Machine Learning and Pattern Recognition

东北大学“智能+X”新工科课程系列

机器学习与模式识别

东北大学 信息科学与工程学院
人工智能系 、智能感知与机器人研究所
陈东岳

Convex Optimization-Math Foundation

凸优化数学基础

CHAPTER ONE

几何解释

Geometric Interpretation

一、几何解释

▶ 有约束凸优化问题

- 问题定义

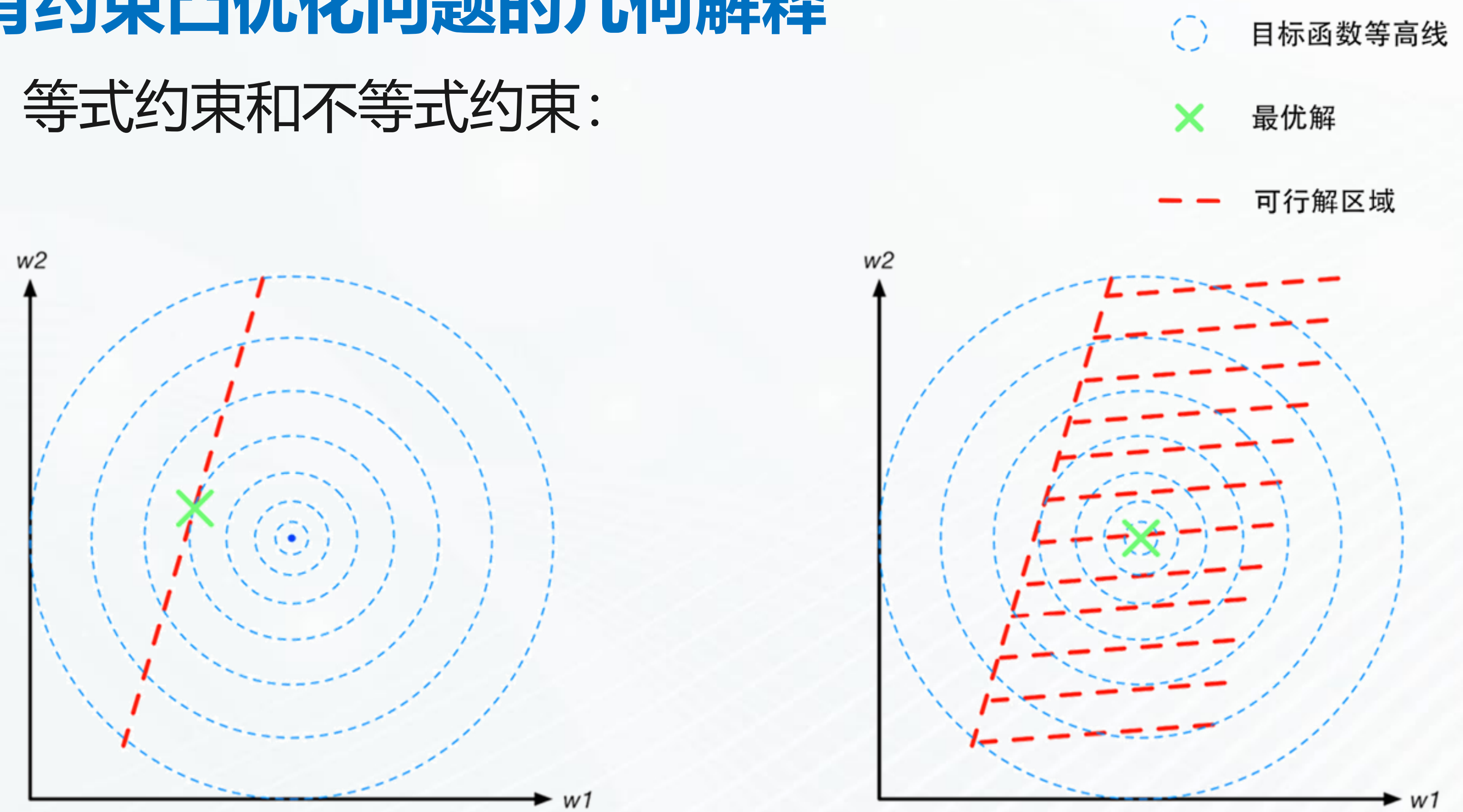
$$
\begin{aligned}
&\min_{\boldsymbol{x}} f(\boldsymbol{x}) \\
&\text{s.t.}\ \ h(\boldsymbol{x}) = 0 \\
&\qquad\ \ g(\boldsymbol{x}) \le 0
\end{aligned}
$$

- $f(\boldsymbol{x})$：目标函数
- $h(\boldsymbol{x}) = 0$：等式约束条件
- $g(\boldsymbol{x}) \le 0$：不等式约束条件
- 可行解区域：满足所有约束条件的解的集合

一、几何解释

▶ 有约束凸优化问题的几何解释

- 等式约束和不等式约束：



等式约束情况
$g(\boldsymbol{x}) = 0$

不等式约束情况
$g(\boldsymbol{x}) \leq 0$

CHAPTER TWO

拉格朗日乘子法

Lagrange Multiplier Method

二、拉格朗日乘子法

▶ 基本描述

- 等式约束的凸优化问题：
  - 目标函数：$f(\boldsymbol{x})$
  - 等式约束：$g(\boldsymbol{x}) = 0$；

- 基本思想
  - 将约束条件$g(\boldsymbol{x}) = 0$转化为目标函数的一部分
  - 将有约束优化问题转化为无约束优化问题

- 拉格朗日乘子式

拉格朗日乘子

$$L(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) + \boxed{\lambda} g(\boldsymbol{x})$$

二、拉格朗日乘子法

▶ 求解

- 根据新目标函数$L(\boldsymbol{x},\lambda)$求取最优解$\boldsymbol{x}^*$ ：

- 目标函数

$$L(\boldsymbol{x},\lambda) = f(\boldsymbol{x}) + \lambda g(\boldsymbol{x})$$

- 最优解

$$\frac{\partial L(\boldsymbol{x},\lambda)}{\partial \boldsymbol{x}} = 0 \quad \Rightarrow \quad \nabla f(\boldsymbol{x}^*) + \lambda \nabla g(\boldsymbol{x}^*) = 0$$

$$\frac{\partial L(\boldsymbol{x},\lambda)}{\partial \boldsymbol{\lambda}} = 0 \quad \Rightarrow \quad g(\boldsymbol{x}^*) = 0$$

CHAPTER THREE

KKT条件

KKT Conditions

三、KKT条件

为什么讨论KKT条件

- KKT条件不是对可行解的约束；
- KKT条件是最优解的充分必要条件；
- 解决优化问题可以转化为寻找满足KKT条件的解的过程

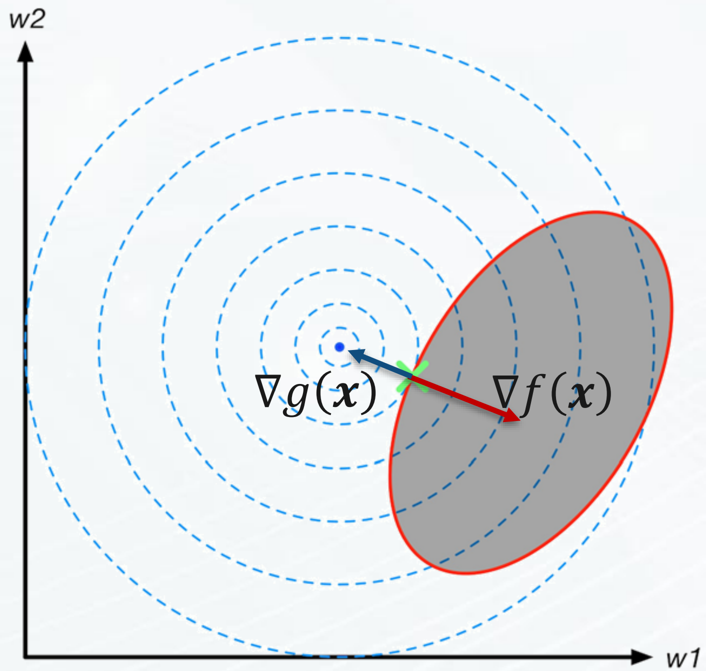
KKT条件的证明

- 情况1：最优解$\boldsymbol{x}^*$所在位置在边界$g(\boldsymbol{x}) = 0$上
- 情况2：最优解$\boldsymbol{x}^*$所在位置为$g(\boldsymbol{x}) < 0$的地方

三、KKT条件

▶ 情况1：最优解x^*在边界$g(x)=0$上

$$\nabla f(\boldsymbol{x}^*) + \lambda \nabla g(\boldsymbol{x}^*) = 0$$



- 在最优解$\boldsymbol{x}^*$附近:

可行解区域外侧，$g(\boldsymbol{x}) > 0$;

可行解区域内侧，$g(\boldsymbol{x}) < 0$;

$\Rightarrow \nabla g(\boldsymbol{x})$指向可行解区域外侧;

- 目标函数$f(\boldsymbol{x})$

在可行解区域外侧较小;

在可行解区域内侧较大;

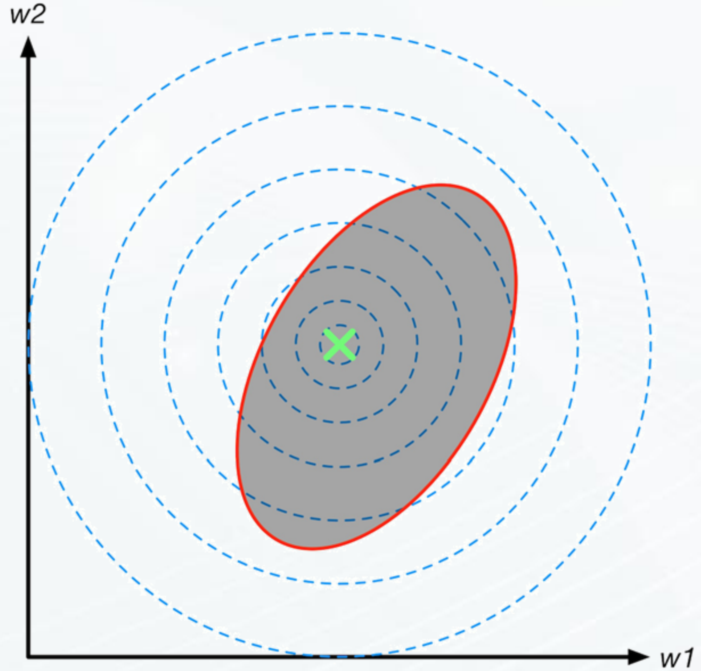
$\Rightarrow \nabla f(\boldsymbol{x})$指向可行解区域内侧;

$\Rightarrow g(\boldsymbol{x}^*) = 0, \lambda > 0$

三、KKT条件

▶ 情况2：最优解$\boldsymbol{x}^*$在$g(\boldsymbol{x})<0$内

$$\nabla f(\boldsymbol{x}^*) + \lambda \nabla g(\boldsymbol{x}^*) = 0$$



- 最优解$\boldsymbol{x}^*$ 在可行解的内部；
- $\boldsymbol{x}^*$不在边界$g(\boldsymbol{x})=0$ 上；

$\Longrightarrow$ $\nabla f(\boldsymbol{x}^*) = 0$，约束条件不起作用

$\Longrightarrow$ $g(\boldsymbol{x}^*) < 0, \lambda = 0$

三、KKT条件

▶ KKT条件的数学描述

情况1：
$$g(\boldsymbol{x}^*) = 0, \lambda > 0$$

情况 2：
$$g(\boldsymbol{x}^*) < 0, \lambda = 0$$

$$\begin{aligned} g(\boldsymbol{x}^*) &\le 0 \\ \lambda &\ge 0 \\ \lambda g(\boldsymbol{x}^*) &= 0 \end{aligned}$$

Karush-Kuhn-Tucker条件

CHAPTER FOUR

拉格朗日对偶

Lagrange Dual

四、拉格朗日对偶

▶ 问题转化

将有约束优化问题转化为无约束优化问题

- 原目标函数（有约束）

$$\min_{\boldsymbol{x}} f(\boldsymbol{x})$$

$$\text{s.t.}\ \ h_i(\boldsymbol{x}) = 0, i = 1,2, \dots, m$$

$$g_j(\boldsymbol{x}) \le 0, j = 1,2, \dots n$$

- 对新目标函数的要求

  - 在可行解区域（定义为Φ）内与原目标函数一致
  - 在可行解区域（Φ）外的数值非常大，甚至无穷大

一、拉格朗日对偶

广义拉格朗日乘子式

$$\theta_P(\boldsymbol{x}) = \max_{\boldsymbol{\alpha},\boldsymbol{\beta};\beta_j \geq 0} L(\boldsymbol{x},\alpha,\beta)$$

其中:

$$L(\boldsymbol{x},\alpha,\beta) = f(\boldsymbol{x}) + \sum_{i=1}^{m} \alpha_i\, h_i(\boldsymbol{x}) + \sum_{j=1}^{n} \beta_j\, g_j(\boldsymbol{x})$$

所以:

$$\theta_P(\boldsymbol{x}) = \max_{\boldsymbol{\alpha},\boldsymbol{\beta};\beta_j \geq 0} L(\boldsymbol{x},\alpha,\beta) = f(\boldsymbol{x}) + \max_{\boldsymbol{\alpha},\boldsymbol{\beta};\beta_j \geq 0}\left[\sum_{i=1}^{m} \alpha_i\, h_i(\boldsymbol{x}) + \sum_{j=1}^{n} \beta_j\, g_j(\boldsymbol{x})\right]$$

一、拉格朗日对偶

广义拉格朗日乘子式

$$\theta_P(\boldsymbol{x}) = \max_{\boldsymbol{\alpha},\boldsymbol{\beta};\beta_j\geq 0} L(\boldsymbol{x},\alpha,\beta) = f(\boldsymbol{x}) + \max_{\boldsymbol{\alpha},\boldsymbol{\beta};\beta_j\geq 0}\left[\sum_{i=1}^{m}\alpha_i h_i(\boldsymbol{x}) + \sum_{j=1}^{n}\beta_j g_j(\boldsymbol{x})\right]$$

- 可行解区域内（ $\forall \boldsymbol{x} \in \Phi$ ）

$$\sum_{i=1}^{m}\alpha_i h_i(\boldsymbol{x})=0，\ \sum_{j=1}^{n}\beta_j g_j(\boldsymbol{x}) \leq 0$$

$$\Rightarrow \max_{\boldsymbol{\alpha},\boldsymbol{\beta};\beta_j\geq 0}\left(\sum_{i=1}^{m}\alpha_i h_i(\boldsymbol{x}) + \sum_{j=1}^{n}\beta_j g_j(\boldsymbol{x})\right) = 0$$

- 可行解区域外（ $\forall \boldsymbol{x} \notin \Phi$ ）

$$\sum_{i=1}^{m}\alpha_i h_i(\boldsymbol{x}) \in (-\infty,+\infty), \sum_{j=1}^{n}\beta_j g_j(\boldsymbol{x}) \in [\mathbf{0},+\infty)$$

$$\Rightarrow \max_{\boldsymbol{\alpha},\boldsymbol{\beta};\beta_j\geq 0}\left(\sum_{i=1}^{m}\alpha_i h_i(\boldsymbol{x}) + \sum_{j=1}^{n}\beta_j g_j(\boldsymbol{x})\right) = +\infty$$

一、拉格朗日对偶

▶ 问题转换结果

- 有约束优化问题

$$\min_{\boldsymbol{x}} f(\boldsymbol{x})$$
$$\text{s.t.}\ \ h_i(\boldsymbol{x}) = 0, i = 1,2,\dots,m$$
$$g_j(\boldsymbol{x}) \le 0, j = 1,2,\dots n$$

- 无约束优化问题

$$\min_{\boldsymbol{x}} \theta_P(\boldsymbol{x}) = \min_{\boldsymbol{x}} \left[\max_{\boldsymbol{\alpha},\boldsymbol{\beta};\beta_j \ge 0} L(\boldsymbol{x},\boldsymbol{\alpha},\boldsymbol{\beta}) \right]$$

$$L(\boldsymbol{x},\boldsymbol{\alpha},\boldsymbol{\beta}) = f(\boldsymbol{x}) + \sum_{i=1}^{m} \alpha_i h_i(\boldsymbol{x}) + \sum_{j=1}^{n} \beta_j g_j(\boldsymbol{x})$$

拉格朗日对偶

一、拉格朗日对偶

拉格朗日对偶方法

将原始问题(Primary)转化为对偶问题(Dual)进行求解

- 原始问题 $\min_{\boldsymbol{x}} \theta_P(\boldsymbol{x}) = \min_{\boldsymbol{x}} \left[\max_{\boldsymbol{\alpha},\boldsymbol{\beta};\beta_j \geq 0} L(\boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \right]$
- 对偶问题 $\max_{\boldsymbol{\alpha},\boldsymbol{\beta};\beta_j \geq 0} \theta_D(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \max_{\boldsymbol{\alpha},\boldsymbol{\beta};\beta_j \geq 0} \left[\min_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \right]$
- 原始问题与对偶问题会有相同的解吗？

一、拉格朗日对偶

拉格朗日对偶方法

- 定理：

- 对于原始问题和对偶问题:

- 假设函数$f(\boldsymbol{x})$和不等式约束条件$g_j(\boldsymbol{x}),\forall j$均为凸函数，

- 等式约束条件中$h_i(\boldsymbol{x})$为$\boldsymbol{x}$的仿射函数$h_i(\boldsymbol{x}) = \mathbf{a}_i^T\boldsymbol{x} + b_i$;

- 至少存在一个$\boldsymbol{x}$使所有不等式约束条件严格成立，即$g_j(\boldsymbol{x}) < 0, \forall j$；

- 则存在$\boldsymbol{x}^*,\boldsymbol{\alpha}^*,\boldsymbol{\beta}^*$使得$\boldsymbol{x}^*$是原始问题的最优解，$\boldsymbol{\alpha}^*,\boldsymbol{\beta}^*$是对偶问题的最优解且有：$d^*=p^*=L(\boldsymbol{x}^*,\boldsymbol{\alpha}^*,\boldsymbol{\beta}^*)$，其充分必要条件如下：

$$\nabla_x(\boldsymbol{x}^*,\boldsymbol{\alpha}^*,\boldsymbol{\beta}^*) = 0$$
$$\nabla_\alpha(\boldsymbol{x}^*,\boldsymbol{\alpha}^*,\boldsymbol{\beta}^*) = 0$$
$$\nabla_\beta(\boldsymbol{x}^*,\boldsymbol{\alpha}^*,\boldsymbol{\beta}^*) = 0$$
$$h_i(\boldsymbol{x}^*) = 0, i = 1,2\ldots,m$$

$$\boxed{\begin{array}{l} g_j(\boldsymbol{x}^*) \le 0, j = 1,2\ldots,n \\ \beta_j^* \ge 0, j = 1,2\ldots,n \\ \beta_j^* g_j(\boldsymbol{x}^*) = 0, j = 1,2\ldots,n \end{array}}$$

原始KKT条件

示 例

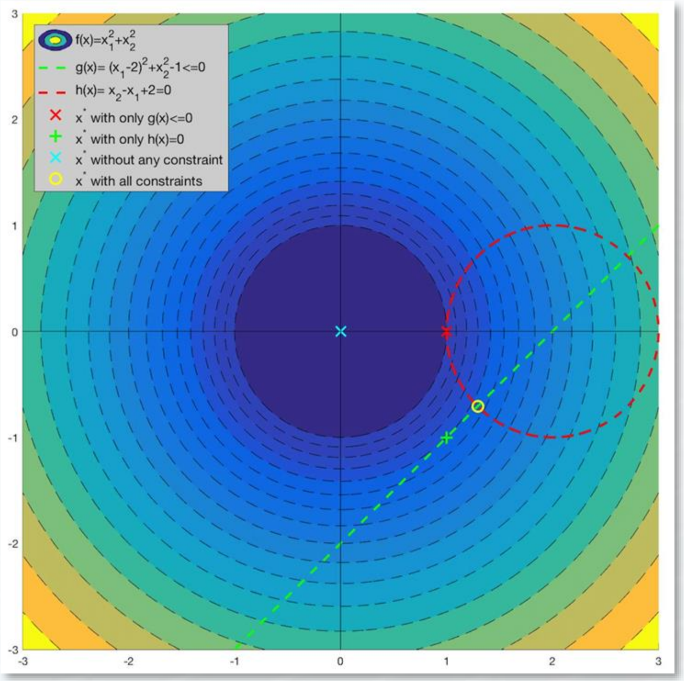
例题——对偶问题求解示例

问题描述

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

$$\text{s.t. } h(\boldsymbol{x}) = x_2 - x_1 + 2 = 0$$

$$g(\boldsymbol{x}) = (x_1 - 2)^2 + x_2^2 - 1 \le 0$$



- 目标函数：二项式（凸函数）
- 等式约束：直线（线性）
- 不等式约束：圆形区域

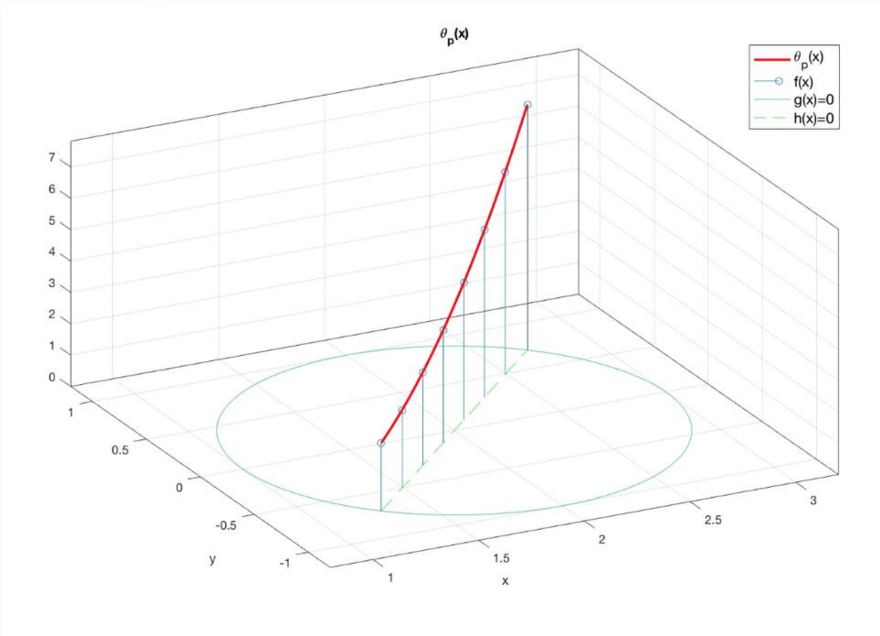
符合拉格朗日对偶条件

例题——对偶问题求解示例

广义拉格朗日乘子式

$$L(\boldsymbol{x},\alpha,\beta)=\left(x_1^2+x_2^2\right)+\alpha(x_2-x_1+2)+\beta[(x_1-2)^2+x_2^2-1]$$

$$\theta_P(\boldsymbol{x})=\max_{\alpha,\beta;\beta_j\geq 0} L(\boldsymbol{x},\alpha,\beta)$$



$$\theta_P(\boldsymbol{x})=\begin{cases} f(\boldsymbol{x}), & \boldsymbol{x}\in\emptyset \\ +\infty, & \text{else}\end{cases}$$

- 可行解区域Φ：
  直线$h(\boldsymbol{x})=0$与圆
  $g(\boldsymbol{x})\leq 0$相交的弦

例题——对偶问题求解示例

对偶问题转化

$$\min_{\boldsymbol{x}} f(\boldsymbol{x}) = x_1^2 + x_2^2$$
$$\text{s.t. } h(\boldsymbol{x}) = x_2 - x_1 + 2 = 0$$
$$g(\boldsymbol{x}) = (x_1 - 2)^2 + x_2^2 - 1$$

⟷ 原始有约束优化问题

↓

$$\min_{\boldsymbol{x}} \theta_P(\boldsymbol{x}) = \min_{\boldsymbol{x}} \left[\max_{\alpha,\beta;\beta\geq 0} L(\boldsymbol{x},\alpha,\beta) \right]$$
$$L(\boldsymbol{x},\alpha,\beta) = \left(x_1^2 + x_2^2\right) + \alpha(x_2 - x_1 + 2) + \beta[(x_1 - 2)^2 + x_2^2 - 1]$$

⟷ 原始无约束优化问题

↓

$$\max_{\alpha,\beta;\beta\geq 0} \theta_D(\alpha,\beta) = \max_{\alpha,\beta;\beta\geq 0} \left[\min_{\boldsymbol{x}} L(\boldsymbol{x},\alpha,\beta) \right]$$
$$L(\boldsymbol{x},\alpha,\beta) = \left(x_1^2 + x_2^2\right) + \alpha(x_2 - x_1 + 2) + \beta[(x_1 - 2)^2 + x_2^2 - 1]$$

⟷ 对偶问题

例题——对偶问题求解示例

▶ 对偶问题求解

- 广义拉格朗日乘子式L对x_1和x_2分别求偏导数：

$$L(\boldsymbol{x},\alpha,\beta)=\left(x_1^2+x_2^2\right)+\alpha(x_2-x_1+2)+\beta[(x_1-2)^2+x_2^2-1]$$

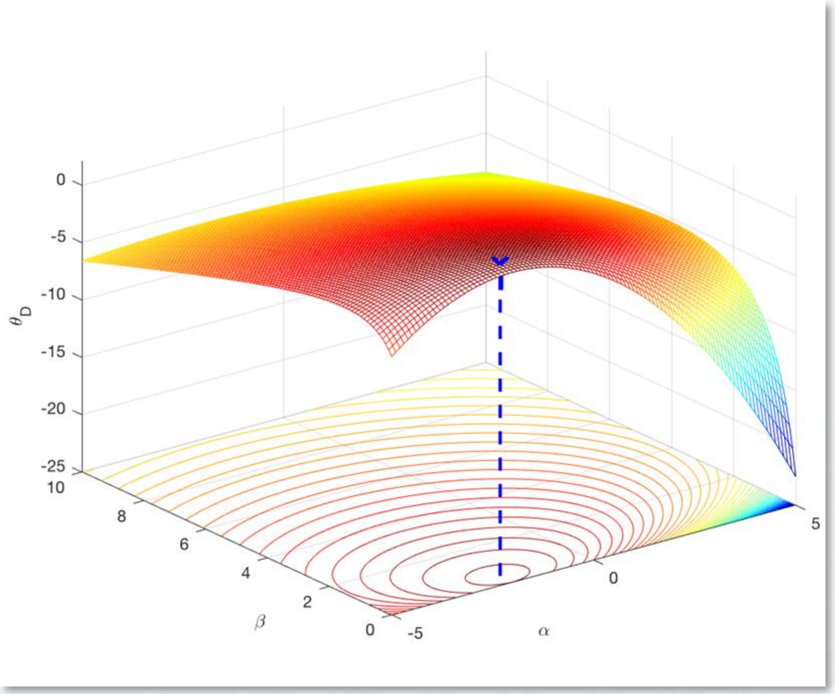
$$\begin{cases}\dfrac{\partial L}{\partial x_1}=\alpha+2x_1+\beta(2x_1-4)=0\\[2mm]\dfrac{\partial L}{\partial x_2}=2x_2-\alpha+2\beta=0\end{cases}\Rightarrow\begin{cases}x_1^*=\dfrac{4\beta-\alpha}{2\beta+2}\\[2mm]x_2^*=\dfrac{\alpha}{2\beta+2}\end{cases}$$

- 令$\theta_D(\alpha,\beta)=\min\limits_{\boldsymbol{x}}L(\boldsymbol{x},\alpha,\beta)=L(\boldsymbol{x}^*,\alpha,\beta)$，得：

$$\theta_D(\alpha,\beta)=-\frac{\alpha^2+4\alpha+2\beta^2-6\beta}{2(\beta+1)}$$

例题——对偶问题求解示例

对偶问题求解



$$\theta_D(\alpha,\beta) = -\frac{\alpha^2 + 4\alpha + 2\beta^2 - 6\beta}{2(\beta+1)}$$

⬇

$$\alpha^*, \beta^* = \underset{\alpha,\beta}{\operatorname{argmax}}(\theta_D(\alpha,\beta))$$

⬇

$$\partial\theta_D/\partial\alpha = 0, \partial\theta_D/\partial\beta = 0$$

⬇

$$\alpha^* = -2, \qquad \beta^* = \sqrt{2} - 1$$

⬇

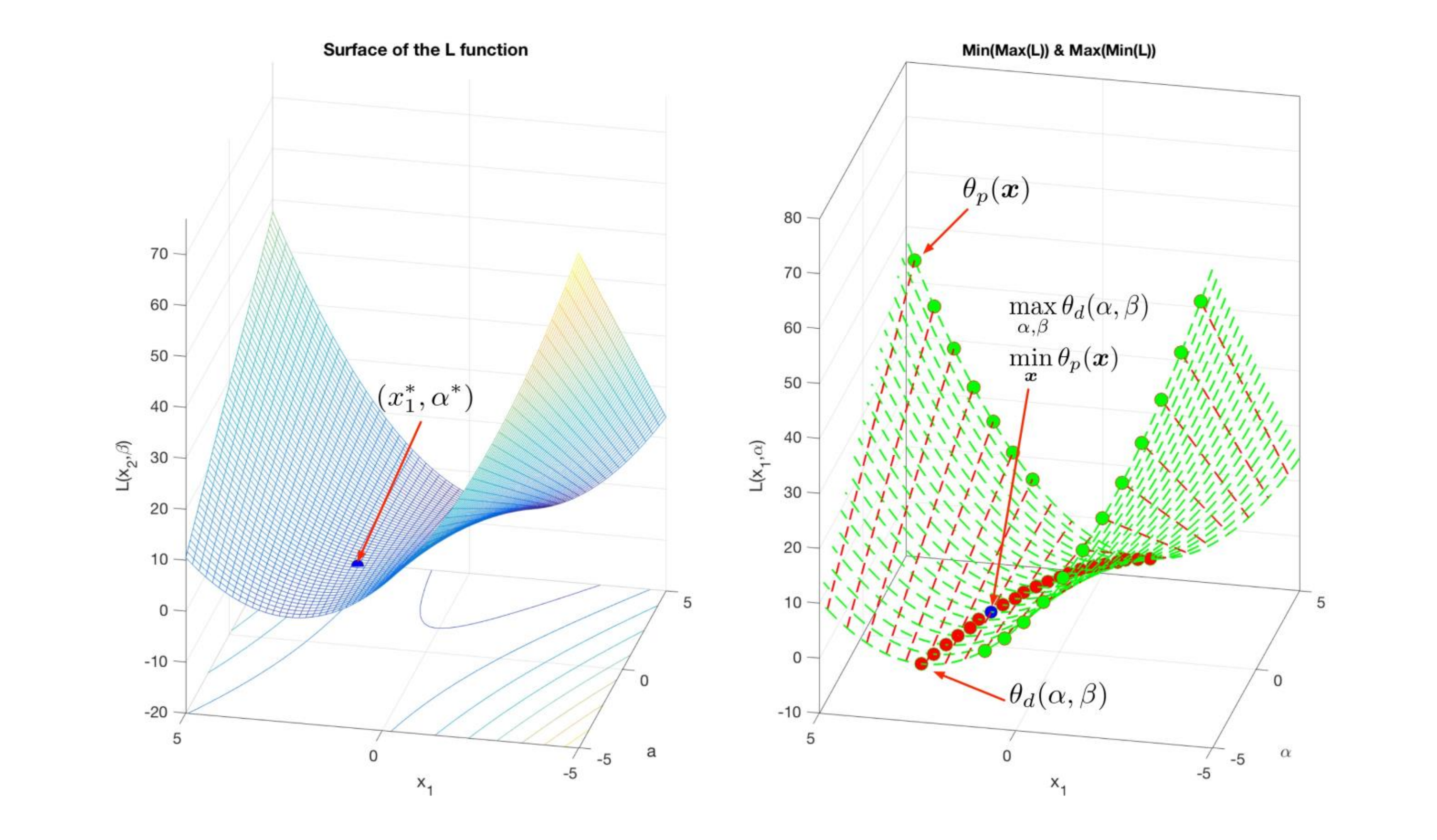
$$x_1 = \frac{4\beta - \alpha}{2\beta + 2}, x_2 = \frac{\alpha}{2\beta + 2}$$

⬇

$$x_1^* = 2 - \sqrt{2}/2\,, x_2^* = -\sqrt{2}/2$$

例题——对偶问题求解示例

拉格朗日对偶的几何解释

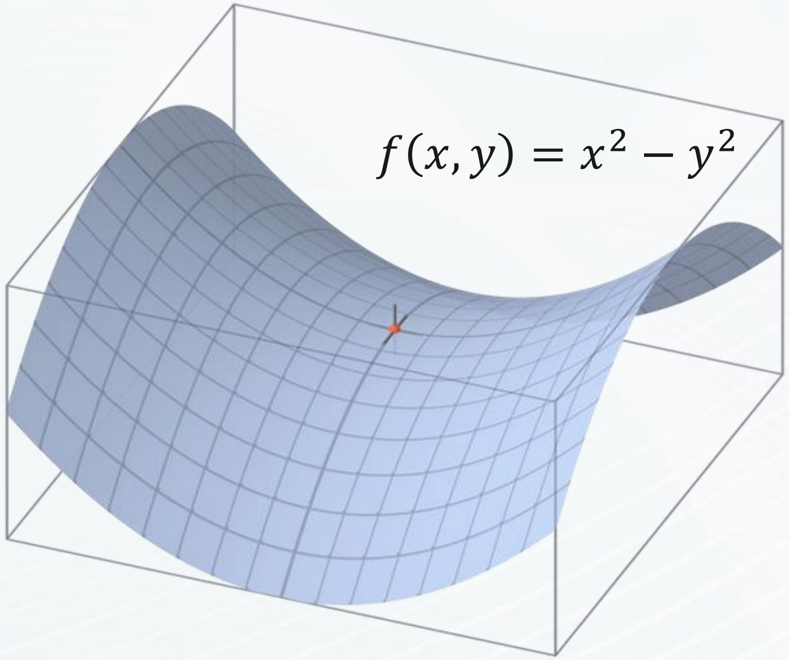


例题——对偶问题求解示例

▶ 拉格朗日对偶的几何解释

- 原始问题与对偶问题的等价解是——

广义拉格朗日乘子式的鞍点



在几何形状上看：

某个方向上：

原始问题：所有最大值中的最小值；

另一方向上：

对偶问题：所有最小值中的最大值。

总 结

- 凸优化数学基础
  - 几何解释
  - 拉格朗日乘子法
    - 基本描述
      - 目标函数
      - 约束条件
    - 拉格朗日乘子式
    - 求解
  - KKT条件
    - KKT条件的定义与用途
    - 解释论证KKT条件的两种情况
    - KKT条件的数学描述
  - 拉格朗日对偶
    - 拉格朗日对偶的定义与用途
    - 原始问题与对偶问题转换
      - 原始目标函数
      - 对构造新目标函数提出要求
      - 广义拉格朗日乘子式
    - 拉格朗日对偶法
      - 原始问题与对偶问题的等价性
      - 对偶问题的规范化表达
    - 示例：二元二次函数的有约束优化问题
      - 问题描述
      - 原始问题属性模型
      - 对偶问题数学模型
      - 对偶问题求解
      - 几何解释
        - 鞍点概念
        - 等价性

S M A R T

THANK YOU

感谢聆听

MADE BY DONGYUE CHEN

三、分类准则

一级标题

最大后验概率分类准则 (for多类问题)

二级标题

分类准则

三级标题

中文：优设标题黑

思源黑体 CN Heavy　　一级标题

思源黑体 CN Medium　　二级标题

思源黑体 CN Normal　　三级标题

英文:Times New Roman

标准字号≥22,内容过多字号可适当调整

公式字体：Cambria Math

主要用色

SOURCEHANSANSCN-MEDIUM.OTF

SOURCEHANSANSCN-NORMAL.OTF

SourceHanSansCN-Heavy.otf