第 5 章：对偶

(2)：最优性条件

授课教师：曹语

课程主页：`https://yucaoyc.github.io/math3806`

从弱对偶到最优性条件

上一讲我们得到：任意对偶可行点都给出原问题的下界。

$$d^\star \leq p^\star$$

本节课回答两个进一步的问题：

- 什么时候下界刚好等于最优值？这就是强对偶；
- 怎样不用先解对偶问题，也能验证一个点是最优的？这就是 KKT 条件。

回顾：原问题与对偶问题

原始问题为

$$\begin{aligned} p^\star = \text{minimize} \quad & f_0(\boldsymbol{x}) \\ \text{subject to} \quad & f_i(\boldsymbol{x}) \le 0, \quad i = 1, \ldots, m, \\ & h_i(\boldsymbol{x}) = 0, \quad i = 1, \ldots, p, \end{aligned}$$

其中优化变量为 $\boldsymbol{x} \in \mathcal{D}$，且 $\mathcal{D}$ 为定义域。Lagrange 函数为

$$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i f_i(\boldsymbol{x}) + \sum_{i=1}^{p} \nu_i h_i(\boldsymbol{x}).$$

对偶问题为

$$d^\star = \sup_{\boldsymbol{\lambda} \succeq \boldsymbol{0},\ \boldsymbol{\nu}} \ g(\boldsymbol{\lambda}, \boldsymbol{\nu}), \qquad g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\boldsymbol{x} \in \mathcal{D}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}).$$

本节目标

学完本节课后，希望大家能够：

- 用 Slater 条件判断凸优化问题是否具有强对偶；
- 准确写出 KKT 条件；
- 用 KKT 条件解简单的等式/不等式约束优化问题；
- 解释 Lagrange 乘子为什么可以理解为约束的“边际价格”。

今天的重心是**会判断、会写、会用**，证明细节只保留主线思想。

课前热身：最大熵问题

考虑凸优化问题

$$
\begin{array}{ll}
\text{minimize} & \sum_{i=1}^{n} x_i \log x_i \\
\text{subject to} & \boldsymbol{A}\boldsymbol{x} \preceq \boldsymbol{b}, \\
& \mathbf{1}^\top \boldsymbol{x} = 1, \\
\text{domain} & \boldsymbol{x} \succeq \mathbf{0}.
\end{array}
$$

这里约定 $0 \log 0 = 0$。[1]

练习：请写出 Lagrange 函数，并尝试计算对偶函数。

[1]非负性在此作为定义域处理，因此计算对偶函数时在 $x_i \geq 0$ 上取下确界，不再配乘子。

热身答案：先写 Lagrange 函数

令 $\boldsymbol{\lambda} \succeq \mathbf{0}$ 对应约束 $\boldsymbol{A}\boldsymbol{x} \preceq \boldsymbol{b}$，令 $\nu \in \mathbb{R}$ 对应 $\mathbf{1}^\top \boldsymbol{x} = 1$。

$$\begin{aligned} L(\boldsymbol{x}, \boldsymbol{\lambda}, \nu) &= \sum_{i=1}^{n} x_i \log x_i + \boldsymbol{\lambda}^\top (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) + \nu(\mathbf{1}^\top \boldsymbol{x} - 1) \\ &= -\boldsymbol{b}^\top \boldsymbol{\lambda} - \nu + \sum_{i=1}^{n} \Big[x_i \log x_i + \underbrace{\big((\boldsymbol{A}^\top \boldsymbol{\lambda})_i + \nu\big)}_{=: s_i} x_i \Big]. \end{aligned}$$

逐个计算可得

$$\inf_{x_i \geq 0} x_i \log x_i + s_i x_i = -e^{-s_i - 1}.$$

热身答案：对偶函数与对偶问题

由上一页得到

$$g(\boldsymbol{\lambda}, \nu) = -\boldsymbol{b}^\top \boldsymbol{\lambda} - \nu - \sum_{i=1}^{n} e^{-(\boldsymbol{A}^\top \boldsymbol{\lambda})_i - \nu - 1}.$$

因此对偶问题为

$$\begin{array}{ll} \text{maximize} & -\boldsymbol{b}^\top \boldsymbol{\lambda} - \nu - \sum_{i=1}^{n} e^{-(\boldsymbol{A}^\top \boldsymbol{\lambda})_i - \nu - 1} \\ \text{subject to} & \boldsymbol{\lambda} \succeq \boldsymbol{0}, \quad \nu \in \mathbb{R}. \end{array}$$

问题：这个对偶问题给出的最优下界什么时候等于原问题最优值?

目录

Slater 条件的直观

强对偶通常需要某种“没有卡在边界上的可行性”。Slater 条件说：

> 对于凸优化问题，只要能找到一个点，使所有**非仿射不等式约束都严格满足**，仿射不等式可仅满足 ≤ 0，且满足等式约束，那么通常就有强对偶。

这里“严格满足”的含义是 $f_i(\boldsymbol{x}) < 0$，不是刚好等于 0。

为什么普通内部不够用?

回到最大熵问题中的约束:

$$\mathbf{1}^\top \boldsymbol{x} = 1, \qquad \boldsymbol{x} \succeq \mathbf{0}.$$

可行点都落在超平面 $\mathbf{1}^\top \boldsymbol{x} = 1$ 上。因此在整个 $\mathbb{R}^n$ 里，任何可行点附近的小球都会跑出这个超平面。

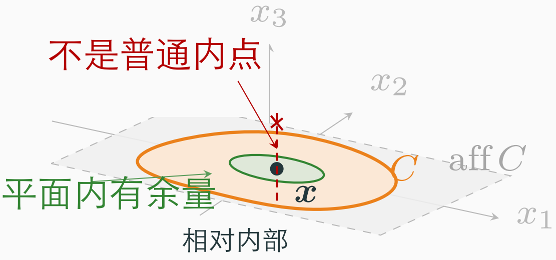
所以

$$\mathrm{int}\{\boldsymbol{x} \succeq \mathbf{0},\ \mathbf{1}^\top \boldsymbol{x} = 1\} = \emptyset.$$

但我们仍然认为 $\boldsymbol{x} \succ \mathbf{0}$ 且 $\mathbf{1}^\top \boldsymbol{x} = 1$ 是“没有贴在边界上”的概率分布。

问题: 应该在什么空间里看“内部”?

相对内点：在自己的平面里是内点



若集合本身落在一条直线或一个平面里，普通内部可能为空；相对内部只要求它在**自己的仿射空间**里像内点。也就是说，相对内部不要求在整个 $\mathbb{R}^n$ 里有活动空间，只要求在允许活动的方向里留有余量。

相对内点：正式定义

给定凸集 $C \subseteq \mathbb{R}^n$，它的仿射包 $\operatorname{aff} C$ 是包含 C 的最小仿射集合。

$$\operatorname{relint} C = \{\boldsymbol{x} \in C : \exists r > 0,\ B(\boldsymbol{x}, r) \cap \operatorname{aff} C \subseteq C\}.$$

也就是说，只沿着 C 自己所在的仿射空间看，$\boldsymbol{x}$ 附近的一小块仍然留在 C 里面。

例子： 概率单纯形的相对内部是

$$\{\boldsymbol{x} : \boldsymbol{x} \succ \mathbf{0},\ \mathbf{1}^\top \boldsymbol{x} = 1\}.$$

Slater 中的“严格”在哪里?

Slater 条件不要求点在整个空间里都能自由移动。

它只要求:

- 仿射结构、定义域造成的低维性是允许的;
- 非仿射不等式约束不能刚好卡在边界上。

因此 Slater 写成

$$\boldsymbol{x} \in \operatorname{relint} \mathcal{D}, \qquad f_i(\boldsymbol{x}) < 0, \qquad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}.$$

一句话:在不可避免的平面里,要留有余量;在真正的不等式边界上,不要贴边。

Slater 定理：常用版本

考虑凸优化问题

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

其中 f_0, f_i 为凸函数，$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ 为仿射等式约束，公共定义域记为 $\mathcal{D}$。

定理 (Slater 定理)

若存在 $\boldsymbol{x} \in \operatorname{relint} \mathcal{D}$ 使得

$$f_i(\boldsymbol{x}) < 0, \quad i = 1, \ldots, m, \qquad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},$$

且 $p^\star$ 有限，则强对偶成立，并且对偶最优值可达到。

Slater 条件的仿射约束版本

如果前 k 个不等式约束 $f_1, \dots, f_k$ 是仿射函数，那么 Slater 条件可以放宽为：

$$
\begin{aligned}
f_i(\boldsymbol{x}) &\le 0, \qquad & i &= 1, \dots, k, \\
f_i(\boldsymbol{x}) &< 0, \qquad & i &= k+1, \dots, m, \\
\boldsymbol{A}\boldsymbol{x} &= \boldsymbol{b}, \qquad & \boldsymbol{x} &\in \operatorname{relint} \mathcal{D}.
\end{aligned}
$$

判断 Slater 时，先问自己：

- 哪些不等式是仿射的？
- 是否存在一个点让所有非仿射不等式都严格小于 0？

简化版的证明见附录。

Slater 判断练习

练习：假设下列问题可行且 $p^\star$ 有限。哪些可以直接用 Slater 推出强对偶？

1. 等式约束二次问题

$$\min \ \boldsymbol{x}^\top \boldsymbol{x} \quad \text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}.$$

2. 线性规划

$$\min \ \boldsymbol{c}^\top \boldsymbol{x} \quad \text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \ \boldsymbol{x} \succeq \boldsymbol{0}.$$

3. 一维问题

$$\min \ x \quad \text{s.t.} \quad x^2 \le 0.$$

4. 最大熵问题（第5页）在什么条件下满足 Slater？

Slater 快速判断：答案

1. 可以。只有仿射等式约束，因此 Slater 条件成立。
2. 可以。这里把 $\boldsymbol{x} \succeq \mathbf{0}$ 写成仿射不等式 $-x_i \leq 0$，因此目标、等式和不等式约束都是仿射的，仿射版 Slater 条件成立。
3. 不可以直接用 Slater。唯一可行点是 $x = 0$，但 $x^2 < 0$ 不可能成立。这个例子提醒我们：**凸问题不等于自动满足 Slater**；不能用 Slater 也不代表强对偶一定失败。
4. 用仿射版 Slater 时，只需存在 $\boldsymbol{x} \succ \mathbf{0}$，满足 $\mathbf{1}^\top \boldsymbol{x} = 1$ 且 $\boldsymbol{A}\boldsymbol{x} \preceq \boldsymbol{b}$；若不使用仿射放宽版，$\boldsymbol{A}\boldsymbol{x} \prec \boldsymbol{b}$ 是更强的充分条件。

目录

原问题与对偶：交换优化顺序

原问题可以写成【证明见附录】

$$\inf_{\boldsymbol{x}\in\mathcal{D}} \sup_{\boldsymbol{\lambda}\succeq\mathbf{0},\ \boldsymbol{\nu}} L(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\nu}).$$

对偶问题则是

$$\sup_{\boldsymbol{\lambda}\succeq\mathbf{0},\ \boldsymbol{\nu}} \inf_{\boldsymbol{x}\in\mathcal{D}} L(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\nu}).$$

因此，弱对偶本质上是：

$$\sup_z \inf_w F(w,z) \le \inf_w \sup_z F(w,z)$$

强对偶就是：在足够好的条件下，交换顺序不改变答案。

鞍点性质

$$\sup_z \inf_w F(w,z) \leq \inf_w \sup_z F(w,z)$$

令左/右侧的外环优化问题的最优解分别为 $z^\star$ 和 $w^\star$，则可得如下形式：

$$\sup_z \inf_w F(w,z) = \inf_w F(w,z^\star) \leq \sup_z F(w^\star,z) \leq \inf_w \sup_z F(w,z).$$

若强对偶性成立，则

$$\inf_w F(w,z^\star) = \sup_z F(w^\star,z).$$

由此可知

$$F(w^\star,z^\star) \geq \inf_w F(w,z^\star) = \sup_z F(w^\star,z) \geq F(w^\star,z^\star).$$

由于上下界是一样的，我们可以得到鞍点性质（见下一页）。

如果存在 $(w^\star, z^\star)$ 满足

$$F(w^\star, z) \leq F(w^\star, z^\star) \leq F(w, z^\star), \qquad \forall w, z \tag{1}$$

则称 $(w^\star, z^\star)$ 是鞍点。

直观解释：

- 固定 $w^\star$ 时，$z^\star$ 让函数尽可能大；
- 固定 $z^\star$ 时，$w^\star$ 让函数尽可能小；

例子： $F(w, z) = w^2 - z^2$ 在 $(0, 0)$ 处有鞍点。

案例：矩阵对策的混合策略

玩家 1 选择 $k \in \{1, \dots, n\}$，玩家 2 选择 $l \in \{1, \dots, m\}$。玩家 1 向玩家 2 支付 $P_{k,l}$。若双方使用随机策略

$$\mathrm{Prob}(k = i) = w_i, \qquad \mathrm{Prob}(l = j) = z_j,$$

其中

$$W = \{\boldsymbol{w} : \boldsymbol{w} \succeq \mathbf{0},\ \mathbf{1}^\top \boldsymbol{w} = 1\}, \quad Z = \{\boldsymbol{z} : \boldsymbol{z} \succeq \mathbf{0},\ \mathbf{1}^\top \boldsymbol{z} = 1\},$$

则期望支付为

$$F(\boldsymbol{w}, \boldsymbol{z}) = \boldsymbol{w}^\top \boldsymbol{P} \boldsymbol{z}.$$

矩阵对策与强对偶

- 玩家 1 希望少付钱:

$$p_1^\star = \min_{\boldsymbol{w} \in W} \max_{\boldsymbol{z} \in Z} \boldsymbol{w}^\top \boldsymbol{P} \boldsymbol{z}.$$

- 玩家 2 希望多挣钱:

$$p_2^\star = \max_{\boldsymbol{z} \in Z} \min_{\boldsymbol{w} \in W} \boldsymbol{w}^\top \boldsymbol{P} \boldsymbol{z}.$$

- 一般总有 $p_2^\star \leq p_1^\star$。矩阵零和博弈中，线性规划对偶性给出 $p_1^\star = p_2^\star$（证明见课本 5.2.5；不要求）。

在 Lagrange 对偶中，$\boldsymbol{x}$ 和 $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ 也可以看成两方的博弈。

博弈视角小结

博弈语言	对偶语言
设计者先选 $\boldsymbol{x}$	原问题：先找一个满足约束的方案
检查者选择 $(\boldsymbol{\lambda}, \boldsymbol{\nu})$	乘子给违反约束的行为定价
$\min\max$ 与 $\max\min$ 的差距	原/对偶最优值之间的间隙
鞍点存在	双方都无法单方面改进，原/对偶值相等

强对偶：
$$\inf_{\boldsymbol{x}\in\mathcal{D}} \sup_{\boldsymbol{\lambda}\succeq 0,\boldsymbol{\nu}} L = \sup_{\boldsymbol{\lambda}\succeq 0,\boldsymbol{\nu}} \inf_{\boldsymbol{x}\in\mathcal{D}} L.$$

下一步：KKT 条件就是鞍点处的“没有单方面改进”的代数形式。

目录

为什么需要 KKT?

Slater 告诉我们**什么时候强对偶成立**；KKT 告诉我们**最优点长什么样**。

对于带约束问题，最优点通常不能只看 $\nabla f_0(\boldsymbol{x}^\star) = \mathbf{0}$。原因是有些方向不可行，例如约束会挡住下降方向。

KKT 是一组最优解满足的可检查的方程/不等式。

KKT 条件从哪里来：互补松弛

假设强对偶成立，且原问题最优解 $\boldsymbol{x}^\star$ 与对偶最优解 $(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$ 都存在。

由于对偶间隙为 0，有

$$
\begin{aligned}
f_0(\boldsymbol{x}^\star) &= g(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star) \\
&\le L(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star) \\
&= f_0(\boldsymbol{x}^\star) + \sum_{i=1}^{m} \lambda_i^\star f_i(\boldsymbol{x}^\star) \le f_0(\boldsymbol{x}^\star).
\end{aligned}
$$

所以所有不等号都必须取等号，得到

$$
\lambda_i^\star f_i(\boldsymbol{x}^\star) = 0, \quad i = 1, \ldots, m.
$$

KKT 条件从哪里来：驻点条件

上一页还说明

$$g(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star) = L(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star) = \inf_{\boldsymbol{x} \in \mathcal{D}} L(\boldsymbol{x}, \boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star).$$

若 L 对 $\boldsymbol{x}$ 可微，且 $\boldsymbol{x}$ 没有额外未写出的边界限制，则一阶必要条件为

$$\nabla_{\boldsymbol{x}} L(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star) = \mathbf{0}.$$

也就是

$$\nabla f_0(\boldsymbol{x}^\star) + \sum_{i=1}^{m} \lambda_i^\star \nabla f_i(\boldsymbol{x}^\star) + \sum_{i=1}^{p} \nu_i^\star \nabla h_i(\boldsymbol{x}^\star) = \mathbf{0}.$$

KKT 条件：检查清单

定理 (KKT 必要条件)

若 f_0, f_i, h_i 可微，没有未显式处理的定义域边界，强对偶成立，且原问题和对偶问题的最优解都达到，则任意一组原/对偶最优解满足：

$$
\begin{aligned}
&\text{原始可行：} && \boldsymbol{x}^\star \in \mathcal{D}, \quad f_i(\boldsymbol{x}^\star) \le 0, \quad h_i(\boldsymbol{x}^\star) = 0, \\
&\text{对偶可行：} && \lambda_i^\star \ge 0, \\
&\text{互补松弛：} && \lambda_i^\star f_i(\boldsymbol{x}^\star) = 0, \\
&\text{驻点条件：} && \nabla f_0(\boldsymbol{x}^\star) + \sum_{i=1}^{m} \lambda_i^\star \nabla f_i(\boldsymbol{x}^\star) + \sum_{i=1}^{p} \nu_i^\star \nabla h_i(\boldsymbol{x}^\star) = \mathbf{0}.
\end{aligned}
$$

这组条件称为**Karush-Kuhn-Tucker（KKT）条件**。

互补松弛怎么理解?

对每个不等式约束都有

$$\lambda_i^\star f_i(\boldsymbol{x}^\star) = 0, \qquad \lambda_i^\star \geq 0, \qquad f_i(\boldsymbol{x}^\star) \leq 0.$$

因此每个约束只能处在两种状态之一：

- **不紧约束：** $f_i(\boldsymbol{x}^\star) < 0$，则必须有 $\lambda_i^\star = 0$；
- **紧约束：** $f_i(\boldsymbol{x}^\star) = 0$，此时 $\lambda_i^\star$ 可以为正。

凸优化中：KKT 也是充分条件

定理 (凸 KKT 充分性)

在与 *KKT* 必要条件相同的光滑和无额外定义域边界假设下，若 f_0, f_i 为凸函数，h_i 为仿射函数，且 $\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star$ 满足 *KKT* 条件，则 $\boldsymbol{x}^\star$ 是原问题最优解，$(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$ 是对偶最优解，且对偶间隙为 0。

证明要点总结：

- 因为 $\lambda_i^\star \geq 0$ 且 f_0, f_i 凸、h_i 仿射，$L(\boldsymbol{x}, \boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$ 关于 $\boldsymbol{x}$ 是凸函数；
- 驻点条件说明 $\boldsymbol{x}^\star$ 最小化这个 Lagrange 函数；
- 原始可行和互补松弛给出 $L(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star) = f_0(\boldsymbol{x}^\star)$。

Slater + 凸性：KKT 充要条件

定理 (常用结论)

对于可微凸优化问题，若等式约束为仿射约束、没有未显式处理的定义域边界、*Slater* 条件满足、$p^\star$ 有限且原问题最优解达到，则 *KKT* 条件是最优性的充要条件。

也就是说：

- 如果 $x^\star$ 最优，则存在乘子让 KKT 成立；
- 如果找到一组变量和乘子满足 KKT，则 $x^\star$ 就是全局最优解。

例子 1：只有等式约束

例子 1： 最小范数解

$$\begin{array}{ll} \text{minimize} & \boldsymbol{x}^\top \boldsymbol{x} \\ \text{subject to} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}. \end{array}$$

KKT 条件为

$$\boldsymbol{A}\boldsymbol{x}^\star = \boldsymbol{b}, \qquad 2\boldsymbol{x}^\star + \boldsymbol{A}^\top \boldsymbol{\nu}^\star = \boldsymbol{0}.$$

可写成线性方程组

$$\begin{bmatrix} 2\boldsymbol{I} & \boldsymbol{A}^\top \\ \boldsymbol{A} & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{x}^\star \\ \boldsymbol{\nu}^\star \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{b} \end{bmatrix}.$$

例子 2：一个不等式约束

例子 2： 求解

$$\begin{array}{ll} \text{minimize} & (x-2)^2 \\ \text{subject to} & x \leq 1. \end{array}$$

写成 $f_1(x) = x - 1 \leq 0$，Lagrange 函数为

$$L(x,\lambda) = (x-2)^2 + \lambda(x-1), \qquad \lambda \geq 0.$$

KKT 条件：

$$\begin{aligned} &x^\star - 1 \leq 0, \qquad \lambda^\star \geq 0, \\ &\lambda^\star(x^\star - 1) = 0, \\ &2(x^\star - 2) + \lambda^\star = 0. \end{aligned}$$

由互补松弛：

- 若约束不紧，$x^\star < 1$，则 $\lambda^\star = 0$。驻点给出 $x^\star = 2$，但它不可行，所以排除；
- 因此约束必须紧：$x^\star = 1$。驻点给出

$$2(1-2) + \lambda^\star = 0 \quad \Rightarrow \quad \lambda^\star = 2.$$

所以最优解为

$$\boxed{x^\star = 1, \qquad \lambda^\star = 2.}$$

例子 3：不紧约束

例子 3： 求解

$$\begin{aligned} \text{minimize} \quad & (x-1)^2 \\ \text{subject to} \quad & x \le 2, \\ & -x \le 0. \end{aligned}$$

无约束最优点 $x = 1$ 已经满足 $0 \le x \le 2$，所以猜测 $x^\star = 1$。

KKT 检查：

$$\begin{aligned} & f_1(1) = 1 - 2 = -1 < 0, \qquad f_2(1) = -1 < 0, \\ & \lambda_1^\star = \lambda_2^\star = 0, \\ & 2(x^\star - 1) + \lambda_1^\star - \lambda_2^\star = 0. \end{aligned}$$

两个约束都不紧，因此乘子都为 0。

应用 4：乘子的价格含义

为了看一个不等式约束的边际价值，把

$$f_1(x) \le 0$$

放宽为

$$f_1(x) \le u.$$

记扰动后问题的最优值为 $p^\star(u)$。先只看一个不等式约束，Lagrange 函数为

$$L(x,\lambda;u) = f_0(x) + \lambda(f_1(x) - u).$$

单位理解： 若 $f_1(x) \le u$ 表示资源上限或安全裕度，则 $\lambda^\star$ 的单位是“目标值/资源单位”。例如 $\lambda^\star = 2$ 表示把约束放宽 1 个单位时，最优成本约下降 2 个单位。

需要的光滑假设： 最优解 $x^\star(u)$、乘子 $\lambda^\star(u)$ 以及最优值 $p^\star(u)$ 在 $u = 0$ 附近可微，并且活跃约束保持活跃（即约束紧）。

乘子的价格含义：一阶变化

若约束在最优点处保持紧约束，则

$$\begin{aligned}
&f_1(x^\star(u)) - u = 0, \\
&p^\star(u) = L(x^\star(u), \lambda^\star(u); u) = f_0(x^\star(u)) + \lambda^\star(u)\big(f_1(x^\star(u)) - u\big).
\end{aligned}$$

对 u 求导，利用驻点条件和紧约束条件，可得

$$\frac{dp^\star(u)}{du} = -\lambda^\star(u).$$

特别地，在 $u = 0$ 处，

$$\frac{dp^\star}{du}(0) = -\lambda^\star.$$

解释： 如果把约束放宽一个小量 $\Delta u > 0$，最优值的一阶变化约为 $-\lambda^\star \Delta u$；若约束不紧，则 $\lambda^\star = 0$，通常没有一阶收益。

课堂快速检查

练习： 对问题

$$\begin{array}{ll} \text{minimize} & x^2 + y^2 \\ \text{subject to} & x + y \geq 1 \end{array}$$

写出 KKT 条件，并判断哪个约束活跃。

提示：先写成标准形式 $f_1(x, y) = 1 - x - y \leq 0$.

课堂快速检查

练习：对问题

$$\begin{aligned} \text{minimize} \quad & x^2 + y^2 \\ \text{subject to} \quad & x + y \geq 1 \end{aligned}$$

写出 KKT 条件，并判断哪个约束活跃。

提示：先写成标准形式 $f_1(x, y) = 1 - x - y \leq 0.$

答案：约束活跃，最优点为 $(1/2, 1/2)$；KKT 乘子为 $\lambda^\star = 1$。

目录

总结：四个核心结论

条件/工具	得到什么
Slater 条件	凸优化中推出强对偶，常还能保证对偶最优解存在
KKT 必要性	强对偶且最优解达到时，最优解必须满足 KKT
凸性 + KKT	满足 KKT 就足以推出全局最优
凸性 + Slater	KKT 成为最优性的充要条件

本节需要掌握：**先检查 Slater，再写 KKT。**

课后整理

- 会解释 Slater 条件的含义，并能判断简单问题是否满足；
- 会完整写出 KKT 条件，尤其不要漏掉互补松弛和对偶可行；
- 会用“紧/不紧”两种情况求解简单不等式约束问题；
- 理解乘子作为边际价格的含义，但知道该解释依赖一定光滑性。

阅读作业 & 参考资料：

- 课本第 5.2–5.5 节。

目录

附录：为什么原问题可写成 min–max？

回顾 Lagrange 函数

$$L(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\nu}) = f_0(\boldsymbol{x}) + \sum_{i=1}^{m}\lambda_i f_i(\boldsymbol{x}) + \sum_{i=1}^{p}\nu_i h_i(\boldsymbol{x}), \qquad \boldsymbol{\lambda}\succeq \mathbf{0},\ \boldsymbol{\nu}\in\mathbb{R}^p.$$

固定一个 $\boldsymbol{x}\in\mathcal{D}$，看内层

$$\sup_{\boldsymbol{\lambda}\succeq\mathbf{0},\ \boldsymbol{\nu}} L(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\nu}).$$

- 若某个不等式约束被违反，即存在 i 使 $f_i(\boldsymbol{x})>0$，令 $\lambda_i\to+\infty$ 且其他乘子固定，则

$$L(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\nu})\to+\infty.$$

- 若某个等式约束被违反，即存在 j 使 $h_j(\boldsymbol{x}) \neq 0$，令 ν_j 沿着 $h_j(\boldsymbol{x})$ 同号方向趋于无穷，则

$$\nu_j h_j(\boldsymbol{x}) \to +\infty, \qquad L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \to +\infty.$$

因此

$$\sup_{\boldsymbol{\lambda}\succeq \mathbf{0},\ \boldsymbol{\nu}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \begin{cases} f_0(\boldsymbol{x}), & \boldsymbol{x} \text{ 原始可行}, \\ +\infty, & \boldsymbol{x} \text{ 原始不可行}. \end{cases}$$

$$\inf_{\boldsymbol{x}\in\mathcal{D}} \sup_{\boldsymbol{\lambda}\succeq \mathbf{0},\ \boldsymbol{\nu}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\boldsymbol{x} \text{ 原始可行}} f_0(\boldsymbol{x}) = p^\star.$$

凸 KKT 充分性的证明

结论见第31页

对任意可行点 $\boldsymbol{x}$，由 $\lambda_i^\star \geq 0$ 和 $f_i(\boldsymbol{x}) \leq 0$ 可知

$$L(\boldsymbol{x}, \boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star) = f_0(\boldsymbol{x}) + \sum_i \lambda_i^\star f_i(\boldsymbol{x}) + \sum_i \nu_i^\star h_i(\boldsymbol{x}) \leq f_0(\boldsymbol{x}).$$

另一方面，由于 $\boldsymbol{x}^\star$ 满足 KKT 条件，因此 $\boldsymbol{x}^\star$ 最小化 $L(\cdot, \boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$，所以

$$f_0(\boldsymbol{x}^\star) = L(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star) \leq L(\boldsymbol{x}, \boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star) \leq f_0(\boldsymbol{x}).$$

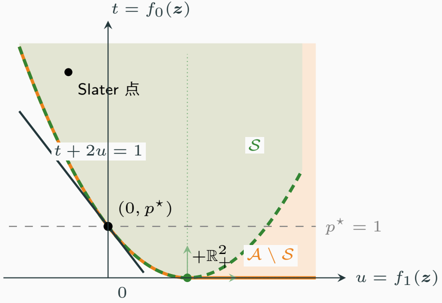
因此 $\boldsymbol{x}^\star$ 优于任意可行点，是全局最优解。

目录

附录：Slater 证明的几何图像

考虑简化问题：

$$\begin{array}{ll} p^\star = \text{minimize} & f_0(\boldsymbol{z}) \\ \text{subject to} & f_1(\boldsymbol{z}) \le 0. \end{array}$$



图中取具体例子 $\boldsymbol{z} = (x, y)$，

$$f_1(\boldsymbol{z}) = x, \qquad f_0(\boldsymbol{z}) = (x-1)^2 + y^2.$$

于是原始图像集合为

$$\mathcal{S} = \{(f_1(\boldsymbol{z}), f_0(\boldsymbol{z}))\} = \{(u, t) \mid t \ge (u-1)^2\}.$$

右上平移后得到

$$\mathcal{A} = \mathcal{S} + \mathbb{R}_+^2 = \{(u, t) \mid t \ge \phi(u)\},$$

$$\phi(u) = \begin{cases} (u-1)^2, & u \le 1, \\ 0, & u \ge 1. \end{cases}$$

可证明这样的集合 $\mathcal{A}$ 是凸集。

附录：Slater 证明草图

由 $p^\star$ 的定义，$(0,p^\star)$ 是 $\overline{\mathcal{A}}$ 的边界点。由支撑超平面定理，存在非零向量 (α,β)，使

$$\alpha u + \beta t \geq \beta p^\star, \qquad \forall (u,t) \in \overline{\mathcal{A}}.$$

因为 $\mathcal{A}$ 向右上方延伸，必须有 $\alpha,\beta \geq 0$。若 $\beta = 0$，则上式推出所有 $(u,t) \in \mathcal{A}$ 都有 $u \geq 0$；但 Slater 点给出 $f_1(\bar{z}) < 0$，矛盾。所以 $\beta > 0$。

令 $\lambda^\star = \alpha/\beta \geq 0$。对任意 $\boldsymbol{z}$，取 $(u,t) = (f_1(\boldsymbol{z}), f_0(\boldsymbol{z})) \in \mathcal{A}$，得到

$$f_0(\boldsymbol{z}) + \lambda^\star f_1(\boldsymbol{z}) \geq p^\star.$$

$$g(\lambda^\star) = \inf_{\boldsymbol{z}\in\mathcal{D}} \left\{ f_0(\boldsymbol{z}) + \lambda^\star f_1(\boldsymbol{z}) \right\} \geq p^\star.$$

再由弱对偶 $g(\lambda^\star) \leq p^\star$，可得 $g(\lambda^\star) = p^\star$。因此强对偶成立，且 $\lambda^\star$ 就是图中支撑线对应的 Lagrange 乘子。