Acceleration Algorithms for Convex Functions

Deeksha Adil
University of Toronto

18 Dec 2020

Preliminaries

- Consider Convex Functions $f : \mathbb{R}^n \to \mathbb{R}$
- Problem: $\min_{x \in Q} f(x)$
- Unconstrained: $Q = \mathbb{R}^n$
  Constrained: $Q =$ Convex Set (Think of linear hyperplanes)

Want Very Fast Algorithms!

Preliminaries

- f is L-smooth if: (Assume norm is Euclidean)

$$\|\nabla f(x) - \nabla f(y)\| \le L\|x - y\|, \quad \forall x, y \in Q\,.$$

- f is p^{th} order L_p-smooth if:

$$\|\nabla^p f(x) - \nabla^p f(y)\| \le L_p\|x - y\|, \quad \forall x, y \in Q\,.$$

- f is μ-strongly convex if:

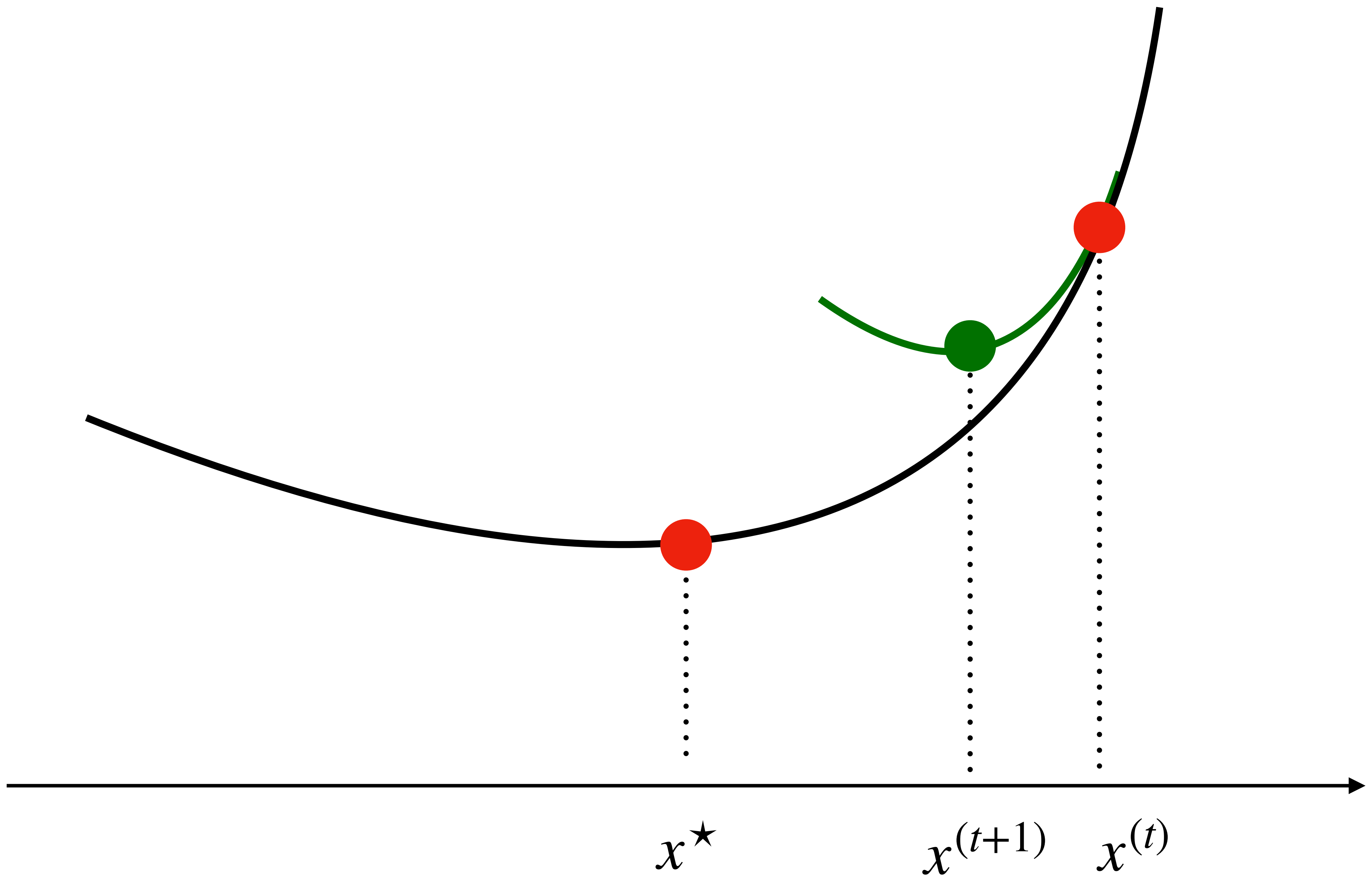
$$\boxed{f(y) - f(x) - \nabla f(x)^\top (y - x)} \ge \mu\|y - x\|^2, \quad \forall x, y \in Q\,.$$

$D_f(y, x)$

First Order Methods

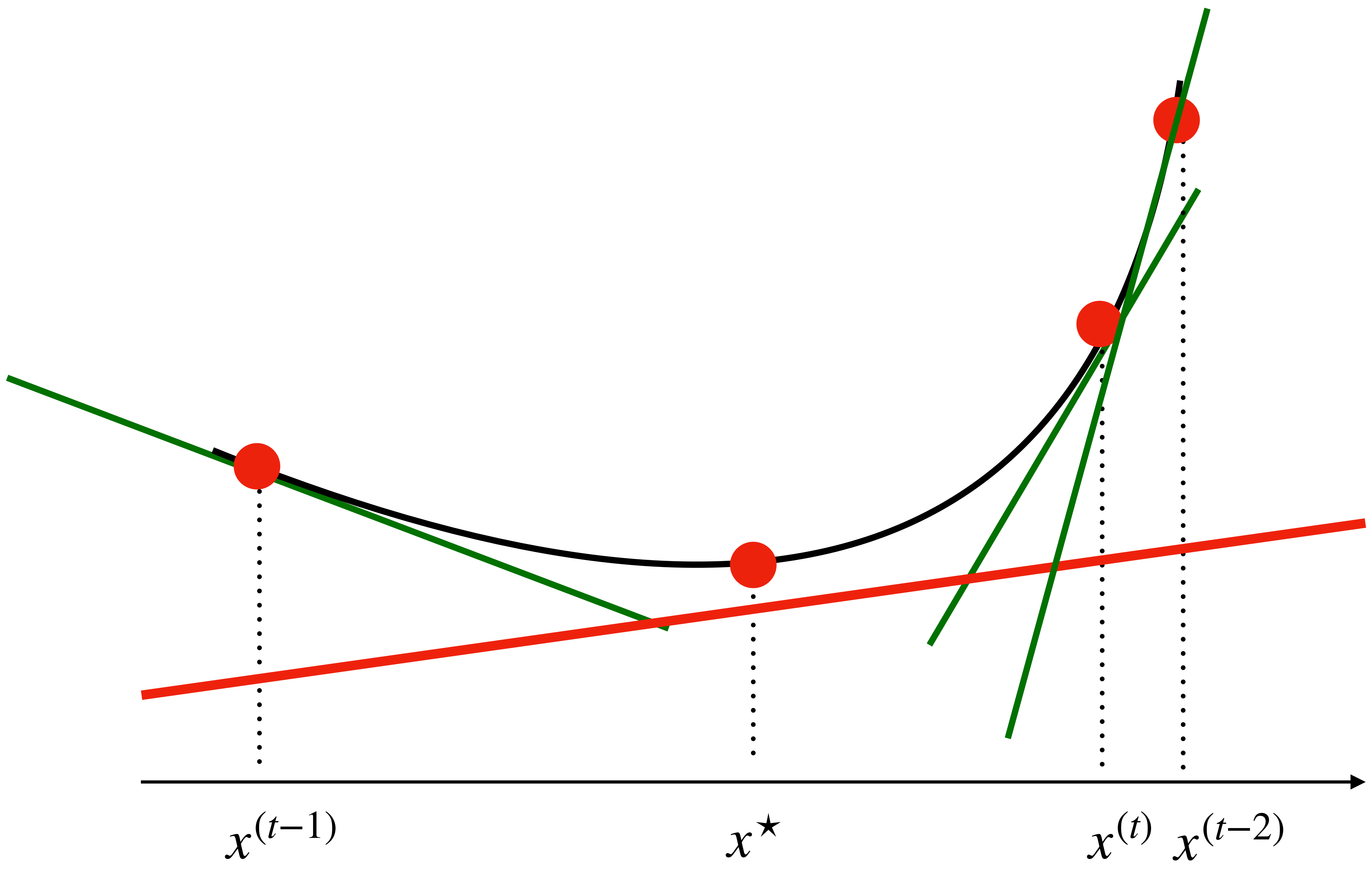
L-smooth function, $Q = \mathbb{R}^n$

Gradient Descent



$$x^{(t+1)} = \arg\min_x \; f(x^{(t)}) + \nabla f(x^{(t)})^\top (x - x^{(t)}) + \frac{L}{2}\|x - x^{(t)}\|^2$$

Mirror Descent



$$x^{(t+1)} = \arg\min_x \quad \alpha\left(f(x^{(t)}) + \nabla f(x^{(t)})^\top (x - x^{(t)})\right) + D_R(x, x^{(t)})$$

First Order Methods

L-smooth function, $Q = \mathbb{R}^n$

Gradient Descent

$$f(x^{(t+1)}) \leq f(x^{t)}) - \frac{1}{2L}\|\nabla f(x^{(t)})\|^2$$

$$f(x^{(T)}) - f(x^\star) \leq O\left(\frac{L\|x^{(0)} - x^\star\|^2}{T}\right)$$

$$T = O\left(\frac{1}{\epsilon}\right)$$

Mirror Descent

$$\alpha\left(f\left(x^{(t)}\right) - f(x^\star)\right) \leq \frac{\alpha^2}{2}\|\nabla f(x^{(t)})\|^2 + D(x^\star, x^{(0)})\,.$$

$$f\left(\sum_{i=1}^{T-1} x^{(i)}\right) - f(x^\star) \leq O\left(\frac{LD(x^\star, x^{(0)})}{\sqrt{T}}\right)$$

$$T = O\left(\frac{1}{\epsilon^2}\right)$$

Accelerated Gradient Descent

[AZO16,DO19]

$$a_0 = 1, \quad \frac{a_i^2}{A_i} = \frac{1}{L}, \quad A_t = \sum_{i=0}^{t-1} a_i$$

$$x^{(t)}$$

$$y^{(t)} = \arg\min_y \; \nabla f(x^{(t)})^\top (y - x^{(t)}) + \frac{L}{2}\|y - x^{(t)}\|^2$$

(GD Step)

$$z^{(t)} = \arg\min_y \; \sum_{i=0}^{t} \frac{a_i}{A_t} \nabla f(x^{(i)})^\top (y - x^{(i)}) + \frac{1}{2A_t}\|y - x^{(0)}\|^2$$

(MD Step)

$$x^{(t+1)} = \frac{A_t}{A_{t+1}} y^{(t)} + \frac{a_t}{A_{t+1}} z^{(t)}$$

$$f(y^T) - f(x^\star) \le O\left(\frac{L\|x^{(0)} - x^\star\|^2}{T^2}\right) \implies T = O\left(\frac{1}{\sqrt{\epsilon}}\right)$$

Higher Order Methods

[Nes19]

Gradient Descent

$$T = O\left(\frac{1}{\epsilon^{1/p}}\right)$$

$$x^{(t+1)} = \arg\min_y \Phi_{x^{(t)},p}(y) + \frac{L_p}{(p+1)!}\|y - x^{(t)}\|^{p+1}$$

Accelerated Gradient Descent

$$T = O\left(\frac{1}{\epsilon^{1/(p+1)}}\right)$$

$x^{(t)}$

$$y^{(t)} = \arg\min_y \ \Phi_{x^{(t)},p}(y) + \frac{L_p}{(p+1)!}\|y - x^{(t)}\|^{p+1}$$

$$z^{(t)} = \arg\min_y \ \sum_{i=0}^{t} \frac{a_i}{A_t} \nabla f(x^{(i)})^\top (y - x^{(i)}) + \frac{C}{(p+1)!A_t}\|y - x^{(0)}\|^{p+1}$$

$$x^{(t+1)} = \frac{A_t}{A_{t+1}} y^{(t)} + \frac{a_t}{A_{t+1}} z^{(t)}$$

$$A_t = O_p\left(\frac{t}{p+1}\right)^{p+1}, \quad a_t = A_{t+1} - A_t$$

Accelerated Taylor Descent

[BJLLS19]

$$x^{(t)} = \frac{A_t}{A_{t+1}} y^{(t)} + \frac{a_{t+1}}{A_{t+1}} z^{(t)}$$

$$y^{(t+1)} = \arg\min_y \ \Phi_{x^{(t)},p}(y) + \frac{L_p}{p!} \|y - x^{(t)}\|^{p+1}$$

$$z^{(t+1)} = \arg\min_y \ \sum_{i=0}^{t} \frac{a_i}{A_t} \nabla f(x^{(i)})^\top (y - x^{(i)}) + \frac{1}{2A_t} \|y - x^{(0)}\|^2$$

$$x^{(t+1)}$$

Constants:

$$a_{t+1}^2 = \lambda_{t+1} A_{t+1}, \quad A_{t+1} = A_t + a_{t+1}$$

$$\frac{1}{2} \leq \lambda_{t+1} \frac{L_p \|y^{(t+1)} - x^{(t)}\|^{p-1}}{(p-1)!} \leq \frac{p}{p+1}$$

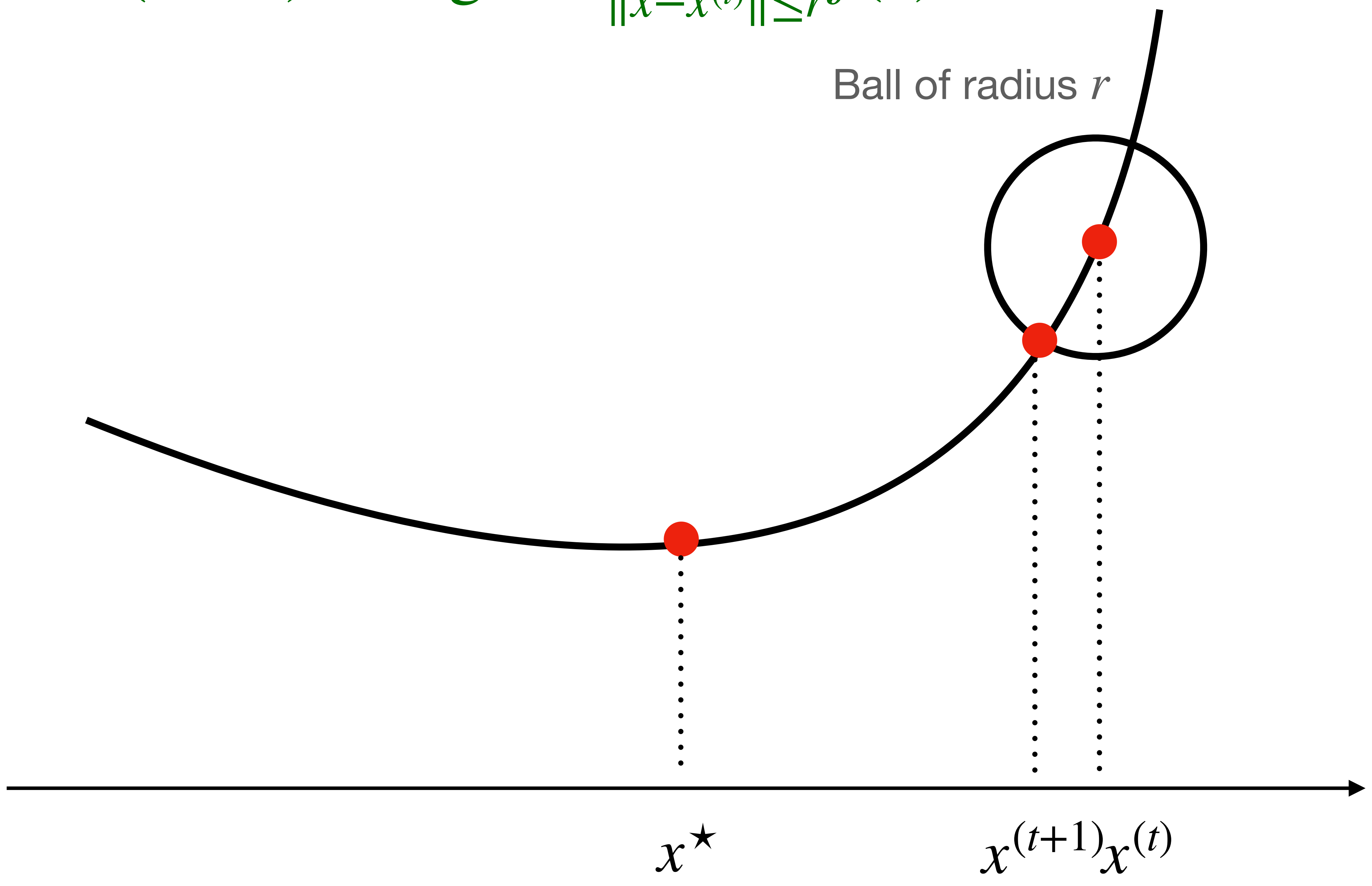
$$f(y^{(T)}) - f(x^\star) \leq O_p\left(\frac{L_p \|x^\star\|^{p+1}}{T^{\frac{3p+1}{2}}}\right)$$

$$p = 1, \quad \lambda_t = \frac{1}{L}$$

Ball Oracle Model

[CJJJLST20]

$B(x^{(t)}, r) = \arg\min_{\|x - x^{(t)}\| \leq r} f(x)$



Simple Algorithm:

$$x^{(t+1)} = B(x^{(t)}, r)$$

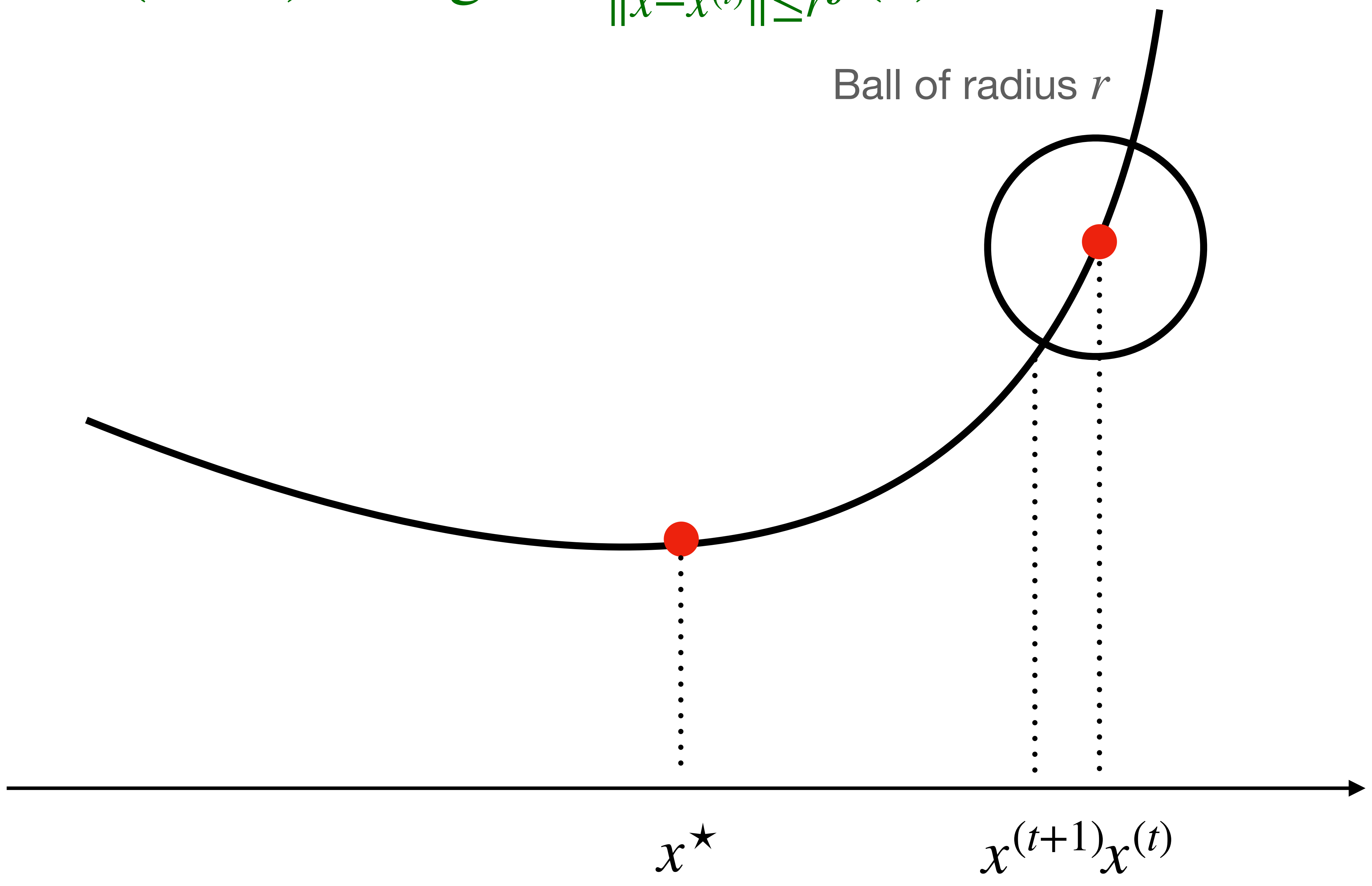
$$\|x^{(t+1)} - x^{(t)}\| = r$$

$$T = O\left(\frac{R}{r}\log\left(\frac{\epsilon_0}{\epsilon}\right)\right)$$

$(f(x^T) - f(x^\star) \leq \epsilon)$

Ball Oracle Model

$B(x^{(t)}, r) = \arg\min_{\|x - x^{(t)}\| \le r} f(x)$



Accelerated Algorithm:

- Similar to ATD, replace p^{th} order taylor expansion with Ball Oracle.

$$\|y^{(t+1)} - x^{(t)}\| = r$$

$$T = O\left(\left(\frac{R}{r}\right)^{2/3} \log\left(\frac{\epsilon_0}{\epsilon}\right)\right)$$

Ball Oracle Model

$f(y^T) - f(x^\star) \leq \epsilon, \quad T = O\left((RM)^{2/3} \log\left(\frac{\epsilon_0}{\epsilon}\right)\right)$

$M - QSC : |\nabla^3 f(x)[h, h, u]| \leq M\|u\| \left(h^\top \nabla^2 f(x) h\right)$

- Logistic Regression: $\sum_i log(1 + exp(-b_i \langle a_i, x \rangle))$ — $O\left(R^{2/3} \log\left(\frac{\epsilon_0}{\epsilon}\right)\right)$
- ℓ_∞-regression: $\|Ax - b\|_\infty$ — $O\left(\frac{m^{1/3}}{\epsilon^{2/3}} \log\left(\frac{\epsilon_0}{\epsilon}\right)\right)$
- ℓ_p-regression: $\|Ax - b\|_p, p \geq 3$ — $O_p\left(m^{1/3} \log\left(\frac{\epsilon_0}{\epsilon}\right)\right)$

Cannot Handle Constrained Setting!

Multiplicative Weights Update

(Original framework: optimises over a simplex)

- ℓ_1-regression: $\min_{Ax=b} \|x\|_1$

  $O\left(\frac{m^{1/3}}{\epsilon^{2/3}} \log\left(\frac{\epsilon_0}{\epsilon}\right)\right)$ [EV19]

- ℓ_∞-regression: $\min_{Ax=b} \|x\|_\infty$

  $O\left(\frac{m^{1/3}}{\epsilon^{2/3}} \log\left(\frac{\epsilon_0}{\epsilon}\right)\right)$ [CKMST11, EV19, AZLOW17]

- ℓ_p-regression: $\min_{Ax=b} \|x\|_p, p \geq 2$

  $O_p\left(m^{1/3} \log\left(\frac{\epsilon_0}{\epsilon}\right)\right)$ [AKPS19]

(Can also handle constrained setting!)

Multiplicative Weights Update

Simple Non-Accelerated Algorithm

- Input $(f, A, b, w^{(0)}, x^{(0)})$
- Initialize $w_e^{(0)} = 1, \forall e$
- For $i = 1 : T$

  $\Delta^{(i)} = Oracle(w^{(i)}, A, b)$

  $x^{(i+1)} = x^{(i)} + \alpha\Delta^{(i)}$

  Update weights
- Return x/T

$Oracle(w, A, b)$ returns Δ such that

1. $A\Delta = b$
2. $\|\Delta\|_\infty \leq \rho$

$$T \approx O(\rho) = O(m^{1/2})$$

Multiplicative Weights Update

Accelerated Algorithm - Width Reduction

- Input $(f, A, b, w^{(0)}, x^{(0)})$
- Initialize $w_e^{(0)} = 1, \forall e$
- For $i = 1 : T$

  $\Delta^{(i)} = Oracle(w^{(i)}, A, b)$

  If $\|\Delta^{(i)}\|_\infty \leq \rho^{2/3}$

  $x^{(i+1)} = x^{(i)} + \alpha \Delta^{(i)}$

  Update weights

  Else, $\forall e, \ s.t., \ |\Delta_e^{(i)}| \geq \rho^{2/3}$

  Double Weights [WIDTH REDUCTiON]

- Return x/T

$Oracle(w, A, b)$ returns Δ such that

1. $A\Delta = b$
2. $\|\Delta\|_\infty \leq \rho$

$$T \approx O(\rho^{2/3}) = O(m^{1/3})$$

Multiplicative Weights Update

Accelerated Algorithm - Analysis

$$Oracle(w, A, b):$$
$$\Delta = \arg\min_{A\Delta=b} \sum_e r_e \Delta_2^2$$
$$\left(r_e \approx f''(w)\right)$$

Track the following Potentials:

$$\Psi(r) = \min_{A\Delta=b} \sum_e r_e \Delta_2^2$$

$$\Phi(w) \approx f^{\star}(w)$$

1. $\Psi(r) \leq O(1)\Phi(w)$
2. $\Phi(w) \leq O(1)$, as long as $T_1 \leq O(\rho^{2/3})$
3. For a width reduction step,
$$\Psi(r') \geq \Psi(r)\left(1 + \Omega\left(\frac{\rho^{4/3}}{m}\right)\right)$$
$$\implies T_2 \leq O\left(\frac{m}{\rho^{4/3}}\right)$$

4. Total iterations:
$$T = O(\rho^{2/3}) + O(m/\rho^{4/3}) = O(m^{1/3})$$

Multiplicative Weights Update

- Can handle constrained settings, unlike optimum higher order methods.
- Rates obtained match optimum higher order rates.
- Very simple oracle used - linear system solver.
- Does not depend on the norm - all higher order methods required Euclidean Norm (whenever needed).

Issue: Mysterious why it achieves such rates. Not known what functions can this work for!!

Expectation: Should work for quasi-self concordant functions at least!

Extensions

- AGD rates for special non-smooth function [AZLOW17]
- Smooth Non-Convex functions: Can use AGD + Movement along Eigenvector to accelerate [CDHS17]
- Want small gradient norms - convex functions: Regularization +AGD+GD used to accelerate [Nes12]
- Convex-Concave functions: Higher order Mirror-Prox can accelerate [BL20]

Thank You!