The University of Texas at Austin
Department of Electrical and Computer Engineering

ECE 381V: Large-Scale Optimization II — Spring 2022

Lecture 14

Caramanis & Mokhtari Monday, March 7, 2022

Goal: In this lecture, we talk about the accelerated proximal gradient method and its convergence properties.

1 Proximal Gradient Method for Composite Optimization

Composite Optimization. In this section, we formally study the composite optimization problem

$$\min f(\mathbf{x}) := \phi(\mathbf{x}) + h(\mathbf{x}) \tag{1}$$

- ϕ is convex and and L-smooth.
- h is convex (possibly nondifferentiable) and its prox operator is easy to compute.
- The optimal solution set is nonempty.

Perhaps the most special case of the above problem and algorithm is the ℓ_1 regularization problem:

$$\min f(\mathbf{x}) := \phi(\mathbf{x}) + \lambda\|\mathbf{x}\|_1$$

Accelerated Proximal Gradient Method (APGM). To solve the above problem, we follow the update of APGM which extends the idea of acceleration to proximal gradient methods.

Initialize: $\mathbf{x}_0 = \mathbf{y}_0 = \mathbf{x}$ and $t_0 = 1$.
Main Loop:

$$\mathbf{x}_{k+1} = \operatorname{prox}_{\eta_k h}(\mathbf{y}_k - \eta_k \nabla\phi(\mathbf{y}_k))$$

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

$$\mathbf{y}_{k+1} = \mathbf{x}_{k+1} + \frac{t_k - 1}{t_{k+1}}(\mathbf{x}_{k+1} - \mathbf{x}_k)$$

Special Case: In the case of ℓ_1 regularized least squares problem, we aim to solve

$$\min f(\mathbf{x}) := \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda\|\mathbf{x}\|_1$$

In this case, APGM is called FISTA and given by

$$\mathbf{x}_{k+1} = \operatorname{shrinkage}_{\eta\lambda}(\mathbf{y}_k - \eta\mathbf{A}^\top(\mathbf{A}\mathbf{y}_k - \mathbf{b}))$$

$$t_{k+1} = (1 + \sqrt{1 + 4t_k^2})/2$$

$$\mathbf{y}_{k+1} = \mathbf{x}_{k+1} + \frac{t_k - 1}{t_{k+1}}(\mathbf{x}_{k+1} - \mathbf{x}_k)$$

Lemma 1. *For the following sequence*

$$t_0 = 1, \qquad t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2}$$

we have

$$t_k \geq \frac{k+2}{2}, \qquad \text{for all } k \geq 0$$

2 Convergence Analysis of Accelerated PGM

First, recall the fundamental lemma that we used in the convergence analysis of PGM.

Lemma 2. *Consider the function $f(\mathbf{x}) = \phi(\mathbf{x}) + h(\mathbf{x})$ under the assumptions that ϕ is L-smooth and h is convex. Then, for any $\eta \leq 1/L$ we have*

$$f(\mathbf{x}) - f(prox_{\eta h}(\mathbf{y}-\eta\nabla\phi(\mathbf{y}))) \geq \frac{1}{2\eta}\|\mathbf{x}-prox_{\eta h}(\mathbf{y}-\eta\nabla\phi(\mathbf{y}))\|^2 - \frac{1}{2\eta}\|\mathbf{x}-\mathbf{y}\|^2 + \phi(\mathbf{x}) - \phi(\mathbf{y}) - \nabla\phi(\mathbf{y})^\top(\mathbf{x}-\mathbf{y})$$

2.1 Convex setting

Next, we use this result to show the convergence rate of APGM for the convex setting.

Theorem 1. *Consider PGM applied to the function $f(\mathbf{x}) = \phi(\mathbf{x}) + h(\mathbf{x})$ with stepsize $\eta_k = \eta \leq 1/L$, for the case that ϕ is L-smooth and convex and h is convex. Then, we have*

$$f(\mathbf{x}_N) - f(\mathbf{x}^*) \leq \frac{2}{\eta}\frac{\|\mathbf{x}^* - \mathbf{x}_0\|^2}{(N+1)^2}$$

Proof. In the result of Lemma 1, replace $\mathbf{x}$ by $t_k^{-1}\mathbf{x}^* + (1-t_k^{-1})\mathbf{x}_k$, $\mathbf{y}$ by $\mathbf{y}_k$ and $\text{prox}_{\eta h}(\mathbf{y}-\eta\nabla\phi(\mathbf{y}))$ by $\mathbf{x}_{k+1}$. Then, we obtain

$$\begin{aligned}
& f(t_k^{-1}\mathbf{x}^* + (1-t_k^{-1})\mathbf{x}_k) - f(\mathbf{x}_{k+1}) \\
& \geq \frac{1}{2\eta}\|t_k^{-1}\mathbf{x}^* + (1-t_k^{-1})\mathbf{x}_k - \mathbf{x}_{k+1}\|^2 - \frac{1}{2\eta}\|t_k^{-1}\mathbf{x}^* + (1-t_k^{-1})\mathbf{x}_k - \mathbf{y}_k\|^2 \\
& \geq \frac{1}{2\eta t_k^2}\|\mathbf{x}^* + (t_k-1)\mathbf{x}_k - t_k\mathbf{x}_{k+1}\|^2 - \frac{1}{2\eta t_k^2}\|\mathbf{x}^* + (t_k-1)\mathbf{x}_k - t_k\mathbf{y}_k\|^2
\end{aligned}$$

Now by convexity of f we can show that

$$f(t_k^{-1}\mathbf{x}^* + (1-t_k^{-1})\mathbf{x}_k) - f(\mathbf{x}_{k+1}) \leq t_k^{-1}f(\mathbf{x}^*) + (1-t_k^{-1})f(\mathbf{x}_k) - f(\mathbf{x}_{k+1}) \leq (1-t_k^{-1})(f(\mathbf{x}_k) - f^*) - (f(\mathbf{x}_{k+1}) - f^*)$$

We further can use $\mathbf{y}_{k+1} = \mathbf{x}_{k+1} + \frac{t_k-1}{t_{k+1}}(\mathbf{x}_{k+1} - \mathbf{x}_k)$ to show that $t_k\mathbf{y}_k = t_k\mathbf{x}_k + (t_{k-1}-1)(\mathbf{x}_k - \mathbf{x}_{k-1})$ and simplify the following expression as

$$\begin{aligned}
\|\mathbf{x}^* + (t_k-1)\mathbf{x}_k - t_k\mathbf{y}_k\|^2 &= \|\mathbf{x}^* + (t_k-1)\mathbf{x}_k - t_k\mathbf{x}_k - (t_{k-1}-1)(\mathbf{x}_k - \mathbf{x}_{k-1})\|^2 \\
&= \|\mathbf{x}^* - \mathbf{x}_k - (t_{k-1}-1)(\mathbf{x}_k - \mathbf{x}_{k-1})\|^2 \\
&= \| - t_{k-1}\mathbf{x}_k + \mathbf{x}^* + (t_{k-1}-1)(\mathbf{x}_{k-1})\|^2 \\
&= \|t_{k-1}\mathbf{x}_k - \mathbf{x}^* - (t_{k-1}-1)(\mathbf{x}_{k-1})\|^2
\end{aligned}$$

Combining the above results leads to

$$\begin{aligned}&(1-t_k^{-1})(f(\mathbf{x}_k)-f^*)-(f(\mathbf{x}_{k+1})-f^*)\\&\geq \frac{1}{2\eta t_k^2}\|t_k\mathbf{x}_{k+1}-\mathbf{x}^*-(t_k-1)\mathbf{x}_k\|^2-\frac{1}{2\eta t_k^2}\|t_{k-1}\mathbf{x}_k-\mathbf{x}^*-(t_{k-1}-1)(\mathbf{x}_{k-1})\|^2\end{aligned}$$

Multiply both sides by t_k^2 to obtain

$$\begin{aligned}&(t_k^2-t_k)(f(\mathbf{x}_k)-f^*)-t_k^2(f(\mathbf{x}_{k+1})-f^*)\\&\geq \frac{1}{2\eta}\|t_k\mathbf{x}_{k+1}-\mathbf{x}^*-(t_k-1)\mathbf{x}_k\|^2-\frac{1}{2\eta}\|t_{k-1}\mathbf{x}_k-\mathbf{x}^*-(t_{k-1}-1)(\mathbf{x}_{k-1})\|^2\end{aligned}$$

Now if we define $\mathbf{u}_k := t_{k-1}\mathbf{x}_k-\mathbf{x}^*-(t_{k-1}-1)(\mathbf{x}_{k-1})$ then we have

$$(t_k^2-t_k)(f(\mathbf{x}_k)-f^*)-t_k^2(f(\mathbf{x}_{k+1})-f^*)\geq \frac{1}{2\eta}\|\mathbf{u}_{k+1}\|^2-\frac{1}{2\eta}\|\mathbf{u}_k\|^2$$

Further, by the update rule of t_k we have $t_k^2-t_k=t_{k-1}^2$ and hence

$$t_{k-1}^2(f(\mathbf{x}_k)-f^*)-t_k^2(f(\mathbf{x}_{k+1})-f^*)\geq \frac{1}{2\eta}\|\mathbf{u}_{k+1}\|^2-\frac{1}{2\eta}\|\mathbf{u}_k\|^2$$

A simple telescope argument for $k\geq 1$ implies that

$$\frac{1}{2\eta}\|\mathbf{u}_N\|^2+t_{N-1}^2(f(\mathbf{x}_N)-f(\mathbf{x}^*))\leq t_0^2(f(\mathbf{x}_1)-f(\mathbf{x}^*))+\frac{1}{2\eta}\|\mathbf{u}_1\|^2$$

Since, $\mathbf{u}_1=\mathbf{x}_1-\mathbf{x}^*$, and $t_0=1$, we have

$$\frac{1}{2\eta}\|\mathbf{u}_N\|^2+t_{N-1}^2(f(\mathbf{x}_N)-f(\mathbf{x}^*))\leq f(\mathbf{x}_1)-f(\mathbf{x}^*)+\frac{1}{2\eta}\|\mathbf{x}_1-\mathbf{x}^*\|^2$$

Now in Lemma 1, if we set $\mathbf{x}=\mathbf{x}^*$ and $\mathbf{y}=\mathbf{y}_0$ then we have

$$f(\mathbf{x}^*)-f(\mathbf{x}_1)\geq \frac{1}{2\eta}\|\mathbf{x}_1-\mathbf{x}^*\|^2-\frac{1}{2\eta}\|\mathbf{y}_0-\mathbf{x}^*\|^2$$

which implies ($\mathbf{y}_0=\mathbf{x}_0$)

$$f(\mathbf{x}_1)-f(\mathbf{x}^*)+\frac{1}{2\eta}\|\mathbf{x}_1-\mathbf{x}^*\|^2\leq \frac{1}{2\eta}\|\mathbf{x}_0-\mathbf{x}^*\|^2$$

Hence, we have

$$\frac{1}{2\eta}\|\mathbf{u}_N\|^2+t_{N-1}^2(f(\mathbf{x}_N)-f(\mathbf{x}^*))\leq \frac{1}{2\eta}\|\mathbf{x}_0-\mathbf{x}^*\|^2$$

which implies that

$$t_{N-1}^2(f(\mathbf{x}_N)-f(\mathbf{x}^*))\leq \frac{1}{2\eta}\|\mathbf{x}_0-\mathbf{x}^*\|^2$$

and since $t_k\geq k+2/2$ we have

$$f(\mathbf{x}_N)-f(\mathbf{x}^*)\leq \frac{2}{\eta(N+1)^2}\|\mathbf{x}_0-\mathbf{x}^*\|^2$$

□

3 Strongly Convex Setting

In the strongly convex setting, we simply used the restarting idea for the convex version of APGM or FISTA. In other words, we run the method for a finite number of iterations, then we use the output of FISTA as the input for the next round of updates by FISTA. We can show that using this procedure, the algorithm converges linearly. In other words, we follow this scheme:

Initialization Pick a point $\mathbf{z}_{-1}$ and set $\mathbf{z}_0 = \text{prox}_{\eta h}(\mathbf{z}_0 - \eta \nabla \phi(\mathbf{z}_0))$.

1. Run FISTA for N iterations with input $\mathbf{z}_k$

2. Set $\mathbf{z}_{k+1}$ as the output of the procedure denoted by $\mathbf{x}_N$

Theorem 2. *Consider the restarting APGM applied to the function $f(\mathbf{x}) = \phi(\mathbf{x}) + h(\mathbf{x})$ with stepsize $\eta_k = \eta \leq 1/L$, for the case that ϕ is L-smooth and μ-strongly convex and h is convex. Hence, after k iterations of Restarted FISTA, which is equivalent to $m = Nk$ iterations of FISTA, an ϵ optimal solution is obtained at the end of last cycle:*

$$f(\mathbf{z}_k) - f^* \leq \epsilon$$

for

$$m \geq \sqrt{8\kappa} \log_2 \frac{\|\mathbf{x}^* - \mathbf{z}_{-1}\|^2}{2\eta\epsilon},$$

Proof. From the previous theorem, we have

$$f(\mathbf{z}_{k+1}) - f(\mathbf{x}^*) \leq \frac{2}{\eta} \frac{\|\mathbf{x}^* - \mathbf{z}_k\|^2}{(N+1)^2}$$

Now using the strong convexity of f we have

$$\frac{\mu}{2}\|\mathbf{x}^* - \mathbf{z}_k\|^2 \leq f(\mathbf{z}_k) - f(\mathbf{x}^*)$$

which implies that

$$f(\mathbf{z}_{k+1}) - f(\mathbf{x}^*) \leq \frac{4}{\eta\mu} \frac{f(\mathbf{z}_k) - f(\mathbf{x}^*)}{(N+1)^2} \leq \frac{4L}{\mu} \frac{(f(\mathbf{z}_k) - f(\mathbf{x}^*))}{(N+1)^2} = \frac{4\kappa(f(\mathbf{z}_k) - f(\mathbf{x}^*))}{(N+1)^2}$$

Now if we set $N = \lceil \sqrt{8\kappa} - 1 \rceil$ then we have

$$f(\mathbf{z}_{k+1}) - f(\mathbf{x}^*) \leq \frac{1}{2}(f(\mathbf{z}_k) - f(\mathbf{x}^*))$$

Hence, after k rounds we have

$$f(\mathbf{z}_k) - f(\mathbf{x}^*) \leq (\frac{1}{2})^k (f(\mathbf{z}_0) - f(\mathbf{x}^*))$$

In the result of Lemma 1, replace $\mathbf{x}$ by $\mathbf{x}^*$, $\mathbf{y}$ by $\mathbf{z}_{-1}$ and $\text{prox}_{\eta h}(\mathbf{y} - \eta\nabla\phi(\mathbf{y}))$ by $\mathbf{z}_0$. Then, we obtain

$$f(\mathbf{x}^*) - f(\mathbf{z}_0) \geq \frac{1}{2\eta}\|\mathbf{x}^* - \mathbf{z}_0\|^2 - \frac{1}{2\eta}\|\mathbf{x}^* - \mathbf{z}_{-1}\|^2$$

and hence

$$f(\mathbf{z}_0) - f(\mathbf{x}^*) \leq \frac{1}{2\eta}\|\mathbf{x}^* - \mathbf{z}_{-1}\|^2$$

Therefore, we have

$$f(\mathbf{z}_k) - f(\mathbf{x}^*) \leq \left(\frac{1}{2}\right)^k \frac{1}{2\eta}\|\mathbf{x}^* - \mathbf{z}_{-1}\|^2$$

Therefore, to obtain an ϵ accurate solution we need to ensure that

$$\left(\frac{1}{2}\right)^k \frac{1}{2\eta}\|\mathbf{x}^* - \mathbf{z}_{-1}\|^2 \leq \epsilon$$

which is equivalent to

$$\left(\frac{1}{2}\right)^k \leq \frac{\epsilon}{\frac{1}{2\eta}\|\mathbf{x}^* - \mathbf{z}_{-1}\|^2}$$

and

$$k \geq \log_2 \frac{\|\mathbf{x}^* - \mathbf{z}_{-1}\|^2}{2\eta\epsilon}$$

Hence, we need the total number of FISTA updates m to be

$$m = Nk \geq \sqrt{8\kappa}\log_2 \frac{\|\mathbf{x}^* - \mathbf{z}_{-1}\|^2}{2\eta\epsilon}$$

□