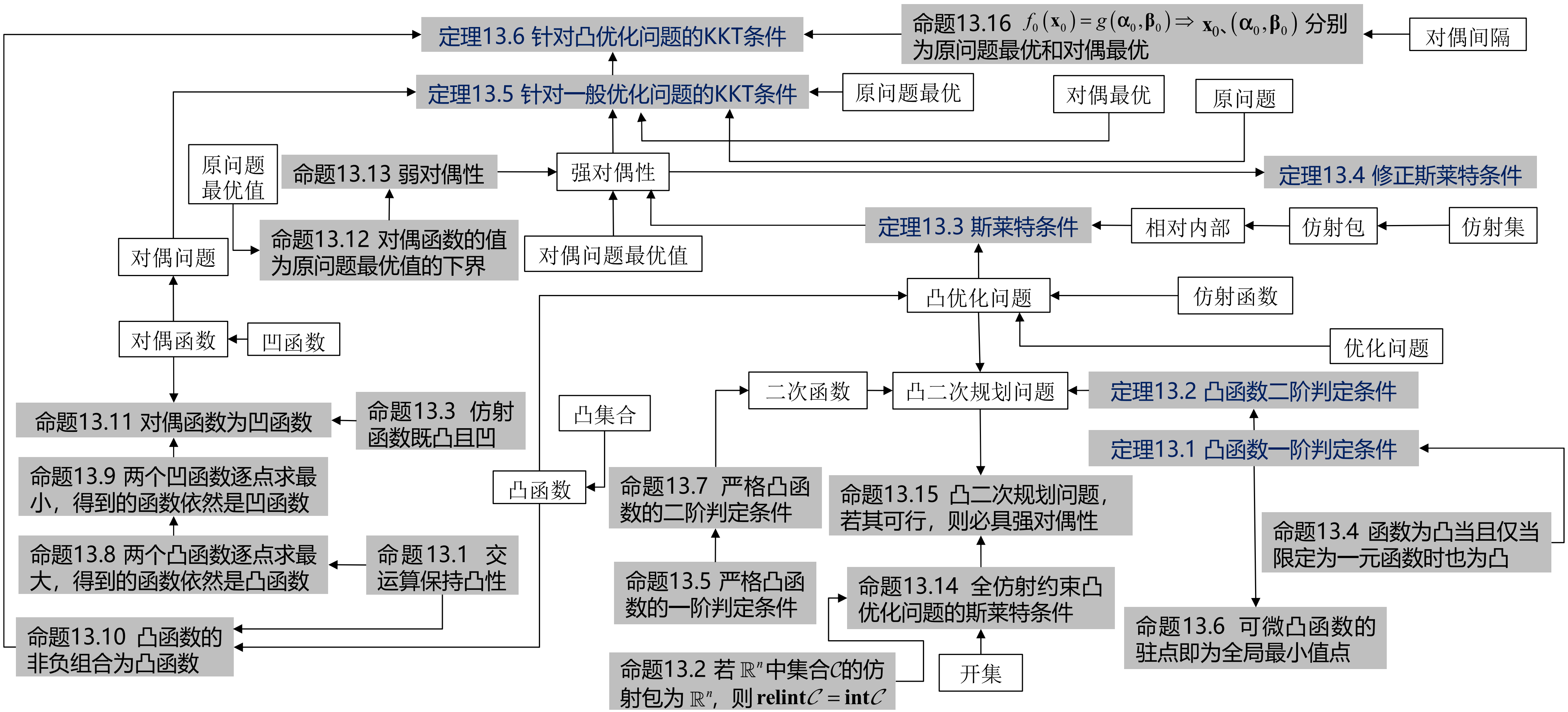
Lecture 07
Fundamentals for Convex Optimization

--From convex sets to KKT optimality conditions

Lin ZHANG, PhD
School of Software Engineering
Tongji University
Fall 2024


定理13.6 针对凸优化问题的KKT条件
命题13.16 $f_0(\mathbf{x}_0)=g(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)\Rightarrow \mathbf{x}_0$、$(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)$ 分别为原问题最优和对偶最优
对偶间隔
定理13.5 针对一般优化问题的KKT条件
原问题最优
对偶最优
原问题
原问题最优值
命题13.13 弱对偶性
强对偶性
定理13.4 修正斯莱特条件
命题13.12 对偶函数的值为原问题最优值的下界
对偶问题最优值
定理13.3 斯莱特条件
相对内部
仿射包
仿射集
对偶问题
凸优化问题
仿射函数
对偶函数
凹函数
优化问题
二次函数
凸二次规划问题
定理13.2 凸函数二阶判定条件
凸集合
命题13.11 对偶函数为凹函数
命题13.3 仿射函数既凸且凹
定理13.1 凸函数一阶判定条件
命题13.9 两个凹函数逐点求最小，得到的函数依然是凹函数
凸函数
命题13.7 严格凸函数的二阶判定条件
命题13.15 凸二次规划问题，若其可行，则必具强对偶性
命题13.4 函数为凸当且仅当限定为一元函数时也为凸
命题13.8 两个凸函数逐点求最大，得到的函数依然是凸函数
命题13.1 交运算保持凸性
命题13.5 严格凸函数的一阶判定条件
命题13.14 全仿射约束凸优化问题的斯莱特条件
命题13.10 凸函数的非负组合为凸函数
命题13.6 可微凸函数的驻点即为全局最小值点
开集
命题13.2 若 $\mathbb{R}^n$ 中集合 C 的仿射包为 $\mathbb{R}^n$，则 $\mathbf{relint}\,C=\mathbf{int}\,C$

Outline

- Convex functions
  - Convex sets and affine sets
  - Convex functions
- Optimization problems
- Convex optimization problems
- Duality

Convex sets and affine sets

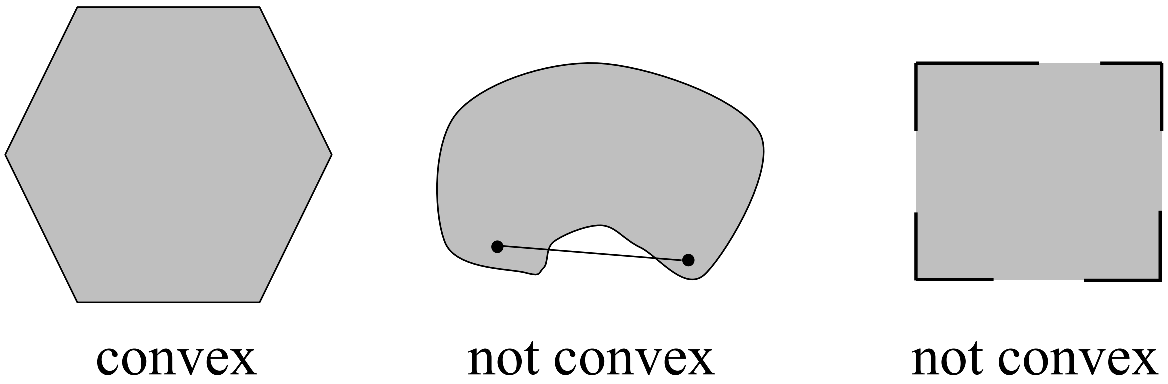
Definition 1: Convex set

A set $\mathcal{C}$ is a **convex set**, if and only if $\forall \mathbf{x}_1 \in \mathcal{C}, \mathbf{x}_2 \in \mathcal{C}, \forall \theta \in [0,1]$, we have,

$$\theta \mathbf{x}_1 + (1-\theta)\mathbf{x}_2 \in \mathcal{C}$$

$\theta \mathbf{x}_1 + (1-\theta)\mathbf{x}_2 \in \mathcal{C}$ is also called the **convex combination** of $\mathbf{x}_1$ and $\mathbf{x}_2$

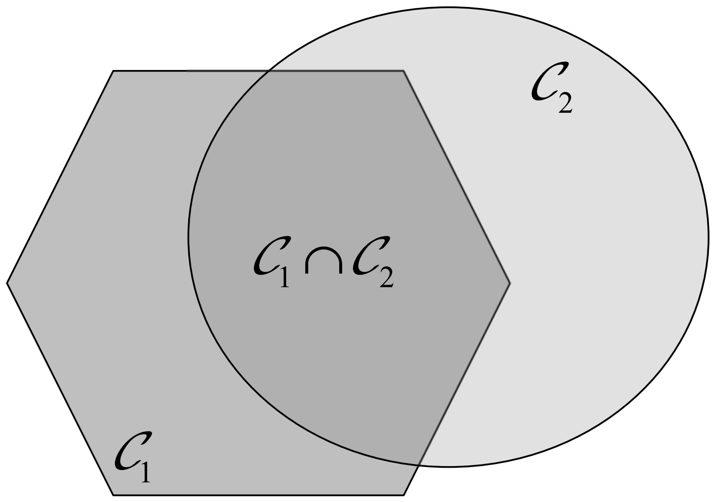
If a set is convex, the line segment linking any two points in that set also belongs to that set



Convex sets and affine sets

Proposition 1:

If $\mathcal{C}_1$ and $\mathcal{C}_2$ are two convex sets, their intersection $\mathcal{C}_1 \cap \mathcal{C}_2$ is also a convex set



An example

Convex sets and affine sets

Definition 2: Affine set

A set $\mathcal{C}$ is an **affine set**, if and only if $\forall \mathbf{x}_1 \in \mathcal{C}, \mathbf{x}_2 \in \mathcal{C}, \forall \theta \in \mathbb{R}$, we have,

$$\theta \mathbf{x}_1 + (1-\theta)\mathbf{x}_2 \in \mathcal{C}$$

$\theta \mathbf{x}_1 + (1-\theta)\mathbf{x}_2 \in \mathcal{C}$ is also called the **affine combination** of $\mathbf{x}_1$ and $\mathbf{x}_2$

If a set is affine, the line passing any two points in that set also belongs to that set

Convex sets and affine sets

Definition 3: Affine hull

The set of all the affine combinations of points in set $\mathcal{C}$ is called the **affine hull** of $\mathcal{C}$, denoted by **aff**$\mathcal{C}$,

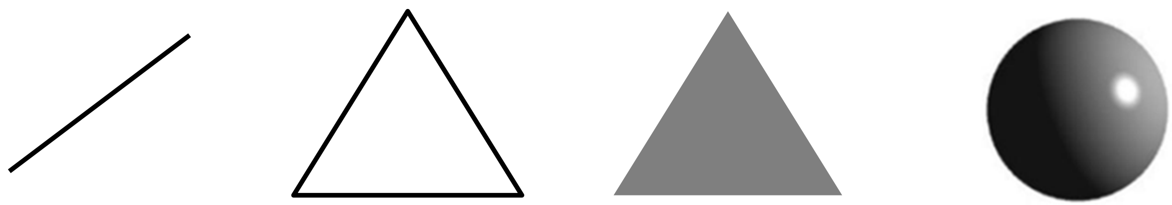
$$\mathbf{aff}\mathcal{C} = \{\theta \mathbf{x}_1 + (1-\theta)\mathbf{x}_2 \mid \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{C}\}$$

where $\theta \in \mathbb{R}$ is any number

No matter whether $\mathcal{C}$ is affine or not, its affine hull **aff**$\mathcal{C}$ is an affine set

aff$\mathcal{C}$ is the smallest affine set that contains $\mathcal{C}$; if $\mathcal{C}$ is an affine set, **aff**$\mathcal{C}$=$\mathcal{C}$

Examples: what are the affine hulls for the following sets?



Convex sets and affine sets

For a set $\mathcal{C} \subseteq \mathbb{R}^n$, we can define its **interior** and **boundary** (relative to $\mathbb{R}^n$)

The interior of $\mathcal{C}$ is defined as, $\mathbf{int}\mathcal{C} = \{\mathbf{x} \in \mathcal{C} \mid \exists \varepsilon > 0, B(\mathbf{x}, \varepsilon) \subseteq \mathcal{C}\}$

where $B(\mathbf{x}, \varepsilon) = \{\mathbf{y} \mid \|\mathbf{y} - \mathbf{x}\| \leq \varepsilon\}$, and $\|\cdot\|$ can be any norm

The closure of $\mathcal{C}$ is defined as, $\mathbf{cl}\mathcal{C} = \{\mathbf{x} \in \mathbb{R}^n \mid \forall \varepsilon > 0, \exists \mathbf{y} \in \mathcal{C}, \|\mathbf{y} - \mathbf{x}\| \leq \varepsilon\}$

The boundary of $\mathcal{C}$ is defined as, $\mathbf{bc}\mathcal{C} = \mathbf{cl}\mathcal{C} \setminus \mathbf{int}\mathcal{C}$

The set $\mathcal{C}$ is an open set if and only if $\mathbf{int}\mathcal{C} = \mathcal{C}$

In $\mathbb{R}^3$, what are the interiors and boundaries for a solid sphere and a thin paper?

Convex sets and affine sets

Definition 4: For a set $\mathcal{C} \subseteq \mathbb{R}^n$, its **relative** (to its affine hull) **interior** is defined as,

$$\mathbf{relint}\mathcal{C} = \left\{ \mathbf{x} \in \mathcal{C} \mid \exists \varepsilon > 0, \textit{such that} \left(B(\mathbf{x}, \varepsilon) \cap \mathbf{aff}\mathcal{C} \right) \subseteq \mathcal{C} \right\}$$

where $B(\mathbf{x}, \varepsilon) = \left\{ \mathbf{y} \mid \|\mathbf{y} - \mathbf{x}\| \leq \varepsilon \right\}$, and $\|\cdot\|$ can be any norm

For a set $\mathcal{C} \subseteq \mathbb{R}^n$, its **relative** (to its affine hull) **boundary** is defined as,

$$\mathbf{cl}\mathcal{C} \setminus \mathbf{relint}\mathcal{C}$$

For a set $\mathcal{C} \subseteq \mathbb{R}^n$, if $\mathbf{aff}\mathcal{C} = \mathbb{R}^n$, then $\mathbf{relint}\mathcal{C} = \mathbf{int}\mathcal{C}$ and $\mathbf{cl}\mathcal{C} \setminus \mathbf{relint}\mathcal{C} = \mathbf{bd}\mathcal{C}$

Convex sets and affine sets

Example: Consider a square in the $(x_1, x_2)-$ plane in $\mathbb{R}^3$, defined as,

$$\mathcal{C} = \left\{ \mathbf{x} \in \mathbb{R}^3 \mid -1 \le x_1 \le 1, -1 \le x_2 \le 1, x_3 = 0 \right\}$$

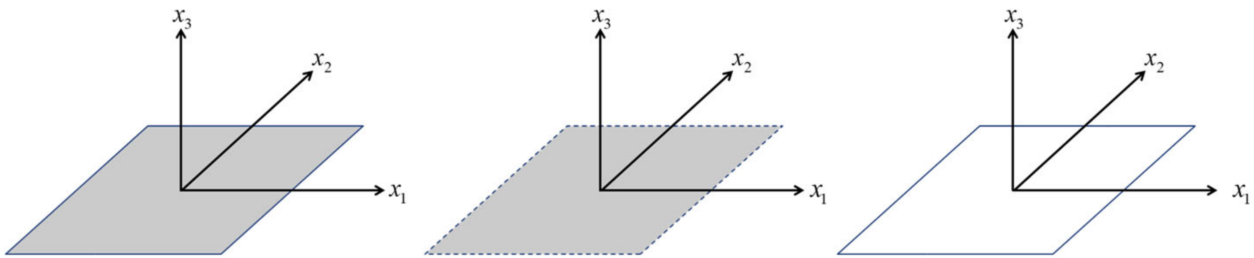
Its affine hull is the (x_1, x_2)-plane, i.e., $\mathbf{aff}\mathcal{C} = \left\{ \mathbf{x} \in \mathbb{R}^3 \mid x_3 = 0 \right\}$

The interior of $\mathcal{C}$ is empty, but its relative interior is,

$$\mathbf{relint}\mathcal{C} = \left\{ \mathbf{x} \in \mathbb{R}^3 \mid -1 < x_1 < 1, -1 < x_2 < 1, x_3 = 0 \right\}$$

The boundary of $\mathcal{C}$ is itself, but its relative boundary is the wire-frame outline,

$$\mathbf{cl}\mathcal{C} \setminus \mathbf{relint}\mathcal{C} = \left\{ \mathbf{x} \in \mathbb{R}^3 \mid \max\left\{ |x_1|, |x_2| \right\} = 1, x_3 = 0 \right\}$$



$\mathcal{C}$ $\mathbf{relint}\mathcal{C}$ $\mathbf{cl}\mathcal{C} \setminus \mathbf{relint}\mathcal{C}$

Outline

- Convex functions
  - Convex sets and affine sets
  - Convex functions
- Optimization problems
- Convex optimization problems
- Duality

Convex functions

Definition 5: Affine function

If the function $f(\mathbf{x}):\mathbb{R}^n \to \mathbb{R}^m$ is the sum of a linear function and a constant, it is called an **affine function**, i.e., it has the form,

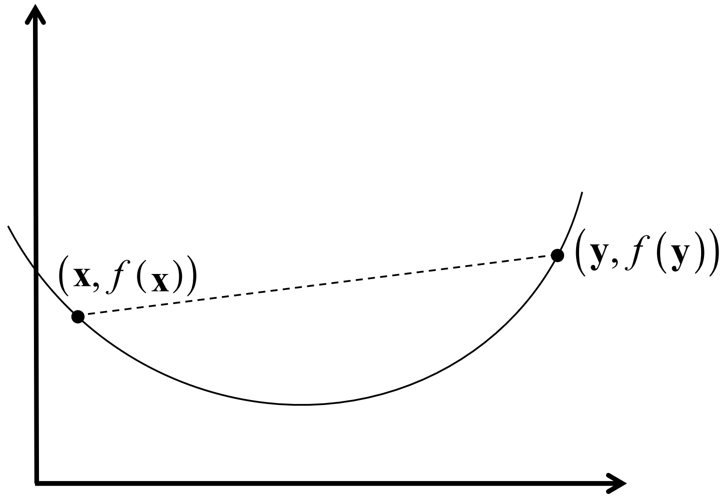
$$f(\mathbf{x}) = A\mathbf{x} + \mathbf{b}, A \in \mathbb{R}^{m\times n}, \mathbf{x} \in \mathbb{R}^{n\times 1}, \mathbf{b} \in \mathbb{R}^{m\times 1}$$

Convex functions

Definition 6: Convex function

A function $f(\mathbf{x}): \mathbb{R}^n \to \mathbb{R}$ is **convex** if its domain $\mathbf{dom}f$ is a convex set, and if for all $\mathbf{x}, \mathbf{y} \in \mathbf{dom}f$ and any $\theta \in [0,1]$, we have,

$$\theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y}) \geq f(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \qquad \textbf{(Eq. 1)}$$



A typical convex function in 2D

Definition 7: Strictly Convex function

In Def. 6, if Eq. 1 changes to,

$$\theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y}) > f(\theta\mathbf{x} + (1-\theta)\mathbf{y})$$

and all the other conditions remain, then the function $f(\mathbf{x})$ is called **strictly convex**

Convex functions

Definition 8: Concave function

If $-f(\mathbf{x})$ is a convex function, we say $f(\mathbf{x})$ is a **concave function**

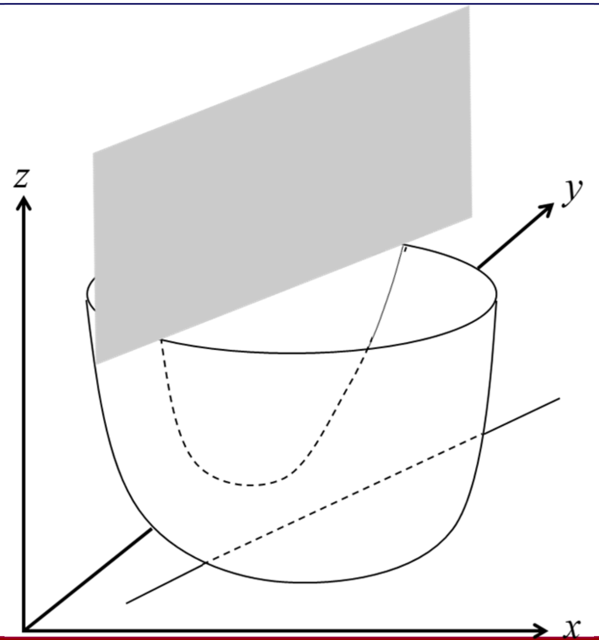
Definition 9: Strictly concave function

If $-f(\mathbf{x})$ is a strictly convex function, we say $f(\mathbf{x})$ is a **strictly concave function**

Convex functions

Proposition 2:

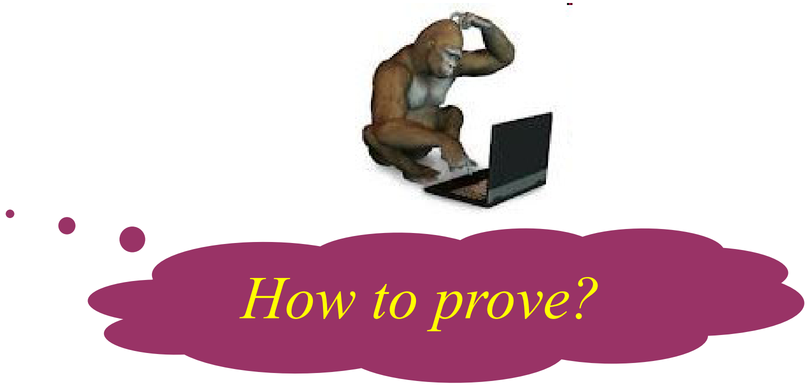
A function is convex if and only if it is convex when restricted to any line that intersects its domain. In other words, f is convex if and only if for all $\mathbf{x} \in \mathbf{dom} f$ and all $\mathbf{v}$, the function $g(t)=f(\mathbf{x}+t\mathbf{v})$ is convex on its domain $\{t \mid \mathbf{x}+t\mathbf{v} \in \mathbf{dom} f\}$



Convex functions

Proposition 3:

The affine function is a convex function and also is a concave function.



Convex functions

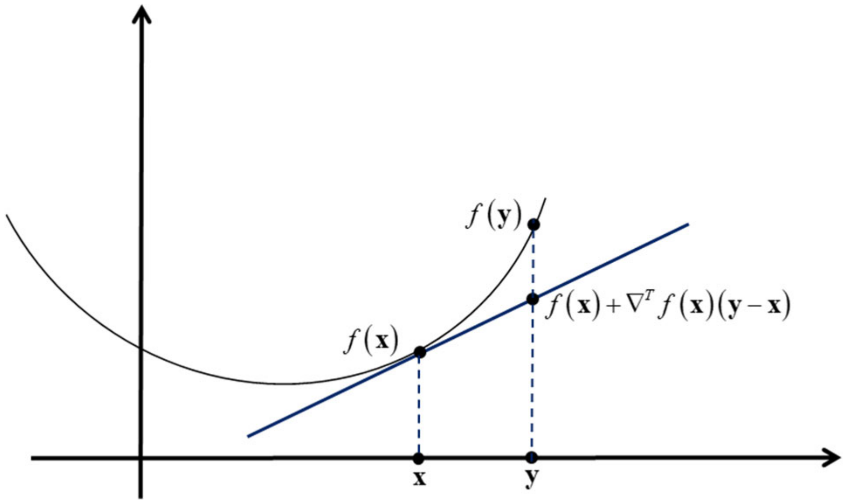
Proposition 4: First-order conditions to determine a convex function

Suppose $f(\mathbf{x})$ is differentiable. Then, $f(\mathbf{x})$ is convex if and only if **dom**f is convex and

$$f(\mathbf{y}) \geq f(\mathbf{x}) + (\nabla f(\mathbf{x}))^T (\mathbf{y} - \mathbf{x})$$

holds for all $\mathbf{x}, \mathbf{y} \in \mathbf{dom} f$

For a formal proof, refer to the textbook



Convex functions

Proposition 5: First-order conditions to determine a strictly convex function

Suppose $f(\mathbf{x})$ is differentiable. Then, $f(\mathbf{x})$ is strictly convex if and only if **dom**f is convex and

$$f(\mathbf{y}) > f(\mathbf{x}) + (\nabla f(\mathbf{x}))^T (\mathbf{y} - \mathbf{x})$$

holds for all $\mathbf{x}, \mathbf{y} \in \mathbf{dom} f$

Convex functions

Proposition 6:

Suppose that $f(\mathbf{x})$ is a differentiable convex function. If $\mathbf{x}_0$ is a stationary point of $f(\mathbf{x})$, i.e., $\nabla_{\mathbf{x}=\mathbf{x}_0} f(\mathbf{x}) = \mathbf{0}$, $\mathbf{x}_0$ is a global minimizer of $f(\mathbf{x})$

Proof:

$f(\mathbf{x})$ is convex and differentiable, according to Prop. 4,

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

At $\mathbf{x} = \mathbf{x}_0$, $f(\mathbf{y}) \geq f(\mathbf{x}_0) + \left(\nabla_{\mathbf{x}=\mathbf{x}_0} f(\mathbf{x})\right)^T (\mathbf{y} - \mathbf{x}_0)$

Since $\nabla_{\mathbf{x}=\mathbf{x}_0} f(\mathbf{x}) = \mathbf{0}$

$f(\mathbf{y}) \geq f(\mathbf{x}_0) \Rightarrow \mathbf{x}_0$ is a global minimizer of $f(\mathbf{x})$

Convex functions

Proposition 7: Second-order conditions to determine a convex function

Suppose $f(\mathbf{x})$ is twice differentiable. Then, $f(\mathbf{x})$ is convex if and only if $\mathbf{dom}f$ is convex and the Hessian matrix $\nabla^2 f(\mathbf{x})$ is positive semidefinite for all $\mathbf{x} \in \mathbf{dom}f$

必要性。我们需要证明：若 $f(\mathbf{x})$是凸函数，则它的定义域是凸集并且它的海森矩阵 $\nabla^2 f(\mathbf{x})$ 为半正定矩阵。由于 $f(\mathbf{x})$是凸函数，根据凸函数定义，$\mathbf{dom}f$ 必然为凸集。由于 $f(\mathbf{x})$二阶可微，$\forall \mathbf{x} \in \mathbf{dom}f$，我们可在 $\mathbf{x}$ 处进行二阶泰勒展开，有，

$$f(\mathbf{x}+\mathbf{h}) = f(\mathbf{x}) + \mathbf{h}^T \nabla f(\mathbf{x}) + \frac{1}{2}\mathbf{h}^T \nabla^2 f(\mathbf{x})\mathbf{h} + O\left(\|\mathbf{h}\|^2\right)$$

其中，$\mathbf{h} \neq \mathbf{0}$。由于已知 $f(\mathbf{x})$为凸函数，根据 prop. 4 可知，$f(\mathbf{x}+\mathbf{h}) \geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})\mathbf{h}$。结合上式有 $\frac{1}{2}\mathbf{h}^T \nabla^2 f(\mathbf{x})\mathbf{h} \geq 0$，因此 $\nabla^2 f(\mathbf{x})$ 为半正定矩阵。

Convex functions

Proposition 7: Second-order conditions to determine a convex function

Suppose $f(\mathbf{x})$ is twice differentiable. Then, $f(\mathbf{x})$ is convex if and only if **dom**f is convex and the Hessian matrix $\nabla^2 f(\mathbf{x})$ is positive semidefinite for all $\mathbf{x} \in \textbf{dom}f$

充分性。我们需要证明：若 $f(\mathbf{x})$的定义域是凸集并且它的海森矩阵$\nabla^2 f(\mathbf{x})$为半正定矩阵，则$f(\mathbf{x})$必为凸函数。$\forall \mathbf{x}, \mathbf{y} \in \mathrm{dom}f$，根据泰勒展开有，

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

其中，$t \in [0,1]$是存在的某个数能使上式成立。由于 dom$f$ 为凸集，根据凸集的性质可知 $\mathbf{x}+t(\mathbf{y}-\mathbf{x}) \in \mathrm{dom}f$。根据已知条件，$f(\mathbf{x})$在其定义域内任意一点处的海森矩阵都为半正定矩阵，则$\nabla^2 f(\mathbf{x}+t(\mathbf{y}-\mathbf{x}))$为半正定矩阵，则$\frac{1}{2}(\mathbf{y}-\mathbf{x})^T \nabla^2 f(\mathbf{x}+t(\mathbf{y}-\mathbf{x}))(\mathbf{y}-\mathbf{x}) \geq 0$。再结合上述泰勒展开结果可知，$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{y}-\mathbf{x})$。根据 prop. 4 可知，$f(\mathbf{x})$为凸函数。

Convex functions

Proposition 8: Second-order conditions to determine a strictly convex function

Suppose $f(\mathbf{x})$ is twice differentiable. If $\mathbf{dom}f$ is convex and $\nabla^2 f(\mathbf{x})$ is positive definite for all $\mathbf{x}\in\mathbf{dom}f$, then $f(\mathbf{x})$ is strictly convex

证明：

$\forall \mathbf{x},\mathbf{y}\in \mathrm{dom}f$ ，根据泰勒展开有，

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

其中， $t\in[0,1]$ 是存在的某个数能使上式成立。由于 dom$f$ 为凸集，根据凸集的性质可知 $\mathbf{x}+t(\mathbf{y}-\mathbf{x})\in \mathrm{dom}f$ 。根据已知条件，$f(\mathbf{x})$在其定义域内任意一点处的海森矩阵都为正定矩阵，则 $\nabla^2 f(\mathbf{x}+t(\mathbf{y}-\mathbf{x}))$ 为正定矩阵，则 $\frac{1}{2}(\mathbf{y}-\mathbf{x})^T\nabla^2 f(\mathbf{x}+t(\mathbf{y}-\mathbf{x}))(\mathbf{y}-\mathbf{x})>0$ 。结合泰勒展开结果则有， $f(\mathbf{y})>f(\mathbf{x})+\nabla^T f(\mathbf{x})(\mathbf{y}-\mathbf{x})$ 。根据 prop. 5 可知，$f(\mathbf{x})$为严格凸函数。

Convex functions

Proposition 8: Second-order conditions to determine a strictly convex function

Suppose $f(\mathbf{x})$ is twice differentiable. If **dom***f* is convex and $\nabla^2 f(\mathbf{x})$ is positive definite for all $\mathbf{x} \in \mathbf{dom} f$, then $f(\mathbf{x})$ is strictly convex

Please note that the above condition is only a sufficient condition for a function being strictly convex but not a necessary condition. In other words, for a strictly convex differentiable function, its Hessian matrix may not be positive definite.

Convex functions

Proposition 9:

Suppose that $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ are convex. Then, their pointwise maximum $f(\mathbf{x})$, defined by,

$$f(\mathbf{x}) = \max\{f_1(\mathbf{x}), f_2(\mathbf{x})\}$$

with the domain $\mathbf{dom}f = \mathbf{dom}f_1 \cap \mathbf{dom}f_2$, is also convex.

设$f_1(\mathbf{x})$、$f_2(\mathbf{x})$的定义域分别为domf_1、domf_2，因为这两个函数均为凸函数，根据凸函数的定义（ Def. 6 ）可知，domf_1、domf_2均为凸集。$f(\mathbf{x})$的定义域dom$f$显然为$f_1(\mathbf{x})$、$f_2(\mathbf{x})$定义域的交集，即$\mathrm{dom}f = \mathrm{dom}f_1 \cap \mathrm{dom}f_2$。根据 Prop. 1, 凸集的交集依然是凸集，因此dom$f$为凸集。另外，$\forall \mathbf{x}, \mathbf{y} \in \mathrm{dom}f$，$\forall \theta \in [0,1]$有，

$$\begin{aligned}
f(\theta\mathbf{x}+(1-\theta)\mathbf{y}) &= \max\{f_1(\theta\mathbf{x}+(1-\theta)\mathbf{y}), f_2(\theta\mathbf{x}+(1-\theta)\mathbf{y})\} \\
&\le \max\{\theta f_1(\mathbf{x})+(1-\theta)f_1(\mathbf{y}), \theta f_2(\mathbf{x})+(1-\theta)f_2(\mathbf{y})\} \\
&\le \theta\max\{f_1(\mathbf{x}), f_2(\mathbf{x})\}+(1-\theta)\max\{f_1(\mathbf{y}), f_2(\mathbf{y})\} \\
&= \theta f(\mathbf{x})+(1-\theta)f(\mathbf{y})
\end{aligned}$$

综合以上信息可知，$f(\mathbf{x})$为凸函数。

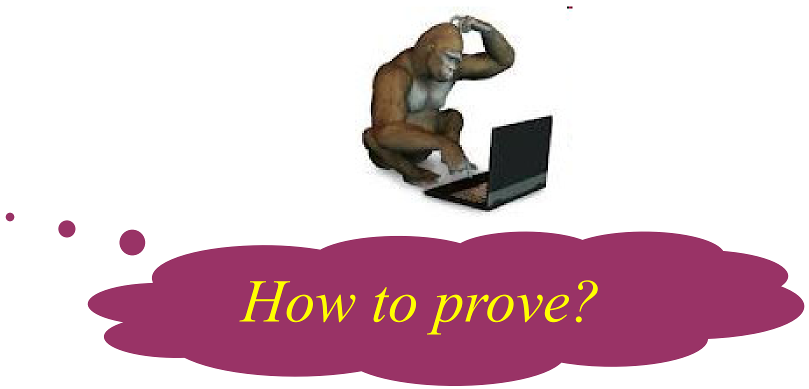
Convex functions

Proposition 10:

Suppose that $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ are concave. Then, their pointwise minimum $f(\mathbf{x})$, defined by,

$$f(\mathbf{x}) = \min\{f_1(\mathbf{x}), f_2(\mathbf{x})\}$$

with the domain $\mathbf{dom}f = \mathbf{dom}f_1 \cap \mathbf{dom}f_2$, is also concave.



Convex functions

Proposition 11:

Suppose that $f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x})$ are convex. Then, their nonnegative combination,

$$f(\mathbf{x}) = \omega_1 f_1(\mathbf{x}) + \omega_2 f_2(\mathbf{x}) + \cdots + \omega_m f_m(\mathbf{x})$$

where $\omega_i \geq 0 (i = 1, 2, \cdots, m)$, is also convex.

证明：

$\mathrm{dom} f = \bigcap_{i=1}^{m} \mathrm{dom} f_i$，由于 $\mathrm{dom} f_i$ 是凸集，则 $\mathrm{dom} f$ 是一系列凸集的交集，根据 Prop. 1， $\mathrm{dom} f$ 为凸集。另外，$\forall \mathbf{x}, \mathbf{y} \in \mathrm{dom} f$，$\forall \theta \in [0,1]$ 有，

$$\begin{aligned} f(\theta\mathbf{x} + (1-\theta)\mathbf{y}) &= \sum_{i=1}^{m} \omega_i f_i(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \\ &\leq \sum_{i=1}^{m} \omega_i \left(\theta f_i(\mathbf{x}) + (1-\theta) f_i(\mathbf{y})\right) = \theta \sum_{i=1}^{m} \omega_i f_i(\mathbf{x}) + (1-\theta) \sum_{i=1}^{m} \omega_i f_i(\mathbf{y}) \\ &= \theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y}) \end{aligned}$$

综上，$f(\mathbf{x})$为凸函数。

Convex functions

Proposition 12:

Suppose that $f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x})$ are concave. Then, their nonnegative combination,

$$f(\mathbf{x}) = \omega_1 f_1(\mathbf{x}) + \omega_2 f_2(\mathbf{x}) + \cdots + \omega_m f_m(\mathbf{x})$$

where $\omega_i \geq 0 \left(i = 1, 2, \cdots, m\right)$, is also concave.

How to prove?

Convex functions

Definition 10: Quadratic function

A function $f(\mathbf{x}): \mathbb{R}^n \to \mathbb{R}$ of the following form,

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T P\mathbf{x} + \mathbf{q}^T\mathbf{x} + r \qquad \textbf{(Eq. 2)}$$

where $P \in \mathbb{R}^{n\times n}$ is a symmetric matrix and $\mathbf{q} \in \mathbb{R}^{n\times 1}$, is called a **quadratic function**

It can be verified that the Hessian matrix of $f(\mathbf{x})$ in Eq. 2 is $\nabla^2 f(\mathbf{x}) = P$

↓

According to Prop. 7, if P is positive semidefinite, $f(\mathbf{x})$ is convex
According to Prop. 8, if P is positive definite, $f(\mathbf{x})$ is strictly convex

Outline

- Convex functions
  - Convex sets and affine sets
  - Convex functions
- Optimization problems
- Convex optimization problems
- Duality

Optimization problems

Definition 11: Optimization problem

A general optimization problem is expressed in the following form,

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} f_0(\mathbf{x})$$

$$\text{subject to } f_i(\mathbf{x}) \le 0, i = 1, \ldots, m$$

$$h_i(\mathbf{x}) = 0, i = 1, \ldots, p$$

$\mathbf{x} \in \mathbb{R}^{n\times 1}$ is the **optimization variable**, $\mathbf{x}^*$ is the **optimal solution;** $f_0(\mathbf{x}): \mathbb{R}^n \to \mathbb{R}$ is the **objective function**

$f_i(\mathbf{x}): \mathbb{R}^n \to \mathbb{R}, i = 1, \ldots, m$ are the **inequality constraint functions**

$h_i(\mathbf{x}): \mathbb{R}^n \to \mathbb{R}, i = 1, \ldots, p$ are the **equality constraint functions**

If m=p=0, we say the problem is **unconstrained**

The set of points for which the objective and all constraint functions are defined,

$$\mathcal{D} = \bigcap_{i=0}^{m} \mathbf{dom} f_i \cap \bigcap_{i=1}^{p} \mathbf{dom} h_i$$

is called the **domain** of the optimization problem Def. 11.

Optimization problems

For a point $\mathbf{x} \in \mathcal{D}$, if it satisfies all the constraints $f_i(\mathbf{x}) \le 0, i = 1, \ldots, m$ and $h_i(\mathbf{x}) = 0, i = 1, \ldots, p$, we say $\mathbf{x}$ is a **feasible point**.

If there exists at least one feasible point, we say the problem Def. 11 is **feasible**, otherwise it is **infeasible**

The set of all feasible points is called the **feasible set**

If the optimal solution $\mathbf{x}^*$ exists, it should be in the feasible set

The **optimal value** of the problem Def. 11 is defined as,

$$v^* = \min\{f_0(\mathbf{x}) \mid f_i(\mathbf{x}) \le 0, i = 1, \ldots, m, h_i(\mathbf{x}) = 0, i = 1, \ldots, p\}$$

It can be easily known that,

$$v^* = \begin{cases} f_0(\mathbf{x}^*), & \text{if the optimal solution } \mathbf{x}^* \text{ exists} \\ +\infty, & \text{if the problem Def. 11 is infeasible} \\ -\infty, & \text{if the problem Def. 11 is unbounded below} \end{cases}$$

Optimization problems

It needs to be noted that even if the feasible set is not empty, the optimal solution of the problem Def. 11 may not exist

Example:

$$x^* = \arg\min_{\mathbf{x}} 2x$$
$$\text{subject to } x \leq 0$$

The feasible set of this problem is not empty, but the objective function is unbounded below and thus the optimal solution does not exist.

Outline

- Convex functions
  - Convex sets and affine sets
  - Convex functions
- Optimization problems
- Convex optimization problems
- Duality

Convex optimization problems

Definition 12: Convex optimization problem

We call the following optimization problem the convex optimization problem,

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} f_0(\mathbf{x})$$
$$\text{subject to } f_i(\mathbf{x}) \le 0, i = 1, \ldots, m$$
$$\mathbf{a}_i^T \mathbf{x} - b_i = 0, i = 1, \ldots, p$$

where $f_i(\mathbf{x})$ (i=0,1,…,m) is convex

Please make a comparison between the definitions of the general optimization problem and the convex optimization prolem

Proposition 13:

The feasible set of a convex optimization problem is convex.

Convex optimization problems

As an example, the convex quadratic program is a typical convex optimization problem

Definition 13: Convex quadratic program problem

The convex quadratic program problem is expressed by,

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} \frac{1}{2}\mathbf{x}^T P\mathbf{x} + \mathbf{q}^T\mathbf{x} + r$$

$$\text{subject to } G\mathbf{x} \le \mathbf{h}, G \in \mathbb{R}^{m\times n}$$

$$A\mathbf{x} = \mathbf{b}, A \in \mathbb{R}^{p\times n}$$

where $P \in \mathbb{R}^{n\times n}$ is positive semidefinite

Can you verify it is really a convex optimization problem?

Outline

- Convex functions
  - Convex sets and affine sets
  - Convex functions
- Optimization problems
- Convex optimization problems
- Duality

Duality

Definition 14: Lagrangian

We define the Lagrangian $l(\mathbf{x},\boldsymbol{\alpha},\boldsymbol{\beta}):\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}^p\rightarrow\mathbb{R}$ associated with the problem Def. 11 as,

$$l(\mathbf{x},\boldsymbol{\alpha},\boldsymbol{\beta})=f_0(\mathbf{x})+\sum_{i=1}^{m}\alpha_i f_i(\mathbf{x})+\sum_{i=1}^{p}\beta_i h_i(\mathbf{x})$$

where $\boldsymbol{\alpha}=\{\alpha_i\}_{i=1}^m$, $\boldsymbol{\beta}=\{\beta_i\}_{i=1}^p$ are **dual variables**, and $\mathbf{dom}\,l=\mathcal{D}\times\mathbb{R}^m\times\mathbb{R}^p$

Definition 15: Lagrange dual function

We define the Lagrange dual function $g(\boldsymbol{\alpha},\boldsymbol{\beta}):\mathbb{R}^m\times\mathbb{R}^p\rightarrow\mathbb{R}$ of the problem Def. 11 as the minimum value of its Lagrangian over $\mathbf{x}$,

$$g(\boldsymbol{\alpha},\boldsymbol{\beta})=\min_{\mathbf{x}\in\mathcal{D}} l(\mathbf{x},\boldsymbol{\alpha},\boldsymbol{\beta})=\min_{\mathbf{x}\in\mathcal{D}}\left(f_0(\mathbf{x})+\sum_{i=1}^{m}\alpha_i f_i(\mathbf{x})+\sum_{i=1}^{p}\beta_i h_i(\mathbf{x})\right)$$

Duality

Proposition 14:

The Lagrange dual function is concave.

证明：

Def. 15 所定义的对偶函数 $g(\boldsymbol{\alpha},\boldsymbol{\beta})$ 可变形为如下形式，

$$g(\boldsymbol{\alpha},\boldsymbol{\beta})=\min_{\mathbf{x}\in\mathcal{D}}\left(\left(f_1(\mathbf{x})\,f_2(\mathbf{x})\cdots f_m(\mathbf{x})\,h_1(\mathbf{x})\,h_2(\mathbf{x})\cdots h_p(\mathbf{x})\right)\begin{pmatrix}\alpha_1\\ \alpha_2\\ \vdots\\ \alpha_m\\ \beta_1\\ \beta_2\\ \vdots\\ \beta_p\end{pmatrix}+f_0(\mathbf{x})\right)$$

显然，$g(\boldsymbol{\alpha},\boldsymbol{\beta})$ 为一系列关于 $\begin{pmatrix}\boldsymbol{\alpha}\\ \boldsymbol{\beta}\end{pmatrix}$ 的仿射函数逐点求最小得到的；而根据 Prop. 3， 仿射函数为凹函数。这也就是说，函数 $g(\boldsymbol{\alpha},\boldsymbol{\beta})$ 为一系列关于 $\begin{pmatrix}\boldsymbol{\alpha}\\ \boldsymbol{\beta}\end{pmatrix}$ 的凹函数逐点求最小而得到的，根据 Prop. 10 可知， $g(\boldsymbol{\alpha},\boldsymbol{\beta})$ 为凹函数。

Duality

Proposition 15:

The Lagrange dual function yields lower bounds on the optimal value v^* of the problem Def. 11, i.e.,

For any $\boldsymbol{\alpha} \geq \mathbf{0}$ and any $\boldsymbol{\beta}$, $g(\boldsymbol{\alpha}, \boldsymbol{\beta}) \leq v^*$

先考虑优化问题 Def. 11 可行的情况。假设 $\tilde{\mathbf{x}}$ 是优化问题 Def. 11 一个可行点，根据可行点的定义可知，

$$f_i(\tilde{\mathbf{x}}) \leq 0, i = 1, 2, \ldots, m, \quad h_i(\tilde{\mathbf{x}}) = 0, i = 1, 2, \ldots, p$$

则有，

$$\sum_{i=1}^{m} \alpha_i f_i(\tilde{\mathbf{x}}) + \sum_{i=1}^{m} \beta_i h_i(\tilde{\mathbf{x}}) \leq 0,$$

因此，

$$l(\tilde{\mathbf{x}}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = f_0(\tilde{\mathbf{x}}) + \sum_{i=1}^{m} \alpha_i f_i(\tilde{\mathbf{x}}) + \sum_{i=1}^{m} \beta_i h_i(\tilde{\mathbf{x}}) \leq f_0(\tilde{\mathbf{x}}),$$

因此，

$$g(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \min_{\mathbf{x} \in \mathcal{D}} l(\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \leq l(\tilde{\mathbf{x}}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \leq f_0(\tilde{\mathbf{x}})$$

由于上式对任意的可行点 $\tilde{\mathbf{x}}$ 都成立，而优化问题 Def. 11 最优值 v^*当然也是在某个可行点处取得的，因此有 $g(\boldsymbol{\alpha}, \boldsymbol{\beta}) \leq v^*$。

再先考虑优化问题 Def. 11 不可行的情况。若优化问题 Def. 11 不可行，则其最优值 $v^* = +\infty$，则显然有 $g(\boldsymbol{\alpha}, \boldsymbol{\beta}) \leq v^*$。

Duality

Based on Prop. 15, we know that the Lagrange dual function $g(\boldsymbol{\alpha},\boldsymbol{\beta})$ yields the lower bounds for the optimal value of the problem Def. 11. Among these lower bounds, the largest one is the most valuable.

That motivates us to solve the following Lagrange dual problem

Definition 16: Lagrange dual problem

$$\boldsymbol{\alpha}^*,\boldsymbol{\beta}^* = \arg\max_{\boldsymbol{\alpha},\boldsymbol{\beta}} g(\boldsymbol{\alpha},\boldsymbol{\beta})$$

$$\text{subject to } \boldsymbol{\alpha} \geq \mathbf{0}$$

The above problem is called the **Lagrange dual problem** of the problem Def. 11, and accordingly, the problem Def. 11 is called the **primal problem**

We say a pair of dual variables $(\boldsymbol{\alpha},\boldsymbol{\beta})$ is dual feasible if they satisfy $\boldsymbol{\alpha} \geq \mathbf{0}$ and $g(\boldsymbol{\alpha},\boldsymbol{\beta}) > -\infty$

It should be noted that no matter whether the primal problem Def. 11 is a convex optimization problem or not, its dual problem Def. 16 is a convex optimization problem!

Duality

Proposition 16: Weak duality

The general optimization problems (Def. 11) have the following **weak duality** property.
The optimal value of the dual problem is,

$$d^* = \max_{\boldsymbol{\alpha},\boldsymbol{\beta}} g(\boldsymbol{\alpha},\boldsymbol{\beta}), \text{subject to } \boldsymbol{\alpha} \geq \mathbf{0}$$

The optimal value of the primal problem (Def. 11) is,

$$v^* = \min\{f_0(\mathbf{x}) \mid f_i(\mathbf{x}) \leq 0, i = 1,\ldots,m, h_i(\mathbf{x}) = 0, i = 1,\ldots,p\}$$

Based on Prop. 15, it can be known that $d^* \leq v^*$

Definition 17: Strong duality

For an optimization problem, if its optimal value is strictly equal to the optimal value of its dual problem, i.e.,

$$d^* = v^*$$

we say this optimization problem has the property of **strong duality**

What are the conditions an optimization problem needs to satisfy to have strong duality?

Duality

Proposition 17: Slater condition

Suppose that the primal problem is convex, i.e., of the form,

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} f_0(\mathbf{x})$$

$$\text{subject to } f_i(\mathbf{x}) \le 0, i = 1, ..., m$$

$$\mathbf{a}_i^T \mathbf{x} - b_i = 0, i = 1, ..., p$$

where $f_i(\mathbf{x})$ (i=0,1,…,m) is convex. If there exists an $\mathbf{x} \in \text{relint}\,\mathcal{D}$ such that,

$$f_i(\mathbf{x}) < 0, i = 1, ..., m, h_i(\mathbf{x}) = 0, i = 1, ..., p \quad (\mathbf{x} \text{ is strictly feasible})$$

Then, such a primal problem has strong duality.

Proposition 18: Refined Slater condition

For a convex optimization problem, if the first k constraint functions $f_1, f_2, \ldots, f_k$ are affine, the strong duality holds provided the following weaker conditions: there is an $\mathbf{x} \in \text{relint}\,\mathcal{D}$ with

$$f_i(\mathbf{x}) \le 0, i = 1, ..., k, f_i(\mathbf{x}) < 0, i = k+1, ..., m, h_i(\mathbf{x}) = 0, i = 1, ..., p$$

Duality

Proposition 19: Slater conditions for a convex problem whose constraint functions are all affine

If the primal problem is convex, all its constraints (both the equality constraints and the inequality constraints) are affine, and $\mathbf{dom}f_0$ is open, then the Slater condition reduces to feasibility. In other words, under such conditions, if the primal problem is feasible, it should satisfy the Slater condition and has strong duality.

证明：

根据已知条件，该凸优化问题的不等式约束函数 $f_1, f_2, \ldots, f_m$ 都为仿射函数，等式约束函数为仿射函数 $A\mathbf{x}\text{-}\mathbf{b}$。仿射函数的定义域均为 $\mathbb{R}^n$。这样，该优化问题的定义域 $\mathcal{D}$ 为 $\mathbf{dom}f_0$ 与所有约束函数定义域的交集，则 $\mathcal{D} = \mathbf{dom}f_0 \cap \mathbb{R}^n \cap \mathbb{R}^n \cdots \cap \mathbb{R}^n = \mathbf{dom}f_0$，而 $\mathbf{dom}f_0$ 又是开集，则 $\mathcal{D}$ 为 $\mathbb{R}^n$ 中的开集，再由开集的定义 可知，$\mathbf{int}\mathcal{D} = \mathcal{D}$。另一方面，由于 $\mathcal{D}$ 为 $\mathbb{R}^n$ 中的开集，则 $\mathbf{aff}\mathcal{D} = \mathbb{R}^n$，则进一步 可知，$\mathbf{relint}\mathcal{D} = \mathbf{int}\mathcal{D}$。则有 $\mathbf{relint}\mathcal{D} = \mathcal{D}$。若该问题可行，则说明至少有一点 $\mathbf{x} \in \mathcal{D} = \mathbf{relint}\mathcal{D}$ 满足所有约束条件，即，

$$f_i(\mathbf{x}) \le 0, i = 1, \ldots, m, A\mathbf{x} = \mathbf{b}$$

根据 Prop. 18 可知，该凸优化问题满足修正斯莱特条件，因此它具有强对偶性。

Duality

Definition 18: Duality gap

Suppose that $\mathbf{x}$ and $(\boldsymbol{\alpha},\boldsymbol{\beta})$ are primal feasible and dual feasible, respectively. The duality gap associated with $\mathbf{x}$ and $(\boldsymbol{\alpha},\boldsymbol{\beta})$ is defined as,

$$f_0(\mathbf{x}) - g(\boldsymbol{\alpha},\boldsymbol{\beta})$$

Proposition 20:

Suppose that $\mathbf{x}^*$ and $(\boldsymbol{\alpha}^*,\boldsymbol{\beta}^*)$ are primal feasible and dual feasible respectively. If the duality gap associated with them is 0, i.e.,

$$f_0(\mathbf{x}^*) - g(\boldsymbol{\alpha}^*,\boldsymbol{\beta}^*) = 0$$

Then, $\mathbf{x}^*$ and $(\boldsymbol{\alpha}^*,\boldsymbol{\beta}^*)$ should be primal optimal and dual optimal, respectively, and the primal problem has strong duality

Duality

Suppose that an optimization problem (the primal problem) has strong duality. If $\mathbf{x}^*$ is the optimal solution for the primal problem and $(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$ is the optimal solution for the dual problem, then we have,

$$
\begin{aligned}
f_0(\mathbf{x}^*) &= g(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) \\
&= \min_{\mathbf{x}} \left(f_0(\mathbf{x}) + \sum_{i=1}^{m} \alpha_i^* f_i(\mathbf{x}) + \sum_{i=1}^{p} \beta_i^* h_i(\mathbf{x}) \right) \\
&\le f_0(\mathbf{x}^*) + \sum_{i=1}^{m} \alpha_i^* f_i(\mathbf{x}^*) + \sum_{i=1}^{p} \beta_i^* h_i(\mathbf{x}^*) \\
&\le f_0(\mathbf{x}^*)
\end{aligned}
$$

$\Rightarrow$

$$
\begin{aligned}
f_0(\mathbf{x}^*) &= g(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) \\
&= \min_{\mathbf{x}} \left(f_0(\mathbf{x}) + \sum_{i=1}^{m} \alpha_i^* f_i(\mathbf{x}) + \sum_{i=1}^{p} \beta_i^* h_i(\mathbf{x}) \right) \\
&= f_0(\mathbf{x}^*) + \sum_{i=1}^{m} \alpha_i^* f_i(\mathbf{x}^*) + \sum_{i=1}^{p} \beta_i^* h_i(\mathbf{x}^*) \\
&= f_0(\mathbf{x}^*)
\end{aligned}
$$

$\Downarrow$

① $\mathbf{x}^*$ is the global minimizer of the Lagrangian $l(\mathbf{x}; \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) = f_0(\mathbf{x}) + \sum_{i=1}^{m} \alpha_i^* f_i(\mathbf{x}) + \sum_{i=1}^{p} \beta_i^* h_i(\mathbf{x})$

② $\alpha_i^* f_i(\mathbf{x}^*) = 0, i = 1, 2, \ldots, m$ $\Rightarrow$ $\begin{cases} \alpha_i^* > 0 \Rightarrow f_i(\mathbf{x}^*) = 0 \\ f_i(\mathbf{x}^*) < 0 \Rightarrow \alpha_i^* = 0 \end{cases}$ (Complementary slackness)

Duality

Proposition 21: KKT conditions for general optimization problems

Suppose that $f_0, f_1, ..., f_m, h_1, h_2, ..., h_p$ are differentiable and the primal problem has strong duality. If $\mathbf{x}^*$ and (α^*, β^*) are primal optimal and dual optimal respectively, they satisfy the following so-called **KKT conditions**,

$$f_i\left(\mathbf{x}^*\right) \le 0, i = 1, ..., m \qquad \textbf{(Eq. 3)}$$

$$h_i\left(\mathbf{x}^*\right) = 0, i = 1, ..., p \qquad \textbf{(Eq. 4)}$$

$$\alpha_i^* \ge 0, i = 1, ..., m \qquad \textbf{(Eq. 5)}$$

$$\alpha_i^* f_i\left(\mathbf{x}^*\right) = 0, i = 1, ..., m \qquad \textbf{(Eq. 6)}$$

$$\nabla_{\mathbf{x}=\mathbf{x}^*} f_0(\mathbf{x}) + \sum_{i=1}^{m} \alpha_i^* \nabla_{\mathbf{x}=\mathbf{x}^*} f_i(\mathbf{x}) + \sum_{i=1}^{p} \beta_i^* \nabla_{\mathbf{x}=\mathbf{x}^*} h_i(\mathbf{x}) = \mathbf{0} \qquad \textbf{(Eq. 7)}$$

Proposition 22: KKT conditions for convex optimization problems

Suppose the primal problem is convex, i.e., f_0, f_1, f_2, …, f_m are convex and h_1, h_2, …, h_p are affine, all its constraint functions are differentiable, and the problem satisfies Slater conditions. Then, $\mathbf{x}_0$ and $(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$ are primal optimal and dual optimal, respectively, if and only if they satisfy KKT conditions,

$$f_i(\mathbf{x}_0) \le 0, i = 1, \dots, m$$

$$h_i(\mathbf{x}_0) = 0, i = 1, \dots, p$$

$$\alpha_{0i} \ge 0, i = 1, \dots, m$$

$$\alpha_{0i} f_i(\mathbf{x}_0) = 0, i = 1, \dots, m$$

$$\nabla_{\mathbf{x}=\mathbf{x}_0} f_0(\mathbf{x}) + \sum_{i=1}^{m} \alpha_{0i} \nabla_{\mathbf{x}=\mathbf{x}_0} f_i(\mathbf{x}) + \sum_{i=1}^{p} \beta_{0i} \nabla_{\mathbf{x}=\mathbf{x}_0} h_i(\mathbf{x}) = \mathbf{0}$$

Duality

Proposition 22: KKT conditions for convex optimization problems

证明：

必要性，即要证明：如果 $\mathbf{x}_0$ 和 (α_0,β_0) 分别为原问题和对偶问题的最优解，则它们满足KKT 条件。由于已知原问题满足斯莱特条件，则可知原问题具有强对偶性；又因为 $\mathbf{x}_0$ 和 (α_0,β_0) 分别为原问题和对偶问题的最优解且原问题所有约束函数都可微，根据 Prop. 21 ，$\mathbf{x}_0$ 和 (α_0,β_0) 满足 KKT 条件。

Proposition 22: KKT conditions for convex optimization problems

充分性，即要证明：如果 $\mathbf{x}_0$ 和 $(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)$ 满足 KKT 条件，则它们必然分别是原问题和对偶问题的最优解。由于 $\mathbf{x}_0$ 和 $(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)$ 满足 KKT 条件，则有，

$$
\begin{aligned}
f_i(\mathbf{x}_0) &\le 0, i=1,\ldots,m\\
h_i(\mathbf{x}_0) &= 0, i=1,\ldots,p\\
\alpha_{0i} &\ge 0, i=1,\ldots,m\\
\alpha_{0i}f_i(\mathbf{x}_0) &= 0, i=1,\ldots,m
\end{aligned}
$$

$$
\nabla_{\mathbf{x}=\mathbf{x}_0} f_0(\mathbf{x})+\sum_{i=1}^{m}\alpha_{0i}\nabla_{\mathbf{x}=\mathbf{x}_0} f_i(\mathbf{x})+\sum_{i=1}^{p}\beta_{0i}\nabla_{\mathbf{x}=\mathbf{x}_0} h_i(\mathbf{x})=\mathbf{0}
$$

由于原问题为凸优化问题，则可知 $f_0,f_1,\ldots,f_m$ 都是凸函数且 $h_1,\ldots,h_p$ 都为仿射函数（见凸优化问题 Def. 12 ）。又从 KKT 条件第 3 条知道，$\alpha_{0i}\ge 0$，则可知拉格朗日函数，

$$
l(\mathbf{x},\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)=f_0(\mathbf{x})+\sum_{i=1}^{m}\alpha_{0i}f_i(\mathbf{x})+\sum_{i=1}^{p}\beta_{0i}h_i(\mathbf{x})
$$

为关于 $\mathbf{x}$ 的凸函数（证明留作练习请读者完成）；同时，KKT 条件的最后一条表明了函数 $l(\mathbf{x};\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)$ 在 $\mathbf{x}_0$ 处的梯度为 $\mathbf{0}$，因此必有 $\mathbf{x}_0$ 为 $l(\mathbf{x};\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)$ 的最小值点。这样我们便有，

$$
g(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)=\min_{\mathbf{x}} l(\mathbf{x},\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)=l(\mathbf{x}_0,\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)=f_0(\mathbf{x}_0)+\sum_{i=1}^{m}\alpha_{0i}f_i(\mathbf{x}_0)+\sum_{i=1}^{p}\beta_{0i}h_i(\mathbf{x}_0)=f_0(\mathbf{x}_0)
$$

其中，最后一个等式应用了 $h_i(\mathbf{x}_0)=0,\alpha_{0i}f_i(\mathbf{x}_0)=0$ 这两个已知条件（KKT 条件中的第 2、4 条）。这说明在 $\mathbf{x}_0$ 和 $(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)$ 处，对偶间隔为 0，再根据 Prop. 20 可知，$\mathbf{x}_0$ 和 $(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)$ 分别是原问题的最优解和对偶问题的最优解。

Duality

Example: Consider the following equality constrained convex quadratic program problem,

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} \frac{1}{2}\mathbf{x}^T P\mathbf{x} + \mathbf{q}^T\mathbf{x} + r$$

$$\text{subject to } A\mathbf{x} = \mathbf{b}, A \in \mathbb{R}^{p\times n}$$

where P is positive semidefinite

The Lagrangian is $l(\mathbf{x},\boldsymbol{\beta}) = \frac{1}{2}\mathbf{x}^T P\mathbf{x} + \mathbf{q}^T\mathbf{x} + r + \boldsymbol{\beta}^T(A\mathbf{x} - \mathbf{b})$

Denote the optimal solution for the primal problem by $\mathbf{x}^*$, and the optimal solution for the dual problem by $\boldsymbol{\beta}^*$

For this specific problem, the last equation in KKT conditions is, $\nabla_{\mathbf{x}=\mathbf{x}^*} l(\mathbf{x};\boldsymbol{\beta}^*) = P\mathbf{x}^* + \mathbf{q} + A^T\boldsymbol{\beta}^* = \mathbf{0}$

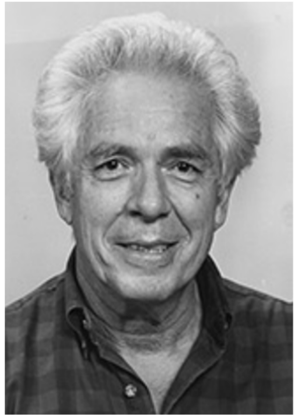
For this specific problem, the 2nd equation in KKT conditions is, $A\mathbf{x}^* = \mathbf{b}$

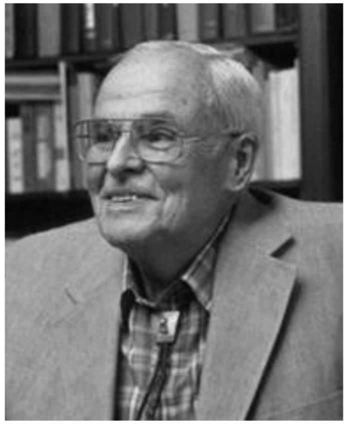
$$\begin{bmatrix} P & A^T \\ A & \mathbf{0} \end{bmatrix}\begin{bmatrix} \mathbf{x}^* \\ \boldsymbol{\beta}^* \end{bmatrix} = \begin{bmatrix} -\mathbf{q} \\ \mathbf{b} \end{bmatrix}$$

By solving the above linear equation system, we can get both the primal optimal and dual optimal solutions

Duality


(a)


(b)


(c)

(a) William Karush (March 1, 1917 to February 22, 1997), a mathematics professor at Northridge, California State University; In his master's thesis, he first proposed the necessary conditions for the optimal solution of inequality constrained problems

(b) Harold W. Kuhn (July 29, 1925 to July 2, 2014), an American mathematician at Princeton University, won the 1980 von Neumann Theory Award together with David Gale and Albert William Tucker; He acted as a math consultant in the movie "Beautiful Mind", which was adapted from Nash's life in 2001

(c) Albert William Tucker (November 28, 1905 to January 25, 1995), a Canadian mathematician, has made important contributions to topology, game theory and nonlinear programming; He had been a professor at Princeton University in 1933 and retired in 1974


Q+A