

FOURIER ANALYSIS OF ITERATIVE AGGREGATION- DISAGGREGATION METHODS FOR NEARLY CIRCULANT STOCHASTIC MATRICES

Ivana Pultarová

Czech Technical University in Prague, Faculty of Civil Engineering
Thákurova 7, 166 29 Praha 6, Czech Republic
ivana@mat.fsv.cvut.cz

Abstract

We introduce a new way of the analysis of iterative aggregation-disaggregation methods for computing stationary probability distribution vectors of stochastic matrices. This new approach is based on the Fourier transform of the error propagation matrix. Exact formula for its spectrum can be obtained if the stochastic matrix is circulant. Some examples are presented.

1. Introduction

Iterative aggregation-disaggregation (IAD) methods are a popular tool for numerical solution of stationary probability distribution vectors of stochastic matrices: they search for a sufficiently good approximation of x fulfilling

$$Bx = x, \quad e^T x = 1, \quad (1)$$

where B is an irreducible column stochastic matrix and e is a vector of all ones. B is column stochastic if $B \geq 0$ and $e^T B = e^T$. It is well known that the solution x exists, is unique and positive [12].

The IAD methods work in a multilevel fashion. A set of aggregation groups of unknowns is chosen. Each group represents one unknown on the coarse level. A solution of the coarse problem is used for improving the approximate solution of the original problem on the fine level. The idea is similar to the classical algebraic multigrid (AMG) used for the solution of symmetric positive definite (SPD) problems [1, 2, 3, 4, 5, 7, 13]. The main difference is caused by the nonsymmetry of stochastic matrices. While for the AMG methods the estimates in a corresponding energy norm are utilized, the theoretical justifying the convergence of the IAD methods exploit completely different approaches. Unfortunately, there are no convergence conditions for general IAD methods and for general stochastic matrices. In spite of this, there are many numerical experiments confirming good efficiency of various

IAD methods. The aim of this paper is to provide a theoretical background for some observations made e.g. in [1, 3, 4, 5].

Let $[B]_{rs}$ denote the element of B in the row r and column s , similarly $[x]_r$ is the r th element of vector x . If B is nonsymmetric, the preferable algorithm of aggregation of unknowns into aggregation groups is according to their strong connection [1, 3, 4, 5]: the unknowns $[x]_r$ and $[x]_s$ are strongly connected if $[B]_{rs} + [B]_{sr} \gg 0$. Then the IAD methods are reported to converge fast. But there is no theoretical background given in the literature. In this paper we consider a special $N \times N$ stochastic matrix B , where

$$[B]_{rs} = 1 \quad \text{if} \quad (r - s - 1) \bmod N = 0, \quad \text{and} \quad [B]_{rs} = 0 \quad \text{otherwise.} \quad (2)$$

Adding small perturbations to B gives rise to typical examples of slowly mixing stochastic matrices for which the stationary iterative methods converge slowly. Such matrices appear for example in queuing network applications. At the same time B is a circulant matrix. While the stationary probability distribution of B is $x = e/N$, the solution for perturbations of B are not known a priori. But from the continuity, similar quality is achieved for perturbations of B . Motivated by the Fourier transform of AMG operators for circulant and Toeplitz SPD matrices [2], we use the Fourier transform for the IAD methods and for circulant matrices. A scope of this paper allows us to consider only two-level IAD methods. Our particular goal is to find the optimal parameters in the IAD methods for B defined by (2).

The paper is organized as follows. In the next section the IAD methods and the error propagation formula are recalled. In Section 3 the Fourier transform is used for the error propagation matrix and its spectrum is computed. The optimal IAD parameters are computed in Section 4. A short discussion concludes the paper.

2. Two-level IAD methods

Let us assume an irreducible $N \times N$ stochastic matrix B . Let pairwise disjoint aggregation groups G_1, \dots, G_n be chosen, $\cup_{k=1}^n G_k = \{1, \dots, N\}$. Then a reduction matrix $R \in \mathcal{R}^{n \times N}$ is given by

$$\begin{aligned} [R]_{ij} &= 1 \quad \text{if} \quad j \in G_i, \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

A prolongation matrix $S(y) \in \mathcal{R}^{N \times n}$ is defined for any positive vector $y \in \mathcal{R}^n$ by

$$\begin{aligned} [S(y)]_{ij} &= \frac{y_i}{\sum_{k \in G_j} y_k} \quad \text{if} \quad i \in G_j, \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

Matrix $B_1 = RBS(y)$ is an aggregated matrix corresponding to B and y . Of course, $P(y) = S(y)R$ is a projection.

On the fine level, μ steps of some stationary iteration (we call it a basic iteration) with matrix T are performed. We use Richardson iteration with $T = \alpha B + (1 - \alpha)I$, where I is the identity matrix and $\alpha \in (0, 1)$. A solution of the coarse problem with matrix B_1 is carried out exactly. One cycle of the IAD method is as follows.

One cycle of the IAD method: input $x^m > 0$; output x^{m+1} .

1. set $B_1 := RBS(x^m)$ and solve $B_1 z = z$, $e^T z = 1$ (coarse step)
2. $y := S(x^m)z$ (prolongation)
3. $x^{m+1} := T^\mu y$ (basic iterations)

It can be easily shown that the exact solution x is a fixed point of this computing process. Moreover, the error of the approximation x^{m+1} is

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

[6], where

$$J(x^m) = T^\mu(I - P(x^m)(B - xe^T))^{-1}(I - P(x^m)). \quad (3)$$

Since spectral radii $\rho(J(x^m))$ are greater than one in general, we can study the asymptotic (local) convergence properties by substituting the exact solution into (3) instead of x^m and computing the spectral radii of $J(x)$. We say that the IAD method is locally convergent if there exists a neighborhood U of x such that for any $x^0 \in U$, the IAD method yields a convergent sequence with a limit x . A sufficient condition for the local convergence is of course $\rho(J(x)) < 1$.

3. Fourier transform of the error propagation formula

The spectral analysis of the AMG methods for circulant and Toeplitz matrices is based on the Fourier transform of the error propagation operator [2]. We apply this idea to the IAD methods and compute spectra of matrices $J(x)$ given by (3) if the stochastic matrix B is circulant. As the first type we consider B defined by (2). According to Theorem 1 a spectrum of $J(x)$ can be expressed exactly which helps us to see what are the values of μ and α resulting in the smallest $\rho(J(x))$. Adding small perturbations to B does not change the convergence rates of the IAD method significantly. Such matrices represent a kind of slowly mixing Markov chains [12]. For the sake of simplicity we consider $n = N/2$ and $G_k = \{2k-1, 2k\}$, $k = 1, \dots, n$, which corresponds to the aggregation of unknowns according to their strong connections.

Let us denote the $N \times N$ Fourier matrix by F_N , where

$$[F_N]_{rs} = \frac{1}{\sqrt{N}} e^{-2\pi(r-1)(s-1)i/N}.$$

The superscript H indicates the adjoint matrix.

Theorem 1. Let B be defined by (2). Assume the IAD method with the basic iteration matrix $T = \alpha B + (1 - \alpha)I$, $\alpha \in (0, 1)$, and with μ steps of basic iterations in each cycle. Let the aggregation groups be $G_k = \{2k - 1, 2k\}$, $k = 1, \dots, n$, $n = N/2$. Then the spectrum of the error propagation matrix $J(x)$ is

$$\sigma(J(x)) = \{0, v_0, v_1, \dots, v_{n-1}\},$$

where

$$v_k = \frac{1}{2} \left((1 - e^{2\pi ki/N}) (1 - \alpha + \alpha e^{-2\pi ki/N})^\mu + (1 + e^{2\pi ki/N}) (1 - \alpha - \alpha e^{-2\pi ki/N})^\mu \right). \quad (4)$$

Proof. The proof aims to compute the spectra of $F_N^H J(x) F_N$. We show only two crucial points of the proof. The first one is the well known formula

$$B = F_N D F_N^H,$$

where D is diagonal and $[D]_{rr} = e^{2\pi(r-1)i/N}$. The second one is that for the exact solution $x = e/N$

$$P(x) = \frac{1}{2} R^T R = \frac{1}{4} F_n \begin{pmatrix} \tilde{D}_1 & 0 \\ 0 & \tilde{D}_2 \end{pmatrix} \begin{pmatrix} I & I \\ I & I \end{pmatrix} \begin{pmatrix} \tilde{D}_1^H & 0 \\ 0 & \tilde{D}_2^H \end{pmatrix} F_n^H, \quad (5)$$

where the matrices \tilde{D}_1 and \tilde{D}_2 are diagonal and $[\tilde{D}_1]_{rr} = 1 + e^{2\pi(r-1)i/N}$ and $[\tilde{D}_2]_{rr} = 1 - e^{2\pi(r-1)i/N}$, $r = 1, \dots, n$. Find more about this technique in [2]. \square

Though the spectrum of $J(x)$ is computable for B defined by (2), it is not straightforward to simplify the term (4) for an arbitrary μ .

4. Optimal parameters μ and α

Under the assumptions of Theorem 1 let $\mu \in \{1, 2, 3\}$. Let the spectra of the corresponding matrices $J(x)$ be $\sigma_1, \sigma_2, \sigma_3$ and the spectral radii ρ_1, ρ_2, ρ_3 . Then

$$\begin{aligned} \sigma_1 &= \{0, 1 - 2\alpha\}, \\ \sigma_2 &= \{0\} \cup \left(\alpha^2 M + (1 - \alpha)(1 - 3\alpha) \right), \\ \sigma_3 &= \{0\} \cup \left((3\alpha^2 - 4\alpha^3)M + (1 - \alpha)^2(1 - 4\alpha) \right), \end{aligned}$$

where $M = \{e^{-4\pi ki/N}\}_{k=0}^{n-1}$.

For $\alpha \approx 1$ we have $\rho_3 < \rho_1 < \rho_2$, see also Figure 1. Thus in case of B nearly of the type (2) and of the aggregation groups with two elements strongly connected, and for $T = \alpha B + (1 - \alpha)I$, $\alpha \approx 1$, the most advantageous number of basic iterations (among 1, 2, 3) in every IAD cycle is $\mu = 3$.

Theorem 1 also allows to find the best parameter α if μ is given. Note that it does not depend on N . For example, for $\mu = 1$ the best is $\alpha = 1/2$ which leads to $\rho_1 = 0$. For $\mu = 2$ the best spectral radius is $\rho_2 = 1/9$ for

$$\alpha = \arg \min_{\alpha \in (0,1)} \max(|(1 - \alpha)(1 - 3\alpha) + \alpha^2|, |(1 - \alpha)(1 - 3\alpha) - \alpha^2|) = 1/3.$$

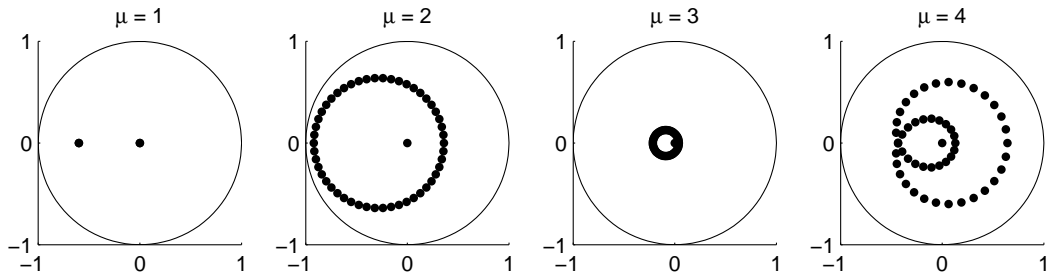


Figure 1: Eigenvalues of $J(x)$ for B defined by (2), $x = e/N$, $N = 100$, aggregation groups $G_k = \{2k - 1, 2k\}$, $k = 1, \dots, N/2$, parameters $\alpha = 0.8$ and $\mu \in \{1, 2, 3, 4\}$. The solid line is a reference unit cycle.

5. Discussion

We contribute to the theory of the IAD methods. Our results are applicable to the theory of the AMG for nonsymmetric problems as well. The introduced approach is based on the Fourier transform.

The introduced analysis can be generalized in several directions. More than two elements in each aggregation group can be considered. Then instead of the 2×2 block form in (5) we get an $m \times m$ block form if m elements are contained in every aggregation group. Also block-circulant matrices can be studied [2].

We would like to emphasize that the local convergence of the IAD methods is not necessarily obtained in general [8]. There are several examples where the spectral radius of $J(x)$ can be arbitrarily large [10]. It was shown that even B in the form (2) can yield the spectral radius of $J(x)$ arbitrarily close to two [9]. These examples should be understood and avoided in the real life computation.

A promising utilization of our approach is in the theory of multi-level IAD methods. Presently we are not able to find any exact criteria for the local convergence of the IAD methods with more than two levels. Our new approach could simplify the involved formulae [11] and help us to find the optimal IAD parameters for at least some special stochastic matrices.

Acknowledgement

The author thanks to the anonymous referee for his/her careful reading the manuscript and for his/her comments.

References

- [1] Bolten, M., Brandt, A., Brannick, J., Frommer A., Kahl K., and Livshits I.: A bootstrap algebraic multilevel method for Markov chains. *SIAM J. Sci. Comput.* **33** (2011), 3425–3446.

- [2] Bolten, M., Donatelli, M., and Huckle, T.: Aggregation-based multigrid methods for circulant and Toeplitz matrices. Preprint BUW-IMACM 12/10, Bergische Universität Wuppertal, 2012.
- [3] Brezina, M., Manteuffel, T., McCormick, S., Ruge, J., and Sanders, G.: Towards adaptive smoothed aggregation (α SA) for nonsymmetric problems. *SIAM J. Sci. Comput.* **32** (2010), 14–39.
- [4] De Sterck, H., Manteuffel, T. A., McCormick, S. F., Miller, K., Ruge, J., and Sanders, G.: Algebraic multigrid for Markov chains. *SIAM J. Sci. Comput.* **32** (2010), 544–562.
- [5] De Sterck, H., Miller, K., Treister, E., and Yavneh, I.: Fast multilevel methods for Markov chains. *Numer. Linear Algebra Appl.* **18** (2011), 961–980.
- [6] Marek, I. and Mayer, P.: Convergence analysis of an iterative aggregation/disaggregation method for computing stationary probability vectors of stochastic matrices. *Numer. Linear Algebra Appl.* **5** (1998), 253–274.
- [7] Notay, Y.: Algebraic analysis of two-grid methods: The nonsymmetric case. *Numer. Linear Algebra Appl.* **17** (2010), 73–96.
- [8] Pultarová, I.: Necessary and sufficient local convergence condition of one class of iterative aggregation-disaggregation methods. *Numer. Linear Algebra Appl.* **15** (2008), 339–354.
- [9] Pultarová, I.: Ordering of matrices for iterative aggregation-disaggregation methods. *Lecture Notes in Control and Inform. Sci.* **389** (2009), 379–386.
- [10] Pultarová, I., Marek, I.: Physiology and pathology of iterative aggregation-disaggregation methods. *Numer. Linear Algebra Appl.* **18** (2011), 1051–1065.
- [11] Pultarová, I.: Error propagation formula of multi-level iterative aggregation-disaggregation methods for non-symmetric problems. *Electron. J. Linear Algebra* **25** (2012), 9–21.
- [12] Stewart, W. J.: *Introduction to the numerical solution of Markov chains*. Princeton University Press, Princeton, 1994.
- [13] Vaněk, P., Brezina, M., and Mandel, J.: Convergence of algebraic multigrid based on smoothed aggregation. *Computing* **56** (1998), 179–196.