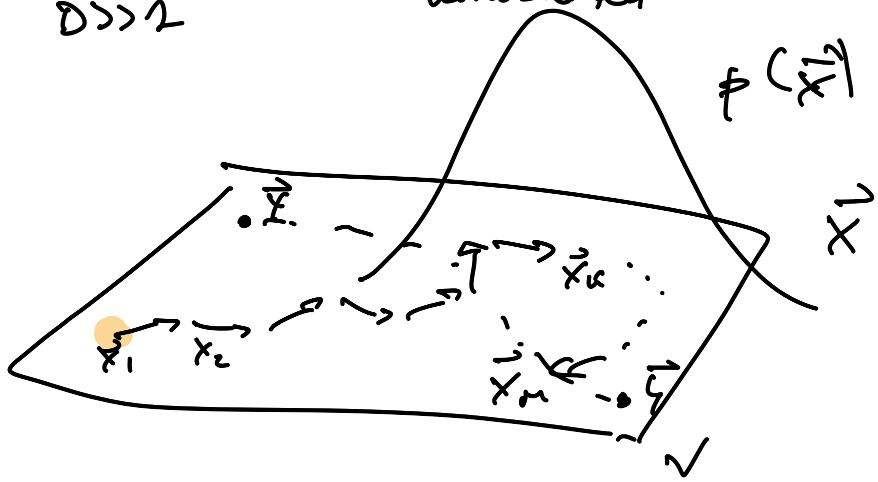
Markov-Chain Monte Carlo (MCMC)

$$I = \int_V d^D x \; p(\vec{x}) \; g(\vec{x}) = \langle g \rangle_p$$

normalized

$D \gg 1$

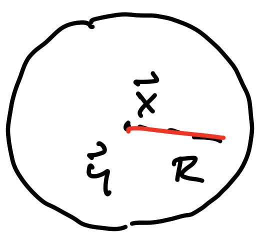


$$T(\vec{x} \to \vec{y}) = T_{prop}(\vec{x} \to \vec{y}) \; A(\vec{x} \to \vec{y})$$

proposal prob.

acceptance prob.

ex.



$$\vec{y} \in B_R(\vec{x})$$

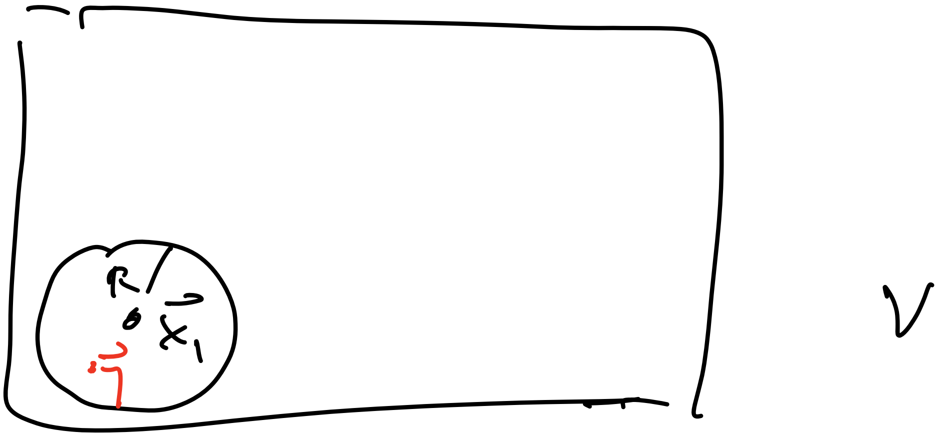
$$T_{prop}(\vec{x} \to \vec{y}) = T_{prop}(\vec{y} \to \vec{x})$$

detailed-balance condition on A

$$A(\vec{x} \to \vec{y}) = \frac{p(\vec{y})}{p(\vec{x})} \; A(\vec{y} \to \vec{x})$$

<u>solution</u> : Metropolis - Hastings algorithm

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



$z \in [0, 1]$ random number

if $z < A(\vec{x} \to \vec{y}) \Rightarrow \vec{x}_2 = \vec{y}$

otherwise $\vec{x}_2 = \vec{x}_1$

After an equilibration phase ($\sim \bar{n}$ steps)

$\vec{x}_n , \vec{x}_{n+1} , \dots , \vec{x}_{n+M}$

$n > \bar{n}$ are distributed according to $\underline{p(\vec{x})}$

$g(\vec{x}_n), g(\vec{x}_{n+1}), \dots, g(\vec{x}_{n+M})$

all distributed according to $p(\vec{x})$

$M \to \infty$

$$\underline{I_M} = \frac{1}{M} \sum_{j=k}^{k+M} g(\vec{x}_k) \simeq \langle g \rangle_p = I$$

in principle dependent

I_M is a sum of random variables

Markov chain has memory

$$\overline{I_M} = I$$

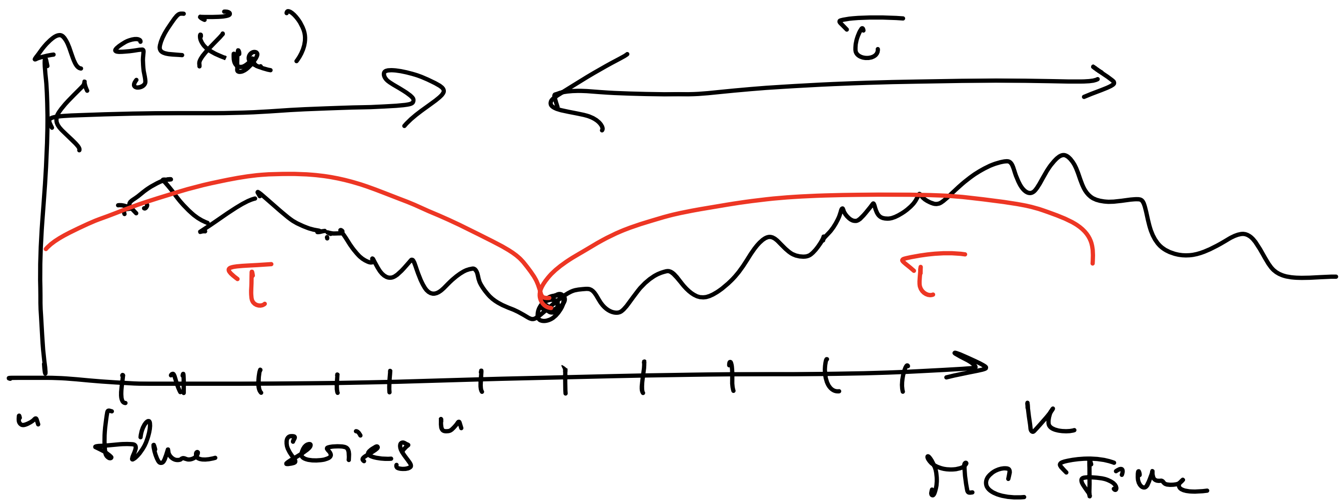
$$|I_M - I| \sim ?$$

$$\sigma^2_{I_M} = c \, \frac{\sigma_g^2}{M} \qquad \sigma_g^2 = \langle g^2 \rangle_p - \langle g \rangle_p^2$$

if $\vec{x}_k$ were independent

$c = 1$

$c > 1$ with dependent variables



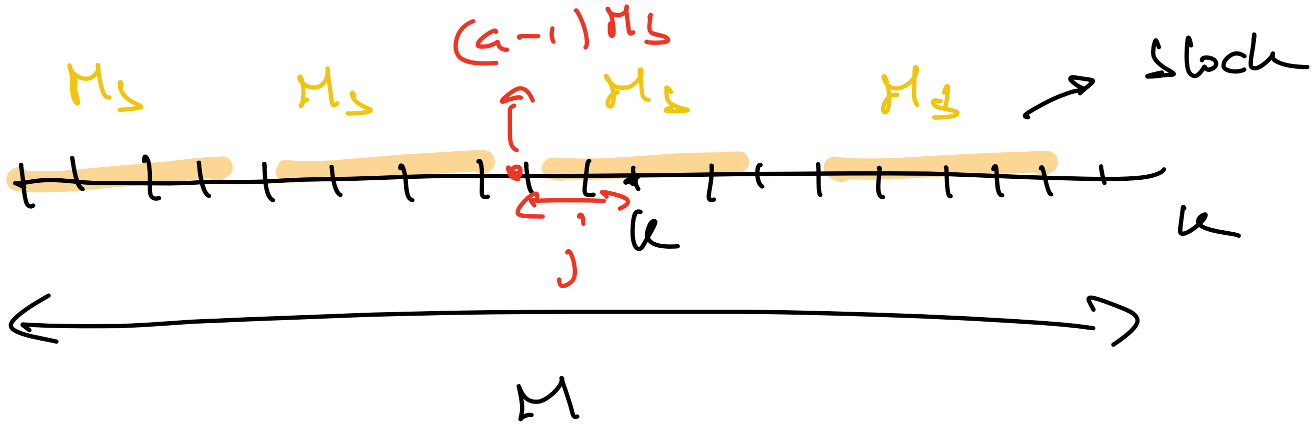
$\tau =$ autocorrelation time (memory)

M points $\rightarrow$ $M_{eff} = \frac{M}{2\tau}$ effective # of statistically independent points

<u>central limit theorem</u>

$$\sigma^2_{I_M} = \frac{\sigma_g^2}{M_{eff}} = 2\tau \frac{\sigma_g^2}{M}$$

How to estimate τ : block variable



$$n_b = \frac{M}{M_b}$$

$$I_M = \frac{1}{M} \sum_{k=1}^{M} g(\vec{x}(k))$$

$$= \frac{1}{n_b}\sum_{a=1}^{n_b} \underbrace{\frac{1}{M_b}\sum_{j=1}^{M_b} g(\vec{x}_{(a-1)M_b+j})}_{\bar{g}_a}$$

block variables

if $M_b > \tau \Rightarrow \bar{g}_a$'s are independent

$\Rightarrow$ apply CLT

$$I_M = \frac{1}{n_b}\sum_a \bar{g}_a$$

$$\sigma_{I_M}^2 = \frac{\sigma_b^2}{n_b}$$

σ_b^2 variance of the block variables

$$= \frac{1}{n_b}\sum_a \bar{g}_a^2 - \underbrace{\left(\frac{1}{n_b}\sum_a \bar{g}_a\right)^2}_{I_M^2}$$

$$= 2\tau \frac{\sigma_g^2}{M} \Rightarrow \boxed{\tau = \frac{1}{2} \underbrace{\frac{M}{n_b}}_{M_b} \frac{\sigma_b^2}{\sigma_g^2}}$$

$M_S \geq \tau \quad \Rightarrow \quad \tau$ becomes independent of M_S

$$|I_M - I| \simeq \sigma_{I_M} = \sigma_g \sqrt{\frac{2\tau}{M}}$$

$$\sim O\left(\sqrt{\frac{\tau}{M}}\right)$$

scaling of τ with D ?

good news : in many situations

$$\boxed{\tau \sim \underline{D}\, L^{z} \qquad z \sim O(1)}$$

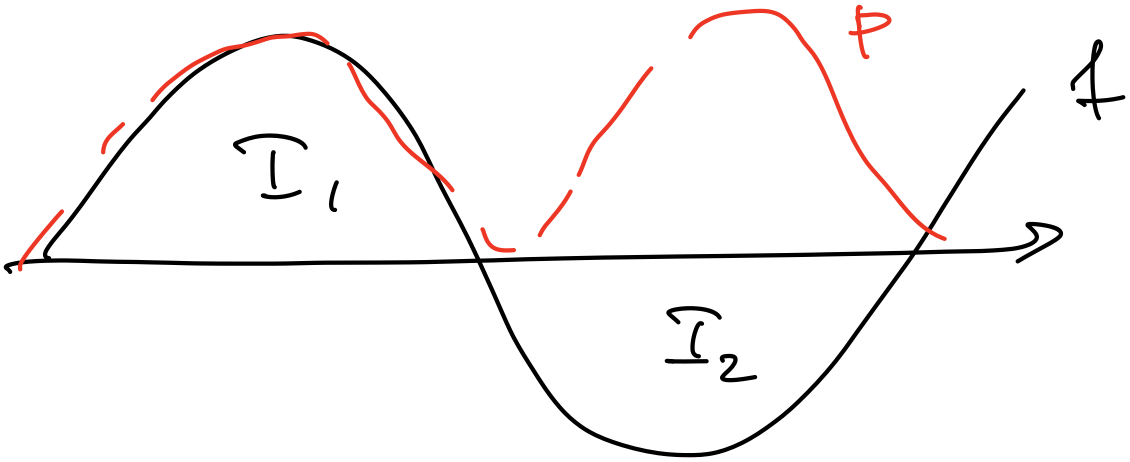
$(z = 2)$

$V = L^D$

polynomial cost in evaluating I

BUT there are exceptions

1) "sign problem"



$$g = \frac{f}{\underline{P}}$$

$$I = \underline{I_1} - \underline{I_2}$$

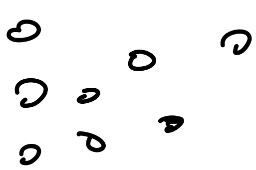
$$I_{1,2} \sim O(\exp(D))$$

$$I \sim (\mathrm{poly}(D))$$

2) τ diverges anomalously with D

$$\tau \sim O(\exp(\underline{D}))$$

<u>glasses</u>



$D \leftrightarrow N$

<u>spin glasses</u>

↓
combinatorial optimization

3) two (as $p_{eq}(D)$) maxima of p

$p(x) \sim \exp(-D)$

NUMERICAL OPTIMIZATION

find minimum of a given function

classical physics

1) minimization of $E(\vec{x}, \vec{p})$ 6N variables

$$\vec{X} = (\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N)$$

$$\vec{p} = (\vec{p}_1, \ldots, \vec{p}_N)$$

2) quantum physics, $\hat{H}$

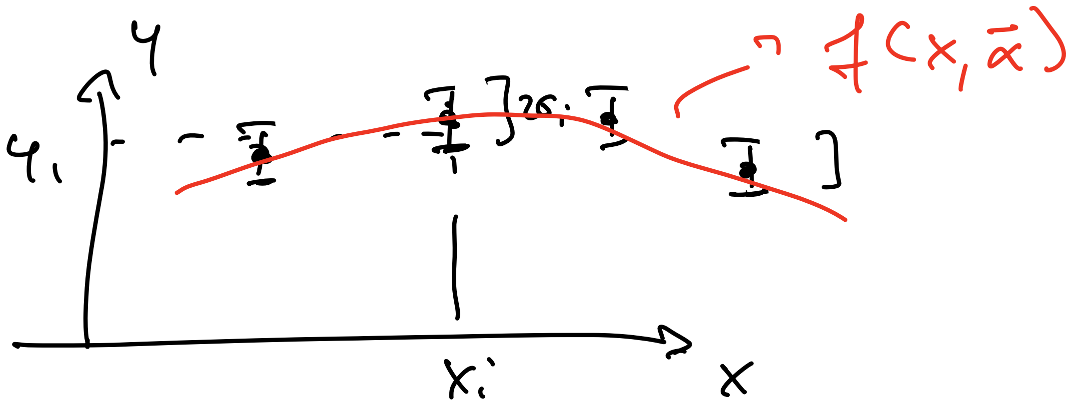
$|\psi(\vec{\alpha})\rangle$ $\qquad$ $\vec{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_P)$

postulate a form of the state $\qquad$ $P \sim poly(N)$

variational principle

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

3)



$\{x_i, y_i, \sigma_i\}$ $\qquad$ $i = 1, \dots, N$

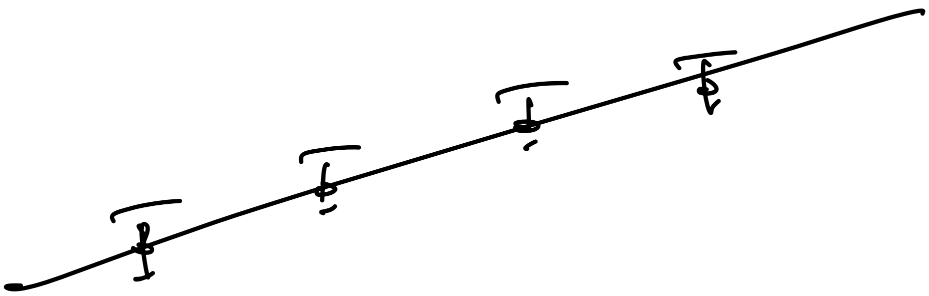
$f(x, \vec{\alpha})$: fit

least-square method

$$\chi^2(\vec{\alpha}) = \sum_{i=1}^{N} \frac{\left(f(x_i, \vec{\alpha}) - y_i\right)^2}{\sigma_i^2}$$

$\min_{\vec{\alpha}} \chi^2(\vec{\alpha}) \rightarrow$ fitting parameters $\vec{\alpha}$

ex.



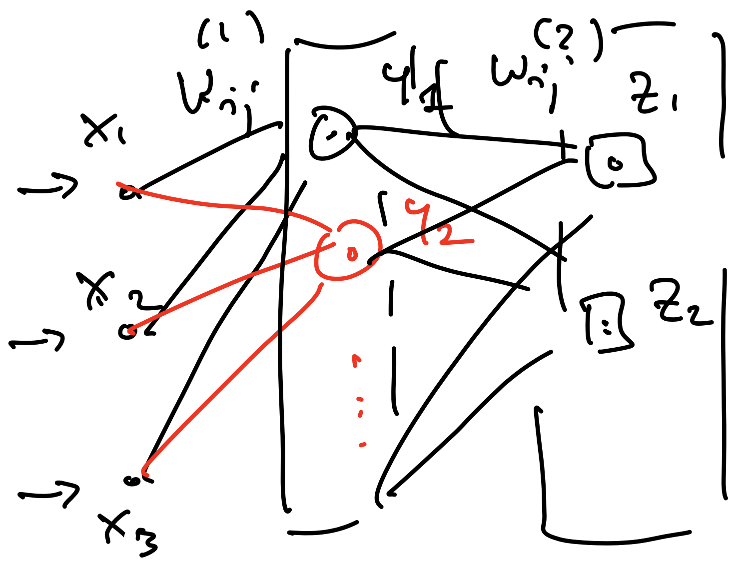
$Ax + b$

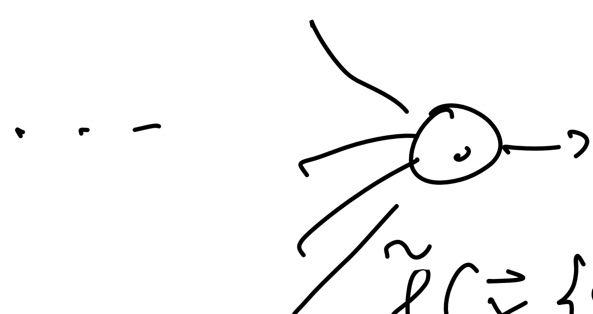
$\alpha_1 = A$

$\alpha_2 = b$

ex. f a complicated fit : MACHINE LEARNING

with neural networks





$\tilde{f}(\vec{x}, \{W_{ij}^{(n)}\})$

$\vec{x}$

mimic some function $\underline{\underline{f(\vec{x})}}$

$\vec{x}$ is a set of pixels $\Rightarrow$ image

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

g nonlinear function

$$z_i = g\left(\sum_j w_{ij}^{(2)} y_j\right)$$

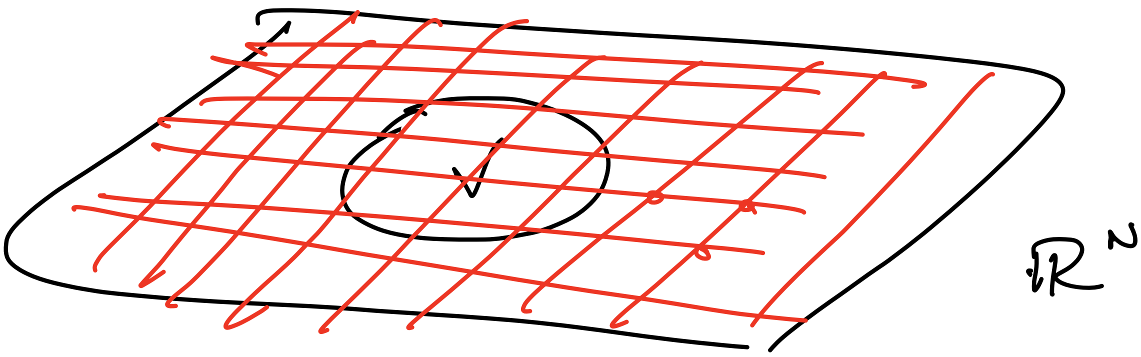
$$\min_{\{w_{ij}^{(k)}\}} \underbrace{\sum_{\vec{x}} \left| f(\vec{x}) - \tilde{f}(\vec{x}, w_{ij}^{(k)}) \right|^2}_{\text{cost function}}$$

Minimization of a function of <u>continuous variables</u>

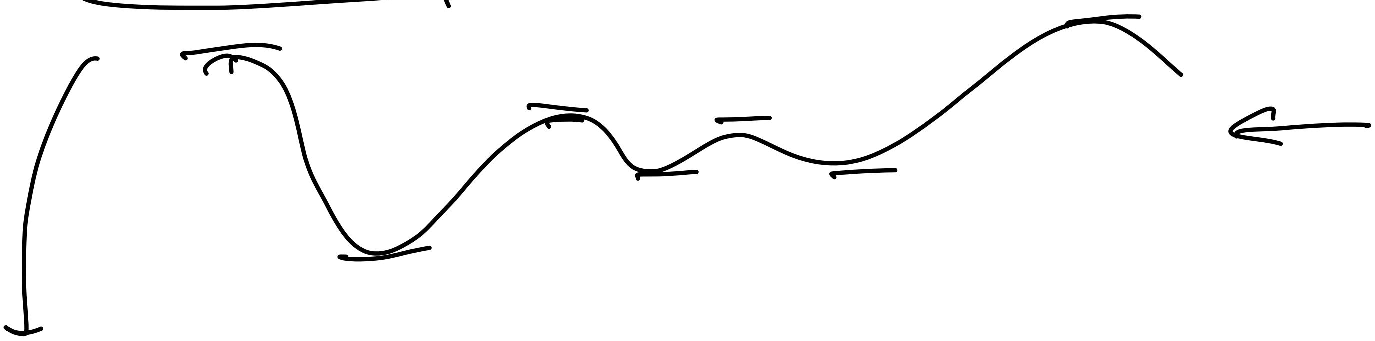
$$f(\vec{x}) : \mathbb{R}^N \rightarrow \mathbb{R}$$

$\underline{N \gg 1}$

problem: $\vec{x}^* : \min_{\vec{x}\in V} f(\vec{x}) = f(\vec{x}^*)$



$\Rightarrow \boxed{\vec{\nabla}_{\vec{x}} f(\vec{x}) = 0}$ extreme



N coupled non-linear equations

solution by enumeration of all possible values of $\vec{x}$ (or a grid thereof)

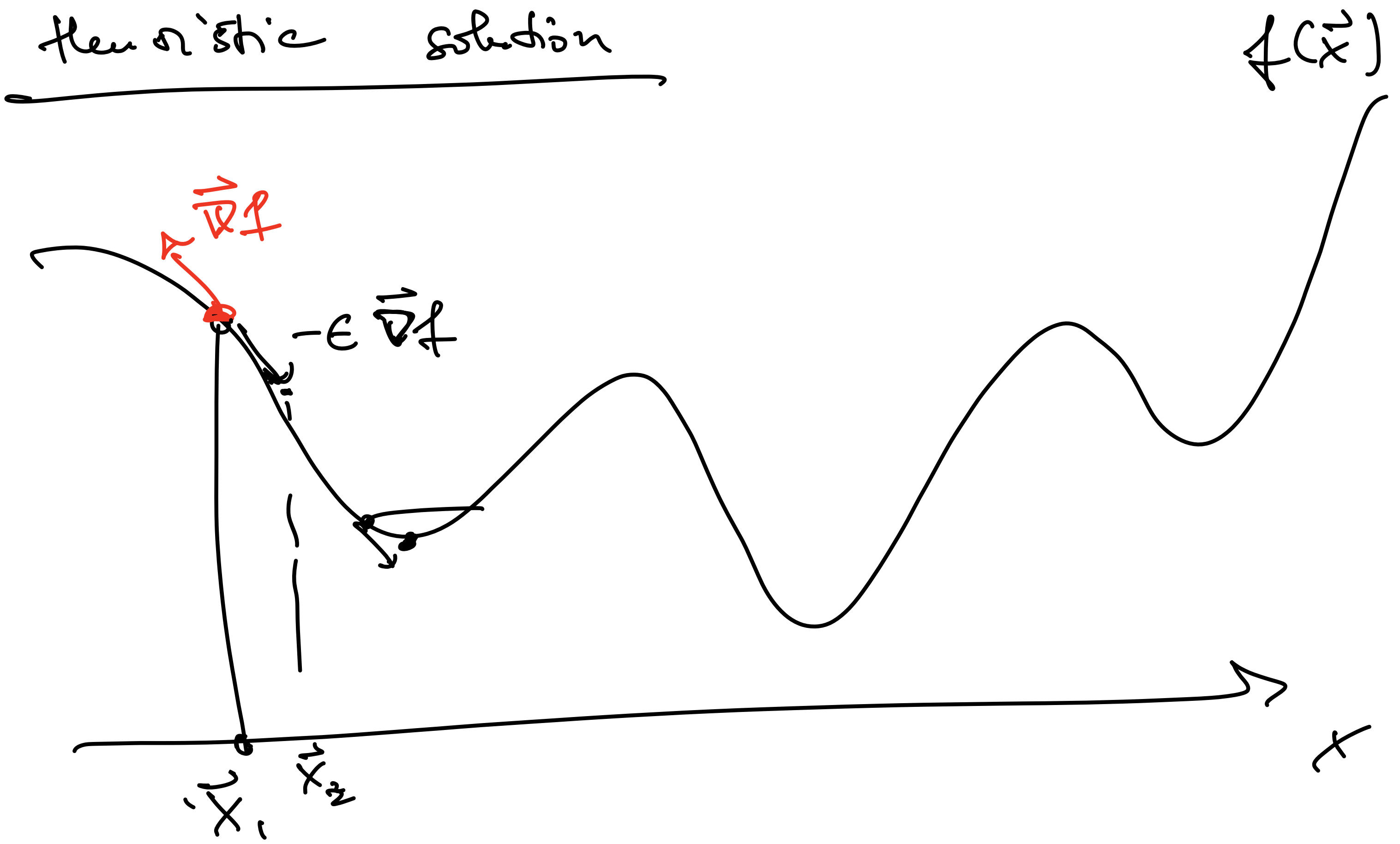
$\downarrow$

$T \sim \exp(N)$

$\Rightarrow$ enumeration is the only sure solution to finding the absolute minimum

$\Rightarrow$ optimization is <u>HARD</u>.

heuristic solution



Gradient descent

$$\vec{x}_1 \to \vec{x}_2 = \vec{x}_1 - \epsilon \vec{\nabla} f(\vec{x}_1)$$

$$(\epsilon \ll 1)$$

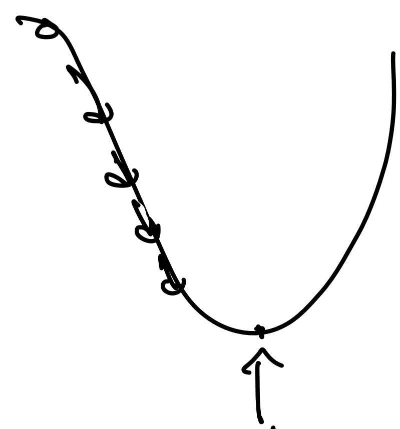
$$\vec{x}_i \to \vec{x}_{i+1}$$

$$f(\vec{x}_{i+1}) = f[\vec{x}_i - \epsilon \underbrace{\vec{\nabla} f(\vec{x}_i)}]$$

$$\simeq f(\vec{x}_i) - \underline{\epsilon \, \|\vec{\nabla} f(\vec{x}_i)\|^2} \quad < 0$$

$$\underbrace{+ \frac{1}{2} \epsilon^2 \left(\vec{\nabla} f(\vec{x}_i)\right)^T H \, \vec{\nabla} f(\vec{x}_i) + o(\epsilon^3)}$$

$$H_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$



$\vec{x}_{min}$

if ϵ is small enough

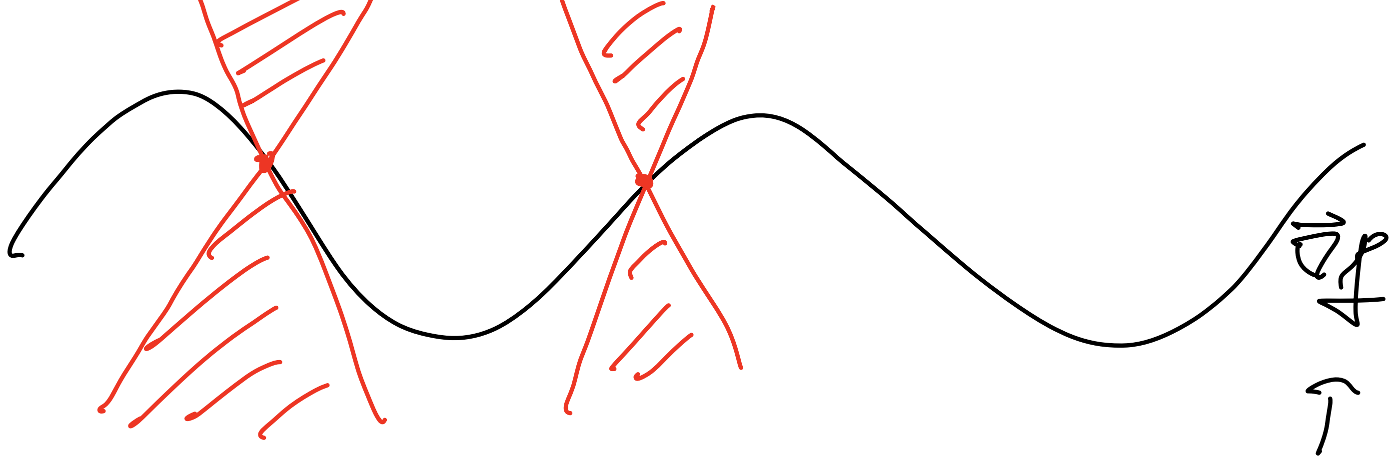
$$f(\vec{x}_{i+1}) - f(\vec{x}_i) < 0$$

$$| f(\vec{x}_k) - f(\vec{x}_{min}) | \simeq \; ?$$

one can show under mild assumption:
(Lipschitz continuity of the gradient)

$\exists \; L > 0$:

$$\| \nabla f(\vec{x}) - \nabla f(\vec{y}) \| \leq L \, \| \vec{x} - \vec{y} \|$$



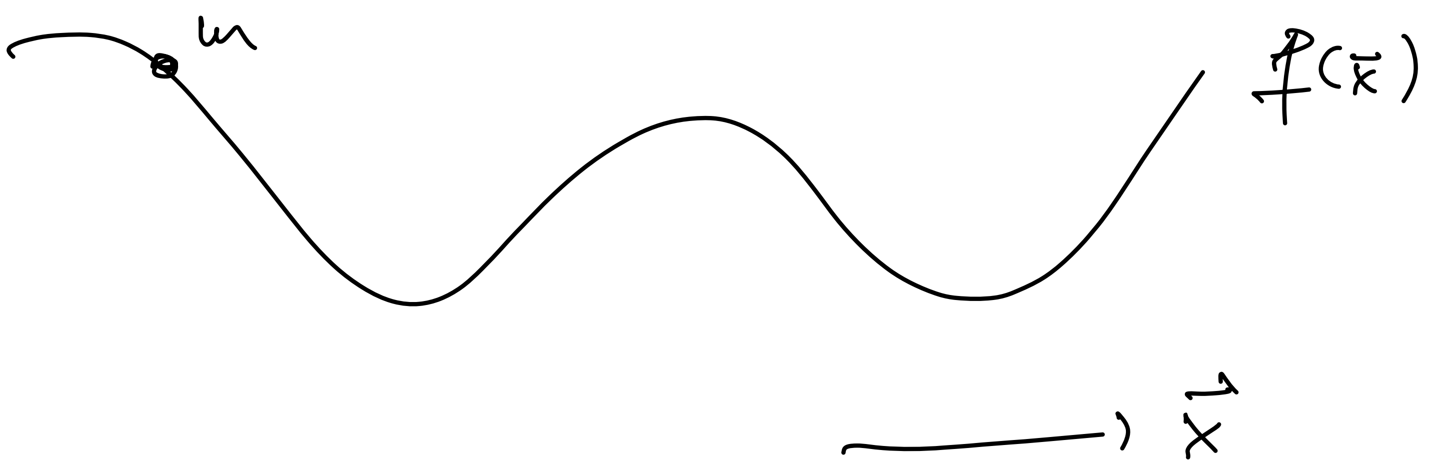
$$|f(\vec{x}_k) - f(\vec{x}_{min})| \sim O\left(\frac{1}{k}\right)$$

worst-case scenario

often ?

$$| \qquad | \sim \exp(-k/k_0)$$

Accelerated gradient descents



$$\cancel{m\ddot{\vec{x}}} = -\gamma\dot{\vec{x}} - \vec{\nabla} f(\vec{x})$$

gradient descent

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

$$\underbrace{\frac{\vec{x}_{n+1} - \vec{x}_n}{\epsilon}}_{\dot{\vec{x}}} = -\vec{\nabla} f$$

n : time

discretized version $t_n = n\Delta t$

$$\underbrace{m\frac{\vec{x}_{n+1} + \vec{x}_{n-1} - 2\vec{x}_n}{(\Delta t)^2}}_{m\ddot{\vec{x}}} + \gamma\frac{\vec{x}_{n+1} - \vec{x}_n}{\Delta t} + \vec{\nabla} f = 0$$

$$(\vec{x}_{n+1} - \vec{x}_n)\underbrace{\cancel{\left(\frac{m}{(\Delta t)^2} + \frac{\gamma}{\Delta t}\right)}}_{\epsilon^{-1}}^{\epsilon} + \overbrace{(\vec{x}_{n-1} - \vec{x}_n)\underbrace{\frac{m\epsilon}{(\Delta t)^2}}_{\beta}} = -\epsilon\vec{\nabla} f$$

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

"momentum"

"momentum term"

Polyak's "heavy-ball" method

Nesterov's accelerated gradient

$\vec{x}_1$

$\hookrightarrow \vec{y}_1 = \vec{x}_1$

$$\vec{x}_2 = \vec{x}_1 - \epsilon \vec{\nabla} f(\vec{x}_1)$$

$$\Rightarrow \vec{y}_2 = \vec{x}_2 + \left(\frac{k-1}{k+2}\right)(\vec{x}_2 - \vec{x}_1) \quad (k=2)$$

Nesterov's point

$$\vec{x}_3 = \vec{y}_2 - \epsilon \nabla f(\vec{y}_2)$$

$\vdots$

$$\boxed{\begin{aligned} \vec{y}_k &= \vec{x}_k + \frac{k-1}{k+2}(\vec{x}_k - \vec{x}_{k-1}) \\ \vec{x}_{k+1} &= \vec{y}_k - \epsilon \vec{\nabla} f(\vec{y}_k) \end{aligned}}$$

effective dynamics

$$\ddot{\vec{x}} + \frac{3}{t}\dot{\vec{x}} + \vec{\nabla} f(\vec{x}) = 0$$

one can

<u>prove</u> $|f(\vec{x}_n) - f(\vec{x}_{min})| \sim O\left(\frac{1}{n^2}\right)$