Solving Linear Systems: Direct, Iterative, and Optimization Methods

N-20260604

§1.1 The computational problem

Solving $Ax = b$ can be done by direct methods, iterative methods, or optimization methods. Structure decides which method is appropriate.

§1.2 Triangular solves

For lower triangular L, forward substitution gives

$$y_i = \frac{1}{l_{ii}} \left(b_i - \sum_{j<i} l_{ij} y_j \right).$$

For upper triangular U, backward substitution gives

$$x_i = \frac{1}{u_{ii}} \left(y_i - \sum_{j>i} u_{ij} x_j \right).$$

§1.3 Gauss elimination

At step k, with pivot $a_{kk}^{(k)} \neq 0$, set

$$l_{ik} = -\frac{a_{ik}^{(k)}}{a_{kk}^{(k)}}.$$

Then update

$$a_{ij}^{(k+1)} = a_{ij}^{(k)} + l_{ik} a_{kj}^{(k)}, \qquad b_i^{(k+1)} = b_i^{(k)} + l_{ik} b_k^{(k)}.$$

After $n-1$ steps one solves an upper triangular system.

§1.4 Pivoting

Small pivots amplify rounding error. Partial pivoting swaps the largest available entry in the current column into the pivot position. Algebraically,

$$PA = LU.$$

§1.5 LU solution

If $A = LU$, solve

$$Ly = b, \qquad Ux = y.$$

If $PA = LU$, solve

$$Ly = Pb, \qquad Ux = y.$$

§1.6 Cholesky and LDL

For Hermitian positive definite A, use

$$A = LL^H.$$

The improved square-root method uses

$$A = LDL^H,$$

where L is unit lower triangular and D is positive diagonal. Then solve $Ly = b$, $Dz = y$, $L^H x = z$.

§1.7 Tridiagonal chasing

A tridiagonal system can be solved in linear time by a specialized LU factorization. The forward sweep computes modified coefficients; the backward sweep recovers $x_n, x_{n-1}, \ldots, x_1$.

§1.8 Stationary iteration

Rewrite $Ax = b$ as

$$x = Bx + f, \qquad x^{(k+1)} = Bx^{(k)} + f.$$

The error satisfies

$$e^{(k+1)} = Be^{(k)}.$$

Thus convergence for every starting vector is equivalent to

$$\rho(B) < 1.$$

§1.9 Jacobi iteration

With $A = D + L + U$, Jacobi uses

$$x^{(k+1)} = -D^{-1}(L + U)x^{(k)} + D^{-1}b.$$

Componentwise,

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

§1.10 Gauss-Seidel iteration

Gauss-Seidel uses the newest available values:

$$x^{(k+1)} = -(D + L)^{-1} U x^{(k)} + (D + L)^{-1} b.$$

Componentwise,

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

§1.11 Diagonal dominance

If

$$|a_{ii}| > \sum_{j \neq i} |a_{ij}|$$

for every row, then Jacobi and Gauss-Seidel converge. Irreducible weak diagonal dominance gives another standard sufficient condition.

§1.12 SOR iteration

SOR introduces a relaxation parameter α:

$$B_\alpha = (D + \alpha L)^{-1}((1 - \alpha)D - \alpha U).$$

For real symmetric positive definite A, SOR converges iff

$$0 < \alpha < 2.$$

§1.13 Optimization viewpoint

If A is Hermitian positive definite, solving $Ax = b$ is equivalent to minimizing

$$\phi(x) = \frac{1}{2} x^H A x - \mathrm{Re}(b^H x).$$

Its gradient is $Ax - b$.

§1.14 Steepest descent

Let $r_k = Ax_k - b$. Steepest descent uses

$$x_{k+1} = x_k - \alpha_k r_k, \qquad \alpha_k = \frac{(r_k, r_k)}{(Ar_k, r_k)}.$$

It decreases the quadratic energy, but may be slow when $\kappa(A)$ is large.

§1.15 Method map

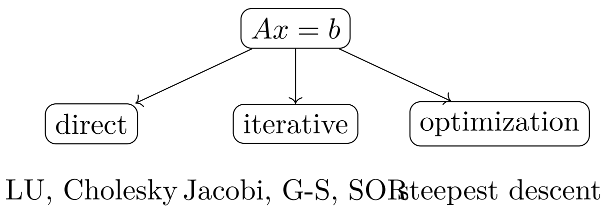

Figure 1.1: Linear systems can be solved by factorization, iteration, or minimization.

§1.16 How the linear-system method is selected

Method	Structure used	Warning
Gaussian elimination	general dense system	watch pivot size
Partial pivoting	row exchange freedom	controls unstable pivots
Cholesky	Hermitian positive definite A	fails without positivity
Chasing method	tridiagonal sparsity	do not destroy the band structure
Jacobi/Gauss–Seidel	$x^{(k+1)} = Bx^{(k)} + c$	need $\rho(B) < 1$
SOR	relaxed Gauss–Seidel	$0 < \alpha < 2$ for SPD convergence
Steepest descent	$\min \frac{1}{2}x^T Ax - b^T x$	zigzags when $\kappa(A)$ is large

Direct methods spend work to factor the matrix and then solve triangular systems; they are preferred when the factorization is affordable or reused for many right-hand sides. Iterative methods are chosen when the matrix is large or sparse, but their convergence is a spectral question about the iteration matrix. The optimization viewpoint is available for symmetric positive definite systems because solving $Ax = b$ is equivalent to minimizing a quadratic. Thus the method map at the end of the note is not a menu of names: it asks what structure the matrix has and which failure mode is most dangerous.