BCD & ADMM 算法收敛理论

Lecture 6: BCD 收敛性

罗自炎

北京交通大学数统学院

E-mail: zyluo@bjtu.edu.cn

参考资料

- 教材与参考文献：
  - 最优化：建模、算法与理论
  - Bolte et al., MP 2014
  - Fazel et al., SIMAX 2013
  - Rockafellar & Wets, Variational Analysis

- 致谢： 北京大学文再文教授; 清华大学张立平教授

Outline of BCD

- 充分下降性
- 子列收敛
- 全序列收敛
- 收敛速率

优化模型

- 考虑如下问题：

$$\min_{x\in\mathbb{R}^n, y\in\mathbb{R}^m} \quad \Psi(x,y) \stackrel{\text{def}}{=} f(x)+g(y)+H(x,y), \tag{1}$$

其中f,g 适当闭函数(可能非凸)，H为定义域上的连续可微函数.

- **假设条件**：

(i) $f:\mathbb{R}^n \to (-\infty,+\infty]$，$g:\mathbb{R}^m \to (-\infty,+\infty]$ 均为适当下半连续函数，$\inf_{\mathbb{R}^n\times\mathbb{R}^m}\Psi > -\infty$, $\inf_{\mathbb{R}^n} f > -\infty$，以及$\inf_{\mathbb{R}^m} g > -\infty$

(ii) $H:\mathbb{R}^n\times\mathbb{R}^m \to \mathbb{R}$是连续可微函数，且$\nabla H$在有界集上是联合Lipschitz连续的．即对于任意的有界集$B_1\times B_2\subset\mathbb{R}^n\times\mathbb{R}^m$，存在$L>0$使得对于任意的$(x_i,y_i)\in B_1\times B_2, i=1,2$有

$$\|(\nabla_x H(x_1,y_1)-\nabla_x H(x_2,y_2), \nabla_y H(x_1,y_1)-\nabla_y H(x_2,y_2))\|$$
$$\le L\|(x_1-x_2, y_1-y_2)\|.$$

PALM

- 近似点交替线性化极小化(Proximal Alternating Linearized Minimization, PALM)：

$$x^{k+1} \in \arg\min_{x}\{f(x) + H(x^k, y^k) + \nabla_x H(x^k, y^k)^T(x - x^k) + \frac{c_k}{2}\|x - x^k\|^2\}$$

$$y^{k+1} \in \arg\min_{x}\{g(y) + H(x^{k+1}, y^k) + \nabla_y H(x^{k+1}, y^k)^T(y - y^k) + \frac{d_k}{2}\|y - y^k\|^2\}$$

其中c_k, d_k为步长参数, 可以取固定步长，或进行动态更新（如Armijo线搜索等）.

非凸函数的邻近算子

Theorem 1 (适当闭函数的邻近算子). *设 h 是适当闭函数(可以非凸)，且具有有限的下界，即满足$\inf_{x\in\mathbf{dom}h} h(x) > -\infty$, 定义 h 的邻近算子为*

$$\mathrm{prox}_h(x) = \arg\min_{u\in\mathbf{dom}\ h} \left\{ h(u) + \frac{1}{2}\|u-x\|^2 \right\}.$$

此时，$\forall x \in \mathbf{dom}\ h$, $\mathrm{prox}_h(x)$ 是$\mathbb{R}^n$ 上的非空紧集.

Proof. 定义 $g(u) = h(u) + \frac{1}{2}\|u-x\|^2$, 设 $\inf_{x\in\mathbf{dom}h} h(x) = l$.
取$u_0 \in \mathbf{dom}\ h$, 由于$\frac{1}{2}\|u-x\|^2$ 无上界，故$\exists R > 0$, 对$\forall$ 满足$\|u-x\| > R$ 的u, 成立$\frac{1}{2}\|u-x\|^2 > g(u_0) - l$，即$g(u) > g(u_0)$.
这说明下水平集$\{u \mid g(u) \leqslant g(u_0)\}$ 含于球$\|u-x\| \leqslant R$ 内, 即g 有一个非空有界下水平集. 显然$g(u)$ 是闭函数, 由Weierstrass 定理可知, $g(u)$ 的最小值点集合$\mathrm{prox}_h(x)$ 是非空紧集. $\square$

非光滑非凸问题函数的次微分

回顾一下非光滑非凸函数的次微分:

Definition 1 (次微分). *设$f: \mathbb{R}^n \to (-\infty, +\infty]$ 是适当下半连续函数.*

- *对给定的$x \in \mathbf{dom}\ f$, 满足如下条件的所有向量$u \in \mathbb{R}^n$ 的集合定义为f 在点x 处的*Fréchet *次微分:*
$$\liminf_{y \to x, y \neq x} \frac{f(y) - f(x) - \langle u, y - x \rangle}{\|y - x\|} \geq 0,$$
记为$\hat{\partial} f(x)$.当$x \notin \mathbf{dom}\ f$ 时, 将$\hat{\partial} f(x)$ 定义为空集$\varnothing$.

- *f 在点$x \in \mathbb{R}^n$ 处的极限次微分(或简称为次微分)定义为*
$$\partial f(x) = \{u \in \mathbb{R}^n \ : \ \exists\, x^k \to x, f(x^k) \to f(x), u^k \in \hat{\partial} f(x^k) \to u\}.$$
*极限次微分通过对x 附近的点处的*Fréchet *次微分取极限得到.*

邻近点算子

Theorem 2. *设h 是适当闭函数(可非凸)且有下界, 若$u \in \mathrm{prox}_h(x)$, 则$x - u \in \partial h(u)$*

Proof. 由一阶最优性条件可知：$0 \in u - x + \partial h(u)$ □

凸函数的邻近点算子： 设h 是适当闭凸函数, 则

- $x - u \in \partial h(u) \Longleftrightarrow u = \mathrm{prox}_h(x)$ (凸优化的一阶充要条件);
- $x = \mathrm{prox}_f(x) + \mathrm{prox}_{f^*}(x)$, 从而$x = \mathrm{prox}_{\beta f}(x) + \beta \mathrm{prox}_{f^*/\beta}(x/\beta)$, $\forall \beta > 0$ (Moreau Identity)
- $x = \Pi_C(x) + \Pi_{C^0}(x)$, 其中$C \subseteq \mathbb{R}^n$ 为非空闭凸锥, Π_C 为集合C 上的投影算子.

  (Hint: $f = \mathrm{I}_C, \mathrm{I}_C^* = \mathrm{I}_{C^0}$)

优化模型的假设条件

- 根据假设(ii)，在有界集上H 关于每个分量都是梯度L-Lipschitz连续的，且参数与另一分量无关．即

$$\begin{aligned}\left\|\nabla_x H\left(x_1, y\right)-\nabla_x H\left(x_2, y\right)\right\| &\leqslant L\left\|x_1-x_2\right\|, \\ \left\|\nabla_y H\left(x, y_1\right)-\nabla_y H\left(x, y_2\right)\right\| &\leqslant L\left\|y_1-y_2\right\|.\end{aligned}$$

- $\Psi(x, y)$ 的次微分：

$$\partial \Psi(x, y)=\left(\nabla_x H(x, y)+\partial f(x), \nabla_y H(x, y)+\partial g(y)\right)$$

其中“$+$”表示为集合间的加法.

全序列收敛的证明梗概

(I) 充分下降：找到一个正常数ρ_1 使得

$$\rho_1\|z^{k+1}-z^k\|^2 \leq \Psi(z^k)-\Psi(z^{k+1})$$

(II) 次梯度上界：假设算法产生的迭代序列有界，找到另一个常数ρ_2，使得次梯度有一个上界估计：

$$\|w^{k+1}\| \leq \rho_2\|z^{k+1}-z^k\|, \quad w^{k+1} \in \hat{\partial}\Psi(z^{k+1})$$

(III) 利用KL 性质证明全序列收敛：假设Ψ 是一个KL函数，证明迭代序列$\{z^k\}_{k\in N}$是一个柯西列.

注：前两个步骤是证明多数算法的基本步骤，当这两个性质成立时，对任意的算法产生的迭代序列的聚点集合都为非空连通紧集，且这些聚点都是Ψ 的临界点.

PALM下降量

Theorem 3. *设*$h:\mathbb{R}^d\to\mathbb{R}$*是连续可微函数，梯度*$\nabla h$ *是*L_h*-Lipschitz连续的,* $\sigma:\mathbb{R}^d\to(-\infty,+\infty]$ *是适当下半连续函数且*$\inf_{\mathbb{R}^d}\sigma>-\infty$*. 固定*$t<\frac{1}{L_h}$*，则对任意的*$u\in\mathbf{dom}\,\sigma$ *和*$\tilde{u}\in\mathrm{prox}_{t\sigma}(u-t\nabla h(u))$*，有*

$$h(\tilde{u})+\sigma(\tilde{u})\le h(u)+\sigma(u)-\frac{1}{2}\left(\frac{1}{t}-L_h\right)\|\tilde{u}-u\|^2. \tag{2}$$

证明：由$\tilde{u}\in\mathrm{prox}_{t\sigma}(u-t\nabla h(u))$可知：

$$\langle\tilde{u}-u,\nabla h(u)\rangle+\frac{1}{2t}\|\tilde{u}-u\|^2+\sigma(\tilde{u})\le\sigma(u).$$

由梯度∇h 是L_h-Lipschitz连续及其二次上界的性质可得：

$$\begin{aligned} h(\tilde{u}) + \sigma(\tilde{u}) &\le h(u) + \langle \tilde{u} - u, \nabla h(u) \rangle + \frac{L_h}{2}\|\tilde{u} - u\|^2 + \sigma(\tilde{u}) \\ &\le h(u) + \frac{L_h}{2}\|\tilde{u} - u\|^2 + \sigma(u) - \frac{1}{2t}\|\tilde{u} - u\|^2 \\ &= h(u) + \sigma(u) - \frac{1}{2}\left(\frac{1}{t} - L_h\right)\|\tilde{u} - u\|^2. \end{aligned}$$

充分下降定理

Theorem 4 (充分下降). *在假设条件下, 设$\{z^k\} = \{(x^k, y^k)\}$为PALM的BCD 算法产生的迭代序列, 且假设$\{z^k\}$有界. 取步长$c_k = d_k = \frac{1}{\gamma L}$, 其中$\gamma > 1$是常数, L为∇H的Lipschitz 常数, 则以下结论成立:*

(i) 迭代点处的函数值序列$\{\Psi(z^k)\}$是单调下降的，且

$$\frac{\rho_1}{2}\|z^{k+1} - z^k\|^2 \leq \Psi(z^k) - \Psi(z^{k+1}), \quad \forall\, k \geq 0,$$

其中$\rho_1 = (\gamma - 1)L$;

(ii) 序列$\{\|z^{k+1} - z^k\|\}_{k=1}^{\infty}$平方可和，即$\sum_{k=1}^{\infty}\|z^{k+1} - z^k\|^2 < +\infty$, 从而$\lim_{k\to\infty}\|z^{k+1} - z^k\| = 0$.

注： $\sum_{k=1}^{\infty}\|z^{k+1} - z^k\|^2 < +\infty \nRightarrow \sum_{k=1}^{\infty}\|z^{k+1} - z^k\| < +\infty$,
e.g., 调和级数: $\sum\limits_{k=1}^{\infty}\frac{1}{k} = +\infty$, $\sum\limits_{k=1}^{\infty}\frac{1}{k^2} = \frac{\pi^2}{6}$

Thm 4 证明

(i) 根据假设条件(ii)，$H(x,y)$关于每个分量都是Lipschitz连续的，由Theorem 3的结论可得到每一步关于x^k和y^k的下降量估计：

$$
\begin{aligned}
& H(x^{k+1}, y^k) + f(x^{k+1}) \\
\le\ & H(x^k, y^k) + f(x^k) - \frac{1}{2}\left(\frac{1}{c_k} - L\right)\|x^{k+1} - x^k\|^2 \\
=\ & H(x^k, y^k) + f(x^k) - \frac{1}{2}(\gamma - 1)L\|x^{k+1} - x^k\|^2,
\end{aligned}
$$

以及

$$
\begin{aligned}
& H(x^{k+1}, y^{k+1}) + g(y^{k+1}) \\
\le\ & H(x^{k+1}, y^k) + g(y^k) - \frac{1}{2}\left(\frac{1}{d_k} - L\right)\|y^{k+1} - y^k\|^2 \\
=\ & H(x^{k+1}, y^k) + g(y^k) - \frac{1}{2}(\gamma - 1)L\|y^{k+1} - y^k\|^2.
\end{aligned}
$$

Thm 4 证明(续)

将上述两个不等式相加，消去$H(x^{k+1}, y^k)$,得到

$$\begin{aligned}
&\Psi(z^k) - \Psi(z^{k+1}) \\
=&H(x^k, y^k) + f(x^k) + g(y^k) - H(x^{k+1}, y^{k+1}) - f(x^{k+1}) - g(y^{k+1}) \\
\geq&\frac{1}{2}(\gamma - 1)L\left(\|x^{k+1} - x^k\|^2 + \|y^{k+1} - y^k\|^2\right).
\end{aligned}$$

由此立即可得

$$\frac{\rho_1}{2}\|z^{k+1} - z^k\|^2 \leq \Psi(z^k) - \Psi(z^{k+1}). \tag{3}$$

综上可知: $\{\Psi(z^k)\}$关于k是单调递减的. 根据假设$\inf \Psi > -\infty$ 可知$\Psi(z^k)$ 单调下降收敛到一个有限的数Ψ^*.

思考： (1) 上述下降性能否推广到多块的情形? (2) 如何保证序列$\{z^k\}$有界?

Thm 4 证明(续)

(ii) 设N为任意的整数，在(3) 式中对k求和，得

$$\sum_{k=0}^{N-1} \|z^{k+1} - z^k\|^2 \le \frac{2}{\rho_1}(\Psi(z^0) - \Psi(z^N)) \le \frac{2}{\rho_1}(\Psi(z^0) - \Psi^*).$$

令$N \to \infty$ 可得$\sum_{k=0}^{\infty} \|z^{k+1} - z^k\|^2 < +\infty$，从而

$$\lim_{k\to\infty} \|z^{k+1} - z^k\| = 0.$$

注：定理表明进行一轮近似点交替线性化迭代后, 函数值下降量的下界可被相邻迭代点之间的距离控制.几乎所有下降类算法在一定条件下都满足这个性质．到此完成了收敛性分析的第一个步骤.

分析

- 已证：函数值$\Psi^k \to \Psi^*$, 但并未证明Ψ^*在某个局部最优解z^*处取得;
- 已证：$\|z^{k+1} - z^k\| \to 0$, 但并不能说明$\{z^k\}$(子)序列收敛;
- 下面：将讨论序列$\{z^k\}$是否会趋于某个临界点z^*, 即$0 \in \partial\Psi(z^*)$, 这是收敛性框架中的第二个步骤.

次梯度上界

Proposition 1. *在假设条件下，设$\{z^k\}$是算法产生的有界序列，$\forall$整数k，记*

$$A_x^k = \frac{1}{c_{k-1}}(x^{k-1} - x^k) + \nabla_x H(x^k, y^k) - \nabla_x H(x^{k-1}, y^{k-1}),$$

$$A_y^k = \frac{1}{d_{k-1}}(y^{k-1} - y^k) + \nabla_y H(x^k, y^k) - \nabla_y H(x^k, y^{k-1}).$$

则有$(A_x^k, A_y^k) \in \hat{\partial}\Psi(x^k, y^k)$ *且*

$$\|(A_x^k, A_y^k)\| \leq \|A_x^k\| + \|A_y^k\| \leq \rho_2 \|z^k - z^{k-1}\|,$$

其中$\rho_2 = (2\gamma + 2)L$.

Proposition 1 证明

由迭代格式中更新x^k的一阶最优性条件可知

$$\nabla_x H(x^{k-1}, y^{k-1}) + \frac{1}{c_{k-1}}(x^k - x^{k-1}) + u^k = 0,$$

其中$u^k \in \hat{\partial} f(x^k)$为$f$的一个次梯度．因此我们有

$$\nabla_x H(x^{k-1}, y^{k-1}) + u^k = \frac{1}{c_{k-1}}(x^{k-1} - x^k).$$

同理，由迭代格式中关于y^k的更新可知

$$\nabla_y H(x^k, y^{k-1}) + v^k = \frac{1}{d_{k-1}}(y^{k-1} - y^k),$$

其中$v^k \in \hat{\partial} g(y^k)$为$g$的一个次梯度．由$A_x^k, A_y^k$的定义和$\partial \Psi$的表达式

可得

$$
\begin{aligned}
A_x^k &= \nabla_x H(x^k, y^k) + u^k \in \hat{\partial}_x \Psi(x^k, y^k), \\
A_y^k &= \nabla_y H(x^k, y^k) + v^k \in \hat{\partial}_y \Psi(x^k, y^k).
\end{aligned}
$$

即有$(A_x^k, A_y^k) \in \hat{\partial}\Psi(x^k, y^k)$.

下面估计A_x^k和A_y^k的模长．由∇H在有界集上关于(x, y)是联合Lipschitz连续, 则

$$
\begin{aligned}
\|A_x^k\| &\leq \frac{1}{c_{k-1}}\|x^{k-1} - x^k\| + \|\nabla_x H(x^k, y^k) - \nabla_x H(x^{k-1}, y^{k-1})\| \\
&\leq \frac{1}{c_{k-1}}\|x^{k-1} - x^k\| + L\|(x^{k-1} - x^k, y^{k-1} - y^k)\| \\
&\leq (\gamma + 1)L\|z^{k-1} - z^k\|.
\end{aligned}
$$

Proposition 1 证明(续)

另一方面，对$\|A_y^k\|$的估计只需要用到∇H关于y的Lipschitz 连续性：

$$
\begin{aligned}
\|A_y^k\| &\leq \frac{1}{d_{k-1}}\|y^k - y^{k-1}\| + \|\nabla_y H(x^k, y^k) - \nabla_y H(x^k, y^{k-1})\| \\
&\leq \frac{1}{d_{k-1}}\|y^k - y^{k-1}\| + L\|y^k - y^{k-1}\| \\
&= \left(\frac{1}{d_{k-1}} + L\right)\|y^k - y^{k-1}\| \\
&\leq (\gamma + 1)L\|z^k - z^{k-1}\|.
\end{aligned}
$$

结合这两个估计我们最终得到

$$
\|(A_x^k, A_y^k)\| \leq \|A_x^k\| + \|A_y^k\| \leq (2\gamma + 2)L\|z^k - z^{k-1}\| = \rho_2\|z^k - z^{k-1}\|.
$$

子列收敛性

- **注：** 由$\|z^{k+1}-z^k\|\to 0$ 可知: $\partial\hat{\Psi}(z^k)\ni(A_x^k,A_y^k)\to 0$

Theorem 5. *定义$\omega(z^0)$为近似点交替线性化方法从点z^0出发产生迭代序列的所有极限点集，且$\{z^k\}$是有界序列，则以下结论成立：*

(1) $\emptyset\neq\omega(z^0)\subset\operatorname{crit}\Psi=\{z:0\in\partial\Psi(z)\}$ *(Ψ所有临界点构成的集合);*

(2) z^k*与集合*$\omega(z^0)$*的距离趋于0，即*

$$\lim_{k\to\infty}\operatorname{dist}(z^k,\omega(z^0))=0;$$

(3) $\omega(z^0)$ *是非空的连通紧集;*

(4) Ψ*在*$\omega(z^0)$*上是一个有限的常数.*

Theorem 5 证明

证明: (1) 由序列$\{z^k\}$ 的有界性, 不妨设$z^* = (x^*, y^*)$ 是$\{z^k\}_{k\in\mathbb{N}} = \{(x^k, y^k)\}_{k\in\mathbb{N}}$ 的一个极限点. 即存在一个子序列$\{(x^{k_q}, y^{k_q})\}_{q\in\mathbb{N}}$, 使得$(x^{k_q}, y^{k_q}) \to (x^*, y^*)$ 当$q \to \infty$. 由假设条件(1)可知f和g都是下半连续的, 因此:

$$\liminf_{q\to\infty} f\left(x^{k_q}\right) \geq f\left(x^*\right), \ \liminf_{q\to\infty} g\left(y^{k_q}\right) \geq g\left(y^*\right). \tag{4}$$

由子问题迭代公式可知, 对于所有正整数k

$$x^{k+1} \in \underset{x\in\mathbb{R}^n}{\operatorname{argmin}} \left\{\langle x - x^k, \nabla_x H\left(x^k, y^k\right)\rangle + \frac{c_k}{2}\|x - x^k\|^2 + f(x)\right\}.$$

在上述不等式中取$x = x^*$, 得到

$$\begin{aligned}&\langle x^{k+1} - x^k, \nabla_x H\left(x^k, y^k\right)\rangle + \tfrac{c_k}{2}\|x^{k+1} - x^k\|^2 + f\left(x^{k+1}\right)\\ &\leq \langle x^* - x^k, \nabla_x H\left(x^k, y^k\right)\rangle + \tfrac{c_k}{2}\|x^* - x^k\|^2 + f\left(x^*\right).\end{aligned}$$

Theorem 5 证明(续)

在上面的不等式中取$k = k_q - 1$, 并让q趋于无穷大, 我们得到

$$\begin{aligned}\limsup_{q\to\infty} f\left(x^{k_q}\right) \quad &\leq \quad \limsup_{q\to\infty}(\langle x^* - x^{k_q-1}, \nabla_x H\left(x^{k_q-1}, y^{k_q-1}\right)\rangle \\ &+\frac{c_k}{2}\|x^* - x^{k_q-1}\|^2) + f\left(x^*\right), \end{aligned} \tag{5}$$

其中使用了以下事实: 序列$\{x^k\}_{k\in\mathbb{N}}$和$\{c_k\}_{k\in\mathbb{N}}$都是有限的,∇H连续, 以及相邻迭代点之间的距离趋近于零(见Theorem 4(ii)).正因为如此，当$q \to \infty$时，$x^{k_q-1} \to x^*$, 所以(5)简化为$\limsup\limits_{q\to\infty} f\left(x^{k_q}\right) \leq f\left(x^*\right)$. 结合(4) 得$f\left(x^{k_q}\right) \to f\left(x^*\right)$ 当$q \to \infty$. 类似可得$g\left(y^{k_q}\right) \to g\left(y^*\right)$, 从而

$$\lim_{q\to\infty} \Psi\left(x^{k_q}, y^{k_q}\right) = \lim_{q\to\infty} \left\{H\left(x^{k_q}, y^{k_q}\right) + f\left(x^{k_q}\right) + g\left(y^{k_q}\right)\right\}$$

$$= H\left(x^*, y^*\right) + f\left(x^*\right) + g\left(y^*\right) = \Psi\left(x^*, y^*\right).$$

Theorem 5 证明(续)

另一方面,由Theorme 4(ii)与Theorem 5 有$\left(A_x^{k_q}, A_y^{k_q}\right) \in \hat{\partial}\Psi\left(x^{k_q}, y^{k_q}\right)$
且$(A_x^{k_q}, A_y^{k_q}) \to (0,0)$, $\left(x^{k_q}, y^{k_q}\right) \to (x^*, y^*)$, 当$q \to \infty$. 由$\partial\Psi$的定义
可知$(0,0) \in \partial\Psi\left(x^*, y^*\right)$, 即$(x^*, y^*)$是$\Psi$ 的一个临界点.

(ii) 记

$$t_k := \mathrm{dist}(z^k, \omega(z^0)), \quad k = 0, 1, \ldots$$

由$w(z^0)$ 的定义以及$\{z^k\}$ 有界性可知: $\{t_k\}$ 是$\mathbb{R}$上的非负有界数列,
从而存在收敛子列. 设t^* 是$\{t_k\}$ 的任一聚点, 则存在子列$K \subseteq \mathbb{N}$, 使得

$$\lim_{k\in K, k\to\infty} t_k = t^*.$$

注意到相应的子列$\{z^k\}_{k\in K} \subseteq \{z^k\}$ 有界, 从而存在子列$K_1 \subseteq K$ 使得

$$\lim_{k\in K_1, k\to\infty} z^k = \bar{z}.$$

由距离函数dist$(\cdot,\omega(z^0))$ 的连续性可知:

$$\begin{aligned} t^* &= \lim_{k\in K,k\to\infty} \mathrm{dist}(z^k,\omega(z^0)) = \lim_{k\in K_1,k\to\infty} \mathrm{dist}(z^k,\omega(z^0)) \\ &= \mathrm{dist}\left(\lim_{k\in K_1,k\to\infty} z^k,\omega(z^0)\right) = \mathrm{dist}(\bar{z},\omega(z^0)) = 0. \end{aligned}$$

由t^* 的任意性可知, 序列$\{t_k\}$ 的任意聚点均为0, 从而全序列收敛到0, 即$\lim_{k\to\infty} \mathrm{dist}(z^k,\omega(z^0)) = 0$.

也可用反证法：假设存在$\epsilon > 0$, 存在无穷子列$\{z^k\}_{k\in K}$, 使得

$$\mathrm{dist}(z^k,\omega(z^0)) \geq \epsilon,\ \ \forall k \in K. \tag{6}$$

由$\{z^k\}_{k\in K}$ 的有界性可知, 存在无穷子列$K_1 \subseteq K$ 使得

$$\lim_{k\in K_1,k\to\infty} z^k = \bar{z} \in \omega(z^0).$$

由距离函数dist$(\cdot,\omega(z^0))$ 的连续性可知:

$$\lim_{k\in K_1,k\to\infty} \mathrm{dist}(z^k,\omega(z^0)) = \mathrm{dist}(\lim_{k\in K_1,k\to\infty} z^k,\omega(z^0)) = \mathrm{dist}(\bar{z},\omega(z^0)) = 0$$

与(6)矛盾. 从而$\lim_{k\to\infty} \operatorname{dist}(z^k, \omega(z^0)) = 0$.

(iii)设$\omega = \omega\left(z^0\right)$. 观察到$\omega$可以看作如下紧集的交集

$$\omega = \bigcap_{q\in\mathbb{N}} \overline{\bigcup_{k\geq q} \{z^k\}} = \lim_{q\to\infty} \overline{\bigcup_{k\geq q} \{z^k\}}$$

所以它也是紧集.

下证连通性. 反证法假设ω是不连通的. 因此存在两个非空闭的不相交子集A和B，使得$\omega = A \cup B$. 对所有$z \in \mathbb{R}^n \times \mathbb{R}^m$, 定义函数$\gamma : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ 如下：

$$\gamma(z) = \frac{\operatorname{dist}(z, A)}{\operatorname{dist}(z, A) + \operatorname{dist}(z, B)}$$

由于A和B的闭性, 函数γ是良定义的连续函数. 注意$A = \gamma^{-1}(\{0\}) = [y = 0]$和$B = \gamma^{-1}(\{1\}) = [\gamma = 1]$.设$U = [\gamma < 1/4]$和$V = [\gamma > 3/4]$，得到$A$ 和B 的两个开邻域. 存在一个整数k_0，使得对于所有$k \geq k_0$ 有z^k要么属于U要么属于V. 否则, 存在一个子

序列$\{z^{k_q}\}_{q\in\mathbb{N}}$收敛到$U\cup V$的补集中, 这将意味着$\{z^k\}$存在一个极限点$z^*$在$\mathbb{R}^n\backslash(U\cup V)$ 中,这是不可能的.

对于每个整数k, 设$r_k=\gamma(z^k)$. 序列$\{r_k\}_{k\in\mathbb{N}}$满足以下条件：

1 . 对于所有$k\geq k_0$，$r_k\notin[1/4,3/4]$.

2 . 存在无限多个k使得$r_k<1/4$.

3 . 存在无限多个k使得$r_k>3/4$.

4 . $|r_{k+1}-r_k|\to 0$, 当$k\to\infty$.

第4点是由于γ在有界集上是一致连续, 且假设条件$\|z^{k+1}-z^k\|\to 0$. 显然，不存在符合上述要求的序列. 因此，集合ω是连通的.

(iv) 设$\Psi^*=\lim_{k\to\infty}\Psi(z^k)$ 是有限值. 任取$z^*\in\omega(z^0)$, 存在一个收敛子序列z_q^k收敛到z^*. 由(1)有$\Psi(z^*)=\Psi^*$. 由z^* 的任意性，Ψ 在$\omega(z^0)$上恒等于Ψ^*.

分析

- 上面的结论表明从点z^0出发产生的点列$\{z^k\}$的极限点都是Ψ的临界点
- 从而迭代序列$\{z^k\}$的是子列收敛的，这至少保证了算法在迭代过程中与临界点越来越接近.
- 一个自然的问题就是：$\{z^k\}$全序列在何种条件下收敛？
- 这就要进入理论分析的第三个步骤：利用函数的KL 性质.

Kurdyka-Łojasiewicz 性质

Definition 2 (KL性质). *定义函数族Φ_η是凹连续函数φ： $[0,\eta) \to \mathbb{R}_+$的集合且满足如下条件：(i) $\varphi(0)=0$；(ii) φ 在$(0,\eta)$ 内连续可微，在点0处连续；(iii) 对任意的$s \in (0,\eta)$，都有$\varphi'(s)>0$.*

设σ： $\mathbb{R}^d \to (-\infty,+\infty]$是适当下半连续函数.

(1) 称σ在点$\bar{u} \in \mathbf{dom}\,\partial\sigma \overset{\text{def}}{=\!=} \{u \mid \partial\sigma(u) \neq \emptyset\}$处具有KL 性质，若存在$\eta \in (0,+\infty]$和$\bar{u}$的一个邻域$U$以及函数$\varphi \in \Phi_\eta$，使得

$$\forall\, u \in U \cap [\sigma(\bar{u}) < \sigma < \sigma(\bar{u}) + \eta],$$

以下不等式成立：

$$\varphi'(\sigma(u) - \sigma(\bar{u})) \cdot \mathrm{dist}(0, \partial\sigma(u)) \geq 1,$$

其中$\mathrm{dist}(x,S)$表示点x到集合S的距离.

(2) 若σ在$\mathbf{dom}\,\partial\sigma$上处处满足KL 性质，则称$\sigma$是一个KL 函数.

KL性质

- 若$\bar{u} \in \mathrm{dom}\sigma$不是适当闭函数$\sigma$的临界点，即$0 \notin \partial\sigma(\bar{u})$，则KL 性质在$\bar{u}$处自然成立.

Proof. 首先证明：$\exists c > 0$ 使得如下推导式成立

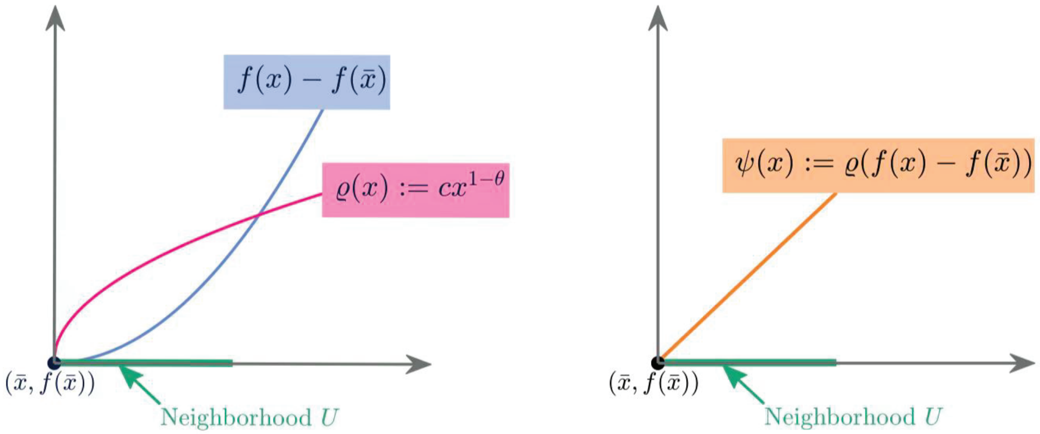
$$\|u - \bar{u}\|_2 + \|\sigma(u) - \sigma(\bar{u})\|_2 < c \implies \mathrm{dist}(0, \partial\sigma(u)) \geq c.$$

反证法假设存在序列$\{c_k\}, c_k > 0$ 且$c_k \to 0$, 存在点列$\{u^k\}$ 使得$\|u^k - \bar{u}\|_2 + \|\sigma(u^k) - \sigma(\bar{u})\|_2 < c_k$, 但是$\mathrm{dist}(0, \partial\sigma(u^k)) < c_k$, 也即$\exists w^k \in \partial\sigma(u^k)$ 满足$\|w^k\|_2 < c_k$. 由$\partial\sigma$ 的闭性以及$u^k \to \bar{u}$, $w^k \to 0$可知，$0 \in \partial\sigma(\bar{u})$. 矛盾.

然后：取$\varphi(t) = c^{-1}t, u \in B(\bar{u}, \frac{c}{2}) \cap [\sigma(\bar{u}) - \frac{c}{2} < \sigma(u) < \sigma(\bar{u}) + \frac{c}{2}]$, 可得KL不等式在$\bar{u}$ 处成立. □

KL性质

- 常用的desingularization 函数: $\varphi(t) = c \cdot t^{1-\theta}, \theta \in [0,1), c > 0;$
- KL 性质的几何图示



$$\varrho'(f(x) - f(\bar{x})) \cdot \|\nabla f(x)\| \geq 1 \quad \Longleftrightarrow \quad \psi'(x) \geq 1$$

KL性质

- 若$0 \in \partial\sigma(\bar{u})$，此时KL 性质保证了“函数$\sigma$可被锐化”. 令

$$\tilde{\varphi}(u) = \varphi(\sigma(u) - \sigma(\bar{u})),$$

KL 性质在某种条件下可以改写成

$$\mathrm{dist}(0, \partial\tilde{\varphi}(u)) \geq 1,$$

其中u的取法需要保证$\sigma(u) > \sigma(\bar{u})$.

- KL不等式表明，无论u多么接近临界点$\bar{u}$，$\tilde{\varphi}(u)$ 次梯度的模长均不小于1. 所以KL 性质也被称为是函数σ 在**重参数化子(reparameterization/desingularization)**φ下的一个**锐化**,这种几何性质在分析一阶算法收敛性时起到关键作用.

特殊KL函数-半代数函数

Definition 3 (半代数集与半代数函数). *称$\mathbb{R}^d$ 的子集S 是一个*半代数集*, 如果存在有限个实多项式函数$g_{ij}, h_{ij} : \mathbb{R}^d \to \mathbb{R}$ 使得*

$$S = \cup_{j=1}^{p} \cap_{i=1}^{q} \{u \in \mathbb{R}^d : g_{ij}(u) = 0, h_{ij}(u) < 0\}$$

*称函数$h : \mathbb{R}^d \to (-\infty, +\infty]$ 为*半代数函数*, 若h 的函数图像*

$$\{(u, t) \in \mathbb{R}^{d+1} : h(u) = t\}$$

是$\mathbb{R}^{d+1}$ 的半代数子集.

- 设$\sigma(u) : \mathbb{R}^d \to (-\infty, +\infty)$ 是下半连续适当函数, 若σ 是半代数函数, 则它在$\mathbf{dom}\,\sigma$中任一点处满足KL性质.

特殊的半代数函数与半代数集

- 实多项式函数.
- 半代数集的指示函数.
- 半代数函数的有限和与有限乘积.
- 半代数函数的复合.
- 上极限/下极限类函数. 例如，当g是半代数函数并且C 是半代数集时，$\sup\{g(u,v) : v \in C\}$ 是半代数的.
- 半正定矩阵锥, Stiefel流形以及恒秩矩阵都是半代数集.
- $S \subseteq \mathbb{R}^d$ 是非空半代数集，则函数$x \to dist(x,S)^2$ 是半代数的.
- $\|\cdot\|_0, \|\cdot\|_p$是半代数函数，其中p 是有理数.

一致KL性质

由于非凸问题有多个临界点，有时单个点$\bar{u}$处的KL 性质是不够的，我们需要引入一致KL 性质：

Lemma 1. *设Ω是紧集，$\sigma:\mathbb{R}^d\to(-\infty,+\infty]$是适当下半连续函数, σ在Ω上为常数且在Ω的每个点处都满足KL 性质，则存在$\varepsilon>0,\eta>0$, $\varphi\in\Phi_\eta$ (该函数族定义见2), 使得对任意$\bar{u}\in\Omega$ 和所有满足以下条件的u：*

$$\{u\in\mathbb{R}^d\ :\ \mathrm{dist}(u,\Omega)<\varepsilon\}\cap[\sigma(\bar{u})<\sigma<\sigma(\bar{u})+\eta],$$

有

$$\varphi'(\sigma(u)-\sigma(\bar{u}))\mathrm{dist}(0,\partial\sigma(u))\geq 1.$$

Lemma 1 证明

- 因为$\mathbb{R}^d$上的紧集可以由有限多个开集覆盖，因此该问题可在有限个点上进行讨论．设μ是σ在Ω上的取值．由于Ω是紧集，根据有限覆盖定理，存在有限多个开球$B(u_i, \varepsilon_i)$ (其中$u_i \in \Omega, i = 1, 2, \cdots, p$) 使得$\Omega \subset \bigcup_{i=1}^{p} B(u_i, \varepsilon_i)$．

- 现在考虑这些点u_i．在点u_i 上KL 性质成立，设$\varphi_i : [0, \eta_i) \to \mathbb{R}_+$是对应的重参数化子，则对任意$u \in B(u_i, \varepsilon_i) \cap [\mu < \sigma < \mu + \eta_i]$，有逐点的KL 性质：

$$\varphi_i'(\sigma(u) - \mu)\mathrm{dist}(0, \partial\sigma(u)) \geq 1.$$

取充分小的$\varepsilon > 0$使得

$$U_\varepsilon \stackrel{\text{def}}{=\!=} \{u \in \mathbb{R}^d \mid \mathrm{dist}(u, \Omega) \leq \varepsilon\} \subset \bigcup_{i=1}^{p} B(u_i, \varepsilon_i).$$

Lemma 1 证明(续)

- 取$\eta = \min_i \eta_i$, 以及
$$\varphi(s) = \sum_{i=1}^{p} \varphi_i(s)$$
容易验证$\varphi \in \Phi_\eta$.

- 对任意的$u \in U_\varepsilon \cap [\mu < \sigma < \mu + \eta]$，$u$必定落在某个球$B(u_{i_0}, \varepsilon_{i_0})$中，我们有
$$\begin{aligned}\varphi'(\sigma(u) - \mu)\mathrm{dist}(0, \partial\sigma(u)) &= \sum_{i}^{p} \varphi_i'(\sigma(u) - \mu)\mathrm{dist}(0, \partial\sigma(u)) \\ &\geq \varphi_{i_0}'(\sigma(u) - \mu)\mathrm{dist}(0, \partial\sigma(u)) \geq 1.\end{aligned}$$
即一致KL 性质成立.

有限长度性质

Theorem 6. *设Ψ是KL 函数，且满足假设条件，则以下结论成立：*

(1) 序列$\{z^k\}$的长度有限，即

$$\sum_{k=1}^{\infty} \|z^{k+1} - z^k\| < +\infty.$$

(2) 序列$\{z^k\}$收敛到Ψ的一个临界点$z^* = (x^*, y^*)$.

注:上述定理的(1) 结论强于$\|z^{k+1} - z^k\|$平方可和的结论, 前者说明从z^0出发，迭代序列的轨迹长度是有限的．这是推导全序列收敛的关键.

Theorem 6 证明

- 由于$\{z^k\}$是有界序列，存在收敛子列$\{z^{k_q}\} \to \bar{z}, q \to \infty$. 和之前的推导类似，对应的函数值列$\{\Psi(z^k)\}$总是收敛的，且

$$\lim_{k\to\infty} \Psi(z^k) = \Psi(\bar{z}). \tag{7}$$

以下不妨设$\Psi(\bar{z}) < \Psi(z^k)$. 否则，若存在$\bar{k}$使得$\Psi(z^{\bar{k}}) = \Psi(\bar{z})$, 由充分下降性可知$z^{\bar{k}+1} = z^{\bar{k}}$，则$z^k = z^{\bar{k}}, \forall\, k > \bar{k}$. 结论成立.

- 由极限(7) 和极限点集$\omega(z^0)$为紧集且$\lim_{k\to\infty} \operatorname{dist}(z^k, \omega(z^0)) = 0$可知, 对任意的$\varepsilon, \eta > 0$, 存在充分大的正整数$l$，使得对任意的$k > l$，

$$\Psi(\bar{z}) < \Psi(z^k) < \Psi(\bar{z}) + \eta, \quad \operatorname{dist}(z^k, \omega(z^0)) < \varepsilon.$$

- 结合Theorem 5(iv) (i.e., $\Psi(z) = \Psi(\bar{z}), \forall z \in \omega(z^0)$), 当$k$充分大时，迭代点序列最终会满足一致 KL 性质的前提.

Theorem 6 证明(续)

(1) 根据临界点的性质，$\omega(z^0)$是非空紧集，且Ψ在$\omega(z^0)$上是常数.
在一致KL性质中令$\Omega = \omega(z^0)$，对任意的$k > l$，

$$\varphi'(\Psi(z^k) - \Psi(\bar{z}))\mathrm{dist}(0, \partial\Psi(z^k)) \geq 1.$$

根据次梯度上界(Proposition 1)可知

$$\mathrm{dist}(0, \partial\Psi(z^k)) \leq \|(A_x^k, A_y^k)\| \leq \rho_2\|z^k - z^{k-1}\|.$$

代入KL 性质有

$$\varphi'(\Psi(z^k) - \Psi(\bar{z})) \geq \frac{1}{\rho_2}\|z^k - z^{k-1}\|^{-1}. \tag{8}$$

另外，由φ的凹性，有

$$\begin{aligned} &\varphi(\Psi(z^k) - \Psi(\bar{z})) - \varphi(\Psi(z^{k+1}) - \Psi(\bar{z})) \\ \geq &\varphi'(\Psi(z^k) - \Psi(\bar{z}))(\Psi(z^k) - \Psi(z^{k+1})). \end{aligned} \tag{9}$$

Theorem 6 证明(续)

定义常数$C = \frac{2\rho_2}{\rho_1} > 0$. 方便起见, 对任意正整数$p$ 与q, 定义

$$\Delta_{p,q} = \varphi(\Psi(z^p) - \Psi(\bar{z})) - \varphi(\Psi(z^q) - \Psi(\bar{z})).$$

根据不等式(9)，使用(8) 式和第一个步骤中的充分下降定理分别估计不等号右边的两项，有

$$\begin{aligned}
\Delta_{k,k+1} &\geq \varphi'(\Psi(z^k) - \Psi(\bar{z}))(\Psi(z^k) - \Psi(z^{k+1})) \\
&\geq \frac{1}{\rho_2}\|z^k - z^{k-1}\|^{-1} \cdot \frac{\rho_1}{2}\|z^{k+1} - z^k\|^2 \\
&= \frac{\|z^{k+1} - z^k\|^2}{C\|z^k - z^{k-1}\|},
\end{aligned}$$

等价于

$$\|z^{k+1} - z^k\| \leq \sqrt{C\Delta_{k,k+1}\|z^k - z^{k-1}\|}.$$

Theorem 6 证明(续)

根据基本不等式$2\sqrt{ab} \le a + b, \forall\, a, b > 0$，我们取$a = \|z^k - z^{k-1}\|$, $b = C\Delta_{k,k+1}$，则

$$2\|z^{k+1} - z^k\| \le \|z^k - z^{k-1}\| + C\Delta_{k,k+1}.$$

对任意的$k > l$，在上式中把k替换成i并对$i = l+1, l+2, \ldots, k$求和，得

$$\begin{aligned} 2\sum_{i=l+1}^{k} \|z^{i+1} - z^i\| &\le \sum_{i=l+1}^{k} \|z^i - z^{i-1}\| + C\sum_{i=l+1}^{k} \Delta_{i,i+1} \\ &\le \sum_{i=l+1}^{k} \|z^{i+1} - z^i\| + \|z^{l+1} - z^l\| + C\Delta_{l+1,k+1}. \end{aligned}$$

最后一个不等式是因为$\Delta_{p,q} + \Delta_{q,r} = \Delta_{p,r}$.

Theorem 6 证明(续)

左右两边抵消一个 $\sum\limits_{i=l+1}^{k} \|z^{i+1} - z^i\|$ 得到

$$
\begin{aligned}
&\sum_{i=l+1}^{k} \|z^{i+1} - z^i\| \\
\leq& \|z^{l+1} - z^l\| + C\Big(\varphi(\Psi(z^{l+1}) - \Psi(\bar{z})) - \varphi(\Psi(z^{k+1}) - \Psi(\bar{z}))\Big) \\
\leq& \|z^{l+1} - z^l\| + C\varphi(\Psi(z^{l+1}) - \Psi(\bar{z})).
\end{aligned}
\tag{10}
$$

不等式右边是有界的数且与k无关，由级数收敛的定义立即可得

$$\sum_{k=1}^{\infty} \|z^{k+1} - z^k\| < +\infty.$$

Theorem 6 证明(续)

(2) 在$\sum_{k=1}^{\infty}\|z^{k+1}-z^k\|<+\infty$的前提下$\{z^k\}$全序列收敛是显然的．这等价于证明$\{z^k\}$是柯西列．对任意$q>p>l$，

$$z^q-z^p=\sum_{k=p}^{q-1}(z^{k+1}-z^k),$$

根据三角不等式，

$$\|z^q-z^p\|=\left\|\sum_{k=p}^{q-1}(z^{k+1}-z^k)\right\|\le\sum_{k=p}^{q-1}\|z^{k+1}-z^k\|,$$

而$\|z^{k+1}-z^k\|$的可和性意味着$\sum_{k=l+1}^{\infty}\|z^{k+1}-z^k\|$趋于0．因此$\{z^k\}$是一个柯西列，算法产生的迭代序列有全序列收敛性.

收敛速率

Theorem 7. *设Ψ是KL 函数, 其中重参数化子为$\varphi(t) = c \cdot t^{1-\theta}$, $\theta \in [0,1)$, 且满足假设条件, 则算法产生的迭代点列$\{z^k\}$ 收敛到临界点z^* 且满足如下性质:*

(1) 若$\theta = 0$, 则$\{z^k\}$ 有限步收敛到z^*;

(2) 若$\theta \in (0, \frac{1}{2}]$, 则$\{\Phi(z^k)\}$ Q-线性收敛到$\Phi(z^*)$, 且$\{z^k\}$ R-线性收敛到z^*;

(3) 若$\theta \in (\frac{1}{2}, 1)$, 则存在$\omega > 0$ 使得对于充分大的k, $\|z^k - z^*\| \leq \omega k^{-\frac{1-\theta}{2\theta-1}}$, 即$\{z^k\}$ R- 次线性收敛到z^*.

Theorem 7 证明

证明: 对任意$k \geq 0$，令$\Delta_k = \sum_{i=k}^{\infty} \|z^{i+1} - z^i\|$, 由Thm 5知它是有限的. 根据三角不等式有$\Delta_k \geq \|z^k - z^*\|$, 因此要证明(2) 和(3) 我们只需证明$\Delta_k$有对应的上界. 不失一般性，假设对任意$k \geq 0$都有$\Delta_k > 0$. 根据KL不等式和次梯度上界, 我们有

$$\varphi'(\Psi(z^k) - \Psi(z^*)) \geq \frac{1}{\operatorname{dist}(0, \partial\Psi(z^k))} \geq \frac{1}{\rho_2\|z^k - z^{k-1}\|},$$

将$\varphi(t) = c \cdot t^{1-\theta}$代入上式有

$$\frac{c(1-\theta)}{(\Psi(z^k) - \Psi(z^*))^{\theta}} \geq \frac{1}{\rho_2\|z^k - z^{k-1}\|}, \tag{11}$$

整理可得

$$\Psi(z^k) - \Psi(z^*) \leq (c(1-\theta)\rho_2\|z^k - z^{k-1}\|)^{\frac{1}{\theta}}. \tag{12}$$

Theorem 7 证明(续)

利用不等式(10), 目标函数$\Psi(z^k)$单调递减到$\Psi(z^*)$和上述不等式，对充分大的k，我们有

$$\begin{aligned}
\Delta_k \leq& \|z^k - z^{k-1}\| + C\varphi(\Psi(z^k) - \Psi(z^*)) \\
=& \Delta_{k-1} - \Delta_k + C \cdot c(\Psi(z^k) - \Psi(z^*))^{1-\theta} \\
\leq& \Delta_{k-1} - \Delta_k + Cc(c(1-\theta)\rho_2\|z^k - z^{k-1}\|)^{\frac{1-\theta}{\theta}} \\
=& \Delta_{k-1} - \Delta_k + Cc(c(1-\theta)\rho_2)^{\frac{1-\theta}{\theta}}(\Delta_{k-1} - \Delta_k)^{\frac{1-\theta}{\theta}}.
\end{aligned} \tag{13}$$

$\diamond$ 考虑$\theta \in (\frac{1}{2}, 1)$.则$\frac{1-\theta}{\theta} < 1$. 因为当$k \to \infty$时$\Delta_k \to 0$，由上述不等式我们可知存在正整数$N_1$和正常数$C_1$使得

$$\Delta_k^{\frac{\theta}{1-\theta}} \leq C_1(\Delta_{k-1} - \Delta_k), \forall k \geq N_1.$$

Theorem 7 证明(续)

定义$h:(0,+\infty)\to\mathbb{R}$, $h(t)=t^{-\frac{\theta}{1-\theta}}$, 令$R\in(1,+\infty)$. 对$k\geq N_1$,

- 若$h(\Delta_k)\leq Rh(\Delta_{k-1})$. 上面的不等式可以写为

$$1\leq\frac{C_1(\Delta_{k-1}-\Delta_k)}{\Delta_k^{\frac{\theta}{1-\theta}}}$$

因此

$$\begin{aligned}1\leq&C_1(\Delta_{k-1}-\Delta_k)h(\Delta_k)\\ \leq&RC_1(\Delta_{k-1}-\Delta_k)h(\Delta_{k-1})\\ \leq&RC_1\int_{\Delta_k}^{\Delta_{k-1}}h(t)dt\\ \leq&RC_1\frac{1-\theta}{1-2\theta}[\Delta_{k-1}^{\frac{1-2\theta}{1-\theta}}-\Delta_k^{\frac{1-2\theta}{1-\theta}}]\end{aligned}$$

Theorem 7 证明(续)

因此如果令$\mu = \frac{2\theta-1}{(1-\theta)RC_1} > 0, \nu = \frac{1-2\theta}{1-\theta} < 0$我们有

$$0 < \mu \leq \Delta_k^\nu - \Delta_{k-1}^\nu.$$

- 若$h(\Delta_k) > Rh(\Delta_{k-1})$. 令$q = (\frac{1}{R})^{\frac{1-\theta}{\theta}} \in (0,1)$，则$\Delta_k \leq q\Delta_{k-1}$，再根据$\nu < 0$我们有

$$\Delta_k^\nu \geq q^\nu \Delta_{k-1}^\nu$$

$$\Delta_k^\nu - \Delta_{k-1}^\nu \geq (q^\nu - 1)\Delta_{k-1}^\nu.$$

由于$q^\nu - 1 > 0$且当$p \to +\infty$时$\Delta_p \to 0^+$，所以存在$\bar{\mu} > 0$使得$(q^\nu - 1)\Delta_{p-1}^\nu > \bar{\mu}, \forall p \geq N_1$. 即我们有

$$\Delta_k^\nu - \Delta_{k-1}^\nu \geq \bar{\mu} > 0.$$

Theorem 7 证明(续)

综上所述，若令$\hat{\mu} = \min\{\mu, \bar{\mu}\} > 0$，则

$$\Delta_k^\nu - \Delta_{k-1}^\nu \geq \hat{\mu} > 0, \forall k \geq N_1.$$

通过将上述不等式从N_1加到比N_1大的任意N我们可以得到$\Delta_N^\nu - \Delta_{N_1}^\nu \geq \hat{\mu}(N - N_1) > 0$. 因此，存在$\omega > 0$使得

$$\Delta_N \leq [\Delta_{N_1}^\nu + \hat{\mu}(N - N_1)]^{\frac{1}{\nu}} \leq \omega N^{\frac{1}{\nu}} = \omega N^{-\frac{1-\theta}{2\theta-1}}$$

结合$\Delta_N \geq \|z^N - z^*\|$ 可知$\|z^N - z^*\| \leq \omega N^{-\frac{1-\theta}{2\theta-1}}$, 即结论(3) 成立.

Theorem 7 证明(续)

◇ 考虑$\theta \in (0, \frac{1}{2}]$，由不等式(13)知对充分大的$k$，存在正常数$C_2 > 0$使得

$$\Delta_k \le C_2(\Delta_{k-1} - \Delta_k).$$

整理可得$\Delta_k \le \frac{C_2}{1+C_2}\Delta_{k-1}$，再根据证明开始时我们得到的$\Delta_k \ge \|z^k - z^*\|$有

$$\|z^k - z^*\| \le \Delta_k \le \frac{C_2}{1+C_2}\Delta_{k-1} \le \cdots \le (\frac{C_2}{1+C_2})^k\Delta_0,$$

即$\{z^k\}$ R-线性收敛到z^*.
结合不等式(12)整理可得

$$\|z^k - z^{k-1}\|^2 \ge \frac{(\Psi(z^k) - \Psi(z^*))^{2\theta}}{(c(1-\theta)\rho_2)^2}.$$

Theorem 7 证明(续)

再根据定理4中目标函数值的充分下降性，存在$C_3 > 0$使得对充分大的k有

$$\begin{aligned}
\Psi(z^{k+1}) - \Psi(z^k) &\leq -\frac{\rho_1}{2}\|z^{k+1} - z^k\|^2 \\
&\leq -\frac{\rho_1}{2}\frac{(\Psi(z^{k+1}) - \Psi(z^*))^{2\theta}}{(c(1-\theta)\rho_2)^2} \\
&\leq -\frac{\rho_1}{2}\frac{(\Psi(z^{k+1}) - \Psi(z^*))}{(c(1-\theta)\rho_2)^2} \qquad (\theta \in (0, 1/2]) \\
&= -C_3(\Psi(z^{k+1}) - \Psi(z^*)).
\end{aligned}$$

整理可得

$$\Psi(z^{k+1}) - \Psi(z^*) \leq \frac{1}{1+C_3}(\Psi(z^k) - \Psi(z^*)),$$

即$\{\Phi(z^k)\}$ Q-线性收敛到$\Phi(z^*)$. 综上可知结论(2)成立.

Theorem 7 证明(续)

◇ 考虑$\theta = 0$，令$I := \{k \in \mathbb{N} : z^{k+1} \neq z^k\}$，取$k$充分大, 在(11)代入$\theta = 0$得

$$\|z^k - z^{k-1}\| \geq \frac{1}{c\rho_2} > 0,$$

再由定理4中目标函数值的充分下降性得

$$\Psi(z^{k+1}) \leq \Psi(z^k) - \frac{\rho_1}{2}\|z^{k+1} - z^k\|^2 \leq \Psi(z^k) - \frac{\rho_1}{2c^2\rho_2^2}.$$

若I是无限集, 则上式两端令$k \to \infty$得$0 < -\frac{\rho_1}{2c^2\rho_2^2}$，矛盾. 故$I$为有限集，因此结论(1)成立.

KL性质与凸函数(with $\varphi(t) = c \cdot t^{1-\theta}$)

Example 1 (一致凸函数). *若f 是一致凸函数, i.e., 存在$p \geq 1$, 使得$\forall x, y \in \mathrm{dom} f, \forall u \in \partial f(x)$,*

$$f(y) \geq f(x) + u^T(y - x) + c\|y - x\|^p,$$

则f 在dom$f$ 上任意点处满足KL性质，且$\varphi(t) = pc^{-1/p} \cdot t^{1/p}$. ($p = 2$时为强凸函数)

Example 2 (带有增长条件的凸函数). *设$f : \mathbb{R}^n \to (-\infty, \infty]$ 是适当闭凸函数. 若f 满足如下增长条件: 取$\bar{x} \in \arg\min f \neq \emptyset$, 存在$p \geq 1$, 使得$\forall x, y \in \mathrm{dom} f, \bar{x}$ 的邻域$U, \exists \eta > 0, c > 0, r \geq 1$ 使得:*

$$f(x) \geq f(\bar{x}) + c \cdot (\mathrm{dist}(x, \arg\min f))^r, \forall x \in U \cap [f(\bar{x}) < f(x) < f(\bar{x}) + \eta],$$

则f 在$\bar{x}$ 处满足KL性质, 且$\varphi(t) = rc^{-1/r} \cdot t^{1/r}$.

Thank you!