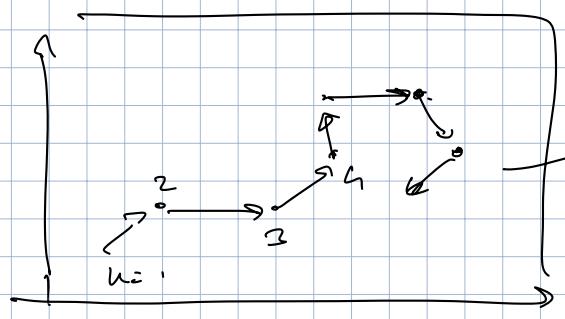
$p(\vec{x})$ important où $f(\vec{x})$ est importante

$$I = \frac{1}{V}\int_V d^D x \; f(\vec{x}) = \frac{N}{V}\int d^D x \; \frac{f(\vec{x})}{p(\vec{x})}\,\frac{p(\vec{x})}{N} = \langle g \rangle_{p/N}$$

$g(\vec{x})$

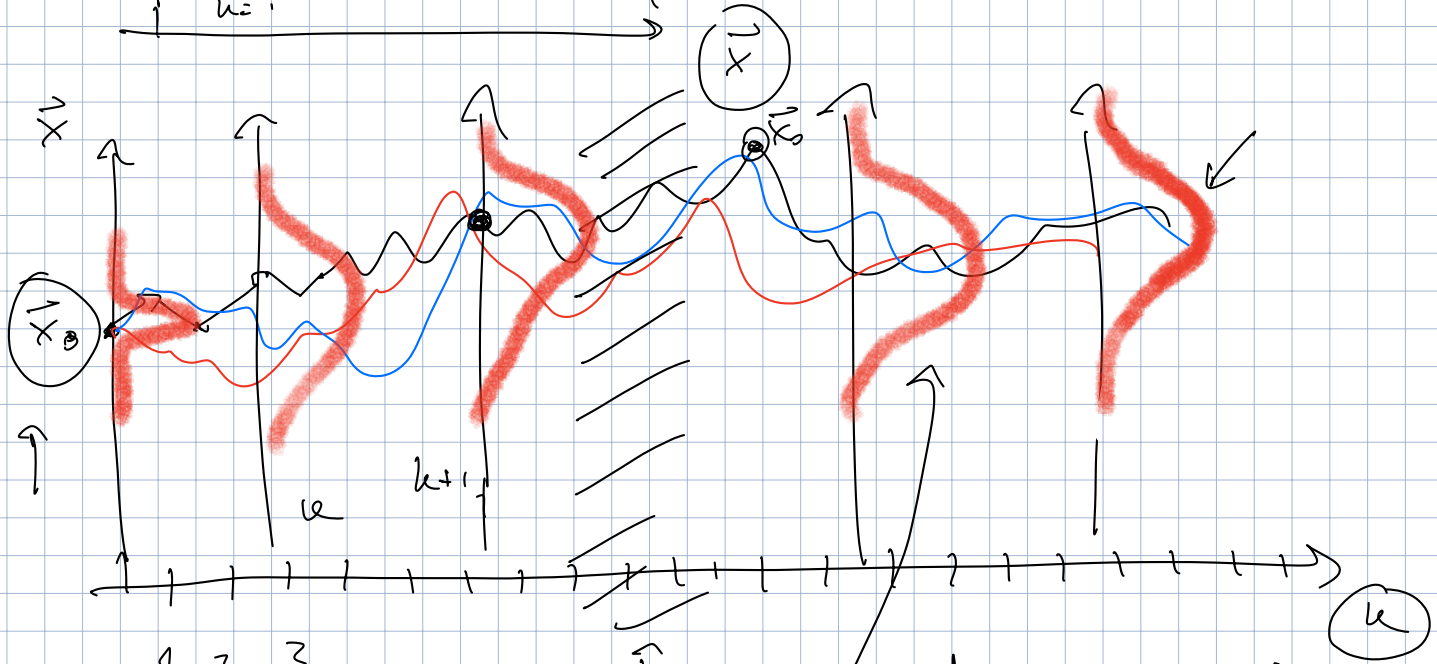
$$p(\vec{x}) = |f(\vec{x})| + \epsilon$$

MARKOV - CHAIN MONTE CARLO



$T(\vec{x} \to \vec{\eta})$

transition probability



1 2 3

equilibration

$\bar{K}$

$P(\vec{x})$

stationary regime

$P_k(\vec{x}; \vec{x}_0) \to$ $k \geq \bar{K}$

k

independent of k and $\vec{x}_0$

Build $T(\vec{x} \to \vec{\eta})$: $P(\vec{x}) = \frac{p(\vec{x})}{N}$

$$P_{\mu+1}(\vec{x}; \vec{x}_0) - P_\mu(\vec{x}; \vec{x}_0) = \frac{dP_\mu}{dt}$$

$$= \sum_{\vec{y}} \Big(P_\mu(\vec{y}; \vec{x}_0)\, T(\vec{y} \to \vec{x}) - P_\mu(\vec{x}; \vec{x}_0)\, T(\vec{x} \to \vec{y}) \Big) = 0$$

$$\sum_{\vec{y}} \Big[p(\vec{x})\, T(\vec{x} \to \vec{y}) - p(\vec{y})\, T(\vec{y} \to \vec{x}) \Big] = 0$$

$\forall\ \vec{y}$

$$p(\vec{x})\, T(\vec{x} \to \vec{y}) = p(\vec{y})\, T(\vec{y} \to \vec{x})$$

detailed balance condition

$$T(\vec{x} \to \vec{y}) = T_{prop}(\vec{x} \to \vec{y})\, A(\vec{x} \to \vec{y})$$

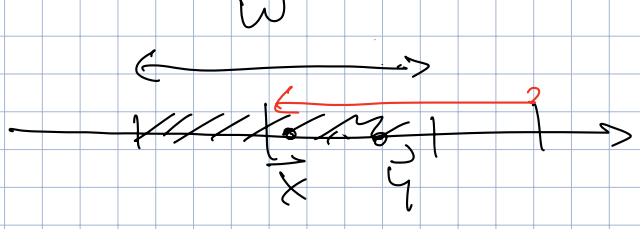
A depends on p

T_{prop}: proposal probability

A: acceptance probability

(purely geometrical factor)

ex.:

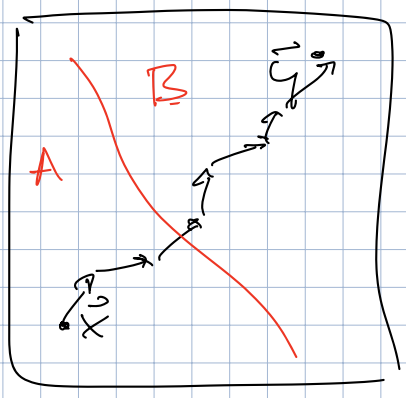


$\vec{y} \in$ interval of width W around $\vec{x}$

$$T_{prop}(\vec{x} \to \vec{y}) = T_{prop}(\vec{y} \to \vec{x})$$

symmetric

ergodicity :



poss. of going from any $\vec{x}$ to any $\vec{y}$ in a finite # of steps should be finite

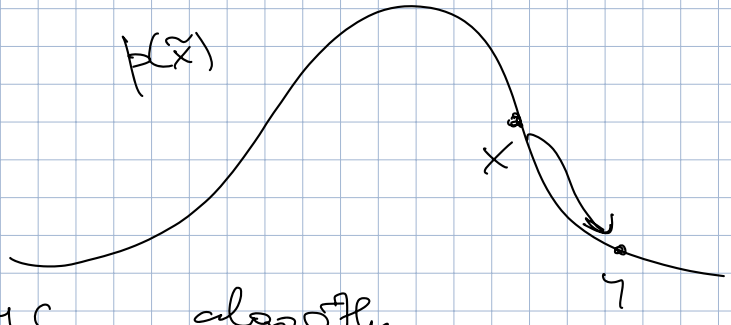
(if $p(\vec{x})$ and $p(\vec{y}) \neq 0$)

$$T_{prop}(\vec{x} \to \vec{y})\, A(\vec{x} \to \vec{y})\, p(\vec{x}) = T_{prop}(\vec{y} \to \vec{x})\, A(\vec{y} \to \vec{x})\, p(\vec{y})$$

Solution: Metropolis - Hastings algorithm

$$A(\vec{x} \to \vec{y}) = \min\left(1, \frac{p(\vec{y})}{p(\vec{x})}\right)$$

acceptance probability



MCMC algorithm

→ pick $\vec{x}_0$ at random

* → with $T_{prop}(\vec{x}_0 \to \vec{y})$ you pick a point $\vec{y}$

→ extract $z \in [0, 1]$ (random #)

if $z < A(\vec{x}_0 \to \vec{y})$ ⇒ $\vec{x}_1 = \vec{y}$

otherwise ⇒ $\vec{x}_1 = \vec{x}_0$

→ goto *

record the values $g_n = g(\vec{x}_n)$ only for $n > \bar{n}$

$\vec{x}_0 \to \vec{x}_1 \to \vec{x}_2 \to \dots \to \vec{x}(\bar{n}) \to \dots \to \vec{x}_n \to \dots \to \vec{x}_{\bar{n}+M}$

↑ stationary regime

$$I_M = \frac{1}{M} \sum_{n=\bar{n}}^{\bar{n}+M} g(\vec{x}_n) \xrightarrow[M\to\infty]{} I$$

$|I - I_M| \sim ?$

I_M is a random variable, sum of $g(\vec{x}_n) = g_n$ random variables

$n > \bar{n}$

since $\vec{x}_n$ is distributed according to $p(\vec{x})/N$

all $\vec{x}_n$'s $\forall\, n > \bar{n}$ are equally distributed

g_n's $\quad n \quad n \quad n$

$$\overline{g_n} = I = \int d^D x \, \frac{p(\vec{x})}{N} \, g(\vec{x}) = \langle g \rangle_{p/N}$$

$$\sigma_g^2 = \langle g^2 \rangle - \langle g \rangle^2$$

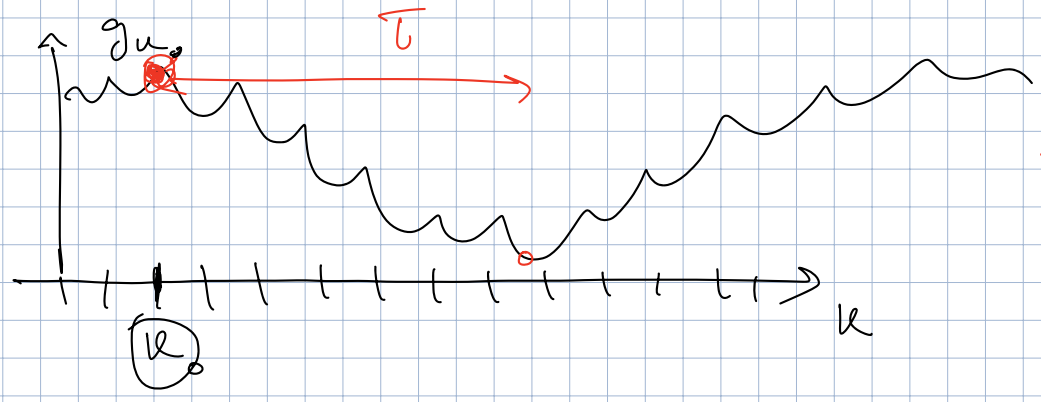
$$\overline{I_M} = \frac{1}{M} \sum_{n=\bar{n}}^{\bar{n}+M} \overline{g_n} = I$$

$$= \frac{1}{M} \sum_n \left(\overline{g_n} \right) I$$

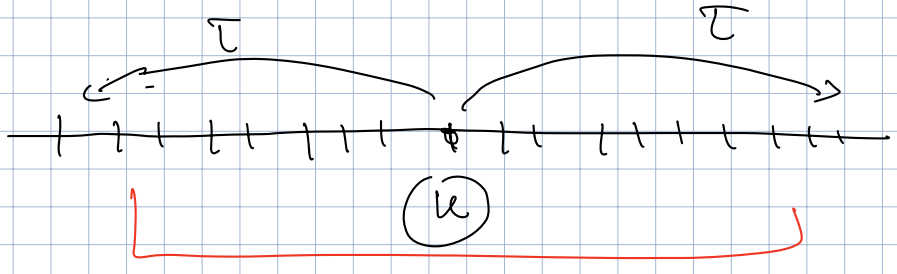
$$|I_M - I| \simeq \sigma_{I_M}$$

$\sigma^2_{I_M}$ $\overset{?}{=}$ $\frac{\sigma_g^2}{M}$ ✗ central limit theorem

are g_n's independent ? NO



Estimate τ : argument

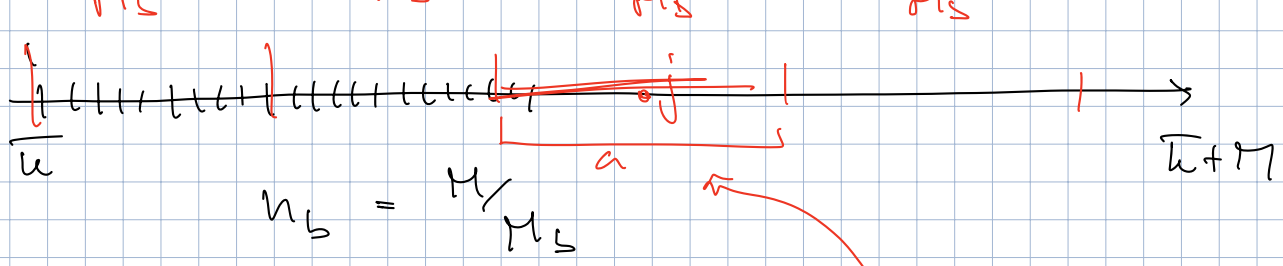


out of M steps (M values of g_n) only

$M_{eff} = \frac{M}{2\tau}$ values are independent

idea : $\sigma^2_{I_M} = \frac{\sigma_g^2}{(M/2\tau)} = (2\tau)\frac{\sigma_g^2}{M}$

way to estimate properly $\sigma^2_{I_M}$: "blocking"



$n_b = M/M_b$

$$I_M = \frac{1}{M} \sum_{u=\bar{u}+1}^{M+\bar{u}} g_u$$

$$= \frac{1}{n_b} \sum_{a=1}^{n_b} \frac{1}{M_b} \sum_{j=1}^{M_b} g_{\bar{u}+(a-1)M_b+j} = \frac{1}{n_b} \sum_a \bar{g}_a$$

$\bar{g}_a$ — block average

If $M_b \geq \tau$ ←

$\bar{g}_a$ are independent variables

⇒ apply CLT to

$$\sigma^2_{I_M} = \frac{\sigma^2_{\bar{g}}}{n_b} = (2\tau)\frac{\sigma^2_g}{M}$$

$$\tau = \frac{1}{2} M_b \frac{\sigma^2_{\bar{g}}}{\sigma^2_g}$$

estimate

provided that $\tau \leq M_b$

Result

$$\epsilon = |I_M - \bar{I}| \simeq \sigma_{I_M} \simeq \sqrt{\frac{2\tau}{M}}\,\sigma_g \sim O\left(\frac{1}{\sqrt{M}}\right)$$

Scaling ? when changing D, L

$$V = L^D$$

τ generically scales as: $\tau \simeq A^{z(1)} D\, L^z$

typically $z \sim O(1)$

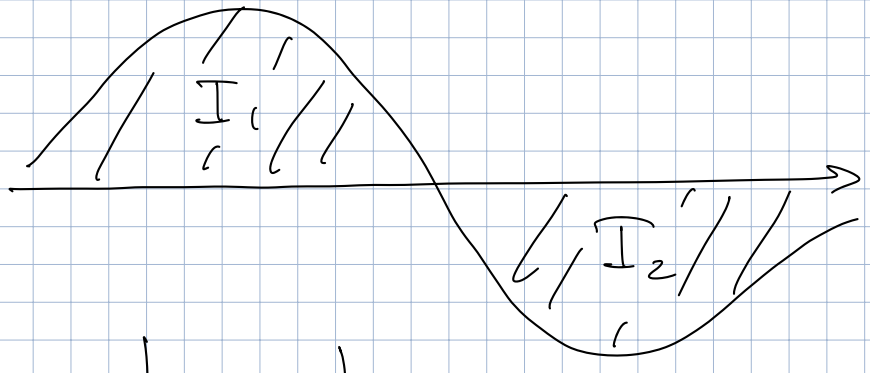
(typically $z = 2$)

$$M_\varepsilon \sim \frac{1}{\varepsilon^2} \, 2\tau$$ ✓

↑

But not always :

1) Sign problem



$$I = I_1 - I_2$$

$I_1, I_2 \simeq \mathcal{O}(\exp(D))$ but $I \sim \text{poly}(D)$

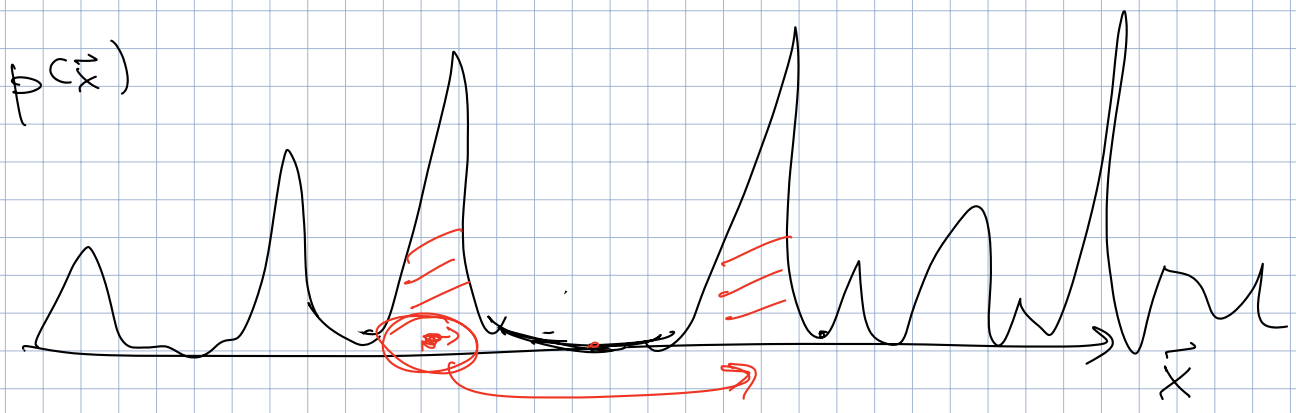
$$\sigma_{I_M} \sim \frac{\mathcal{O}(\exp(D))}{\sqrt{M}}$$

2) ciclomatic divergence of τ

it can happen $\tau \sim \exp(\mathcal{O}(D))$

$\exp(L)$

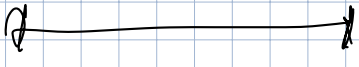
ex:



equilibrium properties of systems with disorder and competitive interactions

⇒ glass

(spin glasses)



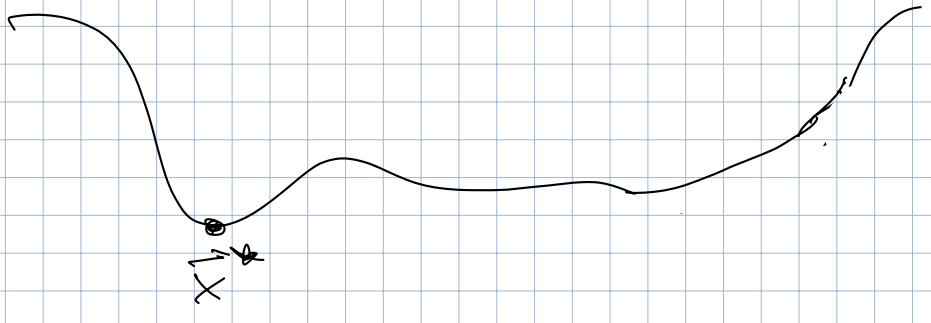
③

Numerical Optimization

$f: \vec{x} \in V \rightarrow \mathbb{R}$

$(\mathbb{R}^N)$

Find $\min_{\vec{x}} f(\vec{x}) \rightarrow \vec{x}^* = \arg\min(f(\vec{x}))$



examples in Physics

→ equilibrium configuration for N interacting classical particles

minimize $U(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N)$

potential energy

→ variational method in quantum mechanics

$\hat{H}$ Hamiltonian

guess the ground state as $|\psi(\vec{\alpha})\rangle$

$\vec{\alpha} = \alpha_1, \dots, \alpha_M$

parameters

ex. $|\psi\rangle = \alpha_1 |\psi_1\rangle + \alpha_2 |\psi_2\rangle$

$$E(\vec{\alpha}) = \frac{\langle \psi(\vec{\alpha}) | \hat{H} | \psi(\vec{\alpha}) \rangle}{\langle \psi(\vec{\alpha}) | \psi(\vec{\alpha}) \rangle}$$

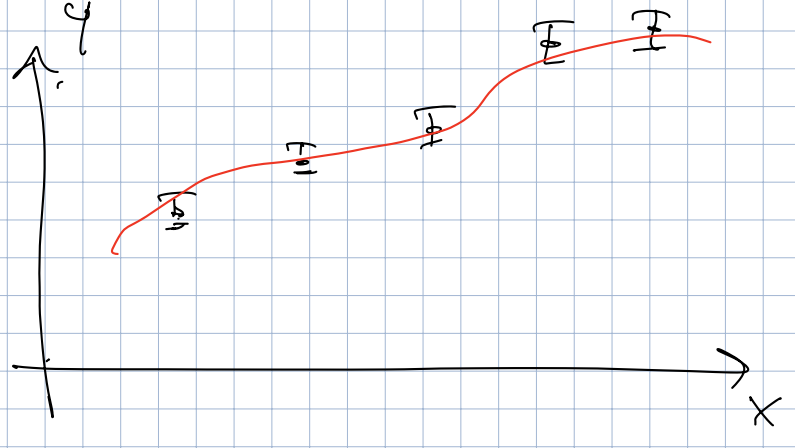
$\min_{\vec{\alpha}} E(\vec{\alpha})$

→ fit of a function to some data

→ x_1, y_1, σ_{y_1}

→ x_2, y_2, σ_{y_2}

- - -



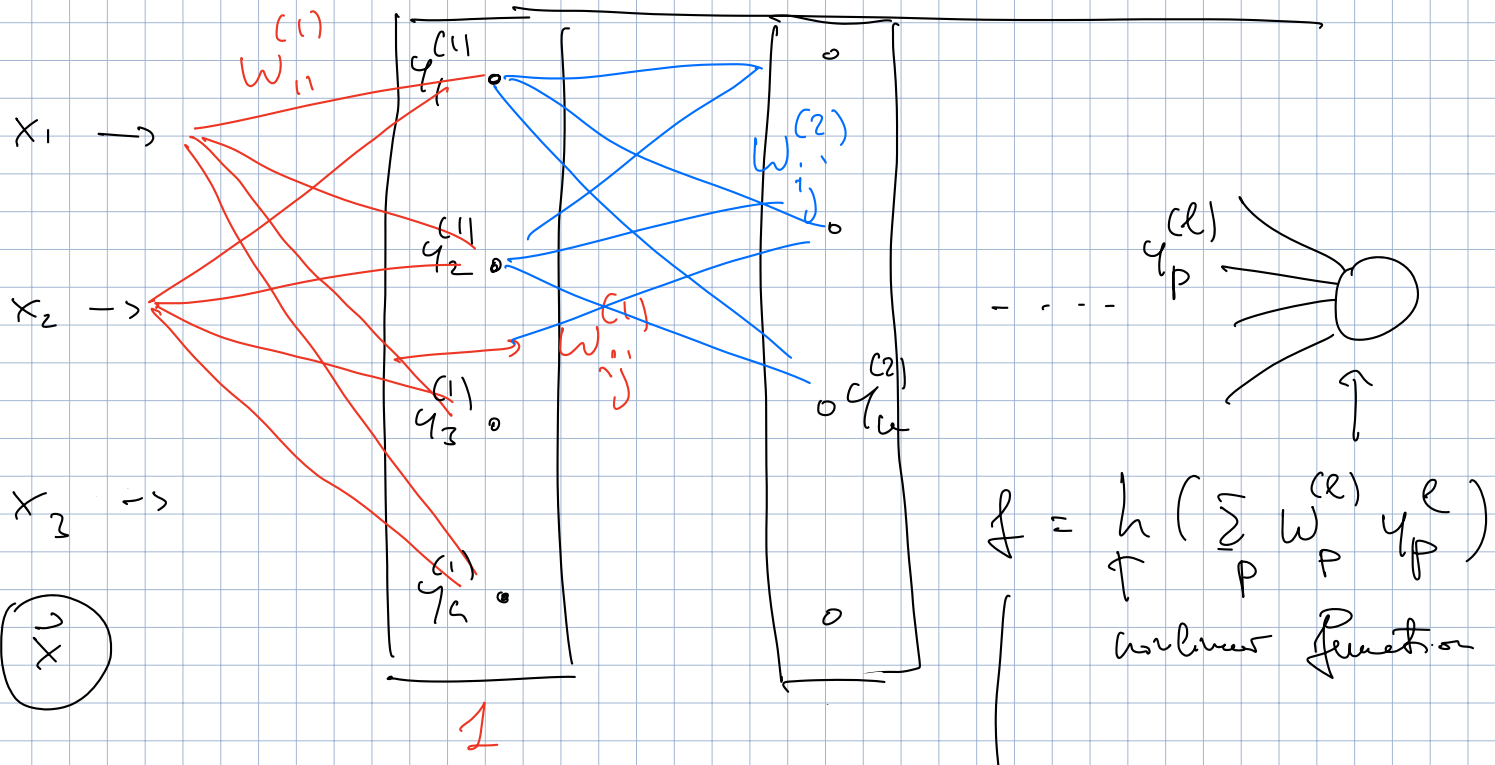
$$y = f(x; \vec{\alpha})$$

$$\chi^2(\vec{\alpha}) = \sum_u \frac{|y_u - f(x_u; \vec{\alpha})|^2}{\sigma_{y_u}^2}$$

$$\min_{\vec{\alpha}} \chi^2(\vec{\alpha})$$ — least-square method

→ example of a very complicated fit:

training a neural network



$$f = h\left(\sum_p w_p^{(l)} y_p^{l}\right)$$

nonlinear function

$$y_i^{(1)} = g\left(\sum_j w_{ij}^{(1)} x_j\right)$$

nonlinear function

$$f = f(\vec{x}; \{w_{ij}^{(m)}\})$$

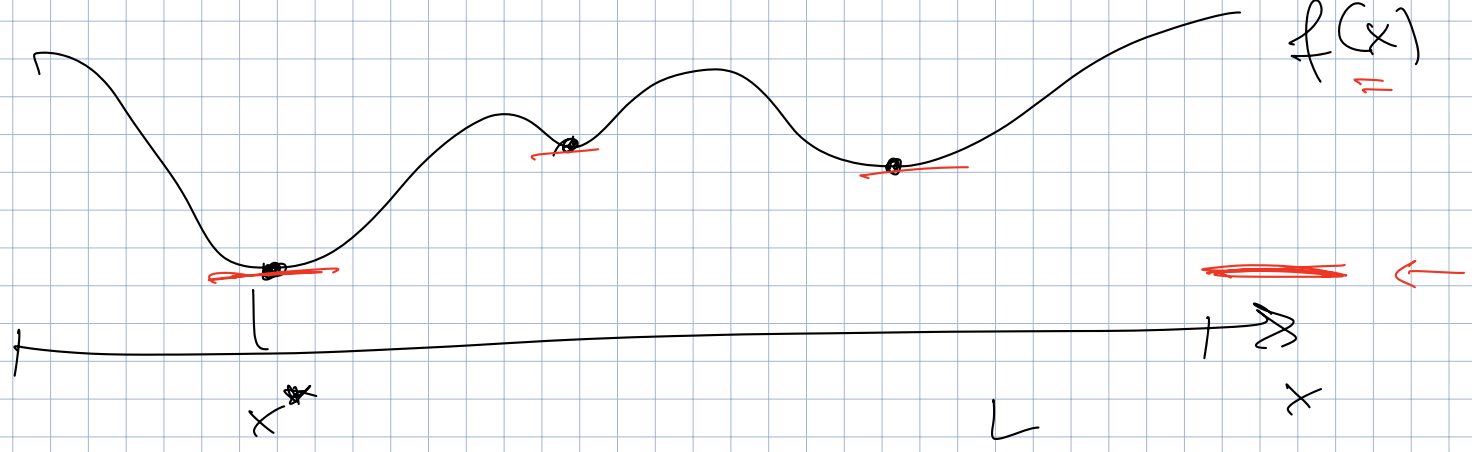
$\vec{\alpha}$

$\vec{x}_1, \vec{x}_2, \vec{x}_3 \cdots$

$y_1 \quad y_2 \quad y_3 \cdots \longrightarrow f$ ←

Numerical optimization : generic features

it is exponentially hard !



Enumeration : time (cost) to find the min

$\sim L^N$

$\vec{x} \in \mathbb{R}^N$

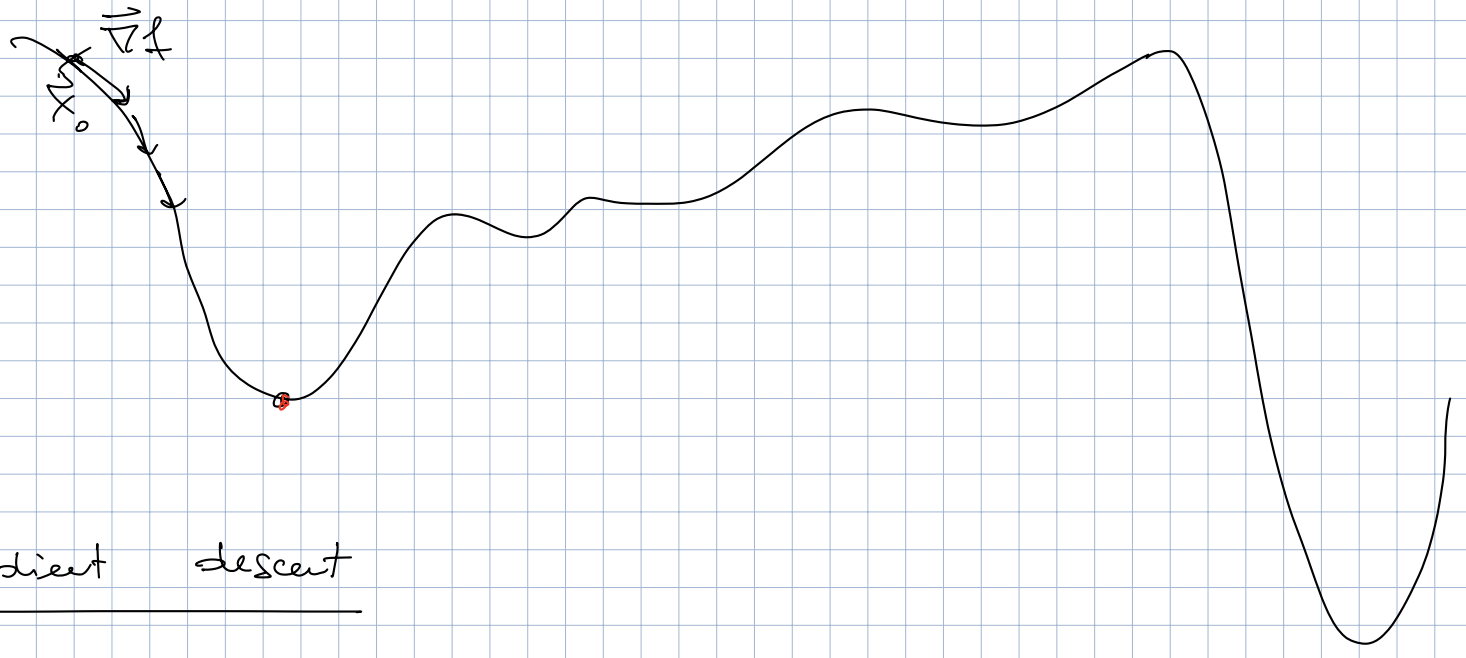
if $\vec{x}$ is a continuous variable

$\Rightarrow \vec{\nabla} f(\vec{x}) = 0$ ← compact nonlinear N V equations

problem of finding the absolute minimum is hard.

Problem : finding a local minimum

Find one $\vec{x}^*$: $\vec{\nabla} f(\vec{x}^*) = 0$ $\quad \frac{\partial^2 f}{\partial x^2} > 0$



gradient descent

$$\vec{x}_0 \to \vec{x}_1 = \vec{x}_0 - \epsilon \vec{\nabla} f(\vec{x}_0)$$

$$\vec{x}_{n+1} = \vec{x}_n - \epsilon \vec{\nabla} f(\vec{x}_n)$$

$$f(\vec{x}_{n+1}) \cong f(\vec{x}_n) - \epsilon \vec{\nabla} f(\vec{x}_n) \cdot \vec{\nabla} f(\vec{x}_n) + \mathcal{O}(\epsilon^2)$$

$$\|\vec{\nabla} f\|^2$$

$$< 0$$

How many steps do you need to hit a local minimum $\frac{2}{5}$

$$|f(\vec{x}_k) - f(\vec{x}^*)| \sim \mathcal{O}\left(\frac{1}{k}\right)$$

worst-case scenario

for some functions it can be as fast as exponential $\sim \exp(-\frac{k}{k_0})$