LECTURE 3: OPTIMALITY CONDITIONS

1. First order and second order information
2. Necessary and sufficient conditions of optimality
3. Convex functions
4. Elementary solution methods

General setting

- General form nonlinear programming problem

$$\begin{aligned} \text{Min} \quad & f(x) \\ \text{s. t.} \quad & x \in S \subset E^n \end{aligned}$$

where S can be a "simple" set

$$\text{or} \quad S \triangleq \{x \in E^n \mid g_i(x) \le 0,\ i = 1, \ldots, m; \\ h_j(x) = 0,\ j = 1, \ldots, n; \\ x \in X\}$$

Local minimum

Definition A point $x^* \in S$ is said to be a *relative minimum point or a local minimum point* of f over S if there is an $\epsilon > 0$ such that $f(x) \geq f(x^*)$ for all $x \in S \cap N(x^*, \epsilon)$, where $N(x^*, \epsilon)$ is the neighborhood of x^* of radius ϵ. If $f(x) > f(x^*)$ for all $x \in S \cap N(x^*, \epsilon)$ and $x \neq x^*$, then x^* is said to be a *strictly relative minimum point* of f over S.

Global minimum

Definition A point $x^*\in S$ is said to be a *global minimum point* of f over S if $f(x)\geq f(x^*)$ for all $x\in S$. If $f(x)>f(x^*)$ for all $x\in S$, $x\neq x^*$, then x^* is said to be a *strictly global minimum point* of f over S.

Comments

- We always intend to seek a global minimum when formulating an optimization problem.

- In most situations, optimization theory and methodologies only enable us to locate local minimums.

- Global optimality can be achieved when certain convexity conditions are imposed.

A general iterative scheme

- A general scheme of an iterative solution procedure:

  Step 1: Start from a feasible solution *x* in S.

  Step 2: Check if the current solution is optimal.
  If the answer is Yes, stop.
  If the answer is No, continue.

  Step 3: Move to a better feasible solution and
  return to Step 2.

What are the feasible moves that lead to a better solution?

- Feasible direction
  - Along any given direction, the objective function can be regarded as a function of a single variable.
  - Given $x \in S \subset E^n$, a vector $d \in E^n$ is a *feasible direction* at x if there is an $\overline{\alpha} > 0$ such that $x + \alpha d \in S$ for all α, $0 \leq \alpha \leq \overline{\alpha}$.
  - A feasible direction is a good direction, if the objective function is reduced along the direction.

How do we know we have attained a minimum solution?

- First order necessary condition

- **Proposition.** Let S be a subset of E^n and let $f \in C^1$ be a function on S. If x^* is a relative minimum point of f over S, then for any $d \in E^n$ that is a feasible direction at x^*, we have $\nabla f(x^*)d \geq 0$.

- **Corollary (Unconstrained case).** Let S be a subset of E^n and let $f \in C^1$ be a function on S. If x^* is a relative minimum point of f over S and if x^* is an interior point of S, then $\nabla f(x^*) = 0$.

Example 1

Example: Constrained problem:

$$\begin{aligned} \min \quad & f(x_1, x_2) = x_1^2 - x_1 + x_2 + x_1 x_2 \\ \text{s. t.} \quad & x_1, x_2 \geq 0 \end{aligned}$$

Check if $x^* = [1/2, 0]$ satisfies the first-order necessary condition or not.

$$\begin{aligned} \nabla f(x) \,|_{x^*} &= [2x_1 - 1 + x_2, 1 + x_1]\big|_{x_1=1/2, x_2=0} \\ &= [0, 3/2] \end{aligned}$$

$\Rightarrow \nabla f(x^*) d \geq 0$ for all d with $d_2 \geq 0$ (feasible direction at x^*).

Example 2

Example: Unconstrained problem:

$$\min f(x_1, x_2) = x_1^2 - x_1 x_2 + x_2^2 - 3x_2$$

Global minimum is known at $x_1 = 1$, $x_2 = 2$.

At this point,

$$\nabla f(x) = [2x_1 - x_2,\ -x_1 + 2x_2 - 3]$$

$$= [0, 0]$$

Comments

- The necessary conditions in the pure unconstrained case lead to a system of *n* equations in *n* unknowns.

- Is the condition a sufficient condition? Why?

- How about the condition of

$$\nabla f(x^*)d > 0?$$

Proof of the proposition

If $\exists$ a feasible direction $d \in E^n$ at x^* with $\nabla f(x^*)d < 0$, then $\exists\ \bar{\alpha} > 0$ s.t. $x(\alpha) = x^* + \alpha d \in S$ with $0 < \alpha < \bar{\alpha}$ and

$$\begin{aligned} f(x(\alpha)) &= f(x^*) + \nabla f(x^*)(x(\alpha) - x^*) + O(\alpha^2) \\ &= f(x^*) + \alpha \nabla f(x^*)d + O(\alpha^2) \\ &< f(x^*)\ , \quad \text{if } \alpha \text{ is sufficiently small.} \end{aligned}$$

This contradicts to the fact that x^* is a local minimum point of f over S.

Corollary – Variational Inequalities

- **Proposition:** Let $S \subset E^n$ be convex and $f : E^n \rightarrow R$ be $C^1(S)$. If x^* is a relative minimum point of f over S, then x^* is a solution of the following variational inequality problem:

$$
\begin{array}{lll}
 & \text{Find} & x \in S \\
(VI) & \text{s. t.} & \langle x' - x, \nabla f(x) \rangle \geq 0, \\
 & & \forall\, x' \in S.
\end{array}
$$

Second order conditions

Proposition (Second-order necessary conditions). Let S be a subset of E^n and let $f \in C^2$ be a function on S. If x^* is a relative minimum point of f over S, then for any $d \in E^n$ that is a feasible direction at x^*, we have

(i) $\nabla f(x^*)d \geq 0$.

(ii) if $\nabla f(x^*)d = 0$, then $d^T \nabla^2 f(x^*)d \geq 0$.

Proof:

$$f(x(\alpha)) = f(x^*) + \tfrac{1}{2}(x(\alpha) - x^*)^T \nabla^2 f(x^*)(x(\alpha) - x^*) + O(\alpha^3)$$
$$= f(x^*) + \tfrac{1}{2}\alpha^2 d^T \nabla^2 f(x^*)d + O(\alpha^3)$$

Example 3

Example: Constrained problem:

$$\begin{aligned} \min \quad & f(x_1, x_2) = x_1^2 - x_1 + x_2 + x_1 x_2 \\ \text{s. t.} \quad & x_1, x_2 \geq 0 \end{aligned}$$

Check if $x^* = [1/2,\ 0]$ satisfies the second-order necessary condition or not.

$\nabla f(x) \mid_{x^*} = [0,\ 3/2]$, since $\nabla f(x^*)d = 3/2 d_2 = 0$

$\Rightarrow d_2 = 0$

$\Rightarrow d^T \nabla^2 f(x^*) d = 2d_1^2 \geq 0$

Second order necessary condition

- **Proposition (Second-order necessary conditions – unconstrained case).** Let x^* be an interior point of the set S, and suppose x^* is a relative minimum point of $f \in C^2$. Then

  (i) $\nabla f(x^*) = 0$.

  (ii) $F(x^*)$ is positive semidefinite.

Example 4

Example: Unconstrained problem:

$$\min f(x_1, x_2) = x_1^2 - x_1 x_2 + x_2^2 - 3x_2$$

Global minimum is known at $x_1 = 1,\ x_2 = 2.$

At this point,

$$\begin{aligned}\nabla f(x) &= [2x_1 - x_2,\ -x_1 + 2x_2 - 3] \\ &= [0, 0]\end{aligned}$$

and $F(x)$ is positive definite.

Example 5

<u>**Example:**</u> Constrained problem:

$$\begin{aligned} \min \quad & f(x_1, x_2) = x_1^3 - x_1^2 x_2 + 2x_2^2 \\ \text{s. t.} \quad & x_1, x_2 \geq 0 \end{aligned}$$

$x^* = [6, 9]$ is a solution to the first-order necessary condition:

$$\nabla f(x) \mid_x = [3x_1^2 - 2x_1 x_2, -x_1^2 + 4x_2] = 0$$

But, x^* does not satisfy the second-order necessary condition,

$$F = \begin{bmatrix} 6x_1 - 2x_2 & -2x_1 \\ -2x_1 & 4 \end{bmatrix} \Big|_{x^*} = \begin{bmatrix} 18 & -12 \\ -12 & 4 \end{bmatrix}$$

Second order sufficient condition

- **Proposition (Second-order sufficient conditions – unconstrained case).** Let $f \in C^2$ be a function on a region in which the point x^* is an interior point. Suppose in addition that

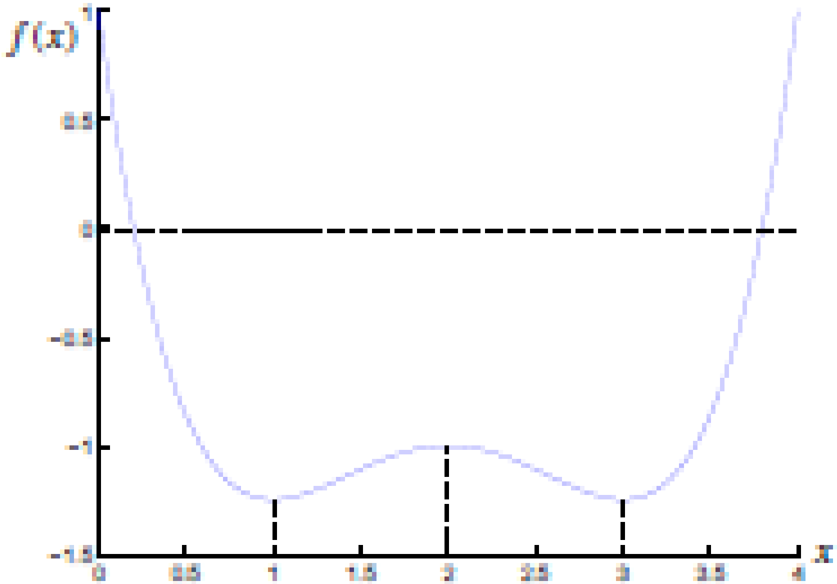
  (i) $\nabla f(x^*) = 0$.

  (ii) $F(x^*)$ is positive definite.

  Then x^* is a strictly relative minimum point of f.

Example 6

$$\begin{array}{ll} \text{Min} & f(x) = \frac{1}{4}x^4 - 2x^3 + \frac{11}{2}x^2 - 6x + 1 \\ \text{s. t.} & 0 \le x \le 4. \end{array}$$



Continue

- First-order information:

$$f'(x) = x^3 - 6x^2 + 11x - 6 = (x-1)(x-2)(x-3).$$

$$f'(0) = -6, \; f'(1) = f'(2) = f'(3) = 0, \; f'(4) = 6.$$

- Second-order information:

$$f''(x) = 3x^2 - 12x + 11$$

$$\Rightarrow \; f''(1) > 0, \; f''(2) < 0, \; f''(3) > 0.$$

By checking the 1st-order necessary conditions, only $x = 1$, $x = 2$ and $x = 3$ are satisfied.

By checking the 2nd-order necessary conditions, only $x = 1$ and $x = 3$ are satisfied.

By checking the 2nd-order sufficient conditions, we know $x^* = 1$ or 3 with $f(x^*) = -1.25$.

Convex functions - definition

- Let $\Omega \subset E^n$ be a convex set and $f : \Omega \to R$ be a real-valued function. Then f is convex on Ω, if

$$f(\alpha x^1 + (1-\alpha)x^2) \;\leq\; \alpha f(x^1) + (1-\alpha) f(x^2)$$

$$\forall\, x^1, x^2 \in \Omega \quad \text{and} \quad \alpha \in [0,1].$$

Moreover, f is strictly convex on Ω, if

$$f(\alpha x^1 + (1-\alpha)x^2) \;<\; \alpha f(x^1) + (1-\alpha) f(x^2)$$

$$\forall\, x^1 \neq x^2, \quad x^1, x^2 \in \Omega \quad \text{and} \quad \alpha \in (0,1).$$

Concave functions

- $g : \Omega \longrightarrow R$ is (strictly) concave on Ω, if $f = -g$ is (strictly) convex on Ω.

Graph and epigraph of a function

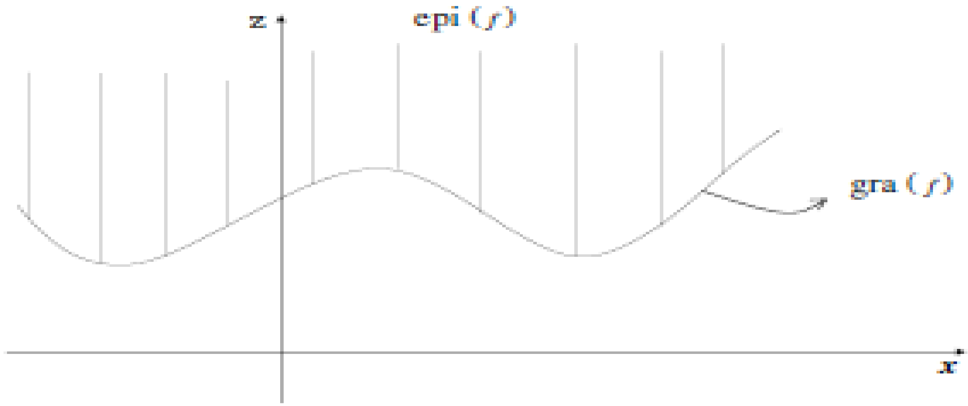
- Let $\Omega \subset E^n$ and $f : \Omega \to R$.

  The graph of f is

  $$gra(f) \triangleq \{(x, z) \in E^{n+1} \mid x \in \Omega \ and \ f(x) = z\}$$

  The epigraph of f is

  $$epi(f) \triangleq \{(x, z) \in E^{n+1} \mid x \in \Omega \ and \ f(x) \leq z\}$$



Set based definition of convex functions

- Definition

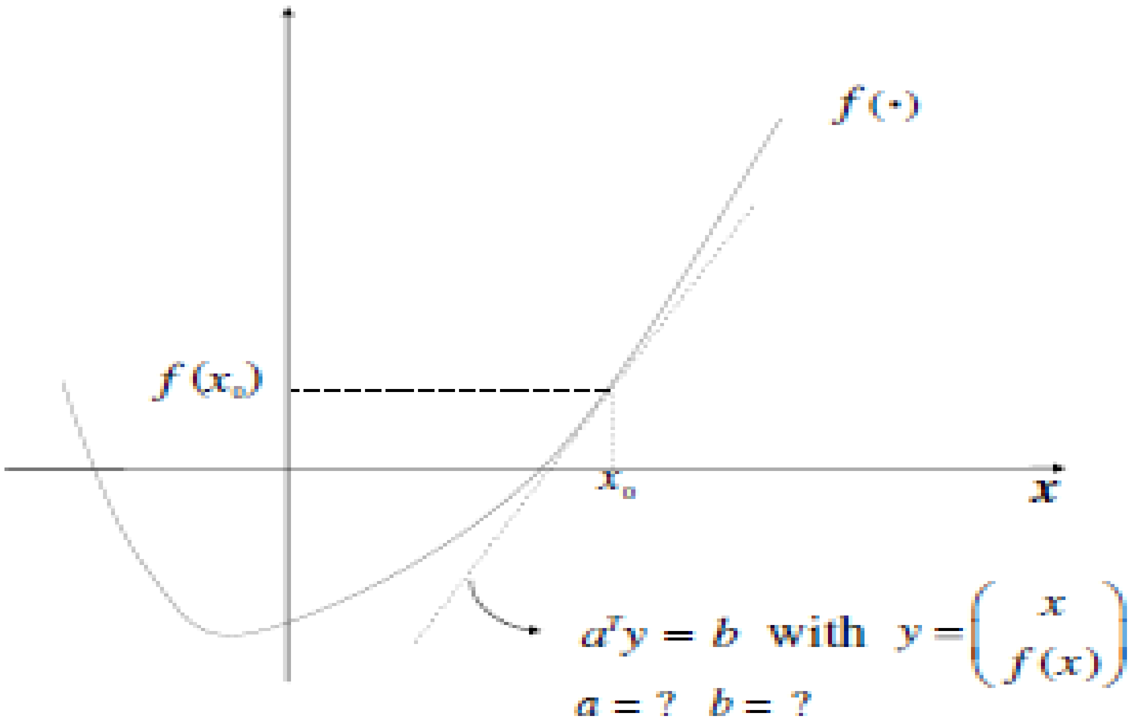
- A function $f : \Omega \subset E^n \rightarrow R$ is convex if $epi(f)$ is a convex subset of E^{n+1}.

- **Theorem:**

  For a convex function f, if each point in $gra(f)$ is an extreme point of $epi(f)$, then the function f is strictly convex.

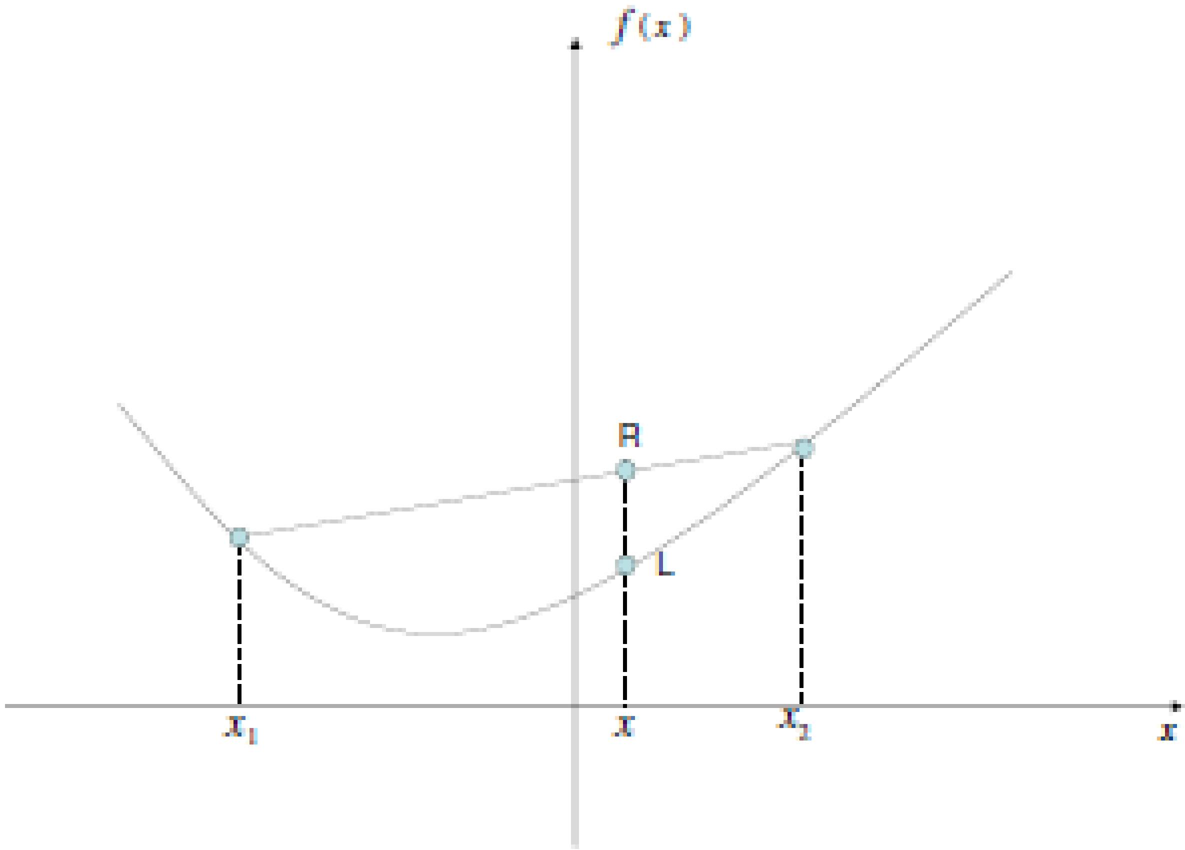
Question

Let $f : \Omega \subset E^n \rightarrow R$ be convex and $f \in C^1(\Omega)$.
For $x^0 \in \Omega$, what's the supporting hyperplane of $epi(f)$ at $(x^0, f(x^0))$



Basic property - 1

- Overestimate by two-point information



Basic property - 2

- **Theorem:**

  Let f be a convex function on a convex set $\Omega \subset E^n$.

  Then

$$f\left(\sum_{i=1}^{m} \alpha_i x^i\right) \leq \sum_{i=1}^{m} \alpha_i \, f(x^i)$$

  $\forall x^i \in \Omega, \quad \alpha_i \in [0,1] \quad and \quad \sum_{i=1}^{m} \alpha_i = 1$

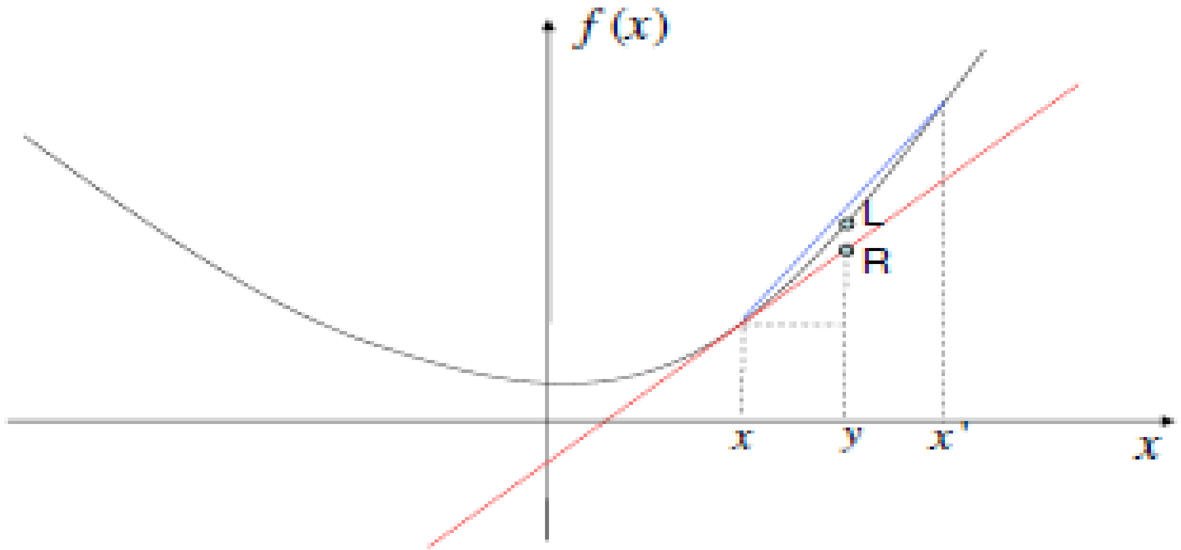
  (Jensen's inequality)

Basic property - 3

- **Theorem:**

  Let $f \in C^1$. Then f is convex on a convex set $\Omega \subset E^n$ if, and only if,

$$f(y) \geq f(x) + \nabla f(x)(y - x), \quad \forall\, x, y \in \Omega$$



(underestimate by one-point information)

Proof

$(\Rightarrow)$ If f is convex, then for $x, y \in \Omega$,

$f(\alpha y + (1-\alpha)x) \le \alpha f(y) + (1-\alpha) f(x), \; \forall \alpha \in [0,1]$

For $\alpha \neq 0$,

$$\frac{f(x + \alpha(y - x)) - f(x)}{\alpha} \le f(y) - f(x)$$

As $\alpha \longrightarrow 0$, we have

$$\nabla f(x)(y - x) \le f(y) - f(x)$$

Proof

($\Leftarrow$) Assume that

$$f(y) \;\geq\; f(x) + \nabla f(x)(y - x), \quad \forall x, y \in \Omega$$

Given $x^1, x^2 \in \Omega$, and any $\bar{\alpha} \in [0, 1]$.

Consider $\bar{x} = \bar{\alpha}x^1 + (1 - \bar{\alpha})x^2$, then

$$f(x^1) \;\geq\; f(\bar{x}) + \nabla f(\bar{x})(x^1 - \bar{x})$$

$$f(x^2) \;\geq\; f(\bar{x}) + \nabla f(\bar{x})(x^2 - \bar{x})$$

Multiplying the first by $\bar{\alpha}$ and the second by $1 - \bar{\alpha}$ and adding up, we have

$$\begin{aligned}\bar{\alpha}f(x^1)+(1-\bar{\alpha})f(x^2) &\geq f(\bar{x})+\nabla f(\bar{x})(\bar{\alpha}x^1+(1-\bar{\alpha})x^2-\bar{x}) \\ &= f(\bar{\alpha}x^1 + (1 - \bar{\alpha})x^2) + \nabla f(\bar{x})(0) \\ &= f(\bar{\alpha}x^1 + (1 - \bar{\alpha})x^2)\end{aligned}$$

Basic properties - 4 and 5

- **Theorem:**

  Let $\Omega \subset E^n$ be a convex set, $f_1, f_2 : \Omega \rightarrow R$ be convex functions.

  Then (i) $f_1 + f_2$ is convex on Ω

  (ii) βf_1 is convex on Ω, $\forall\ \beta \geq 0$

- **Theorem:**

  Let f be a convex function on a convex set $\Omega \subset E^n$. Then the set

  $I_c \triangleq \{x \in \Omega \mid f(x) \leq c\}$ is convex, $\forall\ c \in R$.

Basic property - 6

- **Theorem:**

  Let $f \in C^2$ and $\Omega \subset E^n$ is convex with $int(\Omega) \neq \phi$. Then f is convex on Ω, if and only if, the Hessian matrix F is positive semidefinite over Ω.

Proof

By Taylor's Theorem,

$$f(y) = f(x) + \nabla f(x)(y - x)$$
$$+ \frac{1}{2}(y-x)^T F(x+\alpha(y-x))(y-x)$$

for some $\alpha \in [0, 1]$.

Additional properties

- **Theorem:**
  Let $S \subset E^n$ be convex and $f : S \to R$.
  Then f is (strictly) convex if, and only if, $g(s) \triangleq f(x^0 + sd)$ is (strictly) convex on $I \triangleq \{s \in R \mid x^0 + sd \in S\}$ for any given $x^0 \in S$ and $d \in E^n$.

- **Theorem:**
  Let f be (strictly) convex on $S \subset E^n$ and $x = My + b$ is an affine transformation from E^m to E^n. Then $g(y) \triangleq f(My + b)$ is (strictly) convex on $\{y \in E^m \mid My + b \in S\}$, if M has full rank.

Additional properties

- **Theorem:**

  Let f_j, $j = 1, \ldots, p$, be convex on $S \subset E^n$ and $\alpha_j \geq 0$. Then $f \triangleq \sum_{j=1}^{p} \alpha_j f_j$ is convex on S. In addition, if $\exists$ i such that f_i is strictly convex on S and $\alpha_i > 0$, then $f \triangleq \sum_{j=1}^{p} \alpha_j f_j$ is strictly convex on S.

Additional properties

- **Theorem:**
  Let $f_j,\ j = 1, 2, \ldots,$ be convex on $S \subset E^n$.
  If $\lim\limits_{j\to\infty} f_j(x)$ exists for each $x \in S$, then
  $f(x) \triangleq \lim\limits_{j\to\infty} f_j(x)$ is convex on S.
- **Theorem:**
  Let Ω be an index set and $\{f_w \mid w \in \Omega\}$ be a family of convex functions on $S \subset E^n$.
  Then, $f(x) \triangleq \sup\limits_{w\in\Omega} f_w(x)$ is convex on $\{x \in S \mid \sup\limits_{w\in\Omega} f_w(x) < +\infty\}$. In addition, if Ω is finite and f_w is strictly convex for each $w \in \Omega$, then f is strictly convex on S.

Additional properties

- **Theorem:**

  Let f_1 be convex on $S_1 \subset E^n$ and f_2 be convex and non-decreasing on a set $T \supset f_1(S_1)$. Then the composition function $f_2 \circ f_1\ (x) \triangleq f_2(\ f_1(x)\ )$ is convex on S_1.

  In addition, if f_1 is strictly convex on S_1 and f_2 is increasing, then $f_2 \circ f_1$ is strictly convex on S_1.

Minimization of convex functions

- **Theorem:**

  Let f be a convex function defined on the convex set S. Then any relative minimum of f is a global minimum and the set τ where f achieves its minimum is convex.

Proof

(*i*) If $x^*\in\Omega$ is a local minimum and $\exists\ y\in\Omega$ with $f(y)<f(x^*)$, then

$f(\alpha y+(1-\alpha)x^*)\leq \alpha f(y)+(1-\alpha)f(x^*)<f(x^*)$

for $\alpha\in(0,1)$

This contradicts to the fact that x^* is a local minimum.

(*ii*) $\tau=\{x\mid f(x)\leq f(x^*),\quad x\in\Omega\}$ is obviously convex.

Sufficient and necessary conditions

- For convex functions, the first order necessary condition is also a sufficient condition.
- <u>**Theorem:**</u>

  Let $f \in C^1$ be convex on a convex set $\Omega \subset E^n$. If $\exists\, x^* \in \Omega$, s.t.

$$\nabla f(x^*)(y - x^*) \geq 0, \quad \forall y \in \Omega$$

  then x^* is a global minimum of f over Ω

Proof

Proof: Since

$$f(y) \geq f(x^*) + \nabla f(x^*)(y - x^*) \geq f(x^*), \ \forall y \in \Omega,$$

and any $y \in \Omega$ can be reached from x^* along

a feasible direction $y - x^*$.

Example

- **Example:** Check the convexity of the following optimization problem and find its (global) minimum.

$$\min f(x_1, x_2, x_3) = 4x_1^2 + 3x_2^2 + 5x_3^2 + 6x_1x_2 + x_1x_3 - 3x_1 - 2x_2$$

Maximization of convex functions

- **Theorem:**

  Let f be a convex function defined on the bounded, closed convex set $\Omega \subset E^n$. If f achieves global maximum on Ω, then one maximizer falls in $bdry\ (\Omega)$.

Proof

Assume $x^* \in \Omega$ is a global maximizer of f. If x^* is not a boundary point of Ω, then

$$\exists\, x^1, x^2 \in \ bdry(\Omega)$$

s.t.

$$x^* = \alpha x^1 + (1-\alpha)x^2 \text{ for some } \alpha \in (0,1)$$

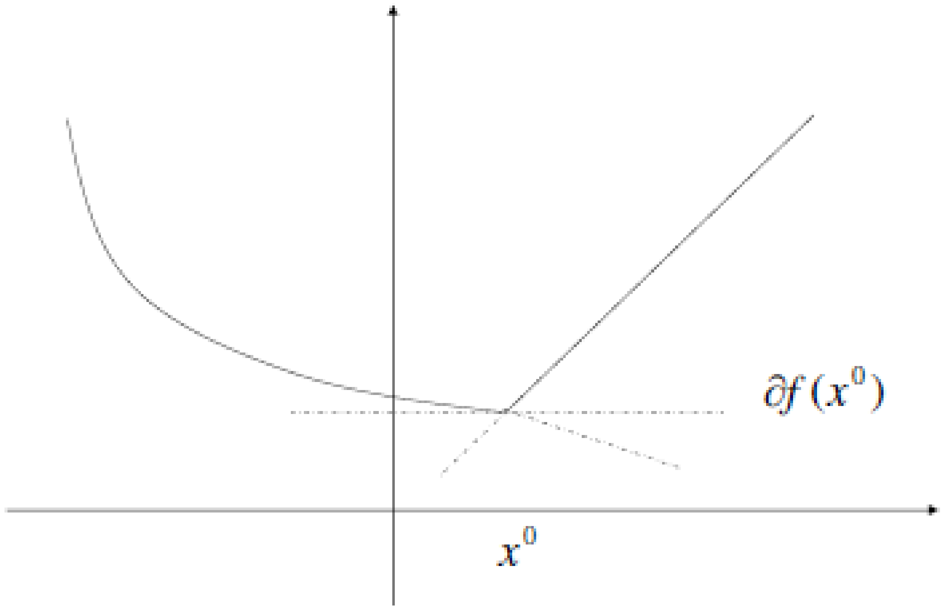
By convexity of f,

$$\begin{aligned} f(x^*) &\le \alpha f(x^1) + (1-\alpha) f(x^2) \\ &\le \max\{f(x^1), f(x^2)\} \end{aligned}$$

Therefore either x^1 or x^2 is a global maximizer.

Non-differentiable convex functions

- Where is the first order information?
  - subgradient and subdifferential



Subgradient and subdifferential

- Definition

  A vector *y* is said to be a subgradient of a convex function *f* (over a set *S*) at a point x^0 if

$$f(x) \geq f(x^0) + \langle y, x - x^0 \rangle, \forall x \in S$$

- Definition

  The set of all subgradients of *f* at x^0 iis called the subdifferential of *f* at x^0 and is denoted by

$$\partial f(x^0) = \{ y \in E^n \mid f(x) \geq f(x^0) + \langle y, x - x^0 \rangle, \forall x \in S \}$$

Properties

1. The graph of the affine function

$$h(x) = f(x^0) + \langle y, x - x^0 \rangle$$

is a non-vertical supporting hyperplane to the convex set *epi*(*f*) at the point of (x^0, $f(x^0)$).

2. The subdifferential set $\partial f(x^0)$ is closed and convex.

3. $\partial f(x^0)$ can be empty, singleton, or a set with infinitely many elements. When it is not empty, *f* is said to be subdifferentiable at x^0.

4. $\nabla f(x^0) \in \partial f(x^0)$ if f is differentiable at x^0.

   $\{\nabla f(x^0)\} = \partial f(x^0)$ if f is convex and differentiable at $x^0 \in int(S)$.

Examples

- In *R*, $f(x) = |x|$ is subdifferentiable at every point and

  $$\partial f(0) = [-1, 1].$$

- In E^n, the Euclidean norm $f(x) = ||x||$ is subdifferentiable at every point and $\partial f(0)$ consists of all the vectors *y* such that

  $$||x|| \geq \langle y, x \rangle \quad \text{for all } x.$$

  This means the Euclidean unit ball !

Elementary Solution Methods

- Two elementary solution methods for the unconstrained optimization problem

$$Minimize\ f(\boldsymbol{x}) \text{ for } \mathbf{x} \in \boldsymbol{E}^n$$

1. First order information based

   Gradient descent method

   (Steepest decent method)

2. Second order information based

   Newton's method

   (Newton-Raphson method)

Gradient Decent Method

- Motivation:

  - Negative gradient direction is the steepest decent direction to reduce the function value.

- Gradient Decent Method

  1. Start from a guess point $\boldsymbol{x_0} \in E^n$ (or S) and set $k = 0$.
  2. At iterate $\boldsymbol{x}_k$, move along the negative gradient direction with an appropriate step size $\gamma_k > 0$ to generate the next iterate

     $$\boldsymbol{x}_{k+1} = \mathbf{x_k} - \gamma_k[\nabla f(x_k)]^T$$

     and reset $k \leftarrow k + 1$
  3. Return to 2 until $\boldsymbol{x}_{k+1} \approx \boldsymbol{x}_k$

Newton's Method

- Motivation – one variable

$$f(x) : E^1 \rightarrow R \text{ and } f \in C^2$$

$$\min_{x \in R} f(x)$$

- At a point $x_k \in E^1$, make the 2nd order Taylor expansion of f

$$f(x_k + t) \approx f(x_k) + f'(x_k)t + \frac{1}{2} f''(x_k)t^2$$

- If $f''(x_k) \geq 0$, then the quadratic approximation is a convex function of variable $t \in R$, and its minimum can be found by setting the derivative to be 0.

Newton's Method

- $$0 = \frac{d}{dt}\left(f(x_k) + f'(x_k) + \frac{1}{2}f''(x_k)t^2\right.$$
$$= f'(x_k) + f''(x_k)t$$

- The minimum is achieved at $t = -\frac{f\prime(x_k)}{f''(x_k)}$.

  Hence, we take

$$x_k = x_k + t = x_k - \frac{f\prime(x_k)}{f''(x_k)}$$

Newton's Method

- Extension to multi-variable

$$f(\boldsymbol{x}) : E^n \to R \text{ and } f \in C^2$$

$$\min_{x \in E^n} f(\boldsymbol{x})$$

- Newton step

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - F(\boldsymbol{x}_k)^{-1} \, [\nabla f(\boldsymbol{x}_k)]^T$$

- Newton's Method

1. Start from a guess point $x_0 \in E^n$ (or S), and set $k = 0$.
2. At iterate x_k, take a Newton step to generate the next iterate $\boldsymbol{x}_{k+1}$, and reset $k \leftarrow k + 1$
3. Return to 2 until $\boldsymbol{x}_{k+1} \approx \boldsymbol{x}_k$

Damped Newton's Method

-
- Replace the Newton step

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - F(\boldsymbol{x}_k)^{-1} [\nabla f(\boldsymbol{x}_k)]^T$$

by a modified Newton step

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \gamma F(\boldsymbol{x}_k)^{-1} [\nabla f(\boldsymbol{x}_k)]^T$$

for $\gamma \in (0,1]$

where γ is taken to satisfy the Wolfe conditions or Armijio's condition.

Performance

1. Newton's method is not necessarily convergent.

2. When it converges, it may converge to a local minimum, or a saddle point.
3. Depends on the existence and computations of the inverse of Hessian.
4. Convergence result:

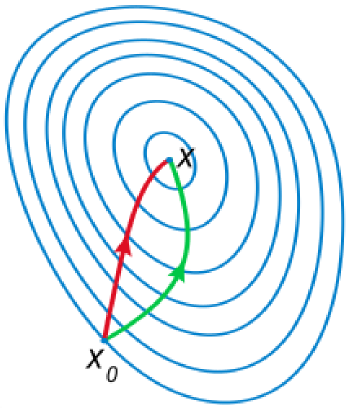
   When f *is strongly convex with Lipschitz Hessian*, provided that x_0 *is close enough to the minimizer* x^*, the Newton's method converges to the minimizer in a quadratic rate, i.e., $\|x_{k+1} - x^*\| \leq \frac{1}{2}\|x_k - x^*\|^2, \forall\, k \geq 0$

Comments

- Newton step takes a "conditioned" negative gradient direction, when the second-order information is available.

- When it works, the Newton's method take "more computational efforts" to determine a "better direction" that may move in a "more straight-forward" path to reach the destination.

See gradient decent method
vs. Newton's method



https://en.wikipedia.org/wiki/File:Newton_optimization_vs_grad_descent.svg