Classic Iterative Methods

\( 1 \)

4. Description of the method. The classic iterative methods to solve the matrix equation \( Ax = b \) use a splitting of the matrix \( A \), namely,

\[ A = N - P \]

and use it to generate a sequence \( \{x^n\}_{n=0}^{\infty} \) that should converge to the solution \( x \). The sequence is generated as follows:

\[
\begin{cases}
(1) & \text{Pick } x^0 = x_0. \\
(2) & \text{If } r^k = A x^k - b = 0 \text{ stop, otherwise compute } x^{k+1} \text{ by solving } \\
& \quad N x^{k+1} = P e^k + b.
\end{cases}
\]

Note that if \( e^k = x - x^k \), we get that the error equation is

\[ N e^{k+1} = P e^k \quad k \geq 0. \]

The matrix \( M = N P \) is called the iteration matrix. The convergence of the method \((1)\) depends only on properties of \( M \).

* From: Introduc. à l'analyse numérique et à l'optimisation, J.G. Iaccalet
The classic iterative methods are described in the table below; we use the fact that $D$ is the diagonal part of $A$, $U$ is the upper part and $L$ the lower part.

<table>
<thead>
<tr>
<th>Name</th>
<th>Iteration matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jacobi</td>
<td>$- D^{-1} (A-D)$</td>
</tr>
<tr>
<td>Gauss-Seidel</td>
<td>$- (D+L)^{-1} U)$</td>
</tr>
<tr>
<td>SOR</td>
<td>$(D+L)^{-1} (D-\omega (A-L)) = H_{\text{SOR}} (L)$</td>
</tr>
<tr>
<td>SSOR</td>
<td>$H_{\text{SOR}} (U) H_{\text{SOR}} (L)$</td>
</tr>
</tbody>
</table>

Note that:

1. $H_{\text{SOR}} (L) \big|_{\omega=1} = H_{\text{GS}}$.

2. The SOR can be rewritten as:

$$
\begin{cases}
D \tilde{x}^{k+1} = -L \tilde{x}^{k+1} - U \tilde{x}^k + b \\
\tilde{x}^{k+1} = (1-\omega) \tilde{x}^k + \omega \tilde{x}^{k+1}
\end{cases}
$$

3. $H_{\text{SSOR}}$ is symmetric if $A$ is symmetric.

$H_{\text{SOR}}$ is not symmetric even if $A$ is.
(2) **Strictly diagonal dominant matrices.**

A matrix $A$ is said to be strictly diagonal dominant if

$$p = \max_{1 \leq i \leq N} \sum_{j=1}^{N} \frac{|a_{ij}|}{|a_{ii}|} < 1.$$ 

For these matrices, the error of both the Jacobi and the Gauss-Seidel methods satisfy

$$\|e^n\|_\infty \leq p^n \|e^0\|_\infty.$$

(3) **Symmetric and positive definite matrices.**

For these matrices, we have that

(i) the Jacobi method converges if $a_{ij} \leq 0$ for $i \neq j$.

(ii) the SOR method converges if $\omega \in (0,2)$. It diverges if $\omega \in (\infty,0) \cup (2,\infty)$.

To show this result, we prove the following theorem.
Theorem. Let $A$ be symmetric and positive definite. Suppose that

$$Q = N + N^T - A$$

is positive definite. Then $\rho(A) < 1$.

Proof. Let $\lambda$ be an eigenvalue of $M$ and let $u$ be its eigenvector. Then

$$Mu = \lambda u$$

$$\Rightarrow (I - N^T A) u = \lambda u$$

$$\Rightarrow Au = (1 - \lambda) Nu$$

Since $A$ is positive definite $Au \neq 0$ and hence $\lambda \neq 1$.

Next, let us exploit the fact that $Q$ is positive definite:

$$0 < u^T Q u = u^T (N + N^T - A) u = u^T N u + u^T N^T u - u^T A u = \frac{u^T A u}{1 - \gamma} + \frac{u^T A u}{1 - \xi} - u^T A u$$

$$= \frac{1 - \lambda \xi}{(1 - \gamma)(1 - \xi)} \Rightarrow |\lambda| < 1.$$
Let us prove the claims about the convergence of the Jacobi and SOR methods. We assume that $A$ is real.

Let us begin with the Jacobi method. In this case $N = D$ and

$$Q = N + N^T - A = 2D - A = |A|. $$

Since $A$ is positive definite, so is $|A|$. Hence $Q$ is positive definite.

Let us now consider the SOR method. In this case, we have

$$N = \frac{1}{\omega} (D + \omega L).$$

Hence

$$Q = \frac{1}{\omega} (D + \omega L + D + \omega U) - A$$
$$= \left( \frac{2}{\omega} - 1 \right) D,$$

which is positive definite if $0 < \omega < 2$, as claimed. Now, note that

$$\det M_{SOR} = \det \left( \left( D + \omega L \right)^{-1} \left( D - \omega (D + U) \right) \right)$$
$$= \det \left( (I + \omega DL)^{-1} (Id - \omega (Id + D'U)) \right)$$
$$= (-\omega)^N$$
and so, \( P(H_{SO}) > |1 - \omega| \). This implies that the SOR method diverges if \( \omega \notin [0, 2] \).

Note that

\[
P(H^2_{SSOR}) \leq P(H_{SO}).
\]

and so, the SSOR method converges for \( \omega \in (0, 2) \).

An argument similar to that used for the SOR method shows that for \( \omega \notin [0, 2] \), the SSOR method diverges.

### 4. Symmetric, positive definite matrices that are block-tridiagonal.

For these matrices, we have the following result:

the Jacobi method, the Gauss-Seidel method and the SOR with \( \omega \in (0, 2) \) converge. Moreover, we have:

\[
P(M_{SO}^{(\omega_{opt})}) = \min_{\omega \in \mathbb{R}} P(H_{SO}^{(\omega)})
\]

\[= \omega_{opt} - 1\]

\[< P(H_{GS})\]

\[= \rho^2(M_J) < P(H_J),\]
where
\[ \omega_{opt} = \frac{2}{1 + \sqrt{1 - \rho^2(M_\omega)}} \]

Note that
\[ \rho(M_{SO2}(\omega)) \]

to prove this result, we use a fundamental fact for block-tridiagonal matrices. Consider the matrix
\[
A(\mu) = \begin{bmatrix}
D_1 & \mu U_1 & 0 \\
\mu L_2 & D_2 & \mu U_2 \\
0 & \mu L_3 & D_3 \\
& & & \ddots & \mu U_{N-1} \\
& & & & \mu U_N & D_N
\end{bmatrix}
\]

\[
= D + \mu L + \mu U
\]

then
\[ \det A(\mu) = \det A(0) \]
\( \sigma, \text{ equivalently,} \)

\[
\det (D + \mu L + \vec{\mu} U) = \det (D + L + U).
\]

To prove this result, it is enough to realize that

\[
A(\mu) = \phi(\mu) \ A(1) \ \phi(\mu)^{-1}
\]

where

\[
\phi(\mu) = \begin{bmatrix}
\mu I_1 & 0 \\
0 & \mu I_2 \\
\vdots & \ddots \\
0 & \cdots & \mu I_n
\end{bmatrix}
\]

Now, to prove the convergence result, we proceed as follows. First, let us consider the following polynomial

\[
\phi_j(\lambda) = \det \left( - \frac{1}{\lambda} \left( \frac{1}{\lambda} + L - \lambda I \right) \right).
\]

By definition, the zeros of \( \phi_j \) are the eigenvalues of the iteration matrix for the Jacobi method. Then

\[
\phi_j(\lambda) = \det \left( - \frac{1}{\lambda} \right) \cdot \det \left( \frac{1}{\lambda} + L + \lambda D \right).
\]
Now, let us consider the SOR method:

\[ p_{\text{SOR}}(\lambda) = \det \left( (D + \omega L)^{-1} (1 - \omega) D - U \right) - \lambda \text{Id} \]

\[ = \det \left( \left( \frac{D}{\omega} + L \right)^{-1} \left( \frac{1 - \omega}{\omega} D - U \right) - \lambda \text{Id} \right) \]

\[ = \det \left( \frac{D}{\omega} + L \right)^{-1} \cdot \det \left( \frac{1 - \omega}{\omega} D - U - \lambda L \right) \]

\[ = \det \left( \frac{D}{\omega} + L \right)^{-1} \cdot \det \left( \frac{1 - \omega}{\omega} D - U - \lambda L \right) \]

\[ = \det \left( \frac{D}{\omega} + L \right)^{-1} \cdot \left( -\sqrt{\lambda} \right)^N \cdot \det \left( \frac{\lambda + \omega - 1}{\omega \sqrt{\lambda}} D + \frac{1}{\sqrt{\lambda}} U + \sqrt{\lambda} L \right) \]

by the property (*)

\[ p_{\text{SOR}}(\lambda) = \det \left( \frac{D}{\omega} + L \right)^{-1} \left( -\sqrt{\lambda} \right)^N \cdot \det \left( \frac{\lambda + \omega - 1}{\omega \sqrt{\lambda}} D + \frac{1}{\sqrt{\lambda}} U + \sqrt{\lambda} L \right) \]

and hence

\[ p_{\text{SOR}}(\lambda) = C(\lambda) \cdot p_J \left( \frac{\lambda + \omega - 1}{\omega \sqrt{\lambda}} \right) \]

\[ C(\lambda) = \det \left( \frac{D}{\omega} + L \right)^{-1} \cdot \det \tilde{D}^{\frac{1}{2}} \cdot \lambda^N \]

this means that

\[ \lambda \in \text{spectra of } N_J \]

\[ \iff \lambda : \frac{\lambda + \omega - 1}{\omega \sqrt{\lambda}} = \alpha \in \text{spectra of } H_{\text{SOR}}(\omega)! \]
If \( \omega = 1 \), \( \sqrt{\lambda} = \alpha \) and \( \rho \)

\[
P(M_{\text{SOR}}(\omega=1)) = P(M_{\text{GS}}) = P(M_J).
\]

To prove the remainder of the result, let us show that \( \alpha \) is an eigenvalue of \( M_J \). Indeed, if \( \alpha \) is an eigenvalue of \( M_J \), we get that

\[
-\lambda (U + L) v = \alpha v,
\]

then

\[
0 = (U + L + \lambda \alpha) v = (A + (\alpha - 1) D) v
\]

and no

\[
\overrightarrow{u}^T A \overrightarrow{u} = (1 - \alpha) \overrightarrow{u}^T D \overrightarrow{u},
\]

which proves the claim since \( A \) is symmetric and positive definite. Moreover, \( \alpha \in [0,1) \).

As a consequence, we must study the number

\[
P(M_{\text{SOR}}(\omega)) = \max_{\lambda \in \text{eigenvalues of } M_J} \left\{ |\lambda| : \frac{\lambda + \omega - 1}{1 + \omega} = \alpha \right\}.
\]

assuming that spectrum \( M_J \subset [0,1) \).
the theorem follows from such study.

(5) The model problem.

Let \( \Omega \) be the unit square and let \( U_{ij} \) denote the classic finite difference approximation to the exact solution of

\[
(P) 
\begin{align*}
- \Delta u &= f \ \text{in} \ \Omega ; \quad u = 0 \ \text{on} \ \partial \Omega ,
\end{align*}
\]

given by

\[
\begin{cases}
- \frac{1}{h^2} \left( U_{i+1,j} + U_{i-1,j} + U_{i,j+1} + U_{i,j-1} - 4U_{ij} \right) = f_{ij} \\
U_{ij} = 0 \quad \text{if} \quad i = 0, j = 0, \quad i = N, \quad j = N.
\end{cases}
\]

It is not difficult to show that if \( A \) denotes the matrix associated to \((P)_h\), it is positive definite and symmetric. Moreover,

\[
AU = \lambda U
\]

\[
\iff U = U^{nm}, \quad \lambda = \lambda^{nm}
\]

\[
U_{ij} = \left( \sin \frac{m\pi x_i}{h} \right) \left( \sin \frac{n\pi y_j}{h} \right)
\]

\[
\lambda^{nm} = \frac{4}{h^2} \left( \sin^2 \frac{m\pi h}{2} + \sin^2 \frac{n\pi h}{2} \right)
\]

where \( x_i = ih, \ y_j = jh \).
In this case, we can write
\[ e^k = \sum_{m,n=1}^{N-1} C_{mn} U^{mn}, \]
where \( e^k = x - x^k \) and \( x \) is the \( k \)-th iterate of the Jacobi method. Then, since
\[ e^{k+1} = D^{-1} (U + L) e^k \]
\[ = D^{-1} (D - A) e^k \]
\[ = (I - D^{-1} A) e^k \]
we get that
\[ C_{mn}^{k+1} = (1 - \frac{h^2}{4} \lambda^{mn}) \ e_{mn}^k. \]

This implies that
\[ C_{mn}^{k+1} = \left( 1 - \sin^2 \frac{m \pi h}{2} - \sin^2 \frac{n \pi h}{2} \right) C_{mn}^k. \]

If \( m = n = N/2 \),
\[ C_{mn}^{k+1} = \left( 1 - 2 \sin^2 \frac{N \pi h}{4} \right) C_{mn}^k = 0, \]
and this implies that the components of \( U^{mn} \), \( m,n = N/2 \), are damped extremely fast.
If \( m=n=1 \),

\[
C_{mn}^{k+1} = \left( 1 - 2 \sin^2 \frac{\pi h}{2} \right) C_{mn}^k
\]

\[
= \left( 1 - \frac{\pi h^2}{2} + O(h^4) \right) C_{mn}^k,
\]

and if \( m=n=N-1 \),

\[
C_{mn}^{k+1} = \left( 1 - 2 \sin^2 \frac{(N-1)\pi h}{2} \right) C_{mn}^k
\]

\[
= \left( 1 - 2 \sin^2 \frac{(N-1)\pi h^2}{2} \right) C_{mn}^k
\]

\[
= \left( -1 + \frac{\pi h^2}{2} + O(h^4) \right) C_{mn}^k,
\]

which means that for \( m=n=1 \) and \( m=n=N-1 \),

the corresponding frequencies are barely damped.