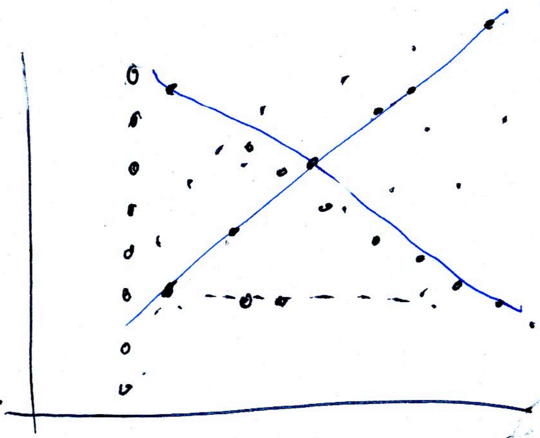
Finding co-linear points

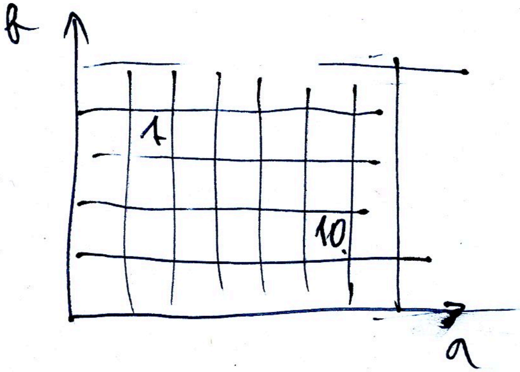
$y = ax + b$



Brute force:

take all i,j pairs:

$$\underline{a_{ij}} = \frac{y_i - y_j}{x_i - x_j} \Rightarrow \underline{b_{ij} = y_i - a_{ij}x_j}$$

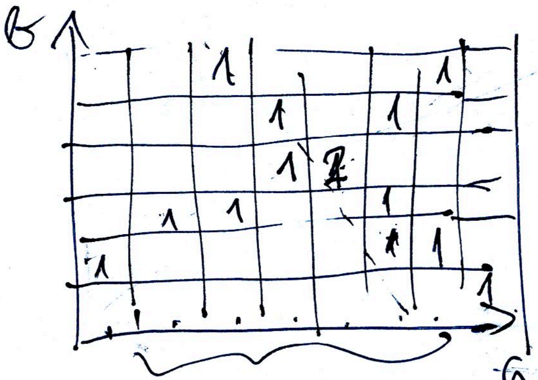


# of pairs $\binom{N}{2} = \frac{N(N-1)}{2} = O(N^2)$

HOUGH-TRANSFORM

$$y_0 = a x_0 + b \Rightarrow \boxed{b = -a x_0 + y_0}$$

Loop through all points

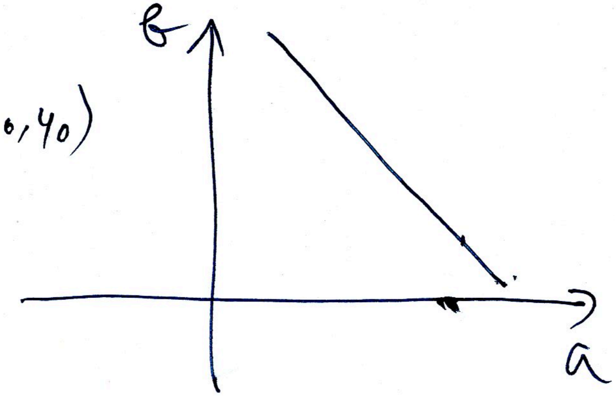


m bins

$a_k = t\ (k = 1..m) \Rightarrow b_k(x_0, y_0)$

$x_1\, y_1$

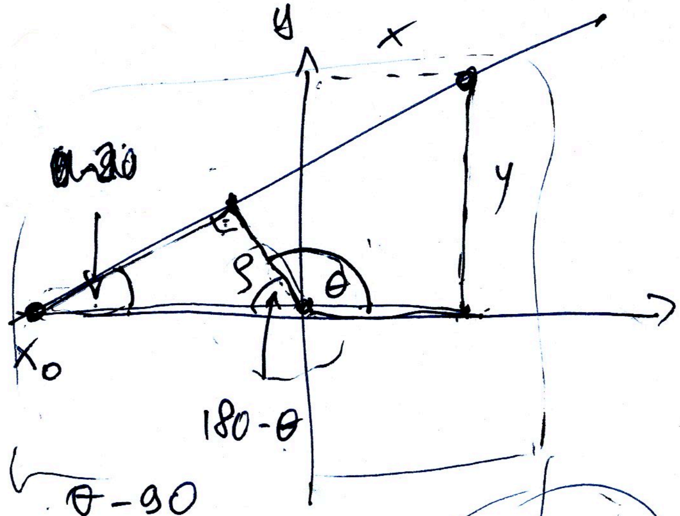
$O(N.m)$ ← fixed



10^8 or 10^{12} ↑ cost is LINEAR in N

Better parametrization of the line

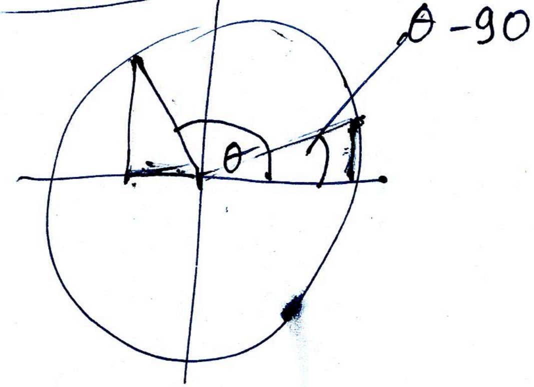
$$x\cos\theta + y\sin\theta = \rho \qquad (\rho,\theta)$$



$$\frac{y}{x-x_0} = \tan(\theta-90) = \frac{\sin(\theta-90)}{\cos(\theta-90)} =$$

$$= -\frac{\cos\theta}{\sin\theta} = \boxed{-\frac{1}{tg\,\theta}}$$

$$\rho = -x_0\sin(\theta-90) = x_0\cos\theta$$



$$y = ax + b.$$

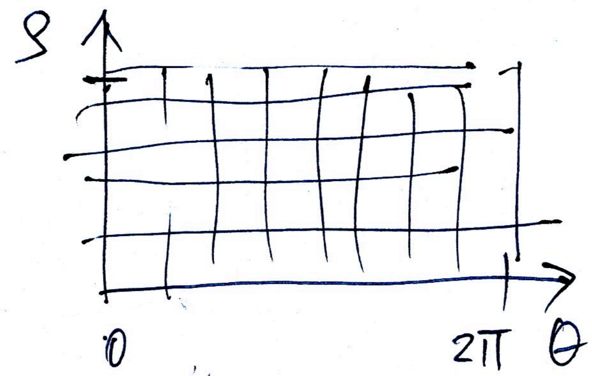
$$y = (x-x_0)\cdot\left(-\frac{1}{tg\,\theta}\right) = -\frac{1}{tg\,\theta}\cdot x + x_0\frac{1}{tg\,\theta}$$

$$y = -\frac{\cos\theta}{\sin\theta}\cdot x + \frac{\rho}{\cancel{\cos\theta}}\cdot\frac{\cancel{\cos\theta}}{\sin\theta} \qquad /\times\sin\theta$$

cost

$$y\sin\theta = -x\cos\theta + \rho$$

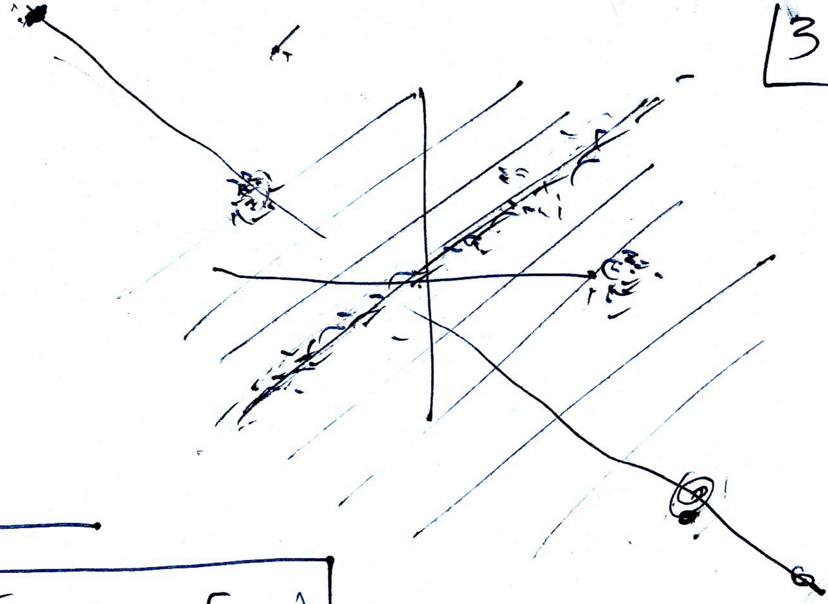
$$\boxed{\rho = x\cos\theta + y\sin\theta}$$



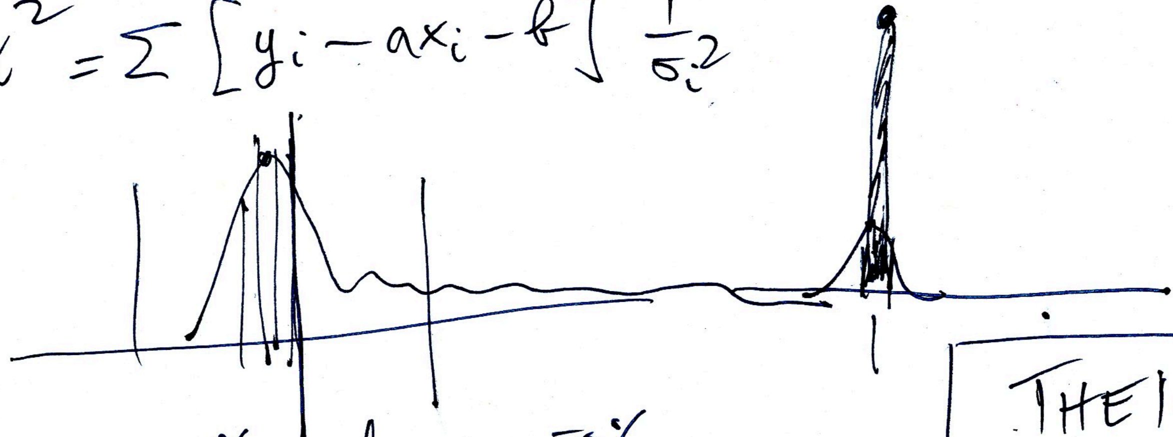
$\theta(N.m)$

$N\log N$ $\quad N^3 \quad \ldots \quad \cancel{2^N}$

Least-squares (LS) regression



$$\chi^2 = \sum \left[y_i - a x_i - b \right]^2 \frac{1}{\sigma_i^2}$$



$\boxed{\text{THEIL-SUN}}$

$$a_{ij} = \left(\frac{y_j - y_i}{x_j - x_i} \right) \qquad \binom{N}{2} = \frac{N(N-1)}{2} \text{ different slopes}$$

$$a = \operatorname*{median}_{i \neq j} (a_{ij})$$

$$\operatorname{median}(y_i - a x_i) = b$$

Robust fit

~30% points are outliers, ~~they~~
right answer!

$\boxed{\text{scipy.stats}}$

$\boxed{\text{Siegel}}$ For every pair i, j we can compute

$$a_{ij} = \left(\frac{y_j - y_i}{x_j - x_i}\right)$$

$$a_i = \text{median}_{j \neq i}(a_{ij})$$

$$b_i = \text{median}_{j \neq i}\left(\frac{x_j y_i - x_i y_j}{x_j - x_i}\right)$$

$$a = \text{median}_i(a_i)$$

$$b = \text{median}_i(b_i)$$

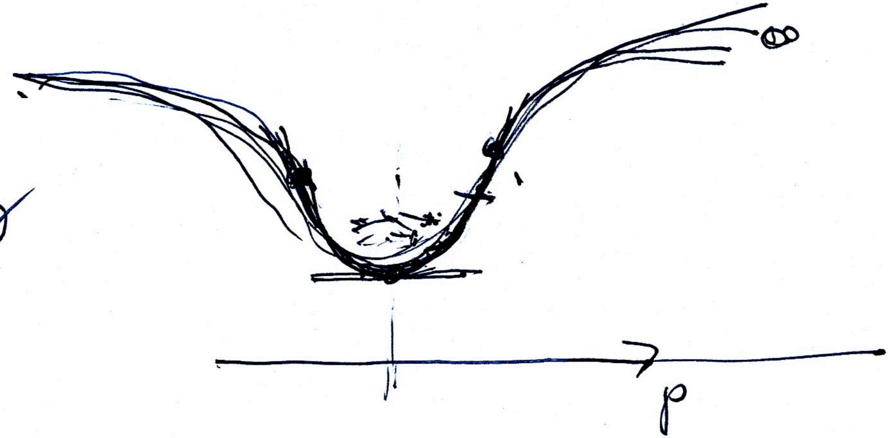
siegel slopes

50% points can be bad...

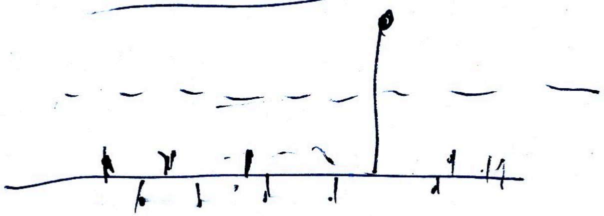
Robust least squares

$$\chi^2 = \sum_{i=0}^{N} r_i^2 .$$

$$\frac{\partial f}{\partial p} = 0$$



Gradient descent.



$$\ln(1+x) \approx x + \dots \qquad x \ll 1$$

Robust cost function:

$$\ln\left(1 + x_i^2/\sigma^2\right)$$

$$\sum \ln\left(1 + x_i^2/\sigma^2\right)$$

$$\ln(100) \sim 5$$

σ is xover between parabolic cost function and the log.

L6

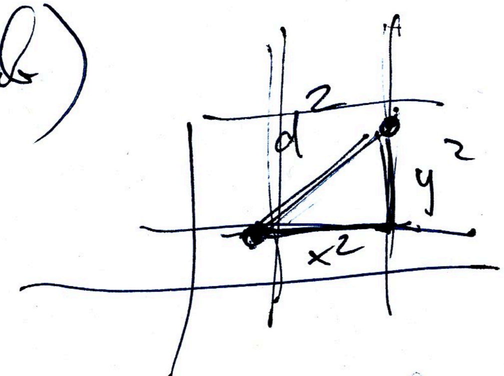
Vector Norms

$\vec{x} = (x_1, \dots\, x_n)$ $\|\vec{x}\| = \sqrt{x_1^2 + x_2^2 + \dots x_n^2}$ Euclidean norm

L2 norm.

L2 $\|x\|_2$

L1 $\|x\|_1 = \sum |x_i|$ Manhattan norm (taxicab)



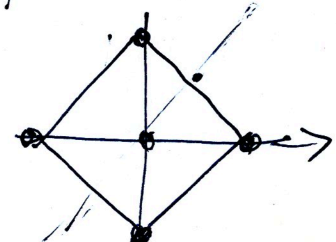
Lp $\|x\|_p = \left[\sum_{i=1}^{n} (x_i)^p\right]^{1/p}$ p - norm

$[1, -2, 4, 3]$

$\vec{x} = (x, y)$ ~~L0~~

$L1(\vec{x}) = 1$

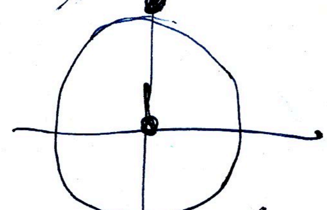
L0



L1 : $1 + 2 + 4 + 3 = 10$

L2 : $\sqrt{1 + 4 + 16 + 9} = \sqrt{30}$

L2



L∞ $\lim_{p \to \infty} (4^p)^{1/p} = 4 \Rightarrow \max(x_i)$

L∞