Generalized Benders Decomposition 学习笔记

2019 年 11 月 13 日

1 原始问题

原始问题如下：

$$\begin{aligned} &\max_{x,y} f(x,y) \\ &s.t. \quad G(x,y) \succeq 0, x \in X, y \in Y \end{aligned} \tag{1}$$

其中，$G(x,y)$ 所得结果为一个 m 维向量（也即有 m 个不等式约束）, 假设在(1)中，y 是复杂变量，即当 y 是固定值的时候，(1)就会变得很好解，这主要体现在一下几个可能的方面：

(a) 对于固定的 y，问题(1)可以分解为若干个独立的优化问题，每一个子优化问题都是 x 的一个子向量，且互相不交叠；

(b) 问题(1)可能是一个可以有效求解的著名的问题结构；

(c) 问题(1)对于 x 和 y 联合起来并不是 concave 的，但是固定 y 之后对于 x 就是 concave 的。

为了将 y 固定，进行的操作是将问题投影到 y 空间上，(projection/partitioning)，经过映射或者是投影之后，问题变为了如下的形式：

$$\begin{aligned} &\max_{y} \quad v(y) \\ &s.t. \quad y \in Y \cap V \end{aligned} \tag{2}$$

其中，$v(y)$ 与 V 的定义如下：

$$\begin{aligned} v(y) &:= \sup_{x} f(x,y) \\ s.t. \quad & G(x,y) \succeq 0, x \in X \end{aligned} \tag{3}$$

$$V := \{y | G(x,y) \succeq 0, \exists x \in X\} \tag{4}$$

实际上，Y 是 y 原来的取值范围，而 V 则是可行的 y 的取值范围，两者求交集，则是在限定取值范围内的可行解 y 的范围。注意到，$v(y)$ 是在给定 y 的情况下，(1)的最优取值，因为(3)太常用了，所以用（1-y）来指代(3)中的优化问题。

$$\max_{x \in X} f(x,y), s.t. G(x,y) \succeq 0 \qquad (1-y)$$

在 Benders 原始的工作中，只考虑了通过线性对偶理论来定义目标函数和限制条件，这样 v 和 V 就可以用有限数量的近似就能得到有限数量的近似。

$$X := \{x|x \succeq 0\} \tag{5a}$$

$$f(x,y) := c^t x + \phi(y) \tag{5b}$$

$$G(x,y) := Ax + g(y) - b \tag{5c}$$

但是本文的主要工作是对 Benders 工作的扩展，扩展到非线性对偶理论的一般情况下, 这样可以解决的问题的范围就增加扩大了。

2 泛化的 Benders 分解

这一章主要从 5 个方面来讨论(1)这样的问题的 Benders 分解。第 1 部分讨论的是如何推导出 master 问题，核心的思想就在变为(2)之后利用对偶的形式来标识 v 和 V；第 2 部分主要是通过解决一系列的子问题 subproblem 来解决 master 问题（主要是通过计算（1-y）不同试验 y 值的最优乘子向量）；第 3 部分主要陈述了 benders 分解的主要步骤；第 4 部分讨论了收敛性的问题，第 5 部分讨论了计算上的一些问题。

2.1 Master Problem 的推导

master problem 的获得主要通过三步来达到：

(i) 将(1)映射到 y 空间，得到(2)的形式

(ii) 触发 V 的 natural dual representation（自然对偶表示?），也就是一些包含 V 区域的重合的部分，invoke the natural representation of V in terms of the intersection of a collection of regions that contain it

(iii) 触发 v 的 natural dual representation（自然对偶表示?），invoke the natural dual representation of v in terms of the pointwise infimum of a collection of functions that dominate it.

2.1.1 步骤 1，Projection

定理 *1*，*Projection* 当且仅当问题(2)不可行或者是无上界时，问题(1)才会不可行或者是无上界。如果 x^*, y^* 是(1)的最优解，那么 y^* 一定是(2)的最优解。如果 y^* 是(2)的最优解，并且在 $y = y^*$ 时通过 x^* 达到了(3)的最大值，则 (x^*, y^*) 同样也是(1)的最优解。如果说 $\bar{y}$ 是(2)的 $\epsilon_1 - optimal$,$\bar{x}$ 是（1-y）的 $\epsilon_2 - optimal$，则 $(\bar{x}, \bar{y})$ 是问题(1)的 $(\epsilon_1 + \epsilon_2) - optimal$

2.1.2 步骤 2，V-Representation

定理 *2*，*V-Representation* 假设 1：X 是非空凸集，对于给定的 $y \in Y$，G 在 X 上是 concave 的；假设 2：对于给定 $y \in Y$，集合 $Z_y := \{z \in R^m | G(x,y) \succeq z, \exists x \in X\}$ 是闭集（这一假设实际上

假设的是对于给定的 y，G(x,y) 的每一个维度的最大值都可以取到，而不是开集，最大值也不是无穷的（?)）。那么，对于点 $\bar{y} \in Y$，当且仅当(6)成立时，其也在集合 V 中。

$$
\begin{aligned}
&[\sup_{x \in X} \lambda^t G(x, \bar{y})] \succeq 0, \forall \lambda \in \Lambda \\
where& \\
&\Lambda := \{\lambda \in R^m | \lambda \succeq 0 \quad and \quad \sum_{i=1}^{m} \lambda_i = 1\}
\end{aligned} \tag{6}
$$

证明，如果点 $\bar{y}$ 在 V 中，(6)自然是成立的，也不必证明。反过来，通过(6)来证明 $\bar{y}$ 则可以通过非线性对偶理论。假设 $\bar{y}$ 满足(6)，那么

$$
\inf_{\lambda \in \Lambda} [\sup_{x \in X} \lambda^t G(x, \bar{y})] \geqslant 0
$$

进一步推导，得出

$$
\inf_{\lambda \geqslant 0} [\sup_{x \in X} \lambda^t G(x, \bar{y})] = 0 \tag{7}
$$

通过(7), 可以断言以下 concave 优化(8) 关于 G 约束的对偶函数（$inf_{\lambda \geqslant 0}[\sup_{x \in X} \lambda^t G(x, \bar{y})]$）是可以达到最优值 0 的（即(8)的对偶是有可行解的，对于给定 $\bar{y}$，存在 x 使得 $G(x, \bar{y}) \geqslant 0$）。

$$
\begin{aligned}
\max_{x \in X} \quad & 0^t x \\
s.t. \quad & G(x, \bar{y}) \geqslant 0
\end{aligned} \tag{8}
$$

根据参考定理 5.1，如果(8)的对偶是有可行解的，那么原问题(8)也必定有可行解，因此，可以证明 $\bar{y} \in V$。另外，关于 Z_y 是闭集的限制也不是严格的，有些情况可以不必是闭集。

2.1.3 步骤 3，v-Representation

定理 *3* 假设 1：X 是非空凸集，f 和 G 在 X 上对于给定的 $y \in Y$ 是 concave 的；假设 2：对于给定的 $\bar{y} \in Y \cap V$，以下三种情况至少有一种会出现：

(1) $v(\bar{y})$ 有限，且 (1-$\bar{y}$) 拥有**最佳乘子向量**（定义见附录，大意是在拉格朗日对偶的情况下，最佳乘子向量与最优解条件下的限制条件满足互补松弛性）；该情况是常见的情况

(2) $v(\bar{y})$ 有限，$G(x, \bar{y})$ 与 $f(x, \bar{y})$ 在 X 上连续（X 是闭集），同时 $\exists \epsilon \geqslant 0$ 使得 (1-$\bar{y}$) 问题的 $\epsilon - optimal$ 解集是非空并且有限的；

(3) $v(\bar{y}) = +\infty$

如此一来，(1-y) 的最优值等于它的对偶函数在 $Y \cap V$ 上的最优取值，即

$$
\begin{aligned}
v(y) = \inf_{u \succeq 0} [\sup_{x \in X} f(x, y) + u^t G(x, y)] \\
\forall y \in Y \cap U
\end{aligned} \tag{9}
$$

2.1.4 总结

经过上述三个步骤的定理与证明，问题(1)可以被转变为如下等价的 master problem 形式：

$$\max_{y\in Y}[\inf_{u\succeq 0}[\sup_{x\in X} f(x,y)+u^tG(x,y)]], \quad s.t.(6)$$

或者，另一种表述形式是利用 infimum 是最小下界的定义，将上述问题转化为如下的形式：

$$\max_{y\in Y,y_0} y_0 \tag{10a}$$

$$s.t. \quad y_0 \leqslant \sup_{x\in X}\{f(x,y)+u^tG(x,y)\}, \forall u\succeq 0 \tag{10b}$$

$$\sup_{x\in X} \lambda^t G(x,y) \geqslant 0, \forall \lambda \in \Lambda \tag{10c}$$

其中，(10a)是将上式中的 $\max_y(\cdot)$ 换成了一个变量，然后(10b)是将 inf 的定义转化为最小下界，即令 $\sup_{x\in X}\{f(x,y)+u^tG(x,y)\}$ 取得最小值的 u 带入之后所得的最小值，与 inf 一致，最后(10c)则是保证(6)的限制被满足，即 $y\in V$。

2.2 求解 master problem

因为 master problem 一般情况下是有大量的限制条件的，因此为了提升求解的效率，直接的做法是进行松弛 (relaxation)。最开始首先忽略大部分的约束条件，只考虑(10b)(10c)中的部分限制条件，进行求解，将所求的“最优解”（次出的最优解既可以指准确的全局最优，也可以指 $\epsilon - optimal$）代入到先前被忽略的限制条件中去检验是否满足所有限制，如果满足，则达到了最优解；如果不满足，则加入被违背的限制条件，重新求解，直到求得满足所有限制条件的最优解。上述思想比较简单，但是在这个过程中存在着两个问题，第一个问题是针对松弛问题所求得的最优解，如何去验证它们满足所有限制条件；第二个是如果是不满足所有限制条件的话，哪些限制条件需要被加到新的松弛优化问题。解决上述问题的一个思路如下：假设 $(\hat{y},\hat{y_0})$ 是松弛问题(10)的最优解，那么我们需要验证其满足(10b)(10c)的限制。从定理 2 与 V 的定义来看，当且仅当 (1-$\hat{y}$) 有可行解的时候，$\hat{y}$ 才能满足(10c)的限制要求（因为 (1-$\hat{y}$) 有可行解的限制条件就是对于指定的 $\hat{y}$，可行解满足 $G(x,\hat{y}\succeq 0)$, 自然有 $\sup_x G(x,\hat{y})\succeq 0$）。另外，如果 (1-$\hat{y}$) 有可行解，定理 3 表明当且仅当松弛最优解小于等于考虑了所有限制条件的最优解即 $\hat{y_0}\leqslant v(\hat{y})$ 时，$(\hat{y},\hat{y_0})$ 满足约束(10b)的要求（定理 3 指明在给定的 $\hat{y}$ 的情况下，考虑了所有约束条件的 (1-$\hat{y}$) 的最优取值为 $v(\hat{y})$，如果 $\hat{y_0}\geqslant v(\hat{y})$ 的话，则说明其一定违背了某些限制条件）。如此一来，(1-$\hat{y}$) 便成了验证松弛最优解是否可行的非常合适的子问题 subproblem，而在给定松弛最优解的自变量 $\hat{y}$ 时，(1-$\hat{y}$) 就变得非常容易求解。假设松弛最优解 $(\hat{y},\hat{y_0})$ 违背了被忽略的限制条件，在求解 (1-$\hat{y}$) 的过程中，如果 (1-$\hat{y}$) 没有可行解，则大多数的对偶类型的解法都会计算出向量 $\hat{\lambda}$ 使得(11b)被满足；如果 (1-$\hat{y}$) 有可行解，且能计算出一个有限的最优解，则会算出一个向量 $\hat{u}$ 满足(11a), 实际上大多数现在的解法都能解出 $\hat{u}$，如果解不出来，则一定是有无界的最优，对应着原问题(1)也是有着无界最优，或者是有有限最优解，但是处于一种异常状态，使得通过(9)算出的 $\hat{u}$ 当其足够靠近最优的时候满足(11a)，这样的 $\hat{u}$ 被称为 near-optimal 乘子向量。

$$\hat{y_0} > \sup_{x\in X}\{f(x,\hat{y})+\hat{u}^tG(x,\hat{y})\} \tag{11a}$$

$$\sup_{x\in X}\{\hat{\lambda}^tG(x,\hat{y})\} < 0 \tag{11b}$$

总之，通过对偶的方式求解 (1-$\hat{y}$) 可以有效的验证松弛最优解是否合理：如果 (1-$\hat{y}$) 不可行，则会算出 $\hat{\lambda} \in \Lambda$ 满足违背条件(11b)；如果可行，可以算出最佳乘子向量 $\hat{u}$ 满足(11a); 或者算不出最佳乘子向量，但是可以算出 near-optimal 的乘子向量满足(11a)（$\hat{y_0}$ 超过了 $v(\hat{y})$）。这里可能会有个疑问，为什么(10b)里边的限制条件明明限制了 y_0 的取值不能大于 $v(y)$，为什么还会算出 $\hat{y_0} > v(\hat{y})$ 呢？这是因为在实际的求解 master problem 过程中，并没有考虑完整的限制条件，而是只是考虑了部分的限制条件，所以才可能出现算出的 $\hat{y_0}$ 大于考虑所有限制条件的全局最优解 $v(\hat{y})$

2.3 求解过程描述

假设：定理 2 与定理 3 的假设依旧成立，同时假设问题(1)有有限的最优解。定义函数如下：

$$L^*(y;u) := \sup_{x\in X}\{f(x,y) + u^t G(x,y)\}, \qquad y \in Y, u \succeq 0 \tag{12a}$$

$$L_*(y;\lambda) := \sup_{x\in X}\{\lambda^t G(x,y)\}, \qquad y \in Y, \lambda \succeq 0 \tag{12b}$$

具体步骤如下：

Step 1 选择 $\bar{y} \in Y \cap U$，解决子问题 (1-$\bar{y}$)，获得最佳或者 near-optimal 乘子向量 $\bar{u}$，获得函数 $L^*(y;\bar{u})$。令 $p = 1, q = 0, u^1 = \bar{u}, LBD = v(\bar{y})$，选择收敛容忍度参数 $\epsilon > 0$。其中，LBD 是可行解能取得的最优值的下界，即全局最优值至少能达到 LBD。

Step 2 通过适当的方法求解当前迭代阶段的松弛主问题(13)，所得到的解 $(\hat{y}, \hat{y_0})$ 为当前松弛最优解，$\hat{y_0}$ 为问题(1)的最优值的上界（即考虑了部分的限制条件下的最优解，拥有更多限制条件的原始问题不可能取得比 $\hat{y_0}$ 更好的解了），如果 $LBD \geqslant \hat{y_0} - \epsilon$, 则迭代停止。

$$\begin{aligned} \max_{y\in Y, y_0} \quad & y_0 \\ s.t. \quad & y_0 \leqslant L^*(y;u^j), j = 1,2,\cdots,p, \\ & L_*(y;\lambda^j) \geqslant 0, j = 1,2,\cdots,q, \end{aligned} \tag{13}$$

Step 3 求解更新过的子问题 (1-$\hat{y}$)，然后检查下列情形：

Step 3A $v(\hat{y})$ 的值是有限的，如果 $v(\hat{y}) \geqslant \hat{y_0} - \epsilon$, 则迭代终止，达到了 $\epsilon - optimal$。否则，说明当前的松弛最优解 $\hat{y}$ 导致部分限制条件被违背，所以原问题的最优解(1)无法通过 $\hat{y}$ 来达到合理的程度，通过再次求解 (1-$\hat{y}$) 获得新的最优乘子向量 $\hat{u}$（如果求不到，则找 near-optimal）进而得到新的函数 $L^*(y;\hat{u})$。将 p 加 1，令 $u^p = \hat{u}$，如果 $v(\hat{y}) > LBD$，则令 $LBD = v(\hat{y})$，更新 LBD，（$v(\hat{y})$ 必定是可达到的），然后重复 step 2 进行检查。

Step 3B 问题 *(1-$\hat{y}$)* 不可行，则算出满足(11b)的 $\hat{\lambda} \in \Lambda$，同时得到函数 $L_*(y;\hat{\lambda})$，令 q 加 1，然后令 $\lambda^q = \hat{\lambda}$，然后返回 Step 2.

有几点值得注意。（1）在 Step2 中，连续找到的 $\hat{y_0}$ 必须是单调非增的，而在 Step3 中，找到的 $v(\hat{y})$ 则不必也未必是单调非减的，这也是引入 LBD 的原因。（2）令人欣慰的是，可以证明，在 Step3A 中，所得到的最优乘子向量 $\hat{u}$ 所对应的约束条件也是被违背最多的约束条件（直观来看，$\hat{u}$ 应该是会在 $G(x,\hat{y})$ 为负的维度上数值最大，以便实现 $\hat{u}^t G(x,\hat{y})$ 的最小），而 near-optimal 乘子向量能够在多大程度上刻画限制条件的被违背程度，则取决于其能够在多大程度上接近 (1-$\hat{y}$) 对偶的最优。同理，$\hat{\lambda}$ 则取决于当 $\bar{y} = \hat{y}$ 时其多大程度达到(8)对偶的最优（对偶变量正规化到 Λ 上）。

2.4 收敛性分析

2.5 计算细节