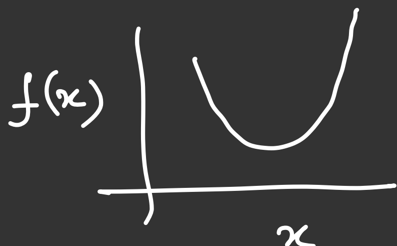
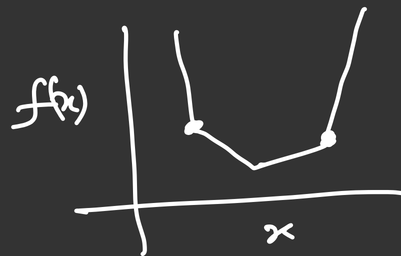
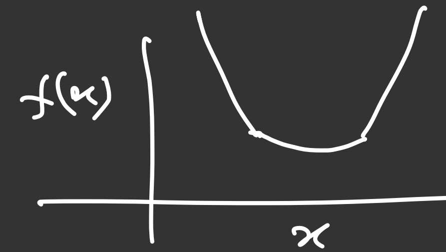
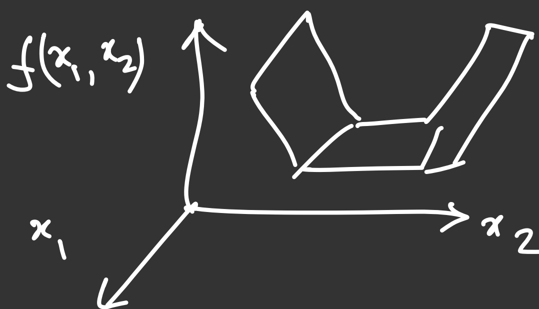
Lec. 4 (10/04/2022)

High level idea of convex function:





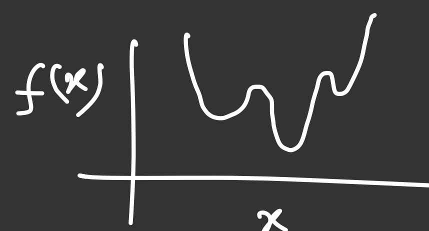


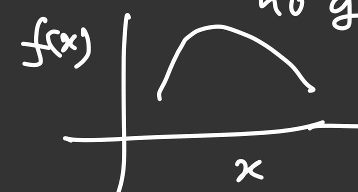


"Convex functions" $\approx$ "bowl-shaped functions"

Not convex is called "Nonconvex"

↑ no gap





Example of f where by changing $\mathrm{dom}(f)$, the same f may become convex or nonconvex :

$$f(\underline{x}) = \sum_{i=1}^{n} x_i \log x_i - (1-x_i)\log(1-x_i)$$

If $\underline{x} \in \mathrm{dom}(f) := \Delta^{n-1} \subset \mathbb{R}^n$, then f is convex

If $\underline{x} \in \mathrm{dom}(f) := [0,1]^n \subset \mathbb{R}^n$, " f " nonconvex

In other words, this f is simplex-convex.

Misconception #2: Convex $\Rightarrow$ Easy

Convex optimization problems can be "hard", e.g., NP-hard!!

Example:

$$\underbrace{\min_{X \in S^n_{copo}} \langle C, X \rangle}_{} \quad \leftarrow \text{This is a convex optimization problem}$$

$$\langle C, X \rangle = \mathrm{tr}(C^T X)$$

$$\text{s.t.} \quad \langle A_i, X \rangle = b_i \quad \text{for } i = 1, \ldots, m$$

This set is convex

$$\underbrace{S^n_{copo}}_{\text{set of } n\times n \text{ co-positive matrices}} := \left\{ X \in S^n \;\middle|\; \underline{x}^T X \underline{x} \geq 0 \;\; \forall\, \underline{x} \in \mathbb{R}^n_+ \right\}$$

$$\boxed{S^n_+ \subset S^n_{copo} \subset S^n}$$

Misconception #3: Nonconvex $\Rightarrow$ hard

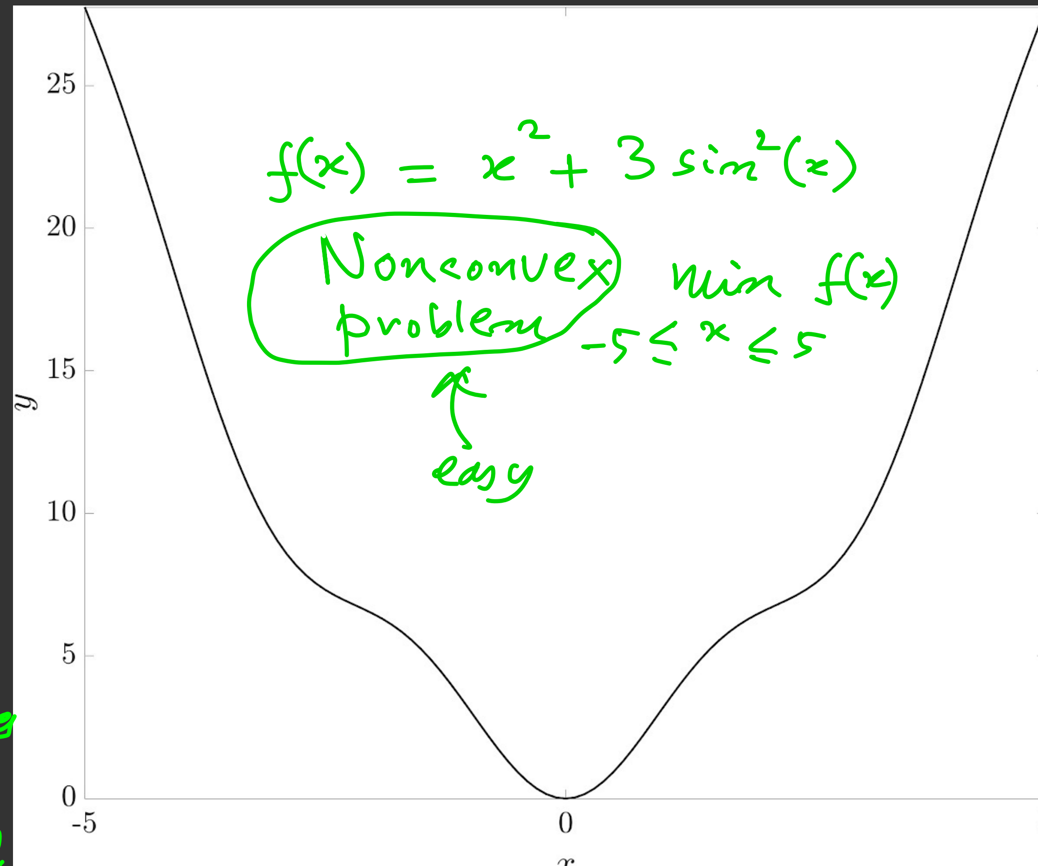
$\Leftrightarrow$ can be easy!!

Example:

This kind of function is called invex (not convex)

The reason $x=0$ is global minimizer is because $f(x) = $ sum of squares ≥ 0

But zero is achieved at $x=0$.



$$f(x) = x^2 + 3\sin^2(x)$$

Nonconvex problem

$$\min_{-5 \leq x \leq 5} f(x)$$

easy

Existence results (No convexity assumptions):

#1 Weirstrass Extreme Value Theorem:

Consider continuous $f : \underbrace{\mathcal{X}}_{\text{non-empty compact}} \subset \mathbb{R}^n \longmapsto I \subseteq \mathbb{R}$.

Such a function achieves extrema (i.e., both max & min)

Non-example:

$$\min_{\underline{x} \in \mathbb{R}^n_+} \langle \underline{a}, \underline{x} \rangle$$

does NOT exist

NOT compact!!

#2 Theorem: Let $f : \mathcal{X} \subseteq \mathbb{R}^n \mapsto I \subseteq \mathbb{R}$.

Need not be compact

Suppose f is $\boxed{\text{continuous}}$ AND 1-$\boxed{\text{coercive}}$.

Then, the global minimizer for $\min_{\underline{x} \in \mathcal{X}} f(\underline{x})$

exists (but may not be unique)

Defn: r-coercive: A continuous function $f(\cdot)$ is called r-coercive for some integer $r \geq 0$, if

$$\lim_{\|\underline{x}\|_2 \to \infty} \frac{f(\underline{x})}{\|\underline{x}\|_2^r} = +\infty.$$

0 - coercive $\Longleftrightarrow$ radially unbounded

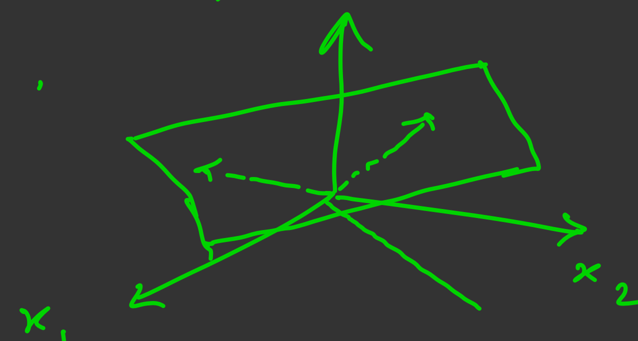
1 - coercive $\Longleftrightarrow$ superlinear

shows up in optimal control

shows up in nonlinear control/ Lyapunov function

Non-example:

Linear/affine function: $f(\underline{x}) = \langle \underline{a}, \underline{x} \rangle$ is NOT r-coercive for any $r \geq 0$ because along $\langle \underline{a}, \underline{x} \rangle = 0$, the function $f(\underline{x}) = 0 \neq +\infty$.



Example: $f(\underline{x}) = \|\underline{x}\|_2^2 = \underline{x}^T\underline{x}$

0 - coercive? $\lim_{\|\underline{x}\|_2 \to \infty} \frac{f(\underline{x})}{1} \stackrel{?}{=} +\infty$ Yes

1 - coercive? $\lim_{\|\underline{x}\|_2 \to \infty} \frac{f(\underline{x})}{\|\underline{x}\|_2} \stackrel{?}{=} +\infty$ Yes

r - coercive? $\lim_{\|\underline{x}\|_2 \to \infty} \frac{f(\underline{x})}{\|\underline{x}\|_2^2} = \lim_{\|\underline{x}\|_2 \to \infty} 1$

$r > 2$

$= 1 \neq +\infty$

No

Not r-coercive for $\boxed{r > 2}$

#3 Linear objective over compact set:

$$\underset{\underline{x} \in \mathcal{X}}{\max \text{ or } \min} \ \langle \underline{c}, \underline{x} \rangle$$

$\mathcal{X}$ ← compact

A linear objective is maximized/minimized at the boundary of $\mathcal{X}$.

i.e., not only existence is guaranteed, it will happen at (possibly non-unique) points

$$\underline{x} \in \underbrace{\partial \mathcal{X}}_{\text{boundary of } \mathcal{X}}.$$

Example:

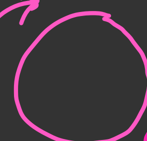
$$\max_{t\in[0,2\pi)} \underbrace{\alpha\cos t + \beta \sin t}_{\text{nonconvex in } t} \quad \left\{ \begin{array}{l} x_1 := \cos t \\ x_2 := \sin t \end{array} \right.$$

nonconvex problem

$$= \max_{(x_1,x_2)\in \underbrace{\mathcal{X} := \{\underline{x}\in\mathbb{R}^2 \mid x_1^2+x_2^2=1\}}_{\text{nonconvex compact set}}} \left\langle \begin{pmatrix}\alpha\\ \beta\end{pmatrix}, \begin{pmatrix}x_1\\ x_2\end{pmatrix} \right\rangle$$

linear objective

still nonconvex problem



circle

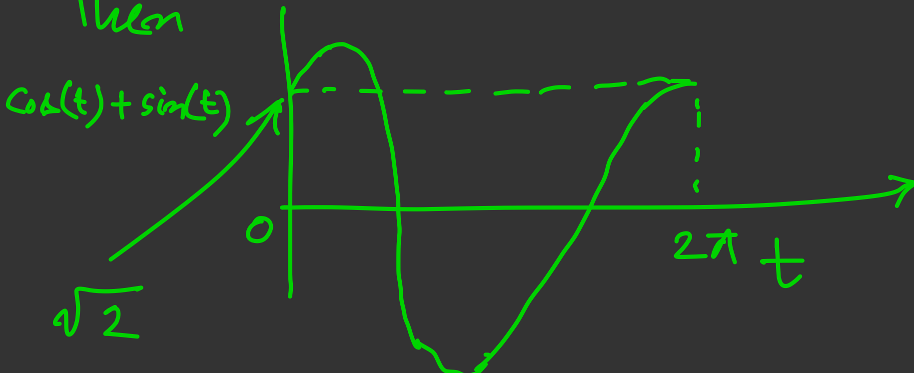
$$= \max_{(x_1,x_2)\in \underbrace{\mathrm{conv}(\mathcal{X})}_{\text{convex hull of } \mathcal{X}}} \left\langle \begin{pmatrix}\alpha\\ \beta\end{pmatrix}, \begin{pmatrix}x_1\\ x_2\end{pmatrix} \right\rangle$$

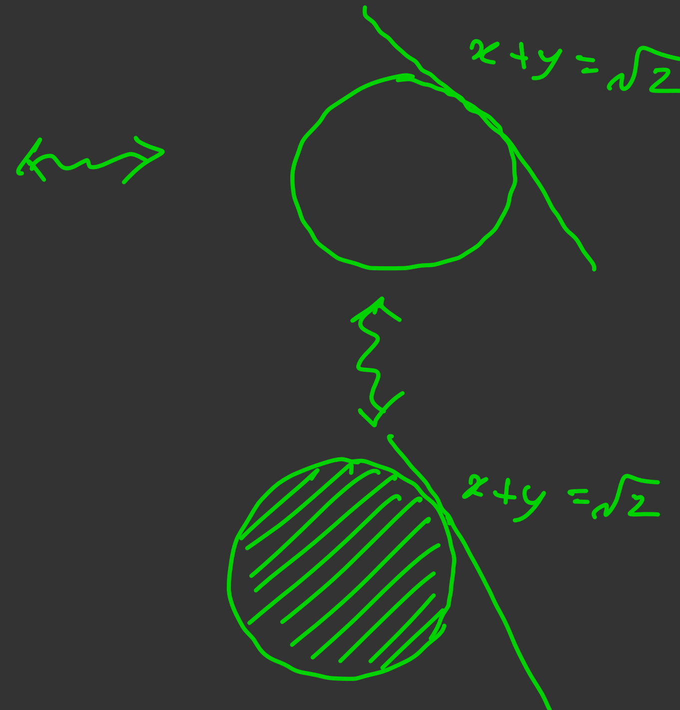
convex problem

circular disc

Suppose we fix $\begin{pmatrix}\alpha\\ \beta\end{pmatrix} = \begin{pmatrix}1\\ 1\end{pmatrix}$.

Then





Some ideas on matrix differential calculus :

$$f : \mathbb{R}^{m \times n} \longmapsto \mathbb{R}$$

(sometimes $m = n$)

Given $f(X)$,
how to compute $\underbrace{\dfrac{\partial f}{\partial X}}$

returns a matrix

Gradient (matricial)

Directional derivative :

Gradient can be extracted from directional derivative.

For vectors:

Directional derivatives: $$D_{\underline{z}} f(\underline{x}) = \lim_{h \to 0} \frac{f(\underline{x} + h\underline{z}) - f(\underline{x})}{h}$$

Directional Derivative of $f(\cdot)$ at $\underline{x}$ in the direction $\underline{z}$

$$= \langle \nabla_{\underline{x}} f(\underline{x}), \underline{z} \rangle$$

vector inner product

$$= \underbrace{\left(\nabla_{\underline{x}} f(\underline{x})\right)}^{T} \underline{z}$$

This object is $\nabla_{\underline{x}} f(\underline{x})$

$$\equiv \underbrace{\frac{\partial f}{\partial \underline{x}}}_{\text{vector gradient}}$$

You can think this as the definition of gradient

For the matrix case, we do the same:

$$\underbrace{D_Z f(X)}_{\text{Matrix directional derivative}} = \lim_{h \to 0} \frac{f(X+hZ) - f(X)}{h}$$

$$= \langle \underbrace{\nabla_X f}, Z \rangle \leftarrow \text{Frobenius/ Matrix inner product}$$

$$= \left\langle \frac{\partial f}{\partial X}, Z \right\rangle$$

$$= \mathrm{tr}\left(\left(\frac{\partial f}{\partial X}\right)^T Z\right)$$

extract this object $\leftarrow$ matricial gradient $\dfrac{\partial f}{\partial X}$.

Example: $f(X) = \text{trace}(AX), \quad X \in \mathbb{R}^{m \times n}$

constant

What is $\dfrac{\partial f}{\partial X}$?

$$D_Z f(X) = \lim_{h \to 0} \frac{f(AX + hAZ) - f(AX)}{h}$$

$$= \lim_{h \to 0} \frac{tr(AX + hAZ) - tr(AX)}{h}$$

$$= \lim_{h \to 0} \frac{\cancel{h}}{\cancel{h}} tr(AZ) = tr(AZ)$$

$$= tr((A^T)^T Z)$$

$$\boxed{\therefore \frac{\partial f}{\partial X} = A^T}.$$

__Example:__ $f(X) = \operatorname{tr}(X^{-1}), \quad X \in \underbrace{GL(n)}_{\text{invertible } n \times n}$

$$D_Z f(X) = \lim_{h \to 0} \frac{\operatorname{tr}\left((X+hZ)^{-1}\right) - \operatorname{tr}(X^{-1})}{h}$$

$$= \lim_{h \to 0} \frac{\operatorname{tr}\left[(I+hX^{-1}Z)^{-1}X^{-1}\right] - \operatorname{tr}(X^{-1})}{h}$$

$$\boxed{(1+hx)^{-1} \approx 1-hx \text{ for } h \text{ small}}$$

$$= \lim_{h \to 0} \frac{\operatorname{tr}\left[(I-hX^{-1}Z)X^{-1}\right] - \operatorname{tr}(X^{-1})}{h}$$

$$= \lim_{h \to 0} \frac{\operatorname{tr}\left[X^{-1} - hX^{-1}ZX^{-1}\right] - \operatorname{tr}(X^{-1})}{h}$$

$$= \lim_{h \to 0} \frac{\cancel{tr(X^{-1})} - h\, tr(X^{-1} Z X^{-1}) - \cancel{tr(X^{-1})}}{\cancel{h}}$$

$$= -tr(X^{-1} Z X^{-1})$$

because $tr(\cdot)$ is invariant under cyclic permutation:

$$tr(ABC) = tr(CAB) = tr(BCA)$$

$$= -tr(X^{-2} Z)$$

$$= tr\left(\left((-X^{-2})^T\right)^T Z\right)$$

$$\boxed{\therefore \frac{\partial f}{\partial X} = -(X^{-2})^T}$$