ST509 Computational Statistics

Lecture 2: Matrix and Linear Equations

Seung Jun Shin

Department of Statistics
Korea University

E-mail: sjshin@korea.ac.kr

Introduction I

- Linear algebra becomes as fundamental a mathematical tool as calculus.
- Numerical analysts always talk about the solution to

$$\mathbf{A}\mathbf{x} = \mathbf{B}.$$

- Assuming non-singularity of $\mathbf{A}$, the solution is

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{B}.$$

- However, $\mathbf{A}^{-1}$ is a bad operator in computer language such as `R`.

Introduction II

- For handling matrix computation, two things are matter:
  - Storage
  - Multiplication
- $\mathbf{A}$ requires $O(n^2)$ *flops* to save.

  $\mathbf{Ab}$ requires $O(n^2)$ *flops* to compute
- $\mathbf{A} = \mathbf{I} + \mathbf{u}\mathbf{v}^T$ requires $O(n)$ *flops* to save.

$$\begin{aligned}\mathbf{Ab} &= (\mathbf{I} + \mathbf{u}\mathbf{v}^T)\mathbf{b} \\ &= \mathbf{b} + \mathbf{u}\mathbf{v}^T\mathbf{b} \\ &= \mathbf{b} + \mathbf{u}(\mathbf{v}^T\mathbf{b})\end{aligned}$$

  and hence $O(n)$ flops to compute.

Introduction III

- Lower/Upper Triangular matrix
- Diagonal matrix
- Symmetric matrix
- Positive definite matrix: Symmetric and $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0, \forall \mathbf{x} \neq= 0$.
- Orthogonal matrix: $\mathbf{A}\mathbf{A}^T = \mathbf{I}$.

Introduction IV

- Permutation matrix:

  (row permutation)

$$\mathbf{PA} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{11} & a_{12} & a_{13} \end{pmatrix}$$

  (column permutation)

$$\mathbf{AP} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} a_{13} & a_{11} & a_{12} \\ a_{23} & a_{21} & a_{22} \\ a_{33} & a_{31} & a_{32} \end{pmatrix}$$

- Elementary Permutation matrix: Switching two rows/columns only.

Solving System of Equations

- System of Linear Equations:

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

- For a diagonal $\mathbf{A} = \text{Diag}(\mathbf{a})$,

```
x <- b/a
```

- For a Upper-triangular $\mathbf{A}$,

```
x <- backsolve(A, b)
```

- For a Lower-triangular $\mathbf{A}$,

```
x <- forwardsolve(A, b)
```

Gaussian Elemination I

- Note that

$$\mathbf{Ax} = \mathbf{b} \quad \Leftrightarrow \quad \mathbf{BAx} = \mathbf{Bb}$$

  and GE seeks a matrix $\mathbf{B}$ such that $\mathbf{BA}$ is triangular.

ex. Apply GE for

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

Gaussian Elemination II

Step 1 Let $\mathbf{A}^{(0)} = \mathbf{A}$,

$$\mathbf{M}^{(1)}\mathbf{A}^{(0)} = \begin{pmatrix} 1 & 0 & 0 \\ -\frac{a_{21}}{a_{11}} & 1 & 0 \\ -\frac{a_{31}}{a_{11}} & 0 & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & a_{13}^{(1)} \\ 0 & a_{22}^{(1)} & a_{23}^{(1)} \\ 0 & a_{32}^{(1)} & a_{33}^{(1)} \end{pmatrix}$$

Step 2 Then

$$\mathbf{M}^{(2)}\mathbf{A}^{(1)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -\frac{a_{32}^{(1)}}{a_{22}^{(1)}} & 1 \end{pmatrix} \begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & a_{13}^{(1)} \\ 0 & a_{22}^{(1)} & a_{23}^{(1)} \\ 0 & a_{32}^{(1)} & a_{33}^{(1)} \end{pmatrix} = \begin{pmatrix} a_{11}^{(2)} & a_{12}^{(2)} & a_{13}^{(2)} \\ 0 & a_{22}^{(2)} & a_{23}^{(2)} \\ 0 & 0 & a_{33}^{(2)} \end{pmatrix}$$

Step 3 $\underbrace{\mathbf{M}^{(2)}\mathbf{M}^{(1)}}_{\mathbf{B}}\mathbf{A}$ is now upper-triangular, and `backsolve` can be applied.

Gaussian Elemination III

1. Initialize $\mathbf{A}^{(0)} = \mathbf{A}$.
2. For $k = 1, \cdots, n-1$
   2.1 Let $\mathbf{M}^{(k)} = \mathbf{I}$ and its elements for the kth column below the diagonal are
   $$m_{ik}^{(k)} = -a_{ik}^{(k-1)} / a_{kk}^{(k-1)}, \qquad i = k+1, \cdots, n.$$
   or equivalently
   $$\mathbf{M}^{(k)} = \mathbf{I} - \mathbf{m}^{(k)} \mathbf{e}_k^T$$
   where $m_i^{(k)} = a_{ik}^{(k-1)} / a_{kk}^{(k-1)}$ for $i = k+1, \cdots, n$, and 0 for $i \leq k$.
   2.2 Compute
   $$\mathbf{A}^{(k)} = \mathbf{M}^{(k)} \mathbf{A}^{(k-1)}$$
3. Then $\mathbf{BAx} = \mathbf{Bb}$ can be solved by `backsolve` where
   $$\mathbf{B} = \mathbf{M}^{(k)} \cdots \mathbf{M}^{(1)}$$

Algorithm 1: Gaussian elimination algorithm for solving a system of equations.

- (`fn_ge.R` and `02_ge_example.R`)

Gaussian Elemination IV

- GE fails when any of diagonal elements of $A_{kk}^{(k-1)}$ are zero.

ex. Consider

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

- Using permutation matrix $\mathbf{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, we have $\mathbf{PAx} = \mathbf{Pb}$ for which GE works.
- Partial Pivoting suggests to always switch $A_{kk}^{(k-1)}$ to $A_{j_k k}^{(k-1)}$ where

$$j_k = \underset{k \le j \le n}{\operatorname{argmax}} |A_{jk}^{(k-1)}|.$$

That is,

$$\mathbf{A}^{(k)} = \mathbf{M}^{(k)} \mathbf{P}(k, j_k) \mathbf{A}^{(k-1)}$$

where $\mathbf{P}(k, j_k)$ denotes the elementary permutation matrix such that $\mathbf{P}(k, j_k)\mathbf{A}$ switches the jth and j_kth rows of $\mathbf{A}$.

Gaussian Elemination V

- Finally, we have

$$\mathbf{BA} = \mathbf{Bb}$$

where $\mathbf{B} = \mathbf{M}^{(n-1)}\mathbf{P}(n-1, j_{n-1}) \cdots \mathbf{M}^{(2)}\mathbf{P}(2, j_2)\mathbf{M}^{(1)}\mathbf{P}(1, j_1)$.

- GE via partial pivoting yields what we call LU-decomposition

$$\mathbf{PA} = \mathbf{LU} \tag{1}$$

where

$$\begin{aligned} \mathbf{P} &= \mathbf{P}(n-1, j_{n-1}) \cdots \mathbf{P}(2, j_2)\mathbf{P}(1, j_1), \\ \mathbf{U} &= \mathbf{A}^{(n-1)}, \\ \mathbf{L} &= \mathbf{I} + \mathbf{m}_*^{(1)}\mathbf{e}_1^T + \cdots + \mathbf{m}_*^{(n-1)}\mathbf{e}_{n-1}^T, \end{aligned}$$

where

$$\mathbf{m}_*^{(k)} = \mathbf{P}(n-1, j_{n-1}) \cdots \mathbf{P}(k+1, j_{k+1})\mathbf{m}^{(k)}$$

Gaussian Elemination VI

- Under the LU-decomposition (1), we can get $\mathbf{x}$ as follows:

1. We have $\mathbf{PAx} = \mathbf{LUx} = \mathbf{Pb}$
2. Solve $\mathbf{Ly} = (\mathbf{Pb})$ for $\mathbf{y}$.
3. Solve $\mathbf{y} = \mathbf{Ux}$ for $\mathbf{x}$.

Cholesky Decomposition I

- In statistics, $\mathbf{A}$ is mostly positive definite.
- GE assumes p.d. of $\mathbf{A}$, but does not actually use the property.
- Instead of $\mathbf{LU}$-decomposition, a p.d. matrix $\mathbf{A}$ can be factored into $\mathbf{LL}^T$, which exploits the symmetry of the matrix.
- The Cholesky decomposition is inductive as you can see the following example.

Cholesky Decomposition II

ex. Cholesky decomposition of $\mathbf{A} = \begin{pmatrix} 4 & 2 & 2 \\ 2 & 5 & 7 \\ 2 & 7 & 19 \end{pmatrix}$.

1. For $k = 1$, $L_{11} = \sqrt{A_{11}} = 2$.
2. For $k = 2$,

$$\begin{pmatrix} 4 & 2 \\ 2 & 5 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ L_{21} & L_{22} \end{pmatrix} \begin{pmatrix} 2 & L_{21} \\ 0 & L_{22} \end{pmatrix} = \begin{pmatrix} 4 & 2L_{12} \\ 2L_{12} & L_{12}^2 + L_{22}^2 \end{pmatrix}$$

which yields $L_{12} = 1$ and $L_{22} = 2$.

3. For $k = 3$,

$$\begin{aligned} \begin{pmatrix} 4 & 2 & 2 \\ 2 & 5 & 7 \\ 2 & 7 & 19 \end{pmatrix} &= \begin{pmatrix} 2 & 0 & 0 \\ 1 & 2 & 0 \\ L_{31} & L_{32} & L_{33} \end{pmatrix} \begin{pmatrix} 2 & 1 & L_{31} \\ 0 & 2 & L_{32} \\ 0 & 0 & L_{33} \end{pmatrix} \\ &= \begin{pmatrix} 4 & 2 & 2L_{31} \\ 2 & 5 & L_{31} + 2L_{32} \\ 2L_{31} & L_{31} + 2L_{32} & L_{31}^2 + L_{32}^2 + L_{33}^2 \end{pmatrix} \end{aligned}$$

which yields $L_{31} = 1$, $L_{32} = 3$, and $L_{33} = 3$.

Cholesky Decomposition III

- The kth step of CD:

$$\begin{pmatrix} \mathbf{A}^{(k-1)} & \mathbf{a}^{(k)} \\ \mathbf{a}^{(k)T} & A_{kk} \end{pmatrix} = \begin{pmatrix} \mathbf{L}^{(k-1)} & \mathbf{0} \\ \ell^{(k)T} & L_{kk} \end{pmatrix} \begin{pmatrix} \mathbf{L}^{(k-1)} & \ell^{(k)} \\ \mathbf{0}^T & L_{kk} \end{pmatrix}$$

1. Initialize $\mathbf{L}^{(1)} = L_{11} = \sqrt{A_{11}}$.
2. For $k = 2, \cdots, n$
   - 2.1 Update $\ell^{(k)} = (L_{k1}, \cdots, L_{kk-1})^T$ by solving

     $$\mathbf{L}^{(k-1)}\ell^{(k)} = \mathbf{a}^{(k)}$$

     where $\mathbf{a}^{(k)} = (a_{1k}, \cdots, a_{kk})^T$.
   - 2.2 Update L_{kk}

     $$L_{kk}^2 = A_{kk} - \ell^{(k)T}\ell^{(k)}.$$

Algorithm 2: Cholesky Decomposition (square-root algorithm)

Cholesky Decomposition IV

- Computing $\mathbf{x}\mathbf{A}^{-1}\mathbf{x}$ for $\mathbf{x} \in \mathbb{R}^n$
  1. Apply CD to $\mathbf{A} = \mathbf{L}\mathbf{L}^T$.
  2. Solve $\mathbf{L}\mathbf{y} = \mathbf{x}$ for $\mathbf{y}$.
  3. Compute $\mathbf{y}^T\mathbf{y}$ since $\mathbf{y}^T\mathbf{y} = \mathbf{x}^T\mathbf{L}^{-T}\mathbf{L}^{-1}\mathbf{x} = \mathbf{x}^T\mathbf{A}^{-1}\mathbf{x}$.
- Computing $\mathbf{B}\mathbf{A}^{-1}\mathbf{B}$ for $\mathbf{B} \in \mathbb{R}^{n \times m}$
  1. Apply CD to $\mathbf{A} = \mathbf{L}\mathbf{L}^T$.
  2. Solve $\mathbf{L}\mathbf{C} = \mathbf{B}$ for $\mathbf{C}$.
  3. Compute $\mathbf{C}^T\mathbf{C}$ since $\mathbf{C}^T\mathbf{C} = \mathbf{B}^T\mathbf{L}^{-T}\mathbf{L}^{-1}\mathbf{B} = \mathbf{B}^T\mathbf{A}^{-1}\mathbf{B}$.

Reference

- Monahan, J. F. (2011). Numerical Methods of Statistics, Cambridge University Press. Chapter 3.