Scientific Computing WS 2019/2020

Lecture 9

Jürgen Fuhrmann

juergen.fuhrmann@wias-berlin.de

Iterative methods: Recap

Elements of iterative methods (Saad Ch.4)

Let $V = \mathbb{R}^n$ be equipped with the inner product $(\cdot, \cdot)$. Let A be an $n \times n$ nonsingular matrix.

Solve $Au = b$ iteratively. For this purpose, two components are needed:

- **Preconditioner**: a matrix $M \approx A$ "approximating" the matrix A but with the property that the system $Mv = f$ is easy to solve
- **Iteration scheme**: algorithmic sequence using M and A which updates the solution step by step

Simple iteration with preconditioning

Idea: $A\hat{u} = b \Rightarrow$

$$\hat{u} = \hat{u} - M^{-1}(A\hat{u} - b)$$

$\Rightarrow$ iterative scheme

$$u_{k+1} = u_k - M^{-1}(Au_k - b) \quad (k = 0, 1 \dots)$$

1. Choose initial value u_0, tolerance ε, set $k = 0$
2. Calculate *residuum* $r_k = Au_k - b$
3. Test convergence: if $||r_k|| < \varepsilon$ set $u = u_k$, finish
4. Calculate *update*: solve $Mv_k = r_k$
5. Update solution: $u_{k+1} = u_k - v_k$, set $k = k + 1$, repeat with step 2.

The Jacobi method

- Let $A = D - E - F$, where D: main diagonal, E: negative lower triangular part F: negative upper triangular part
- Preconditioner: $M = D$, where D is the main diagonal of $A \Rightarrow$

$$u_{k+1,i} = u_{k,i} - \frac{1}{a_{ii}}\left(\sum_{j=1\dots n} a_{ij}u_{k,j} - b_i\right) \quad (i = 1 \dots n)$$

- Equivalent to the succesive (row by row) solution of

$$a_{ii}u_{k+1,i} + \sum_{j=1\dots n, j\neq i} a_{ij}u_{k,j} = b_i \quad (i = 1 \dots n)$$

- Already calculated results not taken into account
- Alternative formulation with $A = M - N$:

$$\begin{aligned} u_{k+1} &= D^{-1}(E+F)u_k + D^{-1}b \\ &= M^{-1}Nu_k + M^{-1}b \end{aligned}$$

- Variable ordering does not matter

The Gauss-Seidel method

- Solve for main diagonal element row by row
- Take already calculated results into account

$$a_{ii}u_{k+1,i} + \sum_{j<i} a_{ij}u_{k+1,j} + \sum_{j>i} a_{ij}u_{k,j} = b_i \qquad (i = 1 \dots n)$$

$$(D - E)u_{k+1} - Fu_k = b$$

- May be it is faster
- Variable order probably matters
- Preconditioners: forward $M = D - E$, backward: $M = D - F$
- Splitting formulation: $A = M - N$
  forward: $N = F$, backward: $M = E$
- Forward case:

$$\begin{aligned} u_{k+1} &= (D - E)^{-1}Fu_k + (D - E)^{-1}b \\ &= M^{-1}Nu_k + M^{-1}b \end{aligned}$$

Convergence

- Let $\hat{u}$ be the solution of $Au = b$.
- Let $e_k = u_k - \hat{u}$ be the error of the k-th iteration step

$$
\begin{aligned}
u_{k+1} &= u_k - M^{-1}(Au_k - b) \\
&= (I - M^{-1}A)u_k + M^{-1}b \\
u_{k+1} - \hat{u} &= u_k - \hat{u} - M^{-1}(Au_k - A\hat{u}) \\
&= (I - M^{-1}A)(u_k - \hat{u}) \\
&= (I - M^{-1}A)^k(u_0 - \hat{u})
\end{aligned}
$$

  resulting in

$$e_{k+1} = (I - M^{-1}A)^k e_0$$

- So when does $(I - M^{-1}A)^k$ converge to zero for $k \to \infty$?
- Let $B = I - M^{-1}A$

Back to iterative methods

Sufficient condition for convergence: $\rho(I - M^{-1}A) < 1$.

Matrix preconditioned Richardson iteration

M, A spd.

- Scaled Richardson iteration with preconditoner M

$$u_{k+1} = u_k - \alpha M^{-1}(Au_k - b)$$

- Spectral equivalence estimate

$$0 < \gamma_{min}(Mu, u) \leq (Au, u) \leq \gamma_{max}(Mu, u)$$

- $\Rightarrow \gamma_{min} \leq \lambda_i \leq \gamma_{max}$
- $\Rightarrow$ optimal parameter $\alpha = \frac{2}{\gamma_{max}+\gamma_{min}}$
- Convergence rate with optimal parameter: $\rho \leq \frac{\kappa(M^{-1}A)-1}{\kappa(M^{-1}A)+1}$
- This is one possible way for convergence analysis which at once gives convergence rates
- But ... how to obtain a good spectral estimate for a particular problem ?

1D heat conduction: spectral bounds estimate

- For $i = 1 \dots n$, the argument of cos is in $(0, \pi)$
- cos is monotonically decreasing in $(0, \pi)$, so we get λ_{max} for $i = 1$ and λ_{min} for $i = n = \frac{1+h}{h}$
- Therefore:

$$\lambda_{max} = \frac{2}{h}\left(1 + \cos\left(\pi\frac{h}{1+2h}\right)\right) \approx \frac{2}{h}\left(2 - \frac{\pi^2 h^2}{2(1+2h)^2}\right)$$

$$\lambda_{min} = \frac{2}{h}\left(1 + \cos\left(\pi\frac{1+h}{1+2h}\right)\right) \approx \frac{2}{h}\left(\frac{\pi^2 h^2}{2(1+2h)^2}\right)$$

Here, we used the Taylor expansion

$$cos(\delta) = 1 - \frac{\delta^2}{2} + O(\delta^4) \quad (\delta \to 0)$$

$$cos(\pi - \delta) = -1 + \frac{\delta^2}{2} + O(\delta^4) \quad (\delta \to 0)$$

and $\frac{1+h}{1+2h} = \frac{1+2h}{1+2h} - \frac{h}{1+2h} = 1 - \frac{h}{1+2h}$

Jacobi preconditioned Richardson for 1D heat conduction

- The Jacobi preconditioner just multiplies by $\frac{h}{2}$, therefore for $M^{-1}A$:

$$\mu_{max} \approx 2 - \frac{\pi^2 h^2}{2(1+2h)^2}$$
$$\mu_{min} \approx \frac{\pi^2 h^2}{2(1+2h)^2}$$

- Optimal parameter: $\alpha = \frac{2}{\lambda_{max}+\lambda_{min}} \approx 1 \ (h \to 0)$
- Good news: this is independent of h resp. n
- No need for spectral estimate in order to work with optimal parameter.
- Is this true beyond this special case ?

Eigenvalue analysis for more general matrices

- For 1D heat conduction we had a very special regular structure of the matrix which allowed exact eigenvalue calculations
- We need a generalization to varying coefficients, nonsymmetric problems, unstructured grids . . .
  $\Rightarrow$ what can be done for general matrices ?

The Gershgorin Circle Theorem (Semyon Gershgorin,1931)

(everywhere, we assume $n \geq 2$)

Theorem (Varga, Th. 1.11) Let A be an $n \times n$ (real or complex) matrix. Let Λ_i be the sum of the absolute values of the i-th rowoff-diagonal entries:

$$\Lambda_i = \sum_{\substack{j=1\dots n \\ j \neq i}} |a_{ij}|$$

If λ is an eigenvalue of A, then there exists r, $1 \leq r \leq n$ such that λ lies on the disk defined by the circle of radius Λ_r around a_{rr}:

$$|\lambda - a_{rr}| \leq \Lambda_r.$$

Gershgorin Circle Theorem, Proof

Proof: Assume λ is an eigenvalue, $\mathbf{x} = (x_1 \dots x_n)$ is a corresponding eigenvector. Assume $\mathbf{x}$ is normalized such that

$$\max_{i=1\dots n} |x_i| = |x_r| = 1.$$

From $A\mathbf{x} = \lambda\mathbf{x}$ it follows that

$$\begin{aligned}
\lambda x_i &= \sum_{j=1\dots n} a_{ij} x_j \\
(\lambda - a_{ii}) x_i &= \sum_{\substack{j=1\dots n \\ j\neq i}} a_{ij} x_j \\
|\lambda - a_{rr}| &= \Big| \sum_{\substack{j=1\dots n \\ j\neq r}} a_{rj} x_j \Big| \leq \sum_{\substack{j=1\dots n \\ j\neq r}} |a_{rj}||x_j| \leq \sum_{\substack{j=1\dots n \\ j\neq r}} |a_{rj}| = \Lambda_r
\end{aligned}$$

□

Gershgorin Circle Corollaries

Corollary: Any eigenvalue of A lies in the union of the disks defined by the Gershgorin circles

$$\lambda \in \bigcup_{i=1\dots n} \{\mu \in \mathbb{V} : |\mu - a_{ii}| \leq \Lambda_i\}$$

Corollary: The Gershgorin circle theorem allows to estimate the spectral radius $\rho(A)$:

$$\rho(A) \leq \max_{i=1\dots n} \sum_{j=1}^{n} |a_{ij}| = ||A||_\infty,$$

$$\rho(A) \leq \max_{j=1\dots n} \sum_{i=1}^{n} |a_{ij}| = ||A||_1.$$

Proof

$$|\mu - a_{ii}| \leq \Lambda_i \quad \Rightarrow \quad |\mu| \leq \Lambda_i + |a_{ii}| = \sum_{j=1}^{n} |a_{ij}|$$

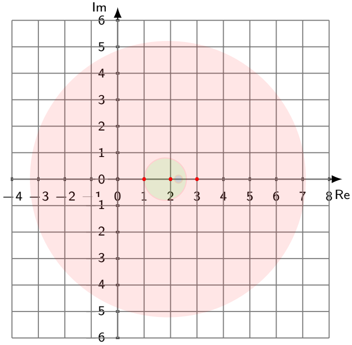
Furthermore, $\sigma(A) = \sigma(A^T)$. □

Gershgorin circles: example

$$A = \begin{pmatrix} 1.9 & 1.8 & 3.4 \\ 0.4 & 1.8 & 0.4 \\ 0.05 & 0.1 & 2.3 \end{pmatrix}$$

$$\lambda_1 = 1, \lambda_2 = 2, \lambda_3 = 3$$

$$\Lambda_1 = 5.2, \Lambda_2 = 0.8, \lambda_3 = 0.15$$



Gershgorin circles: heat example I

$$A = \begin{pmatrix} \frac{2}{h} & -\frac{1}{h} & & & & & \\ -\frac{1}{h} & \frac{2}{h} & -\frac{1}{h} & & & & \\ & -\frac{1}{h} & \frac{2}{h} & -\frac{1}{h} & & & \\ & \ddots & \ddots & \ddots & \ddots & & \\ & & -\frac{1}{h} & \frac{2}{h} & -\frac{1}{h} & & \\ & & & -\frac{1}{h} & \frac{2}{h} & -\frac{1}{h} \\ & & & & -\frac{1}{h} & \frac{2}{h} \end{pmatrix}$$

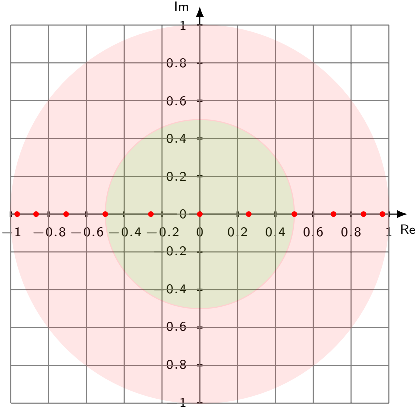
$$B = (I - D^{-1}A) = \begin{pmatrix} 0 & \frac{1}{2} & & & & \\ \frac{1}{2} & 0 & \frac{1}{2} & & & \\ & \frac{1}{2} & 0 & \frac{1}{2} & & \\ & \ddots & \ddots & \ddots & \ddots & \\ & & \frac{1}{2} & 0 & \frac{1}{2} & \\ & & & \frac{1}{2} & 0 & \frac{1}{2} \\ & & & & \frac{1}{2} & 0 \end{pmatrix}$$

We have $b_{ii} = 0$, $\Lambda_i = \begin{cases} \frac{1}{2}, & i = 1, n \\ 1 & i = 2 \dots n-1 \end{cases} \Rightarrow$ estimate $|\lambda_i| \leq 1$

Gershgorin circles: heat example II

Let n=11, h=0.1:

$$\lambda_i = \cos\left(\frac{ih\pi}{1+2h}\right) \quad (i = 1 \ldots n)$$



⇒ the Gershgorin circle theorem is too pessimistic, we need a better theory . . .

Permutation matrices

- Permutation matrices are matrices which have exactly one non-zero entry in each row and each column which has value 1.
- There is a one-to-one correspondence permutations π of the the numbers $1 \dots n$ and $n \times n$ permutation matrices $P = (p_{ij})$ such that

$$p_{ij} = \begin{cases} 1, & \pi(i) = j \\ 0, & \text{else} \end{cases}$$

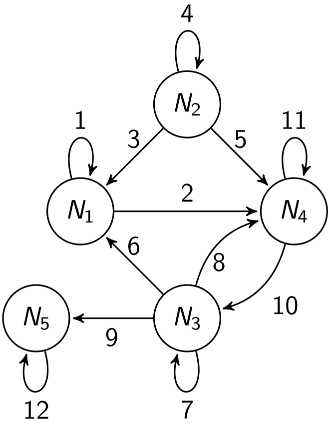
- Permutation matrices are orthogonal, and we have $P^{-1} = P^T$
- $A \to PA$ permutes the rows of A
- $A \to AP^T$ permutes the columns of A

Weighted directed graph representation of matrices

Define a directed graph from the nonzero entries of a matrix $A = (a_{ik})$:

- Nodes: $\mathcal{N} = \{N_i\}_{i=1\dots n}$
- Directed edges: $\mathcal{E} = \{\overrightarrow{N_k N_l} | a_{kl} \neq 0\}$
- Matrix entries $\equiv$ weights of directed edges

$$A = \begin{pmatrix} 1. & 0. & 0. & 2. & 0. \\ 3. & 4. & 0. & 5. & 0. \\ 6. & 0. & 7. & 8. & 9. \\ 0. & 0. & 10. & 11. & 0. \\ 0. & 0. & 0. & 0. & 12. \end{pmatrix}$$



- 1:1 equivalence between matrices and weighted directed graphs
- Convenient e.g. for sparse matrices

Reducible and irreducible matrices

Definition A is *reducible* if there exists a permutation matrix P such that

$$PAP^T = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}$$

A is *irreducible* if it is not reducible.

Theorem (Varga, Th. 1.17): A is irreducible $\Leftrightarrow$ the matrix graph is connected, i.e. for each *ordered* pair (N_i, N_j) there is a path consisting of directed edges, connecting them.

Equivalently, for each i, j there is a sequence of consecutive nonzero matrix entries $a_{ik_1}, a_{k_1k_2}, a_{k_2k_3} \ldots, a_{k_{r-1}k_r} a_{k_rj}$.

□

Taussky theorem (Olga Taussky, 1948)

Theorem (Varga, Th. 1.18) Let A be irreducible. Assume that the eigenvalue λ is a boundary point of the union of all the disks

$$\lambda \in \partial \bigcup_{i=1\dots n} \{\mu \in \mathbb{C} : |\mu - a_{ii}| \leq \Lambda_i\}$$

Then, all n Gershgorin circles pass through λ, i.e. for $i = 1 \dots n$,

$$|\lambda - a_{ii}| = \Lambda_i$$

Taussky theorem proof

Proof Assume λ is eigenvalue, $\mathbf{x}$ a corresponding eigenvector, normalized such that $\max_{i=1\dots n}|x_i| = |x_r| = 1$. From $A\mathbf{x} = \lambda\mathbf{x}$ it follows that

$$(\lambda - a_{rr})x_r = \sum_{\substack{j=1\dots n\\ j\neq r}} a_{rj}x_j \tag{1}$$

$$|\lambda - a_{rr}| \leq \sum_{\substack{j=1\dots n\\ j\neq r}} |a_{rj}|\cdot|x_j| \leq \sum_{\substack{j=1\dots n\\ j\neq r}} |a_{rj}| = \Lambda_r \tag{2}$$

λ is boundary point $\Rightarrow |\lambda - a_{rr}| = \sum_{\substack{j=1\dots n\\ j\neq r}} |a_{rj}|\cdot|x_j| = \Lambda_r$

$\Rightarrow$ For all $p \neq r$ with $a_{rp} \neq 0$, $|x_p| = 1$.

Due to irreducibility there is at least one p with $a_{rp} \neq 0$. For this p, $|x_p| = 1$ and equation (2) is valid (with p in place of r) $\Rightarrow |\lambda - a_{pp}| = \Lambda_p$

Due to irreducibility, this is true for all $p = 1\dots n$. $\square$

Consequences for heat example from Taussky theorem

- $B = I - D^{-1}A$
- We had $b_{ii} = 0$, $\Lambda_i = \begin{cases} \frac{1}{2}, & i = 1, n \\ 1 & i = 2 \dots n-1 \end{cases} \Rightarrow$ estimate $|\lambda_i| \leq 1$
- Assume $|\lambda_i| = 1$. Then λ_i lies on the boundary of the union of the Gershgorin circles. But then it must lie on the boundary of both circles with radius $\frac{1}{2}$ and 1 around 0.
- Contradiction $\Rightarrow |\lambda_i| < 1$, $\rho(B) < 1$!

Diagonally dominant matrices

Definition Let $A = (a_{ij})$ be an $n \times n$ matrix.

- A is *diagonally dominant* if
  (i) for $i = 1 \dots n$, $|a_{ii}| \geq \sum_{\substack{j=1\dots n \\ j \neq i}} |a_{ij}|$
- A is *strictly diagonally dominant* (sdd) if
  (i) for $i = 1 \dots n$, $|a_{ii}| > \sum_{\substack{j=1\dots n \\ j \neq i}} |a_{ij}|$
- A is *irreducibly diagonally dominant* (idd) if
  (i) A is irreducible
  (ii) A is diagonally dominant –
  for $i = 1 \dots n$, $|a_{ii}| \geq \sum_{\substack{j=1\dots n \\ j \neq i}} |a_{ij}|$
  (iii) for at least one r, $1 \leq r \leq n$, $|a_{rr}| > \sum_{\substack{j=1\dots n \\ j \neq r}} |a_{rj}|$

A very practical nonsingularity criterion

Theorem (Varga, Th. 1.21): Let A be strictly diagonally dominant or irreducibly diagonally dominant. Then A is nonsingular.

If in addition, $a_{ii} > 0$ is real for $i = 1 \dots n$, then all real parts of the eigenvalues of A are positive:

$$\mathrm{Re}\lambda_i > 0, \quad i = 1 \dots n$$

A very practical nonsingularity criterion, proof I

Proof:

- Assume A strictly diagonally dominant. Then the union of the Gershgorin disks does not contain 0 and $\lambda = 0$ cannot be an eigenvalue $\Rightarrow$ A is nonsingular.
- As for the real parts, the union of the disks is

$$\bigcup_{i=1\dots n} \{\mu \in \mathbb{C} : |\mu - a_{ii}| \leq \Lambda_i\}$$

and $\mathrm{Re}\mu$ must be larger than zero if μ should be contained.

A very practical nonsingularity criterion, proof I

- Assume A irreducibly diagonally dominant. Then, if 0 is an eigenvalue, it sits on the boundary of one of the Gershgorin disks.

  By Taussky theorem, we have $|a_{ii}| = \Lambda_i$ for all $i = 1 \dots n$.

  This is a contradiction as by definition there is at least one i such that $|a_{ii}| > \Lambda_i$

- Assume $a_{ii} > 0$, real. All real parts of the eigenvalues must be ≥ 0.

  Therefore, if a real part is 0, it lies on the boundary of at least one disk.

  By Taussky theorem it must be contained at the same time in the boundary of all the disks and in the imaginary axis.

  This contradicts the fact that there is at least one disk which does not touch the imaginary axis as by definition there is at least one i such that $|a_{ii}| > \Lambda_i$ □

Corollary

Theorem: If A is complex hermitian or real symmetric, sdd or idd, with positive diagonal entries, it is positive definite.

Proof: All eigenvalues of A are real, and due to the nonsingularity criterion, they must be positive, so A is positive definite.

□

Heat conduction matrix

$$A = \begin{pmatrix} \alpha + \frac{1}{h} & -\frac{1}{h} & & & & & \\ -\frac{1}{h} & \frac{2}{h} & -\frac{1}{h} & & & & \\ & -\frac{1}{h} & \frac{2}{h} & -\frac{1}{h} & & & \\ & \ddots & \ddots & \ddots & \ddots & & \\ & & -\frac{1}{h} & \frac{2}{h} & -\frac{1}{h} & & \\ & & & -\frac{1}{h} & \frac{2}{h} & -\frac{1}{h} & \\ & & & & -\frac{1}{h} & \frac{1}{h} + \alpha \end{pmatrix}$$

- A is idd $\Rightarrow$ A is nonsingular
- $\mathrm{diag} A$ is positive real $\Rightarrow$ eigenvalues of A have positive real parts
- A is real, symmetric $\Rightarrow$ A is positive definite

Perron-Frobenius Theorem (1912/1907)

Definition: A real n-vector $\mathbf{x}$ is

- positive ($\mathbf{x} > 0$) if all entries of $\mathbf{x}$ are positive
- nonnegative ($\mathbf{x} \geq 0$) if all entries of $\mathbf{x}$ are nonnegative

Definition: A real $n \times n$ matrix A is

- positive ($A > 0$) if all entries of A are positive
- nonnegative ($A \geq 0$) if all entries of A are nonnegative

Theorem(Varga, Th. 2.7) Let $A \geq 0$ be an irreducible $n \times n$ matrix. Then

(i) A has a positive real eigenvalue equal to its spectral radius $\rho(A)$.
(ii) To $\rho(A)$ there corresponds a positive eigenvector $\mathbf{x} > 0$.
(iii) $\rho(A)$ increases when any entry of A increases.
(iv) $\rho(A)$ is a simple eigenvalue of A.

Proof: See Varga. □

Perron-Frobenius for general nonnegative matrices

Each $n \times n$ matrix can be brought to the normal form

$$PAP^T = \begin{pmatrix} R_{11} & R_{12} & \dots & R_{1m} \\ 0 & R_{22} & \dots & R_{2m} \\ \vdots & & \ddots & \\ 0 & 0 & \dots & R_{mm} \end{pmatrix}$$

where for $j = 1 \dots m$, either R_{jj} irreducible or $R_{jj} = (0)$.

Theorem(Varga, Th. 2.20) Let $A \geq 0$ be an $n \times n$ matrix. Then

(i) A has a nonnegative eigenvalue equal to its spectral radius $\rho(A)$. This eigenvalue is positive unless A is reducible and its normal form is strictly upper triangular

(ii) To $\rho(A)$ there corresponds a nonzero eigenvector $\mathbf{x} \geq 0$.

(iii) $\rho(A)$ does not decrease when any entry of A increases.

Proof: See Varga; $\sigma(A) = \bigcup_{j=1}^{m} \sigma(R_{jj})$, apply irreducible Perron-Frobenius to R_{jj}. $\square$

Theorem on Jacobi matrix

Theorem: Let A be sdd or idd, and D its diagonal. Then

$$\rho(|I - D^{-1}A|) < 1$$

Proof: Let $B = (b_{ij}) = I - D^{-1}A$. Then

$$b_{ij} = \begin{cases} 0, & i = j \\ -\frac{a_{ij}}{a_{ii}}, & i \neq j \end{cases}$$

If A is sdd, then for $i = 1 \dots n$,

$$\sum_{j=1\dots n} |b_{ij}| = \sum_{\substack{j=1\dots n \\ j\neq i}} |\frac{a_{ij}}{a_{ii}}| = \frac{\Lambda_i}{|a_{ii}|} < 1$$

Therefore, $\rho(|B|) < 1$.

Theorem on Jacobi matrix II

If A is idd, then for $i = 1 \dots n$,

$$\sum_{j=1\dots n} |b_{ij}| = \sum_{\substack{j=1\dots n \\ j\neq i}} |\frac{a_{ij}}{a_{ii}}| = \frac{\Lambda_i}{|a_{ii}|} \leq 1$$

$$\sum_{j=1\dots n} |b_{rj}| = \frac{\Lambda_r}{|a_{rr}|} < 1 \text{ for at least one } r$$

Therefore, $\rho(|B|) <= 1$. Assume $\rho(|B|) = 1$. By Perron-Frobenius, 1 is an eigenvalue. As it is in the union of the Gershgorin disks, for some i,

$$|\lambda| = 1 \leq \frac{\Lambda_i}{|a_{ii}|} \leq 1$$

and it must lie on the boundary of this union. By Taussky then one has for all i

$$|\lambda| = 1 \leq \frac{\Lambda_i}{|a_{ii}|} = 1$$

which contradicts the idd condition. □

Jacobi method convergence

Corollary: Let A be sdd or idd, and D its diagonal. Assume that $a_{ii} > 0$ and $a_{ij} \leq 0$ for $i \neq j$. Then $\rho(I - D^{-1}A) < 1$, i.e. the Jacobi method converges.

Proof In this case, $|B| = B$ $\square$.

- Here, we made assumptions on the sign pattern and the diagonal dominance of the matrix. No additional information on the nonzero pattern or the symmetry has been used.
- Does this generalize to other iterative methods ?

Regular splittings

- $A = M - N$ is a regular splitting if
  - M is nonsingular
  - M^{-1}, N are nonnegative, i.e. have nonnegative entries
- Regard the iteration $u_{k+1} = M^{-1}Nu_k + M^{-1}b$.
- $B = I - M^{-1}A = M^{-1}N$ is a nonnegative matrix.

Convergence theorem for regular splitting

Theorem: Assume A is nonsingular, $A^{-1} \geq 0$, and $A = M - N$ is a regular splitting. Then $\rho(M^{-1}N) < 1$.

Proof: Let $B = M^{-1}N$. Then $A = M(I - B)$, therefore $I - B$ is nonsingular.

In addition

$$A^{-1}N = (M(I - M^{-1}N))^{-1}N = (I - M^{-1}N)^{-1}M^{-1}N = (I - B)^{-1}B$$

By Perron-Frobenius (for general matrices), $\rho(B)$ is an eigenvalue with a nonnegative eigenvector $\mathbf{x}$. Thus,

$$0 \leq A^{-1}N\mathbf{x} = \frac{\rho(B)}{1 - \rho(B)}\mathbf{x}$$

Therefore $0 \leq \rho(B) \leq 1$.
Assume that $\rho(B) = 1$. Then there exists $\mathbf{x} \neq \mathbf{0}$ such that $B\mathbf{x} = \mathbf{x}$.
Consequently, $(I - B)\mathbf{x} = \mathbf{0}$, contradicting the nonsingularity of $I - B$.
Therefore, $\rho(B) < 1$. □