Statistics and Data Science for Engineers
E178 / ME276DS

Optimization theory

Probability theory

univariate

1. Random variable: sample space Ω, P ... probability measure

$\star$ p ... density function

Φ ... cdf.

- Expected value:
  - $E[X]$, μ_X
  - LLN.
  - linear: $E\left[\begin{matrix}\text{linear}\\\text{combination}\end{matrix}\right] = \begin{matrix}\text{linear}\\\text{comb.}\end{matrix}\left(E[\ldots]\right)$.
- Variance:
  - $Var[X]$, σ_X^2 ... σ_X ... standard deviation.
  - Not linear $Var\left[\begin{matrix}\text{linear}\\\text{comb}\end{matrix}\right] = \ldots \alpha^2$, Cov

2. Multivariate R.V. (vectors)

- Marginal distribution
- Conditional RV ⟶ Bayes' theorem.
- Independence , correlation.

3. Parametric families

- Bernoulli
- Binomial.
- Poison
- Exponential

- Uniform
- Gaussian (CLT).

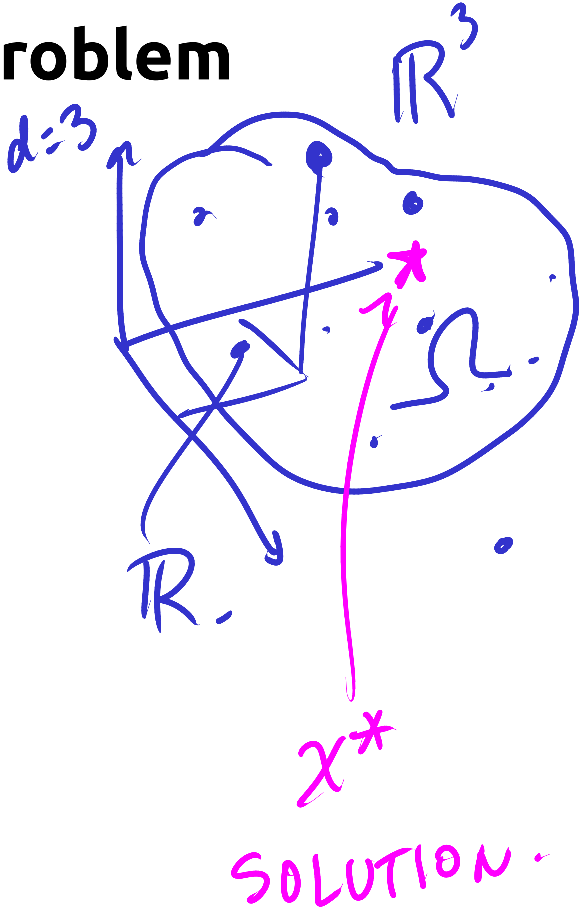
Components of an optimization problem

1. The **decision vector** : $x \in \mathbb{R}^d$ dimension

not the sample space!

2. The **search set** or **feasible set** : $\boxed{\Omega} \subseteq \mathbb{R}^d$

3. The **objective function** : $J : \Omega \to \mathbb{R}$



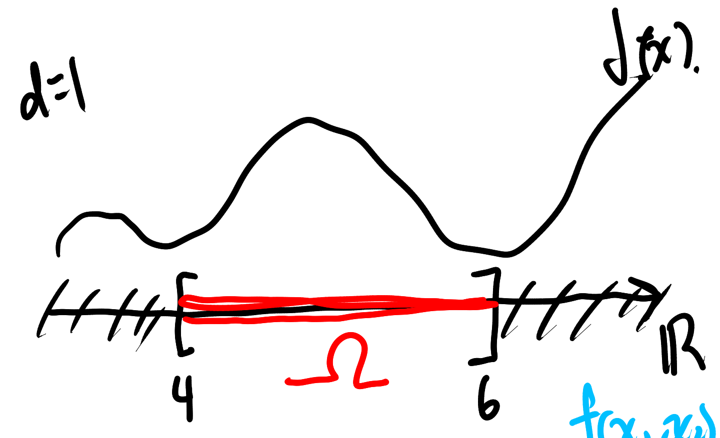
Notation.

$$\underset{x}{\text{minimize}} \quad J(x)$$
$$\text{subject to:} \quad x \in \Omega$$

Constraint specification

$$\underset{x}{\text{maximize}} \; J(x) \;\; \text{s.t.} \;\; x \in \Omega \quad \Longleftrightarrow \quad \underset{x}{\text{minimize}} \; -J(x) \;\; \text{s.t.} \;\; x \in \Omega.$$

$$\underset{x}{\text{minimize}} \quad J(x)$$
$$\text{subject to:} \quad \begin{cases} f_i(x) = 0 & i = 1 \ldots n \;\text{... equality} \\ g_j(x) \leq 0 & j = 1 \ldots m \;\text{... inequality.} \end{cases}$$

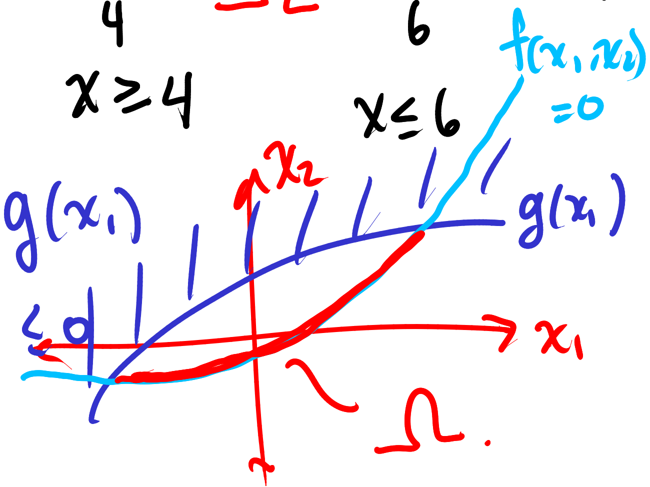


$x \geq 4$ $\quad$ $x \leq 6$

$$\boxed{x^* = \underset{x}{\text{argmin}} \quad J(x)}$$
$$\text{subject to:} \quad f_i(x) = 0 \quad i = 1 \ldots n$$
$$g_j(x) \leq 0 \quad j = 1 \ldots m$$

$x_2 \leq g(x_1)$

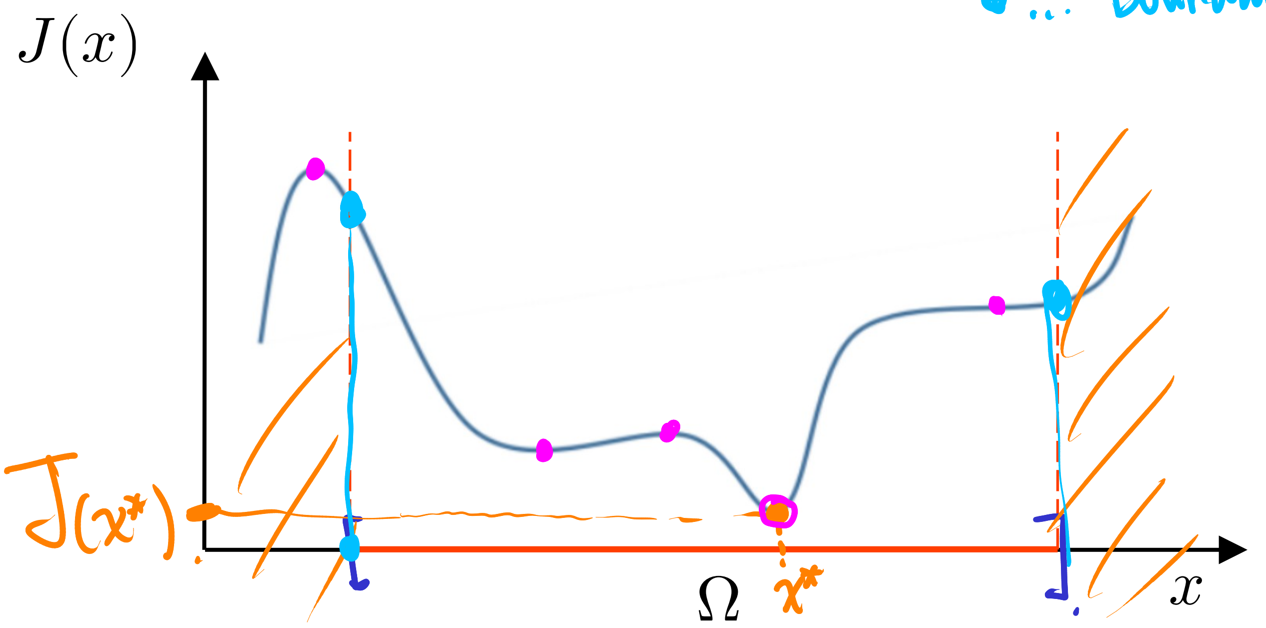
$x_2 - g(x_1) \leq 0$



$d=1$

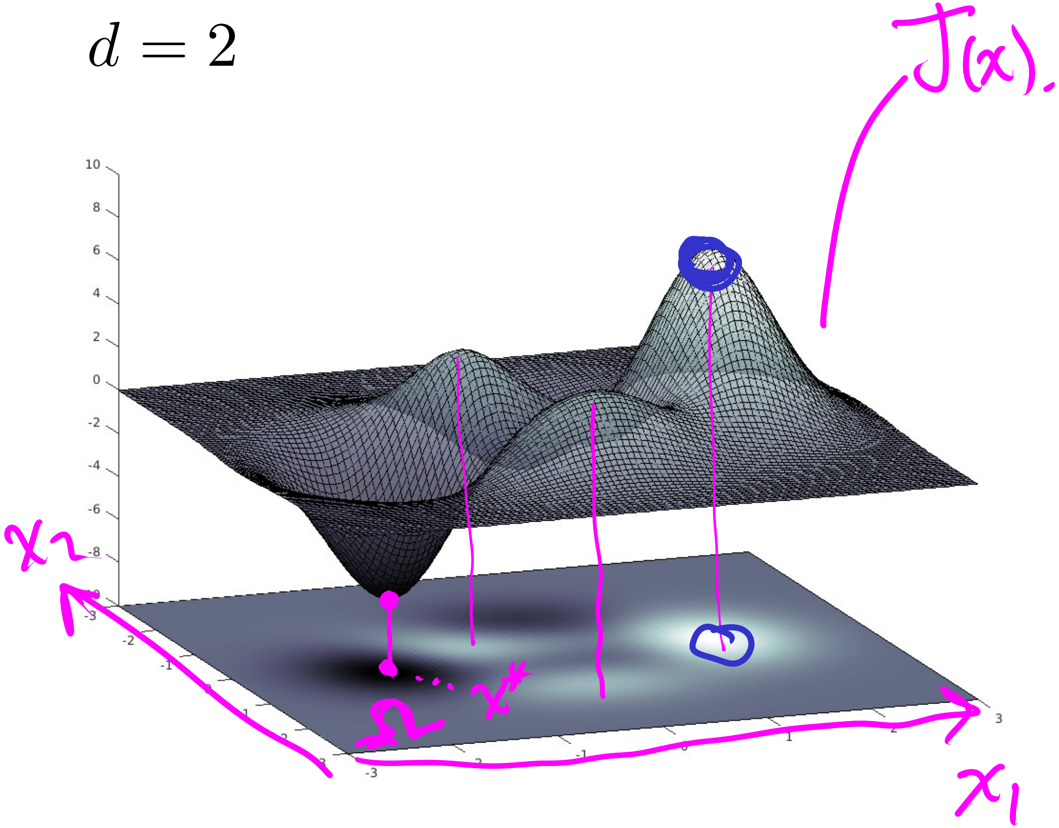
• ... stationary points : $J'(x)=0$

• ... boundary points.

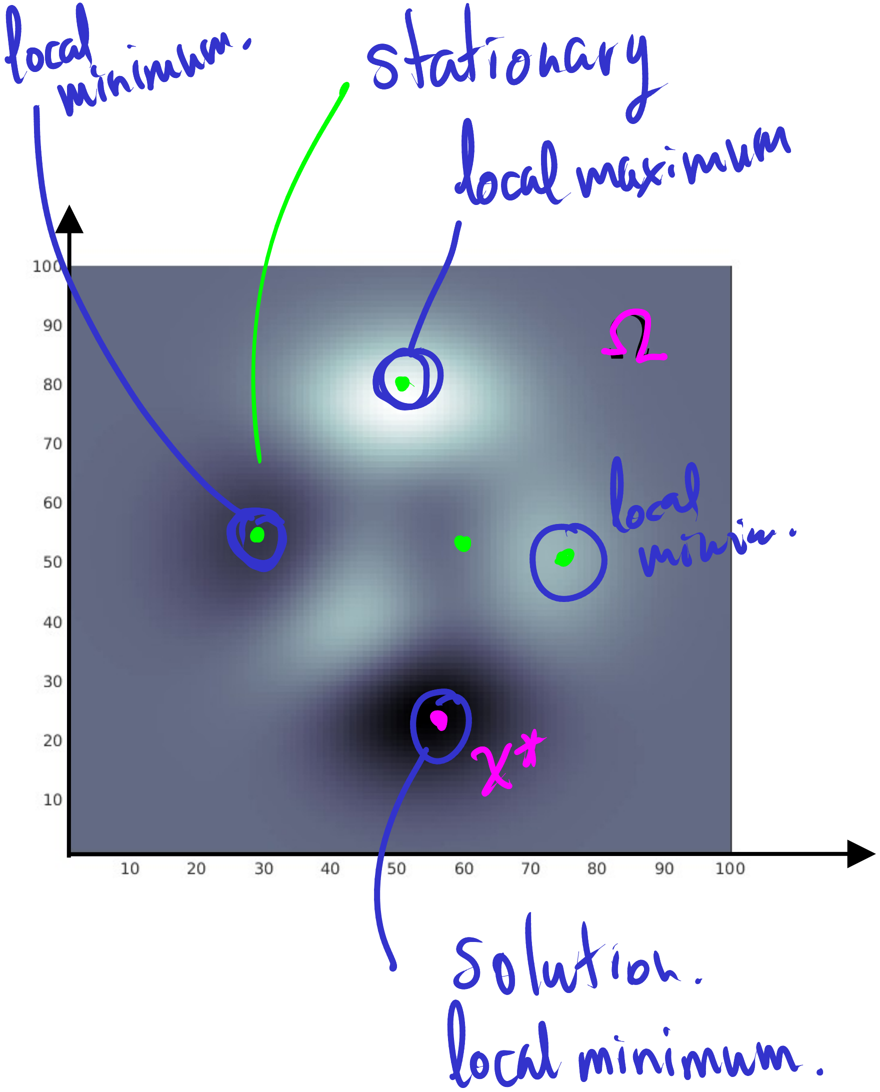


Solution x^* is a stationary point.

$d = 2$

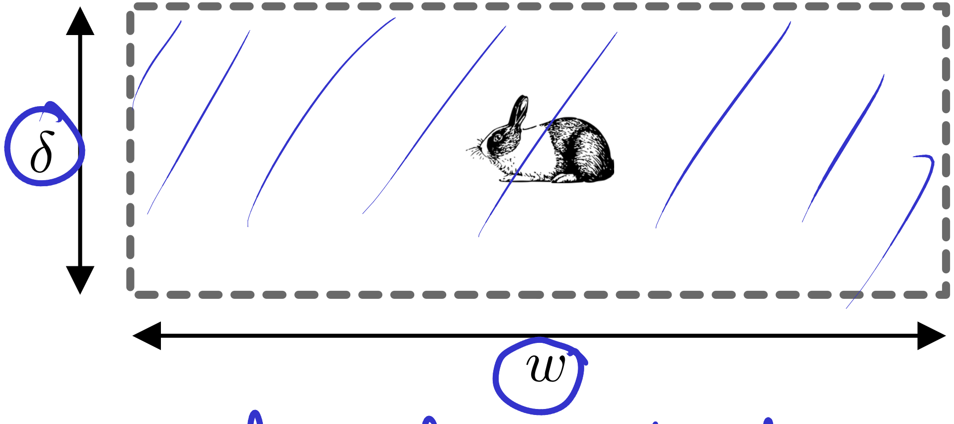


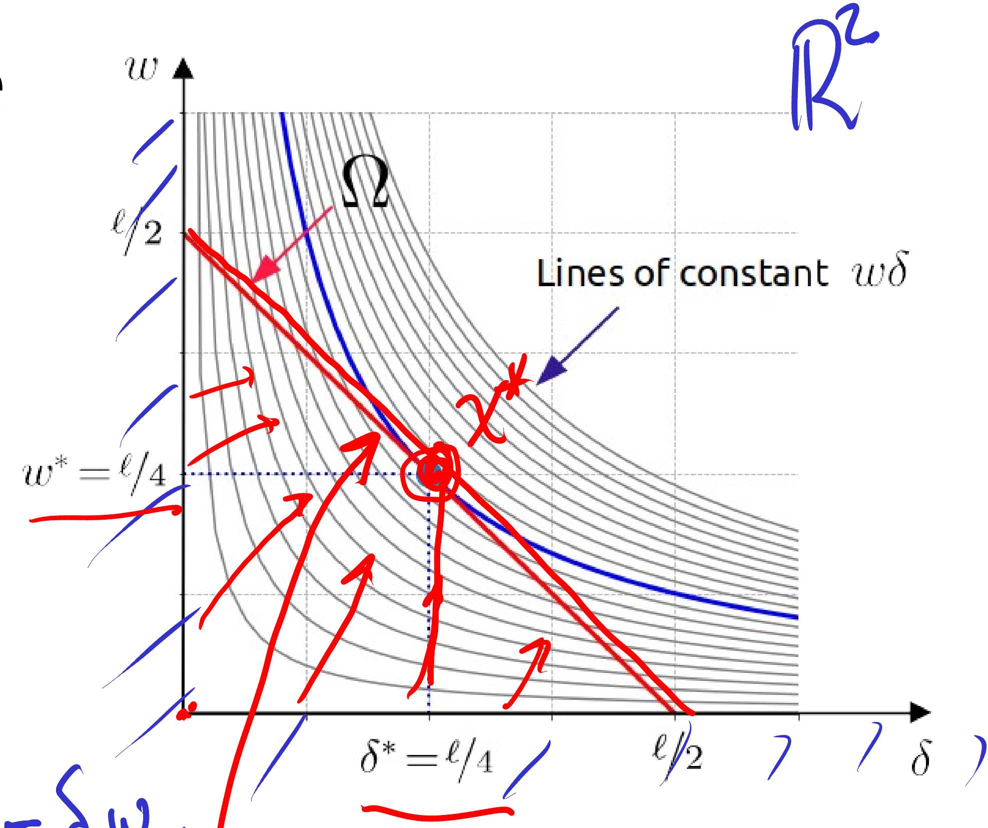
J(x).
10
8
6
4
2
0
-2
-4
-6
-8
x2
x1
-3
-2
-1
0
1
2
3
Ω
x*



local minimum.
stationary
local maximum
Ω
local minim.
x*
solution.
local minimum.
100
90
80
70
60
50
40
30
20
10
10
20
30
40
50
60
70
80
90
100

Example: Design a rabbit enclosure





Length ℓ of fence material

$d = 2$

$x = (\delta, w)$.

$J(\delta, w) = -\delta w$.

$$\underset{\delta, w}{\text{minimize}} \quad -\delta w.$$

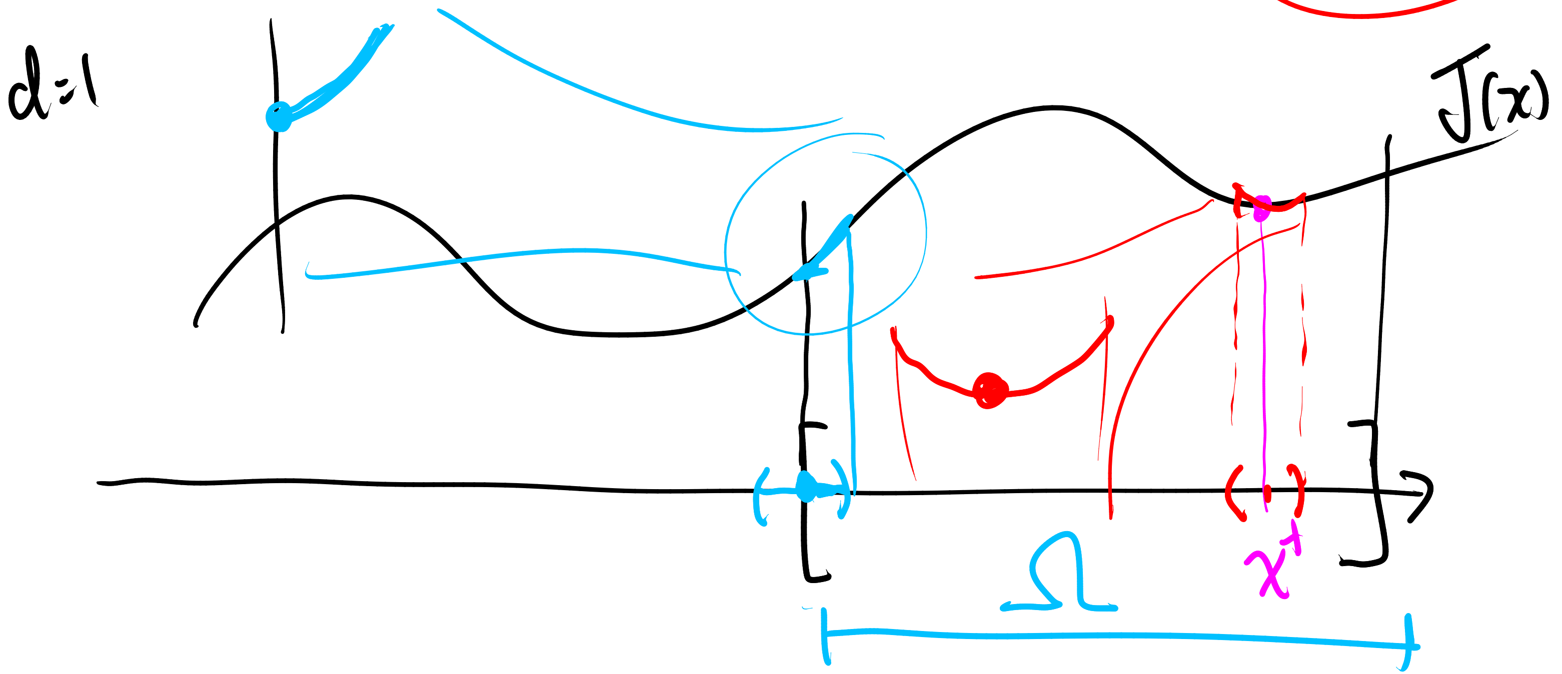
$$\text{s.t.} \quad \left.\begin{array}{l} \delta \geq 0 \\ w \geq 0 \end{array}\right] \text{ 2 inequality constraints}$$

$$2\delta + 2w = \ell \ \big] \text{ 1 equality constraint.}$$

Global vs. local solutions

small ball centered on x^+

- Global solution: $J(x^*) \leq J(x) \qquad \forall\, x \in \Omega$
- Local solution : $J(x^+) \leq J(x) \qquad \forall\, x \in \Omega \cap B_\epsilon(x^+)$

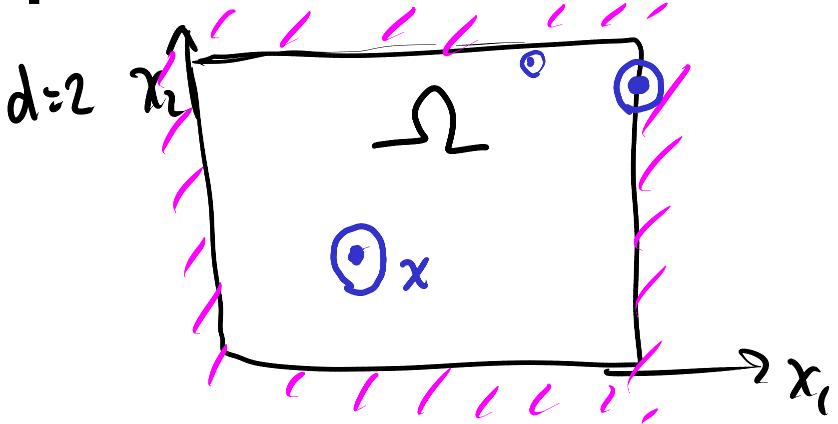


Types of feasible points

- Interior vs. non-interior points ... Ω.

  There is a ball that is contained in Ω

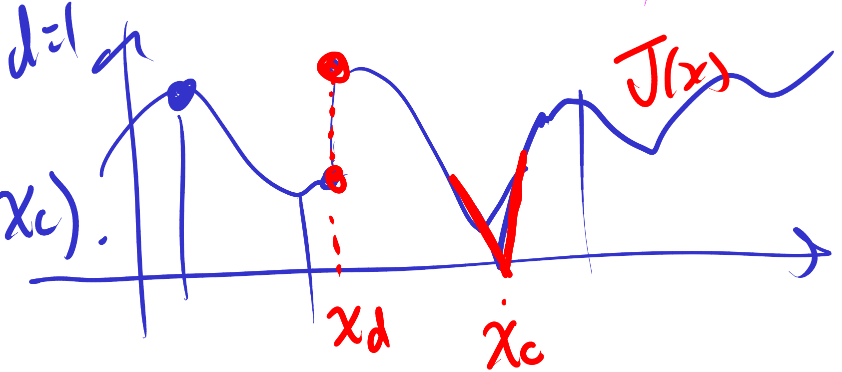
  a.k.a. boundary
  There is no ball



- Differentiable vs. non-differentiable points

  everything else.

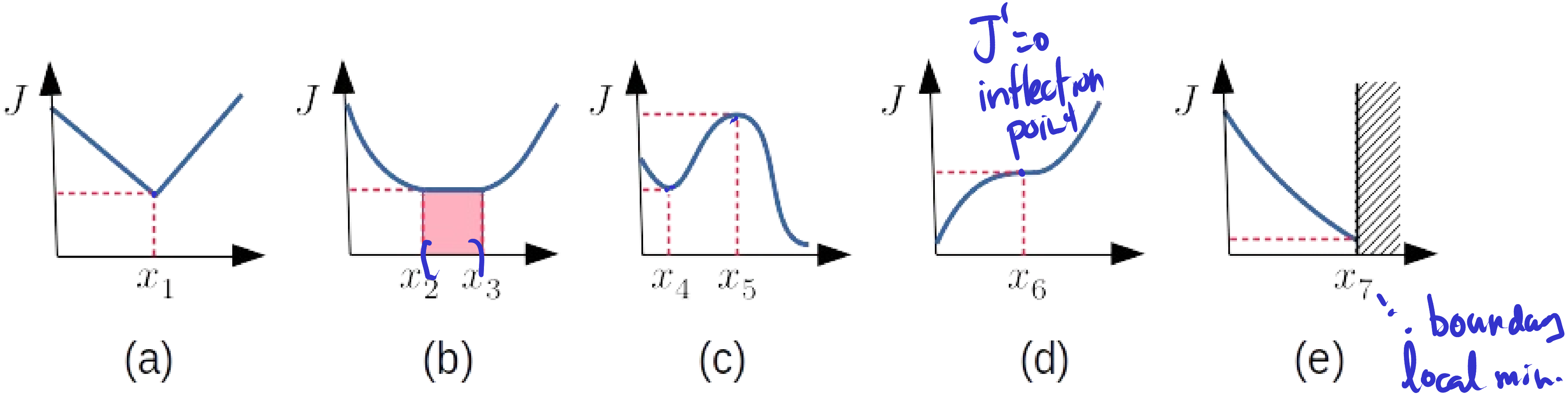
  discontinuous (e.g x_d)
  non-differentiable (e.g. x_c).



- Stationary vs. non-stationary points

  $\nabla J = 0$

  $\nabla J \neq 0$



Stationary points: local minima : J'' (Hessian) > 0.

local maxima $J'' < 0$.

inflection points. otherwise : $J''=0$.

First order optimality condition

x is a differentiable, interior, local solution $\Rightarrow$ x is stationary

All points x:

→ 1. Differentiable and interior .... All solutions in 1st category are stationary.

→ 2. Non-differentiable

3. Non-interior

All local solutions are stationary, Non-diff., or Non-interior

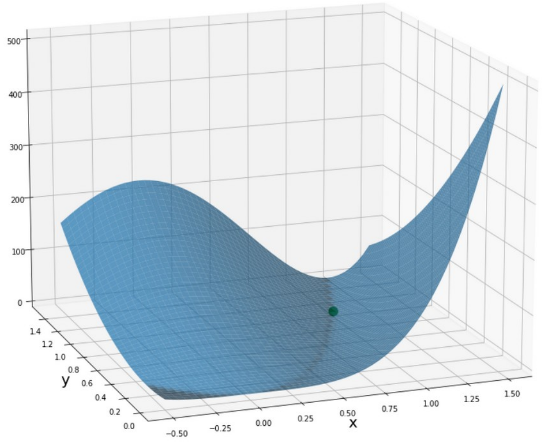
boundar.

Corollary.

If J is smooth. and no constraints.

$\Rightarrow$ all local solutions are amongst the stationary points.

$$\underset{x,y}{\text{minimize}} \quad 100(y-x^2)^2 + (1-x)^2$$



1) $\frac{\partial J}{\partial x} = \cancel{200(y-x^2)(-2x)} + \underline{2(1-x)(-1)} = 0$

$\frac{\partial J}{\partial y} = 200(y-x^2) = 0. \longrightarrow y = x^2$

$\rightarrow x = 1 \Rightarrow y = 1$

$\underline{(x,y) = (1,1)}$

Unique stationary point.

2) Verify local minimum....

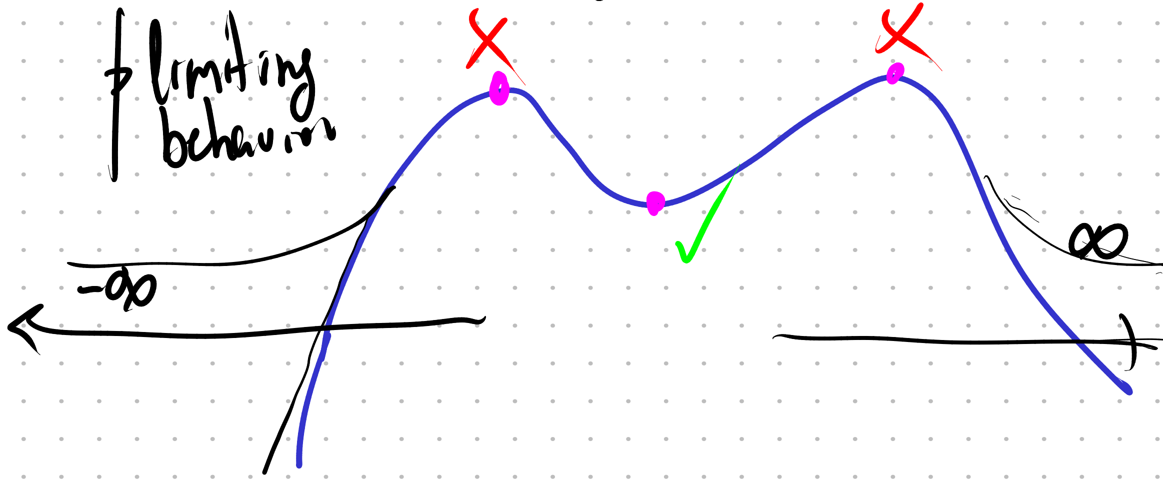
- Check the 2nd derivative.
- Plot it.

---

local minimum $\not\Longrightarrow$ global minimum.

3. Check that a global solution exists.

limiting behavior



No solution here.

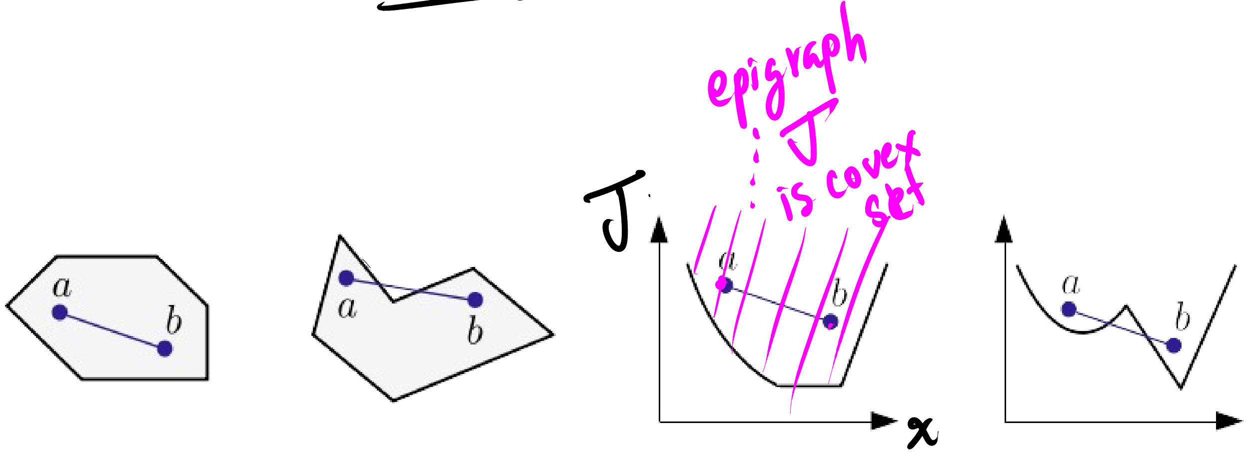
$\cancel{J(x^*) \leq J(x) \quad \forall x \in \Omega}.$

Convex optimization problems

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & J(x) \\ \text{subject to:} \quad & x \in \Omega \end{aligned}$$

convex function

convex set



Convex set Non-convex set Convex function Non-convex function

Properties of convex problems

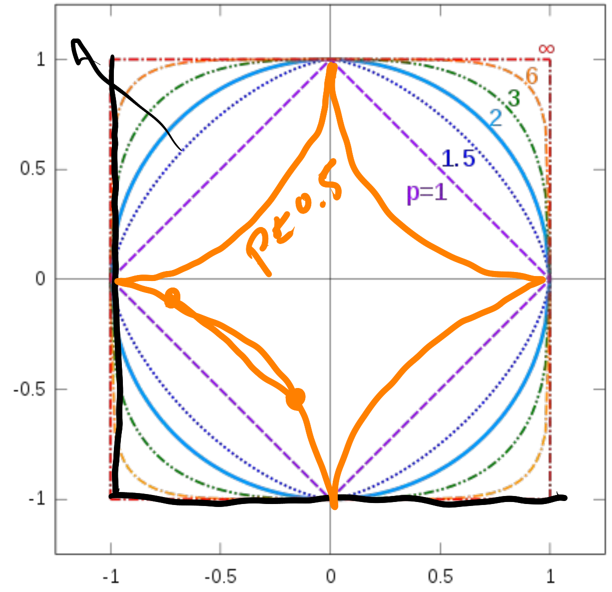
1) For convex problems, every local solution is a global solution.

2) For unconstrained convex problems with continuously differentiable cost function, every stationary point is a global solution.

Examples of convex sets

- p-norm ball : $\{x \in \mathbb{R}^{d} : \|x\|_p \leq r\}$ with $\|x\|_p = \left(\sum_{i=1}^{d} |x_i|^p\right)^{1/p}$

$p \geq 1$



2 norm ball :

$x_1^2 + x_2^2 + x_3^2 \leq 100$ (r^2)

3 norm $\sqrt[3]{|x_1^3| + |x_2^3|} \leq \#$

∞ norm $\max(|x_1|, |x_2|, \dots, |x_d|) \leq \#$

Examples of convex sets

- Affine equality constraints: $Ax = b$

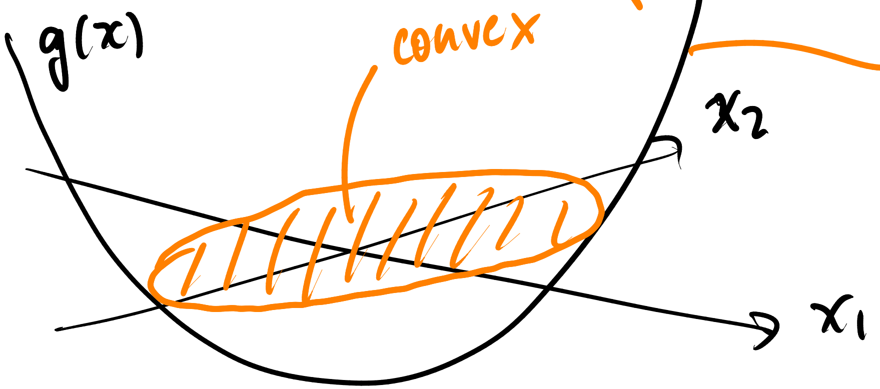
Hyperplane in $\mathbb{R}^d$

convex function.

- Convex inequality constraints: $\boxed{g(x) \leq 0}$

convex.

$d=2$



Examples of convex functions

- Affine functions: $J(x) = a^T x + b$ $\quad a \in \mathbb{R}^d, b \in \mathbb{R}$

- p-norms: $J(x) = \|x\|_p$ $\quad p \geq 1$

- Function composition: $J(x) = g(h(x))$ $\quad$ h convex

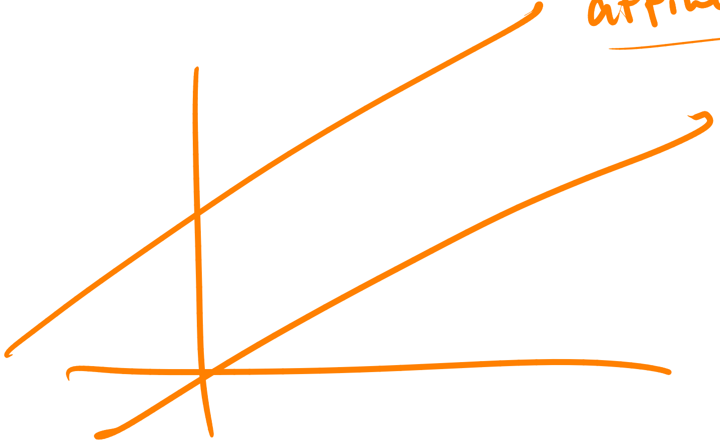
  g ~~nondecreasing~~ affine.

Convex optimization problems

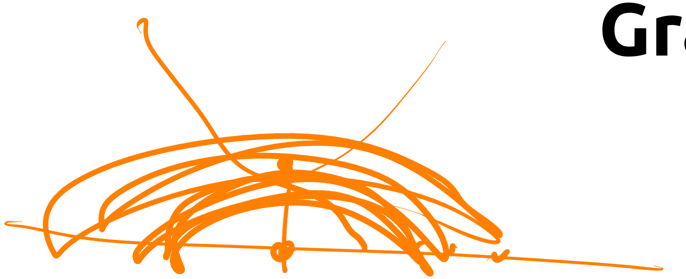
$$
\begin{aligned}
\underset{x}{\text{minimize}} \quad & J(x) && \text{... convex cost function} \\
\text{subject to:} \quad & f_i(x) = 0 \quad i = 1 \dots n && \text{... affine equality constraints} \\
& g_j(x) \leq 0 \quad j = 1 \dots m && \text{... convex inequality constraints}
\end{aligned}
$$

affine : $y = ax + b$

linear : $y = ax$

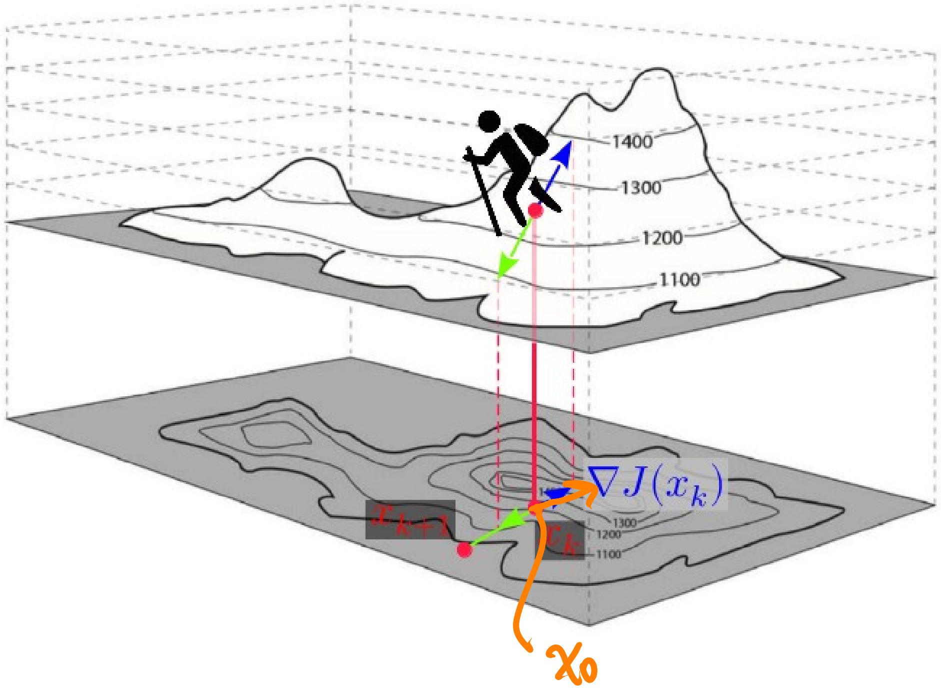


Gradient descent ... finds local minima.



$$x^* = \underset{x}{\text{argmin}}\, J(x)$$

k... step counter.



0. Initialize: $\underline{x_0 \in \Omega}$, $k = 0$
1. Loop until convergence:
   - $x_{k+1} = x_k - \underbrace{\boxed{\gamma}}_{\text{stepsize}} \nabla J(x_k)$
   - $k \leftarrow k + 1$

Stochastic optimization

$\underline{\theta} = (\theta_1, \ldots, \theta_D)$.

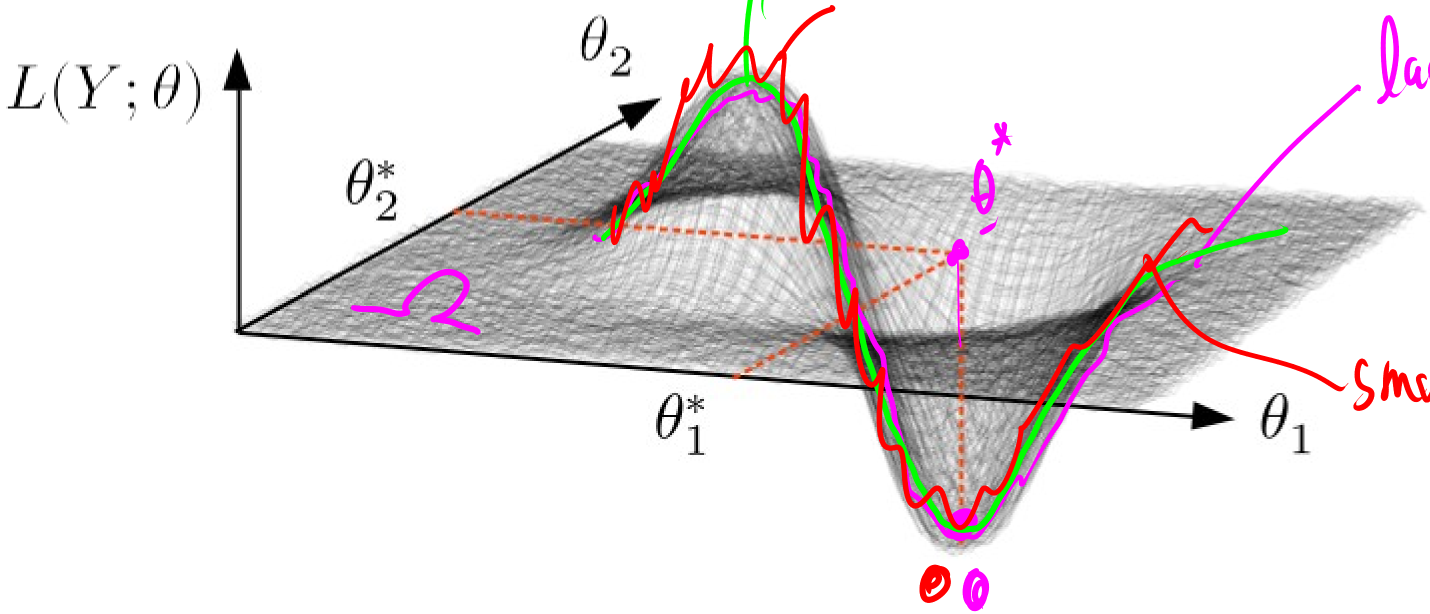
θ ... scalar.

$$\underline{\theta}^* = \underset{\underline{\theta}}{\text{argmin}}\; E[L(Y\,;\,\underline{\theta})]$$

loss.

N is small

Example: $\underline{\theta} = (\theta_1, \theta_2)$



larger N.

smaller N.

Full problem : $$\underline{\theta}^* = \underset{\underline{\theta}}{\text{argmin}}\; E[L(Y\,;\,\underline{\theta})]$$

$$\mathcal{D} = \{y_i\}_N \overset{iid}{\sim} Y$$

Approximate expectation : $$E[L(Y\,;\,\underline{\theta})] \approx \frac{1}{N}\sum_{i=1}^{N} L(y_i\,;\,\underline{\theta})$$

Approximate problem : $$\underline{\theta}^* = \underset{\underline{\theta}}{\text{argmin}}\; \overbrace{\sum_{i=1}^{N} L(\underline{\theta}\,;\,y_i)}^{J(\underline{\theta})}$$

Gradient descent : $$\underline{\theta}_{k+1} = \underline{\theta}_k - \gamma \nabla_{\underline{\theta}} \left(\sum_{i=1}^{N} L(\underline{\theta}_k\,;\,y_i) \right)$$

$$= \underline{\theta}_k - \gamma \sum_{i=1}^{N} \nabla_{\underline{\theta}}\, L(\underline{\theta}_k\,;\,y_i)$$

$\mathcal{D}$.

Stochastic gradient descent (SGD)

$$\underline{\theta}_{k+1} = \underline{\theta}_k - \gamma \sum_{i \in \mathcal{B}} \nabla_{\underline{\theta}} L(\underline{\theta}\,;\ y_i)$$

$\mathcal{B} \subseteq \{1 \dots N\}$

a *batch* of samples

$\mathcal{D} =$

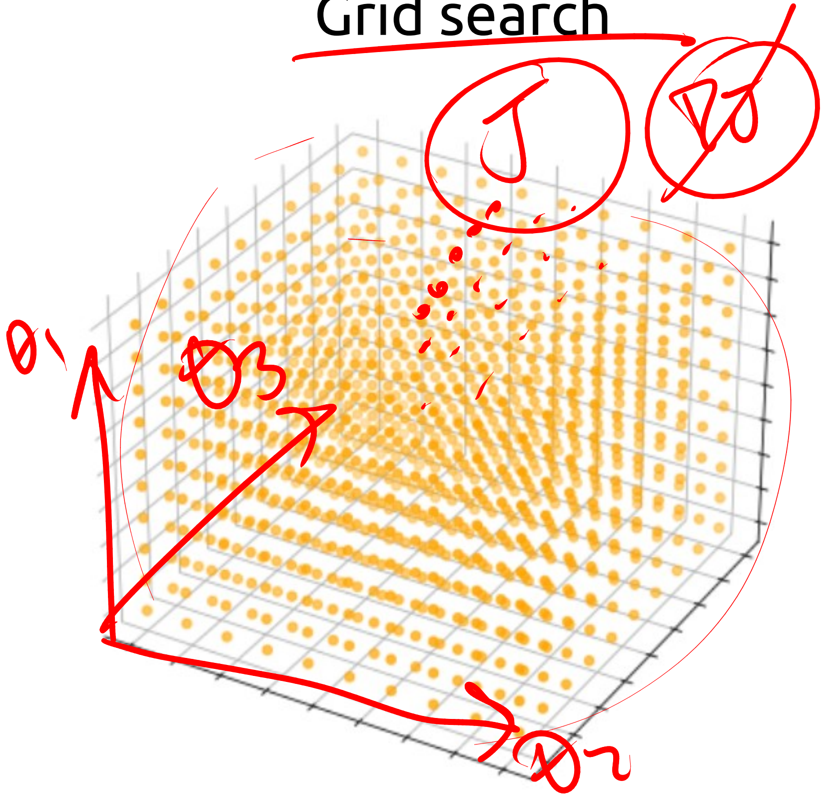
1	2	3	4	5	6																		N

epoch.

- $|\mathcal{B}| = 1$ (pure SGD)
  - $\mathcal{B}$ chosen *with replacement*
  - $\mathcal{B}$ chosen *without replacement*
- $1 < |\mathcal{B}| < N$ (minibatch)
  - partition
  - shuffle / split
- $|\mathcal{B}| = N$ (regular gradient descent)

Gradient-less optimization

Grid search



Genetic algorithms