应用数学和力学编委会编
重 庆 出 版 社 出 版

文章编号: 1000_0887(2005) 09_1067_09

解非光滑方程组的 Krylov 子空间迭代法*

孟泽红, 张建军

(上海大学 数学系, 上海 200444)

(程昌钧推荐)

摘要: 给出了求解非光滑方程组的 Newton_FOM 算法和 Newton_GMRES 算法, 证明了这些 Krylov 子空间方法的局部平方收敛性· 数值结果表明了算法的有效性·

关 键 词: 非光滑方程组; Newton_FOM 算法; Newton_GMRES 算法
中图分类号: O241.7 文献标识码: A

引 言

考虑非光滑方程组

$$F(\boldsymbol{y}) = 0 \tag{1}$$

的求解问题, 这里 $F: D \subseteq R^n \to R^n$· 1993 年, QI 和 SUN[1] 给出了一个求解(1)的广义 Newton 法

$$\boldsymbol{y}_{k+1} = \boldsymbol{y}_k - \boldsymbol{V}_k^{-1} F(\boldsymbol{y}_k), \tag{2}$$

其中 $\boldsymbol{V}_k \in \partial F(\boldsymbol{y}_k)$ 是 F 的广义 Jacobi 矩阵[2], $\partial F(\boldsymbol{y}) = \mathrm{Co}\left\{\lim_{\substack{\boldsymbol{y}_i \to \boldsymbol{y} \\ \boldsymbol{y}_i \in D_F}} \nabla F(\boldsymbol{y}_i)\right\}$, 并证明了算法的超线性收敛性·

由于许多规划问题可以转化为非光滑方程组来求解, 非光滑方程组的求解问题因而引起了广泛的关注, 并且形成了一个新的研究热点[1],[3~8]·

然而, 在实际应用中, 利用(2)求解非光滑方程组, $\boldsymbol{V}_k$ 特别难于估计· 尤其对于大型问题更是如此· 此外, 如何有效地求解广义 Newton 方程组

$$\boldsymbol{V}_k \boldsymbol{x}_k + F(\boldsymbol{y}_k) = \boldsymbol{0} \tag{3}$$

也是值得深入研究的问题·

注意到, 求解线性方程组(3)的 FOM 算法和 GMRES 算法, 只需要矩阵 $\boldsymbol{V}_k$ 与向量 $\boldsymbol{u}$ 的乘积, 不需直接计算 $\boldsymbol{V}_k$· 因此, 在非光滑情形下, 可用 $F'(\boldsymbol{y}_k, \boldsymbol{u}) \approx [F(\boldsymbol{y}_k + \delta\boldsymbol{u}) - F(\boldsymbol{y}_k)]/\delta$ 近似代替 $\boldsymbol{V}_k\boldsymbol{u}$, 结合 FOM 算法或 GMRES 算法可得到一种求解(1)的嵌套算法· 该算法一方面克服了计算 $\boldsymbol{V}_k$ 的困难, 另一方面有效地求解了(2), 因此特别适用·

已有很多作者考虑了非线性方程组的 Krylov 子空间迭代法[9~11], 但是他们只是把问题局

* 收稿日期: 2003_07_13; 修订日期: 2005_05_08
基金项目: 上海高校发展基金资助项目(214348)
作者简介: 孟泽红(1978-), 女, 河北曲阳人, 博士(E_mail: zehongmeng78@163.com);
张建军(1967—), 男(联系人. Tel: +86_21_66135519; E_mail: jjzhang@staff.shu.edu.cn)·

限在光滑的情况• 本文, 我们考虑了非光滑的情形, 实际上是对光滑情形的推广•

1 非光滑方程组的不精确牛顿法

我们首先引进非光滑问题的一些概念和性质•

定义 1.1 假如 $\partial F(\boldsymbol{y})$ 中的所有元素是非奇异的, 则称 F 在 $\boldsymbol{y}$ 点是正则的• 若在 $\boldsymbol{y}$ 的邻域内的任意一点都是正则的, 则称 F 在 $\boldsymbol{y}$ 的邻域内是正则的•

定义 1.2[1] 假如 F 是局部 Lipschitz 函数, 并且对任意的 $\boldsymbol{h} \in R^n$, $\boldsymbol{h} \neq \boldsymbol{0}$, $\lim_{h' \to h, t \downarrow 0}\left\{ \boldsymbol{V}\boldsymbol{h}' : \boldsymbol{V} \in \partial F(\boldsymbol{y} + t\boldsymbol{h}')\right\}$ 存在, 则称 F 在 $\boldsymbol{y}$ 处是半光滑的•

定义 1.3[1] 设 F 在点 $\boldsymbol{y}$ 处是半光滑的• 若对于任意的 $\boldsymbol{V} \in \partial F(\boldsymbol{y}+\boldsymbol{h})$, $\boldsymbol{h} \to \boldsymbol{0}$, 有 $\| \boldsymbol{V}\boldsymbol{h} - F'(\boldsymbol{y}, \boldsymbol{h}) \| = O(\| \boldsymbol{h} \|^2)$, 则称 F 在 $\boldsymbol{y}$ 处是 1_阶半光滑的• 若在 $\boldsymbol{y}$ 的邻域内的任意一点都是 1_阶半光滑的, 则称 F 在 $\boldsymbol{y}$ 的邻域内 1_阶半光滑•

引理 1.1[5] 假如 F 在 $\boldsymbol{y}$ 处是正则的, 则存在 $M > 0$ 使得 $M = \max\left\{ \| \boldsymbol{V}^{-1} \| : \boldsymbol{V} \in \partial F(\boldsymbol{y})\right\}$• 并且对于任意给定的 $\epsilon > 0$, 存在 $\boldsymbol{y}$ 的一个邻域 $N(\boldsymbol{y})$, 使得 F 在 $N(\boldsymbol{y})$ 内是正则的, 且对任意 $\boldsymbol{V} \in \partial F(\boldsymbol{x})$, $\boldsymbol{x} \in N(\boldsymbol{y})$, 有 $\| \boldsymbol{V}^{-1} \| \leqslant M + \epsilon$•

引理 1.2[1] 若 F 在 $\boldsymbol{y}$ 点处 1_阶半光滑, 则对于任意的 $\boldsymbol{h} \to \boldsymbol{0}$ 有

$$\| F(\boldsymbol{y}+\boldsymbol{h}) - F(\boldsymbol{y}) - F'(\boldsymbol{y}, \boldsymbol{h}) \| = O(\| \boldsymbol{h} \|^2)•$$

为研究 Newton_FOM 算法和 Newton_GMRES 算法的收敛性, 首先考虑解非光滑方程组(1)的不精确牛顿法:

$$\begin{cases} \boldsymbol{y}_{k+1} = \boldsymbol{y}_k + \boldsymbol{x}_k, \\ \boldsymbol{V}_k \boldsymbol{x}_k = -F(\boldsymbol{y}_k) + \boldsymbol{r}_k, \end{cases} \tag{4}$$

这里 $\boldsymbol{V}_k \in \partial F(\boldsymbol{y}_k)$, $\boldsymbol{r}_k$ 是残向量•

定理 1.1 $F: D \subseteq R^n \to R^n$ 是开凸集 $D \subseteq R^n$ 上的局部 Lipschitz 映射• 设 F 在 $\boldsymbol{y}^* \in D$ 处是 1_阶半光滑的和正则的, 其中 $F(\boldsymbol{y}^*) = \boldsymbol{0}$• 若迭代序列 $\left\{\boldsymbol{y}_k\right\}$, $k = 0, 1, 2, \cdots$, 满足下述不等式:

$$\| \boldsymbol{V}_k \boldsymbol{x}_k + F(\boldsymbol{y}_k) \| \leqslant \alpha \| F(\boldsymbol{y}_k) \|^2, \tag{5}$$

其中 $\boldsymbol{V}_k \in \partial F(\boldsymbol{y}_k)$, $\alpha > 0$, 则不精确牛顿法(4)局部平方收敛•

证明 因为 F 在 $\boldsymbol{y}^*$ 处是正则的, 且是局部 Lipschitz 映射, 由引理 1.1, 对任意的 $\epsilon > 0$, 存在 $\theta_1 > 0$, $M > 0$ 和 $L > 0$, 当 $\| \boldsymbol{y} - \boldsymbol{y}^* \| < \theta_1$, $\boldsymbol{V} \in \partial F(\boldsymbol{y})$ 时, 有 $\| \boldsymbol{V}^{-1} \| \leqslant M$, $\| F(\boldsymbol{y}) \| \leqslant L \| \boldsymbol{y} - \boldsymbol{y}^* \|$•

因为 F 是 1_阶半光滑的, 故存在 $\delta \in (0, \theta_1)$ 和常数 $c > 0$, 当 $\| \boldsymbol{y} - \boldsymbol{y}^* \| \leqslant \delta$, $\boldsymbol{V} \in \partial F(\boldsymbol{y})$ 时, 有

$$\| \boldsymbol{V}(\boldsymbol{y} - \boldsymbol{y}^*) - F'(\boldsymbol{y}^*, \boldsymbol{y} - \boldsymbol{y}^*) \| \leqslant \frac{c}{2} \| \boldsymbol{y} - \boldsymbol{y}^* \|^2, \tag{6}$$

$$\| F(\boldsymbol{y}) - F(\boldsymbol{y}^*) - F'(\boldsymbol{y}^*, \boldsymbol{y} - \boldsymbol{y}^*) \| \leqslant \frac{c}{2} \| \boldsymbol{y} - \boldsymbol{y}^* \|^2• \tag{7}$$

利用(6)、(7)可得

$$\begin{aligned} &\| F(\boldsymbol{y}) - F(\boldsymbol{y}^*) - \boldsymbol{V}(\boldsymbol{y} - \boldsymbol{y}^*) \| \leqslant \\ &\quad \| F(\boldsymbol{y}) - F(\boldsymbol{y}^*) - F'(\boldsymbol{y}^*, \boldsymbol{y} - \boldsymbol{y}^*) \| + \\ &\quad \| \boldsymbol{V}(\boldsymbol{y} - \boldsymbol{y}^*) - F'(\boldsymbol{y}^*, \boldsymbol{y} - \boldsymbol{y}^*) \| \leqslant \end{aligned}$$

$$c\|\boldsymbol{y}-\boldsymbol{y}^*\|^2,$$

于是, 当 $\|\boldsymbol{y}_k-\boldsymbol{y}^*\|\leqslant\delta$ 时,

$$\begin{aligned}\|\boldsymbol{y}_{k+1}-\boldsymbol{y}^*\| &= \|\boldsymbol{y}_k-\boldsymbol{y}^*+\boldsymbol{x}_k\|\leqslant\|\boldsymbol{V}_k^{-1}\|\,\|\boldsymbol{V}_k(\boldsymbol{y}_k-\boldsymbol{y}^*)+\boldsymbol{V}_k\boldsymbol{x}_k\|\leqslant\\ &\|\boldsymbol{V}_k^{-1}\|[\|\boldsymbol{V}_k\boldsymbol{x}_k+F(\boldsymbol{y}_k)\|+\|F(\boldsymbol{y}_k)-F(\boldsymbol{y}^*)-\boldsymbol{V}_k(\boldsymbol{y}_k-\boldsymbol{y}^*)\|]\leqslant\\ &M[\alpha\|F(\boldsymbol{y}_k)\|^2+c\|\boldsymbol{y}_k-\boldsymbol{y}_*\|^2]\leqslant\\ &M(\alpha L^2+c)\|\boldsymbol{y}_k-\boldsymbol{y}^*\|^2=\beta\|\boldsymbol{y}_k-\boldsymbol{y}^*\|^2,\end{aligned}$$

因此, 利用归纳法可得不精确牛顿法局部平方收敛•

2 Newton_Krylov 子空间迭代法

2.1 Newton_FOM 算法

首先考虑线性方程组

$$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b} \tag{8}$$

的求解, 其中 $\boldsymbol{A}=\boldsymbol{V}_k,\ \boldsymbol{V}_k\in\partial F(\boldsymbol{y}_k),\ \boldsymbol{b}=-F(\boldsymbol{y}_k)$• 求解(8)的 FOM 算法如下:

算法 1(FOM 算法)

步骤 1 计算 $\boldsymbol{r}_0=\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}_0$, 令 $\boldsymbol{u}_1=\boldsymbol{r}_0/\|\boldsymbol{r}_0\|_2$, 其中 $\boldsymbol{x}_0$ 是初始近似•

步骤 2 对于 $j=1,2,\cdots,m$ 计算:

$$h_{ij}=(\boldsymbol{A}\boldsymbol{u}_j,\boldsymbol{u}_i)\qquad(i=1,\cdots,j),$$

$$\boldsymbol{w}_{j+1}=\boldsymbol{A}\boldsymbol{u}_j-\sum_{i=1}^{j}h_{ij}\boldsymbol{u}_i,\quad h_{j+1,j}=\|\boldsymbol{w}_{j+1}\|_2,\quad \boldsymbol{u}_{j+1}=\boldsymbol{w}_{j+1}/h_{j+1,j}\text{•}$$

步骤 3 计算 $\boldsymbol{x}_m=\boldsymbol{x}_0+\boldsymbol{U}_m\boldsymbol{H}_m^{-1}\boldsymbol{U}_m^{\mathrm{T}}\boldsymbol{r}_0$, 其中 $\boldsymbol{U}_m=[\boldsymbol{u}_1,\cdots,\boldsymbol{u}_m]$, $\boldsymbol{H}_m$ 是非零元为 h_{ij} 的 $m\times m$ 上 Hessenberg 矩阵•

由于在算法 1 中不需要直接计算 $\boldsymbol{V}_k$, 而是需要 $\boldsymbol{V}_k\boldsymbol{u}$, 于是可以利用 $[F(\boldsymbol{y}+\sigma\boldsymbol{u})-F(\boldsymbol{y})]/\sigma$ 来近似代替 $\boldsymbol{V}_k\boldsymbol{u}$, 结合(2), 可得求解非光滑方程组(1)的 Newton_FOM 算法:

算法 2(Newton_FOM(N_F)算法)

步骤 1 给定 $\epsilon>0,\ \alpha>0$• 选择初始近似 $\boldsymbol{y}_0\in R^n$, $k=0$•

步骤 2 令 $\boldsymbol{q}_0=(F(\boldsymbol{y}_k+\sigma_0\hat{\boldsymbol{x}}_0)-F(\boldsymbol{y}_k))/\sigma_0$, $\sigma_0>0$, 其 $\hat{\boldsymbol{x}}_0$ 是线性方程组(8)的解的初始近似•

$$\hat{\boldsymbol{r}}_0=-F(\boldsymbol{y}_k)-\boldsymbol{q}_0,\quad \beta=\|\hat{\boldsymbol{r}}_0\|_2,\quad \hat{\boldsymbol{u}}_1=\hat{\boldsymbol{r}}_0/\beta,\quad \boldsymbol{q}_1=\hat{\boldsymbol{u}}_1,\qquad j=0\text{•}$$

步骤 3 $j=j+1$•

$$\boldsymbol{q}_{j+1}=(F(\boldsymbol{y}_k+\sigma_j\hat{\boldsymbol{u}}_j)-F(\boldsymbol{y}_k))/\sigma_j,\quad \sigma_j>0,\quad \hbar_{ij}=(\boldsymbol{q}_{j+1},\hat{\boldsymbol{u}}_i),\qquad i=1,\cdots,j,$$

$$\hat{\boldsymbol{w}}_{j+1}=\boldsymbol{q}_{j+1}-\sum_{i=1}^{j}\hbar_{ij}\hat{\boldsymbol{u}}_i,\quad \hbar_{j+1,j}=\|\hat{\boldsymbol{w}}_{j+1}\|_2,\quad \hat{\boldsymbol{u}}_{j+1}=\hat{\boldsymbol{w}}_{j+1}/\hbar_{j+1,j}\text{•}$$

令 $\boldsymbol{H}_j$ 是非零元为 $\hbar_{ij}$ 的 $j\times j$ 的上 Hessenberg 矩阵, $\boldsymbol{U}_j=[\hat{\boldsymbol{u}}_1,\cdots,\hat{\boldsymbol{u}}_j]$• 计算 $\hat{\boldsymbol{x}}_j=\hat{\boldsymbol{x}}_0+\boldsymbol{U}_j\boldsymbol{H}_j^{-1}\boldsymbol{U}_j^{\mathrm{T}}\hat{\boldsymbol{r}}_0$, $\rho_j=\|\hat{\boldsymbol{r}}_j\|_2=\|F(\boldsymbol{y}_k)+(F(\boldsymbol{y}_k+\sigma_j\hat{\boldsymbol{x}}_j)-F(\boldsymbol{y}_k))/\sigma_j\|_2$•

步骤 4 若 $\rho_j\leqslant\alpha\|F(\boldsymbol{y}_k)\|_2^2$ 或 $j=n$, 则令 $m=j$ 且做下一步, 否则转步骤 3•

步骤 5 令 $\boldsymbol{y}_{k+1}=\boldsymbol{y}_k+\boldsymbol{x}_m$•

步骤 6 若 $\|F(\boldsymbol{y}_{k+1})\|_2\leqslant\epsilon$, 结束• 否则, $k=k+1$ 转步骤 2•

2.2 算法的收敛性分析

为了书写方便, 省去下标 k• 令 $\varepsilon_i=\boldsymbol{q}_{i+1}-\boldsymbol{A}\hat{\boldsymbol{u}}_i$, $i=1,\cdots,m$• 定义 $n\times n$ 矩阵 $\boldsymbol{E}_m$ 为 $\boldsymbol{E}_m$

$= \varepsilon^m \boldsymbol{U}_m^{\mathrm{T}}$, $\varepsilon^m = [\varepsilon_1, \cdots, \varepsilon_m] \in R^{n\times m}$, 于是有

$$(\boldsymbol{A} + \boldsymbol{E}_m)\hat{\boldsymbol{u}}_i = \boldsymbol{A}\hat{\boldsymbol{u}}_i + \varepsilon^m \boldsymbol{U}_m^{\mathrm{T}}\hat{\boldsymbol{u}}_i = \boldsymbol{A}\hat{\boldsymbol{u}}_i + \varepsilon_i = \boldsymbol{q}_{i+1} \quad (i = 1, \cdots, m).$$

引理 2.1 设 $F: D \subseteq R^n \to R^n$ 是局部 Lipschitz 映射. 设 $F(\boldsymbol{y})$ 在 $\boldsymbol{y}$ 处 1_阶半光滑. 则存在 $\gamma > 0$, $\delta > 0$, 使得当 $|\sigma_0| \ \|\hat{\boldsymbol{x}}_0\|_2 < \delta$ 且 $\boldsymbol{y} + \sigma_0\hat{\boldsymbol{x}}_0 \in D$ 时, 有 $\|\varepsilon_0\|_2 = \|\boldsymbol{q}_0 - \boldsymbol{A}\hat{\boldsymbol{x}}_0\|_2 \leqslant (\gamma/2)|\sigma_0| \ \|\hat{\boldsymbol{x}}_0\|_2^2$ 和

$$\rho_m \leqslant \left(\frac{\gamma}{2}\right)|\sigma_0| \ \|\hat{\boldsymbol{x}}_0\|_2^2 + \rho_m + \left(\frac{\gamma}{2}\right)\|\sigma^m\|_2\|\hat{\boldsymbol{z}}_m\|_2, \tag{9}$$

其中 $\rho_m = \|\hat{\boldsymbol{r}}_0 - (\boldsymbol{A} + \boldsymbol{E}_m)\hat{\boldsymbol{z}}_m\|_2$, $\rho_m = \|\boldsymbol{b} - \boldsymbol{A}\hat{\boldsymbol{x}}_m\|_2$.

证明 由于 $F(\boldsymbol{y})$ 在 $\boldsymbol{y}$ 处是 1_阶半光滑的, 所以存在 $\delta > 0$, 当 $\|h\|_2 < \delta$ 时有

$$\|\boldsymbol{Vh} - F'(\boldsymbol{y}, \boldsymbol{h})\|_2 = O(\|\boldsymbol{h}\|_2^2), \tag{10}$$

$$\|F(\boldsymbol{y} + \boldsymbol{h}) - F(\boldsymbol{y}) - F'(\boldsymbol{y}, \boldsymbol{h})\|_2 = O(\|\boldsymbol{h}\|_2^2), \tag{11}$$

所以存在 $\gamma > 0$, 当 $\|\boldsymbol{h}\|_2 < \delta$ 时,

$$\begin{aligned}&\|F(\boldsymbol{y} + \boldsymbol{h}) - F(\boldsymbol{y}) - \boldsymbol{Vh}\|_2 \leqslant \\ &\quad \|F(\boldsymbol{y} + \boldsymbol{h}) - F(\boldsymbol{y}) - F'(\boldsymbol{y}, \boldsymbol{h})\|_2 + \|\boldsymbol{Vh} - F'(\boldsymbol{y}, \boldsymbol{h})\|_2 \leqslant \\ &\quad (\gamma/2)\|\boldsymbol{h}\|_2^2,\end{aligned} \tag{12}$$

令 $\varepsilon_0 = \boldsymbol{q}_0 - \boldsymbol{A}\hat{\boldsymbol{x}}_0$. 因为 $|\sigma_0| \ \|\hat{\boldsymbol{x}}_0\|_2 < \delta$, 所以由(12)有

$$\|\varepsilon_0\|_2 = \|\boldsymbol{q}_0 - \boldsymbol{A}\hat{\boldsymbol{x}}_0\|_2 \leqslant (\gamma/2) \times |\sigma_0| \ \|\hat{\boldsymbol{x}}_0\|_2^2.$$

记 $\hat{\boldsymbol{z}}_m = \boldsymbol{U}_m\boldsymbol{H}_m^{-1}\boldsymbol{U}_m^{\mathrm{T}}\hat{\boldsymbol{r}}_0 = \beta\boldsymbol{U}_m\boldsymbol{H}_m^{-1}e_1$, 其中 $\beta = \|\hat{\boldsymbol{r}}_0\|_2$, 则由算法 2 可得 $\hat{\boldsymbol{x}}_m = \hat{\boldsymbol{x}}_0 + \hat{\boldsymbol{z}}_m$,

$$\hat{\boldsymbol{r}}_m = \boldsymbol{b} - \boldsymbol{A}\hat{\boldsymbol{x}}_m = \boldsymbol{b} - \boldsymbol{A}(\hat{\boldsymbol{x}}_0 + \hat{\boldsymbol{z}}_m) = \varepsilon_0 + (\hat{\boldsymbol{r}}_0 - (\boldsymbol{A} + \boldsymbol{E}_m)\hat{\boldsymbol{z}}_m) + \boldsymbol{E}_m\hat{\boldsymbol{z}}_m,$$

$$\begin{aligned}\rho_m = \|\hat{\boldsymbol{r}}_m\|_2 &\leqslant \|\varepsilon_0\|_2 + \|\hat{\boldsymbol{r}}_0 - (\boldsymbol{A} + \boldsymbol{E}_m)\hat{\boldsymbol{z}}_m\|_2 + \|\boldsymbol{E}_m\hat{\boldsymbol{z}}_m\|_2 \leqslant \\ &(\gamma/2)|\sigma_0| \ \|\hat{\boldsymbol{x}}_0\|_2^2 + \|\hat{\boldsymbol{r}}_0 - (\boldsymbol{A} + \boldsymbol{E}_m)\hat{\boldsymbol{z}}_m\|_2 + \|\boldsymbol{E}_m\hat{\boldsymbol{z}}_m\|_2,\end{aligned}$$

因为 F 是 1_阶半光滑的, 所以由(12), 当 $(|\sigma_1|^2 + \cdots + |\sigma_m|^2)^{(1/2)} < \delta$ 时, 有

$$\|F(\boldsymbol{y} + \sigma_i\hat{\boldsymbol{u}}_i) - F(\boldsymbol{y}) - \sigma_i\boldsymbol{V}_k\hat{\boldsymbol{u}}_i\|_2 \leqslant (\gamma/2)\sigma_i^2\|\hat{\boldsymbol{u}}_i\|_2^2 \quad (i = 1, \cdots, m),$$

由于 $\|\hat{\boldsymbol{u}}_i\|_2 = 1$, 所以有

$$\|\varepsilon_i\|_2 = \|\boldsymbol{q}_{i+1} - \boldsymbol{A}\hat{\boldsymbol{u}}_i\|_2 \leqslant (\gamma/2)|\sigma_i|,$$

$$\begin{aligned}\|\varepsilon^m\|_2 &\leqslant \|\varepsilon^m\|_F = (\|\varepsilon_1\|_2^2 + \cdots + \|\varepsilon_m\|_2^2)^{1/2} \leqslant \\ &(\gamma/2)(\sigma_1^2 + \cdots + \sigma_m^2)^{1/2} = (\gamma/2)\|\sigma^m\|_2,\end{aligned}$$

其中 $\sigma^m = (\sigma_1, \cdots, \sigma_m)^{\mathrm{T}} \in R^m$. 于是

$$\begin{aligned}\rho_m &\leqslant (\gamma/2)|\sigma_0| \ \|\hat{\boldsymbol{x}}_0\|_2^2 + \rho_m + \|\varepsilon^m\|_2\|\boldsymbol{V}_m^{\mathrm{T}}\|_2\|\hat{\boldsymbol{z}}_m\|_2 \leqslant \\ &(\gamma/2)|\sigma_0| \ \|\hat{\boldsymbol{x}}_0\|_2^2 + \rho_m + (\gamma/2)\|\sigma^m\|_2\|\hat{\boldsymbol{z}}_m\|_2.\end{aligned}$$

引理 2.2[9] 设 $\boldsymbol{A}$ 非奇异, 若在算法 1 中 $\boldsymbol{w}_{m+1} = \boldsymbol{0}$, 则 $\boldsymbol{H}_m^{-1}$ 存在且 $\boldsymbol{z}_m$ 是 $\boldsymbol{Az} = \boldsymbol{r}_0$ 的精确解.

定理 2.1 $F: D \subseteq R^n \to R^n$ 是局部 Lipschitz 映射, 假设 F 在 $\boldsymbol{y}$ 处是 1_阶半光滑的, 且是正则的. 设 $\hat{\boldsymbol{x}}_0$ 是 $\boldsymbol{x}^*$ 的初始近似, 其中 $\boldsymbol{x}^* = \boldsymbol{A}^{-1}\boldsymbol{b}$. 给定 $\alpha > 0$, 选 β, δ 足够小, 使得

$$\delta\gamma\|\boldsymbol{A}^{-1}\|_2 < 1, \tag{13}$$

$$\begin{aligned}&(\beta\gamma/2)\|\hat{\boldsymbol{x}}_0\|_2^2 + \delta\gamma\|\boldsymbol{A}^{-1}\|_2(\|\boldsymbol{b} - \boldsymbol{A}\hat{\boldsymbol{x}}_0\|_2 + (\beta\gamma/2)\|\hat{\boldsymbol{x}}_0\|_2^2) \leqslant \\ &\quad \frac{1}{2}\alpha\|F(\boldsymbol{y})\|_2^2,\end{aligned} \tag{14}$$

$$\boldsymbol{y} + \boldsymbol{v} \in D, \ \|\boldsymbol{v}\|_2 \leqslant \delta, \qquad \forall \boldsymbol{v} \in R^n, \tag{15}$$

选 $\sigma^n = (\sigma_1, \ldots, \sigma_n)^{\mathrm{T}}$, 使 $\|\sigma^n\|_2 \leqslant \delta$, $|\sigma_0| \leqslant \beta$, $|\sigma_0| \|\hat{\boldsymbol{x}}_0\|_2 \leqslant \delta$, 则存在 $m \in \{0, \ldots, n\}$ 使得

$$\rho_m = \|\boldsymbol{b} - \boldsymbol{A}\hat{\boldsymbol{x}}_m\|_2 \leqslant \alpha \|F(\boldsymbol{y})\|_2^2.$$

证明　由假设条件(13)

$$\|\boldsymbol{A}^{-1}\boldsymbol{E}_m\|_2 \leqslant \|\boldsymbol{A}^{-1}\|_2 \|\boldsymbol{E}_m\|_2 \leqslant \|\boldsymbol{A}^{-1}\|_2 \|\varepsilon^m\|_2 \leqslant \|\boldsymbol{A}^{-1}\|_2 \frac{\delta\gamma}{2} < \frac{1}{2},$$

所以 $\boldsymbol{A} + \boldsymbol{E}_m = \boldsymbol{A}(\boldsymbol{I} + \boldsymbol{A}^{-1}\boldsymbol{E}_m)$ 非奇异, 且

$$\|(\boldsymbol{A} + \boldsymbol{E}_m)^{-1}\|_2 \leqslant \frac{\|\boldsymbol{A}^{-1}\|_2}{1 - \|\boldsymbol{A}^{-1}(\boldsymbol{A} - (\boldsymbol{A} + \boldsymbol{E}_m))\|_2} \leqslant$$
$$\frac{\|\boldsymbol{A}^{-1}\|_2}{1 - \|\boldsymbol{A}^{-1}E_m\|_2} \leqslant 2\|\boldsymbol{A}^{-1}\|_2.$$

记 $\hat{\boldsymbol{r}}_0 = \boldsymbol{b} - \boldsymbol{q}_0$, 则由引理2.1

$$\|\hat{\boldsymbol{r}}_0\|_2 = \|\boldsymbol{b} - \boldsymbol{q}_0\|_2 = \|(\boldsymbol{b} - \boldsymbol{A}\hat{\boldsymbol{x}}_0) - \varepsilon_0\|_2 \leqslant$$
$$\|\boldsymbol{b} - \boldsymbol{A}\hat{\boldsymbol{x}}_0\|_2 + \frac{\gamma}{2}|\sigma_0| \|\hat{\boldsymbol{x}}_0\|_2^2,$$

设 $\boldsymbol{z}_m^*$ 为 $(\boldsymbol{A} + \boldsymbol{E}_m)\boldsymbol{z} = \hat{\boldsymbol{r}}_0$ 的解, 则

$$\|\boldsymbol{z}_m^*\|_2 = \|(\boldsymbol{A} + \boldsymbol{E}_m)^{-1}\hat{\boldsymbol{r}}_0\|_2 \leqslant 2\|\boldsymbol{A}^{-1}\|_2 \|\hat{\boldsymbol{r}}_0\|_2,$$
$$\hat{\boldsymbol{z}}_m - \boldsymbol{z}_m^* = (\boldsymbol{A} + \boldsymbol{E}_m)^{-1}[(\boldsymbol{A} + \boldsymbol{E}_m)\hat{\boldsymbol{z}}_m - \hat{\boldsymbol{r}}_0],$$
$$\|\hat{\boldsymbol{z}}_m\|_2 \leqslant 2\|\boldsymbol{A}^{-1}\|_2 \|\hat{\boldsymbol{r}}_0\|_2 + 2\|\boldsymbol{A}^{-1}\|_2 \rho_m = 2\|\boldsymbol{A}^{-1}\|_2 (\|\hat{\boldsymbol{r}}_0\|_2 + \rho_m),$$

其中 $\rho_m = \|\hat{\boldsymbol{r}}_0 - (\boldsymbol{A} + \boldsymbol{E}_m)\hat{\boldsymbol{z}}_m\|_2$. 所以由引理2.1

$$\rho_m \leqslant \frac{\gamma}{2}|\sigma_0| \|\hat{\boldsymbol{x}}_0\|_2^2 + \rho_m + \frac{\gamma}{2}\|\sigma^m\|_2 \|\hat{\boldsymbol{z}}_m\|_2 \leqslant$$
$$\frac{\gamma}{2}|\sigma_0| \|\hat{\boldsymbol{x}}_0\|_2^2 + \rho_m + \frac{\delta\gamma}{2}\|\boldsymbol{A}^{-1}\|_2 (\|\hat{\boldsymbol{r}}_0\|_2 + \rho_m) \leqslant$$
$$\frac{\beta\gamma}{2}\|\hat{\boldsymbol{x}}_0\|_2 + \delta\gamma\|\boldsymbol{A}^{-1}\|_2 (\|\boldsymbol{b} - \boldsymbol{A}\hat{\boldsymbol{x}}_0\|_2 + \frac{\beta\gamma}{2}\|\hat{\boldsymbol{x}}_0\|_2^2) + \rho_m(1 + \delta\gamma\|\boldsymbol{A}^{-1}\|_2) \leqslant$$
$$\frac{1}{2}\alpha\|F(\boldsymbol{y})\|_2^2 + \rho_m(1 + \delta\gamma\|\boldsymbol{A}^{-1}\|_2),$$

下证

$$\rho_m(1 + \delta\gamma\|\boldsymbol{A}^{-1}\|_2) \leqslant \frac{1}{2}\alpha\|F(\boldsymbol{y})\|_2^2. \tag{16}$$

1) 若 $\hat{\boldsymbol{r}}_0 = \boldsymbol{0}$, 则 $\boldsymbol{b} = \boldsymbol{q}_0$,

$$\rho_0 = \|\boldsymbol{b} - \boldsymbol{A}\hat{\boldsymbol{x}}_0\|_2 = \|\boldsymbol{q}_0 - \boldsymbol{A}\hat{\boldsymbol{x}}_0\|_2 = \|\varepsilon_0\|_2 \leqslant$$
$$\frac{\gamma}{2}|\sigma_0| \|\hat{\boldsymbol{x}}_0\|_2^2 \leqslant \frac{\gamma\beta}{2}\|\hat{\boldsymbol{x}}_0\|_2^2 \leqslant \frac{1}{2}\alpha\|F(\boldsymbol{y})\|_2^2,$$

结论成立.

2) 若 $\hat{\boldsymbol{r}}_0 \neq \boldsymbol{0}$, $\hat{\boldsymbol{u}}_1 = \hat{\boldsymbol{r}}_0 / \|\hat{\boldsymbol{r}}_0\|_2$, $\boldsymbol{E}_1 = \varepsilon_1\hat{\boldsymbol{u}}_1^{\mathrm{T}}$, 其中 $\varepsilon_1 = \boldsymbol{q}_2 - \boldsymbol{A}\hat{\boldsymbol{u}}_1$, 由以上证明可知, $\boldsymbol{A} + \boldsymbol{E}_1$ 非奇异. 所以 $\boldsymbol{q}_2 = (\boldsymbol{A} + \boldsymbol{E}_1)\hat{\boldsymbol{u}}_1 \neq \boldsymbol{0}$, 于是得到 $\hat{\boldsymbol{w}}_2$.

(a) 若 $\hat{\boldsymbol{w}}_2 = \boldsymbol{0}$, 则 $\hbar_{21} = 0$, 由引理2.2得 $\boldsymbol{H}_1 = \hat{\boldsymbol{u}}_1^{\mathrm{T}}(\boldsymbol{A} + \boldsymbol{E}_1)\hat{\boldsymbol{u}}_1 = \hbar_{11} \neq 0$, $\hat{\boldsymbol{z}}_1 = \|\hat{\boldsymbol{r}}_0\|_2 \hbar_{11}^{-1}\hat{\boldsymbol{u}}_1$ 是 $(\boldsymbol{A} + \boldsymbol{E}_1)\boldsymbol{z} = \hat{\boldsymbol{r}}_0$ 的精确解. 即 $\rho_1 = \|\hat{\boldsymbol{r}}_0 - (\boldsymbol{A} + \boldsymbol{E}_1)\hat{\boldsymbol{z}}_1\|_2 = \boldsymbol{0}$, 所以(16)成立.

(b) 若 $\hat{\boldsymbol{w}}_2 \neq \boldsymbol{0}$, 得 $\hat{\boldsymbol{u}}_2 \neq \boldsymbol{0}$, 此时得到 $\boldsymbol{x}_1$, 若 $\boldsymbol{x}_1$ 满足算法2的步骤4, 则定理结论成立, 否则定义 $\boldsymbol{E}_2 = \varepsilon^2\boldsymbol{V}_2$, 这里 $\varepsilon^2 = [\varepsilon_1, \varepsilon_2]$, $\varepsilon_2 = \boldsymbol{q}_3 - \boldsymbol{A}\hat{\boldsymbol{u}}_2$, 同理可得 $\boldsymbol{A} + \boldsymbol{E}_2$ 非奇异, 因此 $\boldsymbol{q}_3 = (\boldsymbol{A}$

$+ E_2)\hat{u}_2 \neq \mathbf{0}$, 得到 $\hat{w}_3$•

一般地, 若 $U_m = [\hat{u}_1, \cdots, \hat{u}_m]$, $\hat{w}_{m+1}$ 由算法 2 确定, 则 $A + E_m$ 非奇异•

若 $\hat{w}_{m+1} = \mathbf{0}$, 则 H_m^{-1} 存在, 且 $\hat{z}_m$ 是精确解, 这样 $\rho_m = 0$, 则(16) 成立•

若 $\hat{w}_i \neq \mathbf{0}$, $i = 1, \cdots, n$, 且存在某个 i 满足算法 2 中步骤 4, 则定理结论成立; 否则 $\hat{z}_n$ 是 $(A + E_n)z = \hat{r}_0$ 的精确解, 所以 $\rho_n = 0$•

定理 2.2　设 $F: R^n \to R^n$ 是开集 $D \subseteq R^n$ 上局部 Lipschitz 映射, 设 F 在 y^* 的邻域内 1_阶半光滑, 且是正则的, 其中 $F(y^*) = \mathbf{0}$•　则由 Newton_FOM 算法产生的序列 $\{y_k\}$ 局部平方收敛于 y^*•

证明　对任意的 y_0, $\|y_0 - y^*\|_2 \leqslant \theta_1$, 其中 θ_1 满足定理 1.1 的条件•　选取 δ_0 充分小使得 $\delta_0\gamma\|A^{-1}\|_2 < \min\{1, (\alpha\|F(y_0)\|_2)/2\}$, $y_0 + v \in D$, $\|v\|_2 \leqslant \delta_0$ 成立•　选 $\sigma^n = (\sigma_1, \cdots, \sigma_n)^{\mathrm{T}} \in R^n$, 使得 $\|\sigma^n\|_2 < \delta_0$•　用 $\hat{x}_{0,m}$ 表示由算法 2 计算得到的第 m 步时对 x_0 的近似•为了简单起见, 设 $\hat{x}_{0,0} = \mathbf{0}$, 则由定理 2.1 知, 存在 $m_0 \in \{1, \cdots, n\}$ 使得 $\|b - A\hat{x}_{0,m_0}\|_2 \leqslant \alpha\|F(y_0)\|_2^2$•　定义 $x_0 = \hat{x}_{0,m_0}$, 则 $y_1 = y_0 + x_0$, 利用定理 1.1 知

$$\|y_1 - y^*\|_2 \leqslant \beta\|y_0 - y^*\|_2^2,$$

其中 $\beta > 0$•　当 $k = n$ 时, 可类似地证明

$$\|y_k - y^*\|_2 \leqslant \beta\|y_{k-1} - y^*\|_2^2,$$

由数学归纳法可得 $\{y_k\}$ 局部平方收敛于 y^*•

2.3　Newton_GMRES 算法

与 FOM 算法相似, 求解非光滑方程组(1) 的 Newton_GMRES 算法为:

算法 3(Newton_GMRES(N_G) 算法)

步骤 1 给定 $\epsilon > 0$, $\alpha > 0$•　选择初始近似 $y_0 \in R^n$, 令 $k = 0$•

步骤 2 令 $q_0 = (F(y_k + \sigma_0\hat{x}_0) - F(y_k))/\sigma_0$, $\sigma_0 > 0$, 其中 $\hat{x}_0$ 是线性方程组(8) 的解的初始近似•　$\hat{r}_0 = -F(y_k) - q_0$, $\beta = \|\hat{r}_0\|_2$, $\hat{u}_1 = \hat{r}_0/\beta$, $q_1 = \hat{u}_1$, $j = 0$•

步骤 3 $j = j + 1$•

$$q_{j+1} = (F(y_k + \sigma_j\hat{u}_j) - F(y_k))/\sigma_j,\quad \sigma_j > 0,\quad \hbar_{ij} = (q_{j+1}, \hat{u}_i),\qquad i = 1, \cdots, j,$$

$$\hat{w}_{j+1} = q_{j+1} - \sum_{i=1}^{j}\hbar_{ij}\hat{u}_i,\quad \hbar_{j+1,j} = \|\hat{w}_{j+1}\|_2,\quad \hat{u}_{j+1} = \hat{w}_{j+1}/\hbar_{j+1,j}•$$

定义 $(j + 1) \times j$ Hessenberg 矩阵 $\hat{H}_j$, 其非零元为 $\hbar_{ij}$, $U_j = [\hat{u}_1, \hat{u}_2, \cdots, \hat{u}_j]$•　求解极小化问题

$$\min_{d \in R^j}\|\beta e_1 - \hat{H}_j d\|_2 \tag{17}$$

其中 $e_1 = (1, 0, \cdots, 0)^{\mathrm{T}}$ 是 $j + 1$ 维单位向量, 记 d_j 为(17) 的解•　计算 $x_j = x_0 + U_j d_j$, $\rho_j = \|\hat{r}_j\|_2 = \|F(y_k) + (F(y_k + \sigma_j\hat{x}_j) - F(y_k))/\sigma_j\|_2$•

步骤 4 若 $\rho_j \leqslant \alpha\|F(y_k)\|_2^2$ 或 $j = n$, 则令 $m = j$ 且做下一步, 否则转步骤 3•

步骤 5 令 $y_{k+1} = y_k + x_m$•

步骤 6 若 $\|F(y_{k+1})\|_2 \leqslant \epsilon$, 结束•　否则, $k = k + 1$, 转步骤 2•

类似于定理 2.2, 可证明:

定理 2.3　设 $F: R^n \to R^n$ 是开集 $D \subseteq R^n$ 上局部 Lipschitz 映射, 设 F 在 y^* 的邻域内 1_阶

半光滑，且是正则的，其中 $F(\boldsymbol{y}^*)=\mathbf{0}$• 则由Newton_GMRES算法产生的序列$\{\boldsymbol{y}_k\}$局部平方收敛于 $\boldsymbol{y}^*$•

3 数值例子

考虑非线性互补问题：求 $\boldsymbol{y}\in R^n$ 使得

$$\boldsymbol{y}\geqslant\mathbf{0},\ H(\boldsymbol{y})\geqslant\mathbf{0},\ \boldsymbol{y}^{\mathrm{T}}H(\boldsymbol{y})=\mathbf{0}, \tag{18}$$

可以证明[4]，问题(18)等价于下述非线性方程组：

$$F(\boldsymbol{y})=\min(\boldsymbol{y},H(\boldsymbol{y}))=\mathbf{0} \tag{19}$$

的求解问题•

方程组(19)是一个半光滑方程组，因此可利用求解非光滑方程组的算法来求解互补问题• 下面，基于等价非光滑方程组(19)，用算法2和算法3求解两个互补问题• 在数值计算中，我们取 $\epsilon=10^{-6}$, $\sigma_m=10^{-4}(\|\boldsymbol{y}_k\|_2/\|\boldsymbol{u}_m\|_2)$, $\alpha=1/2$•

例3.1 非线性互补问题

$H(\boldsymbol{y})=c(\boldsymbol{y})+\boldsymbol{A}\boldsymbol{y}+\boldsymbol{b}$，其中 $\boldsymbol{b}=(-n/2,-n/2+1,\cdots,-n/2+n-1)^{\mathrm{T}}$, $c(\boldsymbol{y})=(c_1(y_1),c_2(y_2),\cdots,c_n(y_n))^{\mathrm{T}}$, $c_i(y_i)=10\times\arctan(y_i)$, $i=1,2,\cdots,n$,

$$\boldsymbol{A}=\begin{pmatrix}2.5 & -1 & & & \\ -1 & 2.5 & -1 & & \\ & & \ddots & \ddots & \\ & & -1 & 2.5 & -1 \\ & & & -1 & 2.5\end{pmatrix}•$$

对不同的规模和初值，我们计算了上述问题• 计算结果见表1，其中 $\boldsymbol{y}_0$ 是给定的初始值，D_n 是问题的维数•

表1 不同规模问题的迭代步数

$\boldsymbol{y}_0$	(1, …, 1)		(10, …, 10)		(100, …, 100)		(1, 1, 0, …, 0, 1, 1)		(1 000, …, 1 000)	
	N_F	N_G	N_F	N_G	N_F	N_G	N_F	N_G	N_F	N_G
$D_n=50$	6	6	8	8	18	18	7	7	18	18
$D_n=100$	6	6	8	8	18	18	7	7	18	18
$D_n=200$	8	7	8	8	18	18	7	7	18	18
$D_n=500$	9	8	9	10	19	19	7	8	20	19

特别对于100维问题，取初值为(1, …, 1)，(1 000, …, 1000)，图1，图2显示了 $\lg(\|F(\boldsymbol{y})\|_2)$ 与迭代步数的关系•

例3.2 线性互补问题

$H(\boldsymbol{y})=\boldsymbol{M}\boldsymbol{y}+\boldsymbol{q}$，其中 $\boldsymbol{q}=(-1,0,\cdots,-1)^{\mathrm{T}}$,

$$\boldsymbol{M}=\begin{pmatrix}4 & -1 & & & \\ -1 & 4 & -1 & & \\ & & \ddots & \ddots & \\ & & -1 & 4 & -1 \\ & & & -1 & 4\end{pmatrix}•$$

表2给出了不同规模的迭代步数，其中 $\boldsymbol{y}_0$ 是给定的初始值，D_n 是问题的维数，“*”表示不

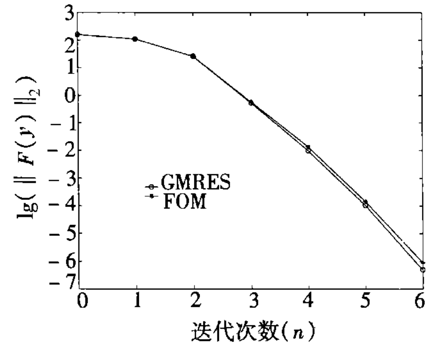

图 1 $D_n = 100,\ y_0 = (1, \cdots, 1)$

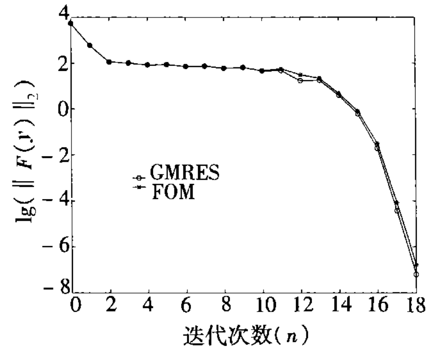

图 2 $D_n = 100,\ y_0 = (1\,000, \cdots, 1\,000)$

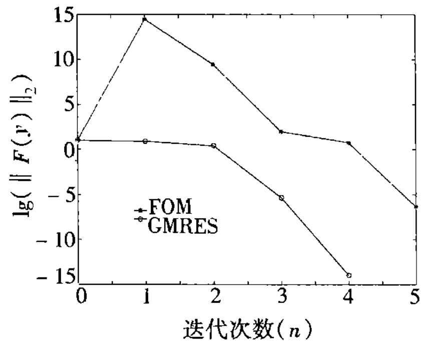

图 3 $D_n = 100,\ y_0 = (1, \cdots, 1)$

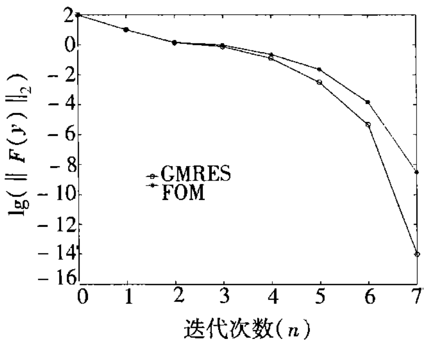

图 4 $D_n = 100,\ y_0 = (10, \cdots, 10)$

收敛.

表 2 不同规模问题的迭代步数

y_0	(1, …, 1)		(10, …, 10)		(100, …, 100)		(1, 1, 0, …, 0, 1, 1)		(1 000, …, 1 000)	
	N_F	N_G	N_F	N_G	N_F	N_G	N_F	N_G	N_F	N_G
$D_n = 50$	*	3	7	7	6	6	3	3	6	6
$D_n = 100$	5	4	7	7	36	6	3	3	6	6
$D_n = 200$	20	4	7	6	26	*	3	3	6	6
$D_n = 500$	5	4	7	6	70	6	3	3	6	6

特别对于 100 维问题, 取初值为(1, …, 1), (10, …, 10), 图 3, 图 4 显示了例 3.2 的 $\lg(\|F(y)\|_2)$ 与迭代步数的关系.

[参 考 文 献]

[1] QI Li_qun, SUN Ji_e. A nonsmooth version of Newton's method[J]. Mathematical Programming, 1993, **58**(3): 353—367.

[2] Clarke Frank H. Optimization and Nonsmooth Analysis[M]. New York: Wiley, 1983, 69—70.

[3] Harker Patrick T, XIAO Bai_chun. Newton's method for the nonlinear complementarity problem: a B_differentiable equation approach[J]. Mathematical Progamming, 1990, **48**(3): 339—357.

[4] IP Chi_ming, Kyparisis Jerzy. Local convergence of quasi_Newton methods for B_differentiable equations[J]. Mathematical Progamming, 1992, **56**(1): 71—89.

[5] Martinez Jos mario, QI Li_qun. Inexact Newton methods for solving nonsmooth equations[J]. Journal of Computational and Applied Mathematics, 1995, **60**(1/2): 127—145.

[6] PANG Jong_shi, QI Li_qun. Nonsmooth equations: motivation and algorithms[J]. SIAM Journal on Optimization, 1993, **3**(2): 443—465.

[7] PANG Jong_shi. Newton's method for B_differentiable equations[J]. Mathematical of Operations Research, 1990, **15**(2): 311—341.

[8] QI Li_qun. Convergence analysis of some algorithms for solving nonsmooth equations[J]. Mathematical of Operations Research, 1993, **18**(1): 227—244.

[9] Brown Peter N. Local convergence theory for combined inexact Newton/finite_difference projection methods[J]. SIAM Journal on Numerical Analysis, 1987, **24**(2): 407—433.

[10] Brown Peter N, Saad Youcef. Convergence theory of nonlinear Newton_Krylov algorithms[J]. SIAM Journal on Optimization, 1994, **4**(2): 297—330.

[11] Brown Peter N, Saad Youcef. Hybrid Krylov methods for nonlinear systems of equations[J]. SIAM Journal on Scientific Computing, 1990, **11**(3): 450—481.

Nonlinear Krylov Subspace Methods for Solving Nonsmooth Equations

MENG Ze_hong, ZHANG Jian_jun

(Department of Mathematics, Shanghai University, Shanghai 200444, P.R.China)

Abstract: Newton_FOM algorithm and Newton_GMRES algorithm for solving nonsmooth equations are presented. It is proved that these Krylov subspace algorithms have locally quadratic convergence. Numerical experiments demonstrate the effectiveness of the algorithms.

Key words: nonsmooth equations; Newton_FOM algorithm; Newton_GMRES algorithm