### Numerical Analysis - Part II

Anders C. Hansen

Lecture 16

# Iterative methods for linear algebraic systems

The general *iterative* method for solving Ax = b is a rule  $\mathbf{x}^{k+1} = f_k(\mathbf{x}^0, \mathbf{x}^1, \dots, \mathbf{x}^k)$ . We will consider the simplest ones: *linear, one-step, stationary* iterative schemes:

$$\mathbf{x}^{k+1} = H\mathbf{x}^k + \mathbf{v}, \qquad \mathbf{x}^0, \mathbf{v} \in \mathbb{R}^n.$$
 (1)

Here one chooses H and v so that  $x^*$ , a solution of  $A\mathbf{x} = \mathbf{b}$ , satisfies  $\mathbf{x}^* = H\mathbf{x}^* + \mathbf{v}$ , i.e. it is the fixed point of the iteration (1) (if the scheme converges). Standard terminology:

▶ the *iteration matrix* H,

• the error 
$$e^k := x^* - x^k$$
,

• the residual 
$$\mathbf{r}^k := A\mathbf{e}^k = \mathbf{b} - A\mathbf{x}^k$$
.

For a given class of matrices A (e.g. positive definite matrices, or even a single particular matrix), we are interested in *convergent* methods, i.e. the methods such that  $\mathbf{x}^k \to \mathbf{x}^* = A^{-1}\mathbf{b}$  for every starting value  $\mathbf{x}^0$ . Subtracting  $\mathbf{x}^* = H\mathbf{x}^* + \mathbf{v}$  from (1) we obtain

$$\boldsymbol{e}^{k+1} = H\boldsymbol{e}^k = \dots = H^{k+1}\boldsymbol{e}^0, \qquad (2)$$

i.e., a method is convergent if  $\mathbf{e}^k = H^k \mathbf{e}^0 \to 0$  for any  $\mathbf{e}^0 \in \mathbb{R}^n$ . (Iterative refinement). This is the scheme

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k - S(A\boldsymbol{x}^k - \boldsymbol{b}).$$

If  $S = A^{-1}$ , then  $\mathbf{x}^{k+1} = A^{-1}\mathbf{b} = \mathbf{x}^*$ , so it is suggestive to choose S as an approximation to  $A^{-1}$ . The iteration matrix for this scheme is  $H_S = I - SA$ .

(Splitting). This is the scheme

$$(A-B)\boldsymbol{x}^{k+1}=-B\boldsymbol{x}^k+\boldsymbol{b}\,,$$

with the iteration matrix  $H = -(A - B)^{-1}B$ . Any splitting can be viewed as an iterative refinement (and vice versa) because

$$(A-B)\mathbf{x}^{k+1} = -B\mathbf{x}^k + \mathbf{b} \quad \Leftrightarrow \quad (A-B)\mathbf{x}^{k+1} = (A-B)\mathbf{x}^k - (A\mathbf{x}^k - \mathbf{b})$$
$$\Leftrightarrow \quad \mathbf{x}^{k+1} = \mathbf{x}^k - (A-B)^{-1}(A\mathbf{x}^k - \mathbf{b}),$$

so we should seek a splitting such that  $S = (A - B)^{-1}$  approximates  $A^{-1}$ .

Theorem 1 Let  $H \in \mathbb{R}^{n \times n}$ . Then  $\lim_{k \to \infty} H^k z = 0$  for any  $z \in \mathbb{R}^n$  if and only if  $\rho(H) < 1$ .

**Proof.** 1) Let  $\lambda$  be an eigenvalue of (the real) H, real or complex, such that  $|\lambda| = \rho(H) \ge 1$ , and let  $\boldsymbol{w}$  be a corresponding eigenvector, i.e.,  $H\boldsymbol{w} = \lambda \boldsymbol{w}$ . Then  $H^k \boldsymbol{w} = \lambda^k \boldsymbol{w}$ , and

$$\|H^{k}\boldsymbol{w}\|_{\infty} = |\lambda|^{k}\|\boldsymbol{w}\|_{\infty} \ge \|\boldsymbol{w}\|_{\infty} =: \gamma > 0.$$
(3)

If  $\boldsymbol{w}$  is real, we choose  $\boldsymbol{z} = \boldsymbol{w}$ , hence  $\|\boldsymbol{H}^{k}\boldsymbol{z}\|_{\infty} \geq \gamma$ , and this cannot tend to zero.

If  $\boldsymbol{w}$  is complex, then  $\boldsymbol{w} = \boldsymbol{u} + i\boldsymbol{v}$  with some real vectors  $\boldsymbol{u}, \boldsymbol{v}$ . But then at least one of the sequences  $(H^k \boldsymbol{u}), (H^k \boldsymbol{v})$  does not tend to zero. For if both do, then also  $H^k \boldsymbol{w} = H^k \boldsymbol{u} + iH^k \boldsymbol{v} \to 0$ , and this contradicts (3).

**Proof. Cont.** 2) Now, let  $\rho(H) < 1$ , and assume for simplicity that H possesses n linearly independent eigenvectors  $(\boldsymbol{w}_j)$  such that  $H\boldsymbol{w}_j = \lambda_j \boldsymbol{w}_j$ . Linear independence means that every  $\boldsymbol{z} \in \mathbb{R}^n$  can be expressed as a linear combination of the eigenvectors, i.e., there exist  $(c_j) \in \mathbb{C}$  such that  $\boldsymbol{z} = \sum_{j=1}^n c_j \boldsymbol{w}_j$ . Thus,

$$H^k \mathbf{z} = \sum_{j=1}^n c_j \lambda_j^k \mathbf{w}_j$$
,

and since  $|\lambda_j| \leq \rho(H) < 1$  we have  $\lim_{k \to \infty} H^k z = 0$ , as required.  $\Box$ 

## Solving linear systems – Convergence

#### Remark 2

The complete proof of case (2) of Theorem 1 exploits the so-called Jordan normal form of the matrix H, namely  $H = SJS^{-1}$ , where J is a block diagonal matrix consisting of the Jordan blocks,



To prove that  $J_i^k \to 0$  if  $|\lambda_i| < 1$  one should split  $J_i = \lambda_i I + P$ , notice that  $P^m = 0$  for  $m \ge n_i$ , and evaluate the terms of the expansion  $(\lambda_i I + P)^k = \sum_{m=0}^{n_i-1} {k \choose m} \lambda_i^{k-m} P^m$ .

### Solving linear systems – Convergence

Applying Theorem 1 to the error estimate (2), we arrive at the following statement.

#### Theorem 3

Let  $\mathbf{x}^*$ , a solution of  $A\mathbf{x} = \mathbf{b}$ , satisfy  $\mathbf{x}^* = H\mathbf{x}^* + \mathbf{v}$  and we are given the scheme

$$\mathbf{x}^{k+1} = H\mathbf{x}^k + \mathbf{v}, \qquad \mathbf{x}^0, \mathbf{v} \in \mathbb{R}^n.$$
 (4)

Then  $\mathbf{x}^k \to \mathbf{x}^*$  for any choice of  $\mathbf{x}^0$  if and only if  $\rho(H) < 1$ .

**Note:** Of course, we would like to know not just convergence but the rate of it. For example, we achieve convergence with

$$H = \left[ \begin{array}{cc} 0.99 & 10^6 \\ 0 & 0.99 \end{array} \right],$$

but it will take quite a long time. We will discuss this topic briefly later on.

Both of these methods are versions of splitting which can be applied to any A with nonzero diagonal elements. We write A as the sum of three matrices  $L_0 + D + U_0$ : subdiagonal (strictly lower-triangular), diagonal and superdiagonal (strictly upper-triangular) portions of A, respectively. 1) Jacobi method. We set A - B = D, the diagonal part of A, and we obtain the next iteration by solving the diagonal system

$$D \mathbf{x}^{(k+1)} = -(L_0 + U_0) \mathbf{x}^{(k)} + \mathbf{b}, \qquad H_{\mathrm{J}} = -D^{-1}(L_0 + U_0).$$

2) Gauss-Seidel method. We take  $A - B = L_0 + D = L$ , the lower-triangular part of A, and we generate the sequence  $(\mathbf{x}^{(k)})$  by solving the triangular system

$$(L_0 + D) \mathbf{x}^{(k+1)} = -U_0 \mathbf{x}^{(k)} + \mathbf{b}, \qquad H_{\rm GS} = -(L_0 + D)^{-1} U_0.$$

There is no need to invert  $(L_0 + D)$ , we calculate the components of  $\mathbf{x}^{(k+1)}$  in sequence by forward substitution:

$$a_{ii}x_i^{(k+1)} = -\sum_{j < i} a_{ij}x_j^{(k+1)} - \sum_{j > i} a_{ij}x_j^{(k)} + b_i, \qquad i = 1..n.$$

As we mentioned above, the sequence  $x^{(k)}$  converges to the solution of Ax = b if the spectral radius of the iteration matrix,

$$H_{
m J} = -D^{-1}(L_0+U_0) ext{ or } H_{
m GS} = -(L_0+D)^{-1}U_0,$$

respectively, is less than one. Our next goal is to prove that this is the case for two important classes of matrices A:

a) diagonally dominant and b) positive definite matrices.

We start with recalling the simple, but very useful Gershgorin theorem.

All eigenvalues of an  $n \times n$  matrix A are contained in the union of the Gershgorin discs in the complex plane:

$$\sigma(A) \subset \cup_{i=1}^n \Gamma_i, \quad \Gamma_i := \{ z \in \mathbb{C} : |z - a_{ii}| \le r_i \}, \quad r_i := \sum_{j \ne i} |a_{ij}|.$$

#### Definition 4 (Strictly diagonally dominant matrices)

A matrix A is called strictly diagonally dominant by rows (resp. by columns) if

$$|\mathbf{a}_{ii}| > \sum_{j \neq i} |\mathbf{a}_{ij}|, \quad i = 1..n$$
 (resp.  $|\mathbf{a}_{jj}| > \sum_{i \neq j} |\mathbf{a}_{ij}|, \quad j = 1..n$ ).

From Gershgorin theorem, it follows that strictly diagonally dominant matrices are nonsingular.

## **Convergence of iterations**

#### Theorem 5

If A is strictly diagonally dominant, then both the Jacobi and the Gauss-Seidel methods converge.

**Proof.** For the Gauss-Seidel method, the eigenvalues of the iteration matrix  $H_{\rm GS} = -(L_0 + D)^{-1}U_0$  satisfy the equation

$$\det[H_{\mathrm{GS}} - \lambda I] = \det[-(L_0 + D)^{-1}U_0 - \lambda I] = 0.$$

Moreover,

$$\det[-(L_0+D)^{-1}U_0-\lambda I]=0 \quad \Rightarrow \quad \det[A_\lambda]:=\det[U_0+\lambda D+\lambda L_0]=0.$$

It is easy to see that if  $A = L_0 + D + U_0$  is strictly diagonally dominant, then for  $|\lambda| \ge 1$  the matrix  $A_{\lambda} = \lambda L_0 + \lambda D + U_0$  is strictly diagonally dominant too, hence it is nonsingular, and therefore the equality det $[A_{\lambda}] = 0$  is impossible. Thus  $|\lambda| < 1$ , hence convergence. The proof for the Jacobi method is the same.  $\Box$ 

#### Theorem 6 (The Householder–John theorem)

If A and B are real matrices such that both A and  $A-B-B^{T}$  are symmetric positive definite, then the spectral radius of  $H = -(A - B)^{-1}B$  is strictly less than one.

**Proof.** Let  $\lambda$  be an eigenvalue of H, so  $H\mathbf{w} = \lambda \mathbf{w}$  holds, where  $\mathbf{w} \neq 0$  is an eigenvector. (Note that both  $\lambda$  and  $\mathbf{w}$  may have nonzero imaginary parts when H is not symmetric, e.g. in the Gauss-Seidel method.) The definition of H provides equality  $-B\mathbf{w} = \lambda(A - B)\mathbf{w}$ , and we note that  $\lambda \neq 1$  since otherwise A would be singular (which it is not). Thus, we deduce

$$\overline{\boldsymbol{w}}^{\mathsf{T}} \boldsymbol{B} \boldsymbol{w} = \frac{\lambda}{\lambda - 1} \overline{\boldsymbol{w}}^{\mathsf{T}} \boldsymbol{A} \boldsymbol{w}, \tag{5}$$

where the bar means complex conjugation.

#### The Householder–John theorem

**Proof. Cont.** Moreover, writing  $\boldsymbol{w} = \boldsymbol{u} + i\boldsymbol{v}$ , where  $\boldsymbol{u}$  and  $\boldsymbol{v}$  are real, we find (for  $C = C^T$ ) the identity  $\overline{\boldsymbol{w}}^T C \boldsymbol{w} = \boldsymbol{u}^T C \boldsymbol{u} + \boldsymbol{v}^T C \boldsymbol{v}$ , so symmetric positive definiteness in the assumption implies  $\overline{\boldsymbol{w}}^T A \boldsymbol{w} > 0$  and  $\overline{\boldsymbol{w}}^T (A - B - B^T) \boldsymbol{w} > 0$ . In the latter inequality, we use relation (5) and its conjugate transpose to obtain

$$0 < \overline{\boldsymbol{w}}^{T} A \boldsymbol{w} - \overline{\boldsymbol{w}}^{T} B \boldsymbol{w} - \overline{\boldsymbol{w}}^{T} B^{T} \boldsymbol{w} = \left(1 - \frac{\lambda}{\lambda - 1} - \frac{\overline{\lambda}}{\overline{\lambda} - 1}\right) \overline{\boldsymbol{w}}^{T} A \boldsymbol{w}$$
$$= \frac{1 - |\lambda|^{2}}{|\lambda - 1|^{2}} \overline{\boldsymbol{w}}^{T} A \boldsymbol{w}.$$

Now  $\lambda \neq 1$  implies  $|\lambda - 1|^2 > 0$ . Hence, recalling that  $\overline{\boldsymbol{w}}^T A \boldsymbol{w} > 0$ , we see that  $1 - |\lambda|^2$  is positive. Therefore  $|\lambda| < 1$  occurs for every eigenvalue of H as required.