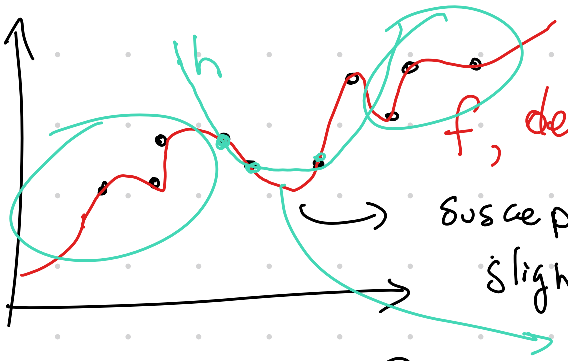
Consider that we have a set of n points.

$P = \{(x_1,y_1), \dots, (x_n,y_n)\}$, $|P| = n$.

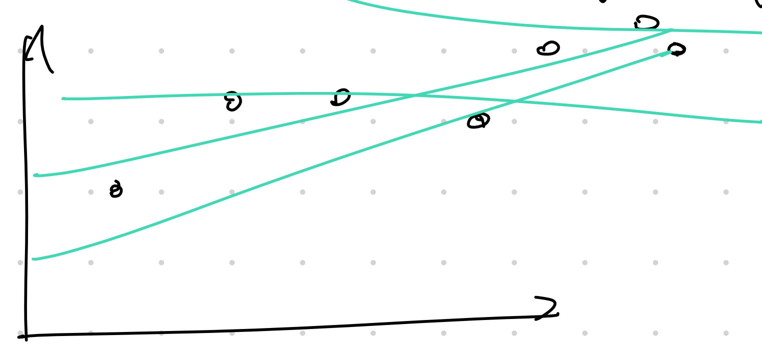


we should lower the degree of this poly.

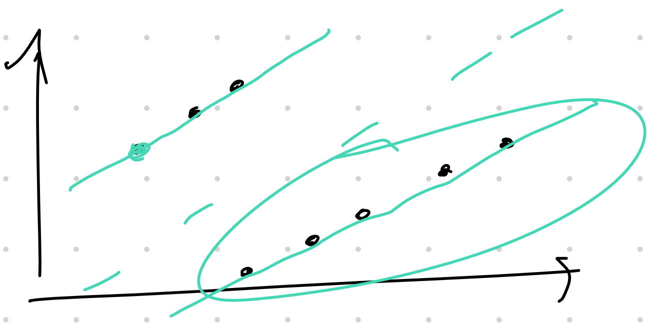
susceptible to slight changes.

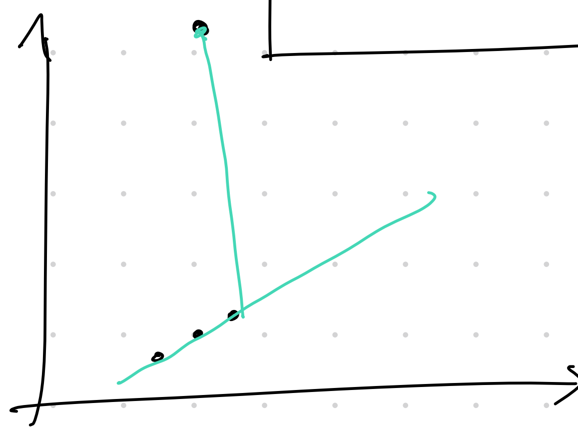
deg h = 3

What is the best lower degree polynomial.



What is a good line?





$y = mx + b$ is a line.

$P = \{(x_1,y_1), \dots, (x_n,y_n)\}$.

$$C(y,P) = \sum_{i=1}^{n} (mx_i + b - y_i)^2$$

sum of the square of the distance to the data.

if the best line exists, then it minimizes the cost function.

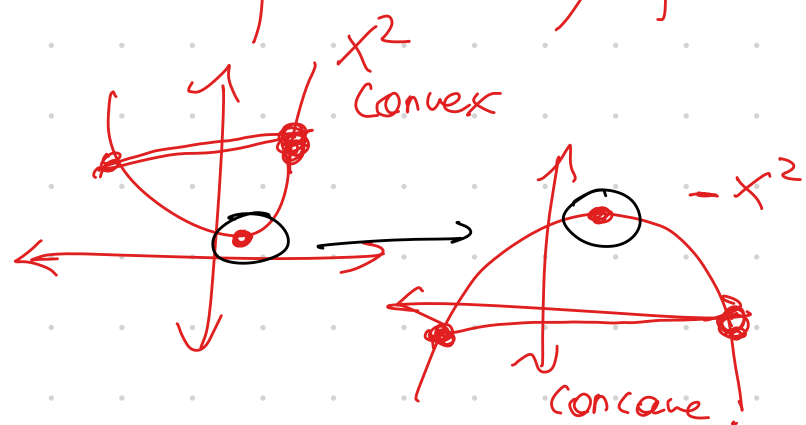
given $f : \mathbb{R} \to \mathbb{R}$, we first find critical points.

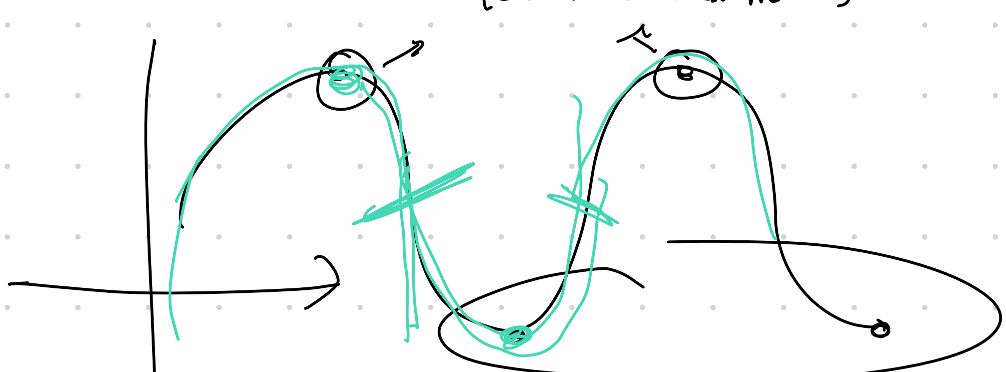
a.k.a. $\frac{df}{dx} = 0$ (critical points).

and a critical point x is a minimum if

$$\left.\frac{d^2f}{dx^2}\right|_x > 0 \qquad \left(\frac{d^2f}{dx^2} < 0\right) \text{ (maximum)}.$$

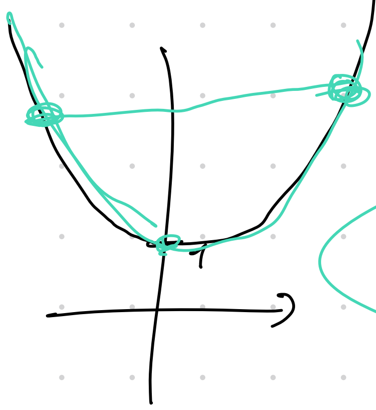
local optimization.





When is a function is globally optimizable?

When $\min_{x \in \mathbb{R}} f(x)$ has a unique minimizer.

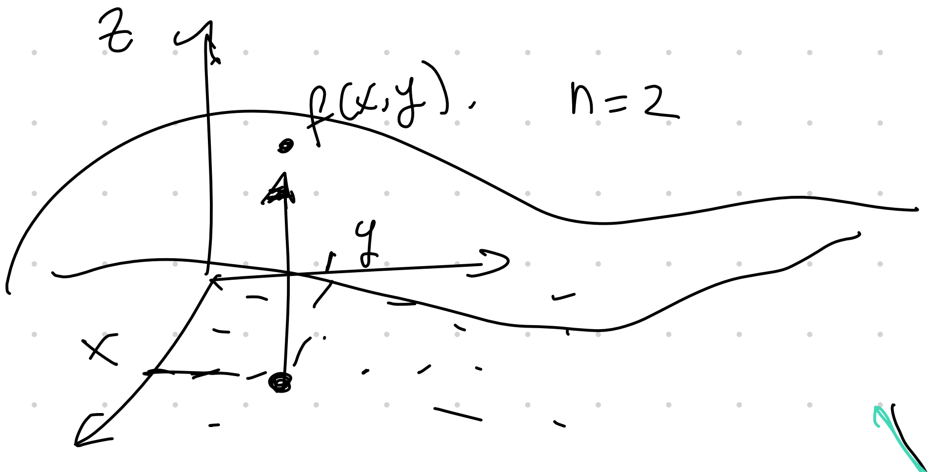


$$\cosh(x) = \frac{e^x + e^{-x}}{2}$$

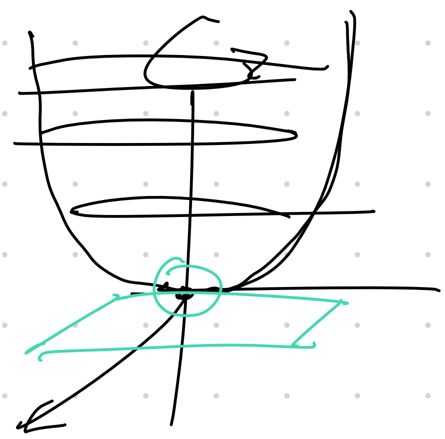
$$\frac{d^2 \cosh(x)}{dx^2} = \cosh(x) > 0.$$

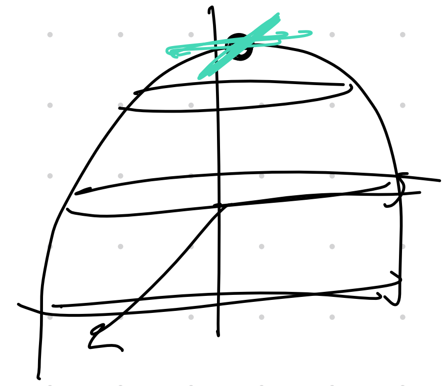
If f is globally concave (convex) and $\frac{df}{dx}(x_0) = 0$, then x_0 is a global maximum (minimum).

$f : \mathbb{R}^n \to \mathbb{R}$. (scalar Field.)

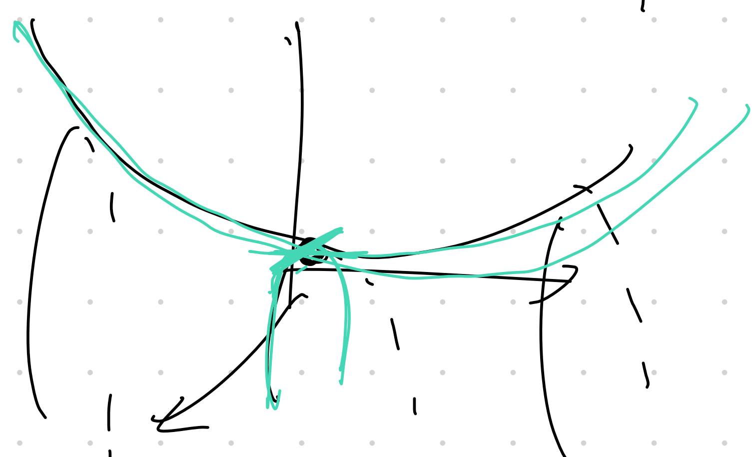


$n = 2$





We first Find critical points.



$f : \mathbb{R}^n \to \mathbb{R}$ ✓

$(x_1, x_2, \ldots, x_n) \mapsto f(x_1, x_2, \ldots, x_n)$.

$X = (x_1, \ldots, x_n)$ is a critical point of f iF

$$\frac{\partial f}{\partial x_j}(X) = 0, \quad \text{For all } j = 1, \ldots, n.$$

$\boxed{\nabla_x f}$ gradient of f. $\nabla_x f : \mathbb{R}^n \to \mathbb{R}^n$.

$$f(x,y) = 2(x^2+y^2) \;;\; \underbrace{\frac{\partial f}{\partial x} = 4x \;;\; \frac{\partial f}{\partial y} = 4y}_{\text{First-order derivatives.}}$$

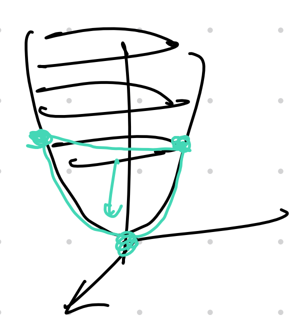
$$\frac{\partial^2 f}{\partial x^2} = 4, \quad \frac{\partial^2 f}{\partial y^2} = 4, \quad \frac{\partial^2 f}{\partial x \partial y} = 0 = \frac{\partial^2 f}{\partial y \partial x}.$$

$$f(x,y) = 2(x^2+y^2), \quad \nabla_x f = (4x, 4y), \quad \underbrace{\nabla_x^2 f}_{H_x(f)} = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$$

hessian matrix

$f : \mathbb{R}^n \to \mathbb{R}$,

$$\nabla_x f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n} \right)$$



$$\nabla_x^2 f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^3 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix} \in M_{n\times n}(\mathbb{R}).$$

<u>Definition</u> (Positive definite matrix): $A \in M_{n\times n}(\mathbb{R})$ is positive definite if

$$\forall \xi \in \mathbb{R}^n ; \quad \xi^T A \xi \geq 0.$$

$n=1, \quad A = r$

$\forall x \in \mathbb{R}, \quad x r x = x^2 r$, ≥ 0

$r \geq 0.$

(negative definite)

$$\forall \xi \in \mathbb{R}^n : \quad \boxed{\xi^T A \xi \leq 0.}$$

(semi-definite) $\xi^T A \xi \underset{(\leq 0)}{\geq} 0$ and $\exists \xi \neq 0$, $\xi \in \mathbb{R}^n$, $\xi^T A \xi = 0$.

$f: \mathbb{R}^n \to \mathbb{R}$ is ~~concave~~ convex at $x_0 \in \mathbb{R}^n$ if

$$\nabla_x^2 f(x_0) < 0$$

$>$ negative definite.

$f(x,y) = 2(x^2+y^2) \qquad \nabla_x^2 f = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix} \to$ is positive definite.

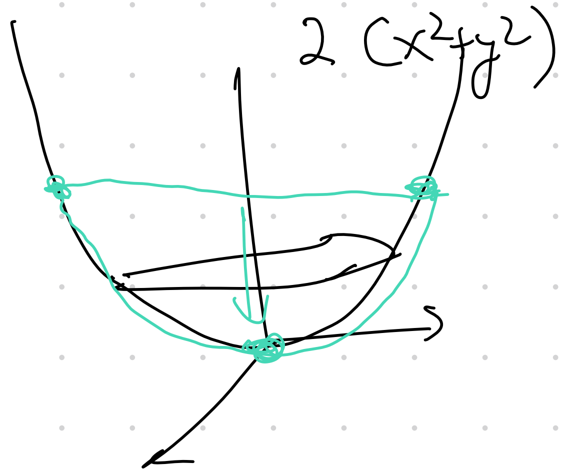
$\forall \xi \in \mathbb{R}^2, \quad \xi = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} ; \quad \xi^T (\nabla_x^2 f) \xi.$

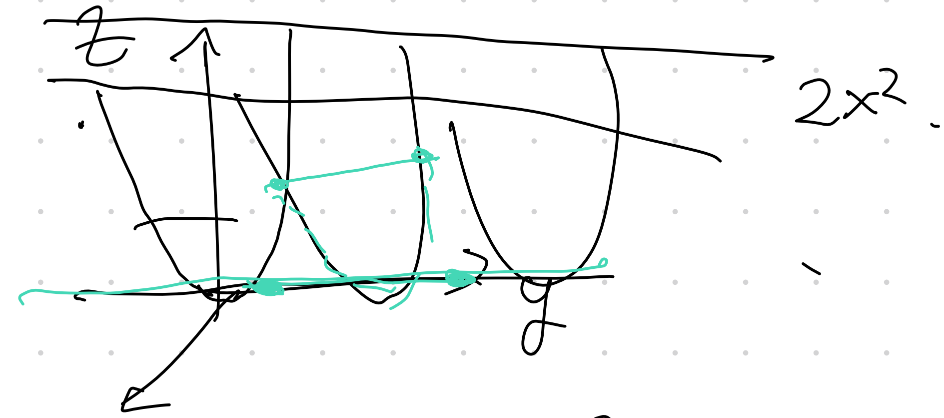
$\xi^T = (x_1, x_2) \qquad \xi^T \nabla_x^2 f \xi = (x_1, x_2) \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

$$= (x_1, x_2) \cdot \begin{pmatrix} 4x_1 \\ 4x_2 \end{pmatrix} = 4x_1^2 + 4x_2^2 = 4(x_1^2 + x_2^2) \geq 0.$$

$f(x,y) = 2x^2, \qquad \nabla_x^2 f = \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix}. \qquad x_2 \in \mathbb{R}. \quad (0, x_2)$

$$(x_1\ x_2) \begin{pmatrix} 4 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (x_1\ x_2) \begin{pmatrix} 4x_1 \\ 0 \end{pmatrix} = 4x_1^2 \geq 0$$





$$\nabla_x^2 f = 0$$

$$f: \mathbb{R} \to \mathbb{R}, \quad \frac{\partial^2 f}{\partial x^2} = 0 \longrightarrow f(x) = mx + b. \to \text{line}$$

$$f: \mathbb{R}^n \to \mathbb{R}, \quad \nabla_x^2 f = 0. \longrightarrow f(x) = C^T X \to \text{plane.}$$

$y = mx+b$ is a line.

$$P = \{(x_1, y_1), \dots, (x_n, y_n)\}.$$

$$\underline{C(y, P)} = \underline{\sum_{i=1}^{n} (mx_i + b - y_i)^2}$$

$$P = \{(x_{11}, x_{12}, x_{13}, \dots, x_{1m}, y_1), \dots, (x_{n1}, x_{n2}, x_{n3}, \dots, x_{nm}, y_n)\}.$$

$$= \{(x_i, y_i) : x_i \in \mathbb{R}^m, y_i \in \mathbb{R}\}.$$

$$\underline{y(x)} = \underline{a_0 + a_1 x_1} + a_2 x_2 + a_3 x_3 + \dots + a_m x_m.$$

$$C(y, P) = \sum_{i=1}^{n} (\underbrace{a_0 + a_1 x_{i1} + \dots + a_m x_{im}}_{\text{predicted value}} - y_i)^2$$

$$y(x) = \underbrace{(a_0, a_1, \dots, a_m)}_{a^T} \cdot \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} = a_0 + a_1 x_1 + \dots + a_m x_m.$$

$$; \quad X_i = (1, x_i)$$

$$C(y, P) = \sum_{i=1}^{n} (a^T \underline{X_i} - y_i)^2$$

$$v = (x_1, \dots, x_n)$$

$$\underline{v^T} v = (\underline{x_1}, \dots, x_n) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

$$= x_1^2 + x_2^2 + \dots + x_n^2$$

$$\underline{y(x) = a_0 + a_1 x_1 + \dots + a_n x_n}.$$

$$X = \begin{pmatrix} 1 & X_{11} & X_{12} & X_{13} & \cdots & X_{1m} \\ 1 & X_{21} & X_{22} & X_{23} & \cdots & X_{2m} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & X_{n3} & \cdots & X_{nm} \end{pmatrix}_{n\times(m+1)} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{pmatrix} = \begin{pmatrix} a_0 + a_1 X_{11} + \cdots + a_m X_{1m} \\ a_0 + a_1 X_{21} + \cdots + \cdots \\ \vdots \\ a_0 + a_1 X_{n1} + \cdots + \cdots \end{pmatrix}$$

Xa, $y = (y_1, y_2, \ldots, y_n)$

$(A+B)^T = A^T + B^T$

$a = (a_0, a_1, \ldots, a_m)$ ← these are the variables.

$(AB)^T = B^T A^T$

they are the same

$$C(y,P) = (Xa - y)^T (Xa - y) = (a^T X^T - y^T)(Xa - y)$$

$$= a^T X^T X a - a^T X^T y - y^T X a + y^T y$$

$$= a^T X^T X a - 2 a^T X^T y + y^T y.$$

Einstein's notation.

$$a X^T X a = \sum_{i,j,k}^{n} a_i X^T_{ij} X_{jk} a_k$$

$$\simeq a_i X^T_{ij} X_{jk} a_k$$

$$\frac{\partial C}{\partial a_l} = \frac{\partial}{\partial a_l}\left(a_i X^T_{ij} X_{jk} a_k - 2 a_i X^T_{ij} y_j + \cancel{y_j y_j} \right)$$

$$\frac{\partial}{\partial a_l}\left(a_i X^T_{ij} X_{jk} a_k \right) = \delta_{il} X^T_{ij} X_{jk} a_k + a_i X^T_{ij} X_{jk} \delta_{kl}$$

$$\boxed{\frac{\partial a_i}{\partial a_l}} = \begin{cases} 0 & ; i \neq l \\ 1 & ; i = l \end{cases} = \delta_{il}$$ Kronecker's delta.

$i = 0, 1, 2, \ldots, m.$

$$\frac{\partial C}{\partial a_l} = X^T_{lj} X_{jk} a_k + a_i X^T_{ij} X_{jl} - 2 X^T_{lj} y_j$$

$$a_i X^T_{ij} X_{jl} = a_i X^T_{ij} X^T_{lj} = a_i X_{ji} X^T_{lj}$$

$$= X_{\ell j}^T \cdot X_{ji} \cdot a_i .$$

$$\frac{\partial C}{\partial a_\ell} = 2 X_{\ell j}^T X_{jk} a_k - 2 X_{\ell j}^T y_j , \quad \ell = 0, 1, 2, \dots, m.$$

$$\nabla_a C = 2X^T X a - 2X^T y = \vec{0}$$

$$X^T X a = X^T y$$

$$\longrightarrow \boxed{a = (X^T X)^{-1} X^T y .}$$

$$\frac{\partial^2 C}{\partial a_\ell \partial a_s} = 2 X_{\ell j}^T X_{js} ; \quad \nabla_a^2 C = 2 X^T X$$

$$\nabla_a^2 C \in M_{(m+1)\times(m+1)}(\mathbb{R})$$

$$\forall \xi \in \mathbb{R}^{m+1} : \quad \xi^T (2X^T X) \xi = 2 \xi^T X^T X \xi .$$

$$= 2 (X\xi)^T \cdot (X\xi) \geq 0 ,$$

$\operatorname{Ker} X \neq \{0\}$.

positive semi-definite

C is convex

$a^* = (X^T X)^{-1} X^T y$ is the global minimizer of C.