Based on Strang’s *Introduction to Applied Mathematics*

**Theory of Iterative Methods**

**The Iterative Idea**

To solve $Ax = b$, write

$$Mx^{(k+1)} = (M - A)x^{(k)} + b, \quad k = 0, 1, 2, \ldots$$

Then the error $e^{(k)} \equiv x^{(k)} - x$ satisfies

$$Me^{(k+1)} = (M - A)e^{(k)}, \quad e^{(k+1)} = Be^{(k)}$$

where the iteration matrix $B = M^{-1}(M - A)$. Now

$$||e^{(k)}|| = ||B^k e^{(0)}|| \sim \rho^k \to 0 \text{ iff } \rho < 1$$

where $\rho = \text{maximum of } |\text{eigenvalues of } B|$ is the spectral radius of $B$.

**Convergence of Iterative Methods**

*Convergence.* Does $x^{(k)} \to x$? If $x^{(k)} \to x^*$, then

$$Mx^* = (M - A)x^* + b$$

and $Ax^* = b$. The rate of convergence is governed by powers of $B = M^{-1}(M - A)$. Since $e^{(k)} = B^k e^{(0)}$, $e^{(k)} \to 0$ and $x^{(k)} \to x$ (convergence) iff $B^k \to 0$ (stability).

*Theorem.* $B^k \to 0$ iff every eigenvalue of $B$ satisfies $|\lambda_i| < 1$. The rate of convergence is governed by the spectral radius $\rho$ of $B$:

$$\rho = \max_i |\lambda_i|.$$  

*Proof using eigenvectors.* Expand the initial error in terms of the eigenvectors of $B$ (assuming a complete set of eigenvectors)

$$e^{(0)} = c_1 v_1 + c_2 v_2 + \cdots + c_n v_n$$
where $Bv_i = \lambda_i v_i$. Then as $k \to \infty$, the error at iteration $k$ is
\[ e^{(k)} = B^k e^{(0)} = c_1 \lambda_1^k v_1 + c_2 \lambda_2^k v_2 + \cdots + c_n \lambda_n^k v_n \to 0 \text{ iff } |\lambda_i| < 1, \quad i = 1, \ldots, n \]
and
\[ \|e^{(k)}\| \to \|c_j \lambda_j^k v_j\| \sim \rho^k. \]

**Proof in matrix form.** Assuming a complete set of eigenvectors, write $B = S^{-1} \Lambda S$, where $S = [v_1 \ v_2 \ \cdots \ v_n]$ is the eigenvector matrix of $B$ and $\Lambda = \text{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ is the eigenvalue matrix of $B$. Then
\[ B^k = \left(S^{-1} \Lambda S\right) \left(S^{-1} \Lambda S\right) \cdots \left(S^{-1} \Lambda S\right) = S^{-1} \Lambda^k S \to 0 \text{ iff } |\lambda_i| < 1, \quad i = 1, \ldots, n. \]
If $B$ does not have a complete set of eigenvectors, the matrix form of the proof simply involves the Jordan form $J$ of $B$ instead of $\Lambda$.

**Choice of $M$**

- $Mx^{(k+1)} = (M - A)x^{(k)} + b$ should be easy to solve. Diagonal or lower triangular $M$’s are good choices.
- $M$ should be close to $A$, so that the eigenvalues of $B = M^{-1}(M - A) = I - M^{-1}A$ are as small in magnitude as possible (must be inside the unit circle in the complex plane).

For $A = L + D + U$, Jacobi takes $M = D$ while Gauss-Seidel takes $M = D + L$. SOR (successive over-relaxation) introduces a relaxation factor $1 < \omega < 2$ in Gauss-Seidel which is adjusted to make the spectral radius $\rho$ as small as possible.

For a wide class of finite-difference matrices, Young’s formula (1950) relates the eigenvalues $\mu$ of Jacobi and the eigenvalues $\lambda$ of SOR:
\[ (\lambda + \omega - 1)^2 = \lambda \omega^2 \mu^2. \]
Minimizing $\rho_{SOR} = \max\{||\lambda||\} = \omega - 1$ (using the quadratic formula) gives
\[ \omega_{opt} = \frac{2 \left(1 - \sqrt{1 - \rho_J^2}\right)}{\rho_J^2}, \quad \rho_{SOR} = \omega_{opt} - 1. \]
For the model Laplace problem on an $N \times N$ grid with $h = 1/N$,
\[ \rho_J = \cos(\pi h), \quad \rho_{GS} = \rho_J^2, \quad \rho_{SOR} = \frac{1 - \sin \pi h}{1 + \sin \pi h}. \]
Model Laplace Problem Spectral Radii

To show that the Jacobi spectral radius $\rho_J = \cos(\pi h)$ for Laplace’s equation on the unit square with second-order accurate central differences, first consider a 1D problem. In 1D, set $A = \text{tridiag}[-1 2 -1]$. Then the iteration matrix $B = \frac{1}{2} \text{tridiag}[1 0 1]$. Then show that $Bv = \cos(\pi h) v$ where the 1D eigenvector

$$v = [\sin(\pi h), \sin(2\pi h), \cdots, \sin(n\pi h)], \quad n = N - 1, \quad h = \frac{1}{N} = \frac{1}{n+1}.$$  

The other eigenvectors of $B$ replace $\pi$ with $2\pi$, $3\pi$, $\ldots$, $n\pi$ in $v$, with eigenvalues $\cos(2\pi h)$, $\cos(3\pi h)$, $\ldots$, $\cos(n\pi h)$.

In 2D, the eigenvector $= [\sin(j\pi h) \sin(k\pi h)] = \sin(\pi h) \sin(\pi h), \sin(2\pi h) \sin(2\pi h), \cdots, \sin(n\pi h) \sin(n\pi h)$, $\sin(2\pi h) \sin(\pi h), \cdots, \sin(2\pi h) \sin(n\pi h), \cdots, \sin(n\pi h) \sin(n\pi h)]$ with the same eigenvalue, so $\rho_J = \cos(\pi h)$.

Using $\rho_J$ in Young’s formula yields the SOR spectral radius

$$\rho_{SOR} = \frac{1 - \sin \pi h}{1 + \sin \pi h}.$$  

Iteration Matrices

Solve $Ax = b$ iteratively. Decompose $A = L + D + U$ and define $r^{(k)} = Ax^{(k)} - b$.

Jacobi Iteration $M = D$

$Dx^{(k+1)} = -(L + U)x^{(k)} + b = Dx^{(k)} - r^{(k)}$

$x^{(k+1)} = x^{(k)} - D^{-1}r^{(k)}$

$B_J = -D^{-1}(L + U)$

Gauss-Seidel Iteration $M = L + D$


\[(D + L)x^{(k+1)} = -Ux^{(k)} + b\]
\[Dx^{(k+1)} = Dx^{(k)} - (Lx^{(k+1)} + Dx^{(k)} + Ux^{(k)} - b) \equiv Dx^{(k)} - \tilde{r}^{(k)}\]
\[x^{(k+1)} = x^{(k)} - D^{-1}\tilde{r}^{(k)}\]
\[B_{GS} = -(L + D)^{-1}U\]

SOR/SUR Iteration \(M = \frac{D}{\omega} + L\)

\[x^{(k+1)} = x^{(k)} - \omega D^{-1}\tilde{r}^{(k)}\]
\[\frac{D}{\omega}x^{(k+1)} = \frac{D}{\omega}x^{(k)} - (Lx^{(k+1)} + Dx^{(k)} + Ux^{(k)} - b)\]
\[\left(\frac{D}{\omega} + L\right)x^{(k+1)} = \left(\frac{D}{\omega} - D - U\right)x^{(k)} + b\]
\[B_{SOR} = \left(\frac{D}{\omega} + L\right)^{-1}\left(\frac{D}{\omega} - D - U\right) = (D + \omega L)^{-1}((1 - \omega)D - \omega U)\]

Note that for SOR/SUR, \(\det\{B\} = (1 - \omega)^n\), \(0 < \omega < 2\), and \(\rho_{SOR} = \omega_{opt} - 1\), \(\rho_{SUR} = 1 - \omega_{opt}\).

2 \times 2 Example

\[\det\{A\} = \ldots \lambda_n, \quad \text{Tr}\{A\} = \lambda_1 + \lambda_2 + \cdots + \lambda_n\]

\[A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}\]

\[M_J = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \quad B_J = \begin{bmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{bmatrix}, \quad \lambda_{\pm} = \pm \frac{1}{2}, \quad \rho_J = \frac{1}{2}\]

\[M_{GS} = \begin{bmatrix} 2 & 0 \\ -1 & 2 \end{bmatrix}, \quad B_{GS} = \begin{bmatrix} 0 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{bmatrix}, \quad \lambda_1 = 0, \quad \lambda_2 = \frac{1}{4}, \quad \rho_{GS} = \frac{1}{4}\]

\[M_{SOR} = \begin{bmatrix} \frac{2}{\omega} & 0 \\ -1 & \frac{2}{\omega} \end{bmatrix}, \quad B_{SOR} = \begin{bmatrix} 1 - \omega & \frac{2}{\omega} \\ \omega(1 - \omega) & (1 - \omega)^2 \end{bmatrix}\]

\[\omega_{opt} = 4(2 - \sqrt{3}), \quad \lambda_1 = \lambda_2 = \omega_{opt} - 1 = \rho_{SOR} \approx 0.0718\]
$3 \times 3$ Examples

(i) Example where Jacobi converges but Gauss-Seidel diverges

\[
A = \begin{bmatrix} 1 & 2 & -2 \\ 1 & 1 & 1 \\ 2 & 2 & 1 \end{bmatrix}, \quad M_J = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad B_J = \begin{bmatrix} 0 & -2 & 2 \\ -1 & 0 & -1 \\ -2 & -2 & 0 \end{bmatrix}
\]

$\lambda_1 = \lambda_2 = \lambda_3 = 0$, $\rho_J = 0$, Jacobi converges. Note that here $B^3 = 0$ and $e_3 = 0$, which is an unusual situation! See the Cayley-Hamilton Theorem in the web notes Conjugate Gradients.

\[
M_{GS} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 2 & 2 & 1 \end{bmatrix}, \quad B_{GS} = \begin{bmatrix} 0 & -2 & 2 \\ 0 & 2 & -3 \\ 0 & 0 & 2 \end{bmatrix}
\]

$\lambda_1 = 0$, $\lambda_2 = \lambda_3 = 2$, $\rho_{GS} = 2$, Gauss-Seidel diverges.

(ii) Example where Jacobi diverges but Gauss-Seidel converges

\[
A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}, \quad M_J = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \quad B_J = \frac{1}{2} \begin{bmatrix} 0 & -1 & -1 \\ -1 & 0 & -1 \\ -1 & -1 & 0 \end{bmatrix}
\]

$\lambda_1 = -1$, $\lambda_2 = \lambda_3 = 0.5$, $\rho_J = 1$, Jacobi diverges.

\[
M_{GS} = \begin{bmatrix} 2 & 0 & 0 \\ 1 & 2 & 0 \\ 1 & 1 & 2 \end{bmatrix}, \quad B_{GS} = \frac{1}{8} \begin{bmatrix} 0 & -4 & -4 \\ 0 & 2 & -2 \\ 0 & 1 & 3 \end{bmatrix}
\]

$\lambda_1 = 0$, $\lambda_\pm = 0.3125 \pm 0.1654i$, $\rho_{GS} = 0.3536$, Gauss-Seidel converges.

**Chebyshev SOR**

By dynamically adjusting $\omega$, the Chebyshev SOR method insures that the norm of the error always decreases. Make two half sweeps on even/odd (black/white) meshes. Odd (even) points depend only on even (odd) mesh values. Define ($\mu$ is the Jacobi spectral radius):

\[
\omega^{(0)} = 1
\]
\[
\omega\left(\frac{1}{2}\right) = \frac{1}{1 - \frac{\mu^2}{2}} \\
\omega\left(k + \frac{1}{2}\right) = \frac{1}{1 - \frac{\mu^2\omega(k)}{4}}, \quad k = \frac{1}{2}, 1, \frac{3}{2}, \ldots
\]

Then \(\omega(\infty) = \omega_{opt}\) and \(||e^{(k)}||\) always decreases.