# COMP7180: Quantitative Methods for Data Analytics and Artificial Intelligence

## Lecture 6: Convex Optimization: Theory I

Jun Qi

Research Assistant Professor in Computer Science @ Hong Kong Baptist University
Affiliated Associate Professor in Electronic Engineering @ Fudan University

香港浸會大學
HONG KONG BAPTIST UNIVERSITY

DEPARTMENT OF
COMPUTER SCIENCE
計算機科學系

22/10/2024

# Self-Introduction (Jun QI)

- Ph.D. in ECE @ Georgia Institute of Technology, Atlanta
  - Speech Signal Processing and Language Processing
  - Quantum Tensor Network in Machine Learning

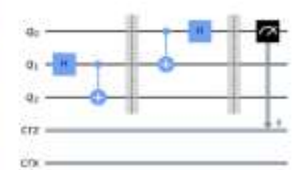
- Research Assistant Professor in CS @ HKBU, and affiliated Associate Professor in EE @ Fudan University, Shanghai
  - Quantum Machine Learning Theory and Algorithms
  - Low-rank Tensor Network Optimization for Machine Learning Systems



AI & Machine Learning

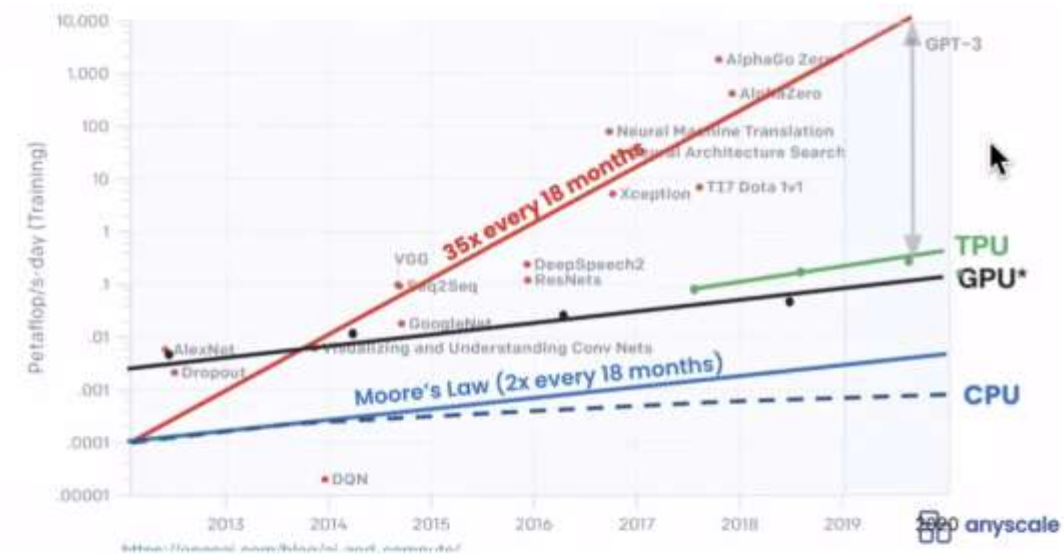
PARK Hajun

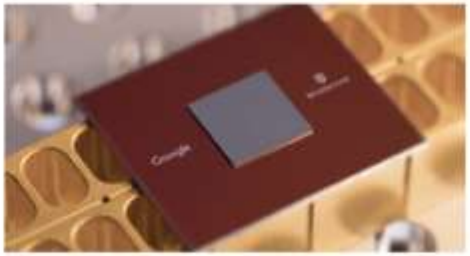
Yusuf Dikeç



Quantum Computing

# Quantum Machine Learning (1/3)

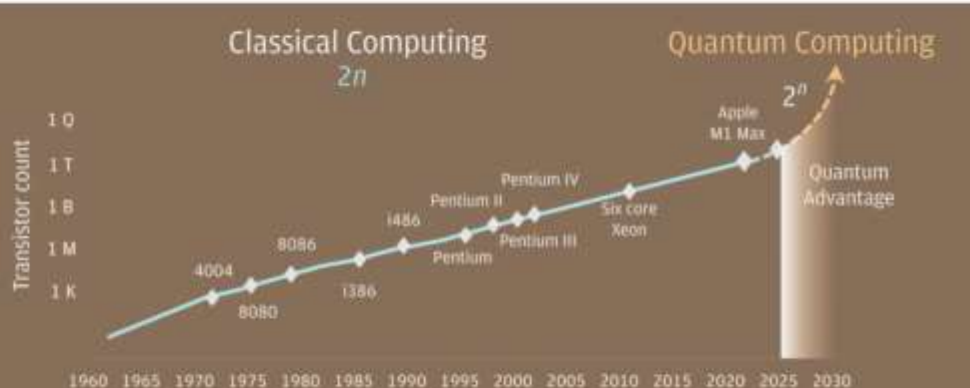



**Exponential growth** in **computing power.**

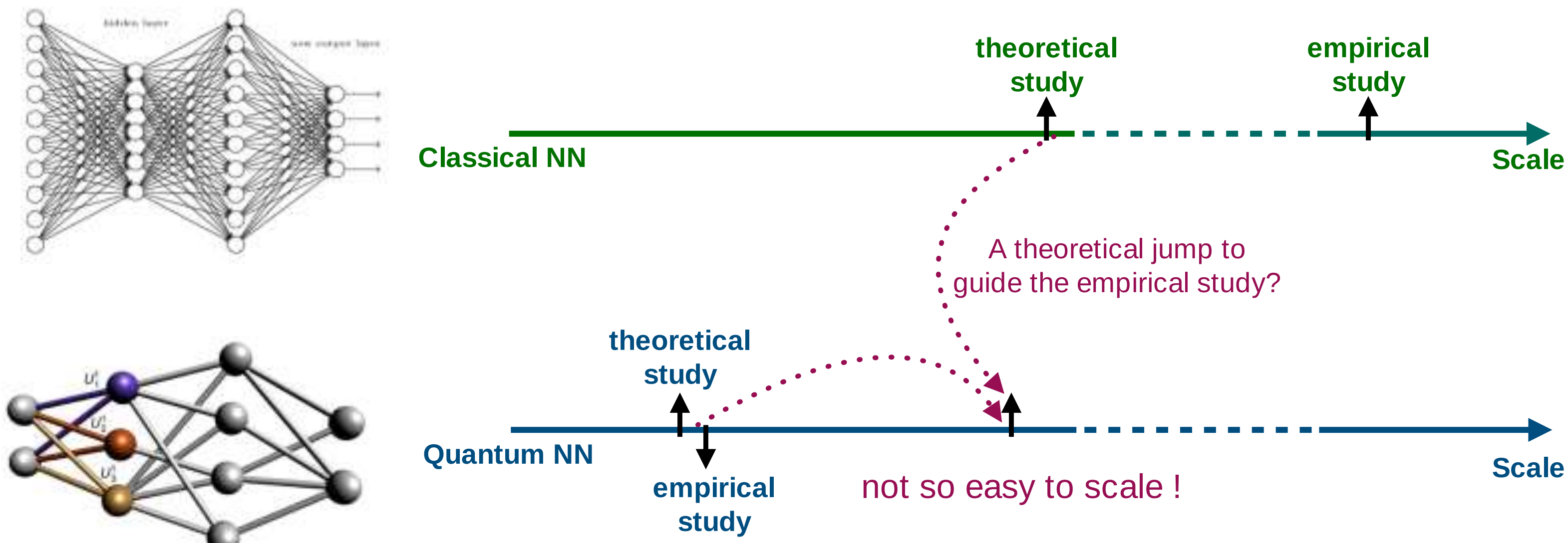
With classical computing, "computational power **doubles** every 2 years"

*- Gordon Moore, Intel Co-Founder & CEO*



# Quantum Machine Learning (2/3)

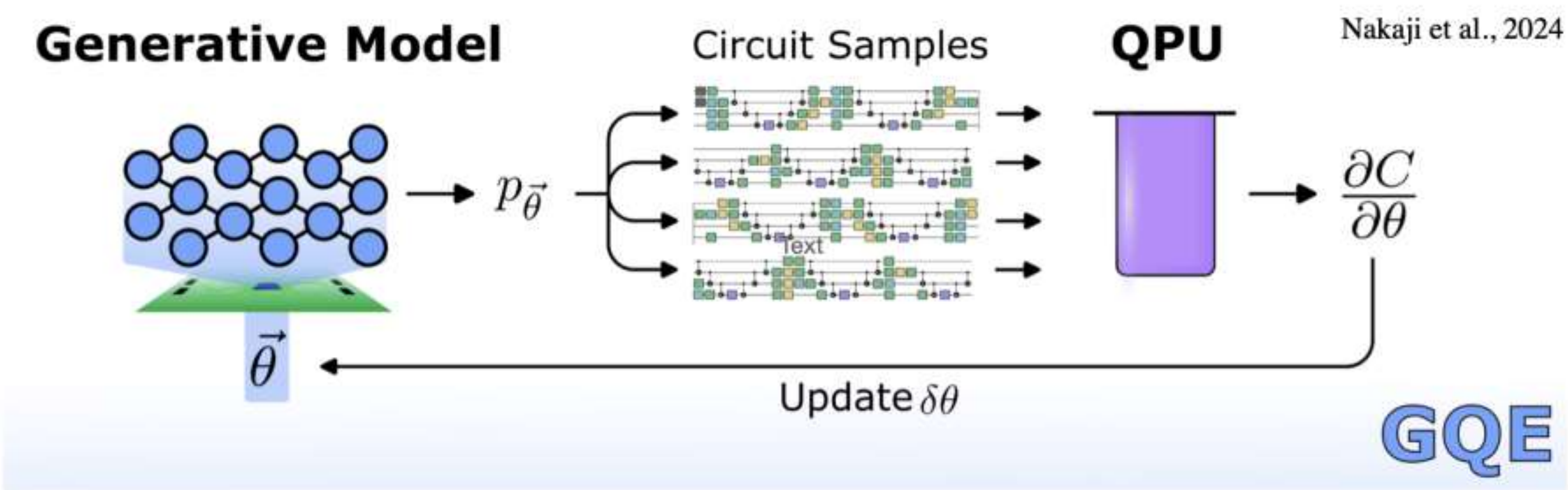
- We leverage machine learning theory to analyze the trainability, expressiveness, and generalization power of quantum machine learning.




- Jun Qi, et al., "Theoretical Error Performance Analysis for Variational Quantum Circuit Based Functional Regression," *npj Quantum Information*, Vol. 9, no. 4, 2023

# Quantum Machine Learning (3/3)

- We leverage machine learning algorithms for quantum circuit architecture search, scaling up the quantum computing for machine learning.



# CVX Course Outline

- Convex Optimization Theory I
  - **Sets and functions** (22 October 2024)
  - Optimization basics (29 October 2024)
- Convex Optimization Algorithms I and Theory II
  - First-order methods (29 October 2024)
  - Duality and optimality (5 November 2024)
- Convex Optimization Algorithms II
  - Second-order methods (5 November 2024)
  - Advanced topics (Optional)

# Optimization in Machine Learning and Statistics

Optimization problems underlie nearly everything we do in Machine Learning and Statistics. In this course, you learn how to:

translate  into $P \; : \; \min_{x \in D} \; f(x)$

*Conceptual idea* *Optimization problem*

Examples of this? Examples of the contrary?

This course: how to solve $P$, and why this is a good skill to have

# Motivation: why do we bother?

Presumably, other people have already figured out how to solve

$$P: \min_{\boldsymbol{x} \in D} f(\boldsymbol{x})$$

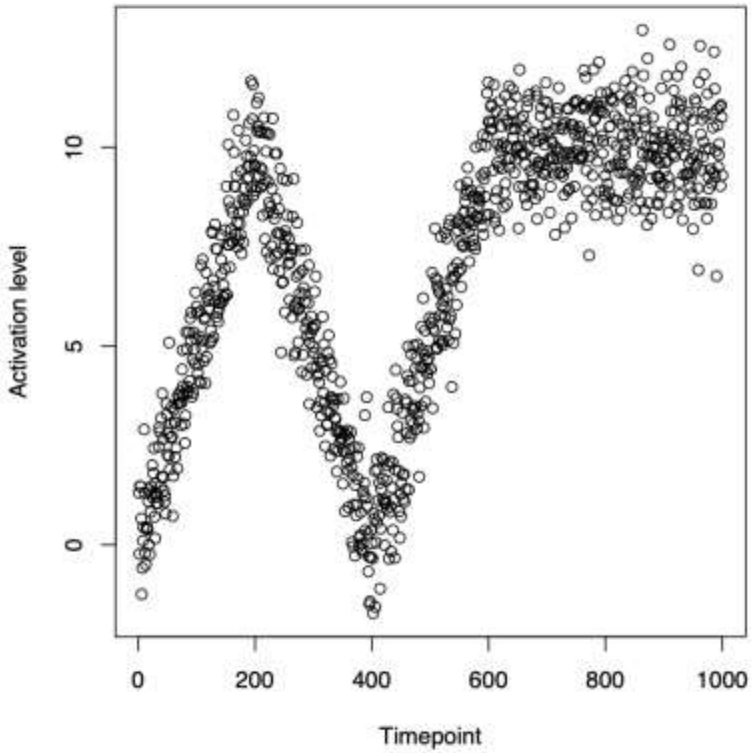
So why bother? Many reasons. Here's three:

- Different algorithms can perform better or worse for different problems $P$ (sometimes drastically so)
- Studying P through an optimization lens can give you a deeper understanding of the task/procedure at hand
- Knowledge of optimization can help you create a new problem $P$ that is even more interesting/useful

Optimization moves quickly as a field. But there is still much room for progress, especially its intersection with ML and Stats

# Example: Algorithms for Linear Trend Filtering

Given observations $y_i \in \mathbb{R}$, $i = 1, \ldots, n$ corresponding to underlying locations $x_i = i$, $i = 1, \ldots, n$

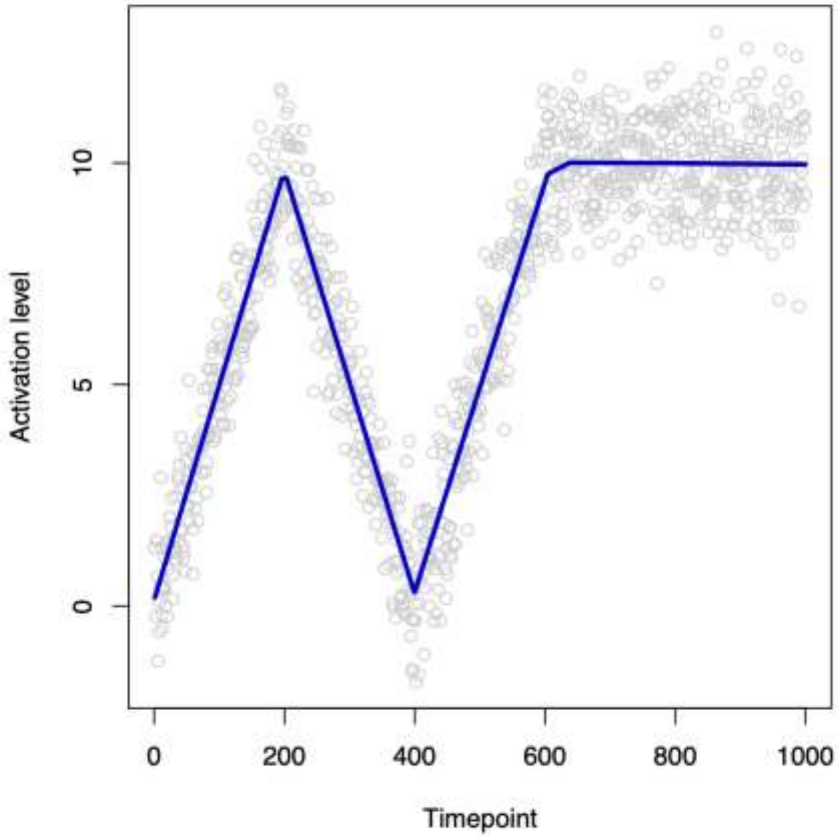



**Linear trend filtering** fits a piecewise linear function, with adaptively chosen knots (Steidl et al. 2006, Kim et al. 2009)

How? By solving $\min_{\theta} \frac{1}{2}\sum_{i=1}^{n}(y_i - \theta_i)^2 + \lambda \sum_{i=1}^{n-2} |\theta_i - 2\theta_{i+1} + \theta_{i+2}|$

# Example: Algorithms for Linear Trend Filtering

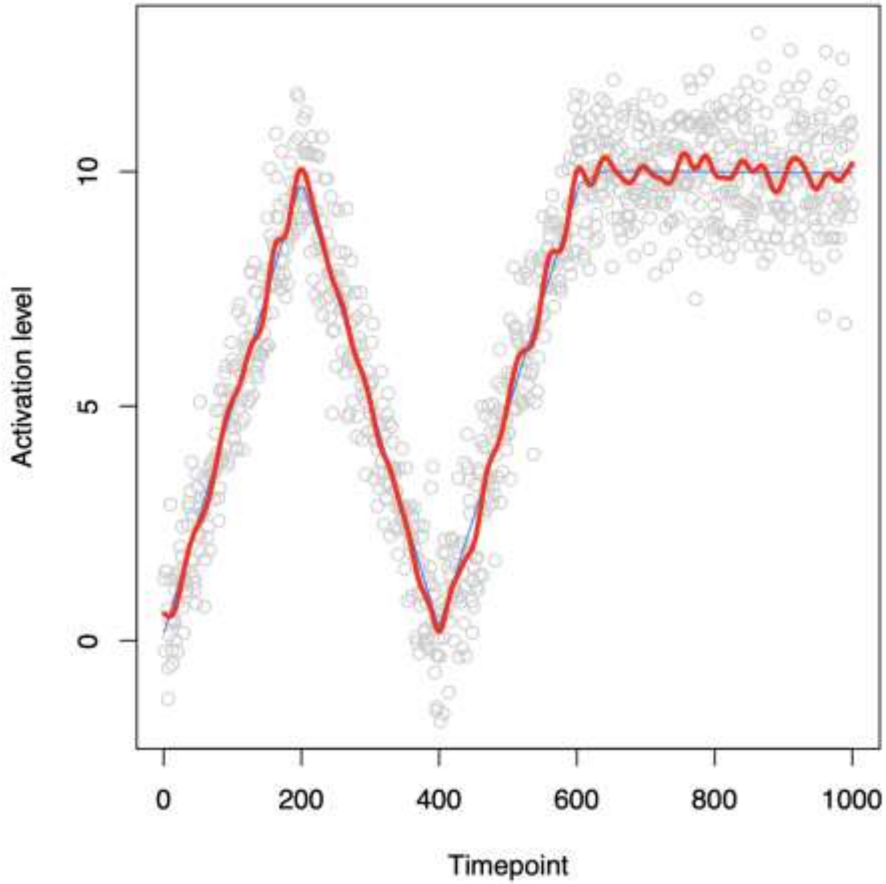
Problem: $$\min_{\theta} \frac{1}{2}\sum_{i=1}^{n}(y_i-\theta_i)^2+\lambda\sum_{i=1}^{n-2}|\theta_i-2\theta_{i+1}+\theta_{i+2}|$$




Interior point method, 20 iterations

## Example: Algorithms for Linear Trend Filtering

Problem: $$\min_{\theta} \frac{1}{2}\sum_{i=1}^{n}(y_i - \theta_i)^2 + \lambda \sum_{i=1}^{n-2} |\theta_i - 2\theta_{i+1} + \theta_{i+2}|$$

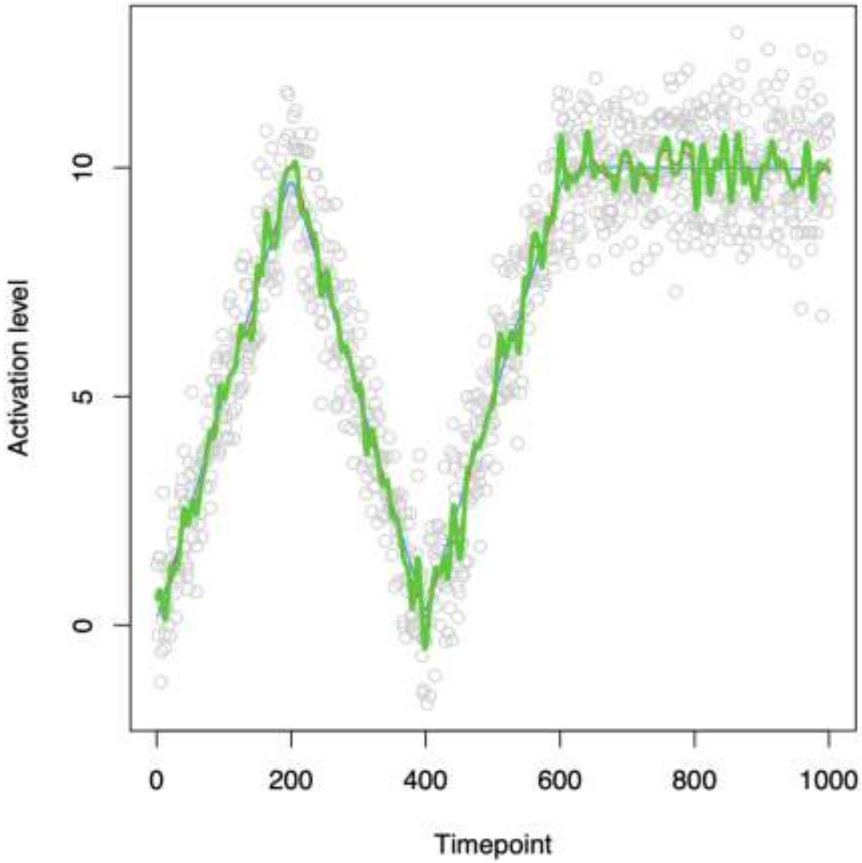



Interior point method, 20 iterations

Proximal gradient descent, 10K iterations

# Example: Algorithms for Linear Trend Filtering

Problem: $$\min_{\theta} \frac{1}{2}\sum_{i=1}^{n}(y_i-\theta_i)^2+\lambda\sum_{i=1}^{n-2}|\theta_i-2\theta_{i+1}+\theta_{i+2}|$$




Interior point method, 20 iterations

Proximal gradient descent, 10K iterations

Coordinate descent, 1000 cycles

# What’s the message here?

So, what’s the proper conclusion here?

Is the primal-dual interior point method better than proximal gradient or coordinate descent? … No

Different algorithms will work better in different situations. We’ll learn details throughout the course

In the linear trend filtering problem:

- Primal-dual: fast (structured linear systems)
- Proximal gradient: slow (conditioning)
- Coordinate descent: slow (large active set)

# Central Concepts: Convexity

Historically, linear programs were the focus of optimization.

Initially, it was thought that the crucial distinction was between linear and nonlinear optimization problems. However, some nonlinear issues were much more challenging than others …

It is widely recognized that the proper distinction is between convex and nonconvex problems.

Your supplementary textbooks for the course:

Boyd and Vandenberghe (2004)


and

Rockafellar (1970)

# Wisdom from Rockafellar (1993)

From Terry Rockafellar's 1993 SIAM Review survey paper:

> a convex set every locally optimal solution is global. Also, first-order necessary conditions for optimality turn out to be sufficient. A variety of other properties conducive to computation and interpretation of solutions ride on convexity as well. In fact the great watershed in optimization isn't between linearity and nonlinearity, but convexity and nonconvexity. Even for problems that aren't themselves of convex type, convexity may enter, for instance, in setting up subproblems as part of an iterative numerical scheme.

Credit to Nemirovski, Yudin, Nesterov, others for formalizing this

This view has dominated the optimization community and many application domains for decades. (… currently being challenged by neural networks' successes?)

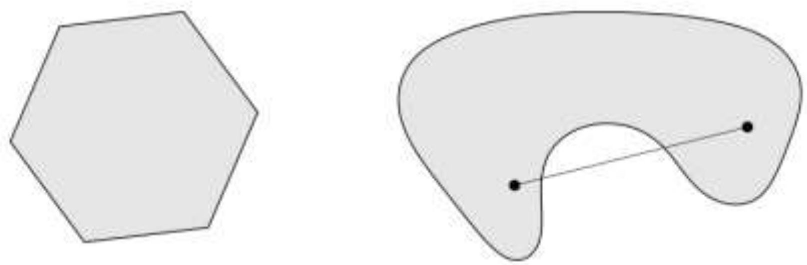
# Convex Sets and Functions

- Convex set: $C \subseteq \mathbb{R}^n$ such that

$$\boldsymbol{x}, \boldsymbol{y} \in C \rightarrow t\boldsymbol{x} + (1-t)\boldsymbol{y} \in C, \ \forall\, 0 \leq t \leq 1$$

In words, line segment joining any two elements lies entirely in set

Convex set



Nonconvex set

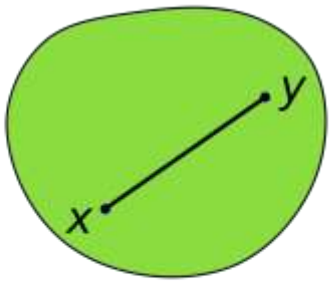
- Convex combination of $x_1, x_2, \ldots, x_k \in \mathbb{R}^n$: any linear combination

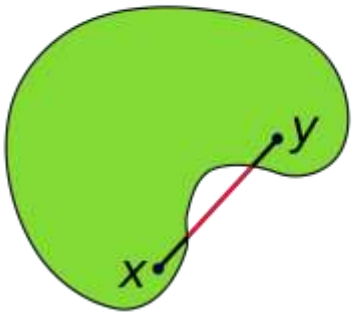
$$\theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k$$

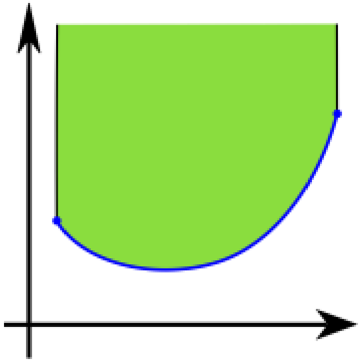
with $\theta_i \geq 0, i = 1, \ldots, k$, and $\sum_{i=1}^{k} \theta_i = 1$. Convex hull of a set $C$, $conv(C)$, is all convex combinations of elements. Always convex

# Convex Set: Small Exercises

- Please answer whether the green area is a convex set or not.

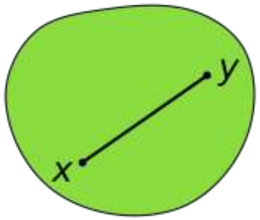


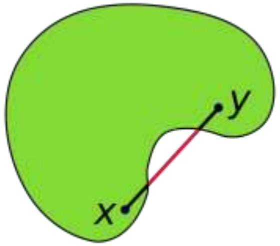



Credits from Wikipedia

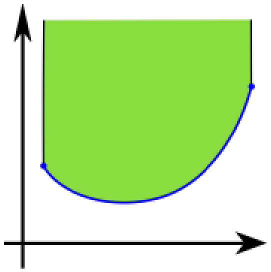
# Convex Set: Small Exercises




- It is a convex set, because the line segment between any two points in the set lies in the set




- It is not a convex set, because the line segment between x and y in the set does not lie in the set



- It is a convex set, because the line segment between any two points in the set lies in the set

# Convex Set: More Examples

- Example: the solution set of linear equations $Ax = b$ is a convex set.

- Why?

suppose $x_1,\ x_2 \in C$, *i.e.*, $Ax_1 = b,\ Ax_2 = b$.

Then for any $\theta$, we have

$$\begin{aligned} A(\theta x_1 + (1-\theta)x_2) &= \theta A x_1 + (1-\theta)Ax_2 \\ &= \theta b + (1-\theta)b \\ &= b \end{aligned}$$

# Examples of Convex Sets

- Trivial ones: empty set, point, line
- Norm ball: $\{\boldsymbol{x}\colon \|\boldsymbol{x}\| \le r\}$, for given norm $\|\cdot\|$, radius $r$
- Hyperplane: $\{x\colon \boldsymbol{a}^T\boldsymbol{x} = b\}$, for given $\boldsymbol{a}, b$
- Halfspace: $\{x\colon \boldsymbol{a}^T\boldsymbol{x} < b\}$
- Affine space: $\{x\colon A\boldsymbol{x} = \boldsymbol{b}\}$, for given $A, \boldsymbol{b}$

DEPARTMENT OF COMPUTER SCIENCE 計算機科學系

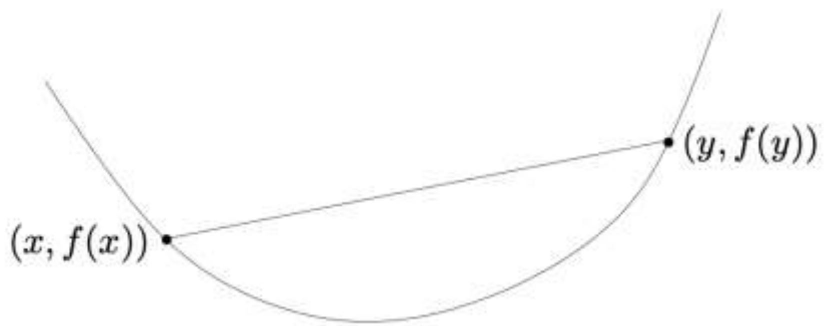
香港浸會大學 HONG KONG BAPTIST UNIVERSITY

# Convex Functions

Convex function: $f\colon \mathbb{R}^n \to \mathbb{R}$ such that $dom(f) \subseteq \mathbb{R}^n$ convex, and

$$f(t\boldsymbol{x} + (1-t)\boldsymbol{y}) \le tf(\boldsymbol{x}) + (1-t)f(\boldsymbol{y}), \text{ for } 0 \le t \le 1$$

and all $\boldsymbol{x}, \boldsymbol{y} \in dom(f)$



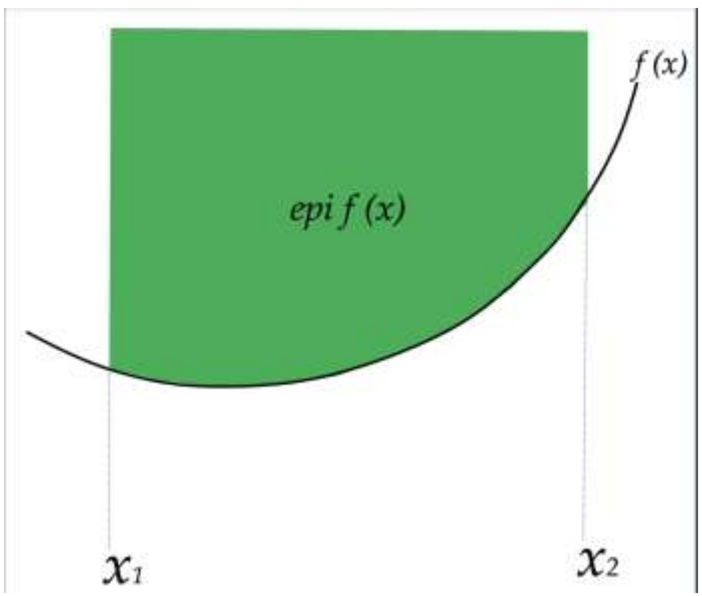

In other words, the function lies below the line segment joining $f(\boldsymbol{x})$, $f(y)$

# Key Properties of Convex Functions

- Epigraph characterization: a function $f$ is convex if and only if its epigraph

$$epi(f) = \{(x, t) \in dom(f) \times \mathbb{R} : f(x) \le t\}$$

is a convex set




We say $f$ is a *convex function* if its epigraph (the set of points above the function) defines a convex set.

# Key Properties of Convex Functions

- Convex sublevel sets: if $f$ is convex, then its sublevel sets

$$\mathrm{C} = \{x \in dom(f): f(x) \leq t\}$$

are convex, for all $t \in \mathbb{R}$. The converse is not true.

Proof: The proof is immediate from the definition of convexity. If $x, y \in C$, then $f(x) \leq t$ and $f(y) \leq t$, and so $f(\theta x + (1-\theta)y) \leq t$, for $0 \leq \theta \leq 1$, and hence $\theta x + (1-\theta)y \in C$.

# CVX Course Outline

- Convex Optimization Theory I
  - Sets and functions (22 October 2024)
  - **Optimization basics** (29 October 2024)
- Convex Optimization Algorithms I and Theory II
  - **First-order methods** (29 October 2024)
  - Duality and optimality (5 November 2024)
- Convex Optimization Algorithms II
  - Second-order methods (5 November 2024)
  - Advanced topics (Optional)