M. SOMBRA-2020

X. ITERATIVE METHODS FOR LINEAR EQUATION SOLVING

<u>Iterative methods</u> for solving

$$\underset{n\times n \text{ non singular}}{A}\, x = \underset{n \text{ vector}}{b}$$

are used when <u>direct methods</u> (Gauss elimination, etc) require too much time and/or space.

These methods <u>do not produce exact answers</u> after a finite number of steps, but rather decrease the error by some amount after each step.

X.1 <u>Basic iterative methods</u> [D, § 6.5]

Given an <u>initial vector</u> x_0, these methods generate a <u>sequence</u>

$$(x_\ell)_{\ell\geq 0}$$

hopefully converging to the solution

$$x = A^{-1} b$$

where each $x_{\ell+1}$ is easy to compute from x_ℓ.

A <u>splitting</u> of A is

$$A = \underset{\text{non singular}}{M} - K$$

Gives an iterative method:

$$\underbrace{A}_{M-K} x = b \text{ implies that } \quad Mx = Kx + b$$

and so

$$x = \underbrace{M^{-1}K}_{R}\, x + \underbrace{M^{-1}b}_{c}$$

We then set

$$\boxed{x_{\ell+1} = R x_\ell + c}$$

The <u>spectral radius</u> of R is

$$\rho(R) = \max_{\lambda \text{ eigenvalue of } R} |\lambda|$$

then the iteration converges <u>for every choice</u> of initial value x_0 if and only if

$$\rho(R) < 1$$

Indeed, if $\rho(R) < 1$ then for every $\varepsilon > 0$ there is a vector norm $\|\cdot\|$ s.t. the associated operator norm verifies that

$$\|R\| \leqslant \rho(R) + \varepsilon$$

In particular we can suppose that $\|R\| < 1$.
We have that

$$
\begin{array}{rcl}
x & = & Rx + c \\
- \quad x_{\ell+1} & = & R x_\ell + c \\
\hline
x - x_{\ell+1} & = & R(x - x_\ell)
\end{array}
$$

and so

$$(*) \qquad \|x - x_{\ell+1}\| = \|R(x - x_\ell)\| \leq \|R\|\cdot\|x - x_\ell\| \leq \underbrace{\|R\|^{\ell+1}}_{\substack{\downarrow\, \ell\to\infty \\ 0}}\cdot\|x - x_0\|$$

The "only if" part is easy to verify:
if $\rho(R) \geq 1$ there is $x_0 \neq x$ st

$x - x_0$ eigenvector of λ with $|\lambda| \geq 1$

Hence

$$x - x_{\ell+1} = R^{\ell+1}(x - x_0) = \lambda^{\ell+1}(x - x_0) \underset{\ell\to\infty}{\not\to} 0$$

From $(*)$

$$\underbrace{-\log_b \|x - x_{\ell+1}\|}_{\text{precision of } x_{\ell+1}} \geq (\ell+1)\cdot(-\log_b \rho(R)) - \log_b \|x - x_0\|$$

The increase of precision is <u>linear</u> with rate

$$-\log_b \rho(R)$$

The smaller $\rho(R)$ the higher the rate of convergence

Our goal is to find $A = M - K$ s.t.

1. $R = M^{-1}K$ and $c = M^{-1}b$ <u>easy to compute</u>
2. $\rho(R)$ <u>small</u>

Suppose that A has no zero entries in the diagonal
Set

$$A = \underbrace{D}_{\text{diagonal}} - \underbrace{\tilde{L}}_{\text{lower triangular}} - \underbrace{\tilde{U}}_{\text{upper triangular}} = D(\mathbb{1}_n - L - U)$$

The <u>Jacobi's method</u> can be interpreted as successively through the equation, changing the variable x_j so that the j-th equation is satisfied:

for $j = 1, \dots, n$

$$x_{l+1,j} \leftarrow \frac{1}{a_{jj}} \left(b_j - \sum_{k \neq j} a_{j,k} x_{l,k} \right)$$

so that

$$a_{j1} x_{l1} + \cdots + \underline{a_{jj} x_{l+1,j}} + \cdots + a_{jn} x_{ln} = b_j$$

In matrix notation

$$R_J = D^{-1}(\tilde{L} + \tilde{U}) = L + U \qquad c_J = D^{-1} b$$

$$\boxed{x_{l+1} = R_J \cdot x_l + c}$$

The <u>Gauss-Seidel method</u> takes advantage of the previously computed $x_{l+1,k}$ for $k = 1, \dots, j-1$:

for $j = 1, \dots, n$

$$x_{l+1,j} \leftarrow \frac{1}{a_{jj}} \left(b_j - \underbrace{\sum_{k=1}^{j-1} a_{j,k} x_{l+1,k}}_{\text{updated } x\text{'s}} - \underbrace{\sum_{k=j+1}^{n} a_{j,k} x_{l,k}}_{\text{older } x\text{'s}} \right)$$

so that

$$a_{j,1} x_{\ell+1,1} + \cdots + a_{j,j} x_{\ell+1,j} + a_{j,j+1} x_{\ell,j+1} + \cdots + a_{j,n} x_{\ell,n} = b_j$$

In matrix notation

$$R_{GS} = (D - \tilde{L})^{-1} \tilde{U} = (\mathbb{1}_n - L)^{-1} U$$

$$c_{GS} = (D - \tilde{L})^{-1} b = (\mathbb{1}_n - L)^{-1} D^{-1} b$$

so that

$$x_{\ell+1} = R_{GS} x_\ell + c_{GS}$$

<u>Successive overrelaxation</u> with parameter $\omega \in \mathbb{R}$ (SOR(ω)) is a weighted average of $x_{\ell+1}$ and x_ℓ from Gauss-Seidel:

$$\underset{\uparrow\,\mathrm{SOR}(\omega)}{x_{\ell+1}} = (1-\omega) x_\ell + \omega \underset{\uparrow\,\mathrm{GS}}{x_{\ell+1}}$$

for $j=1,\dots,n$

$$x_{\ell+1,j} \leftarrow (1-\omega) x_{\ell j} + \frac{\omega}{a_{jj}} \left(b_j - \sum_{k=1}^{j-1} a_{jk} x_{\ell+1,k} - \sum_{k=j+1}^{n} a_{jk} x_{\ell,k} \right)$$

That is

$$(D - \omega \tilde{L}) x_{\ell+1} = ((1-\omega) D + \omega \tilde{U}) x_\ell + \omega b$$

and so

$$x_{\ell+1} = \overbrace{(D - \omega \tilde{L})^{-1} ((1-\omega) D + \omega \tilde{U})}^{R_{SOR(\omega)}} x_\ell + \overbrace{\omega (D - \omega \tilde{L})^{-1} b}^{c_{SOR(\omega)}}$$

$$= (\mathbb{1}_n - \omega L)^{-1} ((1-\omega) \mathbb{1} + \omega U) x_\ell + \omega (\mathbb{1}_n - \omega L)^{-1} D^{-1} b$$

When $\omega=1$ it coincides with Gauss - Seidel

When $\omega<1$ is called underrelaxation

$\omega>1$ ——— overrelaxation

<u>Example</u> $n=2$ $A=\begin{pmatrix}2 & 1\\ 1 & 2\end{pmatrix}$ $b=\begin{pmatrix}1\\ -1\end{pmatrix}$

Then

$$A=\begin{pmatrix}2 & \\ & 2\end{pmatrix}-\begin{pmatrix} & \\ -1 & \end{pmatrix}-\begin{pmatrix} & -1\\ & \end{pmatrix}=D-\tilde{L}-\tilde{U}$$

$$=\begin{pmatrix}2 & \\ & 2\end{pmatrix}\left(\mathbb{1}_2-\begin{pmatrix} & \\ -1/2 & \end{pmatrix}-\begin{pmatrix} & -1/2\\ & \end{pmatrix}\right)=D(\mathbb{1}_2-L-U)$$

We have that

$$R_J=\begin{pmatrix} & -1/2\\ -1/2 & \end{pmatrix}=L+U \qquad c_J=D^{-1}b=\begin{pmatrix}1/2\\ -1/2\end{pmatrix}$$

and so Jacobi writes as

$$\begin{pmatrix}x_{\ell+1,1}\\ x_{\ell+1,2}\end{pmatrix}=\begin{pmatrix} & -1/2\\ -1/2 & \end{pmatrix}\begin{pmatrix}x_{\ell 1}\\ x_{\ell 2}\end{pmatrix}+\begin{pmatrix}1/2\\ -1/2\end{pmatrix}=\begin{pmatrix}-x_{\ell 2}/2+\frac{1}{2}\\ -x_{\ell 1}/2-1/2\end{pmatrix}$$

To compute the spectral radius

$$\chi_{R_J}(t)=\det(R_J-t\mathbb{1}_2)=\det\begin{pmatrix}-t & -1/2\\ -1/2 & -t\end{pmatrix}=t^2-1/4$$

The spectrum of R_J is

$$\lambda(R_J)=(\chi_{R_J}=0)=\{\pm 1/2\}$$

and so $\boxed{\rho(R_J)=1/2}$: the method <u>converges</u> for every choice of x_0:

$$x_\ell\to x=\begin{pmatrix}1\\ -1\end{pmatrix} \quad \text{for } \ell\to\infty$$

Also

$$R_{GS} = (\mathbb{1}_2 - L)^{-1} U = \begin{pmatrix} 1 & \\ 1/2 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 0 & -1/2 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & \\ -1/2 & 1 \end{pmatrix} \begin{pmatrix} & -1/2 \\ & \end{pmatrix} = \begin{pmatrix} 0 & -1/2 \\ 0 & 1/4 \end{pmatrix}$$

$$c_{GS} = (\cancel{D} - \tilde{E})\, b = \begin{pmatrix} 1/2 & \\ -1/4 & 1/2 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1/2 \\ -3/4 \end{pmatrix}$$

and so Gauss-Seidel writes as

$$\begin{pmatrix} x_{l+1,1} \\ x_{l+1,2} \end{pmatrix} = \begin{pmatrix} 0 & -1/2 \\ 0 & 1/4 \end{pmatrix} \begin{pmatrix} x_{l,1} \\ x_{l,2} \end{pmatrix} + \begin{pmatrix} 1/2 \\ -3/4 \end{pmatrix} = \begin{pmatrix} -\frac{x_{l,2}}{2} + 1/2 \\ \frac{x_{l,2}}{4} - \frac{3}{4} \end{pmatrix}$$

We have that

$$\chi_{R_{GS}}(t) = \det \begin{pmatrix} -t & 1/2 \\ 0 & -t + \frac{1}{4} \end{pmatrix} = t^2 - \frac{t}{4}$$

And so $\lambda(R_{GS}) = \{0, \frac{1}{4}\}$ and $\boxed{\rho(R_{GS}) = \frac{1}{4}}$

In this case

$$\rho(R_{GS}) = \rho(R_J)^2$$

GS converges with the double of the speed of convergence of Jacobi.

Now let $\omega \in \mathbb{R}$. Then

$$R_{SOR(\omega)} = (\mathbb{1}_2 - \omega L)^{-1}((1-\omega)\mathbb{1}_2 + \omega U) = \begin{pmatrix} 1 & \\ -\omega/2 & 1 \end{pmatrix} \begin{pmatrix} 1-\omega & -\omega/2 \\ & 1-\omega \end{pmatrix}$$

$$= \begin{pmatrix} 1-\omega & -\omega/2 \\ -\frac{\omega}{2}(1-\omega) & \frac{\omega^2}{4} + 1 - \omega \end{pmatrix} = \begin{pmatrix} 1-\omega & -\omega/2 \\ \frac{\omega^2}{2} - \frac{\omega}{2} & \frac{\omega^2}{4} - \omega + 1 \end{pmatrix}$$

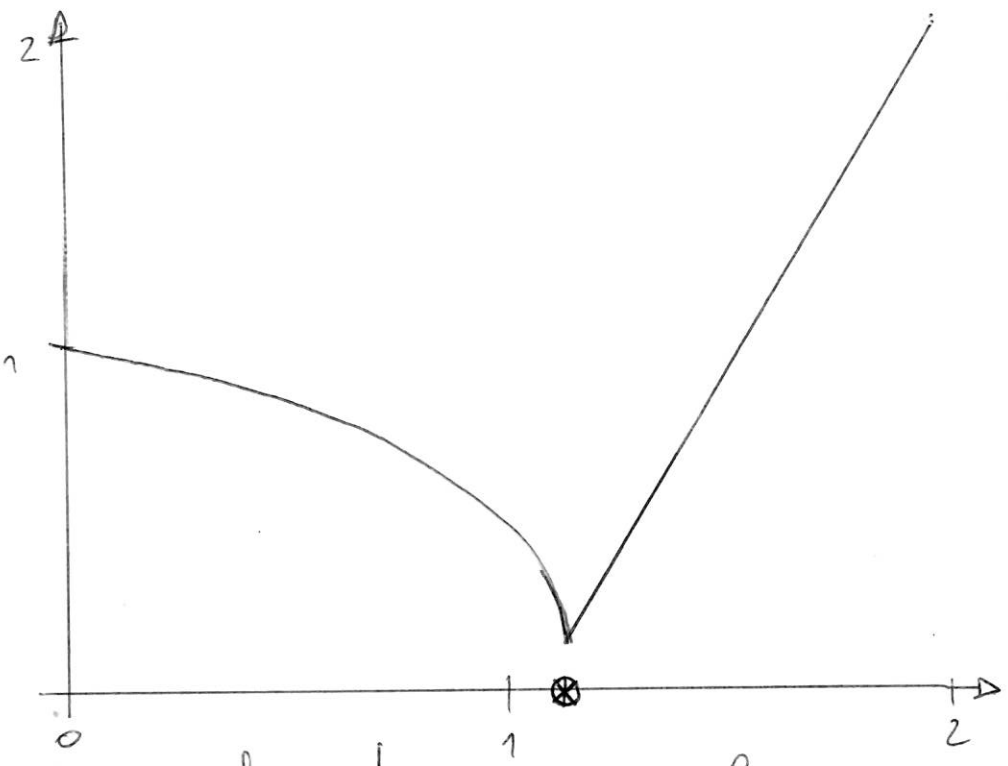
$$c_{SOR(\omega)} = \omega (D - \omega \tilde{E})^{-1} b = \omega \begin{pmatrix} 2 & \\ \omega/2 & 2 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} \omega/2 \\ -\frac{\omega^2}{4} - \frac{\omega}{2} \end{pmatrix}$$

We have that

$$\chi_{R_{SOR(\omega)}}(t) = \det(t\mathbb{1}_2 - R_{SOR(\omega)}) = \det\begin{pmatrix} 1-\omega-t & -\omega/2 \\ \frac{\omega^2}{2}-\frac{\omega}{2} & \frac{\omega^2}{4}-\omega+1-t \end{pmatrix}$$

$$= t^2 + \left(\frac{-\omega^2}{4}+2\omega-2\right)t + (\omega^2-2\omega+1)$$

The spectral radius of $R_{SOR(\omega)}$ for $\omega \in [0,2]$ is



The minimal value appears for $\omega \approx 1{,}0717$.
The corresponding eigenvalues are

$$\lambda = -0.1339\ ,\ -0.1535$$

and so

$$\underline{\rho(R_{SOR(1.0717)}) = 0.1535}$$

X.2 Convergence of the basic iterative schemes

[S] Saad, Iterative methods for sparse linear systems, 1996

[S §4.2.1 and §4.2.2]

Let $A = M - K$ splitting

M non singular

and $x_{l+1} \leftarrow \underbrace{R}_{M^{-1}K} \cdot x_l + \underbrace{c}_{M^{-1}b}$ the associated iteration

If this iteration converges, its limit $x_\infty = \lim_{l \to \infty} x_l$ satisfies

$$x_\infty = R\, x_\infty + c = M^{-1} K\, x_\infty + M^{-1} b$$

which implies that $x_\infty = x$ the solution of

$$Ax = b$$

The convergence of the iteration for an arbitrary initial value x_0 is equivalent to

$$\rho(R) < 1$$

and the rate of convergence is

$$-\log_b \rho(R)$$

b base of floating point system

Since the spectral radius is difficult to compute, sufficient conditions that guarantee convergence can be useful in practice.

Example The Richardson iteration is defined by

$$x_{\ell+1} = (\mathbb{1}_n - \omega A) x_\ell + \omega b$$

for $\omega > 0$: it corresponds to the splitting

$$A = \omega^{-1} \mathbb{1}_n - (\omega^{-1} \mathbb{1}_n - A)$$

The iteration matrix is $R_\omega = \mathbb{1}_n - \omega A$. Suppose that the eigenvalues of A are real and satisfy

$$\lambda_1 \geq \cdots \geq \lambda_n$$

Then the eigenvalues of R_ω are also real and satisfy

$$\mu_1 = 1 - \omega \lambda_1 \leq \cdots \leq \mu_n = 1 - \omega \lambda_n$$

If $\lambda_n \leq 0$ then $\mu_n \geq 1$ and so

$$\rho(R_\omega) \geq 1$$

Else suppose that $\lambda_n > 0$. Then the iteration converges if and only if

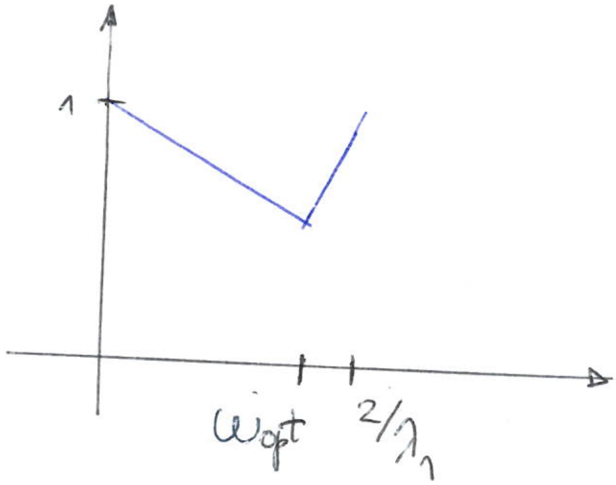
$$-1 \leq 1 - \omega \lambda_1 \quad \text{and} \quad 1 - \omega \lambda_n \leq 1$$

that is, if and only if

$$\boxed{0 < \omega < \frac{2}{\lambda_1}}$$

In this case

$$\rho(R_\omega) = \max(|1 - \omega \lambda_n|, |1 - \omega \lambda_1|)$$



The optimal value satisfies

$$-1+\lambda_1 \omega_{opt} = 1-\lambda_n \omega_{opt}$$

and so

$$\boxed{\omega_{opt} = \frac{2}{\lambda_1+\lambda_n}}$$

in which case, the spectral radius is

$$\boxed{\rho(R_{\omega_{opt}}) = \frac{\lambda_1-\lambda_n}{\lambda_1+\lambda_n}}$$

If the matrix A has very large and very small eigenvalues, the iteration will be slow. Moreover, the determination of ω_{opt} requires knowledge of these eigenvalues.

A matrix A is diagonally dominant if for all j

$$|a_{jj}| > \sum_{i\neq j} a_{ij}$$

For A diagonally dominant, the Jacobi and the Gauss-Seidel iterations converge for any x_0.
Indeed, let

$$R_J = \tilde{D}^{-1}(\tilde{L} + \tilde{U}) \quad \text{and} \quad R_{GS} = (D - \tilde{L})^{-1}\tilde{U}$$

the corresponding iteration matrices.
For an eigenvalue

$$\lambda \in \lambda(R_J)$$

let x be an eigenvector, and m the index of the largest component. We can suppose that

$$x_m = 1 \quad \text{and} \quad |x_j| \leq j \quad \forall j$$

Then from $R_J x = \lambda x$ we deduce that

$$\lambda x_m = - \sum_{j \neq m} \frac{a_{m,j}}{a_{m,m}} x_j$$

and so

$$|\lambda| \leq \sum_{j \neq m} \frac{|a_{m,j}|}{|a_{m,m}|} |x_j| < 1$$

which proves the result for Jacobi's method.
Now let $\lambda \in R_{GS}$ and x an eigenvector for λ.
Let m be the index of the largest component of x
From $R_{GS} \cdot x = \lambda x$ we deduce that

$$\tilde{U}x = \lambda(D-\tilde{L})x$$

and so

$$\sum_{j<m} a_{m,j} x_j = \lambda\left(a_{m,m} x_m + \sum_{j>m} a_{m,j} x_j\right)$$

This implies that

$$|\lambda| \leq \frac{\sum_{j<m} |a_{m,j}||x_j|}{|a_{m,m}| - \sum_{j>m} |a_{m,j}||x_j|}$$

$$\leq \frac{\sum_{j<m} |a_{m,j}|}{|a_{m,m}| - \sum_{j>m} |a_{m,j}|} < 1$$

because $|a_{m}| - \sum_{j>m} |a_{m,j}| > \sum_{j<m} |a_{m,j}|$, which shows the result for the Gauss-Seidel method.

For successive overrelaxation, the condition

$$0 < \omega < 2$$

is necessary for the convergence. Indeed

$$R_{SOR(\omega)} = (\mathbb{1}_n - \omega L)^{-1}((1-\omega)\mathbb{1}_n + \omega U)$$

and so

$$\chi_{R_{SOR(\omega)}}(t) = \det(R_{SOR(\omega)} - t\mathbb{1}_n)$$

Hence

$$|\chi(0)| = \prod_{\lambda} |\lambda| \geq \rho(R_{SOR(\omega)})^n$$

product over the eigenvalues of $R_{SOR(\omega)}$

and

$$\chi(0) = \det(R_{SOR(\omega)}) = (1-\omega)^n$$

which implies that

$$\boxed{\rho(R_{SOR(\omega)}) \geq |1-\omega|}$$

When A is symmetric and positive definite, $SOR(\omega)$ converges when $0 < \omega < 2$.

In particular Gauss-Seidel $(\omega = 1)$ converges