# Homework 6 Foundations of Computational Math 1 Fall 2011

## Problem 6.1

Suppose you are attempting to solve $Ax = b$ using a linear stationary iterative method defined by

$$x_k = Gx_{k-1} + f$$

that is consistent with $Ax = b$.

Suppose the eigenvalues of $G$ are real and such that $|\lambda_1| > 1$ and $|\lambda_i| < 1$ for $2 \leq i \leq n$. Also suppose that $G$ has $n$ linearly independent eigenvectors, $z_i$, $1 \leq i \leq n$.

**6.1.a**. Show that there exists an initial condition $x_0$ such that $x_k$ converges to $x = A^{-1}b$.

**6.1.b**. Does your answer give a characterization of selecting $x_0$ that could be used in practice to create an algorithm that would ensure convergence?

**Solution:**

Let $e^{(k)} = x - x_k$ be the error on step $k$. We know that

$$e^{(0)} = \alpha_1 z_1 + \sum_{i=2}^{n} \alpha_i z_i, \quad \text{uniquely}$$

$$e^{(k)} = \lambda_1^k \alpha_1 z_1 + \sum_{i=2}^{n} \lambda_i^k \alpha_i z_i$$

So if $x_0$ is such that $\alpha_1 = 0$ then

$$e^{(0)} = \sum_{i=2}^{n} \alpha_i z_i, \quad \text{uniquely}$$

$$e^{(k)} = \sum_{i=2}^{n} \lambda_i^k \alpha_i z_i$$

Since $|\lambda_i| < 1$ for $2 \leq i \leq n$ it follows that

$$\lim_{k \to \infty} e^{(k)} = 0$$

This is not a practical characterization for two main reasons. The first is obvious: $e^{(0)}$ and $z_1$ are not known therefore we cannot be sure $\alpha_1 = 0$. However, even if $x_0$ was such that $\alpha_1 = 0$ on each iteration there would be roundoff error that would most likely have a piece in the direction of $z_1$. This would then give an iteration $x_j$ such that its $\alpha_1 \neq 0$ and divergence in practice would follow.

## Problem 6.2

Suppose you are attempting to solve $Ax = b$ using a linear stationary iterative method defined by

$$x_k = M^{-1}Nx_{k-1} + M^{-1}b$$

where $A = M - N$. Suppose further that $M = D + F$ where $D = diag(\alpha_{11}, \dots, \alpha_{nn})$ and $F$ is made up of any subset of the off-diagonal elements of $A$. The matrix $N$ is therefore the remaining off-diagonal elements of $A$ after removing those in $F$.

Show that if $A$ is a strictly diagonally dominant $M$-matrix then the iteration is convergent to $x = A^{-1}b$.

**Solutions:**

Given the method of choosing $F$ it is clear that any sum (row or column) of the magnitudes of off-diagonal elements in $M$ must be smaller than the corresponding sum of the magnitudes of off-diagonal elements in $A$ (since we are simply removing terms). So if $A$ is strictly diagonally dominant then so is $M$. Since $A$ is an $M$-matrix we know that $\alpha_{ij} \leq 0$. It follows that $N \geq 0$ and for the off-diagonal elements of $M$, either $\mu_{ij} = \alpha_{ij} \leq 0$ or $\mu_{ij} = 0$. Since $M$ has nonpositive off-diagonal elements and is strictly diagonally dominant, $M$ is an $M$-matrix. Therefore $A = M - N$ is a regular splitting and the iteration is convergent.

## Problem 6.3

**6.3.a**. Textbook page 241, Problem 2

**6.3.b**. Textbook page 241, Problem 4

**6.3.c**. Textbook page 241, Problem 5

Material in textbook Sections 1.7 and 5.1 is useful for these problems.

**Solution:**

**Textbook page 241, Problem 2:** Since $A$ is symmetric all of its eigenvalues are real by Theorem 5.1 page 184 and the row and column disks are identical.

We have

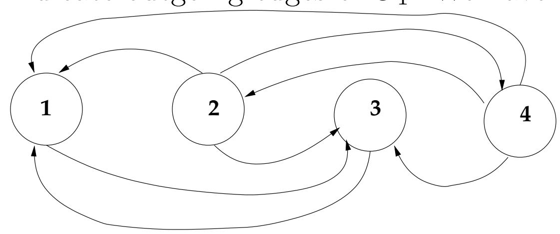
$$\begin{aligned}
\mathcal{R}_1 = \mathcal{C}_1 &= \{z \in \mathbb{C} \;\;:\;\; |z - 1| \leq 3\} \\
\mathcal{R}_2 = \mathcal{C}_2 &= \{z \in \mathbb{C} \;\;:\;\; |z - 7| \leq 2\} \\
\mathcal{R}_3 = \mathcal{C}_3 &= \{z \in \mathbb{C} \;\;:\;\; |z - 5| \leq 1\}
\end{aligned}$$

Since we know the eigenvalues are real we can look at the intersection of the union of the disks and the real axis. This gives the intervals

$$[-2, 4], \quad [4, 6], \quad [5, 9]$$

which in turn imply that the eigenvalues of $A$ are in the interval $[-2, 9]$.

**Textbook page 241, Problem 4:** First we create the graph, $G_1$, that corresponds to the matrix $A_1$. Row elements indicate outgoing edges of $G_1$. We have




We see that the set of nodes {1, 3} and {2, 4} are completely connected and there are edges from {2, 4} into {1, 3} but not the other way. Therefore we have a permutation

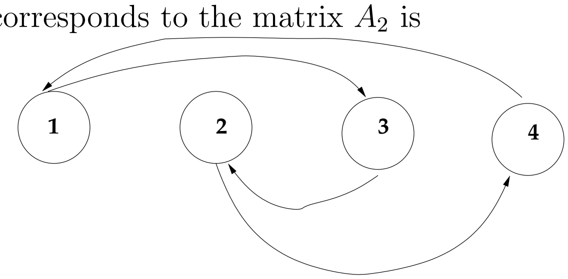
$$P = \begin{pmatrix} e_2 & e_4 & e_1 & e_3 \end{pmatrix}$$

$$\tilde{A}_1 = P^T A_1 P = \left(\begin{array}{cc|cc} 3 & 1 & 2 & -2 \\ -1 & 4 & 1 & 1 \\ \hline 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & -2 \end{array}\right) = \begin{pmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ \tilde{A}_{21} & \tilde{A}_{22} \end{pmatrix}$$

The eigenvalues of $\tilde{A}$ is the union of the eigenvalues of $\tilde{A}_{11}$ and $\tilde{A}_{22}$. These are easily determined by finding the roots of the quadratic equations

$$\det(\tilde{A}_{11} - \lambda I) = (3 - \lambda)(4 - \lambda) + 1$$
$$\det(\tilde{A}_{22} - \lambda I) = (1 - \lambda)(-2 - \lambda) - 1$$

The graph, $G_2$, that corresponds to the matrix $A_2$ is

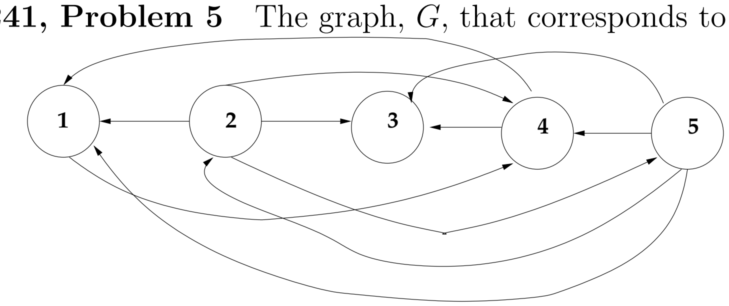



Clearly, there is a cycle connecting all nodes

$$1 \to 3 \to 2 \to 4 \to 1$$

so there is a directed path between any two notes and therefore the matrix is irreducible.

**Textbook page 241, Problem 5** The graph, $G$, that corresponds to the matrix $A$ is



So we can deduce that the matrix is reducible and permute it into block upper triangular form:

$$P = \begin{pmatrix} e_2 & e_5 & e_1 & e_3 & e_4 \end{pmatrix}$$

$$\tilde{A} = P^T A P = \left(\begin{array}{cc|ccc} 2 & 1 & 2 & 4 & -3 \\ 0.5 & 4 & 2 & -1 & 3 \\ \hline 0 & 0 & -4 & 0 & 0.5 \\ 0 & 0 & 0.5 & -1 & 0 \\ 0 & 0 & 0.5 & 0.2 & 3 \end{array}\right) = \begin{pmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{pmatrix}$$

The eigenvalues of $\tilde{A}$ is the union of the eigenvalues of $\tilde{A}_{11}$ and $\tilde{A}_{22}$.

Now consider the eigenvalues of $\tilde{A}_{11}$. The matrix is $2 \times 2$ so we could simply find the roots of

$$\det(\tilde{A}_{11} - \lambda I) = \lambda^2 - 6\lambda + 7.5$$

$$\lambda_{\pm} = 3 \pm \frac{\sqrt{6}}{2}$$

which are both real as desired. It is possible however to deduce this from Gershgorin's theorems. We have

$$\begin{aligned} \mathcal{R}_1 &= \{z \in \mathbb{C} \ : \ |z-2| \leq 1\} \\ \mathcal{R}_2 &= \{z \in \mathbb{C} \ : \ |z-4| \leq 0.5\} \\ \mathcal{C}_1 &= \{z \in \mathbb{C} \ : \ |z-2| \leq 0.5\} \\ \mathcal{C}_2 &= \{z \in \mathbb{C} \ : \ |z-4| \leq 1\} \end{aligned}$$

$$\begin{aligned} \mathcal{S}_R &= \mathcal{R}_1 \cup \mathcal{R}_2 \\ \mathcal{S}_C &= \mathcal{C}_1 \cup \mathcal{C}_2 \\ \mathcal{S}_R \cap \mathcal{S}_C &= \{\mathcal{C}_1 \cup \mathcal{R}_2\} \end{aligned}$$

$\mathcal{S}_R \cap \mathcal{S}_C$ comprises two disjoint disks therefore Gershgorin says that each must have one of the eigenvalues. Since two complex eigenvalues would be in a single disk we must have two distinct and therefore real eigenvalues.

Now consider the eigenvalues of $\tilde{A}_{22}$. We have

$$\begin{aligned}
\mathcal{R}_1 &= \{z \in \mathbb{C} \ : \ |z-(-4)| \leq 0.5\} \\
\mathcal{R}_2 &= \{z \in \mathbb{C} \ : \ |z-(-1)| \leq 0.5\} \\
\mathcal{R}_3 &= \{z \in \mathbb{C} \ : \ |z-3| \leq 0.7\} \\
\mathcal{C}_1 &= \{z \in \mathbb{C} \ : \ |z-(-4)| \leq 1\} \\
\mathcal{C}_2 &= \{z \in \mathbb{C} \ : \ |z-(-1)| \leq 0.2\} \\
\mathcal{C}_3 &= \{z \in \mathbb{C} \ : \ |z-3| \leq 0.5\}
\end{aligned}$$

$$\begin{aligned}
\mathcal{S}_R &= \mathcal{R}_1 \cup \mathcal{R}_2 \cup \mathcal{R}_3 \\
\mathcal{S}_C &= \mathcal{C}_1 \cup \mathcal{C}_2 \cup \mathcal{C}_3 \\
\mathcal{S}_R \cap \mathcal{S}_C &= \{\mathcal{R}_1 \cup \mathcal{C}_2 \cup \mathcal{C}_3\}
\end{aligned}$$

$\mathcal{S}_R \cap \mathcal{S}_C$ comprises three disjoint disks therefore Gershgorin says that each must have one of the eigenvalues and we must have 3 distinct real eigenvalues.

# Problem 6.4

Consider the $n \times n$ matrix

$$T_\alpha = \begin{pmatrix}
\alpha & -1 & 0 & \dots & \dots & \dots & 0 \\
-1 & \alpha & -1 & 0 & \dots & \dots & 0 \\
0 & -1 & \alpha & -1 & 0 & \dots & 0 \\
 & \ddots & \ddots & \ddots & \ddots & \ddots & \\
0 & \dots & 0 & -1 & \alpha & -1 & 0 \\
0 & \dots & \dots & 0 & -1 & \alpha & -1 \\
0 & \dots & \dots & \dots & 0 & -1 & \alpha
\end{pmatrix}$$

(6.4.a) Show that the eigenvalues of $T_\alpha$ are

$$\lambda_j = \alpha - 2\cos j\theta, \quad \theta = \frac{\pi}{n+1}$$

with an associated eigenvector

$$q_j = \big(\sin(j\theta), \ \sin(2j\theta), \dots, \sin(nj\theta)\big)^T$$

(6.4.b) For what values of $\alpha$ is $T_\alpha$ positive definite?

(6.4.c) Show that for $\alpha = 2$ the matrix $T_\alpha$ is an M-matrix.

(6.4.d) What is the rate of convergence for Jacobi's method if $\alpha = 2$?

(6.4.e) What is the rate of convergence for Gauss-Seidel if $\alpha = 2$?

**Solution:**

To show that $T_\alpha q_j = q_j \lambda_j$ we must show that

$$e_1^T T_\alpha q_j = \alpha \sin \theta_j - \sin 2\theta_j = e_1^T q_j \lambda_j = (\alpha - 2\cos\theta_j)\sin\theta_j \tag{1}$$

$$e_n^T T_\alpha q_j = -\sin(n-1)\theta_j + \alpha \sin n\theta_j = e_n^T q_j \lambda_j = (\alpha - 2\cos\theta_j)\sin n\theta_j \tag{2}$$

$$e_i^T T_\alpha q_j = \alpha \sin i\theta_j - \sin(i-1)\theta_j - \sin(i+1)\theta_j = e_i^T q_j \lambda_j = (\alpha - 2\cos\theta_j)\sin i\theta_j \tag{3}$$

where $\theta_j = j\theta$.

These follow from the standard trigonometric identity:

$$2\sin A \cos B = \sin(A+B) + \sin(A-B) \tag{4}$$

For (1) we have

$$\begin{aligned}(\alpha - 2\cos\theta_j)\sin\theta_j &= \alpha \sin\theta_j - 2\cos\theta_j \sin\theta_j \\ &= \alpha\sin\theta_j - \sin 2\theta_j - \sin 0 = e_1^T T_\alpha q_j\end{aligned}$$

by (4).

For (2) we have

$$\begin{aligned}(\alpha - 2\cos\theta_j)\sin n\theta_j &= \alpha \sin n\theta_j - 2\cos\theta_j \sin n\theta_j \\ &= \alpha \sin n\theta_j - \sin(n+1)\theta_j - \sin(n-1)\theta_j \\ &= \alpha\sin n\theta_j - \sin(n-1)\theta_j = e_n^T T_\alpha q_j\end{aligned}$$

by (4) and $\theta_j = (j\pi)/(n+1) \rightarrow \sin(n+1)\theta_j = \sin j\pi = 0$.

For (3) we have

$$\begin{aligned}(\alpha - 2\cos\theta_j)\sin i\theta_j &= \alpha \sin i\theta_j - 2\cos\theta_j \sin i\theta_j \\ &\alpha \sin i\theta_j - \sin(i+1)\theta_j - \sin(i-1)\theta_j \\ &= e_i^T T_\alpha q_j\end{aligned}$$

by (4).

We have $\lambda_j = \alpha - 2\cos\theta_j$ and $|2\cos\theta_j| < 2$ for $1 \le j \le n$. Also $\cos\theta_1 = -\cos\theta_n > 0$.

In order for $A$ to be a symmetric positive definite matrix it must have positive diagonal elements, so we assume $\alpha > 0$. Since $\cos\theta_1 > 0$ and closest to 1, the minimum eigenvalue is

$$\lambda_1 = \alpha - 2\cos\theta_1$$

It follows that $\alpha \ge 2$ implies that $A$ is symmetric positive definite for any $n$. For a specific value of $n$ we need $\alpha - 2\cos\theta_1 > 0$.

The iteration matrix for Jacobi's method is

$$G_J = \alpha^{-1} T_0$$

Since $\cos\theta_1 = -\cos\theta_n > 0$ it follows that

$$\rho(G_J) = \frac{2\cos\theta_1}{\alpha}$$
$$\log\rho = \log\frac{2\cos\theta_1}{\alpha} = \log\frac{2}{\alpha} + \log\cos\theta_1$$

So if $\alpha \geq 2$ we have $\log\rho = -d < 0$ and $d$ is the asymptotic convergence rate. Since $\log\cos\theta_1 < 0$ for all $n$, if $\alpha < 2$ then $n$ must be small enough to have

$$\log\frac{2}{\alpha} > -\log\cos\theta_1$$

in order for $\rho(G_J) < 1$ which is essentially the same as making $A$ positive definite.

So if $\alpha = 2$ then

$$\rho(G_J) = \cos\theta_1$$
$$-\log\rho = -\log\cos\theta_1 = d$$

Since the matrix is tridiagonal

$$\rho(G_{gs}) = \rho^2(G_J)$$

# Problem 6.5

The Gershgorin theorems in Section 5.1 of the text and the idea of irreducibility in Section 5.1 are often valuable in analyzing the convergence of iterative methods. Familiarize yourself with both in order to answer this question.

**6.5.a**. Use the appropriate Gershgorin-related facts to show that if $A$ is symmetric and strictly diagonally dominant then Jacobi converges for all $x_0$.

**6.5.b**. Use the appropriate Gershgorin-related facts to show that Jacobi converges for all $x_0$ for the matrix

$$A = \begin{pmatrix} 2 & -1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 & -1 & 2 \end{pmatrix}$$

**Solutions:**

The Jacobi iteration matrix is $G_J = D^{-1}(L+U)$ where the $i$-th row of $G_J$ has elements

$$\gamma_{ii} = 0 \quad \text{and} \gamma_{ij} = \frac{\alpha_{ij}}{\alpha_{ii}}$$

So the row circles $\mathcal{R}_i$ are centered at the origin and have radii

$$\sum_{j=1, j\neq i}^{n} |\gamma_{ij}| = \sum_{j=1, j\neq i}^{n} \frac{|\alpha_{ij}|}{|\alpha_{ii}|} = \frac{1}{|\alpha_{ii}|}\left(\sum_{j=1, j\neq i}^{n} |\alpha_{ij}|\right) < 1$$

Since the matrix is symmetric the row circles and the column circles are identical. Therefore by the First Gershgorin theorem,

$$\sigma(G_J) \subseteq \mathcal{S}_{\mathcal{R}} = \mathcal{S}_{\mathcal{C}} \subset \mathcal{U}$$

where $\mathcal{U} \subset \mathbb{R}^n$ is the unit sphere. Due to the strict inequality of the radii all eigenvalues have magnitude strictly less than 1 and therefore the method is convergent.

For the second part, we note that

$$A = \begin{pmatrix} 2 & -1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 & -1 & 2 \end{pmatrix}$$

is not strictly diagonally dominant but it is symmetric and therefore we have the row and column circles identical as above. The graph of $A$ is strongly connected there is a directed cycle containing all nodes in the graph and therefore $A$ is irreducible. This and the weak diagonal dominance of rows 2 through $n-1$ and the strict dominance of rows 1 and 2 implies that $A$ is irreducibly diagonally dominant.

Therefore, the Jacobi iteration matrix is also irreducible. So the row (column) circles $\mathcal{R}_i$ are centered at the origin and have radii 1 for $2 \leq i \leq n-1$ and radii $1/2$ for $\mathcal{R}_1$ and $\mathcal{R}_2$. We have that all the eigenvalues are contained in or on the unit sphere by the First Gershgorin theorem, i.e., $|\lambda_i| \leq 1$, since

$$\sigma(G_J) \subseteq \mathcal{S}_{\mathcal{R}} = \mathcal{S}_{\mathcal{C}} \subseteq \mathcal{U}$$

By the Third Gershgorin theorem an eigenvalue of $G_J$ cannot be on the boundary of $\mathcal{S}_{\mathcal{R}}$ unless it is on the boundary of all $\mathcal{R}_i$. Since there are two circles, $\mathcal{R}_1$ and $\mathcal{R}_2$, that do not intersect with the unit sphere $\mathcal{U}$ no eigenvalue can have magnitude 1 therefore, $\rho(G_J) < 1$ and Jacobi is convergent.

## Problem 6.6

Consider the block tridiagonal matrix associated with an $n \times n$ grid discretization of the partial differential $u_{\xi,\xi} + u_{\eta,\eta} = g$ on a two-dimensional domain.

The matrix is $n^2 \times n^2$ with $n \times n$ blocks $T_i \in \mathbb{R}^{n\times n}$ $1 \le i \le n$ $E_i = -I_n \in \mathbb{R}^{n\times n}$ $1 \le i \le n$ with block tridiagonal structure given by

$$A = \begin{pmatrix} T_1 & E_1 & 0 & \cdots & \cdots & \cdots & 0 \\ E_2 & T_2 & E_2 & 0 & & & \vdots \\ 0 & E_3 & T_3 & E_3 & 0 & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & E_{n-1} & T_{n-2} & E_{n-2} & 0 \\ 0 & & \cdots & 0 & E_{n-1} & T_{n-1} & E_{n-1} \\ 0 & & & \cdots & 0 & E_n & T_n \end{pmatrix}$$

where $T_i$ are tridiagonal and $E_i$ are diagonal and dimensions $n \times n$

$$T_i = \begin{pmatrix} 4 & -1 & 0 & 0 & 0 & \dots & 0 \\ -1 & 4 & -1 & 0 & 0 & \dots & 0 \\ 0 & -1 & 4 & -1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & & \vdots \\ 0 & \dots & 0 & -1 & 4 & -1 & 0 \\ 0 & \dots & 0 & 0 & -1 & 4 & -1 \\ 0 & \dots & 0 & 0 & 0 & -1 & 4 \end{pmatrix}$$

$$E_i = -I_n = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & -1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & -1 & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & & \vdots \\ 0 & \dots & 0 & 0 & -1 & 0 & 0 \\ 0 & \dots & 0 & 0 & 0 & -1 & 0 \\ 0 & \dots & 0 & 0 & 0 & 0 & -1 \end{pmatrix}$$

**6.6.a**. Determine the computational complexity of one step of Jacobi for a linear system involving the matrix.

**6.6.b**. Determine the computational complexity of one step of Gauss-Seidel for a linear system involving the matrix.

**Solution:** The matrix is $n^2 \times n^2$ with $n \times n$ blocks $T_i \in \mathbb{R}^{n\times n}$ $1 \le i \le n$ $E_i = -I_n \in \mathbb{R}^{n\times n}$

$1 \le i \le n$ with block tridiagonal structure given by

$$A = \begin{pmatrix} T_1 & E_1 & 0 & \cdots & \cdots & \cdots & 0 \\ E_2 & T_2 & E_2 & 0 & & & \vdots \\ 0 & E_3 & T_3 & E_3 & 0 & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & E_{n-1} & T_{n-2} & E_{n-2} & 0 \\ 0 & & \cdots & 0 & E_{n-1} & T_{n-1} & E_{n-1} \\ 0 & & & \cdots & 0 & E_n & T_n \end{pmatrix}$$

where $T_i$ are tridiagonal and $E_i$ are diagonal and dimensions $n \times n$

$$T_i = \begin{pmatrix} 4 & -1 & 0 & 0 & 0 & \dots & 0 \\ -1 & 4 & -1 & 0 & 0 & \dots & 0 \\ 0 & -1 & 4 & -1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & & \vdots \\ 0 & \dots & 0 & -1 & 4 & -1 & 0 \\ 0 & \dots & 0 & 0 & -1 & 4 & -1 \\ 0 & \dots & 0 & 0 & 0 & -1 & 4 \end{pmatrix}$$

$$E_i = -I_n = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & -1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & -1 & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & & \vdots \\ 0 & \dots & 0 & 0 & -1 & 0 & 0 \\ 0 & \dots & 0 & 0 & 0 & -1 & 0 \\ 0 & \dots & 0 & 0 & 0 & 0 & -1 \end{pmatrix}$$

Given that for this PDE, all $T_i$ are the same and all $E_i$ are the same we have

$$A = \begin{pmatrix} T & -I & 0 & \cdots & \cdots & \cdots & 0 \\ -I & T & -I & 0 & & & \vdots \\ 0 & -I & T & -I & 0 & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -I & T & -I & 0 \\ 0 & & \cdots & 0 & -I & T & -I \\ 0 & & & \cdots & 0 & -I & T \end{pmatrix}$$

For a Jacobi step we assume that we use the form:

$$\begin{aligned} y_k &= (L+U)x_k \\ z_k &= y_k + b \\ x_k &= D^{-1}z_k \end{aligned}$$

$(L+U)$ for this $A$ is trivial. A typical row has 4 nonzeros with the values 1. No multiplications are needed and 3 additions are required to get an element of $y_k$ from such a typical row. There are other rows that have less than 4 nonzeros but there are only $O(n)$ of these and therefore they do not affect the dominant term of the complexity. The complexity is therefore

$$\begin{aligned} y_k &= (L+U)x_k \rightarrow 3n^2 + O(n) \text{ additions} \\ z_k &= y_k + b \rightarrow n^2 \text{ additions} \\ x_k &= D^{-1} z_k \rightarrow n^2 \text{ divisions} \\ \text{Total: } & 5n^2 + O(n) \text{ operations} \end{aligned}$$

If the equation is such that the $-1$'s change to some other values due to discretizing a different PDE then $4n^2 + O(n)$ multiplications must be added to this count.

For a Gauss-Seidel step we assume that we use the form:

$$\begin{aligned} y_k &= U x_k \\ z_k &= y_k + b \\ x_k &= (D-L)^{-1} z_k \end{aligned}$$

Of course the last step is the solution of a banded triangular linear system. The inverse is not computed explicitly.

A typical row of $U$ has 2 nonzeros with the values 1. No multiplications are needed and 1 addition is required to get an element of $y_k$ from such a typical row. There are other rows that have less than 2 nonzeros but there are only $O(n)$ of these and therefore they do not affect the dominant term of the complexity. The solution of $(D-L)x_k = z_k$ has a matrix with a typical row of a diagonal element with the value 4 and two off-diagonal elements with values $-1$. There are other rows that have less than 2 off-diagonal elements with value $-1$ but there are only $O(n)$ of these and therefore they do not affect the dominant term of the complexity. So using a row-oriented solver, the typical row subtracts two elements of $x_k$ from the right-hand side and then divides by 4 for a total of 3 operations. The complexity is therefore

$$\begin{aligned} y_k &= U x_k \rightarrow n^2 + O(n) \text{ additions} \\ z_k &= y_k + b \rightarrow n^2 \text{ additions} \\ x_k &= (D-L)^{-1} z_k \rightarrow 3n^2 + O(n) \text{ operations} \\ \text{Total: } & 5n^2 + O(n) \text{ operations} \end{aligned}$$

If the equation is such that the $-1$'s change to some other values due to discretizing a different PDE then $4n^2 + O(n)$ multiplications must be added to this count.

# Problem 6.7

## 6.7.a

Consider the two matrices:

$$A_1 = \begin{pmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{pmatrix} \quad \textit{and} \quad A_2 = \begin{pmatrix} 1 & -\frac{1}{12} \\ -\frac{3}{4} & 1 \end{pmatrix}$$

Suppose you solve systems of linear equations involving $A_1$ and $A_2$ using Jacobi's method. For which matrix would you expect faster convergence?

**Solution:**

If $A = D - L - U$ then the iteration matrix for Jacobi is $M = D^{-1}(L+U)$. We have $D_1 = D_2 = I$ and

$$\begin{aligned} M_1 &= \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix} \\ M_2 &= \begin{pmatrix} 0 & \frac{1}{12} \\ \frac{3}{4} & 0 \end{pmatrix} \end{aligned}$$

The spectral radii are easily deduced from the determinants of $\lambda I - M_1$ and $\lambda I - M_2$. These are $\rho_1 = 1/2$ and $\rho_1 = 1/4$. So faster convergence is expected for $A_2$ even though it is less diagonally dominant than $A_1$.

## 6.7.b

Consider the matrix

$$A = \begin{pmatrix} 4 & 0 & 0 & -1 \\ -1 & 4 & -1 & 0 \\ 0 & -1 & 4 & 0 \\ -1 & 0 & 0 & 4 \end{pmatrix}$$

(i) Will Jacobi's method converge when solving $Ax = b$?

(ii) Will Gauss-Seidel converge when solving $Ax = b$?

**Solution:** Applying Gershgorin's first theorem to $G_J$ results in $\rho(G_J) \leq 1/2$. Therefore, Jacobi's method is convergent.

For Gauss-Seidel we have

$$A = \begin{pmatrix} 4 & 0 & 0 & -1 \\ -1 & 4 & -1 & 0 \\ 0 & -1 & 4 & 0 \\ -1 & 0 & 0 & 4 \end{pmatrix}$$

$$G_{gs} = \begin{pmatrix} 4 & 0 & 0 & 0 \\ -1 & 4 & 0 & 0 \\ 0 & -1 & 4 & 0 \\ -1 & 0 & 0 & 4 \end{pmatrix}^{-1} \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 1/4 \\ 0 & 0 & 1/4 & 1/16 \\ 0 & 0 & 1/16 & 1/64 \\ 0 & 0 & 0 & 1/16 \end{pmatrix}$$

From solving two triangular systems we see that $G_{gs}$ is upper triangular so $\rho(G_{gs}) = 1/16$. Therefore, Gauss-Seidel converges.