Advances in Applied Mathematics 应用数学进展, 2017, 6(4), 651-658
Published Online July 2017 in Hans. http://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2017.64076

Optimization of Preconditioner in Conjugate Gradient Method

Cunzhu Guo

School of Mathematics & Information, Longnan Teacher's College, Longnan Gansu
Email: guocunzhu@163.com

Received: Jul. 9th, 2017; accepted: Jul. 24th, 2017; published: Jul. 27th, 2017

Abstract

For decreasing the conditional number of the coefficient matrix in solving the linear equations with conjugate gradient methods and accelerating the convergence, it is common to use preconditioned methods to find the equivalent equations, whose conditional numbers are smaller. It is required that the preconditioners should be as close as possible to the original coefficient matrix and their inverse matrices can be easily computed. Starting from diagonal preconditioners, we first compute the eigenvalue decomposition of the coefficient matrix, and obtain the optimal preconditioner. However, it is of high computational complexity to do the eigenvalue decomposition. In this paper, we introduce three *p*-norm preconditioners to approximate the optimal preconditioner. Comparing with the existing preconditioners, the experimental results show that the proposed three diagonal *p*-norm preconditioners converge much faster, which demonstrates the advantages of the proposed family of preconditioners.

Keywords

Preconditioner, Condition Number, Eigenvalue Decomposition, *p*-Norm Preconditioner

共轭梯度法中预条件子的优化

郭存柱

陇南师专数信学院，甘肃 陇南
Email: guocunzhu@163.com

收稿日期：2017年7月9日；录用日期：2017年7月24日；发布日期：2017年7月27日

摘 要

为了降低方程组求解中共轭梯度法系数矩阵的条件数，提高收敛速度，常用预处理方法将原方程进行等

文章引用: 郭存柱. 共轭梯度法中预条件子的优化[J]. 应用数学进展, 2017, 6(4): 651-658.
https://doi.org/10.12677/aam.2017.64076

价转化，同时预条件子既要接近原系数矩阵，又要容易求其逆矩阵。本文从寻求对角预条件子出发，用矩阵的特征值分解方法解出了预处理后系数矩阵特征值矩阵的显式表达，得到对角预条件子矩阵的最优选择，并予以证明。给出了三个*p*-范数预条件子，将之与常用的预条件子进行对比，实例检验表明三个*p*-范数预条件子的作用更优越，且使算法收敛更快。

关键词

预条件子，条件数，特征值分解，*p*-范数预条件子

1. 引言

使用求解方程组 $Ax=b$ (其中 A 为 n 阶对称正定矩阵，x 为 $n\times 1$ 未知向量，b 为 $n\times 1$ 已知向量)，理论上可以在 n 步之内得到精确解。但由于机器的舍入误差，可能需迭代超过 n 步才能达到所需精度的解。尤其当 A 的条件数很大时，迭代求解过程总体收敛会很慢，因之这一优越的方法在一段时间曾被搁置。直到引入预条件处理方法(用 P^{-1} 乘以原方程组，得到同解方程组 $P^{-1}Ax=P^{-1}b$)，来降低系数矩阵的条件数($\mathrm{cond}\left(P^{-1}A\right) << \mathrm{cond}(A)$)，提高了迭代速度和稳定性，才使共轭梯度法在工程、微分方程数值求解和最优化等领域得到广泛应用。

2. 已有成果[1]

预条件子(亦称条件预优矩阵)P 的选取应接近 A，且容易求逆，正对角阵为首选。此前共轭梯度法常用的预条件子有以下三种。

2.1. Jacobi 预条件子

取 $P=D$，其中 D 为 A 的对角阵。

2.2. Gauss-Seidel 预条件子

取 $P=(D+\omega L)D^{-1}(D+\omega U)$，其中 A 被分解为 $A=D+L+U$，D、L、U 为对角阵、下三角阵、上三角阵，$\omega\in(0,2)$。当 $\omega=1$ 时即为 Gauss-Seidel 预条件子。

2.3. 不完全 Cholesky 分解预条件子[2]

由 Meijerink J.A.和 Van der vorst A.1977 年提出[3]。将 A 分解成 $A=LL^{\mathrm{T}}+R$，其中 L 为下三角阵，且与 A 有一样的稀疏性(当 $A_{ij}=0$ 时，有 $L_{ij}=0$)。

取 $P=L^{\mathrm{T}}L$，或 $P=L$。

3. 目标问题

已知矩阵 A，求矩阵 P，使得 $P^{-1}A$ 的条件数最小，即

$$\hat{P}=\arg\min_{P} cond\left(P^{-1}A\right)$$

4. 概念、定理与符号

4.1. 定义(谱条件数)

设正定矩阵 A 的特征值分别为 $\lambda_i > 0$， $n = 1, 2, \cdots, n$， 则 A 的谱条件数为 $condA = \dfrac{\lambda_{\max}}{\lambda_{\min}}$。

4.2. 矩阵的谱分解定理

设 n 阶矩阵 A 的特征值分别为 $\lambda_1, \lambda_2, \cdots, \lambda_n$，相应的特征向量为 $\gamma_i, i = 1, 2, \cdots, n$ 为标准正交向量组，即

$$\gamma_i^{\mathrm{T}} \gamma_j = \left(\gamma_i, \gamma_j\right) = \begin{cases} 1, \ i = j \\ 0, \ i \neq j \end{cases}$$

记 $\Lambda = diag\left(\lambda_1, \lambda_2, \cdots, \lambda_n\right)$， $\Gamma = \left(\gamma_1, \gamma_2, \cdots, \gamma_n\right)$，其中， $diag\left(\lambda_1, \lambda_2, \cdots, \lambda_n\right)$ 表示对角元素分别为 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 的对角矩阵，则 A 可分解为

$$A = \sum_{i=1}^{n} \lambda_i \gamma_i \gamma_i^{\mathrm{T}} = \Gamma \Lambda \Gamma^{\mathrm{T}} \tag{1}$$

4.3. $P^{-1}A$ 的谱分解

设 $P^{-1}A$ 的特征值分别为 $\mu_i > 0$， 相应的特征向量为 $\delta_j, j = 1, 2, \cdots, n$ 为标准正交向量组，即

$$\delta_j^{\mathrm{T}} \delta_i = \left(\delta_j, \delta_i\right) = \begin{cases} 1, \ j = i \\ 0, \ j \neq i \end{cases}$$

记 $\mathrm{M} = diag\left(\mu_1, \mu_2, \cdots, \mu_n\right)$， $\Delta = \left(\delta_1, \delta_2, \cdots, \delta_n\right)$，
由矩阵的谱分解定理

$$P^{-1}A = \sum_{j=1}^{n} \mu_j \delta_j \delta_j^{\mathrm{T}} = \Delta \mathrm{M} \Delta^{\mathrm{T}} \tag{2}$$

5. 求解

首先想到的是 $P = A$，因 $cond\left(P^{-1}A\right) = cond\left(A^{-1}A\right) = cond\left(I\right) = 1$，条件数最小。但求 A^{-1} 和解方程组是同一问题，当 A 的条件数很大时误差会很大，且其时间复杂度为 $O\left(n^3\right)$.

5.1. 最优预条件子

将(1)两边乘以 P^{-1} 得

$$P^{-1}A = \sum_{i=1}^{n} \lambda_i P^{-1} \gamma_i \gamma_i^{\mathrm{T}} = P^{-1} \Gamma \Lambda \Gamma^{\mathrm{T}} \tag{3}$$

由(2)、(3)得

$$\Delta \mathrm{M} \Delta^{\mathrm{T}} = P^{-1} \Gamma \Lambda \Gamma^{\mathrm{T}}$$

解得

$$\mathrm{M} = \Delta^{-1} P^{-1} \Gamma \Lambda \Gamma^{\mathrm{T}} \Delta^{-\mathrm{T}} \tag{4}$$

$P^{-1}A$ 的条件数最小化问题化为

$$\min_P cond\left(P^{-1}A\right) \Leftrightarrow \min_P \frac{\mu_1}{\mu_n} \Leftrightarrow \min_P cond\left(\Delta^{-1} P^{-1} \Gamma \Lambda \Gamma^{\mathrm{T}} \Delta^{-\mathrm{T}}\right) \tag{5}$$

由 $cond(AB) \le cond(A) \cdot cond(B)$，且 AB 与 BA 有相同的非零特征值[4]，

从而， $cond(AB) = cond(BA)$， $cond(ABC) = cond(CAB) = cond(BCA)$

所以 $cond(\Gamma\Gamma^{\mathrm{T}}) = cond(I) = 1$， $cond(\Delta^{-\mathrm{T}}\Delta^{-1}) = cond\left((\Delta\Delta^{\mathrm{T}})^{-1}\right) = cond(I) = 1$

(5)化为

$$\min_P cond\left(P^{-1}\Gamma\Lambda\Gamma^{\mathrm{T}}\right) \tag{6}$$

设 $U = P^{-1}\Gamma\Lambda\Gamma^{\mathrm{T}}$，

由 $\min_P cond(U) \Leftrightarrow \min_P cond(UU^{\mathrm{T}})$，(6)等价于

$$\begin{aligned}
&\min_P cond(U) \Leftrightarrow \min_P cond(UU^{\mathrm{T}}) \\
&\Leftrightarrow \min_P cond\left(P^{-1}\Gamma\Lambda\Gamma^{\mathrm{T}}\Gamma\Lambda\Gamma^{\mathrm{T}}P^{-\mathrm{T}}\right) \Leftrightarrow \min_P cond\left(P^{-1}\Gamma\Lambda\Gamma^{\mathrm{T}}\Gamma\Lambda\Gamma^{\mathrm{T}}P^{-1}\right) \\
&\Leftrightarrow \min_P cond\left(P^{-2}\Gamma\Lambda^{2}\Gamma^{T}\right) \Leftrightarrow \cdots \Leftrightarrow \min_P cond\left(P^{-2m}\Gamma\Lambda^{2m}\Gamma^{\mathrm{T}}\right)
\end{aligned} \tag{7}$$

其中 $m = 1, 2, \cdots$， 要使得(7)条件数最小化为 1， 只需

$$P^{-2m}\Gamma\Lambda^{2m}\Gamma^{\mathrm{T}} \to I, m \to \infty \tag{8}$$

$$\Leftrightarrow P^{-2m}\Gamma\Lambda^{2m} \to \Gamma$$

$$\Rightarrow \left(P^{-2m}\Gamma\Lambda^{2m}\right)_{ii} \to \Gamma_{ii}$$

所以

$$P = \Lambda \tag{9}$$

否则，用反证法。

为简单起见，不妨设 $P_{ii} = p_i$， $p_1 \ge p_2 \ge \cdots \ge p_n > 0$， $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n > 0$， $p_i > \lambda_i > 0$， $i = 1, 2, \cdots, n$，

则 $\left(P^{-2m}\Gamma\Lambda^{2m}\right)_{ij} = \gamma_{ij}\left(\dfrac{\lambda_j}{p_i}\right)^m \to 0$，$(m \to \infty)$， $j \ge i$，

$$\left(P^{-2m}\Gamma\Lambda^{2m}\right) \to \begin{pmatrix} 0 & 0 & \cdots & 0 \\ * & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ * & * & * & 0 \end{pmatrix} \Rightarrow \left(P^{-2m}\Gamma\Lambda^{2m}\Gamma^{\mathrm{T}}\right)_{11} \to 0 \ne 1, m \to \infty$$

若 $\lambda_i > p_i > 0$， $i = 1, 2, \cdots, n$

则 $\left(P^{-2m}\Gamma\Lambda^{2m}\right)_{ij} = \gamma_{ij}\left(\dfrac{\lambda_j}{p_i}\right)^m \to \infty$，$(m \to \infty)$， $j \le i$，

$$\Rightarrow \left(P^{-2m}\Gamma\Lambda^{2m}\Gamma^{\mathrm{T}}\right)_{nn} \to \infty \ne 1, m \to \infty$$

这与(8)矛盾，如此，就证明了(9)成立。

因此，对角预条件子 P 的最优选择是 A 的特征值矩阵 Λ，此时，

$$cond\left(P^{-1}\Lambda\right) = 1$$

为最小。$cond\left(P^{-1}\Gamma\Lambda\Gamma^{\mathrm{T}}\right)$趋向于最小值 1。

5.2. 最优预条件子的近似替代

在数值计算中，当 A 的特征值非常接近于 0 时，由于机器的舍入误差容限，计算机将其按 0 对待，

导致无法求逆。因而，寻找可逆矩阵 P 近似替代 Λ 。

5.2.1. 列范数预条件子

取由 A 的各列的 p-范数为对角元素的矩阵作为预条件子 P，即

$$P_p = diag\left(\left\|A_{,1}\right\|_p, \left\|A_{,2}\right\|_p, \cdots \left\|A_{,n}\right\|_p\right),\ \ p = 1, 2, \cdots$$

其中，$\left\|A_{,j}\right\|_p$ 表示 A 的第 j 列的 p-范数， $j = 1, 2, \cdots, n$ 。

分别取 $p = 1, 2, \infty$ ，可得以下三种不同列范数预条件子。

5.2.2. 列绝对和(1-范数)预条件子

取 $P_1 = diag\left(\left\|A_{,1}\right\|_1, \left\|A_{,2}\right\|_1, \cdots, \left\|A_{,n}\right\|_1\right) = diag\left(\sum_{i=1}^{n}\left|a_{i1}\right|, \sum_{i=1}^{n}\left|a_{i2}\right|, \cdots, \sum_{i=1}^{n}\left|a_{in}\right|\right)$

5.2.3. 列 Euclid(2-范数)预条件子

取 $P_2 = diag\left(\left\|A_{,1}\right\|_2, \left\|A_{,2}\right\|_2, \cdots, \left\|A_{,n}\right\|_2\right) = diag\left(\sqrt{\sum_{i=1}^{n} a_{i1}^2}, \sqrt{\sum_{i=1}^{n} a_{i2}^2}, \cdots, \sqrt{\sum_{i=1}^{n} a_{i2}^2}\right)$

5.2.4. 列最大模(∞-范数)预条件子

取 $P_\infty = diag\left(\left\|A_{,1}\right\|_\infty, \left\|A_{,2}\right\|_\infty, \cdots, \left\|A_{,n}\right\|_\infty\right) = diag\left(\max_{1\le i\le n}\left|a_{i1}\right|, \max_{1\le i\le n}\left|a_{i2}\right|, \cdots, \max_{1\le i\le n}\left|a_{in}\right|\right)$

当 A 为正定对角矩阵时，三个 p-范数预条件子都等于 A。

6. 实例检验

使用 Matlab 的预条件子共轭梯度函数 **pcg** 和双线性共轭梯度函数 **bicg**。

实例 1. 当 A 为 13 阶 Hilbert 矩阵，x = ones(n) (n 个 1); $b = A*x$; cond(A) = 8.3042 × 10^{19}，用 1-范数预条件子 1-步达精确值、∞-范数预条件子 13 步相对误差小于 10^{-16}、2-范数预条件子 17 步相对误差小于 10^{-16}，其次是 Jacobi、Gauss-Seidel 预条件子和不用预条件子，用不完全 Cholesky 预条件子表现最差。如图 1。

实例 2. 当 n = 1000; A = diag((1:n).^2) + diag((1:n-20).^(1/2),20) + diag((1:n − 20).^(1/2),−20); x, b 同例 1(下同); cond(A) = 1.0023 × 10^6，用 1-范数预条件子 2 步达精确值、Gauss-Seidel 预条件子 6 步相对误差小于 10^{-16}，2-范数、∞-范数、Jacobi 预条件子 17 步相对误差小于 10^{-16}，其次是不完全 Cholesky 预条件子，不用预条件子收敛非常慢，如图 2。

实例 3. 当 n = 33; A = log([5:n + 4]')*([1:n].^2); cond(A) = 1.2857 × 10^{20}，用 1-范数、∞-范数预条件子迭代 1 步达精确值，2-范数预条件子迭代 1 步相对误差小于 10^{-16}，无法进行 Cholesky 分解，如图 3。

实例 4. 当 n = 3000; A = diag(1./(3:n + 2)) + diag(1./(n − 1:1),1) + diag(1./(n − 1:1),−1); cond(A) = 1.0007 × 10^3，全部预条件子表现都优越，且用 1-范数、2-范数、∞-范数预条件子表现最好，1 步达到精确值，如图 4。

实例 5. 当 n = 2000; A = magic(n); A 非正定, cond(A) = 8.5285 × 10^{20}，其它方法无法进行，用 1-范数 1 步迭代达精确值，不用预条件子迭代 10 步相对误差小于 10^{-15}，如图 5。

实例 6. 当 n = 1000; A = diag((1:n)) + diag(sin(1:n − 3),3) + diag(sin(1:n − 3), −3) + diag(sin(1:n − 20),20) + diag(sin(1:n − 20),−20) + diag(cos(1:n − 100), 100) + diag(cos(1:n − 100),−100); cond(A) = 1.3568 × 10^3，Gauss-Seidel 预条件子表现优越，迭代 16 步误差即小于 10^{-18}，其次是不完全 2-范数、∞-范数、Jacobi、1-范数、Cholesky 预条件子，不用预条件子收敛非常慢，如图 6。

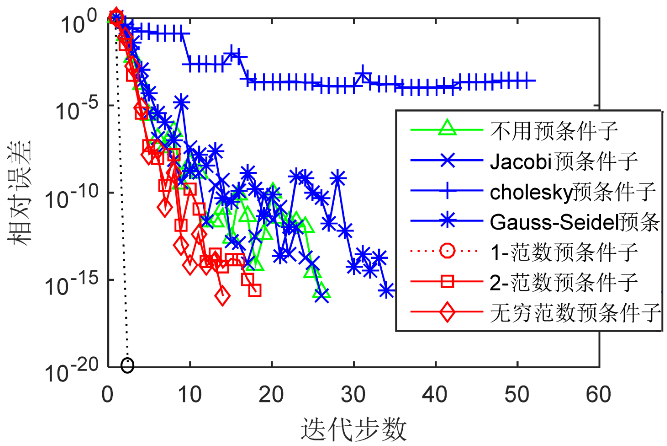

Figure 1. Figure of instance 1
图 1. 实例 1 图

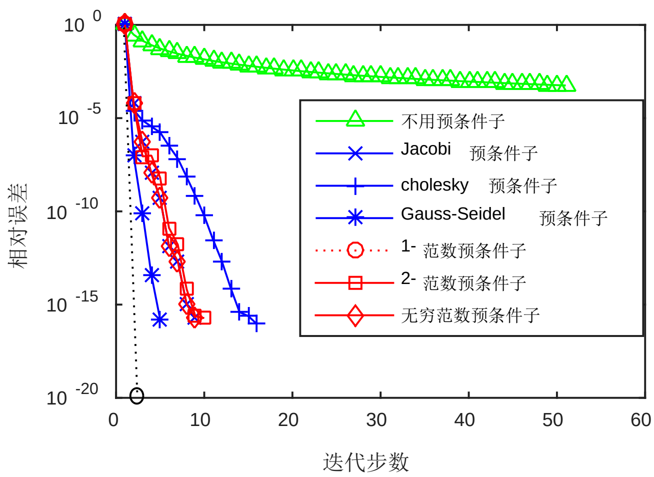

Figure 2. Figure of instance 2
图 2. 实例 2 图

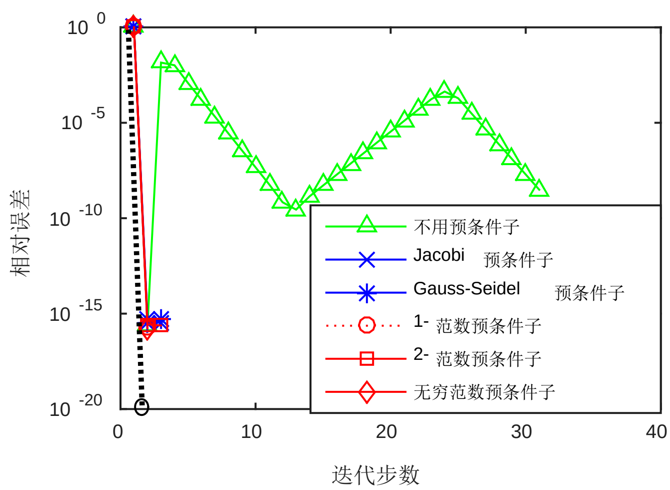

Figure 3. Figure of instance 3
图 3. 实例 3 图

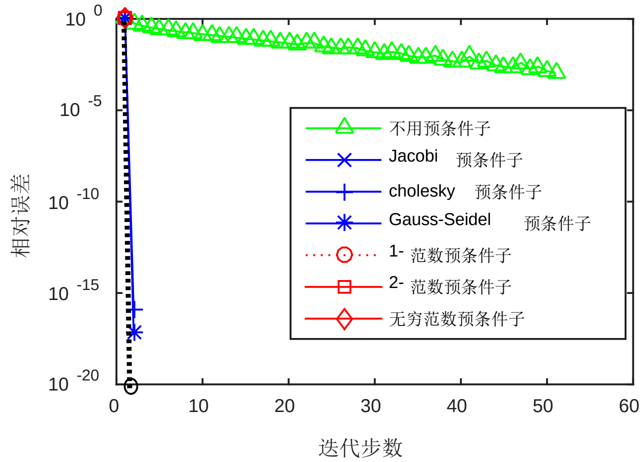

Figure 4. Figure of instance 4
图 4. 实例 4 图

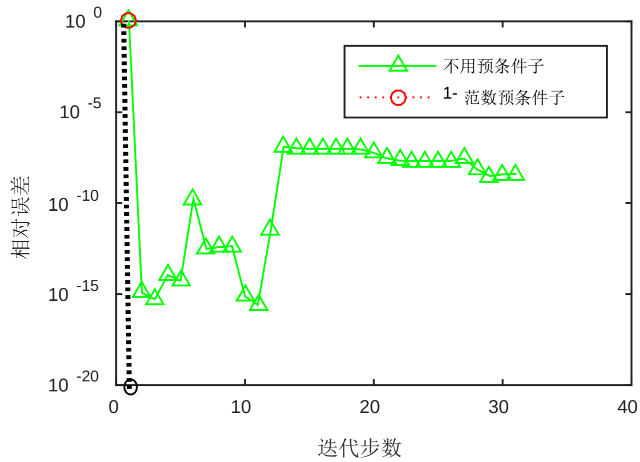

Figure 5. Figure of instance 5
图 5. 实例 5 图

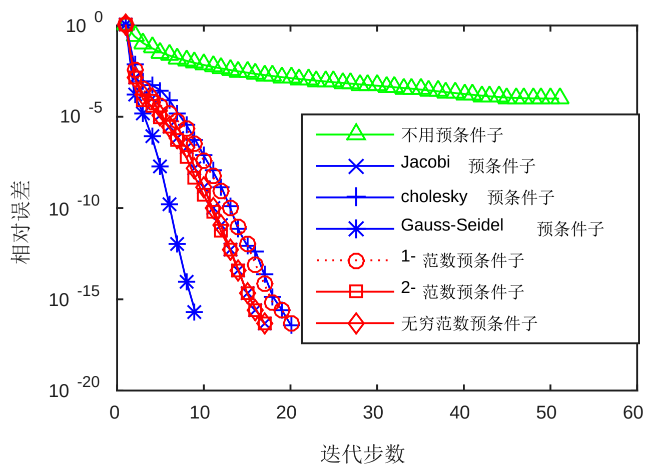

Figure 6. Figure of instance 6
图 6. 实例 6 图

7. 算法复杂度

共轭梯度法的算法复杂度取决于矩阵与向量乘积的次数，如恰当设计程序使每步迭代只有一次矩阵与向量的乘法，在矩阵稀疏时，时间复杂度为$O\left(n^{\frac{4}{3}}\right)$。而使用预条件子的共轭梯度法大大降低了复杂度，节省了计算的时间成本，提高了运算速度和精度。其中用列范数预条件子方法迭代次数远远小于矩阵阶数 n。考虑到两矩阵乘积、矩阵与向量乘积的复杂度，对稀疏矩阵而言，列 Euclid 范数、Gauss-Seidel 范数、不完全 Cholesky 分解预条件子的复杂度都是$O\left(n^2\right)$，而列最大模、绝对和预条件子复杂度为$O\left(n^{\frac{4}{3}}\right)$。

参考文献 (References)

[1] Sauer Timothy. 数值分析[M]. 裴玉茹, 马赓宇, 译. 北京: 机械工业出版社, 2015.

[2] 林成森. 数值计算方法[M]. 第 2 版. 北京: 科学出版社, 2005.

[3] Meijerink, J.V. and Van der Vorst, H.A. (1977) An Iterative Solution Method for Linear Systems of Which the Coefficient Matrix Is a Symmetric M-Matrix. *Mathematics of Computation*, **31**, 148-162.

[4] Horn R.A. and Johnson, C.R. (2015) Matrix Analysis. 2nd Edition, Cambridge University Press, London.