

# CS794/CO673: Optimization for Data Science

## Lec 03: Projection

Yaoliang Yu



UNIVERSITY OF  
**WATERLOO**

FACULTY OF MATHEMATICS  
DAVID R. CHERITON SCHOOL  
OF COMPUTER SCIENCE

September 16, 2022

Constrained smooth minimization:

$$f_{\star} = \inf_{\mathbf{w} \in C} f(\mathbf{w}).$$

- Constraint on the domain: closed set  $C \subseteq \mathbb{R}^d$
- $f : \mathbb{R}^d \rightarrow \mathbb{R}$  is smooth, e.g. continuously differentiable
- $f$  can be convex or nonconvex;  $C$  can be convex or nonconvex
- Minimizer may or may not be attained
- Maximization is just negation

# Problem

Constrained smooth minimization:

$$f_{\star} = \inf_{\mathbf{w} \in C} f(\mathbf{w}).$$

- Constraint on the domain: closed set  $C \subseteq \mathbb{R}^d$
- $f : \mathbb{R}^d \rightarrow \mathbb{R}$  is smooth, e.g. continuously differentiable
- $f$  can be convex or nonconvex;  $C$  can be convex or nonconvex
- Minimizer may or may not be attained
- Maximization is just negation

# Problem

Constrained smooth minimization:

$$f_{\star} = \inf_{\mathbf{w} \in C} f(\mathbf{w}).$$

- Constraint on the domain: closed set  $C \subseteq \mathbb{R}^d$
- $f : \mathbb{R}^d \rightarrow \mathbb{R}$  is smooth, e.g. continuously differentiable
- $f$  can be convex or nonconvex;  $C$  can be convex or nonconvex
- Minimizer may or may not be attained
- Maximization is just negation

# Problem

Constrained smooth minimization:

$$f_{\star} = \inf_{\mathbf{w} \in C} f(\mathbf{w}).$$

- Constraint on the domain: closed set  $C \subseteq \mathbb{R}^d$
- $f : \mathbb{R}^d \rightarrow \mathbb{R}$  is smooth, e.g. continuously differentiable
- $f$  can be convex or nonconvex;  $C$  can be convex or nonconvex
- Minimizer may or may not be attained
- Maximization is just negation

# Problem

Constrained smooth minimization:

$$f_{\star} = \inf_{\mathbf{w} \in C} f(\mathbf{w}).$$

- Constraint on the domain: closed set  $C \subseteq \mathbb{R}^d$
- $f : \mathbb{R}^d \rightarrow \mathbb{R}$  is smooth, e.g. continuously differentiable
- $f$  can be convex or nonconvex;  $C$  can be convex or nonconvex
- Minimizer may or may not be attained
- Maximization is just negation

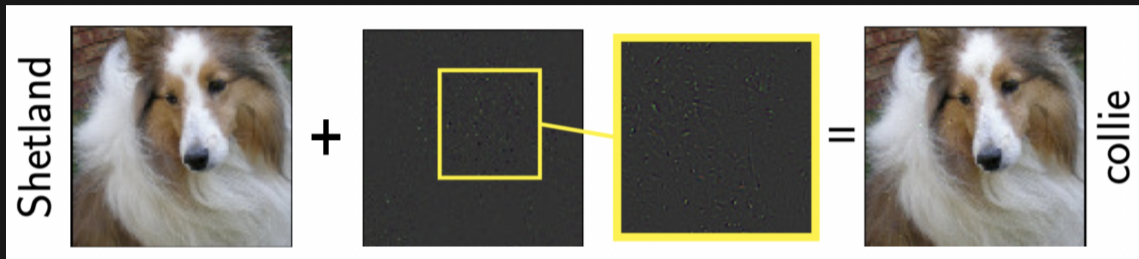
# Problem

Constrained smooth minimization:

$$f_{\star} = \inf_{\mathbf{w} \in C} f(\mathbf{w}).$$

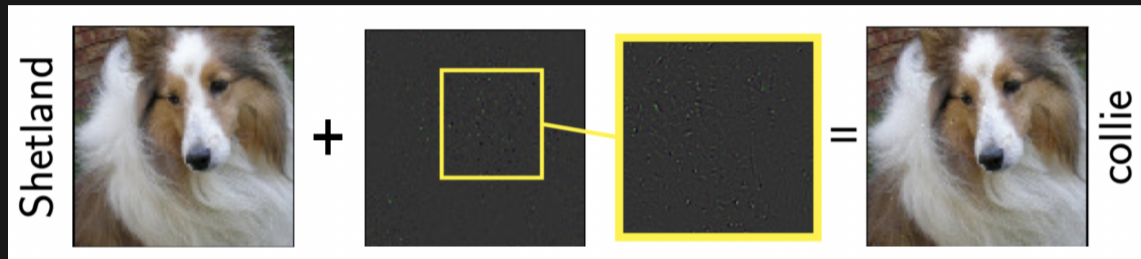
- Constraint on the domain: closed set  $C \subseteq \mathbb{R}^d$
- $f : \mathbb{R}^d \rightarrow \mathbb{R}$  is smooth, e.g. continuously differentiable
- $f$  can be convex or nonconvex;  $C$  can be convex or nonconvex
- Minimizer may or may not be attained
- Maximization is just negation

# White-box Adversarial Attacks





# White-box Adversarial Attacks

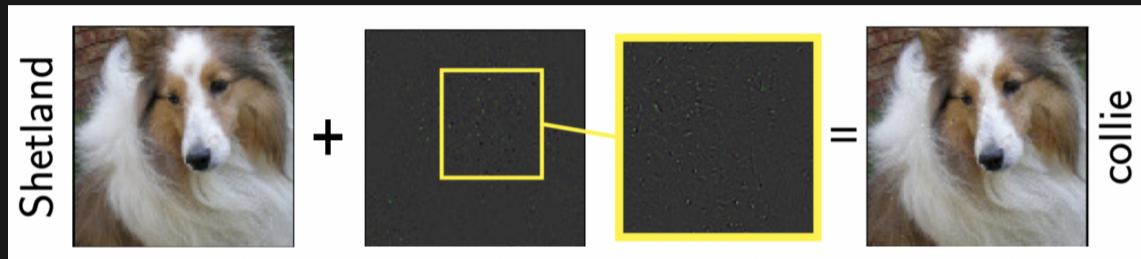


- Mathematically, a neural network is a function  $f(\mathbf{w}; \mathbf{x})$
- Typically, input  $\mathbf{x}$  is given and network weights  $\mathbf{w}$  optimized
- Could also freeze weights  $\mathbf{w}$  and optimize  $\mathbf{x}$ , **adversarially!**

$$\min_{\delta} \text{size}(\delta) \quad \text{s.t.} \quad \text{pred}[f(\mathbf{w}; \mathbf{x} + \delta)] \neq y$$

- More generally:  $\max_{\delta} \ell(\mathbf{w}; \mathbf{x} + \delta, y) \quad \text{s.t.} \quad \text{size}(\delta) \leq \epsilon \quad \text{and} \quad 0 \leq \mathbf{x} + \delta \leq 1$

# White-box Adversarial Attacks

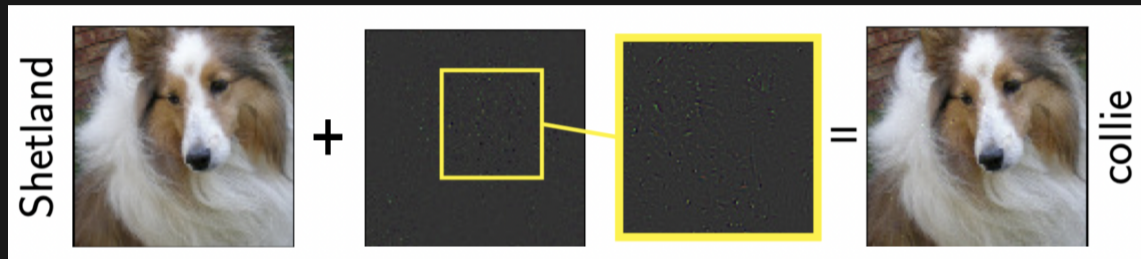


- Mathematically, a neural network is a function  $f(\mathbf{w}; \mathbf{x})$
- Typically, input  $\mathbf{x}$  is given and network weights  $\mathbf{w}$  optimized
- Could also freeze weights  $\mathbf{w}$  and optimize  $\mathbf{x}$ , adversarially!

$$\min_{\delta} \text{size}(\delta) \quad \text{s.t.} \quad \text{pred}[f(\mathbf{w}; \mathbf{x} + \delta)] \neq y$$

- More generally:  $\max_{\delta} \ell(\mathbf{w}; \mathbf{x} + \delta, y) \quad \text{s.t.} \quad \text{size}(\delta) \leq \epsilon \quad \text{and} \quad 0 \leq \mathbf{x} + \delta \leq 1$

# White-box Adversarial Attacks

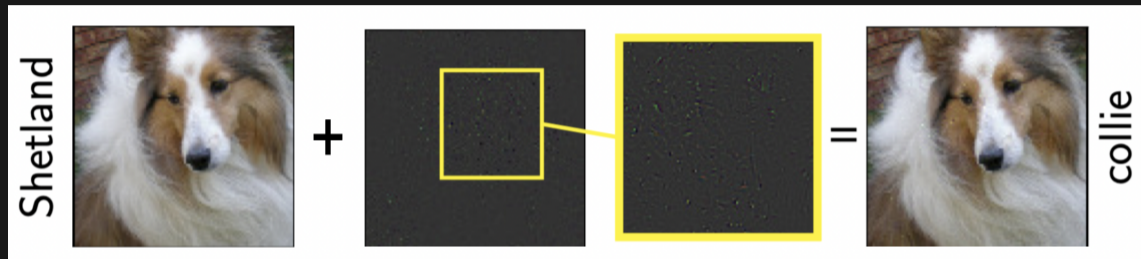


- Mathematically, a neural network is a function  $f(\mathbf{w}; \mathbf{x})$
- Typically, input  $\mathbf{x}$  is given and network weights  $\mathbf{w}$  optimized
- Could also freeze weights  $\mathbf{w}$  and optimize  $\mathbf{x}$ , adversarially!

$$\min_{\delta} \text{size}(\delta) \quad \text{s.t.} \quad \text{pred}[f(\mathbf{w}; \mathbf{x} + \delta)] \neq y$$

- More generally:  $\max_{\delta} \ell(\mathbf{w}; \mathbf{x} + \delta, y) \quad \text{s.t.} \quad \text{size}(\delta) \leq \epsilon \quad \text{and} \quad 0 \leq \mathbf{x} + \delta \leq 1$

# White-box Adversarial Attacks

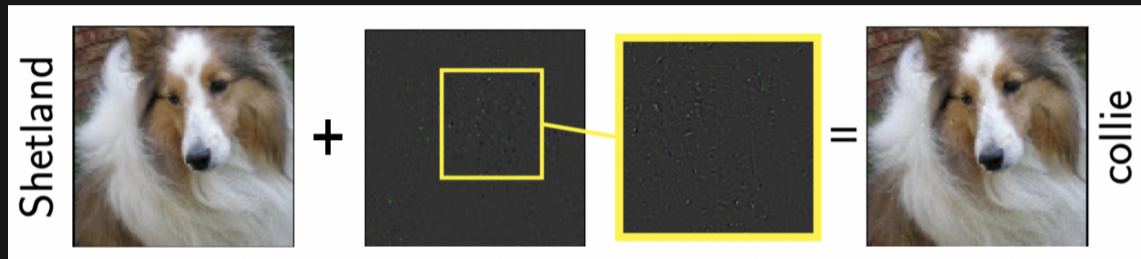


- Mathematically, a neural network is a function  $f(\mathbf{w}; \mathbf{x})$
- Typically, input  $\mathbf{x}$  is given and network weights  $\mathbf{w}$  optimized
- Could also freeze weights  $\mathbf{w}$  and optimize  $\mathbf{x}$ , **adversarially!**

$$\min_{\delta} \text{size}(\delta) \quad \text{s.t.} \quad \text{pred}[f(\mathbf{w}; \mathbf{x} + \delta)] \neq y$$

- More generally:  $\max_{\delta} \ell(\mathbf{w}; \mathbf{x} + \delta, y) \quad \text{s.t.} \quad \text{size}(\delta) \leq \epsilon \quad \text{and} \quad 0 \leq \mathbf{x} + \delta \leq 1$

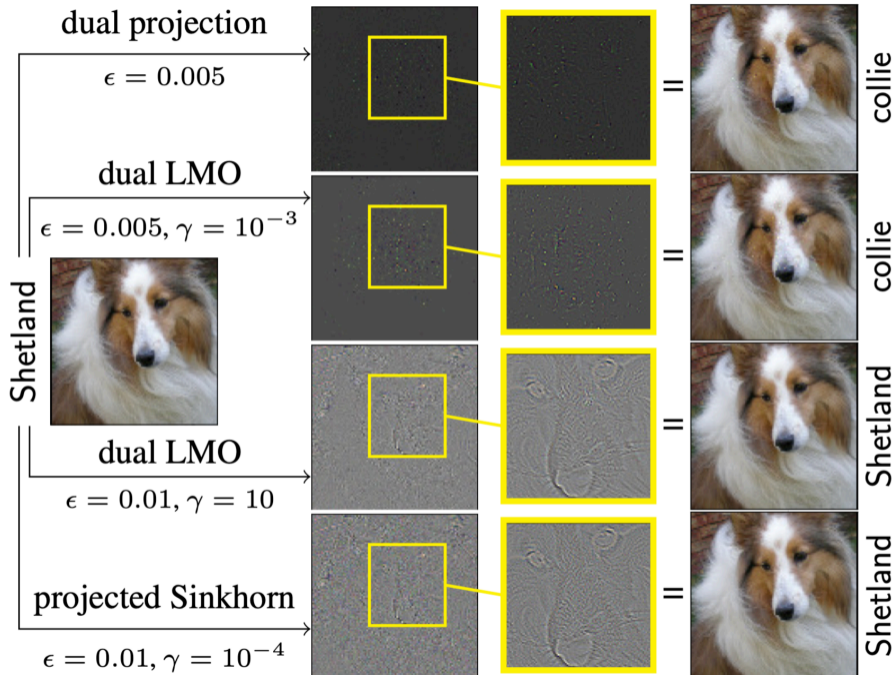
# White-box Adversarial Attacks



- Mathematically, a neural network is a function  $f(\mathbf{w}; \mathbf{x})$
- Typically, input  $\mathbf{x}$  is given and network weights  $\mathbf{w}$  optimized
- Could also freeze weights  $\mathbf{w}$  and optimize  $\mathbf{x}$ , **adversarially!**

$$\min_{\delta} \text{size}(\delta) \quad \text{s.t.} \quad \text{pred}[f(\mathbf{w}; \mathbf{x} + \delta)] \neq y$$

- More generally:  $\max_{\delta} \ell(\mathbf{w}; \mathbf{x} + \delta, y) \quad \text{s.t.} \quad \text{size}(\delta) \leq \epsilon \quad \text{and} \quad \mathbf{0} \leq \mathbf{x} + \delta \leq 1$



# Convexity

A point set  $C \subseteq \mathbb{R}^d$  is **convex** iff for any  $\mathbf{w}, \mathbf{z} \in C$ , the line segment  $[\mathbf{w}, \mathbf{z}] \subseteq C$ .

The **epigraph** of a function  $f : \mathbb{R}^d \rightarrow (-\infty, \infty]$  is defined as the set

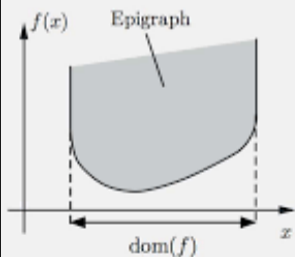
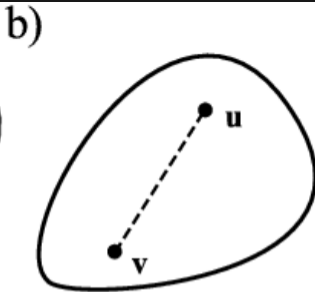
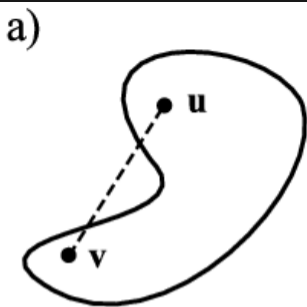
$$\text{epi } f := \{(\mathbf{w}, t) \in \mathbb{R}^{d+1} : f(\mathbf{w}) \leq t\}$$

A function  $f : \mathbb{R}^d \rightarrow (-\infty, \infty]$  is convex iff its epigraph is a convex set, or equivalently

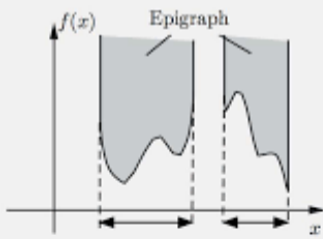
$$\forall \mathbf{w}, \forall \mathbf{z}, \forall \lambda \in [0, 1], \quad f(\lambda \mathbf{w} + (1 - \lambda) \mathbf{z}) \leq \lambda f(\mathbf{w}) + (1 - \lambda) f(\mathbf{z})$$

Theorem: second-order test for convexity

$f$  is convex iff  $\nabla^2 f$  is positive semidefinite.



Convex function



Nonconvex function



# Calculus of Convexity

- $f, g$  convex  $\implies \alpha \cdot f + \beta \cdot g$  is convex for any  $\alpha, \beta \geq 0$
- $f$  convex  $\implies f(A\mathbf{w})$  is convex
- $f$  convex increasing and  $g$  convex  $\implies f \circ g$  is convex
- $f$  convex  $\implies (\mathbf{w}, t > 0) \mapsto tf(\mathbf{w}/t)$  is convex
- $f_t$  convex  $\implies f = \sup_t f_t$  is convex
- $f(\mathbf{w}, \mathbf{z})$  convex  $\implies g = \min_{\mathbf{z}} f(\mathbf{w}, \mathbf{z})$  is convex
- Is  $\log(\sum_j \exp(w_j))$  convex?

# Calculus of Convexity

- $f, g$  convex  $\implies \alpha \cdot f + \beta \cdot g$  is convex for any  $\alpha, \beta \geq 0$
- $f$  convex  $\implies f(A\mathbf{w})$  is convex
- $f$  convex increasing and  $g$  convex  $\implies f \circ g$  is convex
- $f$  convex  $\implies (\mathbf{w}, t > 0) \mapsto tf(\mathbf{w}/t)$  is convex
- $f_t$  convex  $\implies f = \sup_t f_t$  is convex
- $f(\mathbf{w}, \mathbf{z})$  convex  $\implies g = \min_{\mathbf{z}} f(\mathbf{w}, \mathbf{z})$  is convex
- Is  $\log(\sum_j \exp(w_j))$  convex?

# Calculus of Convexity

- $f, g$  convex  $\implies \alpha \cdot f + \beta \cdot g$  is convex for any  $\alpha, \beta \geq 0$
- $f$  convex  $\implies f(A\mathbf{w})$  is convex
- $f$  convex increasing and  $g$  convex  $\implies f \circ g$  is convex
- $f$  convex  $\implies (\mathbf{w}, t > 0) \mapsto tf(\mathbf{w}/t)$  is convex
- $f_t$  convex  $\implies f = \sup_t f_t$  is convex
- $f(\mathbf{w}, \mathbf{z})$  convex  $\implies g = \min_{\mathbf{z}} f(\mathbf{w}, \mathbf{z})$  is convex
- Is  $\log(\sum_j \exp(w_j))$  convex?

# Calculus of Convexity

- $f, g$  convex  $\implies \alpha \cdot f + \beta \cdot g$  is convex for any  $\alpha, \beta \geq 0$
- $f$  convex  $\implies f(A\mathbf{w})$  is convex
- $f$  convex increasing and  $g$  convex  $\implies f \circ g$  is convex
- $f$  convex  $\implies (\mathbf{w}, t > 0) \mapsto tf(\mathbf{w}/t)$  is convex
- $f_t$  convex  $\implies f = \sup_t f_t$  is convex
- $f(\mathbf{w}, \mathbf{z})$  convex  $\implies g = \min_{\mathbf{z}} f(\mathbf{w}, \mathbf{z})$  is convex
- Is  $\log(\sum_j \exp(w_j))$  convex?

# Calculus of Convexity

- $f, g$  convex  $\implies \alpha \cdot f + \beta \cdot g$  is convex for any  $\alpha, \beta \geq 0$
- $f$  convex  $\implies f(A\mathbf{w})$  is convex
- $f$  convex increasing and  $g$  convex  $\implies f \circ g$  is convex
- $f$  convex  $\implies (\mathbf{w}, t > 0) \mapsto tf(\mathbf{w}/t)$  is convex
- $f_t$  convex  $\implies f = \sup_t f_t$  is convex
- $f(\mathbf{w}, \mathbf{z})$  convex  $\implies g = \min_{\mathbf{z}} f(\mathbf{w}, \mathbf{z})$  is convex
- Is  $\log(\sum_j \exp(w_j))$  convex?

# Calculus of Convexity

- $f, g$  convex  $\implies \alpha \cdot f + \beta \cdot g$  is convex for any  $\alpha, \beta \geq 0$
- $f$  convex  $\implies f(A\mathbf{w})$  is convex
- $f$  convex increasing and  $g$  convex  $\implies f \circ g$  is convex
- $f$  convex  $\implies (\mathbf{w}, t > 0) \mapsto tf(\mathbf{w}/t)$  is convex
- $f_t$  convex  $\implies f = \sup_t f_t$  is convex
- $f(\mathbf{w}, \mathbf{z})$  convex  $\implies g = \min_{\mathbf{z}} f(\mathbf{w}, \mathbf{z})$  is convex
- Is  $\log(\sum_j \exp(w_j))$  convex?

# Calculus of Convexity

- $f, g$  convex  $\implies \alpha \cdot f + \beta \cdot g$  is convex for any  $\alpha, \beta \geq 0$
- $f$  convex  $\implies f(A\mathbf{w})$  is convex
- $f$  convex increasing and  $g$  convex  $\implies f \circ g$  is convex
- $f$  convex  $\implies (\mathbf{w}, t > 0) \mapsto tf(\mathbf{w}/t)$  is convex
- $f_t$  convex  $\implies f = \sup_t f_t$  is convex
- $f(\mathbf{w}, \mathbf{z})$  convex  $\implies g = \min_{\mathbf{z}} f(\mathbf{w}, \mathbf{z})$  is convex
- Is  $\log(\sum_j \exp(w_j))$  convex?

# Calculus of Convexity

- $f, g$  convex  $\implies \alpha \cdot f + \beta \cdot g$  is convex for any  $\alpha, \beta \geq 0$
- $f$  convex  $\implies f(A\mathbf{w})$  is convex
- $f$  convex increasing and  $g$  convex  $\implies f \circ g$  is convex
- $f$  convex  $\implies (\mathbf{w}, t > 0) \mapsto tf(\mathbf{w}/t)$  is convex
- $f_t$  convex  $\implies f = \sup_t f_t$  is convex
- $f(\mathbf{w}, \mathbf{z})$  convex  $\implies g = \min_{\mathbf{z}} f(\mathbf{w}, \mathbf{z})$  is convex
- Is  $\log(\sum_j \exp(w_j))$  convex?



# A Nice Univariate Result

## Theorem: constrained univariate convex minimization

For any univariate convex function  $f$  and convex interval  $C = [a, b]$ , we have

$$P_C \left( \underset{w \in \mathbb{R}}{\operatorname{argmin}} f(w) \right) \subseteq \underset{w \in C}{\operatorname{argmin}} f(w),$$

where  $P_C(w) = P_{[a,b]}(w) = (a \vee w) \wedge b$  is the closest point in  $C$  to  $w$ .

- Not true if  $C$  is not an interval (i.e. not convex)
- Not true if  $f$  is not convex
- Not true when dimension  $d \geq 2$ , even when both  $f$  and  $C$  are convex
- Except when  $\underset{w \in \mathbb{R}^d}{\operatorname{argmin}} f(w) \subseteq C$

# A Nice Univariate Result

## Theorem: constrained univariate convex minimization

For any univariate convex function  $f$  and convex interval  $C = [a, b]$ , we have

$$P_C \left( \underset{w \in \mathbb{R}}{\operatorname{argmin}} f(w) \right) \subseteq \underset{w \in C}{\operatorname{argmin}} f(w),$$

where  $P_C(w) = P_{[a,b]}(w) = (a \vee w) \wedge b$  is the closest point in  $C$  to  $w$ .

- Not true if  $C$  is not an interval (i.e. not convex)
- Not true if  $f$  is not convex
- Not true when dimension  $d \geq 2$ , even when both  $f$  and  $C$  are convex
- Except when  $\underset{w \in \mathbb{R}^d}{\operatorname{argmin}} f(w) \subseteq C$

# A Nice Univariate Result

## Theorem: constrained univariate convex minimization

For any univariate convex function  $f$  and convex interval  $C = [a, b]$ , we have

$$P_C \left( \underset{w \in \mathbb{R}}{\operatorname{argmin}} f(w) \right) \subseteq \underset{w \in C}{\operatorname{argmin}} f(w),$$

where  $P_C(w) = P_{[a,b]}(w) = (a \vee w) \wedge b$  is the closest point in  $C$  to  $w$ .

- Not true if  $C$  is not an interval (i.e. not convex)
- Not true if  $f$  is not convex
- Not true when dimension  $d \geq 2$ , even when both  $f$  and  $C$  are convex
- Except when  $\underset{w \in \mathbb{R}^d}{\operatorname{argmin}} f(w) \subseteq C$

# A Nice Univariate Result

## Theorem: constrained univariate convex minimization

For any univariate convex function  $f$  and convex interval  $C = [a, b]$ , we have

$$P_C \left( \underset{w \in \mathbb{R}}{\operatorname{argmin}} f(w) \right) \subseteq \underset{w \in C}{\operatorname{argmin}} f(w),$$

where  $P_C(w) = P_{[a,b]}(w) = (a \vee w) \wedge b$  is the closest point in  $C$  to  $w$ .

- Not true if  $C$  is not an interval (i.e. not convex)
- Not true if  $f$  is not convex
- Not true when dimension  $d \geq 2$ , even when both  $f$  and  $C$  are convex
- Except when  $\underset{w \in \mathbb{R}^d}{\operatorname{argmin}} f(w) \subseteq C$

# A Nice Univariate Result

## Theorem: constrained univariate convex minimization

For any univariate convex function  $f$  and convex interval  $C = [a, b]$ , we have

$$P_C \left( \underset{w \in \mathbb{R}}{\operatorname{argmin}} f(w) \right) \subseteq \underset{w \in C}{\operatorname{argmin}} f(w),$$

where  $P_C(w) = P_{[a,b]}(w) = (a \vee w) \wedge b$  is the closest point in  $C$  to  $w$ .

- Not true if  $C$  is not an interval (i.e. not convex)
- Not true if  $f$  is not convex
- Not true when dimension  $d \geq 2$ , even when both  $f$  and  $C$  are convex
- Except when  $\underset{\mathbf{w} \in \mathbb{R}^d}{\operatorname{argmin}} f(\mathbf{w}) \subseteq C$

# An Algorithm that does NOT work

$$\eta \leftarrow \underset{\eta \geq 0}{\operatorname{argmin}} f(\mathbf{w}_\eta), \quad \text{s.t.} \quad \mathbf{w}_\eta := \mathbf{w} - \eta \cdot \nabla f(\mathbf{w}) \in C$$
$$\mathbf{w} \leftarrow \mathbf{w}_\eta$$

- Does NOT work

- Important to leave the constraint set  $C$

# An Algorithm that does NOT work

$$\eta \leftarrow \underset{\eta \geq 0}{\operatorname{argmin}} f(\mathbf{w}_\eta), \quad \text{s.t.} \quad \mathbf{w}_\eta := \mathbf{w} - \eta \cdot \nabla f(\mathbf{w}) \in C$$
$$\mathbf{w} \leftarrow \mathbf{w}_\eta$$

- Does NOT work

- $f(\mathbf{w}) := \frac{1}{2}(w_1^2 + w_2^2)$

- $C = \{\mathbf{w} \geq \mathbf{0} : w_1 + w_2 = 1\}$

- stuck at  $\mathbf{w} = (1, 0)$  while minimum is at  $\mathbf{w}_* = (\frac{1}{2}, \frac{1}{2})$

- Important to leave the constraint set  $C$

# An Algorithm that does NOT work

$$\eta \leftarrow \underset{\eta \geq 0}{\operatorname{argmin}} f(\mathbf{w}_\eta), \quad \text{s.t.} \quad \mathbf{w}_\eta := \mathbf{w} - \eta \cdot \nabla f(\mathbf{w}) \in C$$
$$\mathbf{w} \leftarrow \mathbf{w}_\eta$$

- Does NOT work

- $f(\mathbf{w}) := \frac{1}{2}(w_1^2 + w_2^2)$

- $C = \{\mathbf{w} \geq \mathbf{0} : w_1 + w_2 = 1\}$

- stuck at  $\mathbf{w} = (1, 0)$  while minimum is at  $\mathbf{w}_* = (\frac{1}{2}, \frac{1}{2})$

- Important to leave the constraint set  $C$



# An Algorithm that does NOT work

$$\eta \leftarrow \underset{\eta \geq 0}{\operatorname{argmin}} f(\mathbf{w}_\eta), \quad \text{s.t.} \quad \mathbf{w}_\eta := \mathbf{w} - \eta \cdot \nabla f(\mathbf{w}) \in C$$
$$\mathbf{w} \leftarrow \mathbf{w}_\eta$$

- Does NOT work

- $f(\mathbf{w}) := \frac{1}{2}(w_1^2 + w_2^2)$

- $C = \{\mathbf{w} \geq \mathbf{0} : w_1 + w_2 = 1\}$

- stuck at  $\mathbf{w} = (1, 0)$  while minimum is at  $\mathbf{w}_* = (\frac{1}{2}, \frac{1}{2})$

- Important to leave the constraint set  $C$

# An Algorithm that does NOT work

$$\eta \leftarrow \underset{\eta \geq 0}{\operatorname{argmin}} f(\mathbf{w}_\eta), \quad \text{s.t.} \quad \mathbf{w}_\eta := \mathbf{w} - \eta \cdot \nabla f(\mathbf{w}) \in C$$
$$\mathbf{w} \leftarrow \mathbf{w}_\eta$$

- Does NOT work

- $f(\mathbf{w}) := \frac{1}{2}(w_1^2 + w_2^2)$

- $C = \{\mathbf{w} \geq \mathbf{0} : w_1 + w_2 = 1\}$

- stuck at  $\mathbf{w} = (1, 0)$  while minimum is at  $\mathbf{w}_* = (\frac{1}{2}, \frac{1}{2})$

- Important to leave the constraint set  $C$

# An Algorithm that does NOT work

$$\eta \leftarrow \underset{\eta \geq 0}{\operatorname{argmin}} f(\mathbf{w}_\eta), \quad \text{s.t.} \quad \mathbf{w}_\eta := \mathbf{w} - \eta \cdot \nabla f(\mathbf{w}) \in C$$
$$\mathbf{w} \leftarrow \mathbf{w}_\eta$$

- Does NOT work

- $f(\mathbf{w}) := \frac{1}{2}(w_1^2 + w_2^2)$

- $C = \{\mathbf{w} \geq \mathbf{0} : w_1 + w_2 = 1\}$

- stuck at  $\mathbf{w} = (1, 0)$  while minimum is at  $\mathbf{w}_* = (\frac{1}{2}, \frac{1}{2})$

- Important to leave the constraint set  $C$

# (Euclidean) Projection

Let  $C \subseteq \mathbb{R}^d$  be a closed set. The Euclidean projection of a point  $\mathbf{w} \in \mathbb{R}^d$  to  $C$  is:

$$P_C(\mathbf{w}) := \operatorname{argmin}_{\mathbf{z} \in C} \|\mathbf{z} - \mathbf{w}\|_2,$$

i.e. the point(s) in  $C$  that are closest to the given point  $\mathbf{w}$ .

- We always have  $P_C(\mathbf{w}) \neq \emptyset$  and compact
- $P_C(\mathbf{w}) = \mathbf{w}$  iff  $\mathbf{w} \in C$
- $P_C(\mathbf{w}) = \operatorname{bd} C$  if  $\mathbf{w} \notin C$
- In  $\mathbb{R}^d$ ,  $P_C$  is unique iff  $C$  is convex

# (Euclidean) Projection

Let  $C \subseteq \mathbb{R}^d$  be a closed set. The Euclidean projection of a point  $\mathbf{w} \in \mathbb{R}^d$  to  $C$  is:

$$P_C(\mathbf{w}) := \operatorname{argmin}_{\mathbf{z} \in C} \|\mathbf{z} - \mathbf{w}\|_2,$$

i.e. the point(s) in  $C$  that are closest to the given point  $\mathbf{w}$ .

- We always have  $P_C(\mathbf{w}) \neq \emptyset$  and compact
- $P_C(\mathbf{w}) = \mathbf{w}$  iff  $\mathbf{w} \in C$
- $P_C(\mathbf{w}) = \operatorname{bd} C$  if  $\mathbf{w} \notin C$
- In  $\mathbb{R}^d$ ,  $P_C$  is unique iff  $C$  is convex

# (Euclidean) Projection

Let  $C \subseteq \mathbb{R}^d$  be a closed set. The Euclidean projection of a point  $\mathbf{w} \in \mathbb{R}^d$  to  $C$  is:

$$P_C(\mathbf{w}) := \operatorname{argmin}_{\mathbf{z} \in C} \|\mathbf{z} - \mathbf{w}\|_2,$$

i.e. the point(s) in  $C$  that are closest to the given point  $\mathbf{w}$ .

- We always have  $P_C(\mathbf{w}) \neq \emptyset$  and compact
- $P_C(\mathbf{w}) = \mathbf{w}$  iff  $\mathbf{w} \in C$
- $P_C(\mathbf{w}) = \operatorname{bd} C$  if  $\mathbf{w} \notin C$
- In  $\mathbb{R}^d$ ,  $P_C$  is unique iff  $C$  is convex

# (Euclidean) Projection

Let  $C \subseteq \mathbb{R}^d$  be a closed set. The Euclidean projection of a point  $\mathbf{w} \in \mathbb{R}^d$  to  $C$  is:

$$P_C(\mathbf{w}) := \operatorname{argmin}_{\mathbf{z} \in C} \|\mathbf{z} - \mathbf{w}\|_2,$$

i.e. the point(s) in  $C$  that are closest to the given point  $\mathbf{w}$ .

- We always have  $P_C(\mathbf{w}) \neq \emptyset$  and compact
- $P_C(\mathbf{w}) = \mathbf{w}$  iff  $\mathbf{w} \in C$
- $P_C(\mathbf{w}) = \operatorname{bd} C$  if  $\mathbf{w} \notin C$
- In  $\mathbb{R}^d$ ,  $P_C$  is unique iff  $C$  is convex

# (Euclidean) Projection

Let  $C \subseteq \mathbb{R}^d$  be a closed set. The Euclidean projection of a point  $\mathbf{w} \in \mathbb{R}^d$  to  $C$  is:

$$P_C(\mathbf{w}) := \operatorname{argmin}_{\mathbf{z} \in C} \|\mathbf{z} - \mathbf{w}\|_2,$$

i.e. the point(s) in  $C$  that are closest to the given point  $\mathbf{w}$ .

- We always have  $P_C(\mathbf{w}) \neq \emptyset$  and compact
- $P_C(\mathbf{w}) = \mathbf{w}$  iff  $\mathbf{w} \in C$
- $P_C(\mathbf{w}) = \operatorname{bd} C$  if  $\mathbf{w} \notin C$
- In  $\mathbb{R}^d$ ,  $P_C$  is unique iff  $C$  is convex



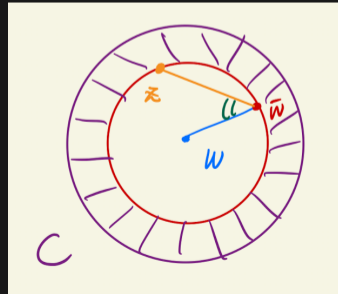
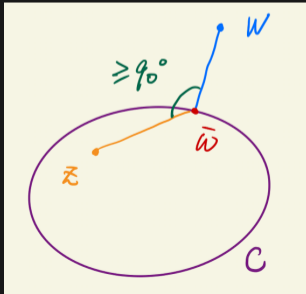
# Geometrically

Theorem:

If  $C$  is convex, then  $\bar{w} = P_C(w)$  iff for all  $z \in C$

$$\langle z - \bar{w}, w - \bar{w} \rangle \leq 0,$$

or equivalently,  $\frac{1}{2}\|z - w\|_2^2 \geq \frac{1}{2}\|z - \bar{w}\|_2^2 + \frac{1}{2}\|\bar{w} - w\|_2^2$ .



## Example: Projection to the hypercube

$$\min_{\mathbf{a} \leq \boldsymbol{\delta} \leq \mathbf{b}} \|\boldsymbol{\delta} - \boldsymbol{\gamma}\|_2 = \min_{\mathbf{a} \leq \boldsymbol{\delta} \leq \mathbf{b}} \|\boldsymbol{\delta} - \boldsymbol{\gamma}\|_2^2$$

- Problem is separable: reduce to each dimension separately
- Apply the nice univariate result  $\boldsymbol{\delta}_* = (\boldsymbol{\gamma} \vee \mathbf{a}) \wedge \mathbf{b}$

- Decompose  $\mathbf{z} = r \cdot \bar{\mathbf{z}}$ , where  $r \geq 0$ ,  $\|\bar{\mathbf{z}}\|_2 = 1$
- Apply the nice univariate result  $\mathbf{w}_* = \left( \frac{\lambda}{\|\mathbf{w}\|_2} \wedge 1 \right) \cdot \mathbf{w}$

## Example: Projection to the hypercube

$$\min_{\mathbf{a} \leq \boldsymbol{\delta} \leq \mathbf{b}} \|\boldsymbol{\delta} - \boldsymbol{\gamma}\|_2 = \min_{\mathbf{a} \leq \boldsymbol{\delta} \leq \mathbf{b}} \|\boldsymbol{\delta} - \boldsymbol{\gamma}\|_2^2$$

- Problem is separable: reduce to each dimension separately
- Apply the nice univariate result  $\delta_* = (\gamma \vee \mathbf{a}) \wedge \mathbf{b}$

- Decompose  $\mathbf{z} = r \cdot \bar{\mathbf{z}}$ , where  $r \geq 0$ ,  $\|\bar{\mathbf{z}}\|_2 = 1$
- Apply the nice univariate result  $\mathbf{w}_* = \left( \frac{\lambda}{\|\mathbf{w}\|_2} \wedge 1 \right) \cdot \mathbf{w}$

## Example: Projection to the hypercube

$$\min_{\mathbf{a} \leq \boldsymbol{\delta} \leq \mathbf{b}} \|\boldsymbol{\delta} - \boldsymbol{\gamma}\|_2 = \min_{\mathbf{a} \leq \boldsymbol{\delta} \leq \mathbf{b}} \|\boldsymbol{\delta} - \boldsymbol{\gamma}\|_2^2$$

- Problem is separable: reduce to each dimension separately
- Apply the nice univariate result  $\boldsymbol{\delta}_* = (\boldsymbol{\gamma} \vee \mathbf{a}) \wedge \mathbf{b}$

## Example: Projection to the ball

$$\min_{\|\mathbf{z}\|_2 \leq \lambda} \|\mathbf{w} - \mathbf{z}\|_2 = \min_{\|\mathbf{z}\|_2 \leq \lambda} \|\mathbf{w} - \mathbf{z}\|_2^2$$

- Decompose  $\mathbf{z} = r \cdot \bar{\mathbf{z}}$ , where  $r \geq 0$ ,  $\|\bar{\mathbf{z}}\|_2 = 1$
- Apply the nice univariate result  $\mathbf{w}_* = \left( \frac{\lambda}{\|\mathbf{w}\|_2} \wedge 1 \right) \cdot \mathbf{w}$

## Example: Projection to the hypercube

$$\min_{\mathbf{a} \leq \boldsymbol{\delta} \leq \mathbf{b}} \|\boldsymbol{\delta} - \boldsymbol{\gamma}\|_2 = \min_{\mathbf{a} \leq \boldsymbol{\delta} \leq \mathbf{b}} \|\boldsymbol{\delta} - \boldsymbol{\gamma}\|_2^2$$

- Problem is separable: reduce to each dimension separately
- Apply the nice univariate result  $\boldsymbol{\delta}_* = (\boldsymbol{\gamma} \vee \mathbf{a}) \wedge \mathbf{b}$

## Example: Projection to the ball

$$\min_{\|\mathbf{z}\|_2 \leq \lambda} \|\mathbf{w} - \mathbf{z}\|_2 = \min_{\|\mathbf{z}\|_2 \leq \lambda} \|\mathbf{w} - \mathbf{z}\|_2^2$$

- Decompose  $\mathbf{z} = r \cdot \bar{\mathbf{z}}$ , where  $r \geq 0$ ,  $\|\bar{\mathbf{z}}\|_2 = 1$
- Apply the nice univariate result  $\mathbf{w}_* = \left( \frac{\lambda}{\|\mathbf{w}\|_2} \wedge 1 \right) \cdot \mathbf{w}$

## Example: Projection to the hypercube

$$\min_{\mathbf{a} \leq \boldsymbol{\delta} \leq \mathbf{b}} \|\boldsymbol{\delta} - \boldsymbol{\gamma}\|_2 = \min_{\mathbf{a} \leq \boldsymbol{\delta} \leq \mathbf{b}} \|\boldsymbol{\delta} - \boldsymbol{\gamma}\|_2^2$$

- Problem is separable: reduce to each dimension separately
- Apply the nice univariate result  $\boldsymbol{\delta}_* = (\boldsymbol{\gamma} \vee \mathbf{a}) \wedge \mathbf{b}$

## Example: Projection to the ball

$$\min_{\|\mathbf{z}\|_2 \leq \lambda} \|\mathbf{w} - \mathbf{z}\|_2 = \min_{\|\mathbf{z}\|_2 \leq \lambda} \|\mathbf{w} - \mathbf{z}\|_2^2$$

- Decompose  $\mathbf{z} = r \cdot \bar{\mathbf{z}}$ , where  $r \geq 0$ ,  $\|\bar{\mathbf{z}}\|_2 = 1$
- Apply the nice univariate result  $\mathbf{w}_* = \left( \frac{\lambda}{\|\mathbf{w}\|_2} \wedge 1 \right) \cdot \mathbf{w}$

---

## Algorithm 1: Projected gradient descent for constrained smooth minimization

---

**Input:**  $\mathbf{w}_0 \in \mathbb{R}^d$ , constraint  $C \subseteq \mathbb{R}^d$ , smooth function  $f : \mathbb{R}^d \rightarrow \mathbb{R}$

```
1 for  $t = 0, 1, \dots$  do
2    $\mathbf{g}_t \leftarrow \nabla f(\mathbf{w}_t)$  // compute the gradient
3    $\mathbf{w}_{t+1} \leftarrow \mathbf{w}_t - \eta_t \mathbf{g}_t$  //  $\eta_t$  is the step size
4    $\mathbf{w}_{t+1} \leftarrow P_C(\mathbf{w}_{t+1})$  // project back to the constraint
```

---

- $C = \mathbb{R}^d$ : reduces to gradient descent
- Motivation from L-smoothness:

$$\begin{aligned} f(\mathbf{w}) &\leq f(\mathbf{w}_t) + \langle \mathbf{w} - \mathbf{w}_t, \nabla f(\mathbf{w}_t) \rangle + \frac{1}{2\eta_t} \|\mathbf{w} - \mathbf{w}_t\|_2^2 \\ &= \frac{1}{2\eta_t} \|\mathbf{w} - (\mathbf{w}_t - \eta_t \nabla f(\mathbf{w}_t))\|_2^2 + f(\mathbf{w}_t) - \frac{\eta_t}{2} \|\nabla f(\mathbf{w}_t)\|_2^2 \end{aligned}$$

---

A. A. Goldstein. "Convex programming in Hilbert space". *Bulletin of the American Mathematical Society*, vol. 70, no. 5 (1964), pp. 709–710, E. S. Levitin and B. T. Polyak. "Constrained Minimization Methods". *USSR Computational Mathematics and Mathematical Physics*, vol. 6, no. 5 (1966), pp. 1–50. [English translation in *Zh. Vychisl. Mat. mat. Fiz.* vol. 6, no. 5, pp. 787–823, 1965].

---

## Algorithm 2: Projected gradient descent for constrained smooth minimization

---

**Input:**  $\mathbf{w}_0 \in \mathbb{R}^d$ , constraint  $C \subseteq \mathbb{R}^d$ , smooth function  $f : \mathbb{R}^d \rightarrow \mathbb{R}$

```
1 for  $t = 0, 1, \dots$  do
2    $\mathbf{g}_t \leftarrow \nabla f(\mathbf{w}_t)$  // compute the gradient
3    $\mathbf{w}_{t+1} \leftarrow \mathbf{w}_t - \eta_t \mathbf{g}_t$  //  $\eta_t$  is the step size
4    $\mathbf{w}_{t+1} \leftarrow P_C(\mathbf{w}_{t+1})$  // project back to the constraint
```

---

- $C = \mathbb{R}^d$ : reduces to gradient descent
- Motivation from L-smoothness:

$$\begin{aligned} f(\mathbf{w}) &\leq f(\mathbf{w}_t) + \langle \mathbf{w} - \mathbf{w}_t, \nabla f(\mathbf{w}_t) \rangle + \frac{1}{2\eta_t} \|\mathbf{w} - \mathbf{w}_t\|_2^2 \\ &= \frac{1}{2\eta_t} \|\mathbf{w} - (\mathbf{w}_t - \eta_t \nabla f(\mathbf{w}_t))\|_2^2 + f(\mathbf{w}_t) - \frac{\eta_t}{2} \|\nabla f(\mathbf{w}_t)\|_2^2 \end{aligned}$$

---

A. A. Goldstein. "Convex programming in Hilbert space". *Bulletin of the American Mathematical Society*, vol. 70, no. 5 (1964), pp. 709–710, E. S. Levitin and B. T. Polyak. "Constrained Minimization Methods". *USSR Computational Mathematics and Mathematical Physics*, vol. 6, no. 5 (1966), pp. 1–50. [English translation in *Zh. Vychisl. Mat. mat. Fiz.* vol. 6, no. 5, pp. 787–823, 1965].



---

### Algorithm 3: Projected gradient descent for constrained smooth minimization

---

**Input:**  $\mathbf{w}_0 \in \mathbb{R}^d$ , constraint  $C \subseteq \mathbb{R}^d$ , smooth function  $f : \mathbb{R}^d \rightarrow \mathbb{R}$

```
1 for  $t = 0, 1, \dots$  do
2    $\mathbf{g}_t \leftarrow \nabla f(\mathbf{w}_t)$  // compute the gradient
3    $\mathbf{w}_{t+1} \leftarrow \mathbf{w}_t - \eta_t \mathbf{g}_t$  //  $\eta_t$  is the step size
4    $\mathbf{w}_{t+1} \leftarrow P_C(\mathbf{w}_{t+1})$  // project back to the constraint
```

---

- $C = \mathbb{R}^d$ : reduces to gradient descent
- Motivation from L-smoothness:

$$\begin{aligned} f(\mathbf{w}) &\leq f(\mathbf{w}_t) + \langle \mathbf{w} - \mathbf{w}_t, \nabla f(\mathbf{w}_t) \rangle + \frac{1}{2\eta_t} \|\mathbf{w} - \mathbf{w}_t\|_2^2 \\ &= \frac{1}{2\eta_t} \|\mathbf{w} - (\mathbf{w}_t - \eta_t \nabla f(\mathbf{w}_t))\|_2^2 + f(\mathbf{w}_t) - \frac{\eta_t}{2} \|\nabla f(\mathbf{w}_t)\|_2^2 \end{aligned}$$

---

A. A. Goldstein. "Convex programming in Hilbert space". *Bulletin of the American Mathematical Society*, vol. 70, no. 5 (1964), pp. 709–710, E. S. Levitin and B. T. Polyak. "Constrained Minimization Methods". *USSR Computational Mathematics and Mathematical Physics*, vol. 6, no. 5 (1966), pp. 1–50. [English translation in *Zh. Vychisl. Mat. mat. Fiz.* vol. 6, no. 5, pp. 787–823, 1965].

