最优性理论

宋晓良

大连理工大学数学科学学院

教材《最优化：建模、算法与理论》配套电子教案

http://bicmr.pku.edu.cn/~wenzw/optbook.html

提纲

最优化问题解的存在性

考虑优化问题

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & f(x), \\ \text{s.t.} \quad & x \in \mathcal{X}, \end{aligned}$$

其中$\mathcal{X} \subseteq \mathbb{R}^n$ 为可行域.

- 首先要考虑的是最优解的存在性, 然后考虑如何求出其最优解.
- 在数学分析课程中, 我们学习过Weierstrass 定理, 即定义在紧集上的连续函数一定存在最大(最小) 值点.
- 而在许多实际问题中, 定义域可能不是紧的, 目标函数也不一定连续, 因此需要将此定理推广来保证最优化问题解的存在性.

最优化问题解的存在性: 推广的Weierstrass 定理

定理 (推广的Weierstrass定理)

若函数$f:\mathcal{X}\to(-\infty,+\infty]$ 适当且闭, 且以下条件中任意一个成立:

1. $\mathrm{dom}f \stackrel{\text{def}}{=\!=} \{x\in\mathcal{X}\ :f(x)<+\infty\}$ 是有界的;
2. 存在一个常数$\bar{\gamma}$ 使得下水平集
$$C_{\bar{\gamma}} \stackrel{\text{def}}{=\!=} \{x\in\mathcal{X}\ :f(x)\le\bar{\gamma}\}$$
是非空且有界的;
3. f 是强制的, 即对于任一满足$\|x^k\|\to+\infty$ 的点列$\{x^k\}\subset\mathcal{X}$, 都有
$$\lim_{k\to\infty} f(x^k)=+\infty,$$

则函数f 的最小值点集$\{x\in\mathcal{X}\mid f(x)\le f(y),\ \forall y\in\mathcal{X}\}$ 非空且紧.

条件(2) 下的证明

假设条件(2) 成立，且$t \overset{\text{def}}{=\!=} \inf_{x\in\mathcal{X}} f(x) = -\infty$.

- 由下确界的定义，存在点列$\{x^k\}_{k=1}^{\infty} \subset C_{\bar{\gamma}}$，使得$\lim\limits_{k\to\infty} f(x^k) = -\infty$.
- 由$C_{\bar{\gamma}}$ 有界知点列$\{x^k\}$ 存在聚点x^*.
- 由f 是闭函数知$\mathbf{epi}f$ 为闭集，因此$(x^*, t) \in \mathbf{epi}f$. 根据上方图的定义知$f(x^*) \le t = -\infty$，这与$f$ 适当矛盾，故$t \overset{\text{def}}{=\!=} \inf_{x\in\mathcal{X}} f(x)$ 有限.
- 由f 为闭函数知$C_{\bar{\gamma}}$ 为闭集，由假设知$C_{\bar{\gamma}}$ 有界，故为紧集.
- 由$f(x^*) = t$ 知最小值点集非空，且为紧集$C_{\bar{\gamma}}$ 的子集，而紧集的子集也是紧集，故最小值点集为非空紧集.

条件(1)(3)下的证明

假设条件(1)成立，则$\mathrm{dom} f$ 是有界的.

- 由f 适当知存在$x_0 \in \mathcal{X}$ 使得$f(x_0) < +\infty$.
- 记$\bar{\gamma} = f(x_0)$，此时f 的$\bar{\gamma}$ 下水平集$C_{\bar{\gamma}}$ 非空有界，条件(2)成立.

假设条件(3)成立，我们证明条件(2)成立.

- 用反证法，假设存在一个无界的下水平集$C_{\bar{\gamma}}$，那么可以取点列$\{x^k\} \subset C_{\bar{\gamma}}$ 使得 $\lim\limits_{k\to\infty} \|x^k\| = +\infty$, $\lim\limits_{k\to\infty} f(x^k) = +\infty$，这与$f(x^k) \le \bar{\gamma}$ 矛盾，故此时f 的任意下水平集都有界，条件(2)成立.

推广的Weierstrass定理:注记

- 推广的Weierstrass定理的三个条件在本质上都是保证$f(x)$ 的最小值不能在无穷远处取到.
- 因此我们可以仅在一个有界的下水平集中考虑$f(x)$ 的最小值.
- 定理仅要求$f(x)$ 为适当且闭的函数, 并不需要$f(x)$ 的连续性, 因此比数学分析中的Weierstrass定理应用范围更广.
- 当定义域不是有界闭集时, 对于强制函数$f(x) = x^2$, $x \in \mathcal{X} = \mathbb{R}$, 其全局最优解一定存在.
- 对于适当且闭的函数$f(x) = e^{-x}$, $x \in \mathcal{X} = \mathbb{R}$, 它不满足三个条件中任意一个, 因此我们不能断言其全局极小值点存在. 事实上, 其全局极小值点不存在.

解的存在唯一性

- 推广的Weierstrass定理给出了最优解的存在性条件, 但其对应的解可能不止一个.
- 当最优解是唯一存在时, 我们可以通过比较不同算法收敛至该解的速度来判断算法好坏.
- 但是如果问题有多个最优值点, 不同的算法收敛到的最优值点可能不同, 那么这些算法收敛速度的比较就失去了参考价值.
- 因此, 最优化问题解的唯一性在理论分析和算法比较中扮演着重要角色.

解的存在唯一性: 拟强凸函数

定义 (拟强凸函数)

给定凸集$\mathcal{X}$ 和函数$f:\mathcal{X}\to(-\infty,+\infty]$.若任取$x\neq y$ 和$\lambda\in(0,1)$, 有

$$f(\lambda x+(1-\lambda)y)<\max\{f(x),f(y)\},$$

那么我们称函数f 是强拟凸的.

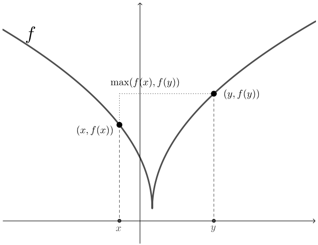

Figure: 一个强拟凸函数

拟强凸函数

- 强拟凸函数不一定是凸函数，但其任意一个下水平集都是凸集，并可以包含一部分性质较好的非凸函数.
- 任意强凸函数均为强拟凸的，但凸函数并不一定是强拟凸的.
- 任何定义在闭有界凸集上的强凸函数(如$f(x)=x^2$), 其最优解都是唯一存在的.
- 对于一般的凸函数，其最优解可能不唯一，比如$f(x)=\max\{x,0\}$，任意$x\leq 0$ 都是$f(x)$ 的最优解.

唯一性定理

定理 (唯一性定理)

设$\mathcal{X}$ 是$\mathbb{R}^n$ 的一个非空, 紧且凸的子集, 如果$f : \mathcal{X} \to (-\infty, +\infty]$ 是适当, 闭且强拟凸函数, 那么存在唯一的x^* 满足

$$f(x^*) < f(x), \quad \forall x \in \mathcal{X} \backslash \{x^*\}.$$

证明:由Weierstrass定理知f 至少存在一个全局极小点x^*.若x^*, y^* 皆为全局极小点, 则有:

$$f(\lambda x^* + (1-\lambda)y^*) < \max\{f(x^*), f(y^*)\} = f(x^*), \quad \forall \lambda \in (0,1).$$

这与x^* 的全局极小性矛盾.

提纲

无约束可微问题的最优性理论:引言

无约束可微优化问题通常表示为如下形式：

$$\min_{x\in\mathbb{R}^n} \quad f(x),$$

其中f 是连续可微函数.

- 给定一个点$\bar{x}$, 我们想要知道这个点是否是函数f 的一个局部极小解或者全局极小解.
- 如果从定义出发, 需要对其邻域内的所有点进行判断, 这不可行.
- 因此, 需要一个更简单的方式来验证一个点是否为极小值点.我们称其为最优性条件, 它主要包含一阶最优性条件和二阶最优性条件.

一阶必要条件:下降方向

定义 (下降方向)

对于可微函数f 和点$x \in \mathbb{R}^n$, 如果存在向量d 满足

$$\nabla f(x)^{\mathrm{T}} d < 0,$$

那么称d 为f 在点x 处的一个下降方向.

- 一阶最优性条件是利用梯度(一阶)信息来判断给定点的最优性.
- 由下降方向的定义, 容易验证: 如果f 在点x 处存在一个下降方向d, 那么对于任意的$T > 0$, 存在$t \in (0, T]$, 使得

$$f(x + td) < f(x).$$

因此, 在局部最优点处不能有下降方向.

一阶必要条件

定理 (一阶必要条件)

假设f 在全空间$\mathbb{R}^n$ 可微.如果x^* 是一个局部极小点, 那么

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

证明:任取$v \in \mathbb{R}^n$, 考虑f 在点$x = x^*$ 处的泰勒展开

$$f(x^* + tv) = f(x^*) + tv^{\mathrm{T}}\nabla f(x^*) + o(t),$$

整理得

$$\frac{f(x^* + tv) - f(x^*)}{t} = v^{\mathrm{T}}\nabla f(x^*) + o(1).$$

根据x^* 的最优性, 在上式中分别对t 取点0 处的左, 右极限可知

$$\lim_{t\to 0^+}\frac{f(x^* + tv) - f(x^*)}{t} = v^{\mathrm{T}}\nabla f(x^*) \geq 0,$$
$$\lim_{t\to 0^-}\frac{f(x^* + tv) - f(x^*)}{t} = v^{\mathrm{T}}\nabla f(x^*) \leq 0,$$

即对任意的v 有$v^{\mathrm{T}}\nabla f(x^*) = 0$, 由$v$ 的任意性知$\nabla f(x^*) = 0$.

二阶最优性条件

- 在没有额外假设时，如果一阶必要条件满足，我们仍然不能确定当前点是否是一个局部极小点.
- 假设f 在点x 的一个开邻域内是二阶连续可微的.类似于一阶必要条件的推导，可以借助当前点处的二阶泰勒展开来逼近该函数在该点附近的取值情况，从而来判断最优性.
- 在点x 附近我们考虑泰勒展开

$$f(x+d) = f(x) + \nabla f(x)^{\mathrm{T}} d + \frac{1}{2} d^{\mathrm{T}} \nabla^2 f(x) d + o(\|d\|^2).$$

- 当一阶必要条件满足时，$\nabla f(x) = 0$，那么上面的展开式简化为

$$f(x+d) = f(x) + \frac{1}{2} d^{\mathrm{T}} \nabla^2 f(x) d + o(\|d\|^2).$$

由此，我们可以导出二阶最优性条件.

二阶最优性条件

定理 (二阶最优性条件)

必要条件:若x^* 是f 的一个局部极小点, 则$\nabla f(x^*) = 0, \nabla^2 f(x^*) \succeq 0$.

充分条件:若$\nabla f(x^*) = 0, \nabla^2 f(x^*) \succ 0$, 则$x^*$ 是f 的一个局部极小点.

- 必要性:若$\nabla^2 f(x^*)$ 有负的特征值$\lambda_- < 0$, 设$\nabla^2 f(x^*)d = \lambda_- d$, 则

$$\frac{f(x^*+d)-f(x^*)}{\|d\|^2} = \frac{1}{2}\frac{d^{\mathrm{T}}}{\|d\|}\nabla^2 f(x^*)\frac{d}{\|d\|} + o(1) = \frac{1}{2}\lambda_- + o(1).$$

当$\|d\|$ 充分小时, $f(x^*+d) < f(x^*)$, 这和点x^* 的最优性矛盾.

- 充分性:由$\nabla f(x^*) = 0$ 时的二阶展开,

$$\frac{f(x^*+d)-f(x^*)}{\|d\|^2} = \frac{\frac{1}{2}d^{\mathrm{T}}\nabla^2 f(x^*)d + o(\|d\|^2)}{\|d\|^2} \geq \frac{1}{2}\lambda_{\min} + o(1).$$

当$\|d\|$ 充分小时有$f(x^*+d) \geq f(x^*)$, 即二阶充分条件成立.

二阶最优性条件:注记

- 设点$\bar{x}$ 满足一阶最优性条件(即$\nabla f(\bar{x}) = 0$), 且该点处的海瑟矩阵$\nabla^2 f(\bar{x})$ 不是半正定的, 那么$\bar{x}$ 不是一个局部极小点.
- 进一步地, 如果海瑟矩阵$\nabla^2 f(\bar{x})$ 既有正特征值又有负特征值, 我们称稳定点$\bar{x}$ 为一个鞍点.
- 事实上, 记d_1, d_2 为其正负特征值对应的特征向量, 那么对于任意充分小的$t > 0$, 我们都有$f(\bar{x} + td_1) > f(\bar{x})$ 且$f(\bar{x} + td_2) < f(\bar{x})$.
- 注意, 二阶最优性条件给出的仍然是关于局部最优性的判断.对于给定点的全局最优性判断, 我们还需要借助实际问题的性质, 比如目标函数是凸的、非线性最小二乘问题中目标函数值为0 等.

- 线性最小二乘:

$$\min_{x\in\mathbb{R}^n} \quad f(x) \stackrel{\text{def}}{=\!=} \frac{1}{2}\|b-Ax\|_2^2,$$

  其中$A\in\mathbb{R}^{m\times n}$, $b\in\mathbb{R}^m$.由f 可微且凸知

$$x^* \text{ 为一个全局最优解} \Leftrightarrow \nabla f(x^*) = A^{\mathrm{T}}(Ax^*-b)=0.$$

- 非线性最小二乘(实数情形的相位恢复):

$$\min_{x\in\mathbb{R}^n} \quad f(x) \stackrel{\text{def}}{=\!=} \sum_{i=1}^m r_i^2(x),$$

  其中$r_i(x)=(a_i^{\mathrm{T}}x)^2-b_i^2, i=1,2,\cdots,m$.

实数情形的相位恢复

$$\nabla f(x) = 2\sum_{i=1}^{m} r_i(x)\nabla r_i(x) = 4\sum_{i=1}^{m}((a_i^\mathrm{T} x)^2 - b_i^2)(a_i^\mathrm{T} x)a_i,$$

$$\nabla^2 f(x) = \sum_{i=1}^{m}(12(a_i^\mathrm{T} x)^2 - 4b_i^2)a_i a_i^\mathrm{T}.$$

- 如果x^* 为局部最优解, 那么其满足一、二阶必要条件

$$\sum_{i=1}^{m}((a_i^\mathrm{T} x^*)^2 - b_i)(a_i^\mathrm{T} x^*)a_i = 0, \quad \sum_{i=1}^{m}(12(a_i^\mathrm{T} x^*)^2 - 4b_i^2)a_i a_i^\mathrm{T} \succeq 0.$$

- 如果一个点$x^\#$ 满足二阶充分条件

$$\sum_{i=1}^{m}((a_i^\mathrm{T} x^\#)^2 - b_i)(a_i^\mathrm{T} x^\#)a_i = 0, \quad \sum_{i=1}^{m}(12(a_i^\mathrm{T} x^\#)^2 - 4b_i^2)a_i a_i^\mathrm{T} \succ 0,$$

那么$x^\#$ 为局部最优解.

提纲

对偶理论: 一般的约束优化问题

$$
\begin{aligned}
\min_{x\in\mathbb{R}^n} \quad & f(x), \\
\text{s.t.} \quad & c_i(x) \le 0, \ i \in \mathcal{I}, \\
& c_i(x) = 0, \ i \in \mathcal{E},
\end{aligned}
$$

其中c_i 为定义在$\mathbb{R}^n$ 或其子集上的实值函数, $\mathcal{I}$ 和$\mathcal{E}$ 分别表示不等式约束和等式约束对应的下标集合且各下标互不相同.

- 这个问题的可行域定义为

$$\mathcal{X} = \{x \in \mathbb{R}^n \mid c_i(x) \le 0, \ i \in \mathcal{I} \text{ 且 } c_i(x) = 0, \ i \in \mathcal{E}\}.$$

- 通过将$\mathcal{X}$ 的示性函数加到目标函数中可以得到无约束优化问题.但是转化后问题的目标函数是不连续的、不可微的以及不是有限的,这导致我们难以分析其理论性质以及设计有效的算法.

拉格朗日函数

一般的约束优化问题:

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & c_i(x) \le 0,\ i \in \mathcal{I},\ |\mathcal{I}| = m \\ & c_i(x) = 0,\ i \in \mathcal{E},\ |\mathcal{E}| = p \end{aligned}$$

变量$x \in \mathbb{R}^n$, 最优值为p^*, 定义域为

$$\mathcal{X} = \{x \in \mathbb{R}^n \mid c_i(x) \le 0,\ i \in \mathcal{I} \text{ 且 } c_i(x) = 0,\ i \in \mathcal{E}\}$$

拉格朗日函数$L : \mathbb{R}^n \times \mathbb{R}^m_+ \times \mathbb{R}^p \ \to \ \mathbb{R}$

$$L(x, \lambda, \nu) = f(x) + \sum_{i \in \mathcal{I}} \lambda_i c_i(x) + \sum_{i \in \mathcal{E}} \nu_i c_i(x)$$

- λ_i 为第i 个不等式约束对应的拉格朗日乘子
- ν_i 为第i 个等式约束对应的拉格朗日乘子

拉格朗日对偶函数

拉格朗日对偶函数$g : \mathbb{R}_+^m \times \mathbb{R}^p \to [-\infty, +\infty)$

$$\begin{aligned} g(\lambda, \nu) &= \inf_{x \in \mathbb{R}^n} L(x, \lambda, \nu) \\ &= \inf_{x \in \mathbb{R}^n} \left(f(x) + \sum_{i \in \mathcal{I}} \lambda_i c_i(x) + \sum_{i \in \mathcal{E}} \nu_i c_i(x) \right) \end{aligned}$$

定理 (弱对偶原理)

若$\lambda \geq 0$, 则$g(\lambda, \nu) \leq p^*$.

证明:若$\tilde{x} \in \mathcal{X}$, 则

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu) \leq L(\tilde{x}, \lambda, \nu) \leq f(\tilde{x}),$$

对$\tilde{x}$取下界得

$$g(\lambda, \nu) \leq \inf_{\tilde{x} \in \mathcal{X}} f(\tilde{x}) = p^*.$$

拉格朗日对偶问题

拉格朗日对偶问题:

$$\max_{\lambda \geq 0, \nu} g(\lambda, \nu) = \max_{\lambda \geq 0, \nu} \inf_{x \in \mathbb{R}^n} L(x, \lambda, \nu)$$

- 称λ 和ν 为对偶变量, 设最优值为q^*
- q^* 为p^* 的最优下界, 称$p^* - q^*$ 为对偶间隙
- 拉格朗日对偶问题是一个凸优化问题
- $\mathrm{dom} g = \{(\lambda, \nu) \mid \lambda \geq 0, g(\lambda, \nu) > -\infty\}$, 称其元素为对偶可行解

例: 标准形式线性规划及其对偶

$$\begin{array}{llcll}
\min & c^T x & \qquad & \max & -b^T \nu \\
\text{s.t.} & Ax = b & & \text{s.t.} & A^T \nu + c \geq 0 \\
 & x \geq 0 & & &
\end{array}$$

实例:线性方程组具有最小模的解

$$\begin{aligned} \min \quad & x^T x \\ \text{s.t.} \quad & Ax = b \end{aligned}$$

对偶函数

- 拉格朗日函数为$L(x,\nu) = x^T x + \nu^T(Ax - b)$
- 求L 关于x 的最小值，由一阶条件:

$$\nabla_x L(x,\nu) = 2x + A^T\nu = 0 \quad \Longrightarrow \quad x = -(1/2)A^T\nu$$

- 将上式代入L 得到对偶函数g:

$$g(\nu) = L((-1/2)A^T\nu, \nu) = -\frac{1}{4}\nu^T AA^T\nu - b^T\nu$$

它是关于ν 的凹函数

弱对偶性: $p^* \geq -(1/4)\nu^T AA^T\nu - b^T\nu, \ \forall\nu$

实例:标准形式线性规划

$$\begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b, \quad x \geq 0 \end{array}$$

对偶函数

- 拉格朗日函数为

$$\begin{aligned} L(x,\lambda,\nu) &= c^T x + \nu^T(Ax - b) - \lambda^T x \\ &= -b^T\nu + (c + A^T\nu - \lambda)^T x \end{aligned}$$

- L 关于x 线性, 因此

$$g(\lambda,\nu) = \inf_x L(x,\lambda,\nu) = \begin{cases} -b^T\nu & A^T\nu - \lambda + c = 0 \\ -\infty & \text{otherwise} \end{cases}$$

g 在仿射集$\{(\lambda,\nu) \mid A^T\nu - \lambda + c = 0\}$ 上线性, 因此是凹函数

弱对偶性: $p^* \geq -b^T\nu$ 若$A^T\nu + c \geq 0$

$$\begin{aligned}\min_{x} \quad & c^{\mathrm{T}}x \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0\end{aligned}$$

- 拉格朗日函数:

$$L(x,s,\nu) = c^{\mathrm{T}}x + \nu^{\mathrm{T}}(Ax-b) - s^{\mathrm{T}}x = -b^{\mathrm{T}}\nu + (A^{\mathrm{T}}\nu - s + c)^{\mathrm{T}}x$$

- 对偶函数:

$$g(s,\nu) = \inf_{x} L(x,s,\nu) = \begin{cases} -b^{\mathrm{T}}\nu, & A^{\mathrm{T}}\nu - s + c = 0 \\ -\infty, & \text{其他} \end{cases}$$

- 对偶问题:

$$\begin{aligned}\max_{s,\nu} \quad & -b^{\mathrm{T}}\nu, \\ \text{s.t.} \quad & A^{\mathrm{T}}\nu - s + c = 0, \\ & s \geq 0.\end{aligned} \quad \overset{y=-\nu}{\Longleftrightarrow} \quad \begin{aligned}\max_{s,y} \quad & b^{\mathrm{T}}y, \\ \text{s.t.} \quad & A^{\mathrm{T}}y + s = c, \\ & s \geq 0.\end{aligned}$$

线性规划问题的对偶

- 若保留约束$x \geq 0$, 则拉格朗日函数为

$$L(x,y) = c^{\mathrm{T}}x - y^{\mathrm{T}}(Ax-b) = b^{\mathrm{T}}y + (c - A^{\mathrm{T}}y)^{\mathrm{T}}x.$$

- 对偶问题需要将$x \geq 0$ 添加到约束里：

$$\max_{y} \left\{ \inf_{x} \; b^{\mathrm{T}}y + (c - A^{\mathrm{T}}y)^{\mathrm{T}}x, \quad \text{s.t. } \; x \geq 0 \right\} \Rightarrow \begin{array}{ll} \max\limits_{y} & b^{\mathrm{T}}y, \\ \text{s.t.} & A^{\mathrm{T}}y \leq c. \end{array}$$

此对偶问题可以通过将上页最后一个问题中的变量s 消去得到.

- 视$\max \; b^{\mathrm{T}}y$ 为$\min \; -b^{\mathrm{T}}y$, 对偶问题的拉格朗日函数为

$$L(y,x) = -b^{\mathrm{T}}y + x^{\mathrm{T}}(A^{\mathrm{T}}y - c) = -c^{\mathrm{T}}x + (Ax-b)^{\mathrm{T}}y.$$

线性规划问题的对偶

- 因此得到对偶函数

$$g(x)=\inf_y L(y,x)=\begin{cases}-c^{\mathrm{T}}x, & Ax=b,\\ -\infty, & \text{其他}.\end{cases}$$

- 相应的对偶问题是

$$\begin{aligned}\max_x \quad & -c^{\mathrm{T}}x,\\ \text{s.t.} \quad & Ax=b,\\ & x\geq 0.\end{aligned}$$

该问题与原始问题完全等价，这表明线性规划问题与其对偶问题互为对偶.

实例:等式约束下的范数最小化

$$\begin{array}{ll} \min & \|x\| \\ \text{s.t.} & Ax = b \end{array}$$

对偶函数

$$g(\nu) = \inf_x (\|x\| - \nu^T Ax + b^T \nu) = \begin{cases} b^T \nu & \|A^T \nu\|_* \leq 1 \\ -\infty & \text{其他} \end{cases}$$

其中 $\|v\|_* = \sup_{\|u\| \leq 1} u^T v$ 是 $\|\cdot\|$ 的对偶范数

证明: 利用 $\inf_x(\|x\| - y^T x)$ 在 $\|y\|_* \leq 1$ 时等于0 否则等于 $-\infty$

- 若 $\|y\|_* \leq 1$, 则 $x - y^T x \geq 0$ 对任意 x 都成立, 当 $x = 0$ 时取等
- 若 $\|y\|_* > 1$, 取 $x = tu$, 其中 $\|u\| \leq 1$, $u^T y = \|y\|_* > 1$:

$$\|x\| - y^T x = t(\|u\| - \|y\|_*) \to -\infty \text{ 当 } t \to \infty$$

弱对偶性: $p^* \geq b^T \nu$ 若 $\|A^T \nu\|_* \leq 1$

拉格朗日对偶与共轭函数

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & Ax \le b, \quad Cx = d \end{aligned}$$

对偶函数

$$\begin{aligned} g(\lambda, \nu) &= \inf_{x \in \mathbf{dom}\, f_0} \left(f_0(x) + (A^T\lambda + C^T\nu)^T x - b^T\lambda - d^T\nu \right) \\ &= -f_0^*(-A^T\lambda - C^T\nu) - b^T\lambda - d^T\nu \end{aligned}$$

- 回顾共轭函数的定义$f^*(y) = \sup_{x \in \mathbf{dom}\, f}(y^Tx - f(x))$
- 在f_0 的共轭函数已知时可以简化对偶函数的推导

例：最大化熵

$$f_0(x) = \sum_{i=1}^n x_i \log x_i, \qquad f_0^*(y) = \sum_{i=1}^n e^{y_i - 1}$$

弱对偶性: $d^* \leq p^*$

- 对凸问题与非凸问题都成立
- 可导出复杂问题的非平凡下界,

强对偶性: $d^* = p^*$

- 对一般问题而言通常不成立
- (通常) 对凸问题成立
- 称保证凸问题强对偶性成立的条件为约束品性

问题形式与对偶性

- 一个问题不同等价形式的对偶可能差异巨大
- 当对偶问题难以推导或没有价值时，可以尝试改写原问题的形式

常用的改写技巧

- 引入新变量与等式约束
- 将显式约束隐式化或将隐式约束显式化
- 改变目标函数或者约束函数的形式
  例如，用 $\phi(f_0(x))$ 取代$f_0(x)$，其中 ϕ 是凸的增函数

引入新变量与等式约束

$$\min \quad f_0(Ax+b)$$

- 对偶函数为常数
- 强对偶性成立，但对偶问题无意义

改写原问题及其对偶

$$\begin{array}{ll} \min & f_0(y) \\ \text{s.t.} & Ax+b-y=0 \end{array} \qquad \begin{array}{ll} \max & b^T\nu - f_0^*(\nu) \\ \text{s.t.} & A^T\nu = 0 \end{array}$$

对偶函数为

$$\begin{aligned} g(\nu) &= \inf_{x,y}(f_0(y) - \nu^T y + \nu^T Ax + b^T\nu) \\ &= \begin{cases} -f_0^*(\nu) + b^T\nu & A^T\nu = 0 \\ -\infty & \text{其他} \end{cases} \end{aligned}$$

范数逼近问题: $\min \|Ax - b\|$

$$\begin{aligned} \min \quad & \|y\| \\ \text{s.t.} \quad & y = Ax - b \end{aligned}$$

由 $\|\cdot\|$ 的共轭函数知其对偶函数为:

$$\begin{aligned} g(\nu) &= \inf_{x,y} (\|y\| + \nu^T y - \nu^T Ax + b^T \nu) \\ &= \begin{cases} b^T \nu + \inf_y (\|y\| + \nu^T y) & A^T \nu = 0 \\ -\infty & \text{其他} \end{cases} \\ &= \begin{cases} b^T \nu & A^T \nu = 0, \quad \|\nu\|_* \le 1 \\ -\infty & \text{其他} \end{cases} \end{aligned}$$

范数逼近问题的对偶

$$\begin{aligned} \max \quad & b^T \nu \\ \text{s.t.} \quad & A^T \nu = 0, \quad \|\nu\|_* \le 1 \end{aligned}$$

隐式约束

带边界约束的线性规划: 原问题与对偶问题

$$\begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b \\ & -\mathbf{1} \le x \le \mathbf{1} \end{array} \qquad \begin{array}{ll} \max & -b^T\nu - \mathbf{1}^T\lambda_1 - \mathbf{1}^T\lambda_2 \\ \text{s.t.} & c + A^T\nu + \lambda_1 - \lambda_2 = 0 \\ & \lambda_1 \ge 0, \quad \lambda_2 \ge 0 \end{array}$$

通过隐式化边界约束改写原问题

$$\begin{array}{ll} \min & f_0(x) = \begin{cases} c^T x & -\mathbf{1} \le x \le \mathbf{1} \\ -\infty & \text{其他} \end{cases} \\ \text{s.t.} & Ax = b \end{array}$$

对偶函数为

$$\begin{aligned} g(\nu) &= \inf_{-\mathbf{1} \le x \le \mathbf{1}} (c^T x + \nu^T(Ax - b)) \\ &= -b^T\nu - \|A^T\nu + c\|_1 \end{aligned}$$

对偶问题: $\max \ -b^T\nu - \|A^T\nu + c\|_1$

ℓ_1 正则化问题的对偶

$$\min_{x\in\mathbb{R}^n} \quad \frac{1}{2}\|Ax-b\|^2+\mu\|x\|_1$$

令$r=Ax-b$, 问题等价于$\min_{x\in\mathbb{R}^n}\frac{1}{2}\|r\|^2+\mu\|x\|_1$, s.t. $r=Ax-b$

- 拉格朗日函数:

$$\begin{aligned} L(x,r,\lambda) &= \frac{1}{2}\|r\|^2+\mu\|x\|_1-\langle\lambda, Ax-b-r\rangle \\ &= \frac{1}{2}\|r\|^2+\lambda^{\mathrm{T}}r+\mu\|x\|_1-(A^{\mathrm{T}}\lambda)^{\mathrm{T}}x+b^{\mathrm{T}}\lambda \end{aligned}$$

- 对偶函数:

$$g(\lambda)=\inf_{x,r}\ L(x,r,\lambda)=\begin{cases} b^{\mathrm{T}}\lambda-\frac{1}{2}\|\lambda\|^2, & \|A^{\mathrm{T}}\lambda\|_\infty\le\mu \\ -\infty, & \text{其他} \end{cases}$$

- 对偶问题:

$$\max \quad b^{\mathrm{T}}\lambda-\frac{1}{2}\|\lambda\|^2, \quad \text{s.t.} \quad \|A^{\mathrm{T}}\lambda\|_\infty\le\mu$$

提纲

带约束凸优化问题

- 稀疏优化问题:

$$\min_{x\in\mathbb{R}^n} \quad \|x\|_1, s.t. \quad Ax = b.$$

- 低秩矩阵恢复问题:

$$\min_{X\in\mathbb{R}^{m\times n}} \|X\|_*, \text{ s.t. } X_{ij} = M_{ij}, (i,j) \in \Omega.$$

- 矩阵分离问题:

$$\min_{X,S\in\mathbb{R}^{m\times n}} \|X\|_* + \mu\|S\|_1, \text{ s.t.} \quad X + S = M.$$

- 回归分析中的问题:

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|Ax - b\|_2^2, \quad \text{s.t.} \quad \|x\|_1 \leqslant \sigma.$$

带约束凸优化问题

- 前述问题都可以写为

$$
\begin{aligned}
\min_{x\in\mathcal{D}} \quad & f(x), \\
\text{s.t.} \quad & c_i(x) \leqslant 0, \quad i = 1, 2, \cdots, m, \\
& Ax = b,
\end{aligned}
$$

  其中$f(x)$ 为适当的凸函数, $\forall i, c_i(x)$ 是凸函数且$\mathrm{dom} c_i = \mathbb{R}^n$.

- $A \in \mathbb{R}^{p\times n}, b \in \mathbb{R}^p$ 是已知的.
- 集合$\mathcal{D}$ 表示自变量x 的自然定义域, 即

$$
\mathcal{D} = \mathrm{dom} f = \{x \mid f(x) < +\infty\}.
$$

- 自变量x 还受约束的限制, 定义可行域

$$
\mathcal{X} = \{x \in \mathcal{D} \ : \ c_i(x) \leq 0, \ i = 1, 2, \cdots, m; \ Ax = b\}.
$$

- 由于凸优化问题的可行域是凸集, 因此等式约束只能是线性约束.

Slater约束品性与强对偶原理

- 凸优化问题有很多好的性质.一个自然的问题是:我们能否像研究无约束问题那样找到该问题最优解的一阶充要条件?如果这样的条件存在, 它在什么样的约束品性下成立?
- 在通常情况下, 优化问题的对偶间隙大于0, 即强对偶原理不满足.
- 但对很多凸优化问题, 在特定约束品性下可以证明强对偶原理.
- 直观的一种约束品性是存在满足所有约束条件的严格可行解.

首先给出集合$\mathcal{D}$的相对内点集relint$\mathcal{D}$的定义.给定集合$\mathcal{D}$, 记其仿射包为

$$\textbf{affine}\,\mathcal{D} = \{x \mid x = \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k,\ x_1, x_2, \cdots, x_k \in \mathcal{D},\ \sum_{i=1}^{k} \theta_i = 1\}.$$

定义 (相对内点)

集合$\mathcal{D}$的相对内点集定义为

$$\mathrm{relint}\mathcal{D} = \{x \in \mathcal{D} \mid \exists\, r > 0,\ \text{使得}\ B(x, r) \cap \textbf{affine}\,\mathcal{D} \subseteq \mathcal{D}\}.$$

相对内点是内点的推广，若$\mathcal{D}$本身的“维数”较低, 则$\mathcal{D}$不可能有内点, 但如果在它的仿射包**affine**$\,\mathcal{D}$中考虑, 则$\mathcal{D}$可能有相对内点.

Slater约束品性

定义 (Slater约束品性)

若对凸优化问题

$$\min_{x\in\mathcal{D}} \quad f(x), \ \text{s.t.}\ c_i(x) \leqslant 0, \ i=1,2,\cdots,m, \quad Ax=b,$$

存在$x \in \operatorname{relint}\mathcal{D}$ 满足

$$c_i(x) < 0, \quad i=1,2,\cdots,m, \quad Ax=b,$$

则称对此问题Slater 约束品性满足.该约束品性也称为 Slater 条件.

- Slater约束品性实际上是要求自然定义域$\mathcal{D}$ 的相对内点中存在使得不等式约束严格成立的点, **affine** $\mathcal{D}=\mathbb{R}^n$ 时相对内点就是内点.
- 不等式约束是仿射函数时, Slater条件可以放宽.设前k 个不等式约束是仿射的, 此时Slater约束品性变为:存在$x \in \operatorname{relint}\mathcal{D}$, 满足

$$c_i(x) \leq 0, \ i=1,2,\cdots,k; \quad c_i(x) < 0, \ i=k+1,k+2\cdots,m; \quad Ax=b,$$

即对线性不等式约束无需要求其存在严格可行点.

Slater约束品性与强对偶原理

定理

若凸优化问题满足*Slater*条件，则强对偶原理成立.

- 当$d^* > -\infty$ 时，对偶问题的最优解可以取到，即存在对偶可行解(λ^*, ν^*)，满足$g(\lambda^*, \nu^*) = d^* = p^*$.
- 假设集合$\mathcal{D}$ 内部非空(即$\mathrm{relint}\mathcal{D} = \mathbf{int}\,\mathcal{D}$)，A 行满秩(否则可以去掉多余的线性等式约束)以及原始问题最优函数值p^* 有限.
- 定义集合

$$
\begin{aligned}
\mathbb{A} = \{(u, v, t) \mid \exists x \in \mathcal{D},\ c_i(x) \le u_i,\ i = 1, 2, \cdots, m,\\
Ax - b = v, f(x) \le t\}.\\
\mathbb{B} = \{(0, 0, s) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R} \mid s < p^*\}.
\end{aligned}
$$

- 可以证明集合$\mathbb{A}$ 和$\mathbb{B}$ 是不相交的.
- 假设$(u, v, t) \in \mathbb{A} \cap \mathbb{B}$.根据$(u, v, t) \in \mathbb{B}$，有$u = 0, v = 0$ 和$t < p^*$.
- 由$(u, v, t) \in \mathbb{A}$，可知$f(x) \le t < p^*$，这与p^* 是原始问题最优值矛盾.

Slater约束品性与强对偶原理

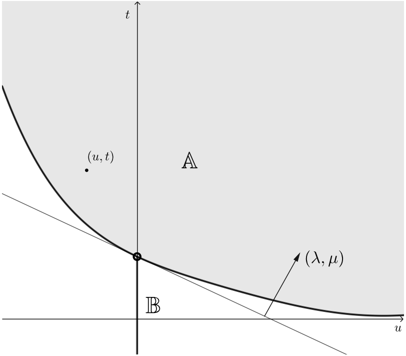

Figure: 集合$\mathbb{A}$ 和$\mathbb{B}$ 在u - t 方向投影的示意图
($\mathbb{A}$ 一般为有内点的凸集, $\mathbb{B}$ 是一条射线且不含点$(0,0,p^*)$)

因为$\mathbb{A}$ 和$\mathbb{B}$ 均为凸集, 由超平面分离定理, 存在$(\lambda,\nu,\mu)\neq 0$ 和α, 使得

$$\lambda^{\mathrm{T}}u+\nu^{\mathrm{T}}v+\mu t\geq\alpha,\quad\forall\,(u,v,t)\in\mathbb{A},$$
$$\lambda^{\mathrm{T}}u+\nu^{\mathrm{T}}v+\mu t\leq\alpha,\quad\forall\,(u,v,t)\in\mathbb{B}.$$

- 我们断言$\lambda \geq 0$ 和$\mu \geq 0$ (否则可以取u_i 和t 为任意大的正实数以及$\nu = 0$, 这会导致$\lambda^{\mathrm{T}} u + \mu t$ 在集合$\mathbb{A}$ 上无下界).
- 同时, 由于$\mu t \leq \alpha$ 对于所有$t < p^*$ 成立, 可得$\mu p^* \leq \alpha$.
- 对任意$x \in \mathcal{D}$, 取$(u, v, t) = (c_i(x), Ax - b, f(x)) \in \mathbb{A}$, 可知

$$\sum_{i=1}^{m} \lambda_i c_i(x) + \nu^{\mathrm{T}}(Ax - b) + \mu f(x) \geq \alpha \geq \mu p^*.$$

- 假设$\mu > 0$, 则

$$L(x, \frac{\lambda}{\mu}, \frac{\nu}{\mu}) \geq p^*.$$

进一步地, 我们有$g(\frac{\lambda}{\mu}, \frac{\nu}{\mu}) \geq p^*$, 根据弱对偶性$g(\frac{\lambda}{\mu}, \frac{\nu}{\mu}) \leq p^*$ 自然成立.因此, 必有$g(\frac{\lambda}{\mu}, \frac{\nu}{\mu}) = p^*$ 成立.说明在此情况下强对偶性满足, 且对偶最优解可以达到.

Slater约束品性与强对偶原理

- 考虑$\mu = 0$ 的情况, 可以从上面得到对于所有的$x \in \mathcal{D}$,

$$\sum_{i=1}^{m} \lambda_i c_i(x) + \nu^{\mathrm{T}}(Ax - b) \geq 0.$$

- 取满足Slater条件的点x_S, 有$\sum_{i=1}^{m} \lambda_i c_i(x_S) \geq 0$.
- 又$c_i(x_S) < 0$ 和$\lambda_i \geq 0$, 我们得到$\lambda = 0$, 上式化为

$$\nu^{\mathrm{T}}(Ax - b) \geq 0, \quad \forall\, x \in \mathcal{D}.$$

根据$(\lambda, \nu, \mu) \neq 0$ 可知$\nu \neq 0$, 结合A 行满秩可以得到$A^{\mathrm{T}}\nu \neq 0$.由于x_S 是可行解, 我们有$\nu^{\mathrm{T}}(Ax_S - b) = 0$.

- 因为$x_S \in \mathbf{int}\, \mathcal{D}$, 则存在点$\tilde{x} = x_S + e \in \mathcal{D}$, 满足$\nu^{\mathrm{T}}(A\tilde{x} - b) < 0$.这与$\nu^{\mathrm{T}}(Ax - b) \geq 0$, $\forall\, x \in \mathcal{D}$ 矛盾.
- 综上所述, Slater条件能保证强对偶性.
- 在定理的证明中, Slater条件保证了$\mu \neq 0$.

一阶充要条件

- 对于一般的约束优化问题，当问题满足特定约束品性时，我们知道KKT条件是局部最优解处的必要条件.
- 而对于凸优化问题，当Slater条件满足时，KKT条件则变为局部最优解的充要条件(根据凸性，局部最优解也是全局最优解).

定理 (凸优化问题的一阶充要条件)

对于凸优化问题，用a_i 表示矩阵A^T 的第i 列，$\partial f, \partial c_i$ 表示次梯度，如果*Slater*条件成立，那么x^*, λ^* 分别是原始，对偶全局最优解当且仅当

$$
\begin{aligned}
&\text{稳定性条件} && 0 \in \partial f(x^*) + \sum_{i\in\mathcal{I}} \lambda_i^* \partial c_i(x^*) + \sum_{i\in\mathcal{E}} \lambda_i^* a_i, \\
&\text{原始可行性条件} && Ax^* = b,\ \forall i \in \mathcal{E}, \\
&\text{原始可行性条件} && c_i(x^*) \le 0,\ \forall i \in \mathcal{I}, \\
&\text{对偶可行性条件} && \lambda_i^* \ge 0,\ \forall i \in \mathcal{I}, \\
&\text{互补松弛条件} && \lambda_i^* c_i(x^*) = 0,\ \forall i \in \mathcal{I}.
\end{aligned}
$$

一阶充要条件:充分性

- 设存在$(\bar{x},\bar{\lambda})$ 满足KKT条件，我们考虑凸优化问题的拉格朗日函数

$$L(x,\lambda)=f(x)+\sum_{i\in\mathcal{I}}\lambda_i c_i(x)+\sum_{i\in\mathcal{E}}\lambda_i(a_i^{\mathrm{T}}x-b_i).$$

- 当固定$\lambda=\bar{\lambda}$ 时，注意到$\bar{\lambda}_i\geq 0, i\in\mathcal{I}$ 以及$\bar{\lambda}_i(a_i^{\mathrm{T}}x), i\in\mathcal{E}$ 是线性函数可知$L(x,\bar{\lambda})$ 是关于x 的凸函数.

- 由凸函数全局最优点的一阶充要性可知，此时$\bar{x}$ 就是$L(x,\bar{\lambda})$ 的全局极小点.根据拉格朗日对偶函数的定义，

$$L(\bar{x},\bar{\lambda})=\inf_{x\in\mathcal{D}}L(x,\bar{\lambda})=g(\bar{\lambda}).$$

- 根据原始可行性条件$A\bar{x}=b$ 以及互补松弛条件$\bar{\lambda}_i c_i(\bar{x})=0, i\in\mathcal{I}$,

$$L(\bar{x},\bar{\lambda})=f(\bar{x})+0+0=f(\bar{x}).$$

- 根据弱对偶原理，

$$L(\bar{x},\bar{\lambda})=f(\bar{x})\geq p^*\geq d^*\geq g(\bar{\lambda})=L(\bar{x},\bar{\lambda})\Rightarrow p^*=d^*,$$

故$\bar{x},\bar{\lambda}$ 分别是原始问题和对偶问题的最优解.

关于充分性的评述

- 定理的充分性说明，若能直接求解出凸优化问题的KKT对，则其就是对应问题的最优解.
- 在充分性部分的证明中，我们没有使用Slater条件，这是因为在证明的一开始假设了KKT点是存在的.
- Slater条件的意义在于当问题最优解存在时，其相应KKT条件也会得到满足.
- 当Slater条件不满足时，即使原始问题存在全局极小值点，也可能不存在 (x^*, λ^*) 满足KKT条件.

光滑凸优化实例:仿射空间的投影问题

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & \frac{1}{2}\|x-y\|_2^2, \\ \text{s.t.} \quad & Ax=b, \end{aligned}$$

其中$A\in\mathbb{R}^{m\times n}$, $b\in\mathbb{R}^m$ 以及$y\in\mathbb{R}^n$ 为给定的矩阵和向量且A 满秩.

- 拉格朗日函数: $L(x,\lambda)=\frac{1}{2}\|x-y\|^2+\lambda^{\mathrm{T}}(Ax-b)$.
- Slater条件成立, x^* 为一个全局最优解当且仅当存在$\lambda^*\in\mathbb{R}^m$ 使得

$$\begin{cases} x^*-y+A^{\mathrm{T}}\lambda^*=0, \\ Ax^*=b. \end{cases}$$

- 由上述KKT 条件第一式, 等号左右两边同时左乘A 可得

$$Ax^*-Ay+AA^{\mathrm{T}}\lambda=0 \Rightarrow \lambda^*=(AA^{\mathrm{T}})^{-1}(Ay-b).$$

- 将λ^* 代回KKT 条件第一式可知

$$x^*=y-A^{\mathrm{T}}(AA^{\mathrm{T}})^{-1}(Ay-b).$$

因此点y 到集合$\{x\mid Ax=b\}$ 的投影为$y-A^{\mathrm{T}}(AA^{\mathrm{T}})^{-1}(Ay-b)$.