Week 4 - L07
Faster than GD: Accelerated Methods

CS 295 Optimization for Machine Learning

Ioannis Panageas

Recap (GD)

Theorem (Gradient Descent)**.** *Let $f : \mathbb{R}^d \to \mathbb{R}$ be differentiable, convex (want to minimize) and L-smooth. Let $R = \|x_1 - x^*\|_2$. It holds for $T = \frac{2R^2L}{\epsilon}$*

$$f(x_{T+1}) - f(x^*) \leq \epsilon,$$

with appropriately choosing $\alpha = \frac{1}{L}$.

Remarks

- Speed of convergence is independent of dimension d.
- This result gives a rate of $O\left(\frac{L}{\epsilon}\right)$.

Recap (GD) cont.

Theorem (Gradient Descent). *Let $f : \mathbb{R}^d \to \mathbb{R}$ be differentiable, μ-strongly convex (want to minimize) and L-smooth. Let $R = \|x_1 - x^*\|_2$. It holds for $T = \frac{2L}{\mu} \ln\left(\frac{R}{\epsilon}\right)$*

$$\|x_T - x^*\|_2 \leq \epsilon,$$

with appropriately choosing $\alpha = \frac{1}{L}$.

Remarks

- Speed of convergence is independent of dimension d.
- This result gives a rate of $O\left(\frac{L}{\mu}\log\frac{1}{\epsilon}\right)$. $\kappa := \frac{L}{\mu}$ is called condition number.

Recap (GD) cont.

Theorem (Gradient Descent). *Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be differentiable, μ-strongly convex (want to minimize) and L-smooth. Let $R = \|x_1 - x^*\|_2$. It holds for $T = \frac{2L}{\mu} \ln\left(\frac{R}{\epsilon}\right)$*

$$\|x_T - x^*\|_2 \leq \epsilon,$$

with appropriately choosing $\alpha = \frac{1}{L}$.

Remarks

- Speed of convergence is independent of dimension d.
- This result gives a rate of $O\left(\frac{L}{\mu}\log\frac{1}{\epsilon}\right)$. $\kappa := \frac{L}{\mu}$ is called condition number.

Can we do better?

Accelerated Gradient Descent (Nesterov)

Definition (Accelerated Gradient Descent). *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a differentiable function. The Accelerated Gradient Descent is defined as follows:*

1. Initialization x_1, $y_1 = x_1$, stepsize η.
2. **For** t=1 ... T **do**
3. $\quad y_{t+1} = x_t - \eta \nabla f(x_t)$
4. $\quad x_{t+1} = (1+\gamma_t) y_{t+1} - \gamma_t y_t = y_{t+1} + \gamma_t (y_{t+1} - y_t)$.
5. **End For**

Remarks

- Introduced by Nesterov in 1983. $y_{t+1} - y_t$ is called momentum.
- γ_t is a sequence independent of x_t and $\gamma_t \geq 0$ for all t.

Analysis for smooth, strongly convex functions

Theorem (Strongly convex case). *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice differentiable function, L-smooth and μ-strongly convex function. Assume that x^* is the minimizer and set $\gamma_t := \frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}$ and $\eta = \frac{1}{L}$. Then it holds that*

$$f(y_{t+1}) - f(x^*) \leq \frac{L+\mu}{2} \|x_1 - x^*\|_2^2 \, e^{-\frac{t}{\sqrt{k}}},$$

hence we reach ϵ-close in ℓ_2 after $T := \sqrt{\frac{L}{\mu}} \log\left(\frac{R^2(L+\mu)}{\epsilon}\right)$ iterations.

Remarks

- This result gives a rate of $O\left(\sqrt{\frac{L}{\mu}} \log \frac{1}{\epsilon}\right)$.

Analysis for smooth, strongly convex functions

Proof. We define the following sequence of functions:

- $\Phi_1(x) = f(x_1) + \frac{\mu}{2} \|x - x_1\|_2^2$
- $\Phi_{s+1}(x) = \left(1 - \frac{1}{\sqrt{\kappa}}\right) \Phi_s(x) + \frac{1}{\sqrt{k}} \left(f(x_s) + \nabla f(x_s)^\top (x - x_s) + \frac{\mu}{2} \|x - x_s\|_2^2\right).$

Analysis for smooth, strongly convex functions

Proof. We define the following sequence of functions:

- $\Phi_1(x) = f(x_1) + \frac{\mu}{2} \|x - x_1\|_2^2$
- $\Phi_{s+1}(x) = \left(1 - \frac{1}{\sqrt{\kappa}}\right) \Phi_s(x) + \frac{1}{\sqrt{k}} \left(f(x_s) + \nabla f(x_s)^\top (x - x_s) + \frac{\mu}{2} \|x - x_s\|_2^2\right).$

Intuitively, $\Phi_s(x)$ is a finer approximation from below of $f(x)$. Formally:

Claim (Approximation)**.**

$$\Phi_{s+1} \leq f(x) + \left(1 - \frac{1}{\sqrt{\kappa}}\right)^s (\Phi_1(x) - f(x)).$$

Analysis for smooth, strongly convex functions

Proof of Claim.

$$\Phi_{t+1}(x) = \left(1 - \frac{1}{\sqrt{\kappa}}\right)\Phi_t(x) + \frac{1}{\sqrt{\kappa}}\left(f(x_t) + \nabla f(x_t)^\top (x - x_t) + \frac{\mu}{2}\|x - x_t\|_2^2\right)$$

Analysis for smooth, strongly convex functions

Proof of Claim.

$$\Phi_{t+1}(x) = \left(1 - \frac{1}{\sqrt{\kappa}}\right)\Phi_t(x) + \frac{1}{\sqrt{\kappa}}\left(f(x_t) + \nabla f(x_t)^\top (x - x_t) + \frac{\mu}{2}\|x - x_t\|_2^2\right)$$

$$\leq \left(1 - \frac{1}{\sqrt{\kappa}}\right)\Phi_t(x) + \frac{1}{\sqrt{\kappa}} f(x) \text{ from strong convexity,}$$

Analysis for smooth, strongly convex functions

Proof of Claim.

$$\begin{aligned}
\Phi_{t+1}(x) &= \left(1 - \frac{1}{\sqrt{\kappa}}\right)\Phi_t(x) + \frac{1}{\sqrt{\kappa}}\left(f(x_t) + \nabla f(x_t)^\top (x - x_t) + \frac{\mu}{2}\|x - x_t\|_2^2\right) \\
&\le \left(1 - \frac{1}{\sqrt{\kappa}}\right)\Phi_t(x) + \frac{1}{\sqrt{\kappa}} f(x) \text{ from strong convexity,} \\
&= f(x) + \left(1 - \frac{1}{\sqrt{\kappa}}\right)(\Phi_t(x) - f(x)).
\end{aligned}$$

Therefore $\Phi_{t+1}(x) - f(x) \le \left(1 - \frac{1}{\sqrt{\kappa}}\right)(\Phi_t(x) - f(x))$.

Analysis for smooth, strongly convex functions

Proof of Claim.

$$\Phi_{t+1}(x) = \left(1 - \frac{1}{\sqrt{\kappa}}\right)\Phi_t(x) + \frac{1}{\sqrt{\kappa}}\left(f(x_t) + \nabla f(x_t)^\top (x - x_t) + \frac{\mu}{2}\|x - x_t\|_2^2\right)$$

$$\leq \left(1 - \frac{1}{\sqrt{\kappa}}\right)\Phi_t(x) + \frac{1}{\sqrt{\kappa}} f(x) \text{ from strong convexity,}$$

$$= f(x) + \left(1 - \frac{1}{\sqrt{\kappa}}\right)(\Phi_t(x) - f(x)).$$

Therefore $\Phi_{t+1}(x) - f(x) \leq \left(1 - \frac{1}{\sqrt{\kappa}}\right)(\Phi_t(x) - f(x))$.

Telescopic product: $\Phi_{t+1}(x) - f(x) \leq \left(1 - \frac{1}{\sqrt{\kappa}}\right)^t (\Phi_1(x) - f(x))$.

Analysis for smooth, strongly convex functions

Proof cont.

We also need to bound f from above. Formally we have:

Analysis for smooth, strongly convex functions

Proof cont.

We also need to bound f from above. Formally we have:

Claim (From above).

$$f(y_s) \leq \min_x \Phi_s(x).$$

Assuming for now the claim is true we have:

Analysis for smooth, strongly convex functions

Proof cont.

We also need to bound f from above. Formally we have:

Claim (From above).

$$f(y_s) \leq \min_x \Phi_s(x).$$

Assuming for now the claim is true we have:

$$\begin{aligned} f(y_s) - f(x^*) &\leq \Phi_t(x^*) - f(x^*) \\ &\leq \left(1 - \frac{1}{\sqrt{\kappa}}\right)^t (\Phi_1(x^*) - f(x^*)) \end{aligned}$$

Analysis for smooth, strongly convex functions

Proof cont.

We also need to bound f from above. Formally we have:

Claim (From above).

$$f(y_s) \leq \min_x \Phi_s(x).$$

Assuming for now the claim is true we have:

$$\begin{aligned} f(y_s) - f(x^*) &\leq \Phi_t(x^*) - f(x^*) \\ &\leq \left(1 - \frac{1}{\sqrt{\kappa}}\right)^t (\Phi_1(x^*) - f(x^*)) \\ &\leq \left(1 - \frac{1}{\sqrt{\kappa}}\right)^t \left(f(x_1) - f(x^*) + \frac{\mu}{2} \|x_1 - x^*\|_2^2\right) \end{aligned}$$

Analysis for smooth, strongly convex functions

Proof cont.

W

Since $f(x_1) - f(x^*) \leq \underbrace{\nabla f(x^*)^\top (x_1 - x^*)}_{=0} + \frac{L}{2} \|x_1 - x^*\|_2^2$, we get

C

$$f(y_s) - f(x^*) \leq \left(1 - \frac{1}{\sqrt{\kappa}}\right)^t \frac{L+\mu}{2} \|x_1 - x^*\|_2^2$$

Assuming for now the claim is true we have:

$$\begin{aligned} f(y_s) - f(x^*) &\leq \Phi_t(x^*) - f(x^*) \\ &\leq \left(1 - \frac{1}{\sqrt{\kappa}}\right)^t (\Phi_1(x^*) - f(x^*)) \\ &\leq \left(1 - \frac{1}{\sqrt{\kappa}}\right)^t \left(f(x_1) - f(x^*) + \frac{\mu}{2} \|x_1 - x^*\|_2^2\right) \end{aligned}$$

Analysis for smooth, strongly convex functions

Proof of second Claim. Let's use GD now and induction. For $s = 1$ we have $f(y_1) \le \min_x \Phi_1(x)$ (why?)

Analysis for smooth, strongly convex functions

Proof of second Claim. Let's use GD now and induction. For $s = 1$ we have $f(y_1) \leq \min_x \Phi_1(x)$ (why?). Set $\min_x \Phi_s(x) = \Phi_s^*$.

$$f(y_{s+1}) \leq f(x_s) - \frac{1}{2L} \left\| \nabla f(x_s) \right\|_2^2 \text{ (descent lemma),}$$

$$= \left(1 - \frac{1}{\sqrt{\kappa}}\right) f(y_s) + \left(1 - \frac{1}{\sqrt{\kappa}}\right) \left(f(x_s) - f(y_s)\right) + \frac{1}{\sqrt{\kappa}} f(x_s) - \frac{1}{2L} \left\| \nabla f(x_s) \right\|_2^2$$

Analysis for smooth, strongly convex functions

Proof of second Claim. Let's use GD now and induction. For $s = 1$ we have $f(y_1) \leq \min_x \Phi_1(x)$ (why?). Set $\min_x \Phi_s(x) = \Phi_s^*$.

$$f(y_{s+1}) \leq f(x_s) - \frac{1}{2L} \|\nabla f(x_s)\|_2^2 \text{ (descent lemma)},$$

$$= \left(1 - \frac{1}{\sqrt{\kappa}}\right) f(y_s) + \left(1 - \frac{1}{\sqrt{\kappa}}\right) (f(x_s) - f(y_s)) + \frac{1}{\sqrt{\kappa}} f(x_s) - \frac{1}{2L} \|\nabla f(x_s)\|_2^2$$

$$\overbrace{\leq}^{\text{induction}} \left(1 - \frac{1}{\sqrt{\kappa}}\right) \Phi_s^* + \left(1 - \frac{1}{\sqrt{\kappa}}\right) (f(x_s) - f(y_s)) + \frac{1}{\sqrt{\kappa}} f(x_s) - \frac{1}{2L} \|\nabla f(x_s)\|_2^2$$

Analysis for smooth, strongly convex functions

Proof of second Claim. Let's use GD now and induction. For $s = 1$ we have $f(y_1) \leq \min_x \Phi_1(x)$ (why?). Set $\min_x \Phi_s(x) = \Phi_s^*$.

$$f(y_{s+1}) \leq f(x_s) - \frac{1}{2L}\|\nabla f(x_s)\|_2^2 \text{ (descent lemma)},$$

$$= \left(1 - \frac{1}{\sqrt{\kappa}}\right) f(y_s) + \left(1 - \frac{1}{\sqrt{\kappa}}\right)(f(x_s) - f(y_s)) + \frac{1}{\sqrt{\kappa}} f(x_s) - \frac{1}{2L}\|\nabla f(x_s)\|_2^2$$

$$\overbrace{\leq}^{\text{induction}} \left(1 - \frac{1}{\sqrt{\kappa}}\right) \Phi_s^* + \left(1 - \frac{1}{\sqrt{\kappa}}\right)(f(x_s) - f(y_s)) + \frac{1}{\sqrt{\kappa}} f(x_s) - \frac{1}{2L}\|\nabla f(x_s)\|_2^2$$

$$\overbrace{\leq}^{\text{convexity}} \left(1 - \frac{1}{\sqrt{\kappa}}\right) \Phi_s^* + \left(1 - \frac{1}{\sqrt{\kappa}}\right) \nabla f(x_s)^\top (x_s - y_s) + \frac{1}{\sqrt{\kappa}} f(x_s) - \frac{1}{2L}\|\nabla f(x_s)\|_2^2$$

Analysis for smooth, strongly convex functions

Proof of second Claim. Let's use GD now and induction. For $s = 1$ we have $f(y_1) \leq \min_x \Phi_1(x)$ (why?). Set $\min_x \Phi_s(x) = \Phi_s^*$.

$$f(y_{s+1}) \leq f(x_s) - \frac{1}{2L} \|\nabla f(x_s)\|_2^2 \text{ (descent lemma)},$$

$$= \left(1 - \frac{1}{\sqrt{\kappa}}\right) f(y_s) + \left(1 - \frac{1}{\sqrt{\kappa}}\right) (f(x_s) - f(y_s)) + \frac{1}{\sqrt{\kappa}} f(x_s) - \frac{1}{2L} \|\nabla f(x_s)\|_2^2$$

$$\overbrace{\leq}^{\text{induction}} \left(1 - \frac{1}{\sqrt{\kappa}}\right) \Phi_s^* + \left(1 - \frac{1}{\sqrt{\kappa}}\right) (f(x_s) - f(y_s)) + \frac{1}{\sqrt{\kappa}} f(x_s) - \frac{1}{2L} \|\nabla f(x_s)\|_2^2$$

$$\overbrace{\leq}^{\text{convexity}} \left(1 - \frac{1}{\sqrt{\kappa}}\right) \Phi_s^* + \left(1 - \frac{1}{\sqrt{\kappa}}\right) \nabla f(x_s)^\top (x_s - y_s) + \frac{1}{\sqrt{\kappa}} f(x_s) - \frac{1}{2L} \|\nabla f(x_s)\|_2^2$$

Pause...

Analysis for smooth, strongly convex functions

Proof of second Claim cont. Observe that $\nabla^2 \Phi_s(x) = \mu I_d$,

$$\text{therefore } \Phi_s(x) = \Phi_s^* + \frac{\mu}{2} \|x - v_s\|_2^2 \text{ for some } v_s.$$

Analysis for smooth, strongly convex functions

Proof of second Claim cont. Observe that $\nabla^2 \Phi_s(x) = \mu I_d$,

therefore $\Phi_s(x) = \Phi_s^* + \frac{\mu}{2}\|x - v_s\|_2^2$ for some v_s.

Since $\nabla \Phi_{s+1}(x) = \mu(1 - \frac{1}{\sqrt{\kappa}})(x - v_s) + \frac{1}{\sqrt{\kappa}} \nabla f(x_s) + (1 - \frac{1}{\sqrt{\kappa}})(x - x_s)$ and v_{s+1} is a minimizer of Φ_{s+1} (that is $\nabla \Phi_{s+1}(v_{s+1}) = 0$) we can find a relation for v_{s+1}, v_s.

Analysis for smooth, strongly convex functions

Proof of second Claim cont. Observe that $\nabla^2 \Phi_s(x) = \mu I_d$,

therefore $\Phi_s(x) = \Phi_s^* + \frac{\mu}{2}\|x - v_s\|_2^2$ for some v_s.

Since $\nabla \Phi_{s+1}(x) = \mu(1 - \frac{1}{\sqrt{\kappa}})(x - v_s) + \frac{1}{\sqrt{\kappa}}\nabla f(x_s) + (1 - \frac{1}{\sqrt{\kappa}})(x - x_s)$ and v_{s+1} is a minimizer of Φ_{s+1} (that is $\nabla \Phi_{s+1}(v_{s+1}) = 0$) we can find a relation for v_{s+1}, v_s.

We conclude that

$$\boxed{v_{s+1} = (1 - \frac{1}{\sqrt{\kappa}})v_s + \frac{1}{\sqrt{\kappa}}x_s - \frac{1}{\mu\sqrt{\kappa}}\nabla f(x_s)}$$

Analysis for smooth, strongly convex functions

Proof of second Claim cont. Observe that $\nabla^2 \Phi_s(x) = \mu I_d$,

therefore $\Phi_s(x) = \Phi_s^* + \frac{\mu}{2}\|x - v_s\|_2^2$ for some v_s.

Since $\nabla \Phi_{s+1}(x) = \mu(1 - \frac{1}{\sqrt{\kappa}})(x - v_s) + \frac{1}{\sqrt{\kappa}} \nabla f(x_s) + (1 - \frac{1}{\sqrt{\kappa}})(x - x_s)$ and v_{s+1} is a minimizer of Φ_{s+1} (that is $\nabla \Phi_{s+1}(v_{s+1}) = 0$) we can find a relation for v_{s+1}, v_s.

We conclude that

$$\boxed{v_{s+1} = (1 - \frac{1}{\sqrt{\kappa}})v_s + \frac{1}{\sqrt{\kappa}} x_s - \frac{1}{\mu\sqrt{\kappa}} \nabla f(x_s)}$$

Evaluating Φ_{s+1} at x_s we have

$$\boxed{\Phi_{s+1}^* + \frac{\mu}{2}\|x_s - v_{s+1}\|_2^2 = (1 - \frac{1}{\sqrt{\kappa}})\Phi_s^* + \frac{\mu}{2}(1 - \frac{1}{\sqrt{\kappa}})\|x_s - v_s\|_2^2 + \frac{1}{\sqrt{\kappa}} f(x_s)}$$

Analysis for smooth, strongly convex functions

Proof of second Claim cont.

Evaluating Φ_{s+1} at x_s we have

$$\boxed{\Phi_{s+1}^* + \frac{\mu}{2}\|x_s - v_{s+1}\|_2^2 = (1 - \frac{1}{\sqrt{\kappa}})\Phi_s^* + \frac{\mu}{2}(1 - \frac{1}{\sqrt{\kappa}})\|x_s - v_s\|_2^2 + \frac{1}{\sqrt{\kappa}} f(x_s)}$$

Two last observations:

- $\|x_s - v_{s+1}\|_2^2$ is equal to

$$= (1 - \tfrac{1}{\sqrt{\kappa}})\|x_s - v_s\|_2^2 + \tfrac{1}{\mu^2\kappa}\|\nabla f(x_s)\|_2^2 - \tfrac{2}{\mu\sqrt{\kappa}}\nabla f(x_s)^\top (v_s - x_s),$$

Analysis for smooth, strongly convex functions

Proof of second Claim cont.

Evaluating Φ_{s+1} at x_s we have

$$\boxed{\Phi_{s+1}^* + \frac{\mu}{2}\left\|x_s - v_{s+1}\right\|_2^2 = (1-\frac{1}{\sqrt{\kappa}})\Phi_s^* + \frac{\mu}{2}(1-\frac{1}{\sqrt{\kappa}})\left\|x_s - v_s\right\|_2^2 + \frac{1}{\sqrt{\kappa}}f(x_s)}$$

Two last observations:

- $\left\|x_s - v_{s+1}\right\|_2^2$ is equal to

$$= (1-\tfrac{1}{\sqrt{\kappa}})\left\|x_s - v_s\right\|_2^2 + \tfrac{1}{\mu^2\kappa}\left\|\nabla f(x_s)\right\|_2^2 - \tfrac{2}{\mu\sqrt{\kappa}}\nabla f(x_s)^\top(v_s - x_s),$$

- Assume $v_s - x_s = \sqrt{\kappa}(x_s - y_s)$ then by induction

$$\begin{aligned} v_{s+1} - x_{s+1} &= (1-\tfrac{1}{\sqrt{\kappa}})v_s + \tfrac{1}{\sqrt{\kappa}}x_s - \tfrac{1}{\mu\sqrt{\kappa}}\nabla f(x_s) - x_{s+1} \\ &= \sqrt{\kappa}x_s - (\sqrt{\kappa}-1)y_s - \tfrac{\sqrt{\kappa}}{L}\nabla f(x_s) - x_{s+1} \end{aligned}$$

Analysis for smooth, strongly convex functions

Proof of second Claim cont.

Evaluating Φ_{s+1} at x_s we have

$$\boxed{\Phi_{s+1}^* + \frac{\mu}{2}\|x_s - v_{s+1}\|_2^2 = (1 - \frac{1}{\sqrt{\kappa}})\Phi_s^* + \frac{\mu}{2}(1 - \frac{1}{\sqrt{\kappa}})\|x_s - v_s\|_2^2 + \frac{1}{\sqrt{\kappa}} f(x_s)}$$

Two last observations:

- $\|x_s - v_{s+1}\|_2^2$ is equal to

$$= (1 - \tfrac{1}{\sqrt{\kappa}})\|x_s - v_s\|_2^2 + \tfrac{1}{\mu^2 \kappa}\|\nabla f(x_s)\|_2^2 - \tfrac{2}{\mu\sqrt{\kappa}} \nabla f(x_s)^\top (v_s - x_s),$$

- Assume $v_s - x_s = \sqrt{\kappa}(x_s - y_s)$ then by induction

$$\begin{aligned} v_{s+1} - x_{s+1} &= (1 - \tfrac{1}{\sqrt{\kappa}})v_s + \tfrac{1}{\sqrt{\kappa}} x_s - \tfrac{1}{\mu\sqrt{\kappa}} \nabla f(x_s) - x_{s+1} \\ &= \sqrt{\kappa} x_s - (\sqrt{\kappa} - 1) y_s - \tfrac{\sqrt{\kappa}}{L} \nabla f(x_s) - x_{s+1} \\ &= \sqrt{\kappa} y_{s+1} - (\sqrt{\kappa} - 1) y_s - x_{s+1} = \sqrt{\kappa}(x_{s+1} - y_{s+1}) \end{aligned}$$

Analysis for smooth convex functions

Theorem (L-smooth case). *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice differentiable function, L-smooth. Assume that x^* is the minimizer and set $\eta = \frac{1}{L}$, $\gamma_t := \frac{\lambda_t - 1}{\lambda_{t+1}}$ where $\lambda_0 = 0$ and $\lambda_t = \frac{1+\sqrt{1+4\lambda_{t-1}^2}}{2}$. Then it holds that*

$$f(y_t) - f(x^*) \leq \frac{2L \left\| x_1 - x^* \right\|_2^2}{t^2},$$

hence we reach ϵ-close in value after $T := \left(\sqrt{\frac{2LR^2}{\epsilon}} \right)$ iterations.

Remarks

- This result gives a rate of $O\left(\sqrt{\frac{L}{\epsilon}}\right)$.
- The proof follows similar arguments to the classic GD smooth case.

Analysis for smooth convex functions

Theorem (*L*-smooth case). *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice differentiable function, L-smooth. Assume that x^* is the minimizer and set $\eta = \frac{1}{L}$, $\gamma_t := \frac{\lambda_t - 1}{\lambda_{t+1}}$ where $\lambda_0 = 0$ and $\lambda_t = \frac{1+\sqrt{1+4\lambda_{t-1}^2}}{2}$. Then it holds that*

$$f(y_t) - f(x^*) \leq \frac{2L \|x_1 - x^*\|_2^2}{t^2},$$

hence we reach ϵ-close in value after $T := \left(\sqrt{\frac{2LR^2}{\epsilon}}\right)$ iterations.

Remarks

- This result gives a rate of $O\left(\sqrt{\frac{L}{\epsilon}}\right)$.
- The proof follows similar arguments to the classic GD smooth case.

Remark, this is the best you can do provably!

Conclusion

- Introduction to Accelerated Methods.
  - L-smooth and strongly convex cases.
  - Better rates of convergence (tight)

- Next lecture we will talk about non-convex optimization.