Set 7: Iterative Methods for Solving Equations: Part 1

Kyle A. Gallivan

Department of Mathematics

Florida State University

Foundations of Computational Math 1

Fall 2012

Overview

The second half of the course deals with the application of the ideas of convergent iteration and optimization of a scalar cost function to solve:

1. linear sytems of equations
2. nonlinear equations
3. systems of nonlinear equations
4. unconstrained optimization problems

Overview

Two parts:

1. Iterations that are contraction mappings for solving:
   - linear sytems of equations
   - nonlinear equations
   - systems of nonlinear equations
2. Iterations that minimize scalar cost functions for solving:
   - linear sytems of equations
   - unconstrained optimization problems

Overview

- Iteration $x_k = F(x_0, \ldots, x_{k-1})$ in general
- particular form of F depends on problem and other constraints such as efficiency
- $x_0,\ x_1,\ x_2, \ldots$ must converge to a solution of the problem
- We will consider:
  1. construction of the iteration
  2. necessary and sufficient conditions for convergence to the desired solution
  3. efficiency of the iteration

Iterative Methods for Linear Systems

Additonal References and Source Material:

- Iterative Methods for Sparse Linear Systems, Yousef Saad, SIAM Press, Second Edition.
- Matrix Iterative Analysis, Richard Varga, Prentice Hall.
- Applied Iterative Methods, L. A. Hageman and D. M. Young, Academic Press.
- Iterative Solution Methods, O. Axelsson, Cambridge University Press.
- Analysis of Numerical Methods, E. Isaacson and H. Keller, Wiley.

Outline for Iterative Methods for Linear Systems

- Motivation
- Linear Stationary Methods
  - Examples
  - Convergence Analysis
  - Convergence Behavior
- Implementation Issues
- Projection and Optimization Approaches : Conjugate Gradient

Motivation

Iterative methods produce a series of approximations to the solution of $Ax = b$, i.e.,

$$x_0, x_1, x_2, \ldots x_k, \ldots$$

such that $x_k \to x = A^{-1}b$

- A is an $n \times n$ and n very large.
- A is a sparse matrix and the fill-in in the factorization is unacceptably large.
- A good guess at x is avaiable and we wish to improve it.
- High accuracy is not required and we want to save computations.
- We want to improve the accuracy of a direct method that was degraded due to saving computations.

Motivation

- The matrix A is not available
  - finite element discretization of a continuous domain
  - analysis of a discrete network of devices – circuit simulation
  - action of $Av \rightarrow z$ is the sum of actions of elements or devices on pieces of v
- Assume that there is structure that makes computation of $Av \rightarrow z$ $O(n)$ or $O(n \log n)$ typically, e.g., sparse or Toeplitz.
- If A is available then it is stored in an efficient data structure.

Linear Stationary Methods

Definition 7.1. Let $G \in \mathbb{R}^{n \times n}$ $\quad f \in \mathbb{R}^n$ and $x_0 \in \mathbb{R}^n$ be given. G and f define a linear stationary method and the initial condition specifies a particular sequence by

$$\text{initial condition given:} \quad x_0 \in \mathbb{R}^n$$

$$\text{iteration:} \quad x_{i+1} = Gx_i + f$$

This is a form that is convenient for analysis. There are two others of interest for computational reasons and for designing iterations.

Linear Stationary Methods

- There are many other approaches that are more sophisticated, e.g., projection and optimization-based.
- Linear Stationary Methods and their acceleration can be used in concert with other methods as preconditioners.
- Linear Stationary Methods and their acceleration can be used effectively also when applied to problems with appropriate structure.
- The derivations and analyses of Linear Stationary Methods as contraction mappings is crucial to understanding iterations to solve nonlinear systems of equations.

Splitting Form

Definition 7.2. Let $A \in \mathbb{R}^{n \times n}$ and $M \in \mathbb{R}^{n \times n}$ be two nonsingular matrices and let $N \in \mathbb{R}^{n \times n}$, be such that

$$A = M - N$$

Define the iteration:

$$x_{i+1} = M^{-1} N x_i + M^{-1} b$$

Stationary Richardson Family of Methods

Stationary Richardson's Method

$$x_{i+1} = x_i + P^{-1} r_i, \quad r_i = b - Ax_i$$

- P^{-1} is called the preconditioner.
- Later we will introduce an acceleration or relaxation parameter α and the preconditioner will be replaced with αP^{-1}.

Equivalence

Richardson and Linear Stationary

$$\{A, b, G, f\} \to P = A(I - G)^{-1}$$

$$\{A, b, P\} \to G = (I - P^{-1}A), \quad f = P^{-1}b$$

Richardson and Splitting

$$\{A, b, M, N\} \to P = M$$

$$\{A, b, P\} \to P = M, \quad N = P - A$$

Linear Stationary and Splitting

$$\{A, b, M, N\} \to G = (I - M^{-1}A) = M^{-1}N, \quad f = M^{-1}b$$

$$\{A, b, G, f\} \to M = A(I - G)^{-1}, \quad N = M - A$$

Complexity

Splitting method:

1. Solve $Mf = b$ and compute $r_0 = b - Ax_0$
2. Form $v = Nx_k$
3. Solve $Mz_k = v$
4. $x_{k+1} = z_k + f$
5. $r_{k+1} = b - Ax_{k+1}$

Complexity

Richardson's method:

1. Solve $Pz_k = r_k$
2. $x_{k+1} = x_k + z_k$
3. $r_{k+1} = r_k - Az_k$

- Note alternate form of residual computation.
- Recall, solving systems with M and matrix vector products with A are assumed to be of low computational complexity.

Linear Stationary Methods

In order for the iteration to solve a problem such as $Ax = b$, we must satisfy two requirements: consistency and convergence. Respectively, these are

- The solution to the problem of interest must be a fixed point of the iteration:

$$x = Gx + f$$

- There must be some set of initial conditions, $\mathcal{S}$, preferably all of $\mathbb{R}^n$, such that $x_0 \in \mathcal{S}$ implies

$$x_k \to x$$

  i.e., the sequence converges to the desired fixed point.

Linear Stationary Methods

We will discuss:

- creating a linear stationary iteration that has the desired fixed point $x = A^{-1}b$ via splitting and Richardson's methods
- conditions that are required for convergence and those that guarantee convergence
- the rate of convergence of the iteration

Linear Stationary Methods

Definition 7.3. The linear stationary iteration defined by $G \in \mathbb{R}^{n \times n}$, $f \in \mathbb{R}^n$ and $x_0 \in \mathbb{R}^n$ is consistent with the linear system $Ax = b$ where $A \in \mathbb{R}^{n \times n}$ is nonsingular and $b \in \mathbb{R}^n$ if

$$f = (I - G)A^{-1}b$$

This condition implies that $x = Gx + f$, i.e., the solution of the linear system is a fixed point.

Create a Fixed Point via Richardson's Method

$x = A^{-1}b$ must be a fixed point of the iteration.

For Richardson's Methods this is obviously true.

$$x_{i+1} = x_i + P^{-1}r_i, \quad r_i = b - Ax_i$$

$$x = A^{-1}b \to r = b - A(A^{-1}b) = 0$$

$$x_i = x \to x_{i+1} = x$$

Create a Fixed Point via Splitting

A method with a fixed point $x = A^{-1}b$ can be created by **splitting**.

Nonsingular $M \in \mathbb{R}^{n\times n}$ and $N \in \mathbb{R}^{n\times n}$ are such that

$$A = M - N$$

$$Ax = b$$

$$(M - N)x = Mx - Nx = b \to Mx = Nx + b$$

$$x = M^{-1}Nx + M^{-1}b \to x = Gx + f$$

Define the iteration: $\quad x_{i+1} = Gx_i + f$

By construction $x = A^{-1}b$ is the unique fixed point.

Convergence Analysis

Definition 7.4. Let $B \in \mathbb{R}^{n \times n}$. $\lambda \in \mathbb{C}$ and $x \in \mathbb{C}^n$ are an eigenvalue and eigenvector of B if $Bx = \lambda x$. There are at most n distinct eigenvalues and linearly independent eigenvectors.

Definition 7.5. The spectral radius of B, denoted $\rho(B)$, is the maximum magnitude of the eigenvalues of B.

Convergence Analysis

Lemma 7.1. *If $B \in \mathbb{R}^{n \times n}$ with spectral radius $\rho(B)$ then we have the following:*

1. *If $\mathcal{M}$ is the set of all submultiplicative matrix norms then*

$$\rho(B) = \inf_{\|.\| \in \mathcal{M}} \|B\|$$

2. *$\rho(B) \leq \|B\|$ for any submultiplicative matrix norm.*

3. *$\forall \epsilon > 0$ there exists an induced matrix norm dependent on ϵ such that*

$$\|B\|_{B,\epsilon} \leq \rho(B) + \epsilon$$

Convergence Analysis

Definition 7.6. $B \in \mathbb{R}^{n\times n}$ is a convergent matrix if

$$\lim_{k\to\infty} B^k = 0$$

We have a norm-based sufficient condition

Lemma 7.2. $B \in \mathbb{R}^{n\times n}$ *is a convergent matrix if* $\|B\| < 1$ *in some submultiplicative norm.*

Proof.

$$\text{We have } \|B\| = \mu < 1 \text{ and } \|B^k\| \leq \|B\|^k = \mu^k$$

$$\lim_{k\to\infty} \mu^k = 0 \to \lim_{k\to\infty} \|B^k\| = 0$$

$$\therefore \lim_{k\to\infty} B^k = 0$$

$\square$

Fundamental Theorem 1

Theorem 7.3. $B \in \mathbb{R}^{n \times n}$ *is a convergent matrix if and only if* $\rho(B) < 1$.

Proof. $(\rightarrow)$ Let λ, x be any eigenpair of B. If B is convergent then we have

$$\lim_{k\to\infty} B^k = 0 \to \lim_{k\to\infty} B^k x = \lim_{k\to\infty} \lambda^k x = 0 \to \lim_{k\to\infty} \lambda^k = 0$$
$$\therefore |\lambda| < 1 \to \rho(B) < 1$$

$(\leftarrow)$ If $\rho(B) < 1$ then by Lemma 7.1 for some $\epsilon > 0$ and norm $\|B\|_\epsilon < \rho(B) + \epsilon < 1$. Therefore, by Lemma 7.2 B is convergent. $\square$

Fundamental Theorem 2

Theorem 7.4. *If* $B \in \mathbb{R}^{n \times n}$ *is a convergent matrix then*

$$\sum_{k=0}^{\infty} B^k = (I - B)^{-1}$$

Proof. We have $Bx = \lambda x \leftrightarrow (I - B)x = (1 - \lambda)x$. Therefore, since $\rho(B) < 1$ the matrix $I - B$ is nonsingular.

$$S_m = \sum_{k=0}^{m} B^k \rightarrow BS_m = \sum_{k=1}^{m+1} B^k \rightarrow S_m - BS_m = I - B^{m+1}$$

$$\lim_{m\to\infty} S_m - BS_m = \lim_{m\to\infty} (I - B)S_m = I - \lim_{m\to\infty} B^{m+1} = I$$

$$\sum_{k=0}^{\infty} B^k = (I - B)^{-1}$$

□

Convergence Analysis

Theorem 7.5. *Define a linear stationary iterative method by*

$$\textit{initial condition } x_0 \in \mathbb{R}^n$$

$$\textit{iteration } x_{i+1} = Gx_i + f$$

$$G \in \mathbb{R}^{n \times n} \quad f \in \mathbb{R}^n$$

If G *is a convergent matrix then* $\forall x_0, \; x_i \to (I - G)^{-1} f$

Proof.

$$
\begin{aligned}
x_k = Gx_{k-1} + f &= G(Gx_{k-2} + f) + f \\
&= G^2 x_{k-2} + Gf + f \\
&= \cdots = G^k x_0 + (\sum_{i=0}^{k-1} G^i) f
\end{aligned}
$$

$$
\lim_{k\to\infty} x_k = \lim_{k\to\infty} G^k x_0 + (\sum_{i=0}^{k-1} G^i) f = (I - G)^{-1} f
$$

To verify that $(I - G)^{-1} f$ is a fixed point note that $\sum_{i=0}^{\infty} G = (I - G)^{-1} = I + G(I - G)^{-1}$ and

$$
G\big[(I - G)^{-1} f\big] + f = (I + G(I - G)^{-1}) f = (I - G)^{-1} f
$$

$\square$

Convergence Analysis

Corollary 7.6. *Let* $A \in \mathbb{R}^{n\times n}$ *be nonsingular and* $b \in \mathbb{R}^n$*. If* G *is a convergent matrix then* $\forall x_0, \;\; x_i \to x = A^{-1}b$ *where*

$$x_{i+1} = Gx_i + f$$

is a linear stationary method consistent with $Ax = b$*.*

Proof. Recall that consistency requires $f = (I - G)A^{-1}b$. Therefore,

$$(I - G)^{-1}f = (I - G)^{-1}(I - G)A^{-1}b = A^{-1}b = x$$

$\square$

Error Behavior

$$G = (I - M^{-1}A) \text{ and } x = A^{-1}b$$

$$x_k = Gx_{k-1} + f \text{ and } x = Gx + f$$

$$(x_k - x) = G(x_{k-1} - x) \rightarrow e^{(k)} = Ge^{(k-1)}$$

$$\therefore \|e^{(k)}\| = \|Ge^{(k-1)}\| = \cdots = \|G^k e^{(0)}\|$$

$$\|e^{(k)}\| \leq \|G^k\|\|e^{(0)}\| \leq \|G\|^k\|e^{(0)}\|$$

So $\|G\| < 1$ is again seen as a sufficient condition for convergence to a fixed point.

Error Behavior

Suppose $Gz_i = \lambda_i z_i$ for $1 \leq i \leq n$ with linearly independent z_i and $|\lambda_1| \leq |\lambda_2| \leq \cdots \leq |\lambda_{n-1}| \ll |\lambda_n| < 1$

$$e^{(0)} = \sum_{i=1}^{n} \alpha_i z_i, \quad \|z_i\| = 1$$

$$G^k e^{(0)} = \sum_{i=1}^{n} \alpha_i G^k z_i = \sum_{i=1}^{n} \alpha_i \lambda_i^k z_i$$

If $k \to \infty$ then $\|e^{(k)}\| \to |\alpha_n| \rho^k(G)$

Error Behavior

$$\|e^{(k)}\| \approx |\alpha_n|\rho^k(G) \text{ and } \|e^{(k-1)}\| \approx |\alpha_n|\rho^{k-1}(G)$$

$$\frac{\|e^{(k)}\|}{\|x\|} \to \frac{\|e^{(k)}\|}{\|e^{(k-1)}\|} \approx \frac{\rho^k(G)}{\rho^{k-1}(G)} = \rho(G) \approx 10^{-d}$$

$d = -\log(\rho(G))$ new digits in the approximation of x per step. This is the asymptotic convergence rate.

Error Behavior

- Using $d = -\log(\rho(G))$ and

$$\|e^{(k)}\| \approx \rho^k(G)\|e^{(0)}\|$$

  can be optimistic depending on the properties of the eigenvalues and eigenvectors since $\rho(G) \leq \|G\|_2$.

- $G^T = G \rightarrow \rho(G) = \|G\|_2$ so estimates of the matrix 2 norm are useful

- In general, $\rho(G^T G) = \|G\|_2^2$ so estimating the largest eigevalue of a symmetric matrix is useful.

- We do not have $e^{(k)}$ what about the residual r_k?

Residual Behavior

$$x_k = x_{k-1} + M^{-1} r_{k-1}$$

$$b - Ax_k = b - Ax_{k-1} - AM^{-1} r_{k-1}$$

$$r_k = (I - AM^{-1}) r_{k-1}$$

$$r_k = \tilde{G} r_{k-1} \text{ and } e^{(k)} = G e^{(k-1)}$$

$$G = (I - M^{-1}A) \text{ and } \tilde{G} = (I - AM^{-1})$$

Two different conditions, i.e., $\rho(G)$ and $\rho(\tilde{G})$?

Residual Behavior

We have

$$G = (I - M^{-1}A) \text{ and } \tilde{G} = (I - AM^{-1})$$

Does It follow that $\rho(G) = \rho(\tilde{G})$?

$$M^{-1}\tilde{G}M = M^{-1}(I - AM^{-1})M = (I - M^{-1}A) = G$$

$$Gz = \lambda z \rightarrow M^{-1}\tilde{G}Mz = \lambda z$$

$$\rightarrow \tilde{G}(Mz) = \lambda(Mz) \rightarrow \tilde{G}\tilde{z} = \lambda\tilde{z}$$

$$\therefore \rho(G) = \rho(\tilde{G})$$

Error and Residual Behavior

- G and $\tilde{G}$ are called similar matrices. Similar matrices have the same eigenvalues.
- Expect $\|e^{(k)}\|$ and $\|r_k\|$ to have similar behavior for this class of iterative method.
- When G does not have n linearly independent eigenvectors then highly nonmonotonic convergence behavior may result.
- Then the norm of G^k can grow significantly before it converges to 0.

Richardson's Method via Matrix Splitting

$$
\begin{aligned}
A &= I - (I - A) = M - N \\
x_{i+1} &= (I - A)x_i + b \\
&= x_i + b - Ax_i \\
&= x_i + r_i
\end{aligned}
$$

- Splitting method: $M = I$ and $N = I - A$
- Residual form of method: $P = I$
- Linear Stationary method: $G = I - A$ and $f = b$

Matrix Form- Jacobi Method

If $A = D - L - U$ where $D = diag(\alpha_{11}, \cdots, \alpha_{nn})$ and L and U are the strict lower and upper parts of A then

$$\begin{aligned} Dx_{k+1} =& (L+U)x_k + b \\ x_{k+1} =& D^{-1}(L+U)x_k + D^{-1}b \\ M =& D \\ N =& L+U \end{aligned}$$

Note that for D^{-1} to exist it is necessary and sufficient that $\alpha_{ii} \neq 0$ for $1 \leq i \leq n$.

Matrix Form – Forward GS

If $A = D - L - U$ where $D = diag(\alpha_{11}, \cdots, \alpha_{nn})$ and L and U are the strict lower and upper parts of A then

$$
\begin{aligned}
(D - L)x_{k+1} =& Ux_k + b \\
x_{k+1} =& (D - L)^{-1}Ux_k + (D - L)^{-1}b \\
M =& D - L \\
N =& U
\end{aligned}
$$

Each step requires a lower triangular or forward solve.

Note that for $(D - L)^{-1}$ to exist it is necessary and sufficient that $\alpha_{ii} \neq 0$ for $1 \leq i \leq n$.

Matrix Form – Backward GS

If $A = D - L - U$ where $D = diag(\alpha_{11}, \cdots, \alpha_{nn})$ and L and U are the strict lower and upper parts of A then

$$\begin{aligned}(D-U)x_{k+1} =& Lx_k + b \\ x_{k+1} =& (D-U)^{-1}Lx_k + (D-U)^{-1}b \\ M =& D - U \\ N =& L\end{aligned}$$

Each step requires an upper triangular or backward solve.

Note that for $(D-U)^{-1}$ to exist it is necessary and sufficient that $\alpha_{ii} \neq 0$ for $1 \leq i \leq n$.

Matrix Form – Symmetric Gauss-Seidel

One forward sweep followed by one backward sweep gives SGS

$$(D-L)x_{k+1/2} = Ux_k + b \text{ and } (D-U)x_{k+1} = Lx_{k+1/2} + b$$

$$x_{k+1} = \left[(D-U)^{-1}L(D-L)^{-1}U\right]x_k + (D-U)^{-1}\left[I + L(D-L)^{-1}\right]b$$

$$x_{k+1} = G_{sgs}x_k + b_{sgs}$$

Matrix Form – Symmetric Gauss-Seidel

Theorem 7.7. *Let $A = D - L - L^T$ be symmetric positive definite. For Symmetric Gauss-Seidel we have:*

$$M = (D - L)D^{-1}(D - L^T)$$

$$G_{sgs} = (D - L^T)^{-1}L(D - L)^{-1}L^T = I - M^{-1}A$$

It follows that M is symmetric positive definite and G_{sgs} is similar to a symmetric positive semidefinite matrix

$$\tilde{G}_{sgs} = C^T G_{sgs} C^{-T} = I - C^{-1}AC^{-T}$$

where $M = CC^T$ is the Cholesky factorization of M.

Richardson Family of Methods

The main splitting methods can be summarized in "preconditioned" form.

This include

- simple and accelerated Richardson
- Jacobi
- Gauss-Seidel
- Symmetric Gauss-Seidel
- Jacobi Overrelaxation
- Successive Overrelaxation
- Symmetric Successive Overrelaxation

Richardson Family of Methods

Stationary Richardson's Method

$$x_{k+1} = x_k + \alpha P^{-1} r_k$$

P is the preconditioner and α is an acceleration or relaxation parameter.

1. Solve $Pz_k = r_k$
2. $x_{k+1} = x_k + \alpha z_k$
3. $r_{k+1} = r_k - \alpha A z_k$

Note the addition of a relaxation parameter. For splitting methods $P = M$.

Richardson Family of Methods

The earlier methods can be put in this form to show the preconditioner.

- $A = D - L - U$
- Richardson: $x_{k+1} = x_k + \alpha r_k \rightarrow P_{sr} = \alpha^{-1} I$
- Jacobi:

$$x_{k+1} = x_k + D^{-1} r_k = x_k + P_J^{-1} r_k$$
$$\therefore P_J = D$$

- Gauss-Seidel:

$$x_{k+1} = x_k + (D - L)^{-1} r_k = x_k + P_{gs}^{-1} r_k$$
$$\therefore P_{gs} = D - L$$

Richardson Family of Methods

- Symmetric Gauss-Seidel:

$$x_{k+1/2} = x_k + (D - L)^{-1} r_k$$
$$x_{k+1} = x_{k+1/2} + (D - U)^{-1} r_{k+1/2}$$
$$\Downarrow$$
$$x_{k+1} = x_k + (D - U)^{-1} D (D - L)^{-1} r_k = x_k + P_{sgs}^{-1} r_k$$
$$\therefore P_{sgs} = (D - L) D^{-1} (D - U)$$

- Jacobi overrelaxation:

$$x_{k+1} = x_k + \omega D^{-1} r_k = x_k + P_{jor}^{-1} r_k$$
$$\therefore P_{jor} = \omega^{-1} D$$

Successive Overrelaxation

Applying relaxation successively at the scalar equation level in Gauss-Seidel yields with $0 < \omega < 2$

$$\xi_i^{GS} = \frac{1}{\alpha_{ii}} \left(\beta_i - \sum_{j=1}^{i-1} \alpha_{ij} \xi_j^{(k+1)} - \sum_{j=i+1}^{n} \alpha_{ij} \xi_j^{(k)} \right)$$

$$\xi_i^{(k+1)} = \xi_i^{(k)} + \omega(\xi_i^{GS} - \xi_i^{(k)})$$

$$\Downarrow$$

$$(D - \omega L) x_{k+1} = [\omega U + (1-\omega) D]\, x_k + \omega b$$

$$x_{k+1} = x_k + (\omega^{-1} D - L)^{-1} r_k = x_k + P_{sor}^{-1} r_k$$

$$P_{sor} = (\omega^{-1} D - L)$$

Symmetric Successive Overrelaxation

One forward sweep followed by one backward sweep gives SSOR

$$(D - \omega L)x_{k+1/2} = [\omega U + (1-\omega)D]\, x_k + \omega b$$

$$(D - \omega U)x_{k+1} = [\omega L + (1-\omega)D]\, x_{k+1/2} + \omega b$$

$$x_{k+1} = x_k + \frac{2-\omega}{\omega}(\omega^{-1}D - U)^{-1} D (\omega^{-1}D - L)^{-1} r_k$$

$$x_{k+1} = x_k + P_{ssor}^{-1} r_k$$

$$P_{ssor} = \frac{\omega}{2-\omega}(\omega^{-1}D - L) D^{-1} (\omega^{-1}D - U)$$

The iteration matrix G_ω is similar to a symmetric (Hermitian) if A is symmetric (Hermitian).

Component Point of View

- The matrix form of the splitting approach gives insight from the operator point of view.
- These methods can also be derived by examining the system of equations solved at the component level (or subvector level)
- This is useful preparation for deriving contraction mappings for systems of nonlinear equations.
- The linear structure of the equations makes the task easier here.

Example

$$\begin{pmatrix} 4 & -1 & 0 & 0 & 0 \\ -1 & 4 & -1 & 0 & 0 \\ 0 & -1 & 4 & -1 & 0 \\ 0 & 0 & -1 & 4 & -1 \\ 0 & 0 & 0 & -1 & 4 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \\ \xi_5 \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix}$$

Example

Look at Richardson's method at the component level of $x = x + b - Ax$.

$$
\begin{aligned}
\xi_1^{(i+1)} &= \xi_1^{(i)} + \beta_1 - 4\xi_1^{(i)} + \xi_2^{(i)} \\
\xi_2^{(i+1)} &= \xi_2^{(i)} + \beta_2 + \xi_1^{(i)} - 4\xi_2^{(i)} + \xi_3^{(i)} \\
\xi_3^{(i+1)} &= \xi_3^{(i)} + \beta_3 + \xi_2^{(i)} - 4\xi_3^{(i)} + \xi_4^{(i)} \\
\xi_4^{(i+1)} &= \xi_4^{(i)} + \beta_4 + \xi_3^{(i)} - 4\xi_4^{(i)} + \xi_5^{(i)} \\
\xi_5^{(i+1)} &= \xi_5^{(i)} + \beta_5 + \xi_4^{(i)} - 4\xi_5^{(i)}
\end{aligned}
$$

Other assignments of components values to step i and $i + 1$ are possible to derive other methods.

Jacobi's Method

Let $e_i^T x = \xi_i$, $e_i^T b = \beta_i$, $e_i^T A e_j = \alpha_{i,j}$

To get x_{k+1} from x_k solve each of the n equations independently for its corresponding component of x, i.e.,

$$\alpha_{ii}\xi_i + \sum_{j=1,\ j\neq i}^{n} \alpha_{ij}\xi_j = \beta_i$$

$$\alpha_{ii}\xi_i^{(k+1)} = -\sum_{j=1,\ j\neq i}^{n} \alpha_{ij}\xi_j^{(k)} + \beta_i$$

Note that there is no ordering implied in the solution of these systems.

Example

$$\begin{pmatrix} 4 & -1 & 0 & 0 & 0 \\ -1 & 4 & -1 & 0 & 0 \\ 0 & -1 & 4 & -1 & 0 \\ 0 & 0 & -1 & 4 & -1 \\ 0 & 0 & 0 & -1 & 4 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \\ \xi_5 \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix}$$

$$4\xi_1 - \xi_2 = \beta_1 \rightarrow 4\xi_1^{(i+1)} = \xi_2^{(i)} + \beta_1$$

$$-\xi_1 + 4\xi_2 - \xi_3 = \beta_2 \rightarrow 4\xi_2^{(i+1)} = \xi_1^{(i)} + \xi_3^{(i)} + \beta_2$$

$$-\xi_2 + 4\xi_3 - \xi_4 = \beta_3 \rightarrow 4\xi_3^{(i+1)} = \xi_2^{(i)} + \xi_4^{(i)} + \beta_3$$

$$-\xi_3 + 4\xi_4 - \xi_5 = \beta_4 \rightarrow 4\xi_4^{(i+1)} = \xi_3^{(i)} + \xi_5^{(i)} + \beta_4$$

$$4\xi_5 - \xi_4 = \beta_5 \rightarrow 4\xi_5^{(i+1)} = \xi_4^{(i)} + \beta_5$$

Example

$$\begin{pmatrix} 4 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} \xi_1^{(i+1)} \\ \xi_2^{(i+1)} \\ \xi_3^{(i+1)} \\ \xi_4^{(i+1)} \\ \xi_5^{(i+1)} \end{pmatrix} = \begin{pmatrix} 4\xi_1^{(i+1)} \\ 4\xi_2^{(i+1)} \\ 4\xi_3^{(i+1)} \\ 4\xi_4^{(i+1)} \\ 4\xi_5^{(i+1)} \end{pmatrix}$$

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \xi_1^{(i)} \\ \xi_2^{(i)} \\ \xi_3^{(i)} \\ \xi_4^{(i)} \\ \xi_5^{(i)} \end{pmatrix} + \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix} = \begin{pmatrix} \xi_2^{(i)} + \beta_1 \\ \xi_1^{(i)} + \xi_3^{(i)} + \beta_2 \\ \xi_2^{(i)} + \xi_4^{(i)} + \beta_3 \\ \xi_3^{(i)} + \xi_5^{(i)} + \beta_4 \\ \xi_4^{(i)} + \beta_5 \end{pmatrix}$$

Gauss-Seidel Method

Let $e_i^T x = \xi_i$, $e_i^T b = \beta_i$, $e_i^T A e_j = \alpha_{i,j}$

To get x_{k+1} from x_k solve each of the n equations independently for its corresponding component of x using latest guess for each component, i.e.,

$$\sum_{j=1}^{i-1} \alpha_{ij}\xi_j + \alpha_{ii}\xi_i + \sum_{j=i+1}^{n} \alpha_{ij}\xi_j = \beta_i$$

$$\alpha_{ii}\xi_i^{(k+1)} = -\sum_{j=1}^{i-1} \alpha_{ij}\xi_j^{(k+1)} - \sum_{j=i+1}^{n} \alpha_{ij}\xi_j^{(k)} + \beta_i$$

Note the forward ordering implied in the solution of these systems.

Example

$$\begin{pmatrix} 4 & -1 & 0 & 0 & 0 \\ -1 & 4 & -1 & 0 & 0 \\ 0 & -1 & 4 & -1 & 0 \\ 0 & 0 & -1 & 4 & -1 \\ 0 & 0 & 0 & -1 & 4 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \\ \xi_5 \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix}$$

$$4\xi_1 - \xi_2 = \beta_1 \rightarrow 4\xi_1^{(i+1)} = \xi_2^{(i)} + \beta_1$$

$$-\xi_1 + 4\xi_2 - \xi_3 = \beta_2 \rightarrow 4\xi_2^{(i+1)} = \xi_1^{(i+1)} + \xi_3^{(i)} + \beta_2$$

$$-\xi_2 + 4\xi_3 - \xi_4 = \beta_3 \rightarrow 4\xi_3^{(i+1)} = \xi_2^{(i+1)} + \xi_4^{(i)} + \beta_3$$

$$-\xi_3 + 4\xi_4 - \xi_5 = \beta_4 \rightarrow 4\xi_4^{(i+1)} = \xi_3^{(i+1)} + \xi_5^{(i)} + \beta_4$$

$$4\xi_5 - \xi_4 = \beta_5 \rightarrow 4\xi_5^{(i+1)} = \xi_4^{(i+1)} + \beta_5$$

Example

$$\begin{pmatrix} 4 & 0 & 0 & 0 & 0 \\ -1 & 4 & 0 & 0 & 0 \\ 0 & -1 & 4 & 0 & 0 \\ 0 & 0 & -1 & 4 & 0 \\ 0 & 0 & 0 & -1 & 4 \end{pmatrix} \begin{pmatrix} \xi_1^{(i+1)} \\ \xi_2^{(i+1)} \\ \xi_3^{(i+1)} \\ \xi_4^{(i+1)} \\ \xi_5^{(i+1)} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \xi_1^{(i)} \\ \xi_2^{(i)} \\ \xi_3^{(i)} \\ \xi_4^{(i)} \\ \xi_5^{(i)} \end{pmatrix} + \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix}$$

Backward Gauss-Seidel Method

The ordering of Gauss-Seidel need not be $i = 1, \dots n$. Let $e_i^T x = \xi_i$, $e_i^T b = \beta_i$, $e_i^T A e_j = \alpha_{i,j}$

To get x_{k+1} from x_k solve each of the n equations independently for its corresponding component of x using latest guess for each component, i.e.,

$$\sum_{j=1}^{i-1} \alpha_{ij}\xi_j + \alpha_{ii}\xi_i + \sum_{j=i+1}^{n} \alpha_{ij}\xi_j = \beta_i$$

$$\alpha_{ii}\xi_i^{(k+1)} = -\sum_{j=1}^{i-1} \alpha_{ij}\xi_j^{(k)} - \sum_{j=i+1}^{n} \alpha_{ij}\xi_j^{(k+1)} + \beta_i$$

Note the backward ordering implied in the solution of these systems.

Example

$$\begin{pmatrix} 4 & -1 & 0 & 0 & 0 \\ -1 & 4 & -1 & 0 & 0 \\ 0 & -1 & 4 & -1 & 0 \\ 0 & 0 & -1 & 4 & -1 \\ 0 & 0 & 0 & -1 & 4 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \\ \xi_5 \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix}$$

$$4\xi_1 - \xi_2 = \beta_1 \rightarrow 4\xi_1^{(i+1)} = \xi_2^{(i+1)} + \beta_1$$

$$-\xi_1 + 4\xi_2 - \xi_3 = \beta_2 \rightarrow 4\xi_2^{(i+1)} = \xi_1^{(i)} + \xi_3^{(i+1)} + \beta_2$$

$$-\xi_2 + 4\xi_3 - \xi_4 = \beta_3 \rightarrow 4\xi_3^{(i+1)} = \xi_2^{(i)} + \xi_4^{(i+1)} + \beta_3$$

$$-\xi_3 + 4\xi_4 - \xi_5 = \beta_4 \rightarrow 4\xi_4^{(i+1)} = \xi_3^{(i)} + \xi_5^{(i+1)} + \beta_4$$

$$4\xi_5 - \xi_4 = \beta_5 \rightarrow 4\xi_5^{(i+1)} = \xi_4^{(i)} + \beta_5$$