Lecture 3
Optimal First-order Methods

Jie Lu (jielu@kth.se)

Richard Combes
Alexandre Proutiere

Automatic Control, KTH

September 17, 2013

Lower Complexity Bound (Lipschitz Gradient)

- Consider a class $\mathcal{F}_L(\mathbb{R}^n)$ of convex functions that are
  - Continuously differentiable
  - Lipschitz continuous gradient with Lipschitz constant L > 0

- Use iterative first-order method $\mathcal{M}$

$$x_k \in x_0 + \text{Lin}\{\nabla f(x_0), \ldots, \nabla f(x_{k-1})\}, \forall k \geq 1$$

- Goal: find a function $f \in \mathcal{F}_L(\mathbb{R}^n)$ that is "bad" for all $\mathcal{M}$ (lower bound on convergence rate)

Lower Complexity Bound (Lipschitz Gradient)

- Let L > 0. Consider the family of quadratic functions $f_k : \mathbb{R}^n \to \mathbb{R}$

$$f_k(x) = \frac{L}{4}\left\{\frac{1}{2}[(x^{(1)})^2 + \sum_{i=1}^{k-1}(x^{(i)} - x^{(i+1)})^2 + (x^{(k)})^2] - x^{(1)}\right\}, \quad k = 1, \ldots, n.$$

  - $0 \le \nabla^2 f_k(x) \le LI \Rightarrow f_k \in \mathcal{F}_L$
  - $\nabla^2 f_k(x) = \frac{L}{4}A_k$, where $A_k = \left(\begin{array}{cc} \left.\begin{array}{ccccccc} 2 & -1 & 0 & & & & \\ -1 & 2 & -1 & & 0 & & \\ 0 & -1 & 2 & & & & \\ & \cdots & & & \cdots & & \\ & & & & -1 & 2 & -1 \\ & 0 & & & 0 & -1 & 2 \end{array}\right\} & k \text{ lines} \\ 0_{n-k,k} & 0_{n-k,n-k} \end{array}\right)$
  - $\nabla f_k(x) = \frac{L}{4}(A_k x - e_1)$
  - $\nabla f_k(x_k^\star) = 0 \Rightarrow \quad x_k^\star = \begin{cases} 1 - \frac{i}{k+1}, & i = 1, \ldots, k, \\ 0, & i = k+1, \ldots, n \end{cases}$

$$f^\star = \frac{L}{8}(-1 + \frac{1}{k+1})$$

Lower Complexity Bound (Lipschitz Gradient)

Theorem: For any k, $1 \le k \le \frac{1}{2}(n-1)$ and any $x_0 \in \mathbb{R}^n$, there exists a function $f \in \mathcal{F}_L(\mathbb{R}^n)$ such that for any $\mathcal{M}$,

$$f(x_k) - f^\star \ge \frac{3L\|x_0 - x^\star\|^2}{32(k+1)^2},$$

$$\|x_k - x^\star\|^2 \ge \frac{1}{32}\|x_0 - x^\star\|^2$$

Proof (Sketch). Without loss of generality, assume $x_0 = 0$. Fix k, $1 \le k \le \frac{1}{2}(n-1)$ and apply $\mathcal{M}$ to $f = f_{2k+1}$, generating $\{x_k\}_{k=1}^{\infty}$. Thus, $x^\star = x^\star_{2k+1}$ and $f^\star = f^\star_{2k+1}$.

• Denote $R^{\ell,n} = \{x \in \mathbb{R}^n : x^{(i)} = 0,\ \ell+1 \le i \le n\}$. Let $p \in \{1, 2, \ldots, n\}$. Prove (by induction) that for any $\{y_\ell\}_{\ell=0}^{p}$ where $y_0 = 0$ and each $y_\ell \in \mathcal{L}_\ell \triangleq \mathrm{Lin}\{\nabla f_p(y_0), \ldots, \nabla f_p(y_{\ell-1})\}$, we have $\mathcal{L}_\ell \subset R^{\ell,n}$ $\forall \ell \in \{0, 1, \ldots, p\}$.

$\Rightarrow f_p(y_\ell) = f_\ell(y_\ell) \ge f^\star_\ell$.

• Prove that $\|x^\star_\ell\|^2 \le \frac{\ell+1}{3}$ $\forall \ell \in \{1, 2, \ldots, n\}$.

With the above,

$$\frac{f(x_k) - f^\star}{\|x_0 - x^\star\|^2} \ge \frac{f^\star_k - f^\star_{2k+1}}{\|x^\star_{2k+1}\|^2} \ge \frac{f^\star_k - f^\star_{2k+1}}{(2k+2)/3} \Rightarrow \text{1st inequality.}$$

$$\|x_k - x^\star\|^2 \ge \sum_{i=k+1}^{2k+1} (x_k^{(i)} - x_{2k+1}^{\star(i)})^2 = \sum_{i=k+1}^{2k+1} (x_{2k+1}^{\star(i)})^2 \Rightarrow \text{2nd inequality.}$$

Lower Complexity Bound (Strongly Convex + Lipschitz Gradient)

- Consider a class $\mathcal{S}_{\mu,L}(\mathbb{R}^n)$ of functions that belong to $\mathcal{F}_L(\mathbb{R}^n)$ and are strongly convex with convexity parameter $\mu > 0$
- Let $Q_f = L/\mu$. Consider $f_{\mu,Q_f} : \mathbb{R}^\infty \to \mathbb{R}$

$$f_{\mu,Q_f}(x) = \frac{\mu(Q_f - 1)}{4}\left\{\tfrac{1}{2}[(x^{(1)})^2 + \sum_{i=1}^{\infty}(x^{(i)} - x^{(i+1)})^2] - x^{(1)}\right\} + \tfrac{1}{2}\mu \parallel x \parallel^2 .$$

  - $\nabla^2 f_{\mu,Q_f} = \frac{\mu(Q_f-1)}{4} A + \mu I$, where $A = \begin{pmatrix} 2 & -1 & 0 & \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & \\ & 0 & & \dots \end{pmatrix}$.
  - $\mu I \leq \nabla^2 f_{\mu,Q_f}(x) \leq LI \Rightarrow f_{\mu,Q_f} \in \mathcal{S}_{\mu,L}$, with condition number Q_f
  - $\nabla f_{\mu,Q_f}(x) = \left(\frac{\mu(Q_f-1)}{4} A + \mu I\right) x - \frac{\mu(Q_f-1)}{4} e_1$
  - Let $q = \frac{\sqrt{Q_f}-1}{\sqrt{Q_f}+1}$ and $x^\star \in \mathbb{R}^\infty$ be such that $x^{\star(i)} = q^i$. Then, $\nabla f_{\mu,Q_f}(x^\star) = 0$.

Lower Complexity Bound (Strongly Convex + Lipschitz Gradient)

Theorem: For any $x_0 \in \mathbb{R}^\infty$, there exists $f \in \mathcal{S}_{\mu,L}(\mathbb{R}^\infty)$ s.t. for any $\mathcal{M}$,

$$\|x_k - x^\star\|^2 \geq \left(\frac{\sqrt{Q_f}-1}{\sqrt{Q_f}+1}\right)^{2k} \|x_0 - x^\star\|^2,$$

$$f(x_k) - f^\star \geq \frac{\mu}{2}\left(\frac{\sqrt{Q_f}-1}{\sqrt{Q_f}+1}\right)^{2k} \|x_0 - x^\star\|^2.$$

Proof (Sketch). WLOG, assume $x_0 = 0$. Let $f = f_{\mu,Q_f}$.
Next, prove by induction that $x_k \in R^{k,\infty}$. Consequently,

$$\|x_k - x^\star\|^2 \geq \sum_{i=k+1}^{\infty} (x_k^{(i)} - x^{\star(i)})^2 = \sum_{i=k+1}^{\infty} (x^{\star(i)})^2 = \sum_{i=k+1}^{\infty} q^{2i} = \frac{q^{2(k+1)}}{1-q^2}.$$

Also,

$$\|x_0 - x^\star\|^2 = \sum_{i=1}^{\infty} \left(x^{\star(i)}\right)^2 = \sum_{i=1}^{\infty} q^{2i} = \frac{q^2}{1-q^2}.$$

Therefore, $\|x_k - x^\star\|^2 \geq q^{2k}\|x_0 - x^\star\|^2$, i.e., the 1st inequality holds. The 2nd inequality comes from the strong convexity of f.

Estimate Sequence

- A pair of sequences $\{\phi_k(x)\}_{k=0}^{\infty}$ and $\{\lambda_k\}_{k=0}^{\infty}$, $\lambda_k > 0$ is an *estimate sequence* of $f(x)$ if
  - $\lambda_k \to 0$.
  - $\forall x \in \mathbb{R}^n$, $\forall k \geq 0$, $\phi_k(x) \leq (1-\lambda_k)f(x) + \lambda_k \phi_0(x)$.

- For any $\{x_k\}_{k=0}^{\infty}$, if $f(x_k) \leq \phi_k^\star \triangleq \min_{x\in\mathbb{R}^n} \phi_k(x)\ \forall k \geq 0$, then $f(x_k) - f^\star \leq \lambda_k(\phi_0(x^\star) - f^\star) \to 0$.
  - Convergence rate of $\{\lambda_k\}_{k=0}^{\infty}$ $\Rightarrow$ convergence rate of $\{x_k\}_{k=0}^{\infty}$

- Consider $f \in \mathcal{S}_{\mu,L}(\mathbb{R}^n)$
  - Allow $\mu = 0$ $(\mathcal{S}_{0,L}(\mathbb{R}^n) = \mathcal{F}_L(\mathbb{R}^n))$

- **Task #1:** Construct an estimate sequence $(\{\phi_k(x)\}_{k=0}^{\infty}, \{\lambda_k\}_{k=0}^{\infty})$
  Task #2: Form $\{x_k\}_{k=0}^{\infty}$ that satisfies $f(x_k) \leq \phi_k^\star$

Task #1

- Let ϕ_0 be an arbitrary function and $\{y_k\}_{k=0}^{\infty}$ be an arbitrary sequence. Also let $\{\alpha_k\}_{k=0}^{\infty}$ be s.t. $\alpha_k \in (0,1)$, $\sum_{k=0}^{\infty} \alpha_k = \infty$. Then, $\{\phi_k(x)\}_{k=0}^{\infty}$ and $\{\lambda_k\}_{k=0}^{\infty}$ defined as follows is an estimate sequence:

$$\lambda_0 = 1,$$
$$\lambda_{k+1} = (1-\alpha_k)\lambda_k,$$
$$\phi_{k+1}(x) = (1-\alpha_k)\phi_k(x) + \alpha_k\left(f(y_k) + \nabla f(y_k)^T(x-y_k) + \frac{\mu}{2}\|x-y_k\|^2\right).$$

- Let $\phi_0(x) = \phi_0^\star + \frac{\gamma_0}{2}\|x-v_0\|^2$, $\gamma_0 > 0$. Then, $\phi_k(x) = \phi_k^\star + \frac{\gamma_k}{2}\|x-v_k\|^2$, where

$$\gamma_{k+1} = (1-\alpha_k)\gamma_k + \alpha_k\mu,$$
$$v_{k+1} = \frac{1}{\gamma_{k+1}}\Big((1-\alpha_k)\gamma_k v_k + \alpha_k\mu y_k - \alpha_k\nabla f(y_k)\Big),$$
$$\phi_{k+1}^\star = (1-\alpha_k)\phi_k^\star + \alpha_k f(y_k) - \frac{\alpha_k^2}{2\gamma_{k+1}}\|\nabla f(y_k)\|^2$$
$$+ \frac{\alpha_k(1-\alpha_k)\gamma_k}{\gamma_{k+1}}\left(\frac{\mu}{2}\|y_k-v_k\|^2 + \nabla f(y_k)^T(v_k-y_k)\right).$$

Task #2

- Given $x_0 \in \mathbb{R}^n$, let $v_0 = x_0$ and $\phi_0^\star = f(x_0)$. Then, $f(x_0) = \phi_0^\star$.

- Let $k \geq 0$. Suppose we already have x_k s.t. $f(x_k) \leq \phi_k^\star$.
Because of this and because $f(x_k) \geq f(y_k) + \nabla f(y_k)^T(x_k - y_k)$,

$$\phi_{k+1}^\star \geq \underbrace{f(y_k) - \frac{\alpha_k^2}{2\gamma_{k+1}}\|\nabla f(y_k)\|^2}_{\leq f(x_{k+1})?} + \underbrace{(1-\alpha_k)\nabla f(y_k)^T\left(\frac{\alpha_k\gamma_k}{\gamma_{k+1}}(v_k - y_k) + x_k - y_k\right)}_{=0?}.$$

  - Recall that $f(y_k) - \frac{1}{2L}\|\nabla f(y_k)\|^2 \geq f(y_k - \frac{1}{L}\nabla f(y_k))$. (Gradient descent)

$$\Rightarrow \begin{cases} \frac{\alpha_k^2}{2\gamma_{k+1}} = \frac{1}{2L} \;\Rightarrow\; L\alpha_k^2 = (1-\alpha_k)\gamma_k + \alpha_k\mu, \\ x_{k+1} = y_k - \frac{1}{L}\nabla f(y_k). \end{cases}$$

  - $\frac{\alpha_k\gamma_k}{\gamma_{k+1}}(v_k - y_k) + x_k - y_k = 0$

$$\Rightarrow \; y_k = \frac{\alpha_k\gamma_k v_k + \gamma_{k+1}x_k}{\gamma_k + \alpha_k\mu}$$

Constant Step Scheme

Initialization: Choose $x_0 \in \mathbb{R}^n$ and $\gamma_0 > 0$. Set $v_0 = x_0$.

At each iteration $k \geq 0$:

1) Compute $\alpha_k \in (0,1)$ from $L\alpha_k^2 = (1-\alpha_k)\gamma_k + \alpha_k \mu$.

2) Set $\gamma_{k+1} = (1-\alpha_k)\gamma_k + \alpha_k \mu$.

3) Set $y_k = \frac{\alpha_k \gamma_k v_k + \gamma_{k+1} x_k}{\gamma_k + \alpha_k \mu}$.

4) Set $x_{k+1} = y_k - \frac{1}{L}\nabla f(y_k)$.

5) Set $v_{k+1} = \frac{1}{\gamma_{k+1}}\Big((1-\alpha_k)\gamma_k v_k + \alpha_k \mu y_k - \alpha_k \nabla f(y_k)\Big)$.

- General scheme:
  Step 4) Find x_{k+1} such that $f(x_{k+1}) \leq f(y_k) - \frac{1}{2L}\|\nabla f(y_k)\|^2$.

Convergence Rate

Theorem: Consider the use of the general scheme (thus the constant step scheme) and let $\gamma_0 > 0$ be such that $\gamma_0 \geq \mu$. Then,

$$f(x_k) - f^\star \leq \left(\frac{L+\gamma_0}{2}\right) \min\left\{\left(1 - \sqrt{\frac{\mu}{L}}\right)^k, \frac{4L}{(2\sqrt{L} + k\sqrt{\gamma_0})^2}\right\}\|x_0 - x^\star\|^2, \quad \forall k \geq 0.$$

Proof (sketch). Recall that $\phi_0(x) = f(x_0) + \frac{\gamma_0}{2}\|x - x_0\|^2$ and that $f(x_k) - f^\star \leq \lambda_k(\phi_0(x^\star) - f^\star)$. Since $f(x_0) - f^\star \leq \frac{L}{2}\|x_0 - x^\star\|^2$, we have $f(x_k) - f^\star \leq \lambda_k \cdot \frac{L+\gamma_0}{2}\|x_0 - x^\star\|^2$. Next, derive the convergence rate of λ_k:

- Prove by induction that $\gamma_k \geq \mu$, $\forall k \geq 0$.

$\Rightarrow \alpha_k \geq \sqrt{\frac{\mu}{L}} \Rightarrow \lambda_k \leq \left(1 - \sqrt{\frac{\mu}{L}}\right)^k$

- Prove by induction that $\gamma_k \geq \gamma_0 \lambda_k$, $\forall k \geq 0$.

Using this and $\lambda_{k+1} \leq \lambda_k$, we get $\frac{1}{\sqrt{\lambda_{k+1}}} - \frac{1}{\sqrt{\lambda_k}} \leq \frac{1}{2}\sqrt{\frac{\gamma_0}{L}}$.

$\Rightarrow \lambda_k \leq \frac{4L}{2\sqrt{L}+k\sqrt{\gamma_0}}$.

Therefore, $\lambda_k \leq \min\left\{\left(1 - \frac{\mu}{L}\right)^k, \frac{4L}{(2\sqrt{L}+k\sqrt{\gamma_0})^2}\right\}$, $\forall k \geq 0$.

Why Optimal?

- Let $\gamma_0 = L$. Then, the scheme is optimal for $\mathcal{S}_{\mu,L}(\mathbb{R}^n)$.
  - $\mu > 0$.

    From the lower convexity bound for $\mathcal{S}_{\mu,L}(\mathbb{R}^n)$,

    $$f(x_k) - f^\star \geq \frac{\mu}{2}\Big(\frac{\sqrt{Q_f}-1}{\sqrt{Q_f}+1}\Big)^{2k}\|x_0 - x^\star\|^2 \geq \frac{\mu}{2}\exp\Big(-\frac{4k}{\sqrt{Q_f}-1}\Big)\|x_0 - x^\star\|^2.$$

    $$\Big(\ln(1-q) \geq -\frac{q}{1-q},\ 0 \leq q < 1\Big)$$

    To make $f(x_k) - f^\star < \epsilon$, $k \geq \frac{\sqrt{Q_f}-1}{4}\Big(\ln\frac{1}{\epsilon} + \ln\frac{\mu}{2} + 2\ln\|x_0 - x^\star\|\Big)$.

    From the convergence rate of the scheme,

    $$f(x_k) - f^\star \leq L\Big(1 - \frac{1}{\sqrt{Q_f}}\Big)^k\|x_0 - x^\star\|^2 \leq L\exp\Big(-\frac{k}{\sqrt{Q_f}}\Big)\|x_0 - x^\star\|^2.$$

    For $k \leq \sqrt{Q_f}\Big(\ln\frac{1}{\epsilon} + \ln L + 2\ln\|x_0 - x^\star\|\Big)$, we have $f(x_k) - f^\star < \epsilon$.

  - $\mu = 0$. (Exercise)

Simplification

- Remove $(v_k)_{k=0}^{\infty}$:

$$\begin{cases} y_k = \frac{\alpha_k \gamma_k v_k + \gamma_{k+1} x_k}{\gamma_k + \alpha_k \mu} \\ x_{k+1} = y_k - \frac{1}{L} \nabla f(y_k) \\ v_{k+1} = \frac{1}{\gamma_{k+1}} \Big((1-\alpha_k)\gamma_k v_k + \alpha_k \mu y_k - \alpha_k \nabla f(y_k) \Big) \end{cases}$$

$$\Rightarrow \begin{cases} v_{k+1} = x_k + \frac{1}{\alpha_k}(x_{k+1} - x_k) \\ y_{k+1} = x_{k+1} + \beta_k (x_{k+1} - x_k), \quad \beta_k = \frac{\alpha_{k+1}\gamma_{k+1}(1-\alpha_k)}{\alpha_k(\gamma_{k+1} + \alpha_{k+1}\mu)} \end{cases}$$

$$\begin{cases} \alpha_0^2 L = \gamma_1 = (1-\alpha_0)\gamma_0 + \mu\alpha_0 \\ v_0 = x_0 \end{cases} \quad \Rightarrow y_0 = x_0$$

- Remove $(\gamma_k)_{v=0}^{\infty}$:

$$\begin{cases} \alpha_k^2 L = \gamma_{k+1} \\ \alpha_{k+1}^2 L = (1-\alpha_{k+1})\gamma_{k+1} + \alpha_{k+1}\mu \end{cases} \quad \Rightarrow \begin{cases} \beta_k = \frac{\alpha_k(1-\alpha_k)}{\alpha_k^2 + \alpha_{k+1}} \\ \alpha_{k+1}^2 = (1-\alpha_{k+1})\alpha_k^2 + \alpha_{k+1}\mu/L \end{cases}$$

Simplified Constant Step Scheme

Initialization: Choose $x_0 \in \mathbb{R}^n$ and $\alpha_0 \in (0,1)$. Set $y_0 = x_0$.

At each iteration $k \geq 0$:

1) Set $x_{k+1} = y_k - \frac{1}{L}\nabla f(y_k)$.

2) Compute $\alpha_{k+1} \in (0,1)$ from $\alpha_{k+1}^2 = (1-\alpha_{k+1})\alpha_k^2 + \alpha_{k+1}\mu/L$.

3) Set $\beta_k = \frac{\alpha_k(1-\alpha_k)}{\alpha_k^2+\alpha_{k+1}}$.

4) Set $y_{k+1} = x_{k+1} + \beta_k(x_{k+1} - x_k)$.

- Let $\alpha_0 \geq \sqrt{\frac{\mu}{L}}$ $(\Leftrightarrow \gamma_0 \geq \mu)$. Then, the same convergence rate is derived.
- For $\mu > 0$, if we set $\alpha_0 = \sqrt{\frac{\mu}{L}}$, then $\alpha_k = \sqrt{\frac{\mu}{L}}$, $\beta_k = \frac{\sqrt{L}-\sqrt{\mu}}{\sqrt{L}+\sqrt{\mu}}$, $\forall k \geq 0$.

Initialization: Choose $x_0 \in \mathbb{R}^n$ and set $y_0 = x_0$.

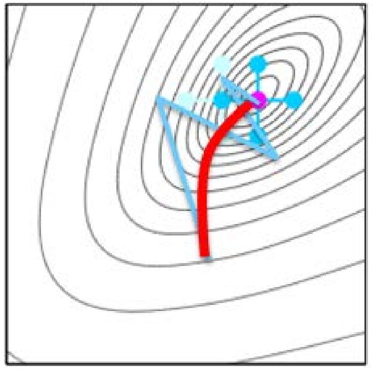
At each iteration $k \geq 0$:

1) Set $x_{k+1} = y_k - \frac{1}{L}\nabla f(y_k)$.

2) Set $y_{k+1} = x_{k+1} + \frac{\sqrt{L}-\sqrt{\mu}}{\sqrt{L}+\sqrt{\mu}}(x_{k+1} - x_k)$.

Heavy Ball Method

- Problem with gradient descent method: cannot avoid zig-zags



- Heavy ball method

$$x_{k+1} = x_k - \alpha \nabla f(x_k) + \beta(x_k - x_{k-1})$$

  - add robustness by accounting for successive moves

- Physical meaning: heavy ball in a potential field under the force of friction

$$m\frac{d^2x(t)}{dt^2} = -\nabla f(x(t)) - p\frac{dx(t)}{dt}$$

- Match Nesterov's lower complexity bound for $S_{\mu,L}(\mathbb{R}^n)$, $\mu > 0$ with optimal parameters

Convergence of Heavy Ball

Theorem: Let $f \in S_{\mu,L}(\mathbb{R}^n) \cap \mathcal{C}^2(\mathbb{R}^n)$. If $0 \le \beta < 1$ and $0 < \alpha < 2(1+\beta)/L$,

$$\|x_k - x^\star\| \le q^k \|x_0 - x^\star\|, \quad \forall k \ge 0,$$

where $q \in (0,1)$ reaches minimum $\frac{\sqrt{L/\mu}-1}{\sqrt{L/\mu}+1}$ for $\alpha = \frac{4}{\sqrt{L}+\sqrt{\mu}}$ and $\beta = \left(\frac{\sqrt{L}-\sqrt{\mu}}{\sqrt{L}+\sqrt{\mu}}\right)^2$.

Proof (Sketch).

$$\begin{bmatrix} x_{k+1} - x^\star \\ x_k - x^\star \end{bmatrix} = \underbrace{\begin{bmatrix} (1+\beta)I_n - \alpha\nabla^2 f(x^\star) & -\beta I_n \\ I_n & 0 \end{bmatrix}}_{A} \begin{bmatrix} x_k - x^\star \\ x_{k-1} - x^\star \end{bmatrix} + o\left(\begin{bmatrix} x_k - x^\star \\ x_{k-1} - x^\star \end{bmatrix}\right)$$

The eigenvalues of A are the eigenvalues of $\begin{bmatrix} 1+\beta-\alpha\lambda_i(\nabla^2 f(x^\star)) & -\beta \\ 1 & 0 \end{bmatrix}$, $i = 1, \dots, n$

i.e., the roots of $\rho^2 - (1+\beta-\alpha\lambda_i(\nabla f(x^\star)))\rho + \beta = 0$.

Since $0 \le \beta < 1$ and $0 < \alpha < 2(1+\beta)/L$, each root ρ_i satisfies $|\rho_i| < 1$.

$\Rightarrow q =$ spectral radius of $A \in (0,1)$.

By solving $\min_{\alpha,\beta} \max_{i=1,\dots,2n} |\rho_i|$, we find the optimal α and β.

Performance of First-order Methods

Problem class	First-order method	Complexity	e=1%
Lipschitz-continuous function	Gradient	$\mathcal{O}(1/\varepsilon^2)$	10,000
Lipschitz-continuous gradient	Gradient	$\mathcal{O}(1/\varepsilon)$	100
	Optimal gradient	$\mathcal{O}(1/\sqrt{\varepsilon})$	10
Strongly convex, Lipschitz gradient	Gradient	$\ln(1/\varepsilon)$	2.3
	Optimal gradient	$\ln(1/\varepsilon)$	

Summary

- Lower complexity bounds
  - Lipschitz gradient
  - Lipschitz gradient+strongly convex
- Nesterov's optimal methods
  - Achieve both lower complexity bounds
- Heavy ball method
  - Achieve lower complexity bound for strongly convex function with Lipschitz gradient with optimal parameter s
- References

  - Y. Nesterov, *Introductory lectures on Convex Optimization: A Basic Course*. Norwell, MA: Kluwer Academic Publishers, 2004.
  - B. Polyak, *Introduction to Optimization*. New York, NY: Optimization Software - Inc, Publications Division, 1987. (available online)