Linear Algebra 20
Symmetric Matrices and Quadratic Forms

Wei-Shi Zheng,
wszheng@ieee.org, 2011

1 What Do You Learn from This Note

Do you still remember the following equation or something like that I have mentioned in the class

$$f = \vec{x}^T A \vec{x}. \tag{1}$$

It is exactly a quadratic form function which is going to detail in the following. This connects the eigen-analysis and orthogonal matrix very well.

Basic Concept：symmetric matrix(对称矩阵), quadratic form, quadratic function, positive definitely matrix(正定矩阵), Spectral Decomposition (谱分解), singular value decomposition (奇异值分解)

2 What is symmetric matrix

Definition 1 (symmetric matrix)**.** A square matrix A is said to be symmetric iff $A^T = A$ (or equivalently, $[A]_{ij} = [A]_{ji}$).

Examples: $0_{n\times n}$, I_n, diagonal matrices, etc..

Theorem 2. *All square matrices in* $\mathbb{R}^{n\times n}$ *form a vector space. Denote this vector space as* M_n*.*

Theorem 3. The set of all symmetric matrices of size n, denoted by Sym_n, forms a subspace of M_n.

Proof. Let A and B be symmetric matrices of size n and c a scalar. Then
1. $0_{n\times n}$ is symmetric obviously;
2. $A+B$ is symmetric since $(A+B)^T = A^T + B^T = A + B$;
3. cA is symmetric since $(cA)^T = cA^T = cA$.
So Sym_n is a subspace of M_n. $\square$

3 Eigenvectors for Symmetric Matrix

Theorem 4. Let $A \in \mathrm{Sym}_n(\mathbb{R})$. Also let $\lambda_1, \lambda_2 \in \mathbb{R}$ be distinct eigenvalues of A with eigenvectors $\vec{v}_1$, $\vec{v}_2$ respectively. Then $\vec{v}_1 \perp \vec{v}_2$.

Proof. We have

$$\lambda_1(\vec{v}_1\cdot\vec{v}_2) = \lambda_1\vec{v}_1\cdot\vec{v}_2 = Av_1\cdot\vec{v}_2 = \vec{v}_1^T A^T v_2 = \vec{v}_1^T A v_2 = \vec{v}_1\cdot Av_2 = \vec{v}_1\cdot\lambda_2\vec{v}_2 = \lambda_2(\vec{v}_1\cdot\vec{v}_2).$$

So $(\lambda_1 - \lambda_2)(\vec{v}_1\cdot\vec{v}_2) = 0$. But $\lambda_1 - \lambda_2 \neq 0$ since they are distinct, which forces that $\vec{v}_1\cdot\vec{v}_2 = 0$. $\square$

Theorem 5. Let $A \in \mathrm{Sym}_n(\mathbb{R})$ and λ a complex eigenvalue of A. Then $\lambda \in \mathbb{R}$, that is λ is a real eigenvalue of A.

Proof. Let $v \in \mathbb{C}^n$ be a λ–eigenvector. Then $A\vec{v} = \lambda\vec{v}$. Also, $\overline{A}^T = A$. So

$$\lambda(\overline{v}^T\vec{v}) = \overline{v}^T(\lambda\vec{v}) = \overline{\vec{v}}^T A\vec{v} = \overline{A\vec{v}}^T\vec{v} = (\overline{\lambda\vec{v}}^T)\vec{v} = \overline{\lambda}(\overline{\vec{v}}^T\vec{v}).$$

So $(\lambda - \overline{\lambda})\overline{\vec{v}}^T\vec{v} = 0$. Since $\vec{v} \neq \vec{0}$, it follows that

$$\overline{\vec{v}}^T\vec{v} = \overline{\vec{v}_1}\vec{v}_1 + \cdots + \overline{\vec{v}_n}\vec{v}_n = |\vec{v}_1|^2 + \cdots + |\vec{v}_n|^2 > 0$$

where $(\vec{v}_1 \ \cdots \ \vec{v}_n)^T = \vec{v}$, which forces that $\lambda - \overline{\lambda} = 0$. So λ is real. $\square$

4 Orthogonality & Symmetry

Theorem 6. Let A be any symmetric function and U any orthogonal matrix. Then $U^{-1}AU$ is still a symmetric matrix.

Proof. Notice that $U^T = U^{-1}$. So

$$(U^{-1}AU)^T = U^T A^T (U^{-1})^T = U^{-1}AU.$$

$\square$

We can now prove the major result on symmetric matrices.

Theorem 7 (orthogonally diagonalizable). Let A be a square matrix. Then A is a symmetric matrix if and only if there exists an orthogonal matrix U such that $U^{-1}AU$ is diagonal.

Proof. Firstly, it is easy to verify that if there exists an orthogonal matrix U such that $U^{-1}AU$ is diagonal, then A is symmetric. It is because let the diagonal matrix be D, then $A = UDU^{-1} \in \mathrm{Sym}_n(\mathbb{R})$.

Secondly, or say reversely, if A is symmetric, then we need to prove there exists an orthogonal matrix U such that $U^{-1}AU$ is diagonal. It can be done by induction on n as follows.

1. Step 1: Base Case: $n = 1$. Trivial.

2. Step 2: Inductive Hypotheses: Assume the result holds for $n - 1$.

3. Step 3: Inductive Step: Let λ_1 be an eigenvalue of A, which is real by Theorem 5, and $\vec{u}_1$ a unit λ_1–eigenvector. Then we can extend $\{\vec{u}_1\}$ to a basis of $\mathbb{R}^n$, say $\{\vec{u}_1, \vec{v}_2, \ldots, \vec{v}_n\}$. Furthermore, by applying the Gram–Schmidt process, we obtain an orthonormal basis $\mathcal{U}_1 = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n\}$ of $\mathbb{R}^n$. Define $U_1 = (\vec{u}_1 \ \cdots \ \vec{u}_n) \in \mathrm{O}_n(\mathbb{R})$, which is the matrix from basis $\mathcal{U}_1$ to basis $\mathcal{E}$. Let

   $$F = U_1^{-1}AU_1$$

   It is not hard to see that the first column of F is $(\lambda_1 \ 0 \ \cdots \ 0)$ since $U_1^{-1} = U^T$. So F has the form

   $$F = \begin{pmatrix} \lambda_1 & \vec{0}_{n-1}^T \\ \vec{0}_{n-1} & A' \end{pmatrix},$$

where $A' \in \mathrm{Sym}_{n-1}(\mathbb{R})$. By INDUCTIVE HYPOTHESES, there exists $U' \in \mathrm{O}_{n-1}(\mathbb{R})$ such that $U'^{-1}A'U'$ is diagonal. Define $U_2 = \begin{pmatrix} 1 & \vec{0}_{n-1}^T \\ \vec{0}_{n-1} & U' \end{pmatrix}$, which is an orthogonal matrix. Then

$$U_2^{-1}U_1^{-1}AU_1U_2 = U_2^{-1}FU_2 = \begin{pmatrix} \lambda_1 & \vec{0}_{n-1}^T \\ \vec{0}_{n-1} & U'^{-1}A'U' \end{pmatrix},$$

which is diagonal. So take $U = U_1U_2$, then U is an orthogonal matrix and $U^{-1}AU$ is diagonal.

□

Remarks (Spectral Decomposition,谱分解): If A is orthogonally diagonalizable, and

$$A = PDP^T = \begin{pmatrix} \vec{u}_1 & \cdots & \vec{u}_n \end{pmatrix} \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \begin{pmatrix} \vec{u}_1^T \\ \vdots \\ \vec{u}_n^T \end{pmatrix}$$

, then $A = \lambda_1\vec{u}_1\vec{u}_1^T + \cdots + \lambda_n\vec{u}_n\vec{u}_n^T$ is the spectral decomposition of matrix A.

5 Singular Value Decomposition

Question: If matrix A is not symmetric, it may not be orthogonally diagonalizable. However, A^TA is symmetric and can be orthogonally diagonalized as PDP^-1. Then is it possible to find some orthogonal matrix U and Σ such that

$$PDP^-1 = P\Sigma^TU^TU\Sigma P^-1,\ A = U\Sigma P^-1 = U\Sigma V^T?$$

Answer: Yes. It can be. The Singular Value Decomposition can realize this!!!

Question: What is singular values?

Definition 8. *Let A be $m \times n$ matrix. Then A^TA can be orthogonally diagonalized. Let $\{\vec{v}_1, \cdots, \vec{v}_n\}$ be the orthogonal eigenvector basis of matrix A^TA, where their corresponding eigenvalues are $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \lambda_n \geqslant 0$. Then we say $\sigma_i = \sqrt{\lambda_i}$ are the singular values of matrix A.*

Theorem 9. *Let A be $m \times n$ matrix, where $r = rank(A)$. Then there exists a $m \times m$ orthogonal matrix U and a $n \times n$ orthogonal matrix V and a $m \times n$ matrix $\Sigma = \begin{pmatrix} D & O_{r\times(n-r)} \\ O_{(m-r)\times r} & O_{(m-r)\times(n-r)} \end{pmatrix}$, where $D = diag(\sigma_1, \sigma_2, \cdots, \sigma_r)$ $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r > 0$ are singular values of A, such that*

$$A = U\Sigma V.$$

Proof.

- STEP 1: Let $\{\vec{v}_1, \cdots, \vec{v}_n\}$ be the orthogonal eigenvector basis of matrix A^TA, where their corresponding eigenvalues are $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \lambda_r > \lambda_{r+1} = \lambda_{r+2} = \cdots = \lambda_n = 0$. Then we can prove that $\{A\vec{v}_1, \cdots, A\vec{v}_r\}$ are orthogonal basis of Col$A$.

- STEP 2: Normalize all vectors in the set $\{A\vec{v}_1, \cdots, A\vec{v}_r\}$ to obtain orthonormal basis $\{\vec{u}_1, \cdots, \vec{u}_r\}$, where

$$\vec{u}_i = \frac{1}{||A\vec{v}_i||}A\vec{v}_i = \frac{1}{\sigma_i}A\vec{v}_i.$$

  That is

$$A\vec{v}_i = \sigma_i\vec{u}_i,\ 1 \leqslant i \leqslant r.$$

- STEP 3: Extend $\{\vec{u}_1, \cdots, \vec{u}_r\}$ to be the orthonormal basis of $\mathbb{R}^m$ $\{\vec{u}_1, \cdots, \vec{u}_r, \vec{u}_{r+1}, \cdots, \vec{u}_n\}$.

- STEP 4: Let $U = [\vec{u}_1, \cdots, \vec{u}_n]$ and $V = [\vec{v}_1, \cdots, \vec{v}_n]$. Then they are orthogonal matrices and

$$AV = [A\vec{v}_1, \cdots, A\vec{v}_r, 0\ \cdots\ 0] = [\sigma_1\vec{u}_1, \cdots, \sigma_r\vec{u}_r, 0 \cdots 0].$$

  Then we can easily have

$$A = U\Sigma V^T.$$

□

Example: Compute the singular value decomposition for matrix $A = \begin{pmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{pmatrix}$.

Applications: Image Processing (见演示)

6 Application: Quadratic Form(二次型)

Definition 10 (quadratic form)**.** A real quadratic form $q(x_1, \dots, x_n)$ in n variables $x_1, \dots, x_n$ is a polynomial over $\mathbb{R}$ having the form

$$q(x_1, \dots, x_n) = \sum_{1 \leqslant i \leqslant j \leqslant n} c_{ij} x_i x_j.$$

Examples: 0, $8x_1^2$, $-x_2x_7$, $x_1^2 + x_2^2 + x_3^2$, $x_1x_2 + 12x_5^2$, $2x_3x_4 - 4x_1^2$ are all quadratic forms.

Now we are given a quadratic form $q(x_1, \dots, x_n) = \sum_{1 \leqslant i \leqslant j \leqslant n} c_{ij} x_i x_j$. We can define a symmetric matrix A such that

$$[A]_{ij} = \begin{cases} c_{ij} & \text{if } i = j; \\ \frac{1}{2} c_{ij} & \text{if } i < j; \\ \frac{1}{2} c_{ji} & \text{if } i > j. \end{cases}$$

Then it is easy to verify that $q(x_1, \dots, x_n) = (x_1 \ \cdots \ x_n) A (x_1 \ \cdots \ x_n)^T$.

Conversely, for any $A \in \mathrm{Sym}_n(\mathbb{R})$,

$$q(x_1, \dots, x_n) = (x_1 \ \cdots \ x_n) A (x_1 \ \cdots \ x_n)^T = \sum_{1 \leqslant i \leqslant n} a_{ii} x_i^2 + \sum_{1 \leqslant i < j \leqslant n} 2a_{ij} x_i x_j$$

is a quadratic form, which is called the quadratic form corresponding to A and is denoted by $q_A(x_1, \dots, x_n)$.

Thus, any quadratic form in n variables can be represented by a symmetric matrix of size n uniquely.

Definition 11 (quadratic form function)**.** Let $q(x_1, \dots, x_n)$ be a real quadratic form. Then the function

$$f_q : \mathbb{R}^n \to \mathbb{R}, \quad f_q(\vec{x}) = q(x_1, \dots, x_n),$$

where $\vec{x} = (x_1 \ \cdots \ x_n)^T$, is called the quadratic form function induced by q. If q corresponds to $A \in \mathrm{Sym}_n$, then we write f_A for f_q and $f_A(\vec{x}) = \vec{x}^T A \vec{x}$.

Question: Can we have a more simplified quadratic form for the same quadratic function?

Recall that for any $x \in \mathbb{R}^n$, we have $x = [x]_{\mathcal{E}}$, where $\mathcal{E}$ is the standard basis. Suppose that $\mathcal{B}$ is another basis of $\mathbb{R}^n$. Then $\vec{x} = [\vec{x}]_{\mathcal{E}} = P[\vec{x}]_{\mathcal{B}}$, where $P = [\mathcal{B}]_{\mathcal{E}}$. So for any quadratic form function f_A we have

$$f_A(\vec{x}) = f_A(P\vec{y}) = \vec{y}^T P^T A P \vec{y} = f_{P^T AP}(\vec{y}),$$

where $\vec{y} = [\vec{x}]_{\mathcal{B}}$. So f_A and $f_{P^T AP}$ can be regarded as the same quadratic form function with respect to different bases. Therefore for the sake of convenience, it is quite natural to choose an invertible matrix P such that $f_{P^T AP}$ has a simple form. Also, under the postulates of Euclidean Geometry, only orthogonal coordinate transformation is admitted, that is P is required to be orthogonal.

Now by Theorem 6, we can choose an orthogonal matrix U such that $U^T AU$ is diagonal. So we have

Theorem 12. Let $f_A(\vec{x}) = \vec{x}^T A\vec{x}$ be a quadratic form function. Then there exists $U \in \mathrm{O}_n$ such that

$$f_A(\vec{x}) = f_{U^T AU}(\vec{y}) = \vec{y}^T U^T AU\vec{y} = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2,$$

where $\vec{y} = (y_1 \ \cdots \ y_n)^T$ and $\lambda_1, \ldots, \lambda_n$ are all eigenvalues of A counting multiplicity.

6.1 Positive definite quadratic form

Definition 13 (positive definite quadratic form)**.** Let q be a real quadratic form in n variables. Then q is said to be
1. positive definite (正定的) if $f_q(\vec{x}) > 0$ for all $x \neq \vec{0}$;
2. negative definite (负定的) if $f_q(\vec{x}) < 0$ for all $x \neq \vec{0}$;
3. indefinite (不定的) if the values of $f_q(\vec{x})$ can be both positive and negative.

Theorem 14. Let q_A be the quadratic form corresponding to $A \in \mathrm{Sym}_n(\mathbb{R})$ and $\lambda_1, \ldots, \lambda_n$ all the eigenvalues of A. Then
1. q_A is positive definite iff $\lambda_i > 0$ for all $i = 1, \ldots, n$;
2. q_A is negative definite iff $\lambda_i < 0$ for all $i = 1, \ldots, n$;
3. q_A is indefinite iff there exist i, j such that $\lambda_i > 0$ and $\lambda_j < 0$.

Proof. Let U be an orthogonal matrix such that

$$f_A(\vec{x}) = f_{U^TAU}(\vec{y}) = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2.$$

Then the range of $f_A(\vec{x})$ is the same as that of f_{U^TAU}. So the result follows easily by arguing the range of $\lambda_1 y_1^2 + \cdots + \lambda_n y_n^2$. □

Theorem 15. Let $A \in \mathrm{Sym}_2(\mathbb{R})$ and $\lambda_1 \geqslant \lambda_2$ be eigenvalues of A. Then equation $f_A(\vec{x}) = 1$ represents a conic (圆锥) section, which can be classified as:
1. an ellipse (椭圆) if $\lambda_1 \geqslant \lambda_2 > 0$;
2. a hyperbola (双曲线) if $\lambda_1 > 0 > \lambda_2$;
3. empty set if $0 \geqslant \lambda_1 \geqslant \lambda_2$;
4. a pair of parallel lines if $\lambda_1 > \lambda_2 = 0$.

Proof. Let $U \in \mathrm{O}_2$ be such that

$$f_A(\vec{x}) = f_{U^TAU}(\vec{y}) = \lambda_1 y_1^2 + \lambda_2 y_2^2.$$

Then the original equation turns out to be $\lambda_1 y_1^2 + \lambda_2 y_2^2 = 1$ and the result follows easily. □

7 Geometric Understanding of Symmetric Matrix

Theorem 16. *Let A be a symmetric matrix, then we have*

$$M = \max\{\vec{x}^T A\vec{x}|\ ||\vec{x}|| = 1\} = \textit{the maximum eigenvalue of } A,$$

$$n = \min\{\vec{x}^T A\vec{x}|\ ||\vec{x}|| = 1\} = \textit{the minimum eigenvalue of } A.$$

The Cafe Terrace on the Place du Forum,

by Van Gogh