又红又正
Ph.D.
223 人赞同了该回答

抛砖引玉, 说一下(Lagrangian) duality是怎么来的。先考虑下面的nonlinear programming:

$$\min\{f(\mathbf{x}) : g_i(\mathbf{x}) \leq 0,\ i = 1, 2, \ldots, m\} \quad (1)$$

现在的问题是如何找到问题(1) 的最优值的一个最好的下界? 首先我们知道若方程组

$$\begin{aligned} & f(\mathbf{x}) < v \\ & g_i(\mathbf{x}) \leq 0, i = 1, 2, \ldots, m \end{aligned} \quad (2)$$

无解，则v是问题(1)的一个下界。注意到方程组(2)有解可以推出对于任意的$\boldsymbol{\lambda} \geq \mathbf{0}$, 以下方程

$$f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) < v \quad (3)$$

有解。因此根据逆否命题，方程组(2)无解的充分条件是存在$\boldsymbol{\lambda} \geq \mathbf{0}$，让方程(3)无解。方程(3)无解的充要条件是

$$\min_{\mathbf{x}} f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) \geq v \quad (4)$$

因为我们要找最好的下界，所以这个时候的v和$\boldsymbol{\lambda}$ 应该取

$$v = \max_{\boldsymbol{\lambda} \geq \mathbf{0}} \min_{\mathbf{x}} f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) \quad (5)$$

由此引入了dual problem. 证明逻辑是根据式(5)取v和$\boldsymbol{\lambda}$, 则(4)成立，从而导出(3)无解，然后可以知道(2)无解，因此v是问题(1)的下界

编辑于 2017-04-19
▲赞同 223 ▼ 25 条评论
分享
收藏 感谢

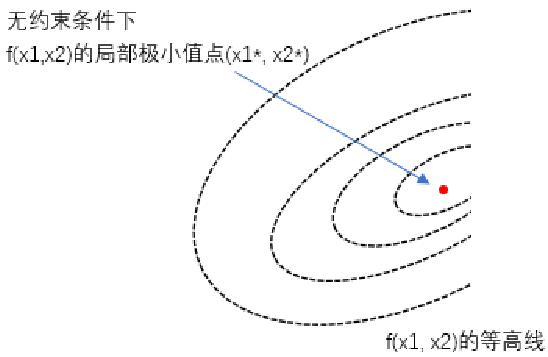
彭一洋
十里挑九的灵魂
290 人赞同了该回答

最近也在看关于优化的东西，题主在问题补充里问了好多，我暂且以二维空间 R^2 举例，从简单的无约束的优化（0梯度条件），到等式约束优化(拉格朗日条件)，再到不等式约束优化（KKT条件），写点对于优化问题自己能写的理解，权当做抛砖引玉。

1. 无约束的优化问题

$$minf(x)$$

其中，$x = (x_1, x_2)$



注意我在图里画了等高线。此时 $f(x)$ 在局部极小值点 $x^* = (x_1^*, x_2^*)$ 处的梯度必然为0，比较容易理解。这个梯度为零的条件是局部极小值点的必要条件。这样，优化问题的求解变成了对该必要条件解方程组。

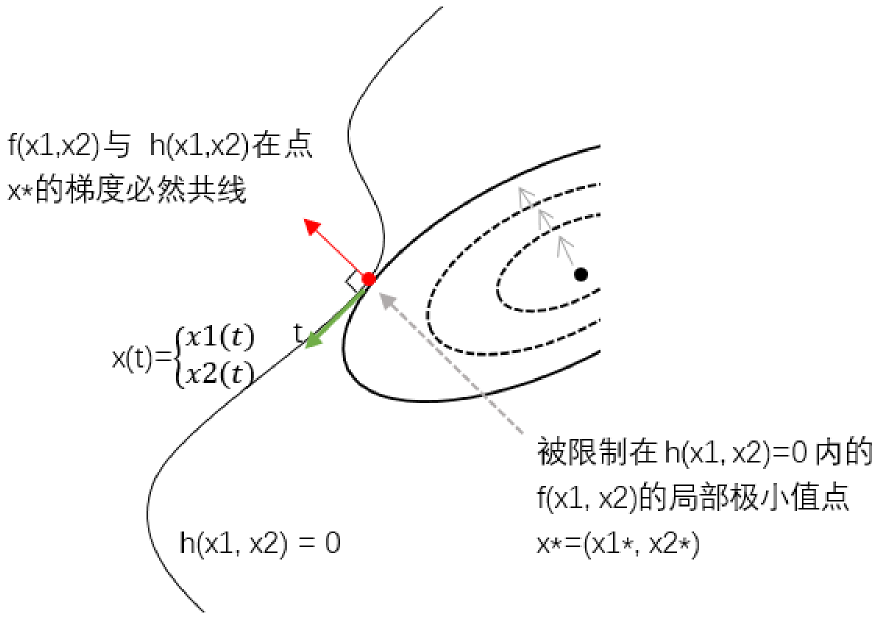
2.带等式约束的优化问题

$minf(x)$,

s.t.

$h(x) = 0$.

与无约束的问题不同。我们所要求的极小值点被限制在曲线 $h(x) = 0$ 上，我们将 $\{x|h(x) = 0\}$ 称为可行域, 解只能在这个可行域里取。如下图所示，曲线 $h(x) = 0$ （黑色实曲线）经过无约束极小值点（黑点）附近。那么满足约束的极小值点应该与黑点尽可能近。我们将 $f(x)$ 的等高线不断放大，直到与曲线 $h(x) = 0$ 相切，切点即为所求。相切是关键，是极小值点的必要条件。



把 $h(x) = 0$ 沿着曲线方向参数化为 $x(t)$, $x^* = x(t^*)$ 。必有 $f(x)$ 在红点 x^* 的梯度方向与 $x(t)$ 的切线方向垂直，即

$\nabla f(x^*) \cdot \dot{x}(t^*) = 0$

另外，由于 $h(x) = 0$ 为常数，那么也有复合函数 $h(x(t)) = 0$, 因此 $h(x(t))$ 在 t 的导数必为0，根据链式法则有

$\nabla h(x) \cdot \dot{x}(t) = 0$ （内积为0，说明 $\nabla h(x^*)$ 与 $\dot{x}(t^*)$ 垂直）

因为 $\nabla f(x^*)$ 垂直于 $\dot{x}(t^*)$ ， $\nabla h(x^*)$ 垂直于 $\dot{x}(t^*)$ ，所以 $\nabla f(x^*)$ 与 $\nabla h(x^*)$ 共线，有 $\nabla f(x^*) + \lambda \nabla h(x^*) = 0$

x^* 若为最小值点就必须满足上式和问题中的约束 $h(x^*) = 0$,这个必要条件就叫作拉格朗日条件，为了好记，定义一个拉格朗日函数

$$L(x, \lambda) = f(x) + \lambda h(x)$$

令其偏导为0，正好就得到拉格朗日条件。

如此，带等式约束的优化问题转化为了无约束的优化问题，只需要对拉格朗日条件解方程组即可。这里λ就是拉格朗日乘子，有多少个等式约束就有多少个拉格朗日乘子。

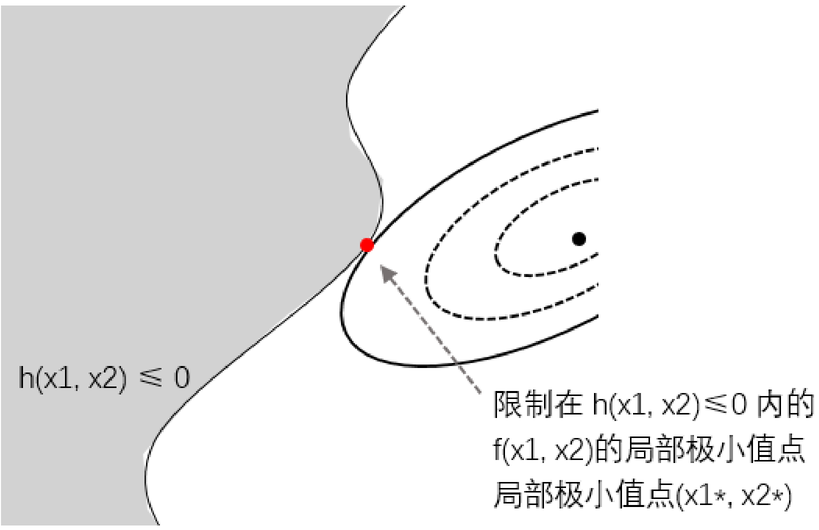
3. 带不等式约束的优化问题

$minf(x)$,

s.t.

$h(x) \leq 0$.

当只有一个不等式起作用时, 如我们把问题2里的等式约束 $h(x) = 0$ 改为 $h(x) \leq 0$ ，如下图所示，可行域变成了阴影部分，最小值点还是切点，情况和问题2完全一样，只需要把不等号当做等号去求解即可。



当两个不等式起作用时，那么问题就来了

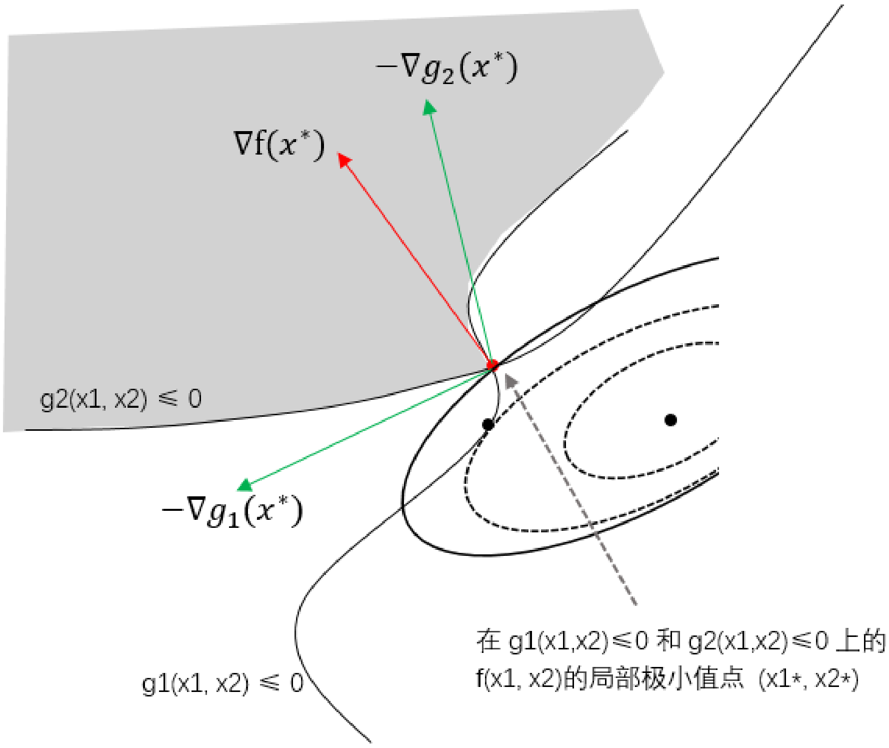
$minf(x)$,

s.t.

$g_1(x) \leq 0$,

$g_2(x) \leq 0$.

如下图，当 $f(x)$ 的等高线慢慢扩大时，等高线与可行域(阴影部分)第一次相遇的点是个顶点，2个不等式同时起作用了。满足约束的最小值点从原来的黑点位置(切点)移动到了红点位置，现在跟哪条约束函数曲线都不相切。这时候就需要用到kkt条件了。这里的“条件”是指：某一个点它如果是最小值点的话，就必须满足这个条件（在含不等式约束的优化问题里）。这是个必要条件，前面说的也全部是必要条件。



这个问题的解 x^* 应满足的KKT（卡罗需-库恩-塔克）条件为：

1. $\mu_1 \geq 0$，$\mu_2 \geq 0$；

2. $\nabla f(x^*) + \mu_1 \nabla g_1(x^*) + \mu_2 \nabla g_2(x^*) = 0$；

3. $\mu_1 g_1(x^*) + \mu_2 g_2(x^*) = 0$。

其中，μ叫KKT乘子，有多少个不等式约束就有多少个KKT乘子。加上问题3中的约束部分，就是完整版的KKT条件。对于有等式的情况，你把其中一个不等式约束换成等式，可行域变成了半条曲线，最小值点还是那个红点，和下面这种情况是一样的。

下面看看KKT条件是怎么来的。在问题2中我们知道了约束曲线的梯度方向与曲线垂直，我在上图画出了两条约束曲线的负梯度方向（绿色箭头）和等高线的梯度方向（红色箭头）。如果这个顶点是满足约束的最小值点，那么该点处（红点），红色箭头一定在两个绿色箭头之间($-\nabla g(x)$ 方向一定指向 $g(x)$ 减小的方向，即 $g(x) < 0$ 的那一边)。即 $\nabla f(x^*)$ 能被 $-\nabla g_1(x^*)$ 和 $-\nabla g_2(x^*)$ 线性表出（ $\nabla f(x^*) = -\mu_1 \nabla g_1(x^*) - \mu_2 \nabla g_2(x^*)$ ），且系数必非负（ $\mu_1 \geq 0$，$\mu_2 \geq 0$ ）。也就是kkt条件中的1和2

1. $\mu_1 \geq 0$，$\mu_2 \geq 0$；

2. $\nabla f(x^*) + \mu_1 \nabla g_1(x^*) + \mu_2 \nabla g_2(x^*) = 0$。

有时候，有的不等式约束实际上不起作用，如下面这个优化问题

$min f(x)$，

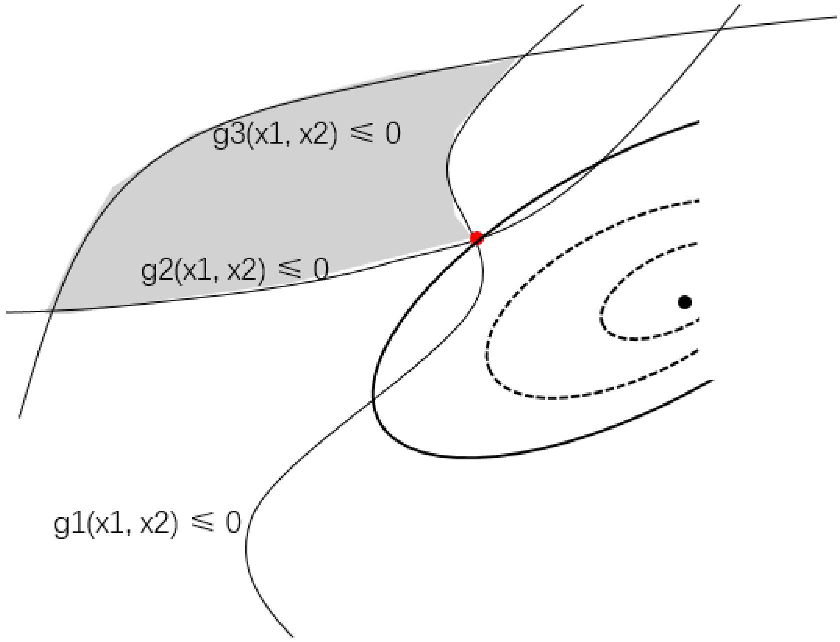
s.t.

$g_1(x) \leq 0$；

$g_2(x) \leq 0$；

$g_3(x) \leq 0$。

如下图的 $g_3(x_1, x_2) \leq 0$ 是不起作用的



对于最小值点 x^* ,三个不等式约束的不同在于

$g_1(x^*) = 0$ （起作用）

$g_2(x^*) = 0$ （起作用）

$g_3(x^*) < 0$ （不起作用, 最小值点不在 $g_3(x) = 0$ 上）

这时，这个问题的KKT条件1，2成了：

1. $\mu_1 \geq 0$， $\mu_2 \geq 0$， $\mu_3 \geq 0$；

2. $\nabla f(x^*) + \mu_1 \nabla g_1(x^*) + \mu_2 \nabla g_2(x^*) + \mu_3 \nabla g_3(x^*) = 0$.

条件2中的 $\mu_3 \nabla g_3(x^*)$ 这一项让我们很苦恼啊， $g_3(x^*)$ 的绿色箭头跟我们的红色箭头没关系。要是能令 $\mu_3 = 0$ 就好了。加上条件3：

3. $\mu_1 g_1(x^*) + \mu_2 g_2(x^*) + \mu_3 g_3(x^*) = 0$

恰好能使 $\mu_3 = 0$ 。由于 $g_1(x^*) = 0$ ， $g_2(x^*) = 0$ ，所以前两项等于0，第三项 $g_3(x^*) < 0$, 在条件3的作用下使得 $\mu_3 = 0$. 正好满足需求。如果再多几项不起作用的不等式约束，比如 $g_4(x) \leq 0$ 。要使

$$\mu_1 g_1(x^*) + \mu_2 g_2(x^*) + \mu_3 g_3(x^*) + \mu_4 g_4(x^*) = 0$$

就只能有 $\mu_3 g_3(x^*) + \mu_4 g_4(x^*) = 0$

同样地， $g_3(x^*) < 0$, $g_4(x^*) < 0$, 只能出现 $\mu_3 = \mu_4 = 0$ 或者 μ_3 和 μ_4 异号的情况。但注意条件1限制了 $\mu_3 \geq 0$, $\mu_4 \geq 0$ ，所以只能有 $\mu_3 = \mu_4 = 0$ 。因此不管加了几个不起作用的不等式约束，条件2都能完美实现：目标函数 $f(x)$ 的梯度 $\nabla f(x)$ 被起作用的不等式约束函数 $g(x)$ 的负梯度（ $-\nabla g(x)$ ）线性表出且系数 μ 全部非负（红色箭头被绿色箭头夹在中间）。这样，优化问题的求解就变成对所有KKT条件解方程组。

如果再定义一个拉格朗日函数

$$L(x, \mu) = f(x) + \mu_1 g_1(x) + \mu_2 g_2(x) + \ldots$$

令它对x的偏导为0，就是KKT条件中的条件2了。

最后说明一下，以上所有都是局部极小值点的必要条件。据此求得的解不一定是局部极小值点（更别提全局了），原因是上图中我所画的等高线也许根本就不闭合，也就是说我们一直想要靠近的等高线中间的黑点压根就是个鞍点或者近似鞍点！

------------------------2017.6.6--------------------------

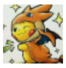
顺带一提，李航老师《统计学习方法》第一版105页式(7.27)中的第1，2行就是这里的KKT条件2（我这里把偏置b算在x里了），第3行是KKT条件3，第4行是问题中的不等式约束，第5行是KKT条件1。

编辑于 2017-06-06

▲赞同 290 ▼ 30 条评论

分享

收藏 感谢 收起



2prime

applied math, data science, numerical analysis

9 人赞同了该回答

换一个问题来看对偶问题吧

大家知道变分问题吧

把极小化函数变成极小化一个泛函 $\int L(\dot{x},x)dx$

然后可以写出来他的最优性条件就是Euler - Lagrange方程

$$-\frac{d}{ds}(D_qL(\dot{x}(s),x(s))+D_xL(\dot{x}(s),x(s)))=0$$

我们尽可能把他化成一个Hamilton ODE（Hamilton - Jacobi方程的特征线）

这里我们要构造一个函数

$$H(p,x):=p\cdot q(p,x)-L(q(p,x),x)$$

这里suppose $p(s):=D_qL(\dot{x}(s),x(s))$

于是对于H我们有hamliton ODE写出来就是E - L方程了

奇妙的是这里的H 和L 就是凸优化里的对偶的关系

L=H*,H=L*

具体可以看evans

还有一个很好的几何解释

过两天有空我来写写在

《凸优化理论》［MIT那本啊］这本书里详细介绍了

最近还是有很多pde和优化关系的文章的都很不错

Yu Mao, Bin Dong and Stanley Osher, A nonlinear PDE-based method for sparse deconvolution, Multiscale Modeling and Simulation: A SIAM Interdisciplinary Journal, **8(3)**, 965-976, 2010.

A variational perspective on accelerated methods in optimization. A. Wibisono, A. Wilson, and M. I. Jordan. *Proceedings of the National Academy of Sciences*, 133, E7351-E7358, 2016. [ArXiv version]

W. Su, S. Boyd and E. J. Candès. A differential equation for modeling Nesterov's accelerated gradient method: theory and insights. *Journal of Machine Learning Research* **17**(153), 1--43. (This is the long form or journal version of the NIPS paper.) (pdf)

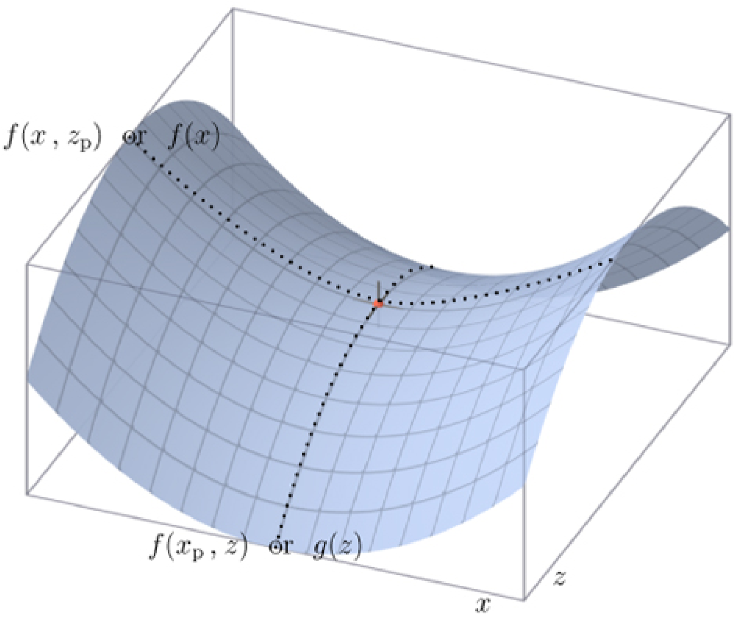
Pratik Chaudhari, Adam Oberman, Stanley Osher, Stefano Soatto, and Guillame Carlier, *Deep Relaxation: Partial Differential Equations for Optimizing Deep Neural Networks*, April 2017

编辑于 2017-05-19
▲赞同 9 ▼ 添加评论
分享
收藏 感谢 收起

atom Native
历史探寻者掉进理工大世界
11 人赞同了该回答

Dual problem 跟primal problem 可以看成本来是两个问题，因为优化的顺序不同而会得出两个不一定相关的值（但是 $\min_x \max_y f(x,y) \geq \max_y \max_x f(x,y)$ 还是成立的，直观理解的话高中经常用的二次函数就可以了）。

两者的差值就是duality gap，描述了我用另一种方式刻画问题的时候所造成的误差，强对偶的情况下最优值没有差别。



在最优点处将会满足KKT 条件，但是KKT条件本身并不需要问题满足强对偶。

关于KKT条件什么时候不满足，有一种另外的理解是他要求各个函数的梯度张成足够大的空间（因为KKT的最后一条本质上是一个Ax=0的问题），希望有助于理解。

建议题主阅读Boyd Convex Optimization Chap. 5.5

编辑于 2017-04-25
▲赞同 11 ▼ 3 条评论
分享
收藏 感谢

覃含章
计算科学与工程博士在读
105 人赞同了该回答

谢谢题主

@孤云独去闲
邀请。前段时间就看到这个问题了，现在看到很多答主都答得不错，那我给一个基于column geometry（翻译成列几何？）的解释吧。拉格朗日对偶有很多种理解和推导的方式，这一种是我比较喜欢的，几何和代数的方法结合，也比较有intuition。关于

KKT条件的几何解释很多答主都提了，那个也是比较经典的，我这个回答就先不涉及了。

我们考虑优化问题如下，记作问题（P）。（知乎编辑器似乎出了点问题，函数括号圆括号显示不了，我就用方括号了）

$$z^* = \min_x f[x] \text{s.t. } g_i[x] \leq 0, \ \forall\, i = 1, \ldots, m, x \in X$$

大家都知道（P）的拉格朗日对偶问题（D）写作

$$v^* = \max_{u \geq 0} \min_{x \in X} \underbrace{f[x] + u^T g[x]}_{L[x,u]}$$

其中的函数L[x,u]就是我们熟知的拉格朗日函数。

这边先给一个小note，实际上原问题和拉格朗日对偶的代数形式就是一组**max-min关系式**（只有max和min的顺序换一下）。具体说明如下。

引理：(P)也可以写成 $z^* = \min_{x \in X} \max_{u \geq 0} L[x, u]$.

证明是很容易的，留作练习。那么这里其实我们看到所谓的拉格朗日对偶从代数上看很简单，就是研究这一对max-min优化问题之间的关系。

好了，之前都是预热。接下来我们来看如何通过column geometry来理解这对关系式。这一段首先介绍一些符号和定义。首先注意到给定一个 $x \in X$,实际上(P)可以用一个在 $\mathbb{R}^{m+1}$ 里的向量来描述：

$$[s, z] = [s_1, s_2, \ldots, s_m, z] = [g_1[x], g_2[x], \ldots, g_m[x], f[x]] = [g[x], f[x]].$$

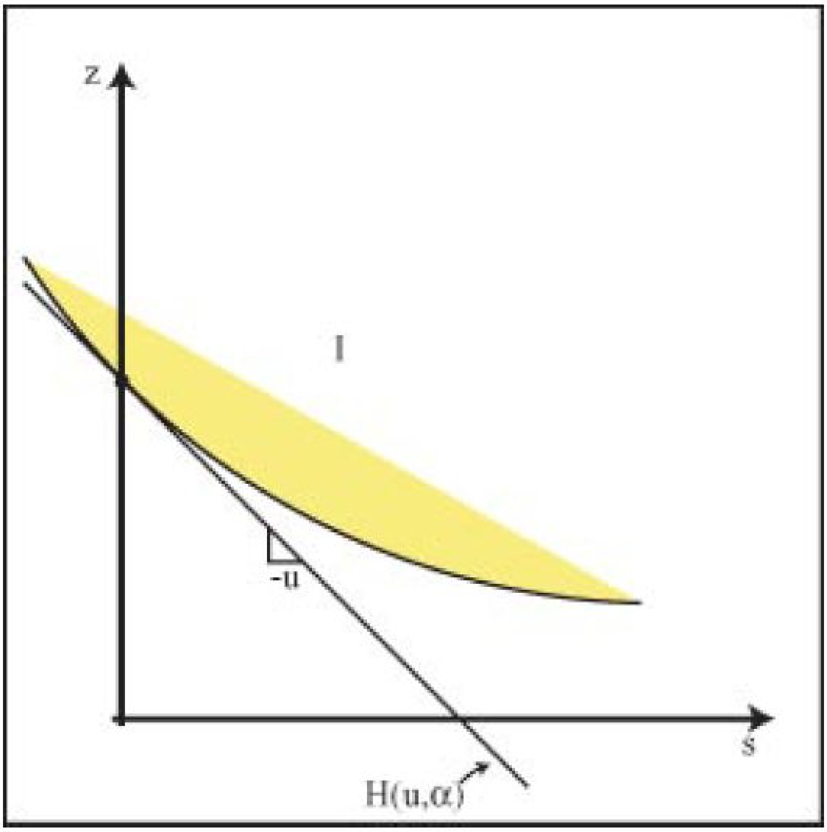
这样我们把问题转换到 $[s, z]$ 定义的空间上，定义集合

$$I := \{[s, z] \in \mathbb{R}^{m+1} : \exists x \in X \text{ s.t. } s \geq g[x], z \geq f[x]\}.$$

然后我们引入“支撑”（support）的概念，我们称一个超平面 $H_{u,\alpha} := \{[s, z] \in \mathbb{R}^{m+1} : u^T s + z = \alpha\}$ 是 I 的下支撑（lower support）仅当

$$u^T s + z \geq \alpha, \forall\, [s, z] \in I.$$

直观的意思就是指 $H_{u,\alpha}$ **在 I 的下方**。下面给一张图片加以解释（黄色部分就是 I ，那条线就代表超平面 $H_{u,\alpha}$ ）。



注意到这里我们图上看到 I 画的是一个凸集（convex set），我们指出在凸优化的一般情况下这是一个必然的事实。

引理：如果 X 是凸集，$f, g_1, \ldots, g_m$ 都是 X 上定义的凸函数，那么 I 是一个凸集。

证明也很容易，利用凸函数epigraph的性质就能立即得出，这里从略。

好了，准备工作到这里，我们给出对偶问题（D）的column geometry解释：拉格朗日对偶问题（D），在空间 $[s, z]$ 中的几何含义是：**找到 I 的下支撑超平面 $H_{u,\alpha}$ 中与z轴交点最“高”（即 α 最大）的那个超平面。**注意到在 $m = 1$ 的情况下（$[s, z]$ 是二维的）我们可以知道 $-u$ 是直线 $H_{u,\alpha}$ 的斜率，α 则是截距。这个intuition到高维情况也是成立的。

把上面提到的几何含义用代数表示出来则是：

$$\max_{u,\alpha}\ \alpha\ \text{s.t.}\ \ u^T s + z \geq \alpha,\ \forall\, [s, z] \in I$$

注意到我们其实只需要考虑 $u \geq 0$ 的情况，因为如果 u 存在一个coordinate是负的，那么我们总可以找到一个无限大的 s（注意 I 的定义，这样的 s 永远是存在的）使得 $u^T s + z < \alpha$，所以以上问题也等于

$$\max_{u \geq 0, \alpha}\ \alpha\ \text{s.t.} u^T s + z \geq \alpha,\ \forall\, [s, z] \in I$$

$$= \max_{u \geq 0, \alpha}\ \alpha$$

$$\text{s.t.}\ \underbrace{u^T g[x] + f[x]}_{L[x,u]} \geq \alpha,\ \forall\, x \in X$$

$$= \max_{u \geq 0, \alpha} \alpha$$

$$\text{s.t.}\ \min_{x \in X} L[x, u] \geq \alpha$$

$= \max_{u\geq 0}\min_{x\in X} L[x,u]$

我们于是发现这就是问题（D）。

这就是拉格朗日对偶基于column geometry的几何解释。 顺便说一下，如果我们考虑一个更特殊的情况， f,g 都是线性函数，即线性规划问题，column geometry会给出更多很有意思的几何解释，这边就留给感兴趣的同学自己去琢磨了。

版权声明：这里的内容（包括那张图）基本都来自于Robert M. Freund和Jorge R. Vera合写的Fundamentals of Nonlinear Optimization: a Constructive Approach的草稿（这书应该还没出版）。

我的另一个相关回答：（讨论更抽象的对偶优化问题，不过核心思路都是一致的，对偶的核心思想就是要找原问题的线性majorization）

线性空间的对偶空间和优化里的拉格朗日对偶有什么关系？ - 知乎
编辑于 2017-05-10
▲赞同 105 ▼ 26 条评论
分享
收藏 感谢 收起

知乎用户
对Convex problem略知一二。
27 人赞同了该回答

这里笔者用**博弈论**里面的**直观结论**解释问题，抛砖引玉。

1. 构造Lagrangian：

对于下面的优化问题，

$$\begin{aligned} &\min_x f(x) \\ s.t.\quad & g(x)\leqslant 0 \;\; and \;\; \exists g(x)<0 \\ & x\in\mathbb{R}^n,\;\; g(x)\in\mathbb{R}^m \end{aligned}$$

我们**构造**一个新问题，

$$\begin{aligned} \min_{x\in\mathbb{R}^n}\max_{\lambda\in\mathbb{R}^m} \mathcal{L}(x,\lambda) = f(x)+\lambda^T g(x) \\ s.t.\quad \lambda\in\mathbb{R}^m\geqslant 0 \\ x\in\mathbb{R}^n,\;\; g(x)\in\mathbb{R}^m \end{aligned}$$

非常易证，**这个问题等同于原问题**。那我们把它叫做**primal problem**。

2. 博弈论中的重要结论：

2.1 同时我们有**Min-Max 不等式**

$$\sup_{y\in Y}\inf_{x\in X} f(x,y)\leqslant \inf_{x\in X}\sup_{y\in Y} f(x,y)$$

2.2 和**Minimax Theroem**：当 $f(x,y)$ 对 x 为凸函数， $-f(x,y)$ 对 y 为凸函数（即对 y 凹 谢谢

@又红又正
指正）时候，上述不等式**取得等号**。

2.3 下面我来解释一下这两个结论：

不等式：如果 $f(x,y)$ 是一个零和游戏中的payoff函数，x 目的是减小此函数值（函数值对他来说是**成本**），y 想要增大此函数值（函数值对他来说是**利润**），则上述不等式中左边为 y 的 “**保底利润**” （最坏情况下，他也一定至少能获得这么多利润），右边为 x 的 “**保底损失**” （最坏情况下他损失也就这么多）。那么显而易见，上述不等式成立，**不然与“保底”的概念矛盾。**

Minimax定理：即说明了此函数存在saddle point，博弈双方有了一个**纳什均衡（谁也无法单方面地获得好处）。**

3. 定义dual problem：

定义**dual problem**：$\max_{\lambda\in\mathbb{R}^m}\min_{x\in\mathbb{R}^n}\mathcal{L}(x,\lambda)=\max_{\lambda\in\mathbb{R}^m}\min_{x\in\mathbb{R}^n}f(x)+\lambda^T g(x)$

由2.中的结论，$\mathcal{L}(x,\lambda)$ 天然对 λ 凹，那么只要 $f(x)$ 与 $g(x)$ 对 x 凸，那么 $\mathcal{L}(x,\lambda)$ 就对 x 凸对 λ 凹，于是**等式成立。**这个结论即是convex problem**最显著最重要的特点。**

如果**等式不成立**，那么**差值**就叫**dual gap**。

希望能帮助到您。

编辑于 2017-05-06

▲赞同 27 ▼ 14 条评论

分享

收藏 感谢 收起

孤云独去闲

咨询业EM，Data Scientist，信号处理博士

11 人赞同了该回答

我提了workshop的想法，我自己不写点什么实在是有点偷懒。不过我很不擅长优化，我的背景更多是统计，概率，分布的那些东西。

在Andrew Ng 教授讲解支持向量机（svm）的那篇教案（note）里，他讲过svm的primal和dual的形式，我觉得是我见过的最好的svm教案。下面是链接：

教案
http://cs229.stanford.edu/notes/cs229-notes3.pdf

视频
http://m.open.163.com/movie?plid=M6SGF6VB4&rid=M6SGJVMC6

请一定要看英文的那个版本，有个中文的，比较浅，没谈到duality。
发布于 2017-04-16

▲赞同 11 ▼ 3 条评论

分享

收藏 感谢

Hello曜

喜欢机器学习的半吊子软件工程师

15 人赞同了该回答

一些粗浅的个人理解：有时候求原始问题是复杂的，尤其当原始问题是一个nonconvex problem时，这个时候我们便把原始问题转化为它的对偶问题。

相比直接计算原始问题，求解对偶问题会为我们的计算带来相当的便利，比如楼主提到的SVM引入对偶形式便是一个很好的例子。

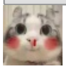
求出对偶问题的解至少能为原始问题找到一个下界（Weak duality always holds），如果运气好的话（满足KKT），此时求出的便是原始问题的最优解。

编辑于 2017-04-21

▲赞同 15 ▼ 1 条评论

分享

收藏 感谢



Cyber

没有什么是遗传算法不能解决的，如果有，那就加钱

12 人赞同了该回答

最近再看线性规划，涉及到对偶，恰好前段时间在学习SVM机的时候也有对偶理论，两者对照起来学习，感觉稍微有点融会贯通了。

自身学在习的时候一直力求直观易于理解，能够透过数学公式看透背后的原理，希望写出来也能通俗易懂。

线性规划中的对偶理论：

原问题：

$$\begin{gathered} max \quad c^T x \\ s.t. Ax \leq b, x \geq 0 \end{gathered}$$

对偶问题：

$$\begin{gathered} min \quad b^T y \\ A^T y \geq c, y \geq 0 \end{gathered}$$

其中，*A,b,c,x,y*均为向量。

用拉格朗日数乘法把原问题转换成无约束问题：

$$L(x,\lambda) = c^T x + \lambda^T (b - Ax)$$

其中$\lambda > 0$, 即λ的每个元素都大于0，维数和b的维数相等，即约束条件个数

$$\text{因为}(b - Ax) \geq 0\text{恒成立，且}\lambda \geq 0,$$

$$\text{令}z = \min_{\lambda} L(x,\lambda)$$

$$\text{则有：}\ z = \min_{\lambda} L(x,\lambda) = c^T x, \text{此时}\lambda = 0$$

$$\text{于是，原问题等价于：}$$

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

好了，重点来了，所谓对偶就是把*max*和*min*的顺序换一下，

对偶问题的表述如下：

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

注意下标！

接下来证明为什么可以由对偶问题解出原问题，这是最关键的一步，也是最精彩的地方

$$-------------------$$

$$令 g(\lambda) = \max_{x} L(x,\lambda),$$

由于$g(\lambda)$的表达形式中λ未作限制，因此与原问题相比，少了一个约束条件，

$$因此, g(\lambda) \geq max\ c^T x$$

从另一方便看，$\lambda(b - Ax)$始终大于等于0，也可以证明$g(\lambda) \geq max\ c^T x$成立，

上式说明$g(\lambda)$是原问题的上界

求得原问题的上界有什么用呢

我们一步步往下推，

假设$s = max\ c^T x,$我们的目的就是要求出s值。

在上一步中我们看到，$g(\lambda) \geq s,$去不同的λ值就能得到一系列的g值，

其中有的$g(\lambda)$比s大很多，

有的比s大的不多，

甚至可能和s一样大。

我们去和s值最接近的$g(\lambda)$值作为s值的近似，

于是通过求得$g(\lambda)$中最接近s的值作为原问题的解。

那么在一系列的$g(\lambda)$中，到底哪一个最接近s呢？

前面已经证明了$g(|lambda)$是s的上界，

那么$g(\lambda)$中的最小值就是最接近s的上界了，

写成数学公式如下即为

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

这就是原问题的对偶表示，

至此，已经说明了为什么对偶问题的解可以作为原问题的解（近似解）

下面，讨论等式$g(\lambda) \geq s$中何时等号成立

当等号成立时，

对偶问题的解就是原问题的精确解，

即可以找到一个λ，使得$g(\lambda) = s,$这种情况就叫强对偶

当等号不成立时，

对偶问题的解和原问题的解总存在一定偏差，

只是原问题解的一个近似，这种情况叫弱对偶

第一次用Tex，排版不是很熟练，蛋疼ing。。。

我们把最原始的s表示成了max,min的形式，好了，现在需要做的工作就是把min,max形式的对偶问题反退回s那种形式，即从

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

$$\text{这种形式反推回}$$

$$min \quad b^T y$$

$$s,t \quad A^T y \geq c, y \geq 0$$

$$\text{这种形式}$$

其实只是意思是一样的，只是表达方式不一样。

$$\begin{aligned} s^{'} &= (s^{'})^T = (\min_{\lambda}\max_{x} L(x,\lambda))^T \\ &= (\min_{\lambda^T}\max_{x^T}[c^T x + \lambda(b - Ax)])^T \\ &= \min_{\lambda^T}\max_{x^T}[x^T c + b^T\lambda^T - x^T A^T \lambda^T] \\ &= \min_{\lambda^T}\max_{x^T}[b^T\lambda^T + x^T(c - A^T\lambda^T)] \end{aligned}$$

令

$$\lambda^T = y, x^T = \beta$$

代入可得

$$\begin{aligned} s^{'} &= \min_{\lambda^T}\max_{x^T}[b^T\lambda^T + x^T(c - A^T\lambda^T)] \\ &= \min_{y}\max_{\beta}[b^T y + \beta(c - A^T y)] \end{aligned}$$

原问题中，

$$x \geq 0 \Rightarrow \beta \geq 0$$
$$\lambda \geq 0 \Rightarrow y \geq 0$$

若满足约束条件

$$c - A^T y \geq 0, \text{即} A^T y \geq c$$

则可以推出如下结论：

$$\max_{\beta}[b^T y + \beta(c - A^T y)] = b^T y \text{（因为后面一项总是大于等于0）}$$

$$\Rightarrow s^{'} = \min_{x} b^T y$$

在加上约束条件写成最开始的形式就是：

$$min \quad b^T y \qquad s.t. \quad A^T y \geq c, y \geq 0$$

$$-------------------------------$$

至此，线性规划中的对偶问题就证明完毕。

以上证明对于数学专业的同学来说可能不够严谨，对于原理讲解应该已经够了。

抛开线性规划的形式，通过寻找原问题的最小上界（最大下界）来近似原问题解的思想对其他形式的凸优化也同样适用。

主要参考来源：

支持向量机：Duality

感兴趣的可以参考那篇，作者其他文章也很不错

发布于 2017-09-15

▲赞同 12 ▼ 1 条评论

分享

收藏 感谢 收起

力学渣
NLP&ML新手/不懂CFD...
6 人赞同了该回答

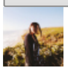
最近也在研究支持向量机，这学期也正在学最优化理论，上周刚刚总结了一下KKT条件并写了一篇分享，算是一个抛砖引玉吧~

浅谈最优化问题的KKT条件 https://zhuanlan.zhihu.com/p/26514613

发布于 2017-04-24
▲赞同 6 ▼ 1 条评论
分享
收藏 感谢

cin Gou
小说家
2 人赞同了该回答
陈宝林的《最优化理论》中有一章专门讲对偶问题
发布于 2017-04-23
▲赞同 2 ▼ 添加评论
分享
收藏 感谢

Stiefel
老师
1 人赞同了该回答
根据Rockafellar的一篇综述论文，Lagrange对偶的思想源于von Neumann的博弈论，min_x max_y L(x,y) 转换为 max_y min_x L(x,y)。要彻底搞明白对偶理论还得看一下Rockafellar的凸分析那本书。
发布于 2017-06-01
▲赞同 1 ▼ 1 条评论
分享
收藏 感谢



西瓜
OR/CS/Microeconomics
1 人赞同了该回答
cost/price of constraint.
发布于 2017-04-17
▲赞同 1 ▼ 8 条评论
分享
收藏 感谢

胡爱珍
学生

有没有 关于学凸优化 学习的视频的 学生党一枚 自己看凸优化那本书 看的不是很懂 希望有个合适的视频可以看看

编辑于 2018-01-12
▲赞同 ▼ 1 条评论
分享
收藏 感谢

知乎用户

谁能给解释下对偶可行与原始可行之间的关系?

发布于 2017-05-11
▲赞同 ▼ 3 条评论
分享
收藏 感谢

匿名用户

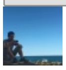
partial dirivative

发布于 2017-04-17

▲赞同 ▼ 添加评论

分享

收藏 感谢



樊二烁

哎，好好学习好好看~ 表示我可能对通俗的理解有问题

编辑于 2017-05-19

▲赞同 ▼ 添加评论

分享

收藏 感谢