# 一些算法收敛性的证明

Zhenwei Lin

日期：2020 年 8 月 8 日

**摘　要**

本文主要为几种算法的收敛性介绍及证明，包括的算法介绍主要是：

**关键词：** constraints;GLM;$\beta$-smooth function;$\alpha$-strong convex function;Projected GD;proximal gradient descent; subgradient;$G_\gamma$ function; Acclerated GD; ADMM; Duality;

## 1　常见的约束条件

1. $\Omega_1$ =nonnegative
2. $\Omega_2$ =Box
3. $\Omega_3 = \{l_2 - ball\}$
4. $\Omega_4$ =Affine 以上适合用 projceted GD, 下面的适合用 proximal GD
5. $\Omega_5 = \left\{x|x \geq 0, 1^T x = 1\right\}$
6. $\Omega_6 = \left\{x|\, \|x\|^2\right\} \leq s$
7. $\Omega_7 = \|x\|_1 \leq s$
8. $\Omega_8 = \left\{x|\, \|x\|_q \leq 1, 0 < q < 1\right\}$

## 2　GLM

首先先给出 GLM 的模型推导，因为这个模型后面经常用。

Dataset is$\{(a_i, b_i)\}^N, a_i \in R^p, b_i \in R, f : R^p \to R$,logistic regression:$min_x\{\sum_{i=1}^N log(1+exp(a_i^T x))-b^T Ax\}$

上面这个模型是怎么样得到的，因为我们有假设

$$log(\frac{p_i}{1-p_i}) = a_i^T x \Rightarrow p_i = \frac{1}{1+exp(-a_i^T)} \Rightarrow 1-p_i = \frac{1}{1+exp(a_i^T x)}$$

logistic 模型本质是一个似然估计思想，将其看成一个 Bernoulli distribution

$$f(bi|p_i) = p_i^{b_i}(1-p_i)^{1-b_i} \Rightarrow log(f) = -log(1+exp(a_i^T x)) + b_i a_i^T x$$

$$\Rightarrow \Pi_{i=1}^N log(b_i|p_i) = \Sigma_{i=1}^N(-log(1+exp(a_i^T x)) + b_i a_i^T x)$$

对于上式取极大，即得到取极小的定理

$$min_x\{\sum_{i=1}^N log(1+exp(a_i^T x)) - b^T Ax\} \tag{1}$$

# 3 General Gradient Descent(略)

# 4 GD for $\beta-$smooth function

**Definition 4.1.** a function f is called $\beta-$smooth function, if $\forall x, y \in dom(f), ||\nabla f(x) - \nabla f(y)||^2 \leq \beta||x-y||^2, \beta > 0$

1. ex1:$f(x) = b^T Ax \Rightarrow \nabla f(x) = A^T b, \forall x, y, ||\nabla f(x) - \nabla f(y)|| = 0$, 因此其 $\beta = 0$ 为 0-smooth function.
2. ex2:$f(x) = \frac{1}{2}||Ax||^2, \nabla f(x) = A^T Ax \Rightarrow ||A^T A(x-y)|| \leq \lambda_{max} A^T A \|x-y\|$, 因此其为一个 $\lambda_{max}(A^T A)-$smooth function(where$\lambda_{max}(A^T A)$ 表示是矩阵 $A^T A$ 的最大特征值).

**引理 4.1.** $\forall x, y \in dom(f)$ f is $\beta$-smooth function $f(y) \leq f(x) + (\nabla f(x), y-x) + \frac{\beta}{2}\|x-y\|^2$

**证明.**

$$g(t) = f(x+t(y-x)), g(0) = f(x), g(1) = f(y)$$

$$\therefore f(y) - f(x) = g(1) - g(0) = \int_0^1 g'(t)dt = \int_0^1 (\nabla f(x+t(y-x)), (y-x))dt$$

$$f(y) - f(x) - (\nabla f(x), y-x) = \int_0^1 (\nabla f(x+t(y-x)) - \nabla f(x), y-x)dt$$

$$\stackrel{CS-inequality}{\leq} \int_0^1 \|\nabla f(x+t(y-x)) - \nabla f(x)\| \, \|y-x\| \, \text{det}$$

$$\leq \int_0^1 t\beta \|y-x\|^2 \, dt = \frac{\beta}{2}\|y-x\|^2$$

□

**Definition 4.2.** $M_t(x) = f(x_t) + (\nabla f(x_t), x-x_t) + \frac{\beta}{2}\|x_t - x\|^2$

1. from lemma, we know$f(x) \leq m_t(x)$
2. $\nabla m_t(x) = \beta(x_t - x) + \nabla f(x_t) \Rightarrow x_{t+1} = x_t - \frac{1}{\beta}\nabla f(x_t)$(根据凸函数的非负加权合仍为凸函数, 所以导数为 0 的地方，其达到值最小)
3. $\Rightarrow f(x_{t+1}) \leq m_t(x_{t+1}) \leq m_t(x_t) = f(x_t)$

   上面的证明回答了**为什么对于 $\beta$-smooth function$\frac{1}{\beta}$ 是一个较好的步长选择**

   因此对于 linear regression 来说

$$min\frac{1}{2}\sum_i^N (a_i^T x) - b^T Ax \Rightarrow \beta = \lambda_{max}(A^T A)$$

$$\Rightarrow x_{t+1} = x_t - 1/\lambda_{max}(A^T A)\nabla L(x_t)$$

**定理 4.2.** Suppose that f is a$\beta$- smooth function$x^{t+1} = x^t - \frac{1}{\beta}\nabla f(x^t)$then

$$min_t \|\nabla f(x_t)\| \leq \sqrt{\frac{2\beta f(x_0)}{T}} \xrightarrow{T} 0$$

**定理 4.3.** f is a convex and$\beta$-smooth function,$x_{t+1} = x_t - \frac{1}{\beta}\nabla f(x_t)$, then

$$f(x_T) - f(x^*) \le \frac{\beta}{2T}\left\|x^T - x^*\right\|^2$$

**证明.**

$$M_t(x) = f(x_t) + (\nabla f(x_t), x - x_t) + \frac{\beta}{2}\|x_t - x\|^2$$

$$\nabla m_t(x) = \nabla f(x_t) + \beta(x - x_t)$$

$$\nabla^2 m_t(x) = \beta I \ge \beta I$$

因此 $m_t(x)$ 为一个 $\beta$-strong convex function

$$\Rightarrow m_t(y) \ge m_t(x) + (\nabla m_t(x), y - x) + \frac{\beta}{2}\|x_t - x\|^2$$

denote$y = x^*, x = x^{t+1}$

$$m_t(x^*) \ge m_t(x_{t+1}) + (\nabla m_t(x_{t+1}), x^* - x_{t+1}) + \frac{\beta}{2}\|x^* - x_{t+1}\|^2$$

$$\ge m_t(x_{t+1}) + \frac{\beta}{2}\|x^* - x_{t+1}\|^2\,(x_t + 10)$$

由 $m_t(x)$ 的性质 (1) 可以知道

$$\begin{aligned} f(x_{t+1}) &\le m_t(x_{t+1}) \le (m_t(x^*) = m_t(x_t)) - \frac{\beta}{2}\|x^* - x_{t+1}\| \\ &\overset{defintion-m_t(x)}{\le} f(x_t) + (\nabla f(x_t), x^* - x) + \frac{\beta}{2}\{\|x^* - x_t\|^2 - \|x^* - x_{t+1}\|^2\} \\ &\overset{convex}{\le} f(x^*) + \frac{\beta}{2}\{\|x^* - x_t\|^2 - \|x^* - x_{t+1}\|^2\} \end{aligned}$$

$$\begin{aligned} \therefore f(x_{t+1}) - f^* &\le \frac{\beta}{2}\{\|x^* - x_t\|^2 - \|x^* - x_{t+1}\|^2\}) \\ \therefore f(x_t) - f^* &\le \frac{\beta}{2}\{\|x^* - x_{t-1}\|^2 - \|x^* - x_t\|^2\}) \\ &\vdots \\ \sum_{t=0}^{T} f(x_t) - Tf(x_0) &\le \frac{\beta}{2}\{\|x^* - x_{t-1}\|^2 - \|x^* - x_t\|^2\}) \end{aligned}$$

之后再两边同时除以 T 就得证了。 □

# 5 $\alpha$-strong convex function

**Definition 5.1.** $\alpha$-strong convex function iff

$$\forall x, y \in dom(f) \qquad f(y) \ge f(x) + (\nabla f(x), y - x) + \frac{\alpha}{2}\|y - x\|^2$$

**引理 5.1.** f(x) is$\alpha$-strong convex function $\Longleftrightarrow f(x) - \frac{\alpha}{2}\|x\|^2$is convex

**定理 5.2.** $f \in C^1$[continuous and differentiable],then the following arguments are equalient:

1. f is$\alpha$- strong convex
2. $(\nabla f(y) - \nabla f(x), y - x) \geq \alpha \|x - y\|^2$
3. $f \in C^2, \nabla^2 f(x) \geq \alpha I$

**定理 5.3.** Assume that f is$\alpha$- strong convex and$\beta$-smooth function

$$x_{t+1} = x_t - \frac{1}{\beta}\nabla f(x_t)$$

then $\|x_T - x^*\| \leq exp(-\frac{T}{\tau}) \left\|x^0 - x^*\right\|^2$ where $\tau = \frac{\lambda_{max}(A^TA)}{\lambda_{min}(A^TA)}$

**证明.**

$$\begin{aligned} \|x_{t+1} - x^*\|^2 &= \left\|x_t - \frac{1}{\beta}\nabla f(x_t) - x^*\right\|^2 \\ &= \|x_t - x^*\|^2 + \frac{1}{\beta^2}\|\nabla f(x_t)\|^2 - \frac{2}{\beta}(\nabla f(x_t), x_t - x^*) \end{aligned} \tag{2}$$

$$f(x^*) \geq f(x_t) + (\nabla f(x_t), x^* - x_t) + \frac{\alpha}{2}\|x^* - x_t\|^2$$

$$\therefore (\nabla f(x_t), x^* - x_t) \leq f(x^*) - f(x_t) - \frac{\alpha}{2}\|x^* - x_t\|^2$$

式 **??** $\leq \|x_t - x^*\|^2 + \frac{1}{\beta^2}\|\nabla f(x_t)\|^2 + \frac{2}{\beta}(f(x^*) - f(x_t) - \frac{\alpha}{2}\|x^* - x_t\|^2)$

$$\leq (1 - \frac{\alpha}{\beta})\|x_t - x^*\|^2 - \frac{2}{\beta}(f(x_t) - \frac{1}{2\beta}\|\nabla f(x_t)\|^2 - f^*)$$

denote

$$\begin{aligned} m_t(x) &\overset{\Delta}{=} f(x_t) + (\nabla f(x_t), x - x_t) + \frac{\beta}{2}\|x - x_t\|^2 \\ x_{t+1} &= argmin_x(m_t(x)) = -\frac{1}{\beta}\nabla f(x_t) + x_t \end{aligned} \tag{3}$$

$$\therefore m_t(x_{t+1}) = f(x_t) + (\nabla f(x_t), -\frac{1}{\beta}\nabla f(x_t)) + \frac{\beta}{2}\left\|-\frac{1}{\beta}\nabla f(x_t)\right\|^2 = f(x_t) - \frac{1}{2\beta}\|\nabla f(x_t)\|^2$$

we know

$$f^* \leq f(x_{t+1}) \leq m_t(x_{t+1}) \Rightarrow (f(x_t) - \frac{1}{2\beta}\|\nabla f(x_t)\|^2 - f^*) > 0$$

$$\|x_{t+1} - x^*\|^2 \overset{\text{由最开始的式子}}{\leq} (1 - \frac{\alpha}{\beta})\|x_t - x^*\|^2$$

递推得到

$$\|x_T - x^*\|^2 \leq (1 - \frac{\alpha}{\beta})^{t+1}\|x_0 - x^*\|^2$$

$$\overset{by (1-x)^t \leq exp(-xt) (0<x<1)}{\Rightarrow} done$$

□

# 6 Projected GD

some constraints

1. non-convex$\Omega = \{x| \|x\|_q \leq R\}(0 < q < 1)$where$\|x\|_q = (\sum |x_i|^q)^{\frac{1}{q}}$
2. group sparse constraints:$\Omega = \{x| \|x\|^2 \leq R\}$

**定理 6.1.**

$$min \quad f(x) s.t. x \in \Omega$$

$$\text{step1: given } x_0 \text{ ,t = 0}$$
$$\text{step2: } y_{t+1} = x_t + \gamma_t h_t, h_t = -\nabla f(x_t)$$
$$x_t + 1 = \Pi_\Omega(y_t) \in \Omega$$

**Definition 6.1.** the projection of a point y on the set C is

$$\Pi_\Omega(y) = argmin_{x\in C} \|x - y\|$$

ex1:(nonnegative)

$$\Omega = \{x | x \geq 0\} \Pi_\Omega(y) = max(y, 0)$$

ex2:(l_2-ball)

$$\Omega = \{x | \ \|x\|^2 \leq 1\}$$

$$\prod_\Omega(y) = \begin{cases} y, \|y\|^2 \leq 1 \\ \frac{y}{\|y\|^2}, otherwise \end{cases}$$

ex3:(affine constrain)

$$\Omega = \{A_{n\times p} x_{p\times 1} = b_{n\times 1}\}$$

$$y^\perp = y - \Pi_\Omega(y) = y - Ax^*$$

$$(y^\perp, Ax) = 0(for \forall x) \Rightarrow (y - Ax^*)^T A = 0 \Rightarrow A^T A x^* = A^T y$$

if $A^T A$ is invertable $x^* = (A^T A)^{-1} A^T y$

example:

$$\min \frac{1}{2} \|Ax - b\|^2 s.t. x \geq 0$$
$$x_{t+1} = max(x_t - \frac{1}{\beta}(Ax_t - b), 0)$$
$$(where \beta = \lambda_{max}(A^T A))$$

**定理 6.2.** f is convex, differentiable, we denote an algorithms

$$\begin{cases} y_{t+1} = x_t - \gamma \nabla f(x_t) (\gamma = \frac{R}{L\sqrt{T}}) \\ x_{t+1} = \Pi_\Omega(y_{t+1}) \end{cases}$$

$\{x_t\}_{t=1}^T$if we suppose

$$\|\nabla f(x_t)\|^2 \leq L, \|x_1 - x^*\| = R$$

then

$$f(\frac{1}{T}\sum_{t=1}^T x_t) - f^* \leq \frac{RL}{\sqrt{T}}$$

**证明.** convex function so we have

$$f(x_t) - f^* \leq (\nabla f(x_t), x_t - x^*) = \frac{1}{\gamma}(x_t - y_{t+1}, x_t - x^*)$$

by$\|a\|^2 + \|b\|^2 = \|a - b\|^2 + 2(a, b)$

上式 =

$$\frac{1}{2\gamma}(\|x_t - y_{t+1}\|^2 + \|x_t - x^*\|^2 - \|x^* - y_{t+1}\|^2)$$
$$= \frac{1}{2\gamma}(\|x_t - x^*\|^2 - \|x^* - y_{t+1}\|^2) + \frac{\gamma}{2}\|\nabla f(x_t)\|^2$$

by$\|x^* - y_{t+1}\|^2 \geq \|x^* - x_{t+1}\|^2$(由于 $x^*$ 在 $\Omega$ 空间内故距离在减小) 故上式

$$\leq \frac{1}{2\gamma}(\|x_t - x^*\|^2 - \|x^* - x_{t+1}\|^2 + \frac{L^2\gamma}{2})$$
$$\therefore f(x_1) \leq \frac{1}{2\gamma}(\|x_1 - x^*\|^2 - \|x_2 - x^*\|^2 + \frac{L^2\gamma}{2})$$
$$\vdots$$
$$f(x_T) - f^* \leq \frac{1}{2\gamma}(\|x_T - x^*\|^2 + \|x_{t+1} - x^*\|^2 + \frac{L^2\gamma}{2})$$

累加即可得到

$$\sum_{t=1}^{T} f(x_t) - Tf^* \leq \frac{1}{2\gamma}(\|x_1 - x^*\|^2 - \|x_{T+1} - x^*\|^2) + \frac{L^2\gamma T}{2}$$
$$\Rightarrow f(\frac{1}{T}\sum_{t=1}^{T} x_t) - f^* \leq \frac{1}{T}\sum_{t=1}^{T} f(x_t) - f^* \leq \frac{R}{2\gamma T} + \frac{L^2\gamma}{2} \overset{C-S}{\leq} \dots$$

□

# 7 proximal gradient descent

在考虑 proximal gradient descent 前，需要先考虑如何将 constrainted optimization covert to **non-constrainted optimization** by using indicator function, we have:

$$min_{x\in R^d} f(x) + \delta_\Omega(x)$$

**indicator function**

$$\delta_\Omega(x) = \begin{cases} 0, & if \quad x \in \Omega \\ +\infty, & if \quad x \notin \Omega \end{cases}$$

所以在 constraint 里面在的 x 可以认为其就是在做原始的优化问题，不在 constraint 里面的 x，则必不可能是最优解。

ex : min $\frac{1}{2}\|Ax - b\|^2\ s.t.\ \|x\|_1 \leq S \iff min\left\{\frac{1}{2}\|Ax - b\|^2 + \lambda\|x\|_1\right\}$ 当 s 足够大（$\to \infty$）时，$\lambda \to 0$ 取小，反之 s 取大时，$\lambda$ 取大，故二者等价。

$$min_{x\in R^d} h(x) = f(x) + g(x)$$

where f is a convex-$\beta$-function, g is a convex function define

$$m_t(x) = (\nabla f(x_t), x - x_t) + \frac{\beta}{2}\|x - x_t\|^2 + g(x)$$
$$= f(x_t) - \frac{1}{2\beta}\|\nabla f(x_t)\|^2 + \frac{\beta}{2}\left\|x - (x_t - \frac{1}{\beta}\nabla f(x_t))\right\|^2 + g(x)$$
$$\therefore min_x m_t(x) \iff min_x\left\{\frac{\beta}{2}\left\|x - (x_t - \frac{1}{\beta}\nabla f(x_t))\right\|^2 + g(x)\right\}$$
$$\because h(x_{t+1}) \le m_t(x_{t+1}) \le m_t(x_t) = h(x_t)$$
$$x_{t+1} = argmin_x\left\{\frac{\beta}{2}\left\|x - (x_t - \frac{1}{\beta}\nabla f(x_t))\right\|^2 + g(x)\right\}$$

是有道理的，使函数值减小

记

$$x_{t+1} = prox_{\frac{1}{\gamma}g}(x_t - \frac{1}{\beta}\nabla f(x_t)) = argmin_x\left\{\frac{1}{2\gamma}\left\|x - (x_t - \frac{1}{\beta}\nabla f(x_t))\right\|^2 + g(x)\right\}$$
$$prox_{\gamma g} = argmin_x\left\{\frac{1}{2\gamma}\|x - z\|^2 + g(x)\right\}$$

ex1:

$$min_x\left\{\frac{1}{2\gamma}\|x - z\|^2 + \delta_\Omega(x)\right\} \iff min_{x\in\Omega}\left\{\frac{1}{2\gamma}\|x - z\|^2\right\}$$
$$\therefore prox_{\gamma\delta_{\Omega(x)}(z)} = argmin_{x\in\Omega}\left\{\frac{1}{2\gamma}\|x - z\|^2\right\} = \Pi_\Omega(z)$$

ex2:(lasso)

$$min\frac{1}{2}\|Ax - b\|^2 \quad s.t. \|x\|_1 \le s \iff min_{x\in R^d}\left\{\frac{1}{2}\|Ax - b\|^2 + \lambda\|x\|_1\right\} \iff min_x\sum_{j=1}^{P}(\frac{1}{2}(x_j - z_j)^2 + \lambda|x_i|)$$

等价于求解 p 个最小化问题

$$\Rightarrow f(x) = \frac{1}{2}\|Ax - b\|^2 \, g(x) = \lambda|x| \, \beta = \lambda_{\max}(A^T A)$$

# 8 subgradient

**Definition 8.1.** A subgradient of a convex possibble non-smooth function $h : R^T \to R$ at point x is vector $v \in R^p$ such that $h(y) \ge h(x) + (v, y - x), \forall y \in dom(h)$; for differentiable function, $f(y) \ge f(x) + (\nabla f(x), y - x)$

**Definition 8.2.** subdifferential of h at point x is a set that containted all the subgradient of h at point x i.e.denote as $\partial h(x)$ =v|v is the subgradient of h at point x

ex1:$h(x) = |x|$ it's $\partial h(x)|_0 = [-1, 1]$

ex2:$h(x) = max(x, 0)$ it's $\partial h(x)|_0 = [0, 1]$

**引理 8.1.** $0 \in \partial h(x) \iff$ x is a global minimum

**证明.** if $0 \in \partial h(x) \Rightarrow h(y) \ge h(x) + (0, y - x)$ □

关于如何求解这个一维优化问题

$$min_x \sum_{j=1}^{P}(\frac{1}{2}(x_j - z_j)^2 + \lambda |x_i|)$$

1. when $x_j > 0$,we have $\frac{\partial L(x_j)}{\partial x_j} = \beta(x_j - z_j) + \lambda = 0 \Rightarrow x_j = z_j - \frac{\lambda}{\beta}$
2. when $x_j < 0$,we have $\frac{\partial L(x_j)}{\partial x_j} = \beta(x_j - z_j) - \lambda = 0 \Rightarrow x_j = z_j + \frac{\lambda}{\beta}$
3. when $x_j = 0$ ,$0 \in \partial L(0) \iff 0 \in \beta(x_j - z_j) + \lambda\partial\left|x_j\right| \iff \beta(z_j - x_j) \in \lambda\partial\left|x_j\right|$，当 $x_j = 0$ 时，$z_j \in [-\frac{\lambda}{\beta}, \frac{\lambda}{\beta}]$

**Definition 8.3.**

$$x_j^* = \begin{cases} z_j - \frac{\lambda}{\beta} & , z_j > \frac{\lambda}{\beta} \\ 0 & , z_j \in \left[-\frac{\lambda}{\beta}, \frac{\lambda}{\beta}\right] \\ z_j + \frac{\lambda}{\beta} & , z_j < -\frac{\lambda}{\beta} \end{cases} \tag{4}$$

记 soft thresholding 为 $\quad x_j^* = sgn(z)(|z| - \frac{\lambda}{\beta})_+ = S_{\frac{\lambda}{\beta}}(z_j)$

$$sgn(z) = \begin{cases} 1 & , z > 0 \\ -1 & , z < 0 \\ 0 & , z = 0 \end{cases}$$

**因此要解决 lasso 问题采用的梯度下降法为** $x_j^* = sgn(z)(|z| - \frac{\lambda}{\beta})_+$ **其中** $z_j = x_t - \frac{1}{\lambda_{\max}(A^T A}A^T(Ax_t - b)$ **即为** $x^* = S_{\frac{\lambda}{\beta}}(x_t - \frac{1}{\lambda_{\max}(A^T A}A^T(Ax_t - b))$

# 9 $G_\gamma$ function

**Definition 9.1.** $G_\gamma(x) = \frac{1}{\gamma}(x - x^*)$, where $x^* = prox_{\gamma g}(z) = argmin\left\{\frac{1}{2\gamma}\|x - z\|^2 + g(x)\right\}$ where $z = x - \gamma\nabla f(x)$

若用迭代，则有 $x^* = prox_{\gamma g}(x - \gamma\nabla f(x)) \Rightarrow x^* = x - \gamma\nabla f(x) \Rightarrow \nabla f(x) = \frac{1}{\gamma}(x - x^*)$(当 g 为 0 不存在时)

**引理 9.1.** $G_\gamma(x) = 0 \iff 0 \in \partial h(x)$ furthermore we have that $G_\gamma(x) - \nabla f(x) \in \partial g(x)$

**证明.**

$$x^+ = argmin_y\left\{\frac{1}{2\gamma}\|y - (x - \gamma\nabla f(x))\|^2 + g(y)\right\}$$

when it's minimum, we have

$$0 \in \partial L(x^+) = \frac{1}{\gamma}(x^+ - (x - \gamma\nabla f(x))) + \partial g(x^+) \Rightarrow \frac{1}{\gamma}(x - x^+) \in \nabla f(x) + \partial g(x^+)$$

if

$$x = x^+, G_\gamma(x) = 0 \Rightarrow 0 \in \nabla f(x^+) + \partial g(x^+) = \partial L(x^+)$$

□

下面的一系列定理都非常重要 (主定理的延伸)

**定理 9.2.** $h(x) = f(x) + g(x)$,$f(x)$ is a $\alpha$-strong convex, $\beta$-smooth function.

$g(x)$ is a convex function,

$G_\gamma(x) = \frac{1}{\gamma}(x - x^+)$,

where $x^+ = argmin_y \left\{ \frac{1}{2\gamma} \|y - (x - \gamma \nabla f(x))\|^2 + g(x) \right\}$,

then $\forall y \in dom(h)$

$h(y) \geq h(x^+) + (G_\gamma, y - x) + \frac{\alpha}{2} \|y - x\|^2 + \gamma(1 - \frac{\gamma\beta}{2}) \left\| G_\gamma(x) \right\|^2$

**推论.** $\alpha$ could be zero,$h(y) \geq h(x^+) + (G_\gamma, y - x) + \gamma(1 - \frac{\gamma\beta}{2}) \left\| G_\gamma(x) \right\|^2$

**推论.** $\gamma = \frac{1}{\beta}, x = y \Rightarrow h(x^+) \leq h(x) - \frac{1}{2\beta} \left\| G_\gamma(x) \right\|^2$

recall $\beta$-smooth function $f(x_{t+1}) \leq f(x_t) - \frac{1}{2\beta} \|\nabla f(x_t)\|^2$

**推论.** $\gamma = \frac{1}{\beta}, G_\gamma = \beta(x - x^+), \alpha = 0$

$h(y) \geq h(x^+) + \beta(x - x^+, y - x) + \frac{\beta}{2} \|x - x^+\|^2$

x 与 y 互换则有

$h(x) \geq h(x^+) + \beta(y - x^+, x - y) + \frac{\beta}{2} \|y - x^+\|^2$

**证明.**

$$h(x^+) = f(x - \gamma G_\gamma(x)) = f(x - \gamma G_\gamma(x)) + g(x^+)$$

by $\beta$ - smooth convex function lemma1,

$$\leq f(x) - \gamma(\nabla f(x), G_\gamma(x)) + \frac{\beta\gamma^2}{2} \left\| G_\gamma \right\|^2 + g(x^+)$$

by $\alpha$ - strong convex function Definition

$$\begin{aligned} &\leq f(y) + (\nabla f(x), x - y) - \frac{\alpha}{2} \|x - y\|^2 - \gamma(\nabla f(x), G_\gamma) + \frac{\beta\gamma^2}{2} \left\| G_\gamma(x) \right\|^2 + g(x^+) \\ &= f(y) + (\nabla f(x), x^+ - y) - \frac{\alpha}{2} \|x - y\|^2 + \frac{\beta\gamma^2}{2} \left\| G_\gamma(x) \right\|^2 + g(x^+) \end{aligned} \tag{5}$$

by $G_\gamma(x) - \nabla f(x) \in g(x^+)$,we have

$$g(y) \geq g(x^+) + (G_\gamma(x) - \nabla f(x), y - x^+)$$

$$\begin{aligned} \therefore \ &\leq f(y) + (\nabla f(x), x^+ - y) - \frac{\alpha}{2} \|x - y\|^2 + \frac{\beta\gamma^2}{2} \left\| G_\gamma(x) \right\|^2 + g(y) + (G_\gamma(x) - \nabla f(x), x^+ - y) \\ &\overset{\text{合并同类项}}{\leq} h(y) + (G_\gamma, x^+ - y) - \frac{\alpha}{2} \|x - y\|^2 + \frac{\beta\gamma^2}{2} \left\| G_\gamma(x) \right\|^2 \\ &\overset{x^+ = x - \gamma G_\gamma(x)}{=} h(y) + (G_\gamma(x), x - y) - \frac{\alpha}{2} \|x - y\|^2 + \gamma(\frac{\beta\gamma}{2} - 1) \left\| G_\gamma(x) \right\|^2 \end{aligned} \tag{6}$$

□

**定理 9.3.** If h = f + g, f is $\beta$ -smooth ,convex g is convex >

$$h(x_T) - h(x^*) \leq \frac{\beta}{2T} \|x_1 - x^*\|^2$$

in addition, if f is $\alpha$ - strong convex funciton ,we have

$$\|x_T - x^*\|^2 \leq exp(-\frac{\alpha}{\beta}T) \|x_- x^*\|^2$$

**证明.**

$$h(x^*) \geq h(x_{t+1}) + (G_\gamma(x_t), x^* - x_t) + \frac{1}{2\beta}\left\|G_\gamma(x_t)\right\|^2$$

where $G_\gamma(x_t) = \beta(x_t - x_{t-1})$

$$\begin{aligned} h(x_{t+1}) - h^* &\leq \frac{\beta}{2}\left\{2(x_t - x_{t-1}, x^* - x_t) + \|x_t - x_{t-1}\|^2\right\} \\ &\leq \frac{\beta}{2}\left\{\|x^* - x_{t-1}\|^2 - \|x_t - x_{t-1}\|^2\right\} \\ &\vdots \end{aligned}$$

累加求平均

$$h(x_T) - h^* \leq \frac{\beta}{2T}\|x_1 - x^*\|^2$$

□

证明 2.

$$\|x_{t+1} - x^*\|^2 = \left\|x_t - \frac{1}{\beta}G_\gamma(x_t) - x^*\right\|^2 = \|x_t - x^*\|^2 + \frac{1}{\beta^2}\left\|G_\gamma(x_t)\right\|^2 - \frac{2}{\beta}(G_\gamma(x_t), x_t - x^*)$$

由推论第三条知:

$$\begin{aligned} 0 \geq h^* - h(x_{t+1}) &\geq (G_\gamma(x_t), x^* - x_t) + \frac{\alpha}{2}\|x^* - x_t\|^2 + \frac{1}{2\beta}\left\|G_\gamma(x)\right\|^2 \\ \Rightarrow (G_\gamma(x_t), x_t - x^*) &\geq \frac{\alpha}{2}\|x^* - x_t\|^2 + \frac{1}{2\beta}\left\|G_\gamma(x)\right\|^2 \end{aligned}$$

$$\begin{aligned} \|x_{t+1} - x^*\|^2 &\leq \|x_t - x^*\|^2 + \frac{1}{\beta^2}\left\|G_\gamma(x)\right\|^2 - \frac{2}{\beta}(\frac{1}{2\beta}\left\|G_\gamma(x)\right\|^2 + \frac{\alpha}{2}\|x_t - x^*\|^2) \\ &= (1 - \frac{\alpha}{\beta})\|x_t - x^*\|^2 \end{aligned}$$

$$\Rightarrow \|x_T - x^*\|^2 \leq (1 - \frac{\alpha}{\beta})^T\|x_1 - x^*\|^2 \overset{T\to\infty}{=} exp(1 - \frac{\alpha}{\beta}T)\|x_1 - x^*\|^2$$

□

# 10 Acclerated GD

**定理 10.1.** (optional rate for $\beta$- smooth function) let T $\leq \frac{P-1}{2}, \beta \geq 0$ ,there exists a $\beta$-smooth convex quadratic f such that

$$min_{1\leq t\leq T}(f(x_t) - f^*) \geq \frac{3\beta\|x_0 - x^*\|^2}{32(1+T)^2}$$

$min_x h(x)$ **AGD**: $x_0, y_1 = x_0, a_0 = 1$

1. step1:$x_t = y_t - \frac{1}{\beta}\nabla h(y_t)$
2. step2: $a_{t+1} = \frac{1+\sqrt{1+4a_t^2}}{2}$
3. step3: $y_{t+1} = x_t + \frac{a_t - 1}{a_{t+1}}(x_{t-1} - x_t)$

more general: **A proximal GD**

$min\{h(x) = f(x) + g(x)\}$ where f is a convex $\beta$ - smooth function and g is a convex function

1. step1: $x_t = Prox_{g/\beta}(y_t - \frac{1}{\beta}\nabla y_t)$ 若 prox 为 soft thresholding 称之为 FISTA
2. step2: $a_{t+1} = \frac{1+\sqrt{1+4a_t^2}}{2}$

3. step3: $y_{t+1} = x_t + \frac{a_t - 1}{a_{t+1}}(x_{t-1} - x_t)$

**引理 10.2.** $\{a_t\}$ 单调递增 $a_t > \frac{t+1}{2}$

**证明.** we know $a_{t+1}^2 - a_{t+1} = a_t^2$ and $a_0 = 1$ $a_0 > \frac{t+1}{2} = \frac{1}{2}$

suppose $a_t > \frac{t+1}{2}$ $a_{t+1}^2 - a_{t+1} - a_t^2 = 0 \Rightarrow a_{t+1}^2 - a_{t+1} > \frac{(1+t)^2}{4} \Rightarrow a_{t+1} \geq \frac{1+\sqrt{1+(1+t)^2}}{2} \geq \frac{1+1+t}{2} = \frac{t+2}{2}$ □

**定理 10.3.** let $\{(x_t, y_t)\}_{t=1}^T$ be generated APGD, then for any $T \geq 1$

$$h(x_T) - h^* \leq 2\beta \|x_0 - x^*\|^2 / (1+T)^2$$

ex:

$$min \sum_{i=1}^N log(1 + exp(a_i^T x)) - b^T Ax$$
$$s.t. \|x\|_1 \leq s$$
$$\Longleftrightarrow \ min \left\{ \sum_{i=1}^N log(1 + exp(a_i^T x)) - b^T Ax + \lambda \|x\|_1 \right\}$$

**引理 10.4.** 1. $h(x) \geq h(x^+) + \beta(y - x^+, x - y) + \frac{\beta}{2}\|y - x^+\|^2$ ,$G_\gamma$function theorem 1 推论三

2. $a_{t+1}^2 - a_{t+1} = a_t^2$

3. $a_t \geq \frac{t+1}{2}$

**引理 10.5.** $\{(x_t, y_t)\}$generated by AGD with constant step size $\frac{1}{\beta}$, then for every $t \geq 1$,we have

$$a_t^2 v_t - a_{t+1}^2 v_{t+1} \geq \frac{\beta}{2}(\|u_{t+1}\|^2 - \|u_t\|^2)$$

where $v_t = h(x_t) - h^*$ and $u_t = a_t x_t - (a_t - 1)x_{t-1} - x^*$

**证明.** in (1) let $x = x_t, y = y_{t+1} \Rightarrow x^+ = x_{t+1}$

$$\begin{aligned} h(x_t) &\geq h(x_{t+1}) + \beta(y_{t+1} - x_{t+1}, x_t - y_{t+1}) + \frac{\beta}{2}\|y_{t+1} - x_{t+1}\|^2 \\ h(x_t) - h^* &\geq h(x_{t+1}) - h^* + \beta(y_{t+1} - x_{t+1}, x_t - y_{t+1}) + \frac{\beta}{2}\|y_{t+1} - x_{t+1}\|^2 \\ \Rightarrow v_t &\geq v_{t+1} + \beta(y_{t+1} - x_{t+1}, x_t - y_{t+1}) + \frac{\beta}{2}\|y_{t+1} - x_{t+1}\|^2 \\ \Rightarrow \frac{2}{\beta}(v_t - v_{t+1}) &\geq 2(y_{t+1} - x_{t+1}, x_t - y_{t+1}) + \|y_{t+1} - x_{t+1}\|^2 \end{aligned} \tag{2}$$

let $x = x^*, y = y_{t+1} \Rightarrow x^+ = x_{t+1}$

$$\begin{aligned} h^* &\geq h(x_{t+1}) + \beta(y_{t+1} - x_{t+1}, x^* - y_{t+1}) + \frac{\beta}{2}\|y_{t+1} - x_{t+1}\|^2 \\ h^* - h(x_{t+1}) &\geq \beta(y_{t+1} - x_{t+1}, x^* - y_{t+1}) + \frac{\beta}{2}\|y_{t+1} - x_{t+1}\|^2 \\ \Rightarrow -\frac{2}{\beta}v_{t+1} &\geq 2(y_{t+1} - x_{t+1}, x^* - y_{t+1}) + \|x_{t+1} - y_{t+1}\|^2 \end{aligned} \tag{3}$$

$$(2) \times (a_t - 1) + (3) \Rightarrow (4)$$

$$\frac{2}{\beta}\{(a_{t+1} - 1)v_t - a_{t+1}v_{t+1}\} \geq a_{t+1}\|x_{t+1} - y_{t+1}\|^2 + 2(x_{t+1} - y_{t+1}, a_t y_{t+1} - (a_{t+1} - 1)x_t - x^*) \tag{4}$$

$$(4) \times a_{t+1}$$

$$\frac{2}{\beta}[\underbrace{(a_{t+1}^2 - a_{t+1})}_{a_t^2} v_t - a_{t+1}^2 v_{t+1}] \geq \ldots \tag{7}$$

□

**引理 10.6.** let $\{c_t, b_t\}$ be positive real numbers, that satisfy

$$c_t - c_{t+1} \geq b_{t+1} - b_t, \forall t \geq 1$$

and

$$c_1 + b_1 \leq c(where \quad c \quad is \quad constant)$$

then

$$c_t \leq c \Leftarrow \{c_t + b_t \leq c\}$$

数学归纳法自己证

recall theorem1 is

**定理 10.7.** let $\{(x_t, y_t)\}_{t=1}^T$ be generated APGD, then for any $T \geq 1$

$$h(x_T) - h^* \leq 2\beta \|x_0 - x^*\|^2 / (1+T)^2$$

**证明.** let $c_t = a_t^2 v_t, \quad b_t = \frac{\beta}{2}\|u_t\|^2 \Rightarrow c_t - c_{t+1} \geq b_{t+1} - b_t$

$$\Rightarrow c_t \leq c, \quad c_1 + b_1 \leq cc_1 = a_1^2 v_1 = h(x_1) - h^* b_1 = \frac{\beta}{2}\|u_1\|^2 = \frac{\beta}{2}\|x_1 - x^*\|^2$$

let $x = x^*, y = y_1 \Rightarrow x^+ = x_1$

$$\begin{aligned} h^* &\geq h(x_1) + \beta(y_1 - x_1, x^* - y_1) + \frac{\beta}{2}\|y_1 - x_1\|^2 - c_1 \\ &\geq \frac{\beta}{2}[2(y_1 - x_1, x^* - y_1) + \|x_1 - y_1\|^2] \\ \overset{by 2(a,b) = \|a+b\|^2 - \|a\|^2 - \|b\|^2}{=} -c_1 &\geq \frac{\beta}{2}(\|x_1 - x^*\|^2 - \|x^* - y_1\|^2) \\ &= b_1 - \frac{\beta}{2}\|x^* - y_1\|^2 \\ b_1 + c_1 &\leq \frac{\beta}{2}\|x^* - y_1\|^2 = \frac{\beta}{2}\|x^* - x_0\|^2 \end{aligned} \tag{8}$$

$$\begin{aligned} \because c_t \leq c \therefore a_t^2 v_t \leq \frac{\beta}{2}\|x^* - x_0\|^2 &\Rightarrow v_t \leq \frac{\beta\|x^* - x_0\|^2}{2a_t^2} \\ by \quad \overset{a_t \geq \frac{t+1}{2}}{\Rightarrow} \quad &h(x_T) - h^* \leq \ldots \end{aligned} \tag{9}$$

□

## 11 Dual Gradient Method

facing problems like

$$\begin{aligned}
min \quad & f(x) \\
s.t. \quad & c(x) \leq 0 \\
L(x, \lambda) =& f(x) + \lambda^T c(x) \\
g(\lambda) =& \inf_x L(x, \lambda) = L(x^*(\lambda), \lambda) \\
\Rightarrow \underset{x}{sup} \quad & g(\lambda) \\
s.t. \quad & \lambda \geq 0 \\
\frac{\partial L(x^*(\lambda), \lambda)}{\partial \lambda} =& \frac{\partial L}{\partial x^*}\frac{\partial x^*}{\partial \lambda} + \frac{\partial L}{\partial \lambda} = 0 + \frac{\partial L}{\partial \lambda}
\end{aligned}$$

projeted gradient ascent

$$\lambda_{t+1} = max(\lambda_t + \gamma \nabla g(\lambda_t), 0)$$

所以最终的迭代格式为

1. $\lambda_0, t = 0$ (initialize)
2. $x^*(\lambda) = \underset{x}{argmin} L(x, \lambda_t)$
3. $x_{t+1} = max(\lambda_t + \gamma \frac{\partial L}{\partial \lambda}, 0)$
4. $t = t + 1$

## 12 Decomposite Dual Gradient Ascent

$$\begin{aligned}
\underset{x}{min} \quad & f(x) \\
s.t. \quad & Ax = b
\end{aligned}
\Rightarrow
\begin{aligned}
L(x, v) \quad & = f(x) + v^T(Ax - b) \\
g(v) \quad & = \underset{x}{inf} L(x, v) = L(x^*(\lambda), v)
\end{aligned}$$

$$max \quad g(v) = \left[ \begin{aligned} & step1 : x^*(v_t) = argmin L(x, v_t) \\ & step2 : v_{t+1} = v_t + \gamma(Ax^*(v_t) - b) \end{aligned} \right.$$

如果目标函数可以进行分解则有：Decomposite:$f(x) = f(x_1) + ... + f(x_p) = \sum_{j=1}^{p} f(x_j)$

$$\begin{aligned}
L(x, v) &= \sum_{j=1}^{p} f(x_j) + v^T(Ax - b) \\
&= \sum_{j=1}^{p} f(x_j) + (A^T v)^T x - v^T b \\
&= \sum_{j=1}^{p} \left\{ f(x_j) + a_j^T v x_j \right\} - v^T b
\end{aligned}$$

1. $x^*(v_t) = argmin_{x_j} \left\{ f(x_j) + a_j^T v_t x_j \right\}$

2. $v_{t+1} = v_t + \gamma(Ax^*(v_t) - b)$

因此可以进行并行计算。

## 13 Dual proximal Gradient Ascent

$$\begin{array}{ll} min & f(x) \\ s.t. & Ax = b \end{array} \Rightarrow \max_v g(v) = \inf_x \left\{ f(x) + v^T(Ax - b) \right\}$$

$$\max_v \quad h(v) + g(v)$$

where $h(v) \equiv 0$ and $g(x)$ is concave.So

$$v_{t+1} = prox_{\gamma g}(v_t) = argmax_v \left\{ g(v) - \frac{1}{2\gamma} \|v - v_t\|^2 \right\}$$

$$d^* = \sup_v \left\{ \inf_x f(x) + v^T(Ax - b) - \frac{1}{2\gamma} \|v - v_t\|^2 \right\} \le \inf_x \left\{ \sup_v f(x) + v^T(Ax - b) - \frac{1}{2\gamma} \|v - v_t\|^2 \right\} = p^*$$

when $\frac{1}{\gamma}(v - v_t) = Ax - b$, we have

$$\begin{aligned} p^* &= \inf_x \left\{ f(x) + v_t^T(Ax - b) + \gamma \|Ax - b\|^2 - \frac{\gamma}{2} \|Ax - b\|^2 \right\} \\ &= \inf_x \left\{ f(x) + v_t^T(Ax - b) + \frac{\gamma}{2} \|Ax - b\|^2 \right\} \end{aligned}$$

**Definition 13.1.** Augmented Lagrangian Function

$$L(x, v) = f(x) + v^T(Ax - b) + \frac{\rho}{2} \|Ax - b\|^2$$

1. $x^*(v_t) = \inf_x L(x, v_t)$
2. $v_{t+1} = v_t + \gamma(Ax^*(v_t) - b)$

## 14 ADMM(Alternateive Direction Method of Multiplies)

$$\begin{aligned} min &\quad f(x) + g(z) \\ s.t. &\quad Ax + Bz = C \\ augmented &\quad L(x, z, y) = f(x) + g(z) + y^T(Ax + Bz - C) + \frac{\rho}{2} \|Ax + Bz - C\|^2 \end{aligned}$$

1. $x_{t+1} = \inf_x L(x, g_t, z_t)$
2. $z_{t+1} = \inf_z L(x_{t+1}, y_t, z)$
3. $y_{t+1} = y_t + \rho(Ax_{t+1} + Bz_{t+1} - C)$

证明其收敛性较为复杂，这里直接给出结论

**定理 14.1.** f,g have a closed, non-empty convex epigraph and L(augment) has a saddle point $x^*, y^*, z^*, i.e.$

$$\forall x, y, z, \quad L(x^*, z^*, y) \le L(x^*, y^*, z^*) \le L(x, z, y^*)$$

then, let $t \to \infty$

$$f(x_t) + g(z_t) \to p^*(stationarypoint)$$

$$y_t^* \to y^*$$

ex: 用 ADMM 算法来解 lasso，

$$\begin{aligned}
&\min_x \left\{ \frac{1}{2} \|Ax-b\|^2 + \lambda \|x\|_1 \right\} \\
\Rightarrow &\min_{x,z} \left\{ \frac{1}{2} \|Ax-b\|^2 + \lambda \|z\|_1 \right\} \\
&s.t. \quad x - z = 0
\end{aligned} \tag{10}$$

augmented lagrangian function:

$$\begin{aligned}
L(x,y,z) &= \frac{1}{2} \|Ax-b\|^2 + y^T(x-z) + \frac{\rho}{2} \|x-z\|^2 + \lambda \|z\|_1 \\
\frac{\partial L}{\partial x} &= A^T(Ax-b) + y + \rho(x-z) = 0 \\
\Rightarrow &(A^TA+\rho I)x - A^Tb + y - \rho z = 0 \\
\Rightarrow &x = (A^TA+\rho I)^{-1}(A^Tb + \rho z - y)
\end{aligned} \tag{11}$$

1. $x_{t+1} = (A^TA+\rho I)^{-1}(A^Tb+\rho z_t - y_t)$
2. $\because \min_z \left\{ \frac{\rho}{2} \|z-x\|^2 + \lambda \|z\|_1 + y^T(x-z) \right\} \iff \min_z \left\{ \frac{\rho}{2\lambda} \left\| z - x - \frac{y}{\rho} \right\| + \|z\|_1 \right\} \Rightarrow z = Soft_{\rho/\lambda}(x - y/\rho) \Rightarrow z_{t+1} = Soft_{\rho/\lambda}(x_{t+1} - y_{t/\rho})$
3. $y_{t+1} = y_t + \rho(x_{t+1} - z_{t+1})$

# 15 Template Method

**Definition 15.1.** Conjugate:$f : R^n \to R$ and $R^n \to R$

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

about indicator function's Conjugate function.

firstly, primal funciton is

$$\delta_c(x) = \begin{cases} 0 & x \in C \\ \infty & x \notin C \end{cases} \tag{12}$$

so the conjugate function of $\delta_c(x)$ is

$$\sigma_c(z) = \sup_{x \in c} (z, x)$$

so when $c = R_+^d$

$$\sigma_{R_+^d}(z) = \sup_{x \in R_+^d} (z, x) = \begin{cases} 0 & z \in R_-^d \\ \infty & z \notin R_-^d \end{cases} = \delta_{R_-^d}(z)$$

因为如果有一个位置上的数不为 0, 则将其对应的数置为无穷大，则最终其为无穷。

ex:

$$\begin{aligned} \min \quad & \frac{1}{2}\|Ax-b\|^2 \\ s.t. \quad & c_{n\times p}x \le b_{n\times 1} \end{aligned}$$

$$L(x,\lambda) = \frac{1}{2}\|Ax-b\|^2 + \lambda^T(cx-b) \quad \lambda \ge 0$$

$$L(x) = \frac{1}{2}\|Ax-b\|^2 + \delta_{R_-^d}(cx-b)$$

by Conjugate transformation $f^*(z) = \sup_x \left\{ z^T x - f(x) \right\}$

$$\begin{aligned} &\Rightarrow \frac{1}{2}\|Ax-b\|^2 + \sup_{\lambda\in R^n} \left\{ (\lambda, cx-b) - \delta_{R_+^d}(\lambda) \right\} \\ &= \sup_{\lambda} \left\{ \frac{1}{2}\|Ax-b\|^2 + \lambda^T(cx-d) - \delta_{R_+^d}(\lambda) \right\} = \sup_{\lambda\ge 0} L(x,\lambda) \end{aligned} \tag{13}$$

现在我们开始考虑下面的问题：

$$\min_x \underbrace{(c,x)}_{\text{linear}} + \underbrace{h(Ax-b)}_{\text{ex:}\frac{1}{2}\|Ax-b\|^2} + \underbrace{k(x)}_{\text{nonlinear}}$$

$$\begin{aligned} &\min_x \quad (c,x) + h(Ax-b) + k(x) \\ p^* &= \min_x \left\{ (c,x) + \sup_y \left\{ (y, Ax-b) - h^*(y) \right\} + \sup_z \left\{ (z,x) - k^*(z) \right\} \right\} \\ &= \min_x \sup_{y,z} \left\{ (c + A^T y + z, x) - h^*(y) - k^*(z) \right\} \\ &\ge \sup_{y,z} \min_z \left\{ (c + A^T y + z, x) - h^*(y) - k^*(z) \right\} = d^* \end{aligned} \tag{14}$$

$$q^* = \begin{array}{l} \sup_{y,z} \{-h^*(y) - k^*(z)\} \\ s.t. \quad C + A^T y + z = 0 \end{array} \iff \begin{array}{l} \min_{y,z} \{h^*(y) + k^*(z)\} \\ s.t. \quad C + A^T y + z = 0 \end{array} \iff \min_y h^*(y) + k^*(-c - A^T y)$$

ex:

$$\frac{1}{2}\|Ax-b\|^2 + \lambda\|x\|_1 \tag{15}$$

则有 $(c,x) = 0, h(Ax-b) = \frac{1}{2}\|Ax-b\|^2, k(x) = \lambda\|x\|_1$ 故接下来求 $h^*(y)$ 和 $k^*(z)$

Conjugate function $h^*(y) = \frac{1}{2}\|y\|$ and $k^*(y) = \delta_{B\|.\|}(z)$