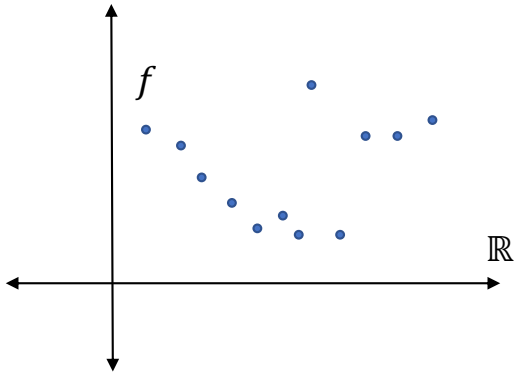
Introduction to Optimization Theory

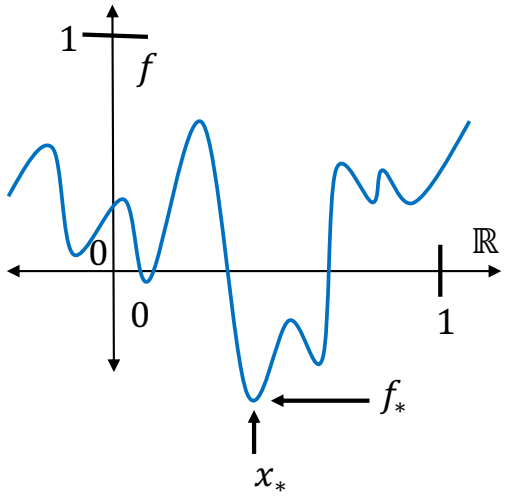
Lecture #6 - 10/1/20

MS&E 213 / CS 2690



Aaron Sidford

sidford@stanford.edu



Plan for Today

Recap

- Gradient descent for smooth function (μ-strongly) convex functions

Geometry

- Linear systems, eigenvalues, and ellipsoids

Acceleration

- Motivation
- Method derivation

Thursday

Proof and extensions

Recap

Problem

$$\min_{x \in \mathbb{R}^n} f(x)$$

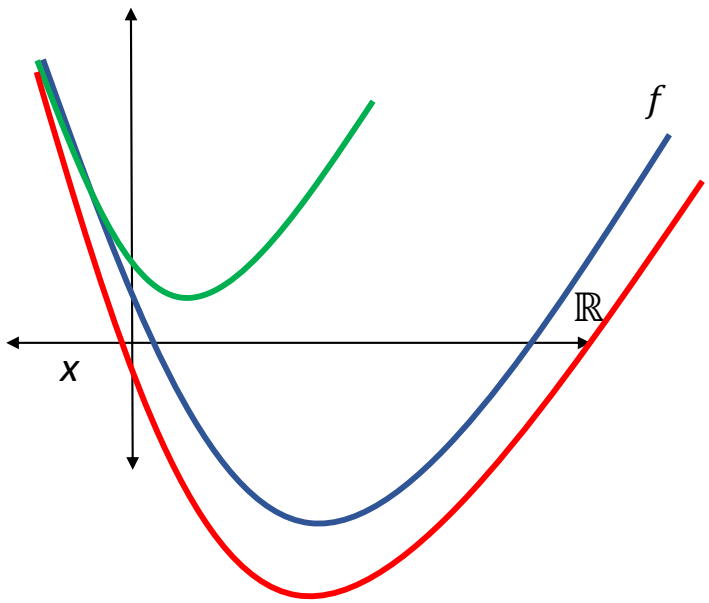
Regularity	Oracle	Goal	Algorithm	Iterations
$n = 1, f(x) \in [0,1], x_* \in [0,1]$	value	½-optimal	anything	∞
$n = 1, x_* \in [0,1]$, L-Lipschitz	value	ϵ-optimal	ϵ-net	$\Theta(L/\epsilon)$
$x_* \in [0,1]$, L-Lipschitz in $\|\cdot\|_\infty$	value	ϵ-optimal	ϵ-net	$\left(\Theta(L/\epsilon)\right)^n$
L-smooth and bounded	value, gradient	ϵ-optimal	ϵ-net	exponential
L-smooth	gradient	ϵ-critical	gradient descent	$O\left(\frac{L(f(x_0) - f_*)}{\epsilon^2}\right)$
L-smooth μ-strongly convex	gradient	ϵ-optimal	gradient descent	$\left(\frac{L}{\mu}\log\left(\frac{f(x_0) - f_*}{\epsilon}\right)\right)$
L-smooth convex	gradient	ϵ-optimal	gradient descent	$O\left(\frac{L\|x_0 - x_*\|_2^2}{\epsilon}\right)$

What does this look like and can we improve?

Recap

Theorem: $f: \mathbb{R}^n \to \mathbb{R}$ is L-smooth and μ-strongly convex (with respect to $\|\cdot\|_2$) if and only if the following hold for all x, y

- $f(y) \leq \boldsymbol{U_x(y)} \stackrel{\text{def}}{=} f(x) + \nabla f(x)^\top (y - x) + \frac{L}{2}\|y - x\|_2^2$
- $f(y) \geq \boldsymbol{L_x(y)} \stackrel{\text{def}}{=} f(x) + \nabla f(x)^\top (y - x) + \frac{\mu}{2}\|y - x\|_2^2$



Algorithm: gradient descent

- Compute upper bound, minimize, repeat
- $x_{k+1} = x_k - \frac{1}{L}\nabla f(x_k)$

Progress Measures

- $\|\nabla f(x)\|_2^2$ - *norm of gradient*
- $f(x) - f_*$ *where* $f_* = \inf_{x \in \mathbb{R}^n} f(x)$ - *function error*
- $\|x - x_*\|_2^2$ - *for minimizer* x_* *(i.e.* $f(x_*) = f_*$*)*

Lemma If $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable then $\nabla f(x_*) = 0$. If f is L-smooth then

$$\frac{1}{2L}\|\nabla f(x)\|_2^2 \leq f(x) - f(x_*) \leq \frac{L}{2}\|x - x_*\|_2^2$$

Lemma If $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable and μ-strongly convex

$$\frac{1}{2\mu}\|\nabla f(x)\|_2^2 \geq f(x) - f(x_*) \geq \frac{\mu}{2}\|x - x_*\|_2^2$$

Lemma If f is differentiable and convex then for all minimizers x_*

$$f(x) - f_* \leq \|\nabla f(x)\|_2 \cdot \|x - x_*\|_2^2$$

Note
Minimizer unique for strongly convex ($\mu > 0$) functions.

Note
All progress measures decrease by $1 - \frac{\mu}{L}$ with one gradient descent step when twice differentiable (in ℓ_2).

Recap

Problem

$$\min_{x \in \mathbb{R}^n} f(x)$$

Regularity	Oracle	Goal	Algorithm	Iterations
$n = 1, f(x) \in [0,1], x_* \in [0,1]$	value	½-optimal	anything	∞
$n = 1, x_* \in [0,1]$, L-Lipschitz	value	ϵ-optimal	ϵ-net	$\Theta(L/\epsilon)$
$x_* \in [0,1]$, L-Lipschitz in $\|\cdot\|_\infty$	value	ϵ-optimal	ϵ-net	$\left(\Theta(L/\epsilon)\right)^n$
L-smooth and bounded	value, gradient	ϵ-optimal	ϵ-net	exponential
L-smooth	gradient	ϵ-critical	gradient descent	$O\left(\frac{L(f(x_0) - f_*)}{\epsilon^2}\right)$
L-smooth μ-strongly convex	gradient	ϵ-optimal	gradient descent	$\left(\frac{L}{\mu}\log\left(\frac{f(x_0) - f_*}{\epsilon}\right)\right)$
L-smooth convex	gradient	ϵ-optimal	gradient descent	$O\left(\frac{L\|x_0 - x_*\|_2^2}{\epsilon}\right)$

What does this look like and can we improve?

Plan for Today

Recap

- Gradient descent for smooth function (μ-strongly) convex functions

Geometry

- Linear systems, eigenvalues, and ellipsoids

Acceleration

- Motivation
- Method derivation

Thursday

Proof and extensions

Illustrative Example

Goal: solve $Ax = b$ where $A \in \mathbb{R}^{n \times n}$ is symmetric, $A = A^\top$, and A is positive definite (PD), i.e. $z^\top Az > 0$ for all $z \neq 0$

Step #1: turn into optimization problem

- $\min_{x \in \mathbb{R}^n} f(x)$ for $f(x) = \frac{1}{2} x^\top Ax - b^\top x$

- One of simplest smooth, strongly convex problems
- Good to look at for method intuition

Why?

- $\nabla f(x) = Ax - b$
- $\nabla^2 f(x) = A$

Critical point ⇒ linear system solution!

Strongly convex!
($z^\top Az > 0$ for all $z \neq 0$)

approximate min ⇒ approximmate system solution!

Illustrative Example

- *Goal:* $Ax = b \Rightarrow \min_{x \in \mathbb{R}^n} f(x)$ *for* $f(x) = \frac{1}{2} x^\top Ax - b^\top x$
- *Symmetric:* $A = A^\top$
- *PD:* $z^\top Az > 0$ *for all* $z \neq 0$

Step #2: look at structure

- What does this look like?
- What are smoothness and strong convexity parameters?

Lemma: If $f: \mathbb{R}^n \to \mathbb{R}$ is twice differentiable and $\nabla^2 f(x) = M$ for all $x \in \mathbb{R}^n$
Then $f(y) = f(x) + \nabla f(x)^\top (y - x) + \frac{1}{2} (y - x)^\top M (y - x)$ for all $x, y \in \mathbb{R}^n$.

Proof: $x_t = x + t(y - x)$

- $f(y) = f(x) + \nabla f(x)^\top (y - x) + \int_0^1 \int_0^t (y - x)^\top \nabla^2 f(x_\alpha)(y - x)\, d\alpha\, dt$

$$= \frac{1}{2} (x - x_*)^\top M (x - x_*)$$

Illustrative Example

- *Goal:* $Ax = b \Rightarrow \min_{x \in \mathbb{R}^n} f(x)$ *for* $f(x) = \frac{1}{2} x^\top Ax - b^\top x$
- *Symmetric:* $A = A^\top$
- *PD:* $z^\top Az > 0$ *for all* $z \neq 0$

Step #2: look at structure?

- What does this look like?
- What are smoothness and strong convexity parameters?

Lemma: If $f: \mathbb{R}^n \to \mathbb{R}$ is twice differentiable and $\nabla^2 f(x) = M$ for all $x \in \mathbb{R}^n$
Then $f(y) = f(x) + \nabla f(x)^\top (y - x) + \frac{1}{2} (y - x)^\top M (y - x)$ for all $x, y \in \mathbb{R}^n$.

Implication: for all x and unique minimizer x_*

- $f(x) = f(x_*) + \nabla f(x_*)^\top (x - x_*) + \frac{1}{2} (x - x_*)^\top A (x - x_*)$
- $f(x) - f(x_*) = \frac{1}{2} \|x - x_*\|_A^2$ where $\|z\|_A \stackrel{\text{def}}{=} \sqrt{z^\top Az}$

Note: this is a norm whenever A *PD (*$z^\top Az > 0$ *for all* $z \neq 0$*)*
Note: semi-norm whenever A *is PSD (*$z^\top Az > 0$ *for all* $z \in \mathbb{R}^n$*)*

What is $z^\top Az$?

- *Goal:* $Ax = b \Rightarrow \min_{x \in \mathbb{R}^n} f(x)$ *for* $f(x) = \frac{1}{2} x^\top Ax - b^\top x$
- *Symmetric:* $A = A^\top$
- *PD:* $z^\top Az > 0$ *for all* $z \neq 0$

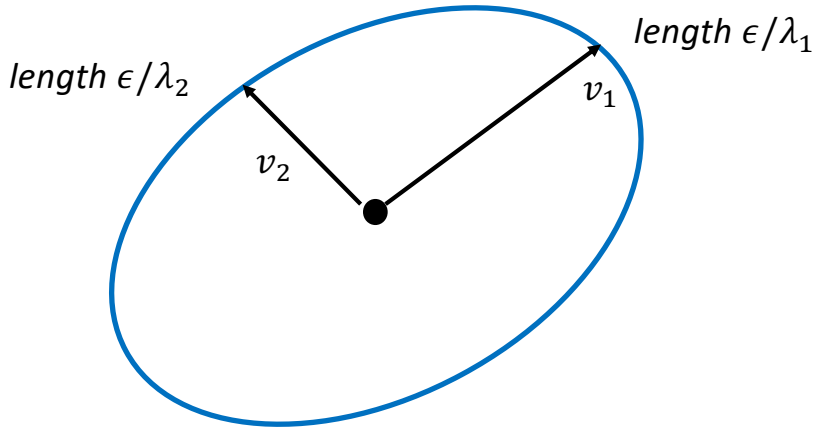
Theorem (spectral): If $A \in \mathbb{R}^{n \times n}$ is symmetric then there exists an orthonormal basis of eigenvectors, $v_1, \ldots, v_n \in \mathbb{R}^n$ with real eigenvalues.

- $v_i^\top v_j = 0$ if $i \neq j$ and $v_i^\top v_i = 1$
- $Av_i = \lambda_i v_i$ for all i with $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$

Implications

- $z = \sum_{i \in [n]} \left(v_i^\top z\right) v_i$
- $z^\top z = \sum_{i \in [n]} \left(v_i^\top z\right)^2$
- $z^\top Az = \sum_{i \in [n]} \lambda_i \left(v_i^\top z\right)^2$

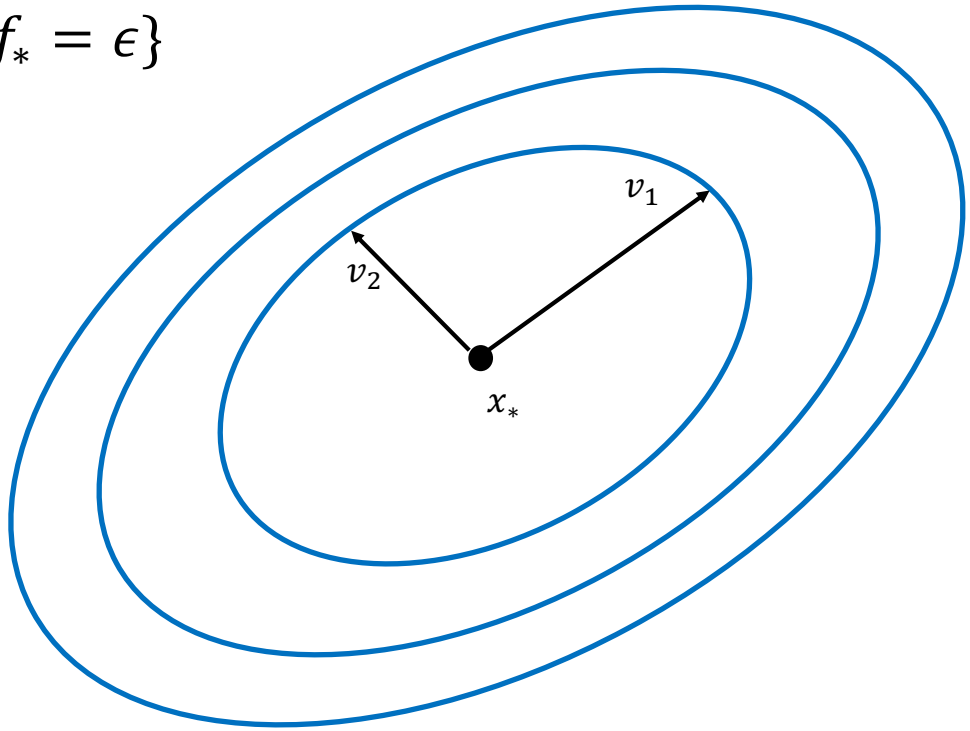
$\boldsymbol{E_A^\epsilon} \stackrel{\text{def}}{=} \{\boldsymbol{z} \mid \|\boldsymbol{z}\|_{\boldsymbol{A}} = \boldsymbol{\epsilon}\}$ *is an Ellipse*



What is $f(x)$?

- *Goal:* $Ax = b \Rightarrow \min_{x \in \mathbb{R}^n} f(x)$ *for* $f(x) = \frac{1}{2} x^\top A x - b^\top x$
- *Symmetric:* $A = A^\top$
- *PD:* $z^\top A z > 0$ *for all* $z \neq 0$

- $f(x) - f_* = \frac{1}{2} \|x - x_*\|_A^2 = \frac{1}{2} (x - x_*)^\top A (x - x_*)$
- "ϵ-level set" = $\{x \in \mathbb{R}^n \mid f(x) - f_* = \epsilon\}$
- $= \left\{x - x_* \mid \|x - x_*\|_A = \sqrt{2\epsilon}\right\}$



$$\boldsymbol{E_A^{\sqrt{2\epsilon}}} \overset{\text{def}}{=} \left\{\boldsymbol{z} \mid \|\boldsymbol{z}\|_{\boldsymbol{A}} = \sqrt{\boldsymbol{2\epsilon}}\right\}$$

Illustrative Example

- *Goal:* $Ax = b \Rightarrow \min_{x \in \mathbb{R}^n} f(x)$ *for* $f(x) = \frac{1}{2} x^\top A x - b^\top x$
- *Symmetric:* $A = A^\top$
- *PD:* $z^\top A z > 0$ *for all* $z \neq 0$

Step #2: look at structure?

- What does this look like?
- What are smoothness and strong convexity parameters?

Lemma: If $f: \mathbb{R}^n \to \mathbb{R}$ is twice differentiable and $\nabla^2 f(x) = M$ for all $x \in \mathbb{R}^n$
Then $f(y) = f(x) + \nabla f(x)^\top (y - x) + \frac{1}{2} (y - x)^\top M (y - x)$ for all $x, y \in \mathbb{R}^n$.

Implication: for all x and unique minimizer x_*

- $f(x) = f(x_*) + \nabla f(x_*)^\top (x - x_*) + \frac{1}{2} (x - x_*)^\top A (x - x_*)$
- $f(x) - f(x_*) = \frac{1}{2} \|x - x_*\|_A^2$ where $\|z\|_A \stackrel{\text{def}}{=} \sqrt{z^\top A z}$

Note: this is a norm whenever A PD ($z^\top A z > 0$ for all $z \neq 0$)
Note: semi-norm whenever A is PSD ($z^\top A z > 0$ for all $z \in \mathbb{R}^n$)

L, μ ?

- *Goal:* $Ax = b \Rightarrow \min_{x \in \mathbb{R}^n} f(x)$ *for* $f(x) = \frac{1}{2} x^\top Ax - b^\top x$
- *Symmetric:* $A = A^\top$
- *PD:* $z^\top Az > 0$ *for all* $z \neq 0$

- **Recall**: $f: \mathbb{R}^n \to \mathbb{R}$ is L-smooth and μ-strongly convex if and only if $\mu \|z\|_2^2 \leq z^\top \nabla^2 f(x) z \leq L \|z\|_2^2$ *for all* $x, z \in \mathbb{R}^n$
  - $\Leftrightarrow \mu \leq \frac{z^\top Az}{z^\top z} \leq L$ *for all* $z \neq 0$
  - $\Leftrightarrow \mu \leq \frac{\sum_{i \in [n]} \lambda_i (v_i^\top z)^2}{\sum_{i \in [n]} (v_i^\top z)^2} \leq L$ *for all* $z \neq 0$
  - $\Leftrightarrow \lambda_i \in [\mu, L]$

- $z = \sum_{i \in [n]} (v_i^\top z) v_i$
- $z^\top z = \sum_{i \in [n]} (v_i^\top z)^2$
- $z^\top Az = \sum_{i \in [n]} \lambda_i (v_i^\top z)^2$

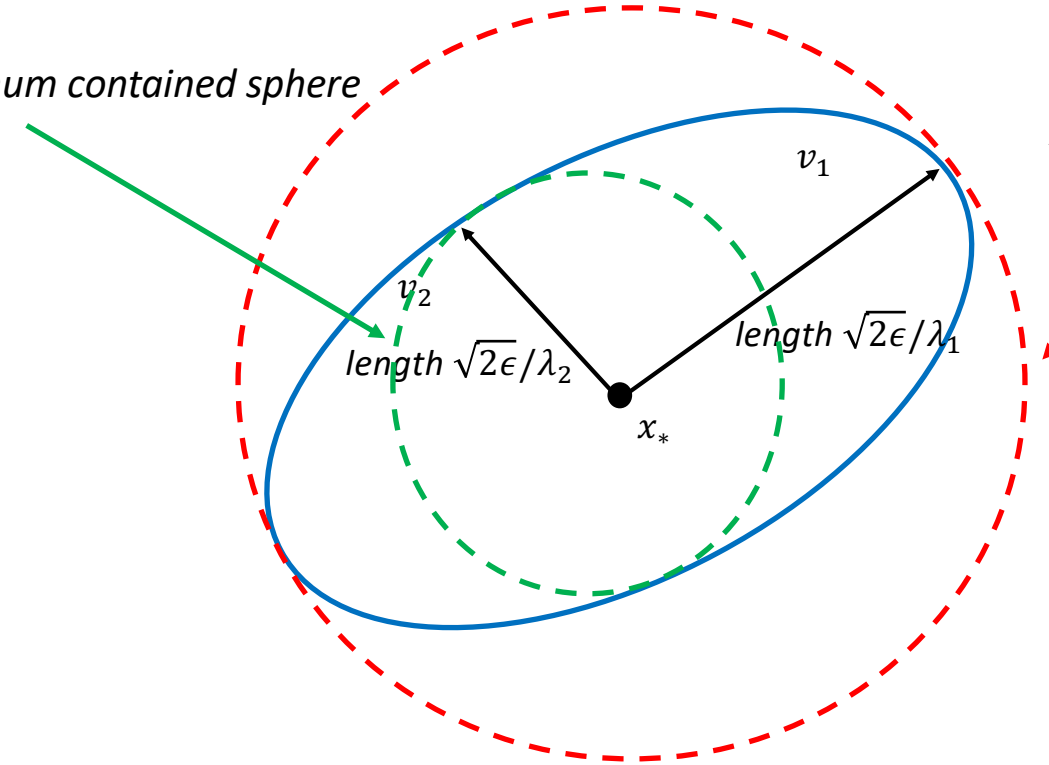
- $\Rightarrow f$ is $\lambda_{max}(A)$-smooth and $\lambda_{min}(A)$-strongly convex

L, μ ?

- *Goal:* $Ax = b \Rightarrow \min_{x \in \mathbb{R}^n} f(x)$ *for* $f(x) = \frac{1}{2} x^\top Ax - b^\top x$
- *Symmetric:* $A = A^\top$
- *PD:* $z^\top Az > 0$ *for all* $z \neq 0$

$E_{LI}^{\sqrt{2\epsilon}}$ *- maximum contained sphere*

$E_{\mu I}^{\sqrt{2\epsilon}}$ *- minimum enclosing sphere*



$$\epsilon - level\ set\ \boldsymbol{E_A^{\sqrt{2\epsilon}}} \overset{\text{def}}{=} \left\{ \boldsymbol{z} \mid \|\boldsymbol{z}\|_{\boldsymbol{A}} = \sqrt{\boldsymbol{2\epsilon}} \right\}$$

Illustrative Example

- *Goal:* $Ax = b \Rightarrow \min_{x \in \mathbb{R}^n} f(x)$ *for* $f(x) = \frac{1}{2} x^\top A x - b^\top x$
- *Symmetric:* $A = A^\top$
- *PD:* $z^\top A z > 0$ *for all* $z \neq 0$

Step #2: look at structure?

- What does this look like?
- What are smoothness and strong convexity parameters?

Lemma: If $f: \mathbb{R}^n \to \mathbb{R}$ is twice differentiable and $\nabla^2 f(x) = M$ for all $x \in \mathbb{R}^n$
Then $f(y) = f(x) + \nabla f(x)^\top (y - x) + \frac{1}{2} (y - x)^\top M (y - x)$ for all $x, y \in \mathbb{R}^n$.

Implication: for all x and unique minimizer x_*

- $f(x) = f(x_*) + \nabla f(x_*)^\top (x - x_*) + \frac{1}{2} (x - x_*)^\top A (x - x_*)$
- $f(x) - f(x_*) = \frac{1}{2} \|x - x_*\|_A^2$ where $\|z\|_A \stackrel{\text{def}}{=} \sqrt{z^\top A z}$

Note: this is a norm whenever A PD ($z^\top A z > 0$ for all $z \neq 0$)
Note: semi-norm whenever A is PSD ($z^\top A z > 0$ for all $z \in \mathbb{R}^n$)

Illustrative Example

- *Goal:* $Ax = b \Rightarrow \min_{x \in \mathbb{R}^n} f(x)$ *for* $f(x) = \frac{1}{2} x^\top A x - b^\top x$
- *Symmetric:* $A = A^\top$
- *PD:* $z^\top A z > 0$ *for all* $z \neq 0$

Step #3: iterative method

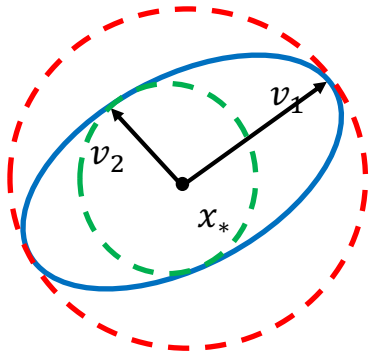
- Gradient descent
- $x_{k+1} = x_k - \frac{1}{L} \nabla f(x_k)$
- $x_{k+1} = x_k - \frac{1}{\lambda_{\max}(A)} [A x_k - b]$ *(Richardson iteration)*
- Obtain ϵ-optimal point in $O\left(\frac{L}{\mu} \log\left(\frac{f(x_0) - f_*}{\epsilon}\right)\right) = O\left(\frac{\lambda_{\max}}{\lambda_{\min}} \log\left(\frac{\|x_0 - x_*\|_A^2}{\epsilon}\right)\right)$
- Can convert between different error measures ($\|x - x_*\|_2$ and $\|Ax - b\|$ losing $\frac{\lambda_{\max}}{\lambda_{\min}}$)

For linear systems, can improve to $O(\sqrt{\kappa} \log(\epsilon_0/\epsilon))$ through Chebyshev iteration / polynomials and CG. What about general smooth strongly convex functions?

Note

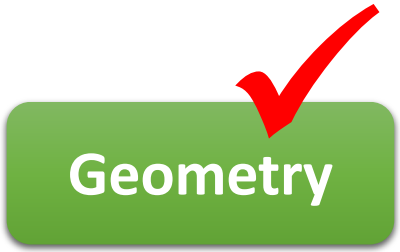
- $\kappa \stackrel{\text{def}}{=} \frac{\lambda_{\max}}{\lambda_{\min}}$ is known as "condition number"
- Is axis-ratio / ellipse ratio
- Natural iteration bound. Can we improve?

Note: for linear systems, can also obtain as Taylor expansion of inverse. Gradient descent is essentially a non-quadratic generalization.



Plan for Today

- Gradient descent for smooth function (μ-strongly) convex functions



- Linear systems, eigenvalues, and ellipsoids

- Motivation
- Method derivation

Proof and extensions

Accelerating Smooth Convex Minimization

Question

- Are $\mathrm{O}\left(\frac{L}{\mu}\log\left(\frac{\epsilon_0}{\epsilon}\right)\right)$ and $\mathrm{O}\left(\frac{LD^2}{\epsilon}\right)$ optimal?

Answer

- No! “Acceleration” is possible
- $\mathrm{O}\left(\sqrt{\frac{L}{\mu}}\log\left(\frac{\epsilon_0}{\epsilon}\right)\right)$ and $\mathrm{O}\left(\sqrt{\frac{LD^2}{\epsilon}}\right)$ are optimal (*if dimension independent*)

Why is Acceleration Possible?

One of the more mysterious theory of optimization phenomena.

Can be very useful in theory and in practice.

- **Many perspectives**
  - Momentum
  - Gradient descent and mirror descent
  - Algebra
  - Primal dual
  - Chebyshev
  - Ellipsoid
  - Continuous time
  - …

Warning

Probably the most mysterious and algebraically intensive proof in class.

2 Key Barriers

- Beyond greedy
  - Analysis use more than function value to analyze progress
- Beyond one point
  - Algorithms use more than one point as state

Some Key Ideas

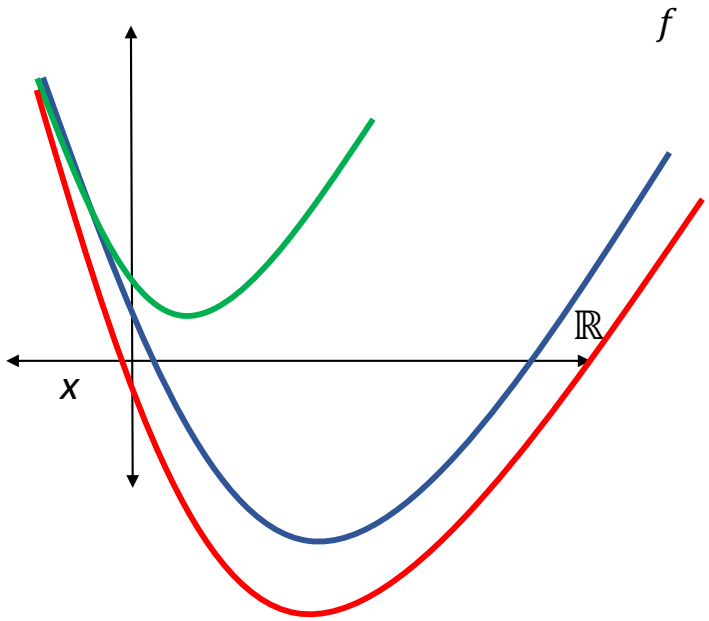
- Use both upper and lower bounds to make progress
- Maintain model (lower bound) and make progress

- *Today: develop and motivate method*
- *Tuesday: prove, study, extend*

Today: Smooth Strongly Convex Functions

Theorem: $f: \mathbb{R}^n \to \mathbb{R}$ is L-smooth and μ-strongly convex (with respect to $\|\cdot\|_2$) if and only if the following hold for all x, y

- $f(y) \leq \boldsymbol{U_x(y)} \stackrel{\text{def}}{=} f(x) + \nabla f(x)^\top (y - x) + \frac{L}{2}\|y - x\|_2^2$
- $f(y) \geq \boldsymbol{L_x(y)} \stackrel{\text{def}}{=} f(x) + \nabla f(x)^\top (y - x) + \frac{\mu}{2}\|y - x\|_2^2$



***Goal**: improve*

$O\left(\frac{L}{\mu}\log\left(\frac{f(x_0)-f_*}{\epsilon}\right)\right)$ *to* $\sim O\left(\sqrt{\frac{L}{\mu}}\log\left(\frac{f(x_0)-f_*}{\epsilon}\right)\right)$*?*

Approach

- *Are many to choose from*
- *This one is intuitive and mechanical and will let us touch on ideas of other proof*
- *Will lose a logarithmic factor and will explain how to remove*

Approach

- Maintain point ($x_k \in \mathbb{R}^n$)
- Maintain lower bound
  - $L_k: \mathbb{R}^n \to \mathbb{R}$ s.t.
  - $L_k(x) \leq f(x)$ for all $x \in \mathbb{R}^n$
- Update both in each iteration

Progress Measure

- $f(x_k) - \min_{x \in \mathbb{R}^n} L_k(x)$
- Why?
- $\geq f(x_k) - L_k(x_*) \geq f(x_k) - f_*$

Question: what lower bound?

- Idea: Quadratic!
- $L_k(x) = \psi_k + \frac{\mu}{2}\|x - v_k\|_2^2$

Why?

- Strong convexity gives quadratic lower bounds
- Linear combinations of quadratics are quadratic

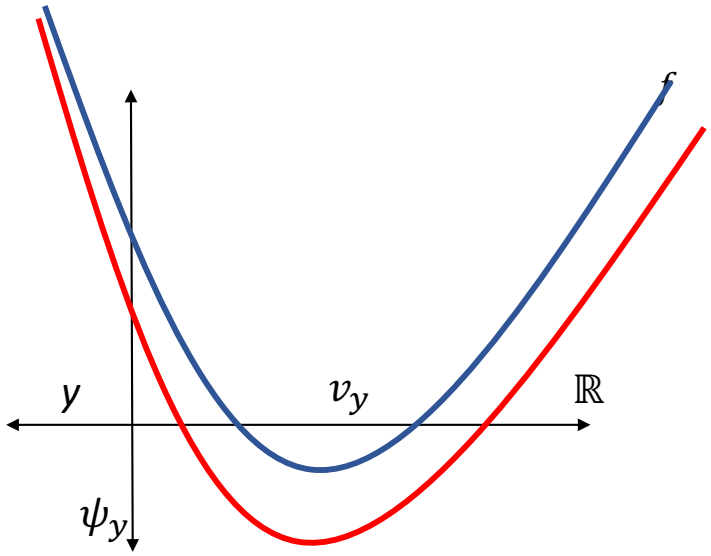
Tool 1: Quadratic Lower Bounds

Lemma $L_y(x) = \psi_y + \frac{\mu}{2}\|x - v_y\|_2^2$ for $\psi_y = f(y) - \frac{1}{2\mu}\|\nabla f(y)\|_2^2$ and $v_y = y - \frac{1}{\mu}\nabla f(y)$.

Proof

- $L_y(x) = f(y) + \nabla f(y)^\top (x - y) + \frac{\mu}{2}\|x - y\|_2^2$
- $\nabla L_y(x) = \nabla f(y) + \frac{\mu}{2}(x - y)$
- $\nabla L_y(v_y) = 0$
- $\nabla^2 L_y(x) = \mu I$
- $L_y(x) = L_y(v_y) + \nabla L_y(v_y)^\top (x - v_y) + \frac{\mu}{2}\|x - v_y\|_2^2$

ψ_y $\qquad$ 0



Tool 2: Combining Lower Bounds

Lemma If $f_1, f_2 \colon \mathbb{R}^n \to \mathbb{R}$ are defined for all $x \in \mathbb{R}^n$ by

$$f_1(x) = \psi_1 + \frac{\mu}{2}\|x - v_1\|_2^2 \quad \text{and} \quad f_2(x) = \psi_2 + \frac{\mu}{2}\|x - v_2\|_2^2$$

Then for all $\alpha \in [0,1]$ we have

$$f_\alpha(x) = \alpha \cdot f_1(x) + (1-\alpha) \cdot f_2(x) = \psi_\alpha + \frac{\mu}{2}\|x - v_\alpha\|_2^2$$

Where

- $v_\alpha = \alpha \cdot v_1 + (1-\alpha) \cdot v_2$
- $\psi_\alpha = \alpha \cdot \psi_1 + (1-\alpha) \cdot \psi_2 + \frac{\mu}{2}\alpha(1-\alpha)\|v_1 - v_2\|_2^2$

Proof: see notes

Note: if $f(x) \geq f_1(x)$ and $f(x) \geq f_2(x)$ then $f(x) \geq f_\alpha(x)$ (*can combine lower bounds*)

Tool 2: Picture

Accelerated Gradient Descent (AGD)

- Initial $x_0 \in \mathbb{R}^n$, $L_0(x) = \psi_0 + \frac{\mu}{2}\|x - v_0\|_2^2$ s.t. $f(x) \geq L_0(x)$ for all x
- Repeat for $k = 0,1,2, \ldots$
  - $y_k = \alpha \cdot x_k + (1-\alpha) \cdot v_k$
  - $L_{y_k}(x) = f(y_k) + \nabla f(y_k)^\top (x - y_k) + \frac{\mu}{2}\|x - y_k\|_2^2$
  - $L_{k+1}(x) = \psi_{k+1} + \frac{\mu}{2}\|x - v_{k+1}\|_2^2 = \beta L_k(x) + (1-\beta) L_{y_{k+1}}(x)$
  - $x_{k+1} = y_k - \frac{1}{L}\nabla f(y_k)$

$$v_{k+1} = \beta v_k + (1-\beta)\left[y_k - \frac{1}{\mu}\nabla f(y_k)\right]$$

Theorem: if $\kappa = \sqrt{L/\mu}$, $\alpha = \frac{\sqrt{\kappa}}{\sqrt{\kappa}+1}$, and $\beta = 1 - \frac{1}{\sqrt{\kappa}}$, then

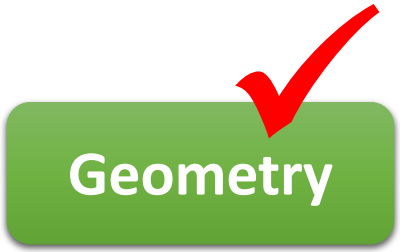
$$f(x_{k+1}) - \psi_{k+1} \leq \left(1 - \frac{1}{\sqrt{\kappa}}\right)[f(x_{k+1}) - \psi_k]$$

and $\sim\sqrt{\kappa}$ iterations suffices

Accelerated Gradient Descent (AGD)

Plan for Today

- Gradient descent for smooth function (μ-strongly) convex functions



- Linear systems, eigenvalues, and ellipsoids

- Motivation
- Method derivation

Proof and extensions