加速邻近梯度算法及其应用

报告人：张金超

河北工业大学理学院

指导教师： 刘新为 教授

2020. 6. 28

目录

- 研究背景
- 基本概念
- 研究现状
- 问题总结
- 参考文献

研究背景

主要关注的问题是两个函数和求极小的组合形式:

$$\min_{\mathbf{x}\in\mathbb{E}}\{F(\mathbf{x})\equiv f(\mathbf{x})+g(\mathbf{x})\} \tag{1}$$

例1：图像去噪中的线性逆问题$Ax=b+w$，$A\in\mathbb{R}^{m\times n}(m\leq n)$ $b\in\mathbb{R}^m$，$w\in\mathbb{R}^m$，$x\in\mathbb{R}^n$ w,x都是未知的. 比如ℓ_1正则化，即

$$\min_{\mathbf{x}\in\mathbb{R}^n} F(x)=\|Ax-b\|_2^2+\ell_1\|x\|_1. \tag{2}$$

例2：具有单纯形约束的非凸二次规划问题:

$$\begin{aligned}\min_{x\in\mathbb{R}^n}\ & \frac{1}{2}x^\top Ax-b^\top x\\ \text{s.t. }\ & e^T x=s,\quad x\geq 0\end{aligned} \tag{3}$$

其中$f(x)=\frac{1}{2}x^\top Ax-b^\top x$，$g(x)=\delta_S(x)$，$S=\{x\in\mathbb{R}^n:e^\top x=s,x\geq 0\}$
应用：在逆问题、图像处理、压缩感知、投资组合优化、统计和机器学习等一些大规模问题都有重要的应用.

研究背景

上述结构的问题可以分为:

- 凸问题：函数f,g均为正常闭凸函数,通常f是Lipschitz连续的,g可能是非光滑函数.

- 非凸问题：函数f,g均为非凸函数;或者g是正常闭凸函数，f可能是一个非凸函数且具有Lipschitz连续.

基本概念

$f:\mathbb{R}^n \to \mathbb{R}$的邻近算子$prox_f$:

$$prox_f(u) = arg\ min_{x\in\mathbb{R}^n}\{f(x) + \frac{1}{2}\|x-u\|^2\}$$

其中t_k是在迭代k步中的步长,迭代步骤如下:

$$x^{k+1} = argmin_{x\in\mathbb{E}}\{f(x^k) + \langle\nabla f(x^k), x - x^k\rangle + g(x) + \tfrac{1}{2t_k}\|x-x^k\|^2\}. \tag{4}$$

临近梯度迭代:

$$x^{k+1} = prox_{t_k g}(x^k - t_k\nabla f(x^k)). \tag{5}$$

研究现状

- 1983：*Nesterov*[1]对于光滑凸问题，提出的加速方案使得收敛率由$\mathbf{O}(1/k)$提升到了$O(1/k^2)$.

$$\alpha_k = \frac{1+\sqrt{1+4\alpha_k^2}}{2}; \tag{6}$$

$$y_{k+1} = x_k + \frac{\alpha_k - 1}{\alpha_{k+1}}(x_k - x_{k-1}). \tag{7}$$

- 1988：*Nesterov*[2]提出了另外一个更一般的加速方案关于光滑函数.

$$x_{k+1} = y_k - (1/L)\nabla f(y_k); \tag{8}$$

$$y_{k+1} = x_{k+1} + \frac{k}{k+3}(x_{k+1} - x_k). \tag{9}$$

[1] Y. Nesterov, A method of solving a convex programming problem with convergence rate $O(1/k^2)$, Soviet Math. Dokl., 27 (1983), pp. 372 – 376.

[2] Y .Nesterov, "An approach to constructing optimal methods for minimization of smooth convex functions". èkonom. i mat. metody 3(1988):509-517.

- 2008：*Tseng, P*[3]对于凸组合光滑和非光滑问题,对加速技巧进行了统一的分析；即加速临近梯度算法的三个变形.

Algorithm 1 (Accelerated proximal gradient algorithm variant 1)

Initialization： Choose $\theta_0 = \in (0,1]$， $x_0, z_0 \in dom(g)$.

General step： for any k=0,1,2,...excute the following steps:

Step 1: Choose a nonempty closed convex set $X_k \subseteq \mathbb{E} with X_k \cap dom(g) \neq \varnothing$. Let:

$$y^k = (1-\theta_k)x^k + \theta_k z_k;$$
$$z_{k+1} = argmin_{x \in X_k} \ell_f(x;y_k) + \theta_k L D(x;z_k);$$
$$\hat{x}_{k+1} = (1-\theta_k)x^k + \theta_k z_{k+1};$$

Step 2: Choose x_{k+1} to be worse than$\hat{x}_{k+1}$ in $\ell_f(\cdot;y_k) + \frac{L}{2}\|\cdot - y_k\|^2$ value,i,e:

$$\ell_f(x_{k+1};y_k) + \frac{L}{2}\|x_{k+1} - y_k\|^2 \leq \ell_f(\hat{x_{k+1}};y_k) + \frac{L}{2}\|\hat{x}_{k+1} - y_k\|^2$$

Step 3: Choose $\theta_{k+1} \in (0,1]$ satisfying :

$$\frac{1-\theta_{k+1}}{\theta_{k+1}^2} \leq \frac{1}{\theta_k^2};$$

Step 4:$k \leftarrow k+1$,and go to step 1.

until: stopping criterion satisfied.

---

[3] Tseng, P.. (2008). On accelerated proximal gradient methods for convex-concave optimization. submitted to SIAM Journal on Optimization.

研究现状

- Accelerated proximal gradient algorithm variant 2:

Algorithm 2 (Accelerated proximal gradient algorithm variant 2)

Initialization: Choose $\theta_0 = \theta_{-1} = \in (0,1]$, $x_0 = x_{-1} \in dom(g) \neq \varnothing, z_0 \in dom(g)$.

General step: for any k=0,1,2,...excute the following steps:

Step 1: Choose a nonempty closed convex set $X_k \subseteq \mathbb{E}$ with $X_k \cap dom(g) \neq \varnothing$. Let:

$y^k = x_k + \theta_k(\theta_{k-1}^{-1})(x_k - x_{k-1})$;

$x_{k+1} = argmin_{x\in X_k} \ell_f(x;y_k) + \frac{L}{2}\|x-y_k\|^2$;

Step 2: Choos $\theta_{k+1} \in (0,1]$ satisfying $\frac{1-\theta_{k+1}}{\theta_{k+1}^2} \leq \frac{1}{\theta_k^2}$

Step 3: $k \leftarrow k+1$ and go to step 1.

until: stopping criterion satisfied.

其中ℓ_f:

$$\ell_f(x;y) := f(y) + < \nabla f(y), x-y > + g(x) \tag{10}$$

- Accelerated proximal gradient algorithm variant 3;

Algorithm 3 (Accelerated proximal gradient algorithm variant 3)

Initialization: Choose $0 < \theta_0 \le 1$, $x_0 \in dom(g)$.Let $z_0 = argmin_{x \in dom(g)} h(x)$,$k=0$;

General step: for any k=0,1,2,...excute the following steps:

Step 1: Choose a nonempty closed convex set $X_k \subseteq X_{k-1}$ with $X_k \cap dom(g) \neq \varnothing$. Let:

$$y^k = (1-\theta_k)x^k + \theta_k z_k;$$

$$\psi_{k+1}(x) = \sum_{i=0}^{k} \frac{\ell_f(x;y_i)}{\vartheta_i}, \forall x;$$

$$z_{k+1} = argmin_{x \in X_k} \psi_{k+1}(x) + Lh(x);$$

$$\hat{x}_{k+1} = (1-\theta_k)x^k + \theta_k z_{k+1};$$

Step 2: Choose x_{k+1} to be worse than$\hat{x}_{k+1}$ in $\ell_f(\cdot;y_k) + \frac{L}{2}\|\cdot - y_k\|^2$ value,i,e:

$$\ell_f(x_{k+1};y_k) + \frac{L}{2}\|x_{k+1} - y_k\|^2 \le \ell_f(\hat{x_{k+1}};y_k) + \frac{L}{2}\|\hat{x}_{k+1} - y_k\|^2;$$

Step 3: Choose $\theta_{k+1} \in (0,1]$,$\vartheta_{k+1} \ge \theta_{k+1}$satisfying :

$$\frac{1-\theta_{k+1}}{\vartheta_{k+1}} = \frac{1}{\theta_k \vartheta_k};$$

Step 3: $k \leftarrow k+1$,and go to step 1.

until: stopping criterion satisfied.

研究现状

- 2009：*Beck and Teboulle*[4]进一步扩展了加速邻近梯度下降法求解复合问题(含有光滑部分和非光滑部分)，其中步长可以基于固定步长或者回溯线搜索(大规模问题) 求取.

Algorithm 4 (The original FISTA scheme)

Initialization: $y_1 = x_0 \in R^n, t_1 = 1, L_k > 0$

Step 1: $t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2}, \alpha_k = \frac{t_{k-1}-1}{t_k}$

Step 2: $y_{k+1} = x_k + \alpha_k(x_k - x_{k-1})$

Step 3: $x_{k+1} = prox_{\frac{1}{L_k}g}(y_k - \frac{1}{L_k}\nabla f(y_k)).$

Step 4: $k = k+1;$

until: stopping criterion satisfied.

---

[4] Beck A , Teboulle M . A fast iterative shrinkage-thresholding algorithm for linear inverse problems[J]. SIAM Journal on Imaging Sciences, 2009, 2(1):183-202.

研究现状

- 2015： *A.Chambolle*[5]等对FISTA算法的迭代序列进行收敛性分析,当$t_k = \frac{k+\alpha-1}{\alpha}, \alpha \geq 2$,则序列$\{x_n\}_{n\in\mathbb{N}}$弱收敛于F的极小值点.

- 2018： *Kim,Donghwan*[6]等使用松弛的PEP对FISTA算法进行worst-case bound分析；在此基础上，提出了新的算法FISTA-OCG，根据组合梯度映射的一个worst-case bound来优化临近梯度算法的步长系数,使得目标函数的界满足收敛率为$O(1/k^2)$.

- 2019： *Vien V Mai*[7]等提出了临近梯度算法的Anderson加速，将光滑不动点映射的Anderson加速算法的局部收敛结果推广到非光滑条件下；并推广到了Bregman邻近梯度算法.

[5] Chambolle, A., Dossal, C. On the convergence of the iterates of the “Fast Iterative Shrinkage/Thresholding Algorithm”. J Optim Theory Appl 166, 968 - 982 (2015).
[6] Kim D, Fessler JA. Another look at the fast iterative shrinkage/thresholding Algorithm (FISTA). SIAM J Optim. 2018;28(1):223-250. doi:10.1137/16M108940X.
[7] Mai V V, Johansson M. Anderson acceleration of proximal gradient methods.[J]. arXiv: Optimization and Control, 2019.

Guarded AA-PGA:

Algorithm 5 (Guarded AA-PGA)	
Input：	$y_0 = x_0, m > 0.$
Initialization：	$y_1 = x_0 - \gamma \nabla f(x_0), x_1 = prox_{\gamma g}(y_1), g_0 = y_1, r_0 = g_0 - y_0$
Step 1: for	$k = 1, ..., K-1$ do
Step 2:	$m_k = min(m, k)$
Step 3:	$g_k = x_k - \gamma \nabla f(x_k), r_k = g_k - y_k$
Step 4:	$R_k = [r_k, ..., r_{k-m_k}]$
Step 5:	$\alpha_k = argmin_{\alpha^T \mathbf{1}=1} \|R_k \alpha\|$
Step 6:	$y_{test} = \sum_{i=0}^{m_k} \alpha_i^k g_{k-i}$
Step 7:	$x_{test} = prox_{\gamma h}(y_{test})$
Step 8:	**if** $f(x_{test}) \leq f(x_k) - \frac{\gamma}{2} \|\|\|\nabla f(x_k)_2^2$ 保证收敛
Step 9:	$x_{k+1} = x_{test}; y_{k+1} = y_{test}$
Step 10:	**else**
Step 11:	$x_{k+1} = prox_{\gamma h}(g_k); y_{k+1} = g_k$
Step 12:	**end**
Step 13:end	

研究现状- Nolips

- 2016： *Heinz H Bauschke*[8]等提出了不需要Lipschitz梯度连续的下降引理,利用初等凸性去等价代替,证明了NoLips与通常的邻近梯度具有相似的收敛性和复杂性.
- 2019：*Gutman D H*[9]等给出了在邻近梯度和加速邻近梯度算法在凸问题中关于收敛速度的新证明,核心是通过目标函数的凸共轭构造一个上界。该证明不需要$\nabla\varphi$是Lipschitz continuous.

选择步长t_k时需要满足下降条件：

$$f(x_{k+1}) \le min_{x\in\mathbb{R}^n}\{\varphi(y_k)+\langle\nabla\varphi(y_k),x-y_k\rangle+\frac{1}{2t_k}\|x-y_k\|^2+\psi(x)\} \tag{11}$$

$$=\varphi(y_k)+\psi(x_{k+1})+<\langle g_k^{\psi},y_k-x_{k+1}\rangle-\frac{t_k}{2}\|g_k\|^2.$$

其中$g_k^{\psi}:=\nabla\varphi(y_k)$; $g_k^{\psi}\in\partial\varphi(x_{k+1})$; $g_k:=g_k^{\varphi}+g_k^{\psi}$

[8] Bauschke H H, Bolte J, Teboulle M, et al. A descent lemma beyond lipschitz gradient continuity: First-Order Methods Revisited and Applications[J]. Mathematics of Operations Research, 2017, 42(2): 330-348.

[9] Gutman D H, Pena J F. Convergence rates of proximal gradient methods via the convex conjugate[J]. SIAM Journal on Optimization, 2019, 29(1): 162-174.

对于函数g是正常凸下半连续的，在具有Lipschitz连续的假设下才会有：

$$g(x) \leq g(y) + \langle x-y, \nabla g(y) \rangle + \frac{L}{2}\|x-y\|^2 \tag{12}$$

通过范数展开等价于凸函数的梯度不等式：

$$\left(\frac{L}{2}\|x\|^2 - g(x)\right) - \left(\frac{L}{2}\|y\|^2 - g(y)\right) \geq \langle Ly - \nabla g(y), x-y \rangle \tag{13}$$

新的下降引理：

令$h : X \to (-\infty, \infty]$是Legendre函数，令$g : X \to (-\infty, \infty]$是凸函数且$dom\ g \supset dom\ h$,根据新的LC条件，则有：

$$\forall (x,y) \in int\ dom(h) \times int\ dom(h) g(x) \leq g(y) + \langle \nabla g(y), x-y \rangle + LD_h(x,y). \tag{14}$$

[8] Bauschke H H, Bolte J, Teboulle M, et al. A descent lemma beyond lipschitz gradient continuity: First-Order Methods Revisited and Applications[J]. Mathematics of Operations Research, 2017, 42(2): 330-348.

研究现状-Nolips

定义 1 (凸共轭函数)

若$h:\mathbb{R}^n \to \mathbb{R}\cup\{\infty\}$是凸函数，则凸共轭函数$h^*:\mathbb{R}^n \to \mathbb{R}\cup\{\infty\}$

$$h^*(z) = sup_{x\in\mathbb{R}^n}\{\langle z,x\rangle - h(x)\}.$$

定理 1

假设$\theta_k \in (0,1], k=0,1,2\ldots$，步长$t_k > 0$，使得假设(16)成立，令$z_k \in \mathbb{R}^n, k=1,\ldots,$有以下：

$$z_k := \frac{\sum_{i=0}^{k-1} \frac{t_i}{\theta_i} g_i}{\sum_{i=0}^{k-1} \frac{t_i}{\theta_i}}. \tag{15}$$

$$LHS_k \le -f^*(z_k) + < z_k, x_0 > - \frac{\sum_{i=0}^{k-1} \frac{t_i}{\theta_i}}{2} \|z_k\|^2, \tag{16}$$

其中LHS_k是取决于$\theta_k \in (0,1]$的选取

研究现状-Nolips

定理 1（续）

(a) 当$\theta_k = 1, k = 0,1,2...$，令

$$LHS_k := \frac{\sum_{i=0}^{k} t_i f(x_{i+1})}{\sum_{i=0}^{k} t_i}, \tag{17}$$

(b) 当$t_k = 1/L, k = 0,1,2...$对所有正的常数$L, \theta_k, \theta_0 = 1$，且$\theta_{k+1}^2 = \theta_k^2(1-\theta_{k+1})$

$$LHS_k := f(x_k). \tag{18}$$

$$f(x_k) - f^* \leq \frac{L \cdot dist(x_0, X^*)^2}{2k} \tag{19}$$

$$f(x_k) - f^* \leq \frac{2L \cdot dist(x_0, X^*)^2}{(k+1)^2} \tag{20}$$

研究现状-Nolips

定理 2

假设$\bar{f} = min_{x\in\mathbb{R}^n f(x)}$是有限的，$\theta_k \in (0,1], k = 0,1,2...$满足$\theta_0 = 1$和$\theta_{k+1}^2 \geq \theta_k^2(1-\theta_{k+1})$且步长$t_k > 0$是非增的其假设(16)成立，令$z_k$

$$z_k := \frac{\theta_{k-1}^2}{t_{k-1}} \sum_{i=0}^{k-1} \frac{t_i}{\theta_i} g_i. \tag{21}$$

$$f(x_k) - F^* \leq -(R_k \cdot (f - F^*))^*(z_k) + < z_k, x_0 > - \frac{t_{k_1}}{2\theta_{k-1}^2} \|z_k\|^2. \tag{22}$$

$$f(x_k) - f^* \leq \frac{2L \cdot dist(x_0, X^*)^2}{(k+1)^2} \tag{23}$$

研究现状(非凸问题)

2015年, *HuanLi, ZhouchenLin*[10]给出了加速邻近梯度法在非凸优化上的收敛性

- 通过引入monitor来满足充分下降性,给出了详细的收敛性分析，证明了一般非凸非光滑优化的每个聚点都是一个临界点,收敛率分析主要是在K L性质下建立了渐近收敛率:
- 依据上述的收敛分析提出了单调的APG和非单调的APG.

[10] Li, Huan, and Zhouchen Lin. "Accelerated proximal gradient methods for nonconvex programming." Advances in neural information processing systems. 2015.

研究现状(非凸问题)

单调的APG:

Algorithm 3 (Monotone APG)

Initialization: $z_1 = x_1 = x_0, t_1 = 1, t_0 = 0, \alpha_x < \frac{1}{L}, \alpha_y < \frac{1}{L}$

Step 1: $y_k = x_k + \frac{t_k - 1}{t_{k+1}}(z_k - x_k) + \frac{t_{k-1} - 1}{t_k}(x_k - x_{k-1})$

Step 2: $z_{k+1} = prox_{\alpha_y g}(y_k - \alpha_y \nabla f(y_k))$

Step 3: monitor: $v_{k+1} = prox_{\alpha_x g}(x_k - \alpha_x \nabla f(x_k))$.

Step 4: $t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}$

Step 5:if $F(\mathbf{z}_{k+1}) \leq F(\mathbf{v}_{k+1})$,修正$z_{k+1}$

$x_{k+1} = \mathbf{z}_{k+1}$

else

$x_{k+1} = \mathbf{v}_{k+1}$.

until: stopping criterion satisfied.

非单调的APG:

Algorithm 4 (Nonmonotone APG)

Input: $\eta \in [0,1], \delta > 0, c_1 = F(x_1), q_1 = 1.$

Initialization: $z_1 = x_1 = x_0, t_1 = 1, t_0 = 0, \alpha_x < \frac{1}{L}, \alpha_y < \frac{1}{L}$

Step 1: $y_k = x_k + \frac{t_k - 1}{t_{k+1}}(z_k - x_k) + \frac{t_{k-1} - 1}{t_k}(x_k - x_{k-1})$

Step 2: $z_{k+1} = prox_{\alpha_y g}(y_k - \alpha_y \nabla f(y_k))$

Step 3: **if** $F(z_{k+1}) \leq c_k - \delta \|z_{k+1} - y_k\|^2$.(该步骤尽可能避免了$v_{k+1}$的计算)

$x_{k+1} = \mathbf{z}_{k+1}$

else

$v_{k+1} = prox_{\alpha_x g}(x_k - \alpha_x \nabla f(x_k)).$

Step 4: **if** $F(\mathbf{z}_{k+1}) \leq F(\mathbf{v}_{k+1})$

$x_{k+1} = \mathbf{z}_{k+1}$

else

$x_{k+1} = \mathbf{v}_{k+1}.$

end

$t_{k+1} = \frac{\sqrt{4t_k^2 + 1} + 1}{2}$

$q_{k+1} = \eta q_k + 1$

$c_{k+1} = \frac{\eta q_k c_k + F(x_{k+1})}{q_{k+1}}.$

until: stopping criterion satisfied.

研究现状(非凸问题)

- 2016： *Quanming Yao*[11]等提出了算法在每次迭代中只需要一个邻近步，邻近步可以不精确,使得每次迭代所需的邻近步数量显著减少，但仍然保证收敛到临界点,由于非单调不同,没有用到KL性质进行收敛分析，且算法具有$O(1/k)$的收敛速度.

Algorithm 5 (Noconvex inexact APG (niAPG) algorithm)

Input： $\eta \in (0,1/L), \delta \in (0,1/\eta - L)$.

Initialization： $x_1 = x_0$

Step 1: $y_k = x_k + \frac{k-1}{k+2}(x_k - x_{k-1})$

Step 2:$\Delta_k = max_{t=max(1,k-q),\dots,k} F(x_t)$

Step 3: **if** $F(y_k) \leq \Delta_k$.

$v_k = y_k$

else

$v_k = x_k$.

Step 4: **end if**

$z_k = v_k - \eta \nabla f(v_k)$

$x_{k+1} = prox_{\eta g}(z_k)$.// **posibly inexact**

end

until: stopping criterion satisfied.

[11] Yao Q . Efficient inexact proximal gradient algorithm for nonconvex problems[J]. 2016.

研究现状(非凸问题)

2019年, *PeterOchs*, *ThomasPock*[12]提出了一种自适应FISTA应用于非凸优化

- 提出了基于FISTA的自适应外推邻近梯度法-$\alpha FISTA$，其中外推参数在每次迭代中自适应优化.

- 当目标函数光滑部分为二次的特殊情况下(也可能是非凸)，$\alpha FISTA$等价于某个proximal quasi-Newton method，并且proximal mapping具有有效的解.

[12] Ochs, Peter, and Thomas Pock. ”Adaptive FISTA for nonconvex optimization.” SIAM Journal on Optimization 29.4 (2019): 2482-2503.

研究现状(非凸问题)

算法流程：

Algorithm 6 (adaptive FISTA for nonconvex problems with backtracking)

Input： $x_0 = \bar{x}$ for $\bar{x} \in \mathbb{N}$ Let $\alpha > 0$.

Initialization： Select extrapolation directions $(D_k)_{k\in\mathbb{N}}$ with $D_k \in \mathbb{N}^{N\times R}$

Step 1: Find x_{k+1} and $\beta_k$uch that the following holds:

Step 2: $(x_{k+1}, \beta_k) \in \underset{x\in\mathbb{R}^N}{\operatorname{argmin}} \min_{\beta\in\mathbb{R}^R} \ell_f^g\left(x; y^{(\beta)}\right) + \frac{1}{2}\left\|x - y^{(\beta)}\right\|_{\boldsymbol{T}_k}^2$

$y^{(\beta)} = x_k + (D_k)\beta$,where$\mathbf{T}_k - \mathbf{L}_k - \alpha\mathbf{I} \in \mathbb{S}_+(N)$,

and $\mathbf{L}_k$satisfies **Step 3**

Step 3:$f_{k+1} \leq f(y^{(\beta_k)}) + \langle \nabla f(y^{(\beta_k)}), x_{k+1} - y^{(\beta_k)}\rangle + \frac{1}{2}\|x - y^{(\beta_k)}\|_{\mathbf{L}_k}^2$.

until: stopping criterion satisfied.

其中$\ell_f^g(x; y^{(\beta)}) := g(x) + f(y^{(\beta)}) + \langle \nabla f(y^{(\beta)}), x - y^{(\beta)}\rangle$

正则化稀疏非线性逆问题的数值实验:

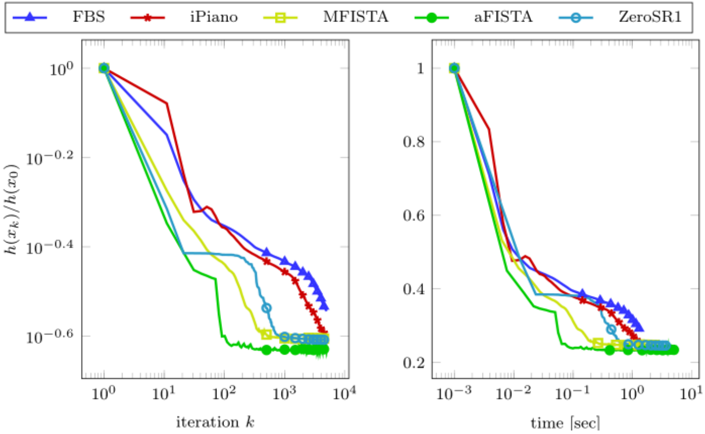

Figure: In this experiment, aFISTA outperforms the other methods with respect to the actual computation time and the final objective value.

未来展望

- 凸问题中以及非凸问题中，非Lipschitz gradient continuity条件下的下降引理和扩展下降引理是否可以根据其他的度量来给出？

- 对于非凸问题,非Lipschitz gradient continuity条件下，对FISTA算法是否能进行收敛性分析？

参考文献

[1] Y. Nesterov, A method of solving a convex programming problem with convergence rate$O(1/k^2)$, Soviet Math. Dokl., 27 (1983), pp. 372 – 376.

[2] Y .Nesterov, ”An approach to constructing optimal methods for minimization of smooth convex functions”. èkonom. i mat. metody 3(1988):509-517.

[3] Tseng, P.. (2008). On accelerated proximal gradient methods for convex-concave optimization. submitted to SIAM Journal on Optimization.

参考文献

[4] Beck A , Teboulle M . A fast iterative shrinkage-thresholding algorithm for linear inverse problems[J]. SIAM Journal on Imaging Sciences, 2009, 2(1):183-202.

[5] Chambolle, A., Dossal, C. On the convergence of the iterates of the "Fast Iterative Shrinkage/Thresholding Algorithm" . J Optim Theory Appl 166, 968 – 982 (2015).

[6] Kim D, Fessler JA. Another look at the FAST ITERATIVE SHRINKAGE/THRESHOLDING ALGORITHM (FISTA). SIAM J Optim. 2018;28(1):223-250. doi:10.1137/16M108940X.

参考文献

- [7] Mai V V, Johansson M. Anderson acceleration of proximal gradient methods.[J]. arXiv: Optimization and Control, 2019.
- [8] Bauschke H H, Bolte J, Teboulle M, et al. A descent lemma beyond lipschitz gradient continuity: First-Order Methods Revisited and Applications[J]. Mathematics of Operations Research, 2017, 42(2): 330-348.
- [9] Gutman D H, Pena J F. Convergence rates of proximal gradient methods via the convex conjugate[J]. SIAM Journal on Optimization, 2019, 29(1): 162-174.
- [10] Yao Q . Efficient inexact proximal gradient algorithm for nonconvex problems[J]. 2016.

谢谢!