

# 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

# 一阶算法困难

但尽管一阶算法(近似点梯度法, Nesterov加速算法等等)有非常多的优点, 比如易于实现, 容易并行并且可以很快的计算低精度的解, 但是收敛到高精度的解往往很慢.

为了解决该问题, 可以考虑应用牛顿方法来得到更快的收敛速度. 然而应用牛顿方法有很多的困难.

- 在很多应用中, 问题不可微, 海瑟矩阵不存在.
- 并不能保证全局收敛性.
- 如何合理控制住计算牛顿方向的计算代价.

**目标:** 如何对带结构不可微问题引入具有全局收敛性的半光滑牛顿算法?

# 问题引入

考虑下面的带 $l_1$ 范数正则的优化问题

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

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

如何定义海瑟矩阵？

- 经典牛顿法的更新格式为：

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

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

$$\nabla f(x) = A^T(Ax - b),$$

$$\text{prox}_{t_k h}(x) = \text{sign}(x) \max\{|x| - t_k \mu, 0\}.$$

- 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:

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

- 近似点梯度法收敛到不动点方程的解:

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

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

## 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
- 谱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
- 半定规划

## 5 收敛性分析

# 雅可比矩阵

## 定义 (梯度)

给定函数  $f: \mathbb{R}^n \rightarrow \mathbb{R}$ , 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

$$\lim_{p \rightarrow 0} \frac{f(x+p) - f(x) - g^T p}{\|p\|} = 0, \quad (3)$$

就称  $f$  在点  $x$  处可微. 此时  $g$  称为  $f$  在点  $x$  处的梯度.

当  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$  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

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

# 广义雅可比

假定  $\Omega \subseteq \mathbb{R}^n$  是开集,  $F: \Omega \rightarrow \mathbb{R}^m$  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  $F: \Omega \rightarrow \mathbb{R}^m$  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

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

克拉克广义雅可比定义为  $B$ -微分的凸包

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

容易看出, 如果  $F: \Omega \rightarrow \mathbb{R}^m$  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

## 定理

设  $F: \Omega \rightarrow \mathbb{R}^m$  是局部利普希茨连续的,  $D_F$  是  $F$  是可微点组成的集合, 取  $S$  是一个零测集, 定义

$$\partial_S F(x) = \text{conv} \left( \left\{ \lim_{k \rightarrow \infty} \nabla F(x^k) \mid x^k \in D_F, x^k \notin S, x^k \rightarrow x \right\} \right).$$

则我们有对于任意的  $v \in \mathbb{R}^n$  和  $w \in \mathbb{R}^m$ ,

$$\partial F(x)v = \partial_S F(x)v,$$

$$\partial F(x)^* w = \partial_S F(x)^* w,$$

其中  $*$  表示集合中每一个元素都转置.



# 广义雅克比的性质

## 定理

设  $f = g \circ F$ ，其中  $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$ ，并且在  $x$  附近是利普希茨连续的， $g: \mathbb{R}^m \rightarrow \mathbb{R}$  在  $F(x)$  附近是利普希茨的。那么  $f$  在  $x$  附近是利普希茨的且我们有

$$\partial f(x) \subset \text{conv} \{ \partial g(F(x)) \partial F(x) \}.$$

如果  $g$  在  $F(x)$  点是可微的，那么等式成立，即有

$$\partial f(x) = \partial g(F(x)) \partial F(x).$$

注：实际上  $g$  只需要在  $F(x)$  严格可微即可，见 [cla\[87\], P73](#)

# 广义雅克比的性质

## 定理

假设  $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$  在  $x$  附近是利普希茨的,  $G: \mathbb{R}^m \rightarrow \mathbb{R}^k$  在  $F(x)$  附近是利普希茨的. 那么我们有

$$\partial(G \circ F)(x)v \subset \text{conv} \{ \partial G(F(x)) \partial F(x)v \}.$$

如果  $G$  在  $F(x)$  附近是连续可微的, 那么有

$$\partial(G \circ F)(x)v = \partial G(F(x)) \partial F(x)v.$$

复合映射的广义雅克比矩阵是在矩阵向量乘意义下的, 但这并不会影响实际的使用.

## 定理

对于利普希茨连续映射  $F: \mathbb{R}^n \rightarrow \mathbb{R}^n$ , 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.

# 广义雅克比的性质

## Proof.

首先用反证法证明对于任何的可微点 $\bar{x}$ ,  $\nabla F(\bar{x})$ 是半正定的. 假设存在常数 $a > 0$  和单位向量 $d \in \mathbb{R}^n$ 使得

$$\langle d, \nabla F(\bar{x})d \rangle = -a$$

对任意的常数 $t > 0$ , 定义函数 $\Phi(t) := F(\bar{x} + td) - F(\bar{x}) - t\nabla F(\bar{x})d$ . 由于 $F$ 在点 $\bar{x}$ 处是可微的, 有 $\|\Phi(t)\| = o(t)$ . 映射 $F$ 的单调性可以推出

$$\begin{aligned} 0 &\leq \langle td, F(\bar{x} + td) - F(\bar{x}) \rangle = \langle td, tF'(\bar{x})d + \Phi(t) \rangle \\ &\leq -at^2 + t\|d\|\|\Phi(t)\| = -at^2 + o(t^2). \end{aligned}$$

当 $t$ 充分小时, 有 $-at^2 + o(t^2) < 0$ , 矛盾. 故所有可微点的雅克比矩阵都是半正定的.

对 $\forall x \in \mathbb{R}^n$  和任意的广义雅克比矩阵 $J \in \partial_B F(x)$ , 存在一个收敛到 $x$ 的可微点序列 $x^k \rightarrow x$ 使得 $\nabla F(x^k) \rightarrow J$ . 因为每一个 $\nabla F(x^k)$ 都是半正定的,  $J$ 也是半正定的. □

# 广义雅克比的性质

## 定理

设 $g$ 是 $\mathbb{R}^n$ 上的适当的闭凸函数, $x \in \mathbb{R}^n$ .则对任意的 $J \in \partial(\text{prox}_{\gamma g}(x))$ ,我们都有 $J$ 是对称半正定矩阵,且 $\|J\|_2 \leq 1$ .

## 命题

设 $g: \mathbb{R}^n \rightarrow \mathbb{R}$ 是(块)可分的,即 $g$ 可表示成: $g(x) = \sum_{i=1}^n g_i(x_i)$ ,  
( $g(x) = \sum_{i=1}^k g_i(x_i)$ ,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 $\partial_B(\text{prox}_{\gamma g}(x))$ 和 $\partial(\text{prox}_{\gamma g}(x))$ 中的所有元素均为(块)对角矩阵.

注:该命题给出了可分函数的近似点算子的广义雅克比的性质.可分性可以导致具有对角结构的广义雅克比矩阵,这对算法中降低运算量很有意义.

# 广义雅克比的性质

## Proof.

我们只证明 $g$ 可分的情形， $g$ 块可分的情形的证明完全类似。根据近似点算子的定义，可分函数 $g$ 的近似点算子具有可分的结构，即有

$$\mathbf{prox}_{\gamma g}(x) = (\mathbf{prox}_{\gamma g_1}(x_1), \dots, \mathbf{prox}_{\gamma g_n}(x_n)), \quad (4)$$

于是显然 $\partial_{\mathbf{B}}(\mathbf{prox}_{\gamma g})(x)$ 由对角矩阵组成。再由

$$\partial(\mathbf{prox}_{\gamma g})(x) = \text{conv}(\partial_{\mathbf{B}}(\mathbf{prox}_{\gamma g})(x)), \quad (5)$$

即得命题成立。 □

## 命题

设  $g$  是  $\mathbb{R}^n$  上适当的闭凸函数,  $g^*$  为其共轭函数, 则

$$\begin{aligned}\partial_B(\mathbf{prox}_{\gamma g^*}(x)) &= \{J = I - Q \mid Q \in \partial_B(\mathbf{prox}_{g/\gamma}(x/\gamma))\}, \\ \partial(\mathbf{prox}_{\gamma g^*}(x)) &= \{J = I - Q \mid Q \in \partial(\mathbf{prox}_{g/\gamma}(x/\gamma))\}.\end{aligned}\tag{6}$$

## Proof.

利用 Moreau 分解我们有

$$\mathbf{prox}_{\gamma g^*}(x) = x - \gamma \mathbf{prox}_{g/\gamma}(x/\gamma).\tag{7}$$

利用上式以及定义(2) 立即得到第一个等式 (因为这里我们将  $\mathbf{prox}_{\gamma g^*}$  表示成两个连续可导函数的差). 再由

$$\text{conv}\{I - Q \mid Q \in \partial_B(\mathbf{prox}_{g/\gamma}(x/\gamma))\} = I - \text{conv}(\partial_B(\mathbf{prox}_{g/\gamma}(x/\gamma))),\tag{8}$$

我们就得到第二个等式. □

## 例 (超平面)

设超平面  $D = \{x | Ax = b\}$ , 其中  $A \in \mathbb{R}^{m \times n}$ . 于是其投影映射  $\mathcal{P}_D(x) = x - A^\dagger(Ax - b)$ , 其中  $A^\dagger$  为  $A$  的 *Moore-Penrose* 广义逆. 显然  $\mathcal{P}_D$  是线性映射, 因此其处处可导, 故其广义雅克比为

$$\partial(\mathcal{P}_D(x)) = \partial_B(\mathcal{P}_D(x)) = \nabla \mathcal{P}_D(x) = \{I - A^\dagger A\}. \quad (9)$$

## 例 (半空间)

记  $x_+ = \max\{0, x\}$ . 对于半空间  $D = \{x | a^\top x \leq b\}$ , 我们有

$$\mathcal{P}_D(x) = x - \left( \frac{(a^\top x - b)_+}{\|a\|_2^2} \right) a, \quad (10)$$

以及

$$\partial(\mathcal{P}_D(x)) = \begin{cases} \left\{ I - \frac{aa^\top}{\|a\|_2^2} \right\}, & \text{若 } a^\top x > b, \\ \{I\}, & \text{若 } a^\top x < b, \\ \text{conv} \left\{ I, I - \frac{aa^\top}{\|a\|_2^2} \right\}, & \text{若 } a^\top x = b. \end{cases} \quad (11)$$

# 广义雅可比例子

## 例 (单位球)

设  $B = \{x \mid \|x\|_2 = 1\}$  为单位球, 则其投影映射为

$$\mathcal{P}_B(x) = \begin{cases} x/\|x\|_2, & \text{若 } \|x\|_2 > 1, \\ x, & \text{若 } \|x\|_2 \leq 1. \end{cases} \quad (12)$$

若定义  $w = x/\|x\|_2$ , 则我们有

$$\partial(\mathcal{P}_B)(x) = \begin{cases} \left\{ \frac{I - ww^\top}{\|x\|_2} \right\}, & \text{若 } \|x\|_2 > 1, \\ \{I\}, & \text{若 } \|x\|_2 < 1, \\ \text{conv} \left\{ \frac{I - ww^\top}{\|x\|_2}, I \right\}, & \text{若 } \|x\|_2 = 1. \end{cases} \quad (13)$$



# 广义雅可比例子

## 例 (二次锥)

设  $(t, x) \in \mathbb{R} \times \mathbb{R}^n$ , 定义二次锥  $K = \{(t, x) \mid \|x\|_2 \leq t\}$ . 则对任意  $(t, x) \in \mathbb{R} \times \mathbb{R}^n$ , 若

$$V \in \partial_B(\mathcal{P}_K)((t, x)), \quad (14)$$

则或者  $V = 0$ , 或者  $V = I_{n+1}$ , 或者  $V$  可以表示成

$$V = \begin{pmatrix} 1 & w \\ w & H \end{pmatrix}, \quad (15)$$

其中  $w \in \mathbb{R}^n$  为单位向量, 而  $H \in \mathbb{R}^{n \times n}$  有如下形式

$$H = (1 + \alpha)I_n - \alpha w w^\top, \quad |\alpha| \leq 1. \quad (16)$$

# 广义雅可比例子

## 例 ( $l_2$ 范数)

设  $g = \|x\|_2$ , 则

$$\mathbf{prox}_{\gamma g}(x) = \begin{cases} (1 - \gamma/\|x\|_2)x, & \text{若 } \|x\|_2 \geq \gamma, \\ 0, & \text{若 } \|x\|_2 < \gamma. \end{cases} \quad (17)$$

注意到  $\mathbf{prox}_{\gamma g}(x)$  是分片光滑的, 故其  $B$ -次微分可以通过分片求其雅可比矩阵得到, 特别地, 若令  $w = x/\|x\|_2$ , 则

$$\partial_B(\mathbf{prox}_{\gamma g}(x)) = \begin{cases} \{I - \gamma/\|x\|_2(I - ww^\top)\}, & \text{若 } \|x\|_2 \geq \gamma, \\ \{0\}, & \text{若 } \|x\|_2 < \gamma, \\ \{I - \gamma/\|x\|_2(I - ww^\top), 0\}, & \text{若 } \|x\|_2 = \gamma. \end{cases} \quad (18)$$

# 广义雅可比例子

## 例 ( $l_1$ 范数)

设  $g = \|x\|_1$ , 则

$$\mathbf{prox}_{\gamma g}(x) = \mathbf{sign}(x) \max(|x| - \gamma, 0), \quad 1 \leq i \leq n. \quad (19)$$

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  $\alpha = \{i \mid |x_i| > \gamma\}$ ,  $\beta = \{i \mid |x_i| = \gamma\}$ ,  $\delta = \{i \mid |x_i| < \gamma\}$ , 若  $J \in \partial_B(\mathbf{prox}_{\gamma g}(x))$ , 则我们有

$$J_{ii} = \begin{cases} 1, & \text{若 } i \in \alpha, \\ \in \{0, 1\}, & \text{若 } i \in \beta, \\ 0, & \text{若 } i \in \delta. \end{cases} \quad (20)$$

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

# 广义雅克比例子

- 谱函数:  $F: \mathbb{S}^n \rightarrow \mathbb{R} \cup \{\infty\}$

$$F(X) = f(\lambda(X)), \quad X \in \mathbb{S}^n, \quad (21)$$

其中  $\mathbb{S}^n$  表示所有  $n$  阶实对称矩阵组成的集合,  $\lambda: \mathbb{S}^n \rightarrow \mathbb{R}^n$  为对应矩阵的特征值 (从大到小排列).  $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$  是一个适当的闭凸函数, 且它是绝对对称的, 即对于任意的  $x \in \mathbb{R}^n$  和任意的置换矩阵  $P \in \mathbb{R}^{n \times n}$  都有  $f(x) = f(Px)$ .

- 特别地, 如果假定  $f$  有形式  $f(x) = g(x_1) + \dots + g(x_n)$ , 那么, 其  $F$  近似点算子可以写为

$$\mathbf{prox}_{\gamma F}(X) = Q \mathbf{diag}(\mathbf{prox}_{\gamma g}(\lambda_1(X)), \dots, \mathbf{prox}_{\gamma g}(\lambda_n(X))) Q^T. \quad (22)$$

# 广义雅克比例子

- 谱函数 $F$ 继承了 $f$ 的许多性质. 令 $X = Q\text{diag}(\lambda(X))Q^\top$  是实对称矩阵 $X$ 的谱分解. 如果 $f$ 是可微的, 那么 $F$ 也是可微的, 且有

$$\nabla F(X) = Q\nabla f(\lambda(X))Q^\top.$$

- $F$ 的近似点算子也可以从 $f$ 继承过来, 我们有

$$\text{prox}_{\gamma F}(X) = Q\text{diag}(\text{prox}_{\gamma f}(\lambda(X)))Q^\top, \quad (23)$$

其中 $X = Q\text{diag}(\lambda(X))Q^\top$  是实对称矩阵 $X$ 的谱分解.

# 广义雅克比例子

## 定理

设  $h: \mathbb{R} \rightarrow \mathbb{R}$  是局部利普希茨连续的, 假定实对称矩阵的特征分解为  $X = Q \text{diag}(\lambda_1, \dots, \lambda_n) Q^\top$ , 算子  $T: \mathbb{S}^n \rightarrow \mathbb{S}^n$  定义为

$$H(X) = Q(h(\lambda_1), \dots, h(\lambda_n))Q^\top.$$

则对于任意的  $X \in \mathbb{S}^n$ ,  $B$ -次微分  $\partial_B H$  存在且非空, 且对于任意的  $J \in \partial_B H$ , 我们有

$$J(S) = Q(\Omega \odot (Q^\top S Q))Q^\top, \quad \forall S \in \mathbb{S}^n,$$

其中  $\odot$  表示 *Hadamard* 积, 而矩阵  $\Omega \in \mathbb{R}^{n \times n}$  的各个元素按如下方式定义

$$\Omega_{ij} = \begin{cases} \frac{h(\lambda_i) - h(\lambda_j)}{\lambda_i - \lambda_j}, & \text{若 } \lambda_i \neq \lambda_j, \\ \in \partial h(\lambda_i), & \text{若 } \lambda_i = \lambda_j. \end{cases} \quad (24)$$

## 广义雅克比例子

- 由上述定理可以写出 $\mathbf{prox}_{\gamma F}$ 的B-次微分. 对于任意 $X \in \mathbb{S}^n$ 和 $P \in \partial_{\mathbb{B}}(\mathbf{prox}_{\gamma F})(X)$ , 有

$$P(S) = Q(\Omega \odot (Q^{\top} S Q)) Q^{\top}, \quad \forall S \in \mathbb{S}^n, \quad (25)$$

矩阵 $\Omega \in \mathbb{R}^{n \times n}$ 的各个元素按如下方式定义

$$\Omega_{ij} = \begin{cases} \frac{\mathbf{prox}_{\gamma g}(\lambda_i) - \mathbf{prox}_{\gamma g}(\lambda_j)}{\lambda_i - \lambda_j}, & \text{若 } \lambda_i \neq \lambda_j, \\ \in \partial(\mathbf{prox}_{\gamma g}(\lambda_i)), & \text{若 } \lambda_i = \lambda_j. \end{cases} \quad (26)$$

## 例 (半正定锥的指示函数)

令  $g = \delta_{\mathbb{R}_+}$  且  $f, F$  如前文定义, 则此时  $F$  可以看作是正定锥  $\mathbb{S}_+^n$  的指示函数. 此时有  $\text{prox}_{\gamma g}(x) = \mathcal{P}_{\mathbb{R}_+}(x) = \max\{x, 0\}$ . 于是可得

$$\mathcal{P}_{\mathbb{S}_+^n}(X) = Q \text{diag}((\lambda_1)_+, \dots, (\lambda_n)_+) Q^\top, \quad (27)$$

其中  $X = Q \text{diag}(\lambda(X)) Q^\top$ . 再定义  $\alpha = \{i | \lambda_i > 0\}$  和  $\bar{\alpha} = \{i | \lambda_i \leq 0\}$ , 则可以得到其中一个雅克比矩阵  $P \in \partial_B \mathcal{P}_{\mathbb{S}_+^n}(X)$  的形式

$$P(S) = Q(\Omega \odot (Q^\top S Q)) Q^\top, \quad \forall S \in \mathbb{S}^n,$$

其中

$$\Omega = \begin{pmatrix} \Omega_{\alpha\alpha} & k_{\alpha\bar{\alpha}} \\ k_{\alpha\bar{\alpha}}^\top & 0 \end{pmatrix},$$

$\Omega_{\alpha\alpha} \in \mathbb{R}^{|\alpha| \times |\alpha|}$  的元素全为1, 而  $k_{\alpha\bar{\alpha}} \in \mathbb{R}^{|\alpha| \times |\bar{\alpha}|}$  且其第  $i$  行第  $j$  列元素为

$$\frac{\lambda_i}{\lambda_i - \lambda_j}, \quad i \in \alpha, j \in \bar{\alpha}.$$



## 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
- 谱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
- 半定规划

## 5 收敛性分析

# 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 $F: \Omega \rightarrow \mathbb{R}^m$ 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

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

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

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

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  $\phi(x) = \text{sign}(x) \max(|x| - \mu t, 0)$ , 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

$$\partial\phi(x) = \begin{cases} \{1\}, & \text{若 } |x| > \mu t, \\ [0, 1], & \text{若 } |x| = \mu t, \\ \{0\}, & \text{若 } |x| < \mu t. \end{cases} \quad (28)$$

# $l_1$ 范数的强半光滑性

## 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  $-\mu t < d < 0$ , 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. □

## 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
- 谱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
- 半定规划

## 5 收敛性分析

# 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

$$F(z) = 0, \quad (29)$$

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(29) 即能求解原始的复合优化问题。

## 定义 (求解非光滑方程组(29)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

$$z^{k+1} = z^k - J_k^{-1} F(z^k). \quad (30)$$

- 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(29)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.



# 半光滑牛顿法的局部收敛性

## 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 $\|y - z^*\| \leq \kappa \|F(y)\|$ .

下面的定理给出了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(30)是良定义的, 且对于任意的 $k$ 有 $z^k \in N(z^*, \epsilon)$ , 迭代(30)是超线性收敛的. 如果 $F$ 是强半光滑的, 迭代(30)是二次收敛的.

# 收敛性证明

## Proof.

根据引理1，迭代(30)是良定义的。我们可以简单的推出

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

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 $O(\|z^k - z^*\|^2)$  □

# 半光滑牛顿法的全局收敛性

## 定理

设 $F$ 是在集合 $S = \{x \in \mathbb{R}^n : \|x - x^0\| \leq r\}$ 是半光滑的. 假定对任意的雅克比矩阵 $J \in \partial F(x)$ 是非奇异的且有 $\|J^{-1}\| \leq C$ . 如果对任意的 $x, y \in S$ 和 $J \in \partial F(x)$ 有

$$\|F(y) - F(x) - J(y - x)\| \leq \beta \|y - x\|,$$

其中 $\alpha = \beta C$  且 $C\|F(x^0)\| \leq r(1 - \alpha)$ . 那么迭代(30) 可以保持落在集合 $S$ 中, 且收敛到 $F$ 在 $S$ 中的唯一解. 更具体的, 我们有对任意 $k \geq 1$ ,

$$\|x^k - x^*\| \leq \frac{\alpha}{1 - \alpha} \|x^k - x^{k-1}\|.$$

注: 上述定理更多的是理论上的意义, 其条件对于实际应用往往不容易验证, 在实际设计算法过程中, 需要提出针对问题的全局性策略.

# 半光滑牛顿法

- 由于非线性方程组(29) 在应用中都有比较好的性质，可作出如下假设：

## 假设

映射 $F$  具有半光滑性，单调性和 $\alpha$ -平均性

- 根据定理4，我们知道单调映射的B-微分中的每一个元素都是半正定的，因此我们可以运用正则化的牛顿算法确定迭代方向，即对于当前迭代点 $z_k$ ，选择任意的 $J_k \in \partial_B F(z^k)$ ，计算线性方程组

$$(J_k + \mu_k I)d = -F^k, \quad (31)$$

其中 $F^k = F(z^k)$ ， $\mu_k = \lambda_k \|F^k\|$ 和 $\lambda_k > 0$ 是一个正则化参数。

# 半光滑牛顿法

- 为了提高算法的效率，牛顿系统不需要精确求解。定义

$$r^k := (J_k + \mu_k I)d^k + F^k. \quad (32)$$

在迭代中，可用共轭梯度算法等迭代算法近似求解线性方程组以得到  $d^k$ ，使其满足

$$\|r^k\| \leq \tau \min\{1, \lambda_k \|F^k\| \|d^k\|\}. \quad (33)$$

得牛顿步更新：

$$u^k = z^k + d^k.$$

## 全局收敛保护策略

- 选择参数  $0 < \nu < 1$  和一个固定整数  $\zeta > 0$ . 如果残差  $\|F(u^k)\|$  在最近几步中有充分的下降量, 即,

$$\|F(u^k)\| \leq \nu \max_{\max(1, k-\zeta+1) \leq j \leq k} \|F(z^j)\|, \quad (34)$$

则可做牛顿方向, 即  $z^{k+1} = u^k$ . 否则, 做一个不动点迭代步  $z^{k+1} = z^k - \beta F(z^k)$ , 其中  $\beta \in (0, 1/\alpha)$ .

- 当  $\lambda_k$  比较大时,  $d_k$  非常接近一个一阶算法的方向并且会导致残差  $\|F(u^k)\|_F$  的下降. 但是这样可能导致一个比较慢的收敛速率.
- 当  $\lambda_k$  是比较小的, 收敛速率可能是比较快的, 但是  $d_k$  有可能是一个非常不好的方向, 导致不能收敛.

为此定义比率  $\rho_k$  去决定如何更新  $\lambda_k$ .

$$\rho_k = \frac{-\langle F(u^k), d^k \rangle}{\|d^k\|^2} \quad (35)$$

## 全局收敛保护策略

- 如果比率 $\rho_k$ 比较小, 表明可能出现坏的牛顿方向, 则增大 $\lambda_k$ . 否则倾向于减小它. 设置如下更新规则:

$$z^{k+1} = \begin{cases} u^k, & \text{如果 } u_k \text{ 满足(34) [牛顿步]} \\ z^k - \beta F(z^k), & \text{否则. [不动点迭代步]} \end{cases} \quad (36)$$

- 并且参数 $\lambda_{k+1}$  按如下更新:

$$\lambda_{k+1} = \begin{cases} \max\{\underline{\lambda}, \gamma_0 \lambda_k\}, & \text{如果 } \rho_k \geq \eta_2, \\ \gamma_1 \lambda_k, & \text{如果 } \eta_1 \leq \rho_k < \eta_2, \\ \min\{\bar{\lambda}, \gamma_2 \lambda_k\}, & \text{否则,} \end{cases} \quad (37)$$

其中 $0 < \gamma_0 < 1, \gamma_2 > \gamma_1 \geq 1$  和 $0 < \eta_1 < \eta_2$  是预先选定的参数,  $\underline{\lambda}, \bar{\lambda}$  是两个正的常数.

# 半光滑牛顿法

下将上述半光滑牛顿算法总结在算法1中.

---

## Algorithm 1 非线性单调方程自适应半光滑牛顿算法 (ASSN)

---

- 1: 给定参数  $0 < \tau, \nu < 1$ ,  $0 < \gamma_0 < 1$ ,  $\gamma_2 > \gamma_1 \geq 1$  和  $0 < \eta_1 < \eta_2$  选择  $z^0$ , 设置  $k = 0$
  - 2: **while** 没有停机时 **do**
  - 3:   选择  $J_k \in \partial_B F(z^k)$
  - 4:   近似求解线性系统(40)使得  $d^k$  满足(33)
  - 5:   计算  $u^k = z^k + d^k$  并且计算  $\rho_k$  如(35)
  - 6:   更新  $z^{k+1}$  和  $\lambda_{k+1}$  根据(36) 和(37).
  - 7:   设置  $k = k + 1$
  - 8: **end while**
-



# 半光滑牛顿法

- 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

$$\min_x f(x), \quad (38)$$

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(38)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

$$x^{k+1} = x^k - J_k^{-1} \nabla f(x^k). \quad (39)$$

# 半光滑牛顿法收敛性

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7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(38)的最优解。那么迭代(39)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(39)是超线性收敛的。如果 $\nabla f(x)$ 是强半光滑的，迭代(39)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

$$(J_k + \mu_k I) d^k = -F^k \quad (40)$$

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

$$f(x^k + \rho^m d^k) \leq f(x^k) + \sigma \rho^m \nabla f(x^k)^T \nabla f(x^k), \quad (41)$$

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

$$x^{k+1} = x^k + \rho^{m_k} J_k^{-1} d^k. \quad (42)$$

# 半光滑牛顿法

上述相应的算法总结在算法2.

---

## Algorithm 2 求解优化问题的半光滑牛顿算法

---

- 1: 给定参数  $0 < \sigma, \rho < 1$
  - 2: 选择  $x^0$ , 设置  $k = 0$
  - 3: **while** 没有停机时 **do**
  - 4:   选择  $J_k \in \partial_B F(z^k)$
  - 5:   选取  $\mu_k > 0$  求解牛顿方向  $d^k$  满足(40)
  - 6:   选取最小的非负常数  $m_k$  满足(41)
  - 7:   更新  $x^{k+1}$  根据(42)
  - 8:   设置  $k = k + 1$
  - 9: **end while**
-

## 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
- 谱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
- **半定规划**

## 5 收敛性分析

# 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

$$\min_x \mu \|x\|_1 + \frac{1}{2} \|Ax - b\|^2. \quad (43)$$

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

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

$F(x)$ 的一个广义雅克比矩阵可以表示为

$$J(x) = I - M(x)(I - t\nabla^2 h(x)), \quad (44)$$

其中 $M(x) \in \partial \text{prox}_{tf}(x - t\nabla h(x))$  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

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

# 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

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

- 为了方便表示，我们定义指标集合

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

则雅克比矩阵可以被表示成

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

- 利用雅克比矩阵的分块结构，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

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

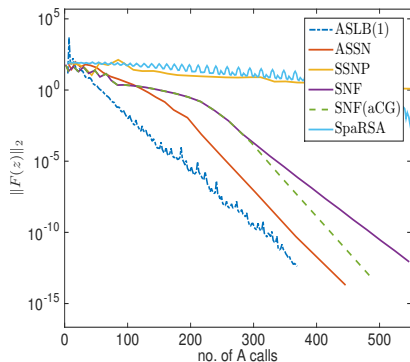
- 可以得到方程组的等价形式

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

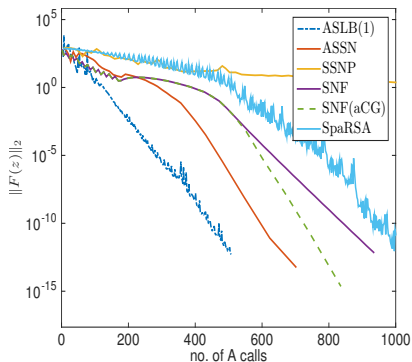
实际上并不需要求解原始比较大的线性方程组，只需要求解规模为  $|\mathcal{I}|$  的线性方程组即可。由于  $\ell_1$  范数能够保证问题(43)的解是稀疏的，而指标集  $\mathcal{I}$  恰好是解非零元的位置，因此在实际问题  $|\mathcal{I}|$  是非常小的。这表明上述算法很好的利用了问题的稀疏结构，求解牛顿方向的代价是比较小的。



# LASSO问题



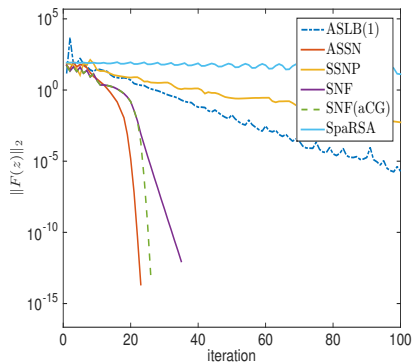
(a) 20dB



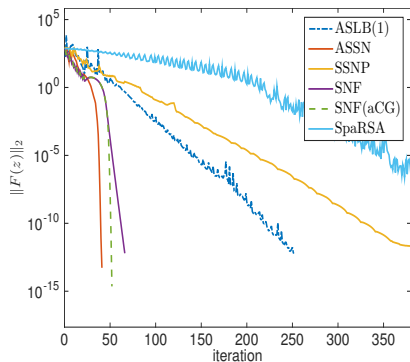
(b) 40dB

Figure: 关于A- and  $A^T$ 取用总次数 $N_A$ 的残差变化

# LASSO问题



(a) 20dB



(b) 40dB

Figure: 关于迭代次数的残差变化

# 基追踪问题

考虑基追踪(BP) 问题

$$\min \|x\|_1, \quad \text{s.t. } Ax = b, \quad (45)$$

其中 $A \in \mathbb{R}^{m \times n}$  是行满秩的,  $b \in \mathbb{R}^m$ . 令 $f(x) = \mathbf{1}_\Omega(Ax - b)$  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
- 基于近似点梯度法的半光滑牛顿算法

## 回顾DRS算法

考虑复合优化问题

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

其中 $g, h$ 是闭凸函数.

**Douglas-Rachford迭代:** 从任意初始点 $z^0$ 开始,

$$x^{(k)} = \text{prox}_{th}(z^{k-1})$$

$$y^{(k)} = \text{prox}_{tg}(2x^k - z^{k-1})$$

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

- $t$ 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

$$F(z) = \text{prox}_{th}(z) - \text{prox}_{tf}(2\text{prox}_{th}(z) - z) = 0. \quad (46)$$

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

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

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

$$D = I - A^\top A. \quad (47)$$

函数  $h(x)$  的近似点算子是

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

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

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

## 基追踪问题：基于DRS的半光滑牛顿算法

因此，映射 $F(z)$ 的一个广义雅可比矩阵有形式

$$J(z) = M(z) + D(I - 2M(z)). \quad (48)$$

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

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

写出矩阵 $W$ 和 $H$ 的对角元：

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

## 基追踪问题：基于DRS的半光滑牛顿算法

那么  $WH^{-1} = \frac{1}{1+\mu}I - S$ , where  $S$  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

$$\mathcal{I}(x) := \{i : |(z)_i| > t\} = \{i : M_{ii}(x) = 1\},$$

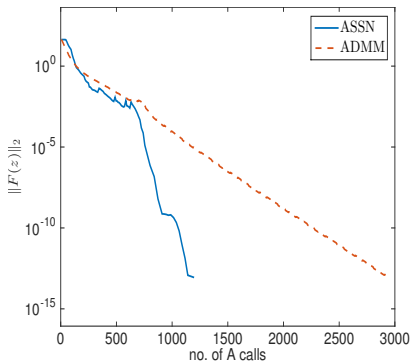
$$\mathcal{O}(x) := \{i : |(z)_i| \leq t\} = \{i : M_{ii}(x) = 0\}$$

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

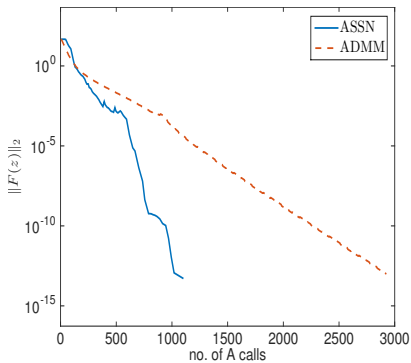
$$ASA^\top = \left(\frac{1}{\mu} + \frac{1}{1+\mu}\right)A_{\mathcal{I}(x)}A_{\mathcal{I}(x)}^\top. \quad (49)$$

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

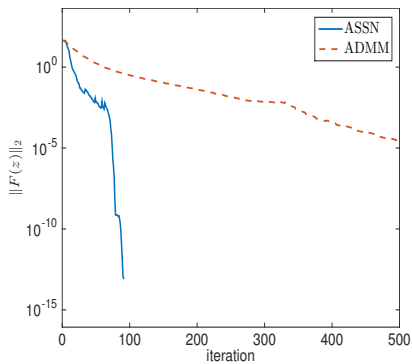


(b) 80dB

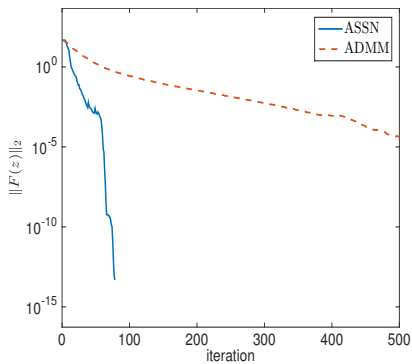
Figure: 关于A- and  $A^T$ 取用总次数 $N_A$ 的残差变化



# 基追踪问题



(a) 60dB



(b) 80dB

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

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

其对偶问题：

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

- 通过引入变量 $s$ ，上述问题可以等价地写成

$$\min_{y \in \mathbb{R}^m, s \in \mathbb{R}^n} b^T y, \quad \text{s.t.} \quad A^T y - s = 0, \|s\|_\infty \leq 1. \quad (50)$$

引入拉格朗日乘子 $\lambda$ 和罚因子 $\sigma$ ，对偶问题的增广拉格朗日函数为

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

- 那么，增广拉格朗日函数法的迭代格式为：

$$\begin{cases} (y^{k+1}, s^{k+1}) = \operatorname{argmin}_{y, \|s\|_\infty \leq 1} \left\{ b^T y + \frac{\sigma}{2} \|A^T y - s + \frac{\lambda}{\sigma_k}\|_2^2 \right\}, \\ \lambda^{k+1} = \lambda^k + \sigma (A^T y^{k+1} - s^{k+1}) \end{cases} \quad (51)$$

## 增广拉格朗日乘子法

- 固定 $y$ 求解只关于 $s$ 的最小化问题得到

$$s = \mathcal{P}_{\|s\|_\infty \leq 1} \left( A^T y + \frac{\lambda}{\sigma} \right). \quad (52)$$

将 $s$ 的表达式代入增广拉格朗日函数中，我们得到

$$L_\sigma(y, \lambda) = b^T y + \frac{\sigma}{2} \left\| \psi \left( A^T y + \frac{\lambda}{\sigma} \right) \right\|_2^2 - \frac{\lambda^2}{2\sigma}, \quad (53)$$

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

- 消去 $s$ 的增广拉格朗日函数法为：

$$\begin{cases} y^{k+1} = \underset{y}{\operatorname{argmin}} \left\{ b^T y + \frac{\sigma}{2} \left\| \psi \left( A^T y + \frac{\lambda}{\sigma} \right) \right\|_2^2 \right\}, \\ \lambda^{k+1} = \sigma \psi \left( A^T y^{k+1} + \frac{\lambda^k}{\sigma} \right) \end{cases} \quad (54)$$

# 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

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

- 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

$$J_k = A D_k A^T,$$

其中  $D_k$  是对角矩阵，对角元为

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

这样通过半光滑牛顿算法，我们就可以很容易的求解(54) 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。

# 半定规划

考虑下面的半定规划问题:

- 令  $\mathcal{S}^n = \{X \in \mathbb{R}^{n \times n} \mid X^\top = X\}$ . 对于两个对称矩阵  $C, X \in \mathcal{S}^n$ , 它们的内积定义为  $\langle C, X \rangle = \text{tr}CX$ . 对于  $m$  个给定对称矩阵  $A_1, \dots, A_m \in \mathcal{S}^n$ , 定义如下的线性算子  $\mathcal{A}: \mathcal{S}^n \rightarrow \mathbb{R}^m$  为  $\mathcal{A}X = (\langle A_1, X \rangle, \dots, \langle A_m, X \rangle)^T$ .
- 线性算子  $\mathcal{A}: \mathcal{S}^n \rightarrow \mathbb{R}^m$  的共轭算子定义为  $\mathcal{A}^*y = \sum_{p=1}^m A_p y_p$ . 利用上面的符号, 对于一个给定的向量  $b \in \mathbb{R}^m$ , 我们可以定义半定规划的原问题为

$$\begin{aligned} \max_{X \in \mathcal{S}^n} \quad & \langle C, X \rangle \\ \text{s.t.} \quad & \mathcal{A}X = b, \\ & X \succeq 0. \end{aligned} \tag{55}$$

# 基于DRS 算法的半光滑牛顿算法

- 令

$$f(X) = -\langle C, X \rangle + 1_{\{AX=b\}}(X) \text{ 和 } h(X) = 1_K(X), \quad (56)$$

其中  $K = \{X : X \succeq 0\}$ .

- DRS 算法可能被表示成一个不动点迭代的形式, 从而可以转化成求解一个非线性方程组

$$F(Z) = \text{prox}_{th}(Z) - \text{prox}_{tf}(2\text{prox}_{th}(Z) - Z) = 0, \quad (57)$$

其中  $Z \in \mathcal{S}^n$ .

- 容易得到  $F$  是强半光滑的和单调的, 且  $F$  的广义雅克比  $\partial F(x)$  的每个元素都是半正定的, 故我们可以使用半光滑牛顿法求解非线性方程组(57). 为了使用半光滑牛顿法, 我们考虑函数  $F(Z)$  的广义雅克比矩阵的结构.

# 基于DRS 算法的半光滑牛顿算法

- 利用(56)中给出的函数 $f(x)$ 和 $h(x)$ 的定义, 我们写出 $\text{prox}_{f^*}(Y)$ 和 $\text{prox}_{th}(Z)$ 的显示形式为

$$\begin{aligned}\text{prox}_{f^*}(Y) &= (Y + tC) - \mathcal{A}^*(\mathcal{A}Y + t\mathcal{A}C - b), \\ \text{prox}_{th}(Z) &= Q_\alpha \Sigma_\alpha Q_\alpha^T,\end{aligned}$$

- 其中

$$Q\Sigma Q^T = (Q_\alpha \quad Q_{\bar{\alpha}}) \begin{pmatrix} \Sigma_\alpha & 0 \\ 0 & \Sigma_{\bar{\alpha}} \end{pmatrix} \begin{pmatrix} Q_\alpha^T \\ Q_{\bar{\alpha}}^T \end{pmatrix}$$

是矩阵 $Z$ 的特征分解,  $\Sigma = \text{diag}(\lambda_1, \dots, \lambda_n)$ 是以特征值 $\lambda_i$ 为对角的对角矩阵,  $\alpha = \{i | \lambda_i > 0\}$ 和 $\bar{\alpha} = \{1, \dots, n\} \setminus \alpha$ 是矩阵 $Z$ 的正特征值和非正特征值对应的指标集合.

# 基于DRS算法的半光滑牛顿算法

- 令 $\mathcal{I}$ 是恒同算子，那么算子

$$\mathcal{D} = I - \mathcal{A}^* \mathcal{A} \quad (58)$$

是近似点算子 $\partial \text{prox}_{tf}(\cdot)$ 在 $(2\text{prox}_{th}(Z) - Z)$ 处的一个广义雅克比。

- 定义一个广义雅克比 $\mathcal{M}(Z) \in \partial \text{prox}_{th}(Z)$ ，其作用在 $n \times n$ 对称矩阵 $S$ 上为

$$\mathcal{M}(Z)[S] = Q(\Omega \odot (Q^T S Q))Q^T, \text{ for all } S \succeq 0, \quad (59)$$

其中记号 $\odot$ 表示Hadamard积，

$$\Omega = \begin{bmatrix} E_{\alpha\alpha} & k_{\alpha\bar{\alpha}} \\ k_{\alpha\bar{\alpha}}^T & 0 \end{bmatrix},$$

$E_{\alpha\alpha}$ 是元素全为1的矩阵， $k_{ij} = \frac{\lambda_i}{\lambda_i - \lambda_j}, i \in \alpha, j \in \bar{\alpha}$ .



# 基于DRS 算法的半光滑牛顿算法

通常去定义广义雅克比 $\partial F(Z)$ 的精确形式并不是一件非常容易的事情.

- 定义如下的替代形式:

$$\hat{\partial}F(Z) = \partial\text{prox}_{th}(Z) + \mathcal{D}(I - 2\partial\text{prox}_{th}(Z)) \quad (60)$$

- 并可以选择其中一个元素 $\mathcal{J}(Z) \in \hat{\partial}F(Z)$ :

$$\mathcal{J}(Z) = \mathcal{M}(Z) + \mathcal{D}(I - 2\mathcal{M}(Z)). \quad (61)$$

根据cla[87],p75,我们有

$$\hat{\partial}F(Z)[S] = \partial F(Z)[S], \text{ 对于所有的 } S \succeq 0. \quad (62)$$

## 基于DRS 算法的半光滑牛顿算法

- 因为广义雅克比矩阵  $\mathcal{J}$  是非对称的，并且其维度是非常大，故运用SMW 公式将求解牛顿方向的线性方程组转换成一个更小的对称线性方程组。通过向量化矩阵  $S$ ，算子  $\mathcal{M}(Z)$  和  $\mathcal{D}$  可以写成矩阵形式：

$$M(Z) = \tilde{Q}\Lambda\tilde{Q}^T \text{ 和 } D = I - A^T A,$$

其中  $\tilde{Q} = Q \otimes Q$ ,  $\Lambda = \text{diag}(\text{vec}(\Omega))$ ,  $I$  是恒同矩阵且  $A$  是算子  $\mathcal{A}$  的矩阵形式。令  $W = I - 2M(Z) = \tilde{Q}(I - 2\Lambda)\tilde{Q}^T$  和  $H = \tilde{Q}((\mu_k + 1)I - \Lambda)\tilde{Q}^T$ 。那么算子  $\mathcal{J}_k + \mu_k I$  的矩阵形式可以被写为  $J_k + \mu_k I = H - A^T A W$ 。

- 根据SMW 公式，有

$$\begin{aligned} (J_k + \mu_k I)^{-1} &= (H - A^T A W)^{-1} \\ &= H^{-1} + H^{-1} A^T (I - A W H^{-1} A^T)^{-1} A W H^{-1}. \end{aligned}$$

# 基于DRS算法的半光滑牛顿算法

- 定义

$$T = \tilde{Q}L\tilde{Q}^T, \quad (63)$$

其中 $L$ 是一个对角矩阵, 其对角元为 $L_{ii} = \frac{\Lambda_{ii}\mu_k}{\mu_k+1-\Lambda_{ii}}$ . 通过运用等式 $H^{-1} = \frac{1}{\mu_k+1}I + \frac{1}{\mu_k(\mu_k+1)}T$ 和 $WH^{-1} = \frac{1}{1+\mu_k}I - (\frac{1}{\mu_k} + \frac{1}{\mu_k+1})T$ , 我们有

$$\begin{aligned} & (J_k + \mu_k I)^{-1} \\ &= \frac{\mu_k I + T}{\mu_k(\mu_k + 1)} \left( I + A^T \left( \frac{\mu_k^2}{2\mu_k + 1} I + ATA^T \right)^{-1} A \left( \frac{\mu_k}{2\mu_k + 1} I - T \right) \right). \end{aligned} \quad (64)$$

- 因此, 牛顿方向的求解可以通过求解下面的对称线性方程组

$$\left( \frac{\mu_k^2}{2\mu_k + 1} I + ATA^T \right) d_s = a, \quad (65)$$

其中 $a = -A(\frac{\mu_k}{2\mu_k+1}I - T)\text{vec}(F^k)$ 且 $\text{vec}$ 把一个矩阵表示成向量的向量化算子.

# 基于DRS算法的半光滑牛顿算法

- 方程(65)的求解可以通过迭代算法，比如共轭梯度（CG）算法和对称QMR算法。并且方程(65)的系数矩阵是 $m \times m$ 的，而原方程组的系数矩阵是 $n^2 \times n^2$ 的。通常 $m$ 要远小于 $n^2$ 。在求解出方程(65)的解后，可下面的公式求得原方程组的解：

$$S^k = \frac{1}{\mu_k(\mu_k + 1)} (\mu_k \mathcal{I} + \mathcal{T}) [-F^k + \mathcal{A}^* d_s],$$

其中 $\mathcal{T}$ 是(63)中矩阵 $T$ 的算子形式。

- 特别地，作用 $\mathcal{T}$ 在对称矩阵 $S$ 上有

$$\mathcal{T}(Z)[S] = Q(\Omega_0 \odot (Q^T S Q)) Q^T, \forall S \succeq 0,$$

其中 $\Omega_0 = \begin{bmatrix} E_{\alpha\alpha} & l_{\alpha\bar{\alpha}} \\ l_{\alpha\bar{\alpha}}^T & 0 \end{bmatrix}$ ,  $l_{ij} = \frac{\mu_k k_{ij}}{\mu_k + 1 - k_{ij}}$ .

# 基于DRS算法的半光滑牛顿算法

- 令  $\Upsilon = \mathcal{T}(Z)[S]$ . 我们可以将  $\Upsilon$  表达为

$$\Upsilon = [Q_\alpha Q_{\bar{\alpha}}] \begin{bmatrix} Q_\alpha^T S Q_\alpha & l_{\alpha\bar{\alpha}} \odot Q_\alpha^T S Q_{\bar{\alpha}} \\ l_{\alpha\bar{\alpha}}^T \odot Q_{\bar{\alpha}}^T S Q_\alpha & 0 \end{bmatrix} \begin{bmatrix} Q_\alpha^T \\ Q_{\bar{\alpha}}^T \end{bmatrix} = G + G^T, \quad (66)$$

其中  $G = Q_\alpha(\frac{1}{2}(UQ_\alpha^T) + l_{\alpha\bar{\alpha}} \odot (UQ_{\bar{\alpha}}))$  且  $U = Q_\alpha^T S$ .

- 计算  $\Upsilon$  需要的浮点数为  $8|\alpha|n^2$ . 如果  $|\alpha|$  比较大, 我们也可以计算  $\Upsilon$  通过等价形式  $\Upsilon = S - Q((E - \Omega_0) \odot (Q^T S Q))Q^T$ , 其要求  $8|\bar{\alpha}|n^2$  浮点数. 因此, 当  $|\alpha|$  或  $|\bar{\alpha}|$  比较小时, 使用表达(66)允许我们比较有效的得到线性方程组的解.

---

## Algorithm 3 求解牛顿方向

---

- 1: 计算  $a = -\mathcal{A}(\frac{\mu_k}{2\mu_k+1}\mathcal{I} - \mathcal{T})F^k$
  - 2: 使用CG 或对称QMR 方法近似求解
  - 3:  $(\frac{\mu_k^2}{2\mu_k+1}\mathcal{I} + \mathcal{A}\mathcal{T}\mathcal{A}^*)d_s = a$ , 其中矩阵向量乘通过方法(66)
  - 4: 计算牛顿方向  $S^k = \frac{1}{\mu_k(\mu_k+1)}(\mu_k\mathcal{I} + \mathcal{T})(-F^k + \mathcal{A}^*d_s)$ .
-

# 基于增广拉格朗日乘子法的半光滑牛顿算法

下面我们介绍基于增广拉格朗日乘子法的半光滑牛顿算法.

- 从一个初始点 $X^0$ 开始, 增广拉格朗日算法求解半定规划对偶问题的格式为

$$y^{k+1} = \arg \min \tilde{L}_{\sigma^k}(y, X^k) \quad (67)$$

$$X^{k+1} = \Pi_{S_+^n}(X^k - \sigma(\mathcal{A}^*y^{k+1} - C)), \quad (68)$$

- 其中

$$\tilde{L}_{\sigma}(y, X) = b^T y + \frac{1}{2\sigma} (\|\Pi_{S_+^n}(X - \sigma(\mathcal{A}^*y - C))\|_F^2 - \|X\|_F^2), \quad (69)$$

$\Pi_{S_+^n}$  是到半正定矩阵的投影. 然后子问题(67)通过半光滑牛顿算法进行求解.

# 基于增广拉格朗日乘子法的半光滑牛顿算法

- 函数(69) 关于 $y$  的梯度和广义海瑟矩阵为

$$\nabla_y \tilde{L}_\sigma(y, X) = b - \mathcal{A} \Pi_{S_+^n}(X - \sigma(\mathcal{A}^*y - C)), \quad (70)$$

$$V \in \sigma \mathcal{A} \partial \Pi_{S_+^n}(X - \sigma(\mathcal{A}^*y - C)) \mathcal{A}^*. \quad (71)$$

- 对于固定的 $y$  和 $X$ , 相应的半光滑牛顿步为

$$(V + \epsilon I)d = \nabla_y L_\sigma(y, X), \quad (72)$$

其中 $\epsilon$  是一个比较小的常数.

## 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
- 谱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
- 半定规划

## 5 收敛性分析



# 算法1的收敛性分析

## 假设

最优解的集合 $Z^*$  是非空的. 函数 $F: \mathbb{R}^n \rightarrow \mathbb{R}^n$  是单调的, 相应的不动点算子是 $\alpha$ -平均的 ( $\alpha \in (0, 1]$ ).

- 该假设对于常见的算子分裂算法都是成立的, 比如近似点梯度法和DRS 算法等.
- 如果 $T$  是 $\alpha$ -平均的, 可以推出 $F$  是全局利普希茨连续的且利普希茨常数为 $L \leq 2\alpha$ . 故对于 $\forall k \geq 0$  和 $\forall J_k \in \partial_B F(z^k)$ , 有 $\|J_k\| \leq L$ .
- 由于假设3对于实现全局收敛是充分的,  $F$ 的半光滑性实际上并不会再这小节中被使用.

## 定理

设假设3 成立并且 $\{z^k\}$  是算法1 产生的迭代序列. 那么 $\{z^k\}$ 的残差收敛到0, 即 $\lim_{k \rightarrow \infty} \|F(z^k)\| = 0$ .

# 算法1的收敛性分析

## Proof.

如果牛顿步的数目是有限的，那么在整个迭代过程中，从某一步开始均为不动点迭代。由根据文献Dav[16]定理1有 $\lim_{k \rightarrow \infty} \|F(z^k)\| = 0$ 。因此，只需要考虑牛顿步的数目是无限的情形。根据Dav[16]定理1有在不动点迭代步后，残差都是不增的，即有 $\|F(z^{k+1})\| \leq \|F(z^k)\|$ 。

定义

$$\bar{F}^k = \max_{k-\zeta+1 \leq j \leq k} \|F(z^j)\|, \text{ for all } k \geq \zeta.$$

对于不动点迭代步，有 $\|F(z^{k+1})\| \leq \|F(z^k)\| \leq \bar{F}^k$ 。对于牛顿步，有 $\|F(z^{k+1})\| \leq \nu \bar{F}^k$ 。

因此可以证明 $\bar{F}^k$ 是不增的，即

$$\bar{F}^{k+1} \leq \max\{\|F(z^{k+1})\|, \bar{F}^k\} \leq \max\{\nu, 1\} \bar{F}^k = \bar{F}^k.$$

最后证明如果第 $k$ 步是一个成功的牛顿步，则 $\bar{F}^{k+\zeta} \leq \nu \bar{F}^k$ 。 □

## Proof.

由上面分析可知：只需要证明

$$\|F(z^{k+j})\| \leq v\bar{F}^k, \quad \forall 1 \leq j \leq \zeta. \quad (73)$$

当 $j=1$ 时，断言(73)成立由于牛顿步成立的条件。当 $j=2$ 时，我们考虑下面两个情形：如果在第 $(k+1)$ 步是一步不动点迭代步，那么

$$\|F(z^{k+2})\| \leq \|F(z^{k+1})\| \leq v\bar{F}^k.$$

如果在第 $(k+1)$ 步是一步牛顿步，那么

$$\|F(z^{k+2})\| \leq v\bar{F}^{k+1} \leq v\bar{F}^k,$$

其中最后一个不等式是由于 $\bar{F}^k$ 是不增的。这表明断言(73)对于 $j=2$ 成立。继续上面的论证方式，可得对于 $2 < j \leq \zeta$ ，断言(73)都成立。结合成功牛顿步是无限的以及 $\bar{F}^k$ 是不增的，有 $\lim_{k \rightarrow \infty} \bar{F}^k = 0$ ，其推出 $\lim_{k \rightarrow \infty} \|F(z^k)\| = 0$ 。 □

为了分析局部收敛性，首先做如下假设.

## 假设

定义在(29)中的映射 $F$ 是半光滑的和BD-正则的，即广义雅可比的所有元素都是非奇异的.

由于近似点算子在很多应用中都是(强)半光滑的，因此，算子分裂算法诱导的映射 $F$ 的半光滑性通常是满足的. **BD-正则性是一个非光滑方法局部收敛性分析的普遍假设.**

## 定理

设假设3和4成立并且正则参数 $\lambda_k$ 有上界 $\bar{\lambda}$ 对于任何的 $z^k \in N(z^*, \varepsilon_1)$ ，其中 $\varepsilon_1 := \min\{\varepsilon_0, 1/(2Lc\tau\bar{\lambda})\}$ .

所以对于充分大的 $k$ ，有 $\|F(u^k)\| \leq \nu\|F(z^k)\|$ 和 $z^{k+1} = u^k$ ，以及 $\{z^k\}$ 超线性收敛到 $z^*$ . 如果 $F$ 在 $z^*$ 是强半光滑的，那么 $\{z^k\}$ 二次收敛到 $z^*$ .

## Proof.

$\forall z^k \in N(z^*, \varepsilon_1)$ , 由  $F$  的 Lipschitz 连续性,  $\|F^k\| \leq L\|z^k - z^*\| \leq L\varepsilon_1$  因此,  $\varepsilon_1$  的定义推出

$$c\tau\bar{\lambda}\|F^k\| \leq 1/2. \quad (74)$$

根据 (40), (33), (74) 和  $\|J_k^{-1}\| \leq c$ , 得到对于牛顿步有

$$\begin{aligned} \|d^k\| &\leq \|(J_k + \mu_k I)^{-1}F^k\| + \|(J_k + \mu_k I)^{-1}r^k\| \\ &\leq cL\|z^k - z^*\| + c\tau\bar{\lambda}\|F^k\|\|d^k\| \\ &\leq 2cL\|z^k - z^*\|, \end{aligned} \quad (75)$$

通过直接计算可以得到

$$\begin{aligned} \|u^k - z^*\| &= \|z^k + d^k - z^*\| \\ &= \|z^k + (J_k + \mu_k I)^{-1}(F^k + (J_k + \mu_k I)d^k - F^k) - z^*\| \\ &\leq \|z^k - z^* - (J_k + \mu_k I)^{-1}F^k\| + \|(J_k + \mu_k I)^{-1}\| \cdot \|F^k + (J_k + \mu_k I)d^k\| \\ &\leq \|(J_k + \mu_k I)^{-1}\| \cdot [\|F^k - F(z^*) - J_k(z^k - z^*)\| + \mu_k\|z^k - z^*\| + \|r^k\|]. \end{aligned}$$

## Proof.

使用  $\|(J_k + \mu_k I)^{-1}\| \leq c$ ,  $\mu_k = \lambda_k \|F^k\|$  和  $\|r^k\| \leq \tau \lambda_k \|F^k\| \|d^k\|$ , 我们有

$$\|u^k - z^*\| \leq c(\|F^k - F(z^*) - J_k(z^k - z^*)\| + \bar{\lambda} \|F^k\| \|z^k - z^*\| + \tau \bar{\lambda} \|F^k\| \|d^k\|). \quad (76)$$

根据  $F$  的  $L$ -利普希茨连续性和(75), 有

$$\bar{\lambda} \|F^k\| \|z^k - z^*\| + \tau \bar{\lambda} \|F^k\| \|d^k\| \leq L \bar{\lambda} (1 + 2cL\tau) \|z^k - z^*\|^2. \quad (77)$$

映射  $F$  在  $z^*$  处的半光滑性推出

$$\|F^k - F(z^*) - J_k(z^k - z^*)\| = o(\|z^k - z^*\|), \quad \text{当 } \|z^k - z^*\| \rightarrow 0,$$

其结合(76)和(77)推出  $\|u^k - z^*\| = o(\|z^k - z^*\|)$ .

## Proof.

故对于充分大的 $k$ ,  $L\|u^k - z^*\| \leq \frac{\nu}{\kappa}\|z^k - z^*\|$ . 局部误差界条件推出对于充分大的 $k$ ,  $\|z^k - z^*\| \leq \kappa\|F(z^k)\|$ . 结合 $F$ 的利普希茨连续性, 有对于充分大的 $k$ ,

$$\|F(u^k)\| \leq L\|u^k - z^*\| \leq \frac{\nu}{\kappa}\|z^k - z^*\| \leq \nu\|F(z^k)\|.$$

那么更新准则(36)产生 $\|F(u^k)\| \leq \nu\|F(z^k)\|$  并且 $z^{k+1} = u^k$ .

当 $F$ 是强半光滑的, 二次收敛 $\|u^k - z^*\| = O(\|z^k - z^*\|^2)$  被建立由于 $\|F^k - F(z^*) - J_k(z^k - z^*)\| = O(\|z^k - z^*\|_2^2)$ .



尽管BD-正则性是比较强的条件，但可能在一些情形中不满足，这里仍然有很多方法去解决这个问题。

- 设 $\partial_B F(z^*)$ 中存在一个非奇异的元素，它的其它元素可能是奇异的。通过探索 $\partial_B F(z)$ 的结构，当 $z$ 非常接近 $z^*$ ，我们能很容易的选择非奇异的广义雅可比矩阵。因此，如果 $z^*$ 是孤立的，我们可以通过上面类似的证明建立快速的局部收敛性。
- 文献MIL[14]中的另外一种方法是Levenberg-Marquardt (LM) 方法。LM方法是解决非奇异系统的正则化高斯牛顿方法。
- 通过文献FAN[05]可知，LM方法在局部误差界条件下，保持超线性或者二次收敛速度，其要弱于非奇异条件。算法在局部误差界条件中的收敛性是一个值得探索的方向。



## 假设

定义在(38)中的函数 $f(x)$ 是凸函数并且具有利普希茨连续的梯度. 梯度 $\nabla f(x)$ 是半光滑的和BD-正则的, 即广义雅可比的所有元素都是非奇异的.

## 引理

设假设5成立. 如果 $\nabla f(x)$ 是半光滑的, 那么对于任意的 $x$ 有

$$\lim_{\substack{J \in \partial(\nabla f(x+d)) \\ \|d\| \rightarrow 0}} \frac{f(x+d) - f(x) - \nabla f(x)^T d - \frac{1}{2} d^T J d}{\|d\|^2} = 0. \quad (78)$$

## 引理

设假设5成立并且 $x^*$ 是优化问题(38)的最优解. 对于任意的 $\delta \in (0, 1)$ , 存在 $x^*$ 的一个领域 $N(x^*, \epsilon)$ 和常数 $\bar{\mu}$ , 使得对于任意的 $x \in N(x^*, \epsilon)$ 和雅克比矩阵 $J \in \partial(\nabla f(x))$ , 以及常数 $\mu \in [0, \bar{\mu}]$ , 方程(40)的解 $d$ 满足

$$\|x + d - x^*\| \leq \delta \|x - x^*\|.$$

## 定理

设假设5成立并且 $x^*$ 是优化问题(38)的最优解. 如果 $\sigma < \frac{1}{2}$ , 那么算法2产生的序列 $\{x^k\}$ 满足下面的结论:

- 1 存在整数 $k_0$ , 使得对所有的 $k \geq k_0$ 有 $m_k = 0$ .
- 2 整个序列 $\{x^k\}$ 收敛到 $x^*$ 且具有二次收敛性.

# References

- Frank H. Clarke, *Optimization and nonsmooth analysis*, Society for Industrial and Applied Mathematics(1987)
- Xiao Xiantao, Li Yongfeng, Wen Zaiwen, Zhang Liwei; *A Regularized Semi-Smooth Newton Method with Projection Steps for Composite Convex Programs*; *Journal of Scientific Computing*; 2018, Vol 76, No. 1, pp 364-389
- PATRINOS P, STELLA L, BEMPORAD A., *Forward-backward truncated newton methods for convex composite optimization*. arXiv preprint arXiv:1402.6655(2014)
- PANG J S, QI L. *Nonsmooth equations: motivation and algorithms* SIAM Journal on optimization, 1993, 3(3): 443-465.
- ZHAO X Y, SUN D, TOH K C. *A newton-CG augmented Lagrangian method for semidefinite programming[J]*. SIAM Journal on Optimization, 2010, 20(4): 1737-1765.

# References

- QI L. *Convergence analysis of some algorithms for solving nonsmooth equations*[J] 1993, 18(1):227-244.
- MENG F, SUN D, ZHAO G. *Semismoothness of solutions to generalized equations and the Moreau-Yosida regularization*[J]. *Mathematical programming*, 2005, 104(2):561-581.
- FAN J Y, YUAN Y X. *On the quadratic convergence of the levenbergmarquardt method without nonsingularity assumption*[J]. *Computing*, 2005, 74(1): 23-39.