拉格朗日对偶与KKT条件

布衣之莠

November 22, 2018

1 问题提出:

1.1 问题1:

有下面最优化问题:

$$\min f(x) \tag{1.1.1}$$

$$s.t. \quad h_j(x) = 0 \tag{1.1.2}$$

$$g_k(x) \leq 0 \tag{1.1.3}$$

1.2 问题2:

在问题1的情况下,定义

$$L(x, \mu, \lambda) = f(x) + \sum_j \mu_j h_j(x) + \sum_k \lambda_k g_k(x) \tag{1.2.1}$$

其中μ_j任意取值，$\lambda_k \geq 0$，那么，把L看做x, μ, λ的三元函数，将有

$$f(x) = \max_{\mu,\lambda;\lambda_k \geq 0} L(x, \mu, \lambda) \tag{1.2.2}$$

所以问题 1 的解决可由下式给出:

$$\min_x f(x) = \min_x \max_{\mu,\lambda;\lambda_k \geq 0} L(x, \mu, \lambda) \tag{1.2.3}$$

1.3 问题3:

定义函数$J(x, \mu, \lambda)$,那么有下面弱对偶性质(weak duality):

$$\max_{\mu,\lambda} \min_x J \leq \min_x \max_{\mu,\lambda} J \tag{1.3.1}$$

2 问题讨论:

2.1 先讨论问题 1 和问题 2 之间的关系:

问题 2 中只有 1 个约束就是$\lambda_k \geq 0$。那么怎么得出(1.2.2)呢?
考虑

$$\max_{\mu,\lambda;\lambda_k\geq 0} L(x,\mu,\lambda) = \max_{\mu,\lambda;\lambda_k\geq 0}[f(x)+\sum_j \mu_j h_j(x)+\sum_k \lambda_k g_k(x)] \tag{2.1.1}$$

如果最大值存在，那么公式(2.1.1)中$h_j(x) = 0$,否则由于μ_j的任意性，总能使$L(x,\mu_j,\lambda_k) \to \infty$(只要取$\mu_j$和$h_j(x)$同号且足够大)。而且有$g_k(x) \leq 0$,否则若$g_k(x) > 0$,在符合约束条件下取$\lambda_k$无限大，总能使$L(x,\mu_j,\lambda_k) \to \infty$，所以公式2.1.1的第三项小于等于 0，考虑到求$L(x,\mu,\lambda)$最大值,第三项只有取零。这样一来，可以看到如果公式(2.1.1)存在最大值，那么第二项和第三项都是0,即可得到

$$f(x) = \max_{\mu,\lambda;\lambda_k\geq 0} L(x,\mu,\lambda) \tag{2.2.2}$$

这样问题 1 就转化到问题 2，可以得到公式(1.2.3)

2.2 单独讨论问题 3：

令

$$D_1(\mu,\lambda) = \min_x J(x,\mu,\lambda) \tag{2.2.1}$$

$$D_2(x) = \max_{\mu,\lambda} J(x,\mu,\lambda) \tag{2.2.2}$$

如果上面两式定义的最值都存在，根据上面两式定义:

$$D_1(\mu,\lambda) \leq J(x,\mu,\lambda) \leq D_2(x) \tag{2.2.3}$$

公式(2.2.3)表面对于任意的μ,λ，任意的x都有$D_1(\mu,\lambda) \leq L(x,\mu,\lambda) \leq D_2(x)$,所以有下面:

$$\max_{\mu,\lambda} D_1(\mu,\lambda) \leq \min_x D_2(x) \tag{2.2.4}$$

即:

$$\max_{\mu,\lambda}\min_x J \leq \min_x \max_{\mu,\lambda} J \tag{2.2.5}$$

这样弱对偶性质就解释了。为了说明问题 3 和问题 1 ,2无关，我们用J代替L，下面我们替代回去，即:

$$\max_{\mu,\lambda}\min_x L \leq \min_x \max_{\mu,\lambda} L \tag{2.2.5}$$

2.3 问题 3 的结论应用于解决问题 2，间接也就解决问题 1：

求解原问题公式(1.2.3)的对偶形式为:

$$\max_{\mu,\lambda;\lambda_k\geq 0} D_1(\mu,\lambda)=\max_{\mu,\lambda;\lambda_k\geq 0}\min_x L(x,\mu,\lambda) \tag{2.3.1}$$

我们讨论问题 2 时只是给出了原问题(1.2.3)的下界是公式(2.3.1)。引入强对偶性，强对偶性是指公式(2.2.5)取等号，这样求解原问题就完全转化为求解对偶问题了。

slater条件:(1)原始问题为凸优化问题(2)$\exists x \quad s.t. \quad g_k(x)<0$

对于问题 1，如果slatter条件成立，那么满足强对偶性，这是强对偶的充分条件。

对于问题 1，假设x^*是原问题(不一定是凸)最优解,μ^*,λ^*是对偶问题的最优解那么，如果强对偶性成立，有下式

$$\begin{aligned} f(x^*)=D_2(x^*)&=D_1(\mu^*,\lambda^*)\\ &=\min_x f(x)+\sum_j\mu_j^*h_j(x)+\sum_k\lambda_k^*g_k(x)\\ &\leq f(x^*)+\sum_j\mu_j^*h_j(x^*)+\sum_k\lambda_k^*g_k(x^*)\leq f(x^*)\end{aligned} \tag{1}$$

分析上式，第一个等号成立来源于问题(2)的结论，第二个等号成立源于满足强对偶性，第三个等号源于μ^*,λ^*是$D_1(\mu,\lambda)$的解，第一个$\leq$是源于$D_1(\mu,\lambda)$的定义，第二个$\leq$是因为问题 1 的约束条件$h_j(x)=0, g_j(x)\leq 0$以及讨论问题 2 给出的$\lambda_k\geq 0$。所以公式(1)不等号可以全换成等号，因为x^*是$D_2(x)$的极值点，所以有

$$\nabla_{x^*}L(x,\mu^*,\lambda^*)=0 \tag{2.3.2}$$

根据公式(1)第二个不等号以及$h_j(x^*)=0$,有:

$$\lambda_k^*g_k(x^*)=0 \tag{2.3.3}$$

和问题 1 给出的约束条件联合起来，得到kkt条件:

$$\nabla_x L(x,\mu,\lambda)=0 \tag{2.3.4}$$

$$\lambda_k g_k(x)=0 \tag{2.3.5}$$

$$h_j(x)=0 \tag{2.3.6}$$

$$g_k(x)\leq 0 \tag{2.3.7}$$

$$\lambda_k\geq 0 \tag{2.3.8}$$

这样我们在满足强对偶性时推出kkt条件，也就是说kkt条件是强对偶性的必要条件。

当原始问题是凸优化问题时，设μ^*,λ^*是D_1的解，x^*是D_2的解，结合kkt条件可以推出强对偶性。如下:

$$D_1(\mu^*,\lambda^*)=\min_x(x,\mu,\lambda)=L(x^*,\mu^*,\lambda^*)=f(x^*)+\sum_j\mu_j^*h_j(x^*)+\sum_k\lambda_k g_k^*(x^*)=f(x^*) \tag{2.3.9}$$

分析上式，第一个等号是因为μ^*, λ^*是D_1的解，第二个等式是因为x^*是D_2的解，第三个等式是定义，第四个是带入kkt条件。
对于问题 1 ，以上说明了kkt条件是强对偶性的充要条件。

3 参考资料:

https://www.cnblogs.com/ooon/p/5723725.html