*Numerical Linear Algebra*

Maya Neytcheva, TDB, February-March 2021



## Basic Iterative Solution methods



# Basic Iterative Solution methods

Introduction:
The ideas to use iterative methods for solving linear systems of equations go back to Gauss (1823), Liouville (1837) and Jacobi (1845).

After deriving an iterative procedure, in 1823, Gauss has written in a letter the following:

*"... You will hardly eliminate directly anymore, at least not when you have more than two unknowns. The indirect method can be pursued while half asleep or while thinking about other things."*



# Introduction:

Before considering iterative solution methods for linear systems of equations, we recall how do we solve nonlinear problems
Let $f(x) = 0$ have to be solved and $f(x)$ is a nonlinear function in $x$.
The usual way to approach the problem is:

$$F(x) \equiv x - f(x).$$

If $x^*$ is the solution of $f(x) = 0$, then $x^*$ is a stationary point for

$$x = F(x). \tag{1}$$

Then we proceed with finding the stationary point for (1) and this is done iteratively, namely,

$$x^{(k+1)} = F(x^{(k)}), k = 0, 1, \cdots, x^{(0)} \text{ given}. \tag{2}$$

## Convergence of the fixed point iteration:

For any initial guess $x^{(0)}$, there exists a unique fixed point $x^*$ for $F(x)$, $x^* = \lim_{k\to\infty} x^{(k)}$ if and only if $F$ is a contracting mapping, i.e.

$$\|F(x) - F(y)\| \leq q\|x - y\|$$

for some $q \in (0, 1)$.



## Fixed point for linear problems:

Let now $f(\mathbf{x}) \equiv A\mathbf{x} - \mathbf{b}$ be linear. We use the same framework:

$$F(\mathbf{x}) = \mathbf{x} - (A\mathbf{x} - \mathbf{b})$$
$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - (A\mathbf{x}^{(k)} - \mathbf{b}) = \mathbf{x}^{(k)} + \mathbf{r}^{(k)}$$

where $\mathbf{r}^{(k)} = \mathbf{b} - A\mathbf{x}^{(k)}$ is called the residual at iteration $k$.
In this way we obtain the simplest possible iterative scheme to solve

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

namely,

$$\boxed{\begin{array}{l} \mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - (A\mathbf{x}^{(k)} - \mathbf{b}), \quad k = 0, 1, \cdots \\ \mathbf{x}^{(0)} \text{ given.} \end{array}}$$



## Fixed point for linear problems - when does it converge?

$$\begin{aligned} \|F(\mathbf{x}) - F(\mathbf{y})\| &= \|\mathbf{x} - (A\mathbf{x} - \mathbf{b}) - \mathbf{y} - (A\mathbf{y} - \mathbf{b}\|) \\ &= \|(I - A)(x - y)\| \leq \|(I - A)\|\|(x - y)\| \end{aligned}$$

The simple iteration will not converge in general.



## Simple iteration

For many reasons the latter form of the simple iteration is replaced by

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \tau\mathbf{r}^{(k)}, \qquad (3)$$

where $\tau$ is some properly chosen method parameter.

Relation (3) defines the so-called *stationary basic iterative method of first kind*.

UPPSALA UNIVERSITET

# Stationary iterative methods ...

If we permit $\tau$ to change from one iteration to the next, we get

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \tau_k \mathbf{r}^{(k)}, \tag{4}$$

which latter defines the so-called
*non-stationary basic iterative method of first kind.*

---

So far $\tau$ and $\tau_k$ are some scalars. Nothing prevents us to replace the method parameter by some matrix, however, if this would improve the convergence of the iterative method.



UPPSALA UNIVERSITET

Nothing prevents us to replace the method parameter by some matrix, however, if this would improve the convergence of the iterative method. Thus, we can consider

$$\begin{aligned} &\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + C^{-1}(\mathbf{b} - A\mathbf{x}^{(k)}) \\ &\text{or} \\ &\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + C^{-1}\mathbf{r}^{(k)}, \end{aligned} \tag{5}$$

It is easy to see that we obtain (5) by replacing $A\mathbf{x} = \mathbf{b}$ with

$$C^{-1}A\mathbf{x} = C^{-1}\mathbf{b}$$

and use the simple iteration framework.



UPPSALA UNIVERSITET

In this case the iterative scheme takes the form

$$\begin{aligned} C\mathbf{d}^{(k)} &= \mathbf{r}^{(k)}, \\ \mathbf{x}^{(k+1)} &= \mathbf{x}^{(k)} + \mathbf{d}^{(k)} \end{aligned} \tag{6}$$

The scheme

$$\begin{aligned} C\mathbf{d}^{(k)} &= \mathbf{r}^{(k)}, \\ \mathbf{x}^{(k+1)} &= \mathbf{x}^{(k)} + \mathbf{d}^{(k)} \end{aligned}$$

has in general a higher computational complexity than (4), since a solution of a system with the matrix $C$ is required at each iteration.



# Concerns: I

**C1:** Does the iteration process converge to the solution, i.e. does $\mathbf{x}^{(k)} \to \mathbf{x}^*$?

**C2:** If 'yes', how fast does it converge?
The number of iterations *it* needed for the iterative method to converge with respect to some convergence criterion, is a function of the properties of $A$.
Say, $it = it(n)$, where $n$ is the size of $A$.
If it turns out that $it = O(n^2)$, we haven't gained anything compared to the direct solution methods.
The best one can hope for is to get $it \leq Const$, where $Const$ is independent of $n$.

## Concerns: II

Since the computational complexity of one iteration is in many cases proportional to $n$ (for sparse matrices, for instance) then the complexity of the whole solution process will be

$$O(n).$$



## Concerns: I

**C3:** Is the method *robust* with respect to the method parameters $(\tau,\ \tau_k)$?

**C4:** Is the method *robust* with respect to various problem parameters?

$$A = A(\rho, \nu, E, \cdots)$$

**C5:** When we are using the scheme $C^{-1}A\mathbf{x} = C^{-1}\mathbf{b}$, it must be easy to solve systems with $C$.

**C6:** Is the method parallelizable?
Parallelization aspects become more and more important since $n$ is XXL.



## Concerns (cont.): I

Suppose the method converges to the exact solution $\mathbf{x}^*$.
Then more questions arise:

**C7:** When do we stop the iterations?

- → We want $\|\mathbf{x}^* - \mathbf{x}^{(k)}\| \leq \varepsilon$ but $\mathbf{x}^*$ is not known.
- → What about checking on $\mathbf{r}^{(k)}$?
- → Is it enough to have $\|\mathbf{r}^{(k)}\| \leq \widetilde{\varepsilon}$?
  Will the latter guarantee that $\|\mathbf{x}^* - \mathbf{x}^{(k)}\| \leq \varepsilon$?
  Denote $\mathbf{e}^{(k)} = \mathbf{x}^* - \mathbf{x}^{(k)}$ (the error at iteration $k$). Then

$$\mathbf{r}^{(k)} = \mathbf{b} - A\mathbf{x}^{(k)} = A(\mathbf{x}^* - \mathbf{x}^{(k)}) = A\mathbf{e}^{(k)}.$$

In other words $\mathbf{e}^{(k)} = A^{-1}\mathbf{r}^{(k)}$.
Scenario: Suppose $\|A^{-1}\| = 10^8$ and $\widetilde{\varepsilon} = 10^{-4}$. Then
$\boxed{\|\mathbf{e}^{(k)}\| \leq \|A^{-1}\|\|\mathbf{r}^{(k)}\| \leq 10^4,}$ which is not very exiting.
Example: Discrete Laplace $\Delta_h^5$:
$\|A^{-1}\| \approx \lambda_{min} = \frac{1}{2}(\pi h)^2 \approx 10^4 \text{ for } h = 10^{-2}$.



## Concerns (cont.): II

**C8:** How do we measure (estimate) the convergence rate?

**C9:** How do we find good method parameters $(\tau,\ \tau_k,\ C)$, which will speed up the convergence?

---

We start our considerations with [C9].

## Choosing $C$:

Intuitively, $C$ has to do something with $A$.
Note that if $C = A$, then $C^{-1} = A^{-1}$ and we will get convergence in one step!
However, the computational effort to construct $A^{-1}$ is higher than to use a direct solution method.

We try the following choice. Consider the following so-called splitting of $A$,

$$A = C - R,$$

where $C$ is nonsingular and $R$ can be seen as an error matrix.



## Choosing $C$ (cont.)

Then $C^{-1}A = C^{-1}(C - R) = I - C^{-1}R = I - B$.

- The matrix $\boxed{B = C^{-1}R}$ is referred to as the iteration matrix.
- $\|B^m\|$ is the convergence factor for $m$ steps
- $(\|B^m\|)^{1/m}$ is called the average convergence factor.



## Equivalent formulation using the splitting:

Using the splitting $A = C - R$ we obtain the following equivalent form of the iterative procedure:

$$\begin{aligned} & A = C - R \longrightarrow R = C - A \\ \mathbf{x}^{(k+1)} &= \mathbf{x}^{(k)} + C^{-1}(\mathbf{b} - A\mathbf{x}^{(k)}) \\ &= \mathbf{x}^{(k)} + C^{-1}\mathbf{b} - C^{-1}(C - R)\mathbf{x}^{(k)} \\ &= C^{-1}\mathbf{b} + C^{-1}R\mathbf{x}^{(k)} \end{aligned}$$

$$C\mathbf{x}^{(k+1)} = R\mathbf{x}^{(k)} + \mathbf{b} \tag{7}$$

The matrix $C$ is called a preconditioner to $A$. Its general purpose is to improve the properties of $A$ in order to achieve a better (faster) convergence of the method.



## A general convergence result:

**Theorem**

*The sequence $\{\mathbf{x}^{(k)}\}$ from $C\mathbf{x}^{(k+1)} = R\mathbf{x}^{(k)} + \mathbf{b}$ converges to the solution $\mathbf{x}^*$ of $A\mathbf{x} = \mathbf{b}$ for any initial guess $\mathbf{x}^{(0)}$ if and only if there holds*

$$\rho(B) \equiv \rho(C^{-1}R) < 1$$

*where $\rho(\cdots)$ denotes the spectral radius.*

**Proof.**

Let $\mathbf{e}^{(k)} = \mathbf{x}^* - \mathbf{x}^{(k)}$, $A = C - R$. Then

$$\left.\begin{aligned} C\mathbf{x}^* &= R\mathbf{x}^* + \mathbf{b} \\ C\mathbf{x}^{(k)} &= R\mathbf{x}^{(k-1)} + \mathbf{b} \end{aligned}\right] -$$

$$C\mathbf{e}^{(k)} = R\mathbf{e}^{(k-1)}$$

$$\mathbf{e}^{(k)} = B\mathbf{e}^{(k-1)} = B^2\mathbf{e}^{(k-2)} = \cdots = B^k\mathbf{e}^{(0)}.$$

If $\rho(C^{-1}R) < 1$ then $\lim_{k\to\infty} B^k = 0$ and $\mathbf{e}^{(k)} \to 0$. □

## If $\rho(C^{-1}R) \geq 1$:

Let $\lambda_i = eig(B)$ and $\rho(B) = |\lambda_j|$ i.e., $\lambda_j$ is the eigenvalue of $B$, such that $\rho(B) = |\lambda_j|$. Let $\mathbf{v}^j$ be the corresponding eigenvector. Then $\left(\mathbf{v}^j\right)^m = B^m\mathbf{v}^j = \lambda_j^m\mathbf{v}^j \nrightarrow 0$.

$$\mathbf{e}^{(0)} = \sum_{k=1}^{n} \beta_k \mathbf{v}^k = \beta_j \mathbf{v}^j + \cdots$$

$$B^m\mathbf{e}^{(0)} = \widetilde{\beta}_j B^m \mathbf{v}^j + \cdots$$

and at least one component of $\mathbf{e}^{(m)}$ does not converge to zero.



## Remarks on the proof:

The basic argument in the latter proof is that if $\rho(B) < 1$ then $B^k \to 0$.. This can be shown in the following way.

**Lemma**

*Let $T$ be a nonsingular matrix and let $\|\mathbf{x}\|_T = \|T\mathbf{x}\|_\infty$. Let $\|A\|_T = \sup_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|_T}{\|\mathbf{x}\|_T}$ be the induced matrix norm. Then.*

(a) $\|A\|_T = \|TAT^{-1}\|_\infty$

(b) *For any $\varepsilon > 0$ and matrix $A$, there exists a nonsingular matrix $T$ such that $\|A\|_T \leq \rho(A) + \varepsilon$.*
*In other words, there exist matrix norms, which are arbitrary close to the spectral radius of a given matrix.*



## Remarks on the proof:

**Proof.**

(a) $\|\mathbf{x}\|_T$ is a vector norm and $T^{-1}$ exists.

$$\|A\|_T = \sup_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|_T}{\|\mathbf{x}\|)_T} = \sup_{\mathbf{x}\neq\mathbf{0}} \frac{\|TA\mathbf{x}\|_\infty}{\|T\mathbf{x}\|_\infty} = \sup_{\mathbf{y}\neq\mathbf{0}} \frac{\|TAT^{-1}\mathbf{y}\|_\infty}{\|\mathbf{y}\|_\infty} = \|TAT^{-1}\|_\infty.$$

□



## *Proof* (cont.)

(b) We use Schur's lemma: There exists a unitary matrix $U$, such that

$$UAU^{-1} = W = \begin{bmatrix} w_{11} & * & * & \cdots & * \\ 0 & w_{22} & * & \cdots & * \\ & & \ddots & \cdots & \vdots \\ & & & & w_{nn} \end{bmatrix},$$

where $w_{ii} = \lambda_i \in S(A)$; $S(A)$ denotes the spectrum of $A$. Let $\delta > 0$ and define $D = D(\delta) = diag\{\delta^{-1}, \delta^{-2}, \cdots, \delta^{-n}\}$. Then $DWD^{-1}$ is also upper triangular and $(DWD^{-1})_{ij} = \begin{cases} 0, & j < i \\ w_{ii}, & j = i \\ w_{ij}\delta^{j-1}, & j > i. \end{cases}$

$$\Rightarrow \quad \|DWD^{-1}\|_\infty \leq \max_i \left\{ |w_{ii}| + n \max_{j>i} |w_{ij}| \delta^{j-1} \right\}.$$

## *Proof* (cont.)

We see that for any given $\varepsilon > 0$ we can choose $\delta > 0$ small enough so that $n \max_{j>i} |w_{ij}\delta^{j-1} < \varepsilon$. Hence,

$\|DWD^{-1}\|_\infty \leq \rho(A) + \varepsilon$

$\|A\|_T = \|TAT^{-1}\|_\infty = \|DUAU^{-1}D^{-1}\|_\infty$

$= \|DWD^{-1}\|_\infty \leq \rho(A) + \varepsilon.$

(for $T = DU$, nonsingular).



## Convergence (cont.)

**Lemma**

*For any square matrix there holds*

*(a)* $\lim_{k\to\infty} A^k = 0 \Leftrightarrow \rho(A) < 1,$

*(b)* *If* $\rho(A) < 1$ *then* $(I - A)^{-1} = I + A + A^2 + \cdots$ *is convergent.*

**Proof.**

(a)'$\Rightarrow$': If $\rho(A) < 1$ then choose $\varepsilon > 0$: $\rho(A) + \varepsilon < 1$. Then there exists a nonsingular $T$ (which depends on $A$), such that $\|A\|_T \leq \rho(A) + \varepsilon < 1$.

$\Rightarrow \quad \|A^k\|_T \leq \|A[|_T^k \to 0 \quad \Rightarrow \lim_{k\to\infty} A^k = 0$

(a)'$\Leftarrow$': If $\lim_{k\to\infty} A^k = 0$, let $\{\lambda, \mathbf{v}\}$ be an eigensolution of $A$, then $\lambda^k \mathbf{v} = A^k \mathbf{v} \to 0$. This is true for all eigenvalues, thus $\rho(A) = \max|\lambda| < 1$.

(b) $(I - A)(I + A + A^2 + \cdots) = I - A^{k+1}$. If $\rho(A) < 1$ then $A^k \to 0 \Rightarrow$ (b) follows. □



## Rate of convergence

Theorem 1 shows both convergence and rate of convergence ($\mathbf{e}^{(k)} = B^k \mathbf{e}^{(k)}$). The latter is difficult to compute. Also the convergence may not be monotone.

**Theorem**

*Consider* $C\mathbf{x}^{(k+1)} = R\mathbf{x}^{(k)} + \mathbf{b}$, $B = C^{-1}R$ *and let* $\rho(B) < 1$. *Then*

$$\|\mathbf{x}^* - \mathbf{x}^{(k)}\| \leq \frac{\|B\|}{1 - \|B\|}\|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\|$$



## Rate of convergence, cont.

**Proof.**

$\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)} = B(\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)})$ and

$\mathbf{x}^{(k+m+1)} - \mathbf{x}^{(k+m)} = B^{m+1}(\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)})$. We have

$\|\mathbf{x}^{(k+s)} - \mathbf{x}^{(k)}\| = \|\sum_{j=0}^{s-1}(\mathbf{x}^{(k+j+1)} - \mathbf{x}^{(k+j)})\| \leq$

$\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\| + \|\mathbf{x}^{(k+2)} - \mathbf{x}^{(k+1)}\| + \cdots$

Therefore

$$\|\mathbf{x}^{(k+m+1)} - \mathbf{x}^{(k+m)}\| \leq \sum_{j=0}^{m} \|B^j\| \, \|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\| = \frac{\|B\| - \|B\|^{m+1}}{1 - \|B\|}\|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\|$$

We let now $m \to \infty$, i.e., $\mathbf{x}^{k+m} \to \mathbf{x}^*$, $\|B\|^m \to 0$.

$$\Rightarrow \|\mathbf{x}^* - \mathbf{x}^{(k)}\| \leq \tfrac{\|B\|}{1-\|B\|}\|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\|.$$

□

UPPSALA UNIVERSITET

## Stopping tests:

Theorem 4 can be used to get information whether the iteration error $\mathbf{e}^{(k)} = \mathbf{x}^* - \mathbf{x}^{(k)}$ is small enough.

---

In practice, most used stopping tests are:

**(S1)** $\|\mathbf{r}^{(k)}\| \le \varepsilon$, residual based, absolute



UPPSALA UNIVERSITET

## Stopping tests:

Theorem 4 can be used to get information whether the iteration error $\mathbf{e}^{(k)} = \mathbf{x}^* - \mathbf{x}^{(k)}$ is small enough.

---

In practice, most used stopping tests are:

**(S1)** $\|\mathbf{r}^{(k)}\| \le \varepsilon$, residual based, absolute

**(S2)** $\|\mathbf{r}^{(k)}\| \le \varepsilon \|\mathbf{r}^{(0)}\|$, residual based, relative



UPPSALA UNIVERSITET

## Stopping tests:

Theorem 4 can be used to get information whether the iteration error $\mathbf{e}^{(k)} = \mathbf{x}^* - \mathbf{x}^{(k)}$ is small enough.

---

In practice, most used stopping tests are:

**(S1)** $\|\mathbf{r}^{(k)}\| \le \varepsilon$, residual based, absolute

**(S2)** $\|\mathbf{r}^{(k)}\| \le \varepsilon \|\mathbf{r}^{(0)}\|$, residual based, relative

**(S3)** $\|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\| \le \varepsilon$



UPPSALA UNIVERSITET

## Stopping tests:

Theorem 4 can be used to get information whether the iteration error $\mathbf{e}^{(k)} = \mathbf{x}^* - \mathbf{x}^{(k)}$ is small enough.

---

In practice, most used stopping tests are:

**(S1)** $\|\mathbf{r}^{(k)}\| \le \varepsilon$, residual based, absolute

**(S2)** $\|\mathbf{r}^{(k)}\| \le \varepsilon \|\mathbf{r}^{(0)}\|$, residual based, relative

**(S3)** $\|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\| \le \varepsilon$

**(S4)** $\|\mathbf{x}^* - \mathbf{x}^{(k)}\| \le \varepsilon_0 \|\mathbf{x}^* - \mathbf{x}^{(0)}\|$.

If the latter is wanted, then we must check on (S3) and choose $\varepsilon$ such that $\varepsilon \le \frac{\|B\|}{1-\|B\|}\varepsilon_0 \|\mathbf{x}^* - \mathbf{x}^{(0)}\|$.

Either estimate of $\|A^{-1}\|$ or of $\|B\|$ is required.

## Choices of the matrix $C$

Choice 'J'

Let $A = D - L - U$, where $D$ is diagonal, $U$ is strictly upper triangular and $L$ is strictly lower triangular.

Let $C \equiv D$, $R = L + U$. The iterative scheme is known as Jacobi iteration:

$$D\mathbf{x}^{(k+1)} = (L + U)\mathbf{x}^{(k)} + \mathbf{b}$$

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

For the method to converge: $B = D^{-1}(L + U)$

$$\rho(B) \leq \|D^{-1}(L + U)\|_\infty = \max_{1 \leq i \leq n} \sum_{j=1, j \neq i}^{n} \left|\frac{a_{ij}}{a_{ii}}\right|$$

We want $\rho(B) < 1$. One class of matrices, for which Jacobi method converges is when $A$ is strictly diagonally dominant.



## Choices of the matrix $C$



## Choices of the matrix $C$

1. Choice GS-B Choose $C \equiv D - U$, $R = L$

   Backward Gauss-Seidel $\boxed{(D - U)\mathbf{x}^{(k+1)} = L\mathbf{x}^{(k)} + \mathbf{b}}$



## Choices of the matrix $C$

1. Choice GS-B Choose $C \equiv D - U$, $R = L$

   Backward Gauss-Seidel $\boxed{(D - U)\mathbf{x}^{(k+1)} = L\mathbf{x}^{(k)} + \mathbf{b}}$

2. Choice GS-F Choose $C \equiv D - L$, $R = U$

   Forward Gauss-Seidel $\boxed{(D - L)\mathbf{x}^{(k+1)} = U\mathbf{x}^{(k)} + \mathbf{b}}$

## Choices of the matrix $C$

1. Choice GS-B Choose $C \equiv D - U$, $R = L$
   Backward Gauss-Seidel $\boxed{(D-U)\mathbf{x}^{(k+1)} = L\mathbf{x}^{(k)} + \mathbf{b}}$
2. Choice GS-F Choose $C \equiv D - L$, $R = U$
   Forward Gauss-Seidel $\boxed{(D-L)\mathbf{x}^{(k+1)} = U\mathbf{x}^{(k)} + \mathbf{b}}$
3. G-S is convergent for s.p.d. matrices.



## Choices of the matrix $C$

1. Choice GS-B Choose $C \equiv D - U$, $R = L$
   Backward Gauss-Seidel $\boxed{(D-U)\mathbf{x}^{(k+1)} = L\mathbf{x}^{(k)} + \mathbf{b}}$
2. Choice GS-F Choose $C \equiv D - L$, $R = U$
   Forward Gauss-Seidel $\boxed{(D-L)\mathbf{x}^{(k+1)} = U\mathbf{x}^{(k)} + \mathbf{b}}$
3. G-S is convergent for s.p.d. matrices.
4. make it more fancy: $A = D - L - U$. Then
   $$\begin{aligned}\omega A &= \omega D - \omega L - \omega L + D - D \leftarrow \text{ overrelaxation}\\ &= (D - \omega L) - (\omega U + (1-\omega)D)\end{aligned}$$
   Choose $C \equiv D - \omega L$, $R = \omega U + (1-\omega)D$:
   SOR $\boxed{(D-\omega L)\mathbf{x}^{(k+1)} = [\omega U + (1-\omega)D]\mathbf{x}^{(k)} + \omega\mathbf{b}}$



## SOR - back to 1940

One can see SOR as a generalization of G-S ($\omega = 1$). Rewrite
$(D-\omega L)\mathbf{x}^{(k+1)} = [\omega U + (1-\omega)D]\mathbf{x}^{(k)} + \omega\mathbf{b}$
as $\left(\frac{1}{\omega}D - L\right)\mathbf{x}^{(k+1)} = \left[\left(\frac{1}{\omega} - 1\right)D + U\right]\mathbf{x}^{(k)} + \mathbf{b}$
For the iteration matrix $B_\omega = \left(\frac{1}{\omega}D - L\right)^{-1}\left[\left(\frac{1}{\omega} - 1\right)D + U\right]$
One can show that $\rho(B_\omega) < 1$ for $0 < \omega < 2$. Furthermore, there is an optimal value of $\omega$, for which $\rho(B_\omega)$ is minimized:

$$\omega_{opt} = \frac{2}{1 + \sqrt{1 - \rho(\widehat{B})^2}}, \quad \widehat{B} = I - D_A^{-1}A.$$



## SOR - cont.

Rate of convergence: Let $\lambda_i = eig(B_\omega)$.
$\left|\prod_{i=1}^{n}\lambda_i\right| = \left|det\left((1-\omega)I + \omega D^{-1}U\right)\right| = |1-\omega|^n$. $\Rightarrow$ at least one $\lambda_i \geq |1-\omega|$.
$\Rightarrow \quad \rho(B_\omega) \geq |1-\omega|$.
We want $\rho(B_\omega) < 1$, i.e. $|1-\omega| \leq \rho(B_\omega) < 1$, $\Rightarrow \quad 0 < \omega < 2$.

## Splittings of $A$

Let $A, C, R \in \mathbf{R}^{n\times n}$ and consider $A = C - R$. A splitting of $A$ is called

- regular if $C$ is monotone and $R \geq 0$ (elementwise)
- weak regular if $C$ is monotone and $C^{-1}R \geq 0$
- nonnegative if $C^{-1}$ exists and $C^{-1}R \geq 0$
- convergent if $\rho(C^{-1}R) < 1$.

Recall: A matrix is called *monotone* if $A\mathbf{x} > 0$ implies $\mathbf{x} > 0$.
Theorem: $A$ - monotone $\Leftrightarrow A^{-1} \geq 0$.



## The Second Order Chebyshev iteration method

Let $A$ be symmetric matrix.

$$\begin{array}{rl} \mathbf{x}_0 \text{ given}, & \mathbf{x}_1 = \mathbf{x}_0 + \frac{1}{2}\beta_0\mathbf{r}_0 \\ \text{For} & k = 0, 1, \cdots \text{ until convergence} \\ & \mathbf{x}_{k+1} = \alpha_k\mathbf{x}_k + (1-\alpha_k)\mathbf{x}_{k-1} + \beta_k\mathbf{r}_k. \\ & \mathbf{r}_k = \mathbf{b} - A\mathbf{x}_k. \end{array}$$

$$\alpha_k = \frac{a+b}{2}\beta_k, \quad \frac{1}{\beta_k} = \frac{a+b}{2} - \left(\frac{b-a}{4}\right)^2\beta_{k-1}, \quad \beta_0 = \frac{4}{a+b}.$$

Note that $\alpha_k > 1, k \geq 1$.

Modifications for nonsymmetric matrices exist.