18.9.24

opti : Etienne Tanré
— Etienne.Tanre@inria.fr
- Fizeau (5^{th} floor) - 005

1

Goal: Optimize.

Max or Min. Inf

$$\inf_{c \in K} f(c) \in \mathbb{R}$$

smallest value

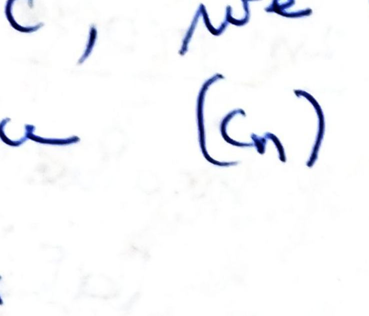
$$c^* = \inf_{c \in K} f(c)$$

- If we don't know c^*, we try to find a sequence (c_n) such that $f(c_n) \xrightarrow[\infty]{} A$

⚠ It does not mean that

$$c_n \longrightarrow c^*$$

Ex: $f : \mathbb{R}_+ \longrightarrow \mathbb{R}$
$t \longmapsto \frac{1}{t}$



$$A = \inf_{t \geq 0} f(t) = 0$$

$f(t_n) = \frac{1}{n} \xrightarrow[\infty]{} 0 = A$

But $t_n \underset{\infty}{\not\to} 0 = t^*$.

Q: Goal $\inf_{c \in K} f(c)$

1) We know an explicit expression of $f(c)$ and are able to study the fonction.

2) In practice: We choose one value c_0. $(f(c_0))$.

- take $c_1 \sim c_0$
- compare $f(c_1)$.

If $f(c_1) < f(c_0) \to$ happy. we're

else $(f(c_1) > f(c_0)) \to$ stay and c_0.

iterate.

After N_1 steps you have final c_N such that, for any trial $\tilde{c}$ around c_{N_1}, $f(\tilde{c}) > f(c_{N_1})$

⚠ local optimal

↳ convexity remove the risk to stay at a local optimal.

Example of f such that the evaluation ~~of~~ $f(c)$ is not easy (or is impossible exactly)

- $f(c) := \mathbb{E}[\varphi(c, X)]$ where the law of X is known, the function φ is known

But the value $f(c)$ in

$\boxed{\text{Algo}}$ • $X^1, \dots, X^{N_{MC}}$ iid

$$\hat{f}^{N_{MC}}(c_0) = \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} \varphi(c_0, X^i)$$

then $\hat{f}^{N_{MC}}(c_0) \xrightarrow[N_{MC} \to \infty]{a.s} f(c_0)$

[thanks to the strong law large Number).

- Size of the error (rate of convergence)

$$\sqrt{N_{MC}} \left(\hat{f}^{N_{MC}}(c_0) - f(c_0) \right) \xrightarrow[N_{MC} \to \infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2)$$

$\hookrightarrow \sigma^2 = \mathbb{V}\left(\varphi(c_0, X)\right)$

$$\frac{\sqrt{N_{MC}}}{\sigma} \left(\hat{f}^{N_{MC}}(c_0) - f(c_0) \right) \xrightarrow[N_{MC} \to \infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

$$\hat{f}^{N_{MC}}(c_0) \simeq f(c_0) + \frac{\sigma}{\sqrt{N_{MC}}} C.$$

18.8.24 opti

②

Berry-Essence théo

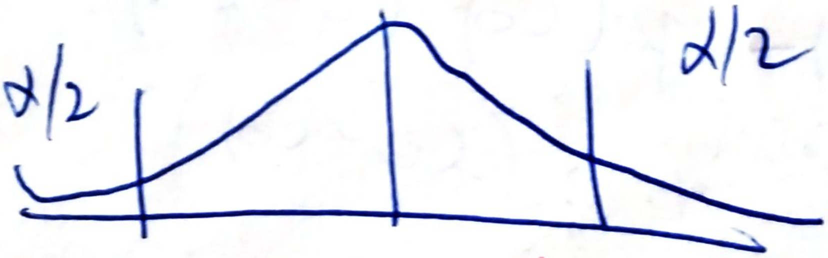
confidence interval

α = level of confidence. (0,95%

$$f(c_0) \in \left[\hat{f}^{N_{MC}}(c_0) - \frac{\sigma\, y_\alpha}{\sqrt{N_{MC}}}, \hat{f}^{N_{MC}}(c_0) + \frac{\sigma}{\sqrt{N_{MC}}} y_\alpha\right]$$

such that:

$\mathbb{P}(|\mathcal{N}(0,1)| \le y_\alpha) = \alpha$.



NB: • the result of MC procedure is an intervalle under a level of intervalle

• It's necessary to estimate the variance.

- If now we know that the function f is smooth.

- from c_0, calculate $f(c_0)$ et $\nabla f(c_0)$

and, $c_1 = c_0 - h f'(c_0)$

and interate.

- Go back: if $f(c) = \mathbb{E}(\varphi(c, X))$

$$f(c) = f(c_0) + f'(c_0)(c - c_0) + o((c - c_0))$$

$$f'(c_0) \simeq \frac{f(c) - f(c_0)}{c - c_0}$$

- $f(c) \simeq \hat{f}^{N_{MC}}(c) \pm \frac{2\sigma^c}{\sqrt{N_{MC}}}$

$f(c_0) \simeq \hat{f}^{N_{MC}}(c_0) \pm \frac{2\sigma c_0}{\sqrt{N_{MC}}}$

$$\hat{f}'(c_0) \approx \frac{\hat{f}^{N_{MC}}(c) - \hat{f}^{N_{MC}}(c_0) \pm \frac{2}{\sqrt{N_{MC}}}(\sigma^c - \sigma^{c_0})}{c - c_0}$$

such that

$\frac{1}{(c-c_0)\sqrt{N_{MC}}}$ is veray smal.

which that $N_{MC} \gg \frac{1}{(c-c_0)^2}$.

Robbins - Monroe Algo .

But imp a perfect oracle.

$$f(c) = \mathbb{E}(\varphi(c, X)) = \int \varphi(c,\theta) f_X(\theta)\, d\theta$$

$$f'(c) = \int \frac{\partial \varphi}{\partial c}(c,\theta) f_X(\theta)\, d\theta$$

$$= \mathbb{E}\left(\frac{\partial \varphi}{\partial c}(c, X)\right)$$

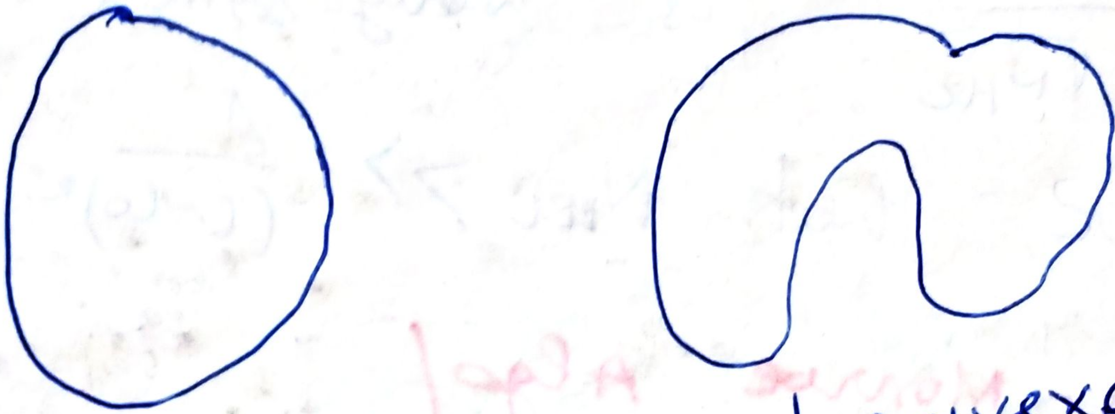
Then.

$$c_1 = c_0 - h \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} \frac{\partial \varphi}{\partial c}(c_0, X^i)$$

Chap 1 : Convexity

I/ Convex sets

Examples



convexe   not convexe.

Def: A set $C \subset \mathbb{R}^d$ is convexe iff

$\forall x, y \in C, \ \forall \theta \in [0,1],$

$\theta x + (1-\theta) y \in C$.

Def of convexe Hull

The convexe hull of a set C is defined by

$$conv(C) = \left\{ \sum_{i=1}^{K} \theta_i x_i, \ \theta_i \in [0,1], \ \sum_{i=1}^{K} \theta_i = 1, \quad K \geq 1, \ x_i \in C \right.$$

18.9.24

(3)

Properties

(1): conv(C) is a convex set

(2): conv(C) is the smallest convex set containing C.

Proof

(1)

$$\begin{cases} x = \sum_{i=1}^{K} \theta_i x_i , \; x_i \in C , \; \theta_i \geq 0 \quad \sum_{i=1}^{K} \theta_i = 1 \\ y = \sum_{i=1}^{L} \eta_i y_i , \; y_i \in C, \; \eta_i \geq 0 \; \sum_{i=1}^{L} \eta_i = 1 \end{cases}$$

$z = \lambda x + (1-\lambda) y$, $\lambda \in [0,1]$, $x,y \in conv(C)$

$= \sum_{i=1}^{K} (\lambda \theta_i) x_i + \sum_{j=1}^{L} (1-\lambda) \eta_j y_j$

posons $n = K+L$

$\forall i \in [\![1, K]\!]$; $z_i = x_i , \gamma_i = \lambda \theta_i$

$\forall i \in [\![K+1, K+L]\!]$, $z_i = y_{i-K} , \gamma_i = (1-\lambda) \eta_{i-K}$

$$z = \sum_{i=1}^{K+L} \gamma_i z_i$$

on a:

$z_i \in C$

$\gamma_i \geq 0$, $\displaystyle\sum_{i=1}^{K+L} \gamma_i = \sum_{i=1}^{K} \lambda \theta_i + \sum_{i=K+1}^{K+L} (1-\lambda) \eta_{i-k}$

$(j = i-k)$

$$\sum_{i=1}^{K+L} \gamma_i = \lambda \underbrace{\sum_{i=1}^{K} \theta_i}_{1} + (1-\lambda) \underbrace{\sum_{j=1}^{L} \eta_j}_{1} = 1$$

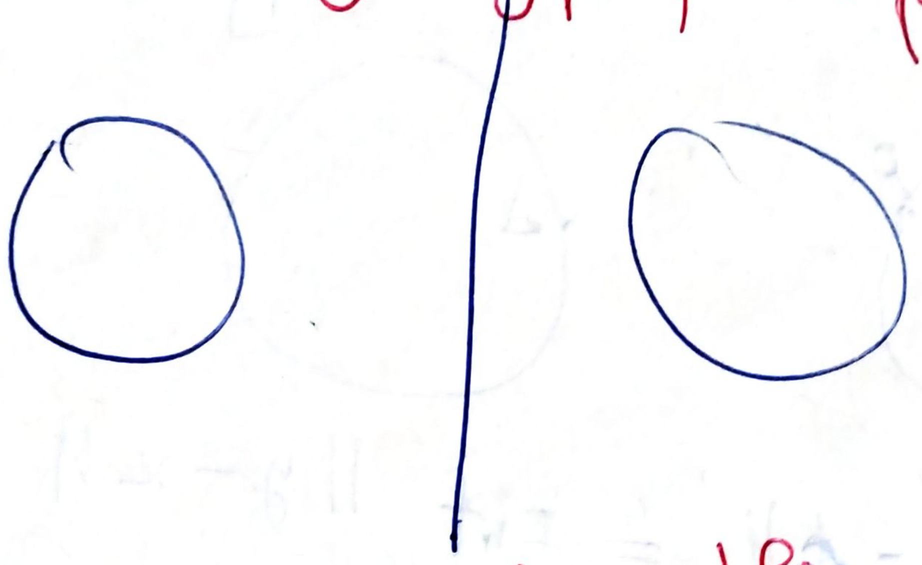
Properties :

Operation preserving convexity

Assume C, C_1, C_2 are convexe sets

- $C_1 + C_2$ is convexe
- $C_1 \times C_2$ is convexe
- $\bigcap_{i \in I} C_i$ is convexe
- f affine applicat° $f(C)$ and $f^{-1}(C)$ are convexe.

P) Separating hyperplane property.



separating hyperplan theo

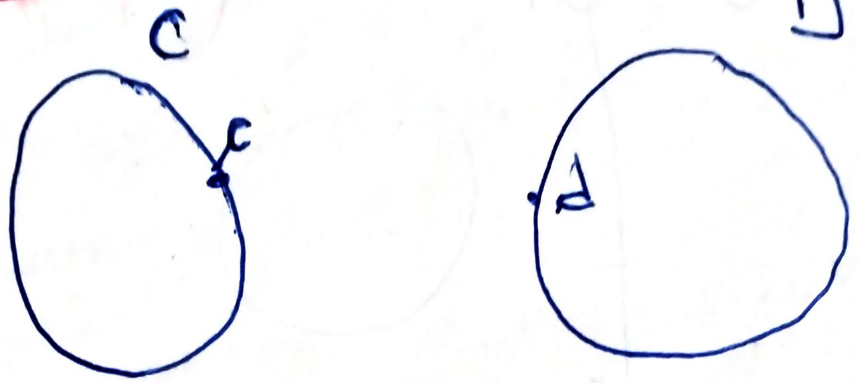
Consider C and D 2 convexes sets closed such that $C \cap D = \phi$

$C \neq \phi$ and $D \neq \phi$; $C, D \subset \mathbb{R}^d$

Then there exists $a \in \mathbb{R}^d$ and $b \in \mathbb{R}$ such that $\forall x \in C$, $a \cdot x \leq b$

$y \in D$, $a \cdot y \geq b$

Proof



$$\|d-c\| = \inf_{\substack{y\in D\\ x\in C}} \|y-x\|$$

Set $a = d-c$

$$b = \frac{\|d\|^2 - \|c\|^2}{2}$$

Prove that

$f(x) = a\cdot x - b$, $f(x) \le 0, \forall x \in C$

$\geq 0 \quad \forall x \in D$?

$$f(x) = (d-c)\cdot\left(x - \tfrac{1}{2}(d+c)\right)$$

Goal: prove that

$$f(y) \geq 0, \quad \forall y \in D.$$

Assume it is not true

$\exists\, u \in D, \ (d-c)\cdot\left(u - \tfrac{1}{2}(d+c)\right) < 0.$

18.9.24

(4)

$$f(u) = (d-c)\cdot\left(u-d+\frac{1}{2}(d-c)\right) < 0.$$

$$= (d-c)\cdot(u-d) + \frac{1}{2}\|d-c\|^2.$$

$$\Rightarrow (d-c)\cdot(u-d) < 0.$$

$$\frac{\partial}{\partial t}\|d-c+t(u-d)\|^2$$

$$= 2(d-c)(u-d) < 0.$$

For small t (<1),

$$\|d-c+t(u-d)\|^2 < \|d-c\|^2.$$

consider $\tilde{d} = d + t(u-d)$

$$= \underbrace{tu}_{\in D} + (1-t)\underbrace{d}_{\in D} \in D$$

so $\tilde{d} \in D$.

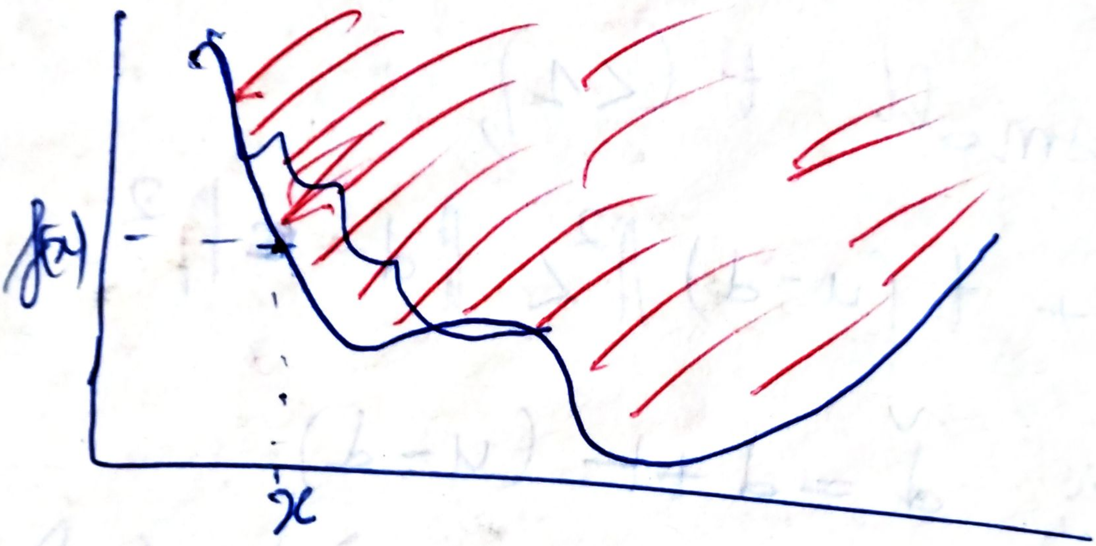
and $\|\tilde{d} - c\|^2 < \|d - c\|^2$

in contradictio

II/ Convexe functions

Def

1) Epigraph of a functi ~~the~~ defined ~~by~~ $C \longrightarrow \mathbb{R}$ by $\{(\lambda, x) \in \mathbb{R} \times C, \lambda \geq f(x)\}$



2) f is a convexe function iff its epigraph is a convexe set

Final EXAM : 18/12/24

24/09/24

course 2: convex function and links with optimization

Def

$f : X \subset \mathbb{R}^d \longrightarrow \overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty ; +\infty\}$

epigraph of f:

$epi(f) = \{ (x, \lambda) \in \mathbb{R}^d : f(x) \leq \lambda \}$

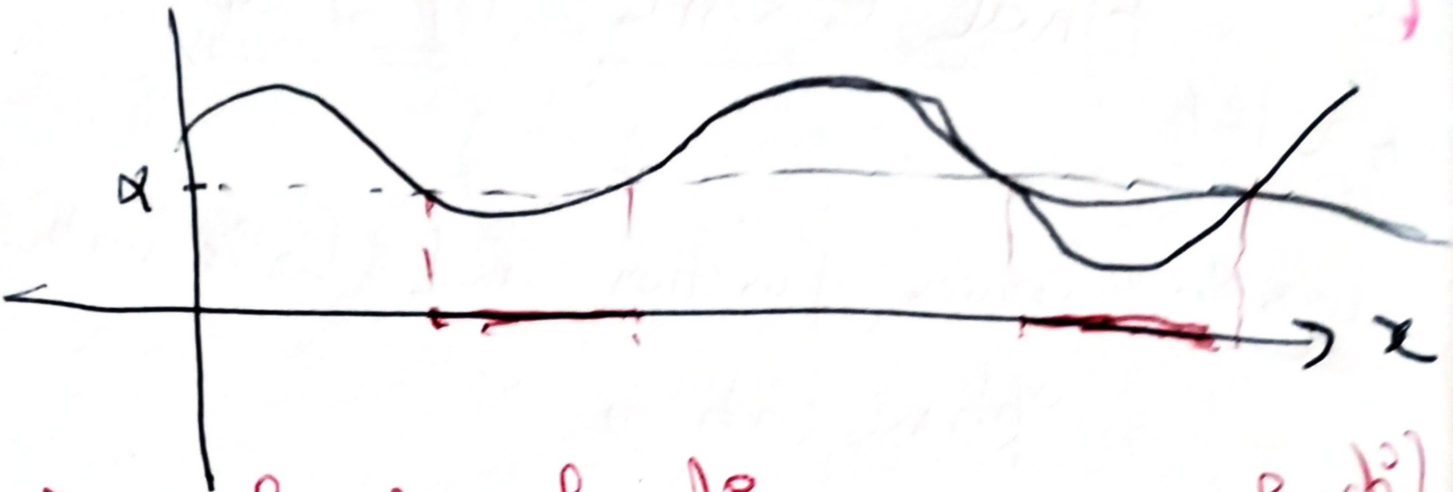
Domain of f

$Dom(f) = \{ x \in X : f(x) < \infty \}$

Sublevel set (of level α)

$lev_\alpha(f) = \{ x \in X : f(x) \leq \alpha \}$

- f is sound to be lower semi-continuous iff $epi(f)$ is closed



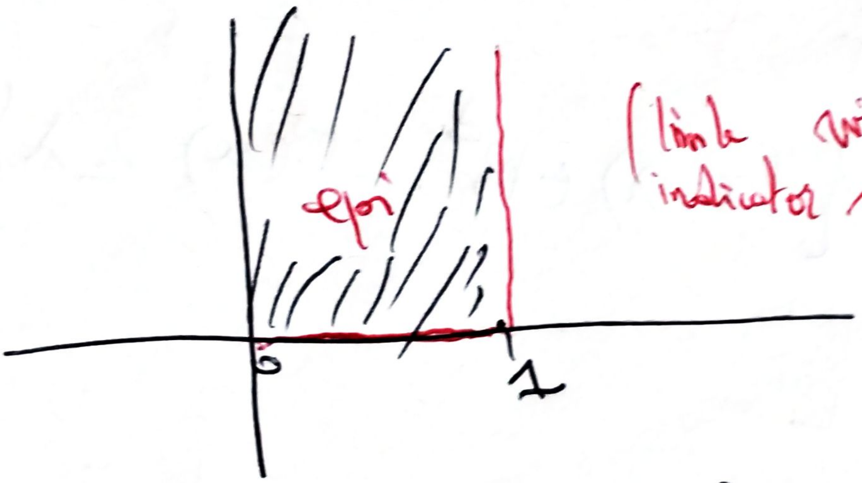
Ex: of lsc functi°

(indicatrice functi°)

• $X = \mathbb{R}$, $f : X \to \mathbb{R}$

$$x \mapsto \begin{cases} 0 & \text{if } x \in [0,1] \\ +\infty & \text{if } x \notin [0,1] \end{cases}$$

(link with "usual" indicator functi° $I = -\log(\text{classical } \mathbb{1})$)



Example of not lsc functi

$g : \mathbb{R} \to \mathbb{R}$

$$x \mapsto \begin{cases} 0 & \text{if } x \in]0,1[\\ +\infty & \text{otherwise} \end{cases}$$

Let be $z_n = (\frac{1}{n}, 1)$.

$\forall n \geq 2$: $z_n \in epi(g)$

$\lim_{n\to+\infty} t_n = (0,1)$

but $(0,1) \notin \mathrm{epi}(g)$

$\mathbb{I}_A : X \longrightarrow \overline{\mathbb{R}}$

$x \longmapsto \begin{cases} 0 & \text{if } x \in A \\ +\infty & \text{if } x \notin A \end{cases}$

$\mathbb{1}_A : X \longrightarrow \overline{\mathbb{R}}$

$x \longmapsto \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$

$\forall x \in X$

$\mathbb{1}_A(x) = \exp(-\mathbb{I}_A(x))$

- f is said to be proper iff it never take value $-\infty$ and it has a non empty domaine.

$\Rightarrow$ $f : X \longrightarrow \overline{\mathbb{R}}$ is convexe iff its epigraph is convex.

Property: The domain of a convex functiᵒ is

Preuve

If $\mathrm{Dom}(f)$ is not convexe then $\mathrm{epi}(f)$ is non convex.

$x, y \in \mathrm{Dom}(f)$

but $z \notin \mathrm{Dom}(f)$

$\hookrightarrow f(z) = +\infty$.

So $\exists \theta \in [0,1]$, $z = \theta x + (1-\theta) y$,

$(x, f(x)) \in \mathrm{epi}(f)$

$(y, f(y)) \in \mathrm{epi}(f)$

$(z, \theta f(x) + (1-\theta) f(y)) \notin \mathrm{epi}(f)$

Sn $\theta f(x) + (1-\theta) f(y) < +\infty = f(z)$

② 10/12/2024

Ex: Value-at-Risk

X is r.v. representing a loss

confidence level α (close to 1)

find $VaR_{\alpha,X}$ such that

$\mathbb{P}(X \leq VaR_{\alpha,X}) = \alpha$.

Naive approch:

F the cdf of X.

$\forall \xi, \; \widehat{F(\xi)}_M = \frac{1}{M} \sum_{k=1}^{M} \mathbb{1}_{X_k \leq \xi}$

We are interested in ξ big.

Remember: iid $Y_k \sim B(p)$.

goal: compute p.

$\hat{p}_M = \frac{1}{M} \sum_{k=1}^{M} Y_k$ error of order. (CLT)

$$\frac{1}{\sqrt{M}} \times \mathbb{E}[Y_1^2]^{1/2} = \frac{\sqrt{p(1-p)}}{\sqrt{M}}$$

relative error

$$\frac{(1-\hat{p}_M)-(1-p)}{1-p}$$ of order

$$\frac{\sqrt{p}}{\sqrt{M}\sqrt{1-p}}.$$

BAD when p close to 1.

suppose

$X = \varphi(Z)$, $Z \sim \mathcal{N}(0,1)$,

then $$\mathbb{P}(X \leq \varepsilon) = \int_{\mathbb{R}} \mathbb{1}_{\varphi(z) \leq \varepsilon} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz$$

[change of variable]

$$= \int_{\mathbb{R}} \mathbb{1}_{\varphi(u+\theta) \leq \varepsilon} \frac{e^{-\frac{(u+\theta)^2}{2}}}{\sqrt{2\pi}} du$$

$$= e^{-\frac{\theta^2}{2}} \mathbb{E}[\mathbb{1}_{\varphi(Z+\theta) \leq \varepsilon} e^{-\theta Z}]$$

②

Properties

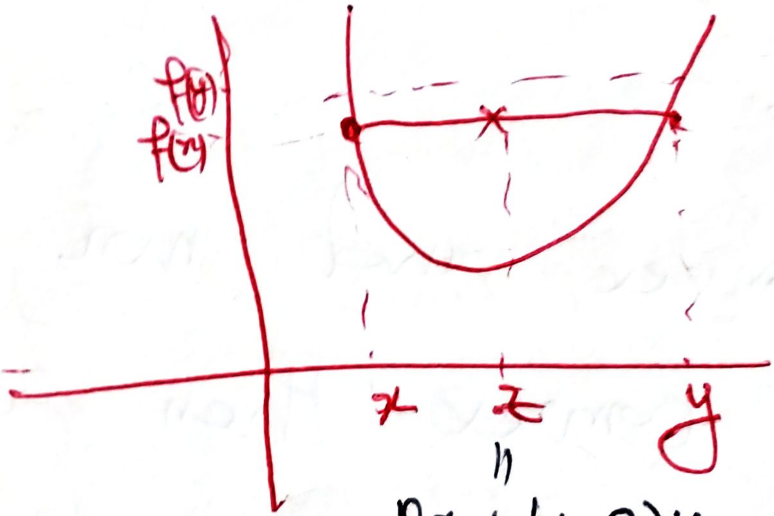
24/09/24

If f is convex, then,

$\forall x, y \in X, \ \forall \theta \in [0,1]$

$$f(\theta x + (1-\theta) y) \leq \theta f(x) + (1-\theta) f(y)$$

Proof



$$z = \theta x + (1-\theta) y$$

Comme l'epi(f) est convexe,

$$(\theta x + (1-\theta) y, \ \theta f(x) + (1-\theta) f(y)) \in \text{epi}(f)$$

$$\iff f(\theta x + (1-\theta) y) \leq \theta f(x) + (1-\theta) f(y)$$

NB: We say that f is ~~convexe~~ concave iff $-f$ is convex

Basic Properties

- Assume f, g are convex and $t \geq 0$ then $tf + g$ is convex.
- $(d=1)$ f convex and non decreasing and g convex then $f \circ g$ convex
- f is convex and a is a "linear" function, then $f \circ g$ is convexe.
- $(f_i)_{i \in I}$ are convex functions, then $\sup_{i \in I} f_i(x)$ is convex

- The domain of a convex functⁿ is convex
- The sublevel sets of a convexe functⁿ are convex (reciprocaly, it is not true
- A convex functⁿ is always above its tangents.
- ## Jensen Inequality

If f is convex, X is a random variable then

$$f(\mathbb{E}(X)) \leq \mathbb{E}(f(X)) \text{ avec}$$

$\mathbb{E}(|X|) < \infty$.

Proof

①: $h = tf + g$

goal: proving that h is convex

i.e $\forall x, y \in X$, $\theta \in [0,1]$,

$$h(\theta x + (1-\theta)y) \leq \theta h(x) + (1-\theta) h(y).$$

— Let $x, y \in X$, $\theta \in [0,1]$.

We know that g, f is convex.

So $f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y)$

, $g(\theta x + (1-\theta)y) \leq \theta g(x) + (1-\theta) g(y)$

Comme $t \geq 0$,

$$t f(\theta x + (1-\theta)y) \leq \theta t f(x) + (1-\theta) t f(y)$$

D'où :

$$h(\theta x + (1-\theta)y) \leq \theta h(x) + (1-\theta) h(y)$$

(3)

24/07 24

- $h(x) = \sup_{i \in I} f_i(x)$

Goal: $\forall x, y \in X$, $\theta \in [0,1]$,

$h(\theta x + (1-\theta)y) \leq \theta h(x) + (1-\theta) h(y)$.

Let $x, y X$ and $\theta \in [0,1]$,

we know that $\forall i \in I$

$f_i(\theta x + (1-\theta)y) \leq \theta f_i(x) + (1-\theta) f_i(y)$

car $\begin{cases} f_i(y) \leq h(y) \Rightarrow (1-\theta) f_i(y) \leq (1-\theta) h(y) \\ f_i(x) \leq h(x) \Rightarrow \theta f_i(x) \leq \theta h(x) \end{cases}$

D'où :

$\sup_{i \in I} f_i(\theta x + (1-\theta)y) \leq \theta h(x) + (1-\theta) h(y)$

$= h(\theta x + (1-\theta)y) \leq \theta h(x) + (1-\theta) h(y)$.

Inégalité de JENSEN

$$\begin{cases} \alpha_i \geq 0 \\ \sum_{i=1}^{n} \alpha_i = 1 \end{cases}, \quad f\left(\sum_{i=1}^{n} \alpha_i x_i\right) \leq \sum_{i=1}^{n} \alpha_i f(x_i)$$

By induction:

$n=1$, $\alpha_1 = 1$, $f(x_1) = 1 \cdot f(x_1)$

$n=2$, $\alpha_2 = 1-\alpha_1$,

$f(\alpha_1 x_1 + (1-\alpha_1) x_2) \leq \alpha_1 f(x_1) + (1-\alpha_1) f(x_2)$

[def of convexi]

Assume we have proved the property for any numbers of points up to $n-1$.

let $\alpha_1, \dots, \alpha_{n-1}, \alpha_n \geq 0$, $\sum_{i=1}^{n} \alpha_i = 1$.

$x_1, \dots, x_n \in X$.

if $\alpha_n = 1$,

$f(\alpha_1 x_1 + \dots \alpha_{n-1} + \alpha_n x_n) = f(x_n) \leq \sum_{i=1}^{n} \alpha_i f(x_i)$

If $\alpha_n \neq 1$.

let $\begin{cases} \tilde{\alpha} = (1-\alpha_n) \\ \tilde{z} = \frac{1}{\tilde{\alpha}} \sum_{i=1}^{n-1} \alpha_i x_i = \sum_{i=1}^{n-1} \underbrace{\frac{\alpha_i}{\tilde{\alpha}}}_{\beta_i} x_i \end{cases}$

then $\sum_{i=1}^{n-1} \beta_i = \frac{\sum_{i=1}^{n-1} \alpha_i}{\tilde{\alpha}} = \frac{1-\alpha_n}{\tilde{\alpha}} = 1.$

I first apply convex inequality

with 2 points $\tilde{z}, x_n$, ~~θt~~

$\theta = \tilde{\alpha}$. $\overbrace{\sum_{i=1}^{n} \alpha_i x_i}$

$f(\theta \tilde{z} + (1-\theta) x_n)$

$\leq \theta f(\tilde{z}) + (1-\theta) f(x_n)$

$\leq \tilde{\alpha} f\left(\sum_{i=1}^{n-1} \beta_i x_i\right) + \alpha_n f(x_n)$ $*$

By assumption in the induction

$f\left(\sum_{i=1}^{n-1} \beta_i x_i\right) \leq \sum_{i=1}^{n-1} \beta_i f(x_i)$

$$\tilde{\alpha} f\left(\sum_{i=1}^{n-1} \beta_i x_i\right) \leq \sum_{i=1}^{n-1} \tilde{\alpha} \beta_i f(x_i) = \sum_{i=1}^{n-1} \alpha_i f(x_i)$$

RHS of ⊛ $\leq \sum_{i=1}^{n} \alpha_i f(x_i)$

That is the property of

② Regularity of convex functi

- $f: \mathbb{R}^d \longrightarrow \mathbb{R} \cup \{-\infty, +\infty\}$, f is convexe. ~~continuo~~

$\Leftrightarrow$ f ist continuous on $\mathbb{R}^n$ iff $\mathrm{dom}(f) = \mathbb{R}^n$.

- f is continuous on the interior of its domain.
- f is lower semi-continuous iff $\mathrm{dom}(f)$ is closed and the restriction of f to its domain is continuous.

(4)

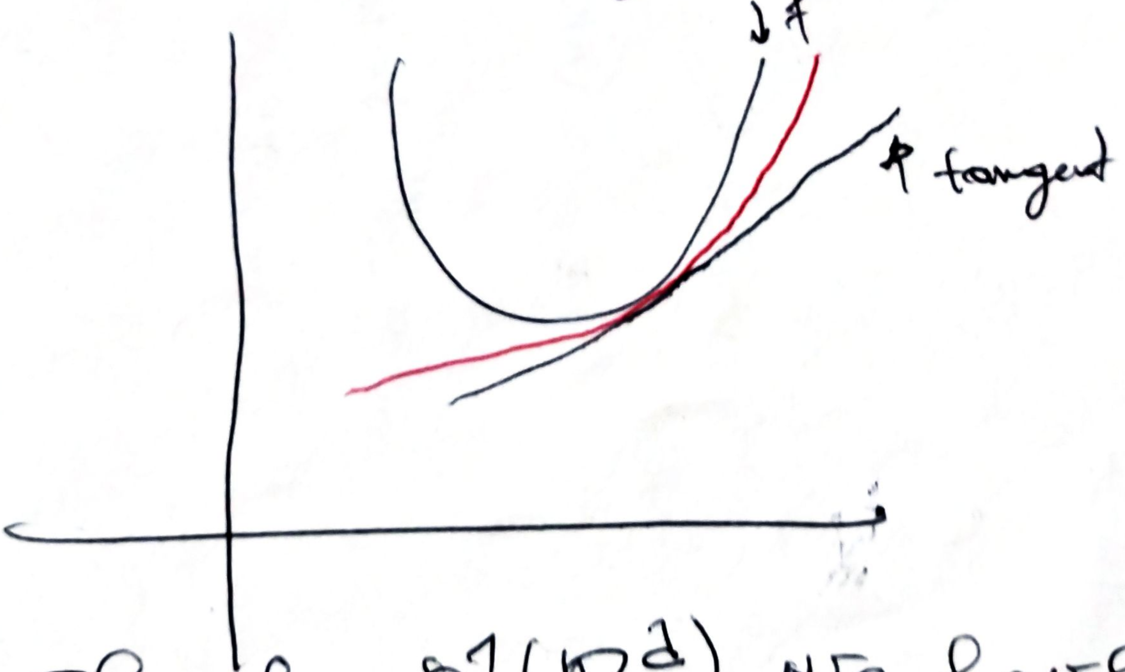
24/09/24

- f is strictly convex iff
$\theta \in]0,1[\; (= (0,1))$, $x \neq y$

$$f(\theta x + (1-\theta)y) < \theta f(x) + (1-\theta) f(y)$$

- f is strongly convex (or α-convex)

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y) - \frac{1}{2}\alpha\theta(1-\theta)\|x-y\|^2.$$



- If $f \in \mathcal{C}^1(\mathbb{R}^d)$, we have
$\langle \nabla f(x) - \nabla f(y) ; x - y \rangle \geq 0$ iff f is convexe.

④ DD

24/09/24

- $f : X \to \mathbb{R} \cup \{+\infty\}$ is α-convexe iff $\forall x, y \in X$

$$f(y) \geqslant f(x) + \langle \nabla f(x), y - \cancel{x} \rangle + \frac{\alpha}{2} \|y - x\|^2.$$

- If $f \in \mathcal{C}^2(\mathbb{R}^d)$,
  - * $\nabla^2 f(x) \geqslant 0$ iff convex
  - if * $\nabla^2 f(x) > 0$, then is strictly convex
  - * If $\nabla^2 f(x) > \alpha I_d$, then f is α-convex.

<u>Exemple</u>

- $\mathbb{I}_C(x) = \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{if } x \notin C \end{cases}$ is Convex iff C convex

- $x \mapsto \exp(ax)$ is convex $\forall a \in \mathbb{R}$.

- $x \longmapsto \|x\|^q$, $\forall q \geq 1$
  and any norm.
  convexe

- $x \longmapsto \log(x)$ Concave

- $x \longmapsto x\log(x)$ convex

- $x \longmapsto \log\left(\sum_{i=1}^{d} \exp(x_i)\right)$ Convex.

Chp 3 : Optimizatio and convex functio

Def: • We say that x^* is a local minimum of f if there exist a neighborhood $(B(x^*, r))$ such that $\forall x \in B(x^*, r)$, $f(x) \geq f(x^*)$

- x^* is a global minimum iff $\forall x \in \mathbb{R}^d$, $f(x) \geq f(x^*)$.

Theo: Assume f is convex.

- A local minimum of f is a global minimum
- The set of global minimum is convexe.

Proof: let x^* be a local minimum.

Assume $\exists z$ such that $f(z) < f(x^*)$.

by convexity assumption, $\forall \theta \in]0,1[$

$$f(\theta x^* + (1-\theta) z) \leq \theta f(x^*) + (1-\theta) f(z) < f(x^*)$$

But any neighborhood of x^* contains points of form

$\theta x^* + (1-\theta) z$

(*) is in contradiction with the fact that x^* is a ~~best min~~ local min.

25-09-24 Chp 2: Gradient descent for convex function

①

Intro

Goal: optimize $\inf_{x \in \mathbb{R}^d} f(x)$

$f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$

Typical algo

1) Initialization: $x^{(0)}$

2) Step k:

$$x^{(k+1)} = x^k - \gamma_k \nabla f(x^k)$$

$\gamma_k > 0$

vocabulary: γ_k is the stepsize or learning rate

Typical value of γ_k

- $\gamma_k = \gamma$ constant
- $\gamma_k = \frac{1}{k}$
- γ_k chosen at each step such that $f(x^k - \gamma_k \nabla f(x^k))$ is as small as possible.

Example : $f(x) = x^2$; $X^* = \{0\}$

$f'(x) = 2x$

- $\gamma_k = \gamma$
- $x^{k+1} = x^k - 2\gamma x^k = (1-2\gamma) x^k$

i) $\gamma = 1$

$x^{(k+1)} = -x^k$

$x^{(2k+1)} = -x^{(0)}$

$x^{(2k)} = x^{(0)}$

suite géo de raison $q=-1$

$U_n = U_0 q^n$

$(x^{(k)})_{k \geq 1}$ does not converg

to $x^* = 0$

ii) $\gamma > 1$; $|x^k| \longrightarrow +\infty$

$$x^{(k)} = (-1)^k \underbrace{|2\gamma - 1|}_{>1}{}^k x^{(0)}$$

(iii) $0 < \gamma < 1$

$$x^{k} = (1-2\gamma)^{k} \, x^{(0)}$$

$$x^{(k)} \xrightarrow[k \to \infty]{} x^{*} = 0$$

and $f(x^{k}) \longrightarrow 0 = f^{*}$

iv) $\gamma_k = \dfrac{1}{3+k}$

$$x^{(k+1)} = \left(1 - \frac{2}{3+k}\right) x^{(k)}$$

d'où: $x^{k} = \prod_{i=0}^{k-1} \left(1 - \frac{2}{3+i}\right) x^{(0)}$

$$= \prod_{i=0}^{k-1} \left(\frac{i+1}{i+3}\right) x^{(0)}$$

(Car $1 - \dfrac{2}{3+k} = \dfrac{k+1}{3+k}$)

$$x^{(k)} = \frac{2}{k(k+1)} \, x^{(0)}$$

(Car $x^{k+1} = (1-2\gamma_k)\, x^{k} = (1-2\gamma_k)(1-2\gamma_{k-1})\, x^{k-1}$
$= (1-2\gamma_k) \cdots (1-2\gamma_0)\, x^{(0)}$

$$x^{(k)} \xrightarrow[k \to +\infty]{} x^* = 0$$

v) $\gamma_k = \dfrac{1}{2(k+2)^2}$

$$x^{(k+1)} = \left(1 - \frac{1}{(k+2)^2}\right) x^{(k)}$$

$$x^{(k)} = \prod_{i=2}^{k+1} \left(1 - \frac{1}{i^2}\right) x^{(0)}$$

<u>Rappel</u> : $A = \prod_{i=2}^{N} \left(1 - \frac{1}{i^2}\right) = \frac{1}{2}$

$$\log A_N = \sum_i \log\left(1 - \frac{1}{i^2}\right)$$

$$\log\left(1 - \frac{1}{i^2}\right) = \log\left(\frac{(i+1)(i-1)}{i \times i}\right)$$

$$= \log(1+i) + \log(1-i) - 2\log(i)$$

$$\sum_{i=2}^{N} \log\left(1 - \frac{1}{i^2}\right) = \sum_{i=3}^{N-1} \log(i) + \log(N+1)$$

$$+ \log N + \sum_{i=3}^{N-1} \log(i) + \log 2$$

$$- 2\log(2) - 2\sum_{i=3}^{N-1} \log i - 2\log N$$

25.09.24

②

$$= -\log(2) + \log\left(1+\frac{1}{N}\right)$$

$$\xrightarrow[N\to+\infty]{} -\log 2$$

So $A_N = \exp\left(\sum_{i=2}^{N} \log\left(1-\frac{1}{i^2}\right)\right) \to \frac{1}{2}$

<u>Conclusion</u>

Even in this very simple case.

- The algo diverges
- — converges to x^*
- — an — to another value.

<u>Propositi°</u>

f: C^1 functi° on $\mathbb{R}^d$.

f is convex iff $\forall x, y \in \mathbb{R}^d$

$$f(y) \geq f(x) + \langle \nabla f(x), y-x \rangle$$

Proof

- Assume f is convex:

$\cancel{f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y)}$

$$f(\theta y + (1-\theta)x) \leq \theta f(y) + (1-\theta) f(x)$$

$$f(\theta y + (1-\theta)x) - f(x) \leq \theta (f(y) - f(x))$$

$$\underbrace{\frac{f(\theta y + (1-\theta)x) - f(x)}{\theta}} \leq (f(y) - f(x))$$

car $\theta > 0$.

$\cancel{\text{et } \lim_{\theta \to 0} \frac{B}{B} =}$

et lorsque $\theta \to 0$,

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$$

- let $x, y \in \mathbb{R}^d$, $\theta \in [0, 1]$

$$z = \theta x + (1-\theta) y$$

$f(x) \geq f(z) + \langle \nabla f(z), x - z \rangle$

$f(y) \geq f(z) + \langle \nabla f(z), y - z \rangle$

$\theta f(x) + (1-\theta) f(y) \geq f(z) + \langle \nabla f(z), \underbrace{\theta(x - z) + (1-\theta)(y - z)}_{B} \rangle$

or $B = z(\theta - 1 - \theta) + \theta x + (1-\theta) y$

$= \theta x + (1-\theta) y - z = 0$

d'où:

$\theta f(x) + (1-\theta) f(y) \geq f(z)$.

Corolary

if f is convex and C^1.

$x^* \in X^*$ iff $\nabla f(x^*) = 0$

Proof: If $\nabla f(x^*) = 0$

$f(y) \geq f(x^*) + \underbrace{\langle \nabla f(x^*), y - x^* \rangle}_{= 0}$.

$\Rightarrow f(y) \geq f(x^*)$

Def

Let $f : \mathbb{R}^d \to \mathbb{R}$ & let $L > 0$, we say that f is L-smooth iff

$\forall x, y \in \mathbb{R}^d$,

$\|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\|$.

Lemma

$f : \mathbb{R}^d \to \mathbb{R}$ L-smooth, then we have

$$f(x) - f(y) - \langle \nabla f(y), x - y \rangle \leq \frac{L}{2} \|x - y\|^2.$$

25-09-24

(3)

$(f\circ g)' = g' \cdot f'(g)$

Proof : x, y are fixed

$h : \mathbb{R}^d \longrightarrow \mathbb{R}$

$h(t) = f(y + t(x-y))$.

$$\begin{cases} h(0) = f(y) \\ h(1) = f(x). \end{cases}$$

$$f(x) - f(y) = h(1) - h(0) = \int_0^1 h'(t)\,dt$$

$$= \int_0^1 \langle \nabla f(y + t(x-y)), x - y \rangle\, dt$$

$$f(x) - f(y) - \langle \nabla f(y), x - y \rangle = \int_0^1 \langle \nabla f(y + t(x-y)) - \nabla f(y), x - y \rangle\, dt$$

$$\leq \int_0^1 \| \nabla f(y + t(x-y)) - \nabla f(y) \| \times \| x - y \|\, dt$$

$$\leq L \int_0^1 \| t(x-y) \| \, \| x - y \|\, dt$$

$$\underbrace{\leq L \|x-y\|^2 \int_0^1 t \, dt}_{= \frac{L}{2} \|x-y\|^2 .}$$

Lemma: If f is convex and L-smooth, then $\forall x, y \in \mathbb{R}^d$,

$$\cancel{f(x) - f(y) \leq \langle} \nabla f(x), x-y \rangle - \frac{1}{2L} \| \nabla f(x) - \nabla f(y)$$

$$f(x) - f(y) \leq \langle \nabla f(x), x-y \rangle$$

$$- \frac{1}{2} \| \nabla f(x) - \nabla f(y) \|^2$$

Corollary : $f \in C^1$ and $\mathbb{R}^d \to \mathbb{R}$

Equivalence between

a) f is L-Smooth and convexe

b) $f(x) - f(y) - \langle \nabla f(x) ; y - x \rangle$

$$\leq \frac{L}{2} \|x - y\|^2 . \quad (*)$$

Théo

$f : \mathbb{R}^d \to \mathbb{R}$ is convexe and L-smooth,

$x^* \in X^*$, $x^{(0)} \in \mathbb{R}^d$

$$x^{(k+1)} = x^{(k)} - \frac{1}{L} \nabla f(x^{(k)})$$

then we have

$$0 \leq f(x^{(k)}) - f(x^*) \leq \frac{2L}{k} \|x^{(0)} - x^*\|^2$$

Proof

1) We apply $(*)$ with ~~x~~

$x = x^{(k+1)} = x^k - \frac{1}{L} \nabla f(x^k)$, $y = x^{(k)}$

$$f(x^{(k+1)}) - f(x^{k}) - \langle \nabla f(x^{k}), -\frac{1}{L}\nabla f(x^{k}) \rangle$$

$$\leq \frac{L}{2} \| -\frac{1}{L} \nabla f(x^{k}) \|^2$$

$$f(x^{(k+1)}) - f(x^{k})$$

$$\leq \underbrace{\frac{1}{2L} \| \nabla f(x^{(k)}) \|^2 - \frac{1}{L} \| \nabla f(x^{k}) \|^2}$$

$$= -\frac{1}{2L} \| \nabla f(x^{k}) \|^2$$

$k \mapsto f(x^{(k)})$ is decreasing

Lemma: ~~f convex and L-smooth~~

$f(x) - f$

(2) $\langle \nabla f(x) - \nabla f(y), x-y \rangle = \langle \nabla f(x), x-y \rangle$

$$- \langle \nabla f(y), x-y \rangle \geq f(x) - f(y) + \frac{1}{2L} \| \nabla f(x) - \nabla f(y) \|^2 - \langle \nabla f(y), x-y \rangle$$

$$\geq \frac{1}{2L} \| \nabla f(x) - \nabla f(y) \|^2.$$

25-09.24

(4)

we apply with $\begin{cases} x = x^k \\ y = x^* \end{cases}$

$$\langle \nabla f(x^k), x^{(k)} - x^* \rangle$$

$$\geq \frac{1}{2L} \| \nabla f(x^k) \|^2 .$$

R_f: $\| x^{(k+1)} - x^* \|^2$

$$= \| x^k - \frac{1}{L} \nabla f(x^k) - x^* \|^2$$

$$= \| x^k - x^* \| + 2 \langle -\frac{1}{L} \nabla f(x^k), x^k - x^* \rangle$$

$$+ \| \frac{1}{L} \nabla f(x^k) \|^2 .$$

$$\leq \| x^k - x^* \|^2 - \frac{1}{L^2} \| \nabla f(x^k) \|^2$$

$$+ \frac{1}{L^2} \| \nabla f(x^{(k)}) \|^2 .$$

$$\leq \| x^k - x^* \|^2 .$$

③ We assume $x^k \neq x^*$

$$\delta_k = f(x^k) - f(x^*)$$

$$\Rightarrow \delta_{k+1} \leq \delta_k - \frac{1}{2L} \| \nabla f(x^{(k)}) \|^2$$

$$\delta_k = f(x^{(k)}) - f(x^*)$$

$$\leq -\langle \nabla f(x^k), x^* - x^k \rangle$$

d'après l'inégalité de C.S

$$\leq \| \nabla f(x^k) \| \, \| x^k - x^* \|$$

$$\leq \| \nabla f(x^k) \| \cdot \| x^0 - x^* \|$$

Donc, $\| \nabla f(x^k) \| \geq \dfrac{\delta_k}{\| x^{(0)} - x^* \|}$

$$\delta_{k+1} \leq \delta_k - \frac{1}{2L} \| \nabla f(x^k) \|^2$$

$$\leq \delta_k - \Big(\frac{1}{\underbrace{2L \| x^{(0)} - x^* \|}_{\omega}} \Big) \delta_k^2$$

$\Leftrightarrow \omega \delta_k^2 + \delta_{k+1} \leq \delta_k$.

$$\cancel{\frac{1}{\delta_k \delta_{k+1}}}$$

(multiplier par $\frac{1}{\delta_k \delta_{k+1}}$)

$$\omega \frac{\delta_k}{\delta_{k+1}} + \frac{1}{\delta_k} \leq \frac{1}{\delta_{k+1}}$$

$$\frac{1}{\delta_{k+1}} - \frac{1}{\delta_k} \geqslant \frac{\delta_{k+1}}{\delta_k}\left(\frac{1}{\delta_{k+1}} - \frac{1}{\delta_k}\right) \geqslant \omega$$

$$\frac{1}{\delta_k} \geqslant \frac{1}{\delta_k} - \frac{1}{\delta_0} = \sum_{j=0}^{k-1}\left(\frac{1}{\delta_{j+1}} - \frac{1}{\delta_j}\right) \geqslant \omega k$$

$$\delta_k \leq \frac{1}{k\omega}.$$

$$0 \leq f(x^k) - f(x^*) \leq \frac{2L \|x^{(0)} - x^*\|^2}{k}.$$

Lemma: $f : \mathbb{R}^d \longrightarrow \mathbb{R}$ is α convex

iff $\varphi(x) = f(x) - \frac{\alpha}{2}\|x\|^2$ is

convex.

Propositi°

$f : \mathbb{R}^d \longrightarrow \mathbb{R}$ is α-convexe and

$\mathcal{C}^1$, there esxists a unig

minimizer.

Lemma: Consider f $\mathcal{C}^1$
equivalence between

(a) f is α- convex

(b) $f(y) \geqslant f(x) + \langle \nabla f(x), y-x \rangle$
$+ \frac{\alpha}{2}\|x-y\|^2$.

Theo: f α-convex and L-smooth

$x^{(0)} \in \mathbb{R}^d$: the initial condition

$\boxed{\gamma = \frac{2}{\alpha+L}}$:on $x^{k+1} = x^k - \gamma \nabla f(x^k)$

25.09.24

5

then we have $L \geqslant \alpha$

$$\forall k \geqslant 1, \quad \|x^{(k)} - x^*\| \leq \left|\frac{L-\alpha}{L+\alpha}\right|^k \|x^{(0)} - x^*\|$$

and

$$0 \leq f(x^{(k)}) - f(x^*) \leq \frac{L}{2}\left|\frac{L-\alpha}{L+\alpha}\right|^{2k} \|x^{(0)} - x^*\|^2$$

Proof: $L \geqslant \alpha$.

$$\|x^{(k+1)} - x^*\|^2 = \|x^{(k)} - \gamma \nabla f(x^{(k)}) - x^* - \nabla f(x^*)\|^2$$

$- \nabla f(x^*)$

$$= \|x^{(k)} - x^*\|^2 + \gamma^2 \|\nabla f(x^{(k)})\|^2$$

$$- 2\gamma \langle \nabla f(x^{(k)}) - \nabla f(x^*), x^{(k)} - x^* \rangle$$

chp: Stochastic optimisation

I/ Goal: $h(y) = \mathbb{E}[H(y, Z)]$

where Z is a random variable with law μ

("it is easy to simulation independent realisations of μ")

Consider $(Z_1, ..., Z_n)$ n random variables i.i.d with law μ.

Goal: Find the minimizer of h

Typical scheme:

Initialization, say Y

and for any n,

$$Y_{n+1} = Y_n - \gamma_n \frac{\partial H}{\partial y}(Y_n, Z_{n+1})$$

We assume that Y_0 is independent

of $(Z_k)_{k\geq 1}$

$(\gamma_n)_{n\geq 1}$ is a deterministic sequen

Ex 1

$$Y_{n+1} = Y_n - \gamma_{n+1}(Y_n - Z_{n+1}).$$

It corresponds to

$$H(y, z) = \frac{1}{2}(y - z)^2.$$

$$h(y) = \frac{1}{2}\mathbb{E}[(y - Z)^2]$$

$$= \frac{1}{2}[y^2 - 2y\,\mathbb{E}(Z) + \mathbb{E}(Z^2)]$$

Minimizes y^*,

$$h'(y) = 0 \iff y^* = \mathbb{E}(Z)$$

$$\boxed{Y_n - y^* = (1-\gamma_n)(Y_{n-1} - y^*) + \gamma_n(Z_n - y^*)}$$

By induction

$$Y_n - y^* = \overbrace{\prod_{k=1}^{n} (1-\gamma_k)\overset{a_0}{(Y_0 - y^*)}}^{(1)} + \underbrace{\sum_{i=1}^{n} \underbrace{\gamma_i \prod_{k=i+1}^{n} (1-\gamma_k)(Z_i - y^*)}_{A_i}}_{(2)}$$

- If Y_0 is deterministic, (1) is deterministic : it is the deterministic ERROR.
- (2) is the RANDOM ERROR.

Notation and Reminder

$$\mathbb{E}(Z_i) = y^*$$

$$\mathbb{E}\left(\|Z_i - y^*\|^2\right) = \sigma^2 > 0.$$

$$\mathbb{E}\left[\|Y_n - y^*\|^2\right] = \prod_{k=1}^{n} (1-\gamma_k)^2 \|Y_0 - y^*\|^2$$

$$+\sum_{i=1}^{n} \gamma_i^2 \prod_{k=i+1}^{n} (1-\gamma_k)^2 \sigma^2 .$$

Notati° $\quad Y_n - y^* = a_0 + \sum_{i=1}^{n} A_i$

$$(Y_n - y^*)^2 = a_0^2 + \sum_{i=1}^{n} A_i^2$$

$$+ 2a_0 \sum_{i=1}^{n} A_i + 2 \sum_{i<\tilde{i}} A_i A_{\tilde{i}}$$

$\forall i \ \mathbb{E}(A_i) = 0$

for $i < \tilde{i}$

$$\mathbb{E}[A_i A_{\tilde{i}}] = \overbrace{\gamma_i \prod_{k=i+1}^{n} (1-\gamma_k) \gamma_{\tilde{i}} \prod_{e=\tilde{i}+1}^{n} (1-\gamma_e)}^{B}$$

$$\mathbb{E}\left[(Z_i - y^*)(Z_{\tilde{i}} - y^*)\right]$$

$$= B \, \mathbb{E}(Z_i - y^*) \, \mathbb{E}(Z_{\tilde{i}} - y^*) = 0$$

If we hope to have

$Y_n \longrightarrow y^*$ in L^2

16.10.2024

②

iff $\prod_{k=1}^{n} (1-\gamma_k)^2 \longrightarrow 0$

and $\sum_{i=1}^{n} \gamma_i^2 \prod_{k=i+1}^{n} (1-\gamma_k)^2 \xrightarrow[n\to+\infty]{} 0.$

* $\log\left(\prod_{k=1}^{n}(1-\gamma_k)^2\right)$

$\gamma_k \xrightarrow[k\to+\infty]{} 0$

$= 2\sum_{k=1}^{n} \log(1-\gamma_k) \sim -2\sum_{k=1}^{n}\gamma_k$

We deduce that we need

$$\boxed{\sum_{k=1}^{n} \gamma_k \xrightarrow[n\to\infty]{} \infty}$$

* For the second term [the "noise" term]

$\sum_{i=1}^{n} \gamma_i^2 \prod_{k=i+1}^{n} (1-\gamma_k)^2 \sim \sum_{i=1}^{n} \gamma_i^2 \prod_{k=i+1}^{n} (1-2\gamma_k)$

<u>Assume</u> $\gamma_n \downarrow$ and γ_n is bounded by $\frac{1}{k}$

$$\prod_{i=k+1}^{n} (1-K\gamma_i) \leq \prod_{j=\ell+1}^{n} (1-K\gamma_j) \leq 1$$

$$\log(1-x) \leq -x$$

$$1-x \leq \exp(-x)$$

Rappel

$$\sum_{k=1}^{n} \prod_{i=k+1}^{n} (1-K\gamma_i)\gamma_k^2 \qquad \boxed{m<n}$$

$$= \sum_{k=1}^{m} \prod_{i=k+1}^{n} (1-K\gamma_i)\gamma_k^2$$

$$+ \sum_{k=m+1}^{n} \prod_{i=k+1}^{n} (1-K\gamma_i)\gamma_k^2$$

$$\cancel{\leq \prod_{i=m+1}^{n} (1+K}$$

$$\leq \prod_{i=m+1}^{n} (1-K\gamma_i) \sum_{k=1}^{m} \gamma_k^2 +$$

$$\gamma_m \sum_{k=m+1}^{n} \prod_{i=k+1}^{m} (1-K\gamma_i)\gamma_k$$

$$\le \exp\left(-k\sum_{i=m+1}^{n}\gamma_i\right)\sum_{k=1}^{m}\gamma_k^2$$

$$+ \frac{\gamma_m}{k}\sum_{k=m+1}^{n}\left[\prod_{j=k+1}^{n}(1-k\gamma_i) - \prod_{i=k}^{n}(1-k\gamma_i)\right]$$

$$= \prod_{j=k+1}^{n}(1-k\gamma_i)\,[\underbrace{1-(1-k\gamma_k)}_{k\gamma_k}]$$

$$\gamma_k = \frac{1}{k}\left(1-(1-k\gamma_k)\right)$$

$$\le \exp\left(-k\sum_{i=m+1}^{n}\gamma_i\right)\sum_{k=1}^{m}\gamma_k^2$$

$$+ \frac{\gamma_m}{k}\left(1-\prod_{i=m+1}^{n}(1-k\gamma_i)\right)$$

$$\le \exp\left(-k\sum_{i=m+1}^{n}\gamma_i\right)\sum_{k=1}^{m}\gamma_k^2 + \frac{\gamma_m}{k}$$

It is sufficient to have

$$\lim_{n\to\infty} \exp\left(-k\sum_{i=m+1}^{n}\gamma_i\right)\sum_{k=1}^{m}\gamma_k^2 = 0$$

Take Typically $m = n/2$

$$\gamma_n = \frac{c}{n^\alpha}, \quad \alpha \neq 1.$$

$\boxed{\alpha > 1}$, $\sum_{i=1}^{n} \frac{1}{i^\alpha} < \infty$,

$$\lim_{n \to +\infty} \sum_{i=m+1}^{n} \gamma_i > 0.$$

The algo fails.

$\boxed{\alpha \in [0,1]}$

$$\sum_{i=1}^{n} \frac{1}{i^\alpha} = C\, n^{1-\alpha} + O(1)$$

We have

$$\lim_{m \to +\infty} \exp\left(-K \sum_{i=m+1}^{n} \gamma_i\right) \sum_{k=1}^{m} \gamma_k^2 = 0$$

exponentially fast.

$\boxed{\alpha = 1}$, $\sum_{i=1}^{n} \frac{1}{i} = \log(n) + \gamma_{euler} + O\left(\frac{1}{n}\right)$

But $\sum_{i \geq 1} \frac{1}{i^2} < \infty$

16·10·2024

(9)

$\gamma_k = \frac{c}{k}$

$$\sum_{k=1}^{n} \prod_{i=k+1}^{n} (1 - K\gamma_i)\gamma_k^2$$

$$\leq \sum_{k=1}^{n} \exp\left(-K \sum_{i=k+1}^{n} \gamma_i\right)\gamma_k^2$$

$$\cancel{\simeq \sum_{k=1}^{n} \exp\left(-KC\left[\log(n) + k\log(k)\right)\right)} \times \frac{C^2}{k^2}$$

$$\cancel{\simeq \frac{C^2}{n^{KC}} \sum_{k=1}^{n} \frac{1}{k^{2-CK}}}$$

$$\simeq \sum_{k=1}^{n} \exp\left(-KC\left[\log n - \log k\right]\right)\frac{C^2}{k^2}$$

$$\simeq \frac{C^2}{n^{KC}} \sum_{k=1}^{n} \frac{1}{k^{2-CK}} \simeq O\left(\frac{1}{n}\right)$$

Going back to the rate, we forget the initial conditi°

we find $\frac{1}{n^{KC}}$ and so if c is too small, it can be very slow.

Somme Reminders in Probability

Def 1: Random walk

~~Ran~~

Consider a sequence of iid $(X_i)_{i\geq 1}$, $S_n = \sum_{i=1}^{n} X_i = \sum_{i=1}^{n-1} X_i + X_n$

$$\mathbb{E}(S_n | \mathcal{F}_{n-1}) = \mathbb{E}\left(\sum_{i=1}^{n-1} X_i \Big| \mathcal{F}_{n-1}\right) + \mathbb{E}(X_n | \mathcal{F}_n)$$

$$= \sum_{i=1}^{n-1} X_i + \mathbb{E}(X_n) = S_{n-1}$$

if $\mathbb{E}(X_1) = 0$, then

- Martingale:

$$\mathbb{E}[M_{n+1} | \mathcal{F}_n] = M_n.$$

- Sub Martingale:

$$\mathbb{E}(M_{n+1} | \mathcal{F}_n) \geq M_n.$$

- superMartingal

$$\mathbb{E}[M_{n+1} | \mathcal{F}_n) \leq M_n.$$

<u>Theo</u>

Every positive super Martingale converges almost surely to a non-negative integrale limit.

<u>Proof</u>

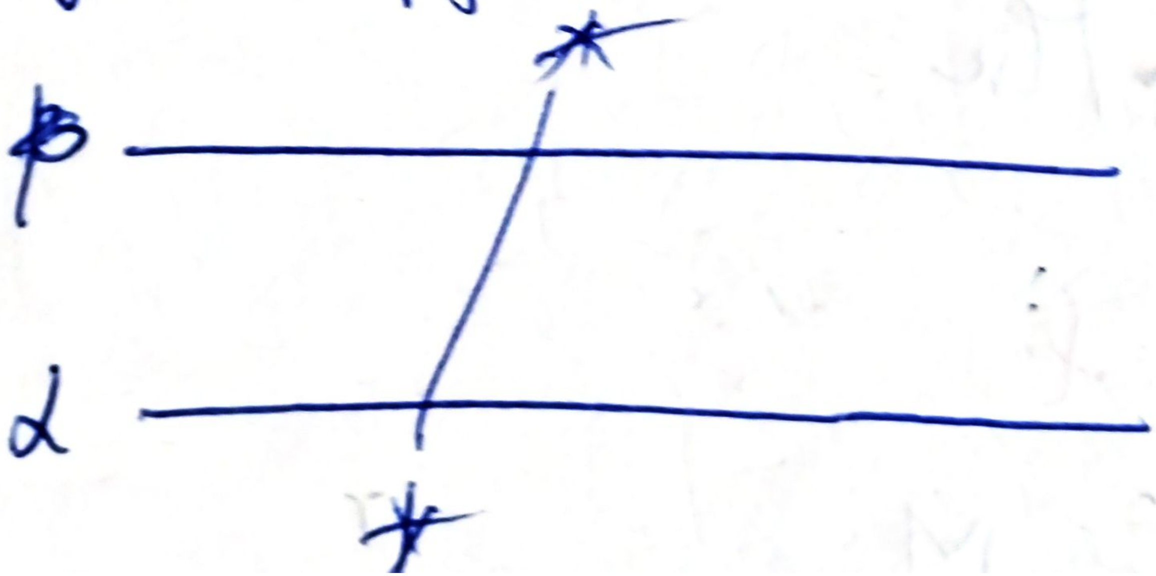
$\forall \omega \in \Omega \setminus N$ (with $\mathbb{P}(N) = 0$)

$$\lim M_n(\omega) = M_\infty(\omega)$$

$\forall \omega \quad M_\infty(\omega) \geq 0$

$\mathbb{E}(M_\infty) < \infty$

- first approach



We count the number of times the Martingale crosses a bound

22.10.2024 Stochastic optimizati° ①

Lemma: Key result (Dubin's inequal + ?)

A positive $\mathcal{F}_n$-supermartingale
$(X_n)_{n \geq 1}$ $\mathbb{E}(X_{n+1} | \mathcal{F}_n) \leq X_n$
cannot oscillate between two levels infinitely often.

We fix two leveles
$0 < a < b < +\infty$.

We define two sequences of stopping times.

Rappel

τ is a stopping time iff $\{\tau \leq n\}$ is $\mathcal{F}_n$-measurable

Example

Consider a sequence (X_n) of Bernoulli random variable with parameter $p > 1/2$.

$$S_n = \sum_{k=1}^{n} 2(X_k - 1)$$

$$\sigma_1 = \operatorname{Inf}\{k \geq 0 ; X_k \leq a\}$$

$$\tau_1 = \operatorname{Inf}\{k \geq \sigma_1 ; X_k \geq b\}$$

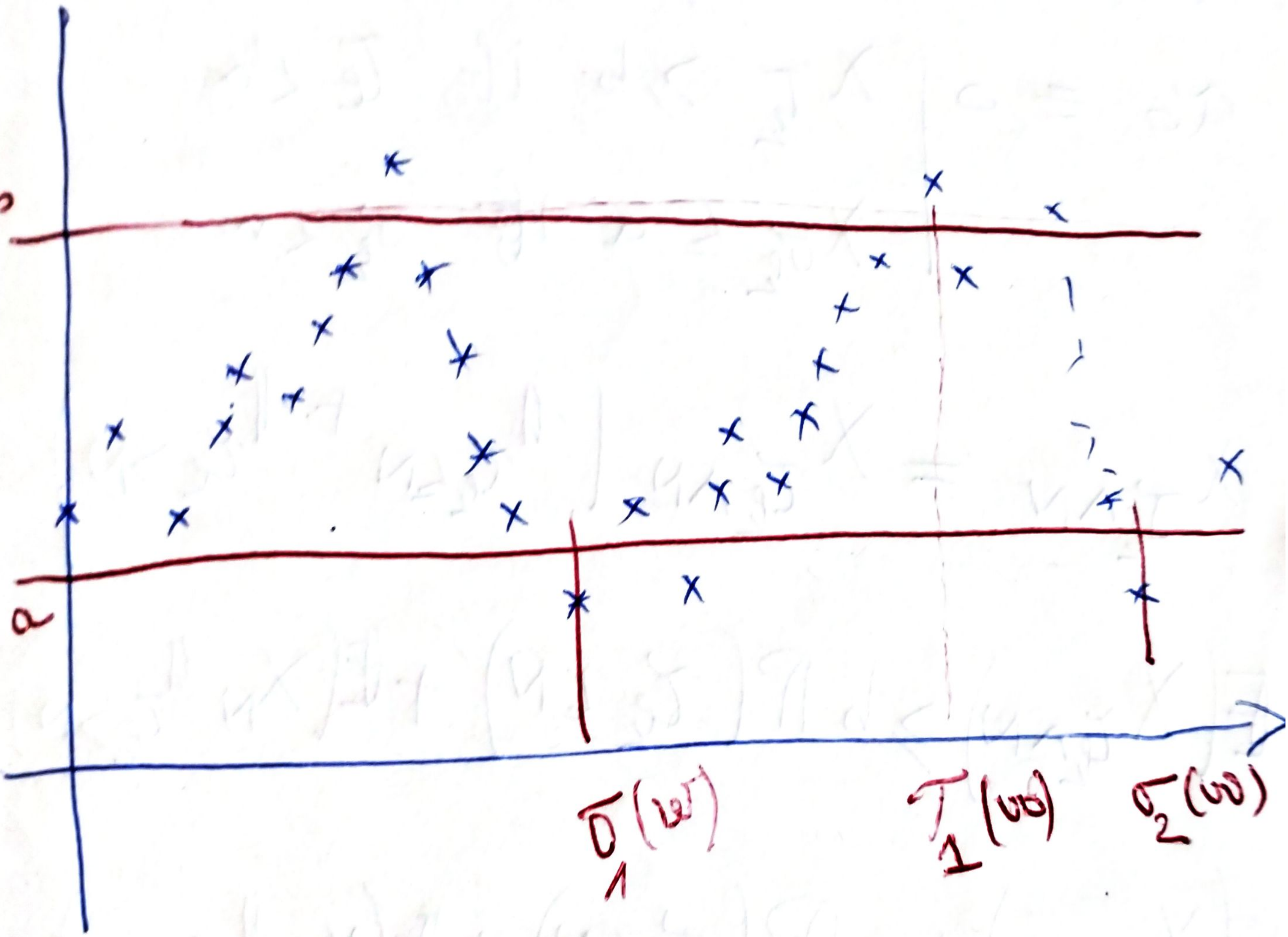
$$\sigma_2 = \operatorname{Inf}\{k \geq \tau_1 , X_k \leq a\}$$

$$\tau_2 = \operatorname{Inf}\{k \geq \sigma_2 ; X_k \geq b\}$$

$$\sigma_{l+1} = \operatorname{Inf}\{k \geq \tau_l ; X_k \leq a\}$$

$$\tau_{l+1} = \operatorname{Inf}\{k \geq \sigma_{l+1} ; X_k \geq b\}$$

$$\operatorname{Inf}(\phi) = +\infty$$



Dubins Inequality

$(X_n)_{n\geq 0}$ a positive supermartingale

$$\mathbb{P}(\tau_k < \infty) \leq \left(\frac{a}{b}\right)^k$$

Proof

First choose a fixed N.

$$\tau_0 \equiv 0 \quad \begin{cases} X_{\tau_k} \geq b & \text{if } \tau_k < \infty \\ X_{\sigma_k} \leq a & \text{if } \sigma_k < \infty \end{cases}$$

$$X_{\tau_k \wedge N} = X_{\tau_k \wedge N} \left(\mathbb{1}_{\tau_k \leq N} + \mathbb{1}_{\tau_k > N} \right)$$

$$\mathbb{E}\left(X_{\tau_k \wedge N}\right) \geq b\, \mathbb{P}\left(\tau_k \leq N\right) + \mathbb{E}\left(X_N \mathbb{1}_{\tau_k > N}\right) \quad (*)$$

$$\mathbb{E}\left(X_{\sigma_k \wedge N}\right) \leq a\, \mathbb{P}\left(\sigma_k \leq N\right) + \mathbb{E}\left(X_N \mathbb{1}_{\sigma_k > N}\right) \quad (**)$$

Rq

if X_n is a supermartingale and $m \leq n$, then $\mathbb{E}(X_n) \leq \mathbb{E}(X_m)$

car $\mathbb{E}\left(\mathbb{E}\left(X_n \mid \mathcal{F}_m\right)\right) \leq \mathbb{E}(X_m)$

Reminder

- For Martingales if σ and τ are 2 bounded stopping times such that $\sigma \leq \tau$, $\mathbb{E}(X_\sigma) = \mathbb{E}(X_\tau)$
- For supermartingales

$$\cancel{\mathbb{E}(X_\sigma) \leq \mathbb{E}(X}$$

$$\mathbb{E}(X_\tau) \leq \mathbb{E}(X_\sigma)$$

The previous reminder gives

$$\mathbb{E}(X_{\tau_k \wedge N}) \leq \mathbb{E}(X_{\sigma_\ell \wedge N})$$

(*)

and (**) $\Rightarrow$ $b\,\mathbb{P}(\tau_\ell \leq N) \leq \mathbb{E}(X_{\tau_\ell \wedge N}) - \mathbb{E}(X_N \mathbb{1}_{\tau_\ell > N})$

$$\leq \mathbb{E}(X_{\sigma_\ell \wedge N}) - \mathbb{E}(X_N \mathbb{1}_{\tau_\ell > N})$$

$$\leq a\,\mathbb{P}(\sigma_\ell \leq N) + \mathbb{E}(X_N \{\underbrace{\mathbb{1}_{\sigma_\ell > N} - \mathbb{1}_{\tau_\ell > N}}_{B}\})$$

$$\mathcal{B} = -\left(1 - \mathbb{1}_{\tau_k \leq N}\right) + \mathbb{1}_{\tau_k > N}$$

$$= \underbrace{\underbrace{\mathbb{1}_{\tau_k > N} + \mathbb{1}_{\tau_k \leq N}}_{\geq 1} - 1}_{\geq 0}.$$

So:

$$b\,\mathbb{P}(\tau_k \leq N) \leq a\,\mathbb{P}(\sigma_k \leq N)$$

$$\leq a\,\mathbb{P}(\tau_{k-1} \leq N)$$

$$\mathbb{P}(\tau_k \leq N) \leq \left(\frac{a}{b}\right) \mathbb{P}(\tau_{k-1} \leq N)$$

$$\leq \left(\frac{a}{b}\right)^k \mathbb{P}(\tau_0 \leq N) = \left(\frac{a}{b}\right)^k$$

<u>Rk</u> If a positive supermartingale hits 0, then it is constant equal to 0 after the first hitting time of 0.

Corollary: The probability that a positive ~~sup~~ supermartingale completes at least k upcrossing before N is less than $\left(\frac{a}{b}\right)^k$.

Theorem

Every positive (X_n) supermartingale converges almost surely to a non-negative limit X_∞ and $\mathbb{E}(X_\infty) < \infty$.

Proof

$$\Delta = \left\{ w : \liminf_n X_n(w) < \limsup_n X_n(w) \right\}$$

Our goal is to prove that $\mathbb{P}(\Delta) = 0$.

$D^{a,b} = \{\omega, \liminf_n X_n(\omega) < a < b < \limsup X_n(\omega)\}$

On $D^{a,b}$, we must have $\tau_k < \infty$ for every $k - 1$.

$\forall k, \mathbb{P}(D^{a,b}) \leq \left(\frac{a}{b}\right)^k \Rightarrow \mathbb{P}(D^{a,b}) = 0$ *

$$D = \bigcup_{\substack{a<b \\ a \in \mathbb{Q} \\ b \in \mathbb{Q} \\ a > -\infty \\ b < +\infty}} D^{a,b}$$

R7

$\forall n \in \mathbb{N}, \mathbb{P}(A_n) = 0 \Rightarrow \mathbb{P}(\bigcup_{n \in \mathbb{N}} A_n) = 0$

but:

$X \sim \mathcal{U}([0,1])$

$\forall a \in [0,1], \mathbb{P}(X \in \{a\}) = 0$

$\mathbb{P}(X \in \bigcup_{a \in [0,1]} \{a\}) = 1.$

We know that $\forall a, b \in \mathbb{Q}, a < b$,

$\mathbb{P}(D^{a,b}) = 0$.

We know that the set $\{a < b, a, b \in \mathbb{Q}\}$ is countable.

So $P(D) = \mathbb{P}\left(\bigcup D^{a,b}\right) = 0$

countable
$a<b$
$a, b \in \mathbb{Q}$

So with proba 1, $w \in D^c$,
that is $\liminf X_n(w) = \limsup X_n(w)$
which means that $X_n(w) \xrightarrow[n \to +\infty]{a.s} X_\infty(w)$

Theo (Robbins - Sigmund lemma)
The proof : Gilles Pages Book "Numerical probability" Chp 6

$h : \mathbb{R}^d \longrightarrow \mathbb{R}^d$

$H : \mathbb{R}^d \times \mathbb{R}^d \longrightarrow \mathbb{R}^d$, $h(y) = \mathbb{E}[H(y, Z)]$

We assume that there exists a functiº $L : \mathbb{R}^d \to \mathbb{R}_+$, C^1 such that
* ∇L is lipschitz continuous,

$\exists\, \hat{C},\ \forall y, \tilde{y} \in \mathbb{R}^d.$

$|\nabla L(y) - \nabla L(\tilde{y})| \leq \hat{C}\, |y - \tilde{y}|$

* $\exists\, C,\ |\nabla L(y)|^2 \leq C\,(1 + L(y))$

* $\nabla L(y) \cdot h(y) \geq 0$

* (sub-linear growth assumption)

$\forall\, y \in \mathbb{R}^d,\ \|H(y, Z)\|_2 \leq C\sqrt{1 + L(y)}.$

Let $(\gamma_n)_{n \geq 1}$ be a sequence of positive real number satisfying

$\sum_{n \geq 1} \gamma_n = \infty$ and $\sum_{n \geq 1} \gamma_n^2 < \infty.$

We assume that Y_0 is independent of the sequence ~~of~~ $(Z_n)_{n \geq 1}$ of iid realizati° of Z

Then the stochastic algo

$Y_{n+1} = Y_n - \gamma_{n+1} H(Y_n, Z_{n+1})$ satisfies the 5 properties.

i) $Y_n - Y_{n-1} \xrightarrow[n\to\infty]{a.s} 0$ and in $L^2(\mathbb{R})$

$\sum_n |Y_n - Y_{n-1}|^2 < \infty$ a.s

ii) The sequence $(L(Y_n))_{n\geq 1}$ is bounded in $L^1(\mathbb{P})$

iii) $L(Y_n) \xrightarrow[n\to\infty]{a.s} L_\infty$ and $\mathbb{E}(L_\infty) < \infty$

iv) $\sum_{n\geq 1} \gamma_n \nabla L(Y_{n-1}) \cdot h(Y_{n-1}) < \infty$ a.s and has a finite expectatio

v) The sequence

$$M_n^{(\gamma)} = \sum_{k=1}^{n} \gamma_k \left(H(Y_{k-1}, Z_k) - h(Y_{k-1}) \right)$$

is an L^2-bounded Martingale and so converges in L^2 and a.s

Rq: Vocabulary

- the sequence $(\gamma_n)_{n\geq 1}$ is called:
  - the step sequence
  - the gain parameter sequence

- If in addition we assume

$$L(y) \xrightarrow[|y|\to\infty]{} \infty,$$ then L is called

a Lyapunov functi

Proof

$\mathcal{F}_n = \sigma(Y_0, Z_1, \dots, Z_n)$

$\Delta Y_n = Y_n - Y_{n-1}$

(1) $\exists\, \xi_{n+1}$ between Y_n and Y_{n+1}

such that

$$L(Y_{n+1}) = L(Y_n) + \nabla L(\xi_{n+1})\cdot \Delta Y_{n+1}$$

$$= L(Y_n) + \big(\nabla L(\xi_{n+1}) - \nabla L(Y_n)\big)\cdot \Delta Y_{n+1}$$

$$+ \nabla L(Y_n)\cdot \Delta Y_{n+1}$$

$$\leq L(Y_n) + \nabla L(Y_n) \cdot \Delta Y_{n+1} + \hat{C} |\Delta Y_{n+1}|^2$$

(∇L is $\hat{C}$ lipschitz continuous

and $(\xi_{n+1} - Y_n) \cdot (Y_{n+1} - Y_n) \leq |Y_{n+1} - Y_n|^2$

$$L(Y_{n+1}) \leq L(Y_n) - \gamma_{n+1} \nabla L(Y_n) \cdot H(Y_n, Z_{n+1}) + \hat{C} \gamma_n^2 |H(Y_n, Z_{n+1})|^2$$

$H(Y_n, Z_{n+1}) - h(Y_n) + h(Y_n)$

$$L(Y_{n+1}) \leq L(Y_n) - \gamma_{n+1} \nabla L(Y_n) \cdot h(Y_n) - \gamma_{n+1} \nabla L(Y_n) \cdot \Delta M_{n+1} + \hat{C} \gamma_{n+1}^2 |H(Y_n, Z_{n+1})|^2 \quad (*)$$

Where $\Delta M_{n+1} = H(Y_n, Z_{n+1}) - h(Y_n)$.

Now, we prove that:

ΔM_n is a sequence of square integrable $\mathcal{F}_n$ - martingale increments which satisfies

$$\mathbb{E}\left(|\Delta M_{n+1}|^2 \mid \mathcal{F}_n\right) \leq \hat{c}\,(1+L(Y_n))$$

for appropriate $\hat{c}$.

- By constructio $\left|\begin{array}{l} Y_n \text{ is } \mathcal{F}_n\text{-measurable} \\ \Delta M_n \text{ is } \mathcal{F}_n\text{-measurable} \end{array}\right.$

. then we prove that (by inducti°)

$\mathbb{E}(L(Y_n)) < \infty$ and $\mathbb{E}[|H(Y_n, Z_{n+1})|^2] < \infty$

We use (A) to prove this property

by inducti°,

$$\mathbb{E}[\nabla L(Y_n) \cdot H(Y_n, Z_{n+1})] \leq \frac{1}{2}\mathbb{E}[|\nabla L(Y_n)|^2] + \frac{1}{2}\mathbb{E}[|H(Y_n, Z_{n+1})|^2]$$

$$a \cdot b \leq \frac{1}{2}\left(|a|^2 + |b|^2\right)$$

- Y_n is $\mathcal{F}_n$-measurable
- Z_{n+1} is independent of $\mathcal{F}_n$

$$\mathbb{E}[|H(Y_n, Z_{n+1})|^2 \mid \mathcal{F}_n) = \mathbb{E}[|H(y, Z_1)|^2]\Big|_{y=Y_n}$$

$$\mathbb{E}\left[\mathbb{E}\left[|H(y,Z_1)|^2\right]\Big|_{y=Y_n}\right] \leq \tilde{C}\left(1+\mathbb{E}[L(Y_n)]\right.$$

using

$$|\nabla L|^2 < C\,(1+L)$$

We obtain

$$\mathbb{E}\left(|\nabla L(Y_n)\cdot H(Y_n, Z_{n+1})|\right) < \tilde{C}^2\left[1+\mathbb{E}[L(Y_n)]\right].$$

Similarly, we prove that

$$\mathbb{E}\left(H(Y_n, Z_{n+1})\,|\,\mathcal{F}_n\right) = \mathbb{E}\left(H(y, Z_1)\right)\Big|_{y=Y_n}$$

$$= h(Y_n)$$

$$\mathbb{E}\left[\Delta M_{n+1}\,|\,\mathcal{F}_n\right] = 0,$$

We have proven

$$\mathbb{E}\left(|\Delta M_{n+1}|^2\,|\,\mathcal{F}_n\right) \leq 2\,\tilde{C}^2\left(1+L(Y_n)\right)$$

$$\left(|\Delta M_{n+1}|^2 \leq 2\,|H(Y_n, Z_{n+1})|^2 + |h(Y_n)|^2\right.$$

$$\mathbb{E}\left[\nabla L(Y_n)\cdot\Delta M_{n+1}\,|\,\mathcal{F}_n\right] =$$

$$\nabla L(Y_n) \cdot \underbrace{\mathbb{E}[\Delta M_{n+1} \mid \mathcal{F}_n]}_{=0} = 0 .$$

$\mathbb{E}[\ \cdot \mid \mathcal{F}_n]$ on $\circledast$.

$$\mathbb{E}[L(Y_{n+1}) \mid \mathcal{F}_n] + \gamma_{n+1} \nabla L(Y_n) \cdot h(Y_n)$$

$$\leq L(Y_n) + \underbrace{C^2 \hat{C}}_{C_L} \gamma_{n+1}^2 (1 + L(Y_n)) .$$

$$\leq L(Y_n)(1 + C_L \gamma_{n+1}^2) + C_L \gamma_{n+1}^2$$

on the left side we add

$$+ \sum_{k=1}^{n} \gamma_k \nabla L(Y_{k-1}) h(Y_{k-1}) + C_L \sum_{k \geq n+2} \gamma_k^2 \quad (\ast\ast)$$

on the right side we add

$$+ (1 + C_L \gamma_{n+1}^2) \Big(\sum_{k=1}^{n} \gamma_k \nabla L(Y_{k-1}) h(Y_{k-1}) + C_L \sum_{k \geq n+2} \gamma_k^2 \Big)$$

22.10.24

⑤

$$S_n = \frac{L(Y_n) + \sum_{k=0}^{n-1} \gamma_{k+1} \nabla L(Y_k) h(Y_k) + C_L \sum_{k \geq n+1} \gamma_k^2}{\prod_{k=1}^{n} (1 + C_L \gamma_k^2)}$$

$$\mathbb{E}(S_{n+1} | \mathcal{F}_n) = \frac{1}{\prod_{k=1}^{n} (1 + C_L \gamma_k^2)} \quad (**)$$

$$\leq \frac{L(Y_n)(1 + C_L \gamma_n^2)}{\prod_{k=1}^{n+1} (1 + C_L \gamma_k^2)} + \frac{1 + C_L \gamma_{n+1}^2}{\prod_{k=1}^{n+2} (1 + C_L \gamma_k^2)} \Big(\sum_{k=1}^{n} \gamma_k \nabla L(Y_{k-1}) h(Y_n) + C_L \sum_{k \geq n+2} \gamma_k^2 \Big)$$

$$+ \frac{C_L \gamma_{n+1}^2}{\prod_{1}^{n+1} (1 + C_L \gamma_k^2)}$$

And

$$\frac{1}{\prod_{k=1}^{n} (1 + C_L \gamma_k^2)} \Big[C_L \sum_{k \geq n+2} \gamma_k^2 + \frac{C_L \gamma_{n+1}^2}{1 + C_L \gamma_{n+1}^2} \Big)$$

$$\leq \frac{1}{\prod^{n} (1 + C_L \gamma_k^2)} \Big(\sum_{k \geq n+2} C_L \gamma_k^2 \Big)$$

We have proved that

$$\mathbb{E}(S_{n+1} | \mathcal{F}_n) \leq S_n .$$

~~$\exists S_\infty$ such the~~

that is S_n is a positive supermartingale.

$\exists S_\infty$ such that $S_n(\omega) \xrightarrow{a.s} S_\infty(\omega)$

$$\tilde{S}_\infty = S_\infty \prod_{k=1}^{\infty} \left(1 + C_L \gamma_k^2\right).$$

We have

$$L(Y_n) + \sum_{k=0}^{n+1} \gamma_{k+1} \nabla L(Y_k) \cdot h(Y_k)$$

$$\xrightarrow[n \to +\infty]{} \tilde{S}_\infty .$$

$$S_n = \frac{L(Y_n) + \sum_{k=0}^{n-1} \gamma_{k+1} \nabla L(Y_k)\cdot h(Y_k) + \boxed{C_L \sum_{k\geq n+2} \gamma_k^2}}{\prod_{k=1}^{n} (1+C_L\gamma_k^2)}$$

We proved that $(S_n)_{n\geq 0}$ is a supermartingale

$$\mathbb{E}[S_{n+1} \mid \mathcal{F}_n] \leq S_n$$

$S_n \geq 0$, so $S_n(\omega) \xrightarrow[n\to\infty]{a.s} S_\infty(\omega)$

$$\boxed{\quad} \xrightarrow[n\to\infty]{} 0$$

$$\bigcirc \xrightarrow[n\to\infty]{} \prod_{k=1}^{\infty} (1+C_L\gamma_k^2)$$

Let $\Gamma_N = \prod_{k=1}^{N} (1+C_L\gamma_k^2)$

$$\beta_N = \log \Gamma_n = \sum_{k=1}^{N} \log(1+C_L\gamma_k^2)$$

then. We know that

$\gamma_k^2 \longrightarrow 0 \quad \left(\sum \gamma_k^2 < \infty \right)$

So $\log(1 + C_L \gamma_k^2) \sim C_L \gamma_k^2$.

$\sum_{k \geq 1} \gamma_k^2 < \infty$ so $\sum_{k \geq 1} \log(1 + C_L \gamma_k^2) < \infty$

$\beta_N \xrightarrow[N \to \infty]{} \beta_\infty$, $\log(\Gamma_N) \xrightarrow[N \to \infty]{} \beta_\infty$

$\Gamma_N = \exp(\log \Gamma_N) \longrightarrow \Gamma_\infty = \exp(\beta_\infty$

Rk $\begin{cases} \text{because } \sum \gamma_k^2 < \infty \\ \exp(\cdot) \text{ is Continuous.} \end{cases}$

We saw that

$\tilde{S}_\infty = S_\infty \prod_{k=1}^{\infty} (1 + C_L \gamma_k^2)$.

$L(Y_n) + \sum_{k=0}^{n-1} \gamma_{k+1} \nabla L(Y_k) \cdot f(Y_k) \xrightarrow[n \to \infty]{a.s} \tilde{S}_\infty$.

Rk 1 :

$$L(Y_n) = S_n \prod_{k=1}^{n} (1 + C_L \gamma_k^2) - \sum \gamma_{k+1} \nabla L(Y_k) h(Y_k) \quad C_L \sum_{k,n} \gamma_k^2$$

$$\leq \prod_{k=1}^{n} (1 + C_L \gamma_k^2) S_n$$

$$\leq \prod_{k=1}^{\infty} (1 + C_2 \gamma_k^2) S_n \quad \cancel{\Rightarrow \mathbb{E}(L(Y_n))} < \infty$$

$$\Rightarrow \mathbb{E}(L(Y_n)) < \infty .$$

≥ 0.

so $\mathbb{E}\left(\sum_{k=0}^{n-1} \gamma_{k+1} \nabla L(Y_k) \cdot h(Y_k) \right) \leq \prod_{k=1}^{n} (1 + C_L \gamma_k^2) \mathbb{E}(S_n)$

$$\leq \prod_{k=1}^{\infty} (1 + C_L \gamma_k^2) \mathbb{E}(S_0)$$

so

$$\mathbb{E}\left(\sum_{k=0}^{n-1} \gamma_{k+1} \nabla L(Y_k) \cdot h(Y_k) \right) < \infty$$

in particular $\sum_{n \geq 0}^{n-1} \gamma_n \nabla L(Y_n) h(Y_n) < \infty$

We deduce that $L(Y_n) \xrightarrow[n\to\infty]{a.s} L_\infty$.

$$\mathbb{E}\Big(\sum_{n\geq 1} |\Delta Y_n|^2\Big) = \sum_{n\geq 1} \mathbb{E}(|\Delta Y_n|^2)$$

$$\leq \sum_{n\geq 1} \gamma_n^2 \mathbb{E}(|H(Y_{n-1}, Z_n)|^2)$$

$$\leq C \sum_{n\geq 1} \gamma_n^2 \big(1 + \mathbb{E}[L(Y_{n-1})]\big) < \infty$$

So $\mathbb{E}(|\Delta Y_n|^2) \longrightarrow 0$ and so

$\sum |\Delta Y_n|^2 < \infty$

$$\Rightarrow \Delta Y_n = Y_n - Y_{n-1} \xrightarrow[a.s]{n\to\infty} 0$$

23/10/24

$\cancel{M_n^{(\gamma)}} =$

$$M_n^{(\gamma)} = \sum_{k=1}^{n} \gamma_k \left(H(Y_{k-1}, Z_k) - h(Y_{k-1}) \right)$$

$$\mathbb{E}[M_{n+1}^{(\gamma)} | \mathcal{F}_n) = \overbrace{\gamma_{n+1} \left(\mathbb{E}(H(Y_n, Z_{n+1}) | \mathcal{F}_n) - h(Y_n) \right]}^{=0}$$

$$+ \underbrace{\sum_{k=1}^{n} \gamma_k \left[H(Y_{k-1}, Z_k) - h(Y_{k-1}) \right)}_{M_n^{(\gamma)}}.$$

We already prove that

$$M_n^{(\gamma)} \xrightarrow[n \to \infty]{a.s \; L^2} M_\infty^{(\gamma)}$$

Reminder

Consider a Martingal $(M_n)_{n \geq 0}$ such that $M_0 = 0$, $\mathbb{E}(M_n^2) < \infty$

$\mathbb{E}(M_{n+1}^2 | \mathcal{F}_n) \geq M_n^2$ (M_n^2 is a sub Martingale.

The bracket ("crochet") is the $\mathcal{F}_{n-1}$-measurable process such that

$M_n^2 - \langle M \rangle_n$ is a martingale

$$\mathbb{E}\left(M_{n+1}^2 - \langle M \rangle_{n+1} \mid \mathcal{F}_n\right) = M_n^2 - \langle M \rangle_n$$

$$\langle M \rangle_{n+1} = \langle M_n \rangle + \mathbb{E}\left[(M_{n+1} - M_n)^2 \mid \mathcal{F}_n\right].$$

We have,

$$\mathbb{E}\left(M_{n+1}^2 \mid \mathcal{F}_n\right) - \langle M \rangle_{n+1} = M_n^2 - \langle M \rangle_n$$

$$\langle M \rangle_{n+1} = \langle M \rangle_n + \mathbb{E}\left(M_{n+1}^2 - M_n^2 \mid \mathcal{F}_n\right)$$

But, $(M_{n+1} - M_n)^2 = M_{n+1}^2 + M_n^2 - 2M_n M_{n+1}$

$$\mathbb{E}\left([M_{n+1} - M_n)^2\right] = \mathbb{E}\left(M_{n+1}^2 \mid \mathcal{F}_n\right) + M_n^2$$

$$-2M_n \underbrace{\mathbb{E}\left(M_{n+1} \mid \mathcal{F}_n\right)}_{M_n}$$

$$= \mathbb{E}(M_{n+1}^2 | \mathcal{F}_n) - M_n^2$$

$$= \mathbb{E}(M_{n+1}^2 - M_n^2 | \mathcal{F}_n)$$

<u>Théo</u>

Let (M_n) be an L^2 martingale then on $\{\langle M\rangle_\infty < \infty\}$, M_n converge a-s.

Using the previous result, we deduce that $M_n^\gamma \xrightarrow[a\text{-}s]{} M_\infty^\gamma$

$$\langle M^\gamma\rangle_n = \sum_{k\geq 1} \gamma_k^2 \mathbb{E}(|\Delta M_k|^2 | \mathcal{F}_{k-1})$$

$$\leq \sum_{k\geq 1} \gamma_k^2 \mathbb{E}(|H(Y_{k-1}, Z_k)|^2 | \mathcal{F}_{k-1})$$

$n \to \infty$

$$= \sum_{k\geq 1} \gamma_k^2 \mathbb{E}(|H(y, Z_1)|)\Big|_{y = Y_{k-1}} < \infty$$

Applications

1) Robbins Monroe Algo

We assume

- the function $h(y)$ is continuous and satisfies $\exists y_*$ such that

$$\forall y \neq y^* \quad (y - y_*) \cdot h(y) > 0.$$

$Y_0 \in L^2(\mathbb{P})$

and H is such $\forall y \in \mathbb{R}^d$

$$\left(\mathbb{E}\left[H(y,Z)^2\right]\right)^{1/2} \leq C(1+|y|)$$

$$\sum_{n\geq 1} \gamma_n = \infty, \quad \sum_{n\geq 1} \gamma_n^2 < \infty.$$

then: $\{y ; h(y)=0\} = \{y_*\}$

and $Y_n \xrightarrow[n\to\infty]{L^p} y_*$

$0 < p < 2$

$(|Y_n - y_*|)$ is bounded in L^2.

23.10.24

3

Consider y such that $h(y)=0$. so $(y-y^*)\cdot h(y)=0$

$\Longrightarrow \{y,\ h(y)=0\} \subset \{y^*\}$.

Let $y = y_* + \varepsilon u$ where $|u|=1$.

$(y-y_*)\,h(y_*+\varepsilon u) = \varepsilon u\cdot h(y_*+\varepsilon u) \geq 0$

Let $\varepsilon \to 0$, $u\cdot h(y_*) \geq 0$. $\Big\} \Rightarrow u\cdot h(y_*)=0$

~~Also~~ $(-u)\cdot h(y_*) \geq 0$

$\Longrightarrow h(y_*) = 0$.

Let z be such that $\forall |u|=1$, $u\cdot z = 0$, then $z=0$.

Assume $|z| \neq 0$. $e = \frac{z}{|z|}$

but $e \cdot z = 0$ so $\frac{z \cdot z}{|z|} = 0 = |z|$

Contradiction.

We have $\nabla L(y) \cdot h(y) > 0$

for $L(y) = |y - y_*|^2$.

Robbins Siegmund Lemma,

$|Y_n - y_*|^2 \longrightarrow L_\infty$

$\sum_{n \geq 0} \gamma_n h(Y_{n-1}) \cdot (Y_{n-1} - y_*) < \infty$

$|Y_n - y_*|^2$ is bounded in $L^1(\mathbb{P})$

Let $\omega \in \Omega$ such $|Y_n(\omega) - y_*|^2$ cv

$\sum \gamma_n h(Y_{n-1}(\omega)) \cdot (Y_{n-1}(\omega) - y_*) < \infty$.

1st result

$$\liminf \left(Y_{n-1}(\omega) - y_*\right) \cdot h\left(Y_{n-1}(\omega)\right) = 0$$

"$\lim$"

(otherwise it is impossible to have

$$\sum \gamma_n h\left(Y_{n-1}(\omega)\right)\left(Y_{n-1}(\omega) - y_*\right) < \infty.$$

We consider a subsequence $\varphi(n)$

such that $\left(Y_{\varphi(n)} - y_*\right) \cdot h\left(Y_{\varphi(n)}\right) \xrightarrow[n\to\infty]{} 0$

and $Y_{\varphi(n)} \longrightarrow y_\infty$.

By continuity, we have

$$\left(y_\infty - y_*\right) \cdot h\left(y_\infty\right) = 0$$

$$\Rightarrow y_\infty = y_*$$

But we know that

$L\left(Y_n(\omega)\right) = \left|Y_n(\omega) - y_*\right|^2$ CV

$$\lim |Y_n(w) - y^*|^2 = 0.$$

② SGD : Stochastic Gradient Descent ($h = \nabla L$)

$L : \mathbb{R}^d \longrightarrow \mathbb{R}_+$ differentiable function

∇L lipschitz continuous.

$$|\nabla L(y)|^2 \leq C(1 + L(y))$$

$$L(y) \xrightarrow[|y| \to +\infty]{} +\infty$$

$$\{y : \nabla L(y) = 0\} = \{y_*\}$$

and we assume that

$$h(y) = \mathbb{E}[H(y, Z)]$$

$$\mathbb{E}[|H(y, Z)|^2] \leq C(1 + L(y))$$

$$\mathbb{E}[L(Y_n)] < \infty$$

$\sum \gamma_n = +\infty$; $\sum \gamma_n^2 < \infty$

then $L(y_*) = \operatorname{Min} L$,

$Y_n \xrightarrow[n \to +\infty]{a.s.} y_*$

$\nabla L(Y_n) \longrightarrow 0$ in $L^p(\mathbb{P})$, $0 < p < 2$; $L(Y_n)$ is bounded in $L^1(\mathbb{P})$

$\nabla L(Y_n)$ is bounded in $L^2(\mathbb{P})$.

$h(y) \cdot \nabla L(y) = |\nabla L(y)|^2 \geq 0$

So we can apply Robbins-Siegmund lemma and we conclude that $L(Y_n) \longrightarrow L_\infty$

$\sum \gamma_n |\nabla L(Y_{n-1})|^2 < \infty$ a.s.

$L(Y_n)$ is $L^1(\mathbb{P})$ bounded.

Consider $\omega \in \Omega$ such that

$L(Y_n(\omega))$ CV and

$\sum_{n\geq 1} \gamma_n |\nabla L(Y_{n-1}(\omega))|^2 < \infty$

and $Y_n(\omega) - Y_{n-1}(\omega) \longrightarrow 0$.

Same argument as for Robbins-Monroe gives

$\liminf |\nabla L(Y_{n-1}(\omega))|^2 = 0$.

Using $L(y) \xrightarrow[|y|\to\infty]{} \infty$, we deduce that Y_n is bounded

So $Y_{\varphi(n)} \longrightarrow \tilde{y}$

$$\left.\begin{array}{l} \nabla L(Y_{\varphi(n)}) \longrightarrow 0 \\ L(Y_{\varphi(n)}) \longrightarrow L_\infty(\omega) \end{array}\right\} \begin{array}{l} \Rightarrow \nabla L(\tilde{y}) = 0 \\ \Rightarrow \tilde{y} = y_* \end{array}$$

$$\{y, L(y) = L(y^*)\} =$$

$$\{y, \nabla L(y) = 0\} = \{y_*\}$$

Proposition

$F : \mathbb{R}^d \longrightarrow \mathbb{R}$ is Lipschitz continuous with constant B.

$$|F(x) - F(y)| \leq B \| x - y \|_2$$

Consider a scheme such that

$$\mathbb{E}[g_k(\alpha_{k-1}) \mid \alpha_{k-1}] = \nabla F(\alpha_{k-1})$$

we set

$$\alpha_k = \alpha_{k-1} - \gamma_k \, g_k(\alpha_{k-1})$$

we assume that F is convexe, lipschitz continuous with constant B.

And has a unique

minimizer α^* and we know that $\|\alpha_0 - \alpha^*\| \leq D$.

We also assume that

$\|g_k(\alpha_{k-1})\|_2 \leq B$ a.s.,

Then, if we choose

$\gamma_k = \frac{D}{B\sqrt{k}}$, we have

~~$\mathbb{E}[\alpha_n] - F(\alpha^*)$~~

$$\mathbb{E}\left(F(\bar{\alpha}_n) - F(\alpha^*)\right) \leq DB\,\frac{2+\log(k)}{2\sqrt{k}}$$

with $\bar{\alpha}_n = \frac{\sum_{k=1}^{n} \gamma_k \alpha_k}{\sum_{k=1}^{n} \gamma_k}$

Proof

$\mathbb{E}\left[\|\alpha_k - \alpha^*\|^2\right]$

$= \mathbb{E}\left[\|\alpha_{k-1} - \gamma_k g_k(\alpha_{k-1}) - \alpha^*\|^2\right]$

$$= \mathbb{E}\left[\|\alpha_{k-1} - \alpha^*\|^2\right]$$

$$- 2\gamma_k \mathbb{E}\left[g_k(\alpha_{k-1}) \cdot (\alpha_{k-1} - \alpha^*)\right]$$

$$+ \gamma_k^2 \mathbb{E}\left[\|g_k(\alpha_{k-1})\|^2\right] \quad \left\{ \leq \gamma_k^2 B^2 \right.$$

- $\mathbb{E}\left[g_k(\alpha_{k-1}) \cdot (\alpha_{k-1} - \alpha^*)\right]$

$$= \mathbb{E}\left[\mathbb{E}\left[g_k(\alpha_{k-1}) \cdot (\alpha_{k-1} - \alpha^*) \mid \alpha_{k-1}\right]\right]$$

$$= \mathbb{E}\left[\mathbb{E}\left[g_k(\alpha_{k-1}) \mid \alpha_{k-1}\right] \cdot (\alpha_{k-1} - \alpha^*)\right]$$

$$= \mathbb{E}\left[\nabla F(\alpha_{k-1}) \cdot (\alpha_{k-1} - \alpha^*)\right]$$

- $\mathbb{E}\left[\|\alpha_k - \alpha^*\|^2\right]$

$$\leq \mathbb{E}\left[\|\alpha_{k-1} - \alpha^*\|^2\right] - 2\gamma_k \mathbb{E}\left[\nabla F(\alpha_{k-1}) \cdot (\alpha_{k-1} - \alpha^*)\right] + \gamma_k^2 B^2$$

$$\mathbb{E}\left[F(\alpha_{k-1}) - F(\alpha^*)\right]$$

$$\leq \mathbb{E}\left[\nabla F(\alpha_{k-1}) \cdot (\alpha_{k-1} - \alpha^*)\right]$$

$$\gamma_k \mathbb{E}\left[F(\alpha_{k-1}) - F(\alpha^*)\right]$$

$$\leq \frac{1}{2}\mathbb{E}\left[\|\alpha_{k-1} - \alpha^*\|^2\right]$$

$$- \frac{1}{2}\mathbb{E}\left[\|\alpha_k - \alpha^*\|^2\right] + \frac{1}{2}\gamma_k^2 B^2.$$

$$\sum_{k=1}^{n} \gamma_k \mathbb{E}\left[F(\alpha_{k-1}) - F(\alpha^*)\right]$$

$$\leq \frac{1}{2}\mathbb{E}\left[\|\alpha_0 - \alpha^*\|^2\right]$$ (somme télescopique)

$$+ \frac{1}{2}B^2 \sum_{k=1}^{n} \gamma_k^2$$

On divise tout par $\sum_{k=1}^{n} \gamma_k$

on trouve,

$$\frac{\sum_{k=1}^{n} \gamma_k \mathbb{E}[F(\alpha_{k-1}) - F(\alpha^*)]}{\sum_{k=1}^{n} \gamma_k} \geq$$

$$\min_{k} \mathbb{E}[F(\alpha_{k-1}) - F(\alpha^*)]$$

$$\frac{\sum_{k=1}^{n} \gamma_k \mathbb{E}[F(\alpha_{k-1}) - F(\alpha^*)]}{\sum_{k=1}^{n} \gamma_k} \geq$$

$$\mathbb{E}[F(\bar{\alpha}_n) - F(\alpha^*)]$$

$$\mathbb{E}[F(\bar{\alpha}_n) - F(\alpha^*)] \leq \frac{D^2}{2\sum_k \gamma_k} + \frac{4}{2} B^2 \frac{\sum_{k=1}^{n} \gamma_k^2}{\sum_{k=1}^{n} \gamma_k}$$

$$\gamma_k = \frac{D}{B\sqrt{k}} \; ; \; \sum_{k=1}^{n} \frac{1}{\sqrt{k}} \xrightarrow[n\to+\infty]{} +\infty$$

$$\int_0^1 \frac{du}{\sqrt{u}} = \left[2\sqrt{u} \right]_0^1 = 2$$

$$2 = \int_0^1 \frac{du}{\sqrt{u}} = \sum_{k=1}^{n} \int_{\frac{k-1}{n}}^{\frac{k}{n}} \frac{du}{\sqrt{u}}$$

$$\leq \sum_{k=1}^{n} \frac{1}{\sqrt{\frac{k}{n}}} \times \frac{1}{n} = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \frac{1}{\sqrt{k}}$$

Rq $u \longmapsto \frac{1}{\sqrt{u}}$ est décroissante sur $[0,1]$.

d'où : $\sum_{k=1}^{n} \frac{1}{\sqrt{k}} \geq 2\sqrt{n}$

$$\gamma_k^2 = \frac{D^2}{B^2} \times \frac{1}{k}$$

$$\sum_{k=1}^{n} \frac{1}{k} \leq \int_1^n \frac{du}{u} = \log(n)$$

Finalement,

$\mathbb{E}[F(\bar{\alpha}_n) - F(\alpha^*)]$

$$\leq \frac{1}{2} \frac{D^2}{2 \times \frac{D}{B} \sqrt{n}} + \frac{1}{2} B^2$$

$$+ \frac{\frac{D^2}{B^2} \log(n)}{2 \frac{D}{B} \sqrt{n}} = \frac{DB}{4\sqrt{n}} (1 + \log(n))$$

Rq:

I used

$\mathbb{E}[\| g_k(\alpha_{k-1}) \|^2] \leq B^2$ and

can relax the assumption

$\| g_k(\alpha_{k-1}) \|^2 \leq B^2$ a.s

---

next course 9h–12h <u>M-14</u>

Proposition

$$\theta_l = \theta_{l-1} - \gamma_l \, g_l(\theta_{l-1})$$

$$\gamma_l = \frac{D}{B\sqrt{l}}$$

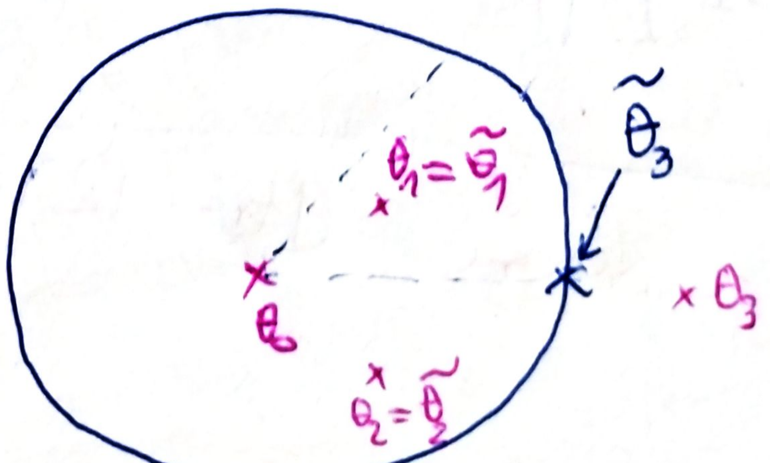
$$\mathbb{E}\left[F(\bar{\theta}_n) - F(\theta^*)\right] \leq DB \, \frac{2 + \log(n)}{2\sqrt{n}}$$

$$\bar{\theta}_l = \frac{\sum_{k=1}^{n} \gamma_k \theta_k}{\sum_{k=1}^{n} \gamma_k}$$

An alternative algo

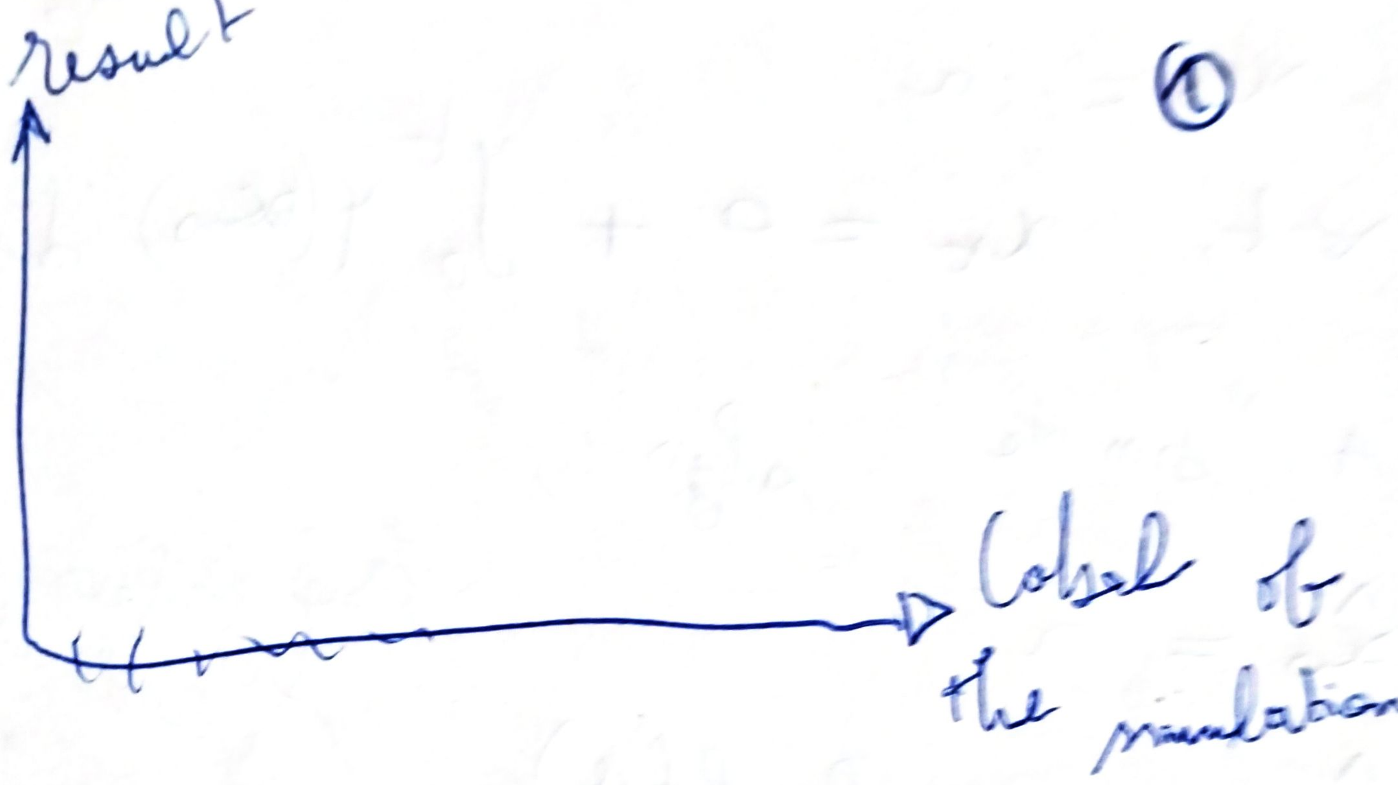
$$\tilde{\theta}_l = \prod_{B(\theta_0, D)} \left[\tilde{\theta}_{l-1} - \gamma_l \, g_l(\tilde{\theta}_{l-1})\right]$$

projection on the ball centered at θ_0 with radius D.



① Cloud of points 12/11/24

①



Rq:

It is a classical way to avoid very large values and the issues in the accuracy of computation for large numbers.

Ex: Dinamical system

$\dot{x}_t = \varphi(x_t)$ $\quad \dot{x}_{\ell h} \simeq \frac{x(\ell+h) - x_{\ell h}}{h}$ (B

$x_0 = a$

$\forall t, \quad x_t = a + \int_0^t \varphi(\cancel{x}_s)\, ds$

A "simple" algo.

$\hat{x}_0 = a$

$\hat{x}_h = a + h\, \varphi(a)$

$\hat{x}_{(\ell+1)h} = \hat{x}_{\ell h} + h\, \varphi(\hat{x}_{\ell h})$

Euler explicit scheme

sufficient condition for convergence:

φ C^1 and lipschitz contin...

$$X_t = X_0 + \int_0^t \varphi(X_s)\, ds + \underbrace{\text{Noise}_t}_{B_t}$$

Brownian motion

$(B_t)_{t\geq 0}$ is a Brownian motion

iff $B_t \sim \mathcal{N}(0,t) = \sqrt{t}\,\mathcal{N}(0,1)$

$0 \leq s < t$, $(B_t - B_s)$ is $\perp\!\!\!\perp$ of

$(B_u, u \leq s)$.

$\mathcal{L}(B_t - B_s)$ only depends on the time spent that is

$B_t - B_s \sim \mathcal{N}(0, t-s)$

$s \longmapsto B_s$ are continuous.

$$\frac{\sqrt{N_{MC}}}{\sigma_X}\left(\frac{1}{N_{MC}}\sum_{i=1}^{N_{MC}} X^{(i)} - \mathbb{E}[X]\right) \xrightarrow[N_{MC}]{\mathcal{L}} \mathcal{N}(0,1)$$

TCL

Considerer (Y_i) iid random variable with mean 0 and variance 1.

$$S_n = S_{n-1} + Y_n$$

$$= \sum_{k=1}^{n} Y_k$$

Donskers theo

$$\hat{Z}^n_t = \frac{1}{\sqrt{n}} S_{\lfloor nt \rfloor}$$

$$\left(\hat{Z}^n_s, 0 \le s \le T\right) \xrightarrow[n]{\mathcal{L}} \left(B_s, 0 \le s \le T\right)$$

$$X_t = X_0 + \int_0^t \varphi(X_s)\, ds + \sigma B_t$$

$$\hat{X}_{(l+1)h} = \hat{X}_{lh} + \varphi(\hat{X}_{lh})\, h + \sigma \underbrace{\left(B_{(l+1)h} - B_{lh}\right)}_{\perp\!\!\!\perp \text{ of the past}}$$

$\mathcal{N}(0,h)$

② 12/11/2024

Ex 1 $\varphi(x) = -x$

$\sigma = 0$. $\frac{dx_t}{dt} = -x_t$

We search a functio

$t \longmapsto x_t$ such that

$\forall t \geq 0, \frac{d}{dt} x_t = -x_t$.

$x_t = a e^{-t}$.

Explanatio

$d(\log(x)) = \frac{dx}{x} = -dt$

$\log x(t) - \log x(t_0) = -(t - t_0)$

$x(t) = x(t_0) \exp(-(t-t_0))$

Exo 2

$$\frac{d x_t}{dt} = -(x_t)^3$$

$x_0 = 1$.

Remind

$$\begin{cases} \frac{d x_t}{dt} = -x_t \iff \\ \log x_u - \log x_0 = \int_0^u \frac{dx}{x} = \int_0^u - dt = -(u-0) \end{cases}$$

1st approche : we have an idea of the solution.

$$y_t = \frac{D}{\sqrt{t+c}} \quad ; \quad x_0 = x_0 = \frac{D}{\sqrt{c}}$$

$$\frac{dy_i}{dt} = \frac{1}{2} \frac{D}{\left(t + \frac{D^2}{x_0^2}\right)\sqrt{t + \frac{D^2}{x_0^2}}}$$

$$\overset{??}{=} \left(\frac{D}{\left(t + \frac{D^2}{x_0^2}\right)^{1/2}} \right)^3$$

if $\frac{D}{2} = D^3 \Rightarrow$ ok

$$y_t = \frac{1}{\sqrt{2t + \frac{1}{x_0^2}}}$$

<u>2 nd approche</u>

$$-\frac{dx}{x^3} = dt$$

$$-\left[-\frac{1}{2} \times \frac{1}{x^2}\right]_0^t = \cancel{dt}\ t$$

$$\frac{1}{2}\frac{1}{x_t^2} - \frac{1}{2x_0} = t$$

$$\Leftrightarrow \frac{1}{2x_t^2} = 2t + \frac{1}{2x_0^2}$$

$$x_t = \operatorname{sing}(x_0)\sqrt{\frac{1}{2t + \frac{1}{x_0^2}}}$$

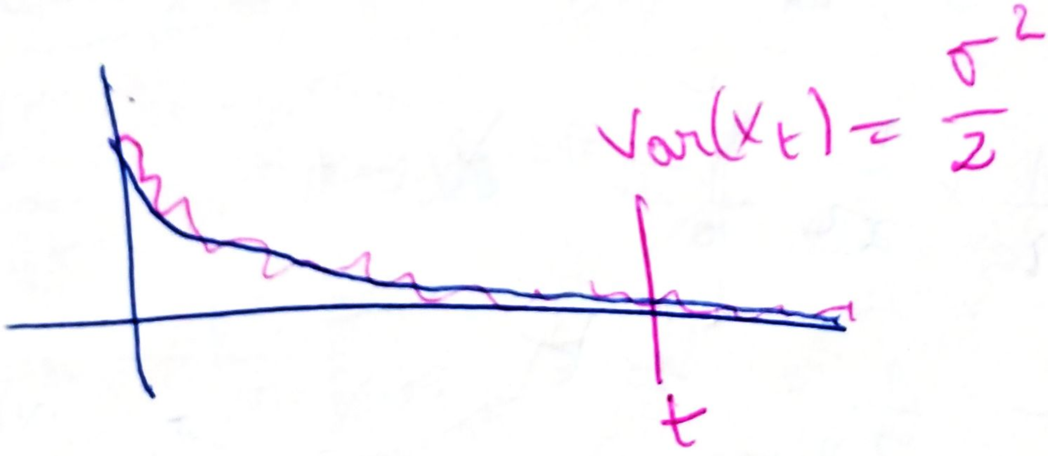
<u>Rq</u>

$$\begin{cases} \dot{x}_t = -x_t & \text{goes to 0 faster} \\ \dot{y}_t = -y_t^3 & \to 0 \text{ slower} \end{cases}$$

$$Y_t = Y_0 - \int_0^t Y_s^3 \, ds + \sigma B_t$$

$$\tilde{Y}_{(l+1)h} = \tilde{Y}_{lh} - \underbrace{\tilde{Y}_{lh} \, h +}_{??? \; > 1} \sigma \sqrt{h} \, G_{l+1}$$

Back to the linear SDE



Go Back

$$\tilde{\theta}_l = \Pi_{B(\theta_0, D)} \left[\tilde{\theta}_{l-1} - \gamma_l \, g_l(\tilde{\theta}_{l-1}) \right]$$

A detailed look on the proof given last time gives you that the conclusion remains true

③ 12/11/24

$\gamma_e = \frac{D}{B\sqrt{e}} \Rightarrow \mathbb{E}[F(\bar{\theta}_n) - F(\theta^*)] \leq$

DB $\frac{2+\log n}{\sqrt{n}}$.

An extension on Robbins Monroe Algo (Chen variatio)

h is continuous

y^* is uniq

$(y-y^*) \cdot h(y) > 0$

$\left(\mathbb{E}[H(y,Z)^2]\right)^{1/2} \leq C(1+|y|)$

$\sum_{n\geq 1} \gamma_n = +\infty$

$\sum_{n\geq 1} \gamma_n^2 < \infty$

then $\{y, h(y)=0\} = \{y^*\}$

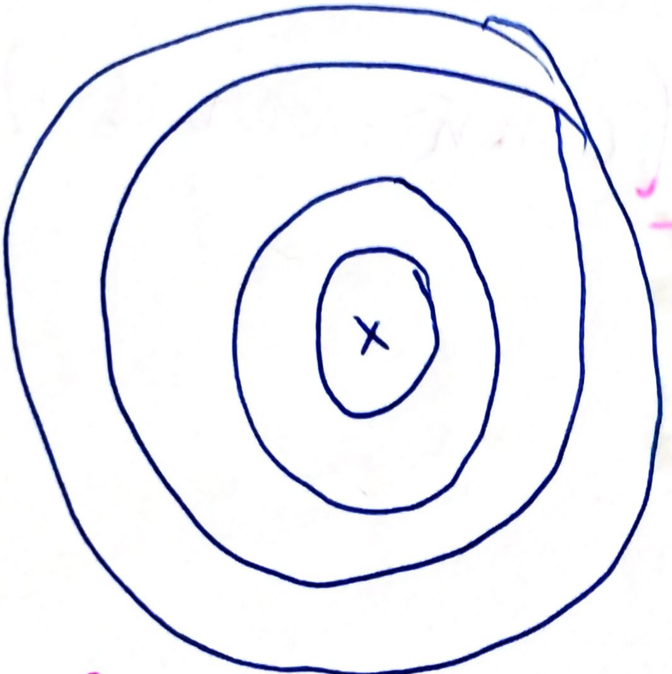
$Y_{n+1} = Y_n - \gamma_{n+1} H(Y_n, Z_{n+1})$

$Y_n \xrightarrow[n\to\infty]{a.s} y^*$

Considere an increasing sequence of compact sets

$K_1 \subset K_2 \subset K_3 \subset \cdots \subset \cdots$

and $\lim_{p \to +\infty} K_p = \mathbb{R}^d$.



Algo

(Y_ℓ, i_ℓ), $Y_\ell \in \mathbb{R}^d$, $i_\ell \in \mathbb{N}$.

At step $\ell+1$,

(a) $z_{\ell+1} \sim Z$

$\hat{Y}_{\ell+1} = Y_\ell - \gamma_{\ell+1} H(Y_\ell, z_{\ell+1})$

• If

$\hat{Y}_{\ell+1} \in K_{i_\ell}$, then $Y_{\ell+1} = \hat{Y}_{\ell+1}$

$i_{\ell+1} = i_\ell$.

• If $\hat{Y}_{\ell+1} \notin K_{i_\ell}$

$\Rightarrow i_{\ell+1} = i_\ell + 1$

$Y_{\ell+1} = Y_0$

$\exists y^*$ such that $h(y^*) = 0$.

$\forall y, (y - y^*) \cdot h(y) \geq 0$.

I assume that there exists a function $V : \mathbb{R}^d \longrightarrow \mathbb{R}^{d \times d}$.

$\lim_{y \to 0} \| V(y) \| = 0$ and a matrix

A with ≥ 0 eigenvalues.

$h(y) = A \cdot (y - y^*) + V(y - y^*)(y - y^*)$

$S\Pi_\ell = \Pi(Y_\ell) - H(Y_\ell, z_{\ell+1})$

+ technical assumptions.

results

① i_ℓ cv a.s to $i_\infty(\omega)$

and $Y_\ell \xrightarrow[\ell \to \infty]{a.s} y^*$

Homework

write and use Chen algo

. Observe how fast it's cv

$\gamma_\ell = \frac{1}{(\ell)^2} \longrightarrow \frac{1}{(100+\ell)^2}$

. observer la loi de i_∞

additional result.

$\frac{Y_\ell - y^*}{\sqrt{\sigma_\ell}} \xrightarrow[\ell \to \infty]{\mathcal{L}}$ Random complexe objects

$\boxed{\text{"Optimal"}}$ → On ne peut pas construire d'algo qui soit "plus rapide" sur une classe ~~sg~~ assez large de prob d'opti.

The setting: ⊕

$\forall \theta \in \Theta$, P_θ a probability distribution.

We generate D data set from this distribution P_θ and construct an "estimator" of θ

$\mathcal{A}(D) \in \Theta$

a distance σ on Θ.

$\forall\, \theta, \tilde{\theta},\ \delta^2(\theta, \tilde{\theta})$

$\sigma(\theta, \mathcal{A}(D))$ is a measure of the performance of our algo. if the "true" parameter is θ.

The testing error of $\mathcal{A}$ is defined as

$$\mathbb{E}_{\theta_*}\left[\sigma(\theta_*, \mathcal{A}(D))^2\right].$$

An algo is effecient if

$$\sup_{\theta^*} \mathbb{E}_{\theta_*}\left[\sigma(\theta_*, \mathcal{A}(D))^2\right]$$

is small

the lower bound is

$$\inf_{\mathcal{A}} \sup_{\theta^* \in \Theta} \mathbb{E}_{\theta^*}\left[\sigma(\theta^*, \mathcal{A}(D))^2\right].$$

this error is usually refered as the Minimax error.

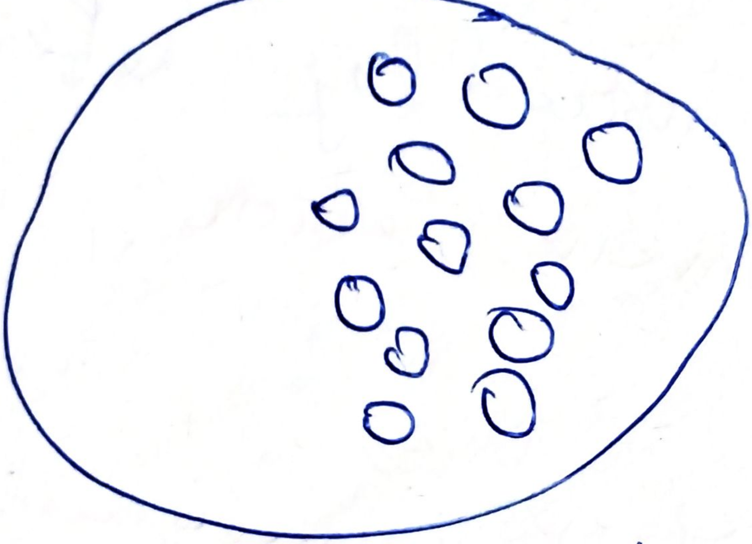
1^ers_ Reminder : Markov-inequality

$$\mathbb{E}_{\theta_*}\left[\sigma^2(\theta_* \mathcal{A}(D))\right]$$
$$\geq A\,\mathbb{P}\left(\sigma^2(\theta_*, \mathcal{A}(D)) \geq A\right).$$

It mease that, a lower bound of

$\inf_{\mathcal{A}}\left[\sup_{\theta^*} \mathbb{P}_{\theta^*}\left(\sigma^2(\theta^*, \mathcal{A}(D))\right]\right.$ gives a lower bound of the ~~max~~ minimax error

Then, we can reduce to an hypothesis test.



Balls of radius $\sqrt{A}$ inside Θ.

$\theta_1, \dots, \theta_M \in \Theta$ such that

$$i \neq j \quad \sigma^2(\theta_i, \theta_j) \geq 4A$$

13/11/2024 Course - Opti

②

$$\sup_{\theta_* \in \Theta} \mathbb{P}\left(\sigma^2(\theta_*, \mathcal{A}(D)) \geq A\right)$$

$$\geq \max_{j \in [\![1,M]\!]} \mathbb{P}_{\theta_j}\left[\sigma^2(\theta_j, \mathcal{A}(D)) \geq A\right].$$

Any algo gives a "test" defined by

$$g(\mathcal{A}(D)) = \operatorname{Argmin}_j \left(\sigma(\theta_j, \mathcal{A}(D))\right)$$

If $g(\mathcal{A}(D)) \neq j$,

$\exists\, \ell \neq j$ such that

$$\sigma(\theta_\ell, \mathcal{A}(D)) \leq \sigma(\theta_j, \mathcal{A}(D))$$

$$\sigma^2(\theta_j, \theta_\ell) \leq 2\left(\sigma^2(\theta_j, \mathcal{A}(D)) + \sigma^2(\theta_\ell, \mathcal{A}(D))\right)$$

$$\sigma^2(\theta_j, \mathcal{A}(D)) \geq \frac{1}{2}\sigma^2(\theta_j, \theta_k) - \sigma^2(\theta_k, \mathcal{A}(D))$$

$$\geq \frac{1}{2}\sigma^2(\theta_j; \theta_k) - \sigma^2(\theta_j, \mathcal{A}(D))$$

$$\sigma^2(\theta_j, \mathcal{A}(D)) \geq \frac{1}{4}\sigma^2(\theta_j, \theta_k) \geq A.$$

$$\mathbb{P}_{\theta_j}\left[\sigma^2(\theta_j, \mathcal{A}(D)) > A\right]$$

$$\geq \mathbb{P}_{\theta_j}\left(g(\mathcal{A}(D)) \neq j\right)$$

$$\inf_{\mathcal{A}} \sup_{\theta^*} \mathbb{E}_{\theta^*}\left[\sigma^2(\theta_*, \mathcal{A}(D))\right]$$

$$\geq A \inf_{h} \max_{j} \mathbb{P}_{\theta_j}\left[h(D) \neq j\right]$$

$$\geq A \inf_{h} \frac{1}{M}\sum_{j=1}^{M} \mathbb{P}_{\theta_j}\left[h(D) \neq j\right]$$

Rf larger is M larger is $\mathbb{P}$ to be far from the true θ

Result on the number of balls

Varshamov Gilbert's Lemma

Consider $\{0,1\}^d$ hypercube

$\forall \alpha \in]0,1[$, there exists a subset $\mathcal{B}$ of $\{0,1\}^d$ such that

a) $\forall x, x' \in \mathcal{B}$, $x \neq x'$

$$\|x - x'\|_1 \geq (1-\alpha)\frac{d}{2}$$

b) $$|\mathcal{B}| \geq \exp\left(d\,\frac{\alpha^2}{2}\right)$$

$\boxed{\text{Proof}}$ Consider the largest family satisfying a)

$$\bigcup_{x \in B} B\left(x, (1-\alpha)\frac{d}{2}\right) \subseteq \{0,1\}^d$$

Proof $\{0,1\}^d = \bigcup_{x \in B} B\left(x, (1-\alpha)\frac{d}{2}\right)$

$\hat{x} \in \{0,1\}^d \setminus \bigcup_{x \in B} B\left(x, (1-\alpha)\frac{d}{2}\right)$

$\hat{B} = B \cup \{\hat{x}\}$, strictly larger than B and has the property that

$\forall\, x, x' \in \hat{B}$, $\|x - x'\|_1 \geq (1-\alpha)\frac{d}{2}$

Which is in contradiction with the assumption that B is the largest family satisfying a)

$$|\{0,1\}^d| = 2^d \leq$$

$$\sum_{x \in B} \left|\left\{y \in \{0,1\}^d, \|y - x\|_1 \geq (1-\alpha)\frac{d}{2}\right\}\right|$$

By symetry,

$$\left|\{y \in \{0,1\}^d, \|y - x\|_1 \geq (1-\alpha)\frac{d}{2}\}\right|$$

does not depend on x (eg: $x=0$)

$$8^d \leq |B| \left|\{y \in \{0,1\}^d, \|y\|_1 \geq (1-\alpha)\frac{d}{2}\}\right|$$

$$\frac{1}{2^d}\left|\{y \in \{0,1\}^d, \|y\|_1 \leq (1-\alpha)\frac{d}{2}\}\right|$$

$$= \mathbb{P}\left(Z \leq (1-\alpha)\frac{d}{2}\right) = \mathbb{P}\left(Z \geq (1+\alpha)\frac{d}{2}\right)$$

where $Z \sim$ Binomial $(d, \frac{1}{2})$

~~wt~~ We conclude with Hoeffding's inequality

$Y_1, \dots, Y_n$ iid random variables

with values in $[0,1]$.

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n Y_i - \frac{1}{n}\sum_{i=1}^n \mathbb{E}[Y_i] \geq t\right)$$

$$\leq \exp\left(-2n t^2\right)$$

$$Z = \sum_{i=1}^d B_i, \quad B_i \sim \text{Bernoulli}\left(\frac{1}{2}\right)$$

$$\mathbb{P}\Big(\frac{1}{d}\sum_{i=1}^d B_i - \frac{1}{d}\sum_{i=1}^d \underbrace{\mathbb{E}[B_i]}_{1/2} \geq \frac{\alpha}{2}\Big)$$

$$\leq \exp\left(-2d\left(\frac{\alpha}{2}\right)^2\right)$$

$$= \exp\left(-\frac{d\alpha^2}{2}\right).$$

But,

$$2^d \leq |B| \, 2^d \underbrace{\frac{1}{2^d}\left|\{y \in \{0,1\}^d, \|y\|_1 \leq (1-\alpha)\frac{d}{2}\}\right|}_{\leq \exp\left(-\frac{d\alpha^2}{2}\right)}$$

So $|B| \geq \exp\left(\frac{d\alpha^2}{2}\right)$

conclude:

$$|B| \geq \exp\left(\frac{d\alpha^2}{2}\right)$$

Proposition

Consider 2 random variables, y and $\hat{y}$ taking values in the same finite set $\mathcal{Y}$ and we have a Markov chain

$y \to z \to \hat{y}$,

For an arbitrary random variable z, then

$$\mathbb{P}(\hat{y} \neq y) \geq \frac{H(y|\hat{y}) - \log 2}{\log|\mathcal{Y}|} \Rightarrow \frac{H(y|z) - \log 2}{\log|\mathcal{Y}|}$$

~~Reminders~~

Entropy: Y a random variable with values in $\mathcal{Y}$

$$H(Y) = -\sum_{\tilde{y} \in \mathcal{Y}} P(Y=\tilde{y}) \log P(Y=\tilde{y})$$

If Y is deterministic $H(Y)=0$.

. $Y \sim \mathcal{U}[1, M]$

$$H(Y) \le -\sum_{i=1}^{M} \frac{1}{M} \log\left(\frac{1}{M}\right) = \log(M)$$

$$\Leftrightarrow 0 \le H(Y) \le \log(M)$$

Joint entropy:

$X \in \mathcal{X}$, $Y \in \mathcal{Y}$ two random variables.

$$H(X,Y) = \sum_{x' \in \mathcal{X}} \sum_{y' \in \mathcal{Y}} \mathbb{P}(X=x', Y=y') \log \mathbb{P}(X=x', Y=y')$$

13/11/2024 Course Opti ④

The conditional entropy

$$H(Y|X) = \sum_{x' \in \mathcal{X}} \mathbb{P}(X = x') H(Y|X = x').$$

$\forall X, Y$, we have

$$H(X, Y) = H(Y|X) + H(X)$$

Proposition

$Y, \hat{Y} \in \mathcal{Y}$,

$$\mathbb{P}(\hat{Y} \neq Y) \geq \frac{H(Y|\hat{Y}) - \log 2}{\log|\mathcal{Y}|}.$$

Proof

$e = \mathbb{1}_{Y \neq \hat{Y}} \in \{0, 1\}$,

$H(e|\hat{Y}) + H(Y|e, \hat{Y}) =$

$$H(e, Y \mid \hat{Y}) = H(Y \mid \hat{Y}) + \underbrace{H(e \mid Y, \hat{Y})}_{=0}$$

$$H(e \mid \hat{Y}) \leq H(e) \leq \log 2 \leftarrow \mathrm{card}(\{0,1\})$$

$$H(Y \mid e, \hat{Y}) = \mathbb{P}(e=1) H(Y \mid \hat{Y}, e=1) + \mathbb{P}(e=0) \underbrace{H(Y \mid \hat{Y}, e=0)}_{=0 \text{ (because deterministic}}$$

$$\leq \mathbb{P}(Y \neq \hat{Y}) H(Y \mid \hat{Y}, e=1)$$

$$\leq \mathbb{P}(Y \neq \hat{Y}) \log |\mathcal{Y}|$$

$$\Rightarrow \mathbb{P}(Y \neq \hat{Y}) \geq \frac{\overbrace{H(Y \mid e, \hat{Y})}}{\log |\mathcal{Y}|}$$

$$= \frac{H(Y \mid \hat{Y}) - H(e \mid \hat{Y})}{\log |\mathcal{Y}|}$$

$$\geq \frac{H(Y \mid \hat{Y}) - \log 2}{\log |\mathcal{Y}|}$$

donc,

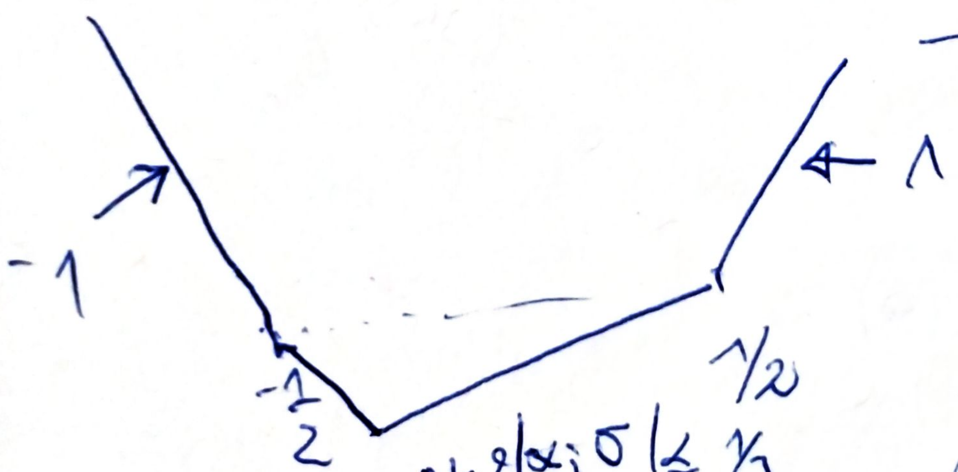
$$P(Y \neq \hat{Y}) \geq \frac{H(Y|\hat{Y}) - \log 2}{\log |\mathcal{Y}|}.$$

An example of stochastic gradient

$\alpha \in \{-1, 1\}^d$, $\sigma \in]0; \frac{1}{4}]$

$$F_\alpha(\theta) = \frac{B}{2d} \sum_{i=1}^{d} \left\{ \left(\frac{1}{2} + \alpha_i \sigma\right) \left|\theta_i + \frac{1}{2}\right| + \left(\frac{1}{2} - \alpha_i \sigma\right) \left|\theta_i - \frac{1}{2}\right| \right\}$$

$\theta_i \mapsto$ piecewise linear.



$\rightarrow$ it is convexe.

It has a ming minimizer.

$2|\alpha_i \sigma| \leq 1/2$

The global minimizer

$\theta = -\frac{\alpha}{2}$

$F_\alpha^* = F_\alpha(-\frac{\alpha}{2}) = \frac{B}{4}(1 - 2\sigma)$

Lemma

~~$\alpha, \beta \in \{0,$~~

$\alpha, \beta \in \{-1, 1\}^d$.

Assume

$F_\alpha(\theta) - F_\alpha^* \leq \varepsilon$.

Then, $F_\beta(\theta) - F_\beta^* \geq \frac{B\sigma}{2d} \|\alpha - \beta\|_1 - \varepsilon$

Proof

~~$F_\beta(\theta) = F_\beta^*(\theta)$~~

13/11/2024

Course : Opti

5

$$F_\beta(\theta) - F_\beta^* = F_\beta(\theta) + F_\alpha(\theta) - F_\beta^* - F_\alpha^* + \underbrace{F_\alpha^* - F_\alpha(\theta)}_{\geq -\varepsilon}.$$

$$F_\beta(\theta) + F_\alpha(\theta) - F_\beta^* - F_\alpha^* \geq$$

$$\frac{B}{2d} \sum_{\alpha_i \neq \beta_i} \left\{ \left|\theta_i + \frac{1}{2}\right| + \left|\theta_i - \frac{1}{2}\right| + 2\sigma - 1 \right\}$$

$$\geq \frac{B}{2d} \sum_{i, \alpha_i \neq \beta_i} (2\sigma) = \frac{2\sigma B}{2d} \underbrace{\sum_{i, \alpha_i \neq \beta_i}}_{\frac{1}{2}\|\beta - \alpha\|_1}$$

$$F_\beta(\theta) - F_\beta^*(\theta) \geq \frac{B\sigma}{2d} \|\beta - \alpha\|_1 - \varepsilon$$

Consider M points

$\alpha^{(1)}, \dots, \alpha^{(M)} \in \{-1,1\}^{n}$,

$\forall i \neq j, \; \| \alpha^{(i)} - \alpha^{(j)} \|_1 \geq \frac{d}{2}$.

thanks to the previous lemma

I can take $M \geq \exp\left(\frac{d}{8}\right)$.

$\varepsilon < \frac{B\sigma}{8}$.

If I minimize with an error less than ε, it means that my "Hypothesis test" works

link between "probability to fail" and conditional entropy.

[This an example of where we

cannot have an error $< \varepsilon$

in $\frac{1}{\sqrt{n\text{-steps}}}$]