[section]

ISyE8813
Composite gradient methods for nonconvex CO problems

Renato D.C. Monteiro

March 27, 2024

1 Unaccelerated CGMs

1.1 Problem description and assumptions

Consider the problem

$$\phi_* = \min_{z\in\mathbb{R}^n} \phi(z) := f(z) + h(z) \tag{1}$$

where

A1) $h \in \overline{\text{Conv}}\,(\mathbb{R}^n)$;

A2) there exists $L > 0$ such that f is L-smooth on $\text{dom}\, h$, i.e., f is differentiable on $\text{dom}\, h$ and

$$\|\nabla f(\tilde{z}) - \nabla f(z)\| \le \|\tilde{z} - z\|, \quad \forall z, \tilde{z} \in \text{dom}\, h \tag{2}$$

A3) f is nonconvex.

Hence,

$$-\frac{M}{2}\|\tilde{z} - z\|^2 \le f(\tilde{z}) - \ell_f(\tilde{z}; z) \le \frac{M}{2}\|\tilde{z} - z\|^2, \quad \forall z, \tilde{z} \in \text{dom}\, h$$

where

$$\ell_f(\cdot; z) := f(z) + \langle \nabla f(z), \cdot - z\rangle \qquad \forall z \in \text{dom}\, h$$

Optimality for (1)**:** If $\bar{x}$ is an local minimum of (1), then

$$0 \in \nabla f(\bar{x}) + \partial h(\bar{x}) \tag{3}$$

A point $\bar{x} \in \mathbb{R}^n$ satisfying (3) is called a stationary point of (1)

Hence, every local min of (1) is a stationary point of (1) but the reverse is not necessarily true.

The method below is the composite gradient method (CGM) for solving (1).

Algorithm 1 Composite Gradient Method (CGM)

0. Let initial point $x_0 \in \operatorname{dom} h$ and $\lambda \in (0, 2/M)$ be given and set $k = 1$
1. Compute

$$x_k \in \operatorname{argmin}_x \left\{ \ell_f(x; x_{k-1}) + h(x) + \frac{1}{2\lambda} \|x - x_{k-1}\|^2 \right\} \tag{4}$$

2. Set $k \leftarrow k + 1$ and go to step 1

Lemma 1.1. *For every $k \geq 1$, the vector $v_k \in \mathbb{R}^n$ defined as*

$$v_k := \frac{1}{\lambda}(x_{k-1} - x_k) + \nabla f(x_k) - \nabla f(x_{k-1})$$

satisfies

$$v_k \in \nabla f(x_k) + \partial h(x_k)$$

and

$$\|v_k\| \leq \left(\frac{1 + \lambda L}{\lambda} \right) \|x_k - x_{k-1}\|$$

Proof: The optimality condition for (4) implies that

$$0 \in \nabla f(x_{k-1}) + \partial h(x_k) + \frac{1}{\lambda}(x_k - x_{k-1})$$

The inclusion in the conclusion of the lemma now follows from the above inclusion and the definition of v_k. Moreover, the definition of v_k and the triangle inequality imply

$$\|v_k\| \leq \lambda^{-1} \|x_{k-1} - x_k\| + \|\nabla f(x_k) - \nabla f(x_{k-1})\| \leq \left(\lambda^{-1} + L\right) \|x_k - x_{k-1}\|$$

where the last inequality is due to (6). Hence, the inequality in the conclusion of the lemma follows. ∎

The size of v_k then tells us how good x_k is as an approximate stationary solution of (1)

Definition: $x \in \operatorname{dom} h$ is said to be a ρ-stationary solution of (1) if there exists $v \in \mathbb{R}^n$ such that

$$v \in \nabla f(x) + \partial h(x), \quad \|v\| \leq \rho$$

Note that the pair $(x, v) = (x_k, v_k)$ generated by CGM satisfies the above inclusion. The question that remains is: how many iterations does CGM need to perform in order to generate a ρ-stationary solution? To answer this question, we will analyze below how $\|v_k\|$ behaves.

Lemma 1.2. *For every $k \geq 1$ and $x \in \mathbb{R}^n$, have*

$$\begin{aligned} \ell_f(x; x_{k-1}) + h(x) + \frac{1}{2\lambda}\|x - x_{k-1}\|^2 - \frac{1}{2\lambda}\|x - x_k\|^2 \\ \geq \ell_f(x_k; x_{k-1}) + h(x_k) + \frac{1}{2\lambda}\|x_k - x_{k-1}\|^2 \end{aligned}$$

Proof: Follows from the popular basic convex analysis result. ■

Lemma 1.3. *For every $k \geq 1$, have*

$$\phi(x_{k-1}) - \phi(x_k) \geq \left(\frac{1}{\lambda} - \frac{M}{2}\right)\|x_k - x_{k-1}\|^2 \geq 0$$

Proof: The first inequality follows from Lemma 1.2 with $x = x_{k-1}$ and the fact that, for every $x \in \mathbb{R}^n$, we have $\ell_f(x; x) = f(x)$ and

$$\ell_f(x; x_{k-1}) \geq f(x) - \frac{M}{2}\|x - x_{k-1}\|^2$$

The second inequality of the lemma follows from the assumption that $\lambda \in (0, 2/M)$ and hence $1/\lambda > M/2$ ■

Lemma 1.4. *For every $K \geq 1$ and $l \in \{0, \dots, K-1\}$, have*

$$\phi(x_l) - \phi(x_K) \geq \frac{\lambda(2 - \lambda M)}{2(1 + \lambda M)^2} \sum_{k=l+1}^{K} \|v_k\|^2$$

As a consequence,

$$\phi(x_l) - \underline{\phi} \geq \frac{\lambda(2 - \lambda M)}{2(1 + \lambda M)^2} \sum_{k=l+1}^{K} \|v_k\|^2$$

Proof: Summing the inequality of Lemma 1.4 from $k = l + 1$ to $k = K$, we have

$$\begin{aligned} \phi(x_l) - \phi(x_K) &\geq \left(\frac{2 - \lambda M}{2\lambda} \right) \sum_{k=l+1}^{K} \|x_k - x_{k-1}\|^2 \\ &\geq \left(\frac{2 - \lambda M}{2\lambda} \right) \frac{\lambda^2}{(1 + \lambda L)^2} \sum_{k=l+1}^{K} \|v_k\|^2 \end{aligned}$$

where the last inequality is due to Lemma 1.1. Hence, the lemma follows. ■

Proposition 1.5. *Assume that the sequence $\{\phi(x_k)\}$ is bounded below by $\underline{\phi}$. Then, for every $K \geq 1$,*

$$\min_{1 \leq k \leq K} \|v_k\|^2 \leq \left[\frac{2(1 + \lambda L)^2}{\lambda(2 - \lambda M)} \right] \left(\frac{\phi(x_0) - \underline{\phi}}{K} \right)$$

Proof: Follows immediately from the previous lemma with $l = 0$. ■

Consequence: The complexity to obtain a ρ-stationary point $x = x_k$ (i.e., a point x such that

$$v \in \partial f(x) + \partial h(x)$$

for some $v \in \mathbb{R}^n$ satisfying $\|v\| \leq \rho$) is

$$\mathcal{O}\left(\frac{\phi(x_0) - \underline{\phi}}{\lambda \rho^2} \right)$$

Under some assumptions, we will next show that

$$\min_{1 \leq k \leq K} \|v_k\|^2 \leq O\left(\frac{mMD^2}{K} + \frac{M^2 d_0^2}{K^2} \right)$$

where m is a weakly convex parameter for f over $\operatorname{dom} h$

1.2 Alternative complexity bound

For this subsection, assume that

A1) $h \in \overline{\mathrm{Conv}}\,(\mathbb{R}^n)$ and $\mathrm{dom}\, h$ is bounded;

A2) f is differentiable on $\mathrm{dom}\, h$ and there exist $m, M > 0$ such that

$$-\frac{m}{2}\|\tilde{z} - z\|^2 \leq f(\tilde{z}) - \ell_f(\tilde{z}; z) \leq \frac{M}{2}\|\tilde{z} - z\|^2, \quad \forall z, \tilde{z} \in \mathrm{dom}\, h \tag{5}$$

A3) f is nonconvex.

It can be shown that (5) implies that $f(\cdot) + (m/2)\|\cdot\|^2$ is convex on $\mathrm{dom}\, h$ and that

$$\|\nabla f(\tilde{z}) - \nabla f(z)\| \leq L\|\tilde{z} - z\|, \quad \forall z, \tilde{z} \in \mathrm{dom}\, h \tag{6}$$

where

$$L := \max\{m, M\}.$$

It can be shown that the set of optimal solutions X_* of (1) is nonempty.

Lemma 1.6. *Assume that $\lambda \in (0, 1/M]$. Then, for every $k \geq 1$ and $x_* \in X_*$, have*

$$\frac{1}{2\lambda}\left(\|x_{k-1}-x_*\|^2 - \|x_k - x_*\|^2\right) + \frac{m}{2}\|x_{k-1}-x_*\|^2 \geq \phi(x_k) - \phi(x_*)$$

Proof: Lemma 1.2 with $x = x_*$ yields

$$\begin{aligned}
&\frac{1}{2\lambda}\|x_* - x_{k-1}\|^2 - \frac{1}{2\lambda}\|x_* - x_k\|^2 \\
&\quad\geq \ell_f(x_k; x_{k-1}) + h(x_k) + \frac{1}{2\lambda}\|x_k - x_{k-1}\|^2 - \ell_f(x_*; x_{k-1}) - h(x_*) \\
&\quad\geq \phi(x_k) + \frac{\lambda^{-1} - M}{2}\|x_k - x_{k-1}\|^2 - \ell_f(x_*; x_{k-1}) - h(x_*) \\
&\quad\geq \phi(x_k) + \frac{\lambda^{-1} - M}{2}\|x_k - x_{k-1}\|^2 - \phi(x_*) - \frac{m}{2}\|x_* - x_{k-1}\|^2
\end{aligned}$$

wehere the second and third inequality is due to (5) ■

Lemma 1.7. *For every $K > k \geq 0$,*

$$\phi(x_k) - \phi(x_*) \geq \frac{\lambda(2-\lambda M)}{2(1+\lambda L)^2}\sum_{k=k+1}^{K}\|v_k\|^2$$

Proof: Follows immediately from Lemma 1.4 with $l = k$ and $\underline{\phi} = \phi(x_*)$ ■

Lemma 1.8. *For every $K \geq 0$,*

$$\Theta_K \leq \frac{(1+\lambda L)^2}{(2-\lambda M)}\left(\frac{d_0^2}{\lambda^2 K(K-1)} + \frac{2mD^2}{\lambda(K-1)}\right)$$

where

$$\Theta_K := \min_{1\leq k\leq K}\|v_k\|^2, \qquad D := \sup\{\|x - x'\| : x, x' \in \operatorname{dom} h\}$$

Proof: Let x_* be the closest point to x_0 in X_*. It follows from the previous lemma that

$$\begin{aligned}
&\frac{\lambda(2-\lambda M)}{2(1+\lambda L)^2}(K-k)\Theta_K \leq \phi(x_k) - \phi(x_*) \\
&\leq \frac{1}{2\lambda}\left(\|x_{k-1}-x_*\|^2 - \|x_k - x_*\|^2\right) + \frac{m}{2}\|x_{k-1}-x_*\|^2 \\
&\leq \frac{1}{2\lambda}\left(\|x_{k-1}-x_*\|^2 - \|x_k - x_*\|^2\right) + \frac{mD^2}{2}
\end{aligned}$$

where the second inequality is due to Lemma 1.6 and the last one is due to the definition of D. Summing the above inequality from $k = 1$ to $k = K$, we conclude that

$$\frac{\lambda(2-\lambda M)}{2(1+\lambda L)^2}\frac{K(K-1)}{2}\Theta_K \leq \frac{1}{2\lambda}\|x_0 - x_*\|^2 + \frac{m}{2}KD^2.$$

and hence that the conclusion of the lemma holds. ■

Hence, if $\lambda = 1/M$, have

$$\Theta_K = \mathcal{O}\left(\frac{L^2 d_0^2}{K(K-1)} + \frac{mL^2D^2}{M(K-1)}\right)$$

and the complexity of finding a ρ-stationary point of (1) is

$$\mathcal{O}\left(\frac{Ld_0}{\rho} + \frac{mL^2D^2}{M\rho^2}\right)$$

Under the assumption that $M \geq m$, and hence $L = M$, the above bound becomes

$$\mathcal{O}\left(\frac{Md_0}{\rho} + \frac{mMD^2}{\rho^2}\right)$$

Exercise: Derive an iteration-complexity bound close to the one above for the case where $\lambda \in (0, 2/M]$

Next lecture, we will present an accelerated CGM whose complexity is

$$\mathcal{O}\left(\left[\frac{Md_0}{\rho}\right]^{2/3} + \frac{mMD^2}{\rho^2} + \sqrt{\frac{mMD^2}{\rho^2}}\right)$$