EE5239 Nonlinear Optimization

Lecture 1: Unconstrained Optimization

Ruoyu Sun

Jan. 2018

Announcement

- Will assign the first HW this Saturday
  1. Cover Lecture 0 and Lecture 1
  2. Due next next Tuesday Jan. 30 3pm
  3. Problem number according to version 2 of the book; scanned version of the homework problems will be provided.
  4. Please print the homework and bring to class

Motivation

- If you are to study optimization, what is the first thing you should study?
- You are given a real-world problem: predicting the income based on data.
  You formulate an optimization problem: linear regression $\min_x \|Ax - b\|^2$.
  Principle in modeling: the optimal solution is desired one (if $\|Ax - b\|^2$ is minimized, you do find the right combination of features to predict income)
- Then what? Understand what is "optimal solution".

Motivation

- If you are to study optimization, what is the first thing you should study?

- You are given a real-world problem: predicting the income based on data.
  You formulate an optimization problem: linear regression $\min_x \|Ax - b\|^2$.
  Principle in modeling: the optimal solution is desired one (if $\|Ax - b\|^2$ is minimized, you do find the right combination of features to predict income)

- Then what? Understand what is ”optimal solution”.

Side: What to Study in the Course

- **Level 1**: Know ”what”. Remember the definitions, formula, algorithms
  In a company: you can start working. Implement what your boss asks you to do, like predict the stock price via linear regression.

- **Level 2**: Know ”how”. How to derive the definitions, formula, algorithms?
  In a company: you can adapt. Your boss asks you to use quadratic regression; you try the algorithm and fail.
  You need to check the derivation/theorem, find the issue and modify the algorithm.

- **Level 3**: Know ”why”. Why do we define the formula/algorithm in this way?
  Assess: is there a fundamentally better way?

Outline

1. Prototype of Unconstrained Optimization

2. Optimality Conditions

Today

- Definitions
- Necessary first/second order optimality conditions
- Sufficient optimality conditions
- Existence of optimal solutions
- Quadratic minimization

Unconstrained Optimization

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in \mathbb{R}^n \end{array}$$

- **Objective function** $f : \mathbb{R}^n \to \mathbb{R}$ is a continuous function
- **Optimization variable** $x \in \mathbb{R}^n$
- **Unconstrained local minimum** x^*: $\exists \epsilon > 0$ s.t. $f(x) \geq f(x^*)$, for all $\|x - x^*\| \leq \epsilon$;
  i.e., x^* is the best in a small enough neighborhood
- **Unconstrained global minimum** $\hat{x}$: $f(x) \geq f(\hat{x})$ for all $x \in \mathbb{R}^n$
- **Graphically...**

Checkable Conditions for Local Min

- Given a point x, how to decide whether it is a local/global min (for a twice continuously differentiable function f)?
- First answer: check $f(x) \geq f(x^*)$ for all $x \in \mathbb{R}^n$
- Good enough?
- We need easily checkable conditions
  Look at the graph. What do you observe?
- **Idea.** Use Taylor expansion to analyze local behavior around x

Checkable Conditions for Local Min

- First let us understand what does being a local min represent
- **Question**: If x^* is local min, what's its property?
- The following necessary conditions

$$\nabla f(x^*) = 0, \quad \text{(first-order condition)},$$
$$\nabla^2 f(x^*) \succeq 0, \quad \text{(second-order condition)}.$$

- Make sense? Check the function $f(x) = x^2$, $f(x_1, x_2) = x_1^2 + x_2^2$
- **Remark.** We call the solutions that satisfy $\nabla f(x^*) = 0$ as stationary solutions, or stationary points

Checkable Conditions for Local Min

- Let's see the formal proof
- First, let us look at the simple one dimensional case (x is scalar)
- The conditions become

$$f'(x) = 0, \quad f''(x) \geq 0 \tag{1}$$

Checkable Conditions for Local Min

- Recall the definition of derivative $f^{'}(x^*)$
- Suppose x^* is a local min, then by definition [graphically...],

$$0 \overset{(i)}{\leq} \lim_{x^r \downarrow x^*} \frac{f(x^r) - f(x^*)}{x^r - x^*} = f'(x^*) = \lim_{x^r \uparrow x^*} \frac{f(x^r) - f(x^*)}{x^r - x^*} \overset{(ii)}{\leq} 0 \tag{2}$$

$$0 \overset{(iii)}{\leq} \lim_{x^r \to x^*} \frac{f(x^r) - f(x^*) - f'(x^*)(x^r - x^*)}{(x^r - x^*)^2} = f''(x^*) \tag{3}$$

- (i): $x^r \geq x^*$, $f(x^r) \geq f(x^*)$ when getting sufficiently close
- (ii): $x^r \leq x^*$, $f(x^r) \geq f(x^*)$ when getting sufficiently close
- Then (2) says $f'(x^*) = 0$
- (iii): $f'(x^*) = 0$, and $f(x^r) \geq f(x^*)$ when $x^r \to x^*$

Checkable Conditions for Local Min

- Recall the definition of derivative $f^{'}(x^*)$
- Suppose x^* is a local min, then by definition [graphically...],

$$0 \overset{(i)}{\leq} \lim_{x^r \downarrow x^*} \frac{f(x^r) - f(x^*)}{x^r - x^*} = f'(x^*) = \lim_{x^r \uparrow x^*} \frac{f(x^r) - f(x^*)}{x^r - x^*} \overset{(ii)}{\leq} 0 \tag{2}$$

$$0 \overset{(iii)}{\leq} \lim_{x^r \to x^*} \frac{f(x^r) - f(x^*) - f'(x^*)(x^r - x^*)}{(x^r - x^*)^2} = f''(x^*) \tag{3}$$

- (i): $x^r \geq x^*$, $f(x^r) \geq f(x^*)$ when getting sufficiently close
- (ii): $x^r \leq x^*$, $f(x^r) \geq f(x^*)$ when getting sufficiently close
- Then (2) says $f'(x^*) = 0$
- (iii): $f'(x^*) = 0$, and $f(x^r) \geq f(x^*)$ when $x^r \to x^*$

Checkable Conditions for Local Min

- Recall the definition of derivative $f^{'}(x^*)$
- Suppose x^* is a local min, then by definition [graphically...],

$$0 \overset{(i)}{\leq} \lim_{x^r \downarrow x^*} \frac{f(x^r) - f(x^*)}{x^r - x^*} = f'(x^*) = \lim_{x^r \uparrow x^*} \frac{f(x^r) - f(x^*)}{x^r - x^*} \overset{(ii)}{\leq} 0 \tag{2}$$

$$0 \overset{(iii)}{\leq} \lim_{x^r \to x^*} \frac{f(x^r) - f(x^*) - f'(x^*)(x^r - x^*)}{(x^r - x^*)^2} = f''(x^*) \tag{3}$$

- (i): $x^r \geq x^*$, $f(x^r) \geq f(x^*)$ when getting sufficiently close
- (ii): $x^r \leq x^*$, $f(x^r) \geq f(x^*)$ when getting sufficiently close
- Then (2) says $f'(x^*) = 0$
- (iii): $f'(x^*) = 0$, and $f(x^r) \geq f(x^*)$ when $x^r \to x^*$

Checkable Conditions for Local Min

- Recall the definition of derivative $f^{'}(x^*)$
- Suppose x^* is a local min, then by definition [graphically...],

$$0 \overset{(i)}{\leq} \lim_{x^r \downarrow x^*} \frac{f(x^r) - f(x^*)}{x^r - x^*} = f'(x^*) = \lim_{x^r \uparrow x^*} \frac{f(x^r) - f(x^*)}{x^r - x^*} \overset{(ii)}{\leq} 0 \tag{2}$$

$$0 \overset{(iii)}{\leq} \lim_{x^r \to x^*} \frac{f(x^r) - f(x^*) - f'(x^*)(x^r - x^*)}{(x^r - x^*)^2} = f''(x^*) \tag{3}$$

- (i): $x^r \geq x^*$, $f(x^r) \geq f(x^*)$ when getting sufficiently close
- (ii): $x^r \leq x^*$, $f(x^r) \geq f(x^*)$ when getting sufficiently close
- Then (2) says $f'(x^*) = 0$
- (iii): $f'(x^*) = 0$, and $f(x^r) \geq f(x^*)$ when $x^r \to x^*$

Checkable Conditions for Local Min

- Recall the definition of derivative $f^{'}(x^*)$
- Suppose x^* is a local min, then by definition [graphically...],

$$0 \overset{(i)}{\leq} \lim_{x^r \downarrow x^*} \frac{f(x^r) - f(x^*)}{x^r - x^*} = f'(x^*) = \lim_{x^r \uparrow x^*} \frac{f(x^r) - f(x^*)}{x^r - x^*} \overset{(ii)}{\leq} 0 \tag{2}$$

$$0 \overset{(iii)}{\leq} \lim_{x^r \to x^*} \frac{f(x^r) - f(x^*) - f'(x^*)(x^r - x^*)}{(x^r - x^*)^2} = f''(x^*) \tag{3}$$

- (i): $x^r \geq x^*$, $f(x^r) \geq f(x^*)$ when getting sufficiently close
- (ii): $x^r \leq x^*$, $f(x^r) \geq f(x^*)$ when getting sufficiently close
- Then (2) says $f'(x^*) = 0$
- (iii): $f'(x^*) = 0$, and $f(x^r) \geq f(x^*)$ when $x^r \to x^*$

Checkable Conditions for Local Min

- Recall the definition of derivative $f^{'}(x^*)$
- Suppose x^* is a local min, then by definition [graphically...],

$$0 \overset{(i)}{\leq} \lim_{x^r \downarrow x^*} \frac{f(x^r) - f(x^*)}{x^r - x^*} = f'(x^*) = \lim_{x^r \uparrow x^*} \frac{f(x^r) - f(x^*)}{x^r - x^*} \overset{(ii)}{\leq} 0 \tag{2}$$

$$0 \overset{(iii)}{\leq} \lim_{x^r \to x^*} \frac{f(x^r) - f(x^*) - f'(x^*)(x^r - x^*)}{(x^r - x^*)^2} = f''(x^*) \tag{3}$$

- (i): $x^r \geq x^*$, $f(x^r) \geq f(x^*)$ when getting sufficiently close
- (ii): $x^r \leq x^*$, $f(x^r) \geq f(x^*)$ when getting sufficiently close
- Then (2) says $f'(x^*) = 0$
- (iii): $f'(x^*) = 0$, and $f(x^r) \geq f(x^*)$ when $x^r \to x^*$

Checkable Conditions for Local Min

- For higher dimensions, derivation is similar (Prop 1.1.1)
- Proof sketch: Consider the one dimensional function $g(\alpha) = f(\mathbf{x}^* + \alpha\mathbf{d})$, where $\mathbf{d} \in \mathbb{R}^n$ is a direction.
- This function of α has a local minimizer $\alpha = 0$ (why?)
- Apply the previous theorem, we have

$$g'(0) = \langle \nabla f(\mathbf{x}^*), \mathbf{d} \rangle = 0, \; g''(0) = \langle \mathbf{d}, \nabla^2 f(\mathbf{x}^*)\mathbf{d} \rangle \geq 0 \tag{4}$$

- Note $\mathbf{d}$ is an arbitrary direction, the first equation means $\nabla f(\mathbf{x}^*) = 0$, the second equation means

$$\nabla^2 f(\mathbf{x}^*) \succeq 0 \tag{5}$$

- For detailed proof, please read Section 1.1 of the text book

Checkable Conditions for Local Min

- **Exercise**: Consider the following functions

$$f(x) = |x|^3, x^3, -|x|^3, x^2, -x^2$$

  Check the necessary condition at $x = 0$; plot f

- What have we done so far?

- For a given solution x^*, I can check whether it is local minimum

- **Question**: Can we have some simple conditions which once they are satisfied by an x^*, then it is local optimum?

- **Question**: Is the necessary condition also sufficient?

Week 2 Tues. Lecture

- Announcement: homework 1 available
- Seats of the course limited. Check homework 1; if too hard for you, you may consider drop the course before Thursday
- How many have started working on homework problems?
- Review of last week:
  Who should take the course?
  Math preparation: SVD, derivative/Hessian, contraction mapping
  Optimality condition: 1st and 2nd order conditions; local-min
- Today: convex/quadratic minimization; gradient descent

Week 2 Tues. Lecture

- Announcement: homework 1 available
- Seats of the course limited. Check homework 1; if too hard for you, you may consider drop the course before Thursday
- How many have started working on homework problems?
- Review of last week:
  Who should take the course?
  Math preparation: SVD, derivative/Hessian, contraction mapping
  Optimality condition: 1st and 2nd order conditions; local-min
- Today: convex/quadratic minimization; gradient descent

Week 2 Tues. Lecture

- Announcement: homework 1 available
- Seats of the course limited. Check homework 1; if too hard for you, you may consider drop the course before Thursday
- How many have started working on homework problems?
- Review of last week:
  Who should take the course?
  Math preparation: SVD, derivative/Hessian, contraction mapping
  Optimality condition: 1st and 2nd order conditions; local-min
- Today: convex/quadratic minimization; gradient descent

Sufficient Conditions for Local Min

- We have the following sufficient conditions [makes sense?]

$\succeq 0$ necessary

$\succ 0$ sufficient

$$\nabla f(x^*) = 0, \quad \text{(first-order condition)},$$
$$\nabla^2 f(x^*) \succ 0, \quad \text{(second-order condition)}.$$

- Together they are "sufficient" for local min
- Why? Check simple scalar case. $f'(x^*) = 0, f''(x^*) > 0$
- Taylor expansion:

$$f(x) - f(x^*) =$$

Sufficient Conditions for Local Min

- We have the following sufficient conditions [makes sense?]

$$\nabla f(x^*) = 0, \quad \text{(first-order condition)},$$
$$\nabla^2 f(x^*) \succ 0, \quad \text{(second-order condition)}.$$

- Together they are "sufficient" for local min
- Why? Check simple scalar case. $f'(x^*) = 0, f''(x^*) > 0$
- Taylor expansion:

$$f(x) - f(x^*) = \underbrace{f'(x^*)}_{=0}(x - x^*) + \frac{1}{2}\underbrace{\overset{=\beta}{f''(x^*)}}_{>0}(x - x^*)^2 + o(\|x - x^*\|^2)$$

when $|x - x^*|$ is small, $\underset{>0}{\beta}(x - x^*)^2 \gg o(\|x - x^*\|^2)$

$\Rightarrow f(x) - f(x^*) > 0$, when $|x - x^*|$ small.

Review: Derive Optimality Conditions in 3 mins

- Write the Taylor expansion:

$$\begin{aligned}\underline{0 \le f(x+\delta) - f(x^*)} &= \cancel{\langle \nabla f(x^*), \delta \rangle} + \frac{1}{2}\delta^T \nabla^2 f(x^*)\delta + o(\|\delta\|^3), \\ &\approx \frac{1}{2}\delta^T \nabla^2 f(x^*)\delta + \text{small term.}\end{aligned}$$

- Necessary condition: $\frac{1}{2}\delta^T \nabla^2 f(x^*)\delta \underline{\ge} 0, \forall \delta$, or PSD Hessian.
- Sufficient conditoin: $\frac{1}{2}\delta^T \nabla^2 f(x^*)\delta \underline{>} 0, \forall \delta$, or PD Hessian.

Why Optimality Conditions?

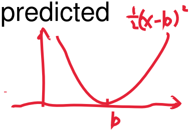
- Optimality conditions are useful because:
  1. provide guarantees for a candidate solution to be optimal (sufficient condition)
  2. indicate when a point is NOT optimal (necessary condition)
- Guide the design of algorithm
  1. Algorithms should look for points achieving the optimality conditions
  2. Algorithm should stop when the optimality condition is approximately satisfied

Use of Optimality Condition

$f(x) =$

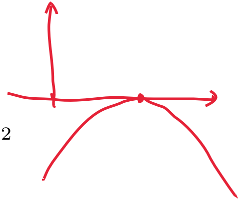
- Let's try to solve the simple problem $\min \ \frac{1}{2}(x-b)^2$ $= 0$
- Clearly the solution is $x = b$; Is this the same solution predicted by the optimality condition?

  $f'(x) = x - b = 0 \Rightarrow x = b;$ $\quad f''(x) = 1 > 0.$



- How about $\min \ -\frac{1}{2}(x-b)^2$? Is the first-order optimality condition sufficient?

  $f'(x) = 0 \Rightarrow x = b;$ $\quad f''(x) = -1 < 0.$



- How about the following problems?

$$\min \ \frac{1}{2}(\mathbf{a}^T\mathbf{x} - b)^2, \quad \min \ \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2$$

Use of Optimality Condition (cont.)

$$\begin{array}{ll} \text{minimize} & f(\mathbf{y}) = e^{y_1} + e^{y_2} + e^{y_3} \\ \text{subject to} & y_1 + y_2 + y_3 = s. \end{array}$$ → constraint

- Eliminate y_3 by substituting $y_3 = s - y_1 - y_2$.

$$g(\mathbf{y}) = e^{y_1} + e^{y_2} + e^{s-y_1-y_2}$$, over y_1, y_2

- First order optimality condition says

$$\frac{\partial g}{\partial y_i} = e^{y_i} - e^{s-y_1-y_2} = 0, \ i = 1, 2.$$

which implies $y_i = s - y_1 - y_2 \Rightarrow y_1 = y_2 = \frac{1}{3}s$.

- The minimum objective value will be $f^* = 3e^{s/3}$
- Plug in the expression for s, we obtain the well-known arithmetic-geometric inequality

$$f(y) = e^{y_1} + e^{y_2} + e^{y_3} \geq 3e^{(y_1+\cdots+y_3)/3}, \ \forall y_1, y_2, y_3 \in \mathbb{R}$$

$$(a_1 + a_2 + a_3)/3 \geq \sqrt[3]{a_1 a_2 a_3}, \ \forall a_i \geq 0.$$

Use of Optimality Condition (cont.)

- How to find minimal solution?
- **Algorithm 1**, the first “algorithm” in this class:

  Step 1: Find all stationary points (candidates) by solving $\nabla f(\mathbf{x}) = 0$;

  >0?

  Step 2 (optional): Find all candidates s.t. $\nabla^2 f(\mathbf{x}) \succeq 0$.

  Step 3: Among all candidates, find the one with minimal value.

- Two issues:
  (i) Solving $\nabla f(x) = 0$ is often as hard as minimizing $f(x)$.

  (ii) The above procedure is WRONG!

Use of Optimality Condition (cont.)

- How to find minimal solution?
- **Algorithm 1**, the first “algorithm” in this class:

  Step 1: Find all stationary points (candidates) by solving $\nabla f(\mathbf{x}) = 0$;

  Step 2 (optional): Find all candidates s.t. $\nabla^2 f(\mathbf{x}) \succeq 0$.

  Step 3: Among all candidates, find the one with minimal value.
- Two issues:

  (i) Solving $\nabla f(x) = 0$ is often as hard as minimizing $f(x)$.

  (ii) The above procedure is WRONG!

Use of Optimality Condition (cont.)

- How to find minimal solution?
- **Algorithm 1**, the first “algorithm” in this class:

  Step 1: Find all stationary points (candidates) by solving $\nabla f(\mathbf{x}) = 0$;

  Step 2 (optional): Find all candidates s.t. $\nabla^2 f(\mathbf{x}) \succeq 0$.

  Step 3: Among all candidates, find the one with minimal value.

- Two issues: $f(x) = (x - \sin x + \cos(x^2))^2$, $\nabla f(x) = 0$.

  (i) Solving $\nabla f(x) = 0$ is often as hard as minimizing $f(x)$.

  (ii) The above procedure is WRONG!

Use of Optimality Condition (cont.)

- How to find minimal solution?
- **Algorithm 1**, the first "algorithm" in this class:

  Step 1: Find all stationary points (candidates) by solving $\nabla f(\mathbf{x}) = 0$;

  Step 2 (optional): Find all candidates s.t. $\nabla^2 f(\mathbf{x}) \succeq 0$.

  Step 3: Among all candidates, find the one with minimal value.

- Two issues:
  (i) Solving $\nabla f(x) = 0$ is often as hard as minimizing $f(x)$.

  (ii) The above procedure is WRONG!

Failure of Optimality Condition

- Example 1: $f(x) = x^3$.
- First order optimality condition: $f'(x) = 3x^2 = 0 \Rightarrow x = 0$.
- Second order optimality condition: $f''(x) = 6x = 0 \big|_{\text{at } x=0}$.

  Is $x = 0$ a local-min? Not sure.

- Since $x = 0$ is the only “candidate”, it should be global-min? Graph.

- Lessons:
  (i) $\nabla^2 f(x) \succeq 0$ is necessary, but NOT sufficient condition.

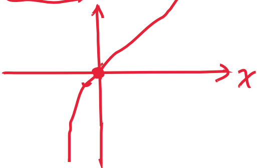
  (ii) Before applying Algorithm 1, need to check ______________

Failure of Optimality Condition

- Example 1: $f(x) = x^3$.
- First order optimality condition:
- Second order optimality condition:

  Is $x = 0$ a local-min? Not sure.

- Since $x = 0$ is the only "candidate", it should be global-min? Graph.



- Lessons:
  (i) $\nabla^2 f(x) \succeq 0$ is necessary, but NOT sufficient condition.

  (ii) Before applying Algorithm 1, need to check ______ existence of global-min

More Examples

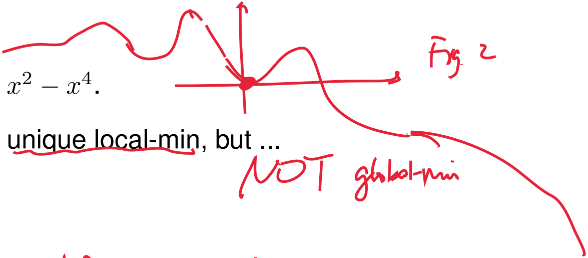


Fig 2

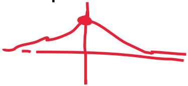
- Example 2: $f(x) = x^2 - x^4$.

  $x = 0$ is indeed the unique local-min, but ...

  NOT global-min

- Example 3:

Fig 3



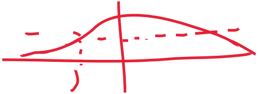
$$\underset{x\in\mathbb{R}}{\overset{\inf}{\cancel{\min}}} \exp(-x^2) = ? \qquad (6)$$

$f'(x) = e^{-x^2} \cdot (-2x) = 0 \Rightarrow x = 0.$

- Unlike Example 1 and Example 2, the function does not go to $-\infty$.

  Is the minimum attained? NOT attainable at $\mathbb{R}$.

- Calculate the gradient:



level set of $\{e^{-x^2} \leq e^{-1}\} \Leftrightarrow \{x^2 \geq 1\}.$

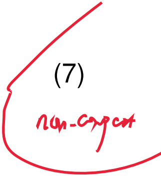
$= (-\infty, -1) \cup [1, +\infty)$

Existence of Optimal Solution

- **Bolzano-Weierstrass Theorem** Every continuous function f attains its infimum over a compact set X. That is, there exists an $x^* \in X$ such that

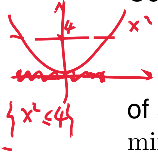
close+bounded

$$f(x^*) = \inf_{x \in X} f(x) \tag{7}$$



What if domain X in $\mathbb{R}^n$?

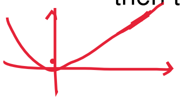
- Consequently, if the level set (for some x^0)



$$\{x \mid f(x) \le f(x^0)\} \tag{8}$$

of a continuous function f is compact, then the global min of $\min\ f(x), \quad \text{subject to } x \in \mathbb{R}^n$ is attained

- **Coercive**: If f is continuous, X is closed, $f(x) \to \infty$ as $\|x\| \to \infty$, then there exists an optimal solution. (not compact, possibly. e.g. $\mathbb{R}^n$)



Unconstrained Quadratic Optimization

$\min f(x) = \frac{1}{2}x^2$

$$\begin{aligned} &\text{minimize} && f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} - \mathbf{b}^T\mathbf{x} \\ &\text{subject to} && \mathbf{x} \in \mathbb{R}^n, \end{aligned}$$

where $\mathbf{Q}$ is a symmetric $n \times n$ matrix. (what if non-symmetric?)

$x^T A x = \frac{1}{2} x^T \underbrace{(A + A^T)}_{Q} x$

- Necessary condition for (local) optimality

$$\underbrace{\mathbf{Q}\mathbf{x} = \mathbf{b}}_{\text{1st order}}, \quad \mathbf{Q} \succeq 0 \tag{9}$$

- **Case 1**: $\mathbf{Q}\mathbf{x} = \mathbf{b}$ has no solution, i.e. $\mathbf{b} \notin R(\mathbf{Q})$. e.g.
  No stationary point, can achieve $-\infty$ (how?).

$Q = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \; b = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

$f(x) = \frac{1}{2}x_1^2 - x_2$

$f(0, +\infty) = 0 - \infty$

- **Case 2**: $\mathbf{Q}$ is not semidefinite.
  No local-min. Can achieve $-\infty$ (how?).
- **Case 3**: $\mathbf{Q} \succeq 0$ and $\mathbf{b} \in R(\mathbf{Q})$.
- **Claim**: Any stationary point is a global optimal solution.

Two examples in machine learning (about $\exists$ of global-min):

1) Gaussian process / Kernel method.

2) Logistic regression & DL.

Unconstrained Quadratic Optimization

$$\begin{array}{ll} \text{minimize} & f(\mathbf{y}) = \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} - \mathbf{b}^T\mathbf{x} \\ \text{subject to} & \mathbf{x} \in \mathbb{R}^n, \end{array}$$

where $\mathbf{Q}$ is a symmetric $n \times n$ matrix. (what if non-symmetric?)

- Necessary condition for (local) optimality

$$\mathbf{Q}\mathbf{x} = \mathbf{b}, \quad \mathbf{Q} \succeq 0 \tag{9}$$

- **Case 1**: $\mathbf{Q}\mathbf{x} = \mathbf{b}$ has no solution, i.e. $\mathbf{b} \notin R(\mathbf{Q})$.
  No stationary point, can achieve $-\infty$ (how?).
- **Case 2**: $\mathbf{Q}$ is not positive semidefinite.
  No local-min. Can achieve $-\infty$ (how?).

$$Q = \begin{pmatrix} 1 & 10 \\ 10 & 1 \end{pmatrix},$$

- **Case 3**: $\mathbf{Q} \succeq 0$ and $\mathbf{b} \in R(\mathbf{Q})$.

Claim: Any stationary point is a global optimal solution.

Unconstrained Quadratic Optimization

$$\begin{array}{ll} \text{minimize} & f(\mathbf{y}) = \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} - \mathbf{b}^T\mathbf{x} \\ \text{subject to} & \mathbf{x} \in \mathbb{R}^n, \end{array}$$

$3x^2 - 2x$

$-3x^2 + 2x$

$\neq (ax+b)^2$

where $\mathbf{Q}$ is a symmetric $n \times n$ matrix. (what if non-symmetric?)

- Necessary condition for (local) optimality

$$\mathbf{Q}\mathbf{x} = \mathbf{b}, \quad \mathbf{Q} \succeq 0 \tag{9}$$

- **Case 1**: $\mathbf{Q}\mathbf{x} = \mathbf{b}$ has no solution, i.e. $\mathbf{b} \notin R(\mathbf{Q})$.
  No stationary point, can achieve $-\infty$ (how?).
- **Case 2**: $\mathbf{Q}$ is not semidefinite.
  No local-min. Can achieve $-\infty$ (how?).
- **Case 3**: $\mathbf{Q} \succeq 0$ and $\mathbf{b} \in R(\mathbf{Q})$.

(even though 3rd order $= 0$, 4th order > 0. .)

Claim: Any stationary point is a global optimal solution.

($Qx = b \Rightarrow$ global-min)

Linear Least Squares

(regression)

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) = \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{c}\|^2 \\ \text{subject to} \quad & \mathbf{x} \in \mathbb{R}^d, \end{aligned}$$

$(2x-1)^2$

where $\mathbf{A} \in \mathbb{R}^{n \times d}$, $\mathbf{x} \in \mathbb{R}^d$

- n number of data points, d number of features
- $\mathbf{A}$ may be fat (under-determined), tall (over-determined), or rank-deficient
- Note that comparing with the previous case, $\mathbf{Q} = \mathbf{A}^T\mathbf{A}$, $\mathbf{b} = \mathbf{A}^T\mathbf{c}$
- Necessary and sufficient optimality condition

$$\mathbf{A}^T\mathbf{A}\mathbf{x}^* - \mathbf{A}^T\mathbf{c} = 0 \Rightarrow x^* = (A^TA)^{\dagger}A^Tc.$$

which always has a solution (why?)

- Review: Similarity and difference between quadratic minimization and least squares?

Convexity and Optimal Conditions

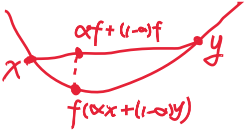
- Sufficient condition for global optimality? Difficult to find.
- Most important conditions:

Convexity + first order condition $\Rightarrow$ global optimal.

- **Convex set** C: $x, y \in C \Rightarrow \alpha x + (1-\alpha) y \in C, \ \forall \alpha \in [0,1]$

- **Convex function**: f is convex in C iff $f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y)$, $\forall x, y \in C$ and $\forall \alpha \in [0,1]$.

  Strictly convex: when $\leq$ becomes $<$



- **Property**: If f is twice differentiable, then f is convex (strictly convex) iff $\nabla^2 f(x) \succeq 0 (\succ 0), \forall x$.

Convexity and Optimal Conditions

- **Proposition** (Prop. 1.1.2 of textbook): Let $f : X \longmapsto \mathbb{R}$ be a convex function over the convex set X.

(a) A local-min of f over X is also a global-min over X. local-mi $\Longleftrightarrow$ global-mi

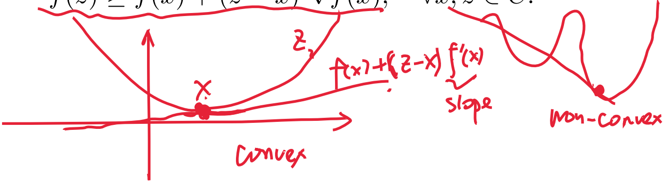
(b) If X is open (e.g. $\mathbb{R}^n$), then $\nabla f(x^*) = 0$ is a necessary and sufficient condition for the global optimality. $\nabla f = 0 \Longleftrightarrow$ global-mi

- Proof based on a property (Prop. B.3): If f is differentiable over C, then f is convex iff

$\nabla f(x^*) = 0 \Rightarrow f(z) \geq f(x^*) + 0, \ \forall z \Rightarrow x^*$ global-mi...

$$f(z) \geq f(x) + (z - x)'\nabla f(x), \quad \forall x, z \in C.$$

Graph:



High-level Understanding

- What is the proposition about?

  A sufficient condition for global optimality. Can you say more?

- Why global-min conditions hard? "Global"!

  What is the nature of derivatives/Hessian/etc.? Local.

- **Mathematical view**: an (everywhere) local property leads to a global property: every local-min is global.
  Graph:

- How to obtain every local-min is global without convexity?
  Open question. One example in homework problem 6.

High-level Understanding

- What is the proposition about?

  A sufficient condition for global optimality. Can you say more?

- Why global-min conditions hard? "Global"!

  What is the nature of derivatives/Hessian/etc.? Local.

- **Mathematical view**: an (everywhere) local property leads to a global property: every local-min is global.
  Graph:

- How to obtain every local-min is global without convexity?
  Open question. One example in homework problem 6.