Numerieke technieken en optimalisatie

Prof. Dr. Jochen Schütz, Manuela Bastidas Olivares

Any questions? Do not hesitate to contact us!

Jochen:	C265	Tel 8315	jochen.schuetz@uhasselt.be
Manuela:	C101A	Tel 8296	manuela.bastidas@uhasselt.be

All computation may be done with Matlab.

Exercise 3.1: (Jacobi, Gauss-Seidel method)

We consider iterative methods based on matrix splittings for the solution of

$$Ax = b. \tag{1}$$

For a matrix splitting $A = M - N$, these schemes are given as

$$Mx^{k+1} = Nx^k + b.$$

Jacobi's method is obtained by setting $M = D$, and Gauss-Seidel's method is obtained by setting $M = D - L$, where D are the diagonals of A and L is the strictly lower left half of the negative of A.

a) Write a programm (in MATLAB) that can perform both Jacobi's and Gauss-Seidel's method. Draw a convergence plot based on your program (putting k on x-axis and $|x^k - x|$ on the y-axis, use the Matlab command 'semilogy'). How fast do these two methods converge?

Answer.
```
function [x,err, vel] = Jacobi_mat(A,b,tol)
D = diag(diag(A));
U = -triu(A)+D;
L = -tril(A)+D;

Cj= inv(D)*(L+U);
bj= inv(D)*b;

sp_j = max(abs(eig(Cj)));
vel  = abs(log(sp_j));
if  sp_j >=1
error('the method does not converge')
end

%% Jacobi's method
x  = zeros(size(A,1),1);
err(1) = norm(A*x-b);
i=1;
while err(i) > tol
x = Cj*x + bj;
```

```
i = i+1;
err(i)= norm(Cj^i*err(1));
end
end

function [x,err,vel] = Gauss_mat(A,b,tol)
D = diag(diag(A));
U = -triu(A)+D;
L = -tril(A)+D;

Cg= inv(D-L)*U;
bg= inv(D-L)*b;

sp_g = max(abs(eig(Cg)));
vel  = abs(log(sp_g));
if sp_g>=1
error('the method does not converge')
end

%% Gauss Method
x = zeros(size(A,1),1);
err(1) = norm(A*x-b);
i=1;
while err(i) > tol
x = Cg*x + bg;
i = i+1;
err(i)= norm(Cg^i*err(1));
end
end
```

b) Consider Jacobi's method and show that this method, applied to the matrix

$$A = \begin{pmatrix} 2 & 4 \\ 1.5 & 3 \end{pmatrix},$$

does in general not converge.

Answer. The Jacobi's iteration matrix is

$$C_j = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{3} \end{pmatrix} \begin{pmatrix} 0 & -4 \\ -1.5 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -2 \\ -0.5 & 0 \end{pmatrix}$$

then

$$\det(C_j - \lambda I) = (-\lambda)^2 + 1 = 0 \iff \lambda = \pm 1 \tag{2}$$

we can conclude: the eigenvalues $\lambda = \pm 1$ then $\rho(C_j) = 1$ and (Theorem 3.) the Jacobi's method does not converges.

Exercise 3.2

Consider the matrix

$$A = \begin{pmatrix} 4 & 3 \\ 2 & 5 \end{pmatrix}.$$

a) Show that both Jacobi's method and Gauss-Seidel's method converge for all possible right-hand sides b and start values x_0.

Answer. - Jacobi: The matrix is

$$C_J = \begin{pmatrix} 1/4 & 0 \\ 1/5 & 0 \end{pmatrix} \begin{pmatrix} 0 & -3 \\ -2 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -3/4 \\ -2/5 & 0 \end{pmatrix}$$

and $\rho(C_J) = \max(|0.54|, |-0.54|) = 0.547 < 1$.

- Gauss-Seidel: The matrix is

$$C_G = \begin{pmatrix} 4 & 0 \\ 2 & 5 \end{pmatrix}^{-1} \begin{pmatrix} 0 & -3 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1/4 & 0 \\ -1/10 & 1/5 \end{pmatrix} \begin{pmatrix} 0 & -3 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -3/4 \\ 0 & -3/10 \end{pmatrix}$$

and $\rho(C_G) = \max(0, |-3/10|) = 3/10 < 1$.

b) Assume your goal is to diminish the initial error by a factor of 10^{-10}, i.e., you seek for a k such that

$$\|x^{k+1} - x\|_2 \leq 10^{-10} \|x^0 - x\|_2.$$

Give for both Gauss-Seidel and Jacobi an estimate for k. Compare to your numerical results. When is this estimate exact?

Answer.

$$\|x^{k+1} - x\|_2 \leq 10^{-10} \|x^0 - x\|_2 \rightarrow \|e^{k+1}\|_2 \leq 10^{-10} \|e^0\|_2$$

Using Lemma 8. we have

$$\max\left(\frac{\|e^{k+1}\|_2}{\|e^0\|_2}\right) \leq 10^{-10} \iff ||C^k|| \leq 10^{-10} \iff ||C^k||^{1/k} \leq (10^{-10})^{1/k}$$

If $k \to \infty$ then $||C^k||^{1/k} \approx \rho(C)$. For this we need k such that $\rho(C) \leq \left(\frac{1}{10^10}\right)^{1/k}$.

- Jacobi:

$$\begin{aligned} 0.541 &\approx \left(\frac{1}{10^{10/k}}\right) \\ \log(0.541) &\approx \frac{1}{k} \log\left(\frac{1}{10^{10}}\right) \\ k &\approx \frac{\log(10^{10})}{-\log(0.541)} = 38.24 \rightarrow k \geq 39 \end{aligned}$$

- Gauss-Seidel:

$$\begin{aligned} 3/10 &\approx \left(\frac{1}{10^{10/k}}\right) \\ k &\approx \frac{\log(10^{10})}{-\log(3/10)} = 19.12 \rightarrow k \geq 20 \end{aligned}$$

Exercise 3.3

Let A be such that

$$|A_{ii}| > \sum_{j=1, j\neq i}^{n} |A_{ij}| \quad \forall i = 1, \dots n.$$

Show that Jacobi's method converges for all possible $x_0, b \in \mathbb{R}^n$.

Answer.

The Jacobi's iteration matrix is $C_j = D^{-1}(L+U)$

$$\begin{pmatrix} 0 & \frac{a_{12}}{a_{11}} & \dots & \frac{a_{1n}}{a_{11}} \\ \frac{a_{21}}{a_{22}} & 0 & \dots & \frac{a_{2n}}{a_{22}} \\ \vdots & & & \vdots \\ \frac{a_{n1}}{a_{nn}} & \dots & \dots & 0 \end{pmatrix}$$

Then

$$\|C_j\|_\infty = \max_i \left(\sum_{j\neq i} \frac{|a_{ij}|}{|a_{ii}|} \right) < 1$$

The jacobi method converges for all $x_0, b \in \mathbb{R}^n$.

Exercise 3.4: (Steepest-descent method)

Let $F : \mathbb{R}^n \to \mathbb{R}$ be defined as

$$F(x) := \frac{1}{2}(x, Ax) - (x, b).$$

We define $\overline{x}$ to be a solution of

$$A\overline{x} = b,$$

where $A \in \mathbb{R}^{n\times n}$ is a symmetric positive definite matrix, and $b \in \mathbb{R}^n$.

a) Show that

$$\overline{x} = \operatorname*{argmin}_{x\in\mathbb{R}^n} F(x).$$

Answer. Using the definition of the inner products we have

$$\begin{aligned} F(x) := \frac{1}{2}(x, Ax) - (x, b) &= \frac{1}{2}x^T Ax - b^T x \\ &= \frac{1}{2}\sum_{i,j=1}^{n} a_{ij}x_i x_j - \sum_{i=1}^{n} b_i x_i \end{aligned}$$

Furthermore

$$\nabla F(x) = \begin{pmatrix} \frac{\partial F(x)}{\partial x_1} \\ \vdots \\ \frac{\partial F(x)}{\partial x_n} \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^n a_{1j}x_j - b_1 \\ \vdots \\ \sum_{j=1}^n a_{nj}x_j - b_n \end{pmatrix}$$
$$= Ax - b$$

Then the minimum will be such that $\nabla F(x) = 0 \iff Ax = b \iff x = \bar{x}$.
We can conclude that $F(x)$ reached the minimum at $x = \bar{x}$.

b) Prove that F, in the neighborhood of a point x_k, (locally) decreases fastest in the direction of $r_k := F'(x_k)^T$. What is $F'(x_k)$ in this case?

Answer. As we already showed $\nabla F(x_k) = Ax_k - b := r_k$. Consider $y := x + tv$ with $t \in \mathbb{R}$ and v a directional vector.
We will show that $F(y_k) \leq F(x_k)$ for certain t_k and v_k and that the best choice of v_k is r_k.

$$\begin{aligned} F(x+tv) &= \frac{1}{2}(x+tv, Ax+tAv) - (x+tv, b) \\ &= \frac{1}{2}(x, Ax) + t(x, Av) + \frac{t^2}{2}(v, Av) - (x, b) - t(v, b) \\ &= \frac{1}{2}(x, Ax) + \frac{t^2}{2}(v, Av) + t\left[(x, Av) - (v, b)\right] \\ &= F(x) + \frac{t^2}{2}(v, Av) + t(Ax - b, v) \longrightarrow \text{ Symmetry } A \end{aligned}$$

Then

$$\frac{\partial F(x+tv)}{\partial t} = t(v, Av) + (Ax - b, v)$$

and

$$\frac{\partial F(x+tv)}{\partial t} = 0 \iff \bar{t} = \frac{(Ax-b, v)}{(v, Av)}$$

In conclusion

- The minimum of $F(x_k + tv_k)$ is located in

$$\bar{t} = \frac{(Ax_k - b, v_k)}{(v_k, Av_k)} = \frac{-(r_k, v_k)}{(v_k, Av_k)} \tag{3}$$

- Evaluate $F(x+tv)$ at $\bar{t}$:

$$\begin{aligned} F(x+\bar{t}v) &= F(x) + \bar{t}\left[(Ax-b, v) + \frac{\bar{t}}{2}(v, Av)\right] \\ &= F(x) + \frac{1}{2}\frac{(Ax-b, v)}{(v, Av)} \end{aligned}$$

then

$$F(x_k + \bar{t}v_k) = F(x_k) + \frac{1}{2}\frac{(Ax_k - b, v_k)}{(v_k, Av_k)}$$
$$= F(x_k) - \frac{1}{2}\frac{(r_k, v_k)}{(v_k, Av_k)}$$

If we choose v_k s.t (r_k, v_k) is maximum, the value of $F(y_k)$ will decrease fastest. This direction is $v_k = r_k$.

c) Show that the function $g(t) := F(x_k + t \cdot r_k)$ has its minimum at $t_k := -\frac{r_k^T r_k}{r_k^T A r_k}$.

Answer. Using (3) we know that the minimum of $F(x_k + tr_k)$ is located in

$$t_k := \bar{t} = \frac{(Ax_k - b, r_k)}{(r_k, Ar_k)} = \frac{-(r_k, r_k)}{(r_k, Ar_k)} = \frac{-r_k^T r_k}{r_k^T A r_k}$$

d) The steepest descent algorithm is defined in the following way: Choose an $x_0 \in \mathbb{R}^n$ and define x_{k+1} recursively as $x_{k+1} := x_k + t_k \cdot r_k$. Show that

$$F(x_l) \leq F(x_k) \quad \forall k < l.$$

Is this a linear iterative method?

Answer.

- We will prove that $F(x_{k+1}) \leq F(x_k)$ for all $k \geq 1$.

  By definition $x_{k+1} = x_k + t_k r_k$ then we have

$$F(x_{k+1}) = F(x_k + t_k r_k) := g(t_k)$$
$$F(x_k) = F(x_k + 0r_k) := g(0)$$

  But we already showed that the minimum is located in t_k (see (c)) then $g(t_k)$ is the minimum and $g(t_k) \leq g(t)$ for all t.

$$F(x_{k+1}) = F(x_k + t_k r_k) := g(t_k) \leq g(0) := F(x_k + 0r_k) = F(x_k)$$

- This is a linear iterative method if and only if exist $C \in \mathbb{R}^{n \times n}$ such that

$$\begin{aligned} x_{k+1} - \bar{x} = C(x_k - \bar{x}) &\iff x_k + t_k r_k - \bar{x} = C(x_k - \bar{x}) \\ &\iff (x_k - \bar{x}) + t_k(Ax_k - b) = C(x_k - \bar{x}) \\ &\iff (x_k - \bar{x}) + t_k(Ax_k - A\bar{x}) = C(x_k - \bar{x}) \\ &\iff (I_n + t_k A)(x_k - \bar{x}) = C(x_k - \bar{x}) \end{aligned}$$

  and this implies that $I_n + t_k A$ should be constant. We can conclude that the method is not a linear iterative method.