Essence of the lecture (19/21/22/23/24)

凸问题：

- 非凸、非拟凸问题的松弛（零范数的松弛）
- 广义凸问题及相关定义
- 问题的等价变换
- 狭义凸问题
- 凸问题的性质：局部最优 = 全局最优，目标函数可微情况下的最优解

零范数的松弛：

在一维条件下，零范数是拟凸函数（非凸），在 $n \geq 2$ 情况下，零范数非拟凸，此时主要有两种松弛方法。

- 松弛为 l_1 范数，该函数为凸函数
- 松弛为 $log(ax^2+1)$，该函数非凸，但是是拟凸函数

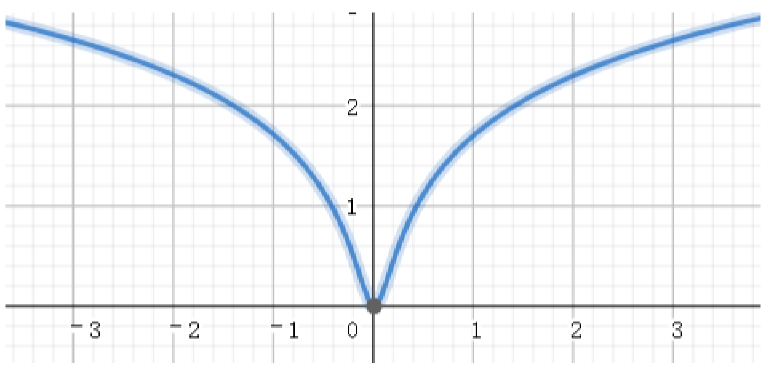

图 1: $f(x) = log(ax^2+1), a \to +\infty$ 时整体趋近零范数，拟凸函数

广义凸问题：凸目标，凸集约束

一般优化问题的描述：

$$\begin{aligned} &\min \ \ f_0(x) \\ &s.t. \ \ f_i(x) \leq 0 \qquad i = 1, \ldots, m \\ &\qquad \ h_i(x) = 0 \qquad i = 1, \ldots, p \end{aligned}$$

x 称为优化变量 (optimization variable)，f_0 称为目标函数、损失函数、效用函数，$f_i(x) \leq 0$ 称为不等式约束 (inequality constraint)，$h_i(x) = 0$ 称为等式约束 (equality constraint)，$m = p = 0$ 时为无约束 (unconstrained) 问题

概念定义 (注意 inf 可以对应 min)：

优化问题的域 (domain)： D=$\bigcap_{i=0}^{m} dom\ f_i\ \cap\ \bigcap_{i=1}^{p} dom\ h_i$

可行解集 (feasible set)： $X_f = \{x \mid x \in D,\ f_i(x) \le 0, i = 1, \ldots, m,\ h_i(x) = 0, i = 1, \ldots, p\}$

最优值 (optimization value)： $p^* = \inf\{f_0(x) \mid x \in X_f\}$，若 $X_f = \emptyset$, $p^* = +\infty$

最优解 (optimization point/solution)： 若 x^* 可行，且 $f_0(x^*) = p^*$，称 x^* 为最优解

ε **次优解集 (ε-suboptional set)：** $X_\varepsilon = \{x \mid x \in X_f,\ f_0(x) \le p^* + \varepsilon\}$

局部最优解 (locally optional)： $f_0(x) = \inf\left\{ f_0(z) \mid \begin{array}{c} f_i(x) \le 0, i = 1, \ldots, m \\ h_i(z) = 0, i = 1, \ldots, p \\ \|z - x\| \le R \end{array} \right\}$，解集为 X_{loc}

不等式约束采用 $\le$ 而非 $<$ 的原因

若 $x \in X_f$, $f_i(x) = 0$，则 $f_i(x) \le 0$ 为活动 (active) 约束

若 $x \in X_f$, $f_i(x) < 0$，则 $f_i(x) \le 0$ 为非活动 (inactive) 约束

而所有的 $<$ 约束都可以转化为 $\le$ 约束

例如 money<100$\Rightarrow -|log(100 - money)| \le 0$

问题的等价变换

例：Box Constraint

$$l_i \le x \le u_i \Rightarrow \begin{cases} l_i - x_i \le 0 \\ x_i - u_i \le 0 \end{cases}$$

例：将问题的量纲做标准化

$$\begin{aligned} &\min\ \ \alpha_0\, f_0(x) \\ &s.t.\ \ \alpha_i\, f_i(x) \le 0 \qquad i = 1, \ldots, m \\ &\qquad\ \beta_i\, h_i(x) = 0 \qquad i = 1, \ldots, p \end{aligned}$$

例：利用函数进行等价变换

ψ_0	$R \to R$	单增
$\psi_1, \ldots, \psi_m$	$R \to R$	$\psi_i(u) \le 0 \Leftrightarrow u \le 0$
$\varrho_1, \ldots, \varrho_p$	$R \to R$	$\varrho_i(u) = 0 \Leftrightarrow u = 0$

$$\begin{aligned} \min \quad & \psi_0\left(f_0(x)\right) \\ s.t. \quad & \psi_i\left(f_i(x)\right) \le 0 \qquad i = 1, \ldots, m \\ & \varrho_i\left(h_i(x)\right) = 0 \qquad i = 1, \ldots, p \end{aligned}$$

$$\text{如 } \min \|Ax - b\|_2 \Leftrightarrow \|Ax - b\|_2^2$$

例：消除等式约束

$\{h_i(x) = 0, i = 1, \ldots, p\}$ 看作一组方程，$X = \varrho(Z)$，$Z \in R^k$，$\varrho :\ R^k \to R^n$

$$\begin{aligned} \min \quad & f_0(\varrho(x)) \qquad \Rightarrow \qquad X^* = \varrho(Z^*) \\ s.t. \quad & f_i(\varrho(x)) \le 0 \qquad i = 1, \ldots, m \end{aligned}$$

例：消除线性等式约束

在线性等式约束的情况下，$h_i(x) = 0 \Rightarrow Ax - b = 0$

此时若方程无解，则问题无解，若 A 可逆，则 $X = A^{-1}b$

否则可将等式约束化为 $X = FZ + X_0$，FZ 在 A 的零空间，即将求解 n 维的 X 化为求解 n-r 维的 Z（因为初等变换 F 不改变 Z 的秩）

只有在有必要的情况下这样做，有时候会带来麻烦（具体麻烦未提及）

狭义凸问题：

$$\begin{aligned} \min \quad & f_0(x) & & f_0(x)\text{为凸} \\ s.t. \quad & f_i(x) \le 0 \qquad i = 1, \ldots, m & & f_i(x)\text{均为凸} \\ & a_i^T x = b_i \qquad i = 1, \ldots, p & & \text{等式约束为仿射函数} \end{aligned}$$

例：引入松弛变量 (slack variable)s_i

虽然是一个升维的情况，但是有时可以将问题化为非常特殊的形式更易求解

$$\begin{aligned} \min \quad & f_0(x) \\ s.t. \quad & s_i \leq 0 \\ & f_i(x) - s_i = 0 \qquad i = 1, \ldots, m \\ & a_i^T x = b_i \qquad\quad\; i = 1, \ldots, p \end{aligned}$$

类似有相关问题（max concave ⇔ min convex）

Quasi-convex optimization f_0 为拟凸函数，f_i，h_i 为凸函数

None-convex optimization f_0 为凹函数，f_i，h_i 为凸函数

求证：凸问题的局部最优即为全局最优

设 x 不是全局最优，$\exists y \in X_f$，$s.t.\ f_0(y) < f_0(x)$

$\because$ x 为局部最优，$\Rightarrow \|y - x\|_2 > R$

令 $z = \theta y + (1-\theta)x$，取 $\theta = \dfrac{R}{2\|y-x\|_2} \in [0,1]$

$\therefore$ z 为凸组合，$z \in X_f$ 且$f_0(z) \leq \theta f_0(y) + (1-\theta) f_0(x)$

$\|z - x\|_2 = \theta\|x - y\|_2 = \frac{R}{2} < R$，由 x 局部最优有 $f_0(x) \leq f_0(z)$

由此应该有$f_0(y) < f_0(x) \leq f_0(z)$ (两条蓝色结论矛盾，因此原假设成立)

分析：目标函数可微情况下的最优解

考虑凸函数的一阶条件：$f(y) \geq f(x) + \nabla f^T(x)(y - x)$

对于全局最优 x^*，此时一定有 $\nabla f_0^T(x^*)(y - x^*) \geq 0$

例：约束仅为等式约束的情况，min $f_0(x)$ $s.t.\ x \geq 0$

此时问题仅有 $Ax = b$ 这一约束条件，由 $Ax^* = b$，$Ay = b \Rightarrow y = x^* + \upsilon$，$\upsilon \in Null(A)$

回代，有 $\nabla f^T(x)\upsilon \geq 0$

要求对 $\forall \upsilon \in Null(A)$均成立 $\Rightarrow \begin{cases} \upsilon = 0 \\ \nabla f(x)\text{与 Null(A) 正交} \end{cases}$

第一种情况下，$x = y$，说明有且仅有一解，$x = A^{-1}b$

例：约束仅为非负约束，$\min f_0(x)\ s.t.\ x \geq 0$

即在给定 x 的情况下，对于 $\forall y \geq 0$, $\nabla f^T(x)(y-x) \geq 0 \Rightarrow \nabla f^T(x)y - \nabla f^T(x)x \geq 0$

$f^T(x)y = \sum a_i x_i$，为了保证对于任意 y 成立，故 $a_i \geq 0 \Leftrightarrow \nabla f(x) \geq 0$

结合 $x \geq 0$，有$\nabla f^T(x)x \geq 0$，又取 y=0 时，有$\nabla f^T(x)x \leq 0$

综上，有 $\nabla f^T(x) = 0 \Rightarrow \begin{cases} x \geq 0 \\ \nabla f(x) \geq 0 \\ (\nabla f(x))_i x_i = 0 \end{cases}$ (互补条件：Complementarity)

Essence of the lecture (25/26/27/28)

典型的凸问题：

- 线性规划
- 线性分数规划
- 二次规划（二次约束的二次规划）
- 半正定规划
- 多目标优化

线性规划：

线性规划要求目标和约束均为线性

linear program 线性问题（名词）

linear programing 线性问题求解（动词）

$$\begin{aligned} min \quad & c^T x + d \\ s.t. \quad & Gx \leq h,\ Ax = b \end{aligned}$$

线性规划的等价变换

$$\begin{aligned} \min \quad & c^T x + d \\ s.t. \quad & Gx + s = h \\ & Ax = b \\ & s \geq 0 \end{aligned} \qquad\qquad \begin{aligned} \min \quad & c^T x^+ - c^T x^- + d \\ s.t. \quad & Gx^+ - Gx^- + s = h \\ & Ax^+ - Ax^- = b \\ & s \geq 0,\ x^+ \geq 0,\ x^- \geq 0 \end{aligned}$$

通过这种变换能够将问题化为线性约束以及非负约束，这方便使用函数 (linprog) 进行求解。

线性分数规划 (linear fractional programing)

BTW，线性分数函数保凸，是拟凸函数但不是凸函数，例如 $1/x$

$$
\begin{aligned}
(P_0)\quad & \min \quad \frac{c^T x + d}{e^T x + f} \\
& s.t. \quad Gx \le h \\
& \qquad\ \ Ax = b \\
& \qquad\ \ e^T x + f > 0
\end{aligned}
\qquad\qquad
\begin{aligned}
(P_1)\quad & \min \quad c^T y + dz \\
& s.t. \quad Gy - hz \le 0 \\
& \qquad\ \ Ax - bz = 0 \\
& \qquad\ \ e^T y + fz = 1 \\
& \qquad\ \ z \ge 0
\end{aligned}
$$

证明两问题等价的思路：P1 的可行解在 P2 可行，P2 的可行解在 P1 可行，同时二者的值相同

证明：1) 若 x 在 P_0 可行，取 $y = \dfrac{x}{e^T x + f}$，$z = \dfrac{1}{e^T x + f}$

$$
\begin{cases}
Gy - hz = \dfrac{Gx - h}{e^T x + f} \le 0 \\
Ay - bz = \dfrac{Ax - b}{e^T x + f} = 0 \\
e^T x + fz = \dfrac{e^T x + f}{e^T x + f} = 1 \\
z > 0
\end{cases}
\Rightarrow \text{x 同时在 P1 可行，同时目标函数值均为 } \frac{c^T x + d}{e^T x + f}
$$

2) 若 y,z 在 P_1 中可行

当 $z > 0 \Rightarrow x = \dfrac{y}{z}$ 在 P_0 可行，且目标函数值均为 $c^T y + dz$

当 $z = 0$，此时有 $Gy \ge 0$, $Ay = 0$, $e^T y = 1$，设 x_0 为 P_0 的可行解，找到射线 $x = x_0 + ty, \forall t \ge 0$ 对 P_0 可行

$$
\begin{cases}
Gx = Gx_0 + Gty \le h \\
Ax = Ax_0 + tAy = b \\
e^T x + f = e^T x_0 + f + te^T y > 0
\end{cases}
\Rightarrow
\begin{array}{l}
\text{目标：在该射线上找到一点使得 } P_1\text{，}P_2 \text{ 同解} \\
f_0(x) = f_0(x_0 + ty) = \dfrac{c^T x_0 + c^T ty + d}{e^T x_0 + e^T ty + f} \xlongequal{t \to \infty} c^T y
\end{array}
$$

综上，P_0，P_1 两问题等价

二次规划 (Quadratic Programing)

$$\begin{aligned} \min \quad & \frac{1}{2}X^TPX + q^Tx + r \\ s.t. \quad & Qx \le h \qquad P \in S_+^n \\ & Ax = b \end{aligned}$$

线性规划问题的最优解只能在边界点取到，二次规划问题的最优解可能在内部取到

二次约束的二次规划 (Quadratically Constrained Quadratic Programing, QCQP)

$$\begin{aligned} \min \quad & \frac{1}{2}X^TPX + q^Tx + r \\ s.t. \quad & \frac{1}{2}X^TP_iX + q_i^Tx + r \le 0 \qquad && P \in S_{++}^n \\ & Ax = b && P_i \in S_+^n,\ i = 1, \dots, m \end{aligned}$$

二次的约束条件其实是椭圆

例：稀疏约束下的最小二乘

之前提到可以使用 1 范数取代 0 范数

$$\begin{aligned} \hat{x} &= \arg\min_x \quad \|b - Ax\|_2^2 + \lambda_0 \|x\|_0 \\ &= \arg\min_x \quad \|b - Ax\|_2^2 + \lambda_1 \|x\|_1 \quad (l_1 - regularized\ least\ squares) \end{aligned}$$

此时由于目标函数内涉及绝对值，是不可导的，同时也不是二次规划问题

令 $x = x^+ - x^-$，回代，由于 $\lambda_1 \|x^+ - x^-\|_1 = \lambda_1 1^T x^+ + \lambda_1 1^T x^-$，从而消除绝对值，化为二次规划的问题

$$\begin{aligned} \hat{x} = \arg\min_x \quad & \|b - Ax^+ - Ax^-\|_2^2 + \lambda_1 1^T x^+ + \lambda_1 1^T x^- \\ s.t. \quad & x^+, x^- \ge 0 \end{aligned}$$

类似有 $(l_2 - regularized\ least\ squares)$，以下两种表述等价，但右边的表述是标准的 QCQP 形式

$$\begin{aligned} \arg\min_x \quad & \|b - Ax\|_2^2 + \lambda_2 \|x\|_2 \qquad\qquad & \arg\min_x \quad & \|b - Ax\|_2^2 \\ \Rightarrow \quad & X^T(A^TA + \lambda_2 I)X + \dots & s.t. \quad & \lambda_2 \|x\|_2 \le \theta \end{aligned}$$

半正定规划 (Semi-definite Program)

半正定规划有两种形式，一种是矩阵形式，一种是向量形式

半正定规划的矩阵形式，这实际上是矩阵空间的线性规划（因为 trace 的操作实际上是线性操作，可从特例对角矩阵理解）

$$\begin{aligned} \min \quad & tr(CX) \\ s.t. \quad & tr(A_iX) = b_i,\ i = 1, \ldots, p \\ & X \succeq 0,\ X \in S_+^n,\ C \in R^{n\times n},\ A_i \in R^{n\times n},\ b_i \in R \end{aligned}$$

半正定规划的向量形式

$$\begin{aligned} \min \quad & c^T x \\ s.t. \quad & x_1A_1 + \cdots + x_nA_n \preceq B \\ & x \in R^n,\ B, A_1, \ldots, A_n \in S^k,\ C \in R^n \end{aligned}$$

例：谱范数 (最大奇异值) 问题的转换

原问题：$\|A(x)\|_2$

由于 $\|A(x)\|_2 \le \sqrt{S} \Leftrightarrow A(x)^T A(x) - SI \preceq 0$

问题化为

$$\begin{aligned}
&\min \quad \sqrt{S}\ (\text{非凸}) \quad \Leftrightarrow \quad S\ (\text{凸}) && (1)\\
&\quad s.t. \quad A(x)^T A(x) \preceq SI \\
\Rightarrow &\min \quad t \\
&\quad s.t. \quad A(x)^T A(x) \preceq t^2 I,\ t \ge 0 && (2)\\
\Rightarrow &\min \quad t \\
&\quad s.t. \quad \begin{bmatrix} tI & A(x) \\ A^T(x) & tI \end{bmatrix} \succeq 0,\ t \ge 0 && (3)\\
\Rightarrow &\min \quad t \\
&\quad s.t. \quad Y = \begin{bmatrix} tI & A(x) \\ A^T(x) & tI \end{bmatrix},\ Y \succeq 0,\ t \ge 0 && (4)
\end{aligned}$$

(1) 通过将目标函数转为 S，使称为凸问题，(2) 到 (3) 的转化是线性代数的内容，(3) 不够之处在于仍旧存在二次项约束，(4) 通过设置 Y，将矩阵正定条件拆出一个等式约束，而后 $Y \succeq 0\ \ t \ge 0$ 变为标准的半定规划形式

多目标优化

帕累托(帕累托曲面/曲线，帕累托最优点/最优值)

帕累托曲面/曲线 (pareto front)：若找到另一解，使其在某指标上更好，必然在某指标上更差。

若 $\{f_0(x)\}$ 在 R^k 中为凸，$f(x)$ 为凸，$h_i(x)$ 为仿射，则必可由如下方法求得 pareto front 中一点

$$\begin{aligned}
\min \quad & \sum_{i=1}^{q} \lambda_i f_{0i}(x) \\
s.t. \quad & \lambda_i \ge 0 \\
& f_i(x) \le 0 \qquad i = 1, \ldots, m \\
& h_i(x) = 0 \qquad i = 1, \ldots, p
\end{aligned}$$

从而通过遍历 $\{\lambda_i\}$ 可以找出所有解，然而现实中由于 $f(x)$ 的凸性无法满足，导致无法找出帕累托曲面上所有点

例：Rideg Regression

关键是展示目标函数和约束可以相互转换

原问题是一个多目标优化问题（左），右边是该问题的等价转换形式，要证明这两种形式等价

$$\begin{cases} \min & \|b-Ax\|^2 \\ \min & \|x\|^2 \end{cases} \Leftrightarrow \begin{array}{lll} \min & \|b-Ax\|^2+\lambda\|x\|^2 & (1) \\ \min & \|b-Ax\|^2 & \\ s.t. & \|x\|^2 \le \epsilon & (2) \end{array}$$

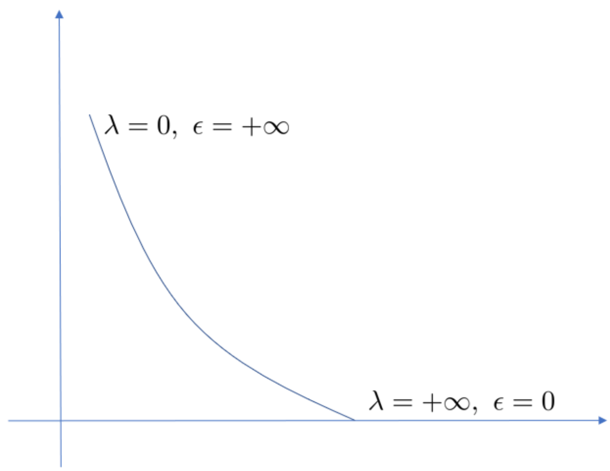

图 2: 这幅图是目标函数的帕累托曲面，横轴表示最小化误差这一目标，纵轴表示能量这一目标，两个端点分别表示两种表述中的不同取值，例如与横轴的交点，意味着能量为 0，即 $\epsilon = 0$，相应的在 (1) 中代表仅考虑能量，$\lambda = +\infty$，另一情况同理