Hrátky se (stochastickou) cyklickou maticí

Ivo Marek
FSv ČVUT v Praze

Seminář Numerické Analýzy

Akce se koná v rámci projektu IET400300415 programu Informační společnost

Sedlec-Prčice, 16. - 20. ledna 2006

1 Iterační procesy

$$Ax = b$$

Konstrukce iterační matice T

$$M^{-1}Ax = M^{-1}x, \quad M^{-1}A = I - T$$

$$A = M - W = M(I - T), \quad \text{splitting}, \quad c = M^{-1}b,$$

$$(1.1)\ Mx^{k+1} = Wx^k + b \quad \text{equivalent with} \quad x^{k+1} = Tx^k + c$$

$$T = P + Z,\ P^2 = P,\ PZ = ZP = 0,$$

Konvergence nastává právě když platí (i) a (ii), kde

$$(i) \quad r(T) = \lim_{k\to\infty} \|T^k\|^{1/k} \le 1,$$

$$(ii) \qquad \lim_{k\to\infty} Z^k = 0$$

1 ***Problém***

Nalézt stacionární vektor pravděpodobnosti stochastické matice B***.***

2 Lemma $0 \in \sigma(A)$ ***právě když*** $1 \in \sigma(T)$

$$0 = Aw$$

$$= M(I - T)w$$

a tedy

$$w = Tw.$$

3 ***Definice***

Index cykličnosti (irreducibilní cyklické matice) p***:***

$\{\lambda \in \mathcal{C}^1 : |\lambda| = 1\} \cap \sigma(T) = \{1, exp\{2\pi i/p\}, ..., exp\{2\pi i(p-1)/p\}\}$

Co dělat v případě, že spektrum $\sigma(T)$ **je cyklické?**

Snadná možnost: Posuv $\tau I + T, 0 < \tau < +\infty$

$$S = \frac{1}{1+\tau}(T + \tau I)$$

obr. 1 **a obr.** 2

Situace je kritická pro $p \geq \kappa 10^s,\ s >> 1,\ \kappa \geq 1$ **natož** $p = O(N^\delta), \delta > 0$**.**

$$\hat{x} = T\hat{x} = S\hat{x},\ (\hat{x})_j > 0, j = 1, ..., N, \sum_{j=1}^{n} (\hat{x})_j = 1.$$

Let $\mathcal{E} = \mathcal{R}^N, \mathcal{F} = \mathcal{R}^n, n < N$**,**

$e^T = e(N)^T = (1, ..., 1) \in \mathcal{R}^N$**.**

$$\mathcal{G} : \{1, ..., N\} \overset{onto}{\to} \{1, ..., n\}$$

IAD communication operators

$$(Rx)_j = \sum_{\mathcal{G}(j)=j} x_j$$

$$S = S(u),\ (S(u)z)_j = \frac{u_j}{(Ru)_j}(Rx)_j$$

We obviously have

$$RS(u) = I_{\mathcal{F}}$$

For the *aggregation projection* $P(x) = S(x)R$

$$P(x)^T e = e \quad \forall x \in \mathcal{R}^N, x_j > 0, j = 1, ..., N$$

and

$$P(x)x = x \quad \forall x \in \mathcal{R}^N, x_j > 0, j = 1, ..., N$$

Define the *aggregated matrix* as

$$\mathcal{B}(x) = RBS(x)$$

How to choose the map $\mathcal{G}$?

Natural, if B is in "suitable" block form

$$B = \begin{pmatrix} B_{11} & B_{12} & \ldots & B_{1n} \\ B_{21} & B_{22} & \ldots & B_{2n} \\ . & . & \ldots & . \\ B_{n1} & B_{n2}, & \ldots & B_{nn} \end{pmatrix}, \text{ with diagonal } n_j \times n_j \text{ block } B_{jj},\ j = 1, \ldots,$$

Then one lets usually,

$$j = \mathcal{G}(j) \text{ for } n_0+n_1+\ldots+n_{j-1}+1 \leq j \leq n_0+n_1+\ldots+n_{j-1}+n_j,\ n_0 = 0.$$

and this means that each of the blocks B_{jk} is aggregated to a 1×1 matrix.

Conversely, the map $\mathcal{G}$ gives a rise to appropriate block form of B (up to a permutation).

4 Example ***Let*** B ***be a stochastic matrix written in a block form***

$$B = \begin{pmatrix} B_{11} & B_{12} & \dots & B_{1n} \\ B_{21} & B_{22} & \dots & B_{2n} \\ . & . & \dots & . \\ B_{n1} & B_{22} & \dots & B_{nn} \end{pmatrix}$$

Take $N = 4$ ***and*** $n = 2$ ***and each block*** B_{jk} ***to be*** 2×2***. Then, choosing***

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

and

$$S(x) = \begin{pmatrix} \frac{1}{x_1+x_2}x_1 & 0 \\ \frac{1}{x_1+x_2}x_2 & 0 \\ 0 & \frac{1}{x_3+x_4}x_3 \\ 0 & \frac{1}{x_3+x_4}x_4 \end{pmatrix}$$

the aggregated matrix becomes

$$\mathcal{B}(x) = \begin{pmatrix} \beta_{11}(x) & \beta_{12}(x) \\ \beta_{21}(x) & \beta_{22}(x) \end{pmatrix}$$

a 2×2 ***matrix with***

$$\beta_{\overline{11}}(x) = \frac{1}{x_1 + x_2}[(b_{11} + b_{21})x_1 + (b_{12} + b_{22})x_2]$$

$$\beta_{\overline{12}}(x) = \frac{1}{x_3 + x_4}[(b_{13} + b_{23})x_3 + (b_{14} + b_{24})x_4]$$

$$\beta_{\overline{21}}(x) = \frac{1}{x_1 + x_2}[(b_{31} + b_{41})x_1 + (b_{32} + b_{42})x_2]$$

$$\beta_{\overline{22}}(x) = \frac{1}{x_3 + x_4}[(b_{33} + b_{43})x_3 + (b_{34} + b_{44})x_4]$$

5 Algorithm SPV(B; M, W, T; s, t; $x^{(0)}$; ε)

Step 1. Set $k = 0$***.***

Step 2. Construct the **aggregated matrix** ***(in case*** $s = 1$ ***irreducibility of*** B ***implies that of*** $\mathcal{B}(x^{(k)})$***)***

$$\mathcal{B}(x^{(k)}) = RB^s S(x^{(k)})$$

Step 3. Find the unique stationary probability vector $z^{(k)}$ ***from***

$$\mathcal{B}(x^{(k)})z^{(k)} = z^{(k)},\ e(p)^T z^{(k)} = 1,\ e(p) = (1, ..., 1)^T \in \mathcal{R}^p.$$

Step 4. Let

$$Mx^{(k+1,m)} = Wx^{(k+1,m-1)},\ x^{(k+1,0)} = x^{(k)},\ m = 1, ..., t,$$

$$x^{(k+1)} = x^{(k+1,t)},\ e(N)^T x^{(k+1)} = 1.$$

Step 5. Test whether

$$\|x^{(k+1)} - x^{(k)}\| < \epsilon.$$

Step 6. If NO in Step 6, then let

$$k + 1 \rightarrow k$$

and GO TO Step 2.
Step 7. If YES in Step 6, then set

$$\hat{x} := x^{(k+1)}$$

and STOP.

A class of stochastic matrices for which IAD converges fast

Another special stochastic matrix

$$B_{dyad} = \begin{pmatrix} B_{11} & B_{12} & . & . & . & B_{1p} \\ B_{21} & B_{22} & . & . & . & B_{2p} \\ . & . & . & . & . & . \\ B_{p1} & B_{p2} & . & . & . & B_{pp} \end{pmatrix}$$

where

$$B_{jj}, j = 1, ..., p, \quad \text{arbitrary substochastic}$$

and

$$B_{jk} = v_j \, u_{jk}^T, j \neq k, \quad \text{rank one matrices}$$

$$v = \begin{pmatrix} v_1 \\ 0 \\ . \\ . \\ . \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ v_2 \\ . \\ . \\ . \\ 0 \end{pmatrix} + ... \begin{pmatrix} 0 \\ 0 \\ . \\ . \\ . \\ v_p \end{pmatrix}, \quad v_j > 0, \; j = 1, ..., p.$$

Let B_{dyad} be irreducible.

Then the IAD algorithm $SPV(B; I - B_{diag}, B_{off}; s, t; x^{(0)}; \varepsilon)$ returns as $x^{(2)}$ the exact stationary probability vector $\hat{x}$ having the form

$$\hat{x} = v = x^{(2)}.$$

Error-vector

$$x^{(k+1)} - \hat{x} = J_t(x^{(k)})\left(x^{(k)} - \hat{x}\right), \tag{4.1}$$

where

$$J_t(x) = T^t\left[I - P(x)Z\right]^{-1}\left(I - P(x)\right), \tag{4.2}$$

where Z **comes from the spectral decomposition of** $B = Q + Z, Q^2 = Q, QZ = ZQ = 0, 1 \notin \sigma(Z)$**. Furthermore,** $J_t(x) = T^{t-1}J_1(x),\ t \geq 1$**, holds for any** x **with all components positive.**

6 Definition **A splitting $\{M, W\}$ of $I - B = M - W$ is called *aggregation-convergent* if $T = M^{-1}W$ is Y-zero-convergent, where $Y = I - P(\hat{x})$ and $P(\hat{x})$ is the aggregation projection .**

7 Example ***Let α, β be complex numbers, $|\alpha| \geq 1$.***

$$T = \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix},$$

and

$$Y = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

We see that T is Y-zero convergent if $|\beta| < 1$ and Y-convergent if $\beta = 1$. If $\alpha \neq 1$ then obviously, $\lim_{k \to \infty} T^k$ does not exist.

8 Proposition ***Let sequence*** $\{x^{(k)}\}$ ***be computed according to*** **Algorithm SPV(B; M, W, T; s, t;** $x^{(0)};\varepsilon$**) ??.**
Then the characteristics of the error estimates

$$\rho_{\mathcal{E}}(S(\hat{x})z^{(k)},\hat{x}) = \min\left\{\sum_{j=1}^{N}\left|\frac{(S(\hat{x})z^{(k)})_j}{(\hat{x})_j} - \mu\right| : \mu \in \mathcal{R}^1\right\}.$$

and

$$\rho_{\mathcal{F}}(z^{(k)}, z^*) = \min\left\{\sum_{\bar{j}=1}^{n}\left|\frac{(z^{(k)})_{\bar{j}}}{(z^*)_{\bar{j}}} - \mu\right| : \mu \in \mathcal{R}^1_+\right\},$$

are related by

$$\underline{d}\,\rho_{\mathcal{F}}(z^{(k)}, z^*) \le \rho_{\mathcal{E}}(S(\hat{x})z^{(k)}, \hat{x}) \le \overline{d}\,\rho_{\mathcal{F}}(z^{(k)}, z^*),$$

with $d_{\overline{j}}$, $\overline{j} = 1, ...n$, $\underline{d}$ ***and*** $\overline{d}$ ***given by***

$$d_{\overline{j}} = \mathrm{card}\left\{ j : G(j) = \overline{j} \right\}.$$

and

$$\underline{d} = \min d_{\overline{j}}, \quad \overline{d} = \max d_{\overline{j}}.$$

respectively.

9 Proposition ***Let*** N ***be a positive integer and*** $v \in \mathcal{R}^N, v^T = (v_1, ..., v_N) = (v_{(1)}^T, ..., v_{(p)}^T), v_{(j)} \in \mathcal{R}^{n_j}, 1 \leq n_j \leq N, \Sigma_{j=1}^p n_j = N$ ***and*** $e^T = (1, ..., 1) = (e(n_1)^T, ..., e_(n_p)^T) \in \mathcal{R}^N$***. With*** $B = B(\alpha) = \alpha B_1 + (1-\alpha) B_2, 0 < \alpha < 1$***, where*** B_1, B_2 ***are column stochastic blockwise written matrices,***

$$B_1 = \begin{pmatrix} B_{11} & B_{12} & . & . & . & B_{1p-1} & B_{1p} \\ B_{21} & B_2 & . & . & . & B_{2p-1} & B_{2p} \\ . & . & . & . & . & . & . \\ B_{p1} & B_{p2} & . & . & . & B_{pp-1} & B_{pp} \end{pmatrix}$$

and

$$B_{2off} = \begin{pmatrix} 0 & 0 & . & . & . & 0 & v_1 e(n_p)^T \\ v_2 e(n_1)^T & 0 & . & . & . & 0 & 0 \\ . & . & . & . & . & . & . \\ 0 & 0 & . & . & . & v_p e(n_{p-1})^T & 0 \end{pmatrix}$$

we define iteration matrices

$$T_1 = (I - B_{1diag})^{-1} B_{10ff},$$

where B_{1diag} ***is the block diagonal of*** B_1***,*** $B_{1off} = B_1 - B_{1diag}$ ***and*** $B_2 = B_{1off} + B_{2off}$***. Thus,***

$$T_2 = (I - B_{1diag})^{-1} B_{2off}.$$

Then, assuming $\alpha > 1/2$***, we have that***

$$\gamma(\alpha T_1 + (1-\alpha)T_2) = \alpha. \tag{4.3}$$

Proof Let us denote $T(\alpha) = \alpha T_1 + (1-\alpha)T_2 = (I - B_{1off})^{-1} \left[\alpha B_{+off} + (1-\alpha)B_{2off}\right]$**.**

The hypotheses guarantee that B **is irreducible for any** $\alpha > 0$ **and thus, the null space** $\mathcal{N}(I - B)$ **is one-dimensional. It follows that the null space** $\mathcal{N}(I - T)$ **is one-dimensional as well.**

It is easy to see that $T(\alpha)$ as a convex combination of two cyclic matrices need not be cyclic. However, we are going to assume that it remains cyclic. Note that the indices of cyclicity q and p of $T(\alpha)$ and T_2 may be different.

It is well known [1] that the Cesaro sums of any homogeneous Markov chain with transition matrix B are convergent. As demonstrated above, matrix $T(\alpha)$ is generated by a regular splitting of $I - B(\alpha)$. Irreducibility of $B(\alpha)$ then implies that also the Cesaro sums

$$C_n = \frac{1}{n} \sum_{k=1}^{n} T(\alpha)^k$$

are convergent and the limit is the Perron projection of $T(\alpha)$:

$$Q_1(\alpha) = \lim_{k\to\infty} C_n.$$

Since the range of $Q_1(\alpha)$ is one-dimensional, we can identify the form of this projection. We derive that

$$Q_1(\alpha) = \hat{y} f^T, \tag{4.4}$$

where $\hat{y}$ is the unique Perron eigenvector of $T(\alpha)$ and $f = (I - B_{1off}^T)e, e^T = (1, ..., 1) \in \mathcal{R}_+^N$.

The validity of relation (4.3) is based on an important result from [2] and the following interesting spectral property formulated in a form of a Lemma.

10 Lemma ***The iteration matrix*** $T(\alpha)$ ***possesses the following spectral decomposition:***

$$T(\alpha) = \sum_{j=0}^{q-1} \mu^j Q_{j+1}(\alpha) + Z(\alpha), \quad \mu = \exp\{2\pi i/q\}, \tag{4.5}$$

where

$$Q_1(\alpha) = \alpha \sum_{t=1}^{r} \tilde{Q}_{1;t,1} + (1-\alpha)\tilde{Q}_{2;1} = \hat{y} f^T, \; f^T \hat{y} = 1, \tag{4.6}$$

with

$$f = \left(I - B_{1off}^{T}\right) e \tag{4.7}$$

$$Q_j(\alpha) = \hat{y}_j f_j^T,\ f_j^T \hat{y}_j = 1,\ j = 2, ..., q, \tag{4.8}$$

with

$$\begin{cases} \hat{y}_j^T & = \left(\mu^j \hat{y}_{j(1)}^T, ..., \mu^{jq} \hat{y}_{j(q)}^T\right) \\ \\ f_j^T & = \left(\mu^{-j} f_{j(1)}^T, ..., \mu^{-jq} f_{j(q)}^T\right) \end{cases} \tag{4.9}$$

and

$$\begin{aligned} Z(\alpha) = {} & \alpha \Sigma_{t=1}^{r} \Sigma_{j=1}^{p_t - 1} \lambda_t^j \left(I - Q_1(\alpha)\right) \tilde{Q}_{1;t,j+1} \left(I - Q_1(\alpha)\right) \\ & + \alpha \left(I - Q_1(\alpha)\right) Z_1 \left(I - Q_1(\alpha)\right) \end{aligned} \tag{4.10}$$

With the above Lemma 4.5 the algorithms investigated in [6] apply and their convergence is guaranteed. With suitable splitting of the transition matrix $B(\alpha)$**,**

e.g. such as shown in our contribution, the speed of convergence can be characterized using the following

11 Corollary ***The spectrum*** $\sigma(T(\alpha))$ ***of the iteration matrix*** $T(\alpha)$ ***is such that***

$$\sigma(T(\alpha)) \subset \{1, \mu, ..., \mu^{q-1}\} \cup \cup_{k=1}^{r} \cup_{j=1}^{p_k} \{\alpha\lambda_k^{j-1}\}$$

$$\cup \{\lambda : |\lambda| \leq \rho, \rho < \alpha\},$$

$$\lambda_k = \exp\{2\pi i/p_k\}, j = 1, ..., r.$$

12 Remark ***It follows that the speed of convergence of the sequence obtained in the process of iterative aggregation/disaggregation method proposed and analyzed***

in [6] is determined by value α, i.e. for any norm on $\mathcal{R}^N$ we have the following estimate

$$\left\| x^{(k)} - \hat{y} \right\| \leq \kappa \alpha^k,$$

with some constant κ independent of k.

5

Konvexní kombinace stochastických matic

$$B(\alpha) = \alpha B_1 + (1-\alpha)B_2,\ 0 < \alpha < 1$$

Pro případ mocninné metody se používá pro poruch B_2 nějaké matice hodnosti 1. Pro ni platí vztahy

$$(5.1)\qquad \begin{cases} \sigma(B_2) = \{0,1\}, \\ \mathcal{N}(I - B_2) = \{\kappa\hat{x} : \kappa \in \mathcal{R}^1\}, \\ \lim_{k\to\infty} B_2^k x^{(0)} = x^{(1)} = \hat{x},\ e^T x^{(0)} = 1, \end{cases}$$

jejichž důsledkem je konvergence mocninné metody pro jakoukoli konvexní kombinace za předpokladu, že parametr $\alpha \in (0,1)$.

Ukážeme, že při volbě IAD pro nalezení $\hat{x}$ je výhodné brát za B_2 "dyadickou" a přitom cyklickou irreducibilní matici. Pro případ této volby existují některé IAD metody jež konvergují rychlostí danou touže veličinou, tedy, rychlostí α^k.

$$B_{dyad} = \begin{pmatrix} B_{11} & 0 & \ldots & 0 & B_{1p} \\ B_{21} & B_{22} & \ldots & 0 & 0 \\ . & . & \ldots & . & . \\ 0 & 0 & \ldots & B_{pp-1} & B_{pp} \end{pmatrix}$$

s libovolnými diagonálními bloky B_{jj} a *dyadickými* mimodiagonálními bloky

$$B_{1p} = \beta_1 v_1 e^T(n_p), v_1 \in \mathbf{Int}(\mathcal{R}^{n_1}), \beta_1 > 0,$$

$$B_{jj-1} = \beta_j v_j e^T(n_{j-1}), v_j \in \mathrm{Int}(\mathcal{R}_+^{n_j}), \beta_j > 0, j = 1, ..., p.$$

Pro vhodné hodnoty parametrů β_j evidentně je takto vybraná matice blokově p-cyklická a irreducibilní.

IAD analogy relací (5.1) jsou tyto vztahy

$$\begin{cases} \sigma\left([I - P(\hat{x})]J(\hat{x})\right) = \{0, 1, \mu, ..., \mu^{p-1}\}, \\ \mathcal{N}(I - B_2) = \{\kappa v : \kappa \in \mathcal{R}^1\}, \\ \lim_{k\to\infty} x^{(k)} = x^{(2)} = v. \end{cases}$$

The above idea how to compute the stationary distribution of an irreducible cyclic matrix will now be applied to computing the so called PageRank of the Google matrix. As well known [5], the PageRank is computed on the basis of a knowledge of the unique stationary probability distribution of the Google matrix. Our main result consists of establishing convergence of an IAD procedure to the PageRank and its efficiency concerning the realization on parallel computer architecture. From theoretical point of view cyclicity not only does not matter, just the opposite: cyclicity of the Google matrix is exploited in order to accelerate the computations in comparison with the methods utilized so far.

Modelem bude konvexní kombinace "původní" Googlovské matice a "vhodné" cyklické matice dyadické

$$B(\alpha) = \alpha B_1 + (1 - \alpha)B_2,\ \alpha > 1/2.$$

Výpočet budeme provádět pomocí SPV$(B(\alpha); I-$**diag**$\{B_{11}, ..., B_{pp}\}, \alpha B_{1off}+(1-\alpha)B_{2off}; t, 1; x^{(0)}; \varepsilon)$**. Iterační proces, na němž je vybudována metoda IAD je dán iterační maticí**

$$T(\alpha) = (I - \text{diag}\{B_{11}, ..., B_{pp}\})^{-1}[\alpha B_{1off} + (1 - \alpha)B_{2off}].$$

neboli,

$$T(\alpha) = \alpha T_1 + (1 - \alpha)T_2,$$

kde

$$T_j = \left(I - \text{diag}\{B_{11}, ..., B_{pp}\})\right)^{-1} B_{joff}, \; j = 1, 2.$$

Acknowledgements **Research on the basis of which this contribution is based has been partly supported by the Program Information Society under Project 1ET400300415. The support is sincerely appreciated.**

[1] **Berman A., Plemmons R.J.** ***Nonnegative Matrices in the Mathematical Sciences.*** **Academic Press New York-San Francisco-London 1979.**

[2] **Courtois P.J., Semal P.** ***Block iterative algorithms for stochastic matrices*****, Linear Algebra Appl. 76, 59-80 (1986).**

[3] **Havelivala T., Kamvar S., Klein D., Manning C., Golub G.** ***Computing PageRank using power extrapolation.*** **Preprint.**

[4] **Ipsen I., Kirkland S.** ***Convergence analysis of an improved pagerank algorithm.*** **Preprint.**

[5] **Langville A.N., Meyer C. D.** ***Deeper inside PageRank.*** **Preprint.**

[6] Marek I., Mayer P. *Convergence theory of a class of aggregation/ disaggregation iterative methods for computing stationary probability vectors of stochastic matrices*. Linear Algebra Appl. 363, 177-200, 2002).

[7] Moler C. *The world's largest matrix computation*. The Math Works, Cleve's Corner, 2004, 3 pp. http://matesco.unican.es/Cleve's/Corner/TheWorld...