机器学习中的优化算法

Lecture08: 复合优化算法–近似点算法

张立平

清华大学数学科学系

办公室：理科楼#A302, Tel: 62798531

E-mail: lipingzhang@tsinghua.edu.cn

Contents and Acknowledgement

- 教材：**最优化：建模、算法与理论**

  http://bicmr.pku.edu.cn/ wenzw/bigdata2021.html

- 致谢：北京大学文再文教授

Outline of Lecture08

- 近似点算法
- 与ALM的关系
- 应用: LASSO问题
- 收敛性分析
- Moreau-Yosida正则化
- 对偶近似点梯度与交替极小化方法

近似点算法

▶ 考虑优化问题

$$\min_{x} \quad \psi(x) \tag{1}$$

其中ψ是一个适当的闭凸函数,并不要求连续或可微.

- 次梯度法, 收敛较慢, 且收敛条件苛刻: $x^{k+1} = x^k - t_k \partial\psi(x^{k+1})$.
- 近似点算法:

$$\begin{aligned} x^{k+1} &= \mathrm{prox}_{t_k\psi}\left(x^k\right) \\ &= \arg\min_{u}\left\{\psi(u) + \frac{1}{2t_k}\left\|u - x^k\right\|_2^2\right\}, \end{aligned} \tag{2}$$

其中t_k为第k步迭代时的步长, 可为固定值或通过某种合适的线搜索策略得到. 近似点算法(2)可看作近似点梯度法的特殊情况, $\psi(x)$的邻近算子一般需要通过迭代求解, 但(2)的目标函数强凸, 相比原问题更利于迭代法的求解.

加速近似点算法

▶针对近似点算法(2)的FISTA加速:

$$x^k = \mathrm{prox}_{t_k\psi}\left(x^{k-1} + \gamma_k \frac{1-\gamma_{k-1}}{\gamma_{k-1}}\left(x^{k-1} - x^{k-2}\right)\right) \tag{3}$$

▶针对近似点算法(2)的第二类Nesterov加速:

$$v^k = \mathrm{prox}_{(t_k/\gamma_k)\psi}\left(v^{k-1}\right), \quad x^k = (1-\gamma_k)\, x^{k-1} + \gamma_k v^k. \tag{4}$$

▶关于算法参数选择的两种策略:

- 固定步长$t_k = t$ 以及$\gamma_k = \dfrac{2}{k+1}$;
- 可变步长t_k, 当$k = 1$ 时取$\gamma_1 = 1$; 当$k > 1$ 时, γ_k 满足:

$$\frac{(1-\gamma_k)\, t_k}{\gamma_k^2} = \frac{t_{k-1}}{\gamma_{k-1}^2}.$$

近似点算法与增广拉格朗日函数法的关系

▶考虑具有如下形式的优化问题:

$$\min_{x\in\mathbb{R}^n} \quad f(x)+h(Ax) \tag{5}$$

其中 f,h 为适当的闭凸函数, $A\in\mathbb{R}^{m\times n}$.

- 当 h 是单点集 $\{b\}$ 的示性函数时, (5) $\iff$ $\min\ f(x)$, s.t. $Ax=b$.
- 当h 是凸集C 上的示性函数时, (5) $\iff$ $\min\ f(x)$, s.t. $Ax\in C$.
- 当$h(y)=\|y-b\|$ 时, (5) $\iff$ $\min\ f(x)+\|Ax-b\|$.

▶ 优化问题(5)的拉格朗日对偶问题为:

$$\max \quad \psi(z)=\inf_{x,y}\{f(x)+h(y)+z^T(Ax-y)\}=-f^*(-A^Tz)-h^*(z) \tag{6}$$

▶ 对对偶问题(6)用近似点算法$\iff$对原问题(5)用增广拉格朗日函数法

Lemma 1. *设 $f(x)$ 是适当的闭凸函数，$f^*(y)$ 是其共轭函数，则对任意的 $y\in\operatorname{dom} f^*$ 和 $x\in\operatorname{dom} f$ 有*

$$y\in\partial f(x)\Leftrightarrow x\in\partial f^*(y). \tag{7}$$

▶ **对对偶问题(6)用近似点算法**$\Longleftrightarrow$对原问题(5)用增广拉格朗日函数法

对原问题(5), 增广拉格朗日函数法迭代格式如下:

$$(x^{k+1}, y^{k+1}) = \arg\min_{x,y} \left\{ f(x) + h(y) + \frac{t_k}{2} \|Ax - y + \frac{1}{t_k} z^k\|^2 \right\},$$

乘子更新 $\quad z^{k+1} = z^k + t_k(Ax^{k+1} - y^{k+1}).$

对对偶问题(6), 近似点算法迭代格式如下:

$$z^{k+1} = \mathrm{prox}_{t\psi}\left(z^k\right) = \arg\min_{z} \left\{ f^*\left(-A^{\mathrm{T}} z\right) + h^*(z) + \frac{1}{2t_k} \left\|z - z^k\right\|_2^2 \right\}$$

事实上也可以写成：$z^{k+1} = \mathrm{prox}_{t\psi}(z^k) = z^k + t^k(A\hat{x} - \hat{y})$, 其中

$$(\hat{x}, \hat{y}) = \operatorname*{argmin}_{x,y} \left(f(x) + h(y) + z^T(Ax - y) + \frac{t}{2}||Ax - y||_2^2 \right). \tag{8}$$

也就是说，$\hat{x}, \hat{y}$ 最小化增广拉格朗日函数，近似点算法迭代格式对应增广拉格朗日函数法中的乘子更新.

Proof. 考虑(8)的等价问题

$$\min_{x,y,w} \quad f(x)+h(y)+\frac{t}{2}\|w\|_2^2$$
$$\text{s.t.} \quad Ax-y+z/t=w.$$

设u为约束$Ax-y+\frac{z}{t}=w$的拉格朗日乘子, 由最优性条件知

$$A\hat{x}-\hat{y}+\frac{z}{t}=w, \quad -A^{\mathrm{T}}u\in\partial f(\hat{x}), \quad u\in\partial h(\hat{y}), \quad tw=u.$$

于是, 根据(7)得:

$$u=z+t(A\hat{x}-\hat{y}), \quad \hat{x}\in\partial f^*\left(-A^{\mathrm{T}}u\right), \quad \hat{y}\in\partial h^*(u).$$

因此,

$$0\in -A\partial f^*(-A^T u)+\partial h^*(u)+\frac{1}{t}(u-z).$$

这正是 $u=\operatorname{prox}_{t\psi}(z)$ 的最优性条件.

另一方面, 若有$u=\operatorname{prox}_{t\psi}(z)$, 则选取$\hat{x}\in\partial f^*\left(-A^{\mathrm{T}}u\right)$ 及 $\hat{y}\in\partial h^*(u)$, 即可恢复出增广拉格朗日函数法中的变量. □

应用: LASSO问题

考虑LASSO问题:

$$\min \mu\|x\|_1 + \frac{1}{2}\|Ax-b\|_2^2 \quad \Longleftrightarrow \quad \begin{cases} \min\limits_{x,y} \quad f(x,y) = \mu\|x\|_1 + \dfrac{1}{2}\|y\|_2^2 \\ \text{s.t.} \quad Ax - y - b = 0. \end{cases} \tag{9}$$

用近似点算法进行求解(9), 第k步迭代为

$$(x^{k+1}, y^{k+1}) \approx \arg \min_{(x,y)\in\mathcal{D}} \left\{ f(x,y) + \frac{1}{2t_k}\left(\|x-x^k\|_2^2 + \|y-y^k\|_2^2\right)\right\}, \tag{10}$$

其中$\mathcal{D} = \{(x,y) \mid Ax - y = b\}$, t_k 为步长.

由于问题(10)没有显式解, 我们需要采用迭代算法(比如罚函数法, 增广拉格朗日方法等)来进行求解. 另一种比较实用的方式是通过对偶问题的解来构造(x^{k+1}, y^{k+1}).

问题(10)的Lagrangian对偶函数为:

$$
\begin{aligned}
\Phi_k(z) =& \inf_x \left\{ \mu\|x\|_1 + z^T A x + \frac{1}{2t_k}\|x - x^k\|_2^2 \right\} \\
&+ \inf_y \left\{ \frac{1}{2}\|y\|_2^2 - z^T y + \frac{1}{2t_k}\|y - y^k\|_2^2 \right\} - b^T z \\
=& \mu \Gamma_{\mu t_k}(x^k - t_k A^T z) - \frac{1}{2t_k}\left(\|x_k - t_k A^T z\|_2^2 - \|x_k\|_2^2 \right) \\
&- \frac{1}{2(t_k+1)}(t_k\|z\|_2^2 + 2(y^k)^T z - \|y^k\|_2^2) - b^T z.
\end{aligned}
$$

其中

$$
\Gamma_{\mu t_k}(u) = \inf_x \left\{ \|x\|_1 + \frac{1}{2\mu t_k}\|x - u\|_2^2 \right\}.
$$

记函数$q_{\mu t_k}: \mathbb{R} \to \mathbb{R}$ 为

$$q_{\mu t_k}(v) = \begin{cases} \frac{v^2}{2\mu t_k}, & |v| \leq \mu t_k, \\ |v| - \frac{\mu t_k}{2}, & |v| > \mu t_k, \end{cases}$$

通过简单地计算, $\Gamma_{\mu t_k}(u) = \sum_{i=1}^n q_{\mu t_k}(u_i)$且为极小点$x = \text{prox}_{\mu t_k \|x\|_1}(u)$处的目标函数值. 易知$\Gamma_{\mu t_k}(u)$ 是关于u 的连续可微函数且

$$\nabla_u \Gamma_{\mu t_k}(u) = \frac{u - \text{prox}_{\mu t_k \|x\|_1}(u)}{\mu t_k}.$$

设对偶问题的逼近最优解为z^{k+1}, 则根据问题(10)的最优性条件,

$$\begin{cases} x^{k+1} = \text{prox}_{\mu t_k \|x\|_1}\left(x^k - t_k A^T z^{k+1}\right), \\ y^{k+1} = \dfrac{1}{t_k + 1}(y^k + t_k z^{k+1}). \end{cases}$$

于是, LASSO (9) 问题的近似点算法的迭代格式为:

$$
\begin{cases}
z^{k+1} \approx \arg\max_z \ \ \Phi_k(z), \\
x^{k+1} = \mathrm{prox}_{\mu t_k \|x\|_1}\left(x^k - t_k A^T z^{k+1}\right), \\
y^{k+1} = \dfrac{1}{t_k+1}(y^k + t_k z^{k+1}).
\end{cases}
\tag{11}
$$

注: 根据$\Phi_k(z)$ 的连续可微性, 可以调用梯度法进行求解. 可以证明$\Phi_k(z)$是半光滑的, 从而可调用半光滑牛顿法来更有效地求解.

为了保证算法(11)的收敛性, 我们采用以下不精确收敛准则:

$$
\begin{aligned}
&\epsilon_k \geq 0, \ \sum_k^{\infty} \epsilon_k < \infty; \quad \delta_k \geq 0, \ \sum_k^{\infty} \delta_k < \infty, \\
&\|\nabla \Phi_k(z^{k+1})\|_2 \leq \sqrt{\alpha/t_k}\epsilon_k, \\
&\|\nabla \Phi_k(z^{k+1})\|_2 \leq \sqrt{\alpha/t_k}\delta_k \|(x^{k+1}, y^{k+1}) - (x^k, y^k)\|^2,
\end{aligned}
\tag{12}
$$

其中ϵ_k, δ_k是人为设定的参数, α 为$-\Phi_k$ 的强凸参数.

► $m = 512; n = 1024; A = \text{randn}(m, n); u = \text{sprandn}(n, 1, r); b = A * u;$ 取固定步长$t = 10^3$, 精度参数设置为$\alpha_k = \dfrac{t}{t+1}$, $\epsilon_k = \delta_k = \dfrac{8}{k^2}$

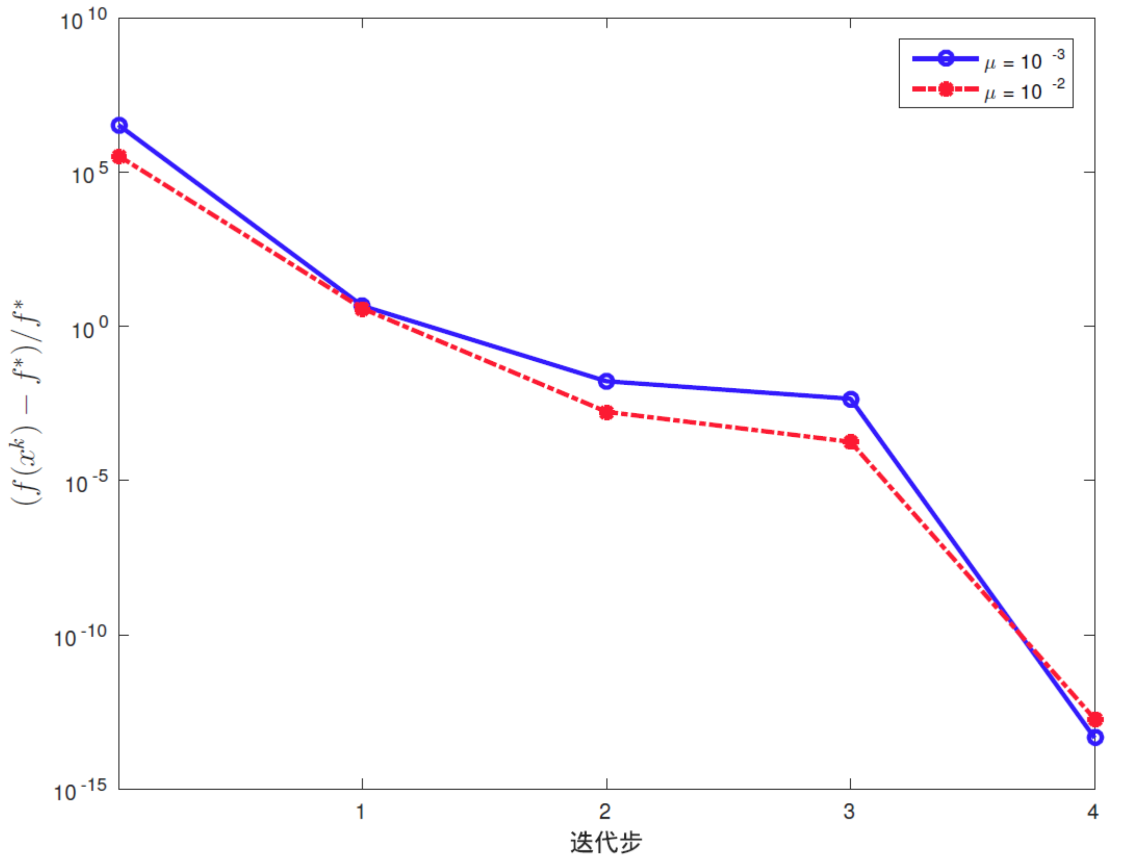


Figure 1: PPA 求解LASSO 问题: PPA收敛所需的外部迭代数很少.

近似点算法(PPA)收敛性分析

基本假设A:

- ψ 是闭凸函数 (因此, 对$\forall x$, $\mathrm{prox}_{t\psi}(x)$ 唯一确定)
- 最优值 ψ^* 有限且在 x^* 取到

结论I: 在**基本假设A**成立的条件下, 近似点算法(2)的收敛性

$$\psi(x^{(k)}) - \psi^* \leq \frac{||x^{(0)} - x^*||_2^2}{2\sum_{i=1}^{k} t_i} \qquad \forall\, k \geq 1.$$

- $\sum_i t_i \to \infty$ 时收敛
- 若 t_i 固定或在一个正下界以上变化，则收敛速率为$1/k$
- t_i 可以任意选取，然而邻近算子的计算代价依赖于 t_i

Proof. PPA就是对原问题应用近似点梯度法(即 $f(x)=0$ 的情形)，则下式在$t>0$的情形下自然满足：

$$f\left(x-tG_t(x)\right) \leqslant f(x)+t\nabla f(x)^{\mathrm{T}}G_t(x)+\frac{t}{2}\left\|G_t(x)\right\|_2^2$$

根据近似点梯度法收敛性的证明，我们有

$$t_i\left(\psi\left(x^i\right)-\psi^*\right) \leqslant \frac{1}{2}\left(\left\|x^{i-1}-x^*\right\|_2^2-\left\|x^i-x^*\right\|_2^2\right)$$

且$\{\psi\left(x^i\right)\}$ 是单调下降的序列. 因此

$$\left(\sum_{i=1}^k t_i\right)\left(\psi\left(x^k\right)-\psi^*\right) \leqslant \sum_{i=1}^k t_i\left(\psi\left(x^i\right)-\psi^*\right) \leqslant \frac{1}{2}\left\|x^0-x^*\right\|_2^2.$$

□

加速近似点算法的收敛性分析

结论II： 在**基本假设A**成立的条件下, 加速近似点算法(3)-(4)的收敛性

$$\psi(x^{(k)}-\psi^*) \leq \frac{2\|x^{(0)}-x^*\|_2^2}{(2\sqrt{t_1}+\sum_{i=2}^k \sqrt{t_i})^2}, \quad k \geq 1.$$

- 若$\sum_i \sqrt{t_i} \to \infty$，则加速近似点算法(3)-(4)保证收敛
- 步长 t_i 取固定值或有正下界时，其收敛速度可达到$\mathcal{O}\left(\frac{1}{k^2}\right)$ (实际上也需控制 t_i 上界以便子问题可快速求解)

Proof. 加速近似点算法(3)-(4)就是在 $f(x)=0$ 的情况下对(1)使用FISTA或第二类Nesterov加速算法, 利用FISTA或第二类Nesterov加速算法收敛性定理的证明.

由于 $f(x)=0$, 对任意的$t>0$,

$$f(x) \leqslant f(y)+\nabla f(y)^{\mathrm{T}}(x-y)+\frac{1}{2t}\|x-y\|_2^2, \quad \forall x, y$$

于是

$$\psi\left(x^{k}\right)-\psi^{*} \leqslant \frac{\gamma_{k}^{2}}{2 t_{k}}\left\|x^{0}-x^{*}\right\|_{2}^{2}. \tag{13}$$

取固定步长$t_k=t$ 和$\gamma_k=\frac{2}{k+1}$, 则

$$\frac{\gamma_k^2}{2t_k}=\frac{2}{(k+1)^2 t}$$

对于变步长，

$$\frac{\gamma_k^2}{2t_k}\leqslant\frac{2}{\left(2\sqrt{t_1}+\sum_{i=2}^{k}\sqrt{t_i}\right)^2}$$

分别由(13) 及第二类Nesterov加速算法收敛性定理的证明，可得结论II. □

Moreau-Yosida 正则化

闭凸函数 f 的Moreau-Yosida正则化 (又称为Moreau envelope) 定义为

$$
\begin{aligned}
f_{(t)}(x) &= \inf_u \left(f(u) + \frac{1}{2t} \|u-x\|_2^2) \quad (t>0) \right) \\
&= f(\mathrm{prox}_{tf}(x)) + \frac{1}{2t}\|\mathrm{prox}_{tf}(x) - x\|_2^2.
\end{aligned}
\tag{14}
$$

- $f_{(t)}$ 是凸函数 (一个以x,u为自变量的凸函数对u取下确界)
- $f_{(t)}$ 的定义域为$\mathbb{R}^n$ ($\mathrm{prox}_{tf}(x)$ 对于任意的x存在且唯一)
- Moreau envelope 分解 $f_{(t)}(x) + f^*_{(t^{-1})}(x) = \dfrac{\|x\|^2}{2t}$

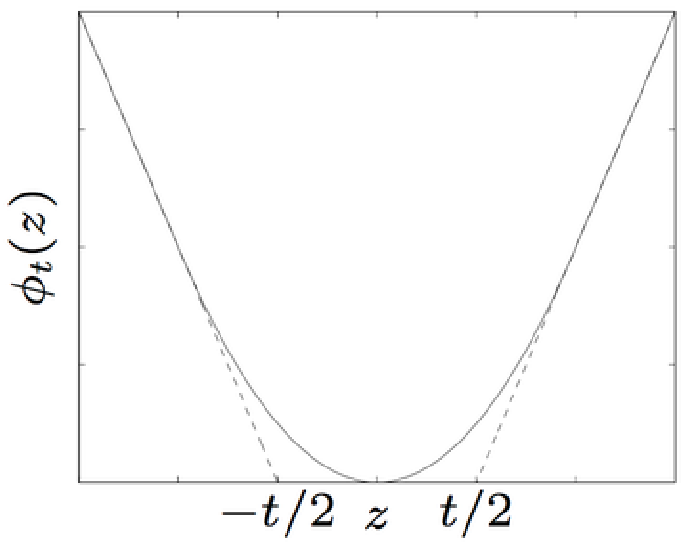
【例1】 示性函数f 的Moreau envelope是Euclid距离的平方

$$
f(x) = I_C(x), \qquad f_{(t)}(x) = \frac{1}{2t} d(x)^2.
$$

【例2】 l_1范数的Moreau envelope是Huber损失函数

$$f(x) = ||x||_1, \qquad f_{(t)}(x) = \sum_{k=1}^{n} \phi_t(x_k),$$

$$\phi_t(z) = \begin{cases} z^2/(2t) & |z| \le t \\ |z| - t/2 & |z| \ge t \end{cases}$$



Moreau envelope的共轭函数

考虑闭凸函数 f 的Moreau envelope

$$f_{(t)}(x) = \inf_{u}\left(f(u) + \frac{1}{2t}||u - x||_2^2\right).$$

- $f_{(t)}$ 是 $f(u)$ 与 $||v||_2^2/(2t)$ 的卷积下确界:

$$f_{(t)}(x) = \inf_{u+v=x}\left(f(u) + \frac{1}{2t}||v||_2^2\right).$$

- $f_{(t)}$ 的共轭是 $f(u)$ 的共轭与 $||v||_2^2/(2t)$ 的和:

$$(f_{(t)})^*(y) = f^*(y) + \frac{t}{2}||y||_2^2.$$

- $f_{(t)}$ 的共轭是 t-强凸的.

  设$f(x)$是适当且闭的强凸函数, 强凸参数为$\mu > 0$, 则$f^*(y)$在全空间$\mathbb{R}^n$ 上有定义, 且是$\dfrac{1}{\mu}$-光滑函数.

Moreau envelope的梯度

$$f_{(t)}(x)=\sup_y(x^Ty-(f_{(t)})^*(y))=\sup_y(x^Ty-f^*(y)-\frac{t}{2}||y||_2^2).$$

- 最大值点 y 唯一且满足:

$$\frac{x}{t}-\partial\frac{f^*(y)}{t}-y=0 \iff y=\arg\min_u\left(\frac{f^*(u)}{t}+\frac{1}{2}||u-\frac{x}{t}||^2\right).$$

- $x\in\partial(f_{(t)})^*(y) \iff y\in\partial f_{(t)}(x)$, 根据唯一性, 最大值点 y 即是 $f_{(t)}$ 的梯度:

$$\nabla f_{(t)}(x)=\mathrm{prox}_{(1/t)f^*}(x/t)=\frac{1}{t}(x-\mathrm{prox}_{tf}(x)).$$

- 梯度 $\nabla f_{(t)}$ 为$(1/t)$-利普希茨连续

  根据邻近算子性质: $\|\mathrm{prox}_h(\boldsymbol{x}_1)-\mathrm{prox}_h(\boldsymbol{x}_2)\|_2\le\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_2$ 可得.

近似点算法的解释

- 使用梯度法来最小化f的Moreau envelope

$$\min \quad f_{(t)}(x) = \inf_u \left(f(u) + \frac{1}{2t} ||u - x||_2^2 \right). \tag{15}$$

这是最小化 $f(x)$ 优化问题的光滑化形式，且满足:

- 问题(15)的最优值点 x 同时也是原问题 f 的最小值点
- $f_{(t)}$ 可微且梯度利普希茨连续 $(L = 1/t)$

- 我们可以使用固定步长：$t_k = 1/L = t$

$$x^{(k)} = x^{(k-1)} - t\nabla f_{(t)}(x^{(k-1)}) = \mathrm{prox}_{tf}(x^{(k-1)}).$$

这就是$\min\ f(x)$的固定步长近似点算法的迭代格式.

增广拉格朗日函数法的解释

- 考虑求解问题(5)的增广拉格朗日函数法迭代:

$$(\hat{x}, \hat{y}) = \underset{x,y}{\operatorname{argmin}} \left(f(x) + h(y) + \frac{t}{2} \|Ax - y + (1/t)z\|_2^2 \right)$$

乘子更新 $\quad z = z + t(A\hat{x} - \hat{y})$

- 固定步长 t , 对偶乘子更新即为对光滑化的对偶问题使用梯度下降

$$z^{k+1} = \operatorname{prox}_{t\psi}\left(z^k\right) = \arg\min_z \left\{ f^*\left(-A^{\mathrm{T}} z\right) + h^*(z) + \frac{1}{2t} \left\| z - z^k \right\|_2^2 \right\}$$

- 如果我们消去 y , 关键变量 x 的更新可以理解为光滑化 h :

$$\hat{x} = \underset{x}{\operatorname{argmin}} \left(f(x) + h_{(1/t)}(Ax + (1/t)z) \right).$$

- 【例】 $\min\ f(x) + \|Ax - b\|_1$

$$\hat{x} = \underset{x}{\operatorname{argmin}} \left(f(x) + \phi_{1/t}(Ax - b + (1/t)z) \right),$$

其中 $\phi_{1/t}$ 表示将Huber损失函数应用到各分量.

交替极小化方法与对偶近似点梯度法

设f, h闭凸函数, 考虑优化问题(5)及其对偶问题(6):

$$
\begin{aligned}
&\text{(P)} \quad \min_{x\in\mathbb{R}^n} \quad \psi(x) = f(x) + h(Ax),\\
&\text{(D)} \quad \max_{z} \quad \phi(z) = -f^*(-A^{\mathrm{T}}z) - h^*(z).
\end{aligned}
$$

- f 是可微$\mu-$强凸函数 $\Rightarrow$ $f^*(-A^{\mathrm{T}}z)$ 是$\dfrac{1}{\mu}-$光滑的

$$
\|A\nabla f^*(-A^{\mathrm{T}}z_1) - A\nabla f^*(-A^{\mathrm{T}}z_2)\|_2 \le \frac{\|A\|_2^2}{\mu}\|z_1 - z_2\|_2.
$$

- PG求解(D)的迭代:

$$
z^{k+1} = \operatorname{prox}_{th^*}\left(z^k + tA\nabla f^*\left(-A^{\mathrm{T}}z^k\right)\right) \tag{16}
$$

引入变量$x^{k+1} = \nabla f^*(-A^{\mathrm{T}}z^k)$, 迭代格式(16)等价于

$$
x^{k+1} = \arg\min_x\left\{f(x) + \left(A^{\mathrm{T}}z^k\right)^{\mathrm{T}}x\right\}, \quad z^{k+1} = \operatorname{prox}_{th^*}\left(z^k + tAx^{k+1}\right)
$$

- 由$h^{**} = h$ 及Moreau分解, 得

$$\begin{aligned} z^k + tAx^{k+1} &= \mathrm{prox}_{th^*}(z^k + tAx^{k+1}) + t\mathrm{prox}_{t^{-1}h}\left(\frac{z^k}{t} + Ax^{k+1}\right) \\ &= z^{k+1} + t\mathrm{prox}_{t^{-1}h}(\frac{z^k}{t} + Ax^{k+1}), \end{aligned}$$

- 对偶近似点梯度法针对原问题(P)的更新:

$$\begin{aligned} x^{k+1} &= \arg\min_x \left\{ f(x) + (z^k)^{\mathrm{T}} Ax \right\}, \\ y^{k+1} &= \mathrm{prox}_{t^{-1}h}\left(\frac{z^k}{t} + Ax^{k+1}\right) \\ &= \arg\min_y \left\{ h(y) - (z^k)^{\mathrm{T}}(y - Ax^{k+1}) + \frac{t}{2}\|Ax^{k+1} - y\|_2^2 \right\}, \\ z^{k+1} &= z^k + t(Ax^{k+1} - y^{k+1}). \end{aligned} \tag{17}$$

交替极小化方法

- (P)的等价问题:

$$\min_{x,y} \ f(x)+h(y), \quad \text{s.t.} \quad y=Ax.$$

- (P)的拉格朗日函数和增广拉格朗日函数:

$$L(x,y,z)=f(x)+h(y)-z^{\mathrm{T}}(y-Ax)$$

$$L_t(x,y,z)=f(x)+h(y)-z^{\mathrm{T}}(y-Ax)+\frac{t}{2}\|y-Ax\|^2$$

- 交替极小化方法迭代格式:

$$x^{k+1}=\arg\min_x L(x,y^k,z^k)$$

$$y^{k+1}=\arg\min_y L_t(x^{k+1},y,z^k)$$

$$z^{k+1}=z^k+t(Ax^{k+1}-y^{k+1})$$

- 对对偶问题(D)用近似点梯度法等价于对原问题(P)使用交替极小化方法

▶ 【例】 利用对偶近似点梯度法(16)求解

$$\min\ f(x) + \|Ax - b\| \qquad \min_{x,y}\ f(x) + \|y\|, \quad \text{s.t. } Ax - b = y$$

- 设f是强凸函数, $\|\cdot\|$是任意一种范数. 令$h(y) = \|y - b\|$, 则

$$h^*(z) = \begin{cases} b^{\mathrm{T}} z & \|z\|_* \le 1, \\ +\infty & \text{otherwise.} \end{cases} \qquad \mathrm{prox}_{th^*}(x) = \mathcal{P}_{\|z\|_* \le 1}(x - tb).$$

应用对偶近似点梯度法(16), 更新如下:

$$x^{k+1} = \arg\min_x \left\{ f(x) + (A^{\mathrm{T}} z^k)^{\mathrm{T}} x \right\}, \quad z^{k+1} = \mathcal{P}_{\|z\|_* \le 1}(z^k + t(Ax^{k+1} - b))$$

- 交替极小化方法:

$$x^{k+1} = \arg\min_x\ f(x) + \|y^k\| + (z^k)^{\mathrm{T}}(Ax - b - y^k),$$

$$y^{k+1} = \arg\min_y f(x^{k+1}) + \|y\| + (z^k)^{\mathrm{T}}(Ax^{k+1} - b - y) + \frac{t}{2}\|Ax^{k+1} - b - y\|_2^2,$$

$$z^{k+1} = z^k + t(Ax^{k+1} - b - y^{k+1})$$

$$L(x, y, z) = f(x) + \|y\| + z^{\mathrm{T}}(Ax - b - y), \quad L_t(x, y, z) = L(x, y, z) + \tfrac{t}{2}\|Ax - b - y\|_2^2$$