CSE 446
Convexity

Natasha Jaques

UNIVERSITY of WASHINGTON

Lecture plan

- Gradient descent algorithm + examples

- Theoretical analysis
  - When does it work?
    - **Key idea: <u>Convexity</u>** ← we are here
  - How quickly does it converge?
  - How do we choose a step size?

Convexity

- Optimization problems are hard to solve in general
- The exception: convex optimization
  - Objective is a convex function
  - Constraints are convex sets

Functions:

Convex   # Concave   # Non-convex

Sets:

Convex   # Non-convex

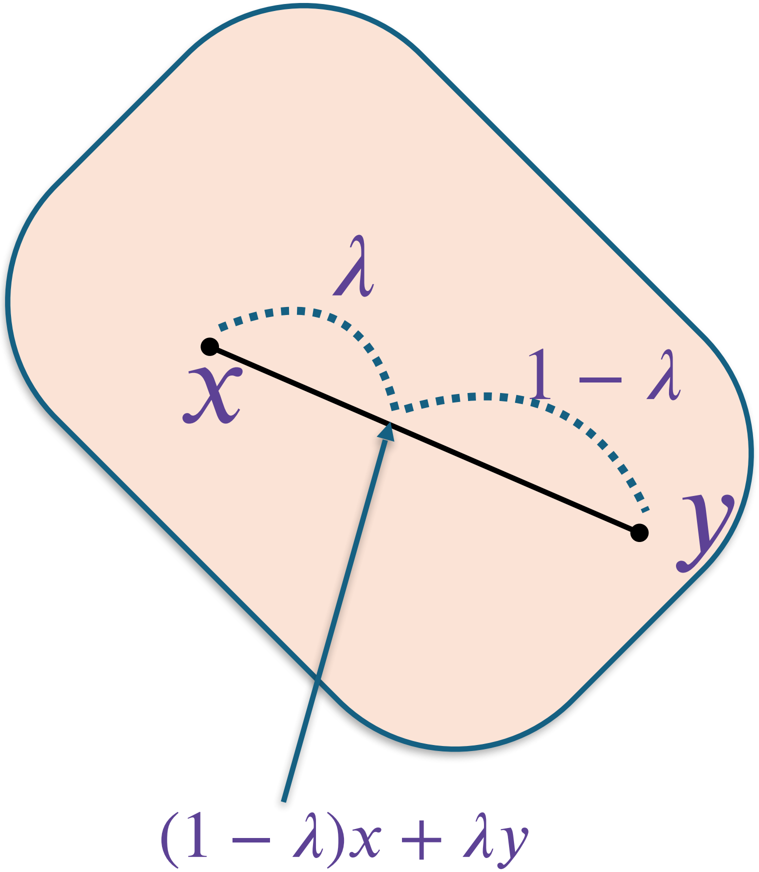
- Special class of problems that can be solved efficiently
- Surprisingly common in practice

What is a convex set?

What is a convex set?

Nothing to do with regularization

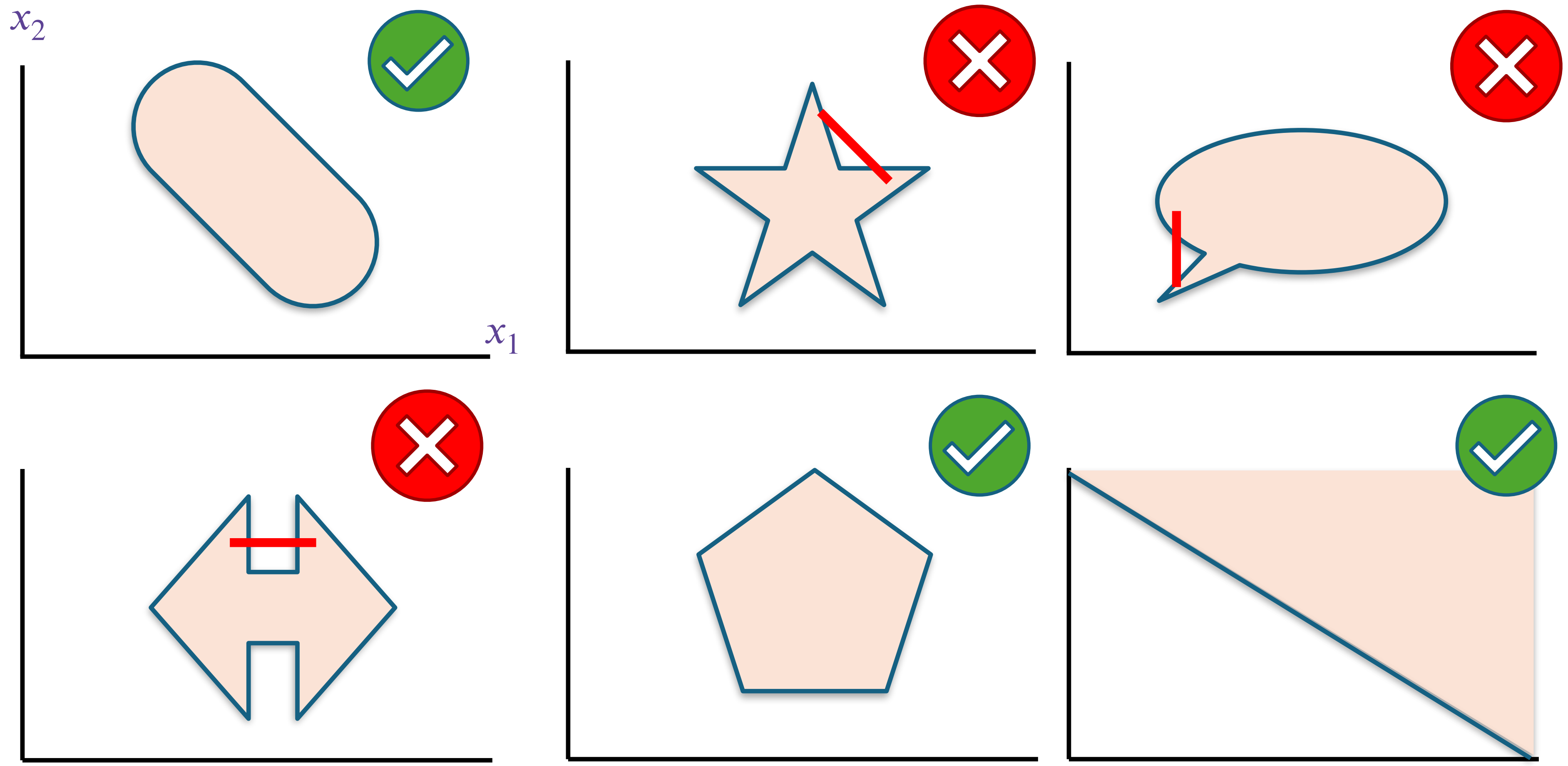
A set $K \subset \mathbb{R}^d$ is convex if $(1-\lambda)x + \lambda y \in K$ for all $x, y \in K$ and $\lambda \in [0,1]$



Translation: every point between 2 points in the set must also be in the set

What is a convex set?

A set $K \subset \mathbb{R}^d$ is convex if $(1-\lambda)x + \lambda y \in K$ for all $x, y \in K$ and $\lambda \in [0,1]$

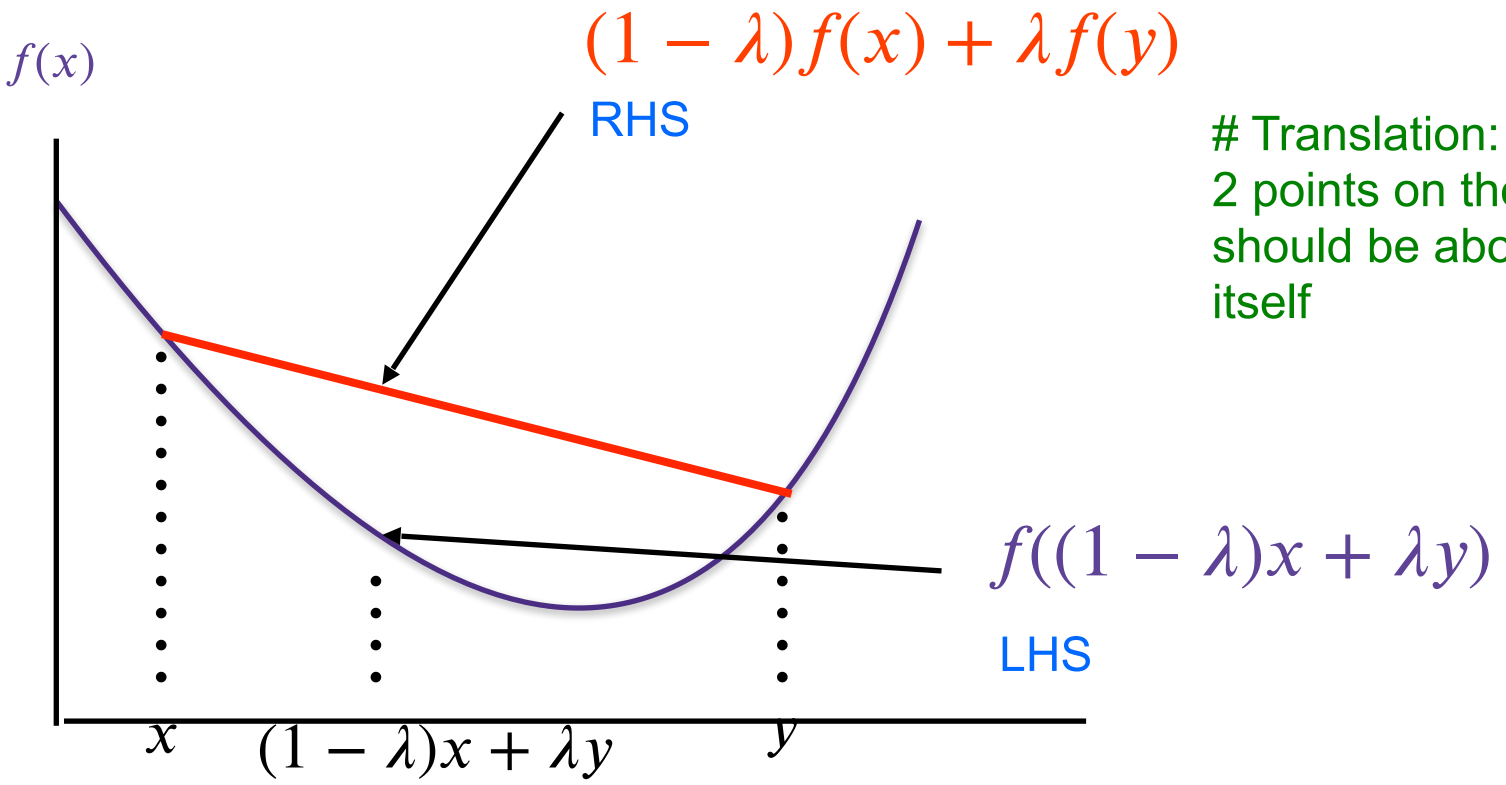


What is a convex function?

What is a convex function?

LHS RHS

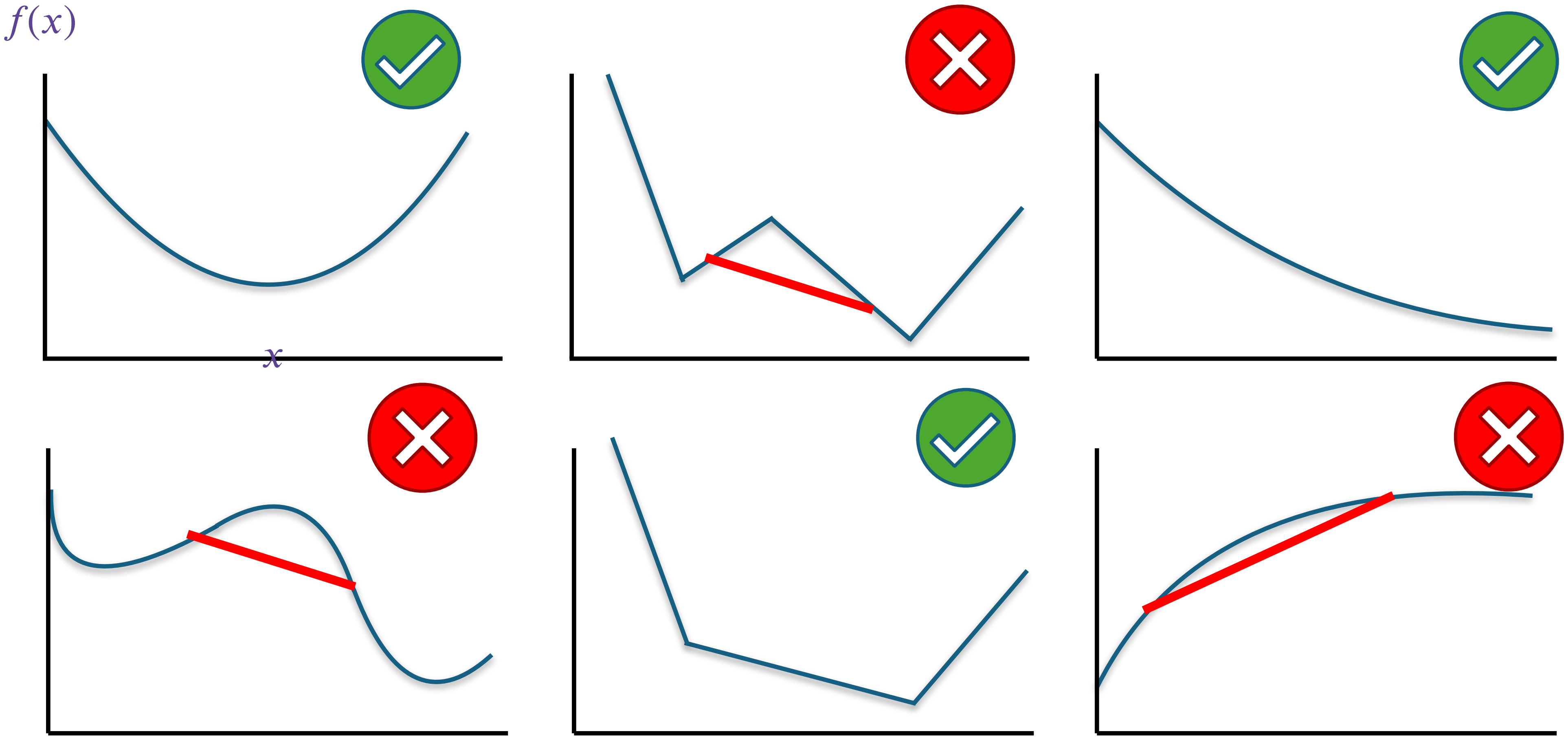
A function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is convex if $f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y)$ for all $x, y \in \mathbb{R}^d$ and $\lambda \in [0, 1]$



Translation: a line between 2 points on the function should be above the function itself

What is a convex function?

A function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is convex if $f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y)$ for all $x, y \in \mathbb{R}^d$ and $\lambda \in [0, 1]$



Convex functions and convex sets

Convex functions and convex sets

A set $K \subset \mathbb{R}^d$ is convex if $(1-\lambda)x + \lambda y \in K$ for all $x, y \in K$ and $\lambda \in [0,1]$ # Our first def'n

A function $f : \mathbb{R}^d \to \mathbb{R}$ is convex if $f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y)$ for all $x, y \in \mathbb{R}^d$ and $\lambda \in [0,1]$ # Our 2nd def'n

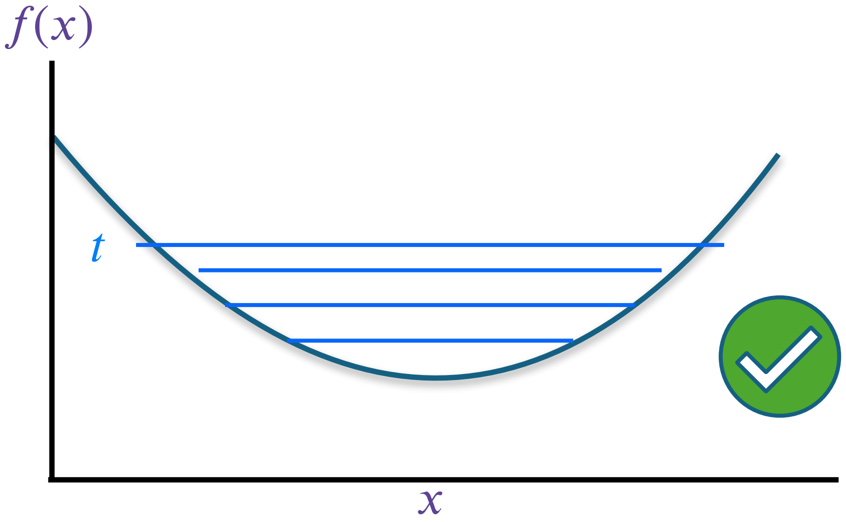
A function $f : \mathbb{R}^d \to \mathbb{R}$ is convex if the set $\{(x,t) \in \mathbb{R}^{d+1} : f(x) \leq t\}$ is convex # New def'n

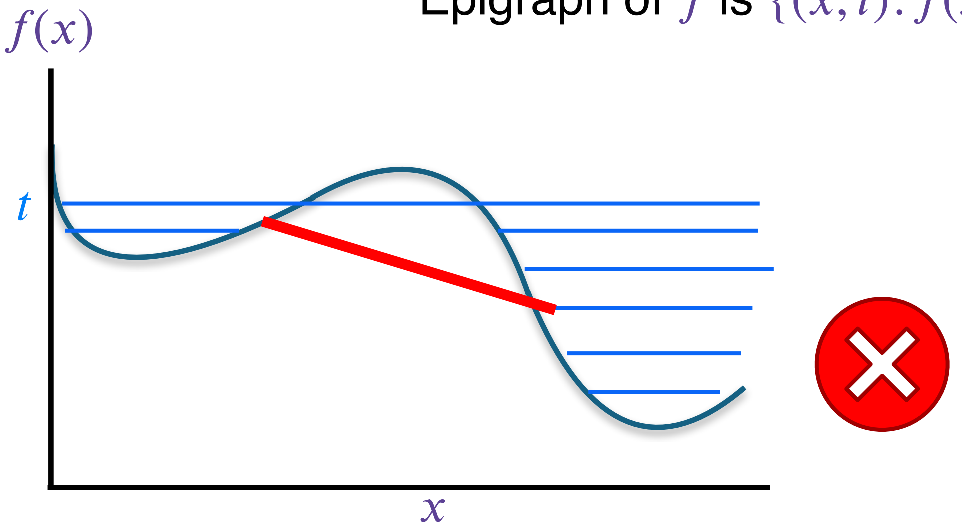
That set is the epigraph

Graph of f is $\{(x,t): f(x) = t\}$
Epigraph of f is $\{(x,t): f(x) \leq t\}$

$\forall t$

Translation: imagine pouring water into the top of the function. Does the water make a convex set like a bowl? Or are there ever disconnected puddles?





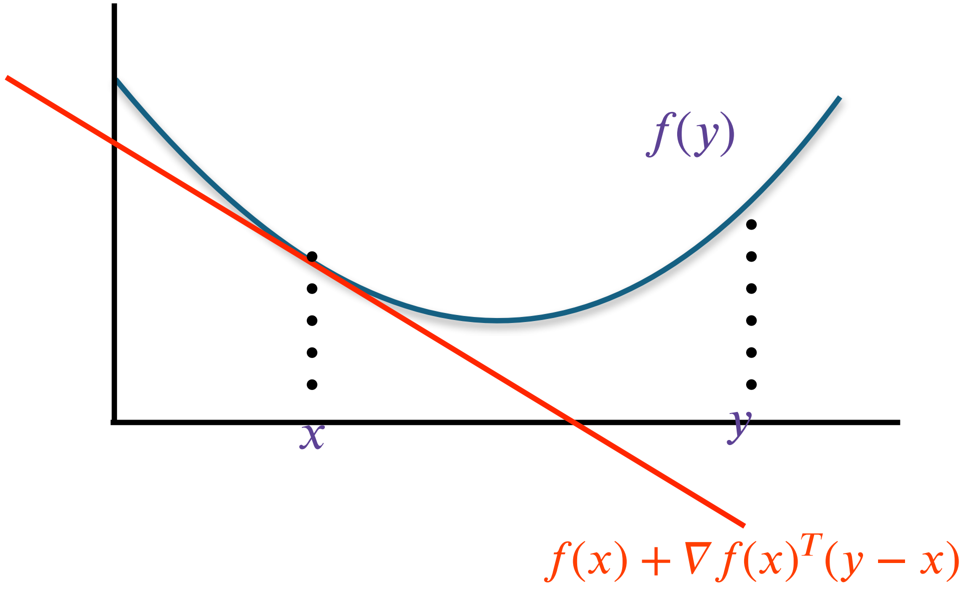
Convexity of differentiable functions

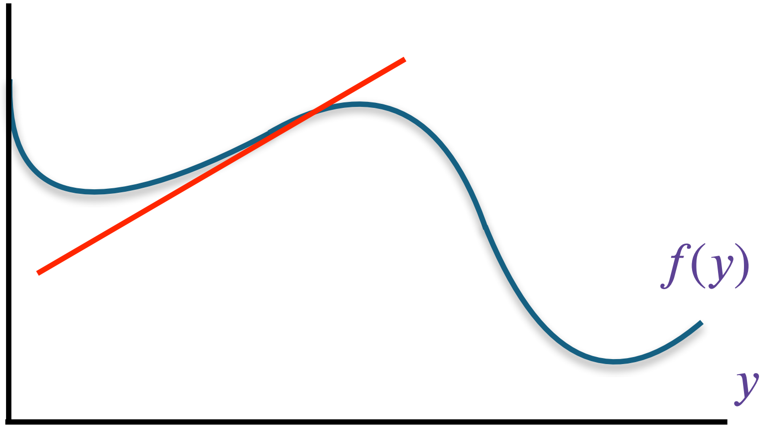
Convexity of differentiable functions

A function $f : \mathbb{R}^d \to \mathbb{R}$ is convex if the set $\{(x,t) \in \mathbb{R}^{d+1} : f(x) \leq t\}$ is convex

A function $f : \mathbb{R}^d \to \mathbb{R}$ that is differentiable everywhere is convex if
$f(y) \geq f(x) + \nabla f(x)^\top (y - x)$ for all $x, y \in dom(f)$

Translation: 1st order Taylor expansion needs to be below the curve everywhere



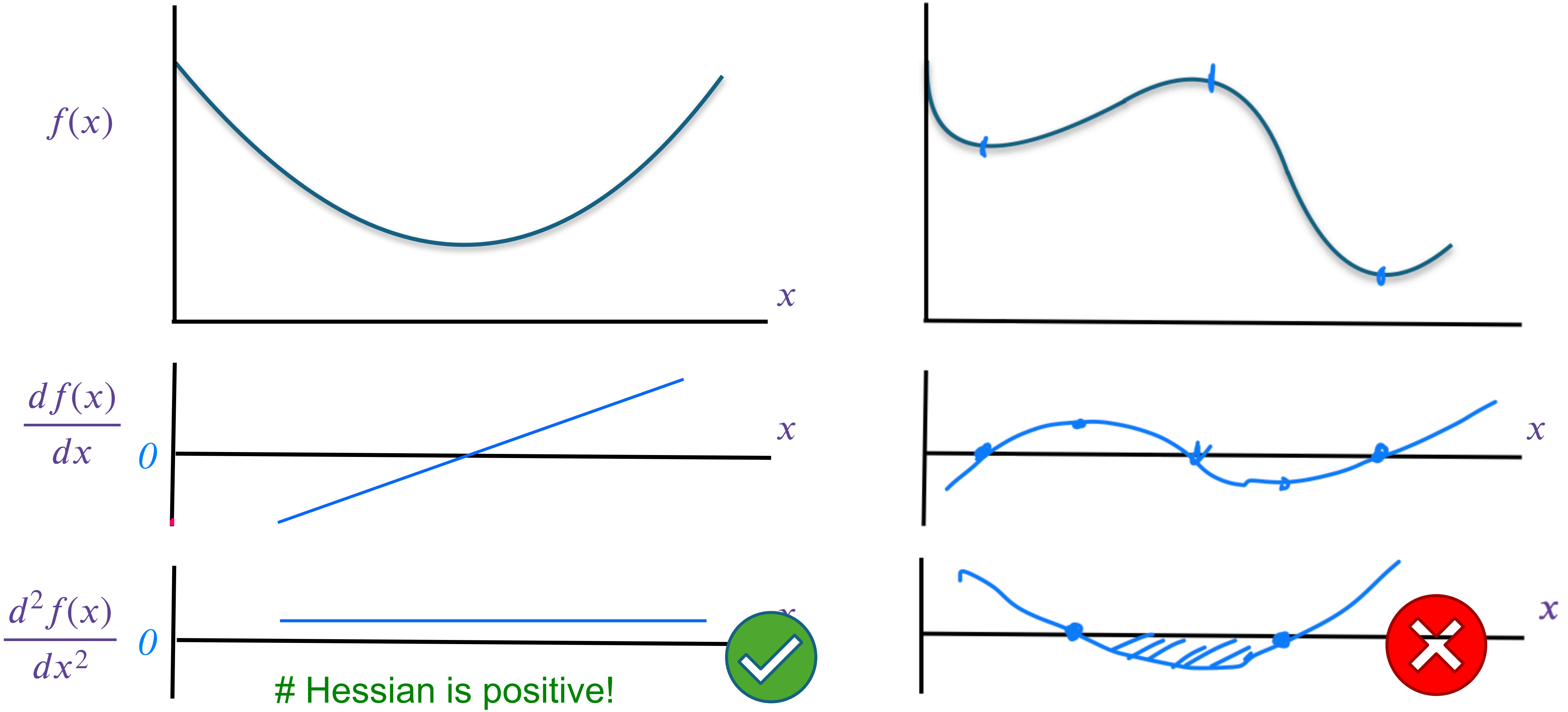


Convexity of twice-differentiable functions

Convexity of twice-differentiable functions

A function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ that is twice-differentiable everywhere is convex if and only if $\nabla^2 f(x) \succeq 0$ for all $x \in \text{dom}(f)$

Hessian must be non-negative



Hessian is positive!

Convexity of twice-differentiable functions

A function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ that is twice-differentiable everywhere is convex if and only if $\nabla^2 f(x) \succeq 0$ for all $x \in \mathrm{dom}(f)$

What is a Hessian?

$$\nabla^2 f(x)_{ij} = \frac{d^2 f(x)}{dx_i dx_j} = A \in \mathbb{R}^{d \times d}$$

Square matrix of second-order partial derivatives

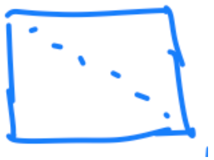
Symmetric: $A_{ij} = A_{ji}$

What is positive semi-definite?

$$A \succeq 0 \iff \forall v \in \mathbb{R}^d, \quad v^T A v \geq 0$$

$$\iff D_{ii} \geq 0 \quad \text{where} \quad A = Q^T D Q$$

Eigendecomposition $A = Q^T D Q$



Eigenvalues are non-negative $D_{ii} \geq 0$

Convexity of twice-differentiable functions

A function $f : \mathbb{R}^d \to \mathbb{R}$ that is twice-differentiable everywhere is convex if and only if $\nabla^2 f(x) \succeq 0$ for all $x \in \mathrm{dom}(f)$

$$A \succeq 0 \iff \forall v \in \mathbb{R}^d,\ \ v^T A v \geq 0 \iff D_{\mathrm{ii}} \geq 0 \ \text{ where } \ A = Q^T D Q$$

Why does having non-negative eigenvalues imply the Hessian is positive semi-definite?

Assume eigenvalues are non-negative

$$D_{\mathrm{ii}} \geq 0 \quad \forall i$$

$$v^T A v = v^T Q^T D Q v \qquad \text{let } u = Qv$$

$$= u^T D u$$

Positive because? Squared

$$= \sum_{i=1}^{d} u_i^2 D_{\mathrm{ii}} \geq 0 \iff A \succeq 0 \iff$$ **Convex!**

Positive because? By assumption

TL;DR: If you need to check if your function is convex, check if the eigenvalues are >= 0

Recap: Definitions of convexity

Sets

A set $K \subset \mathbb{R}^d$ is convex if $(1-\lambda)x + \lambda y \in K$ for all $x, y \in K$ and $\lambda \in [0,1]$

Line above function

A function $f : \mathbb{R}^d \to \mathbb{R}$ is convex if $f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y)$ for all $x, y \in \mathbb{R}^d$ and $\lambda \in [0,1]$

Epigraph

A function $f : \mathbb{R}^d \to \mathbb{R}$ is convex if the set $\{(x,t) \in \mathbb{R}^{d+1} : f(x) \leq t\}$ is convex

1st order Taylor expansion below function

A function $f : \mathbb{R}^d \to \mathbb{R}$ that is differentiable everywhere is convex if $f(y) \geq f(x) + \nabla f(x)^\top (y - x)$ for all $x, y \in dom(f)$

Hessian

A function $f : \mathbb{R}^d \to \mathbb{R}$ that is twice-differentiable everywhere is convex if $\nabla^2 f(x) \succeq 0$ for all $x \in dom(f)$

Example: Ridge regression

$$\underset{w}{\arg\min}\ \underbrace{\|y - Xw\|_2^2 + \lambda\|w\|_2^2}_{f(w)}$$

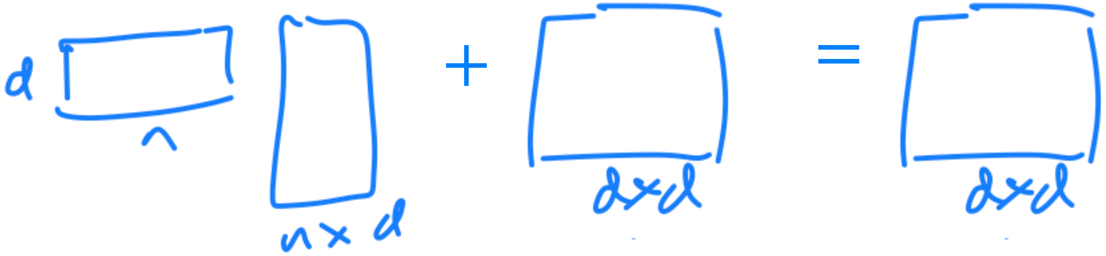
$$\underset{\|w\|_2^2 \le \rho}{\arg\min}\ \|y - Xw\|_2^2$$

How can we prove ridge regression is convex?

Let's use Hessian definition (easiest)

$$\nabla_w f(w) = 2X^T(Xw - y) + 2\lambda w$$

$$\nabla_w^2 f(w) = 2X^T X + 2\lambda I$$



Square and symmetric

Positive definite, convex

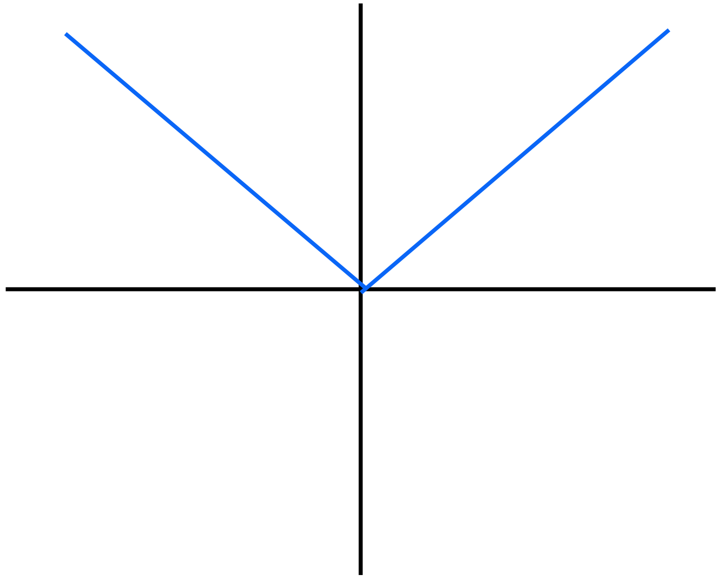
X^TX = PSD

$\lambda >= 0$ Positive

Example: Lasso

$$\underset{w}{\text{argmin}} \underbrace{\|y - Xw\|_2^2}_{f(w)} + \lambda\|w\|_1$$

Not twice differentiable, can't find Hessian

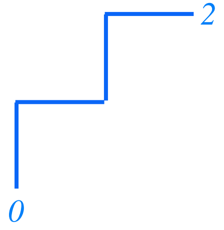


Sparser losses?

If we want sparsity, why not directly solve

$$\underset{w}{\operatorname{argmin}} \|y - Xw\|_2^2 + \lambda \operatorname{card}(w)?$$

Where card(w) = count(w_i) where $w_i > 0$

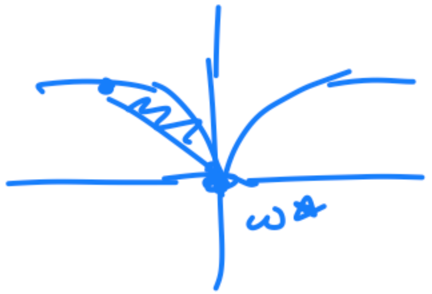


Convex?

Or what about the $l_{0.5}$ norm?

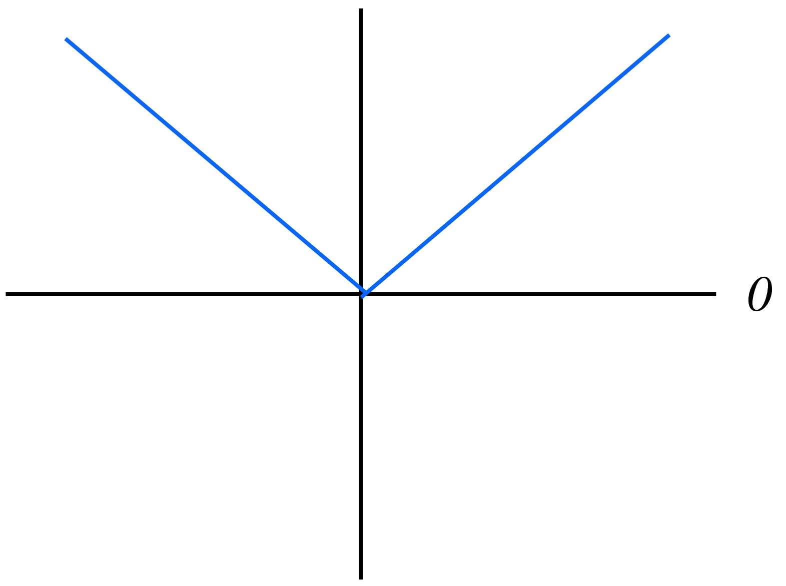
$$\underset{w}{\arg\min} \|y - Xw\|_2^2 + \lambda \|w\|_{1/2}^{1/2}$$

$$+\lambda \sum_{i=1}^{d} \sqrt{|w_i|}$$



Example: Lasso

TL;DR: LASSO is a convex relaxation of cardinality objective

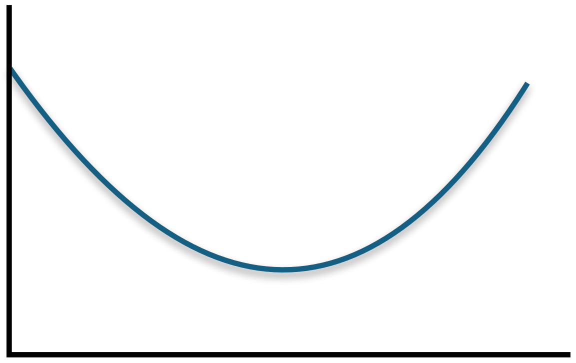


Convexity and gradient descent

Efficient to optimize with gradient descent

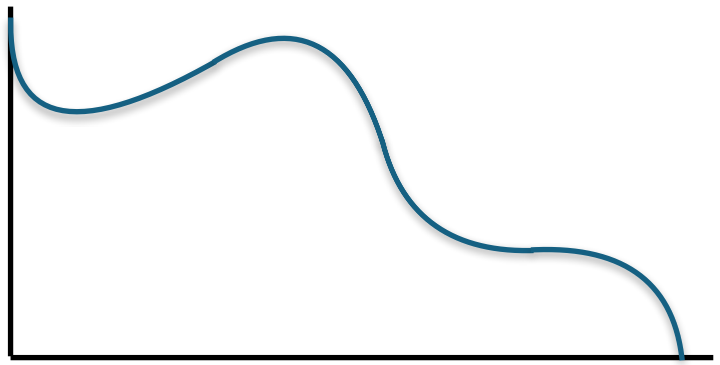
- **Convexity:** All local minima are global minima

Convex function



Stationary points with $\nabla f(x) = 0$
are global minima

Non-convex function



Stationary points with $\nabla f(x) = 0$
could be a local minima,
a local maxima, or a saddle point

- Won't get stuck navigating the parameter constraints

Convexity and gradient descent

- You can always run gradient descent whether $f(w)$ is convex or not!
- But if $f(w)$ is convex, we have guarantees on converging to the global minimum
- Linear regression, ridge regression, Lasso → all convex!