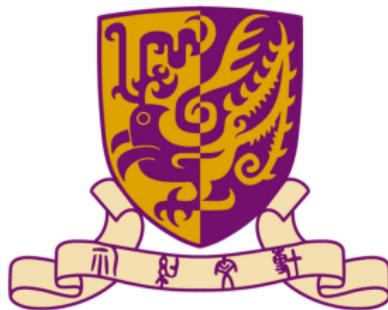


Lecture 2: Unconstrained Optimization

Shi Pu

School of Data Science (SDS)
The Chinese University of Hong Kong, Shenzhen



Outline

- 1 Global and Local Optima
- 2 First Order Optimality Condition
- 3 Classification of Matrices
- 4 Second Order Optimality Conditions
- 5 Global Optimality Condition
- 6 Quadratic Functions

Outline

- 1 Global and Local Optima
- 2 First Order Optimality Condition
- 3 Classification of Matrices
- 4 Second Order Optimality Conditions
- 5 Global Optimality Condition
- 6 Quadratic Functions

Definition (Global Minimum and Maximum)

Let $f: S \rightarrow \mathbb{R}$ be defined on a set $S \subseteq \mathbb{R}^n$. Then

1. $\mathbf{x}^* \in S$ is a *global minimum point* of f over S if $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ for any $\mathbf{x} \in S$.
2. $\mathbf{x}^* \in S$ is a *strict global minimum point* of f over S if $f(\mathbf{x}) > f(\mathbf{x}^*)$ for any $\mathbf{x}^* \neq \mathbf{x} \in S$.
3. $\mathbf{x}^* \in S$ is a *global maximum point* of f over S if $f(\mathbf{x}) \leq f(\mathbf{x}^*)$ for any $\mathbf{x} \in S$.
4. $\mathbf{x}^* \in S$ is a *strict global maximum point* of f over S if $f(\mathbf{x}) < f(\mathbf{x}^*)$ for any $\mathbf{x}^* \neq \mathbf{x} \in S$.

- *global optimum* = global minimum or maximum.
- *maximal value* of f over S :

$$\sup\{f(\mathbf{x}) : \mathbf{x} \in S\}$$

- *minimal value* of f over S :

$$\inf\{f(\mathbf{x}) : \mathbf{x} \in S\}$$

- minimal and maximal values are always unique.

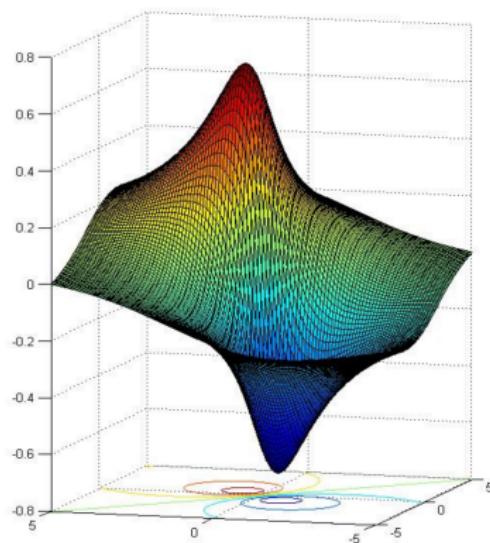
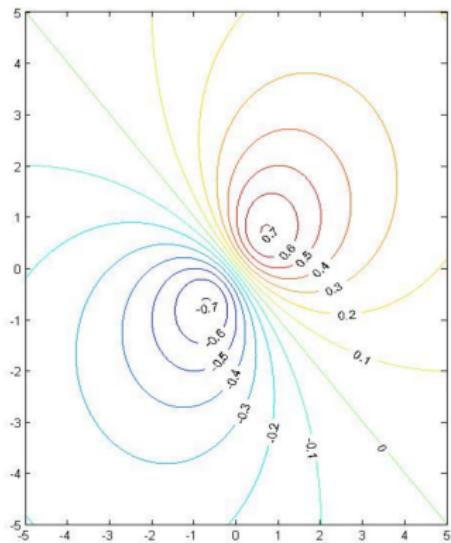
Example 1

Find the global minimum and maximum points of $f(x, y) = x + y$ over the unit ball $S = B[\mathbf{0}, 1] = \{(x, y)^\top : x^2 + y^2 \leq 1\}$

In class.

Example 2

$$\min\{f(x, y) = \frac{x + y}{x^2 + y^2 + 1} : x, y \in \mathbb{R}\}$$



$(1/\sqrt{2}, 1/\sqrt{2})$ - global maximizer $(-1/\sqrt{2}, -1/\sqrt{2})$ - global minimizer.

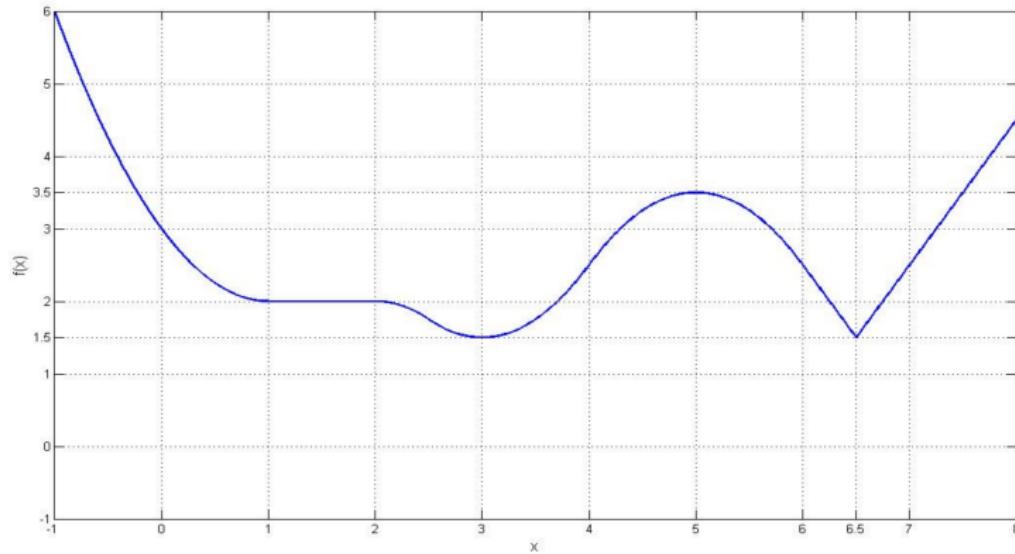
Definition

Let $f: S \rightarrow \mathbb{R}$ be defined on a set $S \subseteq \mathbb{R}^n$. Then

1. $\mathbf{x}^* \in S$ is a *local minimum* of f over S if there exists $r > 0$ for which $f(\mathbf{x}^*) \leq f(\mathbf{x})$ for any $\mathbf{x} \in S \cap B(\mathbf{x}^*, r)$.
2. $\mathbf{x}^* \in S$ is a *strict local minimum* of f over S if there exists $r > 0$ for which $f(\mathbf{x}^*) < f(\mathbf{x})$ for any $\mathbf{x} \neq \mathbf{x}^* \in S \cap B(\mathbf{x}^*, r)$.
3. $\mathbf{x}^* \in S$ is a *local maximum* of f over S if there exists $r > 0$ for which $f(\mathbf{x}^*) \geq f(\mathbf{x})$ for any $\mathbf{x} \in S \cap B(\mathbf{x}^*, r)$.
4. $\mathbf{x}^* \in S$ is a *strict local maximum* of f over S if there exists $r > 0$ for which $f(\mathbf{x}^*) > f(\mathbf{x})$ for any $\mathbf{x} \neq \mathbf{x}^* \in S \cap B(\mathbf{x}^*, r)$.

Of course, a global minimum (maximum) point is also a local minimum (maximum) point.

Example



f described above is defined over $[-1, 8]$. Classify each of the points $x = -1, 1, 2, 3, 5, 6.5, 8$ as strict/nonstrict global/local minimum/maximum points. **In class.**

Outline

- 1 Global and Local Optima
- 2 First Order Optimality Condition
- 3 Classification of Matrices
- 4 Second Order Optimality Conditions
- 5 Global Optimality Condition
- 6 Quadratic Functions

Theorem

Let $f: U \rightarrow \mathbb{R}$ be a function defined on a set $U \subseteq \mathbb{R}^n$. Suppose that $\mathbf{x}^* \in \text{int}(U)$ is a local optimum point and that all the partial derivatives of f exist at \mathbf{x}^* . Then $\nabla f(\mathbf{x}^*) = \mathbf{0}$.

Proof.

- Let $i \in \{1, 2, \dots, n\}$ and consider the 1-D function $g(t) = f(\mathbf{x}^* + t\mathbf{e}_i)$
- \mathbf{x}^* is a local optimum point of $f \Rightarrow t = 0$ is a local optimum of $g \Rightarrow g'(0) = 0$.
- Thus, $\frac{\partial f}{\partial x_i}(\mathbf{x}^*) = g'(0) = 0$.



Definition

Let $f: U \rightarrow \mathbb{R}$ be defined on a set $U \subseteq \mathbb{R}^n$. Suppose that $\mathbf{x}^* \in \text{int}(U)$ and that all the partial derivatives of f are defined at \mathbf{x}^* . Then \mathbf{x}^* is called a *stationary point* of f if $\nabla f(\mathbf{x}^*) = \mathbf{0}$.

Example

$$\min\{f(x, y) = \frac{x + y}{x^2 + y^2 + 1} : x, y \in \mathbb{R}\}$$

Example

$$\min\{f(x, y) = \frac{x + y}{x^2 + y^2 + 1} : x, y \in \mathbb{R}\}$$

■

$$\nabla f(x, y) = \frac{1}{(x^2 + y^2 + 1)^2} \begin{pmatrix} (x^2 + y^2 + 1) - 2(x + y)x \\ (x^2 + y^2 + 1) - 2(x + y)y \end{pmatrix}.$$

- Stationary points are those satisfying:

$$-x^2 - 2xy + y^2 = -1,$$

$$x^2 - 2xy - y^2 = -1.$$

- Hence, the stationary points are $(1/\sqrt{2}, 1/\sqrt{2})$, $(-1/\sqrt{2}, -1/\sqrt{2})$.
■ $(1/\sqrt{2}, 1/\sqrt{2})$ - global maximizer, $(-1/\sqrt{2}, -1/\sqrt{2})$ - global minimizer.

Outline

- 1 Global and Local Optima
- 2 First Order Optimality Condition
- 3 Classification of Matrices
- 4 Second Order Optimality Conditions
- 5 Global Optimality Condition
- 6 Quadratic Functions

Classification of Matrices - Positive Definiteness

1. A symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is called *positive semidefinite*, denoted by $\mathbf{A} \succeq \mathbf{0}$, if $\mathbf{x}^\top \mathbf{A} \mathbf{x} \geq 0$ for every $\mathbf{x} \in \mathbb{R}^n$.
2. A symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is called *positive definite*, denoted by $\mathbf{A} \succ \mathbf{0}$, if $\mathbf{x}^\top \mathbf{A} \mathbf{x} > 0$ for every $\mathbf{0} \neq \mathbf{x} \in \mathbb{R}^n$.

Example 1:

$$\mathbf{A} = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}.$$

In class.

Example 2:

$$\mathbf{B} = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}.$$

In class.

Lemma

Let \mathbf{A} be a positive definite (semidefinite) matrix. Then the diagonal elements of \mathbf{A} are positive (nonnegative).

1. A symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is called *negative semidefinite*, denoted by $\mathbf{A} \preceq \mathbf{0}$, if $\mathbf{x}^\top \mathbf{A} \mathbf{x} \leq 0$ for every $\mathbf{x} \in \mathbb{R}^n$.
2. A symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is called *negative definite*, denoted by $\mathbf{A} \prec \mathbf{0}$, if $\mathbf{x}^\top \mathbf{A} \mathbf{x} < 0$ for every $\mathbf{0} \neq \mathbf{x} \in \mathbb{R}^n$.
3. A symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is called *indefinite* if there exist $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ such that $\mathbf{x}^\top \mathbf{A} \mathbf{x} > 0, \mathbf{y}^\top \mathbf{A} \mathbf{y} < 0$.

Remarks:

- \mathbf{A} is negative (semi)definite if and only if $-\mathbf{A}$ is positive (semi)definite.
- A matrix is indefinite if and only if it is neither positive semidefinite nor negative semidefinite.

Theorem

Let \mathbf{A} be a symmetric $n \times n$ matrix. Then

- (a) \mathbf{A} is positive definite iff all its eigenvalues are positive.
- (b) \mathbf{A} is positive semidefinite iff all its eigenvalues are nonnegative.
- (c) \mathbf{A} is negative definite iff all its eigenvalues are negative.
- (d) \mathbf{A} is negative semidefinite iff all its eigenvalues are nonpositive.
- (e) \mathbf{A} is indefinite iff it has at least one positive eigenvalue and at least one negative eigenvalue.

Eigenvalue Characterization – Proof

Proof of part (a): (other parts follows immediately or by similar arguments)

- There exists orthogonal $\mathbf{U} \in \mathbb{R}^{n \times n}$ such that

$$\mathbf{U}^\top \mathbf{A} \mathbf{U} = \mathbf{D} \equiv \text{diag}(d_1, d_2, \dots, d_n)$$

where $d_i = \lambda_i(\mathbf{A})$.

- Making the linear change of variables $\mathbf{x} = \mathbf{U}\mathbf{y}$, we have

$$\mathbf{x}^\top \mathbf{A} \mathbf{x} = \mathbf{y}^\top \mathbf{U}^\top \mathbf{A} \mathbf{U} \mathbf{y} = \mathbf{y}^\top \mathbf{D} \mathbf{y} = \sum_{i=1}^n d_i y_i^2.$$

- Therefore, $\mathbf{x}^\top \mathbf{A} \mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$ iff

$$\sum_{i=1}^n d_i y_i^2 > 0 \text{ for any } \mathbf{y} \neq \mathbf{0}. \quad (1)$$

- (1) holds iff $d_i > 0$ for all i (why?)

Outline

- 1 Global and Local Optima
- 2 First Order Optimality Condition
- 3 Classification of Matrices
- 4 Second Order Optimality Conditions
- 5 Global Optimality Condition
- 6 Quadratic Functions

Theorem

Let $f : U \rightarrow \mathbb{R}$ be a function defined on an open set $U \subseteq \mathbb{R}^n$. Suppose that f is twice continuously differentiable over U and that \mathbf{x}^* is a stationary point. Then

1. if \mathbf{x}^* is a local minimum point, then $\nabla^2 f(\mathbf{x}^*) \succeq \mathbf{0}$.
2. if \mathbf{x}^* is a local maximum point, then $\nabla^2 f(\mathbf{x}^*) \preceq \mathbf{0}$.

Proof of 1.:

- There exists a ball $B(\mathbf{x}^*, r) \subseteq U$ for which $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ for all $\mathbf{x} \in B(\mathbf{x}^*, r)$.
- Let $\mathbf{d} \in \mathbb{R}^n$ be a nonzero vector. For any $0 < \alpha < \frac{r}{\|\mathbf{d}\|}$, we have $\mathbf{x}_\alpha^* \equiv \mathbf{x}^* + \alpha\mathbf{d} \in B(\mathbf{x}^*, r)$,
- For any such α , $f(\mathbf{x}_\alpha^*) \geq f(\mathbf{x}^*)$.
- On the other hand, there exists a vector $\mathbf{z}_\alpha \in [\mathbf{x}^*, \mathbf{x}_\alpha^*]$ such that

$$f(\mathbf{x}_\alpha^*) - f(\mathbf{x}^*) = \frac{\alpha^2}{2} \mathbf{d}^T \nabla^2 f(\mathbf{z}_\alpha) \mathbf{d}. \quad (2)$$

- \Rightarrow For any $\alpha \in (0, \frac{r}{\|\mathbf{d}\|})$ the inequality $\mathbf{d}^T \nabla^2 f(\mathbf{z}_\alpha) \mathbf{d} \geq 0$ holds.
- Since $\mathbf{z}_\alpha \rightarrow \mathbf{x}^*$ as $\alpha \rightarrow 0^+$, we obtain that $\mathbf{d}^T \nabla^2 f(\mathbf{x}^*) \mathbf{d} \geq 0$.
- $\Rightarrow \nabla^2 f(\mathbf{x}^*) \succeq \mathbf{0}$.

Theorem

Let $f : U \rightarrow \mathbb{R}$ be a function defined on an open set $U \subseteq \mathbb{R}^n$. Suppose that f is twice continuously differentiable over U and that \mathbf{x}^* is a stationary point. Then

1. if $\nabla^2 f(\mathbf{x}^*) \succ \mathbf{0}$, then \mathbf{x}^* is a strict local minimum point of f over U .
2. if $\nabla^2 f(\mathbf{x}^*) \prec \mathbf{0}$, then \mathbf{x}^* is a strict local maximum point of f over U .

Proof. of 1: (2 directly follows)

- There exists a ball $B(\mathbf{x}^*, r) \subseteq U$ for which $\nabla^2 f(\mathbf{x}^*) \succ \mathbf{0}$ for any $\mathbf{x} \in B(\mathbf{x}^*, r)$.
- By LAT, there exists a vector $\mathbf{z}_x \in [\mathbf{x}^*, \mathbf{x}]$ (and hence $\mathbf{z}_x \in B(\mathbf{x}^*, r)$) for which

$$f(\mathbf{x}) - f(\mathbf{x}^*) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^T \nabla^2 f(\mathbf{z}_x)(\mathbf{x} - \mathbf{x}^*).$$

- $\nabla^2 f(\mathbf{z}_x) \succ \mathbf{0} \Rightarrow$ for any $\mathbf{x} \in B(\mathbf{x}^*, r)$ such that $\mathbf{x} \neq \mathbf{x}^*$, the inequality $f(\mathbf{x}) > f(\mathbf{x}^*)$ holds, implying that \mathbf{x}^* is a strict local minimum point of f over U .

Definition

Let $f: U \rightarrow \mathbb{R}$ be a continuously differentiable function defined on an open set $U \subseteq \mathbb{R}^n$. A stationary point $\mathbf{x}^* \in U$ is called a *saddle point* of f over U if it is neither a local minimum point nor a local maximum point of f over U .

Theorem

Let $f : U \rightarrow \mathbb{R}$ be a function defined on an open set $U \subseteq \mathbb{R}^n$. Suppose that f is twice continuously differentiable over U and that \mathbf{x}^* is a stationary point. If $\nabla^2 f(\mathbf{x}^*)$ is an indefinite matrix, then \mathbf{x}^* is a saddle point of f over U .

Proof.

- $\nabla^2 f(\mathbf{x}^*)$ has at least one positive eigenvalue $\lambda > 0$, corresponding to a normalized eigenvector denoted by \mathbf{v} .
- $\exists r > 0$ such that $\mathbf{x}^* + \alpha\mathbf{v} \in U$ for any $\alpha \in (0, r)$.
- By QAT¹ there exists a function $g : \mathbb{R}_{++} \rightarrow \mathbb{R}$ satisfying

$$\frac{g(t)}{t} \rightarrow 0 \text{ as } t \rightarrow 0,$$

such that for any $\alpha \in (0, r)$

$$f(\mathbf{x}^* + \alpha\mathbf{v}) = f(\mathbf{x}^*) + \frac{\lambda\alpha^2}{2} \|\mathbf{v}\|^2 + g(\|\mathbf{v}\|^2\alpha^2).$$

¹quadratic approximation theorem

Proof Contd.

- $f(\mathbf{x}^* + \alpha \mathbf{v}) = f(\mathbf{x}^*) + \frac{\lambda \alpha^2}{2} + g(\alpha^2).$
- $\exists \epsilon_1 \in (0, r)$ such that $g(\alpha^2) > -\frac{\lambda}{2}\alpha^2$ for all $\alpha \in (0, \epsilon_1)$.
- $f(\mathbf{x}^* + \alpha \mathbf{v}) > f(\mathbf{x}^*)$ for all $\alpha \in (0, \epsilon_1)$.
- \mathbf{x}^* *cannot* be a local maximum point of f over U .
- Similarly, *cannot* be a local minimum point of f over $U \Rightarrow$ saddle point.

Outline

- 1 Global and Local Optima
- 2 First Order Optimality Condition
- 3 Classification of Matrices
- 4 Second Order Optimality Conditions
- 5 Global Optimality Condition
- 6 Quadratic Functions

Theorem (Weierstrass Theorem)

Let f be a continuous function defined over a nonempty compact set $C \subseteq \mathbb{R}^n$. Then there exists a global minimum point of f over C and a global maximum point of f over C .

- When the underlying set is not compact, Weierstrass theorem does not guarantee the attainment of the solution, but certain properties of the function f can imply attainment of the solution.

Definition

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a continuous function over \mathbb{R}^n . f is called *coercive* if

$$\lim_{\|\mathbf{x}\| \rightarrow \infty} f(\mathbf{x}) = \infty.$$

Theorem (Attainment of Global Optima Points for Coercive Functions)

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a continuous and coercive function and let $S \subseteq \mathbb{R}^n$ be a nonempty closed set. Then f attains a global minimum point on S .

Proof. In class

Example

Classify the stationary points of the function
 $f(x_1, x_2) = -2x_1^2 + x_1x_2^2 + 4x_1^4.$

Example

Classify the stationary points of the function
 $f(x_1, x_2) = -2x_1^2 + x_1x_2^2 + 4x_1^4.$

$$\nabla f(\mathbf{x}) = \begin{pmatrix} -4x_1 + x_2^2 + 16x_1^3 \\ 2x_1x_2 \end{pmatrix}.$$

⇒ Stationary points are the solutions to

$$-4x_1 + x_2^2 + 16x_1^3 = 0,$$

$$2x_1x_2 = 0.$$

Example

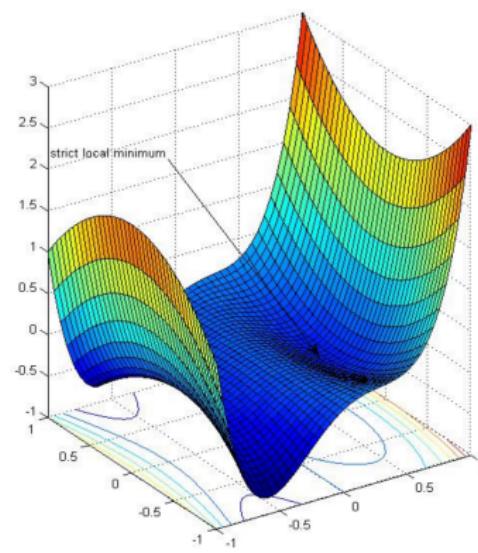
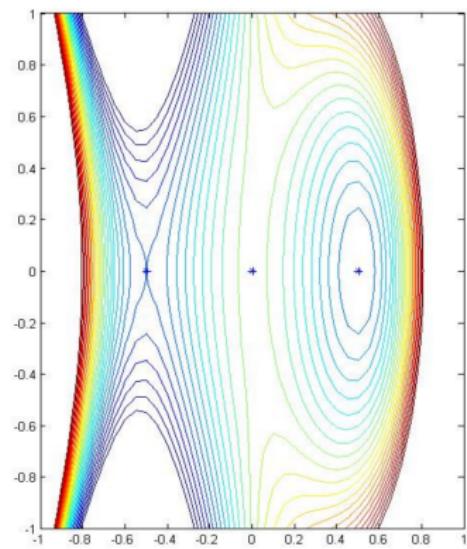
⇒ Stationary points are $(0, 0)$, $(0.5, 0)$, $(-0.5, 0)$

$$\nabla^2 f(x_1, x_2) = \begin{pmatrix} -4 + 48x_1^2 & 2x_2 \\ 2x_2 & 2x_1 \end{pmatrix}.$$

$$\nabla^2 f(0.5, 0) = \begin{pmatrix} 8 & 0 \\ 0 & 1 \end{pmatrix} \quad \nabla^2 f(-0.5, 0) = \begin{pmatrix} 8 & 0 \\ 0 & -1 \end{pmatrix} \quad \nabla^2 f(0, 0) = \begin{pmatrix} -4 & 0 \\ 0 & 0 \end{pmatrix}$$

strict local minimum saddle point saddle point (why?)

Illustration



No free meals...

Theorem

Let f be a twice continuously differentiable function defined over \mathbb{R}^n . Suppose that $\nabla^2 f(\mathbf{x}) \succeq \mathbf{0}$ for any $\mathbf{x} \in \mathbb{R}^n$. Let $\mathbf{x}^* \in \mathbb{R}^n$ be a stationary point of f . Then \mathbf{x}^* is a global minimum point of f .

Proof. In class

Example

$$f(\mathbf{x}) = x_1^2 + x_2^2 + x_3^2 + x_1x_2 + x_1x_3 + x_2x_3 + (x_1^2 + x_2^2 + x_3^2)^2.$$

Example

$$f(\mathbf{x}) = x_1^2 + x_2^2 + x_3^2 + x_1x_2 + x_1x_3 + x_2x_3 + (x_1^2 + x_2^2 + x_3^2)^2.$$

$$\nabla f(\mathbf{x}) = \begin{pmatrix} 2x_1 + x_2 + x_3 + 4x_1(x_1^2 + x_2^2 + x_3^2) \\ 2x_2 + x_1 + x_3 + 4x_2(x_1^2 + x_2^2 + x_3^2) \\ 2x_3 + x_1 + x_2 + 4x_3(x_1^2 + x_2^2 + x_3^2) \end{pmatrix}.$$

$$\nabla^2 f(\mathbf{x}) = \begin{pmatrix} 2 + 4(x_1^2 + x_2^2 + x_3^2) + 8x_1^2 & 1 + 8x_1x_2 & 1 + 8x_1x_3 \\ 1 + 8x_1x_2 & 2 + 4(x_1^2 + x_2^2 + x_3^2) + 8x_2^2 & 1 + 8x_2x_3 \\ 1 + 8x_1x_3 & 1 + 8x_2x_3 & 2 + 4(x_1^2 + x_2^2 + x_3^2) + 8x_3^2 \end{pmatrix}.$$

- $\mathbf{x} = \mathbf{0}$ is a stationary point.
- $\nabla^2 f(\mathbf{x}) \succeq \mathbf{0}$ for all \mathbf{x} . (why?)
- Consequence: $\mathbf{x} = \mathbf{0}$ is the global minimum point.

Outline

- 1 Global and Local Optima
- 2 First Order Optimality Condition
- 3 Classification of Matrices
- 4 Second Order Optimality Conditions
- 5 Global Optimality Condition
- 6 Quadratic Functions

Quadratic Functions

- A *quadratic function* over \mathbb{R}^n is a function of the form

$$f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}\mathbf{x} + 2\mathbf{b}^\top \mathbf{x} + c,$$

where $\mathbf{A} \in \mathbb{R}^{n \times n}$ is symmetric, $\mathbf{b} \in \mathbb{R}^n$ and $c \in \mathbb{R}$.

-

$$\nabla f(\mathbf{x}) = 2\mathbf{A}\mathbf{x} + 2\mathbf{b},$$

$$\nabla^2 f(\mathbf{x}) = 2\mathbf{A}.$$

Consequently,

Lemma

Let $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}\mathbf{x} + 2\mathbf{b}^\top \mathbf{x} + c$ ($\mathbf{A} \in \mathbb{R}^{n \times n}$ sym., $\mathbf{b} \in \mathbb{R}^n$, $c \in \mathbb{R}$).

1. \mathbf{x} is a stationary point of f iff $\mathbf{A}\mathbf{x} = -\mathbf{b}$.
2. if $\mathbf{A} \succeq \mathbf{0}$, then \mathbf{x} is a global minimum point of f iff $\mathbf{A}\mathbf{x} = -\mathbf{b}$.
3. if $\mathbf{A} \succ \mathbf{0}$, then $\mathbf{x} = -\mathbf{A}^{-1}\mathbf{b}$ is a strict global minimum point of f .

Proof. In class

Two Important Theorems on Quadratic Functions

Lemma (coerciveness of quadratic functions)

Let $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}\mathbf{x} + 2\mathbf{b}^\top \mathbf{x} + c$ where $\mathbf{A} \in \mathbb{R}^{n \times n}$ is symmetric, $\mathbf{b} \in \mathbb{R}^n$ and $c \in \mathbb{R}$. Then f is coercive if and only if $\mathbf{A} \succ \mathbf{0}$.

Lemma 2.42 in the book (proof in page 33).

Theorem (characterization of the nonnegativity of quadratic functions)

$f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}\mathbf{x} + 2\mathbf{b}^\top \mathbf{x} + c$ where $\mathbf{A} \in \mathbb{R}^{n \times n}$ is symmetric, $\mathbf{b} \in \mathbb{R}^n$ and $c \in \mathbb{R}$. Then the following two claims are equivalent

(i) $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}\mathbf{x} + 2\mathbf{b}^\top \mathbf{x} + c \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$.

(ii) $\begin{pmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{b}^\top & c \end{pmatrix} \succeq \mathbf{0}$.

Theorem 2.43 in the book (proof in pages 33, 34).