An Average Curvature Accelerated Composite Gradient Method for Nonconvex Smooth Composite Optimization Problems

Jiaming Liang[1] Renato D.C. Monteiro[1]

[1]School of Industrial and Systems Engineering, Georgia Tech

Alice and John Jarvis Ph.D. Student Paper Competition
September 17, 2021

- J. Liang and R. D. C. Monteiro. An average curvature accelerated composite gradient method for nonconvex smooth composite optimization problems. SIAM Journal on Optimization, 31(1):217-243, 2021.

1. The Main Problem
   - Assumptions
   - Approximate solutions
2. Average Curvature ACG Method
   - Motivation
   - AC-ACG method
   - Convergence rate and iteration-complexity
3. Computational Results
4. Concluding Remarks

1. The Main Problem
   - Assumptions
   - Approximate solutions

2. Average Curvature ACG Method
   - Motivation
   - AC-ACG method
   - Convergence rate and iteration-complexity

3. Computational Results

4. Concluding Remarks

The main problem:

$$(P) \qquad \min\left\{f(z)+h(z) : z \in \mathbb{R}^n\right\}$$

where

- $h : \mathbb{R}^n \to (-\infty, \infty]$ is a closed proper convex function such that
$$D := \sup\{\|z'-z\| : z, z' \in \mathrm{dom}\, h\} < \infty$$
- f is differentiable (not necessarily convex) on $\mathrm{dom}\, h$ and there exist $0 < m \leq L$ such that for every $z, z' \in \mathrm{dom}\, h$
$$\|\nabla f(z') - \nabla f(z)\| \leq L\|z'-z\|$$
$$f(z') - \ell_f(z'; z) \geq -\frac{m}{2}\|z'-z\|^2$$

where $\ell_f(z'; z) := f(z) + \langle \nabla f(z), z'-z\rangle$.

A necessary condition for $\bar{z}$ to be a local minimizer of (P) is that

$$0 \in \nabla f(\bar{z}) + \partial h(\bar{z})$$

Goal: for given $\hat{\rho} > 0$, find a $\hat{\rho}$-approximate solution of (P), i.e., a pair $(\hat{z}, \hat{v})$ such that

$$\hat{v} \in \nabla f(\hat{z}) + \partial h(\hat{z}), \quad \|\hat{v}\| \leq \hat{\rho}$$

There are a couple of ACG methods which accomplishes the above goal (e.g., Ghadimi-Lan's method). This talk describes a different and novel ACG method for doing that.

ACG methods compute the next iterate as

$$z_{k+1} = z(\tilde{x}_k; M_k) := \mathrm{argmin}_z \left\{ \ell_f(z; \tilde{x}_k) + h(z) + \frac{M_k}{2}\|z - \tilde{x}_k\|^2 \right\}$$

where

$$M_k \geq \mathcal{C}(z_{k+1}; \tilde{x}_k) := \frac{2[f(z_{k+1}) - \ell_f(z_{k+1}; \tilde{x}_k)]}{\|z_{k+1} - \tilde{x}_k\|^2} \quad (*)$$

Fact: small M_k leads to fast convergence

- Constant estimate: $M_k = L$
- Adaptive estimate: $M_k \leftarrow \tau M_k$ for some $\tau > 1$, if (*) is not satisfied

Matrix completion: function, gradient and resolvent evaluations require SVD

ACG methods compute the next iterate as

$$z_{k+1} = z(\tilde{x}_k; M_k) := \mathrm{argmin}_z \left\{ \ell_f(z; \tilde{x}_k) + h(z) + \frac{M_k}{2}\|z - \tilde{x}_k\|^2 \right\}$$

where

$$M_k \geq \mathcal{C}(z_{k+1}; \tilde{x}_k) := \frac{2[f(z_{k+1}) - \ell_f(z_{k+1}; \tilde{x}_k)]}{\|z_{k+1} - \tilde{x}_k\|^2} \quad (*)$$

Fact: small M_k leads to fast convergence

- Constant estimate: $M_k = L$
- Adaptive estimate: $M_k \leftarrow \tau M_k$ for some $\tau > 1$, if (*) is not satisfied
  Matrix completion: function, gradient and resolvent evaluations require SVD

We will exploit the novel idea of choosing M_k as

$$M_k = \frac{\sum_{i=0}^{k-1} \mathcal{C}(z_{i+1}; \tilde{x}_i)}{k\,\alpha}$$

where $\alpha \in (0,1)$

Note: No line search for M_k is involved here!
$M_k \geq \mathcal{C}(z_{k+1}; \tilde{x}_k)$ might be violated.

Average Curvature ACG (AC-ACG) Method

0. Let $\alpha, \gamma \in (0,1)$, tolerance $\hat{\rho} > 0$ and initial point $z_0 \in \operatorname{dom} h$ be given; set $A_0 = 0$, $x_0 = z_0$, $M_0 = \gamma L$ and $k = 0$
1. compute

$$a_k = \frac{1 + \sqrt{1 + 4M_k A_k}}{2M_k} \qquad A_{k+1} = A_k + a_k \qquad \tilde{x}_k = \frac{A_k z_k + a_k x_k}{A_{k+1}}$$

2. compute

$$x_{k+1} = \operatorname{argmin}_u \left\{ a_k \left(\ell_f(u; \tilde{x}_k) + h(u) \right) + \frac{1}{2} \|u - x_k\|^2 \right\}$$

$$z^g_{k+1} = \operatorname{argmin}_u \left\{ \ell_f(u; \tilde{x}_k) + h(u) + \frac{M_k}{2} \|u - \tilde{x}_k\|^2 \right\}$$

$$v_{k+1} = M_k(\tilde{x}_k - z^g_{k+1}) + \nabla f(z^g_{k+1}) - \nabla f(\tilde{x}_k)$$

3. compute

$$C_k = \max\left\{\mathcal{C}(z^g_{k+1};\tilde{x}_k), \mathcal{L}(z^g_{k+1};\tilde{x}_k)\right\}, \quad \mathcal{L}(z;\tilde{x}) = \frac{\|\nabla f(z) - \nabla f(\tilde{x})\|}{\|z-\tilde{x}\|}$$

$$M_{k+1} = \max\left\{\gamma M, \frac{\sum_{j=0}^{k} C_j}{\alpha(k+1)}\right\},$$

4. set

$$z_{k+1} = \begin{cases} z^g_{k+1} & \text{if } C_k \le 0.9 M_k \quad \leftarrow \text{good iteration} \\ \frac{A_k z_k + a_k x_{k+1}}{A_{k+1}} & \text{otherwise} \quad \leftarrow \text{bad iteration} \end{cases}$$

and $k \leftarrow k+1$, and go to step 1

Remarks:

- Both good and bad iterations perform well-known types of acceleration steps
  Good: mimics the condition in standard ACG methods
  $M_k \geq \mathcal{C}(z_{k+1}; \tilde{x}_k)$
- If
  $$\frac{1}{\alpha} \geq 1 + \frac{1}{\gamma}$$
  then it can be shown that the proportion of good iterations is at least $2/3$
- Every iteration performs two resolvent evaluations

1. The Main Problem
   - Assumptions
   - Approximate solutions
2. Average Curvature ACG Method
   - Motivation
   - AC-ACG method
   - Convergence rate and iteration-complexity
3. Computational Results
4. Concluding Remarks

Theorem

The following statements hold:

(a) *for every $k \geq 1$, we have $v_k \in \nabla f(z_k^g) + \partial h(z_k^g)$*

(b) *for every $k \geq 12$, we have*

$$\min_{1 \leq i \leq k} \|v_i\|^2 \leq \mathcal{O}\left(\frac{M_k^2 D^2}{\gamma k^2} + \frac{\theta_k m M_k D^2}{k}\right)$$

where

$$\theta_k := \max\left\{\frac{M_k}{M_i} : 0 \leq i \leq k\right\} \geq 1.$$

The facts that $\theta_k = \mathcal{O}(1)$ and $M_k/L = \mathcal{O}(1)$ imply that the iteration-complexity bound for AC-ACG to obtain $\hat{\rho}$-approx. sol. is

$$\mathcal{O}\left(\frac{LD}{\hat{\rho}} + \frac{mLD^2}{\hat{\rho}^2} + 1\right)$$

Computational Results

The variant of AC-ACG described above was benchmarked against

- AG method by Ghadimi and Lan (known Lipschitz constant)
- nmAPG method by Li and Lin (known Lipschitz constant)
- UPFAG method by Ghadimi, Lan and Zhang (backtracking)

All methods stop with a pair (z, v) satisfying

$$v \in \nabla f(z) + \partial h(z), \qquad \frac{\|v\|}{\|\nabla f(z_0)\| + 1} \leq \hat{\rho}$$

Constrained Matrix Completion

$$\min_{X\in\mathbb{R}^{m\times n}}\left\{\frac{1}{2}\|\Pi_\Omega(X-O)\|_F^2+\mu\sum_{i=1}^{r}p(\sigma_i(X)) : \|X\|_F\le R\right\}$$

where $O\in\mathbb{R}^\Omega$ is an incomplete observed matrix, $\mu>0$ is a parameter, $r:=\min\{m,n\}$, $\sigma_i(X)$ is the i-th singular value of X and

$$p(t)=p_{\beta,\theta}(t):=\beta\log\left(1+\frac{|t|}{\theta}\right)$$

L	Function Value ×1000 / Iteration Count				Running Time ×1000 seconds				Curvature		Good
	AG	APG	UPFAG	AC	AG	APG	UPFAG	AC	Max	Avg	
4	2.26 3856	**1.81** 1036	2.60 521	2.29 765	4.6	1.0	2.6	**0.8**	1.00	0.31	96%
9	3.89 9158	**3.36** 1617	4.26 576	3.88 968	10.3	1.6	4.3	**1.1**	1.00	0.28	94%
20	4.28 22902	**3.64** 2875	4.64 676	4.27 1079	29.2	2.8	4.6	**1.2**	0.99	0.25	91%
30	5.97 37032	**5.24** 3717	6.75 606	5.97 1085	41.7	4.2	6.8	**1.3**	0.97	0.23	89%

Table: MC — 100K MovieLens dataset ($\alpha = 0.5$ and $\hat{\rho} = 5 \times 10^{-4}$)

Concluding Remarks

- AC-ACG is an ACG method based on the average of the previously observed curvatures.
- AC-ACG does not require any line search for M_k.
- Convergence rate bound for AC-ACG in terms of the average observed curvatures (novel result).
- A follow-up work [1] presents a FISTA variant of AC-ACG that substantially outperforms previous ACG variants as well as AC-ACG variants.

[1] J. Liang and R. D. C. Monteiro. Average Curvature FISTA for Nonconvex Smooth Composite Optimization Problems. Available on arXiv:2105.06436, 2021.

THE END

Thanks!