Numerical Analysis - Part II

Anders C. Hansen

Lecture 16

Iterative methods for linear algebraic systems

Solving linear systems with iterative methods

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

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

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* $\mathbf{e}^k := \mathbf{x}^* - \mathbf{x}^k$, the *residual* $\mathbf{r}^k := A\mathbf{e}^k$

Solving linear systems – Iterative refinement

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

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

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

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

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 $H_S = I - SA$.

Solving linear systems – Splitting

(Splitting). This is the scheme

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

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

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

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

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

$$\|H^k\mathbf{w}\|_\infty = |\lambda|^k \|\mathbf{w}\|_\infty \geq \|\mathbf{w}\|_\infty =: \gamma > 0. \tag{3}$$

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

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

Proof. Cont. 2) Now, let $\rho(H) < 1$, and assume for simplicity that H possesses n linearly independent eigenvectors $(\mathbf{w}_j)$ such that $H\mathbf{w}_j = \lambda_j \mathbf{w}_j$. Linear independence means that every $\mathbf{z} \in \mathbb{R}^n$ can be expressed as a linear combination of the eigenvectors, i.e., there exist $(c_j) \in \mathbb{C}$ such that $\mathbf{z} = \sum_{j=1}^{n} c_j \mathbf{w}_j$. 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 $\lim_{k \to \infty} H^k \mathbf{z} = 0$, as required. $\square$

Remark 2 (Non-examinable)

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,

$$J = \begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_r \end{bmatrix}, \qquad J_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix}, \qquad J_i \in \mathbb{R}^{n_i \times n_i},$$

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 $m \geq n_i$, and evaluate the terms of the expansion $(\lambda_i I + P)^k = \sum_{m=0}^{n_i - 1} \binom{k}{m} \lambda_i^{k-m} P^m$.

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

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

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

$$H = \begin{bmatrix} 0.99 & 10^6 \\ 0 & 0.99 \end{bmatrix},$$

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

Jacobi and Gauss–Seidel

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.

The Jacobi method

1) *Jacobi method*. We set $A - B = D$, the diagonal part of A, and we obtain the next iteration by solving the diagonal system

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

The Gauss–Seidel method

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

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

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

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

Convergence

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

$$H_{\mathrm{J}} = -D^{-1}(L_0 + U_0) \text{ or } H_{\mathrm{GS}} = -(L_0 + D)^{-1}U_0,$$

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.

Revision – Gershgorin theorem

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

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

Strictly diagonally dominant matrices

Definition 4 (Strictly diagonally dominant matrices)

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

$$|a_{ii}| > \sum_{j \neq i} |a_{ij}|, \quad i = 1..n \qquad \text{(resp.} \quad |a_{jj}| > \sum_{i \neq j} |a_{ij}|, \quad j = 1..n\,\text{)}.$$

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

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 $H_{\mathrm{GS}} = -(L_0 + D)^{-1} U_0$ 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 $|\lambda| \geq 1$ 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. $\square$

The Householder–John theorem

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 λ be an eigenvalue of H, so $H\mathbf{w} = \lambda\mathbf{w}$ holds, where $\mathbf{w} \neq \mathbf{0}$ is an eigenvector. (Note that both λ 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

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

where the bar means complex conjugation.

The Householder–John theorem

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

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

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