十月算法班第三讲

管老师

七月在线

Oct, 2016

主要内容 (上)

- 优化与凸优化简介
  - 优化问题基本形式
  - 凸优化问题基本形式
  - 凸优化的应用
- 凸集合与凸函数基本概念
  - 凸集合与凸函数的关系
  - 凸集合与凸函数的性质对应
- 凸集分离定理
  - 凸集分离定理
- 共轭凸函数
  - 共轭凸函数
- 凸优化问题举例

主要内容 (下)

- 共轭函数
  - 共轭函数
- 对偶问题
  - 拉格朗日对偶函数
  - 拉格朗日对偶问题
  - 共轭函数与拉格朗日对偶函数
- 对偶性
  - 弱对偶性与强对偶性
  - 强对偶性成立的几种情况
  - 凸优化问题求解 (KKT)
- 应用举例
  - 支持向量机 (SVM) 的最简单形式
- 总结寄语 (数学部分)

记号

- 本节课常用数学记号

V, W	向量空间
v, w	向量
$\mathbb{R}^n$,$\mathbb{R}^m$	实坐标空间
α,β	V 和 W 的基
$T: V \to W$	向量空间 V 到 W 的线性映射
$A_{\alpha,\beta}(T)$	线性映射 T 在 α 和 β 这两组基下的矩阵
$G(v_1, v_2)$	内积空间 V 上的内积
H_α	G 在基 α 下的矩阵形式

优化问题

优化问题的一般形式

$$\begin{aligned} \text{最小化：} \quad & f_0(x) \\ \text{条件：} \quad & f_i(x) \le b_i, i = 1, \cdots, m. \end{aligned}$$

其中 $f_0(x)$ 为目标函数，条件里的不等式是限制条件.

优化与凸优化简介

举例:

极大似然估计

如果 $L(\mu,\sigma)$ 是一个极大似然估计问题中的似然函数，其中 μ,σ 分别是期望和方差，那么极大似然估计的问题转化为

$$\text{最小化:}\quad -L(\mu,\sigma)$$
$$\text{条件:}\quad \sigma \geq 0$$

最小二乘

如果 $A_{n\times k}$ 是一个矩阵,$b\in\mathbb{R}^n$ 是一个向量, 对于 $x\in\mathbb{R}^k$

$$\text{最小化:}\quad f_0(x) = |Ax-b|^2$$

凸优化问题

凸优化问题的一般形式

$$
\begin{aligned}
\text{最小化:}\quad & f_0(x) \\
\text{条件:}\quad & f_i(x) \le b_i, i = 1, \cdots, m.
\end{aligned}
$$

其中 $f_0(x)$ 为目标函数，条件里的不等式是限制条件.

- 凸优化问题的条件,$f_0, f_1, \cdots, f_m$ 都是凸函数.
- 凸优化问题的特点, 局部最优等价于全局最优.
- 凸优化问题的求解, 几乎总有现成的工具来求解.

优化与凸优化简介

凸优化的应用

- 凸优化问题逼近非凸优化问题，寻找非凸问题的初始点
- 利用对偶问题的凸性给原问题提供下界估计
- 凸优化问题可以给非凸问题带来一些启发

凸集合与凸函数

凸集合定义

如果一个集合 Ω 中任何两个点之间的线段上任何一个点还属于 Ω, 那么 Ω 就是一个凸集合.i.e.

$$\lambda x_1 + (1-\lambda)x_2 \in \Omega, \forall x_1, x_2 \in \Omega, \lambda \in (0,1)$$

凸函数定义

如果一个函数 f 定义域 Ω 是凸集，而且对于任何两点. 以及两点之间线段上任意一个点都有

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2)$$

$\forall x_1, x_2 \in \Omega, \lambda \in (0,1)$

凸集合与凸函数

凸集合与凸函数

函数的上境图

假设 f 是一个定义在 Ω 上的函数，区域 $\{(x,y):y\geq f(x),\forall x\in\Omega\}$ 就是 f 的上境图.

上境图就是函数图像上方的部分区域.

凸集合与凸函数

一个函数是凸函数当且仅当 f 的上境图是凸集合.

凸集合与凸函数有很多相对应的性质可以由这个结论来进行链接。

凸集合与凸函数

凸集合与凸函数

凸组合

对于任何 n 个点 $\{x_i\}_{i=1}^n$，以及权重系数 $\{w_i\}_{i=1}^n$. 若权重系数非负 $w_i \leq 0$ 而且 $\sum\limits_{i=1}^{n} w_i = 1$，则线性组合

$$S = \sum_{i=1}^{n} w_i x_i$$

为一个凸组合.

凸组合的物理意义可以理解成 n 个重量为 w_i 的点的整体重心.

凸集合与凸函数

凸集合与凸函数

集合的凸包

n 个点 $\{x_i\}_{i=1}^n$ 的全部凸组合就构成 $\{x_i\}_{i=1}^n$ 的凸包.

函数的凸闭包

如果 C 是函数 f 的上境图, $\overline{C}$ 是 C 的凸包, 那么以 $\overline{C}$ 为上境图的函数称为 f 的凸闭包.

凸集合与凸函数

集合的凸包的性质

若 $\overline{C}$ 是 C 的凸闭包，那么

- $C \subset \overline{C}$
- C 的支撑平面也是 $\overline{C}$ 的支撑平面，反之亦然.

函数的凸闭包的性质

若 g 是 f 的凸闭包，那么

- $g \leq f$
- $\inf\{g\} = \inf f$

凸集合与凸函数的对应性质 (凸组合)

凸集合性质

假设 Ω 是一个凸集合，那么 Ω 任何子集的凸包仍包含于 Ω.

凸函数性质：琴生 (Jensen) 不等式

如果 $f:\Omega\to\mathbb{R}$ 是一个凸函数，则对于任何 $\{x_i\in\Omega\}_{i=1}^n$，以及凸组合 $\sum\limits_{i=1}^n w_i x_i$ 都有

$$\sum_{i=1}^n w_i f(x_i) \geq f(\sum_{i=1}^n w_i x_i)$$

话外篇：琴生 (Jensen) 不等式的推论

很多常用的不等式都是琴生不等式的推论，事实上，大部分不等式要么来自于 $x^2 \geq 0$ 要么来自于琴生不等式.

算数集合平均不等式

对于正数 $a_1, \cdots, a_n$

$$\frac{\sum_{i=1}^{n} a_i}{n} \geq \left(\prod_{i=1}^{n} a_i\right)^{\frac{1}{n}}$$

柯西不等式

$$\left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{i=1}^{n} b_i^2\right) \geq \left(\sum_{i=1}^{n} a_i b_i\right)^2$$

凸集合与凸函数

凸集合与凸函数的对应性质 (集合相交)

凸集合性质

任意多个凸集合的交集仍是凸集合.

凸函数性质

- 任意多个凸函数的逐点上确界仍是凸函数.
- 固定一个凸函数的若干个变量，所得的函数仍然是凸函数
- 凸函数的 sublevel set 都是凸集合.

凸集合与凸函数

凸集合与凸函数的对应性质 (线性组合)

凸集合性质

假设 $T : V \to W$ 是一个线性映射，则

- 若 Ω_V 是 V 中的凸集合，则 $\Omega_W = T(\Omega_V)$ 是 W 中的凸集合.
- 若 Ω_W 是 W 中的凸集合，则 $\Omega_V = T^{-1}(\Omega_W)$ 是 V 中的凸集合.

凸函数性质

- 凸函数的非负线性组合仍是凸函数,$f_1, \cdots, f_k$ 是凸函数，而且 $w_i \geq 0$，则 $\sum_{i=1}^{k} w_i f_k$ 也是凸函数.
- 若 $f : \mathbb{R}^n \to \mathbb{R}$ 是凸函数，$A \in \mathbb{R}^{n\times m}$,$b \in \mathbb{R}^n$，那么复合函数 $g(x) = f(Ax + b)$ 还是凸函数.

凸集合与凸函数

凸集合与凸函数的对应性质 (微分)

凸集合性质

- 若凸集合 Ω 的边界 C 是一个可微曲线，则 C 在任何一点上的切线（平面）都是这个凸集合的支撑线（平面）
- 若凸集合 Ω 的边界 C 是一个二阶可微曲线，则 C 在任何一点上的曲率向量都指向 Ω 内部.

曲率向量的物理意义是加速方向或者受力方向.

凸函数性质

- 若一个凸函数一阶可微，那么凸函数的一阶近似不大于函数本身.i.e.

$$f(x) \geq f(x_0) + (\nabla f(x_0))^T \cdot (x - x_0)$$

- 若一个凸函数二阶可微，那么这个函数的二阶导数 (Henssen 矩阵) 非负（半正定）.

凸集合与凸函数

凸集合与凸函数的对应性质 (光学投影)

凸集合性质

- 若 Ω 是凸集合，那么 Ω 在任何一个平面上的投影仍是凸集合（平行光源投影）
- 若 $\Omega \subset \mathbb{R}^n$ 是凸集合，那么

$$\Omega_{\hat{n}} = \{(x_1/x_n, \cdot, x_{n-1}/x_n, 1) : (x_1, \cdots, x_n) \in \Omega \text{且} x_n \neq 0\}$$

  也是凸集合.(点光源投影)
- 若 $\Omega \subset \mathbb{R}^n$ 是一个凸集合，那么椎体 $tx : x \in \Omega, t \in \mathbb{R}_+$ 也是个凸集合.(点光源)

凸集合与凸函数的对应性质 (光学投影)

凸函数性质

- 若 $f(x,y)$ 是凸函数, 那么 $g(x) = \inf_{(x,y)\in\Omega} f(x,y)$, 也是凸函数
- 若 $f:\mathbb{R}^n \to \mathbb{R}$ 是凸函数，那么 $g(x,t) = tf(x/t):\mathbb{R}^{n+1} \to \mathbb{R}$ 也是个凸函数.

凸集合与凸函数

凸集分离定理

凸集分离定理

若 C, D 分别为 $\mathbb{R}^n$ 中的两个不交的非空凸集合,i.e.$C \cap D = \emptyset$, 则一定存在向量 $a \in \mathbb{R}^n$ 以及实数 $b \in \mathbb{R}$ 使得任何 $x_C \in C, x_D \in D$ 有 $a^T x_C \le b$ 以及 $a^T x_D \ge b$.

定理中不等式的几何意义在于 C, D 分别位于超平面 $a^T x = b$ 的两边.

凸集分离定理

凸集分离定理

证明

分情况讨论：如果 C 是一个独点集 $\{p\}$，那么考虑集合 $\{|p-q|:q\in D\}$. 因为点之间的距离是非负数，所以根据确界原理存在一个序列 $\{q_i\in D\}_{i=1}^{\infty}$. 使得

$$\lim_{i\to\infty}|p-q_i|=\inf\{|p-q|:q\in D\}$$

而另一方面有界序列一定存在收敛子序列, 所以可以假定 $\lim_{i\to\infty} q_i=q_\infty$.

(1) 如果 $p\neq q_\infty$, 则令 $a=q_\infty-p$,$b=a^T(p+q_\infty)/2$, 我们得到

$$a^Tp=a^T(-\frac{q_\infty-p}{2}+\frac{p+q_\infty}{2})=-\frac{|a|^2}{2}+b\leq b$$
$$a^Tq_\infty=a^T(\frac{q_\infty-p}{2}+\frac{p+q_\infty}{2})=\frac{|a|^2}{2}+b\geq b$$

继续证明

此时对于任何一个 $q_0 \in D$, 考虑线段 $q_\infty + t(q_0 - q_\infty)$，因为 $|q_\infty - p| = \inf\{|q - p| : q \in D\}$, 所以

$$f(t) = |q_\infty + t(q_0 - q_\infty) - p|^2 \geq |q_\infty - p|^2 = f(0), \forall t \in (0,1)$$

所以 $f'(0) \geq 0$, 于是我们有

$$(q_\infty - p)^T(q_0 - q_\infty) \geq 0$$

也就是

$$a^T q_0 \geq a^T q_\infty \geq b$$

.

凸集分离定理

继续证明

(2) 如果 $p = q_\infty$，那么选取序列 $\{p_j\}_{j=1}^{\infty}$, 使得 p_j 不在 D 的闭包里面. 那么根据上一种情况我们知道，对每一个 j，都存在着 a_j 使得 $a_j^T q \geq a_j^T p_j$. 因为 a_j 的长度无关紧要，所以可以假设 a_j 都是单位向量。于是有界序列一定存在收敛子序列，所以可以假定 a_j 收敛于一个单位向量 a_∞. 那么对于任何一个 $q \in D$ 我们有

$$\begin{aligned} a_\infty^T p &= \lim_{j\to\infty} a_j^T p_j \\ &\leq \lim_{j\to\infty} a_j^T q \\ &= \lim_{j\to\infty} a_j^T q \\ &= a_\infty^T q \end{aligned}$$

令 $a = a_\infty$,$b = a_\infty^T p$. 即得所求.

凸集分离定理

继续证明

如果 C, D 是任意凸集合，考虑集合 $C - D = \{x - y : p \in C, q \in D\}$. 那么不难证明 $C - D$ 还是一个凸集合（作业）. 而且零向量 $O \notin C - D$. 于是使用前面证明的独点集与凸集合的分离定理我们有，存在一个向量 a 使得

$$a^T(p-q) \leq q^T O = 0$$

所以对于任意的 $p \in C$,$q \in D$, 都有

$$a^T q \geq a^T p$$

这时候令 $b = \sup\{a^T p : p \in C\}$, 既得所证。

共轭函数

共轭函数

如果 $f:\mathbb{R}^n \to \mathbb{R}$ 是一个函数, 那么 f 的共轭函数

$$f^*(y) = \sup_{x \in \mathbf{dom} f} (y^T x - f(x))$$

其中 $f^*(y)$ 的定义域是使得等式右边有界的那些 y.

共轭函数

共轭函数的性质

- 共轭函数 f^* 是一个凸函数
- 如果 g 是 f 的凸闭包，那么 $g^* = f^*$
- 如果 f 是一个凸函数，那么 $f^{**} = f$

共轭函数

共轭函数的进一步性质

- $f(x) + f^*(y) \geq x^T y$
- 如果 f 是凸函数而且可微，那么 $f^*(y) = x * T\nabla f(x^*) - f(x^*)$，其中 x^* 满足 $\nabla f(x^*) = y$.
- 如果 $g(x) = f(Ax + b)$, 则 $g^*(y) = f^*(A^{-T}y) - b^T A^{-T} y$.
- 如果 $f(u, v) = f_1(u) + f_2(v)$, 那么 $f^*(w, z) = f_1^*(w) + f_2^*(z)$

共轭函数

对偶问题：拉格朗日对偶函数

考虑 $\mathbb{R}^n$ 上的优化问题:

优化问题

$$\begin{aligned} \text{最小化:} & f_0(x) \\ \text{不等条件:} & f_i(x) \le b_i, i = 1, \cdots, m \\ \text{等式条件:} & h_i(x) = 0, i = 1, \cdots, p \\ \text{定义域:} & \mathcal{D} = \bigcap_{i=0}^{m} \mathsf{dom} f_i \cap \bigcap_{i=1}^{p} \mathsf{dom} h_i. \end{aligned}$$

请注意定义域 $\mathcal{D}$ 指的是使得所有函数 f_i, h_i 有定义的区域。而可行域指的是定义域中满足不等条件与等式条件的那些点. 本课中把这个优化问题称为原问题，优化点称为 x^*, 最优化值为 p^*.

对偶问题：拉格朗日对偶函数

根据原函数与限制条件我们定义拉格朗日量 $L(x,\lambda,\nu):\mathbb{R}^{n+m+p}\to\mathbb{R}$

拉格朗日量

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

根据拉格朗日函数我们定义拉格朗日对偶函数 $g(\lambda,\nu):\mathbb{R}^{m+p}\to\mathbb{R}$

拉格朗日对偶函数

$$\begin{aligned}g(\lambda,\nu)&=\inf_{x\in\mathcal{D}}L(x,\lambda,\nu)\\&=\inf_{x\in\mathcal{D}}f_0(x)+\sum_{i=1}^{m}\lambda_i f_i(x)+\sum_{i=1}^{p}\nu_i h_i(x)\end{aligned}$$

拉格朗日量在数学与物理中有极为广泛的应用，感兴趣的同学可以在“信息几何学”，“理论力学”中找到它其他的应用.

对偶问题：拉格朗日对偶函数

对偶函数有如下重要性质

对偶函数为原问题提供下界

如果限制 $\lambda_i \geq 0, \forall i = 1, \cdots, m$, 则

$$g(\lambda, \nu) \leq p^*$$

证明

对任意一个 $x \in \mathcal{D}$, 如果 x 在可行域中，那么

$$\begin{aligned} g(\lambda, \nu) &\leq f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \\ &= f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) \\ &\leq f_0(x) \end{aligned}$$

对偶问题：对偶问题

根据对偶函数，定义对偶问题的一般形式

对偶问题

$$\text{最大化:}g(\lambda,\nu)$$
$$\text{不等条件:}\lambda_i \geq 0, i=1,\cdots,m$$

我们把对偶问题的最大值点称为 (λ^*,ν^*)，相应的最大值称为 d^*，这里面的对偶函数 g 定义域为 $\mathrm{dom}g=\{(\lambda,\nu):g(\lambda,\nu)>-\infty\}$. 在 g 的定义域中满足 $\lambda_i \geq 0$ 的那些 (λ,ν) 全体，叫做对偶可行域. 也就是对偶问题的可行域.

对偶问题：拉格朗日对偶函数

对偶问题举例

原问题，线性规划

$$
\begin{aligned}
&\text{最小化:} c^T x \\
&\text{条件:} Ax = b, \text{而且} x_i \geq 0, \forall i = 1, \cdots, n
\end{aligned}
$$

对偶问题：拉格朗日对偶函数

对偶问题：拉格朗日对偶函数

对偶问题举例

线性规划的对偶问题

$$\text{最小化:} b^T\nu$$
$$\text{条件:} A^T\nu - \lambda + c = 0, \text{而且}, \nu_i \geq 0, \forall i = 1, \cdots, n.$$

对偶问题：共轭函数与拉格朗日对偶函数

当优化问题的限制条件是线性条件时，可以利用共轭函数的一些性质方便的得到对偶问题

线性约束优化问题的对偶问题

$$
\begin{aligned}
\text{最小化:}&f_0(x)\\
\text{不等条件:}&AX \leq b\\
\text{等式条件:}&Cx = d
\end{aligned}
$$

这里面向量的比较 $u < v$ 指的是 u 里边每一个分量都小于 v 里面对应的分量.

对偶问题：共轭函数与拉格朗日对偶函数

$$
\begin{aligned}
g(\lambda,\nu) &= \inf_x \left(f_0(x) + \lambda^T(Ax-b) + \nu^T(Cx-d)\right) \\
&= -b^T\lambda - d^T\nu + \inf_x \left(f_0(x) + (A^T\lambda + C^T\nu)^T x\right) \\
&= -b^T\lambda - d^T\nu - f_0^*(-A^T\lambda - C^T\nu)
\end{aligned}
$$

而这个函数的定义域就是

$$
\mathrm{dom}\, g = \{(\lambda,\nu) : -A^T\lambda - C^T\nu \in \mathrm{dom}\, f_0^*\}.
$$

对偶问题：共轭函数与拉格朗日对偶函数

最小化向量范数

对偶问题举例

$$\text{最小化:}|x|$$
$$\text{条件:}Ax = b$$

若 $f(x) = |x|$ 那么共轭函数是 $f^*(y) = 0, \forall |y| \le 1$, 否则 $f^*(y) = +\infty$. 所以对偶问题就是：

$$\text{最大化:}|-b^T\nu|$$
$$\text{条件:}|A^T\nu| \le 1$$

最大熵问题

对偶问题举例

$$\text{最小化:} f_0(x) = \sum_{i=1}^{n} x_i \ln x_i$$

$$\text{不等条件:} Ax \le b$$

$$\text{等式条件:} \mathbf{1}^T x = 1$$

因为 $(f_1(x) + f_2(y))^* = f_1^*(w) + f_2^*(z)$, 我们有

$$f_0^*(y) = \sum_{i=1}^{n} e^{y_i - 1}$$

所以对偶函数为 $g(\lambda, \nu) = -b^T\lambda - \nu - e^{-\nu-1}\sum_{i=1}^{n} e^{-a_i^T\lambda}$

最大熵问题的对偶问题就是

最大熵问题的对偶问题

$$\text{最大化:} -b^T\lambda - \nu - e^{-\nu-1}\sum_{i=1}^{n} e^{-a_i^T\lambda}$$

不等条件:$\lambda \geq 0$

对偶性

根据对偶函数的性质我们已经知道在对偶可行域中，$g(\lambda, \nu)$ 总是不大于 p^*. 所以就有

弱对偶性

$$d* \leq p*$$

若对偶性总是对的. 相对而言的强对偶性是指一部分优化问题来说，有更好的结论.

强对偶性

$$d* = p*$$

强对偶性不总成立.

强对偶性条件

第一个强对偶性的条件, 几乎所有的凸优化问题都满足强对偶性.

Slater 条件

对于一个凸优化问题

$$\begin{aligned} \text{最小化:} & f_0(x) \\ \text{不等条件:} & f_i(x) \le b_i, i = 1, \cdots, m \\ \text{等式条件:} & h_i(x) = 0, i = 1, \cdots, p \end{aligned}$$

如果存在一个可行域中的点 x 使得 $f_i(x) < 0, i = 1, \cdots, m$, 那么这个凸优化问题就满足强对偶条件.

强对偶性条件

根据 Slater 条件，我们前述的几个例子都是满足强对偶性的.

满足强对偶性的例子

- 线性规划
- 最小二乘
- 最大熵问题

这种情况下我们如果发现对偶问题比原问题更容易解决，那么就可以使用对偶问题来解出 $d^* = p^*$

凸优化问题求解 (KKT)

我们来看一下如果强对偶性满足的话，这些最优化点应该满足何种条件. 这一部分中我们假定所有的函数都是可微函数.
如果 $x^*, (\lambda^*, \nu^*)$ 分别是原问题与对偶问题的最优解，那么首先这些点应该满足可行域条件

- $f_i(x^*) \leq 0$
- $h_i(x^*) = 0$
- $\lambda_i^* \geq 0$

凸优化问题求解 (KKT)

其次我们已经知道

$$\begin{aligned} d^* &= g(\lambda^*, \nu^*) \\ &\leq f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* f_i(x^*) + \sum_{i=1}^{p} \nu_i^* h_i(x^*) \\ &= f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* f_i(x^*) \\ &\leq f_0(x^*) \\ &= p^* \end{aligned}$$

于是 $d^* = p^*$ 意味着上述不等式全都是等式.

凸优化问题求解 (KKT)

所以我们有

$$\sum_{i=1}^{m} \lambda_i^* f_i(x^*) = 0$$

，以及

$$g(\lambda^*, \nu^*) = L(x^*, \lambda^*, \nu^*)$$

而因为

$$g(\lambda^*, \nu^*) = \inf(L(x^*, \lambda^*, \nu^*))$$

所以 x^* 是拉格朗日函数在 x 方向的驻点，所以有

$$\nabla_x L(x^*, \lambda^*, \nu^*) = 0$$

. 综上所述我们就的到了 KKT 条件.

KKT 条件

- $f_i(x^*) \leq 0,\ i = 1, \cdots, m$
- $h_i(x^*) = 0, i = 1, \cdots, p$
- $\lambda_i^* \geq 0,\ i = 1, \cdots, m$
- $\lambda_i^* f_i(x^*) = 0,\ i = 1, \cdots, m$
- $\nabla_x L(x^*, \lambda^*, \nu^*) = 0$

其中第四个条件是由 $\sum\limits_{i=1}^{m} \lambda_i^* f_i(x^*) = 0$ 以及第一个和第三个条件共同得到的.

所以我们有

$$\sum_{i=1}^{m} \lambda_i^* f_i(x^*) = 0$$

, 以及

$$g(\lambda^*, \nu^*) = L(x^*, \lambda^*, \nu^*)$$

而因为

$$g(\lambda^*, \nu^*) = \inf(L(x^*, \lambda^*, \nu^*))$$

所以 x^* 是拉格朗日函数在 x 方向的驻点，所以有

$$\nabla_x L(x^*, \lambda^*, \nu^*) = 0$$

. 综上所述我们就的到了 KKT 条件.

凸优化问题求解 (KKT)

KKT 条件叙述

- $f_i(x^*) \le 0,\ i = 1, \cdots, m$
- $h_i(x^*) = 0, i = 1, \cdots, p$
- $\lambda_i^* \ge 0,\ i = 1, \cdots, m$
- $\lambda_i^* f_i(x^*) = 0,\ i = 1, \cdots, m$
- $\nabla_x L(x^*, \lambda^*, \nu^*) = 0$

其中第四个条件是由 $\sum\limits_{i=1}^{m} \lambda_i^* f_i(x^*) = 0$ 以及第一个和第三个条件共同得到的.

凸优化问题求解 (KKT)

KKT 条件使用

- 对于凸优化问题,KKT 条件是 $x^*, (\lambda^*, \nu^*)$ 分别作为原问题和对偶问题的最优解的充分必要条件.
- 对于非凸优化问题，KKT 条件仅仅是必要而非充分.

使用 KKT 条件解决优化问题例子

$$\text{最小化}:(1/2)x^P x + q^T x + r$$
$$\text{等式条件}:Ax = b$$

KKT 条件是

$$Ax^* = b$$

以及

$$Px^* + q + A^T v^* = 0$$

我们求解这个线性方程即可得到索要的结果.

应用举例：支持向量机最简单形式

支持向量机的最简单形式

空间 $\mathbb{R}^n$ 中有可分的两个点集 C,D. 我们希望找到一个最合适的超平面 $a^Tx = b$ 对他们进行区分. 也就是说对于 $p \in C, q \in D$, 有

$$a^T p > b, \quad a^T q < b$$

也就是说存在一个正数 t 使得

$$a^T p - b \geq t, \quad a^T q - b \leq -t$$

这个正数 t 一定程度上描述了两个集合被分开的程度.

为了体现“最合适”，一个比较好的想法就是希望当 a 是单位向量的时候，t 越大越好. 于是的到了一个优化问题

应用举例：支持向量机最简单形式

SVM 优化问题

$$
\begin{aligned}
\text{最小化:} \quad & -t \\
\text{不等条件 1:} \quad & -a^T p_i + b \leq -t \\
\text{不等条件 2:} \quad & a^T q_i - b \leq -t \\
\text{不等条件 3:} \quad & -t \leq 0 \\
\text{等式条件:} \quad & |a|^2 = 1
\end{aligned}
$$

但是这里面 $|a|^2 = 1$ 并非凸条件，所以我们把它换成 $|a|^2 \leq 1$.
就得到一个跟原问题等价的凸优化问题.

应用举例：支持向量机最简单形式

SVM 凸优化问题

$$
\begin{aligned}
\text{最小化:}\quad & -t \\
\text{不等条件 1:}\quad & -a^T p_i + b \le -t \\
\text{不等条件 2:}\quad & a^T q_i - b \le -t \\
\text{不等条件 3:}\quad & -t \le 0 \\
\text{不等条件 4:}\quad & |a|^2 \le 1
\end{aligned}
$$

具体求解，因为这是一个凸优化问题，所以可以选择使用 KKT 条件来解方程求解.

线性代数：参考资料

参考资料

- 非常清晰完整的经典教材：凸优化，Stephen Boyd, Lieven Vandenberghe

作业

- 教材：凸优化 (英文版页码与题号) p60:2.32.10,2.12,2.16,2.23,2.26; p113:3.1,3.2,3.12,3.21,3.36,3.39; p189:4.2,4.8,4.10,4.21 p273:5.1,5.4,5.11,5.225.24

谢谢大家!