Chapter 0

Introduction to iterative methods and preconditioning

Many important practical problems give rise to systems of linear equations written as the matrix equation

$$\mathbf{A}\mathbf{x} = \mathbf{c}, \tag{0.1}$$

where $\mathbf{A}$ is a given $n \times n$ nonsingular matrix and $\mathbf{c}$ is an n-dimensional vector; the problem is to find an n-dimensional vector $\mathbf{x}$ satisfying equation (0.1).

Such systems of linear equations arise mainly from discrete approximations of partial differential equations. To solve them, two types of methods are normally used: direct methods and iterative methods.

Direct methods approximate the solution after a finite number of floating point operations. Since computer floating point operations can only be obtained to a given precision, the computed solution is usually different from the exact solution. When a square matrix $\mathbf{A}$ is large and sparse, solving $\mathbf{A}\mathbf{x} = \mathbf{c}$ by direct methods can be impractical, and iterative methods become a viable alternative.

Iterative methods, based on splitting $\mathbf{A}$ into $\mathbf{A} = \mathbf{M} - \mathbf{N}$, compute successive approximations $\mathbf{x}^{(t)}$ to obtain more accurate solutions to a linear system at each iteration step t. This process can be written in the form of the matrix equation

$$\mathbf{x}^{(t)} = \mathcal{G}\mathbf{x}^{(t-1)} + \mathbf{g}, \tag{0.2}$$

where an $n \times n$ matrix $\mathcal{G} = \mathbf{M}^{-1}\mathbf{N}$ is the *iteration matrix*. The iteration process is stopped when some predefined criterion is satisfied; the obtained vector $\mathbf{x}^{(t)}$ is an approximation to the solution. Iterative methods of this form are called *linear stationary iterative methods of the first degree*. The method is of the first degree because $\mathbf{x}^{(t)}$ depends explicitly only on $\mathbf{x}^{(t-1)}$ and not on $\mathbf{x}^{(t-2)}, \ldots, \mathbf{x}^{(0)}$. The method is linear because neither $\mathcal{G}$ nor $\mathbf{g}$ depends on $\mathbf{x}^{(t-1)}$, and it is stationary because neither $\mathcal{G}$ nor $\mathbf{g}$ depends on t. In this book, we also consider *linear stationary iterative methods of the second degree*, represented by the matrix equation

$$\mathbf{x}^{(t)} = \mathcal{M}\mathbf{x}^{(t-1)} - \mathcal{N}\mathbf{x}^{(t-2)} + \mathbf{h}. \tag{0.3}$$

The idea of solving large linear systems by iterative methods is not new; it dates back to Gauss [21] in 1823 and Jacobi [31] in 1845. One hundred years later Southwell

[55] in 1946 renewed interest in relaxed variants of the Gauss-Seidel method for numerical solutions of practical and engineering problems; these relaxed variants were the precursor to the SOR (successive overrelaxation) method. In 1950, work by Young [96] led to SOR and its line variants becoming, for many years, the most popular iterative methods for solving large and sparse linear systems arising in many areas of science and engineering. Basic references for stationary iterative methods include books by Varga [61], Young [97], and Hageman-Young [27].

Krylov subspace methods

Research in the past three decades has been dominated by the development of Krylov subspace methods, called *nonstationary iterative methods*. These methods produce approximations $\mathbf{x}_k$ to $\mathbf{A}^{-1}\mathbf{c}$ of the form

$$\mathbf{x}_k \in \mathbf{x}_0 + \mathcal{K}_k(\mathbf{r}_0, \mathbf{A}), \quad k = 1, 2, \ldots, \tag{0.4}$$

where $\mathbf{x}_0$ is any initial guess for the solution of equation (0.1), $\mathbf{r}_0 = \mathbf{c} - \mathbf{A}\mathbf{x}_0$ is the corresponding residual vector, and

$$\mathcal{K}_k(\mathbf{r}_0, \mathbf{A}) = \operatorname{span}\left\{\mathbf{r}_0, \mathbf{A}\mathbf{r}_0, \ldots, \mathbf{A}^{k-1}\mathbf{r}_0\right\} \tag{0.5}$$

is the *Krylov subspace* of dimension k generated by $\mathbf{r}_0$ and $\mathbf{A}$.

To improve the convergence performance of the iteration, *preconditioning* is used, that is, equation (0.1) is replaced by

$$\mathbf{B}\mathbf{A}\mathbf{x} = \mathbf{B}\mathbf{c}. \tag{0.6}$$

This has the same solution as the original system (0.1), but the spectral properties of $\mathbf{BA}$ may be more favorable if the matrix $\mathbf{B}$ is a good approximation to $\mathbf{A}^{-1}$. This modified problem can be solved by generating approximate solutions $\mathbf{x}_1, \mathbf{x}_2, \ldots$ over a different Krylov subspace satisfying

$$\mathbf{x}_k \in \mathbf{x}_0 + \mathcal{K}_k(\mathbf{B}\mathbf{r}_0, \mathbf{B}\mathbf{A}), \quad k = 1, 2, \ldots. \tag{0.7}$$

The matrix $\mathbf{B}$ is called the *left preconditioner* (a similar modification can be done with a right preconditioner or with both left and right preconditioners simultaneously). Thus, preconditioned Krylov subspace methods require choosing the preconditioning matrix $\mathbf{B}$, approximating $\mathbf{A}^{-1}$ well enough so that the algorithm applied to (0.6) will converge faster than for the original system (0.1), and at each iteration solving a linear system of the type

$$\mathbf{B}^{-1}\mathbf{y} = \mathbf{z}. \tag{0.8}$$

Therefore, the preconditioner $\mathbf{B}$ needs to be easily invertible, so that the linear system (0.8) can be solved cheaply.

The combination of preconditioning and Krylov subspace iterations became a central area of research and these new techniques are commonly considered the best methods for solving (0.1) when $\mathbf{A}$ is nonsymmetric.

The *conjugate gradient method* (CG) is a special case of Krylov subspace methods. It is the most efficient and best understood Krylov subspace method, but is suitable only for problems with symmetric positive definite matrices.