第9卷 第1期
1987年2月

数学研究与评论
JOURNAL OF MATHEMATICAL RESEARCH AND EXPOSITION

Vol.9 No.1
Feb., 1989

关于Lemaréchal高阶算法的研究
——概念性算法

黄婉珍 张连生
（上海科技大学应用数学系）

摘要 对于非光滑的极小化问题，C.Lemaréchel在〔1〕中对凸函数的无约束极小化问题提示了一个高阶σ-牛顿型算法的思想，并讨论了某些性质．本文对〔1〕的高阶σ-牛顿型算法作了进一步研究，并提出一个概念性算法，证明了算法的全局收敛性．

在实际问题中会遇到一类非光滑函数的极小化问题，目前对于非光滑函数的算法．已有相当研究．C.Lemarèchl〔1〕对凸函数的无约束极小化问题提出了一个高阶σ-牛顿型算法的思想，并讨论了某些性质．本文对〔1〕的高阶σ-牛顿型算法作了进一步研究，并提出一个概念性算法，证明了算法的全局收敛性．

在讨论之前，我们先给出有关的概念和命题．

一、次梯度与σ-扰动次梯度

设f：$R^n\to R$是 欧氏空间中的一个实值凸函数，对$x\in R^n$，定义f在x处的次梯度为：

$$\partial f(x)=\{x^*\in R^n:\langle x^*,y-x\rangle\le f(y)-f(x),\ \forall y\in R^n\}$$

对$\forall d\in R^n$，定义f在x处沿方向d的方向导数为：

$$f'(x,d)=\max\{\langle x^*,d\rangle,\ x^*\in\partial f(x)\}$$

在我们所讨论的算法中，使用的是σ-扰动次梯度，即对$\sigma\ge 0$，（其中当$\sigma=0$时即为上面定义的次梯度），定义f在x处的σ-扰动次梯度为：

$$\partial_\sigma f(x)=\{x^*\in R^n:\langle x^*,y-x\rangle\le f(y)-f(x)+\sigma,\ \forall y\in R^n\}.$$

相应地定义σ-扰动方向导数为：

$$f'_\sigma(x,d)=\inf\{\frac{f(x+td)-f(x)+\sigma}{t},\ t>0\}\ 或 f'_\sigma(x,d)=\max\{\langle x^*,d\rangle,\ x^*\in\partial_\sigma f(x)\}.$$

命题1.1 对$\forall\sigma\ge 0$，$x\in R^n$，$\partial_\sigma f(x)$是R^n中的一个非空凸紧集．

命题1.2 $\partial_\sigma f(x)$关于(x,σ)是上半连续的．

证明 设$x_i\to x$，$\sigma_i\to\sigma$，$y_i\to y$，且$y_i\in\partial_{\sigma_i}f(x_i)$，即对$\forall z\in R^n$，有$\langle y_i,z-x_i\rangle\le f(z)-f(x_i)+\sigma_i$．令$i\to\infty$，$\langle y,z-x\rangle\le f(z)-f(x)+\sigma$，即$y\in\partial_\sigma f(x)$，即$\partial_\sigma f(x)$关于$(x,\sigma)$是上半连续的．

命题1.3 对R^n中任一有界闭集S，$\partial_\sigma f(x)$在S上关于x是一致有界的．

证明 对S中任一点x^0：由于f是局部Lip连续的，$\exists\mu(x^0)>0$，$L(x^0)>0$，使对一切x'，$x''\in B(x^0,\mu(x^0))$，有

$$|f(x')-f(x'')|\le L(x^0)\|x'-x''\|. \tag{1}$$

* 1986年1月20日收到．

现对 $\forall \widetilde{x}\in B(x^0,\frac{1}{2}\mu(x^0))$：当 $x',x''\in B(\widetilde{x};\frac{1}{2}\mu(x^0))$ 时，有 $x',x''\in B(x^0,\mu(x^0))$.

由（1）有 $|f(x')-f(x'')|\le L(x^0)\|x'-x''\|$. 任取 $y\in\partial_\sigma f(\widetilde{x})$，则对 $\forall x\in B(\widetilde{x},\frac{1}{2}\mu(x^0))$，由（1）有

$$\langle y,x-\widetilde{x}\rangle\le f(x)-f(\widetilde{x})+\sigma\le L(x^0)\|x-\widetilde{x}\|+\sigma.$$

特取 $x=\widetilde{x}+\frac{1}{2}\mu(x^0)\frac{y}{\|y\|}\in B(\widetilde{x},\frac{1}{2}\mu(x^0))$，则有

$$\frac{1}{2}\mu(x^0)\|y\|\le L(x^0)\cdot\frac{1}{2}\mu(x^0)+\sigma,\quad \|y\|\le L(x^0)+\frac{2\sigma}{\mu(x^0)}.$$

这样就证明了对 $\forall x^0\in s$，存在邻域 $B(x^0,\frac{1}{2}\mu(x^0))$，使在 $\forall x\in B(x^0,\frac{1}{2}\mu(x^0))$，$y\in\partial_\sigma f(x)$ 时有 $\|y\|\le L(x^0)+\frac{2\sigma}{\mu(x^0)}$. 利用覆盖定理，存在有限个 $x_1,x_2,\cdots,x_k\in S$，使 $\bigcup_{i=1}^{k}B(x_i,\frac{1}{2}\mu(x_i))\supset S$，从而取 $L=\max\{L(x_i)+\frac{2\sigma}{\mu(x_i)},\ i=1,\cdots,k\}$，则 $\partial_\sigma f(x)$ 在S上有界于 L.

二、Lemaréchal 高阶算法思想介绍

对于凸函数（不一定可微）的极小化问题

$$\min f(x),\quad x\in R^n.$$

Lemaréchal [1] 模仿牛顿型算法的思想，在当前点 x 处对 $D(d)=f(x+d)-f(x)$ 作出高于一阶的逼近，然后对 D 进行极小化，得到搜索方向 d. 下面的引理是Lemaréchal算法的基础.

引理2.1 [1] $f(x)$ 是 $R^n\to R$ 的凸连续函数，则对 $\forall x,d\in R^n$，有

$$f(x+d)-f(x)=\max\{f'_\sigma(x,d)-\sigma,\ \sigma\ge 0\}.\tag{5}$$

并且若记使达到max的 σ 所成的集为 $E(d)$，即

$$E(d)=\{\sigma\ge 0;\ f(x+d)-f(x)=f'_\sigma(x,d)-\sigma\},\tag{6}$$

则 $E(d)=\{f(x)-f(x+d)+\langle x^*\cdot d\rangle,\ x^*\in\partial f(x+d)\}$. 若 $\|d\|$ 充分小，则 $\sigma\le 2L\|d\|$，$\forall\sigma\in E(d)$（其中 L 是Lip常数）. （7）

从而，我们由（5）可得到一个类似于Taylor展开的公式：

$$f(x+d)-f(x)=f'(x,d)+\max\{f'_\sigma(x,d)-f'(x,d)-\sigma,\sigma\ge 0\}.\tag{8}$$

当 $\sigma\to 0^+$ 时，$f'_\sigma(x,d)\to f'(x,d)$，Lemarechal假设 $f'_\sigma(x,d)$ 逼近 $f'(x,d)$ 的速度为 σ^β，而 $0<\beta\le 1$，且

$$\lim_{\sigma\to 0^+}\frac{f'_\sigma(x,d)-f'(x,d)}{\sigma^\beta}=C(d),\quad C(d)\ 是R^n上正的齐次函数.\tag{9}$$

在这样的假设之下，有

$$f'(x,d)-f(x,d)=C(d)\sigma^\beta+r_d(\sigma),\ 其中 r_d(\sigma)=o$$

记 $$f(x+d)-f(x)\approx f'(x,d)+\max\{C(d)\sigma^\beta-\sigma,\ \sigma\ge 0\}\triangleq D(d)\tag{11}$$

由于 σ 是 $O(d)$ 阶的，$C(d)$ 与 d 同阶，$\therefore C(d)\sigma^\beta$ 的阶至少为 $O(\|d\|^{1+\beta})$，且 $\beta>0$.
$\therefore$（11）式是高于一阶的逼近式.

根据Lemarechal的建议，用 $[f'_\sigma(x.d)-f'(x.d)]/\sigma^\beta$ 代替函数 $C(d)$，从而（11）中 $D(d)$ 由下面的 $D_\sigma(d)$ 代替：

$$D_\sigma(d)=f'(x.d)+(\frac{\beta}{\sigma})^{\beta\gamma}(1-\beta)[f'_\sigma(x.d)-f'(x.d)]^\gamma,\ 其中\gamma=\frac{1}{1-\beta}.\tag{12}$$

Lemaréchal提出的σ-牛顿法的步骤就是求 $D_\sigma(d)$ 的极小，作为算法的搜索方向．并且，在一定条件下，Lemarechal得到一个 d 为$D_\sigma(d)$ 的极小的充要条件．

定理2.2 [1] 若$g_0\in G_\sigma\triangleq\partial_\sigma f(x)$， $0\ \overline{\in} G_\sigma$，定义

$$\lambda_\sigma=\min\{\lambda\geq 0,\ \lambda g_0\in G_\sigma\} \tag{13}$$

$$k_{\beta\sigma}(d)=\left(\frac{\beta}{\sigma}\right)^{\beta\gamma}[f'_\sigma(x.d)-f'(x.d)]^{\gamma-1},\ \gamma=\frac{1}{1-\beta},$$

则 d 是$D_\sigma(d)$ 的极小的充要条件是下面三个式子成立：

i) $$\lambda_\sigma<1. \tag{14}$$

ii) $$k_{\beta\sigma}(d)=\frac{1}{1-\lambda_\sigma}, \tag{15}$$

iii) $$\max\{\langle x^*,d\rangle,\ x^*\in G_\sigma\}=\lambda_\sigma\langle g_0,d\rangle<0. \tag{16}$$

从上面的定理中，可以看出，（14）式仅依赖于g_0在G_σ中的位置，（16）式取决于 d 的方向，而 d 的长度则由（15）式确定．由于在我们的算法中，作为搜索方向，我们关心的仅仅是方向，因此Lemaréchal把满足（16）的单位向量称为σ-牛顿方向．即

定理2.3 设$g_0\in G_\sigma$， $0\ \overline{\in} G_\sigma$，称$R^n$中模为 1 的向量$\overline{d}$为关于$G_\sigma$的$\sigma$-牛顿方向，若定理2.2中的（16）式关于$\overline{d}$成立．

Lemaréchal认为，沿着某个σ-牛顿方向去进行线搜索是合理的，并且，可以通过对下列问题的求解来得到σ-牛顿方向：

$(\widetilde{\mathrm{p}})$：$\min p(d):=\langle g_0,d\rangle+\frac{1}{2}\|d\|^2$， $\langle g,d\rangle\leq\lambda_\sigma\langle g_0,d\rangle$， $\forall g\in G_\sigma$，其中

$$\lambda_\sigma=\min\{\lambda\geq 0,\ \lambda g_0\in G_\sigma\}.$$

三、子问题的构造及性质

在我们的算法中，若要将上述子问题$(\widetilde{\mathrm{p}})$的解作为搜索方向，那么这个$(\widetilde{\mathrm{p}})$应当具有这样的性质，当$(\widetilde{\mathrm{p}})$的解为 0 时，应有 $0\in G_\sigma$，(在迭代步中则是 $0\in G_i$)．并且当达到最优点时，$(\widetilde{\mathrm{p}})$的解应为 0．然而，遗憾的是，对于凸集G_σ，上节提到的$(\widetilde{\mathrm{p}})$问题并不具有这种性质．

例 1 设$G_\sigma=\{(x_1,x_2)\mid x_1\geq a-\sqrt{a^2-(x_2-a)^2},\ x_1\leq 2a\}$ 其中$a>0$，取$g_0=(b,0)\in G_\sigma$，$b\in(a,2a)$，则由λ_σ的定义$\lambda_\sigma g_0=(a,0)$．$(\widetilde{\mathrm{p}})$的解$d_\sigma=0$，但$0\ \overline{\in} G_\sigma$．

例 2 设$G_\sigma=\{(x_1,x_2),\ 0\leq x_1\leq 1,\ 0\leq x_2\leq 1\}$．显然 $0\in G_\sigma$，但取$g_0=(1,1)\in G_\sigma$，则$\lambda_\sigma=0$，$(\widetilde{\mathrm{p}})$的解$d_\sigma=-g_0\neq 0$．

为了使子问题得到的解能够作为算法的搜索方向，我们修改$(\widetilde{\mathrm{p}})$为（p）：

（p）： $\min\langle\lambda_\sigma g_0,d\rangle+\frac{1}{2}\|d\|^2$， s.t. $\langle g,d\rangle\leq\lambda_\sigma\langle g_0,d\rangle$， $\forall g\in G_\sigma$，

其中 $\lambda_\sigma=\min\{\lambda\geq 0,\ \lambda g_0\in G_\sigma\}$．

下面的命题保证了修改后的子问题的解仍是Lemaréchal所定义的σ-牛顿方向．

命题3.1 问题（p）有唯一的最优解d_σ，若$d_\sigma=0$，则不存在σ-牛顿方向；若$d_\sigma\neq 0$，则$\overline{d}_\sigma=d_\sigma/\|d_\sigma\|$是一个$\sigma$-牛顿方向．

证明 问题（p）是一个点$-\lambda_\sigma g_0$到闭凸锥

$\{d:\ \langle g,d\rangle\leq\lambda_\sigma\langle g_0,d\rangle,\ \forall g\in G_\sigma\}=(G_\sigma-\lambda_\sigma(g_0))^*$上的垂直投影问题，

故解存在且唯一．

若存在牛顿方向$\overline{d}$，则由定义

$$\langle g,\quad\leq\lambda_\sigma\langle g_0,\overline{d}\rangle,\ \forall g\in G_\sigma,\ 且\ \langle\lambda_\sigma g_0,\overline{d}\rangle<0.$$

从而当$t>0$而充分小时,$\langle \lambda_\sigma g_0,\ t\overline{d}\rangle+\frac{1}{2}\|t\overline{d}\|^2<0$. 故（p）的解$d_\sigma\neq 0$.

若$d_\sigma\neq 0$是（p）的解，则显然有$\langle \lambda_\sigma g_0,\ d_\sigma\rangle+\frac{1}{2}\|d_\sigma\|^2\leq 0$. 再由（p)的约束条件:

$$\max\{\langle g,\ d_\sigma\rangle,\ g\in G_\sigma\}=\lambda_\sigma\langle g_0,\ d_\sigma\rangle\leq -\frac{1}{2}\|d_\sigma\|^2<0.$$

该式显然对$d_\sigma/\|d_\sigma\|$也成立. 故$\overline{d_\sigma}=d_\sigma/\|d_\sigma\|$是$\sigma$-牛顿方向. 下面我们讨论（p）的解$d_\sigma$的性质.

引理3.2 $0\in G_\sigma$的充要条件为$\lambda_\sigma=0$.

命题3.3 若$0\in G_\sigma$，则（p）的解$d_\sigma=0$.

证明 由引理3.2显然.

在算法中，G_σ取为$\partial_\sigma f(x)$或是某种近似，是R^n中的紧凸集，故$Ri(G_\sigma)\neq\phi$. 我们可以取g_0为G_σ的生成锥的一个相对内点，从而得到下面的结果.

命题3.4 若（p）的解$d_\sigma=0$，g_0是G_σ的生成锥的相对内点，则有$0\in G_\sigma$.

证明 若$d_\sigma=0$，则由命题3.1，G_σ无σ-牛顿方向，即问题:

（p′）: $\langle g,d\rangle\leq\lambda_\sigma\langle g_0,d\rangle$, $\forall g\in G_\sigma$, $\langle\lambda_\sigma g_0,d\rangle<0$ 无解.

而这又等价于: $-\lambda_\sigma g_0\in(G_\sigma-\lambda_\sigma\{g_0\})^{**}$. 由二次共轭的定义，$-\lambda_\sigma g_0$在由$G_\sigma-\lambda_\sigma\{g_0\}$生成的闭凸锥中.

下面分两种情况讨论:

i) 若$\lambda_\sigma=0$，则由引理3.2. $0\in G_\sigma$.

ii) 若$\lambda_\sigma>0$. 则由引理3.2. $0\notin G_\sigma$，从而$g_0\neq 0$.

a. 若$-\lambda_\sigma g_0$在由$G_\sigma-\lambda_\sigma\{g_0\}$生成的凸锥中，则$\exists\tilde{g}\in G_\sigma$，$\tilde{\lambda}\geq 0$，使$-g_0=\tilde{\lambda}(\tilde{g}-\lambda_\sigma g_0)$.

因$g_0\neq 0$，故$\tilde{\lambda}>0$. 从而有

$$\tilde{g}=\lambda_\sigma g_0-\frac{1}{\tilde{\lambda}}g_0=(\lambda_\sigma-\frac{1}{\tilde{\lambda}})g_0\in G_\sigma \tag{17}$$

若$\lambda_\sigma-\frac{1}{\tilde{\lambda}}\geq 0$，则与$\lambda_\sigma$的定义矛盾.

若$\lambda_\sigma-\frac{1}{\tilde{\lambda}}<0$，则由$G_\sigma$的凸性

$$0=\left(\frac{\frac{1}{\tilde{\lambda}}-\lambda_\sigma}{1+\frac{1}{\tilde{\lambda}}-\lambda_\sigma}\right)g_0+\left[1-\frac{\frac{1}{\tilde{\lambda}}-\lambda_\sigma}{1+\frac{1}{\tilde{\lambda}}-\lambda_\sigma}\right](\lambda_\sigma-\frac{1}{\tilde{\lambda}})g_0\in G_\sigma.$$

这与$0\notin G_\sigma$的假设矛盾.

b. 若$-\lambda_\sigma g_0$在由$G_\sigma-\lambda_\sigma\{g_0\}$生成的闭凸锥的相对边界上. 设该闭凸锥的维数为$n$(若维数为$m<n$，则在其生成子空间上考虑). 由于$-\lambda_\sigma g_0$是其边界点，故存在$n-1$维的通过$-\lambda_\sigma g_0$和原点的分离超平面$\pi$. 由于$g_0$是$G_\sigma$的生成锥的（相对）内点，故$g_0-\lambda_\sigma g_0$是$G_\sigma-\lambda_\sigma\{g_0\}$的（相对）内点.

由于$-\lambda_\sigma g_0\in\pi$，$0\in\pi$，故$\pi$是一个子空间，故$(1-\lambda_\sigma)g_0\in\pi$. 这与$(1-\lambda_\sigma)g_0=g_0-\lambda_\sigma g_0$是$G_\sigma-\lambda_\sigma\{g_0\}$的内点矛盾. 故$0\in G_\sigma$.

下面我们讨论，当（p）的解d_σ非零时，目标函数$f(x)$沿d_σ方向的下降性. 在下面的定理中，我们认为$G_\sigma=\partial_\sigma f(x)$.

定理3.5 取$G_\sigma=\partial_\sigma f(x)$，若（p）的解$d_\sigma\neq 0$，则$\exists\beta_0>0$，当$\beta\leq\beta_0$时

$$f(x+\beta d_\sigma)-f(x)\leq-\frac{1}{2}\beta\|d_\sigma\|^2. \tag{18}$$

证明 在我们的算法中，假设所有的迭代点都落在一个有界闭集$S_0=\{x\mid f(x)\leq f(x_0)\}$内，由$f$的局部Lip性知，存在$L>0$，使：

$$|f(x')-f(x'')|\leq L\|x'-x''\|,\quad \forall x',x\in S. \tag{19}$$

由$\partial f(x)$在S_0上的一致有界性，$\exists K>0$，使对一切$x\in S_0$，$g\in\partial f(x)$，有

$$\|g\|\leq K. \tag{20}$$

对当前点x，（p）的解d_σ，取$\beta_0=\dfrac{\sigma}{(L+K)\|d_\sigma\|}$，则由中值定理，对$\beta<\beta_0$，

$$f(x+\beta d_\sigma)-f(x)=\langle x^*,\beta d_\sigma\rangle,\text{ 其中 } x^*\in\partial f(x+\alpha\beta d_\sigma),\ \alpha\in(0,1). \tag{21}$$

由于$x^*\in\partial f(x+\alpha\beta d_\sigma)$，故对$\forall y\in R^n$，有

$$\langle x^*,y-x-\alpha\beta d_\sigma\rangle\leq f(y)-f(x+\alpha\beta d_\sigma).$$

移项，由（19）和（20）：

$$\begin{aligned}\langle x^*,y-x\rangle&\leq f(y)-f(x)+[f(x)-f(x+\alpha\beta d_\sigma)+\langle x^*,\alpha\beta d_\sigma\rangle]\\&\leq f(y)-f(x)+L\alpha\beta\|d_\sigma\|+K\alpha\beta\|d_\sigma\|\leq f(y)-f(x)+\sigma.\end{aligned}$$

故$x^*\in\partial_\sigma f(x)=G_\sigma$。由于$d_\sigma$是（p）的解，故有

$$\langle\lambda_\sigma g_0,d_\sigma\rangle+\frac{1}{2}\|d_\sigma\|^2\leq 0 \tag{22}$$

及

$$\langle g,d_\sigma\rangle\leq\lambda_\sigma\langle g_0,d_\sigma\rangle,\quad\forall g\in G_\sigma \tag{23}$$

由（21),(23),(22)：

$$f(x+\beta d_\sigma)-f(x)=\langle x^*,\beta d_\sigma\rangle\leq\beta\lambda_\sigma\langle g_0,d_\sigma\rangle\leq-\frac{1}{2}\beta\|d_\sigma\|^2.$$

证毕。

四、算法的构造及收敛定理

通过上面的讨论，我们知道问题（p）的解d_σ若非零，则可以作为算法的搜索方向，为了使一开始讨论的算法简单，我们对事先取定的$\sigma_0>0$，算法求出的是σ_0-极小，即对$\forall y\in R^n$，有$f(y)\geq f(x^*)-\sigma_0$。显然，若$0\in\partial_{\sigma_0}f(x^*)$，则$x^*$就是$\sigma_0$-极小。

算法步骤：

Step 0： 给出$\sigma_0>0$，$\beta_0>0$，$x^0\in R^n$，置$i=0$；

Step 1： 取适当的$g_i\in Ri(\partial_{\sigma_0}f(x_i))\triangleq Ri(G_i)$，求解问题（$P_i$）：

$$\min\langle\lambda_i g_i,d\rangle+\frac{1}{2}\|d\|^2.\ \text{s.t.}\ \langle g,d\rangle\leq\lambda_i\langle g_i,d\rangle,\quad\forall g\in G_i,$$

其中 $\lambda_i=\min\{\lambda\geq 0,\ \lambda g_i\in G_i\}$。得解$\lambda_i$，$d_i$；

Step 2： 若$d_i=0$转则停止；否则取$\beta_i=\beta_0$，转Step 3；

Step 3： 若$f(x_i+\beta_i d_i)-f(x_i)\leq-\frac{1}{2}\beta_i\|d_i\|^2$。转step4． (24)

否则， Step5；

Step 4： 令$x_{i+1}=x_i+\beta_i d_i$，置$i=i+1$，转step1；

Step 5： 令$\beta_i=\frac{1}{2}\beta_i$，转step3；

定理4.1 设f为R^n到R的下有界的连续凸函数，且$S_0=\{x\mid f(x)\leq f(x_0)\}$有界，$\{x_i\}$ $i=1,2,\cdots$是由算法产生的序列，则

a）若算法在x_k处停滞，则$d_k=0$。

b）若算法产生无穷序列$\{x_i\}$，且$\overline{x}$为$\{x_i\}$的聚点，则$d_i\to 0$。

证明 a）且由$d_k=0$知，$0\in\partial_{\sigma_0}f(x_k)$，故$x_k$是$f$的$\sigma_0$-极小点。

b）由 $\{f(x_i)\}$ 的单调下界性和 f 的下有界性，及不等式（24）知，当 $i\to\infty$ 时，
$$\beta_i\|d_i\|^2\to 0. \tag{25}$$
若 $\overline{x}$ 是 $\{x_i\}$ 的聚点，不妨设 $x_i\to\overline{x}$（否则可取子序列），则由算法的构造中可看出，β_i 或等于 β_0，即Step3中的不等式（24）在第一次选取 β_i 时即满足，或是对某一 $\beta_i<\beta_0$ 满足，而在之前是不满足的，即对 $2\beta_i$，（24）式不成立。

在定理3.5的证明过程中，我们得知与 $\beta_i\le\dfrac{\sigma_0}{(L+K)\|d_i\|}$ 时，Step3中的（24）总能满足，故
$$\beta_i\ge\min\left\{\beta_0,\ \frac{\sigma_0}{2(L+K)\|d_i\|}\right\}. \tag{25}$$
由于 d_i 是（p_i）的解，故
$$\langle\lambda_i g_i,\ d_i\rangle\le-\frac{1}{2}\|d_i\|^2; \tag{27}$$
另一方面：
$$\langle-\lambda_i g_i,\ d_i\rangle\le\|\lambda_i g_i\|\,\|d_i\|. \tag{28}$$
合并（27），（28）得
$$\|d_i\|\le 2\lambda_i\|g_i\|. \tag{29}$$
将（29）代入（26）：
$$\begin{aligned}\beta_i&\ge\min\left\{\beta_0,\ \frac{\sigma_0}{4(L+K)\lambda_i\|g_i\|}\right\}\\&\ge\min\left\{\beta_0,\ \frac{\sigma_0}{4(L+K)\|g_i\|}\right\}\quad(\because\lambda_i\le 1).\\&\ge\min\left\{\beta_0,\ \frac{\sigma_0}{4(L+K)K}\right\}\quad(\text{由于 }\|g_i\|\le K).\\&>0.\end{aligned}$$
从而由（25）得 $\|d_i\|\to 0$.　　〈证毕〉.

上面的结果只是说明 $\|d_i\|\to 0$，而并不是我们所需要的 $0\in\partial_{\sigma_0}f(\overline{x})$，下面我们给出一个条件，使保证由 $\|d_i\|\to 0$ 可得出 $0\in\partial_{\sigma_0}f(\overline{x})$。

对 $a\in(0,\ 1)$，我们取 $g_i\in G_i$ 满足下面的“a-锐角条件”：

假设4.2　在算法的每一迭代步骤中，假设所取的 $g_i\in G_i$ 一致地满足：存在 $\tilde{d}_i\neq 0$，$\tilde{d}_i\in(G_i-\lambda_i\{g_i\})^*$．使
$$\langle-g_i,\ \tilde{d}_i\rangle\ge a\|g_i\|\cdot\|\tilde{d}_i\|. \tag{30}$$
定理4.3　设定理4.1的条件满足，若存在某 $a\in(0,\ 1)$，使上述“a-锐角条件”对一切 g_i 满足，则当 $d_i\to 0$ 时，$0\in\partial_{\sigma_0}f(\overline{x})$。

证明　由定义，存在 $\tilde{d}_i\neq 0$，使 $\langle g,\ \tilde{d}_i\rangle\le\lambda_i\langle g_i,\ \tilde{d}_i\rangle$，$\forall g\in G_i$，且 $\langle g_i,\ \tilde{d}_i\rangle\le-a\|g_i\|\,\|\tilde{d}_i\|$ 对 $\forall i=1,2,\cdots$ 成立。

由于（p_i）的解 d_i 是 $-\lambda_i g_i$ 在凸锥 $(G_i-\lambda_i\{g_i\})^*$ 上的正交投影，故：
$$\langle d_i,\ d_i-(-\lambda_i g_i)\rangle=0 \tag{31}$$
或
$$\|d_i\|^2=-\lambda_i\langle d_i,\ g_i\rangle \tag{32}$$
由 d_i 的定义，对 $\forall d\in(G_i-\lambda_i\{g_i\})^*$，有
$$\langle\lambda_i g_i,\ a_i\rangle+\frac{1}{2}\|d_i\|^2\le\langle\lambda_i g_i,\ d\rangle+\frac{1}{2}\|d\|^2. \tag{33}$$
由于 $\tilde{d}_i\in(G_i-\lambda_i\{g_i\})^*$，取 $d_i^0=\left\langle-\lambda_i g_i,\ \dfrac{\tilde{d}_i}{\|\tilde{d}_i\|}\right\rangle\dfrac{\tilde{d}_i}{\|\tilde{d}_i\|}\in(G_i-\lambda_i\{g_i\})^*$ 代入（33）得

$$\langle \lambda_i g_i, d_i \rangle + \frac{1}{2} \| d_i \|^2 \le -\frac{1}{2} [\langle \lambda_i g_i, \tilde{d}_i \rangle / \| \tilde{d}_i \|]^2. \tag{34}$$

由（32),（34）式左边化为：$-\frac{1}{2}\| d_i \|^2$，故有

$$-\frac{1}{2} \| d_i \|^2 \le -\frac{1}{2} (\langle \lambda_i g_i, \tilde{d}_i \rangle / \| \tilde{d}_i \|)^2 \tag{35}$$

由"a-锐角条件"(30)：

$\| d_i \|^2 \ge (\langle \lambda_i g_i, \tilde{d}_i \rangle / \| \tilde{d}_i \|)^2 \ge (\frac{a\lambda_i \| g_i \| \| \tilde{d}_i \|}{\| \tilde{d}_i \|})^2$，故 $\| d_i \| \ge a\lambda_i \| g_i \|$。当 $\| d_i \| \to 0$ 时，$\lambda_i \| g_i \| \to 0$。而 $\lambda_i g_i \in G_i \triangleq \partial_{\sigma_0} f(x_i)$。由命题1.2 $\partial_\sigma f(x_0)$ 的上半连续性知，$0 \in \partial_{\sigma_0} f(\bar{x})$。

五、关于算法实现性的讨论

1° 在前节提出的算法中，终止点为问题的σ_0-极小，利用$\partial_\sigma f(x)$ 关于σ的上半连续性和我们给出的"a-锐角条件"，若修改算法的Step2．代之以 $\| d_i \| < \sigma_i$ 且在每一迭代步中不断修正σ_i的值，使在整个算法中$\sigma_i \to 0, (i \to \infty)$．则算法产生的点列 $\{x_i\}$ 若有聚点$\bar{x}$的话，就有 $0 \in \partial f(\bar{x})$。

2° 上面提出的算法在实现中的一个主要困难是$G_\sigma = \partial_\sigma f(x)$ 的出现．可以在算法中取G_σ中有限个向量$g_1, \cdots, g_k$的凸包来近似G_σ，这样，问题（p_i）化为一个二次规划问题，并在一定条件下，可使该二次规划问题的解是问题的下降方向，并得到类似于式（24）的下降程度．

上面我们只是对Lemarechal提出的高阶算法在实现性方面作了一些讨论，要使这一算法真正地得以实现，并体现出高阶性的较好效果，还有待于进一步的研究．

参 考 文 献

〔1〕C.Lemarechal and Zowe, Some remarks on the construction of higher order algorithms in convex optimization, Appl. Math. & Opti. Vol.10, (1983), pp51—68.

〔2〕R.T.Rockafeller, Convex Analysis.

〔3〕M.S.Bazaraa,《最优化基础》(上海科大，上海师大译)．

〔4〕F.H, Clarke, Generalized gradient of Lipschitz functions, Adv. in Math., 40, pp.52—67 (1981).

〔5〕F.H.Clarke, A new approach to lagrange maltipliers, Math.Oper.Res., Vol.1 (1976), pp.165—174.

〔6〕F.H.Clarke, Generalized gradients and applications, Trans. Amer. Math. Soc., 205 (1975), 247—262.

Study on Lemaréchal's High Order Algorithm: Conceptual Algorithm

Huang Wan-zhen Zhang Lian-sheng

(Shanghai University of Science and Technology)

Abstract C.Lemaréchal proposed an idea of high order σ-Newton type algorithm for nonsmooth convex functions in (1).some properties were discussed there.

In this paper, we study more about this high order σ-Newton type algorithm, and give a conceptual algorithm. The grobal convergence has been proved.