ICS 6N Computational Linear Algebra
Symmetric Matrices and Orthogonal Diagonalization

Xiaohui Xie

University of California, Irvine

xhx@uci.edu

Symmetric matrices

- An $n \times n$ matrix A is **symmetric** if $A^T = A$.
- Component wise: A is symmetric if

$$a_{ij} = a_{ji}$$

for $i, j = 1, 2, \ldots, n$

Matrix Diagonalization

- Matrix A is diagonalizable if there exists a diagonal matrix Λ such that
$$A = P\Lambda P^{-1}$$
- If A can be diagonalized, then $A^k = P\Lambda^k P^{-1}$
- No all matrices can be diagonalized.
- A matrix can be diagonalized if and only if there exists n linearly independent eigenvectors.
- Some special cases:
  - If an nxn matrix A has n distinct eigenvalues, then it is diagonalizable.
  - If A is symmetric, then it is diagonalizable.

Diagonalization of symmetric matrices

Example: diagonalize the matrix

$$A = \begin{bmatrix} 6 & -2 & -1 \\ -2 & 6 & -1 \\ -1 & -1 & 5 \end{bmatrix}$$

- Characteristic equation of A is

$$0 = -\lambda^3 + 17\lambda^2 - 90\lambda + 144 = -(\lambda - 8)(\lambda - 6)(\lambda - 3)$$

  so we have three distinct eigenvalues $\lambda_1 = 8, \lambda_2 = 6, \lambda_3 = 3$.
- Find corresponding eigenvectors

$$v_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, v_2 = \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix}, v_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

- Note that $v_1^T v_2 = 0, v_1^T v_3 = 0, v_2^T v_3 = 0$, i.e., the eigenvectors are mutually orthogonal.

Diagonalization of symmetric matrices

Example: diagonalize the matrix $A = \begin{bmatrix} 6 & -2 & -1 \\ -2 & 6 & -1 \\ -1 & -1 & 5 \end{bmatrix}$

- Further normalize eigenvector to be unit vectors.

$$u_1 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}, u_2 = \begin{bmatrix} -1/\sqrt{6} \\ -1/\sqrt{6} \\ 2/\sqrt{6} \end{bmatrix}, u_3 = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}$$

- Let

$$P = \begin{bmatrix} -1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 0 & 2/\sqrt{6} & 1/\sqrt{3} \end{bmatrix}, D = \begin{bmatrix} 8 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

- $A = PDP^T$, since P is an orthogonal matrix ($P^{-1} = P^T$).

Spectrum theorem

If A is an $n \times n$ symmetric matrix

1. All eigenvalues of A are real
2. A has exactly n real eigenvalues (counting for multiplicity). But this doesn't mean they are distinct
3. The geometric multiplicity of $\lambda = dim(Null(A - \lambda I))$ = the algebraic multiplicity of λ
4. The eigenspaces are **mutually orthogonal**:
   If $\lambda_1 \neq \lambda_2$ are two distinct eigenvalues, then their corresponding eigenvectors v_1, v_2 are orthogonal.

Proof

1. Let λ be an eigenvalue of A with corresponding eigenvector x, so $Ax = \lambda x$ and $Ax^* = \lambda^* x^*$. Then

$$\lambda^* x^T x^* = x^T A x^* = (Ax)^T x^* = \lambda x^T x^*.$$

$\implies \lambda^* = \lambda$, so λ is real.

2. Let x_1 and x_2 be two eigenvectors corresponding to two distinct eigenvalues λ_1 and λ_2.

$$x_1^T A x_2 = (x_1^T A x_2)^T = x_2^T A^T (x_1^T)^T = x_2^T A x_1$$

$$\implies \lambda_2 x_1^T x_2 = \lambda_1 x_1^T x_2 \implies (\lambda_1 - \lambda_2)(x_1^T x_2) = 0$$

Since $\lambda_1 \neq \lambda_2$, $(x_1^T x_2) = 0$ so they are orthogonal.

Orthogonal diagonalization

- If an $n \times n$ matrix A is symmetric, its eigenvectors $v_1, \cdots, v_n$ can be chosen to be orthonormal.
  - If it has n distinct eigenvalues, then the n eigenvectors are orthogonal. Normalize these vectors to make them orthonormal.
  - If an eigenvalue λ has multiplicity greater than 1, find an orthonormal basis of the corresponding eigenspace, $\mathrm{Null}(\mathrm{A} - \lambda \mathrm{I})$, and use vectors in this basis as eigenvectors.
- In this case, $P = \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix}$ is an orthogonal matrix, that is, $P^{-1} = P^T$.
- And A can be orthogonally diagonalized

$$A = P \Lambda P^T$$

Orthogonal diagonalization: an example

Orthogonally diagonalize the matrix $A = \begin{bmatrix} 3 & -2 & 4 \\ -2 & 6 & 2 \\ 4 & 2 & 3 \end{bmatrix}$

- Characteristic equation:

$$0 = -\lambda^3 + 12\lambda^2 - 21\lambda - 98 = -(\lambda - 7)^2(\lambda + 2)$$

- Produce bases for the eigenspaces by solving linear equations:

$$\lambda = 7: \; v_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \; v_2 = \begin{bmatrix} -1/2 \\ 1 \\ 0 \end{bmatrix}; \quad \lambda = -2: \; v_3 = \begin{bmatrix} -1 \\ -1/2 \\ 1 \end{bmatrix}$$

- Apply Gram-Schdmit to produce an orthogonal basis for the eigenspace of $\lambda = 7$.

Orthogonal diagonalization: an example

- Produce bases for the eigenspaces by solving linear equations:

$$\lambda = 7: \; v_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \; v_2 = \begin{bmatrix} -1/2 \\ 1 \\ 0 \end{bmatrix}; \quad \lambda = -2: \; v_3 = \begin{bmatrix} -1 \\ -1/2 \\ 1 \end{bmatrix}$$

- Apply Gram-Schdmit to produce orthogonal bases
  - The component of v_2 orthogonal to v_1 is

$$z_2 = v_2 - \frac{v_2 \cdot v_1}{v_1 \cdot v_1} v_1 = \begin{bmatrix} -1/4 \\ 1 \\ 1/4 \end{bmatrix}$$

  - Normalize v_1, z_2

$$u_1 = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}, u_2 = \begin{bmatrix} -1/\sqrt{18} \\ 4/\sqrt{18} \\ 1/\sqrt{18} \end{bmatrix}$$

  - Normalize v_3 to obtain u_3.
- $A = PDP^T$ where $P = [u_1, u_2, u_3]$ and $D = \mathrm{diag}(7, 7, -2)$.

Application 1: Quadratic Forms

- Any quadratic function of x can be expressed in the form of

$$Q(x) = x^T Ax$$

where x is a vector in R^n and A is an nxn symmetric matrix.

- More explicitly,

$$x^T Ax = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$$

Example

For example,

$$Q(x) = 2x_1^2 + 3x_2^2 + 4x_3^2 + 5x_2x_3 + 6x_1x_2$$

can be written in quadratic form with matrix

$$A = \begin{bmatrix} 2 & 3 & 0 \\ 3 & 3 & \frac{5}{2} \\ 0 & \frac{5}{2} & 4 \end{bmatrix}$$

Optimizing quadratic functions

Consider the following optimization problem:

$$\begin{aligned} \max \;\; & Q(x) = 2x_1^2 + 3x_2^2 + 4x_3^2 \\ \text{subject to } & \|x\| = 1 \end{aligned}$$

Optimizing quadratic functions

Consider the following optimization problem (without cross-product terms):

$$\max \;\; Q(x) = 2x_1^2 + 3x_2^2 + 4x_3^2$$
$$\text{subject to } \|x\| = 1$$

Solution:
Since $2x_1^2 \le 4x_1^2$ and $3x_2 \le 4x_2^2$, we have

$$Q(x) \le 4x_1^2 + 4x_2^2 + 4x_3^2 = 4$$

In addition, we can choose $x_1 = 0$, $x_2 = 0$, $x_3 = 1$ to reach the maximum.

Optimizing quadratic functions

A more general problem:

$$\begin{aligned} \max \;\; & Q(x) = x^T A x \\ \text{subject to } & \|x\| = 1 \end{aligned}$$

Optimizing quadratic functions

A more general problem:

$$\max \; Q(x) = x^T A x$$
$$\text{subject to } \|x\| = 1$$

Solution: Use $A = P \Lambda P^T$ to transform the problem into an easier form.

- $Q(x) = x^T P \Lambda P^T x = (P^T x)^T \Lambda (P^T x)$
- Use $y = P^T x$ to change variables. Convert the problem to

$$\max \; Q(y) = y^T \Lambda y = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2$$
$$\text{subject to } \|y\| = 1$$

- max $x^T A x$ subject to $\|x\| = 1$: $\lambda_{max}\{A\}$
- min $x^T A x$ subject to $\|x\| = 1$: $\lambda_{min}\{A\}$

Optimizing quadratic functions: example

$$\max \ Q(x) = x_1^2 - 8x_1x_2 - 5x_2^2$$
$$\text{subject to } \|x\| = 1$$

Optimizing quadratic functions: example

Solution:

- The matrix of the quadratic form is

$$A = \begin{bmatrix} 1 & -4 \\ -4 & -5 \end{bmatrix}$$

- Orthogonally diagonalize A:

$$P = \begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ -1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix}, D = \begin{bmatrix} 3 & 0 \\ 0 & -7 \end{bmatrix}$$

- Change variables from x to $y = P^T x$, and rewrite the objective function

$$x_1^2 - 8x_1x_2 - 5x_2^2 = x^T Ax = (Py)^T A(Py) = y^T Dy = 3y_1^2 - 7y_2^2$$

- max Q(x) over $\|x\| = 1$ is 3.

Application 2: Principle Component Analysis (PCA)

Problem: Given a set of data points $\{x^{(1)}, x^{(2)}, \cdots, x^{(m)}\}$ in R^n, find the axis along which the data points have maximal variation.

- Assume the data center around origin. If not, subtract the mean from each data point.

Application 2: Principle Component Analysis (PCA)

Problem: Given a set of data points $\{x^{(1)}, x^{(2)}, \cdots, x^{(m)}\}$ in R^n, find the axis along which the data points have maximal variance.

- Use a unit vector u in R^n denote the direction of the axis.
- Project each data point onto u to obtain $\{y^{(1)}, y^{(2)}, \cdots, y^{(m)}\}$, where $y^{(i)} = u^T x^{(i)}$.
- The variance of projected points

$$\sigma^2 = \frac{1}{m}\sum_{i=1}^{m}(y^{(i)})^2 = \frac{1}{m}\sum_{i=1}^{m} u^T x^{(i)}(x^{(i)})^T u = u^T X u$$

where matrix X is defined by

$$X = \frac{1}{m}\sum_{i=1}^{m} x^{(i)}(x^{(i)})^T$$

called covariance matrix.

Application 2: Principle Component Analysis (PCA)

Problem: Given a set of data points $\{x^{(1)}, x^{(2)}, \cdots, x^{(m)}\}$ in R^n, find the axis along which the data points have maximal variance.

- Reformulate the problem into a quadratic optimization problem

$$\max\ u^T X u$$
$$\text{subject to } \|u\| = 1$$

  where matrix $X = \frac{1}{m}\sum_{i=1}^{m} x^{(i)}(x^{(i)})^T$ is the covariance matrix.
- Solution: u is the eigenvector corresponding to the largest eigenvalue of X. The resulting y points are called the first principle component.