半光滑牛顿算法

文再文

北京大学北京国际数学研究中心

教材《最优化：建模、算法与理论》配套电子教案

http://bicmr.pku.edu.cn/~wenzw/optbook.html

致谢：本教案由李勇锋，邓展望协助准备

问题引入

考虑下面的带ℓ_1范数正则的优化问题

$$\min_x \ \mu\|x\|_1 + \frac{1}{2}\|Ax-b\|^2.$$

- ℓ_1范数不可微，$f(x)$的一个次梯度为$A^{\mathrm{T}}(Ax-b)+\mu\mathrm{sign}(x)$.

  如何定义海瑟矩阵?

- 经典牛顿法的更新格式为:

$$x^{k+1} = x^k - \nabla^2 f(x^k)^{-1}\nabla f(x^k).$$

目标：不可微情形如何定义牛顿法?

问题引入

- 令$f(x)=\frac{1}{2}\|Ax-b\|^2$，$h(x)=\mu\|x\|_1$，则

$$
\begin{aligned}
\nabla f(x) &= A^{\mathrm{T}}(Ax-b), \\
\mathrm{prox}_{t_k h}(x) &= \mathrm{sign}(x)\max\{|x|-t_k\mu, 0\}.
\end{aligned}
$$

- 对于该非光滑问题，可以利用近似点梯度算法给出迭代格式：

$$
x^{k+1}=\mathrm{prox}_{t_k h}(x^k-t_k\nabla f(x^k))
$$

- 近似点梯度法收敛到不动点方程的解：

$$
F(x)=x-\mathrm{prox}_{th}(x-t\nabla f(x))=0
$$

- 如何对上述方程应用牛顿法?
  - 由于prox 算子不可微，如何定义雅可比矩阵?
  - 如何定义牛顿法?

提纲

雅可比矩阵

定义 (梯度)

给定函数$f:\mathbb{R}^n\to\mathbb{R}$，且$f$在点$x$的一个邻域内有意义，若存在向量$g\in\mathbb{R}^n$ 满足

$$\lim_{p\to 0}\frac{f(x+p)-f(x)-g^{\mathrm{T}}p}{\|p\|}=0, \tag{1}$$

就称f 在点x 处可微．此时g 称为f 在点x 处的梯度.

当$f:\mathbb{R}^n\to\mathbb{R}^m$ 是向量值函数时，可以定义它的雅可比（Jacobi）矩阵$J(x)\in\mathbb{R}^{m\times n}$，它的第$i$行是分量$f_i(x)$梯度的转置，即

$$J(x)=\begin{bmatrix}\frac{\partial f_1(x)}{\partial x_1} & \frac{\partial f_1(x)}{\partial x_2} & \cdots & \frac{\partial f_1(x)}{\partial x_n}\\ \frac{\partial f_2(x)}{\partial x_1} & \frac{\partial f_2(x)}{\partial x_2} & \cdots & \frac{\partial f_2(x)}{\partial x_n}\\ \vdots & \vdots & & \vdots\\ \frac{\partial f_m(x)}{\partial x_1} & \frac{\partial f_m(x)}{\partial x_2} & \cdots & \frac{\partial f_m(x)}{\partial x_n}\end{bmatrix}.$$

假定$\Omega \subseteq \mathbb{R}^n$是开集，$F : \Omega \to \mathbb{R}^m$是局部利普希茨连续的，根据Rademacher定理，F几乎处处可微的，故可引入广义微分的概念.

定义

设$F : \Omega \to \mathbb{R}^m$是局部利普希茨连续的, D_F是F是可微点组成的集合, F在x 的B-微分可以被定义为

$$\partial_B F(x) := \left\{ \lim_{k\to\infty} \nabla F(x^k) | x^k \in D_F, x^k \to x \right\}.$$

克拉克广义雅可比定义为B-微分的凸包

$$\partial F(x) = \mathrm{conv}(\partial_B F(x)).$$

容易看出，如果$F : \Omega \to \mathbb{R}^m$ 上局部利普希茨连续，则对任意$x \in \Omega$，广义雅可比$\partial F(x)$ 是非空紧凸集.

广义雅克比的性质

- 对于利普希茨连续映射$F:\mathbb{R}^n\to\mathbb{R}^n$, 如果$F$ 是单调的，那么对于任何的$x\in\mathbb{R}^n$，$\partial_B F(x)$ 中的每个元素都是半正定的.
- 由于邻近算子是单调的，因此其广义雅克比矩阵是半正定。此外，由于其是非扩张的，还可以推出它们的特征值的界.
- 设g 是$\mathbb{R}^n$ 上的适当的闭凸函数, $F(x)=\mathbf{prox}_{\gamma g}(x)$. 则对任意的$J\in\partial F(x)$，我们都有$J$ 是对称半正定矩阵，且$\|J\|_2\leq 1$.
- 设$g:\mathbb{R}^n\to\mathbb{R}$ 是（块）可分的，即g可表示成：$g(x)=\sum_{i=1}^n g_i(x_i)$, 其中$x_i\in\mathbb{R}^{n_i}$，$\sum_{i=1}^k n_i=n$ 是x 的所有分量的块划分)，则$\partial_{\mathrm{B}}(\mathbf{prox}_{\gamma g}(x))$ 和$\partial(\mathbf{prox}_{\gamma g}(x))$ 中的所有元素均为（块）对角矩阵.

例 (ℓ_1 范数的邻近算子)

设$g = \|x\|_1$，则

$$\mathbf{prox}_{\gamma g}(x) = \mathbf{sign}(\mathbf{x}) \max(|x| - \gamma, 0), \quad 1 \le i \le n. \tag{2}$$

注意到$\mathbf{prox}_{\gamma g}(x)$ 是可分的，因此$\partial_B(\mathbf{prox}_{\gamma g}(x))$ 中的每个元素均为对角矩阵，设$\alpha = \{i||x_i| > \gamma\}$，$\beta = \{i||x_i| = \gamma\}$，$\delta = \{i||x_i| < \gamma\}$，若$J \in \partial_B(\mathbf{prox}_{\gamma g}(x))$，则我们有

$$J_{ii} = \begin{cases} 1, & \text{若 } i \in \alpha, \\ \in \{0, 1\}, & \text{若 } i \in \beta, \\ 0, & \text{若 } i \in \delta. \end{cases} \tag{3}$$

且上面定义的对角矩阵是关于广义微分$\mathbf{prox}_{\gamma g}(x)$.

提纲

半光滑性

下面给出本章中最重要的一个概念——半光滑性的定义.

定义

设$F:\Omega\to\mathbb{R}^m$是局部利普希茨连续的，称F在x处是半光滑的，如果其满足

(a) F在x点是方向可微的;

(b) 对于任意的d和$J\in\partial F(x+d)$,下面的关系成立

$$\|F(x+d)-F(x)-Jd\|=o(\|d\|),\quad \text{当 } d\to 0.$$

如果（b）被替换成对于任意的d和$J\in\partial F(x+d)$，下面的关系成立

$$\|F(x+d)-F(x)-Jd\|=O(\|d\|^2),\quad \text{当 } d\to 0.$$

那么称F在x处是强半光滑的.

半光滑性

- 半光滑性和强半光滑性在数乘、求和和复合运算下都是封闭的。
- 光滑函数、所有的凸函数，分段连续可微的函数都是半光滑的。具有李普希兹连续梯度的可微函数，p范数$\|\cdot\|_p$和分段线性函数是强半光滑的。一个向量值函数是半光滑的（或强半光滑的）当且仅当每个元素函数是半光滑的(或强半光滑的)。
- 很多函数的邻近算子具有半光滑性和和强半光滑性
  - $\|x\|_1$ 与 $\|x\|_\infty$ 的邻近算子映射是强半光滑的.
  - 多面体的投影是强半光滑的.
  - 对称锥上的投影是强半光滑的.
  - 在许多应用中,邻近算子是逐段C^1 因此是半光滑的.

ℓ_1 范数邻近算子的强半光滑性

例 (ℓ_1 范数)

ℓ_1 范数 $\|x\|_1$ 的邻近算子为收缩算子 $\phi(x) = \mathrm{sign}(x)\max(|x| - \mu t, 0)$，其是强半光滑的.

Proof.

只证明一维的情形，多维的情形可以类似被证明. 其广义雅克比为

$$\partial\phi(x) = \begin{cases} \{1\}, & 若\ |x| > \mu t, \\ [0,1], & 若\ |x| = \mu t, \\ \{0\}, & 若\ |x| < \mu t. \end{cases} \tag{4}$$

Proof.

当$|x| \neq \mu t$时，函数$\phi(x)$是可微的，因此是强半光滑的．故只需要验证在两个不可微点，强半光滑性是成立的．对于$x = \mu t$，如果$d > 0$，则我们有$x + d > \mu t$，因此其广义雅克比为$\partial\phi(x) = \{1\}$，则对于任意的$J \in \partial\phi(x)$，我们有

$$|\phi(x+d) - \phi(x) - Jd| = 0.$$

如果$-2\mu t < d < 0$，则我们有$-\mu t < x + d < \mu t$，因此其广义雅克比为$\partial\phi(x) = \{0\}$，则对于任意的$J \in \partial\phi(x)$，我们有

$$|\phi(x+d) - \phi(x) - Jd| = 0.$$

因此可得在$x = \mu t$处，$\phi(x)$是强半光滑的．类似的，可以证明$x = -\mu t$也是强半光滑的．故$\phi(x)$是强半光滑的． □

提纲

求解非光滑线性方程组的半光滑牛顿算法

- 许多算子分裂算法，比如近似点梯度法、DRS 算法等，等价于一个不动点迭代，其可以诱导一个非线性方程组

$$F(z) = 0, \tag{5}$$

其中$F(z) = z - T(z)$，T为算子分裂算法对应的不动点映射．求解非线性方程组(5) 即能求解原始的复合优化问题．

定义 (求解非光滑方程组(5)的半光滑牛顿算法)

假定F的是局部利普希茨连续的，则其广义雅克比存在．取F在z^k点任意的广义雅克比矩阵$J_k \in \partial F(z^k)$，若J_k可逆，其基本的迭代格式为

$$z^{k+1} = z^k - J_k^{-1} F(z^k). \tag{6}$$

- 半光滑性也能保证牛顿型算法具有超线性收敛或二次收敛．

BD-正则性

定义

如果F在z点所有B-微分的元素$J \in \partial_B F(z)$都是非奇异的，那么称F在z点是BD-正则的.

BD-正则性是一个非光滑方法局部收敛性分析的普遍假设.

假设

定义在(5)中的映射F在最优点z^*是半光滑的和BD-正则的.

注：BD-正则的条件也可以换成是所有雅克比矩阵是非奇异的.

引理

如果假设1 成立，则存在常数$c > 0$，$\kappa > 0$ 和一个小邻域$N(z^*, \varepsilon_0)$ 使得对于任意的$y \in N(z^*, \varepsilon_0)$ 和$J \in \partial_B F(y)$，下面的结论成立：

(1) z^* 是一个孤立解；

(2) J 是非奇异的并且$\|J^{-1}\| \leq c$；

(3) 局部误差界条件对于$F(z)$ 在邻域$N(z^*, \varepsilon_0)$ 上成立，也就是说$\|y - z^*\| \leq \kappa \|F(y)\|$.

半光滑牛顿法的局部超线性收敛性

定理

设假设1 成立并且z^*是$F(z)=0$的解．那么存在一个小邻域$N(z^*,\epsilon)$，使得迭代(6) 是良定义的，且对于任意的k 有$z^k\in N(z^*,\epsilon)$，迭代(6) 是超线性收敛的．如果F是强半光滑的，迭代(6) 是二次收敛的．

Proof.

根据引理1，迭代(6) 是良定义的．我们可以简单的推出

$$\begin{aligned}
\|z^{k+1}-z^*\| &= \|z^k - J_k^{-1}F(z^k) - z^*\| \\
&\le \|J_k^{-1}\|\cdot\|F(z^k)-F(z^*)-J_k(z^k-z^*)\| \\
&= o(\|z^k-z^*\|).
\end{aligned}$$

其中最后一个等式来源于半光滑性．如果F是强半光滑的，则其可以改为$O(\|z^k-z^*\|^2)$ □

求解优化问题的半光滑牛顿算法

- 半光滑牛顿法也可以用于求解优化问题，考虑如下优化问题

$$\min_{x} f(x), \tag{7}$$

  其中$f(x)$是可微的，但不是二阶可微的．其最优性条件为$\nabla f(x) = 0$．
- 可以看到求解优化问题(7) 可以看成是求解一个非光滑的方程组．如果$\nabla f(x)$是半光滑的，并且假定$J^k \in \partial_B(\nabla f(x^k))$是非奇异的，则可以做如下迭代

$$x^{k+1} = x^k - J_k^{-1} \nabla f(x^k). \tag{8}$$

可以看到半光滑牛顿算法求解优化问题实际上是求解方程的半光滑牛顿方法一个简单的推广，因此求收敛性可以简单的从定理1 推广过来，故有如下局部收敛性定理.

定理

设假设*1* 对$\nabla f(x)$ 成立并且x^*是优化问题(7) 的最优解. 那么迭代(8) 是良定义的，且存在一个小邻域$N(x^*,\epsilon)$，使得对于任意的k有$x^k\in N(x^*,\epsilon)$，迭代(8) 是超线性收敛的. 如果$\nabla f(x)$是强半光滑的，迭代(8) 是二次收敛的.

半光滑牛顿法

- 针对凸优化问题，其广义海瑟矩阵是对称半正定的，因此可以选择任意的$J_k \in \partial_B(\nabla f(x^k))$，计算线性方程组

$$(J_k + \mu_k I)d^k = -F^k \tag{9}$$

来得到半光滑牛方向.

- 此时得到的牛顿方向为优化问题的下降方向，因此可以利用Armijo 线搜索准则选取步长，即选取最小的非负常数m_k，满足

$$f(x^k + \rho^m d^k) \leq f(x^k) + \sigma\rho^m \nabla f(x^k)^T \nabla f(x^k), \tag{10}$$

其中$\rho, \sigma \in (0,1)$ 是给定的常数. 取下一步的迭代点为

$$x^{k+1} = x^k + \rho^{m_k} J_k^{-1} d^k. \tag{11}$$

半光滑牛顿法

上述相应的算法总结在算法1.

Algorithm 1 求解优化问题的半光滑牛顿法

给定参数$0 < \sigma, \rho < 1$
选择x^0，设置$k = 0$
while 没有停机时 **do**
    选择$J_k \in \partial_B F(z^k)$
    选取$\mu_k > 0$ 求解牛顿方向d^k 满足(9)
    选取最小的非负常数m_k 满足(10)
    更新x^{k+1} 根据(11)
    设置$k = k + 1$
end while

提纲

1. 广义雅可比
   - 广义雅可比介绍
   - 常见凸函数的近似点算子的广义雅可比
2. 半光滑性质
3. 半光滑牛顿算法
   - 求解非光滑线性方程组的半光滑牛顿算法
   - 求解优化问题的半光滑牛顿算法
4. 应用举例
   - LASSO问题
   - 基追踪问题

LASSO问题：基于近似点梯度法的半光滑牛顿算法

考虑下面的带ℓ_1范数正则的优化问题

$$\min_x \ \mu\|x\|_1 + \frac{1}{2}\|Ax-b\|^2. \tag{12}$$

令$f(x) = \mu\|x\|_1$ 和$h(x) = \frac{1}{2}\|Ax-b\|^2$.

- 利用近似点梯度法可以将需要求解的问题表示成求解下面的非线性方程组

$$F(x) = x - \mathrm{prox}_{tf}(x - t\nabla h(x)) = 0.$$

  $F(x)$的一个广义雅克比矩阵可以表示为

$$J(x) = I - M(x)(I - t\nabla^2 h(x)), \tag{13}$$

  其中$M(x) \in \partial\mathrm{prox}_{tf}(x - t\nabla h(x))$ 和$\nabla^2 h(x)$ 是$h(x)$ 的广义海瑟矩阵（梯度$\nabla h(x)$ 的一个广义雅克比矩阵）.

- 特别的，$f(x)$的邻近算子为收缩算子

$$\left(\mathrm{prox}_{tf}(x)\right)_i = \mathrm{sign}(x_i)\max(|x_i| - \mu t, 0).$$

LASSO问题：基于近似点梯度法的半光滑牛顿算法

- 由上述分析可知广义雅克比矩阵$M(x)$可以为一个对角矩阵，它的对角元素是

$$(M(x))_{ii} = \begin{cases} 1, & \text{如果} |(x - t\nabla h(x))_i| > \mu t, \\ 0, & \text{否则}. \end{cases}$$

- 为了方便表示，我们定义指标集合

$$\begin{aligned} \mathcal{I}(x) &:= \{i : |(x - t\nabla h(x))_i| > t\mu\} = \{i : (M(x))_{ii} = 1\}, \\ \mathcal{O}(x) &:= \{i : |(x - t\nabla h(x))_i| \le t\mu\} = \{i : (M(x))_{ii} = 0\}. \end{aligned}$$

则雅克比矩阵可以被表示成

$$J(x) = \begin{pmatrix} t(\partial^2 h(x))_{\mathcal{I}(x)\mathcal{I}(x)} & t(\partial^2 h(x))_{\mathcal{I}(x)\mathcal{O}(x)} \\ 0 & I \end{pmatrix}.$$

- 利用雅克比矩阵分块结构，可以降低求解线性方程组的复杂度.

LASSO问题：基于近似点梯度法的半光滑牛顿算法

- 令$\mathcal{I} = \mathcal{I}(x^k)$和$\mathcal{O} = \mathcal{O}(x^k)$，然后需求解方程组可以被表示为

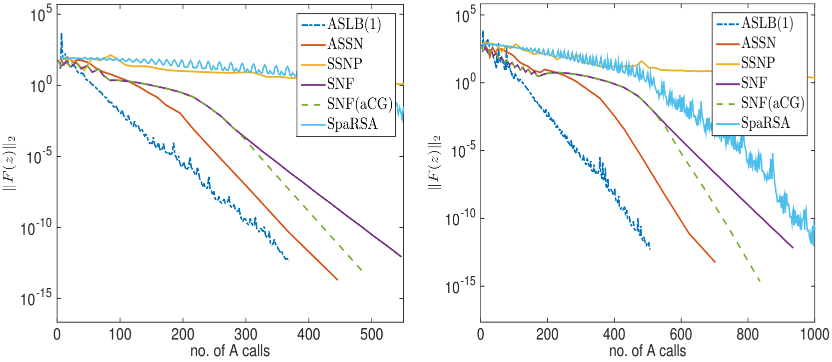
$$
\begin{aligned}
&(1+\mu_k)s^k_{\mathcal{O}} = -F_{k,\mathcal{O}},\\
&(t(\partial^2 h(x))_{\mathcal{I}\mathcal{I}} + \mu I)s^k_{\mathcal{I}} + t(\partial^2 h(x))_{\mathcal{I}\mathcal{O}} s^k_{\mathcal{O}} = -F_{k,\mathcal{I}}.
\end{aligned}
$$

- 可以得到方程组的等价形式

$$
\begin{aligned}
&s^k_{\mathcal{O}} = -\frac{1}{1+\mu_k}F_{k,\mathcal{O}},\\
&(t(\partial^2 h(x))_{\mathcal{I}\mathcal{I}} + \mu I)s^k_{\mathcal{I}} = -F_{k,\mathcal{I}} - t(\partial^2 h(x))_{\mathcal{I}\mathcal{O}} s^k_{\mathcal{O}}.
\end{aligned}
$$

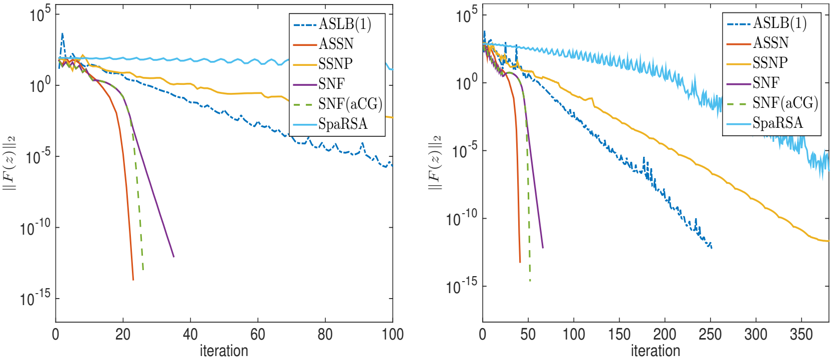
实际上并不需要求解原始比较大的线性方程组，只需要求解规模为$|\mathcal{I}|$的线性方程组即可．由于ℓ_1范数能够保证问题(12) 的解是稀疏的，而指标集$\mathcal{I}$恰好是解非零元的位置，因此在实际问题$|\mathcal{I}|$是非常小的．这表明上述算法很好的利用了问题的稀疏结构，求解牛顿方向的代价是比较小的．

LASSO问题



(a) 20dB　　(b) 40dB

Figure: 关于A- and A^T取用总次数N_A的残差变化



(a) 20dB (b) 40dB

Figure: 关于迭代次数的残差变化

考虑基追踪(BP) 问题

$$\min \ \|x\|_1, \ \text{s.t.} \ Ax = b, \tag{14}$$

其中$A \in \mathbb{R}^{m \times n}$ 是行满秩的，$b \in \mathbb{R}^m$. 令$f(x) = 1_\Omega(Ax - b)$ 和$h(x) = \|x\|_1$，其中设置$\Omega = \{0\}$，则其可以表示为两项均不光滑的复合优化问题.

下面我们讨论两种求解基追踪问题的半光滑牛顿算法.

- 基于DRS 算法的半光滑牛顿算法
- 基于增广拉格朗日函数法的半光滑牛顿算法

考虑复合优化问题

$$\min \quad f(x) = g(x) + h(x)$$

其中g, h 是闭凸函数.

Douglas-Rachford迭代: 从任意初始点z^0开始,

$$\begin{aligned} x^{(k)} &= \mathrm{prox}_{th}(z^{k-1}) \\ y^{(k)} &= \mathrm{prox}_{tg}(2x^k - z^{k-1}) \\ z^{(k)} &= z^{k-1} + y^k - x^k \end{aligned}$$

- t为正常数
- 通常用于g, h 的proximal算子计算代价较小的场景
- 在较弱的条件下(如极小点存在), 迭代点列x^k收敛

基追踪问题：基于DRS的半光滑牛顿算法

DRS 算法不动点映射对应的非线性方程组为

$$F(z) = \mathrm{prox}_{th}(z) - \mathrm{prox}_{tf}(2\mathrm{prox}_{th}(z) - z) = 0. \tag{15}$$

为了简化近似点算子的求解，我们假定$AA^\top = I$．那么函数$f(x)$ 的近似点算子可以写为

$$\begin{aligned}\mathrm{prox}_{tf}(z) &= (I - A^\top A)z + A^\top \left(\mathrm{prox}_{1_\Omega}(Az - b) + b\right) \\ &= z - A^\top(Az - b).\end{aligned}$$

其一个广义雅可比矩阵$D \in \partial\mathrm{prox}_{tf}((2\mathrm{prox}_{th}(z) - z))$ 为

$$D = I - A^\top A. \tag{16}$$

函数$h(x)$ 的近似点算子是

$$(\mathrm{prox}_{th}(z))_i = \mathrm{sign}(z_i)\max(|z_i| - t, 0).$$

我们可以取一个广义雅可比矩阵$M(z) \in \partial\mathrm{prox}_{th}(z)$，其为对角矩阵，且对角元为

$$M_{ii}(z) = \begin{cases} 1, & |(z)_i| > t, \\ 0, & \text{否则}. \end{cases}$$

因此，映射$F(z)$ 的一个广义雅可比矩阵有形式

$$J(z) = M(z) + D(I - 2M(z)). \tag{17}$$

令$W = (I - 2M(z))$ 和$H = W + M(z) + \mu I$· 使用Sherman-Morrison-Woodbury (SMW)公式，我们有逆矩阵

$$\begin{aligned}(J(z) + \mu I)^{-1} &= (H - A^\top AW)^{-1} \\ &= H^{-1} + H^{-1}A^\top(I - AWH^{-1}A^\top)^{-1}AWH^{-1}.\end{aligned}$$

写出矩阵W 和H 的对角元：

$$W_{ii}(z) = \begin{cases} -1, & |(z)_i| > t, \\ 1, & \text{否则} \end{cases} \quad \text{and} \quad H_{ii}(z) = \begin{cases} \mu, & |(z)_i| > t, \\ 1 + \mu, & \text{否则}. \end{cases}$$

那么$WH^{-1} = \frac{1}{1+\mu}I - S$, 其中$S$也是对角矩阵，其对角元为

$$S_{ii}(z) = \begin{cases} \frac{1}{\mu} + \frac{1}{1+\mu}, & |(z)_i| > t, \\ 0, & \text{otherwise}. \end{cases}$$

因此，$I - AWH^{-1}A^\top = (1 - \frac{1}{1+\mu})I + ASA^\top$. 定义指标集合

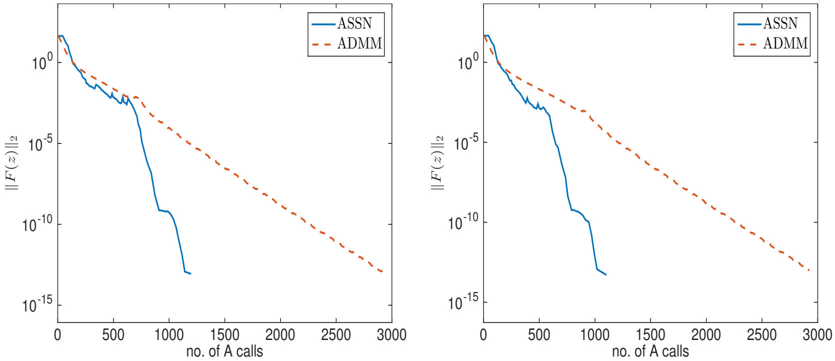
$$\begin{aligned} \mathcal{I}(x) &:= \{i : |(z)_i| > t\} = \{i : M_{ii}(x) = 1\}, \\ \mathcal{O}(x) &:= \{i : |(z)_i| \le t\} = \{i : M_{ii}(x) = 0\} \end{aligned}$$

和$A_{\mathcal{I}(x)}$表示一个矩阵，其包含A的$\mathcal{I}(x)$列. 则有

$$ASA^\top = (\frac{1}{\mu} + \frac{1}{1+\mu})A_{\mathcal{I}(x)}A_{\mathcal{I}(x)}^\top. \tag{18}$$

上面的性质可以推出$I - AWH^{-1}A^\top$是半正定的. 如果子矩阵$A_{\mathcal{I}(x)}$是容易获得的，可以避免在求解牛顿方向时使用更大的矩阵A，从而降低求解牛顿方向的计算复杂度.

基追踪问题



(a) 60dB (b) 80dB

Figure: 关于A- and A^T取用总次数N_A的残差变化

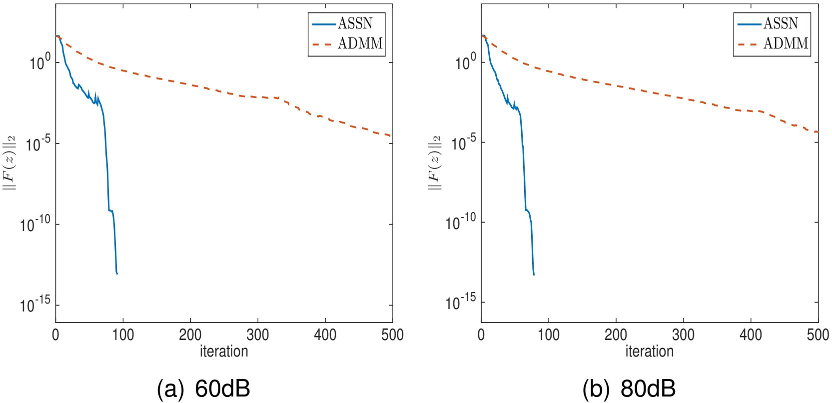


Figure: 关于迭代次数的残差变化

增广拉格朗日乘子法

考虑问题

$$\min_{x\in\mathbb{R}^n} \quad \|x\|_1 \quad \text{s.t.} \quad Ax=b.$$

其对偶问题：

$$\min_{y\in\mathbb{R}^m} \quad b^{\mathrm{T}}y, \quad \text{s.t.} \quad \|A^{\mathrm{T}}y\|_\infty \le 1.$$

- 通过引入变量s，上述问题可以等价地写成

$$\min_{y\in\mathbb{R}^m, s\in\mathbb{R}^n} \quad b^{\mathrm{T}}y, \quad \text{s.t.} \quad A^{\mathrm{T}}y-s=0,\ \|s\|_\infty \le 1. \tag{19}$$

引入拉格朗日乘子λ和罚因子σ，对偶问题的增广拉格朗日函数为

$$L_\sigma(y,s,\lambda)=b^{\mathrm{T}}y+\lambda^{\mathrm{T}}(A^{\mathrm{T}}y-s)+\frac{\sigma}{2}\|A^{\mathrm{T}}y-s\|_2^2, \quad \|s\|_\infty\le 1.$$

- 那么，增广拉格朗日函数法的迭代格式为：

$$\begin{cases} (y^{k+1},s^{k+1})=\displaystyle\operatorname*{argmin}_{y,\|s\|_\infty\le 1}\left\{b^{\mathrm{T}}y+\frac{\sigma}{2}\|A^{\mathrm{T}}y-s+\frac{\lambda}{\sigma_k}\|_2^2\right\}, \\ \lambda^{k+1}=\lambda^k+\sigma(A^{\mathrm{T}}y^{k+1}-s^{k+1}) \end{cases} \tag{20}$$

- 固定y求解只关于s的最小化问题得到

$$s = \mathcal{P}_{\|s\|_\infty \le 1}\left(A^{\mathrm{T}}y + \frac{\lambda}{\sigma}\right). \tag{21}$$

将s 的表达式代入增广拉格朗日函数中，我们得到

$$L_\sigma(y,\lambda) = b^{\mathrm{T}}y + \frac{\sigma}{2}\left\|\psi\left(A^{\mathrm{T}}y + \frac{\lambda}{\sigma}\right)\right\|_2^2 - \frac{\lambda^2}{2\sigma}, \tag{22}$$

其中$\psi(x) = \mathrm{sign}(x)\max\{|x|-1, 0\}$ 。

- 消去s的增广拉格朗日函数法为：

$$\begin{cases} y^{k+1} = \operatorname*{argmin}_y \left\{ b^{\mathrm{T}}y + \frac{\sigma}{2}\left\|\psi\left(A^{\mathrm{T}}y + \frac{\lambda}{\sigma}\right)\right\|_2^2 \right\}, \\ \lambda^{k+1} = \sigma\psi\left(A^{\mathrm{T}}y^{k+1} + \frac{\lambda^k}{\sigma}\right) \end{cases} \tag{23}$$

半光滑牛顿法

- 函数$L_\sigma(y,\lambda^k)$ 关于y 是连续可微的，且其梯度为

$$\nabla_y L_{\sigma_k}(y,\lambda^k) = b + \sigma_k A\psi\left(A^\mathrm{T} y + \frac{\lambda^k}{\sigma_k}\right).$$

- 函数$L_{\sigma_k}(y,\lambda^k)$并不是二阶可微的，但是$\nabla_y L_{\sigma_k}(y,\lambda^k)$是广义可微且半光滑，其一个广义雅可比矩阵为

$$J_k = AD_kA^\mathrm{T},$$

其中D_k是对角矩阵，对角元为

$$(D_k)_{ii} = \begin{cases} 1 & 若 |(A^\mathrm{T} y^{k+1} + \frac{\lambda^k}{\sigma_k})_i| > 1, \\ 0 & 若 |(A^\mathrm{T} y^{k+1} + \frac{\lambda^k}{\sigma_k})_i| \le 1. \end{cases}$$

这样通过半光滑牛顿算法，我们就可以很容易的求解(23) 中子问题，且上面广义雅可比矩阵有很好的稀疏性，能用比较小的代价求解牛顿步.