EECS 275 Matrix Computation

Ming-Hsuan Yang

Electrical Engineering and Computer Science
University of California at Merced
Merced, CA 95344
http://faculty.ucmerced.edu/mhyang

Lecture 10

Overview

- Gaussian elimination
- LU decomposition
- Solving linear systems
- Cholesky decomposition

Reading

- Chapter 20, 21 and 23 of *Numerical Linear Algebra* by Llyod Trefethen and David Bau
- Chapter 3 and 4 of *Matrix Computations* by Gene Golub and Charles Van Loan
- Chapter 3 of *Matrix Analysis and Applied Linear Algebra* by Carl Meyer

Projection

- Recall let $S \subset \mathbb{R}^n$ be a subspace, $P \in \mathbb{R}^{n \times n}$ is the orthogonal projection (projector) onto S if $\text{ran}(P) = S$, $P^2 = P$, and $P^\top = P$
- If $\mathbf{v} \in \text{ran}(P)$, then $P\mathbf{v} = \mathbf{v}$

$$\text{As } \mathbf{v} \in \text{ran}(P), \mathbf{v} = P\mathbf{x}, \text{ and thus } P\mathbf{v} = P^2\mathbf{x} = P\mathbf{x} = \mathbf{v}$$

  ($\mathbf{v}$ lies exactly on its own shadow).
- Likewise, if $\mathbf{v} \in \text{null}(P)$, then $P\mathbf{v} = \mathbf{0}$
- For least squares, $P = A(A^\top A)^{-1}A^\top$, and for $\mathbf{v} \in \text{ran}(A)$, $P\mathbf{v} = \mathbf{v}$

$$\text{As } \mathbf{v} \in \text{ran}(A), \mathbf{v} = A\mathbf{x}, \text{ and thus } P\mathbf{v} = A(A^\top A)^{-1}A^\top A\mathbf{x} = \mathbf{v}$$

- Recall if $\mathbf{u} \in \mathbb{R}^m$, then $\frac{\mathbf{u}\mathbf{u}^\top}{\mathbf{u}^\top\mathbf{u}}$ is an orthogonal projection, and $I - \frac{\mathbf{u}\mathbf{u}^\top}{\mathbf{u}^\top\mathbf{u}}$ is an orthogonal projection to $\text{null}(A)$
- It follows

$$P_A = A(A^\top A)^{-1}A^\top \quad P_{\perp A} = I - A(A^\top A)^{-1}A^\top$$

Quadratic form

- A function $f : \mathbb{R}^n \to R$ has quadratic form

$$f(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}x_i x_j$$

- Often assume A is symmetric

$$\mathbf{x}^\top A\mathbf{x} = \mathbf{x}^\top((A + A^\top)/2)\mathbf{x}$$

where $((A + A^\top)/2)$ is called the symmetric part of A

- Examples:

$$\|B\mathbf{x}\|^2 = \mathbf{x}^\top B^\top B\mathbf{x}$$

$$d_M^2 = (\mathbf{x} - \boldsymbol{\mu})^\top \mathcal{C}^{-1}(\mathbf{x} - \boldsymbol{\mu})$$

$$f(x, y) = ax^2 + bxy + cy^2,\ f(\mathbf{x}) = \mathbf{x}^\top M\mathbf{x},\ M = \begin{bmatrix} a & b/2 \\ b/2 & c \end{bmatrix}$$

- Uniqueness: If $\mathbf{x}^\top A\mathbf{x} = \mathbf{x}^\top B\mathbf{x}$ for $\mathbf{x} \in \mathbb{R}^n$ and $A = A^\top$, $B = B^\top$, then $A = B$
- $\{\mathbf{x}|f(\mathbf{x}) = a\}$ is called a quadratic surface
- $\{\mathbf{x}|f(\mathbf{x}) \leq a\}$ is called a quadratic region

Positive definite

- Recall a matrix $A \in R^{n \times n}$ is positive definite if $\mathbf{x}^\top A\mathbf{x} > 0$ for all nonzero $\mathbf{x} \in R^n$
- Consider a 2-by-2 symmetric case, if

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

is positive definite then

$$\begin{aligned}
\mathbf{x} &= (1,0)^\top & \Rightarrow \quad \mathbf{x}^\top A\mathbf{x} &= a_{11} > 0 \\
\mathbf{x} &= (0,1)^\top & \Rightarrow \quad \mathbf{x}^\top A\mathbf{x} &= a_{22} > 0 \\
\mathbf{x} &= (1,1)^\top & \Rightarrow \quad \mathbf{x}^\top A\mathbf{x} &= a_{11} + 2a_{12} + a_{22} > 0 \\
\mathbf{x} &= (1,-1)^\top & \Rightarrow \quad \mathbf{x}^\top A\mathbf{x} &= a_{11} - 2a_{12} + a_{22} > 0
\end{aligned}$$

- The last two equations imply $\|a_{12}\| \leq (a_{11} + a_{22})/2$ and the largest entry in A is on the diagonal and that is positive
- A symmetric positive definite matrix has a weighty diagonal

Matrix decomposition

- LU decomposition: $A = LU$, applicable to square matrix A
- Cholesky decomposition: $A = U^\top U$ where U is upper triangular with positive diagonal entries, applicable to square, symmetric, positive definite matrix A
- QR decomposition: $A = QR$, where Q is an m-by-m orthogonal matrix and R is an m-by-n upper triangular matrix, applicable to m-by-n matrix A
- Eigendecomposition: $A = QDQ^{-1}$ where D is a diagonal matrix formed from the eigenvalues of A, and columns of Q are the corresponding eigenvectors of A, applicable to square matrix A
- Schur decomposition: $A = QTQ^\top$ where Q is an orthogonal matrix, and T is a block upper triangular matrix, applicable to square matrix A
- Singular value decomposition: $A = U\Sigma V^\top$, where Σ is a non-negative diagonal matrix of singular values, and the columns of U are eigenvectors of $AA^\top$, and V are eigenvectors of $A^\top A$

Gaussian elimination

- For the linear system

$$\begin{array}{rcrcr} 3x_1 & + & 5x_2 & = & 9 \\ 6x_1 & + & 7x_2 & = & 4 \end{array}$$

- Multiply the first equation by 2 and subtract it from the second equation, we get

$$\begin{array}{rcrcr} 3x_1 & + & 5x_2 & = & 9 \\ & - & 3x_2 & = & -14 \end{array}$$

  which is the Gaussian elimination for $A\mathbf{x} = \mathbf{b}$

- In general form, we want to factorize A into a lower triangular and upper triangular matrices, $A = LU$

$$\begin{bmatrix} 3 & 5 \\ 6 & 7 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 3 & 5 \\ 0 & -3 \end{bmatrix}$$

Gaussian elimination (cont'd)

- With $A = LU$, the solution of $A\mathbf{x} = \mathbf{b}$ is found by two step triangular solve process

$$\begin{aligned} A\mathbf{x} &= LU\mathbf{x} = \mathbf{b} \\ L\mathbf{y} &= \mathbf{b} \\ \mathbf{y} &= L^{-1}\mathbf{b} \\ \mathbf{x} &= U^{-1}\mathbf{y} \end{aligned}$$

- Back substitution

$$x_i = (b_i - \sum_{j=i+1}^{n} u_{ij}x_j)/u_{ii}$$

Gauss transformation

- Need a zeroing process for Gaussian elimination, e.g., for $m = 2$, if $x_1 \neq 0$ and $\tau = x_2/x_1$,

$$\begin{bmatrix} 1 & 0 \\ -\tau & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ 0 \end{bmatrix}$$

- More generally, for $\mathbf{x} \in \mathbb{R}^n$ with $x_i \neq 0$, let

$$\boldsymbol{\tau}^\top = (\underbrace{0, \ldots, 0}_{k}, \tau_{k+1}, \ldots, \tau_n), \quad \tau_i = \frac{x_i}{x_k}, \quad i = k+1, \ldots, n$$

where τ_k is the pivot, and define $M_k = I - \boldsymbol{\tau}\mathbf{e}_k^\top$, then

$$M_k\mathbf{x} = \begin{bmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & & 1 & 0 & & 0 \\ 0 & & -\tau_{k+1} & 1 & & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & -\tau_n & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_k \\ x_{k+1} \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} x_1 \\ \vdots \\ x_k \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$(-\tau_{k+1}x_k + x_{k+1} = 0, \tau_{k+1} = x_{k+1}/x_k)$

Gauss transformation (cont'd)

- $M_k = I - \boldsymbol{\tau} \mathbf{e}_k^\top$ is a Gauss transformation
- The first k components of $\boldsymbol{\tau} \in \mathbb{R}^m$ are zero
- The Gaussian transformation is unit lower triangular
- The vector $\boldsymbol{\tau}$ is called the Gauss vector, and the components of $\boldsymbol{\tau}(k+1:n)$ are called multipliers
- Assume $A \in \mathbb{R}^{n \times n}$, Gaussian transformations $M_1, \ldots, M_{n-1}$ can usually be found such that $M_{n-1} \ldots M_2 M_1 A = U$ is upper triangular, e.g.,

$$A = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 10 \end{bmatrix}, \boldsymbol{\tau}_1 = \begin{bmatrix} 0 \\ 2 \\ 3 \end{bmatrix}, M_1 = I - \boldsymbol{\tau}_1 \mathbf{e}_1^\top = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix}$$

Gauss transformation (cont'd)

- Upper triangularizing

$$M_1 A = \begin{bmatrix} 1 & 4 & 7 \\ 0 & -3 & -6 \\ 0 & -6 & -11 \end{bmatrix}$$

likewise

$$M_2 = I - \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix} \mathbf{e}_2^\top = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix}, M_2(M_1 A) = \begin{bmatrix} 1 & 4 & 7 \\ 0 & -3 & -6 \\ 0 & 0 & 1 \end{bmatrix}$$

- From this, we have a matrix $A^{(k-1)} = M_{k-1} \cdots M_1 A$ that is upper triangular in columns 1 to $k-1$
- The multipliers in M_k are based on $A^{(k-1)}(k+1:n,k)$. In particular, we need $A_{kk}^{(k-1)} \neq 0$ to proceed
- The entry A_{kk} must be checked to avoid a zero divide. These quantities are referred to as the pivots, and their relative magnitude turns out to be critically important

LU factorization

- With Gauss transforms $M_1, \ldots, M_{n-1}$ such that $M_{n-1} \cdots M_1 A = U$ is upper triangular
- It is easy to verify that if $M_k = I - \boldsymbol{\tau}^{(k)} \mathbf{e}_k^\top$, then its inverse $M_k^{-1} = I + \boldsymbol{\tau}^{(k)} \mathbf{e}_k^\top$
- More importantly,

  $$A = LU$$

  where

  $$L = M_1^{-1} \cdots M_{n-1}^{-1} \qquad U = M_{n-1} \cdots M_1 A$$

  can be uniquely factorized
- It is clear that L is a unit lower triangular matrix as each M_k^{-1} is unit lower triangular
- Solving n-by-n linear questions with back substitutions via triangular matrices

  $$A\mathbf{x} = LU\mathbf{x} = L\mathbf{y} = \mathbf{b} \Rightarrow \mathbf{y} = L^{-1}\mathbf{b} \quad \mathbf{x} = U^{-1}\mathbf{y}$$

Solving linear system

- Given

$$A = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 10 \end{bmatrix}$$

$$L = M_1^{-1} M_2^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 2 & 1 \end{bmatrix}, \text{ and } \quad U = M_2(M_1 A) = \begin{bmatrix} 1 & 4 & 7 \\ 0 & -3 & -6 \\ 0 & 0 & 1 \end{bmatrix}$$

- If $\mathbf{b} = [1, 1, 1]^\top$, then $\mathbf{y} = [1, -1, 0]^\top$ solves $L\mathbf{y} = \mathbf{b}$, and $\mathbf{x} = [-1/3, 1/3, 0]^\top$ solves $U\mathbf{x} = \mathbf{y}$
- Note L is lower triangular with unit diagonal

$$A = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 10 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix}$$

Pivoting

- Consider LU factorization of A

$$A = \begin{bmatrix} 0.0001 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 10000 & 1 \end{bmatrix} \begin{bmatrix} 0.0001 & 1 \\ 0 & -9999 \end{bmatrix} = LU$$

  with relatively small pivots

- Can get around by interchanging rows with a permutation matrix

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

  then

$$PA = \begin{bmatrix} 1 & 1 \\ 0.0001 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0.0001 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 0.9999 \end{bmatrix}$$

  and the triangular factors are composed of acceptably small entries

Permutation matrix

- Permutation matrix: an identity matrix with rows re-ordered

$$P = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

- P is orthogonal and $P^{-1} = P^\top$
- PA is the row permuted version of A, and AP is the column permuted version of A

Partial pivoting

- To get the smallest possible multipliers, need to have A_{11} to be the largest entry in the first column

$$A = \begin{bmatrix} 3 & 17 & 10 \\ 2 & 4 & -2 \\ 6 & 18 & -12 \end{bmatrix}, \quad E_1 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad E_1 A = \begin{bmatrix} 6 & 18 & -12 \\ 2 & 4 & -2 \\ 3 & 17 & 10 \end{bmatrix}$$

  and the Gauss transformation

$$M_1 = \begin{bmatrix} 1 & 0 & 0 \\ -1/3 & 1 & 0 \\ -1/2 & 0 & 1 \end{bmatrix} \Longrightarrow M_1 E_1 A = \begin{bmatrix} 6 & 18 & -12 \\ 0 & -2 & 2 \\ 0 & 8 & 16 \end{bmatrix}$$

- To get the smallest possible multiplier in M_2, we need to swap rows 2 and 3

Partial pivoting (cont'd)

$$E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \text{ and } M_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1/4 & 1 \end{bmatrix}$$

then

$$M_2 E_2 M_1 E_1 A = \begin{bmatrix} 6 & 18 & -12 \\ 0 & 8 & 16 \\ 0 & 0 & 16 \end{bmatrix}$$

- The particular row interchange strategy is called partial pivoting

  $$M_{n-1} E_{n-1} \dots M_1 E_1 A = U$$
  $$PA = LU$$

  where $P = E_{n-1} \cdots E_1$ (also known as LU decomposition with partial pivoting or LUP decomposition)
- Solve $PA\mathbf{x} = LU\mathbf{x} = P\mathbf{b}$, i.e., first solve $L\mathbf{y} = P\mathbf{b}$ and then $U\mathbf{x} = \mathbf{y}$
- See also complete pivoting and numerical stability issues

LDU factorization

- Asymmetry in LU factorization as the lower factor has 1's on its diagonal
- Can be remedied by factorizing the diagonal entries

$$\begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ 0 & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & u_{nn} \end{bmatrix} = \begin{bmatrix} u_{11} & 0 & \cdots & 0 \\ 0 & u_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & u_{nn} \end{bmatrix} \begin{bmatrix} 1 & \frac{u_{12}}{u_{11}} & \cdots & \frac{u_{1n}}{u_{11}} \\ 0 & 1 & \cdots & \frac{u_{2n}}{u_{22}} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

- $A = LU$ can be scaled so that L and U are unit triangular, i.e., all the diagonal entries of L and U are one, so that $A = LDU$ where D is a diagonal matrix
- When A is symmetric,

$$A = LDL^\top$$

Existence and uniqueness

- An invertible matrix has an LU factorization if and only if all its leading principal minors are nonzeros, i.e., $\det(A_{ii}) \neq 0, \forall i = 1, \ldots, n$
- The factorization is unique if we require that the diagonal of L or U consists of ones
- The matrix has a unique LDU factorization under the same conditions
- For a (not necessarily invertible) matrix, the exact necessary and sufficient conditions under which a not necessarily invertible matrix has an LU factorization are unknown
- Every matrix, square or not, has a LUP decomposition where L and P are square matrices but U has the same shape as A

Cholesky decomposition

- Preserving and exploiting matrix symmetry
- A symmetric matrix A possessing an LU factorization where each pivot is positive, i.e., positive definite
- A is positive definite if and only if A can be uniquely factored as $A = R^\top R$ where R is an upper triangular matrix with positive diagonal entries
- For symmetric matrix

$$A = LDL^\top$$

  where $D = \mathrm{diag}(p_1, p_2, \ldots, p_n)$ and $p_i > 0$
- Setting $R = D^{1/2}L^\top$,

$$A = LD^{1/2}D^{1/2}L^\top = R^\top R$$

- Cholesky factorization:

$$A = R^\top R$$

  where R is an upper triangular matrix with positive diagonal entries, and called Cholesky factor

Cholesky decomposition (cont'd)

- Conversely, if $A = R^\top R$ where R is upper triangular matrix with positive diagonal, then one can factor out the diagonal entries R so that $R = DU$ where U is upper triangular matrix with unit diagonal
- Consequently, $A = U^\top D^2 U$ is the LDU factorization of A, and thus the pivots must be positive
- Suppose $A = R_1^\top R_1 = R_2^\top R_2$, and factor out the diagonal entries with $R_1 = D_1 U_1$ and $R_2 = D_2 U_2$
- It follows that

$$A = U_1^\top D_1^2 U_1 = U_2^\top D_2^2 U_2$$

- The uniqueness of LDU factors forces $U_1 = U_2$ as well as $D_1 = D_2$, and therefore $R_1 = R_2$

Cholesky algorithm

- Let $A = R^\top R = LDL^\top$

$$\begin{aligned}
A &= \begin{bmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{bmatrix} \begin{bmatrix} L_{11} & L_{21} & L_{31} \\ 0 & L_{22} & L_{32} \\ 0 & 0 & L_{33} \end{bmatrix} \\
\begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} &= \begin{bmatrix} L_{11}^2 & L_{21}L_{11} & L_{31}L_{11} \\ L_{21}L_{11} & L_{21}^2 + L_{22}^2 & L_{31}L_{21} + L_{32}L_{22} \\ L_{31}L_{11} & L_{31}L_{21} + L_{32}L_{22} & L_{31}^2 + L_{32}^2 + L_{33}^2 \end{bmatrix}
\end{aligned}$$

$$\begin{aligned}
L_{j,j} &= \sqrt{A_{j,j} - \sum_{k=1}^{j-1} L_{j,k}^2} \\
L_{i,j} &= \frac{1}{L_{j,j}}\left(A_{i,j} - \sum_{k=1}^{j} L_{i,k}L_{j,k}\right) \text{ for } i > j
\end{aligned}$$

- No need for pivoting
- Complexity: $(1/3)n^3$ flops, i.e., $O(n^3)$ (only half of the matrix is processed)