrovňových IAD metod např. pro speciální struktury matic a volby skupin.

Některé prvotní úsudky byly vyvráceny. Např. zvýšení počtu kroků základní iterace obecně nezrychluje konvergenci, může způsobit i divergenci v lokálním smyslu. Lokální konvergence pro L úrovní není nutná ani postačující pro lokální konvergenci pro $L+1$-úrovní.

Bylo by vhodné najít snadno ověřitelné postačující podmínky lokální konvergence víceúrovňových IAD metod pro někté další typy metod nebo pro speciální typy stochastických matic. Takové podmínky by mohly např. vycházet pouze ze struktury nenulových prvků matice B, podobně jako u obdobných kriterií pro dvouúrovňové metody.

V případě divergence IAD metody pro danou matici B pozorujeme, že množina hromadných bodů počítané posloupnosti přiblížení je konečná. Nabízí se otázky: Jaká je souvislost mezi touto množinou a Perronovým vektorem? Jak se liší množina hromadných bodů přesně počítané posloupnosti x^n a množina hromadných bodů posloupnosti $\tilde{x}^n$ počítané v konečné aritmetice?

Za nejzajímavější otázku lze pokládat souvislost lokální a globální konvergence IAD metod. Zatím nebylo dokázáno, že z lokální konvergence dané metody pro danou matici plyne konvergence v globálním smyslu. Tato domněnka však nebyla dosud ani vyvrácena.

Zatím nevyužitým prostředkem studia IAD metod zůstává tzv. Fourierova analýza. Tento přístup se osvědčil pro klasický algebraický multigrid pro úlohy se symetrickou pozitivně (semi)definitní maticí, která je současně cirkulentní nebo Toeplitzova [1]. Zdá se, že Fourierova analýza by mohla poskytnout nové teoretické výsledky i pro IAD metody pro cirkulentní nebo Toeplitzovy nesymetrické stochastické matice, např. zdůvodnění výhodnosti agregování "podle síly vazby" a nalezení vhodného počtu kroků základní iterace.

Literatura

[1] M. Bolten, M. Donatelli, T. Huckle. Aggregation-based multigrid methods for circulant and Toeplitz matrices. Preprint BUW-IMACM 12/10, Bergische Universitat Wuppertal, 2012.

[2] M. Brezina, T. Manteuffel, S. McCormick, J. Ruge, G. Sanders. Towards adaptive smoothed aggregation (αSA) for nonsymmetric problems. *SIAM Journal on Scientific Computing*, 32:14–39, 2010.

[3] H. De Sterck, T. A. Manteuffel, S. F. McCormick, K. Miller, J. Pearson, J. Ruge, G. Sanders. Smoothed aggregation multigrid for Markov chains. *SIAM Journal on Scientific Computing*, 32:40–61, 2010.

[4] I. C. F. Ipsen, S. Kirkland. Convergence analysis of a PageRank updating algorithm by Langville and Meyer. *SIAM Journal on Matrix Analysis and Applications*, 27:952–967, 2006.

[5] J. Mandel, B. Sekerka. A local convergence proof for the iterative aggregation method. *Linear Algebra and Its Applications*, 51:163–172, 1983.

[6] I. Marek, P. Mayer. Convergence analysis of an iterative aggregation/disaggregation method for computing stationary probability vectors of stochastic matrices. *Numerical Linear Algebra with Applications*, 5:253–274, 1998.

[7] I. Marek, I. Pultarová. A note on local and global convergence analysis of iterative aggregation-disaggregation methods. *Linear Algebra and Its Applications*, 413:327–341, 2006.

[8] I. Pultarová. Error propagation formula of multi-level iterative aggregation-disaggregation methods for non-symmetric problems. *The Electronic Journal of Linear Algebra*, 25:9–21, 2012.

[9] I. Pultarová. Local convergence analysis of iterative aggregation-disaggregation methods with polynomial correction. *Linear Algebra and Its Applications*, 421:122–137, 2007.

[10] I. Pultarová. Necessary and sufficient local convergence condition of one class of iterative aggregation-disaggregation methods. *Numerical Linear Algebra with Applications*, 15:339–354, 2008.

[11] I. Pultarová, I. Marek. Physiology and pathology of iterative aggregation-disaggregation methods. *Numerical Linear Algebra with Applications*, 18:1051–1065, 2011.

[12] H. A. Simon, A. Ando. Aggregation of variables in dynamic systems. *Econometrica*, 29:111–138, 1961.

[13] W. J. Stewart. *Introduction to the numerical solution of Markov chains*. Princeton University Press, Princeton, 1994.

[14] P. Vaněk, M. Brezina, J. Mandel. Convergence of algebraic multigrid based on smoothed aggregation. *Computing*, 56:179–196, 1998.

[15] R. S. Varga. *Matrix iterative analysis*. Springer Series in Computational Mathematics 27, Springer-Verlag, New York, 2000.

RNDr. Ivana Pultarová, Ph.D.

Vzdělání:

1984 - 1989	Matematicko-fyzikální fakulta UK, obor Přibližné a numerické metody, diplomová práce Aplikace lineárního a diskrétního programování pro potřeby klinické a experimentální medicíny, Cena Rektora UK,
2002 - 2006	doktorské studium, Fakulta stavební ČVUT, obor Matematika ve stavebním inženýrství, disertační práce IAD metody v počítání Markovových řetězců.

Zaměstnání:

1989 - 1991	asistentka, Katedra matematiky, Fakulta stavební ČVUT,
2004 - 2008	vědecká pracovnice, Ústav informatiky AV ČR,
1998 - dosud	odborná asistentka, Katedra matematiky, Fakulta stavební ČVUT.

Odborné zájmy:
víceúrovňové numerické metody pro řešení Markovových řetězců; hierarchické báze konečných prvků; metody rozkladu oblasti; aposteriorní odhady chyby numerického řešení diferenciálních rovnic.

Vlastní řešené projekty:
Víceúrovňové řešení úloh s rychle se měnícími vlastnostmi, interní grant ČVUT, 2003,
Víceúrovňová aditivní a multiplikativní Schwarzova metoda, interní grant ČVUT, 2004,
Víceúrovňové metody v předpodmínění a v odhadech chyb, projekt GAČR, 2009 - 2010.

Účast v daších projektech:
Funkční způsobilost a optimalizace stavebních konstrukcí, výzkumný záměr, 2001 - 2004,
Spolehlivost, optimalizace a trvanlivost stavebních materiálů a konstrukcí, výzkumný záměr, 2005 - 2011,
Modelování a simulace náročných technických problémů, program Informační společnost Národního výzkumného programu, 2004 - 2008,
Matematická teorie iteračních procesů s aplikacemi, GAČR, 2002 - 2004,
Výpočtové aspekty víceúrovňových modelů, GAČR, 2005 - 2007,
Efektivní iterační metody řešení rozsáhlých soustav rovnic předpodmíněné na bázi agregací, GAČR, 2009 - 2011.

Odborné publikace:
Devět článků publikováno a dva články přijaty k publikování v časopisech s kladným impaktním faktorem, dva články publikovány v mezinárodních recenzovaných časopisech. Jedna kapitola v knize Linear Algebra Research Advances. Více než dvacet příspěvků ve sbornících konferencí. Monografie F. Bubeník, M. Pultar a I. Pultarová, Matematické vzorce a metody, skriptum J. Novák, I. Pultarová a P. Novák, Úvod do programování v Matlabu.

Recenze pro časopisy Applied Mathematics and Computation, Electronic Transactions on Numerical Analysis, Numerical Functional Analysis and Optimization, Applications of Mathematics.