

H1 **Uncertainty Modeling and Optimization- 不确定性建模与优化-基础理论篇**

H2 **0. 前言**

已经很久没有写过类似学科知识分享的内容了。之前提到过多次，最近一段时间在学习不确定性问题的处理。虽然学习很不连贯，导致效率一直不高，但也算读过/看过一些资料，对这方面有了大概的了解。感觉自己自主推倒/复习不够，根基很浅，为了巩固学过的一些知识，决定做一点小笔记，在做的过程中复习学过的内容，也顺带与大家分享。

“不确定性优化”这个词应该是没有的。我这样叫，是因为我打算介绍的内容基本上用于处理不确定性问题。这个范围是很广的，因为优化本身是很广的概念，我也不能够、不打算做/学/写这么多内容。所以我改成了“建模与优化”，旨在介绍以优化应用为目的的建模方法，以及一些基本的处理方法。

无需多说，在现实生活中，很多东西都是不确定的。客户的需求，行驶的时间，投资收益率。我们通常用参数的期望来估计其真实值，这在很多情况下会对结果造成巨大的影响。轻则使得目标函数变差，比如错误估计了开会迟到扣工资，重则使得解不可行，比如错过了整个会议，领导让你走人。因此，如何更精确地处理不确定性问题，是很重要的。

简单来说，不确定性的建模方法主要是随机规划(stochastic programming/optimization, 简称SO)，鲁棒优化(robust optimization, 简称RO)和机会约束(chance-constraints programming, 简称CCP)。SO和CPP是传统的处理方法，已经有很多很多年的研究。他们主要基于已知分布的随机变量，这在现实中往往是很难得到的。此外，SO比较多的使用期望这一概念，因此结果往往是平均情况下的，不能反应极端情况的结果。RO主要是这个世纪发展的一门技术，无需多说，近几年在优化领域，无论是OR还是ML都产生了很大的影响。他的目的是为了放松“已知概率分布”的条件，并得到一个鲁棒性够强的解，而不是平均意义上的解。当然，对应的问题就是，由于解需要在任何情况下都可行，所以解的质量可能在大多数情况下不够好，也就是过于conservative(保守)。深究下去，SO也有Apriori paradigm和reoptimization，RO也有classical RO，Distributionally RO，Adaptive/Adjustable RO，ADRO等多种框架。不同的建模范式各有特点。

虽然叫不确定性，但我暂时比较想写/学的内容还是RO相关的，因为SO目前的知识已经够我用了，CCP则暂时不想研究。因此，我只会花一点内容简单介绍SO和CCP，留坑以后再填。因为未来不一定更新，所以暂时决定在近几天边学边写，多写一点内容。

回到作者本身，我是管理学出身，本科没有凸优化、随机过程等课程，只学过皮毛的线性规划，数学基础很差。但是，RO本身是一门很严谨的内容，因为他在一定程度上挑战了一般的线性规划框架，并且涉及概率论，推导的过程自然需要很严谨。学RO的目的如果真的想推导公式，那是比较难的。好在，作者是做algorithm的，我们的目的是应用。所以，我的内容也会很不严谨，以理解为目的，并添加一些应用的内容，比如RO在facility location, portfolio, vehicle-routing等方面的应用，作为学习的实例。不过对应的，作者基础差的好处就是，介绍到内容可能会详细一些。在文章中，对我自己也不确定的内容，我会尽量用<>标出，或者用文字说明。此外，我也会尽量详细地列出知识的来源，因为在我的学习过程中，很多文献/课程/资料都起到很大的作用。有什么问题也希望大家指出。

大家一起学习，一起进步！

H2 **1. 随机规划和机会约束简介**

在第一章里，我们简单了解CCP和SP的基本概念。这些内容与后续RO关系不大，但有些部分有类似的概念，相互结合可能有助于理解。

H3 Chance-constraints programming

前文提到，不确定性的建模方法主要是SO, RO, CCP。这里我们先简单了解SO和CCP。对于一个简单的例子问题（1）：

$$\begin{aligned} \max. \quad & c^T x \\ \text{st.} \quad & Ax \leq b \\ & x \geq 0. \end{aligned} \tag{1}.$$

$$c \in \mathbb{R}^n, A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m.$$

问题（1）是一个常规的线性规划（linear programming, 简称LP）问题。在不确定性环境下，其中的参数A, b, c都可能是随机变量。问题就转化为了：

$$\begin{aligned} \max. \quad & c(\xi)^T x \\ \text{st.} \quad & A(\xi)^T x = b(\xi) \\ & x \geq 0. \end{aligned} \tag{2}.$$

在问题（2）中，原问题的各个参数都由随机变量 ξ 控制。这个问题就已经超出了LP的处理范围。在SO和CCP中，我们通常假定所有随机变量的分布函数已知（似乎不是所有随机变量都有给定的分布函数的，这种说法不一定正确）。所以我们可以尝试使用一些随机变量的数字特征描述问题，比如期望：

$$\begin{aligned}
 & \max_{\xi} E_{\xi} [c(\xi)^T x] \\
 \text{s.t. } & E_{\xi} [A(\xi)^T x] = E_{\xi} [b(\xi)] \quad (3) \\
 & x \geq 0.
 \end{aligned}$$

当然，问题 (3) 是一个确定性问题。这样估计显然有失偏颇，如前文所说，这种方式无法衡量风险的损失。因此，CCP用一个概率为风险兜底。这个概率也可以理解为概率论的参数估计中的置信度：

$$\begin{aligned}
 & \max_{\xi} E_{\xi} [c(\xi)^T x] \\
 \text{s.t. } & P_{\xi} [A(\xi)^T x - b(\xi)] \geq 1 - \varepsilon. \quad (4) \\
 & x \geq 0.
 \end{aligned}$$

问题 (4) 是一个很典型的机会约束模型。简单来说，他确保了满足约束的概率大于某个给定的下界，这个下界也决定了问题的保守程度。虽然概率 P 不一定能转化为 LP 的形式处理，但我们有时可以进行一些简单的处理，得到确定性约束。以下是[1]中关于旅行时间随机的车辆路径规划问题 (SVRP) 的一个例子：

Proposition 1. Consider a SVRPTD and let $R_k = \{r_0 = 0, r_1, r_2, \dots, r_{n_k}, r_{n_k+1} = 0\}$ be the k -th route of the solution of SVRPTD and $V_k = \{r_0, r_1, \dots, r_{n_k}, r_{n_k+1}\}$ be the set of vertex in route R_k . If travel and service times, $t_{r_i r_j}$ and δ_{r_i} , ($r_j, r_k, r_l \in V_i$), and B are independent and normally distributed random variables, then chance constraint $P\{\sum_{i=0}^{n_k} (t_{r_i r_{i+1}} + \delta_{r_i}) \leq B\} \geq \beta$ has the deterministic equivalent

$$\phi^{-1}(\beta) \sqrt{\sum_{i=0}^{n_k} [V(t_{r_i r_{i+1}}) + V(\delta_{r_i})] + \sum_{i=0}^{n_k} [E(t_{r_i r_{i+1}}) + E(\delta_{r_i})]} \leq B$$

where ϕ is the standardized normal distribution function and ϕ^{-1} is the inverse function of ϕ . $E(\cdot)$ and $V(\cdot)$ denote the expected and variance values, respectively.

Proof. We first define variable $y = \sum_{i=0}^{n_k} (t_{r_i r_{i+1}} + \delta_{r_i}) - B$. Because $t_{r_i r_j}$, δ_{r_i} , and B are independently random variables with normal distribution, variable y is also normally distributed, namely $y \sim \mathcal{N}(E(y), V(y))$. The expected and variance values are defined as follows:

$$E(y) = \sum_{i=0}^{n_k} [E(t_{r_i r_{i+1}}) + E(\delta_{r_i})] - B, \quad V(y) = \sum_{i=0}^{n_k} [V(t_{r_i r_{i+1}}) + V(\delta_{r_i})]$$

so we have $\sum_{i=0}^{n_k} (t_{r_i r_{i+1}} + \delta_{r_i}) - B - E(y)/\sqrt{V(y)} \sim \mathcal{N}(0, 1)$. Constraint $\sum_{i=0}^{n_k} (t_{r_i r_{i+1}} + \delta_{r_i}) \leq B$ is equivalent to the following equation:

$$\frac{\sum_{i=0}^{n_k} (t_{r_i r_{i+1}} + \delta_{r_i}) - B - E(y)}{\sqrt{V(y)}} \leq -\frac{E(y)}{\sqrt{V(y)}}$$

so the chance constraint is equivalent to $P\left\{\theta \leq -E(y)/\sqrt{V(y)}\right\} \geq \beta$, where $\theta \sim \mathcal{N}(0, 1)$. So the chance constraint holds if and only if $\phi^{-1}(\beta) \leq -E(y)/\sqrt{V(y)}$, namely

$$\phi^{-1}(\beta) \sqrt{\sum_{i=0}^{n_k} [V(t_{r_i r_{i+1}}) + V(\delta_{r_i})]} + \sum_{i=0}^{n_k} [E(t_{r_i r_{i+1}}) + E(\delta_{r_i})] \leq B \quad \square$$

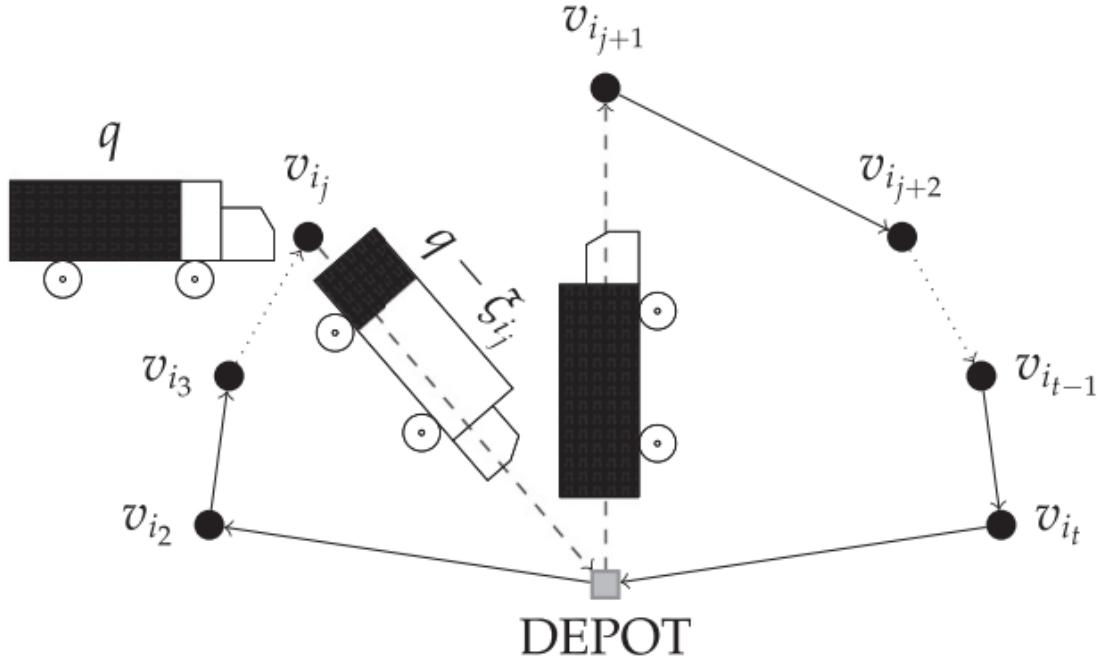
直接看可能有些难懂，其实约束的 P 中所有参数都是已知的。当我们知道随机变量 B 的分布函数（这里满足正态分布）时，可以用分布函数的逆函数表示 P ，从而得到一个非线性的确定性约束。目前这样仍然无法用 LP 处理，并且这个例子把决策变量去掉了。不过，这依旧给出了一个可行的操作方向，如果新公式是凸的，或许可以使用其他的工具解决。因为我对 CCP 的了解不多，这一方向暂时不继续拓展。

H3 Apriori paradigm

[2] 将 SO 在 SVRP 中的应用分为 Apriori paradigm 和 reoptimization。前者也被称为 stochastic programming with recourse。Apriori 将问题改变成两阶段问题：

$$\begin{aligned} \max. \quad & c(\xi)^T x + \bar{E}_{\xi} [Q(x, \xi)] \\ \text{s.t.} \quad & \bar{E}_{\xi} [A(\xi)^T x] = \bar{E}_{\xi} [b(\xi)] \quad (5) \\ & x \geq 0. \end{aligned}$$

其中 $Q(x, \xi)$ 代表 x 和 ξ 实现之后recourse的期望成本。在第一阶段，我们首先根据参数的期望值构造一个解。在第二阶段，如果随机变量的实现值使得该解违背了某些约束，则通过一个给定的recourse修复解，recourse的成本记入总成本。所以，我们可以在求解第一阶段的时候将recourse的期望成本加入成本计算中。



以随机客户需求的VRP问题为例，假设原先路径为 $\{depot, v_{i_2}, v_{i_3}, \dots, v_{i_t}, depot\}$ ，这条路径上所有客户的需求的期望值之和必须小于车辆总容量。但是，如果车辆到达节点 v_{i_j} 后发现容量不足，那么车辆会返回depot，装/卸货后，再前往下一个节点。多出的路途 $d_{v_{i_j},0} + d_{0,v_{i_{j+1}}} - d_{v_{i_j},v_{i_{j+1}}}$ 则作为recourse cost，乘上这种情况发生的概率，记入总成本。这里的recourse可以简单定义为，如果在点*i*发现剩余容量不够服务点*i+1*，则返回仓库补货。通过这种方式，Apriori paradigm考虑了失败的风险带来的成本。

一般不能直接用线性规划求解。但是，由于Apriori paradigm本身可以看作两阶段问题，因此可以使用处理两阶段问题的线性规划算法。常用的求解方式L-shaped algorithm就是基于branch-and-cut，或者说更类似于Benders decomposition，通过对第一阶段的最优解添加feasibility cut来保证解的可行性，添加optimality cut保证最优性。在这里暂时不做更多介绍。此外，由于第二阶段成本可以作为期望成本加入第一阶段，因此也可以使用常规非线性的解法，比如启发式算法、branch-and-price。

H3 Reoptimization

另一种reoptimization的方式则利用了马尔科夫决策过程，动态调整解。在无后效性的马氏过程中，可以通过动态规划（DP）的方式构建模型，然后通过逆向求解的方式求解。计算成本的时候，我们依然使用期望值作为成本。

$$\tilde{\mu}_k(x) = \arg \min_{u_k \in U_k(x_k)} \sum_{x_{k+1} \in S} p_{x_k x_{k+1}}(u_k) \{g(x_k, u_k, x_{k+1}) + \tilde{J}(x_{k+1}) | x_k = x\}.$$

以上式子给了一个DP公式的例子。这里 p 项为概率， \tilde{J} 为DP的cost-to-go函数，代表每个状态下的最优解， g 代表当前决策带来的成本， u 代表决策， x 代表状态。所以，问题的目标值等于，在每一种状态 x_k 下，选择使得期望成本最小的决策 u_k 时的期望成本。由于reoptimization完全建立在DP和马氏过程的基础上，因此相关求解算法也是离不开这两者。这方面我了解不多，因此暂时不做更多探讨。

以上介绍是我学习SVRP相关内容时的了解，SO在很多领域可能有别的建模范式，有待未来补充。一些没有详细介绍的内容，比如求解的技术，待以后再补充。下一章将正式进入RO的内容。

参考文献：

[1] Xiangyong Li, Peng Tian, Stephen C.H. Leung. Vehicle routing problems with time windows and stochastic travel and service times: Models and algorithm. Int. J. Production Economics

[2] Michel Gendreau, Ola Jabali, Walter Rei. 50th Anniversary Invited Article—Future Research Directions in Stochastic Vehicle Routing. Transportation Science. (强烈建议阅读)

H2 2. 经典鲁棒优化求解

前面我们介绍，鲁棒优化有classical RO（这个好像是我自己叫的），Distributionally RO，Adaptive/Adjustable RO，ADRO等。DRO, ARO, ADRO都是对经典RO在某些方面的改进，他们的处理方法都以RO为基础。而RO的求解方法，归根到底，都是基于线性规划中的对偶理论。

H3 Robust Counterpart

我们前面引入了这个问题：

$$\begin{aligned} \max_{\xi} \quad & c(\xi)^T x \\ \text{s.t.} \quad & A(\xi)^T x = b(\xi) \\ & x \geq 0. \end{aligned} \tag{2}$$

注意这里的 A, b, c 关于 ξ 是一种线性关系，可以称为robust linear constraints。为了方便起见，这里的 b 是一个标量而非向量。对于SO而言， ξ 是一个已知分布的随机变量。如果有多个变量，一般还假设iid（独立同分布）。在RO中，我们不知道 ξ 的分布。但我们假设知道 ξ 的取值范围： $\xi \in Z$ 。这里的集合 Z 被称为不确定集（uncertainty set）。这显然放松了对 ξ 的要求。

如名字所说，RO需要解有足够的鲁棒性，也就是说解应该在任何情况下都可行、都由目标函数确保有准确的下界。在SO with recourse中，为了保证在某些情况下解的不可行带来的成本，我们在目标函数中添加了 Q 项。在RO中，我们不添加这样的成本，而是选择让目标函数以下界的形式确保其鲁棒性。我们可以将问题 (2) 写成如下的鲁棒对应形式 (Robust Counterpart, RC)：

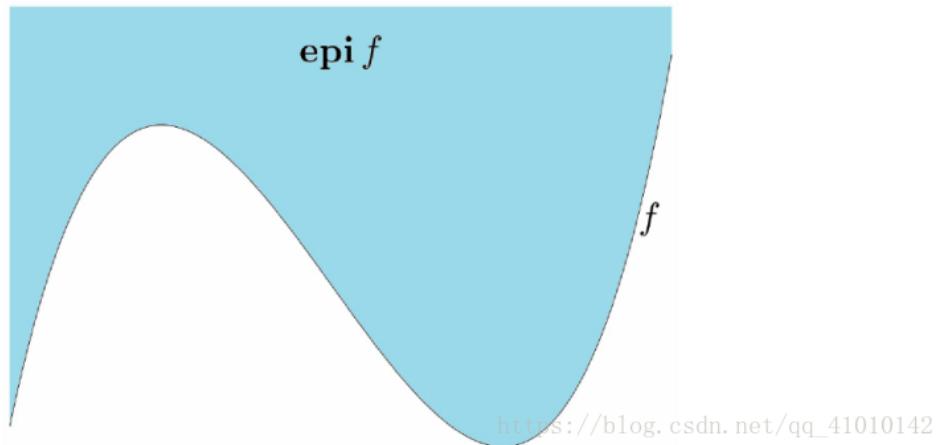
$$\begin{aligned} \max_x \quad & \min_{\xi} c(\xi)^T x \\ \text{s.t.} \quad & A(\xi)^T x = b(\xi), \forall \xi \in Z \quad (\text{Robust Counterpart, RC}) \\ & x \geq 0. \end{aligned}$$

首先保证约束条件对所有情况可行，其次目标函数变成一个max-min形式。我们的优化对象变成了目标函数的下界。但是，max-min并不是线性算子。我们将问题转化为上境图 (epigraph) 形式：

$$\begin{aligned}
 & \max_{x,t} && t \\
 & \text{s.t.} && t \leq c(\xi)^T x, \quad \forall \xi \in Z. \quad (\text{RC of epigraph}) \\
 & && Ax = b(\xi), \quad \forall \xi \in Z \\
 & && x \geq 0.
 \end{aligned}$$

这里的epigraph应该是凸优化中的一个概念。函数 f 的上境图定义为：

$$\text{epi } f = \{(x, t) \mid x \in \text{dom } f, f(x) \leq t\}$$



其中 $\text{dom } f$ 表示 f 的定义域。

回到问题，注意，这种形式应该叫epigraph的RC。RO的epigraph应该是不写成不确定集的。

在epigraph形式下，RO已经转化为了一个线性形式。但是，很多情况下 Z 是无限集合，比如连续型随机变量。一般的LP显然无法处理无限个约束的问题。因此，我们用一种称为reformulation的方法将原问题转化为可以求解的LP问题。

H3 Reformulation

我们考虑一个简化的例子：

$$\begin{aligned}
 & \min && C^T x \\
 & \text{s.t.} && x + P \cdot \xi \leq b, \quad \forall \xi \in Z \quad (0) \\
 & && x \geq 0.
 \end{aligned}$$

这里的不确定性只在约束的一次项系数中出现。当目标项有不确定性时，可以通过如上的epigraph转化到约束中；当常数项 b 中存在不确定性时，可以写成 $b(\xi)^T x_0$ 的形式，固定 $x_0 = -1$ ，将其转化到一次项中。此外，我们还用 $a + P\xi$ 表示了 $A(\xi)$ 。这里的 P 是一个矩阵。这种模型被称为factor model。相比于 $a + \hat{a}\xi$ 的均值+离差形式，这种model能反映出更多的 ξ 的组成。

• Factor model

$$a = a^0 + Bz$$

- Example 1:

- $a_1 = a_1^0 + \hat{a}_1 z_1,$
- $a_2 = a_2^0 + \hat{a}_2 z_2,$
- $a_3 = a_3^0 + \hat{a}_3 z_3.$

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} a_1^0 \\ a_2^0 \\ a_3^0 \end{bmatrix} + \begin{bmatrix} \hat{a}_1 & 0 & 0 \\ 0 & \hat{a}_2 & 0 \\ 0 & 0 & \hat{a}_3 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix}$$

- Example 2:

- $a_1 = a_1^0 + \hat{a}_1 z_1,$
- $a_2 = a_2^0 + \hat{a}_2 z_2,$
- $a_3 = a_3^0 + \hat{a}_3^1 z_1 + \hat{a}_3^2 z_2.$

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} a_1^0 \\ a_2^0 \\ a_3^0 \end{bmatrix} + \begin{bmatrix} \hat{a}_1 & 0 \\ 0 & \hat{a}_2 \\ \hat{a}_3^1 & \hat{a}_3^2 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$$

如图所示，factor model可以表示Example 2的情况，而 $a + \hat{a}\xi$ 不可以。详细可以参考[2]的讲解。（也是原图出处）

第二个约束可以进一步转化为如下形式：

$$\begin{aligned}
 & (a + P\xi)^T x \leq b \\
 \Leftrightarrow & a^T x + (P\xi)^T x \leq b \quad \text{-- 展开} \\
 \Leftrightarrow & x^T (P\xi) \leq b - a^T x \quad \text{-- 对第二项取转置} \\
 \Leftrightarrow & (P^T x)^T \xi \leq b - a^T x. \\
 \Leftrightarrow & \max_{\xi \in Z} (P^T x)^T \xi \leq b - a^T x.
 \end{aligned}$$

注意这里第二步加的转置，是为了后续子问题写成“参数的转置*变量”的形式，没有其他意义。由于是标量，所以可以直接加。

上一步实际上将约束转化为一个子问题。接下来，接下来我们从最简单的Box uncertainty set开始，展示如何通过对偶进行reformulation。首先介绍Box uncertainty set：

$$1. U = \left\{ A \mid \begin{array}{l} A = a + P \xi \\ \|\xi\|_\infty \leq 1 \end{array} \right\}$$

$$2. \left\{ \begin{array}{l} U = \left\{ A \mid A = a + P \xi, \xi \in Z \right\} \\ Z = \left\{ \xi \mid \|\xi\|_\infty \leq 1 \right\} \end{array} \right.$$

e.g. $U = \left\{ (A_1, A_2) \mid \begin{array}{l} -2 \leq A_1 \leq 4, \\ -1 \leq A_2 \leq 3 \end{array} \right\}$

上图展示了Box不确定集的两种写法和一个例子。其中 U 一般代表不确定项的取值。这一不确定集由范数描述，在这个不确定集下，对该问题取对偶，得：

$$(P) \quad \max_{\xi} (p^T x)^T \xi \quad \longrightarrow \quad (D) \quad \min_{p,q} q - p.$$

s.t.

$$\begin{cases} \xi \geq -1 & \cdots p \\ \xi \leq 1 & \cdots q \end{cases}$$

$$\begin{cases} p \leq 0 \\ q \geq 0 \end{cases}$$

最后将对偶问题带入原问题，得：

$$\max_{x,p,q} C^T x$$

s.t.

$$q - p \leq b - a^T x$$

$$p \leq 0, q \geq 0, x \geq 0$$

这样就转化为了一个可用LP直接求解的问题。

在上述步骤中，我们实际上将最初RC中无穷个约束的问题，转化为子问题中的变量。LP本身就可以解决无穷个变量的情况，也就可以解决原问题了。

事实上，上述过程中还有很多可探究的部分。例如，你真的了解怎么求对偶了吗？是仅仅从公式表格，“原问题变量和对偶问题约束符号一致，原问题约束和对偶问题变量符号相反”，还是通过其他思路？是否真的理解强弱对偶的关系、内容、条件？在更广的含义下，例如，非线性的不确定集中，简单的LP的对偶是否还成立？或者说，需要引入线性规划以外的，更广的范围下的规划方法？这些内容，我们暂且放到下一章介绍。

H3 Adversarial approach

回到epigraph形式，LP其实也有办法解决无限个约束的情况。这种处理RO的方法被称为adversarial approach。他的核心思想是，从一个有限的不确定集的子集 $Z' \in Z$ 开始，接触最优解，再判断该解在整个集合Z的环境下是否可行。如果不可行，再将导致不可行的scenario，或者叫 ξ 的可能的取值，加入到 Z' 中。详细介绍见[1]。这种方法其实类似于branch-and-cut的思想。但总体来说，reformulation依然是更常见的解法。

参考资料：

[1] Bram L. Gorissen, İhsan Yanıkoglu, Dick den Hertog. A practical guide to robust optimization. Omega.

[2] 章宇. 鲁棒优化及其在车辆路径问题中的应用简介。 https://www.bilibili.com/video/BV1tK4y1e7fQ?spm_id_from=333.337.search-card.all.click

H2 3. 对偶定理与常见不确定集

前面我们介绍了box不确定集下RO的reformulation，主要利用了对偶定理。在这一章，我们先介绍对偶定理，再介绍更多的不确定集。

H3 对偶定理

我本人上运筹学的时候，老师似乎没有深入解释过对偶问题的构造过程及对偶定理在其中的作用。我们以上次的盒子不确定集转化的过程作为例子。

$$(P) \quad \begin{aligned} & \max_{\xi} (\mathbf{p}^T \mathbf{x})^T \cdot \xi \\ \text{s.t. } & \begin{cases} \xi \geq -1 & \cdots \mathbf{p} \\ \xi \leq 1 & \cdots \mathbf{q} \end{cases} \end{aligned} \quad \longrightarrow \quad (D) \quad \begin{aligned} & \min_{\mathbf{p}, \mathbf{q}} \mathbf{q}^T \mathbf{p} \\ \text{s.t. } & \begin{cases} \mathbf{p} + \mathbf{q} = (\mathbf{p}^T \mathbf{x})^T \\ \mathbf{p} \leq 0 \\ \mathbf{q} \geq 0 \end{cases} \end{aligned}$$

对原问题做如下处理：

$$\max_{\xi} \min_{\substack{\mathbf{p} \geq 0 \\ \mathbf{q} \leq 0}} L(\xi, \mathbf{p}, \mathbf{q}) = (\mathbf{p}^T \mathbf{x})^T \cdot \xi + \mathbf{p}^T (\xi + 1) + \mathbf{q}^T (\xi - 1) \quad \text{--- (1)}$$

$$\min_{\substack{\mathbf{p} \geq 0 \\ \mathbf{q} \leq 0}} L(\xi, \mathbf{p}, \mathbf{q}) = \begin{cases} (\mathbf{p}^T \mathbf{x})^T \cdot \xi, & \text{if } \xi + 1 \geq 0 \text{ & } \xi - 1 \leq 0 \\ -\infty, & \text{else.} \end{cases} \quad \text{--- (2)}$$

$$\therefore (P) \Leftrightarrow \max_{\xi} \min_{\substack{\mathbf{p} \geq 0 \\ \mathbf{q} \leq 0}} L(\xi, \mathbf{p}, \mathbf{q}) \quad \text{--- --- (3)}$$

构造对偶问题的第一步是构造拉格朗日松弛的形式。在步骤 (1) 中，我们通过加入约束的惩罚项构造了一个两阶段的问题。在步骤 (2) 中，我们只关注内部的min问题。当约束被违背时， p 可以取一个很大的值， q 可以取很小的值，导致结果为负无穷，问题无解。因此该问题等价于原问题 (P)。

$$\min_{\substack{p \geq 0 \\ q \leq 0}} \max_{\xi} L(\xi, p, q)$$

$$\max_{\xi} L(\xi, p, q) = (p^T x + p + q)^T \cdot \xi + p^T 1 + q^T (-1) \quad \text{--- (4)}$$

$$= \begin{cases} p - q & , \text{ if } (p^T x + p + q) = 0 \\ +\infty & , \text{ else.} \end{cases} \quad \text{--- (5)}$$

我们考虑交换max和min之后得到的新问题。还是先考虑内部的max。提出和ξ项，依然可以得到类似上文 (2) 的结论。我们发现，该问题等价于 (D)。这样就构造了对偶问题。

$$\min_{\substack{p \geq 0 \\ q \leq 0}} \max_{\xi} L(\xi, p, q) = \begin{array}{l} \min_{\substack{\xi, p, q \\ \text{s.t.}}} p - q \\ (p^T x + p + q) = 0 \\ p \geq 0, q \leq 0. \end{array} \quad \text{--- (6)}$$

注意，在这个过程中 x 只是参数，不是决策变量。同时，(6) 右侧与上图 (D) 的符号不一样，如果在 (1) 中改成减去惩罚项就变成一样的了。

现在的问题是，max-min和min-max是否完全等价。这就是对偶定理告诉我们的内容。

Theorem1: 弱对偶：(P) 的解 \leq (D) 的解。

$$\text{例证: } \min_{\substack{p \geq 0 \\ q \leq 0}} L(\xi, p, q) \leq L(\xi, p, q) \leq \max_{\xi} L(\xi, p, q)$$

$$\therefore \max_{\xi} \min_{p, q} L \leq L \leq \min_{p, q} \max_{\xi} L.$$

Theorem2: 强对偶：若原问题 (P) 有最优解，则对偶问题存在最优且与其相同。

上文给出了弱对偶定理以及简单的证明。强对偶定理证明比较复杂，在这里暂不介绍（我也没看）。根据这两个定理，就能得出 (P) 与 (D) 等价，即max-min与min-max等价的结论。

个人认为，这部分理解是很重要的。因为在后续分布式鲁棒和一些非线性的不确定性集中，用到的对偶定理将不是简单的线性规划下的对偶定理。因此在这里特别介绍。关于这部分内容，可以参考资料[1]中苏文藻老师关于DRO的课程，也可以参考[2]中第二章的内容，里面有关于对偶理论的详细证明。

H3 不确定集

构造不确定集要从可解释性和可解性两个方面度量，即，既要有概率层面的价值，又要可处理。这一章我们介绍其他经典的不确定集合。

H4 Polyhedral Uncertainty Set

我们先介绍Box不确定集的拓展，Polyhedral Uncertainty set。在凸优化中，多面体Polyhedren可以理解为<一个线性规划问题的可行解区域，是一个由超平面围成的凸集>。超平面就是指取等号的线性约束。多面体不确定集可以写成如下形式：

$$Z := \left\{ \xi \in \mathbb{R}^m \mid w_i^\top \xi \leq v_i, \quad \forall i=1, \dots, s \right\}$$

, where $w_i \in \mathbb{R}^m, v_i \in \mathbb{R}$

可以看出，盒子不确定集其实是他的特例，将 v_i 取 ± 1 ， w_{ij} 取1。我们依然从上述步骤的约束子问题开始处理该不确定集：

1. Lagrange 取对偶。

$$\max_{\xi} \min_{\lambda} L(\xi, \lambda) = (p^\top x)^\top \xi + \lambda^\top (v - w^\top \xi)$$

$\lambda \geq 0.$

$$\begin{aligned} \max_{\xi} L(\xi, \lambda) &= (p^\top x - \lambda^\top w)^\top \xi + \lambda^\top v \\ &= \begin{cases} \lambda^\top v & \text{if } p^\top x - \lambda^\top w = 0 \\ +\infty & \text{else.} \end{cases} \end{aligned}$$

2. 代入原问题。

$$\begin{aligned} \max_{x, \lambda} \quad & C^\top x \\ \text{s.t.} \quad & \lambda^\top v \leq b - a^\top x \\ & p^\top x - \lambda^\top w = 0 \\ & x \geq 0, \lambda \geq 0. \end{aligned}$$

只需要机械化地处理，就能得到结论了。当然，构造对偶的方法不一定要用拉格朗日，也可以直接套用公式。

包括盒子不确定集在内的很多经典不确定集都是用p范数定义的。p范数的简单概念如下：

$$\|z\|_\infty : \text{无穷范数} = \max_j |z_j|$$

* 范数 polynomial norm.

$$\|z\|_p = \left(\sum_j |z_j|^p \right)^{1/p}, p \geq 1. \quad \text{一种 "distance" 的度量.}$$

$$\text{e.g., } \|z\|_1 = \sum_j |z_j|, \quad \|z\|_2 = \sqrt{\sum_j z_j^2}, \quad \|z\|_\infty = \max_j |z_j|$$

盒子不确定集使用了无穷范数，我们自然想知道，由1范数和2范数确定的不确定集，是否存在比较好的性质？

H4 Budgeted Uncertainty Set

Budgeted uncertainty set由1范数定义。具体如下：

$$1. Z(P) = \left\{ \xi \in \mathbb{R}^m \mid \|\xi\|_\infty \leq 1, \|\xi\|_1 \leq P \right\}$$

$$2. Z(P) = \left\{ \xi \in \mathbb{R}^m \mid \exists s \in \mathbb{R}^m, -s \leq \xi_i \leq s_i, \forall i=1, \dots, m, s \in \mathbb{R}, \sum_i s_i \leq P \right\}$$

图中展示了两种等价的表现形式。我们依然按套路解决：

$$(a + P\xi)^T x \leq b$$

$$\Leftrightarrow \max_{\xi \in Z(P)} (P^T x)^T \xi \leq b - a^T x. \quad \begin{matrix} \xrightarrow{\text{构造}} & \max_{\substack{\xi, s \\ \text{s.t.}}} (P^T x)^T \cdot \xi \\ & \xi \geq -s \quad \cdots (p) \\ & \xi \leq s \quad \cdots (q) \\ & s \leq 1 \quad \cdots (\tau) \\ & \sum_i s_i \leq P \quad \cdots (\lambda) \end{matrix}$$

1. Lagrange 取对偶.

$$\begin{aligned} \max_{\xi, s \leq 1} \min_{\substack{p \geq 0 \\ q \leq 0 \\ r \leq 0 \\ \lambda \leq 0}} L(\xi, p, q, r, \lambda) &= (p^T x)^T \cdot \xi + p^T (\xi + s) + q^T (\xi - s) \\ &\quad + r^T (s - 1) + \lambda \cdot (I^T s - P) \\ \max_{\xi, s \leq 1} L(\xi, p, q, r, \lambda) &= (p^T x + p + q)^T \xi + (p - q + r + \lambda)^T s - \lambda \cdot P - r^T I \\ &= \begin{cases} -\lambda \cdot P - r^T I, & \text{if } p^T x + p + q = 0, p - q + r + \lambda = 0. \\ +\infty, & \text{else.} \end{cases} \end{aligned}$$

2. 代入原问题.

$$\begin{aligned} \max_{x, p, q, r, \lambda} & C^T x \\ \text{s.t.} & -\lambda \cdot P - r^T I \leq b - a^T x \\ & p^T x + p + q = 0 \\ & p - q + r + \lambda = 0 \\ & x \geq 0, p \geq 0, q \leq 0, r \leq 0, \lambda \leq 0 \end{aligned}$$

提一嘴，提出这个不确定集的论文The price of robustness是很有影响力的文章。（但我还没仔细读过）

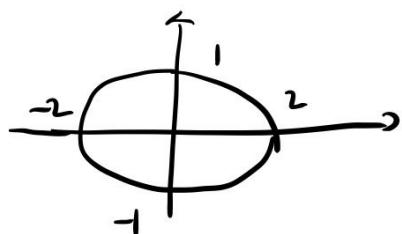
H4 Ellipsoidal Uncertainty Set

椭圆不确定集由2范数定义：

$$\begin{aligned} 1. \quad U(P) &= \left\{ A \mid (A - a)^T \cdot \Sigma \cdot (A - a) \leq P \right\} \\ 2. \quad U(P) &= \left\{ A \mid \| \Sigma^{1/2} \cdot (A - a) \|_2 \leq \sqrt{P} \right\} \\ 3. \quad \begin{cases} U(P) = \left\{ A \mid \exists \xi \in Z(P); A = a + \sum \xi \cdot \xi \right\} \\ Z(P) = \left\{ \xi \mid \| \xi \|_2 \leq \sqrt{P} \right\} \end{cases} \end{aligned}$$

where $\Sigma \in \mathbb{R}^m \times \mathbb{R}^m$, 为正定矩阵.

$$U = \left\{ (A_1, A_2) \mid \frac{A_1^2}{4} + A_2^2 \leq 1 \right\}$$



这里列举了集中表示方法和一个例子。从例子中可以看出其二维图形是一个椭圆。由线性代数中二次型的知识可知，二次形式都可以表示为 $x^T Y x$ 的形式。其中 Y 为对称矩阵。如果 Y 为正定矩阵，则这个值永远大于等于0。所以，这些椭圆都可以表示为3中的factor model。注意这里的求和符号代表一个矩阵。因为可以调整矩阵参数，所以 Z 中的 Γ 可以取1。

回到之前的问题，可以处理如下：

$$\begin{aligned} & \min_x \quad C^T x \\ \text{s.t.} \quad & A(\{\zeta\})^T x \leq b, \quad \forall \zeta \in Z. \quad \dots \quad \textcircled{1} \\ & x \geq 0, \end{aligned}$$

$$\begin{aligned} \textcircled{1} \Leftrightarrow & (a + \sum_{j=1}^k \zeta_j \cdot \zeta_j) ^T x \leq b, \quad \forall \zeta \in Z. \\ \Leftrightarrow & \max_{\|\zeta\|_h \leq 1} (\sum_{j=1}^k x^T \zeta_j)^T \zeta \leq b - a^T x \\ \max_{\zeta} & (\sum_{j=1}^k x^T \zeta_j)^T \zeta \\ \text{s.t.} \quad & \|\zeta\|_h \leq 1 \end{aligned}$$

我们发现，约束对应的子问题变成了一个非线性规划，里面包含二次、根号。这个问题也可以用拉格朗日松弛之后求导解决，但在这里我们用另一个方法：

$$\begin{aligned} (\sum_{j=1}^k x^T \zeta_j)^T \zeta & \leq \|\sum_{j=1}^k x^T \zeta_j\|_2 \cdot \|\zeta\|_h \quad \dots \text{Cauchy-Schwarz.} \\ & \leq \|\sum_{j=1}^k x^T \zeta_j\|_h \end{aligned}$$

$$\textcircled{1} \Leftrightarrow (\sum_{j=1}^k x^T \zeta_j)^T \zeta \leq b - a^T x.$$

也就是我们高中常用的柯西不等式。（实际上盒子不确定集也可以很快得出结论）但是，即使如此，新的约束依然是非线性的情况。熟悉凸优化的同学可能知道，这其实是一个conic programming可解的问题，是一个二阶锥。在后续DRO的内容中，我们再详细介绍这个内容。

H4 CVaR Uncertainty Set

最后介绍一个CVaR的不确定集，并详细介绍求解过程，作为总结。

VaR和CVaR是常用于资产组合的概念。前几章已经提到了一些不确定性问题处理范式的固有缺陷，比如，无法反应风险情况。这两个概念就是从目标函数的角度限制风险（相关处理方法还可以在许多论文中看到）。

VaR指的是在给定的置信区间下，最小化可能发生的最大的损失。例如95%的概率下，最大可能损失1k，剩下的5%则可能超出1k。

CVaR是指，当预期损失高于预期最差情况时的回报率。例如5%的VaR是-25%，指的是95%以外的（最差情况）5%的概率，回报率低于-25%。

简单来说，VaR衡量了95%里的损失，而CVaR衡量了5%里面的损失，因此，CVaR的损失大于VaR。这部分背景可以自己查一查。

VaR会作为后续内容的例子出现。在经典RO中，CVaR的不确定集更常见。定义如下：

$$\mathcal{Z} := \left\{ z \in \mathbb{R}^n \mid \exists \theta \in \mathbb{R}^K, z = \sum_{k=1}^K \theta_k \cdot \tilde{z}_k, \theta \geq 0, \sum_{k=1}^K \theta_k = 1, \theta \leq \frac{1}{K\alpha} \right\}$$

对不同的 \tilde{z}_k 加权 小于给定的上界.

其中 \tilde{z} 是一个给定的向量，可以理解为不同的 scenario (场景)， α 是给定的置信度。相当于对任何可能的场景分配概率，概率和不超过一个上界。

回顾最初的问题，RO可以写成如下的形式：

$$\begin{aligned} & \max_{x \in \mathbb{R}^n} \quad \tilde{r}^T \cdot x \\ & \text{s.t.} \quad \sum x_i = 100\% \quad \cdots \cdots \quad (1) \\ & \quad 0 \leq x \leq 1 \end{aligned}$$

这个问题是一个资产组合选择的例子，其中 r 为不确定的回报率， x 为投资选择。将它写出 robust counterpart 形式（问题2），用 factor model 表示：

$$\begin{aligned} & \text{step 1} \quad \max_{x \in \mathbb{R}^n} \min_{z \in \mathcal{Z}} \quad (\mu + p_z)^T \cdot x \\ & \text{s.t.} \quad \sum x_i = 100\% \quad \cdots \cdots \quad (2) \\ & \quad 0 \leq x \leq 1 \end{aligned}$$

进一步写出 epigraph 的 RC 形式（问题3），并处理 robust linear constraints：

step 2.

$$\max_x t$$

$$t \leq (\mu + p \cdot z)^T \cdot x, \quad \forall z \in \mathbb{Z} \quad \text{--- (1)}$$

$$\sum x_i \leq 1 \quad \text{--- (2)}$$

$$0 \leq x_i \leq 1.$$

step 3.

$$\text{Q} \Leftrightarrow t \leq \mu^T x + (p^T z)^T x, \quad \forall z \in \mathbb{Z}$$

$$\Leftrightarrow (p^T x)^T z \geq t - \mu^T x, \quad \forall z \in \mathbb{Z}$$

$$\Leftrightarrow \min_{z \in \mathbb{Z}} (p^T x)^T z \geq t - \mu^T x.$$

写出对应子问题。这里可以替换 z ，简化问题（问题4）。然后用得出对偶问题（问题5）：

step 4

$$\begin{aligned} & \min_{\tilde{z}, \theta} (p^T x)^T \cdot \tilde{z} \\ \text{s.t. } & \tilde{z} = \tilde{z}^T \theta, \quad \left. \right\} \Rightarrow \min_{\theta} (p^T x)^T \cdot \tilde{z}^T \theta \\ & \theta \leq \frac{1}{K\alpha}, \quad \text{s.t. } \theta \leq \frac{1}{K\alpha} \quad \text{--- (4)} \\ & \theta^T \cdot 1 = 1, \quad \theta \geq 0, \\ & \theta \geq 0. \end{aligned}$$

step 5.

$$\begin{aligned} & \min_{\theta \geq 0} \max_{\lambda \geq 0} L(\theta, \lambda, \rho) = (p^T x)^T \tilde{z}^T \theta + \lambda^T (\theta - \frac{1}{K\alpha}) + \rho \cdot (1^T \theta - 1) \\ & \min_{\theta \geq 0} L(\theta, \lambda, \rho) = (\tilde{z} \cdot p^T x + \lambda^T + \rho)^T \theta - \lambda^T \frac{1}{K\alpha} - \rho \end{aligned}$$

最后带入主问题（问题6）：

step 6.

$$\begin{aligned} & \max_{x, \lambda, \rho} t \\ & \tilde{z} \cdot p^T x + \lambda^T + \rho \geq 0, \\ & t - \mu^T x + \lambda^T \cdot \frac{1}{K\alpha} + \rho \leq 0 \quad \text{--- (5)} \\ & \sum x_i \leq 1 \\ & 0 \leq x_i \leq 1, \lambda \geq 0. \end{aligned}$$

就能成功解决了。这部分内容可以参考[3]。

这一章介绍了一些常用的不确定集，并引出了RO和椎规划的关系。下一章可能还会写一点不确定集的内容，然后就会进入分布式鲁棒优化（DRO）的内容。

参考资料：

[1] 苏文藻. 分布鲁棒优化简介. “数据驱动的优化理论、算法与应用”暑期学校. https://www.bilibili.com/video/BV1bM4y1L7R7?spm_id_from=333.337.search-card.all.click

[2] Erick Delage. Quantitative Risk Management Using Robust Optimization Lecture Note. https://tin.hec.ca/pages/erick.delage/MATH80624_LectureNotes.pdf

[3] 运小筹. 鲁棒优化 | 利用rome求解鲁棒优化简单案例：以CVaR不确定性集为例.

https://mp.weixin.qq.com/s?_biz=MzI3NTI2NzcxOA==&mid=2247486373&idx=1&sn=7bead25d63d72555bd8ac9e4fb9dff&cksm=eb06224cdc71ab5a5c83637b0393dbc34338bbff8bea5154834d44f6e5de025f29e386821567&scene=178&cur_album_id=2003418663093059585#rd

H2 4. [选读] 不确定集参数的选择

其实在各个不确定集中，都有一些参数用于调整不确定集的保守性。可以说，box不确定集是最保守的，他的图形就是一个盒子；相比之下，budgeted、ellipsoidal不确定集就分别砍掉了长方体的某个部分、圆整了边界。选择这些不确定集就是为了在保守性上做一定取舍。

参数调整肯定是基于已有数据的。这部分内容在[1]中被称为Data-driven Uncertainty Set Design，就是从这个角度考虑。因为这部分内容比较多应用于portfolio，有些内容也比较深入，所以暂时只是简单带过，作为番外内容。

H3 不保守的不确定集

上一章提到了，VaR指的是在给定的置信区间下，最小化可能发生的最大的损失。他的定义和之前第一章提到的CCP很类似。这里给一个VaR的数学定义：

Definition 3.1. : Mathematically, for a revenue calculated as $c(z)^T x + d(z)$, it can be defined as

$$\text{VaR}_{1-\epsilon}(c(Z)^T x + d(Z)) := - \sup\{y \in \mathbb{R} \mid \mathbb{P}(c(Z)^T x + d(Z) \geq y) \geq 1 - \epsilon\}$$

Hence, when maximizing an uncertain revenue one can formulate a VaR minimization problem as follows:

$$\begin{aligned} & \underset{x,y}{\text{minimize}} && -y \\ & \text{subject to} && \mathbb{P}(c(Z)^T x + d(Z) \geq y) \geq 1 - \epsilon \\ & && x \in \mathcal{X}. \end{aligned}$$

<我猜测，这里用 $-y$ 是因为VaR的定义是损失，而这里的 y 如果理解成收益，把下面问题的 $\min -y$ 换成 $\max y$ 也是等价的>。可以看到，下面的问题形式就是CCP。如第一章所说，要将CCP转化为线性约束，就需要对已知的密度/累计函数做反函数处理，总体来说比较难算。因此，通过RO处理这个问题，会是一个有趣的方法。虽然，按我们之前的说法，所谓的“robust”应该让 ϵ 等于0。但在很多情况下，这确实过于保守。下面的定理给出一个将CCP转化为RO的方法：

Theorem 4.1:

对给定的 $\epsilon > 0$, Z 为某随机向量, \mathcal{Z} 为随机向量的集合, 使得

$$P(Z \in \mathcal{Z}) \geq 1 - \epsilon.$$

则对 $\forall x$ 满足

$$a(z)^T x \leq b(z), \forall z \in \mathcal{Z}. \quad \text{--- ①}$$

也会满足

$$P(a(z)^T x \leq b(z)) \geq 1 - \epsilon.$$

Proof: 对 $\forall x$, 若满足条件 ①, 则有

$$P(a(z)^T x \leq b(z)) \geq P(Z \in \mathcal{Z}) \geq 1 - \epsilon.$$

这个定理将约束的概率转化为了不确定集的概率。证明也很简单, 如果后面的情况下前面必定满足, 所以前面的概率大于后面。

H3 参数选择的例子

假设有一组2000-2009年10只股票的月度回报数据。我们希望解决以VaR为优化目标的CCP问题:

$$\begin{aligned} & \underset{x,y}{\text{minimize}} && -y \\ & \text{subject to} && \mathbb{P}(r^T x \geq y) \geq 1 - \epsilon \\ & && \sum_{i=1}^n x_i = 1 \\ & && x \geq 0, \end{aligned}$$

其中 $\epsilon = 5\%$ 。

根据上面的定理, 我们可以考虑去掉CCP约束, 考虑这样的RO问题:

$$\begin{aligned} & \underset{x,y}{\text{minimize}} && -y \\ & \text{subject to} && r^T x \geq y, \forall r \in \mathcal{U} \\ & && \sum_{i=1}^n x_i = 1 \\ & && x \geq 0, \end{aligned}$$

假设使用ellipsoidal不确定集：

$$\mathcal{U}(r_0, \gamma) := \{r \in \mathbb{R}^{10} \mid \|r - r_0\|_2 \leq \gamma\}$$

接下来要做的，就是确定 $\mathcal{U}(r_0, \gamma)$ ，使 $\mathbb{P}(Z \in \mathcal{Z}) \geq 1 - \epsilon$ 。

很明显， r_0 肯定要用期望 $E[r]$ 近似，这里肯定用样本期望近似真实期望；剩下的 γ ，可以用一下步骤确定：

1. 对所有历史数据 r_k ，计算 $\gamma_k := \|r_k - r_0\|_2$ 。
2. 对 $\{\gamma_k\}$ 排序，取第 $(1 - \epsilon) \times K$ 大的那个，作为估计值。

事实上，在一定情况下，可以选择很合适的参数使得RO和CCP等价。上述RO问题的reformulation形式为：

$$\begin{aligned} & \underset{x,y}{\text{minimize}} && -y \\ & \text{subject to} && r_0^T x - \gamma \|x\|_2 \geq y \\ & && \sum_{i=1}^n x_i = 1 \\ & && x \geq 0. \end{aligned}$$

如果 γ 满足正态分布，那么CCP约束 $\mathbb{P}(Z \in \mathcal{Z}) \geq 1 - \epsilon$ 可以转化为：

$$r_0^T x - \Phi^{-1}(1 - \epsilon)\sigma \|x\|_2 \geq y,$$

$\phi^{-1}(x)$ 是累计概率函数的反函数， σ 是标准差。只要令 $\gamma = \phi^{-1}(1 - \epsilon)\sigma$ ，两个约束就完全等价了。

以上内容只是参数选择的一个例子，由于和RO具体理论关系小一些，需要用的同学还是更多参考文献或者详细阅读[1]的内容。

参考资料：

[1] Erick Delage. Quantitative Risk Management Using Robust Optimization Lecture Note. https://tin.hec.ca/pages/erick.delage/MATH80624_LectureNotes.pdf

H2 5. 适应性鲁棒优化建模与NP-hard介绍

从这一章开始，将进入适应性鲁棒优化（adaptive/adjust robust optimization, ARO）的学习。ARO实际上处理的是多阶段决策问题，这和我们第一章提到的stochastic programming with recourse有相似之处，因为他也是一个两阶段问题。因此，联系前后的内容理解，可能会更有帮助。本章内容基本参考自[1]。

H3 一个例子

本章将从一个类似multi-stage newsvendor的问题出发，展示使用鲁棒优化建模多阶段问题时遇到的困难，以此引出ARO。首先思考一个如下的多阶段问题：

$$\begin{aligned}
& \min_{\{x_t\}_{t=1}^T, \{y_t\}_{t=1}^T} && \sum_{t=1}^T c_t \cdot x_t + h_t (y_{t+1})^+ + b_t (-y_{t+1})^+ \\
& \text{s.t.} && y_{t+1} = y_t + x_t - d_t, \quad \forall t=1, \dots, T \quad \text{--- ①} \\
& && x_t \geq 0, \quad \forall t.
\end{aligned}$$

这里每个值都是标量。其中 x_t 是每个阶段的进货量， y_t 为正数表示每个阶段库存量，为负数表示缺货量， d_t 表示需求量， h_t 表示库存成本， b_t 表示缺货成本。 c_t 表示进货成本。很明显问题是一个 multi-stage newsvendor，如果不太清楚背景可以看一看相关资料。其中 $(\cdot)^+$ 表示 max 算子。

max 算子显然不是线性的。可以线性化为如下问题 2：

$$\begin{aligned}
& \min_{\{x_t\}_{t=1}^T, \{y_t\}_{t=1}^T} && \sum_{t=1}^T c_t \cdot x_t + h_t \cdot s_t^+ + b_t \cdot s_t^- \\
& \text{s.t.} && y_{t+1} = y_t + x_t - d_t, \quad \forall t=1, \dots, T \quad \text{--- ②} \\
& && s_t^+ \geq y_{t+1} \quad \forall t \\
& && s_t^- \geq -y_{t+1} \quad \forall t \\
& && s_t^+ \geq 0, s_t^- \geq 0, x_t \geq 0, \quad \forall t.
\end{aligned}$$

我们关注的是问题的鲁棒形式，也就是 robust counterpart。一个很 naive 的方法是直接对每个参数加上随机项 z ：

$$\begin{aligned}
& \min_{\{x_t\}_{t=1}^T, \{y_t\}_{t=1}^T} && \sup_{z \in \mathbb{Z}} \sum_{t=1}^T c_t(z) \cdot x_t + h_t(z) \cdot s_t^+ + b_t(z) \cdot s_t^- \\
& \text{s.t.} && y_{t+1} = y_t + x_t - d_t(z), \quad \forall t=1, \dots, T, \forall z \in \mathbb{Z} \\
& && s_t^+ \geq y_{t+1} \quad \forall t \\
& && s_t^- \geq -y_{t+1} \quad \forall t \\
& && s_t^+ \geq 0, s_t^- \geq 0, x_t \geq 0, \quad \forall t. \quad \text{--- ③}
\end{aligned}$$

这里的 sup 是上确界，可以简单理解为 max，不明白区别的同学可以自己查查。

这里出现一个问题。第一个约束中，除了 $d_t(z)$ 项都是确定值。也就是说，这一组约束对不同的 z 都要满足，相当于一个确定值等于很多个不同的 z 。这显然是错的，因此我们在写成 RC 之前去掉 y_t 项，然后得到：

$$\begin{aligned}
 & \min_{\{x_t\}_{t=1}^T} \quad \sup_{z \in Z} \sum_{t=1}^T c_t(z) \cdot x_t + h_t(z) \cdot s_t^+ + b_t(z) \cdot s_t^- \\
 \text{s.t.} \quad & s_t^+ \geq y_1 + \sum_{t'=1}^t [x_{t'} - d_{t'}(z)], \quad \forall t, \forall z \in Z \\
 & s_t^- \geq -y_1 - \sum_{t'=1}^t [x_{t'} - d_{t'}(z)], \quad \forall t, \forall z \in Z \\
 & s_t^+ \geq 0, s_t^- \geq 0, x_t \geq 0, \quad \forall t.
 \end{aligned}
 \tag{④}$$

这样就初步得到了一个可能正确的RC形式。但是，这样处理有两大问题：

1. 这里的 x_t 是predefined的，是offline的。也就是说，我们在第一阶段开始之前就决定了每个阶段的订货量。实际上，在全局最优的角度看，每个订货量应该根据之前的实际销售情况决定，也就是说，需要实时地，online地调整。
2. s_t^+, s_t^- 这样建模是错误的。正确的形式应该是：

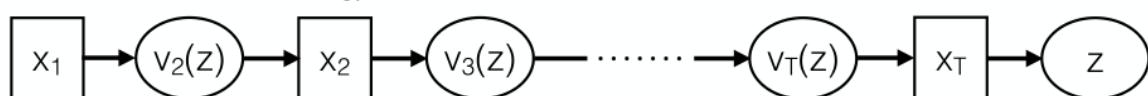
$$\begin{aligned}
 & \min_{\{x_t\}_{t=1}^T} \quad \sup_{z \in Z} \sum_{t=1}^T c_t(z) \cdot x_t + h_t(z) \cdot (y_1 + \sum_{t'=1}^t [x_{t'} - d_{t'}(z)])^+ \\
 & \quad + b_t(z) \cdot (-y_1 - \sum_{t'=1}^t [x_{t'} - d_{t'}(z)])^+ \\
 \text{s.t.} \quad & x_t \geq 0, \quad \forall t.
 \end{aligned}
 \tag{⑤}$$

这两个问题的区别在于，在问题4中，对每一个 z^* 求最优解时， s_t^+, s_t^- 需要对每个 z 都满足约束，而在问题5中，他们只需要对 z^* 满足就够了。因此，问题应该写成多阶段问题，这里令 $T = 2$ ，以两阶段为例：

$$\begin{aligned}
 & \min_{x_1} \quad \sup_{z \in Z} \quad c_1(z) x_1 + h(x_1, d_1) \\
 \text{s.t.} \quad & x_1 \geq 0 \\
 h(x_1, d_1) := & \min s_1^+ + s_1^- \\
 \text{s.t.} \quad & s_1^+ \geq x_1 - d_1, \\
 & s_1^- \geq -x_1 + d_1, \\
 & s_1^+, s_1^- \geq 0.
 \end{aligned}
 \tag{⑥}$$

H3 建模方式

这样我们大概就明白了建模的方法。回顾整个多阶段问题的处理过程，如图：



当我们做出 x_1 决策时，时间轴往后走，一个随机性事件 $v_2(z)$ 发生了，我们可以观察到他的实现值，然后做出下一个决策 x_2 的具体值，以此类推。在多阶段问题里，我们通常把第一阶段直接做出的决策变量成为here-and-now decision，而把后续做出的决策成为wait-and-see decision。这也和我们前面讲的一样，每个 $x_t, t \geq 2$ 时online决策的，根据前面的 $v_t(z)$ 调整，所以需要wait and see。

总结对multi-stage问题的RC，下面问题成为Multi-Stage ARC (adaptive robust counterpart):

$$\begin{aligned} & \min_{x_1, \{x_t(\cdot)\}_{t=2}^T} \inf_{z \in Z} c(z)^T \cdot x_1 + \sum_{t=2}^T \left[c_t(z)^T \cdot x_t(v_t(z)) + d_t(z) \right] \\ \text{s.t. } & a_{j1}(z)^T \cdot x_1 + \sum_{t=2}^T a_{jt}(z)^T \cdot x_t(v_t(z)) \leq b_j(z), \forall z \in Z, \forall j \end{aligned} \quad \text{--- (7)}$$

但是，这里又隐含了一个条件：时间的一致性。在这个问题中，不确定集 Z 是没有更新的。但在某些时候，这是需要更新的。我们举一个小例子。

- 假设你经营一家咖啡馆，从早上7点经营到11点，需要预估咖啡的进货量。你有理由依据经验假设7-9点有100人来，9-11点有100人来，7-11点最多300人来，或许因为咖啡馆的坐位不超过300个（这就类似budgeted uncertainty set）。那么，如果7-9点真的有200人来了，依据假设9-11点不超过100人，7-11点一共不超过300人，你依然不需要更改自己的期望。
- 但如果你打算在9点把咖啡铺子搬到另一个地方（假设是移动摊位）。你依然可以假设一共不超过200杯，因为两个地方是独立的， $100+100=200$ 。但是，如果你发现7-9点出售了150杯咖啡，按总共200的思路，9-11点应该预估50；但是，你有理由调整预期到100，因为两地是独立的。那么，原先的预期就需要调整。

对于上面的例子来说，第一个例子中不确定集（也就是预期销量）是前后一致的，不随时间而改变。但第二个例子中，不确定集是会改变的。这时我们就需要调整问题。这里依然用一个两阶段问题为例：

$$\begin{aligned} & \underset{x_1, x_2(\cdot), s(\cdot)}{\text{minimize}} \quad \sup_{d \in \mathcal{U}} x_1 + 4x_2(d_1) + 10s(d) \\ & \text{subject to} \quad s(d) \geq 0, \forall d \in \mathcal{U} \\ & \quad s(d) \geq d_1 + d_2 - x_1 - x_2(d_1), \forall d \in \mathcal{U} \\ & \quad x_1 \geq 0 \\ & \quad x_2(d_1) \in \underset{x \geq 0}{\text{argmin}} \max_{d'_2 \in \tilde{\mathcal{U}}_2(d_1)} c_2 x + 10(d_1 + d'_2 - x_1 - x)^+, \forall d_1 \in \mathcal{U}_1, \end{aligned}$$

在这里， x_1 是第一阶段， x_2 是第二阶段，随机因素是 $d = [d_1, d_2]^T$ 。这与问题7最大的区别就是最后一个约束，不确定集 $\tilde{\mathcal{U}}_2(d_1)$ 是随 d_1 改变的集合，在第二阶段决策中， d'_2 需要属于这个不确定集，但是从总体角度， d 依然要属于 \mathcal{U} 。

上述问题很难解决。目前我们不考虑这类问题。事实上，问题7已经很难解决了，我们用一个定理结束这一章：

Theorem 5.1: 即使 $v_t(z) = z$, Z 为 polyhedral,
问题⑦ 依然是 NP-hard 的.

proof: 可以 reduce 到 3-SAT 问题. 暂略.

这就足以说明问题求解的困难了。

H3 NP-hard介绍

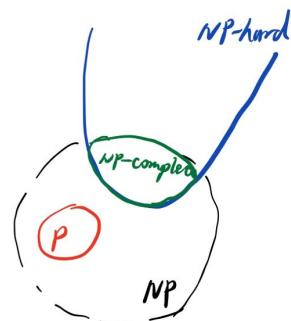
关于NP-hard, 放一点很久之前的笔记:

P: 问题可以在多项式时间内解决

NP: 可以在多项式时间内判断解是否正确(最优)

NP-hard: 不可以在多项式时间内解决

NP-complete: $NP \wedge NP\text{-hard}$.



NP-hard / complete 的证明:

使用 reduction 的方法.

reduction: 将两个问题通过多项式算法建立联系.

TSP: Hamiltonian cycle.

对不完全图 $G'(V', E')$, 存在 Hamiltonian cycle. ($|V'|=N$)

↓ reduction: 1. 为 E' 赋值 1 的权重
2. 对 V' 建立完全图 $G(V, E)$. 对 $\forall e \in E, e \notin E'$, 赋值 2 的权重 } 多项式时间 $O(N^2)$

在完全图 $G(V, E)$ 中找到 TSP, 则 $c(TSP) = N$.

因此, 求解 Hamiltonian cycle 可以通过求解 TSP 得到.

Hamiltonian cycle 是 NP-hard, 故 TSP NP-hard.

第一个 NP-hard: SAT

布尔可满足性问题: 指定一个布尔表达式, 问是否存在一个赋值使结果为 true.

所有算法都可以编码为布尔表达式.

其实, ARO并没有非常通用的解法。下一章中, 我们介绍目前ARO在一些特殊情况下的解法。

参考资料:

[1] Erick Delage. Quantitative Risk Management Using Robust Optimization Lecture Note. https://tin.hec.ca/pages/erick.delage/MATH80624_LectureNotes.pdf

[2] Ihsan Yaniko glu, Bram L. Gorissen, Dick den Hertog. A survey of adjustable robust optimization. European Journal of Operational Research.

H2 6. 适应性鲁棒优化求解

上一章我们通过从多阶段问题中引出了ARO的概念:

$$\begin{aligned}
 & \min_{\substack{x_1, \{x_t(\cdot)\}_{t=2}^T}} \inf_{z \in Z} C(z)^T \cdot x_1 + \sum_{t=2}^T [C_t(z)^T \cdot x_t(v_t(z)) + d_t(z)] \\
 \text{s.t. } & a_{j,t}(z)^T \cdot x_t \leq b_j(z), \quad \forall z \in Z, \quad \forall j
 \end{aligned} \quad \text{--- ①}$$

并说明了他的求解难度。在这一章里我们介绍一些可以求解的情况。

H3 情况一：等价于经典RO

在一定条件下，多阶段问题与单阶段问题等价，这时就不需要求解ARO问题。当最优解 $x_t(v_t(z)) = x_t(v_t(z'))$, $\forall z, z' \in Z$ 时， x_t 就与 v_t 无关了，只需要求解如下的RO问题：

$$\begin{aligned}
 & \min_{\{x_t(\cdot)\}_{t=1}^T} \inf_{z \in Z} \sum_{t=1}^T [C_t(z)^T \cdot x_t + d_t(z)] \\
 \text{s.t. } & \sum_{t=1}^T a_{j,t}(z)^T \cdot x_t \leq b_j(z), \quad \forall z \in Z, \quad \forall j
 \end{aligned} \quad \text{--- ②}$$

有如下定理：

Theorem 6.1：如果存在一个 Z 的划分，使 $Z = [z_0, z_1, \dots, z_J]^T$ ，满足如下条件，则问题①等价于问题②：

1. 存在非空凸紧集 $Z_j \subset R^{d_{m,z_j}}$ 使得

$$Z := Z_0 \times Z_1 \times \dots \times Z_J = \{z \mid \exists z_i \in Z_i, \forall i, z = [z_0, z_1, \dots, z_J]^T\}$$

2. 存在 $M > 0$ ，对任意 $\{x_t(\cdot)\}_{t=1}^T$ 的可行解，都满足

$$\|x_t(v_t(z))\|_\infty \leq M, \quad \text{对 } \forall z \text{ 成立.}$$

3. 目标函数是关于 z_0 的函数，i.e., $C_t(z) = C_t(z_0)$, $d_t(z) = d_t(z_0)$

4. 每个约束 j 的函数仅与 z_j 有关，i.e., $a_{j,t}(z) = a_{j,t}(z_j)$, $b_{j,t}(z) = b_{j,t}(z_j)$

条件1, 3, 4将每个约束与随机项 z 的对应情况分开了，条件2表示每个 x_t 都有一个上界。这个定理并不是最完整的定理，与之相关的论文似乎还有更新。定理的证明暂时挖坑。

H3 情况二：可精确求解

[3]提出了一种行列生成的方法，并在文章里介绍了之前多用的benders分解算法。这一节里，我们结合[1]和[3]的内容介绍这两种方法。

[3]解决的其实是一个非常简单的问题：

$$\min_y c^T y + \max_{u \in U} \min_{x \in F(y, u)} b^T x$$

s.t. $Ay \geq d$, $y \in S_y$. —— (*).

其中, $F(y, u) = \{x \in S_x : Gx \geq h - Ey - Mu\}$.

$S_x \subseteq \mathbb{R}_+^m$, $S_y \subseteq \mathbb{R}_+^n$.

这里的符号和前面略有区别。在这里 y 是第一阶段决策变量, x 是第二阶段决策变量, u 是随机变量, $F(y, u)$ 是第二阶段问题。可以看到, 第二阶段问你题和第一阶段决策相关的项 Gx 是一个线性函数, 和不确定因素 u 相关的项也只是线性项 Mu , 所以非常非常 special。此外, 还要求不确定集 U 为 polyhedron 或 discrete。在后面我们会拓展这些限制。

使用 benders 分解的方法很粗暴。首先需要满足一个 relative complete recourse。在后面我们再解释这个内容, 大概可以理解为, 针对属于该集合的第一阶段的 x , 第二阶段 y 永远有可行解。因此在 benders 中不需要加入 feasibility cut, 只需要加 optimality cut。看到第二阶段子问题是 min, 但不确定项是 max, 所以对 min 问题取对偶, 合并两个 max, 得到子问题:

$$\begin{aligned} Q(y) &= \max_{u \in U} \min_x b^T x \\ \text{s.t. } &Gx \geq h - Ey - Mu. \quad (\pi) \end{aligned} \xrightarrow{\text{dual}} \begin{aligned} Q(y) &= \max_{u, \pi} (h - Ey - Mu)^T \pi \\ \text{s.t. } &G^T \pi \leq b. \quad — SP1. \\ &u \in U \\ &\pi \geq 0. \end{aligned}$$

虽然 SP1 不是一个线性规划问题 (注意目标函数的 $u^T \pi$) , 但还是有很多办法求解, 这里不展开解释。因此, 对于一个主问题的最优解 y_k^* , 可以解上述子问题 SP1 得到最优子问题解 (u_k^*, π_k^*) 。构造 optimality cut: $\eta \geq (h - Ey - Mu_k^*)^T \pi_k^*$ 加入主问题:

$$\begin{aligned} \min_{y, \eta} \quad &c^T y + \eta \\ \text{s.t. } &Ay \geq d. \\ &y \geq (h - Ey - Mu_k^*)^T \pi_k^*, \quad \forall i \quad — MP1. \\ &y \in S_y. \end{aligned}$$

加入后重新解主问题 MP1, 得到最优解 $(y_{k+1}^*, \eta_{k+1}^*)$ 。很明显 $c^T y_k^* + Q(y_k^*)$ 会是一个 upper bound, 因为第一阶段和第二阶段是分开优化的, 局部最优之和肯定不如全局最优; 而 $c^T y_{k+1}^* + \eta_{k+1}^*$ 则是一个 lower bound, 因为没有考虑第二阶段完整的极点, 或者说, 没有考虑所有的约束。因此, 如果两个解相等, 则得到最优解, 算法终止。这一段也算是对 benders 的复习了。

容易理解，迭代求解子问题的次数和第二阶段的极点数量有关。记 p, q 分别为不确定集 Z 和 x 的定义域的极点数（若 U 离散，则为离散节点数），算法的迭代次数的复杂度应该是 $O(pq)$ 。

第二种方法是行列生成。我们通过[1]中更general的描述来介绍。

针对以下的两阶段鲁棒优化形式，可以用上述方法和以下方法求解：

$$\begin{aligned} \max_{x, y} \quad & \inf_{z \in Z} c_1(z)^T x + c_2^T y(z) + d(z) \\ \text{s.t.} \quad & a_{j1}(z)^T x + a_{j2}^T y \leq b_j(z), \quad \forall z \in Z, \forall j \\ & x \in \mathcal{X}. \end{aligned} \quad \text{--- (3)}$$

可以看到，我们去掉了第二阶段决策变量 y 的系数的不确定扰动 z 。这种情况被称为fixed recourse。注意recourse这个词，虽然中文没有很好的翻译，但是我们在第一章介绍SP的时候遇到过。当初我们用VRP做了一个例子，recourse代表的是一种在第一阶段决策给定后的，针对第二阶段的补救措施。在SP中，我们将recourse导致的期望成本(expected cost of recourse)加入目标函数。在这里， a_{j2}^T, c_2^T 实际上也可以理解为recourse，是用于描述第二阶段的参数，只不过描述成了线性规划形式。在RO中，我们不考虑期望，而是考虑worst case。这个例子很好的展现了SO和RO的不同。

回到问题3，问题3的fixed recourse简化了第二阶段的不确定因素。此外，我们还规定 $x \in \mathcal{X}$ 。这个集合的意思是，针对属于该集合的第一阶段的 x ，第二阶段 y 永远有可行解。所以集合可以表示为如下形式：

$$\mathcal{X} \subseteq \{x \in \mathbb{R}^n \mid \forall z \in Z, \exists y \in \mathbb{R}^n, a_{j1}(z)^T x + a_{j2}^T y \leq b_j(z), \forall j = 1, \dots, J\}.$$

这被称为relatively complete recourse。与之相对的还有一个complete recourse，<意思是，在约束中， y 之外的参数取任何值，都存在可行的 y 。>所以complete recourse是强于relatively complete recourse的。

问题可以写作如下的两阶段形式：

$$(TSARC) \quad \underset{x \in \mathcal{X}}{\text{maximize}} \quad \inf_{z \in Z} h(x, z),$$

where we have that

$$\begin{aligned} h(x, z) := \max_y \quad & c_1(z)^T x + c_2^T y + d(z) \\ \text{subject to} \quad & a_{j1}(z)^T x + a_{j2}^T y \leq b_j(z), \quad \forall j = 1, \dots, J. \end{aligned}$$

注意，第一阶段没有目标函数和约束（其实加入模型后也很容易求解，我不知道为什么不加）。

在上述情况下，当 Z 为由有限个节点构成的凸包（注意，这拓展了[3]的范围）时，可以有可行的解法。有如下定理：

Theorem 6.2: 当 Z 为有限个节点组成的凸包时, i.e.,
 $Z := \text{Convex Hull}(\{\bar{z}_1, \dots, \bar{z}_k\})$.

问题① 等价于如下问题:

$$\begin{aligned} & \max_{x, \{y_k\}_{k=1}^K} \min_k G(\bar{z}_k)^T x + c^T y_k + d(\bar{z}_k) \\ \text{s.t. } & a_{j_1}(\bar{z}_k)^T x + a_{j_2}^T y_k \leq b_j(\bar{z}_k), \forall k, \forall j. \\ & x \in X. \end{aligned} \quad \text{--- (4)}$$

熟悉单纯形法的同学可能已经明白了, 如果不确定集 Z 有凸包的性质, 那就可以利用类似单纯形法枚举 vertex(凸包的定点, 忘了中文怎么叫) 来得到最优解。定理的证明很简单, 首先有引理: 凸函数的最优值必定在凸包的定点上。然后利用对偶定理, 将问题3关于第二阶段的子问题求对偶。由于 relatively complete recourse 可知, 强对偶成立。再根据引理, 就能得到问题4等价于问题3。具体证明参考[1]。

但是, 凸包的节点数是关于维数的级数, 枚举所有节点是非常困难的。单纯形法通过一个特殊的规则枚举, 从而缩短计算时间。针对这个问题, 可以使用一种行列生成的算法求解。下面介绍该方法的思路。

考虑枚举 Z 的子集 Z' 中的 K' 个定点, 称其为问题5。记这个解为 \hat{s} 。很明显这个解会是最优解 s^* 的上界, 因为 \min 的定义域范围更小。那我们就可以利用行列生成的方式, 不断地检验 \hat{s} 是否为最优解, 然后拓展 Z' 的集合。记 $h(x, z)$ 为给定 x 和 z 下以 y 为目标函数的子问题。算法步骤如下:

1. 从任意 $\hat{x} \in X$ 开始;
 2. 得到 $\hat{z} := \arg\min_{z \in Z} h(\hat{x}, z)$, 并构建 $Z' := \{\hat{z}\}$ 。
 3. 求解针对不确定集 Z' 的问题5, 得到最优的 \hat{x} 和 \hat{s} 。
 4. 得到 $\hat{z} := \arg\min_{z \in Z} h(\hat{x}, z)$, 如果 $h(\hat{x}, \hat{z}) = \hat{s}$, 则算法终止, 否则将 \hat{z} 加入集合 Z' , 返回3。
- 注意, 这里加入 Z' 在模型中的体现是分别在目标函数和约束中加入一项。

第4步的内在逻辑是: $s^* = \max_{x \in X} \min_{z \in Z} h(x, z) \geq \min_{z \in Z} h(\hat{x}, z) = \hat{s} \geq s^*$ 。所以如果 $h(\hat{x}, \hat{z}) = \hat{s}$, \hat{s} 必定等于 s^* 。相反, 如果 $h(\hat{x}, \hat{z}) < \hat{s}$, 说明这个 \hat{z} 是会导致约束被违背的, 需要加入集合中。这一段和前面benders那一段分析类似。

行列生成的方法迭代次数最多是不确定集 U 的极点数量, 也就是 $O(p)$, 要小于benders。

我们已经知道问题5是相对好解的, 那么影响算法效率的核心部分就是第4步中求解 \hat{z} 的过程。这个问题可以写成问题6:

$$\begin{aligned} & \inf_{z \in Z} h(\hat{x}, z) := \inf_{z \in Z} \max_y G(z)^T \hat{x} + c^T y + d(z) \\ \text{s.t. } & a_{j_1}(z)^T \hat{x} + a_{j_2}^T y \leq b_j(z), \forall j. \\ & z \in Z \end{aligned} \quad \text{--- (6)}$$

这个问题里有两个变量, z, y 。要注意这个问题和RO的区别。在RO中, z 不是决策变量, 而是要对所有 z 满足条件。这里的问题只是一个双决策变量的问题。我们用KKT定理处理内层的max问题。KKT定理也是OR, OM的同学应该很熟悉的了(虽然我也忘了), 推广了拉格朗日乘数法到不等式形式, 可以处理凸函数的约束。简单回顾一下:

Lemma . (K-K-T condition):

对 $\max_x f(x)$ 若 $f(x)$ 为凸且可微, $g_0(x)$ 凸可微,
st. $g_j(x) \leq 0, \forall j$.

且满足强对偶, 则其最优解 \tilde{x} 和变量 $\tilde{\lambda}$ 满足:

$$\begin{aligned} g_0(\tilde{x}) &\leq 0 \\ \tilde{\lambda} &\geq 0 \\ \tilde{\lambda}_j \cdot g_j(\tilde{x}) &= 0, \forall j \\ \nabla f(\tilde{x}) &= \sum_j \tilde{\lambda}_j \nabla g_j(\tilde{x}) \end{aligned}$$

其中 ∇ 是梯度符号。

所以, 问题6就可以reformulation成如下问题7, 进而求解:

$$\begin{aligned} \min_{z \in Z} h(x, z) &:= \min_{z \in Z, y, \lambda, u} c(z)^T x + c^T y + d(z) \\ \text{st. } a_{j1}(z)^T x + a_{j2}^T y &\leq b_j(z), \forall j \\ \lambda &\geq 0 \\ \lambda_j &\leq M u_j, \forall j \\ b_j(z) - a_{j1}(z)^T x - a_{j2}^T y &\leq m(u_j), \forall j \\ c = \sum_j a_{j2} \lambda_j & \\ u_j &\in \{0, 1\}. \end{aligned} \quad \text{--- (7)}$$

这就是一个容易求解的问题了。

其实我有一点不太明白, 为什么这里要用KKT呢? 为什么不考虑对内层做对偶? 在线性约束下, KKT应该是比较慢的方法。

H3 情况三：可近似求解

ARO最经典的做法应该是第三种。前文的两种解法各有其局限性, 第一种适用范围不广, 第二种求解速度不快。因此有了第三种方法, 采用一些近似手段。这种近似主要基于affine decision rule。下面我们将具体介绍这一内容。

H4 Affine decision rule

所谓的affine decision rule(简称ADR), 就是在问题1中, 将 $x_t(v_t(z))$ 关于 z 的形式固定为仿射函数的形式。实际上这个函数形式应该是不固定的, 这也是处理的难点。在第2种情况中, 我们用枚举节点的形式对 z 赋值, 避免了不同复杂函数形式带来的问题。在ADR中, 通过处理成仿射函数形式, 将其转化为线性约束。但是, 转化为线性约束的过程本身就缩小了搜索范围, 因此这个解是原先最优解的下界。其实, 我们只需要转化为凸函数形式, 但可以证明, 凸函数形式只能写成仿射函数形式。问题1变成解决以下 Affinely Adjustable Robust Counterpart (AARC) 问题:

$$\begin{aligned} & \max_{\{x_t\}_{t=1}^T, \{x_t\}_{t=2}^T} \inf_{z \in Z} C(z)^T x_1 + \sum_{t=2}^T C_t(z)^T (x_t + x_t v_t(z)) + d(z) \\ \text{s.t. } & a_{j1}(z)^T x_1 + \sum_{t=2}^T a_{jt}(z)^T (x_t + x_t v_t(z)) \leq b_j(z), \forall z, \forall j. \end{aligned} \quad \text{--- (8)}$$

此外，为了得到robust linear constraints，我们还需要使用fixed recourse，不然 x 的参数上还会有 z 项，导致二次项出现；并令 $v_t(z)$ 本身为线性形式。有如下定理：

Theorem 6.3: 若一个 α -多阶段AARC(问题)满足如下两点，则
对应AARC模型可以表示成 Robust linear constraints.(问题9)

1. 对 $\forall t=2, \dots, T, j=1, \dots, J$, $C_t(z)$ 和 $a_{jt}(z)$ 为常数。(fixed recourse)
2. 对 $\forall t=2, \dots, T$, $v_t(\cdot)$ 可以表示为 $v_t(z) = V_t z$.

$$\begin{aligned} & \max_{\{x_t\}_{t=1}^T, \{x_t\}_{t=2}^T} \inf_{z \in Z} C^T x_1 + \sum_{t=2}^T C_t^T (x_t + x_t V_t z) + d(z) \\ \text{s.t. } & a_{j1}(z)^T x_1 + \sum_{t=2}^T a_{jt}(z)^T (x_t + x_t V_t z) \leq b_j(z), \forall z, \forall j. \end{aligned} \quad \text{--- (9)}$$

问题9可以被 reformulation 为有限维、有限个约束的问题。

感觉挺明显的，就不需要证明了吧。

H4 Piecewise affine decision rule

尽量考虑更多可处理的 $x_t(v_t(z))$ 的形式总是好事，因为这样能拓展我们的可解范围。分段放射函数是有必要考虑的。可以考虑以下例子。

在库存管理inventory中（管科出身同学唯一懂的东西），一种常用的库存管理策略是make to stock（另一种是make to order）。意思是，给出一个best的库存下界，每次检查库存量时，若低于下界，则补货，否则不补货。那么补货量 x_t 就会被描述为如下形式：

$$x_t(d_{t+1}, y_{t-1}, \theta_t) := \max(0, \theta_t - y_{t-1} + d_{t+1})$$

其中 θ_t 为 t 时期的最佳库存量， d_t 为需求， y_t 为实际库存量。很明显这是一个分段函数。尤其是当 t 变大时，将较小的 t 带入后续函数，会分成更多个分段。因此， x_t 可以描述为以下形式，并进一步简化为后续形式：

$$x_t(v_t(z)) := \bar{x}_t + \bar{x}_t v_t(z) + \sum_{k=1}^V [\theta_{tk}^+ \max(0, v_{tk}(z)) + \theta_{tk}^- \max(0, -v_{tk}(z))]$$

$$\Rightarrow x_t(v_t(z), \bar{x}_t, x_t^+, x_t^-) := \bar{x}_t + \bar{x}_t^+ V_t z^+ + \bar{x}_t^- V_t z^-,$$

$$\text{where } z_i^+ := \max(0, z_i), z_i^- := \max(0, -z_i), z_i := z_i^+ - z_i^-.$$

这时AARC中的 Z' 就拓展为三部分了。这个不确定集被称为lifted uncertainty set，对应的AARC也被称为Lifted AARC (LAARC)：

$$\begin{aligned}
 & \max_{\{x_t\}_{t=1}^T, \{x_t^+, x_t^-\}_{t=1}^T} \inf_{(z, z^+, z^-) \in Z'} C(z)^T x_i + \sum_{t=2}^T C_t(z)^T (x_t + x_t^+ V_t z^+ + x_t^- V_t z^-) + d(z) \\
 \text{s.t. } & a_{j1}(z)^T x_i + \sum_{t=2}^T a_{jt}(z)^T (x_t + x_t^+ V_t z^+ + x_t^- V_t z^-) \leq b_j(z), \\
 & (z, z^+, z^-) \in Z', b_j.
 \end{aligned}
 \quad \text{—— (10).}$$

但是这个不确定集 Z' 并不是凸的。为了解决这个问题，我们考虑fixed recourse下的情况，记为问题11；并令 $Z'' := ConvexHull(Z')$ ，记为问题12。有如下定理：

Theorem 6.4: 如果问题12中的凸包 Z'' 可以用有限个
线性约束描述，则该LAARC可以被解决。

Proof: 首先， $\because Z'' \supseteq Z'$ ， \therefore 问题12的解必然在问题11中可行。

其次， \because 每个约束的参数对 (z, z^+, z^-) 为线性。

\therefore 问题11中可行解在问题12中的最差情况的实现，
对每个约束都会出现在 Z'' 的顶点处，也就出现在 Z' 中。

类似可证明目标函数等价。

<其实我也不是特别懂第二步>。详细请看[1]。总而言之，求解目标变为问题12。 Z' 可以表述为如下形式：

$$\begin{aligned}
 & \max_{z, z^+, z^-, u^+, u^-} a_{j1}^T x_i + \sum_{t=2}^T a_{jt}^T (x_t + x_t^+ V_t z^+ + x_t^- V_t z^-) - b_j(z) \\
 \text{s.t. } & z = z^+ + z^- \\
 & 0 \leq z^+ \leq M \cdot u^+ \\
 & 0 \leq z^- \leq -M \cdot u^- \\
 & u^+ + u^- = 1 \\
 & z \in Z \\
 & u^+, u^- \in \{0, 1\}^m
 \end{aligned}
 \quad \text{—— (11).}$$

这个表示的好处在于，我们可以进一步将验证问题12中的约束转化为如下MILP问题：

Proposition 6.1: 若 $\exists M > 0$ 使 $Z \subseteq [-M, M]^m$ ，则 Z' 可表示为：

$$Z' = \left\{ (z, z^+, z^-) \in \mathbb{R}^{3m} \mid \exists u^+, u^- \in \{0, 1\}^m, \begin{array}{l} z \in Z \\ z = z^+ - z^- \\ 0 \leq z^+ \leq M \cdot u^+ \\ 0 \leq z^- \leq -M \cdot u^- \\ u^+ + u^- = 1 \end{array} \right\}$$

最后，我们可以给出在 Z 为budgeted不确定集下一个很好处理的性质：

Proposition 6.2. 若 Z 为 budgeted uncertainty set, 则有:

$$\text{Convex Hull}(Z') = \left\{ (\bar{z}, \bar{z}^+, \bar{z}^-) \in \mathbb{R}^{3m} \mid \begin{array}{l} \bar{z}^+ + \bar{z}^- \leq 1 \\ \sum_{i=1}^m (\bar{z}_i^+ + \bar{z}_i^-) \leq P \\ \bar{z} = \bar{z}^+ - \bar{z}^- \\ \bar{z}^+ \geq 0 \\ \bar{z}^- \geq 0 \end{array} \right\}$$

Proof: 将上述右边的集合记为 Z'_2 . 在 Z 为 budgeted 的条件下, 根据 Proposition 6.1, 有:

$$Z' = \left\{ (\bar{z}, \bar{z}^+, \bar{z}^-) \in \mathbb{R}^{3m} \mid \exists u^+, u^- \in \{0, 1\}^m, \begin{array}{l} \bar{z}^+ + \bar{z}^- \leq 1 \\ \sum_{i=1}^m (\bar{z}_i^+ + \bar{z}_i^-) \leq P \\ \bar{z} = \bar{z}^+ - \bar{z}^- \\ 0 \leq \bar{z}^+ \leq u^+ \\ 0 \leq \bar{z}^- \leq u^- \\ u^+ + u^- = 1 \end{array} \right\}$$

容易看出 $Z' \subseteq Z'_2$, 因为 Z' 有多重约束.

从而有 $\text{Convex Hull}(Z') \subseteq Z'_2$, 因为 Z'_2 也是 convex 的.

问题转化为证明 $Z'_2 \subseteq \text{Convex Hull}(Z')$.

事实上, 可以将凸包定义为: 由所覆盖的点对应的支撑的超平面所描述的半空间的交集 (有点绕).

则, 若 Z'_2 被包含在所有 Z' 对应的半空间中, 则有

$Z'_2 \subseteq \text{Convex Hull}(Z')$.

对任意 (c, c^+, c^-) 定义的方向, 可以通过以下方式判断出 $(\bar{z}, \bar{z}^+, \bar{z}^-)$ 对应支撑的半空间:

$$c^T \bar{z} + c^{+T} \bar{z}^+ + c^{-T} \bar{z}^- \leq \beta,$$

$$\beta = \sup_{(\bar{z}, \bar{z}^+, \bar{z}^-) \in Z'} c^T \bar{z} + c^{+T} \bar{z}^+ + c^{-T} \bar{z}^-.$$

为了说明 Z_2' 在所有半空间内，需要证明，对 $\forall C \in \mathbb{R}^m$, $C^T, C \in \mathbb{R}^m$ ，有：

$$\max_{(z, z^+, z^-) \in Z_2'} C^T z + C^{+T} z^+ + C^{-T} z^- = \max_{(z, z^+, z^-) \in \text{ConvexHull}(Z_2')} C^T z + C^{+T} z^+ + C^{-T} z^-$$

即证明： $\max_{z, z^+, z^-} C^T z + C^{+T} z^+ + C^{-T} z^-$

s.t. $\begin{aligned} z^+ + z^- &\leq 1 \\ \sum_{i=1}^m (z_i^+ + z_i^-) &\leq P \\ z &= z^+ - z^- \\ z^+, z^- &\geq 0 \end{aligned}$ --- (P1)

的解，是等于零子

$$\max_{z, z^+, z^-, u^+, u^-} C^T z + C^{+T} z^+ + C^{-T} z^-$$

s.t. $\begin{aligned} z^+ + z^- &\leq 1 \\ \sum_{i=1}^m (z_i^+ + z_i^-) &\leq P \\ z &= z^+ - z^- \\ 0 \leq z^+ &\leq u^+ \\ 0 \leq z^- &\leq u^- \\ u^+ + u^- &= 1 \\ u^+, u^- &\in [0, 1]^m \end{aligned}$ --- (P2)

的解。

然而，对 P1 的最优解 (z, z^+, z^-) ，总有对应的解

$$u_j^+ := \begin{cases} 1, & \text{if } z_j^+ > 0 \\ 0, & \text{else} \end{cases}$$

$$u_j^- := 1 - u_j^+$$

$$z_j^+ := (z_j^+ + z_j^-) \cdot u_j^+$$

$$z_j^- := (z_j^+ + z_j^-) \cdot u_j^-$$

$$z_j := z_j^+ + z_j^-$$

为 P2 的可行解，因此结论成立。#.

证明不算复杂，但非常有趣。

H4 Dual decision rules

以上内容出自[1]。在[2]中，还记录了一些近似求下界的方法。其中，对情况二的精确解算法中，如果对 Z 非凸包的情况，如果依旧节选一些点进行操作，也能得到近似下界。当然，可以选择更有效的选择有限子集的替代方法。此外，[2]还提到了一种Dual decision rules。不过还没看，暂且留个坑。

TODO

下一章将进入分布式鲁棒优化DRO的学习，这是最有趣的一部分。在此之中还会穿插conic programming的知识。可惜的是，博主下周三要期中考试，得先停更复习。呜呼，万恶的考试！

参考资料：

[1] Erick Delage. Quantitative Risk Management Using Robust Optimization Lecture Note. <https://tinyurl.com/yx2qzv4n>

[2] Ihsan Yaniko glu, Bram L. Gorissen, Dick den Hertog. A survey of adjustable robust optimization. European Journal of Operational Research.

[3] Bo Zeng, Long Zhao. Solving two-stage robust optimization problems using a column-and-constraint generation method. Operations Research Letters.

H2 7. 分布式鲁棒优化和锥规划

虽然还没开始考试，但我懒得复习了。

从这章开始，将进入RO里面比较难的一块内容，分布式鲁棒优化（DRO）。

H3 分布式鲁棒优化概述

我们在前面已经了解过基础的RO了。他解决的是一个这样的约束：

$$\begin{aligned} \min_{x} \quad & \max_{\xi \in Z} C(\xi)^T x \\ \text{s.t.} \quad & A(\xi)^T x \leq b(\xi), \forall \xi \in Z. \quad \text{--- ①.} \end{aligned}$$

问题1是一个RO的RC形式。当我们不知道随机变量 ξ 属于什么分布的时候，给定一个 ξ 可能出现的不确定集 Z ，保证最优解对其中任意的 ξ 鲁棒。

DRO尝试解决类似如下的问题：

$$\begin{aligned} \min_x \quad & \sup_{P \in \mathcal{P}} \mathbb{E}_{\xi \sim P} [C(\xi)^T x] \\ \inf_{P \in \mathcal{P}} \quad & P \left[A(\xi)^T x \leq b(\xi) \right] \geq 1 - \delta. \quad \text{--- ②.} \end{aligned}$$

问题2是一个DRO问题的例子（不是普遍形式）。有很多内容值得说明。首先，目标函数变成了期望值，这与我们第一章将的SP类似。其次，约束中出现了概率的形式，这与CCP类似。此外，概率 ξ 不再属于某个不确定集 Z ，而是服从针对某个分布 P ，这个分布 P 又属于一个分布的集合 \mathcal{P} 。注意，这里的sup是下确界，对于max，不了解的同学可以自行百度。至于这里为什么要用sup而不是max，说实话我也是很不清楚，包括后续的测度、Borel σ -代数等实变函数概念我也没学过，只能暂时忽略这些细节内容。

这样看DRO的特点就很明显了。RO中的随机变量是完全未知分布的，SP中随机变量是完全已知分布的，DRO则介于两者之间，给定一个被称为模糊集(ambiguity set)的随机分布集合 \mathcal{P} ，使得问题的解对其中所有分布都保持鲁棒。针对每个特定的分布，处理方式与传统的SP或者CCP大致类似。我们之前提到过，RO的缺点是过于保守，SP则过于不保守；DRO则处于二者之间。

依照我们第二到四章关于RO的内容，针对这种范式，需要解决的问题很明显是以下两个：1. 如何解决这类问题？2. 如何构造模糊集 \mathcal{P} ？接下来，我们从一个熟悉的问题引入。

H3 VaR的DRO例子（上）

在第四章中我们已经提到过VaR的概念了。大概意思是在一个给定的置信度下，最大的可能损失值。以此为目标的问题可以写出如下形式：

$$\min_{w, \gamma} \gamma$$

$$\text{s.t. } P(\gamma \leq -w^T x) \leq \varepsilon, \quad \text{--- ①.}$$

$$\sum w_i = 1,$$

$$w \geq 0.$$

这里 γ 是VaR值， w 可以理解为一组股票的投资组合也是决策变量， x 为损失值。与之前的讨论相同， x 一般为随机值。在RO中，我们构建以下RC：

$$\min_{w, \gamma} \gamma$$

$$\text{s.t. } \max_{x \in U} P(\gamma \leq -w^T x) \leq \varepsilon, \quad \text{--- ④}$$

$$\sum w_i = 1,$$

$$w \geq 0.$$

得到问题4的RC后，我们一般通过股票的历史数据，进一步确定不确定集 U 的形式（见第三、四章）。这个问题通过DRO建模，则为以下形式：

$$\begin{aligned}
 & \inf_{w, \gamma} \gamma \\
 \text{s.t. } & \sup_{P \in \mathcal{P}} P(\gamma \leq -w^T x) \leq \epsilon, \quad \text{--- ⑤.} \\
 & \sum w_i = 1, \\
 & P \in \mathcal{P}
 \end{aligned}$$

很明显，构建模糊集 \mathcal{P} 也需要通过历史数据，或者说，data-driven。假设我们计算得到样本的均值 \bar{x} 和协方差矩阵 Γ ，以此来估计真实值的均值和方差，我们可以通过这两个最经典的数字特征来描述模糊集 \mathcal{P} 中的随机变量，规定他们的均值和方差等于已知值。注意 Γ 应该是一个半正定矩阵。

在RO中，我们接下来会对问题4中的约束1构造子问题，再通过reformulation得到可解的形式。对问题5，我们依然对约束1构造子问题。约束1的左侧等价于如下子问题：

$$\begin{aligned}
 & \sup \int_{\mathbb{R}^n} \mathbb{1}_{\{\gamma \leq -w^T x\}}(x) dP(x). \\
 \text{s.t. } & \int_{\mathbb{R}^n} dP(x) = 1, \quad \text{--- ⑥.} \\
 & \int_{\mathbb{R}^n} x \cdot dP(x) = \bar{x}, \\
 & \int_{\mathbb{R}^n} (x - \bar{x})(x - \bar{x})^T dP(x) = \bar{\Gamma}, \\
 & P \geq 0.
 \end{aligned}$$

DRO和RO的一个区别就在于，因为要描述随机变量的数字特征，所以需要使用积分形式。应该很好理解几个约束对应的内容。在这个问题中，决策变量不再是某个参数，而是 P 的分布函数。
<这里最好不要写成密度函数，似乎不是所有概率测度都有密度函数。>为了方便理解，可以写出 P 为离散分布时的情况：

$$\sup \sum_{j: y \leq -w^T x} p_j$$

$$\begin{aligned} & \text{s.t. } \sum_j p_j = 1, \\ & \sum_j p_j \cdot \hat{x}_j = \bar{x}, \\ & \sum_j p_j \cdot (\hat{x} - \bar{x})(\hat{x} - \bar{x})^T = P, \\ & p_j \geq 0, \forall j. \end{aligned} \quad \text{--- ⑦.}$$

其中 \hat{x} 为不同情况下的 x 的实现值。这样就很好理解了。而且，这个问题其实已经是一个有限维度、有限约束的LP问题，可以直接求解。这种模型实际上可以称为一种Scenario-based model。scenario这个词经常出现在不确定性问题的文献中，基本上用来表示不同的可能情景。

当然，问题7不是我们的重点。在RO中，求解的难点是不确定集为无限集的情况，也就是问题6。此时决策变量是无限维的。在处理这个问题前，可以先对问题6做一点简化：

$$\left. \begin{array}{l} \int_{\mathbb{R}^n} dP(x) = 1, \\ \int_{\mathbb{R}^n} x \cdot dP(x) = \bar{x}, \\ \int_{\mathbb{R}^n} (x - \bar{x})(x - \bar{x})^T dP(x) = P. \end{array} \right\} \quad \begin{array}{l} \int_{\mathbb{R}^n} \begin{bmatrix} x \\ 1 \end{bmatrix} \begin{bmatrix} x \\ 1 \end{bmatrix}^T dP(x) = \begin{bmatrix} P + \bar{x}\bar{x}^T & \bar{x} \\ \bar{x}^T & 1 \end{bmatrix} \\ \begin{bmatrix} \text{"} & x \\ x\bar{x}^T & x \\ \bar{x}^T & 1 \end{bmatrix} = \sum \in S_{++}^{n+1} \end{array}$$

应该容易看懂这个变换，并不影响求解，只是为了简化问题。注意这里的 \sum 是一个 $n + 1$ 维的正定矩阵， S_{++}^{n+1} 为 $n + 1$ 维正定矩阵的集合。所以，问题变为如下形式：

$$\begin{aligned} & \sup \int_{\mathbb{R}^n} \mathbb{1}_{\{y \leq -w^T x\}}(x) dP(x). \\ & \int_{\mathbb{R}^n} \begin{bmatrix} x \\ 1 \end{bmatrix} \begin{bmatrix} x \\ 1 \end{bmatrix}^T dP(x) = \sum, \quad \text{--- ⑧.} \\ & P \geq 0. \end{aligned}$$

如何对问题8进行reformulation呢？答案依旧是对偶定理。但是，这里需要用到的不再是简单的LP的对偶，而是锥规划的对偶。我们暂时把问题8搁置，简单介绍一下锥规划。

H3 锥线性规划基础

我第一次了解锥(cone)的概念是在在benders分解里。简单来说，锥是一个满足一个条件的集合：

锥 Cone:

存在 $C \subset R^n$, 对 $\forall x \in C, \lambda \geq 0, \lambda x \in C$.



上图右边是一个二维平面内的例子。这个定义和例子能直观地说明锥的定义，但是我们接触锥的最终目的是为了求解对偶，而不是单纯了解一个概念。因此我们重新从LP的视角认识锥。

实际上，对偶理论以及线性规划的整个框架都基于不等式符号 \leq ，基于这个符号可以证明Farkas定理和强对偶理论，进而得到我们已知的体系。我们认识中的 \leq 是定义在实数集上、从纯代数的角度理解的。 $x \geq y$ 可以进一步写成 $x - y \geq 0$ 。这又可以理解为， $x - y \in \mathbb{R}$ 。换句话说，这个定义的本质是实数集 \mathbb{R} 。而我们的理论体系，Farkas定理和强对偶理论等定理的证明，则是依据下图左边所示的五个性质：

$$x \geq y \rightarrow x - y \geq 0, \quad \text{IR.}$$

$$K = \left\{ x \in \mathbb{R}^n : x_i \geq 0, \forall i \right\} = \mathbb{R}_+^n$$

$$x \geq 0 \Leftrightarrow x \in \mathbb{R}_+^n.$$

(1) Reflexivity. $u \geq u, \forall u$.

(2) Anti-Symmetry. $\begin{cases} u \geq v \\ v \geq u \end{cases} \Rightarrow u = v$.

K 为 pointed cone, if:

(3) Transitivity. $\begin{cases} u \geq v \\ v \geq w \end{cases} \Rightarrow u \geq w$.

a). $K \neq \emptyset$, and
 $u, v \in K \Rightarrow u + v \in K$

(4) Homogeneity. $\begin{cases} u \geq v \\ \alpha \geq 0 \end{cases} \Rightarrow \alpha \cdot u \geq \alpha \cdot v$.
 $\forall \alpha \in \mathbb{R}$.

b). $u \in K, \alpha \geq 0 \Rightarrow \alpha \cdot u \in K$

(5) Additivity. $\begin{cases} u \geq v \\ w \geq z \end{cases} \Rightarrow u + w \geq v + z$.

c). $u - u \in K \Rightarrow u = 0$.

仔细想想，确实是这五个简单的性质构建了整个基于实数的LP的运算体系。可以进一步思考，是否可以赋予 \leq 其他含义，在保持这些性质的基础上，进一步拓展LP的范围？这就是右边的内容，从几何意义上理解。

在右图中，我们将集合 K 定义为了 n 维非负实数集 \mathbb{R}_+^n 。将所有 ≥ 0 改为了 $\in K$ ，所有元素改为向量。如果一个 K 有a, b, c三个性质，称其为pointed cone。这三个性质分别被称为closed under addition, conic, pointedness。性质如同字面意思一样很好理解。<[1]中只说明了 \mathbb{R}_+^n 在几何意义下满足这三个性质，而没有说明左边5个都满足，我不太清楚为什么。也非常容易证明这些都是能满足的。>

根据上面的分析，我们就有机会对 \mathbb{R}_+^n 进一步推导得到类似LP的结论。那么，除了 \mathbb{R}_+^n 和实数集外，还有其他类似性质的集合吗？

答案是肯定的，不然也不会介绍这些。下图是两种符合该条件的锥

(1) Lorentz cone / second-order cone. (洛伦兹锥/二阶锥)

$$Q^{n+1} = \{(t, x) \in \mathbb{R} \times \mathbb{R}^n : t \geq \|x\|_2\}.$$

(2) Semidefinite cone. (半正定锥)

$$S_t^n = \{Y \in S^n : u^T Y u \geq 0, \forall u \in \mathbb{R}^n\}.$$

可以证明, Q^{n+1}, S_t^n 为 pointed cone, 且 $\text{int}(Q^{n+1})$, $\text{int}(S_t^n) \neq \emptyset$.

* $\text{int}(\cdot)$ 表示内部, interior, $\text{int}(k) = \{y \in k : \exists \varepsilon > 0, B(y, \varepsilon) \subseteq k\}$.
 Ball, y 为中心
 ε 为半径的球.

这两个锥都是由很熟悉的两个概念引申出的, 分别由2范数和半正定矩阵定义。三个性质的证明也非常简单:

Q^{n+1}

$$(a). (t, x), (s, y) \in Q^{n+1}.$$

$$\therefore t \geq \|x\|_2, s \geq \|y\|_2.$$

$$\therefore t+s \geq \|x\|_2 + \|y\|_2 \geq \|x+y\|_2 \quad (\text{三角不等式}).$$

$$\therefore (t+s, x+y) \in Q^{n+1}.$$

$$(b). (t, x) \in Q^{n+1}$$

$$\therefore t \geq \|x\|_2$$

$$\therefore \alpha \cdot t \geq \alpha \cdot \|x\|_2 = \|\alpha x\|_2 \quad (\alpha \geq 0).$$

$$\therefore (\alpha \cdot t, \alpha \cdot x) \in Q^{n+1}.$$

$$(c). (t, x) \in Q^{n+1}, (t, -x) \in Q^{n+1}$$

$$\therefore t \geq \|x\|_2, -t \geq \|-x\|_2 = \|x\|_2$$

$$\therefore \|x\|_2 \geq 0, \|x\|_2 \leq t \leq \|x\|_2, \therefore t=0, x=0.$$

S_t^n

$$(a). X, Y \in S_t^n$$

$$\therefore \forall u, u^T X u \geq 0, u^T Y u \geq 0$$

$$\therefore u^T X u + u^T Y u = u^T (X+Y) u \geq 0.$$

$$\therefore X+Y \in S_t^n. \quad (\text{半正定/半定/正定/负定矩阵相加依旧不变})$$

$$(b). \because X \in S_t^n$$

$$\therefore u^T X u \geq 0$$

$$\therefore \alpha \cdot u^T X u \geq 0 \quad (\alpha \geq 0)$$

$$\therefore u^T (\alpha X) u \geq 0$$

$$\therefore \alpha X \in S_t^n.$$

$$(c). X \in S_t^n, -X \in S_t^n$$

$$\therefore u^T X u \geq 0, u^T (-X) u \geq 0$$

$$\therefore 0 \leq u^T X u \leq 0, X=0.$$

这两个锥分别对应了两种不同的 \leq 的体系, 可以理解为, 定义了两种有一定优良性质的集合($Q^{(n+1)}$ 和 S_t^n)和排序(两个集合对应的 \leq), 满足pointed cone的三条性质。下面我们进一步说明满足这三条性质的集合存在类似对偶理论的结论。

结论：令 E 表示欧几里得空间， $\langle \cdot, \cdot \rangle$ 为内积函数。
 令 $K \subseteq E$ 为内部非空的闭 pointed cone.
 (e.g., \mathbb{Q}^{n+1} , S_+^n , \mathbb{R}_+^n).
 主问题和对偶问题分别为：

$(P) \quad \inf_{\substack{x \in K \\ c, a_i \in E}} \langle c, x \rangle$ $\text{s.t. } \langle a_i, x \rangle = b_i, \quad \forall i \quad -(y_i)$	$(D) \quad \sup_{\substack{c \in E \\ y_i \in \mathbb{R}}} \sum_{i=1}^m b_i y_i$ $\text{s.t. } c - \sum y_i a_i \in K^*$
--	--

$y_i, b_i \in \mathbb{R}^m$

上面的结论拓展了LP的对偶到pointed cone的对偶。其中 $\langle \cdot, \cdot \rangle$ 是内积符号，定义如下：

$$\langle A, B \rangle = \text{tr}(AB) = \sum_{i,j} A_{ij} \cdot B_{ij}$$

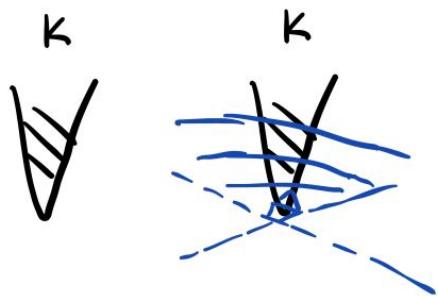
$$\langle a, b \rangle = \text{tr}(a^\top b) = a^\top b.$$

对于向量而言，内积和平时是一样的，但这里的内积是针对 E 的，而不是针对 \mathbb{R}^m ，因此在对偶问题中没有写成内积的形式。

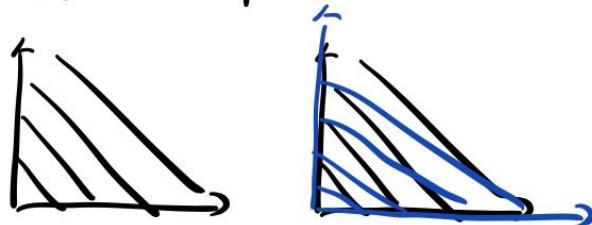
问题D引入了一个新符号 K^* ，称为 K 的对偶锥。其定义如下：

$$K^*(K) = \{ y \in E : \langle x, y \rangle \geq 0, \forall x \in K \}$$

e.g. ①.



$$\textcircled{2}. (\mathbb{R}_+^2)^* = \mathbb{R}_+^2.$$



$$\textcircled{3}. (\mathbb{R}_+^n)^* = \mathbb{R}_+^n, (\mathbb{Q}^{n+1})^* = \mathbb{Q}^{n+1}, \\ (\mathcal{L}_+^n)^* = \mathcal{L}_+^n.$$

这里我们给出了3个例子。第一个例子是二维平面中的一个夹角在90度内的锥。向量 y 与其中某个向量 x 内积非负，就意味着 x 与 y 的夹角小于90度。于是我们找这个锥的边界形成直角的边，在右边用蓝色虚线表示，这之间的区域就是 K 的对偶锥。对于二维非负区域而言，两条边正好构成直角，因此其对偶锥就是自己，成为self-dual。还能进一步证明，前面提到的二阶锥、半正定锥也都是自对偶的。（这个证明不算复杂，但我没怎么看懂，想了解的可以自行搜索）可以进一步拓展结论如下：

Proposition. 7.1. 若 K 为闭 pointed cone, 且其内部非空, 则 K^ 也是.*

到这里我们就基本建立了基于pointed cone的线性规划框架。问题D约束1其实只是说明 y 的仿射函数属于一种类型的锥。如果这种类型的锥是二阶锥，问题就是二阶锥规划，SOCP；如果锥是半正定锥，问题就是半正定锥规划，SDP。回到第三章关于椭圆不确定集最后的子问题：

$$\max_{\xi} (\sum_{i=1}^k \alpha_i^\top x)^\top \cdot \xi$$

$$\text{s.t. } \|\xi\|_2 \leq 1.$$

$$(1, \xi) \in Q^{n+1}$$

其实约束本身也可以写成SCOP的形式。而这些问题，都是锥规划领域研究已久的问题，有现成的求解器和求解手段处理。

但我们的目的还是求对偶，即从问题P到问题D。回顾第三章，求对偶就需要构造拉格朗日松弛。拉格朗日松弛的目的是为了将约束加入目标函数，对违背约束的情况给予惩罚。问题P中约束1的惩罚容易构造，但约束2的 $x \in K$ 又该如何惩罚呢？

$$(P) \quad \inf_{x \in K} \langle c, x \rangle \\ \text{s.t. } \langle a_i, x \rangle = b_i, \quad \forall i \quad -(y_i)$$

$$L(x, y, w) = \langle c, x \rangle + \sum_i y_i \cdot (b_i - \langle a_i, x \rangle) + ? \\ (\text{希望: 若 } x \notin K, ? \rightarrow +\infty). \\ (LP: ? = -w^\top x, w \geq 0)$$

看到括号中LP的情形，大家肯定有所猜测了。迄今为止我们的一起操作都是基于LP的情况，做某种映射，这里也不例外。结论如下：

结论: $? = -\langle w, x \rangle$, $w \in K^*$.

即, 需证明:

若 $x \notin K$, 则 $\exists w \in K^*$, st. $\langle w, x \rangle < 0$.
否则对 $\forall w \in K^*$, $\langle w, x \rangle \geq 0$.

证明:

先证明了引理.

1. K 为凸集.

由于性质(a)和(b),

对 $\forall u \in K$, $0 < \alpha < 1$, $(1-\alpha)u, \alpha u \in K$.
 $\therefore \alpha u + (1-\alpha)v \in K$.

2. 对 $x \notin K$, $\exists w$, 使

$\langle w, x \rangle < \inf_{y \in K} \langle w, y \rangle$.

通过分离定理 (separation theorem) 得.

3. $\langle w, x \rangle < \langle w, o \rangle = 0$.

由 K^* 的定义, $w \in K^*$, 有

$\inf_{y \in K} \langle w, y \rangle = 0$.

由 2. 有 $\langle w, x \rangle < \langle w, o \rangle = 0$.

下面证明结论. 假设 $\exists y \in K$, 使 $\langle w, y \rangle < 0$.

对 $\forall \alpha$, $\alpha \cdot y' \in K$. 则当 α 较大时, 有:

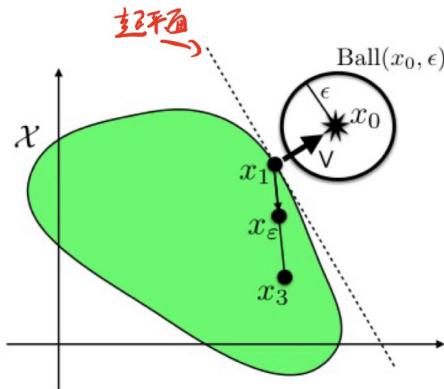
$\langle w, x \rangle < \inf_{y \in K} \langle w, y \rangle = -\infty$.

与矛盾. 所以结论成立.

Theorem 7.0: (strict separating hyperplane theorem)

令 $X \in \mathbb{R}^n$ 为一个闭凸集, $x_0 \notin X$. 则存在超平面参数为 $v \in \mathbb{R}^n$, $b \in \mathbb{R}$,
严格分离 X 与 x_0 . 即:

$$v^T x \leq b, \forall x \in X \quad \text{且} \quad v^T x_0 > b.$$



证明思路.

构造出拉格朗日对偶之后, 就可以按类似第三章的操作构造对偶问题:

$$(P) \Leftrightarrow \inf_{x \in E} \sup_{\substack{y \in \mathbb{R}^m \\ w \in K^*}} L(x, y, w) = \langle c, x \rangle + \sum_i y_i (b_i - \langle a_i, x \rangle) - \langle w, x \rangle$$

其拉格朗日对偶为:

$$\begin{aligned} \sup_{\substack{y \in \mathbb{R}^m \\ w \in K^*}} \inf_{x \in E} L(x, y, w) &= \langle c - \sum_i y_i \cdot a_i - w, x \rangle + \sum_i y_i \cdot b_i \\ &= \begin{cases} 0, & \text{if } c - \sum_i y_i \cdot a_i - w \neq 0. \\ -\infty, & \text{else} \end{cases} \end{aligned}$$

$$\Leftrightarrow (D) \quad \sup_{y \in \mathbb{R}^m} \sum_i y_i \cdot b_i \\ \text{s.t.} \quad c - \sum_i y_i \cdot a_i \in K^*$$

在第三章中为了证明 $\min\max = \max\min$, 我们使用了强对偶定理, 但没有证明; 这里也可以证明锥线性规划下的强对偶:

Theorem 7.1 (锥线性规划下的强对偶定理).

若 (P) 是有 下界 的, 并且 严格可行, 即.
 (D)

存在可行解 \bar{x} , s.t. $\bar{x} \in \text{int}(K)$
 $c - \sum_i y_i a_i \in \text{int}(K^*)$

那么, (P) 和 (D) 最优解相同, 且存在最优解 \hat{x}^* .

但也暂时挖一个坑, 日后再写。

本节所有证明均来自[1], 说明可能都不够严谨, 有不清楚的可以参考[1]。

H3 VaR的DRO例子 (下)

有了处理锥线性规划的工具, 就可以继续处理之前的DRO了。回顾一下之前的问题, 我们需要对问题8进行reformulation处理:

$$\begin{aligned} \sup_{\mathbb{P}} & \int_{\mathbb{R}^n} \mathbb{1}_{\{y \leq -w^T x\}}(x) d\mathbb{P}(x). \\ & \int_{\mathbb{R}^n} [x]_1 [x]_1^T d\mathbb{P}(x) = \Sigma, \quad \text{--- ⑧.} \\ & \mathbb{P} \geq 0. \end{aligned}$$

其中 Σ 是一个半正定矩阵。很明显, 接下来需要将其转化为SDP求解。构造对偶问题:

$$\mathcal{L}(\mathbb{P}, M) = \int_{\mathbb{R}^n} \mathbb{1}_{\{y \leq -w^T x\}}(x) d\mathbb{P}(x) + \langle M, \Sigma - \int_{\mathbb{R}^n} [x]_1 [x]_1^T d\mathbb{P}(x) \rangle \quad (M \in \mathbb{S}^{n+1}).$$

$$\begin{aligned} \text{其中 } \langle M, \Sigma - \int_{\mathbb{R}^n} [x]_1 [x]_1^T d\mathbb{P}(x) \rangle &= \int_{\mathbb{R}^n} \langle M, \Sigma - [x]_1 [x]_1^T \rangle d\mathbb{P}(x) \\ &= \langle M, \Sigma \rangle - \int_{\mathbb{R}^n} M \cdot [x]_1 [x]_1^T d\mathbb{P}(x) = \langle M, \Sigma \rangle - \int_{\mathbb{R}^n} g(x) d\mathbb{P}(x). \quad (g(x) = [x]_1^T M [x]_1) \end{aligned}$$

$$\text{问题 8} \Leftrightarrow \sup_{\mathbb{P} \geq 0} \inf_{M \in \mathbb{S}^{n+1}} \mathcal{L}(\mathbb{P}, M).$$

$$\begin{aligned}
& \langle M, [x] [x]^T \rangle \\
&= \text{tr}(M \cdot [x] [x]^T) = \text{tr}([x]^T M [x]) \\
&= [x] M [x]^T.
\end{aligned}$$

这里利用了我们前一小节证明的内容。注意我们这里把 $\mathbb{P} \geq 0$ 拿到外面了，因为放到里面还需要再定义一个内积。

$$\begin{aligned}
\sup_{P \geq 0} \mathcal{L}(P, M) &= \langle M, \Sigma \rangle + \sup_{P \geq 0} \int_{\mathbb{R}^n} \left[\mathbf{1}_{\{y \leq -w^T x\}}(x) - q(x) \right] dP(x) \\
&= \begin{cases} \langle M, \Sigma \rangle, & \text{if } \mathbf{1}_{\{y \leq -w^T x\}}(x) \leq q(x), \forall x. \\ +\infty, & \text{else.} \end{cases}
\end{aligned}$$

回忆构造对偶，就是把 \sup 和 \inf 换个位置，再提取出决策变量相关的约束。需要注意的是，在这里的子问题中， \mathbb{P} 是没有求和等于1的约束的，这条约束隐含在里面，所以只需要示性函数那条约束大于0出现， \mathbb{P} 就可以把那个点的密度调到无穷大，甚至大于1，使得解趋向正无穷。因此能得到大括号中的结论。

但是，这里又有个新的问题：新得到的约束依然是无限的（示性函数那条中的 $\forall x$ ）。因此我们还需要进一步处理示性函数的约束：

$$\mathbf{1}_{\{y \leq -w^T x\}}(x) \leq q(x), \forall x. \Leftrightarrow \begin{cases} q(x) \geq 0, \forall x & \text{--- A} \\ q(x) \geq 1, y + w^T x \leq 0. & \text{--- B} \end{cases}$$

$$1. A \Leftrightarrow M \in S_+^{n+1}$$

$$2. B \Leftrightarrow \inf_{\substack{x \in \mathbb{R}^n \\ s.t. y + w^T x \leq 0}} q(x) \geq 1.$$

$$\begin{aligned} \text{构造对偶. } & \inf_{\tau \geq 0} \sup_{x \in \mathbb{R}^n} \{ q(x) + \tau(y + w^T x) \} \\ & \Leftrightarrow \sup_{\tau \geq 0} \inf_{x \in \mathbb{R}^n} \{ q(x) + \tau(y + w^T x) \} \end{aligned}$$

问题转化为, $\exists \tau \geq 0$, 对 $\forall x \in \mathbb{R}^n$, 有:

$$\begin{aligned} q(x) \geq 1 + \tau(y + w^T x) &= \begin{bmatrix} x \\ 1 \end{bmatrix}^T \begin{bmatrix} 0 & -\tau w \\ -\tau w^T & 1-\tau y \end{bmatrix} \begin{bmatrix} x \\ 1 \end{bmatrix} \\ \Leftrightarrow & \begin{bmatrix} x \\ 1 \end{bmatrix}^T \left(M + \begin{bmatrix} 0 & -\tau w \\ -\tau w^T & 1-\tau y \end{bmatrix} \right) \begin{bmatrix} x \\ 1 \end{bmatrix} \geq 0. \\ \Leftrightarrow & \left(M + \begin{bmatrix} 0 & -\tau w \\ -\tau w^T & 1-\tau y \end{bmatrix} \right) \in S_+^{n+1}. \end{aligned}$$

约束可以写出A, B两个约束, 并分别按1. 和2. 处理。2中又用到了一次强对偶, 但这次的对偶只是LP的对偶。最后构造 $q(x)$ 的正定矩阵的时候可以写出 $[x, 1]$ 的二次型来构造。最终问题就能转化成如下的SDP形式:

$$\begin{aligned} \inf_{M \in S^{n+1}} \sup_{B \geq 0} L(B, M) &\Leftrightarrow \inf_{\substack{M, \tau \\ M \in S_+^{n+1}, \tau \geq 0}} \left\langle M, \Sigma \right\rangle \\ &\quad \text{s.t. } M + \begin{bmatrix} 0 & \tau w \\ \tau w^T & 1-\tau y \end{bmatrix} \in S_+^{n+1} \\ &\quad \text{SDP} \end{aligned}$$

最后的最后, 还要证明 $\sup\inf = \inf\sup$ 。这似乎与前面锥规划的强对偶又不完全一样, 因为合适针对概率测度 \mathbb{P} 的。[3]中的Proposition 3.4告诉我们一个证明方法。我们只需要说明如下内容:

考虑概率测度 $(\mathbb{R}^n, \mathcal{F})$ 中的集合：

$$\mathcal{M} = \{P \mid P \geq 0, P(\mathbb{R}^n) = 1\}.$$

令 $C = \text{cone}(\mathcal{M}) = \bigcup_{\lambda > 0} \lambda \cdot \mathcal{M}$ 为 \mathcal{M} 生成的凸锥。

定义如下的锥性映射：

$$P \mapsto A(P) = \int_{\mathbb{R}^n} [x] [x]^T dP(x) \in S_+^{n+1}$$

则问题 8 的约束可写作：

$$A(P) \in \Sigma, P \in C$$

因为 $\Sigma \in S_+^{n+1}$, 因此有 $\Sigma \in \text{int } A(C)$.

则根据 [3] 中的 Proposition 3.4, 上述问题的强对偶成立。

定理证明的关键在于：1. 将约束写出第一个红字的线性形式；2. 证明结果有第二个红字的形式，类似于说明这个解为上一部分中锥线性规划对偶里的严格可行解。由于定理的具体证明还需要很多篇幅（主要是我也没看...）所以暂时不继续深入。到这里，我们终于将问题 5 转化成了可以用 SDP 求解的形式。属实麻烦了。

以上例子来自 OR 的一篇文章 [2]。也算是 DRO 的一个 application 了。日后做实践篇的时候，有时间再翻原文出来详细看看。

本章类似于第二章 RO 中的求解方法概述。与 RO 中的不确定集相同，DRO 中最重要的还是模糊集。下一章我们将考虑其他模糊集。

参考资料：

[1] 苏文藻. 分布鲁棒优化简介. “数据驱动的优化理论、算法与应用”暑期学校. https://www.bilibili.com/video/BV1bM4y1L7R7?spm_id_from=333.337.search-card.all.click

[2] Laurent El Ghaoui, Maksim Oks, Francois Oustry. Worst-Case Value-At-Risk and Robust Portfolio Optimization: A Conic Programming Approach. Operations Research.

[3] Alexander Shapiro. On duality theory of conic linear problems.

8. 矩约束下的分布式鲁棒优化

上一章从一个例子引入了 DRO，并介绍了一个给定均值和方差的模糊集 DRO 问题例子。对于这类模糊集，我们统称为矩 (moment) 约束，均值和方差分别对应一、二阶矩，这也是一般文献中会考虑的。

上一章中的矩的特点是已知，但实际上，现实生活中一般只能得到样本的均值和方差，用来近似替代实际值。这也会带来很大的不确定性。因此，有必要考虑矩不确定性。这章的第一部分主要参考 [1][3] 介绍这部分内容。

此外，这章的第二部分还会介绍另一种解决矩模糊集下的方法，主要参考 [2]。

H3 矩不确定

回顾上次的内容，我们实际上解决了一个如下模糊集：

$$P(\mu_0, P_0) = \left\{ \begin{array}{l} P_{\mathbb{R}}(\xi \in \mathbb{R}^n) = 1 \\ \text{s.t. } E_{\mathbb{R}}[\xi] = \mu_0 \\ E_{\mathbb{R}}[(\xi - \mu_0)(\xi - \mu_0)^T] = P_0 \end{array} \right\},$$

$$\text{where } \mu_0 = \frac{1}{N} \sum_{i=1}^N \xi_i, \quad P_0 = \frac{1}{N} \sum_{i=1}^N (\xi_i - \mu_0)(\xi_i - \mu_0)^T.$$

其中均值和方差都是样本中估计的。因此我们希望处理不确定均值和方差的情况。[3]提出了一个模糊集：

$$P(C, \mu_0, P_0, \gamma_1, \gamma_2) = \left\{ \begin{array}{l} P_{\mathbb{R}}(\xi \in C) = 1 \\ \text{s.t. } (E_{\mathbb{R}}[\xi] - \mu_0)^T \cdot P_0^{-1} \cdot (E_{\mathbb{R}}[\xi] - \mu_0) \leq \gamma_1 \\ E_{\mathbb{R}}[(\xi - \mu_0)(\xi - \mu_0)^T] \preceq \gamma_2 \end{array} \right\}$$

$$\inf_{x \in U} \sup_{P \in \mathbb{P}} E_{\mathbb{R}}[h(x, \xi)]. \quad \text{--- ①.}$$

其中 C 是随机变量的支撑集（表示该随机变量可以取到的所有值的集合）。第一个约束表示 P 是一个概率测度，支撑集为 C ；第二个约束表示真实均值与样本均值的差在一个椭圆内，这个椭圆的大小由参数 γ_1 决定；第三个约束近似表示真实方差的范围在 γ_2 内，注意 γ_2 是一个半正定矩阵，LHS (left hand side) 也是一个矩阵，符号 \preceq 是定义在半正定锥上的符号，表示RHS-LHS为一个半正定矩阵。

有两点需要注意：1. 第二个约束并没有左侧的lower bound。这是因为加入之后不容易处理。2. 第三个约束并不是真实方差。真实方差应该使用真实均值 μ 而非样本均值 μ_0 。因此是一个近似约束。这点会在后续讨论。

因此，我们得到基于该moment uncertainty的DRO问题1。解决问题的操作与上一章求解VaR例子的方法基本相同。首先我们得到子问题：

$$\begin{aligned} \text{①} &\Leftrightarrow \inf_{x \in U} t. \\ \text{s.t. } t &\geq \sup_{P \in \mathbb{P}} E_{\mathbb{R}}[h(x, \xi)] \Leftrightarrow \begin{array}{l} \text{s.t. } \int_C dP(\xi) = 1, \quad (r) \\ \int_C \left[\begin{array}{cc} P_0 & \xi - \mu_0 \\ (\xi - \mu_0)^T & \gamma_1 \end{array} \right] dP(\xi) \succcurlyeq 0, \quad \left(\begin{bmatrix} P & P^T \\ P^T & s \end{bmatrix} \right) \text{半正定} \\ \int_C (\xi \xi^T - \xi \mu_0^T - \mu_0 \xi) dP(\xi) \preceq \gamma_2 P_0 - \mu_0 \mu_0^T \quad (\alpha) \end{array} \\ &P \geq 0. \end{aligned}$$

上一章中我们忽略了左侧的简单转换。这是在经典RO中的基础操作，转化为epigraph的形式。对于右侧的子问题，需要注意的是约束(1)从关于均值的标量不等式转化为了SDP的不等式。这是为了简化后续操作。这样转化是基于如下的舒尔补：

(1) Theorem. 8.1 : (Schur complement).

(不严谨表述). 分块矩阵 $\begin{bmatrix} A & B^T \\ B & C \end{bmatrix} \succ 0$ 等价于 $\begin{cases} C - B^T A^{-1} B \succ 0 \\ A \succ 0, C \succ 0. \end{cases}$

根据 Schur complement,

$$\left(E_{\mathbb{R}}[\xi] - \mu_0 \right)^T \frac{P_0^{-1}}{A} \cdot \left(E_{\mathbb{R}}[\xi] - \mu_0 \right) \leq \frac{\gamma_1}{C}$$

$$\Leftrightarrow \begin{bmatrix} P_0 & E_{\mathbb{R}}[\xi - \mu_0] \\ (E_{\mathbb{R}}[\xi - \mu_0])^T & \gamma_1 \end{bmatrix} \succ 0.$$

$$\Leftrightarrow E_{\mathbb{R}} \begin{bmatrix} P_0 & \xi - \mu_0 \\ (\xi - \mu_0)^T & \gamma_1 \end{bmatrix} \succ 0.$$

$$(2) \int_C (\xi - \mu_0)(\xi - \mu_0)^T d\mathbb{R}(\xi) \preceq \kappa \cdot P_0$$

$$\Leftrightarrow \int_C (\xi \xi^T - \xi \mu_0^T - \mu_0 \xi^T) d\mathbb{R}(\xi) \preceq \kappa \cdot P_0 - \mu_0 \mu_0^T$$

接下来基于问题2构造拉格朗日对偶，注意对矩阵写成内积形式，对最后一个约束不加入对偶：

$$\begin{aligned} \mathcal{L}(\mathbb{R}, r, \varrho, P, p, s) &= \int_C h(x, \xi) d\mathbb{R}(\xi) + r \cdot \left(1 - \int_C d\mathbb{R}(\xi) \right) \\ &\quad - \langle \varrho, \int_C (\xi \cdot \xi^T - \xi \cdot \mu_0^T - \mu_0 \cdot \xi^T) \cdot d\mathbb{R}(x) - \kappa \cdot P + \mu_0 \cdot \mu_0^T \rangle \\ &\quad + \langle P, P_0 \rangle + 2p^T \cdot \left(\int_C \xi d\mathbb{R}(\xi) - \mu_0 \right) + s \cdot \gamma_1. \end{aligned}$$

$$\textcircled{2} \Leftrightarrow \sup_{\substack{\mathbb{R} \succeq 0 \\ \varrho \succeq 0 \\ [P, P_0] \succ 0}} \inf_{r} \mathcal{L}(\mathbb{R}, r, \varrho, P, p, s)$$

接下来交换sup和inf：

$$\langle Q, \xi \xi^T \rangle = \text{tr}(Q \xi \xi^T) = \text{tr}(\xi^T Q \xi) = \xi^T Q \xi.$$

$$\sup_{P \in \mathbb{R}^{3,0}} L(P, r, Q, P_0, p, s) = \int_C [h(x, \xi) - r - \underbrace{\xi^T Q \xi}_{+ 2 \xi^T Q \cdot M_0 + 2 p^T \xi} + r + \langle Q, P_0 \rangle - \mu_0 \mu_0^T + \langle P, P_0 \rangle - 2 p^T \mu_0 + s \cdot v]$$

最后得到对偶问题：

$$\begin{aligned} \text{dual: } & \inf_{r, Q, P, p, s} r + \langle Q, P_0 \rangle - \mu_0 \mu_0^T + \langle P, P_0 \rangle - 2 p^T \mu_0 + s \cdot v \\ \text{s.t. } & h(x, \xi) - r - \xi^T Q \xi + 2 \xi^T Q \cdot M_0 + 2 p^T \xi \leq 0, \quad \forall \xi \in C. \\ & Q \succeq 0, \quad \begin{bmatrix} P & p \\ p^T & s \end{bmatrix} \succeq 0. \end{aligned}$$

这些应该是驾轻就熟的。对于该问题，可以看到变量 s 和 P 没有出现在第一个约束中，只在定义域上有约束；而定义域的规定半正定性的约束可以使用前面提到的舒尔补化简。因此对偶问题还可以进一步简化：

假设 $\xi^* \geq 0$ 是最优解中 ξ 的取值。

1. $\xi^* \geq 0$. 由舒尔补，对 $\begin{bmatrix} P & p \\ p^T & s \end{bmatrix} \succeq 0$ ，有 $P \succeq \frac{1}{2} P P^T$
对半正定矩阵 A, B 有性质：
若 $A, B \succeq 0$ ，则 $\langle A, B \rangle \geq 0$. (证：由 $u^T A u \geq 0$ ，构造 $U = \sum_j u_j A u_j$)

∴ 由于 $P \succeq 0$, $P \succeq \frac{1}{2} P P^T \succeq 0$. 有：

$$\langle P, P_0 \rangle \geq \langle \frac{1}{2} P P^T, P_0 \rangle. \quad \text{所以对 } P, P^* = \frac{1}{2} P P^T.$$

因此，对 ξ^* ，只需要考虑 $\min_{s \geq 0} \frac{1}{2} P^T P \cdot p + s \cdot v$.

$$\text{求导得 } \xi^* = \frac{1}{\sqrt{P_0}} \cdot \|P_0^{-\frac{1}{2}} \cdot p\|_2$$

令 $q := p + Q \cdot M_0$, 则 $\lambda(\xi^*, P^*)$ 有：

$$\text{obj} = r + \nu_2 \cdot \langle P_0, Q \rangle + M_0^T Q \cdot M_0 - 2 M_0^T \cdot q + 2 \sqrt{P_0} \cdot \|P_0^{-\frac{1}{2}}(q - Q \cdot M_0)\|_2.$$

—— (a)

2. $\xi^* = 0$. 对半正定矩阵 A ，有性质：

若 $A \succeq 0$, $A_{ij} = 0$, 则 对 k, l , $A_{kj} = A_{lj} = 0$.

∴ 当 $\xi^* = 0$ 时, $p^* = 0$, 进而 $P^* = 0$.

此时 (a) 成立。

这里面运用了一些半正定矩阵的性质，看来线性代数还是要加强。

将最优值带入问题3最终得到对偶问题如下：

$$\inf_{r, \alpha, q, t} \quad r + \gamma_2 \cdot \langle P_0, \alpha \rangle + \mu_0^\top \alpha \cdot \mu_0 - 2\mu_0^\top q + 2\sqrt{\nu_1} \cdot t$$

$$\text{s.t.} \quad t \geq \|P_0^{\frac{1}{2}}(q - \alpha \mu_0)\|_2 \quad \text{--- (1)}$$

—— (4).

$$h(x, \xi) - r - \xi^\top \alpha \xi + 2\xi^\top q \leq 0, \forall \xi \in C.$$

$$\alpha \succcurlyeq 0$$

约束1其实是一个二阶锥约束。

记得在上一章中，为了证明强对偶成立，我们使用了[5]中的proposition 3.4。在这里我们也需要做同样的证明：

问题②可写作：

$$\sup_{\alpha} \langle h(x, \xi), \alpha \rangle.$$

$$\text{s.t.} \quad A(\alpha) = \begin{bmatrix} 1 \\ -P_0 + \mu_0^\top \mu_0 \\ \begin{bmatrix} P_0 & \mu_0 \\ \mu_0 & -r \end{bmatrix} \end{bmatrix} \in \underbrace{\{0\} \times S_r^n \times S_r^m}_{\triangleq K},$$

$$\alpha \in \mathcal{C}. \quad (\text{一个锥!})$$

其中， $A(\cdot)$ 是一个锥性映射，从 $\mathcal{C} \rightarrow \left(\int_C d\alpha(\xi), \int_C (\xi \xi^\top + \xi^\top \mu_0 + \mu_0^\top \xi) d\alpha(\xi), \left[\int_C \xi^\top d\alpha(\xi), 0 \right] \right)$

$$\mathcal{G} = \text{cone}(\mathcal{M}), \quad \mathcal{M} = \left\{ \alpha : \alpha \geq 0, \alpha(\alpha) = 1 \right\}.$$

令 $\alpha(\xi) = \delta_{\mu_0}(\xi)$. (一个概率测度，在 μ_0 处 mass 为 1，其余位置都为 0) $\in \mathcal{M}$

$$\therefore \alpha(\delta_{\mu_0}(\xi)) \in \mathcal{G} \quad \therefore A(\mathcal{G}) - K \supseteq A(\alpha(\delta_{\mu_0}(\xi))) - K = \left\{ (0, 0 \cdot \mu_0 \mu_0^\top - S_r^n, 0 \left[\begin{smallmatrix} \mu_0 & 0 \\ 0 & 0 \end{smallmatrix} \right] - S_r^m) \right\}$$

因此， $b \in \text{int}[A(\mathcal{G}) - K]$. 根据 Shapley Proposition 3.4 得，强对偶成立。

证明的核心是两个：一个是写作 A 的线性形式，另一个是证明 b 在 $A(\mathcal{C}) - K$ 内部。证明完成后就可以将 reformulation 的结果带入问题 1：

$$\inf_{r, \alpha, q, t} \quad r + t$$

$$\text{s.t.} \quad h(x, \xi) - r - \xi^\top \alpha \xi + 2\xi^\top q \leq 0, \forall \xi \in C.$$

$$t \geq \gamma_2 \cdot \langle P_0, \alpha \rangle + \mu_0^\top \alpha \cdot \mu_0 - 2\mu_0^\top q + 2\sqrt{\nu_1} \|P_0^{\frac{1}{2}}(q - \alpha \mu_0)\|_2$$

$$r \in \mathbb{R}, \alpha \succcurlyeq 0, q \in \mathbb{R}^n, x \in U, t \in \mathbb{R}.$$

之前我们提到第二个约束是一个二阶锥约束。但其实，真正影响求解的还是第一个约束，因为 $\xi \in C$ 是一个无限集。在[3]中使用了椭球算法证明可以在多项式时间内求解该问题，在这里我们就不进一步说明了。余下的一个问题，如何确定模糊集的参数 γ_1, γ_2 。

对于模糊集来说，很明显两个参数越大，模糊集的范围就越大，里面包含真实分布的概率就越大。但是，由于求解的是鲁棒优化，最终解要符合集合中所有分布，因此太大的模糊集又会造成反效果。如何 balance这两项内容是一个挑战。

实际上确定参数的大小相当于确定真实的均值和方差离样本均值和方差的距离。可以猜测，由大数定理，样本容量很大的时候距离应该很近。进一步bound这之间的距离需要使用概率论中的集中不等式 (concentration inequalities)。

在处理模糊集中的约束2之前，我们先处理约束2中方差为真实方差的情况。并给出一个上界的假设，对约束进行处理：

$$P(C, \mu_0, P_0, \gamma_1, \gamma_2) = \left\{ \begin{array}{l} P \geq 0 : (E_P[\xi] - \mu_0)^T \cdot P^{-1} \cdot (E_P[\xi] - \mu_0) \leq \gamma_1 \\ E_P[(\xi - \mu_0)(\xi - \mu_0)^T] \leq \gamma_2 \end{array} \right\}$$

令 μ 和 P 表示真实均值方差。假设真实分布 P^* 是有界的，即对某个 $R \in (0, +\infty)$ ，有：

$$P^*[(\xi - \mu)^T \cdot P^{-1} \cdot (\xi - \mu) \leq R^2] = 1.$$

则对 $\forall t > 0$ ，有

$$\begin{aligned} P^*[(\mu_0 - \mu)^T \cdot P^{-1} \cdot (\mu_0 - \mu) \leq t] &= P^*[\|P^k(\mu_0 - \mu)\|_2 \leq t] \\ &\stackrel{\text{均值定义}}{=} P^*\left[\left\|\bar{P}^k\left(\frac{1}{N} \sum_{i=1}^N \xi_i - \mu\right)\right\|_2 \leq t\right] = P^*\left[\left\|\frac{1}{N} \sum_{i=1}^N \bar{P}^k(\xi_i - \mu)\right\|_2 \leq t\right] \\ &\stackrel{\xi_i = \bar{P}^k(\xi_i - \mu)}{=} P^*\left[\left\|\frac{1}{N} \sum_{i=1}^N \xi_i\right\|_2 \leq t\right]. \end{aligned}$$

这里我们要求该分布有上界。在后面我们会看到这样操作的目的。经过如上图的处理后，式子从由 ξ 表示的约束变成了由 ζ 表示。 ζ 有一些良好的性质：

均值: $\sum \zeta_i = \bar{P}^k(\sum \xi_i - N \cdot \mu) = 0$.

方差: $E[\zeta \cdot \zeta^T] = E[\bar{P}^k(\xi_i - \mu)(\xi_i - \mu)^T \cdot \bar{P}^k]$
 $= \bar{E}[\bar{P}^k \cdot P \cdot \bar{P}^k] = I$

在此之上，我们对 t 赋值，就能得到想要的 bound:

令 $t = \frac{R^2}{N} \cdot (2 + \sqrt{2 \cdot \ln(1/\delta)})^2$ ，根据 McDiarmid's concentration inequalities，
 有：

$$P^*\left[\left\|\frac{1}{N} \sum_{i=1}^N \zeta_i\right\|_2 \leq \frac{R^2}{N} \cdot (2 + \sqrt{2 \cdot \ln(1/\delta)})^2\right] \geq 1 - \delta.$$

可以观察到 t 的取值中有 R 。这里用到的集中不等式如下：

Theorem 8.2: (McDiarmid).

令 $\{\xi_i\}_{i=1}^N$ 为相互独立的随机向量， g 为实值函数。若对 ζ_j 有：

$$|g(\xi_1, \dots, \xi_j, \dots, \xi_N) - g(\xi_1, \dots, \xi'_j, \dots, \xi_N)| \leq \gamma_j$$

则对 $\forall t > 0$, 有：

$$P\left[g(\xi_1, \dots, \xi_N) - E[g(\xi_1, \dots, \xi_N)] \leq -t\right] \leq \exp\left(-\frac{2t^2}{\sum_{j=1}^N \gamma_j^2}\right)$$

这个不等式告诉我们，如果只改变函数 g 某一项之后，与原函数的差距可控，则对 $g - E[g]$ 会有一个 bound 的界。为了应用到我们的问题上，需要进行如下操作：

$$\frac{1}{n} g(\xi_1, \dots, \xi_n) = \left\| \frac{1}{n} \sum_{i=1}^n \xi_i \right\|_2^2 = (\frac{1}{n} \xi)^T \cdot (\frac{1}{n} \xi).$$

$$\text{则 } E[g(\xi_1, \dots, \xi_n)] = \frac{1}{n} \sum_{i,j} E[\xi_i^T \xi_j]$$

$$\because E[\xi_i^T \xi_j] = E[\xi_i]^T \cdot E[\xi_j] = 0$$

$$\therefore E[g] = \frac{1}{n} \sum_i E\left[\left\|\xi_i\right\|_2^2\right] \stackrel{\text{(同分布)}}{=} \frac{1}{n} E\left[\left\|\xi\right\|_2^2\right]$$

$$\therefore P^*[(\xi - \mu)^T \cdot P^{-1} \cdot (\xi - \mu) \leq R^2] = P^*\left[\left\|\xi\right\|_2^2 \leq R^2\right] = 1$$

$$\therefore E[g] \leq \frac{R^2}{n}. \quad \text{接下来研究 } g \dots$$

这里还没有操作完。挖个坑先。总之，通过之前对 P 的上界假定，我们可以进一步算出 $E[g]$ 的bound，从而带入到不等式中，得到关于 g 的bound。这样我们就证明了，在模糊集的约束2中，如果方差是真实方差，就可以得到一个有保障的 γ_1 。当然，这里要求 P 是有界的，这显然不太好，因为包括正态在内的很多分布是无界的。因此苏老师在一篇MP上拓展了这个定理到一些清尾的分布上，指的是和正态类似， x 趋向无穷时概率趋向无穷的分布。

为了处理约束2为样本方差的情况，有以下基于集中不等式推导得出的定理：

Theorem 8.3:

令 $\{\xi_i\}_{i=1}^N$ 为相互独立的随机向量，有：

$$P\left[\frac{1}{1+\alpha} P_0 \leq P \leq \frac{1}{1-\beta} P_0\right] \geq 1 - \delta.$$

其中 $\alpha = \dots, \beta = \dots$

(偷懒了懒得写了) 结合上述定理, 就可以说明模糊集的约束2是有bound的。同样, γ_2 也可以用类似集中不等式得出。可见, DRO的公式推导型文章真的很麻烦。详细请原文[3]。

矩不确定性应该还有别的表述方式。[6]中就提到一个模糊集:

- Ambiguity set of distributions:

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathcal{U}) \middle| \begin{array}{l} \tilde{\alpha} \sim \mathbb{P}, \\ \mathbb{E}_{\mathbb{P}}[\tilde{\alpha}] = \mu, \\ \mathbb{E}_{\mathbb{P}}[|\tilde{\alpha} - \mu|] \leq \sigma. \end{array} \right\}$$

- *Support set*: $\mathcal{U} = [\underline{a}, \bar{a}]$
- $\mathcal{P}_0(\mathcal{U})$: the set of all probability distributions on \mathcal{U}

- Focus on

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}[\max\{\tilde{\alpha}^\top x, 0\}].$$

这里不再使用方差, 而是离差。据说这种模糊集可以使用lifted uncertainty set很简单地转化为RO。个人认为, 因为模糊集包括RO中的不确定集确实很多, 没必要在初学时全部掌握, 见招拆招、要用时再看即可。

下一部分内容来自[2], 是[3]的作者数年之后写的, 似乎和[3]不完全一样。我认为有必要了解。

H3 另一种方法

在第二种方法中, 我们首先从以下一阶矩约束开始:

$$\min_{x \in X} \sup_{F \in \mathcal{D}} E_F[h(x, \xi)].$$

$$\mathcal{D}(Z, \mu) = \left\{ F \in M \mid \begin{array}{l} P_F(\xi \in Z) = 1 \\ E_F[\xi] = \mu \end{array} \right\} \quad \text{--- } \textcircled{B}$$

这里有些符号和之前不同, 但应该很好认出。 Z 是支撑集, μ 是已知的均值 (可以理解为之前说的样本均值), M 是测度空间, \mathcal{D} 为模糊集。注意, 为了得到更general的结论, 在这里我们不要求 $h(x, \xi)$ 是线性函数。

写成epigraph后写出子问题:

$$\max_{\bar{F}GM} \int_Z h(x, \xi) d\bar{F}(\xi).$$

$$\text{s.t. } \int_Z d\bar{F}(\xi) = 1 \quad (r) \quad \dots \quad \textcircled{7}$$

$$\int_Z \xi \cdot d\bar{F}(\xi) = \mu. (q)$$

求拉格朗日对偶：

$$L(F, r, q) = \int_Z h(x, \xi) d\bar{F}(\xi) + r \cdot (1 - \int_Z d\bar{F}(\xi)) + q^T [\mu - \int_Z \xi d\bar{F}(\xi)]$$

$$\sup_F L(F, r, q) = r + \mu^T q + \int_Z [h(x, \xi) - r - q^T \xi] \cdot d\bar{F}(\xi).$$

由于之类没有矩阵的约束，因此不需要内积形式的惩罚项。对偶问题如下：

$$\begin{aligned} \text{dual: } & \min_{r, q} \mu^T q + r \\ \text{s.t. } & z^T q + r \geq h(x, z), \forall z \in Z. \end{aligned} \quad \text{--- } \textcircled{8}.$$

强对偶定理的证明依旧需要用到[5]中的proposition 3.4。因此我们可以得到问题7的等价问题：

$$\min_q \sup_{z \in Z} h(x, z) + (\mu - z)^T q. \quad \text{--- } \textcircled{9}$$

问题9将问题8中的决策变量 r 简化了。在进一步处理问题9之前，我们讨论一下问题8的性质。

<对于任何凸集中最优解对应的点，也就是多面体的顶点，只需要 $m + 1$ 条约束就可以确定这样的点（其中 m 是 z 的维度）。或者说，对于任何一个点，都可以通过 $m + 1$ 条约束构造出一个多面体，让这个点变成它的定点。由于问题8有无数条约束，最优解应该可以用 $m + 1$ 个 z 刻画出>（其实这里我没怎么看懂）。因此，问题8可以转化为：

$$\begin{aligned}
 & \min_{r, q} \mu^T q + r \\
 \text{s.t. } & z^T q + r \geq h(x, z), \forall z \in \mathbb{Z}^* \quad \text{--- (D).}
 \end{aligned}$$

$$\mathbb{Z}^* = \{z_1^*, z_2^*, \dots, z_{m+1}^*\}$$

同样的，可以得到对应的原问题。但是我们依然不知道 z_i^* 的取值。但是，[5]给出了如下定理：

Theorem 8.4: 问题①等价于如下问题：

$$\begin{aligned}
 & \max_{\{p_i\}_{i=1}^{m+1}, \{z_i\}_{i=1}^{m+1}} \sum_{i=1}^{m+1} p_i \cdot h(x, z_i) \\
 \text{s.t. } & \sum_{i=1}^{m+1} p_i = 1. \quad \text{--- (II)} \\
 & \sum_{i=1}^{m+1} p_i \cdot z_i = \mu \\
 & p_i \geq 0, z_i \in \mathbb{Z}, \forall i.
 \end{aligned}$$

虽然在这里约束2是非线性的，但问题已经变成一个有限维的问题。进一步的，考虑到 $h(x, z_i)$ 并不一定线性，我们进一步考虑特殊的 h 下的结果。有如下定理：

Theorem 8.5:

当 Z 为凸集, $h(x, z) = \max_{k=1, \dots, K} h_k(x, z)$, 其中 $h_k(x, z)$ 为凸函数,
则问题⑪等价于
(如, $h(x, z)$ 为分段凸函数)

$$\begin{aligned} & \max_{\{p_k\}_{k=1}^K, \{\bar{z}_k\}_{k=1}^K} \sum_{k=1}^K p_k h_k(x, \bar{z}_k) \\ \text{s.t. } & \sum_{k=1}^K p_k = 1. \quad \longrightarrow \textcircled{12} \\ & \sum_{k=1}^K p_k \cdot \bar{z}_k = \mu \\ & p_k \geq 0, \bar{z}_k \in Z, \forall k \end{aligned}$$

定理的证明只需要分别证明问题7和问题12的解能相互转换即可。

下面的推论非常有效:

推论: 当 $h(x, z)$ 为凸时, 问题⑬等价于:

$$\max_{x \in X} h(x, \mu). \quad \textcircled{13}$$

这个推论其实已经很普遍了, 可以解决大多数问题。当然, 对前面提到的分段凹的情况, 还可以有如下定理:

Theorem 8.7:

当 Z 为凸集, $h(x, z) = \max_{k=1, \dots, K} h_k(x, z)$, 其中 $h_k(x, z)$ 为

关于 z 的凹函数时, 问题 ⑥ 可被简化为:

$$\begin{array}{ll} \min_{x, q, t, \{v_k\}} & t \\ \text{s.t.} & t \geq \delta^*(v_k | Z) + \mu^T q - h_x^k(x, v_k + q), \quad \forall k, \\ & x \in X. \end{array}$$

其中 $\delta^*(v_k | Z) := \sup_{z \in Z} z^T v_k$, $h_x^k(x, v) = \inf_{z \in Z} v^T z - h_k(x, z)$.

在证明之前需要用到:

Theorem 8.6: $x \in X$ 满足非闭性鲁棒约束

$$g(x, z) \leq 0, \quad \forall z \in Z,$$

其中 Z 为非空闭凸集, 对 $\forall z$, g 对 x 为凸, 对 $\forall x$, g 对 z 为凸.

当且仅当存在 $v \in \mathbb{R}^m$ 满足:

$$\delta^*(v | Z) - g_x^*(x, v) \leq 0.$$

这个我们暂不详细说明。利用 8.6, 可以证明 8.7:

Proof: 问题⑥ reformulation 可得问题⑦:

$$\min_q \sup_{z \in Z} h(x, z) + (\mu - z)^T q. \quad \text{--- ⑨}$$

在满足 Z 和 h_k 性质时, 可进一步转化为 (?)

$$\min_{x, q} \sup_{z \in Z} \max_k h_k(x, z) + (\mu - z)^T q.$$

$$\Leftrightarrow \min_{x, q} \max_k \sup_{z \in Z} h_k(x, z) + (\mu - z)^T q.$$

$$\begin{aligned} & \min_{x, q, t} \quad t \\ & \text{s.t.} \quad t \geq \sup_{z \in Z} h_k(x, z) + (\mu - z)^T q, \quad \forall k. \end{aligned}$$

由定理 8.6, 得来等价于:

$$\exists v \in R^m, \quad t \geq \delta^*(v/z) + \mu^T q - g_*(x, v).$$

$$\begin{aligned} \text{其中 } g_*(x, v) &= \inf_{z \in Z_h} v^T z - h_k(x, z) + q^T z \\ &= \inf_{z \in Z_h} (v + q)^T z - h_k(x, z) = h_*^k(x, v + q). \end{aligned}$$

到这里我们初步结束对问题6的模糊集的处理。接下来考虑上一章中考虑的二阶矩约束下的模糊集:

$$\begin{aligned} & \min_{x \in X} \sup_{F \in D} E_F [h(x, \xi)]. \\ D_2(\mu, \sigma^2) &= \left\{ F \in M \mid \begin{array}{l} P_F(\xi \in R) = 1 \\ E_F[\xi] = \mu \\ E_F[(\xi - \mu)^2] = \sigma^2 \end{array} \right\} \quad \text{--- ⑩} \end{aligned}$$

注意, 为了简化问题, 这里的参数都是标量而不是向量。随机变量 ξ 不是向量, 而是一维变量。

处理这个模糊集, 可以构造一个lifted space和对应的随机变量:

$$\zeta = [\zeta_1 \ \zeta_2]^T := [\xi, (\xi - \mu)^2]^T$$

$$Z' := \{z' \in \mathbb{R}^2 \mid z_2' = (z_1' - \mu)^2\}.$$

构造这个的目的是为了带入之前的模糊集中：

$$\min_{x \in X} \sup_{F' \in \mathcal{D}(Z', [\mu \sigma^2]^T)} E_{F'}[h(x, \zeta_1)].$$

$$\mathcal{D}(Z, \mu) = \left\{ F \in M \mid \begin{array}{l} P_F(\xi \in Z) = 1 \\ E_F[\xi] = \mu \end{array} \right\} \quad \text{--- (15)}$$

神奇的事情发生了：问题14和问题15应该是等价的。注意问题14和15的模糊集中对应参数的含义不一样，要弄清楚。这样做的原因在于，方差的矩约束其实可以转化为均值的约束，都有期望=某个数的形式。接下来就可以用之前分析的方法处理问题15，得到问题14的解：

$$\begin{aligned} & \min_{x \in X, q \in \mathbb{R}^2} \sup_{z \in Z} h(x, z_1) + (\mu - z_1) \cdot q_1 + (\sigma - z_2) \cdot q_2 \\ \Leftrightarrow & \min_{x \in X, q \in \mathbb{R}^2} \sup_{z \in Z} h(x, z_1) + (\mu - z_1) q_1 + [\sigma - (z_2 - \mu)] q_2 \\ \xrightarrow{\text{epigraph}} & \begin{array}{ll} \min_{x \in X, q, t} & t \\ \text{s.t.} & t \geq \sup_{z \in Z} h_k(x, z_1) + \mu q_1 + (\sigma - \mu) q_2 - (q_1 - 2q_2 \mu) z_2 - q_2 z_1^2 \end{array} \end{aligned}$$

由于问题15中考虑了两个随机变量对应的约束，所以reformulation后对偶问题的向量 q 有两个标量 q_1, q_2 。

因为平板不再身边，后续暂时用截图。

再考虑一个如下例子：

$$\underset{x \in \mathcal{X}}{\text{minimize}} \sup_{F \in \mathcal{D}(\mathcal{Z}, \mu, \Sigma)} \mathbb{E}_F[h(x, \xi)],$$

$$\mathcal{D}(\mathcal{Z}, \mu, \Sigma) := \{F \mid \mathbb{P}(\xi \in \mathcal{Z}) = 1, \mathbb{E}[\xi] = \mu, \mathbb{E}[\xi \xi^T] \preceq \Sigma\}.$$

这里对方差的约束变成了 \preceq 。依然构造lifted space:

$$\mathcal{Z}' := \{(z_1, Z_2) \in \mathbb{R}^m \times \mathcal{S}^{m \times m} \mid z_1 \in \mathcal{Z}, Z_2 \succeq z_1 z_1^T\},$$

进行reformulation:

$$\underset{x \in \mathcal{X}, q, Q}{\text{minimize}} \quad \sup_{(z_1; Z_2): z_1 \in \mathcal{Z}, Z_2 \succeq z_1 z_1^T} h(x, z_1) + (\mu - z_1)^T q + (\Sigma - Z_2) \bullet Q,$$

这里最后一项为内积。reformulation的方式与前一节类似。最后转化为epigraph的形式:

$$\begin{aligned} & \underset{x \in \mathcal{X}, q, r}{\text{minimize}} \quad r \\ \text{subject to} \quad & r \geq \sup_{(z_1; Z_2): z_1 \in \mathcal{Z}, Z_2 \succeq z_1 z_1^T} h_k(x, z_1) + (\mu - z_1)^T q + (\Sigma - Z_2) \bullet Q, \forall k. \end{aligned}$$

这个例子的内容就大致结束了。[2]中还有针对本章第一节的内容，可以用相同方式处理，但是因为要用到一些非线性鲁棒的知识，所以暂时不写。可以看出的是，这一种方式与上一节的方法差距还是很大的，即使[2][3]的作者相同。

对此，我也咨询了老师，老师的解释是目前后一种lifted的方法使用的更广泛。看来还需要进一步理解lifted。

在下一章中，我们将讨论另一种模糊集建模下的DRO。

[1] 苏文藻. 分布鲁棒优化简介.“数据驱动的优化理论、算法与应用”暑期学校. https://www.bilibili.com/video/BV1bM4y1L7R7?spm_id_from=333.337.search-card.all.click

[2] Erick Delage. Quantitative Risk Management Using Robust Optimization Lecture Note. https://tin.hec.ca/pages/erick.delage/MATH80624_LectureNotes.pdf

[3] Erick Delage, Yinyu Ye. Distributionally Robust Optimization Under Moment Uncertainty with Application to Data-Driven Problems. Operations Research.

[4] Wolfram Wiesemann, Daniel Kuhn, Melvyn Sim. Distributionally Robust Convex Optimization. Operations Research.

[5] Alexander Shapiro. On duality theory of conic linear problems.

[6] 章宇. 鲁棒优化及其在车辆路径问题中的应用简介. https://www.bilibili.com/video/BV1tK4y1e7fQ?spm_id_from=333.337.search-card.all.click

H2 9. Wasserstein下的分布式鲁棒优化

看到一篇知乎文章“Distributionally Robust Optimization 高引论文”，挺有意思的，里面说“DRO推公式是个大坑，慎入。难度略大，对分析学的要求会高一点”，确实如此，所以我一个管理学学生没学过实分析凸分析泛函分析随机过程，拿什么学数学... 流下了没有数学的眼泪。所以今年水完之后，明年不要急着写文章了，先打好数学基础...

基本上到理论部分最后两章了，打算这一章介绍一个Wasserstein的例子，不深入分析[2]。然后再学一学ADRO，简单写一章介绍。还有很多模糊不清楚的地方，后续再看两篇应用的paper，根据算法设计、应用的需要来决定后续学习。应用篇应该会再开一个系列。

H3 Wasserstein Distance

上两章我们构建了基于矩约束的模糊集，称为moment-based model。这当然是有局限性的，因为矩不能完美地刻画一个分布。最近几年非常火的一种模糊集是针对概率分布之间的距离的，被称为metric-based。也就是说，需要找一种衡量两种概率分布之间距离的手段，距离越近显然越好，用于对比的初始分布可以通过历史数据得到。记如下分布为经验分布(empirical distribution):

对随机向量 ξ ，有*N*个iid的样本 $\xi_i, i=1, \dots, N$ 。定义经验分布：

$$\mathbb{P}_N^1 = \frac{1}{N} \sum_{i=1}^N \delta_{\xi_i}.$$

即，对每一个 ξ 有 \mathbb{P} 的概率为零的分布。

不难猜测，标题的Wasserstein distance就是测度分布的一种方式。它的定义如下：

令 \mathbb{Q}, \mathbb{P} 为两个概率空间 $(\mathcal{X}, \mathcal{B})$ 中的概率测度，则
Wasserstein 距离定义为：

$$W(\mathbb{Q}, \mathbb{P}) = \inf_{\Pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})} \left\{ \int_{\mathcal{X} \times \mathcal{Y}} d(\xi, \xi') \Pi(d\xi, d\xi') : \underbrace{\Pi(\mathcal{X}, d\xi')}_{\text{边缘分布}} = \mathbb{P}(\xi), \underbrace{\Pi(d\xi, \mathcal{Y})}_{\text{边缘分布}} = \mathbb{Q}(\xi) \right\}$$

其中 $d(\xi, \xi')$ 为一个凸函数，称为transport cost.

则可定义经验分布 \mathbb{P} 的邻域 $Ball_\epsilon(\mathbb{P}) = \{\mathbb{Q} : W(\mathbb{Q}, \mathbb{P}) \leq \epsilon\}$.

解释一下这个距离，这个定义做的事情其实就是找到一个近似的联合分布 Π ，满足其边缘分布分别为两个分布 \mathbb{Q} 和 \mathbb{P} 。然后，在计算距离时，对所有点积分，每个点从 \mathbb{Q} 到 \mathbb{P} 的距离通过一个transport cost表示。这个函数 d 也就控制了计算距离的方式。DRO的模糊集就可以通过这个距离定义的邻域得到。

H3 Logistic Regression

Wasserstein distance-based的DRO模型有许多应用，本章介绍一个logistic regression中的应用例子。

特征向量 $x \in \mathbb{R}^n \rightarrow$ 二元标签 $y \in \{1, -1\}$.

$$\text{s.t. } P(y|x) = (1 + \exp(-y \cdot \beta^T \cdot x))^{-1}.$$

回归目标: β .

随机变量 $\xi = (x, y)$ 有 N 个样本 $\{\xi_i = (\hat{x}_i, \hat{y}_i)\}_{i=1}^N$

极大似然估计下, β 为:

$$\beta_{MLE} = \arg \min_{\beta} \frac{1}{N} \sum_{i=1}^N l_p(\xi_i) = \arg \min_{\beta} E_{P^N} [l_p(\xi)] \\ \neq \arg \min_{\beta} E_{P^*} [l_p(\xi)].$$

$$\text{其中 } l_p(\xi) = \log(1 + \exp(-y \cdot \beta^T \cdot x))$$

这个回归问题的输入是一个 n 维特征向量 x , 输出一个二元标签 y 。他们之间满足一个条件概率的关系。需要估计这个关系中的参数 β 。按照传统极大似然估计下得到的 $\hat{\beta}_{MLE}$ 并不是真实分布的期望, 而是经验分布的期望。虽然在样本容量很大的情况下两者近似相同, 但在样本小的情况下不完全一致。因此我们构建如下的DRO问题:

$$\inf_{\beta} \sup_{P \in Ball_{\epsilon}(\mathbb{R}^N)} E_P [l_p(\xi)] \quad \text{--- ①}$$

$$Ball_{\epsilon}(P) = \{\alpha : w(\alpha, P) \leq \epsilon\}.$$

$$d(\xi, \xi') = \|x - x'\|_1 + \frac{\kappa}{2} |y - y'|.$$

$\|\cdot\|$ 表示任意范数。

当模糊集合适时, 真实分布就会包含在模糊集中。模糊集由Wasserstein distance定义。其中transport cost函数如图所示。参数 κ 权衡了 x 和 y 对距离的影响。其中对 x 可以用任何范数来表示, y 是二元变量, 因此用绝对值表示, 若 $y \neq y'$, 绝对值的值等于2, 正好消掉 $\kappa/2$ 的分母。

下面我们专心处理问题1。首先是写成epigraph形式, 根据模糊集重新书写子问题:

$$\sup_{P \in \text{Ball}_\varepsilon(\mathbb{R}^N)} E_P[\ell_p(\xi)] \Leftrightarrow \begin{aligned} & \sup_{\Pi \in \mathcal{P}(\Xi \times \Xi)} \int_{\Xi} \ell_p(\xi) d\Pi(\xi) \\ & \text{s.t. } \int_{\Xi \times \Xi} d(\xi, \xi') \Pi(d\xi, d\xi') \leq \varepsilon, \quad \text{--- ②.} \\ & \Pi(\Xi, d\xi') = \hat{P}_N(d\xi'). \end{aligned}$$

这里的子问题书写可能比以前麻烦一些。注意目标函数中是对真实分布进行积分。问题2可以做一些简化。注意到经验分布其实是一个离散分布，因此可以写成条件分布的形式：

$$\begin{aligned} Q(d\xi) &= \Pi(d\xi, \Xi) \\ &= \int_{\xi' \in \Xi} \Pi d(\xi, \xi') \\ &= \sum_{i=1}^N \underbrace{\Pi(d\xi | \xi' = (\hat{x}_i, \hat{y}_i))}_{\text{红色波浪线}} \cdot P_N^i(\hat{x}_i, \hat{y}_i) \\ &= \frac{1}{N} \sum_{i=1}^N Q^i(d\xi) \end{aligned}$$

$$\begin{aligned} \Pi d(\xi, \xi') &= \Pi(d\xi | \xi' = (\hat{x}_i, \hat{y}_i)) \cdot P_N^i(\xi') \\ &= \frac{1}{N} \cdot \sum_{i=1}^N \delta_{(\hat{x}_i, \hat{y}_i)}(\xi') \cdot Q^i(d\xi) \end{aligned}$$

$$\begin{aligned} & \sup_{Q^i \geq 0} \frac{1}{N} \sum_{i=1}^N \int_{\Xi} \ell_p(\xi) Q^i(d\xi) \\ & \text{s.t. } \frac{1}{N} \cdot \sum_{i=1}^N \int_{\Xi} d(\xi, (\hat{x}_i, \hat{y}_i)) Q^i(d\xi) \leq \varepsilon, \quad \text{--- ①.} \\ & \int_{\Xi} Q^i(d\xi) = 1. \quad \forall i. \end{aligned}$$

红色波浪线部分是 Q^i 的定义。为了后续处理方便，将 Q^i 继续分解：

$$\mathbb{Q}^i(d\zeta) = \mathbb{Q}_{+1}^i(dx) + \mathbb{Q}_{-1}^i(dx).$$

其中 $\mathbb{Q}_j^i(dx) = \mathbb{Q}^i(dx, y=j)$.

$$\begin{aligned} & \sup_{\mathbb{Q}_{+1}^i \geq 0} \frac{1}{N} \sum_{i=1}^N \int_{\Omega} f_p(x_i) \mathbb{Q}_{+1}^i(d\zeta) + f_p(x_{-1}) \mathbb{Q}_{-1}^i(d\zeta) \quad \text{--- ④} \\ \text{s.t. } & \frac{1}{N} \sum_{i=1}^N \int_{\Omega} d(\zeta, (x_i, y_i)) \mathbb{Q}_{+1}^i(d\zeta) + d(\zeta, (x_{-1}, y_{-1})) \mathbb{Q}_{-1}^i(d\zeta) \leq \varepsilon, \\ & \int_{\Omega} \mathbb{Q}_{+1}^i(d\zeta) + \mathbb{Q}_{-1}^i(d\zeta) = 1. \quad \forall i. \end{aligned}$$

最后将 d 带入约束 2:

$$d(\zeta, \zeta') = \|x - x'\| + \frac{\kappa}{2} |y - y'|$$

$$\begin{aligned} & \sup_{\mathbb{Q}_{+1}^i \geq 0} \frac{1}{N} \sum_{i=1}^N \int_{\Omega} f_p(x_i) \mathbb{Q}_{+1}^i(d\zeta) + f_p(x_{-1}) \mathbb{Q}_{-1}^i(d\zeta) \\ \text{s.t. } & \frac{1}{N} \cdot \left[\sum_{i: y_i=+1} \int_{\mathbb{R}^n} \|x - x_i\| \mathbb{Q}_{+1}^i(d\zeta) + (\|x - x_i\| + \kappa) \mathbb{Q}_{-1}^i(d\zeta) \right. \\ & \left. \sum_{i: y_i=-1} \int_{\mathbb{R}^n} (\|x - x_i\| + \kappa) \mathbb{Q}_{+1}^i(d\zeta) + \|x - x_i\| \mathbb{Q}_{-1}^i(d\zeta) \right] \leq \varepsilon. \quad \text{--- ⑤} \\ & \int_{\Omega} \mathbb{Q}_{+1}^i(d\zeta) + \mathbb{Q}_{-1}^i(d\zeta) = 1. \quad \text{(5i)} \end{aligned}$$

注意这里 y 的绝对值被消除了，因为 \mathbb{Q}_{+1}^i 已经确定了 y 的取值，然后又将 y 分成两组。

到这里我们的简化就告一段落了。这个形式已经和我们之前见过的差距不大。只不过，之前的问题只需要对一个分布（向量）积分，而现在的问题需要对多个积分并求和（目标函数中）。

接下来又是熟悉的取对偶环节：

$$\begin{aligned} \textcircled{5} \Leftrightarrow & \sup_{\mathbb{Q}_{+1}^i \geq 0} \inf_{\lambda \geq 0, \delta_i} \frac{1}{N} \sum_{i=1}^N \int_{\Omega} f_p(x_i) \mathbb{Q}_{+1}^i(d\zeta) + f_p(x_{-1}) \mathbb{Q}_{-1}^i(d\zeta) \\ & + \lambda \cdot \left\{ \varepsilon - \frac{1}{N} \cdot \left[\sum_{i: y_i=+1} \int_{\mathbb{R}^n} \|x - x_i\| \mathbb{Q}_{+1}^i(d\zeta) + (\|x - x_i\| + \kappa) \mathbb{Q}_{-1}^i(d\zeta) \right] \right. \\ & \left. \sum_{i: y_i=-1} \int_{\mathbb{R}^n} (\|x - x_i\| + \kappa) \mathbb{Q}_{+1}^i(d\zeta) + \|x - x_i\| \mathbb{Q}_{-1}^i(d\zeta) \right] \\ & + \frac{1}{N} \sum_{i=1}^N \delta_i \cdot [1 - \mathbb{Q}_{+1}^i(d\zeta) + \mathbb{Q}_{-1}^i(d\zeta)]. \end{aligned}$$

为了方便后续化简，这里加入了红色部分的 $1/N$ 。

$$\inf_{\lambda \geq 0, s_i} \sup_{Q_{\pm i} \geq 0} \frac{1}{N} \cdot \left\{ \int_{R^n} \left[\sum_{i=1}^N l_p(x, 1) - \sum_{j:j \neq i} \lambda \cdot K - \lambda \sum_i \|x - \hat{x}_i\| - \sum_i s_i \right] Q_{+i}^i(d\xi) \right. \\ \left. + \int_{R^n} \left[\sum_{i=1}^N l_p(x, -1) - \sum_{j:j \neq i} \lambda \cdot K - \lambda \sum_i \|x - \hat{x}_i\| - \sum_i s_i \right] Q_{-i}^i(d\xi) \right\} \\ + \lambda \cdot \varepsilon + \frac{1}{N} \cdot \sum_{i=1}^N s_i$$

$$\rightsquigarrow \sum_{i=1}^N \int_{R^n} \left[l_p(x, 1) - \mathbb{1}_{\{y_i = 1\}} \lambda \cdot K - \lambda \cdot \|x - \hat{x}_i\| - s_i \right] Q_{+i}^i(d\xi).$$

dual:

$$\inf_{\lambda, s_i} \lambda \cdot \varepsilon + \frac{1}{N} \cdot \sum_{i=1}^N s_i \quad \text{--- ⑥}$$

s.t.

$$l_p(x, 1) - \frac{1}{2} \lambda \cdot K(1, \hat{y}_i) - \lambda \cdot \|x - \hat{x}_i\| - s_i \leq 0, \quad \forall i$$

$$l_p(x, -1) - \frac{1}{2} \lambda \cdot K(-1, \hat{y}_i) - \lambda \cdot \|x - \hat{x}_i\| - s_i \leq 0, \quad \forall i$$

$$\lambda \geq 0.$$

交换inf和sup得到对偶问题。由于红色波浪线部分可以提出求和符号，因此在问题6中约束可以分开写（相对于*i*）。

在进入下一步处理之前，还需要使用之前的定理验证强对偶定理成立。这里暂时不写。

对偶问题6还可以通过一个共轭函数的性质做简化：

Claim: 对 $\forall \lambda > 0$, 由共轭凸函数的性质，有：

$$\sup_{x \in R^n} l_p(x, \pm 1) - \lambda \|x - x'\| = \begin{cases} l_p(x', \pm 1), & \text{if } \|\beta\|_K \leq \lambda, \\ -\infty, & \text{else.} \end{cases}$$

$$\inf_{\lambda, s_i} \lambda \cdot \varepsilon + \frac{1}{N} \cdot \sum_{i=1}^N s_i \quad \text{--- ⑦}$$

s.t.

$$l_p(\hat{x}_i, 1) - \frac{1}{2} \lambda \cdot K(1, \hat{y}_i) - s_i \leq 0, \quad \forall i \text{ --- ①}$$

$$l_p(\hat{x}_i, -1) - \frac{1}{2} \lambda \cdot K(-1, \hat{y}_i) - s_i \leq 0, \quad \forall i \text{ --- ②}$$

$$\|\beta\|_K \leq \lambda.$$

其中 $\|\cdot\|_*$ 表示对偶范数。这里面的内容又很多，我也不熟悉，所以暂时不证明。

最后再做一点简化：

$$\text{约束 (1)} \Leftrightarrow \begin{cases} \ell_p(\hat{x}_i, 1) - \xi_i \leq 0 & , \text{if } \hat{y}_i = 1 \\ \ell_p(\hat{x}_i, 1) - \lambda K - \xi_i \leq 0, & \text{else.} \end{cases}$$

$$\text{约束 (2)} \Leftrightarrow \begin{cases} \ell_p(\hat{x}_i, -1) - \xi_i \leq 0 & , \text{if } \hat{y}_i = -1 \\ \ell_p(\hat{x}_i, -1) - \lambda K - \xi_i \leq 0, & \text{else.} \end{cases}$$

$$\begin{aligned} & \inf_{\lambda, \xi_i} \lambda \cdot \varepsilon + \frac{1}{n} \sum_{i=1}^n \xi_i \\ \text{s.t.} \quad & \ell_p(\hat{x}_i, \hat{y}_i) - \xi_i \leq 0, \quad \forall i \quad \text{--- (P)} \\ & \ell_p(\hat{x}_i, -\hat{y}_i) - \xi_i \leq 0, \quad \forall i \\ & \|\beta\|_2 \leq \lambda. \end{aligned}$$

在inf里加上 β 的决策变量，问题就变成了一个有限维的凸优化问题。全部工作就暂时结束了。

据苏老师所说，一般建模型文章都结束在这一步。之前的章节我们也大致是在这里停止的。不过我个人还是偏向应用、算法类文章，后续应用篇的时候再看看他们的做法吧。

下一章可能会写ADRO，也可能先看两篇应用。

[1] 苏文藻. 分布鲁棒优化简介. “数据驱动的优化理论、算法与应用”暑期学校. https://www.bilibili.com/video/BV1bM4y1L7R7?spm_id_from=333.337.search-card.all.click

[2] Peyman Mohajerin Esfahani, Daniel Kuhn. Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations. Mathematical Programming.

H2 10. 自适应分布式鲁棒优化

时间过的很快，已经来到第十章了。不出意外的话，近期内这章会是理论部分的最后一章。在这章之后我打算新开一帖，读一些应用类论文。

ADRO结合了前面两大块内容，ARO和DRO的知识。今天的内容主要根据[1]介绍。[1]主要基于第八章和第九章前半部分提到的基于一、二矩的模糊集。让我们来看看MS偏数学的文章是什么水平。

H3 特殊情况下的精确求解

在第七章ARO的解时我们提到，ARO没有很general的求解框架，只能针对一些特殊情况求精确解，例如fixed recourse, affine decision rule。ADRO也是一样。[1]的第一部分介绍了一种可精确求解的类型。

首先给出二阶段ADRO的一个表示：

$$Z^* = \min_x \left\{ c^T x + \beta(x) \right\}$$

s.t. $x \in X$

$$\beta(x) = \sup_{P \in \mathcal{P}} E_P [Q(x, \tilde{z})] \quad \text{——①.}$$

$$Q(x, z) = \max_y d^T y$$

s.t. $A(z)x + By \geq b(z)$.

$$y \in \mathbb{R}^{M_z}$$

其中 $x \in \mathbb{R}^N$, $y \in \mathbb{R}^{M_z}$, 支撑集 $w \in \mathbb{R}^{I_z}$,
 $P \in \mathcal{P}_0(\mathbb{R}^{I_z})$ 表示可测空间.

这里直接将问题分开写成多阶段的形式。第一阶段here-and-now决策变量为 x , 第二阶段wait-and-see决策变量为 y 。需要注意的是，这里和我们第五章中提到的最广义的ARO问题不一致的地方是， y 和系数 d, B 不受随机变量 z 的影响。此外，还需要引入一些假设：

$$A(z) = A^0 + \sum_{k \in I_z} A^k \cdot z_k, \quad b(z) = b^0 + \sum_{k \in I_z} b^k \cdot z_k.$$

d, B, y 与 z 无关.

代入 ①, 得到 ②.

以及，问题要满足relative complete recourse。回忆第五章，我们提到了三个概念：fixed recourse, relative complete recourse, affine decision rule。这里的假设和当初不完全一样。fixed是指 y 的系数不受 z 影响，affine是指 y 本身与 z 的关系为仿射函数。relative是指，对任何第一阶段的解 x , 第二阶段都可解。所以要注意一下我们目前讨论的问题的特殊性。

我们再考虑如下的二阶锥模糊集：

$$\tilde{F} = \left\{ \mathbb{R}_{\geq 0}(\mathbb{R}^2) \mid \begin{array}{l} \tilde{z} \sim P \\ E_P[g(\tilde{z})] = \mu \\ E_P[g_i(\tilde{z})] \leq g_i, \forall i \in [L] \\ P[\tilde{z} \in W] = 1 \end{array} \right\}$$

其中 W 和 g_i 的上境图为 SOC 可表示的集合.

$$\text{epi } g_i = \{(z, u) \in \mathbb{R}^2 \times \mathbb{R} \mid g_i(z) \leq u\}.$$

模糊集的第二个约束和第三个约束是主要内容，第二个约束是一个一阶矩的约束，第三个约束的函数 g 不一定要取线性函数，因此可以表示方差、离差的上界等约束。

这里提到一个概念：SOC representable set。我也是很清楚这个概念，找了一些相关资料：

SOC-Representable Sets and Functions

- Definition (SOC-Representable Set)

A set $\mathcal{S} \subseteq \mathbb{R}^n$ is SOC-representable if there exists a linear transformation

$$L(x) = Mx + q \text{ for } x \in \mathcal{S}$$

with $M = (\cdot)_{m \times n}$ and $q \in \mathbb{R}^m$, such that

$$L(\mathcal{S}) = \prod_{i=1}^K \mathcal{L}^{m_i} \text{ for some } K > 0 \text{ and } \sum_{i=1}^K m_i = m.$$

- Note that \mathcal{S} is SOC-representable if and only if \mathcal{S} is the x -space projection of the solution set of $K (> 0)$ SOC inequalities:

$$M_i \begin{pmatrix} x \\ \mu \end{pmatrix} + q^i \in \mathcal{L}^{m_i}, \quad i = 1, \dots, K,$$

$$\text{i.e., } \mathcal{S} = \{x \in \mathbb{R}^n \mid M_i \begin{pmatrix} x \\ \mu \end{pmatrix} + q^i \in \mathcal{L}^{m_i}, \quad i = 1, \dots, K\}.$$

- Definition (SOC-Representable Function)

A function $g(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}$ is SOC-representable if its epigraph

$Epi(g) = \{(x, t) \in \mathbb{R}^{n+1} \mid g(x) \leq t\}$ is an SOC representable set in \mathbb{R}^{n+1} .

我个人认为对后续影响不大，可以先放一放。

可以写出这个模糊集的lifted模糊集，以及对应的lifted支撑集：

$$\mathcal{L} = \left\{ P \in \mathcal{P}_0(\mathbb{R}^{I_1} \times \mathbb{R}^{I_2}) \mid \begin{array}{l} (\tilde{z}, \tilde{u}) \in \bar{P}, \\ E_{\bar{P}}[g_{\tilde{z}}] = \mu, \\ E_{\bar{P}}[u] \leq \delta, \\ P[(\tilde{z}, \tilde{u}) \in W] = 1 \end{array} \right\}$$

$$\bar{W} = \left\{ (\tilde{z}, u) \in \mathbb{R}^{I_1} \times \mathbb{R}^{I_2} \mid \tilde{z} \in W, g(\tilde{z}) \leq u \right\}.$$

lifted的方法我们在上一章DRO的第二部分用过，但当初没有正面与原先的集合之间的关系，[1]针对这个lifted给了一个性质：

Proposition 10.1: 用 $\pi_{\tilde{z}} \mathcal{L}$ 表示当 $\tilde{z} \sim P \in \mathcal{L}$ 时， \tilde{z} 的边缘分布。
有 $f = \pi_{\tilde{z}} \mathcal{L}$.

*. 进而有 $f(x) = \sup_{P \in \mathcal{L}} E_P[\alpha(x, \tilde{z})] = \sup_{P \in \mathcal{L}} E_{\bar{P}}[\alpha(x, \tilde{z})]$.

我也不是非常理解 (-_-||)，关键是注意红星的部分。

在正式处理前，还需要对支撑集 \bar{W} 给一个假设：

Assumption 10.1: (Slater's Condition)

假设 lifted 支撑集有如下显示表示:

$$\bar{W} = \{(z, u) \in \mathbb{R}^{I_1} \times \mathbb{R}^{I_2} \mid \exists v \in \mathbb{R}^{I_3}: Cz + Du + Ev \leq h\}$$

其中 \leq 表示二阶锥下的序号. C, D, E 为矩阵. 则 $\exists z^*, u^*, v^*$, 使:

$$\begin{aligned} Cz^* &= \mu, \\ u^* &\leq 6, \\ Cz^* + Du^* + Ev^* &\leq h. \end{aligned}$$

现在我们就可以正式开始对问题2进行reformulation了。

$$\begin{aligned} Z^* &= \min_x \{C^T x + f(x)\} \\ \text{s.t. } x &\in X \\ f(x) &= \sup_{R \in \mathcal{P}_0} E_R [\varrho(x, \tilde{x})] \\ \varrho(x, z) &= \min_{y \in R^M} d^T y \\ \text{s.t. } A(z)x + By &\geq b(z), \\ y &\in \mathbb{R}^{M_n} \end{aligned}$$

$$\begin{aligned} \mathcal{C}_1 &= \left\{ R \in \mathcal{P}_0(\mathbb{R}^{I_1} \times \mathbb{R}^{I_2}) \mid \begin{array}{l} (\tilde{x}, \tilde{u}) \in \bar{W}, \\ E_R[g(z)] = \mu, \\ E_R[u] \leq 6, \\ R[(\tilde{x}, \tilde{u}) \in \bar{W}] = 1 \end{array} \right\} \\ \bar{W} &= \{(z, u) \in \mathbb{R}^{I_1} \times \mathbb{R}^{I_2} \mid z \in w, g(z) \leq u\} \\ &= \{(z, u) \in \mathbb{R}^{I_1} \times \mathbb{R}^{I_2} \mid \exists v \in \mathbb{R}^{I_3}: Cz + Du + Ev \leq h\} \end{aligned}$$

问题的不确定性集中在第二阶段, 因此将第二阶段的模糊集写成子问题得到问题3:

$$\begin{aligned}
 \beta(x) &\Leftrightarrow \sup_{\mathbb{P}} \int_{\bar{W}} Q(x, z) \mathbb{P}(d(z, u)) \\
 \text{s.t.} \quad &\int_{\bar{W}} G \geq \mathbb{P}(d(z, u)) = \mu, \quad (s) \quad -- \textcircled{1} \\
 &\int_{\bar{W}} u \mathbb{P}(d(z, u)) \leq 6, \quad (t) \\
 &\int_{\bar{W}} \mathbb{P}(d(z, u)) = 1. \quad (r) \\
 &\mathbb{P} \geq 0.
 \end{aligned}$$

$$\begin{aligned}
 \textcircled{1} \Leftrightarrow \sup_{\mathbb{P} \geq 0} \inf_{\substack{t \geq 0 \\ r, s}} \mathcal{L}(\mathbb{P}, \mu, r, s) &= \int_{\bar{W}} Q(x, z) \mathbb{P}(d(z, u)) \\
 &+ s^T [\mu - \int_{\bar{W}} G \geq \mathbb{P}(d(z, u))] \\
 &+ t^T [6 - \int_{\bar{W}} u \mathbb{P}(d(z, u))] \\
 &+ r \cdot [1 - \int_{\bar{W}} \mathbb{P}(d(z, u))]
 \end{aligned}$$

问题3写成拉格朗日对偶形式，因为约束都是线性的，所以对偶不需要内积：

$$\begin{aligned}
 \sup_{\mathbb{P} \geq 0} \mathcal{L}(\mathbb{P}, \mu, r, s) &= s^T \mu + t^T G + r \\
 &+ \int_{\bar{W}} [Q(x, z) - s^T G z - t^T u - r] \mathbb{P}(d(z, u)) \\
 \beta_1(x) &= \inf_{\substack{r, s, t \\ t \geq 0}} r + s^T \mu + t^T G \\
 \text{s.t.} \quad &r + s^T G z + t^T u \geq Q(x, z), \quad \forall (z, u) \in \bar{W}. \quad -- \textcircled{4}
 \end{aligned}$$

其对偶形式记作问题4， β_1 。注意这里我们没证明强对偶。

接下来对约束中的 Q 取对偶：

$$Q(x, z) = \min_{\substack{y \\ \text{s.t.} \\ y \in \mathbb{R}^m}} d^T y$$

$$\text{s.t. } A(z)x + By \geq b(z). \quad (\text{P}) \Leftrightarrow \min_p \max_y L(y, p) = d^T y - p^T [A(z)x + By - b(z)]$$

$$\max_p \min_y L(y, p) = (d - B^T p)^T y + p^T [b(z) - A(z)x] \Leftrightarrow \max_p p^T [b(z) - A(z)x]$$

$$\text{s.t. } B^T p = d.$$

由 LP 的强对偶，上下等价。令 $P_p, p \in [P]$ 表示多面体 $P = \{p \in \mathbb{R}_+^m : B^T p = d\}$ 的极点。有：

$$Q(x, z) = \max_{p \in P} p^T [b(z) - A(z)x] = \max_{p \in [P]} \left\{ p^T [b(z) - A(z)x] \right\} \quad \text{--- (5)}$$

这里用到的只是 LP 的强对偶，可以直接套公式得出。然后对对偶问题的多面体的极点进行枚举，因此问题 5 是 $\max \{\cdot\}$ 的形式，而非 LP 问题的形式。

将上述处理后的 Q 带入原问题，有：

把 (5) 代入 (4)，由 $A(z) = A^\circ + \sum_{k \in [l]} A^k z_k, b(z) = b^\circ + \sum_{k \in [l]} b^k z_k$ ，有

$$(4): P_1(x) = \inf_{r, s, t} r + s^T \mu + t^T \sigma$$

$$\text{s.t. } r \geq \sup_{(z, u) \in \bar{W}} \left\{ \left(\begin{bmatrix} p_p^T(b^\circ - A^\circ x) \\ \vdots \\ p_p^T(b^k - A^k x) \end{bmatrix} - G^T s \right)^T z - t^T u + p_p^T(b^\circ - A^\circ x) \right\}, \forall p \in [P].$$

--- (6)

t 30.

这里的约束有无限个，相当于一个经典 RO 问题。对约束取子问题处理：

$$\sup_{z, u, v} \left(\begin{bmatrix} p_p^T(b' - A'x) \\ \vdots \\ p_p^T(b'' - A''x) \end{bmatrix} - C^T s \right)^T z - t^T u \quad \text{--- ⑦}$$

$$\text{s.t. } Cz + Du + Ev \leq h. \quad (\bar{x}_p)$$

$$\sup_{z, u, v} \inf_{\bar{x}_p \succcurlyeq 0} \mathcal{L}(z, u, v, \bar{x}_p) = \left(\begin{bmatrix} p_p^T(b' - A'x) \\ \vdots \\ p_p^T(b'' - A''x) \end{bmatrix} - C^T s \right)^T z - t^T u + \langle \bar{x}_p, h - Cz - Du - Ev \rangle.$$

$$\begin{aligned} \mathcal{L}(z, u, v, \bar{x}_p) &= \bar{x}_p^T \cdot h \\ &+ \left(\begin{bmatrix} p_p^T(b' - A'x) \\ \vdots \\ p_p^T(b'' - A''x) \end{bmatrix} - C^T s - C^T \cdot \bar{x}_p \right)^T z \\ &+ (t + D^T \bar{x}_p)^T u \\ &- (E \cdot \bar{x}_p)^T v \end{aligned}$$

$$\begin{aligned} \textcircled{7} &\leq \inf_{\bar{x}_p} \bar{x}_p^T \cdot h \\ \text{s.t. } C^T \cdot \bar{x}_p &= \begin{bmatrix} p_p^T(b' - A'x) \\ \vdots \\ p_p^T(b'' - A''x) \end{bmatrix} - C^T s \quad \text{--- ⑧} \end{aligned}$$

$$D^T \bar{x}_p = -t$$

$$E^T \bar{x}_p = 0$$

$$\bar{x}_p \succcurlyeq 0$$

注意，这里的不确定集范围是支撑集 \bar{W} ，而不是 \mathcal{G} 。此处的对偶是之前提到的对二阶锥规划的对偶，因此要取内积，这里的 \preceq 也是定义在二阶锥上的。注意二阶锥的对偶锥依然是二阶锥。此外，我们依然没证明强对偶。将对偶结果带回问题6，得到：

将⑧代入⑥，有：

$$\begin{aligned}\beta_2(x) &= \inf_{r, s, t, \lambda_p} r + s^T \mu + t^T \delta \\ \text{s.t. } &r \geq \lambda_p^T h + p_p^T (b^0 - A^0 x), \quad \forall p \\ C^T \lambda_p &= \begin{bmatrix} p_p^T (b^1 - A^1 x) \\ \vdots \\ p_p^T (b^n - A^n x) \end{bmatrix} - C^T s, \quad \forall p. \quad \text{--- ⑨.}\end{aligned}$$

$$\begin{aligned}D^T \lambda_p &= -t, \quad \forall p. \\ E^T \lambda_p &= 0, \quad \forall p. \\ \lambda_p &\geq 0, \quad \forall p. \\ t &\geq 0.\end{aligned}$$

$$\beta_1(x) \leq \beta_2(x) \leq \beta_3(x). = \beta_3(x) \leq \beta(x).$$

此时问题就是一个有限约束的线性问题了。但是，由于中间两次对偶的强对偶没有证明，每个问题的解之间只有黑字的弱对偶关系。要证明强对偶成立，可以构造红字的关系，证明问题等价。这里构造的 $\beta_3(x)$ 其实是问题9的对偶问题，由前面的Assumption可以证明存在可行解，并进一步由不等式关系得到 $\beta(x)$ 。因为有一些地方没太看懂，所以暂时不往后写了。

H3 Linear Decision Rule近似解

[1]还包含了一种使用linear decision rule的近似解。LDR其实affine decision rule。考虑如下问题：

$$\begin{aligned}\beta_{\text{LDR}} &= \min_y \sup_{\mathbb{E}f} C^T x + \mathbb{E}_P [d^T y(z)] \quad \text{--- ⑩} \\ \text{s.t. } &A(z)x + B(y(z)) \geq b(z), \quad \forall z \in W,\end{aligned}$$

[1]认为这和问题1中的 β 一致，但我觉得不太一样。我们在第五章也提到类似情况。在问题11中，约束要求对所有 z 都满足，但在问题1中，只需要对计算期望值的 z 满足即可。问题11与问题1的另一个差别在于 y 不再是与 z 无关的。考虑如下LDR：

$$\begin{aligned}
\beta_L(x) = \min_y \quad & \sup_{\mathbf{P} \in \tilde{\mathcal{F}}} \mathbf{c}^T x + E_{\mathbf{P}}[d^T y(z)] \\
\text{s.t.} \quad & A(z)x + By(z) \geq b(z), \quad \forall z \in W, \\
& y \in \mathcal{L}^N
\end{aligned}
\quad - (12)$$

$$\mathcal{L}^N = \left\{ y \in \mathbb{R}^{I_1, N} \mid \begin{array}{l} \exists y^0, y^1_i \in \mathbb{R}^N, i \in [I_1], \\ y(z) = y^0 + \sum_i y^1_i z_i \end{array} \right\}$$

但是LDR存在一些问题。例如，一些函数形式，例如 $|z|$ ，无法用仿射函数表示。因此[1]考虑添加辅助变量：

$$\begin{aligned}
\beta_E(x) = \min_y \quad & \sup_{\mathbf{P} \in \tilde{\mathcal{F}}} \mathbf{c}^T x + E_{\mathbf{P}}[d^T y(z, u)] \\
\text{s.t.} \quad & A(z)x + By(z, u) \geq b(z), \quad \forall z \in \bar{W}, \\
& y \in \bar{\mathcal{L}}^N \\
\bar{\mathcal{L}}^N = \left\{ y \in \mathbb{R}^{I_1 + I_2, N} \mid \begin{array}{l} \exists y^0, y^1_i, y^2_j \in \mathbb{R}^N, i \in [I_1], j \in [I_2] \\ y(z, u) = y^0 + \sum_i y^1_i z_i + \sum_j y^2_j u_j \end{array} \right\}
\end{aligned}
\quad - (13)$$

这里的支撑集与前文的 \bar{W} 一致。在这个辅助变量下，可表示的 $y(z)$ 关系就更多了。因此，问题13有性质： $\beta(x) \leq \beta_E(x) \leq \beta_L(x)$ 。

用前文的reformulation方法，可以得到问题13的如下形式：

$$\begin{aligned}
\beta_E(\mathbf{x}) = \min \quad & \{r + \mathbf{s}' \boldsymbol{\mu} + \mathbf{t}' \boldsymbol{\sigma}\} \\
\text{s.t.} \quad & r + \mathbf{s}' (\mathbf{G}z) + \mathbf{t}' \mathbf{u} \geq \mathbf{d}' \mathbf{y}(z, \mathbf{u}), \quad \forall (z, \mathbf{u}) \in \bar{W}, \\
& \mathbf{A}(z)\mathbf{x} + \mathbf{B}\mathbf{y}(z, \mathbf{u}) \geq \mathbf{b}(z), \quad \forall (z, \mathbf{u}) \in \bar{W}, \\
& \mathbf{t} \geq 0, \\
& r \in \mathbb{R}, \mathbf{s} \in \mathbb{R}^{L_1}, \mathbf{t} \in \mathbb{R}^{I_2}, \\
& \mathbf{y} \in \bar{\mathcal{L}}^{N_2}.
\end{aligned}
\quad (16)$$

这个问题是一个AARO问题，去除了分布式的成分，就可以用已知的ARO方法求解了。

上面描述的两种方法可以简单地拓展到multi-stage的形式，只需要将决策变量 y 变成 y_i 即可。[1]中还有对LDR其他性质的描述，这里就不多介绍了。

本来想在这章结束理论篇内容，但在咨询了老师后新增了一些读物，可能在未来几天对过去的内容（包括本章）缝缝补补。敬请期待。

[1] Dimitris Bertsimas, Melvyn Sim, Meilin Zhang. Adaptive Distributionally Robust Optimization. Management Science.

[2] Zhi Chen, Melvyn Sim, Peng Xiong. Robust Stochastic Optimization Made Easy with RSOME. Management Science.