Lecture 21: Convergence of Iterative Methods

July 21, 2025

Lecture 21: Convergence of Iterative Methods - Outline

Convergence of Iterative Methods - Introduction

We will now revisit iterative schemes to analyze aspects of their convergence behaviour in detail. In this lecture we will study the stationary iterative methods:

1. Richardson,
2. Jacobi,
3. Gauss-Seidel, and
4. Successive-Over-Relaxation (SOR).

These methods were first discussed in Lecture 08.

Convergence of Iterative Methods - Introduction

Recall that the stationary iterative methods amount to different choices of M when splitting $A = M - N$. The generic iteration is

$$x^{k+1} = x^k + M^{-1}(b - Ax^k).$$

For each method we have the following splittings of the matrix A:

1. Richardson: $M = \frac{1}{\theta} I$ for scalar $\theta > 0$,
2. Jacobi: $M = D$,
3. Gauss-Seidel: $M = D - L$,
4. SOR: $\frac{1}{\omega} D - L$ for scalar $\omega > 0$.
   - $0 < \omega < 1$ indicates under-relaxation;
   - $1 < \omega$ indicates over-relaxation.

$$A = \begin{bmatrix} \ddots & & -U \\ & D & \\ -L & & \ddots \end{bmatrix}$$

Convergence of Iterative Methods - Introduction

We can rewrite the generic iteration as

$$x^{k+1} = \boxed{(I - M^{-1}A)}\, x^k + M^{-1}b.$$

- Then we call $G = I - M^{-1}A$ the **iteration matrix** for the scheme.
- The method converges if and only if $\rho(I - M^{-1}A) < 1$, where $\rho(\cdot)$ denotes the **spectral radius** of a matrix (i.e., maximum eigenvalue magnitude). See Lecture 08.
- Note that a smaller ρ implies faster convergence to the solution.
- We will now consider the convergence behaviour for SPD matrices.

Convergence of Iterative Methods - Richardson Convergence

The iteration matrix for the Richardson iteration is

$$\begin{aligned} G^{Rich} &= I - M_{Rich}^{-1} A \\ &= I - \theta A, \end{aligned}$$

for scalar $\theta > 0$. Let (λ, v) be an eigenpair of A. Then,

$$\begin{aligned} G^{Rich} v &= (I - \theta A) v \\ &= v - \theta \lambda v \\ &= (1 - \theta \lambda) v. \end{aligned}$$

Therefore, $\mu = 1 - \theta\lambda$ is an eigenvalue for G^{Rich}.

Convergence of Iterative Methods - Richardson Convergence

Lemma 1

Let λ_{min} and λ_{max} satisfy $\lambda_{min} \leq \lambda_i \leq \lambda_{max}, \forall i$. Then $\rho(G^{Rich}) = \max\{|1 - \theta\lambda_{min}|, |1 - \theta\lambda_{max}|\}$.

Proof.

Let i be arbitrary. Then

$$\begin{aligned} \lambda_{min} &\leq \lambda_i \leq \lambda_{max}, \\ 1 - \theta\lambda_{min} &\geq 1 - \theta\lambda_i \geq 1 - \theta\lambda_{max}, \\ \Rightarrow |\mu| &\leq \max\{|1 - \theta\lambda_{min}|, |1 - \theta\lambda_{max}|\}. \end{aligned}$$

□

Convergence of Iterative Methods - Richardson Convergence

Note, if $\lambda_{min} < 0$ and $\lambda_{max} > 0$ then either

$$1 - \theta\lambda_{min} > 1 \text{ if } \theta > 0 \text{ or,}$$
$$1 - \theta\lambda_{max} > 1 \text{ if } \theta < 0.$$

Hence, $\rho(G^{Rich}) > 1$ for this case and Richardson will diverge for such matrices. (Recall the condition on ρ was necessary and sufficient for convergence).
If we assume that A is SPD, then its eigenvalues **cannot** be negative.
Also, we usually assume that $\theta > 0$.

Convergence of Iterative Methods - Richardson Convergence

Theorem 2

Assume all eigenvalues of A are positive (i.e., A is positive definite). Then Richardson converges iff $0 < \theta < \frac{2}{\lambda_{max}}$.

Proof. If $0 < \theta < \frac{2}{\lambda_{max}}$, then multiplying through by λ_{max} (and inserting the obvious $\theta\lambda_{min} \leq \theta\lambda_{max}$) yields

$$\begin{aligned} 0 < \theta\lambda_{min} &\leq \theta\lambda_{max} < 2, \\ -2 < -\theta\lambda_{max} &\leq -\theta\lambda_{min} < 0, \quad \text{(multiply by -1)} \\ -1 < 1 - \theta\lambda_{max} &\leq 1 - \theta\lambda_{min} < 1 \quad \text{(add one)} \end{aligned}$$

Therefore, $|1 - \theta\lambda_{max}| < 1$ and $|1 - \theta\lambda_{min}| < 1 \Rightarrow \rho(G^{Rich}) < 1$.

Convergence of Iterative Methods - Richardson Convergence

For the other direction assume $\rho(G^{Rich}) < 1$, then

$$-1 < 1 - \theta\lambda_{max} \leq \mu \leq 1 - \theta\lambda_{min} < 1. \tag{1}$$

From the left inequality of (1) we have

$$\begin{aligned} -1 &< 1 - \theta\lambda_{max}, \\ -2 &< -\theta\lambda_{max}, \\ \Rightarrow \theta &< \frac{2}{\lambda_{max}}. \end{aligned}$$

The right inequality of (1) gives

$$\begin{aligned} 1 - \theta\lambda_{min} &< 1, \\ -\theta\lambda_{min} &< 0, \\ \Rightarrow \theta &> 0. \qquad (\text{since } \lambda_{min} > 0) \end{aligned}$$

So $0 < \theta < \frac{2}{\lambda_{max}}$. $\square$

Convergence of Iterative Methods - Richardson Convergence - Choosing Optimal θ

Assume A is PD. Assume $\theta > 0$. To optimize convergence speed we must minimize $\rho(G^{Rich})$. Eigenvalues of $A \in [\lambda_{min}, \lambda_{max}]$, so eigenvalues of Richardson iteration matrix $I - \theta A$ are in $[1 - \theta\lambda_{max}, 1 - \theta\lambda_{min}]$. Plotting this range gives the blue region in Figure 1 (left). But to get the minimum spectral radius, we need the absolute value. Reflecting negative parts over the x-axis gives Figure 1 (right).

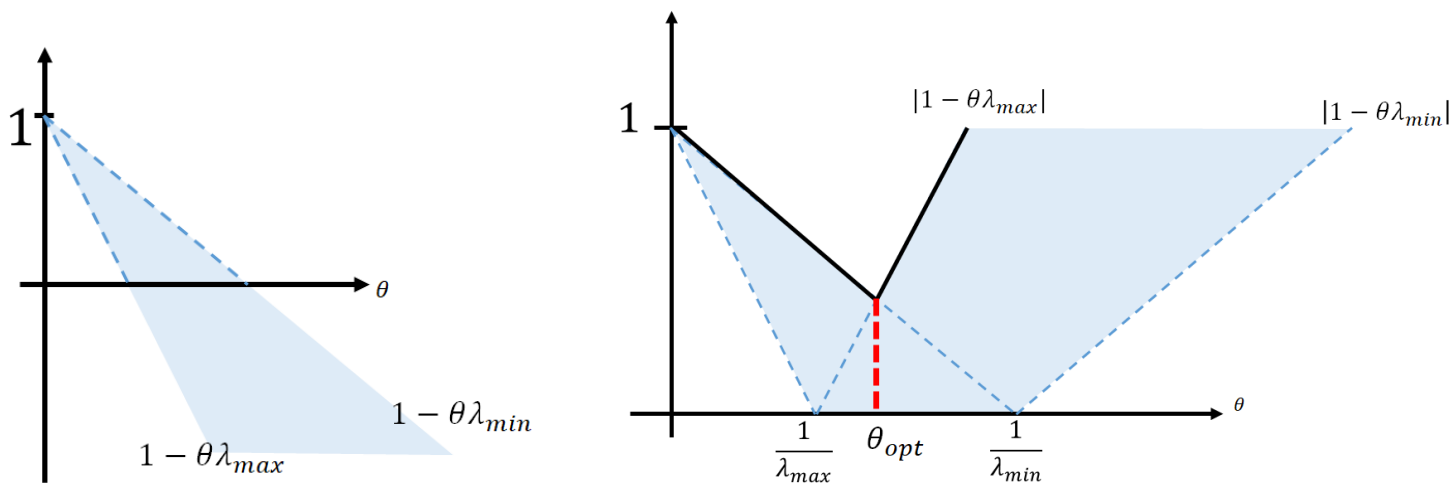


Figure: Finding the optimal θ for Richardson iteration.

Convergence of Iterative Methods - Richardson Convergence - Choosing Optimal θ

Remark: The vertical axis is ρ.
For any choice of θ, the largest magnitude eigenvalue will sit at the top of the blue band, shown by the black line in Figure 1 (right). Thus ρ is minimized where the two lines $|1-\theta\lambda_{min}|$ and $|1-\theta\lambda_{max}|$ intersect. Hence, we must find where $|1-\theta\lambda_{max}| = |1-\theta\lambda_{min}|$ since this is where the largest μ "switches" lines. That is, the optimal θ is when

$$\begin{aligned}
-(1-\theta_{opt}\lambda_{max}) &= 1-\theta_{opt}\lambda_{min} \\
-1+\theta_{opt}\lambda_{max} &= 1-\theta_{opt}\lambda_{min} \\
\theta_{opt}\lambda_{max}+\theta_{opt}\lambda_{min} &= 2 \\
\theta_{opt}(\lambda_{max}+\lambda_{min}) &= 2 \\
\theta_{opt} &= \frac{2}{\lambda_{min}+\lambda_{max}}.
\end{aligned}$$

Convergence of Iterative Methods - Richardson

Convergence - Choosing Optimal θ

Plugging θ_{opt} back in to find corresponding ρ gives

$$\begin{aligned}
\rho_{opt} &= 1 - \theta_{opt}\lambda_{min}, \\
&= 1 - \frac{2\lambda_{min}}{\lambda_{min} + \lambda_{max}}, \\
&= \left(\frac{\lambda_{max} + \lambda_{min} - 2\lambda_{min}}{\lambda_{max} + \lambda_{min}} \right) \\
&= \left(\frac{\lambda_{max} - \lambda_{min}}{\lambda_{max} + \lambda_{min}} \right) \left(\frac{\frac{1}{\lambda_{min}}}{\frac{1}{\lambda_{min}}} \right) \\
&= \frac{\frac{\lambda_{max}}{\lambda_{min}} - 1}{\frac{\lambda_{max}}{\lambda_{min}} + 1} \\
&= \frac{\kappa_2(A) - 1}{\kappa_2(A) + 1}.
\end{aligned}$$

Recall that $\kappa_2(A) = \frac{|\lambda_{max}|}{|\lambda_{min}|}$ and $\lambda > 0$ was assumed in Theorem 2.

Convergence of Iterative Methods - Richardson Convergence - Choosing Optimal θ

Alternatively,

$$\begin{aligned}
\rho_{opt} &= -(1-\theta_{opt}\lambda_{max}) \\
&= -1+\theta_{opt}\lambda_{max} \\
&= -1+\left(\frac{2}{\lambda_{min}+\lambda_{max}}\right)\lambda_{max} \\
&= \frac{-(\lambda_{min}+\lambda_{max})+2\lambda_{max}}{\lambda_{min}+\lambda_{max}} \\
&= \frac{\lambda_{max}-\lambda_{min}}{\lambda_{min}+\lambda_{max}},
\end{aligned}$$

the same expression as in the middle of the previous computation!
Moral: It does not matter which line we take: θ_{opt} was computed using their intersection.

Convergence of Iterative Methods - Richardson Convergence - Choosing Optimal θ

Note that

1. we need eigenvalues (or estimates) to choose optimal θ, and
2. convergence can be slow, depending on λ's. E.g. the convergence can be poor when $\frac{\lambda_{max}}{\lambda_{min}} \approx -1$, so opposite signs.

Convergence of Iterative Methods - Jacobi, Gauss-Seidel, & SOR Convergence

Theorem 3

If A and $2D - A$ are SPD, then the Jacobi iteration converges.

Proof. Let μ be an eigenvalue of $G_J = I - M_J^{-1}A = I - D^{-1}A$, with eigenvector v. Then

$$\begin{aligned}
(I - D^{-1}A)v &= \mu v, \\
D^{-1}(D - A)v &= \mu v, \\
(D - A)v &= \mu Dv, \\
v^T(D - A)v &= \mu v^T Dv, \\
v^T Dv - v^T Av &= \mu v^T Dv, \\
v^T Dv - \mu v^T Dv &= v^T Av \\
(1 - \mu)v^T Dv &= v^T Av \\
&> 0, \text{ since } A \text{ is SPD.}
\end{aligned}$$

So $(1 - \mu)v^T Dv > 0$, which implies $\mu < 1$, because $v^T Dv > 0$, since A is SPD and hence D is also SPD (See Lecture Notes for

Convergence of Iterative Methods - Jacobi, Gauss-Seidel, & SOR Convergence

Similarly, since $2D - A$ is SPD,

$$\begin{aligned} v^T(2D - A)v &> 0 \\ v^T Dv - v^T Av &> -v^T Dv \\ v^T(D - A)v &> -v^T Dv. \end{aligned}$$

Also, as above:

$$v^T (D - A)\, v = \mu v^T Dv,$$

Convergence of Iterative Methods - Jacobi, Gauss-Seidel, & SOR Convergence

and thus we can continue the above sequence of inequalities:

$$\begin{aligned} \mu v^T D v &> -v^T D v \\ (\mu + 1) v^T D v &> 0 \\ \Rightarrow \mu &> -1, \text{ since } D \text{ is SPD.} \end{aligned}$$

Hence, $-1 < \mu < 1 \Rightarrow \rho\left(G^J\right) < 1$, i.e. a Jacobi iteration converges. $\square$

Convergence of Iterative Methods - Jacobi, Gauss-Seidel, & SOR Convergence

Theorem 4

If A is SPD then GS and SOR (for $0 < \omega < 2$) both converge.

- The optimal value of ω for SOR is not known in general.
- It is only known for special cases, e.g., for A that are tridiagonal, and SPD, we have

$$\omega_{opt} = \frac{2}{1 + \sqrt{1 - \rho(G_J)^2}},$$

where G_J is Jacobi's iteration matrix.

[Proof still needed]

Convergence of Iterative Methods - Jacobi, Gauss-Seidel, & SOR Convergence

Gauss-Seidel and Jacobi also converge for another class of matrices, called M-matrices.

Definition 3.1

A is an **M-matrix** *if*

1. $a_{ii} > 0$,
2. $a_{ij} \leq 0$ *for* $i \neq j$,
3. A^{-1} *exists and* $(A^{-1})_{ij} \geq 0, \forall i, j$.

Theorem 5

If A is an M-matrix then Jacobi and GS converge and

$$\rho(I - M_{GS}^{-1}A) \leq \rho(I - M_J^{-1}A) < 1,$$

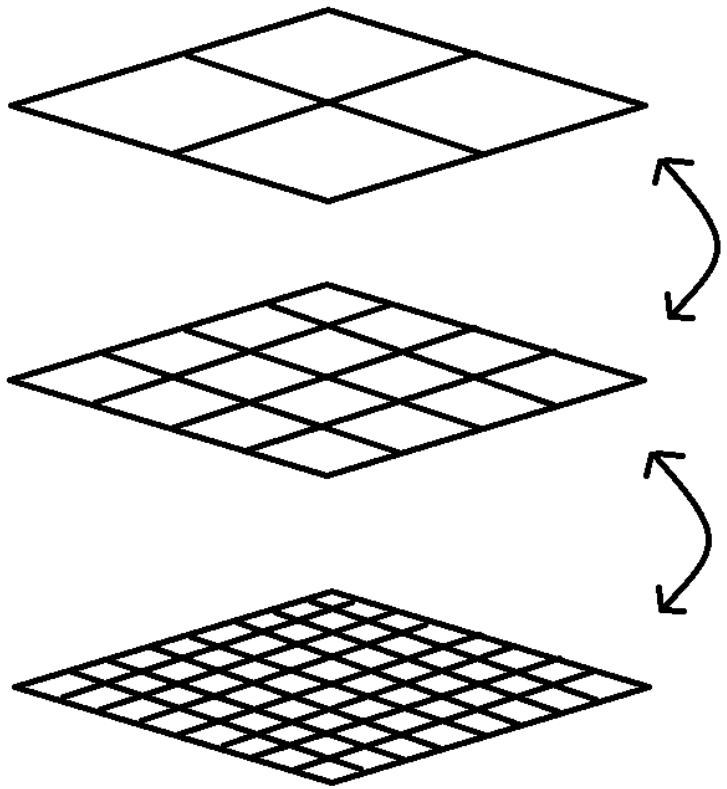
i.e., GS converges at least as rapidly as Jacobi.

[Proof still needed]

Convergence of Iterative Methods - Jacobi, Gauss-Seidel, & SOR Convergence

Stationary iteration methods are often slow for low-frequency error components. Their modern use is often in "multigrid" methods, that use multiple grid levels to find a solution more quickly.

Since low frequency data has relatively high frequency on coarser grids, those errors can be eliminated more quickly, by transferring the solution between levels.



Convergence of Iterative Methods - Convergence on Discrete Poisson Equation

Analytical expressions can be found for eigendecompositions of certain matrices. This can give us a sense for how our iterative schemes fare in practice. We will consider the familiar 2D finite difference Laplacian matrix for the Poisson equation:

$$\begin{aligned} -\nabla \cdot \nabla u &= f, \\ -u_{xx} - u_{yy} &= f. \end{aligned}$$

Note that the negative sign in front of the Laplacian makes the finite difference matrix positive definite (otherwise it is negative definite).

Convergence of Iterative Methods - Convergence on Discrete Poisson Equation

Theorem 6

Let A be the negative of a 2D Laplacian matrix with cell size h and m grid points in each axis. Then the exact eigenvalues are

$$\lambda_{ij} = \frac{4}{h^2}\left[\sin^2\left(\frac{\pi h i}{2}\right) + \sin^2\left(\frac{\pi h j}{2}\right)\right] \text{ for } 1 \leq i, j \leq m.$$

[Proof still needed]

Convergence of Iterative Methods - Convergence on Discrete Poisson Equation

Note that the smallest eigenvalue is

$$\lambda_{min} = \frac{8}{h^2} \sin^2 \left(\frac{\pi h}{2} \right),$$

and the largest eigenvalue is

$$\begin{aligned}
\lambda_{max} &= \frac{8}{h^2} \sin^2 \left(\frac{m\pi h}{2} \right), \\
&= \frac{8}{h^2} \sin^2 \left(\frac{\pi}{2}(1-h) \right) \text{ , using } h = \frac{1}{m+1} \text{ or } mh = 1-h, \\
&= \frac{8}{h^2} \cos^2 \left(\frac{\pi h}{2} \right) \text{ , using } \sin \left(\frac{\pi}{2} - u \right) = \cos(u).
\end{aligned}$$

Convergence of Iterative Methods - Convergence on Discrete Poisson Equation

The matrix A is both SPD and an M-matrix. Furthermore, the conditioning of A gets worse with finer grids, e.g., consider two grid resolution cases

$$
\begin{aligned}
1)\ h &= \frac{1}{10},\ m = 9, & 2)\ h &= \frac{1}{100},\ m = 99,\\
\lambda_{min} &\approx 19.6, & \lambda_{min} &\approx 19.7,\\
\lambda_{max} &\approx 780, & \lambda_{max} &\approx 80000,\\
\kappa_2 &= \frac{\lambda_{max}}{\lambda_{min}} & \kappa_2 &= \frac{\lambda_{max}}{\lambda_{min}}\\
&\approx 40. & &\approx 4000.
\end{aligned}
$$

Finer Resolution $\rightarrow$ Worse conditioning.

Convergence of Iterative Methods - Convergence on Discrete Poisson Equation - Richardson

- We will finish this lecture by showing convergence for the Poisson equation with the stationary iterative methods.
- For the **Richardson** iteration we have

$$\rho(I - \theta A) = \max\left\{\left|1 - \theta\frac{8}{h^2}\sin^2\left(\frac{\pi h}{2}\right)\right|, \left|1 - \theta\frac{8}{h^2}\cos^2\left(\frac{\pi h}{2}\right)\right|\right\}.$$

Hence, Richardson converges for

$$0 < \theta < \frac{2}{\lambda_{max}} = \frac{h^2}{4\cos^2\left(\frac{\pi h}{2}\right)}.$$

Convergence of Iterative Methods - Convergence on Discrete Poisson Equation - Richardson

The optimal θ is

$$\theta_{opt} = \frac{2}{\lambda_{max} + \lambda_{min}} = \frac{2}{\frac{8}{h^2}\left(\cos^2\left(\frac{\pi h}{2}\right) + \sin^2\left(\frac{\pi h}{2}\right)\right)^1} = \frac{h^2}{4},$$

which gives optimal convergence with

$$\begin{aligned}
\rho_{opt} &= \frac{\lambda_{max} - \lambda_{min}}{\lambda_{max} + \lambda_{min}} \\
&= \frac{\frac{8}{h^2}\left(\cos^2\left(\frac{\pi h}{2}\right) - \sin^2\left(\frac{\pi h}{2}\right)\right)}{\frac{8}{h^2}} \\
&= \cos^2\left(\frac{\pi h}{2}\right) - \sin^2\left(\frac{\pi h}{2}\right) \\
&= 1 - 2\sin^2\left(\frac{\pi h}{2}\right), \text{ since } \cos^2 x = 1 - \sin^2 x \\
&= \cos(\pi h), \text{ since } 1 - 2\sin^2(u) = \cos(2u).
\end{aligned}$$

Convergence of Iterative Methods - Convergence on Discrete Poisson Equation - Jacobi

Since

$$\begin{aligned} D &= \frac{4}{h^2} I \\ &= \theta_{opt}^{-1} I, \end{aligned}$$

therefore we have that

$$\begin{aligned} G_J &= I - D^{-1} A \\ &= I - \theta_{opt} A. \end{aligned}$$

Therefore, Jacobi iteration is equivalent to the optimal Richardson iteration for this case, and hence

$$\rho = \cos(\pi h).$$

Convergence of Iterative Methods - Convergence on Discrete Poisson Equation - Jacobi

The Taylor expansion of $\cos(x)$ gives

$$\cos(x) = 1 - \frac{x^2}{2} + \frac{x^4}{4} + \dots .$$

Therefore,

$$\begin{aligned} \rho\left(G^J\right) &= \rho\left(I - D^{-1}A\right) \\ &= \cos(\pi h) \\ &= 1 - \frac{\pi^2 h^2}{2} + O\left(h^4\right). \end{aligned}$$

For small h, Jacobi (and optimal Richardson) has slow convergence, since $\rho\left(G^J\right) \approx 1$.

Convergence of Iterative Methods - Convergence on Discrete Poisson Equation - GS and SOR

The spectral radius for Gauss-Seidel is the square of that for Jacobi **[Proof still needed - See 1948 Paper "On the Solution of Linear Simultaneous Equations By Iteration", by Stein and Rosenberg]**, and so we have

$$\begin{aligned}
\rho\left(I - M_{GS}^{-1}A\right) &= \left[\rho(I - M_J^{-1}A)\right]^2 \\
&= \cos^2(\pi h), \\
&= 1 - \sin^2(\pi h), \\
&= 1 - \pi^2 h^2 + O(h^4). \text{ (Taylor expansion)}
\end{aligned}$$

- Notice there is no division by 2 compared to Jacobi, so GS convergence is 2 times better than Jacobi.
- However, this is only a constant factor, therefore GS is still slow for small h.
- This relationship is typical for SPD systems.

Convergence of Iterative Methods - Jacobi, Gauss-Seidel, & SOR Convergence

For SOR we have that

$$\omega_{opt} = \frac{2}{1 + \sin(\pi h)},$$

and thus

$$\begin{aligned} \rho_{opt} &= \omega_{opt} - 1 \\ &= \frac{1 - \sin(\pi h)}{1 + \sin(\pi h)}, \\ &= 1 - 2\pi h + O(h^2). \end{aligned}$$

Therefore, optimal SOR is <u>much</u> faster than GS/Jacobi/Richardson since the h factor is not squared. For example, with $h = 0.1$ we have

$$\begin{aligned} \rho(G_J) &= 0.95, \\ \rho(G^{GS}) &= 0.9, \\ \rho(G^{SOR}) &= 0.37. \end{aligned}$$

Convergence of Iterative Methods - Jacobi, Gauss-Seidel, & SOR Convergence

Run the demo code `StationaryIterativeConvergence.m` to see a comparison of Jacobi, GS, SOR, and optimal SOR iterations for solving the Laplace equation (i.e., the Poisson equation with $f = 0$). It is apparent from the demo that optimal SOR converges much faster.