# 3. Optimal gradient methods

- lower complexity bounds
- estimate sequence
- optimal gradient methods

# Lower complexity bound for smooth convex optimization

## computational model

- problem formulation: $\text{minimize}_{x \in \mathbf{R}^n} f(x)$
- problem class: $f$ is convex and $\|\nabla f(x) - \nabla f(y)\|_2 \leq L\|x - y\|_2$
- oracle: first-order local black box
- approximate solution: find $\bar{x}$ such that $f(\bar{x}) - f^\star \leq \epsilon$

**assumption:** iterative algorithm generates a sequence $\{x^{(k)}\}$ such that

$$x^{(k)} \in x^{(0)} + \operatorname{span}\left\{\nabla f(x^{(0)}), \nabla f(x^{(1)}), \ldots, \nabla f(x^{(k-1)})\right\}$$

**theorem** (Nesterov): for any integer $k \leq (n-1)/2$ and any $x^{(0)}$, there exists a function in the problem class such that

$$f(x^{(k)}) - f^\star \geq \frac{3L\|x^{(0)} - x^\star\|_2^2}{32(k+1)^2}$$

**proof:** consider the quadratic function

$$f(x) = \frac{L}{4}\left(\frac{1}{2}\left(x_1^2 + \sum_{i=1}^{n-1}(x_i - x_{i+1})^2 + x_n^2\right) - x_1\right)$$

which can be expressed as $f(x) = \frac{L}{4}\left(\frac{1}{2}x^T Ax - e_1^T x\right)$, where

$$A = \begin{bmatrix} 2 & -1 & 0 & & & & \\ -1 & 2 & -1 & 0 & & & \\ 0 & -1 & 2 & -1 & & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & & -1 & 2 & -1 \\ & & & 0 & -1 & 2 \end{bmatrix}, \qquad e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}$$

- $0 \preceq \nabla^2 f(x) \preceq L \Longrightarrow f$ is convex and $\nabla f(x)$ is $L$-Lipschitz continuous
- optimal solution $x_i^\star = 1 - \frac{i}{n+1}$ for $i = 1, \ldots, n$ (by solving $Ax^\star = e_1$)

$$\|x^\star\|_2^2 = \frac{1}{(n+1)^2}(n^2 + \cdots + 1^2) \leq \frac{1}{3}(n+1)$$

- optimal value: $f(x^\star) = \frac{L}{4}\left(\frac{1}{2}x^{\star T}Ax^\star - e_1^T x^\star\right) = -\frac{L}{8}e_1^T x^\star = -\frac{L}{8}\frac{n}{(n+1)}$

without loss of generality, let $x^{(0)} = 0$; by the tri-diagonal form of $A$,

$$\begin{aligned} \nabla f(x^{(0)}) = -\frac{L}{4} e_1 \quad &\Longrightarrow \quad x^{(1)} \in \operatorname{span}\{e_1\} \\ \Longrightarrow \quad \nabla f(x^{(1)}) \in \operatorname{span}\{e_1, e_2\} \quad &\Longrightarrow \quad x^{(2)} \in \operatorname{span}\{e_1, e_2\} \\ \cdots \quad &\Longrightarrow \quad x^{(k)} \in \operatorname{span}\{e_1, \ldots, e_k\} \end{aligned}$$

therefore

$$f(x^{(k)}) \geq \inf_{x^{(k+1)} = \cdots = x^{(n)} = 0} f(x) = -\frac{L}{8} \frac{k}{(k+1)}$$

for $k \approx n/2$ or $n = 2k + 1$

$$f(x^{(k)}) - f^\star \geq -\frac{L}{8} \frac{k}{(k+1)} + \frac{L}{8} \frac{n}{(n+1)} \geq \frac{L}{16(k+1)}$$

finally

$$\frac{f(x^{(k)}) - f^\star}{\|x^{(0)} - x^\star\|_2^2} \geq \frac{L}{16(k+1)} \Big/ \frac{2k+2}{3} = \frac{3L}{32(k+1)^2}$$

# Lower complexity bound for $\mathcal{S}_{\mu,L}(\mathbf{R}^\infty)$

**computational model**

- formulation: $\mathrm{minimize}_{x\in\ell_2} f(x)$, where $\ell_2 = \{x \in \mathbf{R}^\infty \mid \sum_{i=1}^\infty x_i^2 \leq \infty\}$
- problem class: $f$ is $\mu$-strongly convex & $\|\nabla f(x) - \nabla f(y)\|_2 \leq L\|x-y\|_2$
- oracle: first-order local black box
- approximate solution: find $\bar{x}$ such that $f(\bar{x}) - f^\star \leq \epsilon$

**assumption:** iterative algorithm generates a sequence $\{x^{(k)}\}$ such that

$$x^{(k)} \in x^{(0)} + \mathrm{span}\left\{\nabla f(x^{(0)}), \nabla f(x^{(1)}), \ldots, \nabla f(x^{(k-1)})\right\}$$

**theorem** (Nesterov): for any constants $\mu > 0$ and $\kappa \triangleq L/\mu > 1$, and any $x^{(0)} \in \ell_2$, there exist a function in the problem class such that

$$f(x^{(k)}) - f^\star \;\geq\; \frac{\mu}{2}\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^{2k} \|x^{(0)} - x^\star\|_2^2$$

**proof:** consider the quadratic function

$$f(x) = \frac{\mu(\kappa-1)}{4}\left(\frac{1}{2}\left(x_1^2 + \sum_{i=1}^{\infty}(x_i - x_{i+1})^2\right) - x_1\right) + \frac{\mu}{2}\|x\|^2$$

which can be expressed as $f(x) = \frac{\mu(\kappa-1)}{4}\left(\frac{1}{2}x^T A x - e_1^T x\right) + \frac{\mu}{2}\|x\|^2$, where

$$A = \begin{bmatrix} 2 & -1 & 0 & & & \\ -1 & 2 & -1 & 0 & & \\ 0 & -1 & 2 & -1 & 0 & \\ & \ddots & \ddots & \ddots & \ddots & \ddots \end{bmatrix}, \qquad e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \end{bmatrix}$$

- $0 \preceq A \preceq 4I \Longrightarrow \mu I \preceq \nabla^2 f(x) \preceq LI$
- first-order optimality condition: $\nabla f(x^\star) = 0 \Longrightarrow \left(A + \frac{4}{\kappa-1}\right)x^\star = e_1$

  $$x_i^\star = q^i, \qquad i = 1, 2, \ldots \qquad \text{where} \quad q = \frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}$$

  therefore

  $$\|x^\star\|^2 = \sum_{i=1}^{\infty} x_i^{\star 2} = \sum_{i=1}^{\infty} q^{2i} = \frac{q^2}{1-q^2}$$

without loss of generality, let $x^{(0)} = 0$; by the tri-diagonal form of $A$,

$$
\begin{array}{rcl}
\nabla f(x^{(0)}) = -\frac{L}{4}e_1 & \Longrightarrow & x^{(1)} \in \mathrm{span}\{e_1\} \\
\Longrightarrow\ \nabla f(x^{(1)}) \in \mathrm{span}\{e_1, e_2\} & \Longrightarrow & x^{(2)} \in \mathrm{span}\{e_1, e_2\} \\
\cdots & \Longrightarrow & x^{(k)} \in \mathrm{span}\{e_1, \ldots, e_k\}
\end{array}
$$

therefore

$$
\|x^{(k)} - x^\star\|^2 \geq \sum_{i=k+1}^{\infty} x_i^{\star 2} = \sum_{i=k+1}^{\infty} q^{2i} = \frac{q^{2(k+1)}}{1-q^2} = q^{2k}\|x^{(0)} - x^\star\|^2
$$

by strong convexity with parameter $\mu$,

$$
f(x^{(k)}) - f^\star \ \geq\ \frac{\mu}{2}\|x^{(k)} - x^\star\|^2 \ \geq\ \frac{\mu}{2}q^{2k}\|x^{(0)} - x^\star\|_2^2
$$

# Complexity of the gradient method

**gradient method does not match the lower bound**

- for smooth convex functions ($L$-Lipshichz gradient)

$$f(x^{(k)}) - f^\star \leq \frac{L}{2k}\|x^{(0)} - x^\star\|_2^2$$

- for strongly convex and smooth functions

$$f(x^{(k)}) - f^\star \leq \frac{L}{2}\left(\frac{L-\mu}{L+\mu}\right)^{2k} \|x^{(0)} - x^\star\|_2^2$$

**Nesterov's comments:**

- gradient method relied on decreasing objective values ("relaxation"):

$$f(x^{(k+1)}) \leq f(x^{(k)})$$

- optimal methods: don't rely on relaxation (too *"microscopic"* of a property); use some *global* properties of convex functions

# Estimate sequence (Nesterov)

a pair of sequences $\{\lambda_k, \phi_k(x)\}_{k=0}^{\infty}$ is called *estimate sequence* of $f(x)$ if

- $\lambda_k \to 0$
- $\phi_k(x) \leq (1-\lambda_k)f(x) + \lambda_k \phi_0(x)$ for any $x \in \mathbf{R}^n$ and all $k > 0$

**lemma:** if a sequence $\{x^{(k)}\}$ satisfies $f(x^{(k)}) \leq \min_{x \in \mathbf{R}^n} \phi_k(x)$, then

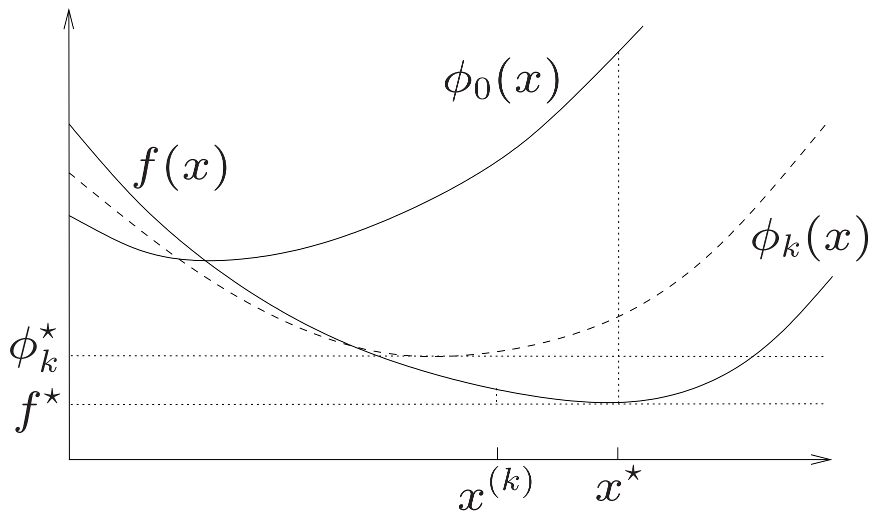
$$f(x^{(k)}) - f^\star \;\leq\; \lambda_k \left(\phi_0(x^\star) - f^\star\right) \;\; \to 0$$

**proof:**

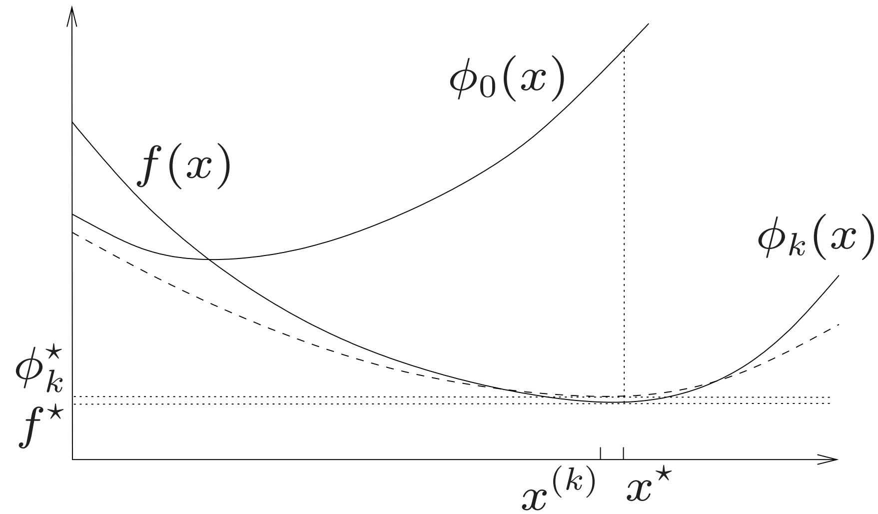
$$\begin{aligned}
f(x^{(k)}) &\leq \min_{x \in \mathbf{R}^n} \phi_k(x) \;\leq\; \min_{x \in \mathbf{R}^n} \{(1-\lambda_k)f(x) + \lambda_k \phi_0(x)\} \\
&\leq (1-\lambda_k)f(x^\star) + \lambda_k \phi_0(x^\star) \\
&= f(x^\star) + \lambda_k \left(\phi_0(x^\star) - f(x^\star)\right)
\end{aligned}$$

**estimate sequence:** pair of sequences $\{\lambda_k, \phi_k(x)\}_{k=0}^{\infty}$ such that

- $\lambda_k \to 0$
- $\phi_k(x) \le (1-\lambda_k) f(x) + \lambda_k \phi_0(x)$ for any $x \in \mathbf{R}^n$ and all $k > 0$




$k$ small




$k$ large

**questions:**

- how to form the estimate sequence?
- how can we ensure $f(x^{(k)}) \le \phi_k^\star \triangleq \min_{x \in \mathbf{R}^n} \phi_k(x)$?

**lemma:** suppose $f \in \mathcal{S}_{\mu,L}(\mathbf{R}^n)$, then for any function $\phi_0(x)$, any sequence $\{y^{(k)}\}_{k=1}^{\infty}$, and $\{\alpha_k\}_{k=0}^{\infty}$ that satisfies

$$\alpha_k \in (0,1), \qquad \sum_{k=1}^{\infty} \alpha_k = \infty$$

the following pair is an estimate sequence

$$\lambda_{k+1} = (1-\alpha_k)\lambda_k, \quad \text{with } \lambda_0 = 1$$

$$\phi_{k+1}(x) = (1-\alpha_k)\phi_k(x) + \alpha_k \left( f(y^{(k)}) + \left\langle \nabla f(y^{(k)}), x - y^{(k)} \right\rangle + \frac{\mu}{2}\|x - y^{(k)}\|^2 \right)$$

**proof:** note $\phi_0(x) \le (1-\lambda_0)f(x) + \lambda_0\phi_0(x) = \phi_0(x)$; use induction

$$\begin{aligned}
\phi_{k+1}(x) &\le (1-\alpha_k)\phi_k(x) + \alpha_k f(x) \\
&= \big(1-(1-\alpha_k)\lambda_k\big) f(x) + (1-\alpha_k)\big(\phi_k(x) - (1-\lambda_k)f(x)\big) \\
&\le \big(1-(1-\alpha_k)\lambda_k\big) f(x) + (1-\alpha_k)\lambda_k\phi_0(x) \\
&= (1-\lambda_{k+1})f(x) + \lambda_{k+1}\phi_0(x)
\end{aligned}$$

$\lambda_k = \lambda_0 \prod_{i=0}^{k}(1-\alpha_i) \to 0$ due to the fact

$$\alpha_k \in (0,1), \quad \sum_{k=1}^{\infty} \alpha_k = \infty \qquad \Longrightarrow \qquad \prod_{k=0}^{\infty}(1-\alpha_k) \to 0$$

proof:

- $\{\lambda_k\}_{k=0}^{\infty}$ monotone decreasing and bounded below, so has limit
- suppose $\lambda_k \to c > 0$
- rewrite iteration as $\lambda_k - \lambda_{k+1} = \alpha_k \lambda_k$, and sum over $k = 0, \ldots, N$

$$\lambda_0 - \lambda_{N+1} = \sum_{k=1}^{N} \alpha_k \lambda_k \geq c \sum_{k=1}^{N} \alpha_k$$

contradiction when $N \to \infty$, so need to have $c = 0$

# Update quadratic approximations

let $\phi_0(x) = \phi_0^\star + \frac{\gamma_0}{2}\|x - v_0\|^2$, then $\{\phi_k(x)\}$ on page 3–11 can be written as

$$\phi_k(x) = \phi_k^\star + \frac{\gamma_k}{2}\|x - v^{(k)}\|^2,$$

where

$$\begin{aligned}
\gamma_{k+1} &= (1-\alpha_k)\gamma_k + \alpha_k\mu \\
v^{(k+1)} &= \frac{1}{\gamma_{k+1}}\Big((1-\alpha_k)\gamma_k v^{(k)} + \alpha_k\mu y^{(k)} - \alpha_k\nabla f(y^{(k)})\Big) \\
\phi_{k+1}^\star &= (1-\alpha_k)\phi_k^\star + \alpha_k f(y^{(k)}) - \frac{\alpha_k^2}{2\gamma_{k+1}}\|\nabla f(y^{(k)})\|^2 \\
&\quad + \frac{\alpha_k(1-\alpha_k)\gamma_k}{\gamma_{k+1}}\Big(\langle\nabla f(y^{(k)}), v^{(k)} - y^{(k)}\rangle + \frac{\mu}{2}\|y^{(k)} - v^{(k)}\|^2\Big)
\end{aligned}$$

(manipulations of simple quadratic functions) assume we already have $\phi_k^\star \geq f(x^{(k)})$, then

$$\begin{aligned}\phi_{k+1}^\star \quad &\geq \quad (1-\alpha_k)f(x^{(k)}) + \alpha_k f(y^{(k)}) - \frac{\alpha_k^2}{2\gamma_{k+1}}\|\nabla f(y^{(k)})\|^2 \\ &\quad\quad + \frac{\alpha_k(1-\alpha_k)\gamma_k}{\gamma_{k+1}}\langle \nabla f(y^{(k)}), v^{(k)} - y^{(k)}\rangle\end{aligned}$$

by convexity, $f(x^{(k)}) \geq f(y^{(k)}) + \langle \nabla f(y^{(k)}), x^{(k)} - y^{(k)}\rangle$,

$$\begin{aligned}\phi_{k+1}^\star \quad &\geq \quad f(y^{(k)}) - \frac{\alpha_k^2}{2\gamma_{k+1}}\|\nabla f(y^{(k)})\|^2 \\ &\quad\quad + (1-\alpha_k)\left\langle \nabla f(y^{(k)}),\ \frac{\alpha_k\gamma_k}{\gamma_{k+1}}(v^{(k)} - y^{(k)}) + x^{(k)} - y^{(k)}\right\rangle\end{aligned}$$

finally, in order to make $\phi_{k+1}^\star \geq f(x^{(k+1)})$,

- choose $x^{(k+1)}$ such that $f(x^{(k+1)}) \leq f(y^{(k)}) - \frac{\alpha_k^2}{2\gamma_{k+1}}\|\nabla f(y^{(k)})\|^2$
- choose $y^{(k)}$ so that $\frac{\alpha_k\gamma_k}{\gamma_{k+1}}(v^{(k)} - y^{(k)}) + x^{(k)} - y^{(k)} = 0$

# Choose $\{y^{(k)}\}$ and $\{x^{(k+1)}\}$

- choose $y^{(k)}$ to eliminate inner-product term

$$y^{(k)} = \frac{1}{\gamma_k + \alpha_k \mu}(\alpha_k \gamma_k v^{(k)} + \gamma_{k+1} x^{(k)})$$

- recall from quadratic upper bound (page 2-6):

$$f\left(y - \frac{1}{L}\nabla f(y)\right) \leq f(y) - \frac{1}{2L}\|\nabla f(y)\|_2^2$$

so we can let

$$x^{(k+1)} = y^{(k)} - \frac{1}{L}\nabla f(y^{(k)})$$

and solve for $\alpha_k$ from the equation $\frac{\alpha_k^2}{\gamma_{k+1}} = \frac{1}{L}$, that is,

$$L\alpha_k^2 = (1-\alpha_k)\gamma_k + \alpha_k \mu$$

# General scheme of optimal method (Nesterov)

- choose $x_0 \in \mathbf{R}^n$ and $\gamma_0 > 0$, and set $v_0 = x_0$
- for $k = 0, 1, 2, \ldots$, repeat

1. find $\alpha_k \in (0,1)$ that satisfies the equation

$$L\alpha_k^2 = (1-\alpha_k)\gamma_k + \alpha_k \mu$$

   and let $\gamma_{k+1} = (1-\alpha_k)\gamma_k + \alpha_k \mu$

2. choose

$$y^{(k)} = \frac{1}{\gamma_k + \alpha_k \mu}(\alpha_k \gamma_k v^{(k)} + \gamma_{k+1} x^{(k)})$$

   and compute $f(y^{(k)})$ and $\nabla f(y^{(k)})$

3. find $x^{(k+1)}$ such that $\quad f(x^{(k+1)}) \leq f(y^{(k)}) - \frac{1}{2L}\|\nabla f(y^{(k)})\|^2$

4. set

$$v^{(k+1)} = \frac{1}{\gamma_{k+1}}\Big((1-\alpha_k)\gamma_k v^{(k)} + \alpha_k \mu y^{(k)} - \alpha_k \nabla f(y^{(k)})\Big)$$

# Bounding $\lambda_k$

**lemma:** if $\gamma_0 \geq \mu$ in the optimal scheme on page 3–16, then

$$\lambda_k = \prod_{i=0}^{k-1}(1-\alpha_i) \leq \min\left\{\left(1-\sqrt{\frac{\mu}{L}}\right)^k, \frac{4L}{(2\sqrt{L}+k\sqrt{\gamma_0})^2}\right\}$$

**proof:**

- $\gamma_k \geq \mu$ and $\alpha_k \geq \sqrt{\mu/L}$ for all $k \geq 0$ because

$$\gamma_{k+1} = L\alpha_k^2 = (1-\alpha_k)\gamma_k + \alpha_k\mu \geq \mu$$

- $\gamma_k \geq \gamma_0\lambda_k$ for all $k \geq 0$, since $\gamma_0 \geq \gamma_0\lambda_0$ and

$$\gamma_{k+1} \geq (1-\alpha_k)\gamma_k \geq (1-\alpha_k)\gamma_0\lambda_k = \gamma_0\lambda_{k+1}$$

- let $a_k = \frac{1}{\sqrt{\lambda_k}}$, then $a_k \geq 1 + \frac{k}{2}\sqrt{\frac{\gamma_0}{L}}$ because

$$\begin{aligned}
a_{k+1} - a_k &= \frac{\sqrt{\lambda_k}-\sqrt{\lambda_{k+1}}}{\sqrt{\lambda_k}\sqrt{\lambda_{k+1}}} = \frac{\lambda_k-\lambda_{k+1}}{\sqrt{\lambda_k}\sqrt{\lambda_{k+1}}(\sqrt{\lambda_k}+\sqrt{\lambda_{k+1}})} \\
&\geq \frac{\lambda_k-\lambda_{k+1}}{2\lambda_k\sqrt{\lambda_{k+1}}} = \frac{\alpha_k\lambda_k}{2\lambda_k\sqrt{\lambda_{k+1}}} = \frac{\alpha_k}{2\sqrt{\lambda_{k+1}}} \geq \frac{1}{2}\sqrt{\frac{\gamma_0}{L}}
\end{aligned}$$

# Rate of convergence

**theorem:** let $\gamma_0 = L$, then the method on page 3–16 generates $\{x^{(k)}\}_{k=0}^{\infty}$ such that

$$f(x^{(k)}) - f^\star \le \min\left\{\left(1 - \sqrt{\frac{\mu}{L}}\right)^k, \frac{4}{(k+2)^2}\right\} L\|x_0 - x^\star\|^2$$

this means the method is *optimal* for functions from class $\mathcal{S}_{\mu,L}(\mathbf{R}^n)$

**proof:** by lemma on page 3–9,

$$f(x^{(k)}) - f^\star \le \lambda_k \left(f(x^{(0)}) - f^\star + \frac{\gamma_0}{2}\|x^{(0)} - x^\star\|^2\right)$$

then use $\gamma_0 = L$ and quadratic upper bound $f(x^{(0)}) - f^\star \le \frac{L}{2}\|x^{(0)} - x^\star\|_2^2$

# Variant of optimal method

eliminate $\{v^{(k)}\}$ and $\{\gamma_k\}$, and use constant step size $t = 1/L$

- choose $x^{(0)} \in \mathbf{R}^n$ and $\alpha_0 \in [\sqrt{\frac{\mu}{L}}, 1)$, set $y^{(0)} = x^{(0)}$ and $q = \mu/L$
- for $k = 0, 1, 2, \ldots$, repeat
  1. compute $f(y^{(k)})$ and $\nabla f(y^{(k)})$, use gradient step update in step 3 in page 3-16, i.e.,
  $$x^{(k+1)} = y^{(k)} - \frac{1}{L}\nabla f(y^{(k)})$$
  2. compute $\alpha_{k+1} \in (0,1)$ from equation
  $$\alpha_{k+1}^2 = (1 - \alpha_{k+1})\alpha_k^2 + q\alpha_{k+1}$$
  and set $\beta_k = \frac{\alpha_k(1-\alpha_k)}{\alpha_k^2+\alpha_{k+1}}$ and
  $$y^{(k+1)} = x^{(k+1)} + \beta_k(x^{(k+1)} - x^{(k)})$$

# A simpler variant

choose $\alpha_0 = \sqrt{\frac{\mu}{L}}$, then

$$\alpha_k = \sqrt{\frac{\mu}{L}}, \qquad \beta_k = \frac{\sqrt{L} - \sqrt{\mu}}{\sqrt{L} + \sqrt{\mu}}$$

- choose $y^{(0)} = x^{(0)} \in \mathbf{R}^n$
- for $k = 0, 1, 2, \ldots$, repeat

$$\begin{aligned} x^{(k+1)} &= y^{(k)} - \frac{1}{L} \nabla f(y^{(k)}) \\ y^{(k+1)} &= x^{(k+1)} + \frac{\sqrt{L} - \sqrt{\mu}}{\sqrt{L} + \sqrt{\mu}} (x^{(k+1)} - x^{(k)}) \end{aligned}$$

however, this scheme does not work for $\mu = 0$

# A simple variant when $\mu = 0$

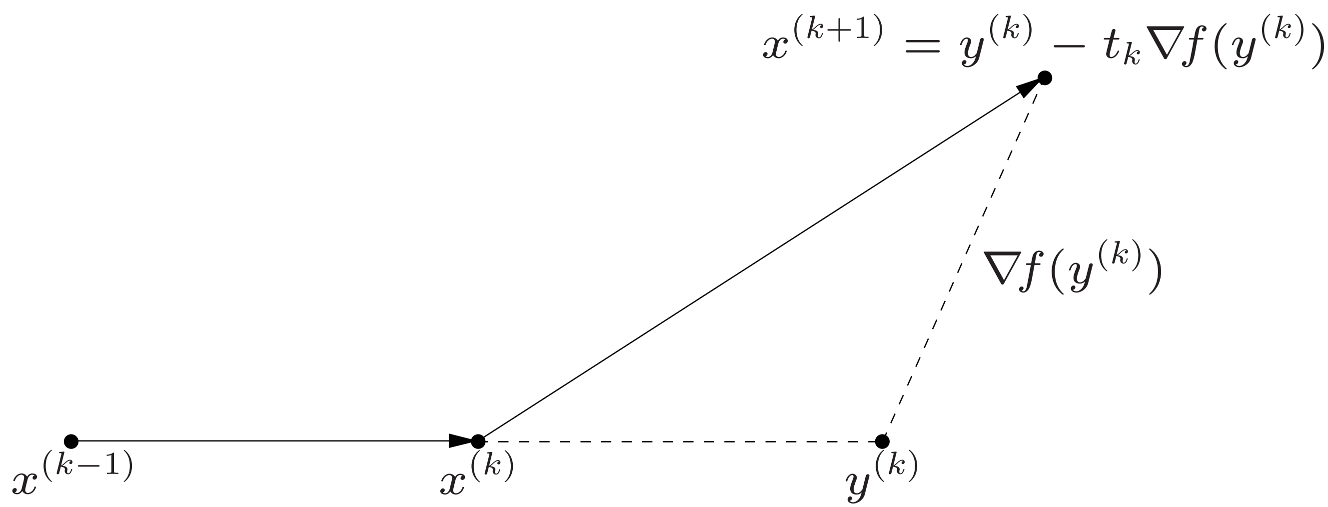
- choose $y^{(0)} = x^{(0)} \in \mathbf{R}^n$
- for $k = 0, 1, 2, \ldots$, repeat

$$
\begin{aligned}
x^{(k+1)} &= y^{(k)} - \frac{1}{L}\nabla f(y^{(k)}) \\
y^{(k+1)} &= x^{(k+1)} + \frac{k}{k+3}(x^{(k+1)} - x^{(k)})
\end{aligned}
$$

when $L$ is unknown, can replace first equation with line search

$$x^{(k+1)} = y^{(k)} - t_k \nabla f(y^{(k)})$$

# Interpretation

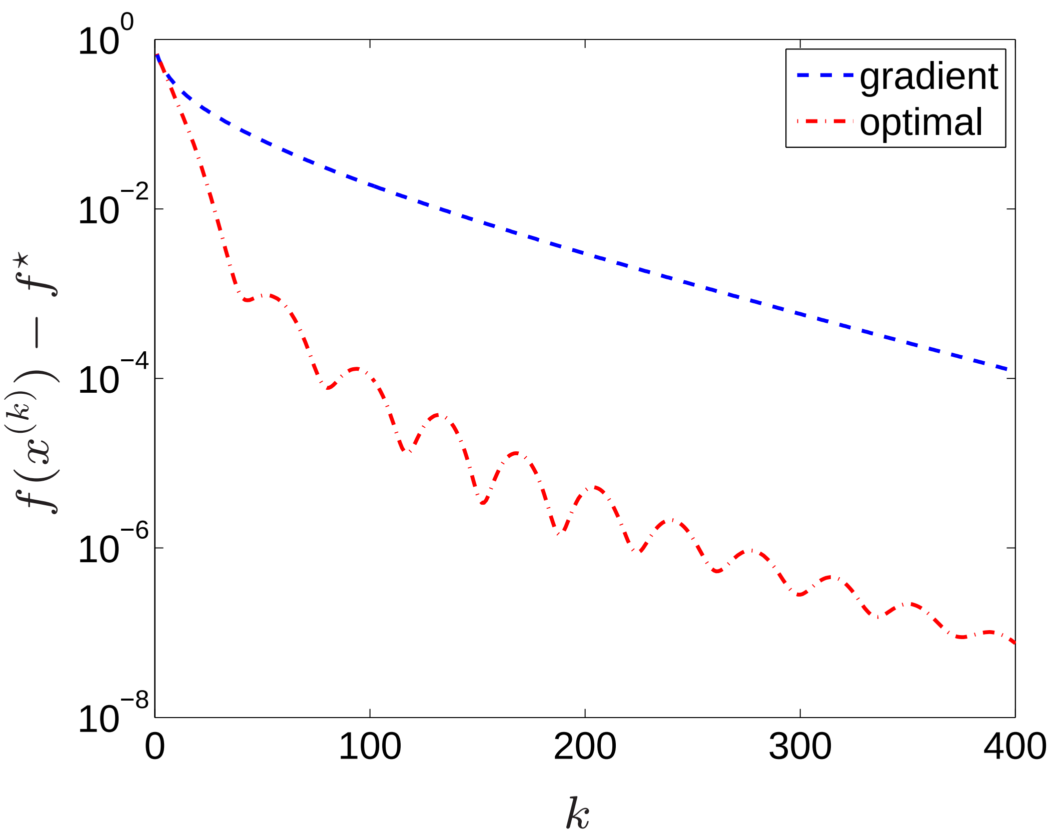



keep the momentum!

# Example

$$\text{minimize} \quad \log\left(\sum_{i=1}^{m} \exp(a_i^T x + b_i)\right)$$

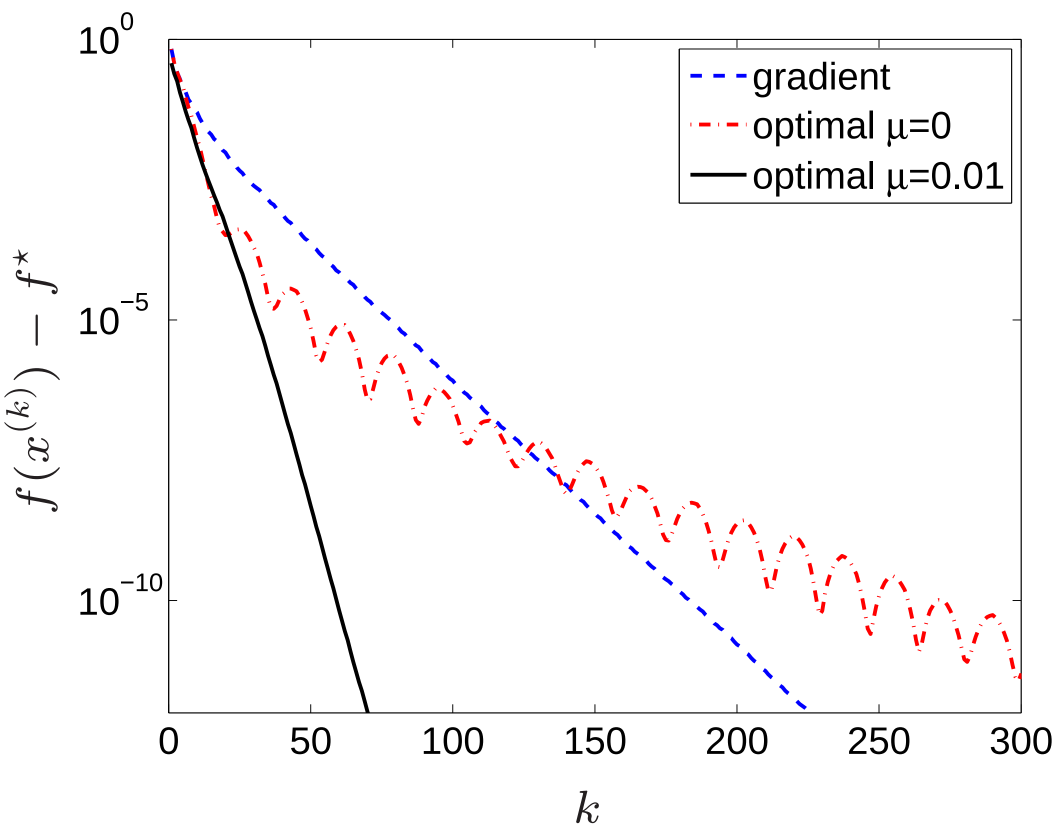
randomly generated data with $m = 500$ and $n = 200$, same fixed step size



# Example

$$\text{minimize} \quad \log\left(\sum_{i=1}^{m} \exp(a_i^T x + b_i)\right)$$

randomly generated data with $m = 500$, $n = 200$, backtracking line search



# References


- Yu. Nesterov, *Introductory Lectures on Convex Optimization. A Basic Course* (2004), Section 2.2.

- P. Tseng, *On accelerated proximal gradient methods for convex-concave optimization* (2008).

- L. Vandenberghe, *Lecture notes for EE236C - Optimization Methods for Large-Scale Systems* (Spring 2011), UCLA.


almost all materials of this lecture are taken from Nesterov's book (2004)
(except the numerical examples)