Lagrangian Duality for Dummies

By David Knowles

Translated by hittle

来源：

https://cs.stanford.edu/people/davidknowles/lagrangian_duality.pdf

我们想解决下面这个优化问题

$$\min f_0(x) \tag{1}$$

$$\text{such that } f_i(x) \leq 0 \qquad \forall i \in 1, \ldots, m \tag{2}$$

我们先不对函数的凸性做任何假设。为了简单起见我们也没有给出不等式约束，但是后面我们会直接加上这些约束条件。一个很直接（也很笨）的解决办法也是优化下面这个函数：

$$J(x) = \begin{cases} f_0(x), & \text{if } f_i(x) \leq 0 \ \forall i \\ \infty, & \text{otherwise} \end{cases} \tag{3}$$

$$= f_0(x) + \sum_i I[f_i(x)] \tag{4}$$

这儿 I[u]是一个无穷阶跃函数：

$$I[u] = \begin{cases} 0, & \text{if } u \leq 0 \\ \infty, & \text{otherwise} \end{cases} \tag{5}$$

如果不满足约束条件，阶跃函数就会给出一个∞的惩罚。如果能最小化 J(x)，就能解决我们的问题。但是呢从优化的角度来说，J(x)是一个非常糟糕的函数，因为它既不可微也不是连续的。能不能用一个更好的函数替换 I[u]呢？表示直线的函数λu 处理起来更简单。选用后者可能看起来很蠢， 但是在λ大于等于 0 的情况下，如果约束条件不满足，惩罚项是在正确的方向上，并且λu 是 I[u]的一个下界(见图 1)。

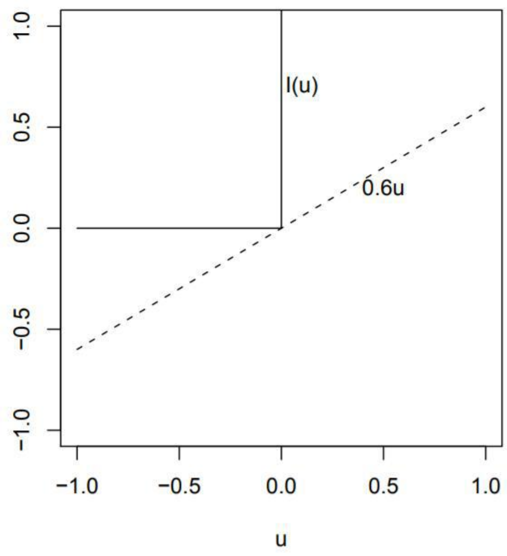

Figure 1: The infinite step function $I(u)$ and the linear relaxation λu. For $\lambda \geq 0$ note that λu is a lower bound on $I(u)$.

如果用λu 替换 J(x)中的 I[u]， 我们就得到一个 x 和λ的拉格朗日函数：

$$L(x, \lambda) = f_0(x) + \sum_i \lambda_i f_i(x) \tag{6}$$

应该看到，在求 L 相对于λ的最大值时，对 L 取λ的偏导数就得到原来的函数 J(x)。对 x 的每一个取值，如果函数满足约束条件，也即

$$f_i(x) \leq 0 \ \forall i$$

那么我们尽量能做的就是让

$$\lambda_i = 0 \ \forall i,$$

从而有

$$L(x, 0) = f_0(x).$$

当然，也可以让 $\lambda_i \to \infty$ ，从页 L(x, λ)也趋于无穷大。综上，我们有

$$\max_\lambda L(x, \lambda) = J(x)$$

这个结论很不错，但是它对解决原来那个优化问题有什么作用呢？注意到我们是要求 J(x)的最小值，现在变成了求

$$\min_{x} \max_{\lambda} L(x, \lambda) \quad (8)$$

求这个问题很难。但是如果我们改变一下求的顺序会怎么样？我们有

$$\max_{\lambda} \min_{x} L(x, \lambda) = \max_{\lambda} g(\lambda) \quad (9)$$

这儿 g(λ) = $\min_x L(x, \lambda)$,即所谓的对偶函数。求对偶函数的极大值问题就变成了所谓的对偶问题（dual problem），这有别于原始问题(primal problem)。因为 g(λ)是仿射函数的点态极小值（L(x, λ)是仿射函数），所以它是一个凹函数。求 L(x, λ)相对于 x 的极小值可能很困难。但是因为 g(λ)是一个凹函数并且在λ大于等于 0 是一个线性约束条件，所以求 g(λ)相对于λ的极大值是一个凸优化问题，也是一个简单的问题。但是解决这个问题和原来那个问题有关系吗？注意到λu 是 I[u]的一个下界。所以，对所有的 $\lambda \geq 0$,L(x, λ)是 J(x)的一个下界。所以有，

$$\begin{aligned} L(x, \lambda) &\leq J(x) \qquad \forall \lambda \geq 0 && (10) \\ \Rightarrow \min_{x} L(x, \lambda) = g(\lambda) &\leq \min_{x} J(x) =: p^\star && (11) \\ \Rightarrow d^\star = \max_{\lambda} g(\lambda) &\leq p^\star && (12) \end{aligned}$$

这儿 p 星和 q 星分别是原始问题和对偶问题的最优解。可以看到，对任意的λ，对偶函数 g(λ)是优化问题的一个下界。我们可将求偶函数的极大值问题解释为找到原始问题最大下界的问题，即

$$\max_{\lambda} \min_{x} L(x, \lambda) \leq \min_{x} \max_{\lambda} L(x, \lambda) \quad (13)$$

这个不等式被称为弱对偶问题，这个不等式对所有光滑函数都是成立的。$p^\star - d^\star$ 称为对偶距离。强对偶意味着这两者相等，也即对偶距离为 0。如果优化问题是凸的并且存在一个严格的可行点(一个满足所有约束条件的点)，那么对偶问题和原始问题的解是等价的。

KKT 条件

当梯度为 0 时，我们可以得到非约束凸优化问题的全局最小值。KKT 条件和约束凸优化问题

的全局最小值条件是等价的。如果强对偶条件成立，并且(x*，λ*)是对应的最优解，那么 x* 可以最小化 L(x，λ*)，这也是第一个 KKT 条件：

$$\nabla_x L(x^\star, \lambda^\star) = \nabla_x f_0(x^\star) + \sum_i \lambda_i^\star \nabla_x f_i(x^\star) = 0 \tag{14}$$

我们可以这样解释这个式子：目标函数和约束条件函数的梯度必须相互平行（并且方向相反）。这也意味着，沿着约束函数曲面移动不会提高目标函数。图 2 中有一个 2D 呈现说明优化一个带不等式约束问题的图。

另外，通过定义我们也可得出：

$$f_0(x^\star) = g(\lambda^\star) = \min_x L(x, \lambda^\star) \leq f_0(x^\star) + \sum_i \lambda_i^\star f_i(x^\star) \leq f_0(x^\star) \tag{15}$$

因为

$$\sum_i \lambda_i^\star f_i(x^\star) \leq 0$$

所以我们有

$$\sum_i \lambda_i^\star f_i(x^\star) = 0 \tag{16}$$

这也是第二个 KKT 条件。

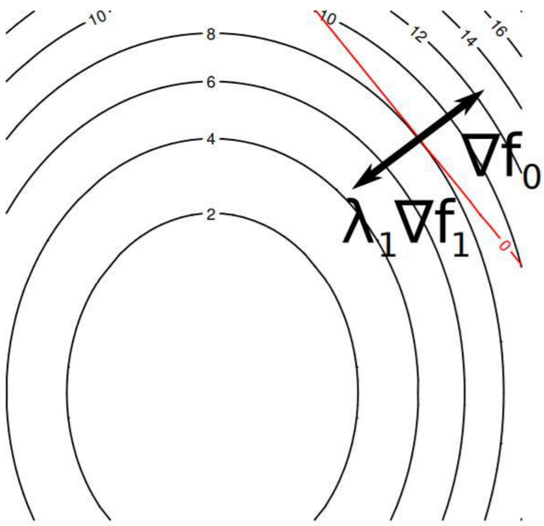

Figure 2: The first KKT condition, $\nabla_x f_0(x^\star) + \lambda_1^\star \nabla_x f_1(x^\star) = 0$. The black contours are the objective function, the red line is the constraint boundary. At the optimal solution the gradient of the objective and constraint must be parallel and opposing so that no direction along the constraint boundary could give an improved objective value.