1. Concept of Optimization

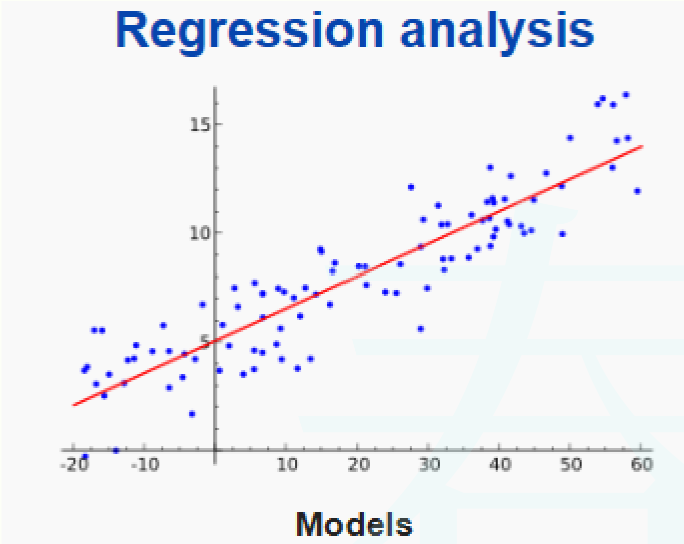
Solve Equation

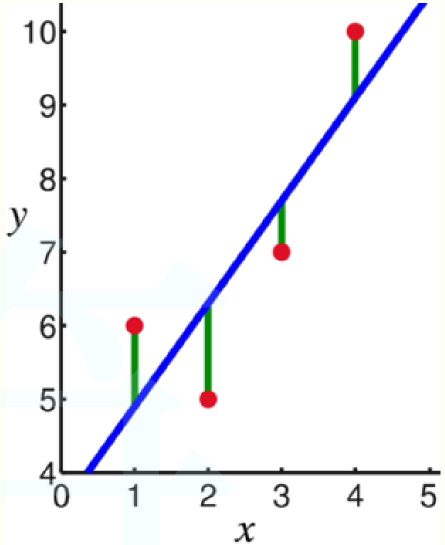
$$\begin{cases} a_{1,1}x_1 + a_{1,2}x_2 + \cdots a_{1,n}x_n = b_1 \\ a_{2,1}x_1 + a_{2,2}x_2 + \cdots a_{2,n}x_n = b_2 \\ \vdots \\ a_{\mathrm{m},1}x_1 + a_{\mathrm{m},2}x_2 + \cdots a_{\mathrm{m},n}x_n = b_m \end{cases}$$

Solve Equation

$$\begin{cases} f_1(\boldsymbol{x}) = b_1 \\ f_2(\boldsymbol{x}) = b_2 \\ \vdots \\ f_m(\boldsymbol{x}) = b_m \end{cases} \longrightarrow \sum_{i=1}^{m} \left(f_i(\boldsymbol{x}) - b_i \right)^2$$

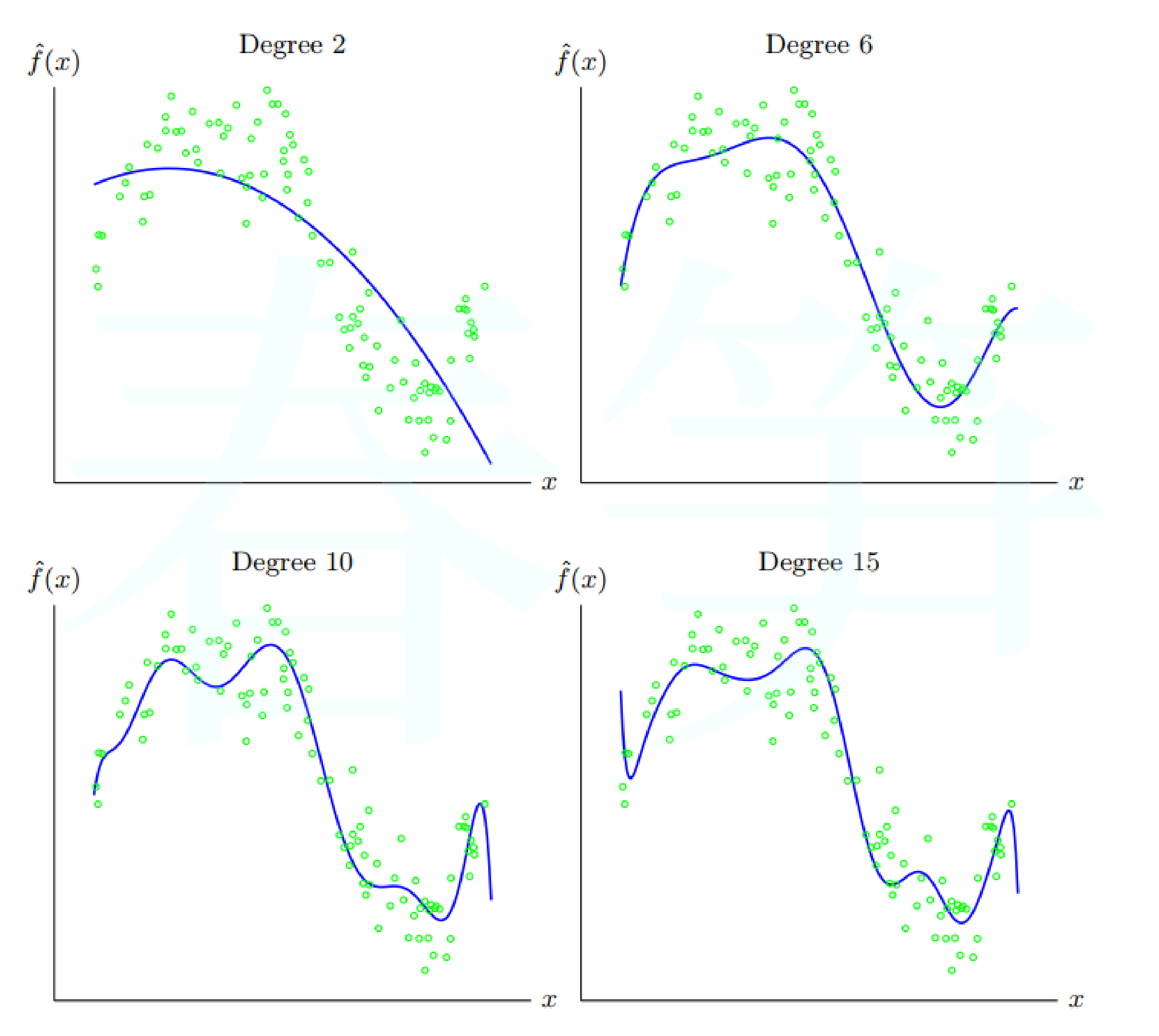
Linear Regression Analysis



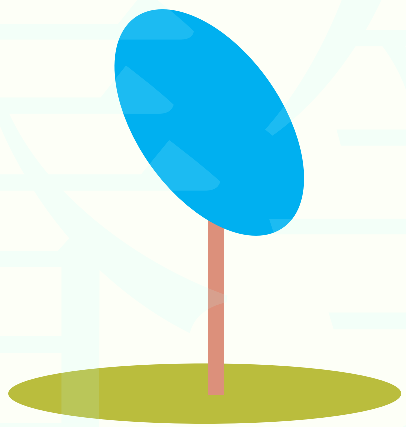


$$y_i = \beta_0 1 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \qquad i = 1, \ldots, n$$

Curve Fitting(Model Selection)



Measure Distance



General Optimization

$$\underset{\boldsymbol{x}\in S}{\text{minimize}}\ f\left(\boldsymbol{x}\right)$$

$f\left(\boldsymbol{x}\right)$: Objective/Cost Function

$\boldsymbol{x}$:Decision Variable

$\boldsymbol{S}$:Constrained Set/Feasible Region

$$\text{minimize } f(\boldsymbol{x})$$

subject to:

$$a_i(\boldsymbol{x}) = 0 \qquad \forall \ i=1,2,\ldots,p$$

$$c_j(\boldsymbol{x}) \geq 0 \qquad \forall \ j=1,2,\ldots,q$$

$a_i(\boldsymbol{x})$: equality constraints

$c_j(\boldsymbol{x})$: inequality constraints

Feasible Region

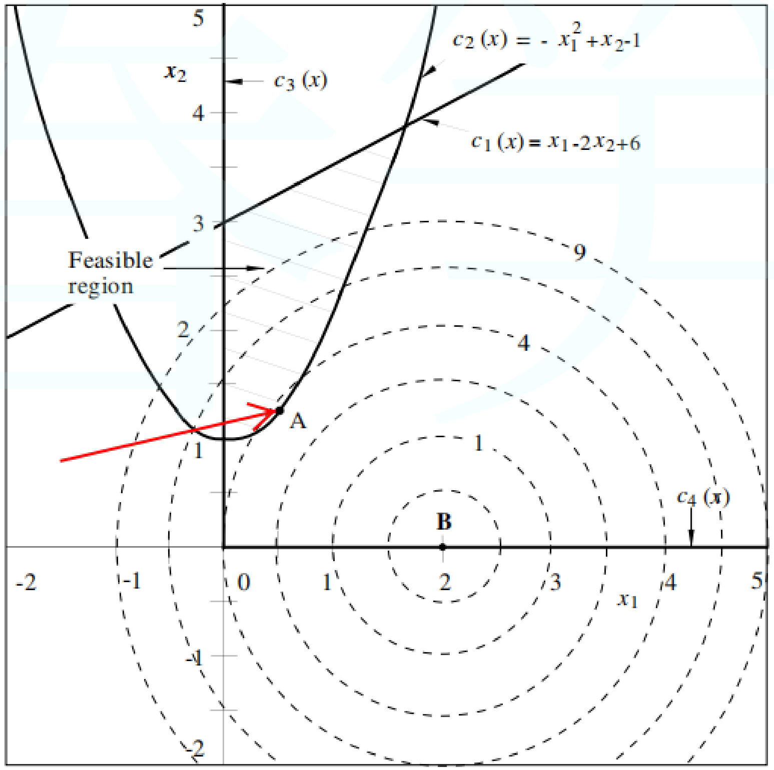
$$\mathcal{R} = \{\mathbf{x} : a_i(\mathbf{x}) = 0 \text{ for } i = 1, 2, \ldots, p \text{ and } c_j(\mathbf{x}) \geq 0 \text{ for } j = 1, 2, \ldots, q\}$$

where $\mathcal{R} \subset E^n$.

1. Interior points
2. Boundary points
3. Exterior points

An*interior point* is a point for which $c_j(\mathbf{x}) > 0$ for all j. A *boundary point* is a point for which at least one $c_j(\mathbf{x}) = 0$, and an *exterior point* is a point for which at least one $c_j(\mathbf{x}) < 0$. Interior points are feasible points, boundary points may or may not be feasible points, whereas exterior points are nonfeasible points.

$$
\begin{aligned}
\text{minimize } f(\mathbf{x}) &= x_1^2 + x_2^2 - 4x_1 + 4 \\
\text{subject to:} \quad c_1(\mathbf{x}) &= x_1 - 2x_2 + 6 \geq 0 \\
c_2(\mathbf{x}) &= -x_1^2 + x_2 - 1 \geq 0 \\
c_3(\mathbf{x}) &= x_1 \geq 0 \\
c_4(\mathbf{x}) &= x_2 \geq 0
\end{aligned}
$$



Gradient

$$\nabla f = \boldsymbol{g}(\boldsymbol{x}) = \begin{bmatrix} \dfrac{\partial f}{\partial x_1} & \dfrac{\partial f}{\partial x_2} & \cdots & \dfrac{\partial f}{\partial x_n} \end{bmatrix}^T$$

Hessian

$$\boldsymbol{H}(\boldsymbol{x}) = \nabla \boldsymbol{g}^T = \nabla \left\{ \nabla^T f(\boldsymbol{x}) \right\} = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_1 \partial x_n} \\ \dfrac{\partial^2 f}{\partial x_1 \partial x_2} & \dfrac{\partial^2 f}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_1 \partial x_n} & \dfrac{\partial^2 f}{\partial x_2 \partial x_n} & \cdots & \dfrac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

The Taylor Series

$$f(\mathbf{x}+\boldsymbol{\delta}) = f(\mathbf{x}) + \mathbf{g}(\mathbf{x})^T\boldsymbol{\delta} + \tfrac{1}{2}\boldsymbol{\delta}^T\mathbf{H}(\mathbf{x})\boldsymbol{\delta} + o(\|\boldsymbol{\delta}\|^2)$$

Linear Approximation

$$f(\mathbf{x}+\boldsymbol{\delta}) \approx f(\mathbf{x}) + \mathbf{g}(\mathbf{x})^T\boldsymbol{\delta}$$

Quadratic Approximation

$$f(\mathbf{x}+\boldsymbol{\delta}) \approx f(\mathbf{x}) + \mathbf{g}(\mathbf{x})^T\boldsymbol{\delta} + \tfrac{1}{2}\delta^T\mathbf{H}(\mathbf{x})\boldsymbol{\delta}$$

Types of Extrema

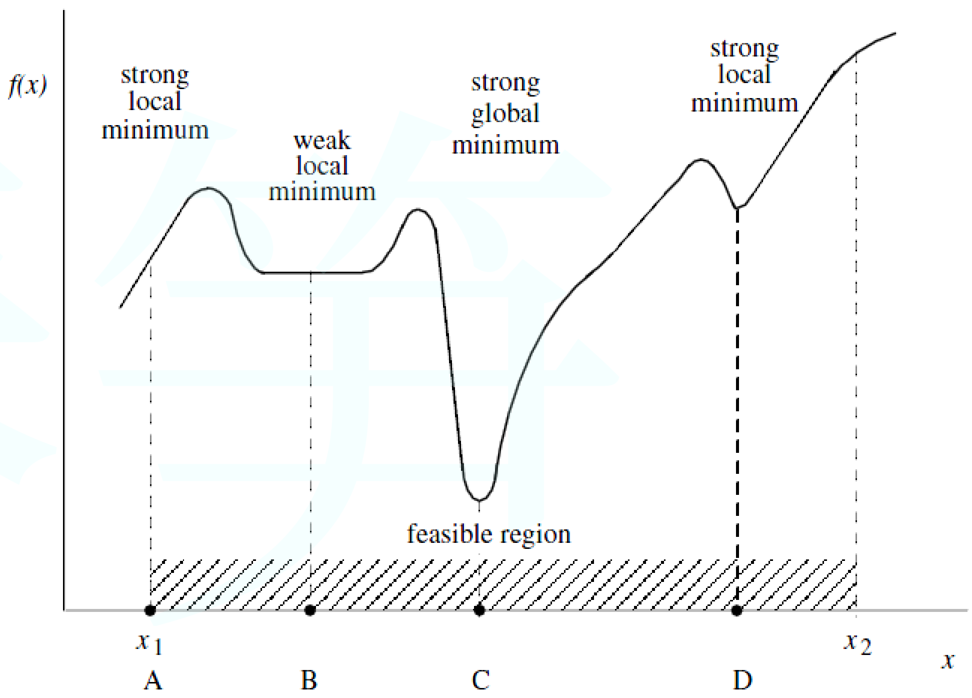
weak local minimizer:

$$f(\mathbf{x}) \geq f(\mathbf{x}^*)$$

$$\mathbf{x} \in \mathcal{R} \quad \text{and} \quad \|\mathbf{x} - \mathbf{x}^*\| < \varepsilon$$

strong local minimizer:

$$f(\mathbf{x}) > f(\mathbf{x}^*)$$



Necessary and Sufficient Conditions for Local Minima

First-order necessary conditions

(a) If $f(\mathbf{x}) \in C^1$ *and* $\mathbf{x}^*$ *is a local minimizer, then*

$$\mathbf{g}(\mathbf{x}^*)^T \mathbf{d} \geq 0$$

for every feasible direction $\mathbf{d}$ *at* $\mathbf{x}^*$.

(b) If $\mathbf{x}^*$ *is located in the interior of* $\mathcal{R}$ *then*

$$\mathbf{g}(\mathbf{x}^*) = \mathbf{0}$$

Proof (a) If $\mathbf{d}$ is a feasible direction at $\mathbf{x}^*$, then from Def. 2.4

$$\mathbf{x} = \mathbf{x}^* + \alpha\mathbf{d} \in \mathcal{R} \qquad \text{for } 0 \leq \alpha \leq \hat{\alpha}$$

From the Taylor series

$$f(\mathbf{x}) = f(\mathbf{x}^*) + \alpha\mathbf{g}(\mathbf{x}^*)^T\mathbf{d} + o(\alpha\|\mathbf{d}\|)$$

If

$$\mathbf{g}(\mathbf{x}^*)^T\mathbf{d} < 0$$

then as $\alpha \to 0$

$$\alpha\mathbf{g}(\mathbf{x}^*)^T\mathbf{d} + o(\alpha\|\mathbf{d}\|) < 0$$

and so

$$f(\mathbf{x}) < f(\mathbf{x}^*)$$

(b) If $\mathbf{x}^*$ is in the interior of $\mathcal{R}$, vectors exist in all directions which are feasible. Thus from part (a), a direction $\mathbf{d} = \mathbf{d}_1$ yields

$$\mathbf{g}(\mathbf{x}^*)^T \mathbf{d}_1 \geq 0$$

Similarly, for a direction $\mathbf{d} = -\mathbf{d}_1$

$$-\mathbf{g}(\mathbf{x}^*)^T \mathbf{d}_1 \geq 0$$

Therefore, in this case, a necessary condition for $\mathbf{x}^*$ to be a local minimizer is

$$\mathbf{g}(\mathbf{x}^*) = 0$$

Second-order necessary conditions

Theorem 2.2 ***Second-order necessary conditions for a minimum***

(*a*) *If* $f(\mathbf{x}) \in C^2$ *and* $\mathbf{x}^*$ *is a local minimizer, then for every feasible direction* $\mathbf{d}$ *at* $\mathbf{x}^*$

(*i*) $\mathbf{g}(\mathbf{x}^*)^T\mathbf{d} \geq 0$

(*ii*) *If* $\mathbf{g}(\mathbf{x}^*)^T\mathbf{d} = 0$, *then* $\mathbf{d}^T\mathbf{H}(\mathbf{x}^*)\mathbf{d} \geq 0$

(*b*) *If* $\mathbf{x}^*$ *is a local minimizer in the interior of* $\mathcal{R}$, *then*

(*i*) $\mathbf{g}(\mathbf{x}^*) = \mathbf{0}$

(*ii*) $\mathbf{d}^T\mathbf{H}(\mathbf{x})^*\mathbf{d} \geq 0$ *for all* $\mathbf{d} \neq \mathbf{0}$

$$f(\mathbf{x}) = f(\mathbf{x}^*) + \alpha \mathbf{g}(\mathbf{x}^*)^T \mathbf{d} + \frac{1}{2}\alpha^2 \mathbf{d}^T \mathbf{H}(\mathbf{x}^*)\mathbf{d} + o(\alpha^2 \|\mathbf{d}\|^2)$$

Now if condition (i) is satisfied with the equal sign, then

$$f(\mathbf{x}) = f(\mathbf{x}^*) + \frac{1}{2}\alpha^2 \mathbf{d}^T \mathbf{H}(\mathbf{x}^*)\mathbf{d} + o(\alpha^2 \|\mathbf{d}\|^2)$$

If

$$\mathbf{d}^T \mathbf{H}(\mathbf{x}^*)\mathbf{d} < 0$$

then as $\alpha \to 0$

$$\frac{1}{2}\alpha^2 \mathbf{d}^T \mathbf{H}(\mathbf{x}^*)\mathbf{d} + o(\alpha^2 \|\mathbf{d}\|^2) < 0$$

and so

$$f(\mathbf{x}) < f(\mathbf{x}^*)$$

This contradicts the assumption that $\mathbf{x}^*$ is a minimizer. Therefore, if $\mathbf{g}(\mathbf{x}^*)^T \mathbf{d} = 0$, then

$$\mathbf{d}^T \mathbf{H}(\mathbf{x}^*)\mathbf{d} \geq 0$$

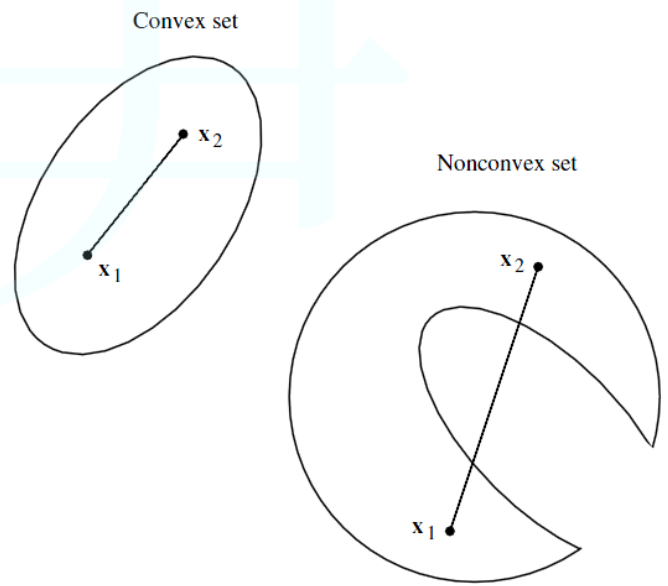
Convex Set and Convex Functions

Definition 2.7

A set $\mathcal{R}_c \subset E^n$ is said to be *convex* if for every pair of points $\mathbf{x}_1,\ \mathbf{x}_2 \subset \mathcal{R}_c$ and for every real number α in the range $0 < \alpha < 1$, the point

$$\mathbf{x} = \alpha \mathbf{x}_1 + (1 - \alpha)\mathbf{x}_2$$

is located in $\mathcal{R}_c$, i.e., $\mathbf{x} \in \mathcal{R}_c$.



Convex Functions

Definition

$f : \mathbf{R}^n \to \mathbf{R}$ is convex if $\mathbf{dom}\, f$ is a convex set and

$$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta) f(y)$$

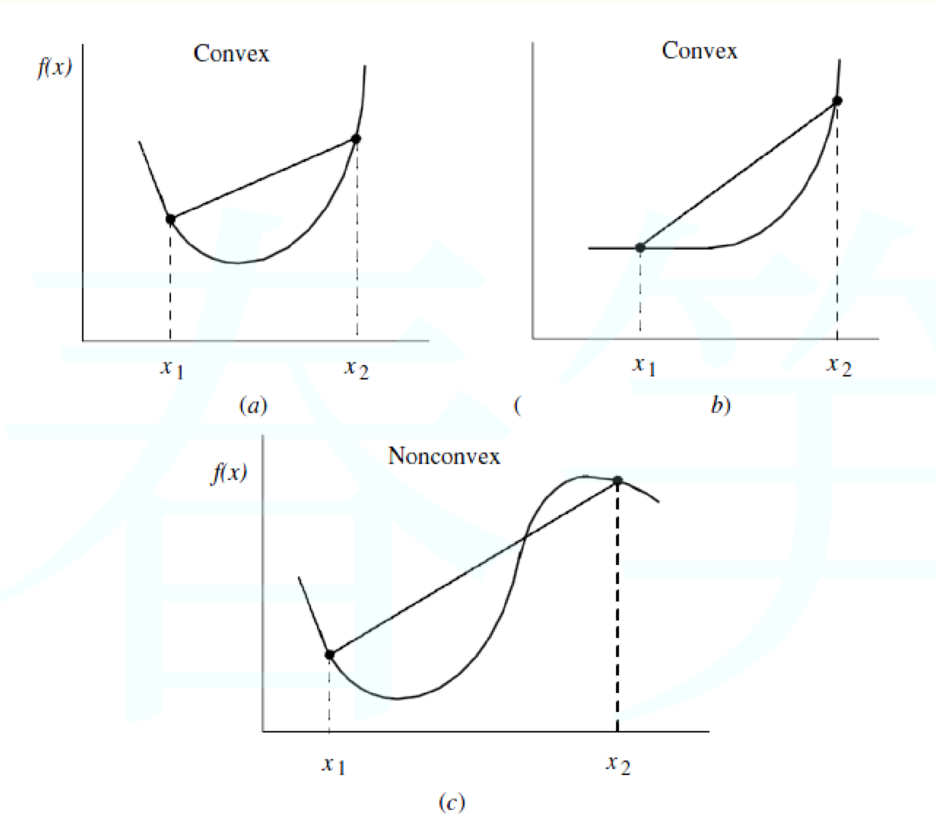
for all $x, y \in \mathbf{dom}\, f$, $0 \le \theta \le 1$

- f is concave if $-f$ is convex
- f is strictly convex if $\mathbf{dom}\, f$ is convex and

$$f(\theta x + (1-\theta)y) < \theta f(x) + (1-\theta) f(y)$$

for $x, y \in \mathbf{dom}\, f$, $x \ne y$, $0 < \theta < 1$

Convex Functions



Examples on R

convex:

- affine: $ax + b$ on **R**, for any $a, b \in$ **R**
- exponential: e^{ax}, for any $a \in$ **R**
- powers: x^α on $\mathbf{R}_{++}$, for $\alpha \geq 1$ or $\alpha \leq 0$
- powers of absolute value: $|x|^p$ on **R**, for $p \geq 1$
- negative entropy: $x \log x$ on $\mathbf{R}_{++}$

concave:

- affine: $ax + b$ on **R**, for any $a, b \in$ **R**
- powers: x^α on $\mathbf{R}_{++}$, for $0 \leq \alpha \leq 1$
- logarithm: $\log x$ on $\mathbf{R}_{++}$

Examples on $\mathbf{R}^n$ and $\mathbf{R}^{m\times n}$

affine functions are convex and concave; all norms are convex

examples on $\mathbf{R}^n$

- affine function $f(x) = a^T x + b$
- norms: $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$ for $p \geq 1$; $\|x\|_\infty = \max_k |x_k|$

examples on $\mathbf{R}^{m\times n}$ ($m \times n$ matrices)

- affine function

$$f(X) = \mathbf{tr}(A^T X) + b = \sum_{i=1}^m \sum_{j=1}^n A_{ij} X_{ij} + b$$

- spectral (maximum singular value) norm

$$f(X) = \|X\|_2 = \sigma_{\max}(X) = (\lambda_{\max}(X^T X))^{1/2}$$

Proposition 6.3.3 *Convex functions satisfy the following:*

(a) *If* $f(\boldsymbol{x})$ *is convex and* $g(\boldsymbol{x})$ *is convex and increasing, then the functional composition* $g \circ f(\boldsymbol{x})$ *is convex.*

(b) *If* $f(\boldsymbol{x})$ *is convex, then the functional composition* $f(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b})$ *of* $f(\boldsymbol{x})$ *with an affine function* $\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$ *is convex.*

(c) *If* $f(\boldsymbol{x})$ *and* $g(\boldsymbol{x})$ *are convex and* α *and* β *are nonnegative constants, then* $\alpha f(\boldsymbol{x}) + \beta g(\boldsymbol{x})$ *is convex.*

(d) *If* $f(\boldsymbol{x})$ *and* $g(\boldsymbol{x})$ *are convex, then* $\max\{f(\boldsymbol{x}), g(\boldsymbol{x})\}$ *is convex.*

(e) *If* $f_m(\boldsymbol{x})$ *is a sequence of convex functions, then* $\lim_{m\to\infty} f_m(\boldsymbol{x})$ *is convex whenever it exists.*

First-order condition

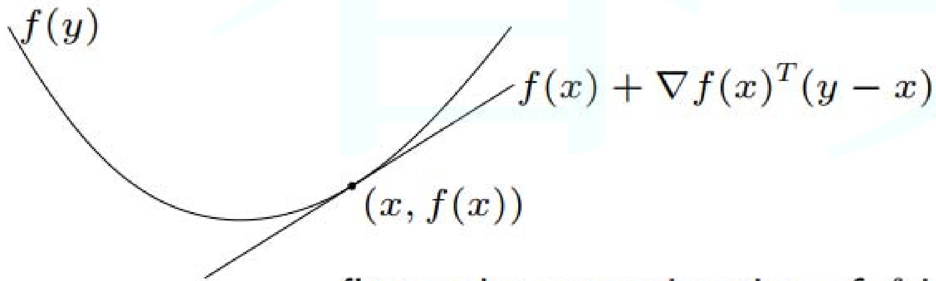
f is **differentiable** if $\mathbf{dom}\, f$ is open and the gradient

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \ldots, \frac{\partial f(x)}{\partial x_n} \right)$$

exists at each $x \in \mathbf{dom}\, f$

1st-order condition: differentiable f with convex domain is convex iff

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \quad \text{for all } x, y \in \mathbf{dom}\, f$$



first-order approximation of f is global underestimator

Second-order conditions

f is **twice differentiable** if $\mathbf{dom}\, f$ is open and the Hessian $\nabla^2 f(x) \in \mathbf{S}^n$,

$$\nabla^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, \quad i, j = 1, \ldots, n,$$

exists at each $x \in \mathbf{dom}\, f$

2nd-order conditions: for twice differentiable f with convex domain

- f is convex if and only if

$$\nabla^2 f(x) \succeq 0 \quad \text{for all } x \in \mathbf{dom}\, f$$

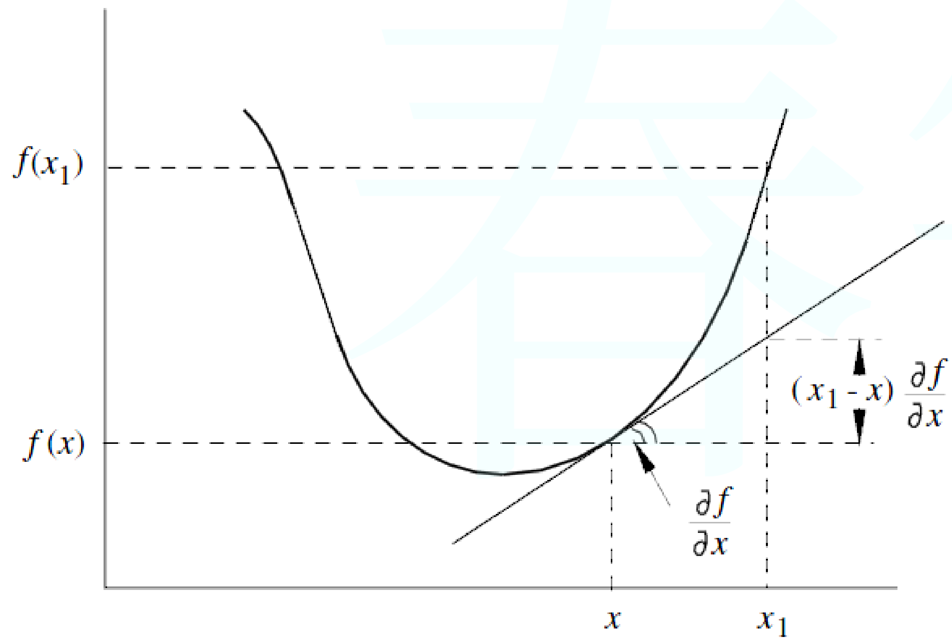
- if $\nabla^2 f(x) \succ 0$ for all $x \in \mathbf{dom}\, f$, then f is strictly convex

Property of convex functions relating to gradient

$$f(\mathbf{x}_1) \geq f(\mathbf{x}) + \mathbf{g}(\mathbf{x})^T(\mathbf{x}_1 - \mathbf{x})$$

$$f(\mathbf{x}_1) = f(\mathbf{x}) + \mathbf{g}(\mathbf{x})^T(\mathbf{x}_1 - \mathbf{x}) + \tfrac{1}{2}\mathbf{d}^T\mathbf{H}(\mathbf{x} + \alpha\mathbf{d})\mathbf{d}$$

$$\tfrac{1}{2}\mathbf{d}^T\mathbf{H}(\mathbf{x} + \alpha\mathbf{d})\mathbf{d} \geq 0$$



Examples

quadratic function: $f(x) = (1/2)x^T Px + q^T x + r$ (with $P \in \mathbf{S}^n$)

$$\nabla f(x) = Px + q, \qquad \nabla^2 f(x) = P$$

convex if $P \succeq 0$

least-squares objective: $f(x) = \|Ax - b\|_2^2$

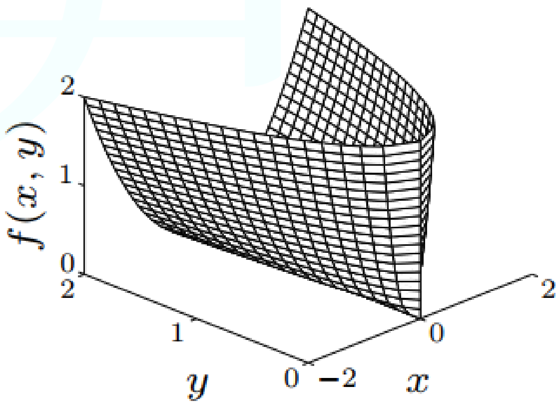
$$\nabla f(x) = 2A^T(Ax - b), \qquad \nabla^2 f(x) = 2A^T A$$

convex (for any A)

quadratic-over-linear: $f(x, y) = x^2/y$

$$\nabla^2 f(x,y) = \frac{2}{y^3} \begin{bmatrix} y \\ -x \end{bmatrix} \begin{bmatrix} y \\ -x \end{bmatrix}^T \succeq 0$$

convex for $y > 0$



Convex optimization problem

standard form convex optimization problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & a_i^T x = b_i, \quad i = 1, \ldots, p \end{array}$$

- $f_0, f_1, \ldots, f_m$ are convex; equality constraints are affine
- problem is *quasiconvex* if f_0 is quasiconvex (and $f_1, \ldots, f_m$ convex)

often written as

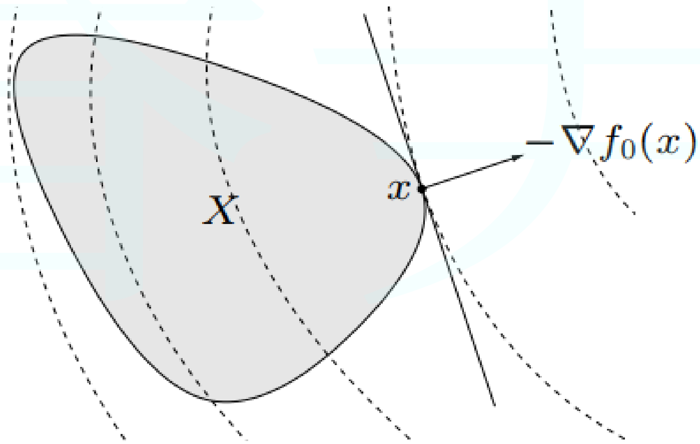
$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{array}$$

important property: feasible set of a convex optimization problem is convex

Optimality criterion for differentiable f_0

x is optimal if and only if it is feasible and

$$\nabla f_0(x)^T(y - x) \geq 0 \quad \text{for all feasible } y$$



Convex optimization

- Linear Programming
- Quadratic Programming
- QCQP
- SOCP
- SDP