Week 10-11. Quadratic forms. Principal axes theorem.

Text reference: this material corresponds to parts of sections 5.5, 8.2, 8.3 8.9.

Section 4.1 Motivation and introduction.

Consider an inner product space which is $\mathbf{R}^n$ equipped with inner product $<\vec{u},\vec{v}>=\vec{u}^T A\vec{v}$, where A is a $n\times n$ positive definite matrix.

Recall that the unit ball is the collection of all vectors $\vec{u}$ such that $||\vec{u}||=1$, or equivalently, $<\vec{u},\vec{u}>=1$. In the low-dimension Euclidian spaces ($A=I$) the unit ball *is* a circle in 2D and a sphere in 3D. That is, unit ball in 2D contains all $\vec{u}=(x,y)^T$ such that $x^2+y^2=1$, and unit ball in 3D contains all $\vec{u}=(x,y,z)^T$ such that $x^2+y^2+z^2=1$.

Question: What could be the geometrical shape of a unit ball in 2D and 3D for A other that I (but positive definite)?

In section 4.5 we will prove that ANY unit ball in 2D is an ellipse, and ANY unit ball in 3D is an ellipsoid.

(Note that we are we are talking only about inner product spaces with inner product defined via a positive definite matrix.)

Definition 1. *Quadratic form* in variables x_1,x_2,x_3 is a linear combination of squares x_1^2,x_2^2,x_3^2 and cross terms x_1x_2,x_1x_3,x_2x_3.

More general, *quadratic form* in variables $x_1,x_2,...,x_n$ is a linear combination of squares $x_1^2,x_2^2,...,x_n^2$ and cross terms x_ix_j, where $i\neq j$ and $1\leq i,j\leq n$.

For example, general quadratic form in two variables is

$$q=ax_1^2+bx_1x_2+cx_2^2,$$

where a,b,c are some numbers, called coefficients. Observe that in 2D quadratic form can be written as $q=\vec{v}^T A\vec{v}$, where $\vec{v}=(x_1,x_2)^T$ and $A=\begin{bmatrix} a & b/2 \\ b/2 & c \end{bmatrix}$.

Similarly, for arbitrary n one can introduce $\vec{v}=(x_1,x_2,...,x_n)^T$ and rewrite quadratic form as $q=\vec{v}^T A\vec{v}$ with some symmetric matric A ($A^T=A$).

Remark: There is no requirement for A other than being symmetric in order to define a quadratic form. If A happen to be positive definite then such quadratic form $q=\vec{v}^t A\vec{v}$ defines the norm of $\vec{v}$ (namely, $q=||\vec{v}||^2$) in the inner product space with inner product defined by A: $<\vec{u},\vec{v}>=\vec{u}^T A\vec{v}$.

We will identify $n\times 1$-matrix X with $\vec{v}=(x_1,x_2,...,x_n)$.

Thus we will write general quadratic form as $q=X^TAX$ instead of $q=\vec{v}^T A\vec{v}$.

Problem 1. Let $q=x_1^2+x_1x_2+x_2^2$.

1) Find A such that $q=X^TAX$, where $X=(x_1,x_2)^T$.

2) Denote $Y=(y_1,y_2)^T$. Let $X=MY$, where $M=\frac{1}{\sqrt{2}}\begin{bmatrix} -1 & -1 \\ 1 & -1 \end{bmatrix}$.

Rewrite q in Y-variables and find B such that $q=Y^TBY$, where $\vec{Y}=(y_1,y_2)^T$.

Solution:

1) $A=\begin{bmatrix} 1 & 1/2 \\ 1/2 & 1 \end{bmatrix}$

2) We have $x_1=\frac{1}{\sqrt{2}}(-y_1-y_2)$, $x_2=\frac{1}{\sqrt{2}}(y_1-y_2)$. Thus $q=x_1^2+x_1x_2+x_2^2=y_1^2/2+(3/2)y_2^2$. Therefore,

$B=\begin{bmatrix} 1/2 & 0 \\ 0 & 3/2 \end{bmatrix}$.

In this problem a change of variables (from (x_1, x_2) to (y_1, y_2)) was made such that in new variables quadratic form has **no cross terms** (matrix B is diagonal). This process is called *diagonalization of quadratic form.*

Question: Is it always possible to find a change of variables such that in new variables quadratic form has no cross terms?

In order to answer this question we have to review some facts about diagonalization of a square matrix and take into account that matrices defining quadratic forms are symmetric. We will answer the question in section 4.5 of this handout.

Section 4.2 Similarity and diagonalization.

Definition 2.

Two $n \times n$ matrices are *similar*, $A \sim B$, if $A = P^{-1}BP$ for some invertible matrix P.

Theorem 1. If matrices A and B are similar then they have the same determinant, rank, trace, characteristic polynomial and eigenvalues.

(Trace of a matrix, tr$A$ is the sum of its diagonal elements. An important property of trace is trAB=trBA, which can be verified by direct computation.)

Theorem 2. If $n \times n$ matrix has n distinct eigenvalues then it is similar to a diagonal matrix with the eigenvalues on the diagonal: $A \sim D$.

Proof: We have $AX_k = \lambda_k X_k$ for $k = 1, 2, .., n$. Compose matrix P whose columns are the eigenvectors: $P = [X_1|X_2|...|X_n]$. Observe that $AP = PD$, where D is a diagonal matrix with the eigenvalues $\lambda_1, ..., \lambda_n$ on the diagonal. It can be shown that the eigenvectors X_1,..,X_n are linearly independent, and P is invertible. Thus $P^{-1}AP = D$. $\square$

Let λ be an eigenvalue of A. Denote by mult(λ) the multiplicity of the eigenvalue, and by E_λ corresponding eigenspace. Note that dimE_λ $\leq$mult(λ). If all eigenvalues are distinct then all eigenvalues have multiplicity one and dimE_λ =mult(λ) $= 1$ for all eigenvalues.

Theorem 3. Let dimE_λ =mult(λ) for each eigenvalue of matrix A. Then A is similar to a diagonal matrix.

Proof: If dimE_λ =mult(λ) $= r$ then there exists basis of r (linearly independent) vectors $X_1, .., X_r$ in this eigenspace, and they can be taken as columns of matrix P. Since the sum of multiplicities of all eigenvalues is n, we will collect n linearly independent vectors and construct an invertible matrix P such that $P^{-1}AP = D$. Multiple eigenvalues will repeat on the diagonal of D according to their multiplicities. $\square$

Section 4.3 Symmetric matrices.

Definition 3. An $n \times n$ matrix A is *symmetric* if $A^T = A$.

Observe that some matrices with real elements may have complex eigenvalues, for example, $\begin{bmatrix} 0 & -2 \\ 2 & 0 \end{bmatrix}$ has characteristic equation $\lambda^2 + 4 = 0$, thus imaginary eigenvalues $\pm 2i$. It is interesting to know that:

Theorem 4. All eigenvalues of a symmetric matrix are real numbers.

Proof: We restrict ourselves to the case $n = 2$. For $n \geq 3$ one needs to operate with complex numbers, so we leave this case for now.

Let $A = \begin{bmatrix} a & b \\ b & d \end{bmatrix}$. Then characteristic equation is $\lambda^2 - (a+c)\lambda + (ac - b^2) = 0$. Its determinant $d = (a-c)^2 + 4b^2 \geq 0$. Since the determinant is never negative the matrix can't have complex eigenvalues. All its eigenvalues are real numbers. $\square$.

Theorem 5. If $n \times n$-matrix A is symmetric then eigenvectors of A corresponding to distinct eigenvalues are orthogonal.

Proof:

1. Let X, Y be any n-vectors ($n \times 1$-matrices) and $A = A^T$.
We have $(AX) \cdot Y = (AX)^T Y = X^T A^T Y = X^T AY = X \cdot (AY)$.
2. Let $AX = \lambda X$, and $AY = \mu Y$, and $\lambda \neq \mu$.
We have $\lambda(X \cdot Y) = (\lambda X) \cdot Y = AX \cdot Y =$(use 1.)$= X \cdot (AY) = X \cdot (\mu Y) = \mu X \cdot Y$.
Thus $(\lambda - \mu)(X \cdot Y) = 0$ but $\lambda \neq \mu$, so $X \cdot Y = 0$. Therefore the vectors are orthogonal. $\square$

Theorem 6. An $n \times n$-matrix A is symmetric if and only if it has an orthogonal set of n eigenvectors.

Remark: This orthogonal set of eigenvectors can be converted to an orthonormal set by normalization. Let matrix P be formed from the orthonormal eigenvectors of matrix A. Then P has property $P^T P = I$ and is invertible. Thus $P^{-1} = P^T$. Matrix P is so called an *orthogonal* matrix.

Definition 4. Matrix M is called *orthogonal* if $M^{-1} = M^T$.

(The columns of an orthogonal matrix form an orthonormal set of linearly independent vectors.)

Theorem 7. (Principal Axes Theorem).

An $n \times n$-matrix A is symmetric if and only if $P^T AP = D$ for some orthogonal matrix P and diagonal matrix D.

Definition 5. The orthonormal set of eigenvectors of a symmetric matrix A (i.e. the columns of matrix P) is called the set of *principle axes* for corresponding quadratic form $q = \vec{v}^T A\vec{v}$.

The geometrical meaning of this term will be seen in section 4.5 of this handout.

Section 4.4 Positive definite matrices.

Recall that we had two definitions of positive definite matrices. Now we are ready to prove that the two definitions are equivalent.

Theorem 8. The following two statements are equivalent:

(I) Matrix A is symmetric and all its n eigenvalues are positive;

(II) quadratic form $q = X^T AX > 0$ for all $X \neq 0$ in $\mathbf{R}^n$.

Proof:

1. By Principal Axes Theorem, A is symmetric means $P^T AP = D$, or $A = PDP^T$.
2. Consider $q = X^T AX = X^T(PDP^T)X = X^T(P^T)^T DP^T X = (P^T X)^T DP^T X = Y^T DY$, where $Y = P^T X$. Thus $q = Y^T DY = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots \lambda_n y_n^2$. This expression is always positive because all eigenvalues λ_k are positive and not all y_k are zeroes (since $X \neq 0$). $\square$

Section 4.5 Diagonalization of quadratic forms.

We use notations $X = (x_1, x_2, ..., x_n)^T$, $Y = (y_1, y_2, ..., y_n)^T$.

Theorem 9. Let $q = X^T AX$ be a quadratic form in variables $x_1, x_2, ..., x_n$, and $A^T = A$. Let P be an orthogonal matrix such that $P^T AP = D$, where D is a diagonal matrix. Define new variables $y_1, y_2, ..., y_n$ via the formula $X = PY$. Then quadratic form q has no cross term in new variables, in other words $q = Y^T DY = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots \lambda_n y_n^2$.

Proof: Substitute $X = PY$ into $q = X^T AX$ and use $P^T AP = D$ to obtain $q = X^T AX = (PY)^T A(PY) = Y^T P^T APY = Y^T DY$. The existence of P follows from the Principle Axes Theorem. Recall that the columns of P are orthonormal eigenvectors of A corresponding to the eigenvalues $\lambda_1, .., \lambda_n$. □

Problem 2. Identify the change of variables that will diagonalize the quadratic form and rewrite the form w.r.t these new variables

$q = 7x_1^2 + x_2^2 + x_3^2 + 8x_1x_2 + 8x_1x_3 - 16x_2x_3$.

Solution:

Rewrite the form as $q = X^T AX$, $X = (x_1, x_2, x_3)^T$ and $A = \begin{bmatrix} 7 & 4 & 4 \\ 4 & 1 & -8 \\ 4 & -8 & 1 \end{bmatrix}$.

This matrix has the characteristic equation $(\lambda - 9)^2(\lambda + 9) = 0$, that is eigenvalues $\lambda_1 = 9$ of multiplicity 2 and $\lambda_3 = -9$ of multiplicity 1.

Corresponding eigenvectors are $\vec{f_1} = (2, 2, -1)^T$, $\vec{f_2} = (2, -1, 2)^T$, $\vec{f_3} = (-1, 2, 2)^T$. This set of eigenvectors is orthogonal. Each vector has length 3. Thus corresponding orthonormal set is $\vec{e}_j = \frac{1}{3}\vec{f}_j$, $j = 1, 2, 3$. Compose matrix P whose columns are the orthonormal eigenvectors. $P = \frac{1}{3}\begin{bmatrix} 2 & 2 & -1 \\ 2 & -1 & 2 \\ -1 & 2 & 2 \end{bmatrix}$. This matrix is orthogonal: $P^{-1} = P^T$. Accidently this matrix happen to be symmetric as well - normally it would not be the case.

The change of variables is defined by $X = PY$ or $Y = P^{-1}X = P^T X$. That is

$x_1 = \frac{1}{3}(2y_1 + 2y_2 - y_3)$,
$x_2 = \frac{1}{3}(2y_1 - y_2 + 2y_3)$,
$x_3 = \frac{1}{3}(-y_1 + 2y_2 + 2y_3)$.

Equivalently:

$y_1 = \vec{e}_1 \cdot X = \frac{1}{3}(2x_1 + 2x_2 - x_3)$,
$y_2 = \vec{e}_2 \cdot X = \frac{1}{3}(2x_1 - x_2 + 2x_3)$,
$y_3 = \vec{e}_3 \cdot X = \frac{1}{3}(-x_1 + 2x_2 + 2x_3)$.

Substitution of $X = PY$ into the form $q = X^T AX$ gives the diagonal form: $q = 9y_1^2 + 9y_2^2 - 9y_3^2$.

In 2D we have the following statement.

Theorem 10. Consider $q = ax_1^2 + bx_1x_2 + cx_2^2$, where $a \neq 0$, $b \neq 0$, $c \neq 0$. There is a counter-clockwise rotation of the coordinate axes about the origin such that in the new coordinate system q has no cross terms.

Proof:

1. Let (x_1, x_2) be coordinate of some vector $\vec{v}$ in the standard basis $E_1 = (1, 0)^T$, $E_2 = (0, 1)^T$. Introduce new basis $F_1 = (\cos\theta, \sin\theta)^T$, $F_2 = (-\sin\theta, \cos\theta)^T$ and let (y_1, y_2) be coordinate of the same vector $\vec{v}$ in the new basis, that is

$$\vec{v} = (x_1, x_2)^T = x_1E_1 + x_2E_2 = y_1F_1 + y_2F_2.$$

This means that the *matrix of coordinate transformation* from basis (E_1, E_2) to basis (F_1, F_2) is $M = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ and $X = MY$, where $X = (x_1, x_2)$, $Y = (y_1, y_2)$ (see section 2.3 from handout 2 (Week 2-3)).

Matrix M makes counterclockwise θ-rotation of the standard basis (coordinate axes) about the origin.

2. From $x_1E_1 + x_2E_2 = y_1F_1 + y_2F_2$ we have

$$x_1 = y_1\cos\theta - y_2\sin\theta, \quad x_2 = y_1\sin\theta + y_2\cos\theta.$$

Substitute these expressions into $q = ax_1^2 + bx_1x_2 + cx_2^2$. The coefficient of the cross term y_1y_2 becomes $(c-a)\sin(2\theta) + b\cos(2\theta)$. To make this coefficient zero we chose θ such that

$$\tan(2\theta) = \frac{b}{a-c}, \quad a \neq c,$$

or $\theta = \pi/4$ for $a = c$. $\square$

Theorem 11. The graph of equation $ax_1^2 + bx_1x_2 + cx_2^2 = 1$ is an ellipse for $b^2 - 4ac < 0$ or a hyperbola for $b^2 - 4ac > 0$.

Proof: Rewrite $q = X^T AX$, where $A = \begin{bmatrix} a & b/2 \\ b/2 & c \end{bmatrix}$.

Note that $A = PDP^T$, thus $\det A = \det D$ or $ac - b^2/4 = \lambda_1\lambda_2$.

But we know that canonical equation of ellipse $\frac{x^2}{\alpha^2} + \frac{y^2}{\beta^2} = 1$ corresponds to the case $\lambda_1\lambda_2 = \alpha^{-2}\beta^{-2} > 0$, thus $ac - b^2/4$ must be positive for the ellipse case. Similarly, canonical equation of hyperbola $\frac{x^2}{\alpha^2} - \frac{y^2}{\beta^2} = 1$ corresponds to the case $\lambda_1\lambda_2 = -\alpha^{-2}\beta^{-2} < 0$, thus $ac - b^2/4$ must be negative for the hyperbola case.

Problem 3. Determine whether the curve $2x^2 + 5xy + y^2 = 1$ is an ellipse or hyperbola. Find the angle between each of its two axes of symmetry and the horizontal OX axis. Sketch the graph of the curve.

Solution: 1. $b^2 - 4ac = 25 - 8 > 0$ thus the curve is hyperbola.

2. the angle θ is found from $\tan(2\theta) = \frac{b}{a-c} = 5$, thus $\theta = 0.5\arctan(5) \approx 40$ (deg); the second axis forms an angle $90 + 0.5\arctan(5) \approx 130$(deg) or (-50 deg).

3. the asymptotes are found by substitution $y = mx$ into the equation with zero RHS:

$2x^2 + 5xy + y^2 = 0$; then we have equation for m: $2 + 5m + m^2 = 0$ and find the slopes of the asymptotes: $m = (-5 \pm \sqrt{17})/2$, or $m_1 \approx -4.56$ (or -77 deg w.r.to horizontal) and $m_2 \approx -0.44$ (or -23 deg w.r.to horizontal).

Note that $\frac{(-23) + (-77)}{2} = -50$, which confirms the axis of the symmetry: asymptotes must be symmetric w.r.to the axes of the symmetry of the hyperbola.

4. To sketch the hyperbola we first plot the asymptotes $y = m_1x$ and $y = m_2x$. We plot the symmetry exes: two lines forming angles θ and $90 + \theta$ with the horizontal axis.

Then we identify few points which lie on the curve $2x^2 + 5xy + y^2 = 1$, for instance: $x = 0$, $y = \pm 1$.

Then we plot the curve according to its symmetry, asymptotes and identified points:

Theorem 12. The unit ball in $\mathbf{R}^2$ with inner product $< \vec{u}, \vec{v} >= \vec{u}^T A\vec{v}$ (defined by a positive definite matrix A) is an ellipse.

Proof: The unit ball contains all vectors X such that $X^T AX = 1$. The graph of equation $X^T AX = 1$, where A has positive eigenvalues, is an ellipse. Note that eigenvectors of A define the axes of symmetry of the ellipse. That explains the name: principal axes of quadratic form $q = X^T AX$.

Exercise Set 6 for Quiz on Wed March 19 (or March 24?).

1 and 2. Give a definition and an example of:
- -unit ball in an inner product space;
- -quadratic form in variables $x_1, ...x_n$;
- -two similar matrices;
- -symmetric matrix;
- -orthogonal matrix;
- -positive definite matrix;
- -characteristic polynomial for a square matrix; (M2050)
- -eigenvalue of a matrix; multiplicity of an eigenvalue;(M2050)
- -eigenspace;
- - matrix of rotation about the origin in 2D;
- - set of principal axes for a quadratic form;
- -ellipse;
- -hyperbola;

3. Let $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$. Find eigenvalues and corresponding eigenvectors. Find an orthogonal matrix P such that $P^T AP$ is diagonal. Find the set of principle axes for quadratic form $q = X^T AX$.

4. Identify the change of variables that will diagonalize the quadratic form and rewrite the form w.r.to these new variables.
(a) $q = x_1^2 + 4x_1x_2 + x_2^2$
(b) $q = x_1^2 - 2x_1x_2 + 2x_2x_3 + x_3^2$

5. Determine whether the following curve is an ellipse or hyperbola. Find the angle between each of its two axes of symmetry and the horizontal OX axis. Sketch the graph of the curve.
(a) $3x^2 + 4xy + y^2 = 1$
(b) $3x^2 + 4xy + 2y^2 = 1$

6. Prove that eigenvectors of a symmetric matrix corresponding to distinct eigenvalues are orthogonal.

7. Consider an inner product space which is $\mathbf{R}^2$ equipped with inner product $< \vec{u}, \vec{v} >= \vec{u}^T A\vec{v}$, where A is a 2×2 positive definite matrix. Explain in details why the unit ball in this space is an ellipse.

8. Consider an inner product space which is $\mathbf{R}^3$ equipped with inner product $< \vec{u}, \vec{v} >= \vec{u}^T A\vec{v}$, where A is a 2×2 positive definite matrix. Explain in details why the unit ball in this space is an ellipsoid.