# Solving Large Linear Equations

Moo K. Chung

University of Wisconsin-Madison, USA

mkchung@wisc.edu

**Abstract.** In this short lecture, we explain how to solve large matrix systems $\mathbf{y} = A\mathbf{b}$, where $A$ can be on the order of hundreds of thousands. We discuss the challenges associated with large-scale computations and explore strategies such as leveraging sparse matrices, utilizing iterative solvers, applying preconditioning techniques. A practical Matlab example is provided to demonstrate these methods in action.

## 1 Introduction

Directly solving such large systems using standard matrix inversion techniques can be computationally intensive and impractical. Storing large matrices in memory can quickly exhaust available resources. For instance, a dense 300,000 $\times$ 300,000 matrix requires approximately 720 GB of memory (assuming double-precision floating-point numbers, which occupy 8 bytes each), which most end users do not have access to. Often, 300,000 is the scale of numbers in brain imaging. There are approximately 300,000 voxels in the brain (covering both white and gray matter) at a 1mm resolution in MRI scans. Additionally, cortical surface meshes typically contain around 300,000 triangles in each hemisphere. Even if memory is sufficient, the computational time for operations like matrix inversion scales poorly with matrix size. Traditional algorithms for solving $y = A\mathbf{b}$ have computational complexities in quadratic run time.

## 2 Least squares estimation

Consider linear equation

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

where $A$ is an $M \times N$ matrix, potentially with $M, N$ in the hundreds of thousands. $\mathbf{b}$ is a vector or matrix we need to estimate. $\mathbf{y}$ is the given data or observation.

If $A$ is square ($M = N$) and invertible, the solution is obtained as

$$\mathbf{b} = A^{-1}\mathbf{y}.$$

However, when $A$ is not square or is rank-deficient, an exact inverse does not exist, and we resort to generalized solutions using the least-squares solution, which minimizes the squared residual $\|\mathbf{y} - A\mathbf{b}\|_2^2$. The solution is given by the normal equations

$$A^\top A\mathbf{b} = A^\top \mathbf{y},$$

where $A^\top$ is the transpose of $A$. The solution is:

$$\mathbf{b} = (A^\top A)^{-1} A^\top \mathbf{y},$$

assuming $A^\top A$ is invertible.

## 3 Generalized Inverse

When $A$ is ill-conditioned or rank-deficient, $A^\top A$ may not invertible. Then we apply Tikhonov regularization, which minimizes

$$\|\mathbf{y} - A\mathbf{b}\|_2^2 + \lambda \|\mathbf{b}\|_2^2,$$

where $\lambda > 0$ is a regularization parameter. The solution is:

$$\mathbf{b} = (A^\top A + \lambda I)^{-1} A^\top \mathbf{y}.$$

Note the eigenvalues of $A^\top A$ are nonnegative. Adding $\lambda I$ with $\lambda > 0$ shifts all eigenvalues of $A^\top A by \lambda$, thus $A^\top A + \lambda I$ is invertible.

Alternatively, the pseudo-inverse (also called the Moore-Penrose inverse) provides a generalized solution. The pseudo-inverse of a matrix $A$, denoted as $A^+$, is used to find the least-squares solution:

$$\mathbf{b} = A^+ \mathbf{y}.$$

The pseudo-inverse $A^+$ satisfies the following properties

$$AA^+A = A, \quad A^+AA^+ = A^+, \quad (AA^+)^\top = AA^+, \quad (A^+A)^\top = A^+A.$$

When $M > N$ (overdetermined system), the pseudo-inverse is given by

$$A^+ = (A^\top A)^{-1} A^\top.$$

When $M < N$ (underdetermined system), the pseudo-inverse is expressed as

$$A^+ = A^\top (AA^\top)^{-1}.$$

These solutions minimize the squared residual $\|\mathbf{y} - A\mathbf{b}\|_2^2$, which is equivalent to solving the normal equations:

$$A^\top A\mathbf{b} = A^\top \mathbf{y}.$$

The pseudo-inverse of a matrix $A$ can be computed using its Singular Value Decomposition (SVD). For a matrix $A$, the SVD decomposes $A$ as

$$A = U \Sigma V^\top,$$

where $U$ and $V$ are orthogonal matrices, and $\Sigma$ is a diagonal matrix containing the singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, with $r = \mathrm{rank}(A)$. The pseudo-inverse $A^+$ is then expressed as

$$A^+ = V\Sigma^+U^\top,$$

where $\Sigma^+$ is the pseudo-inverse of $\Sigma$ given by

$$\Sigma^+ = \mathrm{diag}\left(\frac{1}{\sigma_1}, \frac{1}{\sigma_2}, \ldots, \frac{1}{\sigma_r}, 0, \ldots, 0\right).$$

The SVD-based pseudo-inverse is computationally stable and avoids the numerical issues associated with direct inversion of $A^\top A$ or $AA^\top$, making it especially suitable for large or ill-conditioned matrices.