Advances in Applied Mathematics 应用数学进展, 2019, 8(7), 1228-1238
Published Online July 2019 in Hans. http://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2019.87142

Hans汉斯

Inertial Proximal Alternating Minimization Algorithm for a Class of Nonconvex and Nonsmooth Problems

Mengxia Chen, Haiyan Zheng*

College of Mathematics and Information Science, Guangxi University, Nanning Guangxi
Email: cmxlovefam@outlook.com, *zhenghaiyan166@126.com

Received: Jun. 28th, 2019; accepted: Jul. 16th, 2019; published: Jul. 23rd, 2019

Abstract

In this paper, we propose an inertial proximal alternating minimization algorithm for a class of nonconvex and nonsmooth problems. We show the global convergence by constructing a new merit function *H* with guaranteed descent property. If *H* satisfies the Kurdyka-Lojasiewicz property, we determine the strongly convergence of the whole sequence.

Keywords

Nonconvex and Nonsmooth, Kurdyka-Lojasiewicz Property, Inertial, Alternating Minimization Algorithm

求解一类非凸非光滑问题的惯性邻近交替极小化算法

陈梦霞，郑海艳*

广西大学数学与信息科学学院，广西 南宁
Email: cmxlovefam@outlook.com, *zhenghaiyan166@126.com

收稿日期：2019年6月28日；录用日期：2019年7月16日；发布日期：2019年7月23日

摘 要

本文考虑一类非凸非光滑优化问题，提出了一种惯性邻近交替极小化算法。通过构造一个新的效益函数

*通讯作者。

文章引用：程梦霞，郑海艳. 求解一类非凸非光滑问题的惯性邻近交替极小化算法[J]. 应用数学进展，2019，8(7): 1228-1238. DOI: 10.12677/aam.2019.87142

***H*，并保证其具有下降性，证明了算法的全局收敛性。当*H*为Kurdyka-Lojasiewicz函数时，证明了算法的强收敛性。**

关键词

非凸非光滑，Kurdyka-Lojasiewicz性质，惯性，交替极小化方法

1. 引言

本文考虑如下非凸非光滑优化问题：

$$\min_{x,y} L(x,y)=f(x)+g(y)+R(x,y) \tag{1}$$

其中 $f:\mathbb{R}^n\to\mathbb{R}\cup\{\infty\}$， $g:\mathbb{R}^m\to\mathbb{R}\cup\{\infty\}$ 为正常下半连续函数， $R:\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}$ 连续可微。这一模型广泛应用于信号恢复、图像处理等问题。

求解问题(1)的常用方法是基于 Gauss-Seidel 迭代的交替极小化方法，迭代形式如下：

$$\begin{cases} x^{k+1}\in\arg\min\left\{L\left(x,y^k\right):x\in\mathbb{R}^n\right\}, \\ y^{k+1}\in\arg\min\left\{L\left(x^{k+1},y\right):y\in\mathbb{R}^m\right\} \end{cases} \tag{2}$$

在凸的情况下，[1]证明当 $L(x,y)$ 连续可微，且关于 x 或 y 严格凸时，序列 $\left\{\left(x^k,y^k\right)\right\}$ 的每一个极限点都是 $L(x,y)$ 的稳定点。[2]通过引入正则项去掉了[1]中的严格凸条件，提出邻近交替极小化方法(PAM)，迭代格式如下：

$$\begin{cases} x^{k+1}\in\arg\min\limits_{x}\left\{L\left(x,y^k\right)+\frac{c_k}{2}\left\|x-x^k\right\|^2\right\}, \\ y^{k+1}\in\arg\min\limits_{y}\left\{L\left(x^{k+1},y\right)+\frac{d_k}{2}\left\|y-y^k\right\|^2\right\} \end{cases}$$

对于问题(1)，本文在惯性思想的启发下，结合邻近交替极小化方法，提出一种惯性邻近交替极小化算法。惯性思想最早由 Alvarez [3]提出，用来解决动力系统问题。惯性思想是指为了得到下一步迭代点需要用到当前迭代点与上一步迭代点的信息，其优点是加快收敛速度。近年来，这一惯性类型的算法越来越受关注。如惯性向前向后分裂算法[4] [5]，惯性 Douglas-Rachford 分裂算法[6]，惯性 ADMM [7]等。

2. 预备知识

本节列出本文用到的基本概念及性质。

设 $\eta\in(0,+\infty]$，记 Φ_η 为满足如下条件的凹可微函数 $\varphi:[0,\eta)\to[0,+\infty)$ 的集合

(i) $\varphi=0$；

(ii) φ 在 $(0,\eta)$ 上连续；

(iii) $\varphi'(s)>0$， $\forall s\in(0,\eta)$。

定义 1 (Kurdyka-Lojasiewicz 性质)设 $\sigma:\mathbb{R}^d \to (-\infty,+\infty]$ 是正常下半连续函数。

(i) 设 $\bar{u}\in \operatorname{dom}\partial\sigma := \{u\in\mathbb{R}^d : \partial\sigma(u)\neq\varnothing\}$，若存在 $\eta\in(0,+\infty]$，$\bar{u}$ 的一个邻域 U 及 $\varphi\in\Phi_\eta$ 有

$$\varphi'(\sigma(u)-\sigma(\bar{u}))\operatorname{dist}(0,\partial\sigma(u))\geq 1,\quad \forall u\in U\cap[\sigma(\bar{u})<\sigma(u)<\sigma(\bar{u})+\eta]$$

则称 σ 在 $\bar{u}$ 处有 Kurdyka-Lojasiewicz (KL)性质。

(ii) 若 σ 在 $\operatorname{dom}\partial\sigma$ 中的每一点都满足 KL 性质，则称 σ 为 KL 函数。

引理 1 (一致 Kurdyka-Lojasiewicz 性质)设集合 Ω 是紧集，$\sigma:\mathbb{R}^d\to\mathbb{R}\cup\{\infty\}$ 是正常下半连续函数。若 σ 在 Ω 上为常数，且在 Ω 的每一点处都满足 KL 性质。设 $\bar{u}\in\Omega$，则存在 $\varepsilon>0,\eta>0,\varphi\in\Phi_\eta$ 使得

$$\varphi'(\sigma(u)-\sigma(\bar{u}))\operatorname{dist}(0,\partial\sigma(u))\geq 1,\quad \forall\{u\in\mathbb{R}^d : d(u,\Omega)<\varepsilon\}\cap\{\sigma(\bar{u})<\sigma(u)<\sigma(\bar{u})+\eta\}$$

引理 2 [8] $h:\mathbb{R}^n\to\mathbb{R}$ 是连续可微函数，∇h Lipschitz 连续(Lipschitz 常数为 L)，则对任意 $x,y\in\mathbb{R}^n$，有

$$\left|h(y)-h(x)-\langle\nabla h(x),y-x\rangle\right|\leq\frac{h}{2}\|y-x\|^2$$

3. 算法及假设条件

本文需要如下假设条件：

假设 1 (i) $\inf L(x,y)=f(x)+g(y)+R(x,y)>-\infty$，

(ii) 对任意固定的 y，$\nabla_x R(x,y)$ 关于 x 是 Lipschitz 连续的，$L_1(y)$ 为 Lipschitz 常数，即

$$\|\nabla_x R(x_1,y)-\nabla_x R(x_2,y)\|\leq L_1(y)\|x_1-x_2\|,\ \forall x_1,x_2\in\mathbb{R}^n$$

(iii) 对任意固定的 x，$\nabla_y R(x,y)$ 关于 y 是 Lipschitz 连续的，$L_2(x)$ 为 Lipschitz 常数，即

$$\|\nabla_y R(x,y_1)-\nabla_y R(x,y_2)\|\leq L_2(x)\|y_1-y_2\|,\ \forall y_1,y_2\in\mathbb{R}^m$$

(iv) 存在 $\lambda_i^-,\lambda_i^+>0\quad(i=1,2)$，使得 $\lambda_1^-\leq L_1(y^k)\leq\lambda_1^+$，$\lambda_2^-\leq L_2(x^k)\leq\lambda_2^+$。

(v) ∇R 在 $\mathbb{R}^n\times\mathbb{R}^m$ 的有界子集上 Lipschitz 连续，即对任意有界集 $B_1\times B_2$，$\exists m>0$，使得 $\forall(x_i,y_i)\in B_1\times B_2$，$i=1,2$，有

$$\|(\nabla_x R(x_1,y_1)-\nabla_x R(x_2,y_2),\nabla_y R(x_1,y_1)-\nabla_y R(x_2,y_2))\|\leq m\|(x_1-x_2,y_1-y_2)\|$$

惯性邻近交替极小化算法(IPAM)：

步 0. 给定初始点 $(x^0,y^0)\in\mathbb{R}^n\times\mathbb{R}^m$，选取 $\gamma_1>1$，$\gamma_2>1$。

步 1. 令 $c_k=\gamma_1 L_1(y^k)$，计算

$$x^{k+1}\in\arg\min\left\{L(x,y^k)+\frac{c_k}{2}\|x-\bar{x}^k\|^2\right\}\tag{3}$$

步 2. 取 $\theta\in(0,1)$，并计算

$$\bar{x}^{k+1}=x^{k+1}+\theta(x^{k+1}-\bar{x}^k)\tag{4}$$

步 3. 令 $d_k=\gamma_2 L_2(\bar{x}^{k+1})$，计算

$$y^{k+1}\in\arg\min\left\{L(\bar{x}^{k+1},y)+\frac{d_k}{2}\|y-y^k\|^2\right\}\tag{5}$$

4. 收敛性分析

记 $\{z^k\}=\{(x^k,y^k)\}$ 为惯性邻近交替极小化算法产生的序列。

4.1. 充分下降性

引理 3：若假设 1 成立，则

$$L\left(x^{k+1},y^{k+1}\right)+M_1\left\|\overline{x}^{k+1}-x^{k+1}\right\|^2+M_2\left\|y^{k+1}-y^k\right\|^2\le L\left(x^k,y^k\right)+M_3\left\|\overline{x}^k-x^k\right\|^2 \tag{6}$$

其中 $M_1=\dfrac{\gamma_1\lambda_1^-}{2\theta^2}-\mu m^2$， $M_2=\dfrac{\gamma_2\lambda_2^-}{2}-\lambda_2^+-\mu m^2-\dfrac{1}{3\mu}$， $M_3=\dfrac{\gamma_1\lambda_1^+}{2}$。

证明：由(3)，(5)知，

$$f\left(x^{k+1}\right)+R\left(x^{k+1},y^k\right)+\frac{c_k}{2}\left\|x^{k+1}-\overline{x}^k\right\|^2\le f\left(x^k\right)+R\left(x^k,y^k\right)+\frac{c_k}{2}\left\|x^k-\overline{x}^k\right\|^2 \tag{7}$$

$$g\left(y^{k+1}\right)+R\left(\overline{x}^{k+1},y^{k+1}\right)+\frac{d_k}{2}\left\|y^{k+1}-y^k\right\|^2\le g\left(y^k\right)+R\left(\overline{x}^{k+1},y^k\right) \tag{8}$$

由引理 2 得

$$R\left(x^{k+1},y^k\right)\ge R\left(x^{k+1},y^{k+1}\right)-\left\langle\nabla_y R\left(x^{k+1},y^k\right),y^{k+1}-y^k\right\rangle-\frac{L_2\left(x^{k+1}\right)}{2}\left\|y^{k+1}-y^k\right\|^2$$

$$R\left(\overline{x}^{k+1},y^{k+1}\right)\ge R\left(\overline{x}^{k+1},y^k\right)+\left\langle\nabla_y R\left(\overline{x}^{k+1},y^{k+1}\right),y^{k+1}-y^k\right\rangle-\frac{L_2\left(\overline{x}^{k+1}\right)}{2}\left\|y^{k+1}-y^k\right\|^2$$

将上述不等式代入(7)，(8)得

$$\begin{aligned}&f\left(x^{k+1}\right)+R\left(x^{k+1},y^{k+1}\right)-\left\langle\nabla_y R\left(x^{k+1},y^k\right),y^{k+1}-y^k\right\rangle-\frac{L_2\left(x^{k+1}\right)}{2}\left\|y^{k+1}-y^k\right\|^2+\frac{c_k}{2}\left\|x^{k+1}-\overline{x}^k\right\|^2\\&\le f\left(x^k\right)+R\left(x^k,y^k\right)+\frac{c_k}{2}\left\|x^k-\overline{x}^k\right\|^2\end{aligned}$$

$$g\left(y^{k+1}\right)+\left\langle\nabla_y R\left(\overline{x}^{k+1},y^{k+1}\right),y^{k+1}-y^k\right\rangle-\frac{L_2\left(\overline{x}^{k+1}\right)}{2}\left\|y^{k+1}-y^k\right\|^2+\frac{d_k}{2}\left\|y^{k+1}-y^k\right\|^2\le g\left(y^k\right)$$

以上两式相加得

$$\begin{aligned}L\left(x^{k+1},y^{k+1}\right)\le{}& L\left(x^k,y^k\right)+\left\langle\nabla_y R\left(x^{k+1},y^k\right)-\nabla_y R\left(\overline{x}^{k+1},y^{k+1}\right),y^{k+1}-y^k\right\rangle\\&-\frac{c_k}{2\theta^2}\left\|\overline{x}^{k+1}-x^{k+1}\right\|^2+\frac{c_k}{2}\left\|\overline{x}^k-x^k\right\|^2-\left(\frac{d_k}{2}-\frac{L_2\left(x^{k+1}\right)}{2}-\frac{L_2\left(\overline{x}^{k+1}\right)}{2}\right)\left\|y^{k+1}-y^k\right\|^2\end{aligned}$$

由不等式 $\langle\alpha,\beta\rangle\le\|\alpha\|\cdot\|\beta\|$， $2ab\le\mu a^2+\dfrac{1}{\mu}b^2$， $(\mu>0)$，及假设(v)可知，

$$\begin{aligned}&\left\langle\nabla_y R\left(x^{k+1},y^k\right)-\nabla_y R\left(\overline{x}^{k+1},y^{k+1}\right),y^{k+1}-y^k\right\rangle\\&\le\frac{\mu m^2}{2}\left(\left\|\overline{x}^{k+1}-x^{k+1}\right\|+\left\|y^{k+1}-y^k\right\|\right)^2+\frac{1}{2\mu}\left\|y^{k+1}-y^k\right\|^2\\&\le\mu m^2\left\|\overline{x}^{k+1}-x^{k+1}\right\|^2+\left(\mu m^2+\frac{1}{2\mu}\right)\left\|y^{k+1}-y^k\right\|^2\end{aligned}$$

进一步由 $c_k=\gamma_1L_1\left(y^k\right)$， $d_k=\gamma_2L_2\left(\overline{x}^{k+1}\right)$ 及假设(ii)，得

$$L\left(x^{k+1},y^{k+1}\right)\le L\left(x^k,y^k\right)-\left(\frac{\gamma_1\lambda_1^-}{2\theta^2}-\mu m^2\right)\left\|\overline{x}^{k+1}-x^{k+1}\right\|^2+\frac{\gamma_1\lambda_1^+}{2}\left\|\overline{x}^k-x^k\right\|^2$$
$$-\left(\frac{\gamma_2\lambda_2^-}{2}-\lambda_2^+-\mu m^2-\frac{1}{2\mu}\right)\left\|y^{k+1}-y^k\right\|^2$$

移项后可证得(6)式成立。

注 1. 若参数 θ,γ_1,γ_2 满足下列不等式条件

$$\theta^2<\frac{\gamma_1\lambda_1^-}{\gamma_1\lambda_1^++2\mu m^2},\quad \gamma_1>\frac{2\theta^2\mu m^2}{\lambda_1^-},\quad \gamma_2>\frac{2\left(\lambda_2^++\mu m^2+\frac{1}{2\mu}\right)}{\lambda_2^-}$$

则有 $M_1,M_2,M_3>0$，且 $M_1>M_3$。

注 2. 记 $\left\{\hat{z}^k\right\}=\left\{\left(y^{k-1},x^k,\overline{x}^k,y^k\right)\right\}$，$H\left(\tilde{y},x,\tilde{x},y\right)=L\left(x,y\right)+M_3\left\|x-\tilde{x}\right\|^2+\frac{M_2}{2}\left\|y-\tilde{y}\right\|^2$。则有

$$H\left(\hat{z}^k\right)-H\left(\hat{z}^{k+1}\right)\ge\rho_1\left(\left\|\overline{x}^{k+1}-x^{k+1}\right\|^2+\left\|y^{k+1}-y^k\right\|^2\right)\tag{9}$$

其中 $\rho_1=\min\left\{M_1-M_3,\frac{M_2}{2}\right\}$。

4.2. 次微分的范数估计

引理 4：若假设 1 成立，$\left\{z^k=\left(x^k,y^k\right)\right\}$ 有界，令

$$A_x^{k+1}:=-\frac{c_k}{\theta}\left(\overline{x}^{k+1}-x^{k+1}\right)-\nabla_xR\left(x^{k+1},y^k\right)+\nabla_xR\left(x^{k+1},y^{k+1}\right)$$
$$A_y^{k+1}:=-d_k\left(y^{k+1}-y^k\right)-\nabla_yR\left(\overline{x}^{k+1},y^{k+1}\right)+\nabla_yR\left(x^{k+1},y^{k+1}\right)$$
$$\omega^{k+1}:=\begin{pmatrix}M_2\left(y^k-y^{k+1}\right)\\ A_x^{k+1}+2M_3\left(x^{k+1}-\overline{x}^{k+1}\right)\\ 2M_3\left(\overline{x}^{k+1}-x^{k+1}\right)\\ A_y^{k+1}+M_2\left(y^{k+1}-y^k\right)\end{pmatrix}\tag{10}$$

则 $\left(A_x^{k+1},A_y^{k+1}\right)\in\partial L\left(x^{k+1},y^{k+1}\right)$，$\omega^{k+1}\in\partial H\left(\hat{z}^{k+1}\right)$。更进一步，存在常数 $C_1>0$，使得

$$\left\|\omega^{k+1}\right\|\le C_1\left(\left\|\overline{x}^{k+1}-x^{k+1}\right\|+\left\|y^{k+1}-y^k\right\|\right)$$

证明：由最优性条件可知，

$$0\in\partial_xL\left(x^{k+1},y^k\right)+c_k\left(x^{k+1}-\overline{x}^k\right)=\partial f\left(x^{k+1}\right)+\nabla_xR\left(x^{k+1},y^k\right)+\frac{c_k}{\theta}\left(\overline{x}^{k+1}-x^{k+1}\right)$$
$$0\in\partial_yL\left(\overline{x}^{k+1},y^{k+1}\right)+d_k\left(y^{k+1}-y^k\right)=\partial g\left(y^{k+1}\right)+\nabla_yR\left(\overline{x}^{k+1},y^{k+1}\right)+d_k\left(y^{k+1}-y^k\right)$$

因此，

$$A_x^{k+1}:=-\frac{c_k}{\theta}\left(\overline{x}^{k+1}-x^{k+1}\right)-\nabla_xR\left(x^{k+1},y^k\right)+\nabla_xR\left(x^{k+1},y^{k+1}\right)$$
$$\in\partial f\left(x^{k+1}\right)+\nabla_xR\left(x^{k+1},y^{k+1}\right)=\partial_xL\left(x^{k+1},y^{k+1}\right)$$

$$A_y^{k+1} := -d_k\left(y^{k+1}-y^k\right)-\nabla_y R\left(\overline{x}^{k+1},y^{k+1}\right)+\nabla_y R\left(x^{k+1},y^{k+1}\right)$$
$$\in \partial g\left(y^{k+1}\right)+\nabla_y R\left(x^{k+1},y^{k+1}\right)=\partial_y L\left(x^{k+1},y^{k+1}\right)$$

根据函数 H 的定义，得

$$\partial H\left(\tilde{y},x,\tilde{x},y\right)=\begin{pmatrix} M_2\left(\tilde{y}-y\right) \\ \partial_x L\left(x,y\right)+2M_3\left(x-\tilde{x}\right) \\ 2M_3\left(\tilde{x}-x\right) \\ \partial_y L\left(x,y\right)+M_2\left(y-\tilde{y}\right) \end{pmatrix}$$

故 $\omega^{k+1}\in\partial H\left(\hat{z}^{k+1}\right)$。

由假设 1(v)，得

$$\left\|A_x^{k+1}\right\|\le\frac{c_k}{\theta}\left\|\overline{x}^{k+1}-x^{k+1}\right\|+\left\|\nabla_x R\left(x^{k+1},y^{k+1}\right)-\nabla_x R\left(x^{k+1},y^k\right)\right\|$$
$$\le\frac{\gamma_1\lambda_1^+}{\theta}\left\|\overline{x}^{k+1}-x^{k+1}\right\|+m\left\|y^{k+1}-y^k\right\|$$
$$\left\|A_y^{k+1}\right\|\le d_k\left\|y^{k+1}-y^k\right\|+\left\|\nabla_y R\left(x^{k+1},y^{k+1}\right)-\nabla_y R\left(\overline{x}^{k+1},y^{k+1}\right)\right\|$$
$$\le\gamma_2\lambda_2^+\left\|y^{k+1}-y^k\right\|+m\left\|\overline{x}^{k+1}-x^{k+1}\right\|$$

因此，存在 $C_1>0$，使得 $\left\|\omega^{k+1}\right\|\le C_1\left(\left\|\overline{x}^{k+1}-x^{k+1}\right\|+\left\|y^{k+1}-y^k\right\|\right)$。

引理 5：若假设 1 成立，$\left\{z^k\right\}$ 有界，则 $\sum_{k=0}^{\infty}\left\|z^{k+1}-z^k\right\|^2<\infty$。

证明：由假设 1(i)知 $L(x,y)$ 有下界，故 $H\left(\tilde{y},x,\tilde{x},y\right)$ 也有下界，再由注 2，$\left\{H\left(\hat{z}^k\right)\right\}$ 非增。根据单调有界定理可知，$\left\{H\left(\hat{z}^k\right)\right\}$ 收敛，其极限为 $\overline{H}$。

令 N 为正整数，将(9)式从 $k=0$ 到相加 $N-1$，得

$$H\left(\hat{z}^0\right)-H\left(\hat{z}^N\right)\ge\sum_{k=0}^{N-1}\rho_1\left(\left\|\overline{x}^{k+1}-x^{k+1}\right\|^2+\left\|y^{k+1}-y^k\right\|^2\right)$$

令 N 趋于无穷，得

$$\infty>H\left(\hat{z}^0\right)-\overline{H}\ge\sum_{k=0}^{\infty}\rho_1\left(\left\|\overline{x}^{k+1}-x^{k+1}\right\|^2+\left\|y^{k+1}-y^k\right\|^2\right)$$

故 $\sum_{k=0}^{\infty}\left\|\overline{x}^{k+1}-x^{k+1}\right\|<\infty$，$\sum_{k=0}^{\infty}\left\|y^{k+1}-y^k\right\|<\infty$。

又

$$\sum_{k=0}^{\infty}\left\|x^{k+1}-x^k\right\|^2\le\sum_{k=0}^{\infty}\left(\left\|x^{k+1}-\overline{x}^k\right\|+\left\|\overline{x}^k-x^k\right\|\right)^2$$
$$\le 2\left(\frac{1}{\theta^2}\sum_{k=0}^{\infty}\left\|\overline{x}^{k+1}-x^{k+1}\right\|^2+\sum_{k=0}^{\infty}\left\|\overline{x}^k-x^k\right\|^2\right)<\infty$$

因此，$\sum_{k=0}^{\infty}\left\|z^{k+1}-z^k\right\|^2=\sum_{k=0}^{\infty}\left(\left\|x^{k+1}-x^k\right\|^2+\left\|y^{k+1}-y^k\right\|^2\right)<\infty$。

4.3. 收敛结论

记 $z^*=\left(x^*,y^*\right)$ 为序列 $\left\{z^k\right\}$ 的一个极限点，$\hat{z}^*=\left(y^*,x^*,\overline{x}^*,y^*\right)$。

定理 1：设 Ω^* 为序列 $\{\hat{z}^k\}$ 的极限点集。若假设 1 成立，$\{z^k\}$ 有界，则有以下结论成立

(i) Ω^* 是非空紧的连通集，且 $\lim\limits_{k\to\infty}\operatorname{dist}(\hat{z}^k,\Omega^*)=0$，

(ii) $\Omega^*\subseteq\operatorname{crit}H$，

(iii) 当且仅当 $\bar{x}^*=x^*$ 时，$\hat{z}^*\in\Omega^*$，

(iv) $z^*\in\operatorname{crit}L$，

(v) H 在 Ω^* 上恒为常数。

证明：(i) 证明可参照文献[9]引理 5。

(ii) 对任意 $\hat{z}^*\in\Omega^*$，则存在一个子序列 $\{\hat{z}^{k_j}\}$，有 $\lim\limits_{k_j\to\infty}\hat{z}^{k_j}=\hat{z}^*$。由 f 的下半连续性，有 $\liminf\limits_{k_j\to\infty} f(x^{k_j})\ge f(x^*)$。

再由 x^{k+1} 的定义得，

$$\begin{aligned}&f(x^{k+1})+R(x^{k+1},y^k)+g(y^k)+\frac{c_k}{2}\|x^{k+1}-\bar{x}^k\|^2\\&\le f(x^*)+R(x^*,y^k)+g(y^k)+\frac{c_k}{2}\|x^*-\bar{x}^k\|^2\end{aligned}$$

取 $k=k_j-1$，有

$$\begin{aligned}&f(x^{k_j})+R(x^{k_j},y^{k_j-1})+g(y^{k_j-1})+\frac{c_{k_j-1}}{2}\|x^{k_j}-\bar{x}^{k_j-1}\|^2\\&\le f(x^*)+R(x^*,y^{k_j-1})+g(y^{k_j-1})+\frac{c_{k_j-1}}{2}\|x^*-\bar{x}^{k_j-1}\|^2\end{aligned}\tag{11}$$

在引理 5 的证明中得，$\lim\limits_{k\to\infty}\|\bar{x}^k-x^k\|=0,\lim\limits_{k\to\infty}\|x^{k+1}-x^k\|=0,\lim\limits_{k\to\infty}\|y^{k+1}-y^k\|=0$。

而 $\|x^{k+1}-\bar{x}^k\|=\frac{1}{\theta}\|\bar{x}^{k+1}-x^{k+1}\|,\|\bar{x}^k-x^*\|\le\|\bar{x}^k-x^k\|+\|x^k-x^*\|$，对(11)式中的 k_j 取上极限，得 $\limsup\limits_{k_j\to\infty} f(x^{k_j})\le f(x^*)$。因此，$\lim\limits_{k_j\to\infty} f(x^{k_j})=f(x^*)$。同样地，$\lim\limits_{k_j\to\infty} g(y^{k_j})=g(y^*)$。根据 R 的连续性，$R(x^{k_j},y^{k_j})\to R(x^*,y^*)\quad(k_j\to\infty)$。故

$$\lim_{k_j\to\infty}L(x^{k_j},y^{k_j})=L(x^*,y^*)$$

$$\lim_{k_j\to\infty}H(\hat{z}^{k_j})=\lim_{k_j\to\infty}\left\{L(x^{k_j},y^{k_j})+M_3\|\bar{x}^{k_j}-x^{k_j}\|^2+\frac{M_2}{2}\|y^{k_j}-y^{k_j-1}\|^2\right\}=L(x^*,y^*)=H(\hat{z}^*)$$

由引理 4 及(10)知 $\omega^{k_j}\in\partial H(\hat{z}^{k_j})$，$\omega^{k_j}\to 0\quad(k_j\to\infty)$。再由 ∂H 的闭性，有 $0\in\partial H(\hat{z}^*)$。因此，$\Omega^*\subseteq\operatorname{crit}H$。

(iii) 由引理 5 及 $\hat{z}^k$ 的定义，结论得证。

(iv) 由引理 4 及引理 5 知 $(A_x^{k_j},A_y^{k_j})\in\partial L(x^{k_j},y^{k_j})$，$\|x^{k+1}-x^k\|\to 0$，$\|y^{k+1}-y^k\|\to 0$。则 $(A_x^{k_j},A_y^{k_j})\to 0$ ($k_j\to\infty$)，再由 ∂L 的闭性，有 $0\in\partial L(z^*)$，即 $z^*\in\operatorname{crit}L$。

(v) 记 $\lim\limits_{k\to\infty}H(\hat{z}^k)=h$，($h$ 为常数)。对任意 $\hat{z}^*\in\Omega^*$，则存在一个子序列 $\{\hat{z}^{k_j}\}$，使得 $\lim\limits_{k_j\to\infty}\hat{z}^{k_j}=\hat{z}^*$。再根据(iii)证明过程中所得 $\lim\limits_{k_j\to\infty}H(\hat{z}^{k_j})=H(\hat{z}^*)$。于是，$H(\hat{z}^*)=h$，即 H 在 Ω^* 上恒为常数。

定理 2：假设 H 是 Kurdyka-Lojasiewicz 函数，$\{z^k\}$ 有界，则

(i) $\sum_{k=0}^{\infty}\left\|z^{k+1}-z^{k}\right\|^{2}<\infty$，

(ii)序列 $\left\{z^{k}\right\}$ 收敛到 L 的一个稳定点 $\left(x^{*},y^{*}\right)$。

证明： 由定理 1， $\lim_{k_j\to\infty}H\left(\hat{z}^{k_j}\right)=H\left(\hat{z}^{*}\right)$， $\forall\hat{z}^{*}\in\Omega^{*}$。接下来，将分成两种情况证明。

Case1. 存在一个正整数 k_0，使得 $H\left(\hat{z}^{k_0}\right)=H\left(\hat{z}^{*}\right)$。由注 2 知 $H\left(\hat{z}^{k}\right)=H\left(\hat{z}^{*}\right)$，$\forall k\ge k_0$，则 $\hat{z}^{k}=\hat{z}^{k_0}$，$\forall k\ge k_0$。于是，定理得证。

Case2. 对所有 $k\ge 0$，有 $H\left(\hat{z}^{k}\right)>H\left(\hat{z}^{*}\right)$。根据极限的定义，由 $\lim_{k_j\to\infty}H\left(\hat{z}^{k_j}\right)=H\left(\hat{z}^{*}\right)$，则对给定 $\eta>0$，$\exists k_1>0$，$k_1\in\mathbb{N}$，使得当 $k\ge k_1$ 时，有 $H\left(\hat{z}^{k}\right)<H\left(\hat{z}^{*}\right)+\eta$。由定理 1(i)，即对给定 $\varepsilon>0$，$\exists k_2>0$，$k_2\in\mathbb{N}$，使得当 $k\ge k_2$ 时，有 $\operatorname{dist}\left(\hat{z}^{k},\Omega^{*}\right)<\varepsilon$。于是，对给定 $\eta,\varepsilon>0$，有 $\operatorname{dist}\left(\hat{z}^{k},\Omega^{*}\right)<\varepsilon$，$H\left(\hat{z}^{*}\right)<H\left(\hat{z}^{k}\right)<H\left(\hat{z}^{*}\right)+\eta$，$\forall k>k_3=\max\left\{k_1,k_2\right\}$。

(i) 由引理 1，则 $\forall k>k_3$，有

$$\varphi'\left(H\left(\hat{z}^{k}\right)-H\left(\hat{z}^{*}\right)\right)\operatorname{dist}\left(0,\partial H\left(\hat{z}^{k}\right)\right)\ge 1$$

由 φ 的凹性，

$$\begin{aligned}&\varphi\left(H\left(\hat{z}^{k}\right)-H\left(\hat{z}^{*}\right)\right)-\varphi\left(H\left(\hat{z}^{k+1}\right)-H\left(\hat{z}^{*}\right)\right)\\&\ge\varphi'\left(H\left(\hat{z}^{k}\right)-H\left(\hat{z}^{*}\right)\right)\left(H\left(\hat{z}^{k}\right)-H\left(\hat{z}^{k+1}\right)\right)\\&\ge\frac{H\left(\hat{z}^{k}\right)-H\left(\hat{z}^{k+1}\right)}{\operatorname{dist}\left(0,\partial H\left(\hat{z}^{k}\right)\right)}\end{aligned}\tag{12}$$

记 $\Delta_{k,k+1}=\varphi\left(H\left(\hat{z}^{k}\right)-H\left(\hat{z}^{*}\right)\right)-\varphi\left(H\left(\hat{z}^{k+1}\right)-H\left(\hat{z}^{*}\right)\right)$。利用注 2，结合(12)式，有

$$\Delta_{k,k+1}\ge\frac{\rho_1\left(\left\|\bar{x}^{k+1}-x^{k+1}\right\|^{2}+\left\|y^{k+1}-y^{k}\right\|^{2}\right)}{C_1\left(\left\|\bar{x}^{k}-x^{k}\right\|+\left\|y^{k}-y^{k-1}\right\|\right)}$$

移项后，不等号两边同时开方，根据不等式 $2\sqrt{ab}\le a+b$ 得

$$\sqrt{\left\|\bar{x}^{k+1}-x^{k+1}\right\|^{2}+\left\|y^{k+1}-y^{k}\right\|^{2}}\le\frac{1}{4}\left(\left\|\bar{x}^{k}-x^{k}\right\|+\left\|y^{k}-y^{k-1}\right\|\right)+\frac{C_1}{\rho_1}\Delta_{k,k+1}\tag{13}$$

再由 $\left(a+b\right)^{2}\le 2\left(a^{2}+b^{2}\right)$ 得

$$\left\|\bar{x}^{k+1}-x^{k+1}\right\|+\left\|y^{k+1}-y^{k}\right\|\le\sqrt{2\left(\left\|\bar{x}^{k+1}-x^{k+1}\right\|^{2}+\left\|y^{k+1}-y^{k}\right\|^{2}\right)}$$

将(13)式代入上式，得

$$\left\|\bar{x}^{k+1}-x^{k+1}\right\|+\left\|y^{k+1}-y^{k}\right\|\le\frac{\sqrt{2}}{4}\left(\left\|\bar{x}^{k}-x^{k}\right\|+\left\|y^{k}-y^{k-1}\right\|\right)+\frac{\sqrt{2}C_1}{\rho_1}\Delta_{k,k+1}$$

将上式从 $k=s+1,\cdots,N$ 相加得

$$\begin{aligned}&\sum_{k=s+1}^{N}\left(\left\|\bar{x}^{k+1}-x^{k+1}\right\|+\left\|y^{k+1}-y^{k}\right\|\right)\\&\le\frac{\sqrt{2}}{4}\sum_{k=s+1}^{N}\left(\left\|\bar{x}^{k+1}-x^{k+1}\right\|+\left\|y^{k+1}-y^{k}\right\|\right)+\sum_{k=s+1}^{N}\frac{\sqrt{2}C_1}{\rho_1}\Delta_{k,k+1}+\frac{\sqrt{2}}{4}\left(\left\|\bar{x}^{s+1}-x^{s+1}\right\|+\left\|y^{s+1}-y^{s}\right\|\right)\end{aligned}$$

因为 $\varphi>0$，则

$$\sum_{k=s+1}^{N}\Delta_{k,k+1}=\varphi\left(H\left(\hat{z}^{s+1}\right)-H\left(\hat{z}^{*}\right)\right)-\varphi\left(H\left(\hat{z}^{N+1}\right)-H\left(\hat{z}^{*}\right)\right)\le\varphi\left(H\left(\hat{z}^{s+1}\right)-H\left(\hat{z}^{*}\right)\right)$$

于是，

$$\begin{aligned}&\left(1-\frac{\sqrt{2}}{4}\right)\sum_{k=s+1}^{N}\left(\left\|\overline{x}^{k+1}-x^{k+1}\right\|+\left\|y^{k+1}-y^{k}\right\|\right)\\&\le\frac{\sqrt{2}C_1}{\rho_1}\varphi\left(H\left(\hat{z}^{s+1}\right)-H\left(\hat{z}^{*}\right)\right)+\frac{\sqrt{2}}{4}\left(\left\|\overline{x}^{s+1}-x^{s+1}\right\|+\left\|y^{s+1}-y^{s}\right\|\right)\end{aligned}$$

令 N 趋于无穷，得

$$\sum_{k=0}^{\infty}\left(\left\|\overline{x}^{k+1}-x^{k+1}\right\|+\left\|y^{k+1}-y^{k}\right\|\right)<\infty$$

故 $\sum_{k=0}^{N}\left\|\overline{x}^{k+1}-x^{k+1}\right\|<\infty$， $\sum_{k=0}^{N}\left\|y^{k+1}-y^{k}\right\|<\infty$。

于是，

$$\begin{aligned}\sum_{k=0}^{\infty}\left\|x^{k+1}-x^{k}\right\|&\le\sum_{k=0}^{\infty}\left(\left\|x^{k+1}-\overline{x}^{k}\right\|+\left\|\overline{x}^{k}-x^{k}\right\|\right)\\&\le\frac{1}{\theta}\sum_{k=0}^{\infty}\left\|\overline{x}^{k+1}-x^{k+1}\right\|+\sum_{k=0}^{\infty}\left\|\overline{x}^{k}-x^{k}\right\|<\infty\end{aligned},$$

因此， $\sum_{k=0}^{\infty}\left\|z^{k+1}-z^{k}\right\|=\sum_{k=0}^{\infty}\left(\left\|x^{k+1}-x^{k}\right\|+\left\|y^{k+1}-y^{k}\right\|\right)<\infty$。

(ii) 由(i)可知 $\left\{z^{k}\right\}$ 为柯西序列，故收敛。由定理 1，可知 $\left\{z^{k}\right\}$ 收敛到 L 的一个稳定点。

5. 数值试验

为了验证算法的有效性与可行性，本节应用惯性邻近交替极小化算法(IPAM)及邻近交替极小化算法(PAM)求解 L_0 问题。数值试验过程均由 MATLAB(2014a)实现，程序运行环境为：Windows 7 (64 bite)，RAM：4G，CPU：@ 1.60 GHz 2.30 GHz。

在压缩感知中，基础的问题是从 m 个不完全实验数据中恢复一个 n 维的稀疏信号 x，($m\ll n$)。该问题可转化为如下模型：

$$\begin{aligned}&\min\quad \|x\|_0\\&\text{s.t.}\quad Dx=b\end{aligned}$$

或正则化形式

$$\min\quad \lambda\|x\|_0+\frac{1}{2}\|Dx-b\|^2$$

其中 $D\in\mathbb{R}^{m\times n}$ 是实验矩阵， $b\in\mathbb{R}^{m}$ 是观察信号， $\lambda>0$ 是正则化参数， $\|x\|_0$ 表示 x 中非零元素的个数。一般情况下，上述模型是 NP 难问题。为了克服这一困难，可以用 $L_{1/2}$ 拟范数来松弛 L_0 问题，考虑下述问题

$$\min\quad \lambda\|x\|_{\frac{1}{2}}^{\frac{1}{2}}+\frac{1}{2}\|Dx-b\|^2$$

其中 $\|x\|_{\frac{1}{2}}=\left(\sum_{i=1}^{n}|x_i|^{\frac{1}{2}}\right)^2$。通过引入一个新的变量 $y\in\mathbb{R}^{n}$，将问题转化为带约束的问题

$$\begin{aligned}&\min \quad \lambda\|y\|_{\frac{1}{2}}^{\frac{1}{2}}+\frac{1}{2}\|Dx-b\|^2\\&\text{s.t.} \quad x=y\end{aligned}$$

再用罚函数方法将上述问题转化为问题(1)的形式

$$\min \quad \lambda\|y\|_{\frac{1}{2}}^{\frac{1}{2}}+\frac{1}{2}\|Dx-b\|^2+\frac{\mu}{2}\|x-y\|^2 \tag{14}$$

利用 IPAM 求解问题(14)，有

$$\begin{cases}x^{k+1}=\left(D^{\mathrm{T}}D+\mu I+c_kI\right)^{-1}\left(D^{\mathrm{T}}b+\mu y^k+c_k\bar{x}^k\right)\\ y^{k+1}\in H\left(\dfrac{1}{\mu+d_k}\left(\mu\bar{x}^{k+1}+d_ky^k\right);\dfrac{2\lambda}{\mu+d_k}\right)\end{cases}$$

其中 $H(\cdot,\cdot)$ 被称为 half-shrinkage 算子[10]。

利用 PAM 求解问题(14)，有

$$\begin{cases}x^{k+1}=\left(D^{\mathrm{T}}D+\mu I+c_kI\right)^{-1}\left(D^{\mathrm{T}}b+\mu y^k+c_kx^k\right)\\ y^{k+1}\in H\left(\dfrac{1}{\mu+d_k}\left(\mu x^{k+1}+d_ky^k\right);\dfrac{2\lambda}{\mu+d_k}\right)\end{cases}$$

在试验过程中，选取 $D_{ij}\sim N(0,1)$，并正则化每一列，随机生成含有 100 个非零元素的稀疏向量 x^0，每一个样本都服从正态分布。向量 $b=Dx^0+v$，其中 $v\sim N\left(0,10^{-3}I\right)$，正则化参数 $\lambda=0.01\lambda_{\max}$，其中 $\lambda_{\max}=\left\|D^{\mathrm{T}}b\right\|_\infty$。选取参数如下：$\lambda^+=\lambda^-=\mu$，$m=\mu$，$\theta=0.8$，$\mu=1$，$c_k=c=2$，$d_k=d=3.5$。除 θ 外的其他参数，两种算法选取相同的参数值。终止准则为

$$\frac{\left\|z^{k+1}-z^k\right\|}{\left\|z^k\right\|}<TOL$$

$TOL=10^{-4}$ 为给定的误差值。计算两种算法在不同情况下分别所需的迭代次数，迭代时间及目标函数值。由图 1 可知，IPAM 与 PAM 目标函数都趋于一个固定值，说明两种算法都收敛。且由表 1 知，IPAM 比 PAM 运行时间更少，收敛速度更快。

Table 1. Numerical result
表 1. 数值结果

维数		IPAM			PAM		
m	n	iter	time(s)	I-obj	iter	time(s)	P-obj
900	1000	110	14.96	2.42	123	17.72	2.50
1000	1100	105	19.14	2.25	121	26.35	2.28
1000	1200	97	21.97	2.64	118	27.97	2.64
1000	1300	129	35.33	2.68	149	40.77	2.79
1000	1400	102	31.2	2.67	123	38.86	2.79
1500	2000	99	86.32	2.68	119	102.25	2.72
2000	3000	99	253.72	2.78	121	310.16	2.87

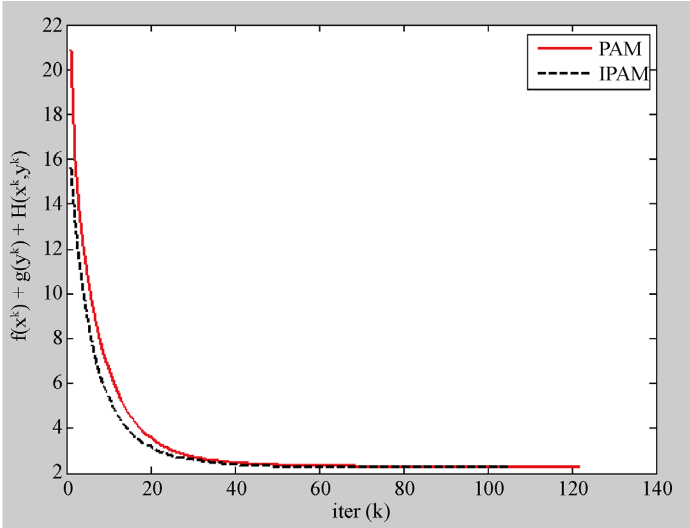

Figure 1. The trend of objective function varies with iterations
图 1. 目标函数随迭代次数变化趋势

基金项目

获国家自然科学基金青年基金(No. 1160195)，国家自然科学基金(No. 71861002)，广西自然科学基金青年基金(2017GXNSFBA380185)，(2017GXNSFBA198238)资助。

参考文献

[1] Bertsekas, D.P. and Tsitsiklis, J.N. (1989) Parallel and Distributed Computation: Numerical Methods (Vol. 23). Prentice Hall, Englewood Cliffs, NJ.

[2] Auslender, A. (1992) Asymptotic Properties of the Fenchel Dual Functional and Applications to Decomposition Problems. *Journal of Optimization Theory and Applications*, **73**, 427-449. https://doi.org/10.1007/BF00940050

[3] Alvarez, F. (2000) On the Minimizing Property of a Second Order Dissipative System in Hilbert Spaces. *SIAM Journal on Control and Optimization*, **38**, 1102-1119. https://doi.org/10.1137/S0363012998335802

[4] Ochs, P., Chen, Y., Brox, T. and & Pock, T. (2014) iPiano: Inertial Proximal Algorithm for Nonconvex Optimization. *SIAM Journal on Imaging Sciences*, **7**, 1388-1419. https://doi.org/10.1137/130942954

[5] Ochs, P., Brox, T. and Pock, T. (2015) iPiasco: Inertial Proximal Algorithm for Strongly Convex Optimization. *Journal of Mathematical Imaging and Vision*, **53**, 171-181. https://doi.org/10.1007/s10851-015-0565-0

[6] Boţ, R.I., Csetnek, E.R. and Hendrich, C. (2015) Inertial Douglas-Rachford Splitting for Monotone Inclusion Problems. *Applied Mathematics and Computation*, **256**, 472-487. https://doi.org/10.1016/j.amc.2015.01.017

[7] Bot, R.I. and Csetnek, E.R. (2014) An Inertial Alternating Direction Method of Multipliers. *Mathematics*, ArXiv preprint arXiv: 1404.4582.

[8] Nesterov, Y. (2013) Introductory Lectures on Convex Optimization: A Basic Course (Vol. 87). Springer Science & Business Media, Berlin, Heidelberg.

[9] Bolte, J., Sabach, S. and Teboulle, M. (2014) Proximal Alternating Linearized Minimization for Nonconvex and Nonsmooth Problems. *Mathematical Programming*, **146**, 459-494. https://doi.org/10.1007/s10107-013-0701-9

[10] Xu, Z., Chang, X., Xu, F. and Zhang, H. (2012) $L_{1/2}$ Regularization: A Thresholding Representation Theory and a Fast Solver. *IEEE Transactions on Neural Networks and Learning Systems*, **23**, 1013-1027. https://doi.org/10.1109/TNNLS.2012.2197412