应用运筹学基础：线性规划 (1) - 极点与基可行解

学校有一门课叫《应用运筹学基础》，是计算机学院唯一教优化的课程，感觉上得还行，这里简单记录一下上课学到的知识。第一节课是线性规划（linear programming）。

凸集

对于集合 S，若任意两元素 $x, y \in S$，且对于任意 $0 \le \theta \le 1$ 有 $\theta x + (1-\theta)y \in S$，那么 S 是凸集（convex set，形象地想象就是凸的图形）。

可以推广：若 S 为凸集，那么对任意 $n \ge 2$ 个元素 $x_1, x_2, \dots, x_n \in S$ 以及任意 $\sum_{i=1}^n \theta_i = 1$，都有 $\sum_{i=1}^n \theta_i x_i \in S$。

可以使用归纳法证明：

(1) 对于 $n = 2$，根据凸集的定义，结论成立。

(2) 若对于 $n = k$ 结论成立，即对任意 $x_1, x_2, \dots, x_k \in S$ 以及任意 $\sum_{i=1}^k \theta_i = 1$，有 $\sum_{i=1}^k \theta_i x_i \in S$。那么对于任意 $y_1, y_2, \dots, y_{k+1} \in S$ 以及任意 $\sum_{i=1}^{k+1} \lambda_i = 1$，有 $\sum_{i=1}^k \lambda_i = 1 - \lambda_{k+1}$，即 $\sum_{i=1}^k \lambda_i / (1-\lambda_{k+1}) = 1$，那么 $t_{k+1} = \sum_{i=1}^k \lambda_i y_i / (1-\lambda_{k+1}) \in S$，根据凸集的定义，自然有 $\sum_{i=1}^{k+1} \lambda_i y_i = (1-\lambda_{k+1}) t_{k+1} + \lambda_{k+1} y_{k+1} \in S$，结论成立。

凸集的交仍然是凸集，容易通过定义证明。

凸函数

对于定义在凸集 S 上的函数 $f(x)$，若对于任意 $0 \le \theta \le 1$ 有 $f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta) f(y)$，那么 $f(x)$ 是凸函数（convex function）。

可以推广为延森不等式（Jensen's inequality，原来一直念琴生不等式，被老师吐槽了一通...）：若 $f(x)$ 是凸函数，那么对于任意 $\sum_{i=1}^n \theta_i = 1$，有 $f(\sum_{i=1}^n \theta_i x_i) \le \sum_{i=1}^n \theta_i f(x_i)$。

若 $-f(x)$ 是凸函数，那么 $f(x)$ 是凹函数（concave function）；根据定义，仿射函数（affine function）即是凸函数又是凹函数。

凸函数美妙的性质是：局部最优就是全局最优。利用反证法证明如下：

若 x_1 与 x_2 满足 $|x_1 - x_2| < \epsilon$，那么 x_1 与 x_2 在对方的邻域内。假设 $\hat{x}$ 是局部最优点而不是全局最优点，设 x^* 为全局最优点，那么 $f(\hat{x}) > f(x^*)$。由于 $f(x)$ 为凸函数，那么对于它们凸组合出来的一点 $x = \theta\hat{x} + (1-\theta)x^*$ 有 $f(x) \leq \theta f(\hat{x}) + (1-\theta)f(x^*) < f(\hat{x})$。只要取 $\theta = 1 - \epsilon/(2|\hat{x} - x^*|)$，就有 $|x - \hat{x}| = \epsilon/2 < \epsilon$，说明 x 在 $\hat{x}$ 的邻域内，而且比它优，与我们开始的假设不相符。

线性规划

线性规划（linear programming，LP）问题指的是如下形式的优化问题：$\begin{aligned} \min_x \quad & c^T x + d \\ \text{s.t.} \quad & Ax \leq b \\ & Px = q \end{aligned}$ 简单来说，就是目标函数和约束函数都是仿射函数的优化问题。

由于仿射函数既是凸函数又是凹函数，所以优化问题是 min 还是 max 问题不大；由于常数 d 对优化问题的解没有影响，一般也可以去掉。课堂上讨论的 LP 问题是如下形式的问题 $\begin{aligned} \max_x \quad & c^T x \\ \text{s.t.} \quad & Ax \leq b \\ & Px = q \\ & x \geq 0 \end{aligned}$ 其实，$Ax \leq b$ 这个约束，可以通过给每个不等式增加一个松弛变量进行松弛：对于 $a_1x_1 + a_2x_2 + \cdots + a_nx_n \leq b$ 这个约束，我们添加 x_{n+1}，把问题变为 $a_1x_1 + a_2x_2 + \cdots + a_nx_n + x_{n+1} = b$。注意到原来的约束取的是小等于号，所以 $x_{n+1} \geq 0$ 这个条件是满足的。

这样，我们就把 LP 问题特化为 $\begin{aligned} \max_x \quad & c^T x \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0 \end{aligned}$

为了发掘 LP 问题的一些性质，我们进行一些定义。

极点（extreme point）：设 S 为凸集，若 $x \in S$ 无法表示为其它两个 S 内元素的凸组合，那么 x 是极点。

LP 问题的可行域实际上是很多超平面的交，最后组成的应该是一个超多面体。在这个超多面体有界的情况下，极点就是这些超多面体的顶点。对于 LP 问题而言，在超多面体有界的情况下，最优解一定可以在极点取到，且极点的数量是有限的（不过不知道怎么证明- -但是感性地想一想好像还是很有道理的，和函数的极值什么的有点像...）

基可行解（basic feasible solution）：我们讨论 $Ax = b$ 有解且行满秩的情况（如果 $Ax = b$ 没有解那这个问题也没得做了，如果行不满秩，那么我们去掉线性相关的限制条件即可）。设 A 是一个 $m \times n$ 的矩阵，根据线性代数的知识，我们可以从 A 中选出最多 m 列线性无关的列向量，其它列向量都和它们线性相关。我们把这 m 个列向量调整到前面去，把 A 分成两部分：$A = [A_B \quad A_N]$，其中 A_B 就是那 m 个线性无关的列向量。我们容易构造出 $Ax = b$ 的一个解：$x = \begin{bmatrix} A_B^{-1}b \\ 0 \end{bmatrix}$ 称这种解为基可行解。显然，基可行解**至多**有 C_n^m 种。

接下来我们要证明一个厉害的定理：**每个极点都对应着一个基可行解，且每个基可行解都对应着一个极点**。有了这个定理，再结合可行域有解情况下最优解一定可以在极点取到，我们只要枚举基可行解，就能找到最优解了（至于如何优雅地枚举下节课再说- -）。

首先证明一个引理：**若 $x = [x_1 \quad x_2 \quad \dots \quad x_k \quad 0 \quad 0 \quad \dots]$ 不是基可行解，那么 x 中非 0 元素对应的 A 中的 k 列是线性相关的**。如果 $k > m$ 显然这 k 列线性相关；如果 $k \le m$ 但这 k 列线性无关，那么我们就可以把这 k 列当作 A_B（如果 $k < m$ 就再选几个线性无关的列，凑成 m 个），x 就成为了一个基可行解。所以这 k 列一定是线性相关的。

首先我们用反证法证明：若 x 为极点，那么 x 也是基可行解。假设 x 并不是基可行解，我们把 x 里的非 0 元素（假设有 k 个）都调整到前面去（相应地，也要把 A 中这 k 个非 0 元素对应的列调到前面去），那么我们可以把 x 写为 $x = [x_1 \quad x_2 \quad \dots \quad x_k \quad 0 \quad 0 \quad \dots]$。根据引理，$A$ 中对应的 k 列是线性相关的。

记线性相关的 k 列为 $p_1, p_2, \ldots, p_k$，我们就有不全为 0 的 λ_i，使得 $\sum_{i=1}^{k} \lambda_i p_i = 0$。记辅助向量 $v = [\lambda_1 \quad \lambda_2 \quad \ldots \quad \lambda_k \quad 0 \quad 0 \quad \ldots]^T$，令 $x' = x + \epsilon v$，$x'' = x - \epsilon v$，显然我们有 $x = (x' + x'')/2$。由于 x_1 至 x_k 均大于 0，当 ϵ 充分小时，$x' \neq x''$，且 $x' \geq 0$ 和 $x'' \geq 0$ 的性质也能得到保证。另外，$Ax' = A(x + \epsilon v) = Ax = b$，$Ax'' = A(x - \epsilon v) = Ax = b$，说明 x' 与 x'' 都是可行解。这就是说，x 可以表示为可行域内其它两点的凸组合，与 x 是极点的假设矛盾，反证法结束。

我们继续用反证法证明：若 x 为基可行解，那么 x 为极点。假设 x 不是极点，那么有 $x = (x' + x'')/2$，而且 $x' \neq x''$，以及 $x' \geq 0$ 与 $x'' \geq 0$。设 $x_i = 0$，注意到 x' 与 x'' 元素非负，那么 $x'_i = x''_i = 0$。设 x 中有 k 个非 0 元素，根据基可行解的定义，这些元素所对应的 A 的列向量是线性无关的，那么想从这些列向量得到 b，线性组合的方式也是唯一的。这就说明了 $x = x' = x''$，则 x 是极点，反证法结束。

这次课上还遇到了一个有趣的转换。给定 $m \times n$ 的矩阵 A 和 m 维向量 b，考虑以下优化问题：

$$\begin{aligned} \max_x \min_j \quad & \sum_{i=1}^{m} b_i x_i - \sum_{i=1}^{m} a_{i,j} x_i \\ \text{s.t.} \quad & x \leq q \\ & x \geq 0 \end{aligned}$$

这个问题第一眼看上去并不像一个线性规划，因为这是一个 max 再套 min 的问题。我们把它改写一下：

$$\begin{aligned} \max_x \quad & \sum_{i=1}^{m} b_i x_i - (\max_j \quad \sum_{i=1}^{m} a_{i,j} x_i) \\ \text{s.t.} \quad & x \leq q \\ & x \geq 0 \end{aligned}$$

注意到第二个 max 针对的变量 j 的取值是有限的（只有 1 到 n），我们就可以把它提出来，变成下面的问题：

$$\begin{aligned} \max_{x,y} \quad & \sum_{i=1}^{m} b_i x_i - y \\ \text{s.t.} \quad & x \leq q \\ & x \geq 0 \\ & y \geq \sum_{i=1}^{m} a_{i,j} x_i \quad \forall j = 1, 2, \ldots, n \end{aligned}$$

一下子就变成了线性规划问题，感觉这个操作非常厉害...