凸优化笔记1.

1. 仿射集 affine set

$\forall x_1, x_2 \in C, \theta \in \mathbb{R}$. 有 $\theta x_1 + (1-\theta)x_2 \in C$.

则C为仿射集

即.如任意x_1,x_2在C中 $\boxed{\text{直线}}$ x_1x_2也在C中,则C是仿射集

2. 仿射组合 affine combination.

设$x_1, x_2 \cdots x_n \in C$. 若$\theta_1 \cdots \theta_n \in \mathbb{R}$ 且 $\sum \theta_i = 1$.

将 $\theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$ 称为仿射组合.

$\sum \theta_i x_i \in C$

3. 仿射包 affine hull.

$$\mathrm{aff}\, C = \{x_1\theta_1 + \cdots + x_n\theta_n \mid x_1 \cdots x_n \in C, \theta_1 + \cdots + \theta_n = 1\}$$

4. 凸集 convex set

一个集合是凸的,若C中任意两点之间的线段都在C中.

$\forall x_1, x_2 \in C$. $\forall \theta \in [0,1]$, $x_1\theta + (1-\theta)x_2 \in C$.

5. 凸组合.

设$x_1 \cdots x_n \in C$. $\theta_1 + \cdots + \theta_n = 1$. 则 $\theta \geq 0$

$\theta_1 x_1 + \cdots + \theta_n x_n$称为凸组合.

6. 凸包 convex hull.

$$\mathrm{conv}\, C = \{\sum \theta_i x_i \mid x_i \in C, \theta_i \geq 0, \sum \theta_i = 1\}$$

7. 锥 cone

如果对$\forall x \in C$, $\theta \geq 0$都有 $\theta x \in C$, 则称C为锥. 锥必过原点

8. 凸锥.

对$\forall x_1, x_2 \in C$, $\theta_1, \theta_2 \geq 0$

都有 $x_1\theta_1 + x_2\theta_2 \in C$. $\Rightarrow$ C为凸锥

凸优化笔记2.

1. 凸函数

$f: \mathbb{R}^n \to \mathbb{R}$ 是凸的，如果

① 定义域 $\mathrm{dom}(f)$ 是凸集

② $\forall x, y \in \mathrm{dom}(f)$，$\theta \in [0,1]$

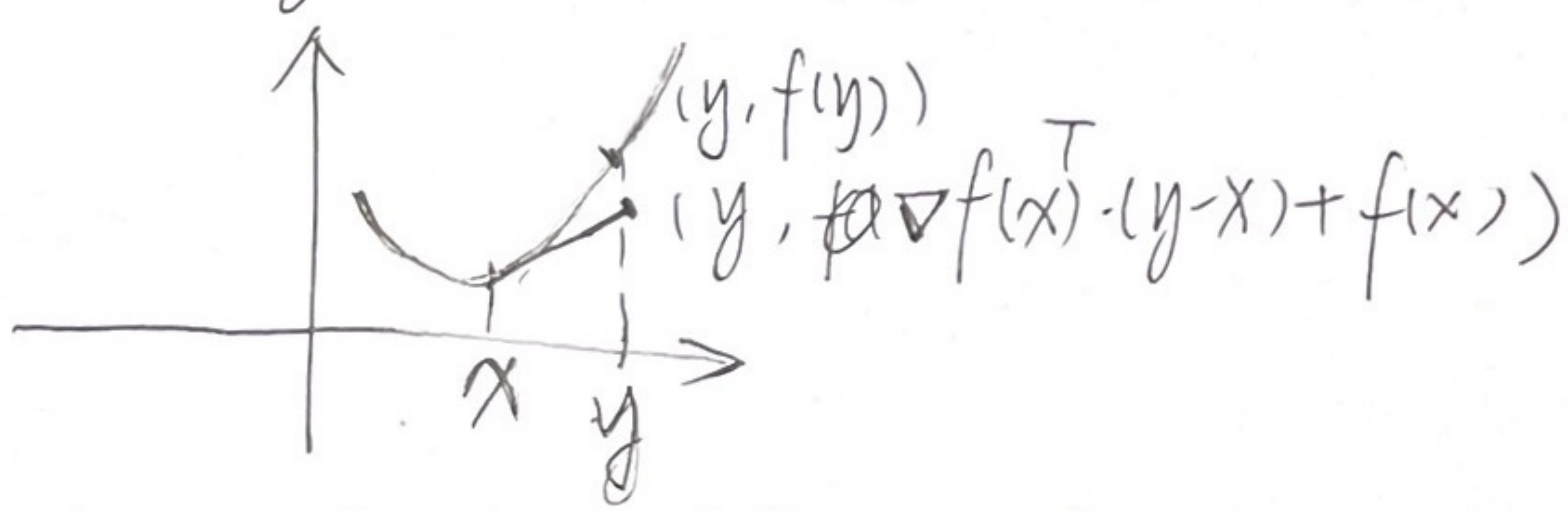
有 $f(\theta x + (1-\theta) y) \le \theta f(x) + (1-\theta) f(y)$

2. 一阶条件.

f是凸函数，iff..

① $\mathrm{dom}(f)$ 为凸集

② $\forall x, y \in \mathrm{dom}(f)$，有 $f(y) \ge f(x) + \nabla f(x)^T (y-x)$



因此，对于可微凸函数 若 $\nabla f(x_0) = 0$，则 x_0 是极小值点.

3. 二阶条件.

f是凸函数，iff ① $\mathrm{dom}(f)$ 为凸集

② $\forall x \in \mathrm{dom}(f)$，$\nabla^2 f(x) \succeq 0$

其中 $\nabla^2 f(x)$ 为Hessian矩阵 $H(x)$.

$$\nabla^2 f(x) = H(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots \\ \vdots & \vdots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{bmatrix}$$

当 $H(x) \succeq 0$ 为(半正定矩阵)时，f是凸函数.

4. 下水平集.

$f: \mathbb{R}^n \to \mathbb{R}$ 的 α 下水平集为 $C_\alpha = \{x \in \mathrm{dom} f \mid f(x) \le \alpha\}$

5. 范数.

满足以下条件的函数 $f: \mathbb{R}^m \to \mathbb{R}$，$\mathrm{dom}(f) = \mathbb{R}^m$ 称为范数

① f非负：$\forall x \in \mathbb{R}^m$，$f(x) \ge 0$

② f正定：仅对 $x=0$，$f(x)=0$

③ f齐次：对所有 $x \in \mathbb{R}^m$，$t \in \mathbb{R}$ 有 $f(tx) = |t| f(x)$

④ f满足三角不等式：$\forall x, y \in \mathbb{R}^m$，$f(x+y) \le f(x) + f(y)$

凸优化笔记3.

1、凸优化问题.

形如 $\min_{x} f(x)$

$\text{s.t.}\quad g_i(x)\le 0 \quad i=1\cdots m$

$h_j(x)=0 \quad j=1\cdots p$

← 此处必是 $g_i(x)\le 0$ 而非 $g_i(x)\ge 0$.
因为凸函数的下水平集总是凸集
若写成 $g_i(x)\ge 0$, 则可行域不一定为凸集
算法不可以保证得到全局最优解.

其中 $x\in\mathbb{R}^n$ 为优化变量.

$f(x)$, $g_i(x)$ 为凸函数

$h_j(x)$ 为仿射函数

定义该凸问题定义域为 $\mathbb{D}=\text{dom}(f)\cap\left(\bigcap_{i=1}^{m}\text{dom}(g_i)\right)\cap\left(\bigcap_{j=1}^{p}\text{dom}(h_j)\right)$

当 $x\in\mathbb{D}$ 且满足 $g_i(x)\le 0$, $h_j(x)=0$ 时 称 x 是可行路, 由这样 x 构成可行域.

可行域中使 $f(x)$ 最小的 x^* 为最优点, 目标函数的值为最优值 $p^*=f(x^*)$

凸优化问题的基础性质: 任意局部最优解就是全局最优解.

2、示性函数.

对于凸集 C. 它的示性函数为 $I_C(x)=\begin{cases}0 & \text{if } x\in C.\\ \infty & \text{if } x\notin C.\end{cases}$

4、共轭函数

设函数 $f: \mathbb{R}^n \to \mathbb{R}$

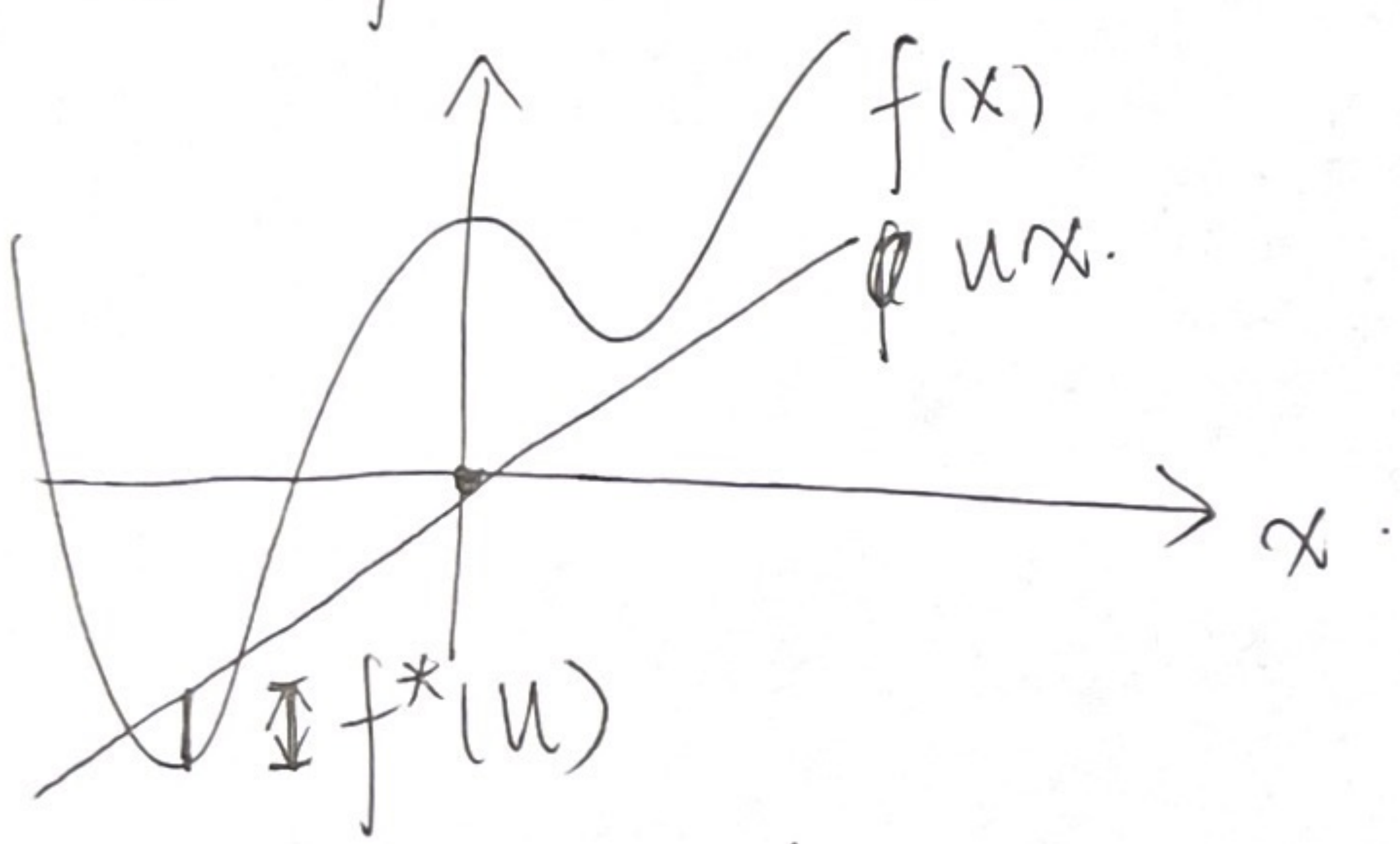
定义函数 $f^*: \mathbb{R}^n \to \mathbb{R}$ 为

$$f^*(u) = \sup_{x \in \mathrm{dom}(f)} \{u^T x - f(x)\}.$$

称 f^* 为 f 的共轭函数，

无论 f 是否为凸函数，f^* 都为凸函数.

eg. 当 $f: \mathbb{R} \to \mathbb{R}$. 给定 u.



eg. $f(x) = -\log(x)$.

$$f^*(u) = \sup_{x>0} \{ux + \log(x)\}$$

对于固定的 u. 令 $ux + \log x$ 对 x 求导后为0 $\Rightarrow u + \frac{1}{x} = 0$

$\Rightarrow f^*(u) = -\log(-u) - 1$

eg. $f(x) = e^x$.

$$f^*(u) = \sup_{x \in \mathbb{R}} \{ux - e^x\}$$

$$\frac{\partial(ux - e^x)}{\partial x} = u - e^x = 0 \Rightarrow x = \ln u.$$

$\therefore f^*(u) = u\ln u - u.$

4. 对偶范数

令 $\|\cdot\|$ 为 $\mathbb{R}^n$ 上的范数，对应的对偶范数用 $\|\cdot\|_*$ 表示

$$\|z\|_* = \sup\{z^T x \mid \|x\| \le 1\}$$

即，对于 $z \in \mathbb{R}^n$，其对偶范数是寻找一个向量 $x \in \mathbb{R}^n$，$\|x\| \le 1$，使得 x 与 z 内积最大，这个最大内积就是 z 的对偶范数。

√ l_1 的对偶范数是 l_∞

l_p 的对偶范数是 l_q.（$\frac{1}{p} + \frac{1}{q} = 1$）

√ 范数的共轭函数是示性函数

若 $f(x) = \|x\|$，则它的共轭函数是

$$f^*(y) = \begin{cases} 0 & \|y\|_* \le 1 \\ \infty & \text{otherwise} \end{cases} = I_{\{z : \|z\|_* \le 1\}}(y)$$

用共轭函数定义，$f(x) = \|x\|$ 的共轭函数为 $f^*(y) = \sup\{y^T x - \|x\|\}$

(1) 如果 $\|y\|_* > 1$，则存在 $z \in \mathbb{R}^n$，$\|z\| \le 1$，使得 $y^T z \ge 1$，

由于 $x \in \mathbb{R}^n$，取 $x = t \cdot z$，则 $f^*(y) = \sup\{t(y^T z - \|z\|)\}$

令 $t \to \infty$，则 $f^*(y) \to \infty$

(2) 若 $\|y\|_* \le 1$，对 $\forall x$，有 $y^T x \le \|x\| \cdot \|y\|_*$

$\therefore y^T x \le \|x\|$，当 $x = 0$ 时，$f^*(y)$ 有最大值 0.

凸优化笔记4.

1、近端梯度下降算法 proximal gradient descent

可以快速求解凸优化问题.

目标函数在某些地方不可微，但可以拆成可微的函数 g 与不可微的函数 h 之和

$$\min f(x) = \min \{ g(x) + h(x) \}$$

$g(x)$：凸函数，可微.　$h(x)$：凸函数，未必可微.

✓ 近端映射

近端映射定义：

$$\mathrm{prox}_{h,t}(x) = \underset{z}{\arg\min}\ \frac{1}{2t}\|x-z\|_2^2 + h(z)$$

即对于给定 x，找到 $z = \mathrm{prox}_{h,t}(x)$ 使得 $\frac{1}{2t}\|x-z\|_2^2 + h(z)$ 最小，

这样的 z 使 h 很小且接近原不可微点 x.

✓ 梯度迭代

选择初始点 $x^{(0)}$，重复

$$x^{(k)} := \mathrm{prox}_{h,t_k}\left(x^{(k-1)} - t_k \nabla g(x^{(k-1)})\right).$$

即梯度下降后再在附近找一个可微的点作为这次更新值

用可微的函数 g.

√近端映射的几个特殊形式 $h(x)=0$ 或 $I_C(x)$ 或 $\|x\|_1$ 时.

① 常规梯度下降.

令 $h(x)=0$

近端映射 $prox_t(x) = \underset{z}{argmin}\ \frac{1}{2t}\|x-z\|_2^2 + 0$

$= x$.

∴更新策略: $x^{(k)} = x^{(k-1)} - t_k \nabla g(x^{(k-1)})$

② 常规投影梯度下降.

$h(x) = I_C(x)$

近端映射: $prox_t(x) = \underset{z}{argmin}\ \frac{1}{2t}\|x-z\|_2^2 + I_C(z)$

$= \underset{z\in C}{argmin}\ \frac{1}{2t}\|x-z\|_2^2$

$= P_C(x)$. ($P_C(\cdot)$为C上的投影算子).

更新: $x^{(k)} = P_C(x^{(k-1)} - t_k \cdot \nabla g(x^{(k-1)}))$.

③ 软阈值算子.

令 $h(x)=\|x\|_1$

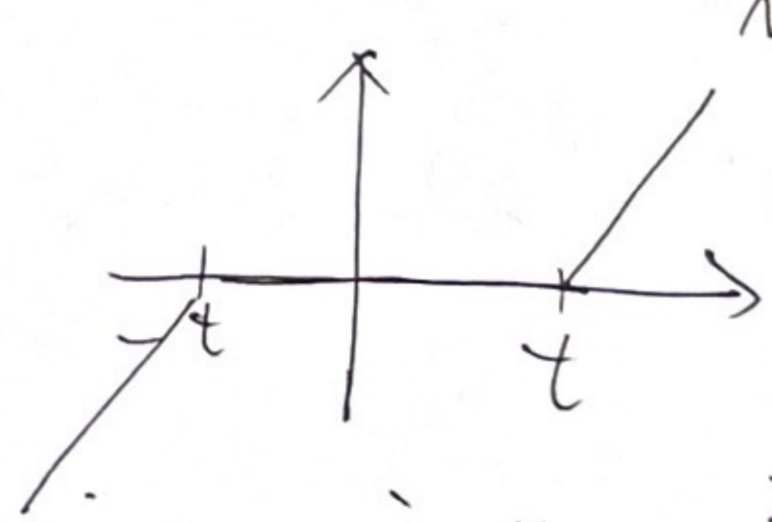
近端映射 $prox_t(x) = \underset{z}{argmin}\ \frac{1}{2t}\|x-z\|_2^2 + \|z\|_1$

$= \underset{z}{argmin}\ \|x-z\|_2^2 + 2t\cdot\|z\|_1$

$= S_t(x)$

用 $[S_t(x)]_i$ 表示 $S_t(x)$ 的第i个分量.

$$[S_t(x)]_i = \begin{cases} x_i - t, & x_i > t \\ 0, & |x_i| \le t \\ x_i + t, & x_i < -t \end{cases}$$



对形如 $\underset{x}{argmin}\ \|x-b\|_2^2 + \lambda\|x\|_1$ 这样的目标函数, 它的解为 x^*, 第i个分量为

$x_i^* = [S_{\frac{\lambda}{2}}(b)]_i$

✓ 近端梯度下降求Lasso.

Lasso问题

$$f(\beta)=\frac{1}{2}\|y-X\beta\|_2^2+\lambda\|\beta\|_1.$$

$f(\beta)$ 可拆为 $\frac{1}{2}\|y-X\beta\|_2^2$ 和 $\lambda\|\beta\|_1$.

$$\therefore \operatorname{prox}_t(\beta)=\underset{z}{\operatorname{argmin}}\ \frac{1}{2t}\|\beta-z\|_2^2+\lambda\|z\|_1$$

$$=\underset{z}{\operatorname{argmin}}\ \|\beta-z\|_2^2+2t\lambda\|z\|_1$$

$$=S_{t\lambda}(\beta)$$

由于 ~~$\nabla g(x)$~~ $\nabla g(\beta)=-X^T(y-X\beta)$

$\therefore$ 更新策略: $\beta^{(k)}=\cancel{\beta^{(k-1)}}\ S_{t\lambda}(\beta^{(k-1)}+t_kX^T(y-X\beta^{(k-1)}))$.

凸优化笔记5. 次梯度下降.

解决目标函数并不总是处处可导的问题.

优势: 比传统方法处理问题范围更大.

缺点: 收敛速度更慢.

√回顾. 若f为凸函数. 则 $\forall x,y \in \mathrm{dom} f$

有 $f(y) \geq f(x) + \nabla f(x)^T (y-x)$

类比一阶条件. 给定函数f, 对 $\forall y$ 如果

$f(y) \geq f(x) + g^T (y-x)$

则g是f在x处的次梯度.

√将f在x处所有次梯度构成的集合称为

f在x处的次微分. Subdifferential.

记为 $\partial f(x)$ (次微分是个集合).

◦凸函数的次微分总是非空.

◦非凸函数即使可微也不一定有次梯度.

◦凹函数的次梯度是空集.

√次梯度计算.

设函数f在 x_0 处不一定可导.

左导数 $a = \lim\limits_{x \to x_0^-} \frac{f(x)-f(x_0)}{x-x_0}$

右导数 $b = \lim\limits_{x \to x_0^+} \frac{f(x)-f(x_0)}{x-x_0}$

凸函数f在 x_0 的次微分 $\partial f(x_0) = [a,b]$

[a,b]中任何一个取值都是次梯度.

√次梯度优化条件.

当 x^* 是函数f的最优解, 当且仅当 x^* 处的次微分中包含0.

即 $f(x^*) = \min\limits_x f(x) \Leftrightarrow 0 \in \partial f(x^*)$

√次梯度优化算法.

$x^{(k)} = x^{(k-1)} - t_k \cdot g^{(k-1)}$.

$g^{(k-1)} \in \partial f(x^{(k-1)})$ 即从次微分中随机

选择一个次梯度作为梯度.

为了保证更新的参数使f(x)下降. 只有当次梯度

计算的x使f更小才更新x.

√ 次梯度求解 lasso 问题.

lasso: $\min_{\beta} \frac{1}{2}\|y-X\beta\|_2^2+\lambda\|\beta\|_1$

由次梯度优化条件，0是该目标函数在最优解上次微分的一个元素.

即. $0 \in \partial \frac{1}{2}\|y-X\beta\|_2^2+\lambda\|\beta\|_1$

$0 \in -X^T(y-X\beta)+\partial\lambda\|\beta\|_1$

$\Rightarrow$ 记 $v=\partial\|\beta\|_1$

$X^T(y-X\beta)=\lambda\cdot v$

由次梯度计算方式、得 v 的第 i 个分量满足

$$v_i=\begin{cases}1, & \beta_i>0\\ [-1,1], & \beta_i=0\\ -1, & \beta_i<0\end{cases}$$

为简单起见. 假设 $X=I$，则 $\beta=y-\lambda v$

当 $\beta_i>0$ 时. $v_i=1$, $\beta_i=y_i-\lambda>0 \Rightarrow y_i>\lambda$

当 $\beta_i=0$ 时 $|v_i|\le 1$. $|y_i|=\lambda|v_i|<\lambda$. $\Rightarrow$

当 $\beta_i<0$ 时 $v_i=-1$ $\beta_i=y_i+\lambda<0 \Rightarrow y_i<-\lambda$.

即. $\beta_i^*=[S_\lambda(y)]_i$

凸优化笔记6.

把问题转化为对偶问题，使其便于求解.

对偶性是凸优化的核心内容.（duality）

1. 原始问题标准转化.

原始问题 $\min_{x} f_0(x)$

$$\text{s.t. } f_i(x) \le 0,\ i=1\cdots m.$$
$$h_j(x)=0,\ j=1\cdots p.$$

2. 拉格朗日函数. Lagrangian function.

$$L(x,\lambda,v)=f_0(x)+\sum_{i=1}^{m}\lambda_i f_i(x)+\sum_{j=1}^{p}v_j h_j(x).$$

3. 拉格朗日对偶函数 Lagrangian dual function.

对 Lagrangian 对 x 取下确界得到.

$$g(\lambda,v)=\inf_{x} L(x,\lambda,v)$$

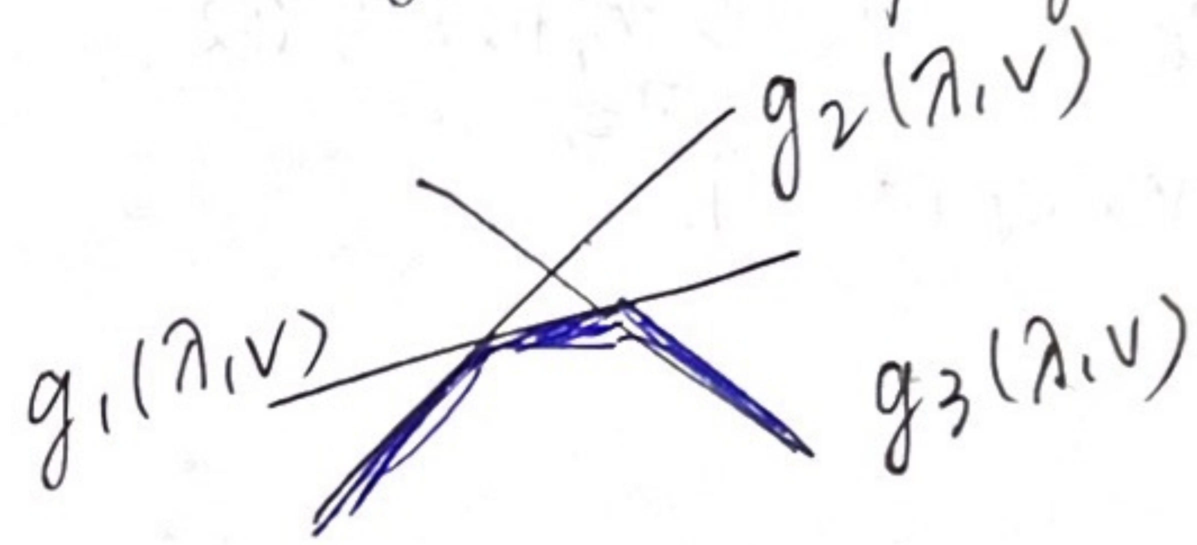
对偶函数有如下的两条重要性质：

① 对偶函数一定是凹函数.（其凹性与原目标函数和约束条件无关）

证明：$L(x,\lambda,v)$ 可看作无限的函数集，其中每个元素都是 $L(x_i,\lambda,v)$，x_i 取遍全部 $\mathcal{D}$ 得到.

$\therefore L(x_i,\lambda,v)$ 退化为 $g_i(\lambda,v)$

$\therefore g(\lambda,v)=\inf\{g_1(\lambda,v),\dots,g_\infty(\lambda,v)\}$



② 对 $\forall\lambda\ge0$，$\forall v$，如果原始问题的最优值为 p^*，则 $g(\lambda,v)\le p^*$

证明：设最优解为 x^*

$$\therefore L(x^*,\lambda,v)=f_0(x^*)+\sum_{i=1}^{m}\lambda_i f_i(x^*)+\sum_{j=1}^{p}v_j h_j(x^*)$$
$$\le f_0(x^*)=p^*$$

假设 x^* 在 x_1 处取得.（①中的证明）

$\therefore L(x^*,\lambda,v)=g_1(\lambda,v)$

由上图知，$g(\lambda,v)$ 总在 $g_1(\lambda,v)$ 下方.

$\therefore g(\lambda,v)<g_1(\lambda,v)=L(x^*,\lambda,v)\le p^*$

4、拉格朗日对偶问题 Lagrange dual Problem.

由 $g(\lambda,v)$ 性质②，

当 $\lambda\succeq0$，$\forall v$ 时，$g(\lambda,v)$ 是 p^* 的下确界。

则最好的下确界就是最大化对偶函数，

因此构造原问题的对偶问题

$$\max_{\lambda,v}\ g(\lambda,v)$$
$$\text{s.t.}\ \ \lambda\succeq0.$$

由于 $g(\lambda,v)$ 为凹函数(性质①)

故该对偶问题一定为凸优化问题，

对应的最优解为 λ^*,v^* (最优Lagrange乘子)

若对应的最优值为 d^*，则有 $d^*\le p^*$

当 $d^*\le p^*$ 时，称为弱对偶。

当 $d^*=p^*$ 时，称为强对偶。

将 p^*-d^* 称为对偶间隙。

✓ Slater条件：用于判断什么情况下强对偶成立.(充分条件)

在原问题是凸问题的条件下，若 $\exists x\in \text{relint}(D)$

使得约束条件满足 $f_i(x)<0$，$h_j(x)=0$

则强对偶成立.(一般大多数原凸问题，强对偶都是成立的)

(relint (D) 表示D除了边界点之外的所有点).

5、KKT条件.

在对偶间隙 p^*-d^* 为0，且L对x可微时，

设 x^*,λ^*,v^* 分别为原问题和对偶问题最优解.

则以下四组条件称为KKT条件：

$$\begin{cases}\left.\dfrac{\partial L(x,\lambda^*,v^*)}{\partial x}\right|_{x=x^*}=0 & \text{(stationary)}\\ \lambda_i^*f_i(x^*)=0 & \text{(complementary slackness)}\\ f_i(x^*)\le0,\ h_j(x^*)=0 & \text{(primal feasibility)}\\ \lambda_i^*\ge0 & \text{(dual feasibility)}\end{cases}$$

i. 稳定性条件：若 x^* 为原问题最优解，则 $L(x,\lambda^*,v^*)$ 在 x^* 处微分为0

ii. 互补松弛条件：

当 $\lambda_i^*>0$ 时，$f_i(x^*)=0$.

当 $f_i(x^*)<0$ 时，$\lambda_i^*=0$.

证明：$f_0(x^*)=g(\lambda^*,v^*)$

$$=\inf_x\left\{f_0(x)+\sum_i^m\lambda_if_i(x)+\sum_j^p v_j h_j(x)\right.$$

由 $g(\lambda^*,v^*)\le g_i(\lambda^*,v^*)$ 得到

$$\le f_0(x^*)+\sum_i^m\lambda_if_i(x^*)+\sum_j^p v_jh_j(x^*)$$

$$\le f_0(x^*).$$

$$\therefore\sum_{i=1}^m\lambda_if_i(x^*)=0$$

iii. 原问题可行性：原问题的最优解必然满足对偶问题的约束条件.

iv. 对偶问题可行性：对偶问题的解必然满足对偶问题的约束条件.

eg. lasso 和 KKT 条件.

i. lasso 问题 $\min\limits_{\beta} \frac{1}{2}\|y-X\beta\|_2^2+\lambda\|\beta\|_1$

ii. 写成标准的优化问题:

$$\min_{\beta,z} \frac{1}{2}\|y-z\|_2^2+\lambda\|\beta\|_1$$

$$\text{s.t.} \quad z-X\beta=0$$

iii. Lagrangian.

$$L(\beta,z,u)=\frac{1}{2}\|y-z\|_2^2+\lambda\|\beta\|_1+u^T(z-X\beta)$$

iv. 写出对偶函数

$$g(u)=\inf_{\beta,z} L(\beta,z,u)$$

$$=\min_{\beta,z} \frac{1}{2}\|y-z\|_2^2+\lambda\|\beta\|_1+u^T(z-X\beta)$$

$$=\min_{z}\left\{\frac{1}{2}\|y-z\|_2^2+u^Tz\right\}+\min_{\beta}\left\{\lambda\|\beta\|_1-u^TX\beta\right\}.$$

利用 $0=\dfrac{\partial\left(\frac{1}{2}\|y-z\|_2^2+u^Tz\right)}{\partial z}$

$$= u^Ty-\frac{1}{2}\|u\|_2^2+\min_{\beta}\left\{\lambda\|\beta\|_1-u^TX\beta\right\}.$$

$$=-\frac{1}{2}\left(\|u\|_2^2-2u^Ty+\|y\|_2^2\right)+\frac{1}{2}\|y\|_2^2+\min_{\beta}\left\{\lambda\|\beta\|_1-u^TX\beta\right\}$$

$$=\frac{1}{2}\|y\|_2^2-\frac{1}{2}\|u-y\|_2^2-\lambda\max_{\beta}\left\{\frac{u^TX\beta}{\lambda}-\|\beta\|_1\right\}.$$

由于 $f^*\left(\frac{X^Tu}{\lambda}\right)=\max\limits_{\beta}\left\{\frac{u^TX\beta}{\lambda}-\|\beta\|_1\right\}$.

是 $f(\beta)=\|\beta\|_1$ 的共轭函数

∵ 范数函数的共轭函数是示性函数

$$\therefore =\frac{1}{2}\|y\|_2^2-\frac{1}{2}\|u-y\|_2^2-\lambda\, \mathbb{I}_{\{v:\|v\|_\infty\le 1\}}\left(\frac{X^Tu}{\lambda}\right)$$

v. 得对偶问题

$$\max_{u} \frac{1}{2}\|y\|_2^2-\frac{1}{2}\|y-u\|_2^2$$

$$\text{s.t.} \quad \|X^Tu\|_\infty\le\lambda.$$

可进一步简化.

$$\min \|y-u\|_2^2$$

$$\text{s.t.} \quad \|X^Tu\|_\infty\le\lambda.$$

vi. KKT条件.

1. 稳定性条件.

$$0=\frac{\partial L(\beta, z, u^*)}{\partial z}\bigg|_{z=z^*}=-(y-z^*)+u^*$$

2. 无互补松弛条件

3. 原问题可行性.

$$z^*-X\beta^*=0$$

4. 对偶问题可行性 $\|X^T u^*\|_\infty \leq \lambda$.

得KKT条件为 $\begin{cases} z^*-y+u^*=0 \\ z^*=X\beta^* \\ \|X^T u\|_\infty \leq \lambda. \end{cases}$

即.要若得到对偶问题最优解 u^*, 则

它与原问题最优解 β^* 应满足 $X\beta^*=y-u^*$

凸优化笔记：交替方向乘子法. Alternating Direction Method of Multipliers.

解决存在2个优化变量的，只含等式约束的优化类问题.

一般形式：

$$\min_{x,z} f(x)+g(z)$$
$$\text{s.t. } Ax+Bz=c.$$

其中 $x\in\mathbb{R}^n$, $z\in\mathbb{R}^m$ 为优化变量.

$A\in\mathbb{R}^{p\times n}$, $B\in\mathbb{R}^{p\times m}$, $c\in\mathbb{R}^p$

√ 为解决这类问题，定义增广拉格朗日 Augmented Lagrangian.

$$L_\rho(x,z,u)=f(x)+g(z)+u^T(Ax+Bz-c)+\frac{\rho}{2}\|Ax+Bx-c\|_2^2$$

L_ρ 在原 Lagrange 中增加了和约束条件有关的惩罚项，其中 $\rho>0$.

√ 算法流程：每次固定二个变量，只更新一个变量，如此往复.

step 1: $x^{(k)}=\arg\min_x L_\rho(x, z^{(k-1)}, u^{(k-1)})$.

step 2: $z^{(k)}=\arg\min_z L_\rho(x^{(k)}, z, u^{(k-1)})$

step 3: $\cancel{u^{(k)}=\arg\min_u L_\rho(x^{(k)}, z^{(k)}, u)}$

$u^{(k)}=u^{(k-1)}+\rho(Ax^{(k)}+Bz^{(k)}-c)$

✓ 为简化形式，令 $w=\frac{u}{\rho}$，则 $L_\rho(x,z,u)$ 可写成

$$L_\rho(x,z,w)=f(x)+g(z)+\frac{\rho}{2}\|Ax+Bz-c+w\|_2^2-\frac{\rho}{2}\|w\|_2^2$$

相应地，更新步骤变为

Step 1：$x^{(k)}=\arg\min_x f(x)+\frac{\rho}{2}\|Ax+Bz^{(k-1)}-c+w^{(k-1)}\|_2^2$

Step 2：$z^{(k)}=\arg\min_z g(z)+\frac{\rho}{2}\|Ax^{(k)}+Bz-c+w^{(k-1)}\|_2^2$

Step 3：$w^{(k)}=w^{(k-1)}+Ax^{(k)}+Bz^{(k)}-c$.

（注意到，$w^{(k)}=w^{(0)}+\sum_{i=1}^{k}(Ax^{(i)}+Bz^{(i)}-c)$ ）.

初值 + 约束条件的累计残差和

✓ ADMM 解 lasso.

i. lasso 问题. $\min_\beta \frac{1}{2}\|y-X\beta\|_2^2+\lambda\|\beta\|_1$

ii. 重写为 $\min_{\beta,\alpha} \frac{1}{2}\|y-X\beta\|_2^2+\lambda\|\alpha\|_1$

s.t. $\beta-\alpha=0$

iii. 构造增广 Lagrangian.

$$L_\rho(\beta,\alpha,u)=\frac{1}{2}\|y-X\beta\|_2^2+\lambda\|\alpha\|_1+u^T(\beta-\alpha)+\frac{\rho}{2}\|\beta-\alpha\|_2^2$$

iv. 首先更新β值.

$$\frac{\partial L}{\partial \beta} = -X^T(y - X\beta) + u + \rho(\beta - \alpha) = 0$$

$$\Rightarrow \beta = (X^TX + \rho I)^{-1}(X^Ty + \rho\alpha - u)$$

然后更新α.

$$\arg\min_{\alpha} L_\rho(\beta, \alpha, u) = \arg\min_{\alpha}\left\{\lambda\|\alpha\|_1 + u^T(\beta - \alpha) + \frac{\rho}{2}\|\beta - \alpha\|_2^2\right\}$$

$$= \arg\min_{\alpha}\left\{\lambda\|\alpha\|_1 + u^T(\beta - \alpha) + \frac{\rho}{2}\|\beta - \alpha\|_2^2\right\} \cdot \frac{2}{\rho}$$

$$= \arg\min_{\alpha}\left\{\frac{2\lambda}{\rho}\|\alpha\|_1 + \frac{2}{\rho}u^T(\beta - \alpha) + \|\beta - \alpha\|_2^2\right\}$$

$$= \arg\min_{\alpha}\left\{\frac{2\lambda}{\rho}\|\alpha\|_1 + \left\|\frac{u}{\rho} + \beta - \alpha\right\|_2^2 - \left\|\frac{u}{\rho}\right\|_2^2\right\}$$

$$= \arg\min_{\alpha}\left\{\frac{2\lambda}{\rho}\|\alpha\|_1 + \left\|\alpha - \left(\frac{u}{\rho} + \beta\right)\right\|_2^2\right\}$$

$$= S_{\frac{\lambda}{\rho}}\left(\frac{u}{\rho} + \beta\right)$$

$$= S_{\frac{u}{\rho}}(w + \beta)$$

∴ Lasso 更新方法为:

$$\beta^{(K)} = (X^TX + \rho I)^{-1}(X^Ty + \rho \cdot \alpha^{(K-1)} - w^{(K-1)})$$

$$\alpha^{(K)} = S_{\frac{\lambda}{\rho}}(w^{(K-1)} + \beta^{(K)})$$

$$w^{(K)} = w^{(K-1)} + \beta^{(K)} - \alpha^{(K)}$$