Second-order optimality

~~Definite matrices~~

Lecture 06

September 14, 2021

Beck, section 2.3

Necessary second-order optimality.

twice continuously differentiable on S.

Proposition: Assume that $f \in C^2(S)$, $S \subset \mathbb{R}^n$. Let S be open.

(i) If $\underline{x}_*$ is a local minimum of f on S, then
$$\nabla^2 f(\underline{x}_*) \succeq \underline{0}.$$

(ii) If $\underline{x}_*$ is a local maximum of f on S, then
$$\nabla^2 f(\underline{x}_*) \preceq \underline{0}.$$

$f(x) = 1$ on $\mathbb{R}$

$x = 0$ local min and $\nabla^2 f(0) \succeq 0$.

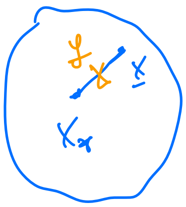
Proof (idea) (i)

$$\underline{x}_* \text{ local min} \implies \nabla f(\underline{x}_*) = \underline{0}.$$

$$\implies \exists\, r > 0 \text{ s.t. } f(\underline{x}) \geq f(\underline{x}_*)$$
$$\forall\, \underline{x} \in B(\underline{x}_*, r)$$

Taylor expansion: Given $\underline{x} \in B(\underline{x}_*, r)$, then

$$f(\underline{x}) = f(\underline{x}_*) + \nabla f(\underline{x}_*)^T(\underline{x} - \underline{x}_*) + \frac{1}{2}(\underline{x} - \underline{x}_*)^T \nabla^2 f(y)(\underline{x} - \underline{x}_*)$$



for some $y \in [\underline{x}_*, \underline{x}]$

$$\Rightarrow \frac{1}{2}(\underline{x} - \underline{x}_*)^T \nabla^2 f(y)(\underline{x} - \underline{x}_*) = f(\underline{x}) - f(\underline{x}_*) \geq 0,$$

Let $\underline{a} \in \mathbb{R}^n$ be arbitrary, $\underline{a} \neq \underline{0}$.

Choose $\underline{x} = \underline{x}_* + \alpha \underline{a}$, $\alpha > 0$ is sufficiently small.

$$\Rightarrow \underline{x} - \underline{x}_* = \alpha \underline{a}$$

$$\Rightarrow \frac{1}{2}\alpha^2 \underline{a}^T \nabla^2 f(y) \underline{a} \geq 0. \rightarrow \underline{a}^T \nabla^2 f(y) \underline{a} \geq 0.$$

$\forall \alpha$ sufficiently small.

Take $\alpha \downarrow 0 \Rightarrow \underline{a}^T \nabla^2 f(\underbrace{\lim_{\alpha \downarrow 0} y}_{\underline{x}_*}) \underline{a} \geq 0$

$$\Rightarrow \underline{a}^T \nabla^2 f(\underline{x}_*) \underline{a} \geq 0 \quad \blacksquare$$

Sufficient second-order optimality

Theorem: Assume $\cancel{\underline{A} \in \mathbb{R}^n}$ $f \in C^2(S)$, $S \subset \mathbb{R}^n$ is open.

(i) If $\underline{x}_*$ is a stationary point and $\nabla^2 f(\underline{x}_*) \succ \underline{0}$, then $\underline{x}_*$ is a local minimum.

(ii) If $\underline{x}_*$ is a stationary point and $\nabla^2 f(\underline{x}_*) \prec \underline{0}$, then $\underline{x}_*$ is a local maximum.

Proof (idea) (i) Again, $\nabla f(\underline{x}_*) = \underline{0}$ since $\underline{x}_*$ is a stationary point.

For $\underline{x}$ "close enough" to x_*:

$$f(\underline{x}) = f(\underline{x}_*) + \nabla f(\underline{x}_*)^T (\underline{x} - \underline{x}_*) + \frac{1}{2} (\underline{x} - \underline{x}_*)^T \nabla^2 f(\underline{y}) (\underline{x} - \underline{x}_*)$$

for some $\underline{y} \in [\underline{x}_*, \underline{x}]$.

$$\Longrightarrow f(\underline{x}) - f(\underline{x}_*) = \frac{1}{2} (\underline{x} - \underline{x}_*)^T \nabla^2 f(\underline{y}) (\underline{x} - \underline{x}_*).$$

claim: if $\nabla^2 f(\underline{x}_*) \succ \underline{0}$ and $f \in C^2(S)$, then $\nabla^2 f(\underline{x}) \succ \underline{0}$ for $\underline{x} \in B(\underline{x}_*, r)$ for some $r > 0$.

$\Longrightarrow$ if $\underline{x}$ is sufficiently close to $\underline{x}_*$

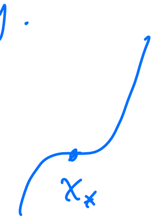
$\Longrightarrow$ $\underline{y} \in [\underline{x}_*, \underline{x}]$ is "close" to $\underline{x}_*$

$\Longrightarrow$ $\nabla f^2(\underline{y}) \succ \underline{0}$ since $\nabla^2 f(\underline{x}_*) \succ \underline{0}$.

$\Longrightarrow$ $f(\underline{x}) - f(\underline{x}_*) = \frac{1}{2} (\underline{x} - \underline{x}_*)^T \nabla^2 f(\underline{y}) (\underline{x} - \underline{x}_*) > 0$. ∎

Ex. semidefiniteness is not sufficient for optimality!

$f(x) = x^3$, $f'(x) = 3x^2$, $f''(x) = 6x$



$x_* = 0$ is a stationary point

$x_* = 0$ has positive semidefinite Hessian.

$x_* = 0$ is not a local optimum.

Saddle points

<u>Definition</u>: Let $f \in C^1(S)$, $S \subset \mathbb{R}^n$ is open. If $\underline{x}_*$ is a stationary point of f but not a local optimum, then it's a saddle point.

Indefinite Hessians

Proposition: Assume $f \in C^2(S)$, $S \subset \mathbb{R}^n$ is open. If f has a stationary point $\underline{x}_*$ whose Hessian is indefinite, then $\underline{x}_*$ is a saddle point.

Proof sketch.

- $\nabla^2 f(\underline{x}_*)$ indefinite $\Longrightarrow$ can choose $\underline{x}$ s.t.

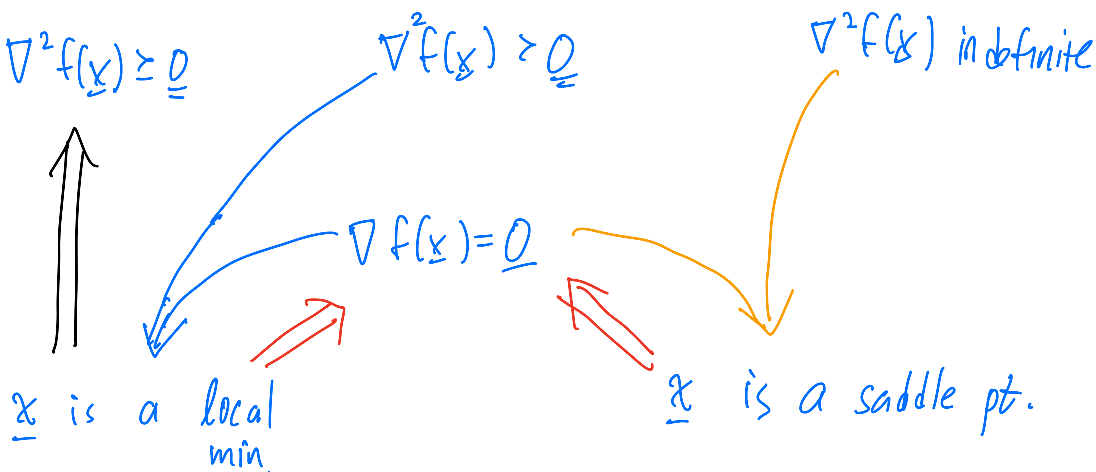
$$(\underline{x} - \underline{x}_*)^T \nabla^2 f(\underline{x}_*)(\underline{x} - \underline{x}_*) > 0 \text{ and } < 0.$$

- i.e. $f(\underline{x}) > f(\underline{x}_*)$ and $f(\underline{y}) < f(\underline{x}_*)$ for carefully chosen $\underline{x}, \underline{y}$.

- i.e. $\underline{x}_*$ cannot be a local optimum. $\blacksquare$

HW #2 due Tuesday

Summary of optimality conditions



First-order optimality

Necessary second-order optimality

Sufficient second-order optimality

Indefinite Hessian.

Existence of global extrema

Theorem (Weierstrass)

If $f \in C(S)$, $S \subset \mathbb{R}^n$. Assume S is compact.

Then f attains its global maximum and minimum on S.

(I.e. f has at least one global maximizer on S and at least one global minimizer on S.)

Optimization on unbounded domains: coercivity

Definition: A function $f \in C(S)$, $S \subset \mathbb{R}^n$ is coercive if

$$\lim_{\|x\| \to \infty} f(\underline{x}) = +\infty$$

any norm

E.g. $f(x) = x^3$ is not coercive.

$f(x) = x^2$ is coercive

Coercivity can be used to remove the boundedness condition on S in the Weierstrass theorem:

<u>Theorem</u>: Assume $f \in C(S)$, $S \subset \mathbb{R}^n$, and that S is closed. If f is coercive, then f attains a global minimum on S.

Examples

Example (Beck, 2.33)

Find the extrema of

$$f(\boldsymbol{x}) = x_1^2 + x_2^2,$$

for $\boldsymbol{x} \in \mathbb{R}^2$.

Compute stationary points: $\nabla f = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} \Longrightarrow \underline{x} = \underline{0}$ is a stationary point.

Hessian: $\nabla^2 f(\underline{x}) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \succ \underline{0} \quad \forall \ \underline{x}$

$\Longrightarrow (0,0)$ is a local minimum.

Since $f(\underline{x}) \geq 0$, and $f(\underline{0}) = 0$, then $\underline{0}$ is a global minimum.

Note also: f is coercive: $\lim_{\|\underline{x}\| \to \infty} \|\underline{x}\|_2^2 = \infty$

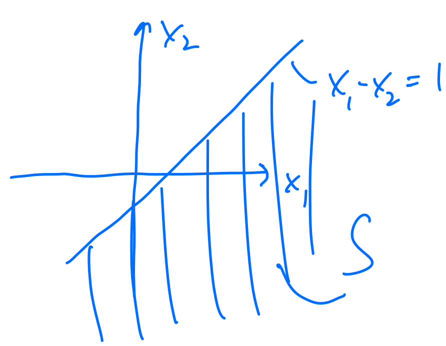
$f(\underline{x})$

I.e., since $\mathbb{R}^n$ is closed in $\mathbb{R}^n$: f attains a global min.

$\|\underline{x}\| = \text{dist}(0, \underline{x})$

Example: Compute extrema of

$$f(\underline{x}) = x_1^2 + x_2^2 \quad \text{on} \quad S = \{\underline{x} \mid x_1 - x_2 \geq 1\}$$



Note: f is coercive, and S is closed $\Rightarrow$ f attains a global minimum.

Two possibilities: (i) global min is in $\text{int}(S)$

(ii) global min is on $\text{bd}(S)$.

(i) $\nabla f(\underline{x}) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} = \underline{0} \Rightarrow (0,0)$ is a stationary point,
but $\underline{0} \notin S$.

i.e.: no stationary points in $\operatorname{int}(S)$.

$\Rightarrow$ no global minimum in $\operatorname{int}(S)$.

(ii) on boundary? $\quad x_1 - x_2 = 1$

$$x_1 = 1 + x_2$$

on $\operatorname{bd}(S)$: $f(\underline{x}) = x_1^2 + x_2^2 = \underbrace{(1+x_2)^2 + x_2^2}_{g(x_2)}, \quad \forall x_2 \in \mathbb{R}$.

$$\left.\begin{array}{l} g'(x_2) = 2(x_2+1) + 2x_2 = 0 \\ \qquad \Rightarrow x_2 = -1/2 \\ g''(x_2) = 4 > 0 \quad \forall x_2 \end{array}\right\} \Rightarrow x_2 = -1/2 \text{ is a global minimum of } g.$$

$x_2 = -1/2 \Rightarrow x_1 = +1/2$

$\Rightarrow$ global min of f on $\operatorname{bd}(S)$ is $(1/2, -1/2)$

$\Rightarrow$ global min of f on S is $(1/2, -1/2)$.

$$f(1/2, -1/2) = 1/2$$

Examples

Example (Beck, 2.33)

Find the extrema of

$$f(\boldsymbol{x}) = x_1^2 + x_2^2,$$

for $\boldsymbol{x} \in \mathbb{R}^2$.

Example (Beck, 2.34)

Compute and analyze the stationary points of

$$f(\boldsymbol{x}) = 2x_1^3 + 3x_2^2 + 3x_1^2 x_2 - 24x_2$$

$$\nabla f = \begin{pmatrix} 6x_1^2 + 6x_1x_2 \\ 6x_2 + 3x_1^2 - 24 \end{pmatrix} = \begin{pmatrix} 6x_1(x_1+x_2) \\ 3(2x_2 + x_1^2 - 8) \end{pmatrix} \overset{?}{=} \underline{0}$$

if $x_1 = 0 \Longrightarrow$ need $2x_2 - 8 = 0 \Rightarrow x_2 = 4$

i.e. $(0,4)$ is a stationary point

if $x_1 = -x_2 \Longrightarrow$ need $2x_2 + (-x_2)^2 - 8 = 0$

$$x_2^2 + 2x_2 - 8 = 0$$

$$(x_2 + 4)(x_2 - 2) = 0$$

$$x_2 = 2, -4$$

$$x_1 = -2, +4$$

3 stationary pts: $(0,4)$, $\cancel{(2,-4), (-2,4)}$ $(-2,2)$, $(4,-4)$

Classify: $\nabla^2 f = \begin{pmatrix} 12x_1 + 6x_2 & 6x_1 \\ 6x_1 & 6 \end{pmatrix}$

$$= 6 \begin{pmatrix} 2x_1 + x_2 & x_1 \\ x_1 & 1 \end{pmatrix}$$

$\nabla^2 f(0,4) = 6 \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix} \succ \underline{\underline{0}} \Longrightarrow (0,4)$ is a local minimum

$$\nabla^2 f(-2,2) = 6\begin{pmatrix} -2 & -2 \\ -2 & 1 \end{pmatrix}$$

"ignore"

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

$(\lambda+2)(\lambda-1)$

eigenvalues of $\underline{\underline{A}}$: $(-2-\lambda)(1-\lambda) - 4 = 0$

$$\lambda^2 + \lambda - 6 = 0$$

$$(\lambda+3)(\lambda-2) = 0$$

$$\lambda = +2, -3$$

$\Longrightarrow$ Eigenvalues of $\nabla^2 f(-2,2)$ are $12, -18$, so $\nabla^2 f$ is indefinite.

$\Longrightarrow$ $(-2,2)$ is a saddle point.

Last point $(4,-4)$: exercise.

Examples, II

Example (Beck, 2.35)

Compute and analyze the stationary points of

$$f(\boldsymbol{x}) = (x_1^2 + x_2^2 - 1)^2 + (x_2^2 - 1)^2$$

Global optimality

Proposition: Assume $f \in C^2(S)$, $S \subset \mathbb{R}^n$. If $\nabla^2 f(\underline{x}) \succeq \underline{\underline{0}}$ $\forall\, \underline{x} \in S$, then any stationary point is a global minimum.

Idea of proof: Assume $\underline{x}_*$ is stationary point. Then for any $\underline{x} \in S$:

$$f(\underline{x}) = f(\underline{x}_*) + \underbrace{\nabla f(\underline{x}_*)}_{\to 0}{}^T(\underline{x} - \underline{x}_*) + \frac{1}{2}(\underline{x} - \underline{x}_*)\nabla^2 f(\underline{y})(\underline{x} - \underline{x}_*),$$

for some $\underline{y} \in [\underline{x}_*, \underline{x}]$.

$$f(\underline{x}) - f(\underline{x}_*) = \frac{1}{2}(\underline{x} - \underline{x}_*)^T \nabla^2 f(\underline{y})(\underline{x} - \underline{x}_*) \geq 0.$$

$$\Rightarrow f(\underline{x}) \geq f(\underline{x}_*). \quad \forall\, \underline{x}.$$

Quadratic functions

Definition: A function $f: S \to \mathbb{R}$, $S \subset \mathbb{R}^n$ is a quadratic function if $\exists$ symmetric matrix $\underline{\underline{A}}$, and $\underline{b} \in \mathbb{R}^n$, and $c \in \mathbb{R}$ such that

$$f(\underline{x}) = \underline{x}^T \underline{\underline{A}}\, \underline{x} + 2\underline{b}^T \underline{x} + c.$$

What if $\underline{\underline{A}}$ is not symmetric?

$$\underline{x}^T \underline{\underline{A}}\, \underline{x} = \frac{1}{2} \underline{x}^T \underline{\underline{A}}\, \underline{x} + \frac{1}{2} \underline{x}^T \underline{\underline{A}}\, \underline{x}$$

$$= \frac{1}{2} \underline{x}^T \underline{\underline{A}}\, \underline{x} + \frac{1}{2} \left(\underline{x}^T \underline{\underline{A}}\, \underline{x} \right)^T$$

$$= \frac{1}{2} \underline{x}^T \underline{\underline{A}} \underline{x} + \frac{1}{2} \underline{x}^T \underline{\underline{A}}^T \underline{x}$$

$$= \frac{1}{2} \underline{x}^T \underbrace{(\underline{\underline{A}} + \underline{\underline{A}}^T)}_{\text{symmetric matrix.}} \underline{x}.$$

replace $\underline{\underline{A}}$ by $\frac{1}{2}(\underline{\underline{A}} + \underline{\underline{A}}^T)$

For quadratic functions: $f(\underline{x}) = \underline{x}^T \underline{\underline{A}} \underline{x} + 2\underline{b}^T \underline{x} + c$

$$\nabla f(\underline{x}) = 2\underline{\underline{A}}\underline{x} + 2\underline{b} \longrightarrow \text{stationary points: } \underline{\underline{A}}\underline{x} = -\underline{b}.$$

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

Optimizing definite quadratic functions

Theorem: Let $f(\underline{x}) = \underline{x}^T \underline{\underline{A}}\, \underline{x} + 2\underline{b}^T \underline{x} + c$ be quadratic. ($\underline{\underline{A}}$ is symmetric). Then:

(i) $\underline{x}$ is stationary point iff $\underline{\underline{A}}\,\underline{x} = -\underline{b}$

(ii) If $\underline{\underline{A}} \succeq \underline{\underline{0}}$: if $\underline{x}$ is stationary, then $\underline{x}$ is a global minimum and satisfies $\underline{\underline{A}}\,\underline{x} = -\underline{b}$.

(iii) If $\underline{\underline{A}} \succ \underline{\underline{0}}$: then $\underline{x} = -\underline{\underline{A}}^{-1}\underline{b}$ is the unique global minimum of f on $\mathbb{R}^n$.

Quadratic functions and coercivity

Characterization of non-negative quadratic functions