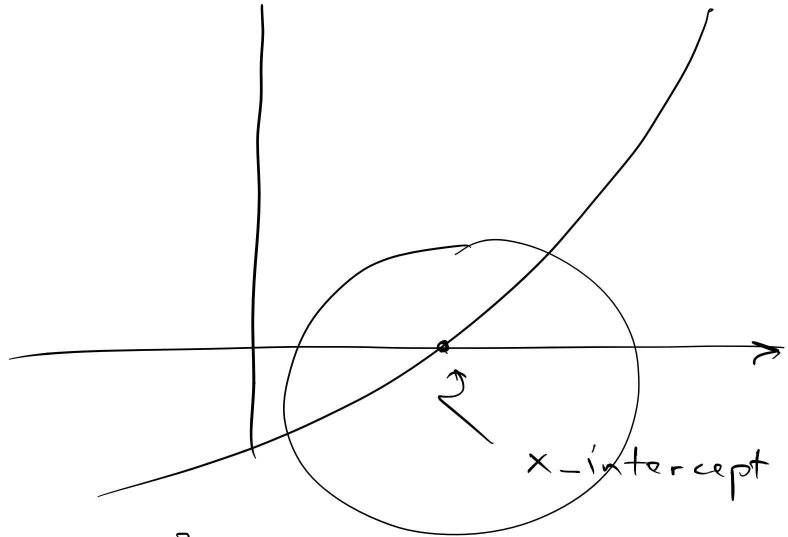
- $F(x) = 0$



- $\sqrt{3}$ $\quad f(x) = x^2 - 9 = 0$
- $\boxed{f(x) = p}$

complex

$g(x) = f(x) - p = 0$

$f(x) > p$

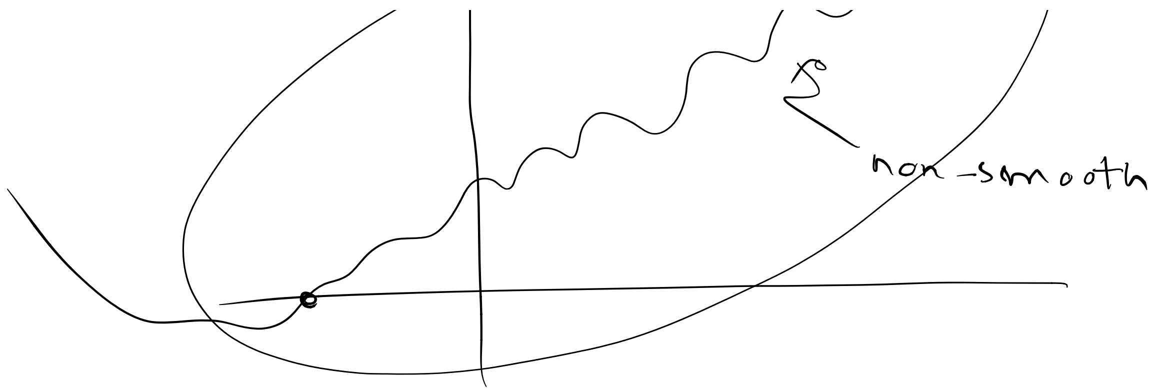
$f(x) < p$

$\boxed{f(x) = p}$

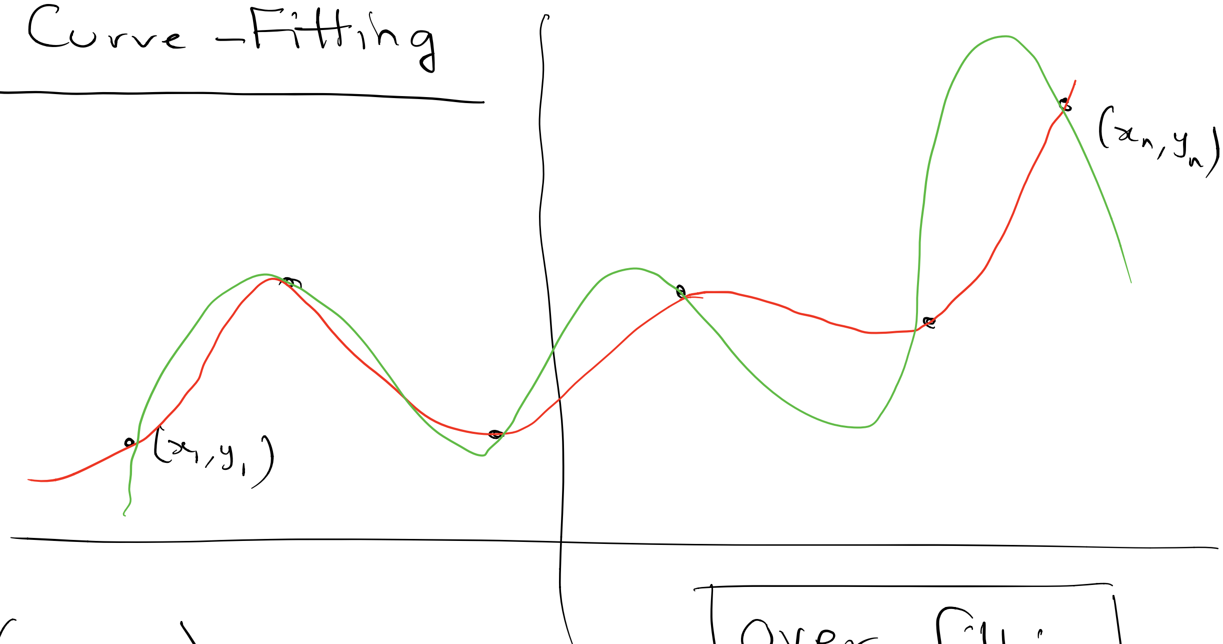
<u>Methods</u>

- Bi-section
- ✓ Newton-Raphson
- ✓ Secant

pros/cons

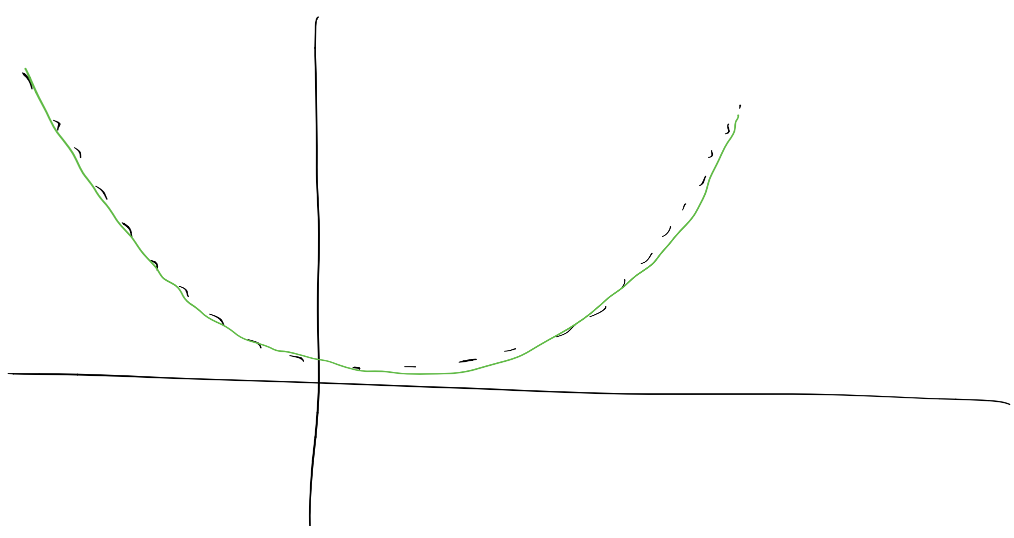


Curve-Fitting



(x_j, y_j)

$\boxed{\text{Over-fitting}}$



|

Family of curves

- line
- parabola
- cubic
- ⋮

$$a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots$$

Distance

$$\sum_j \left| y_j - \hat{y}_j \right|$$

$$\sqrt{\frac{1}{N}\sum_j \left(y_j - \hat{y}_j\right)^2} \quad \text{RMS}$$

given points

fitted points

Brute-Force (Grid search)

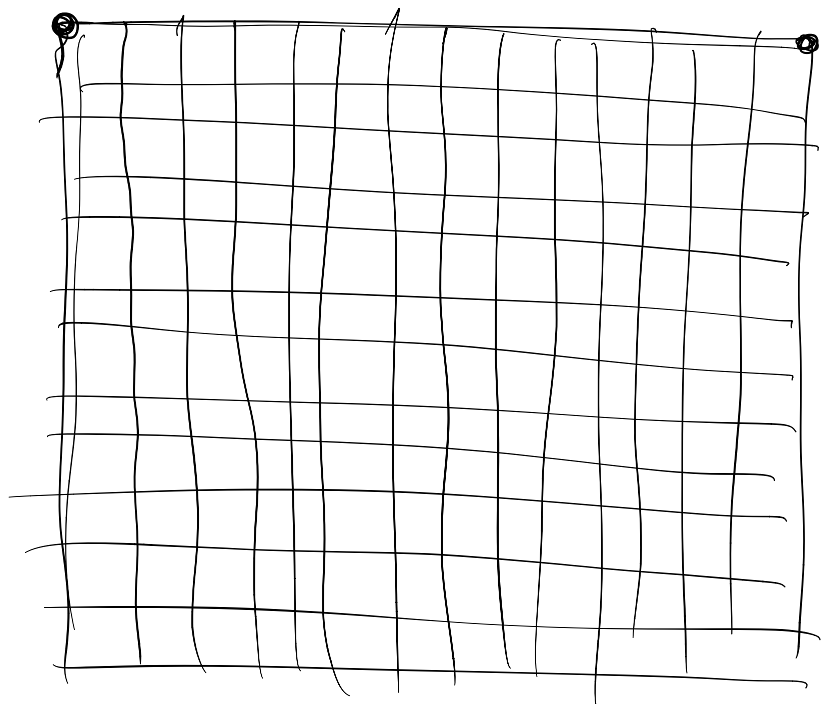
```
for
    for
        for
```

$y = a_0 + a_1 x$

for

a_0

a_1



for $a_0 = -3 : 3$

for $a_1 = -5 : 5$

- calc
- find the min

end for

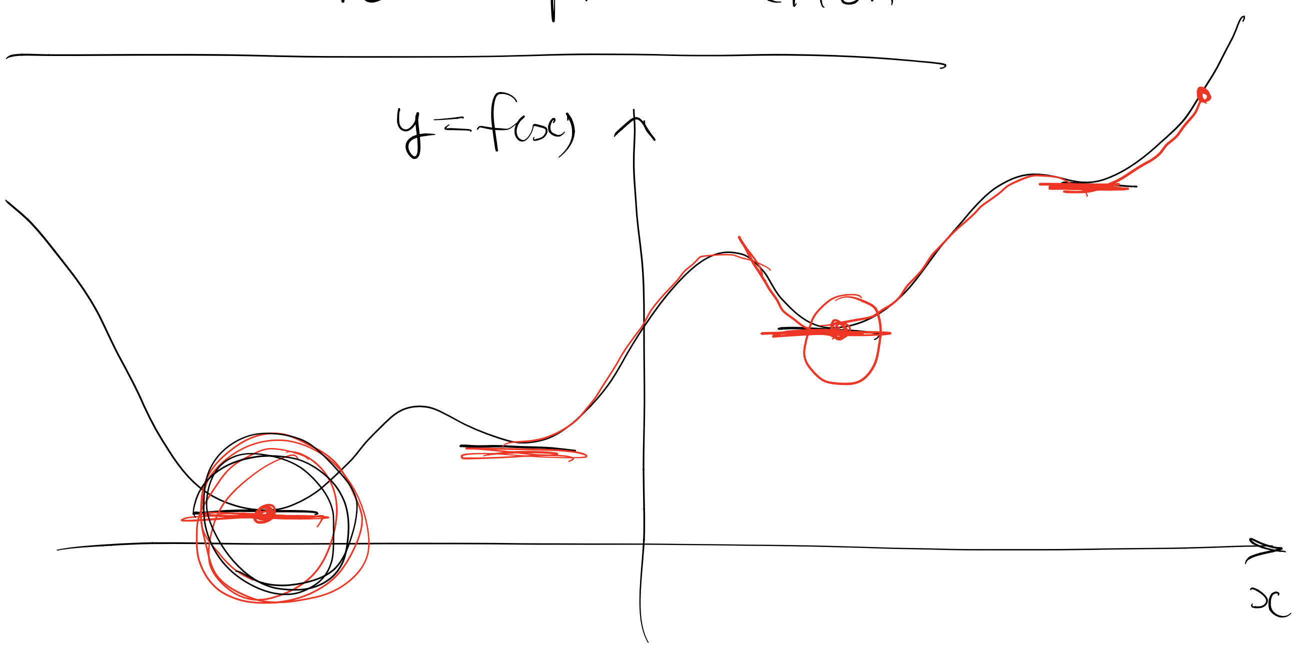
endfor

min

- fminsearch (Matlab) simplex method
    - criteria
    - obj func

hopefully best possible fit

Intro. to Optimization



Q: Looking for the minimum of f.

There are many (local) minima

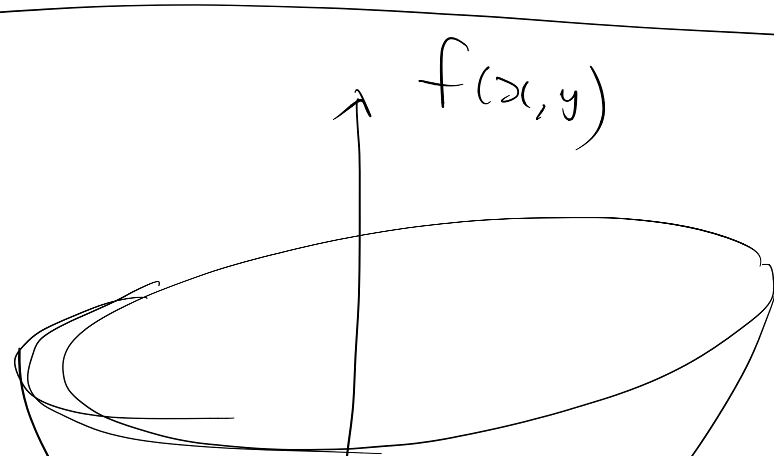
Global Minimum

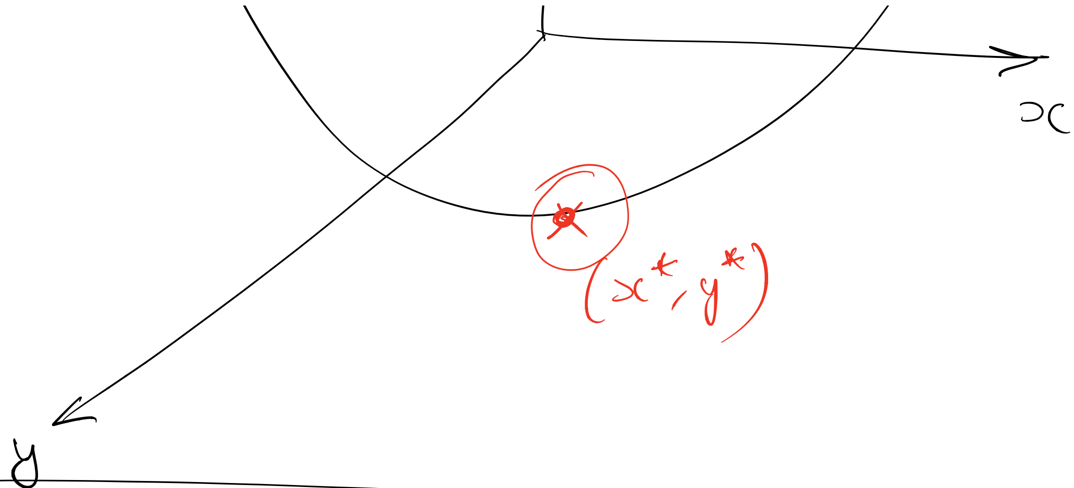
Global Maximum = Global minimum of negative f

Given a function $f(x)$

find x^* s.t.

$$\boxed{f(x^*)} \leqslant \boxed{f(x)} \quad \forall x \text{ in the domain}$$





$$\boxed{f(x,y) = 2\cancel{(x-1)}^2 + 4\cancel{(y+2)}^4 + 13}$$

find (x^*, y^*)

and $\min f(x,y)$

$x^* = 1$

$y^* = -2$

$\min f(x,y) = 13$

$$\cancel{f(x,y) = 2(x-1)^2 + 4(y+2)^3 + 13}$$

$$f(x,y) = \underbrace{\boxed{2(x-1)^2 + 4(y+2)^2}} + (x-1)(y+2)$$

$$+ 13$$

$$\boxed{f(x,y) = 2(x-1)^2 + 4(y+2)^2 + 13}$$

Grid Search

Gradient Descent Method

Def. Gradient

$f(x,y,z)$

$f(x_1, x_2, x_3, \dots, x_n)$

$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{pmatrix}$$

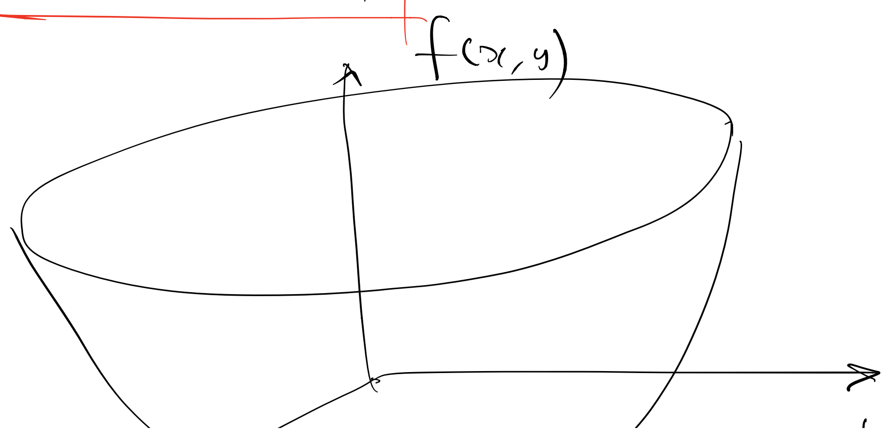
$$\begin{pmatrix} \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}_{n \times 1} \qquad \text{Column vector}$$

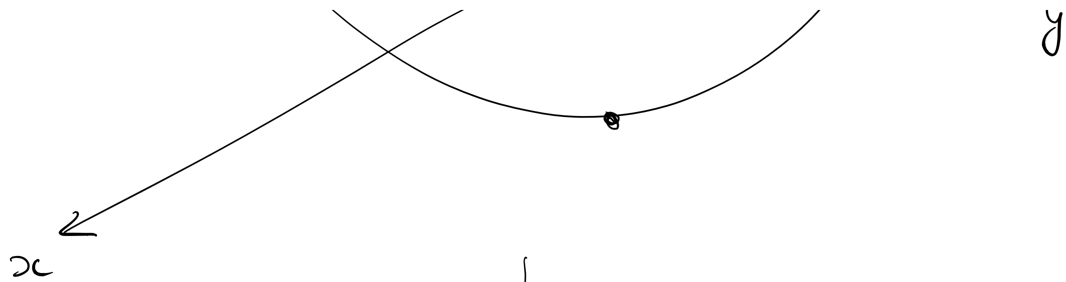
$$f(x,y) = 2xy$$

$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \\ \frac{\partial f}{\partial y} \end{pmatrix} = \begin{pmatrix} 2y \\ \\ 2x \end{pmatrix}$$

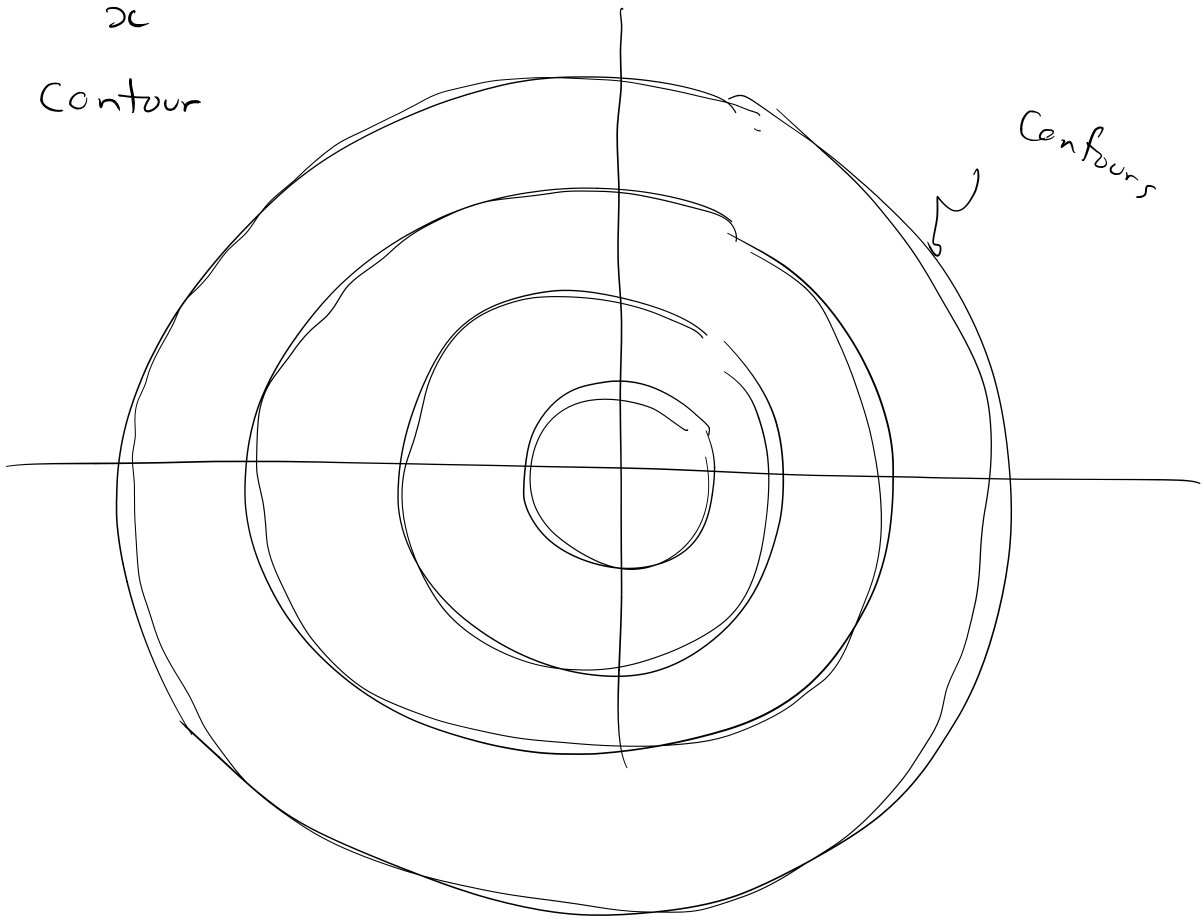
$$f(x,y) = 2x^2y^3$$

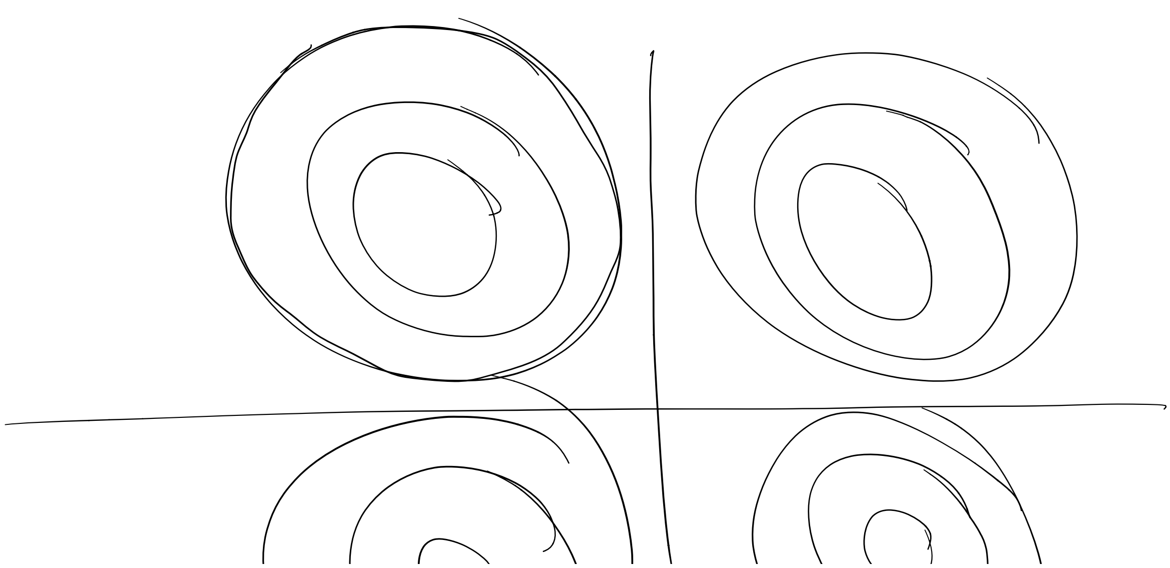
$$\boxed{\nabla f = \begin{pmatrix} 2y^3(2x) \\ \\ 2x^2(3y^2) \end{pmatrix}}$$

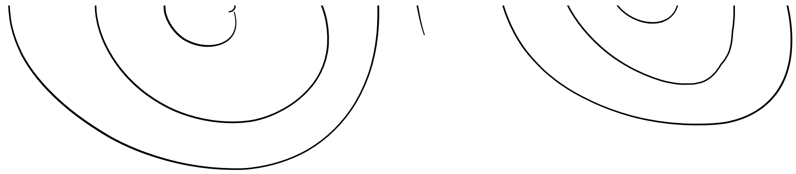


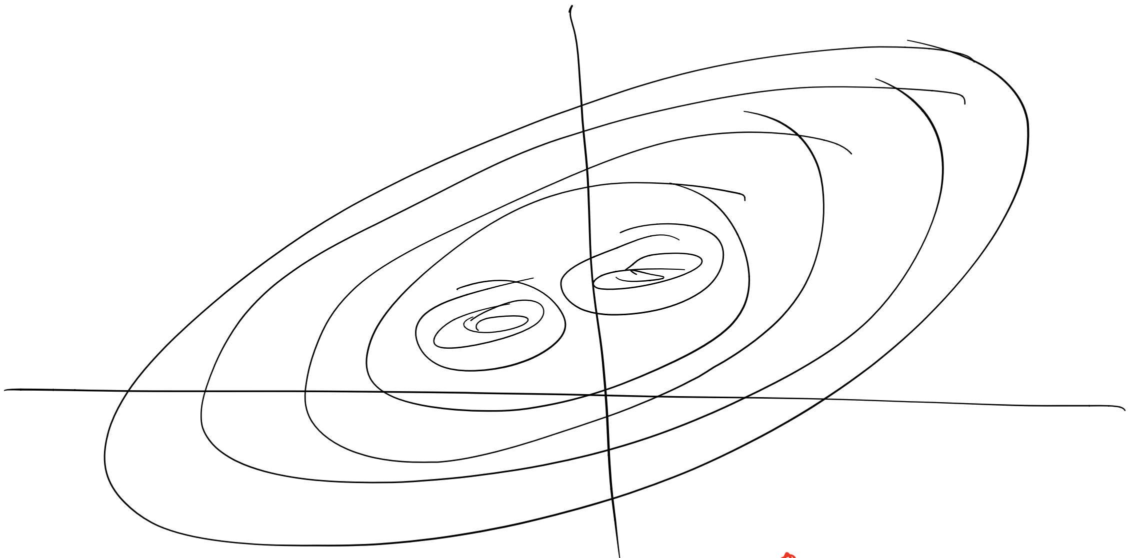


Contour

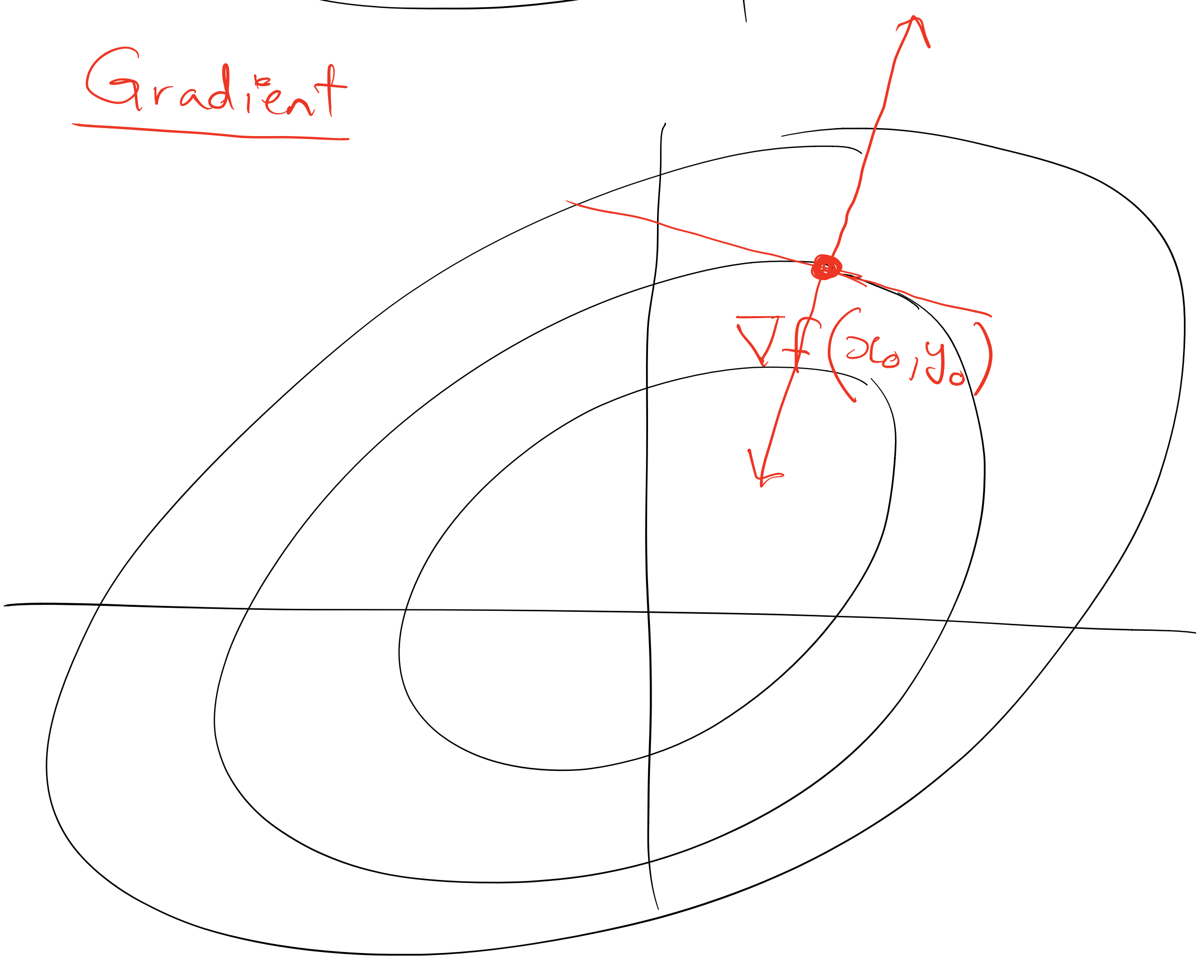








Gradient



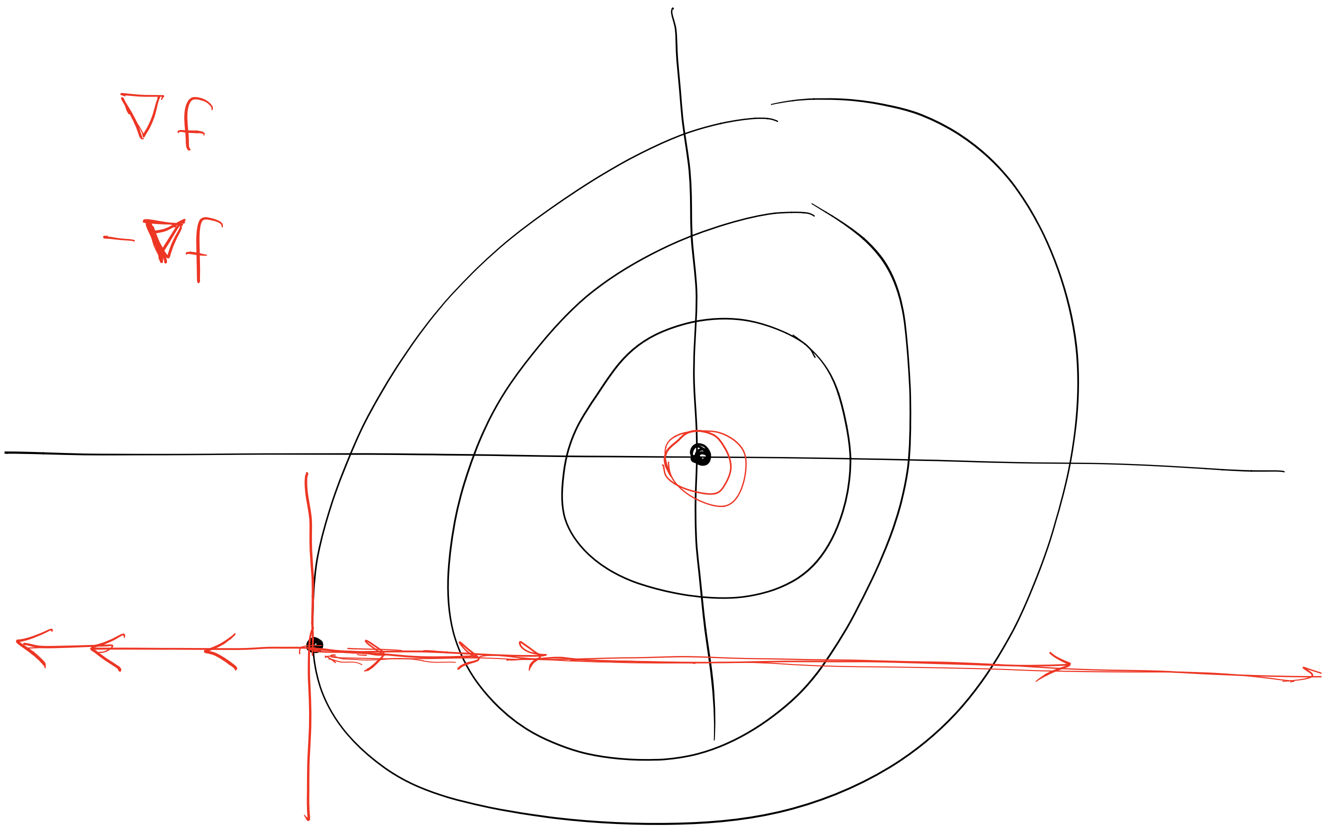
Gradient points in the direction of

maximum increase

greatest rate of increase!

∇f

$-\nabla f$



$$\vec{X} = \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \qquad -\nabla f(\vec{X})$$

$$X = X - \delta \nabla F(X)$$

$$\boxed{X_{new} = X_{old} - \delta \nabla F(X_{old})}$$

learning rate

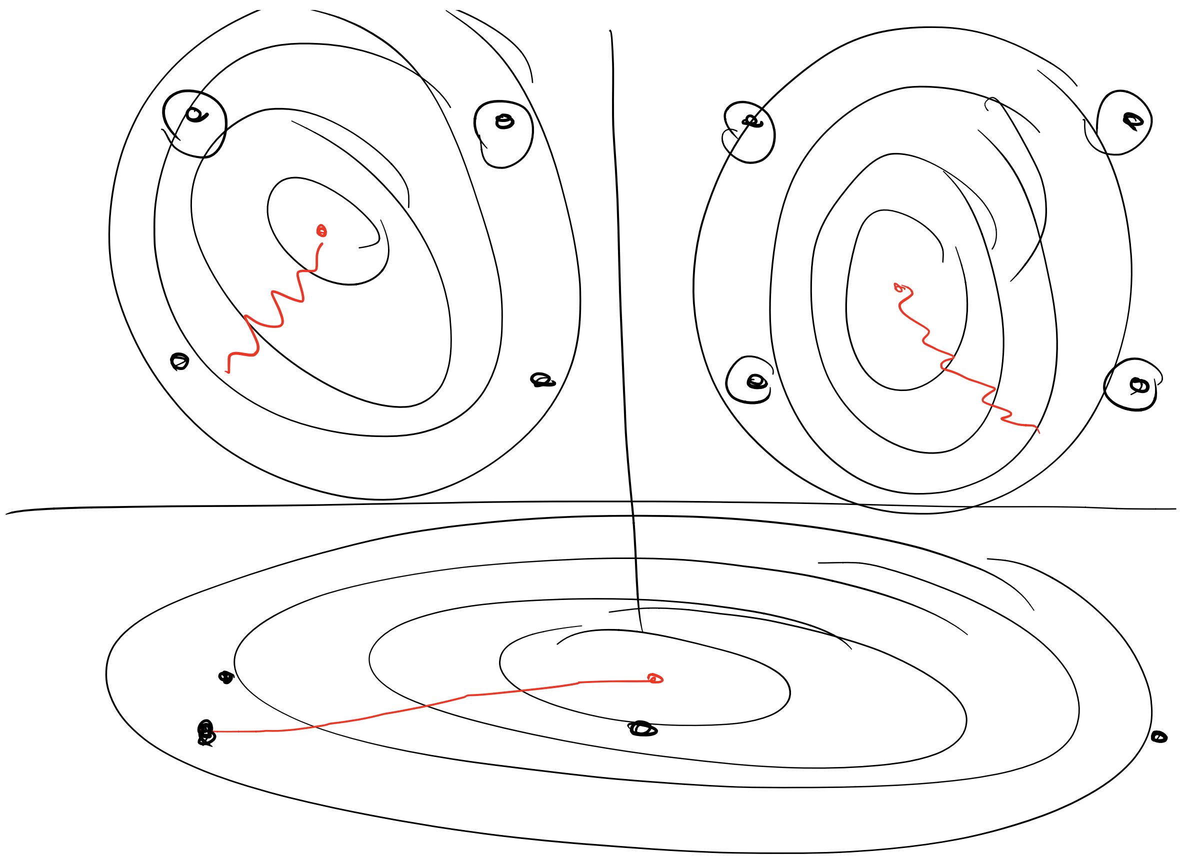
$$X_{old} = X_{new}$$

$$X = X - \delta \nabla F(X)$$

a good choice of δ is a <u>MUST</u>

δ is too large would not converge

δ is too small converges very slowing



Dan Goldfarb

Machine Learning

Loss function

- various different starting points
- randomization (stochastic)
- (F) loss function / different weighting

$$L = \sum (y_j - \hat{y}_j)^2$$

$$L = \boxed{\sum | y_j - \hat{y}_j |}$$

$$L = \sum w_j (y_j - \hat{y}_j)^2$$

$\boxed{j = 10, 13, 14}$

$w_j = 1 \quad \forall j$ except for 10, 13, 14

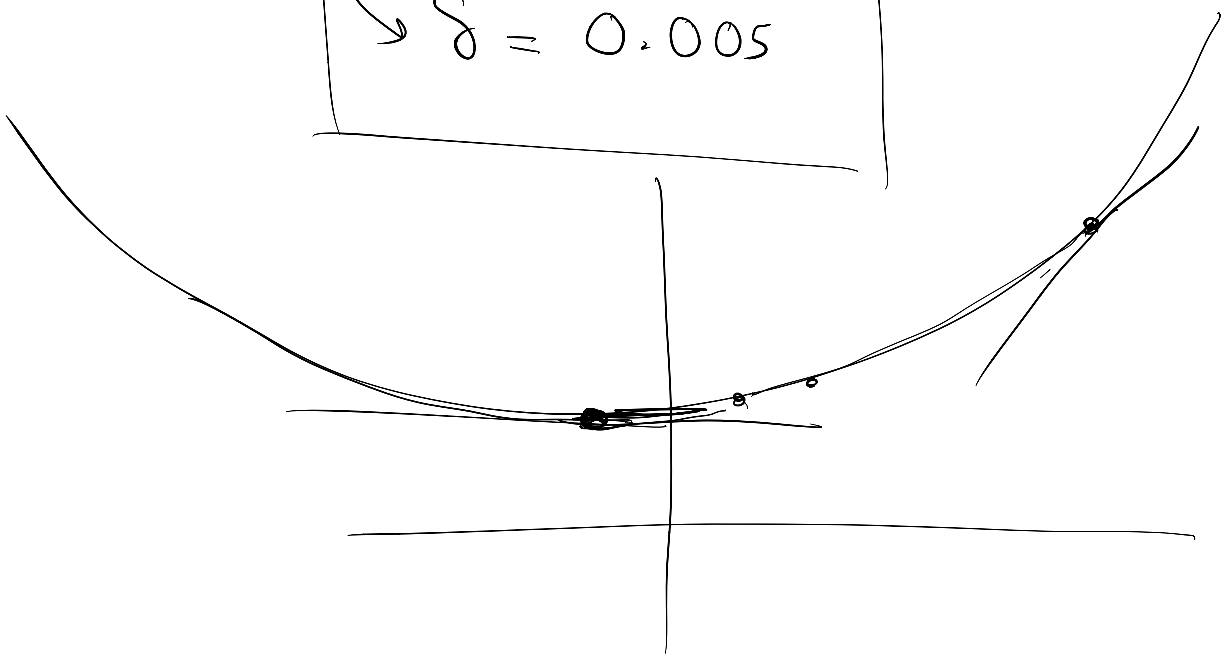
$w_j = 2 \quad$ 10, 13, 14

- how to make δ adaptive?

$\delta = 0.005$

$\delta = 0.2$

$\delta = 0.005$



$$X_{new} = X_{old} - \delta \nabla f(X_{old})$$

✓ $\|X_{new} - X_{old}\| = \sqrt{\sum_i \left(x^{(i)}_{new} - x^{(i)}_{old}\right)^2}$

Criterion to stop.

- you stop if $\|X_{new} - X_{old}\| < \varepsilon = 10^{-5}$

8 $\nabla F(x_{old})$

- Vanilla $\boxed{\text{Gradient Descent}}$

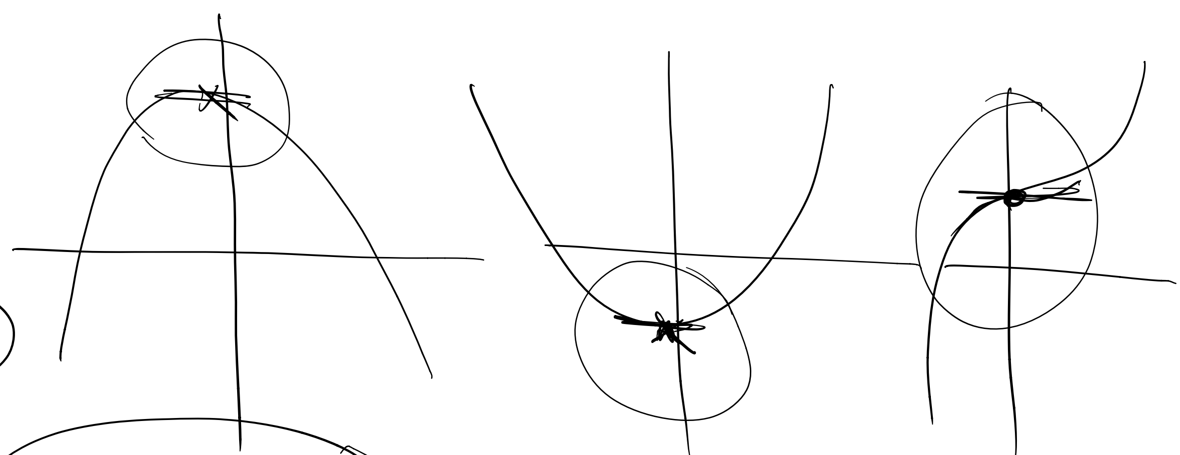
✓ Stochastic

✓ Mini-Batch

-

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

1



1 out of 3 $\frac{1}{3}$

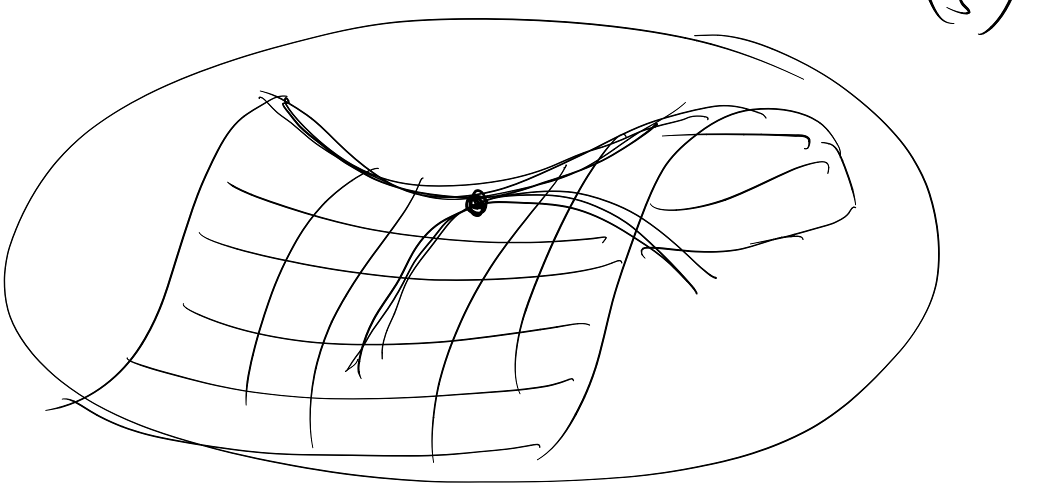
$f(x)$

d-dimensional case $\nexists$ local minimum

1-dim $\frac{1}{3}$

l-dim ? $\boxed{\left(\frac{1}{3}\right)^{l}}$

$\frac{1}{3} \times \frac{1}{3}$ $\frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} \times \dots = \left(\frac{1}{3}\right)^{l}$



$$\left(\frac{1}{3}\right)^{\ell}$$