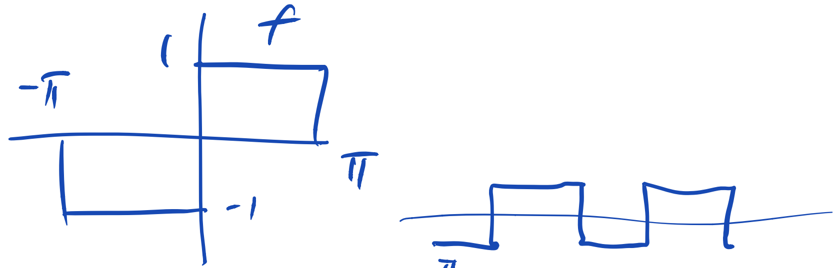
11/11



$$a_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos kt\,dt$$

$$= \frac{1}{\pi}\left[-\int_{-\pi}^{0}\cos kt\,dt + \int_{0}^{\pi}\cos kt\,dt\right]$$

Math 2270-6

Notes of 11/26/19

7.2 Quadratic Forms

Positive Definiteness

- **Definition:** A **quadratic form** on $\mathbb{R}^n$ is a function Q defined on $\mathbb{R}^n$ whose value at a vector $\mathbf{x}$ in $\mathbb{R}^n$ can be computed by an expression of the form

$$Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$$

  where A is an $n \times n$ symmetric matrix. The matrix A is called the **matrix of the quadratic form**.

- In the case of $n = 1$ a quadratic form is a function of the form

$$Q(x) = ax^2$$

- In terms of the entries of A and $\mathbf{x}$ the quadratic form is given by

$$Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j.$$

$$\|x\|^2 = x^T I x$$

- Example 1a. Suppose

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$[x_1 \ x_2][4x_1 \ 3x_2]$ $\downarrow$

$$Q(x) = 4x_1^2 + 3x_2^2$$

- Example 1b. Suppose

$$A = \begin{bmatrix} 3 & -2 \\ -2 & 7 \end{bmatrix}$$

$$Q(x) = 3x_1^2 - 2x_1x_2 - 2x_2x_1 + 7x_2^2$$
$$= 3x_1^2 - 4x_1x_2 + 7x_2^2$$

- Example 2. Suppose for $\mathbf{x}$ in $\mathbb{R}^3$ we have

$$Q(\mathbf{x}) = 5x_1^2 + 3x_2^2 + 2x_3^2 - x_1x_2 + 8x_2x_3.$$

What is the matrix of this quadratic form?

$$A = \begin{bmatrix} 5 & -1/2 & 0 \\ -1/2 & 3 & 4 \\ 0 & 4 & 2 \end{bmatrix}$$

Quadratic Forms and Eigenvalues

- We know that every symmetric matrix has an orthogonal similarity transform to diagonal form:
$$D = P^T A P \tag{1}$$
where D is a diagonal matrix with the eigenvalues along the diagonal, and P is an orthogonal matrix whose columns are the corresponding eigenvectors of A.
- The equation (1) can be rewritten as
$$A = PDP^T.$$
- Thus
$$\mathbf{x}^T Ax = \mathbf{x}^T PDP^T \mathbf{x} = \mathbf{y}^T D\mathbf{y}$$
where
$$\mathbf{y} = P^T\mathbf{x} \qquad \text{or} \qquad \mathbf{x} = P\mathbf{y}.$$
- Thus any quadratic form can be written as a quadratic form in terms of the coordinate vector of $\mathbf{x}$ with respect to the orthogonal basis of eigenvectors. The matrix of that quadratic form is the diagonal matrix of eigenvalues.

- This is the contents of what the textbook calls **The Principal Axes Theorem**. Let A be an $n \times n$ symmetric matrix. Then there is an orthogonal change of variables, $\mathbf{x} = P\mathbf{y}$, that transforms the quadratic form into a quadratic form
$$\mathbf{y}^T D\mathbf{y}$$
with no cross product terms.
- Note that saying "with not cross product terms" is equivalent to saying that D is diagonal. We've seen of course, that P is the matrix of eigenvectors, and D is the diagonal matrix of eigenvalues.

Positive Definiteness

- We come now to a key definition:
- A quadratic form $Q = \mathbf{x}^T A\mathbf{x}$, and its matrix A, is

a. **positive definite** if $\mathbf{x}^T A\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$.

b. **negative definite** if $\mathbf{x}^T A\mathbf{x} < 0$ for all $\mathbf{x} \neq \mathbf{0}$.

c. **indefinite** if $\mathbf{x}^T A\mathbf{x}$ assumes both positive and negative values.

d. **positive semidefinite** if $\mathbf{x}^T A\mathbf{x} \geq 0$ for all $\mathbf{x}$.

e. **negative semidefinite** if $\mathbf{x}^T A\mathbf{x} \leq 0$ for all $\mathbf{x}$.

- Note that if D is diagonal, with the ii entry being λ_i, then

$$\mathbf{x}^T D\mathbf{x} = \sum_{i=1}^{n} \lambda_{ii}\mathbf{x}_i^2. \quad = \quad z^T A z$$

- Thus, by the principal axes theorem we can think of these concepts in terms of the eigenvalues of A.

- A quadratic form $Q = \mathbf{x}^T A\mathbf{x}$, and its matrix A, is

a. positive definite if all e.v. > 0

b. negative definite if all e.v. < 0

c. indefinite if some negative and some positive

d. positive semidefinite if all e.v. ≥ 0

e. negative semidefinite if all e.v. ≤ 0

- Examples:

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad p.d.$$

$$A = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \qquad n.d.$$

$$A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \qquad \text{indef}$$

$$A = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \qquad p.s.d.$$

$$A = \begin{bmatrix} 0 & 0 \\ 0 & -11 \end{bmatrix} \qquad n.s.d.$$

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \qquad \text{not symm.}$$

$$A = \begin{bmatrix} 4 & -1 & -1 & -1 \\ -1 & 4 & -1 & -1 \\ -1 & -1 & 4 & -1 \\ -1 & -1 & -1 & 4 \end{bmatrix} \qquad p.d.$$

- $A = B^T B$ where $B = m \times n$.

$$A^T = (B^T B)^T = B^T B = A$$

$$x^T A x = x^T B^T B x = y^T y \geq 0 \qquad y = Bx$$

p.s.d.

- We now come to the reason why positive definite matrices are important:

Positive Definite Matrices occur when minimizing scalar valued functions.

- To build up to understanding this fact recall a little bit of Calculus.
- Suppose we want to find extreme values of a function $f : \mathbb{R} \longrightarrow \mathbb{R}$.
- We know that extreme values may occur at boundary points of intervals, singular points where the derivative does not exist, and **stationary points** where the derivative is zero.

In the remainder of these notes we ignore singular and boundary points and focus on stationary points.

- Thus, in order to have an extreme value at $\hat{x}$ we must have that $\hat{x}$ is stationary, i.e.,

$$f'(\hat{x}) = 0.$$

- If $\hat{x}$ is in fact stationary, then we have
  - a minimum value if $f''(\hat{x}) > 0$,
  - a maximum value if $f''(\hat{x}) < 0$
- If $f''(x) = 0$ the second derivative test is inconclusive and we may have a minimum, a maximum, or a saddle point.

- in Calc III we considered scalar valued functions of a vector variable. Suppose

$$f : \mathbb{R}^n \longrightarrow \mathbb{R}.$$

- Then the **gradient** of f is the vector of first order partial derivatives

$$\nabla f = [\ \frac{\partial f}{\partial x_1} \quad \frac{\partial f}{\partial x_1} \quad \ldots \quad \frac{\partial f}{\partial x_n}\ ]^T$$

- The **Hessian** of f is the symmetric (!) matrix of second order partial derivatives:

$$\nabla^2 f = \begin{bmatrix} \frac{\partial^2}{\partial x_1 \partial x_1} f & \frac{\partial^2}{\partial x_1 \partial x_2} f & \frac{\partial^2}{\partial x_1 \partial x_3} f & \cdots & \frac{\partial^2}{\partial x_1 \partial x_n} f \\ \frac{\partial^2}{\partial x_2 \partial x_1} f & \frac{\partial^2}{\partial x_2 \partial x_2} f & \frac{\partial^2}{\partial x_2 \partial x_3} f & \cdots & \frac{\partial^2}{\partial x_2 \partial x_n} f \\ \frac{\partial^2}{\partial x_3 \partial x_1} f & \frac{\partial^2}{\partial x_3 \partial x_2} f & \frac{\partial^2}{\partial x_3 \partial x_3} f & \cdots & \frac{\partial^2}{\partial x_3 \partial x_n} f \\ \vdots & \vdots & \vdots & & \vdots \\ \frac{\partial^2}{\partial x_n \partial x_1} f & \frac{\partial^2}{\partial x_n \partial x_2} f & \frac{\partial^2}{\partial x_n \partial x_3} f & \cdots & \frac{\partial^2}{\partial x_n \partial x_n} f \end{bmatrix}$$

- In Calc III we learned that in order for $f(\hat{\mathbf{x}})$ to be an extreme value we must have that

$$\nabla f(\hat{\mathbf{x}}) = \mathbf{0}.$$

- We did have a second derivative test of sorts, but only for the special case that $n = 2$. Specifically, let

$$D = \left(\frac{\partial^2 f}{\partial x_1 \partial x_1} \times \frac{\partial^2 f}{\partial x_2 \partial x_2} - \left(\frac{\partial^2 f}{\partial x_1 \partial x_2} \right)^2 \right) (\hat{\mathbf{x}}).$$

  Supposing that the gradient is zero, we know that $f(\hat{\mathbf{x}})$ is a local maximum value if $D > 0$ and $\frac{\partial^2 f}{\partial x_1 \partial x_1} > 0$, and it is a local minimum if $D > 0$ and $\frac{\partial^2 f}{\partial x_1 \partial x_1} < 0$. If $D < 0$ then $\hat{\mathbf{x}}$ is a saddle point. If $D = 0$ the second derivative test is inconclusive.

- That's pretty contrived, and its scope is extremely limited to $n = 2$.

- There is a much more general and compelling result in terms of the definiteness of the Hessian!

- To build up to the vector case, let's first derive the scalar case.
- Remember the beginning of the Taylor Series of f about the point $\hat{x}$:

$$f(x) = f(\hat{x}) + (x - \hat{x}) f'(\hat{x}) + \frac{1}{2}(x - \hat{x})^2 f''(\hat{x}) + \ldots$$

- low order terms dominate higher order terms

$z^2 < |z|$ for small $z \neq 0$

- $f'(\hat{x}) \neq 0 \Rightarrow$ neither

we can increase or decrease $(x-\hat{x}) f'(\hat{x})$

- $f'(\hat{x}) = 0$ $\quad f''(\hat{x}) > 0$

$\frac{1}{2}(x-\hat{x})^2 f''(x) > 0$

can only increase

have a min

We can do the vector case the same way:

$$f(\mathbf{x}) = f(\hat{\mathbf{x}}) + (\mathbf{x} - \hat{\mathbf{x}})^T \nabla f(\hat{\mathbf{x}}) + \frac{1}{2}(\mathbf{x} - \hat{\mathbf{x}})^T \nabla^2 f(\hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}}) + \dots$$

same

$\nabla^2 f(\hat{x})$ pos. def.

neg. def.

- The key idea is that for small values of $\|\mathbf{x}-\hat{\mathbf{x}}\|$ low order terms dominate higher order terms.

- Thus, if the gradient is non-zero, then we can increase or decrease the function value by picking $\mathbf{x}$ in the right direction from $\hat{\mathbf{x}}$, and sufficiently close to $\hat{\mathbf{x}}$.

- Hence the gradient must be zero for $f(\hat{\mathbf{x}})$ to be an extreme value.

- If the gradient is in fact zero, then the quadratic term will dominate higher order terms. In a small neighborhood of $\hat{\mathbf{x}}$ we can ignore the higher order terms.

- The statements about the second derivatives follow. The local behavior of the function is determined by the behavior of the quadratic form whose matrix is the Hessian at $\hat{\mathbf{x}}$.

- For example, if the Hessian is positive definite then for $\mathbf{x}$ close to $\hat{\mathbf{x}}$ we can only increase the function value, and so $\mathbf{f}(\hat{\mathbf{x}})$ must be a local minimum.

General Second Derivative Test

- To summarize: In order to have an extreme value at some point the gradient at that point must be zero.
- If in addition:
  - The Hessian is positive definite we have a local minimum,
  - The Hessian is negative definite we have a local maximum,
  - The Hessian is indefinite we have a local saddle point,
  - The Hessian is (positive or negative) semidefinite the test is inconclusive.

- Let's revisit the second derivative test for functions of two variables.

$$\begin{bmatrix} f_{x_1x_1} & f_{x_1x_2} \\ f_{x_2x_1} & f_{x_2x_2} \end{bmatrix}$$

- The discriminant

$$D = \left(\frac{\partial^2 f}{\partial x_1 \partial x_1} \times \frac{\partial^2 f}{\partial x_2 \partial x_2} - \left(\frac{\partial^2 f}{\partial x_1 \partial x_2} \right)^2 \right) (\hat{\mathbf{x}}).$$

is actually the **determinant** of the Hessian. Thus it is the product of the eigenvalues. Those eigenvalues are real, and there are two of them. If the product is positive then either both eigenvalues are positive, or both are negative. The Hessian, therefore, is either positive definite, or negative definite. If in addition the second order partial with respect to one of the variables is positive we have a local minimum if we restrict the function to a line in the corresponding coordinate direction, and therefore we must have a global minimum. If the product of the eigenvalues is negative then one eigenvalue is positive and the other negative. We have a saddle point. If the Hessian is positive semidefinite, and the product of the eigenvalues zero, we may have a minimum or a saddle point. If it's negative semidefinite we may a maximum or a saddle point.

$$\det(A - \lambda I) = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \ldots (\lambda_n - \lambda)$$

$\det(A)$ = product of evalues