拉格朗日乘子理解、对偶方法介绍

目 录

Contents

1 方法理解

2 最优控制

3 对偶观点

方法理解

方法理解
最优控制
对偶观点

拉格朗日乘数法是最优化问题求解中, **处理约束**(等式, 不等式)的常见思路

Primal Problem:

$$\min f_0(x)$$
$$s.t. \begin{cases} f_i(x) \le 0, i = 1, \dots, m \\ h_i(x) = 0, i = 1, \dots, p \end{cases}$$

Lagrangian:

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

$\lambda_i \ge 0$称作第i个不等式的拉格朗日乘子, ν_i ⋯第i个等式

Optimal Condition (equations):

$$\frac{\partial L}{\partial x_i} = 0$$

And other KKT conditions…

方法理解

方法理解
最优控制
对偶观点

第一种理解方式: 可行方向与下降方向线性相关.

考虑如下n元m约束最优化问题.

Primal Problem:

$$\min f(x)$$
$$s.t.\ g(x)=0$$

Lagrangian Function:

$$L(x,\lambda,\nu)=f(x)+\lambda^T g(x)$$

这里λ称作等式的拉格朗日乘子

设最优解为x^*

- 在x^*附近的**可行方向**集合为: $\{\Delta x: \nabla g(x^*)^T\Delta x=0\}$
- 沿可行方向, **目标函数不再下降**: $f(x^*+\epsilon\Delta x)\approx f(x^*)+\nabla f(x^*)^T\Delta x+\epsilon\geq f(x^*)$

以上表明: 对任意可行方向, $\nabla f(x^*)^T\Delta x=0$. 从而$\nabla f=\lambda^T\nabla g$, 即$\nabla_x L=0$($n$个方程).
结合已有的m个等式约束, 构成了$m+n$等式/约束方程, 理论上可解.

Remark: λ只是梯度线性相关的一种表示, L没有具体物理含义.

方法理解

方法理解
最优控制
对偶观点

第二种理解方式: 惩罚函数

Primal Problem:

$$\min f_0(x)$$
$$s.t. \begin{cases} f_i(x) \le 0, i = 1, \dots, m \\ h_i(x) = 0, i = 1, \dots, p \end{cases}$$

Lagrangian:

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

$\lambda_i \ge 0$称作第i个不等式的拉格朗日乘子, $\nu_i \cdots$第i个等式

原问题等价于:

$$\min_x \max_{\lambda,\nu} L(x, \lambda, \nu)$$

Proof: 在给定x的前提下, 考虑λ, ν的最佳取值

- 当x在可行域中时,$\nu_i h_i(x) = 0$, $\lambda_i f_i(x) \le 0, \forall \lambda_i \ge 0, \nu_i \in R$. 这表明最佳取值下 $\lambda_i f_i(x) = \nu_i h_i(x) = 0$. 从而此时$\max_{\lambda,\nu} L(x, \lambda, \nu) = f_0(x)$
- 当$h_i(x) \ne 0$时, $\max_{\lambda,\nu} L(x, \lambda, \nu) = L\big(x, \lambda, \infty \times \operatorname{sgn}(h_i(x))\big) = +\infty$
- 当$f_i(x) > 0$时, $\max_{\lambda,\nu} L(x, \lambda, \nu) = L(x, +\infty, \nu) = +\infty$

方法理解

方法理解
最优控制
对偶观点

第二种理解方式: 惩罚函数

Primal Problem:

$$\min f_0(x)$$
$$s.t. \begin{cases} f_i(x) \le 0, i = 1, \ldots, m \\ h_i(x) = 0, i = 1, \ldots, p \end{cases}$$

Lagrangian:

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

$\lambda_i \ge 0$称作第i个不等式的拉格朗日乘子, ν_i ⋯第i个等式

若L关于x, λ, ν连续, 且各决策变量均连续, 则

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

最优解一定满足$\nabla_x L = 0$.

Remark: λ, ν定义为惩罚因子, 但当$f_i(x) = 0$或$h_j(x) = 0$时,
惩罚项均为0, λ, ν可以任意取.

方法理解

基于惩罚函数的举例

举例:

$$\min_x x^2$$
$$s.t.\, x^2 \le 1$$

$$L = x^2 + \lambda(x^2 - 1), \lambda \ge 0$$

对于$\min_x \max_\lambda L$:

$$\lambda = \begin{cases} +\infty, x^2 > 1 \\ 0, x^2 < 1 \\ 任意, x^2 = 1 \end{cases}$$

原问题划为

$$\min \bar{f}(x)$$

其中

$$\bar{f}(x) = \begin{cases} x^2, x^2 \le 1 \\ +\infty, else \end{cases}$$

目 录

Contents

1 方法理解

2 最优控制

3 对偶观点

最优控制

方法理解
最优控制
对偶观点

变分法中的等式约束

优化目标	$\min f(x)$ $s.t.\, h(x)=0$	$\min_x \int_{t_0}^{t_f} g(x,\dot{x},t)\,dt$ $s.t.\, f(x,\dot{x},t)=0$
拉格朗日函数	$L:=f+\lambda^T h$	$L:=\int_{t_0}^{t_f} g+\lambda(t)f\,dt$
等价命题	$\min_x \max_\lambda L$	$\min_x \max_\lambda L$
必要条件	$\nabla_x L=\nabla_\lambda L=0$	$\delta_x L=\delta_\lambda L=0$
数量与索引	有限个λ, 以i为下标	无限个λ, 以t为索引

Remark: 拉格朗日乘子可以处理不等式约束, 可否类比到变分法中处理不等式约束? 有兴趣者可以自行探索.

目 录

Contents

1 方法理解

2 最优控制

3 对偶观点

对偶观点

Review: 拉格朗日乘子/函数的两个理解角度:

$$L = f(x) + \lambda^T h(x)$$

- 梯度角度: $\nabla_x L = 0$. 缺陷: L本身无含义, 其他λ无含义.
- 惩罚角度:$\min_x \max_\lambda L$. 缺陷: $h(x) = 0$时, λ可任意取(实际能求出唯一值).
- 对偶角度:

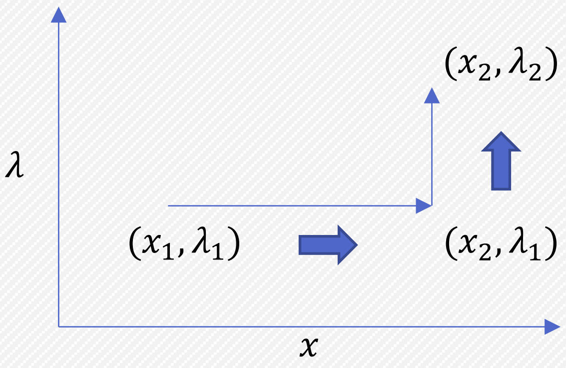
$$\text{Dual Prob} \Rightarrow \max_\lambda \min_x L \le \min_x \max_\lambda L \Leftarrow \text{Primal Prob}$$

Proof:

对于任意一点(x_0, λ_0), 有

$$L(x^*, \lambda_0) = \min_x L(x, \lambda_0) \le L(x_0, \lambda_0) \le \max_\lambda L(x_0, \lambda) = L(x_0, \lambda^*)$$

假设左边最优在(x_1, λ_1)取得, 右边最优在(x_2, λ_2)取到, 则由下图可知命题成立.



对偶观点

方法理解
最优控制
对偶观点

构建对偶问题

$$\min f_0(x)$$
$$s.t.\begin{cases} f_i(x) \le 0, i = 1, \dots, m \\ h_i(x) = 0, i = 1, \dots, p \end{cases}$$

$$\bar{f}(x) = \max_{\lambda \ge 0, \nu} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right)$$

$$g(\lambda, \nu) = \min_x \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right)$$

Primal problem:

$$\min_x \bar{f}(x)$$

Dual problem:

$$\max_{\lambda \ge 0, \nu} g(\lambda, \nu)$$

$$g(\lambda, \nu) \le \max_{\lambda, \nu} g(\lambda, \nu) \le \min_x \bar{f}(x) \le \bar{f}(x)$$

- 对任意λ, ν, $g(\lambda, \nu)$提供了原问题最优解的一个下界, 而$g(\lambda^*, \nu^*)$提供了最优下界.
- 若$\max_{\lambda, \nu} g(\lambda, \nu) = \min_x f(x)$, 则$f_0(x^*) = g(\lambda^*, \nu^*) \le f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* f_i(x^*) + \sum_{i=1}^{p} \nu_i^* h_i(x^*) \le f_0(x^*)$说明上面两个不等号取等.

注: *代表最优解.

对偶观点

方法理解
最优控制
对偶观点

KKT条件、正则项

续上页:
根据

$$f_0(x^*) = g(\lambda^*, v^*)$$
$$\le f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* f_i(x^*) + \sum_{i=1}^{p} v_i^* h_i(x^*)$$
$$\le f_0(x^*)$$

这说明:

$$\lambda_i^* f_i(x^*) = 0, \forall i = 1, \dots, m$$

以及

$$L(\lambda^*, v^*, x^*) = \inf_x L(\lambda^*, v^*, x)$$

表明$L(\lambda^*, v^*, x)$在$x = x^*$导数为0:

$$\nabla f_0\,(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^{p} v_i^* \nabla h_i(x^*) = 0$$

Remark1:
由此得到完整的KKT条件(不严谨表述)

$$f_i(x^*) \le 0,$$
$$h_i(x^*) = 0,$$
$$\lambda_i^* \ge 0,$$
$$\lambda_i^* f_i(x^*) = 0,$$
$$\nabla f_0\,(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^{p} v_i^* \nabla h_{i(x^*)} = 0$$

Remark2:
后者表明, 原问题的最优解也是无约束优化问题

$$\min_x L(\lambda^*, v^*, x)$$

的最优解, 这表明, 先选择合适的λ, v, 原最优化问题可以近似转化为无约束优化问题.

谢谢观看