Numerical Methods and Optimization

Chapter 3: Systems of linear Equations - Iterative Methods
3.2 The Jacobi and Gauss-Seidel method

28.11.2022

The Jacobi method

- simplest method to solve SLE $Ax = b$ is: "Jacobi-Method"
- necessary: $a_{ii} \neq 0 \quad \forall i$
- get algorithm by solving the i^{th} equation for the i^{th} unknown:

$$x_i^{(k+1)} = \frac{1}{a_{ii}}\left(b_i - \sum_{j \neq i} a_{ij} x_j^{(k)}\right) \tag{1}$$

JACOBI METHOD

initialization: given $Ax = b$ with $a_{ii} \neq 0 \; \forall i$. Choose any $x^{(0)} \in \mathbb{R}^n$.

```
for k = 0, 1, 2, ... do
  for i = 1, ..., n do
      x_i^(k+1) = 1/a_ii (b_i - Σ_{j≠i} a_ij x_j^(k))
  end
until stop
```

$$x_i^{(k+1)} = \frac{1}{a_{ii}}\left(b_i - \sum_{j \neq i} a_{ij} x_j^{(k)}\right)$$

- #mult/div: n^2 (per iteration)

- prove convergence using the BFPT:
- for this, decompose

$$A = D - L - U$$

  D: diagonal;
  L or U: "strict" lower/upper triangular
- iterationeqn. $x_i^{(k+1)} = \frac{1}{a_{ii}}\big(b_i - \sum_{j\neq i} a_{ij}x_j^{(k)}\big)$ can be rewritten:

$$\begin{aligned} x^{(k+1)} &= D^{-1}(b + (L+U)x^{(k)}) \\ &= \underbrace{D^{-1}(L+U)}_{=M^{-1}N} x^{(k)} + \underbrace{D^{-1}b}_{=M^{-1}b} \\ &= Tx^{(k)} + c \\ &= \mathcal{J}(x^{(k)}) + c, \end{aligned}$$

  where $\mathcal{J} \coloneqq T = D^{-1}(L+U)$: **Jacobi iteration matrix**

Ex: Compute first 4 Jacobi iterates for $Ax = b$ with $A = \begin{pmatrix} 5 & 1 & 1 \\ 1 & 5 & 0 \\ 1 & 0 & 5 \end{pmatrix}$, $b = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$.

We obtain the iteration rule

$$x_1^{(k+1)} = \frac{1}{5}(1 - x_2^{(k)} - x_3^{(k)})$$
$$x_2^{(k+1)} = \frac{1}{5}(2 - x_1^{(k)})$$
$$x_3^{(k+1)} = \frac{1}{5}(0 - x_1^{(k)})$$

Using this we get

k	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$
0	0.0000	0.0000	0.0000
1	0.2000	0.4000	0.0000
2	0.1200	0.3600	−0.0400
3	0.1360	0.3760	−0.0240
4	0.1296	0.3728	−0.0272

- exact solution rounded to 4 dec. places: $x = \begin{pmatrix} 0.1304 \\ 0.3739 \\ -0.0261 \end{pmatrix}$
- $\|x - x^{(4)}\|_2 = 1.75 \cdot 10^{-3}$
- Jacobi iteration matrix $\mathcal{J}$ is calc. from $A = M - N = D - (L + U)$

  with $D = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{pmatrix}$, $L = \begin{pmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}$, $U = \begin{pmatrix} 0 & -1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$

  to $\mathcal{J} = M^{-1}N = D^{-1}(L + U) = \frac{1}{5}\begin{pmatrix} 0 & -1 & -1 \\ -1 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}$
- easy: $\sigma(\mathcal{J}) = \{0, \frac{1}{5}\sqrt{2}, -\frac{1}{5}\sqrt{2}\}$.
- Convergence follows since $\rho(\mathcal{J}) < 1$.

The Gauss-Seidel method (m. of successive displacement)

Gauss-Seidel method (for $n = 3$):

$$
\begin{aligned}
x_1^{(k+1)} &= \frac{1}{a_{11}}\left(b_1 \qquad\qquad\qquad\quad - a_{12}x_2^{(k)} - a_{13}x_3^{(k)}\right)\\
x_2^{(k+1)} &= \frac{1}{a_{22}}\left(b_2 - a_{21}x_1^{(k+1)} \qquad\qquad\qquad - a_{23}x_3^{(k)}\right)\\
x_3^{(k+1)} &= \frac{1}{a_{33}}\left(b_3 - a_{31}x_1^{(k+1)} - a_{32}x_2^{(k+1)} \qquad\qquad\quad\right)
\end{aligned}
$$

GAUSS-SEIDEL (GS) METHOD

initialization: given $Ax = b$ with $a_{ii} \neq 0 \ \forall i$. Choose any $x^{(0)} \in \mathbb{R}^n$.

```
for k = 0, 1, 2, ... do
  for i = 1, ..., n do
```

$$x_i^{(k+1)} = \frac{1}{a_{ii}}\left(b_i - \sum_{j<i} a_{ij}x_j^{(k+1)} - \sum_{j>i} a_{ij}x_j^{(k)}\right)$$

```
  end
until stop
```

Using $A = D - L - U$ we can write

$$x_i^{(k+1)} = \frac{1}{a_{ii}}\big(b_i - \sum_{j<i} a_{ij}x_j^{(k+1)} - \sum_{j>i} a_{ij}x_j^{k}\big)$$

in matrix form:

$$\begin{aligned} x^{(k+1)} &= D^{-1}\big(b + Lx^{(k+1)} + Ux^{(k)}\big) \\ \Leftrightarrow \quad \underbrace{(D-L)}_{=:M\in \mathrm{Gl}} x^{(k+1)} &= \underbrace{U}_{=:N} x^{(k)} + b \\ \Leftrightarrow \quad x^{(k+1)} &= M^{-1}Ux^{(k)} + M^{-1}b \end{aligned}$$

- operator $\mathcal{L} := (D-L)^{-1}U$ is called **Gauss-Seidel matrix**
- we have $A = (D-L) - U = M - N$
- and $x^{(k+1)} = \mathcal{L}x^{(k)} + c$

Ex. (3.7)**:** Solve $Ax = b$ using GS method; $A = \begin{pmatrix} 5 & 1 & 1 \\ 1 & 5 & 0 \\ 1 & 0 & 5 \end{pmatrix}, b = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$

Calc. $\mathcal{L} = (D-L)^{-1}U$ for previous Ex.:

$$D - L = \begin{pmatrix} 5 & 0 & 0 \\ 1 & 5 & 0 \\ 1 & 0 & 5 \end{pmatrix} = 5\begin{pmatrix} 1 & 0 & 0 \\ 1/5 & 1 & 0 \\ 1/5 & 0 & 1 \end{pmatrix};\ U = \begin{pmatrix} 0 & -1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Thus

$$(D-L)^{-1} = \frac{1}{5}\begin{pmatrix} 1 & 0 & 0 \\ -1/5 & 1 & 0 \\ -1/5 & 0 & 1 \end{pmatrix}$$

and

$$\mathcal{L} = (D-L)^{-1}U = \frac{1}{5}\begin{pmatrix} 0 & -1 & -1 \\ 0 & \frac{1}{5} & \frac{1}{5} \\ 0 & \frac{1}{5} & \frac{1}{5} \end{pmatrix}.$$

Finally, one easily verifies that

$$\sigma(\mathcal{L}) = \{0, \frac{2}{25}\} = \sigma(\mathcal{J})^2.$$

i.e.

$$\rho(\mathcal{L}) = \rho(\mathcal{J})^2.$$

$$\rho(\mathcal{L}) = \rho(\mathcal{J})^2 \tag{2}$$

- according to the prop. concerning the asympt. conv. rate, the GS-method should converge faster than the J-method
- (2) not accidental: this is true for all matrices of certain form (later)
- calc. first 4 iterates of GS-method with $x^{(0)} = 0$:

k	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$
0	0.0000	0.0000	0.0000
1	0.2000	0.4000	0.0000
2	0.1200	0.3600	−0.0400
3	0.1360	0.3760	−0.0240
4	0.1296	0.3728	−0.0272

- $\|x - x_{GS}^{(4)}\|_2 = 3.70{\cdot}10^{-5} < \|x - x_J^{(4)}\|_2 = 1.75{\cdot}10^{-3}$
- $\rightsquigarrow$ almost 2 decimal places better //.

Now we employ Prop. 3.3 ($\|M^{-1}N\|_* < 1 \;\Rightarrow$ convergence) for a compatible matrix norm $\|\cdot\|_*$.

Definition

A matrix $A = (a_{ij})_{i,j=1,\dots,n} \in \mathbb{R}^{n\times n}$ is called **strictly diagonally dominant** *if the following inequality holds for all $i = 1, \dots, n$*

$$\sum_{\substack{j=1\\ j\neq i}}^{n} |a_{ij}| < |a_{ii}|.$$

e.g. $\begin{pmatrix} -4 & 1 & -1 \\ 4 & 8 & 3 \\ 1 & 2 & -4 \end{pmatrix}$ is diagonally dominant.

Proposition

If $A \in \mathbb{R}^{n\times n}$ is strictly diagonally dominant, then the Jacobi and Gauss-Seidel method both converge for each initial guess $x^{(0)} \in \mathbb{R}^n$ to the unique solution of $Ax = b$.

PROOF:

- due to the st. diag. dom. it follows $a_{ii} \neq 0 \; \forall i \Rightarrow$ J-/GS-method are well-defined.
- Jacobi method: Taking into account that the i^{th} row of $D^{-1}(L+U)$ equals to

$$-\frac{1}{a_{ii}} \begin{pmatrix} a_{i,1} & \cdots & a_{i,i-1} & 0 & a_{i,i+1} & \cdots & a_{i,n} \end{pmatrix}$$

  it follows from the str. diag. dom.

$$\|\mathcal{J}\|_\infty = \|D^{-1}(L+U)\|_\infty = \max_{i=1,\dots,n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{|a_{ij}|}{|a_{ii}|} =: q < 1 \qquad (3)$$

- Thus $\|\mathcal{J}\|_* = \|M^{-1}N\|_* < 1$ is satisfied for the row sum norm
- GS method: Proof slightly more difficult
- We want to shoe that

$$\|\mathcal{L}\|_\infty := \max_{\|x\|_\infty = 1} \|\mathcal{L}x\|_\infty \leq q(< 1).$$

- Let $\|x\|_\infty = 1$, and q be defined as in (3)

To estimate $\|\mathcal{L}\|_\infty \coloneqq \max_{\|x\|_\infty=1}\|\mathcal{L}x\|_\infty < 1$, est. components y_i of

$$y \coloneqq \mathcal{L}x = (D-L)^{-1}Ux$$

$$\begin{aligned}
\Leftrightarrow \quad (D-L)y &= Ux \\
\Leftrightarrow \quad y &= D^{-1}(Ly + Ux) \\
\Leftrightarrow \quad y_i &= \frac{1}{a_{ii}}\Big(-\sum_{j<i} a_{ij}y_j - \sum_{j>i} a_{ij}x_j\Big) \qquad \forall i
\end{aligned}$$

Initial case $i=1$: We have

$$\begin{aligned}
|y_1| &= \Big| -\frac{1}{a_{11}}\sum_{j=2}^n a_{1j}x_j \Big| \\
&\le \frac{1}{|a_{11}|}\sum_{j=2}^n |a_{1j}|\|x\|_\infty \\
&= \frac{1}{|a_{11}|}\sum_{j=2}^n |a_{1j}| \\
&\le q.
\end{aligned}$$

$i-1 \to i$:
Assume $|y_k| \le q, \forall 0 \le k \le i-1$.

$$\begin{aligned}
|y_i| &\le \frac{1}{|a_{ii}|}\Big(\sum_{j<i}|a_{ij}||y_j| + \sum_{j>i}|a_{ij}||x_j|\Big) \\
&\le \frac{1}{|a_{ii}|}\Big(\sum_{j<i}|a_{ij}|q + \sum_{j>i}|a_{ij}|\|x\|_\infty\Big) \\
&\le \frac{1}{|a_{ii}|}\Big(\sum_{j<i}|a_{ij}| + \sum_{j>i}|a_{ij}|\Big) \\
&\le q. \qquad \blacksquare
\end{aligned}$$

Ex: Given: SLE $Ax = b$, where $A = \begin{pmatrix} 2 & 0 & 1 \\ 1 & -4 & 1 \\ 0 & -1 & 2 \end{pmatrix}, b = \begin{pmatrix} 1 \\ 4 \\ -1 \end{pmatrix}$. Exakt sln: $x = (1, -1, -1)^T$. A is str. diag. dom. with

$$\max_{i=1,\dots,n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{|a_{ij}|}{|a_{ii}|} = \frac{1}{2} =: q.$$

Thus J and GS method converge as $\|\mathcal{J}\|_\infty = q < 1$ and $\|\mathcal{L}\|_\infty \leq q < 1$. Since for $T = \mathcal{J}$ or $T = \mathcal{L}$ we have $x^{(k+1)} = Tx^{(k)} + c = \Phi(x^{(k)}) \Rightarrow$

$$\begin{aligned} \|x^{(k+1)} - \hat{x}\|_\infty &= \|\Phi(x^{(k)}) - \Phi(\hat{x})\|_\infty \\ &\leq \|Tx^{(k)} - T\hat{x}\|_\infty \\ &\leq \|T\|_\infty \|x^{(k)} - \hat{x}\|_\infty \\ &\leq q\|x^{(k)} - \hat{x}\|_\infty \end{aligned}$$

$\rightsquigarrow$ error correction at least with a factor $q = \frac{1}{2}$ wrt the ∞-norm.

Recalling $x = (1,-1,-1)^T$, we first get for $x^{(0)} = (1,1,1)^T$ that $\|x^{(0)} - \hat{x}\|_\infty = 2$ and after one iteration for the

(a) Jacobi method: Recall: $A = \begin{pmatrix} 2 & 0 & 1 \\ 1 & -4 & 1 \\ 0 & -1 & 2 \end{pmatrix}, b = \begin{pmatrix} 1 \\ 4 \\ -1 \end{pmatrix}$. We obtain

$$x_{\mathcal{J}}^{(1)} = \begin{pmatrix} \frac{1}{a_{11}}(b_1 - a_{12}x_2^{(0)} - a_{13}x_3^{(0)}) \\ \frac{1}{a_{22}}(b_2 - a_{21}x_1^{(0)} - a_{23}x_3^{(0)}) \\ \frac{1}{a_{33}}(b_3 - a_{31}x_1^{(0)} - a_{32}x_2^{(0)}) \end{pmatrix} = \begin{pmatrix} \frac{1}{2}(1-1) \\ -\frac{1}{4}(4-1-1) \\ \frac{1}{2}(-1+1) \end{pmatrix} = \begin{pmatrix} 0 \\ -\frac{1}{2} \\ 0 \end{pmatrix},$$

i.e. $\|x_{\mathcal{J}}^{(1)} - \hat{x}\|_\infty = 1$.

(b) Gauss-Seidel method:

$$x_{\mathcal{L}}^{(1)} = \begin{pmatrix} \frac{1}{a_{11}}(b_1 - a_{12}x_2^{(0)} - a_{13}x_3^{(0)}) \\ \frac{1}{a_{22}}(b_2 - a_{21}x_1^{(1)} - a_{23}x_3^{(0)}) \\ \frac{1}{a_{33}}(b_3 - a_{31}x_1^{(1)} - a_{32}x_2^{(1)}) \end{pmatrix} = \begin{pmatrix} \frac{1}{2}(1-1) \\ -\frac{1}{4}(4-0-1) \\ \frac{1}{2}(-1+0-\frac{3}{4}) \end{pmatrix} = \begin{pmatrix} 0 \\ -\frac{3}{4} \\ -\frac{7}{8} \end{pmatrix},$$

$$\Rightarrow \|x_{\mathcal{L}}^{(1)} - \hat{x}\|_\infty = 1$$

- Thus, wrt to the max-norm the error is reduced by the factor q.
- However, regarding each individual component, the GS method is somewhat better:
- $\|x_{\mathcal{J}}^{(1)} - \hat{x}\|_2 = \frac{3}{2}$ and $\|x_{\mathcal{L}}^{(1)} - \hat{x}\|_2 < 1.0384$

- although there are exa. for which the J method is superior (cf.ex.8.2)
- Gauss-Seidel often converges faster
- can be made more precise for special mat.: e.g. consider for $p, q \neq 0$

$$A = \begin{pmatrix} E_q & -B^T \\ -B & E_p \end{pmatrix} \in \mathbb{R}^{n\times n}, \tag{4}$$

where $B \in \mathbb{R}^{p\times q}$. Here, it is

$$L = \begin{pmatrix} 0 & 0 \\ B & 0 \end{pmatrix},\ U = \begin{pmatrix} 0 & B^T \\ 0 & 0 \end{pmatrix}.$$

For $\mathcal{J}$ and $\mathcal{L}$ we obtain:

$$\mathcal{L} = (D-L)^{-1}U = \begin{pmatrix} E & 0 \\ -B & E \end{pmatrix}^{-1} \begin{pmatrix} 0 & B^T \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} E & 0 \\ B & E \end{pmatrix} \begin{pmatrix} 0 & B^T \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & B^T \\ 0 & BB^T \end{pmatrix} \tag{5}$$

or

$$\mathcal{J} = E^{-1}(L+U) = \begin{pmatrix} 0 & B^T \\ B & 0 \end{pmatrix}$$

hence

$$\mathcal{J}^2 = \begin{pmatrix} B^TB & 0 \\ 0 & BB^T \end{pmatrix}.$$

Goal now: show that $\rho(\mathcal{L}) = \rho(\mathcal{J}^2) = \rho(\mathcal{J})^2$. Saw this in already in Example 3.7.

Lemma

Given $X \in \mathbb{R}^{p\times q}$, $Y \in \mathbb{R}^{q\times p}$ *and* $Z \in \mathbb{R}^{n\times n}$ *with* $p, q, n \in \mathbb{N}$. *Then*

(a) $\sigma(XY)\backslash\{0\} = \sigma(YX)\backslash\{0\}$

(b) $\sigma(Z^2) = \{\lambda^2 \mid \lambda \in \sigma(Z)\} =: \sigma(Z)^2$

PROOF: **Ad (a):**

- If $\lambda \in \sigma(XY)\backslash\{0\} \Rightarrow \exists u \neq 0$ s.t. $XYu = \lambda u \neq 0 \Rightarrow v := Yu \neq 0$
  Thus
$$YXv = Y(XYu) = Y(\lambda u) = \lambda Yu = \lambda v.$$

Consequently, we obtain $\lambda \in \sigma(YX)\backslash\{0\}$ and therefore $\sigma(YX)\backslash\{0\} \subset \sigma(XY)\backslash\{0\}$. The reverse direction is analogously true.
Ad (b): If $\lambda \in \sigma(Z) \Rightarrow \exists x \neq 0: \ Zx = \lambda x$ and from this we deduce
$$Z^2x = Z(\lambda x) = \lambda Zx = \lambda^2 x.$$

Thus, it is $\lambda^2 \in \sigma(Z^2)$. Conversely, if $\mu \in \sigma(Z^2)$ and if $\pm\lambda$ are the (possibly complex) roots of μ, then it follows that

$$0 = \det(Z^2 - \mu E) = \det\big((Z-\lambda E)(Z+\lambda E)\big) = \det(Z-\lambda E)\det(Z+\lambda E).$$

From this we can conclude that $\lambda \in \sigma(Z)$ or $-\lambda \in \sigma(Z)$ with $\mu = (\pm\lambda)^2 \Rightarrow \mu \in \{\lambda^2 \mid \lambda \in \sigma(Z)\}$. ∎

Proposition

If A has the form (4), i.e. $A = \begin{pmatrix} E_q & -B^T \\ -B & E_p \end{pmatrix}$, then $\rho(\mathcal{L}) = \rho(\mathcal{J})^2$.

PROOF: Recall that $\mathcal{L} = \begin{pmatrix} 0 & B^T \\ 0 & BB^T \end{pmatrix}$, $B^T \in \mathbb{R}^{q\times p} \Rightarrow BB^T \in \mathbb{R}^{p\times p}$.

$$\det(\mathcal{L} - \lambda E) = \det\begin{pmatrix} -\lambda E_q & B^T \\ 0 & BB^T - \lambda E_p \end{pmatrix} = (-\lambda)^q \det(BB^T - \lambda E_p).$$

So $\sigma(\mathcal{L}) = \{0\} \cup \sigma(BB^T)$ and thus $\rho(\mathcal{L}) = \rho(BB^T)$.

From the representation $\mathcal{J}^2 = \begin{pmatrix} B^T B & 0 \\ 0 & BB^T \end{pmatrix}$ we conclude

$$\begin{aligned} \sigma(\mathcal{J}^2) &= \{\lambda \mid \det\begin{pmatrix} B^T B - \lambda E & 0 \\ 0 & BB^T - \lambda E \end{pmatrix} = 0\} \\ &= \{\lambda \mid \det(B^T B - \lambda E) \cdot \det(BB^T - \lambda E) = 0\} \\ &= \sigma(B^T B) \cup \sigma(BB^T) \end{aligned}$$

and further using 3.11 (a), i.e. $\sigma(XY)\backslash\{0\} = \sigma(YX)\backslash\{0\}$,

$$\sigma(\mathcal{J}^2) \cup \{0\} = \big(\sigma(B^T B) \cup \{0\}\big) \cup \big(\sigma(BB^T) \cup \{0\}\big) = \sigma(BB^T) \cup \{0\}$$

Employing 3.11 (b) we finally obtain

$$\rho(\mathcal{J})^2 \overset{(b)}{=} \rho(\mathcal{J}^2) = \rho(BB^T) = \rho(\mathcal{L}),$$

completing the proof.