

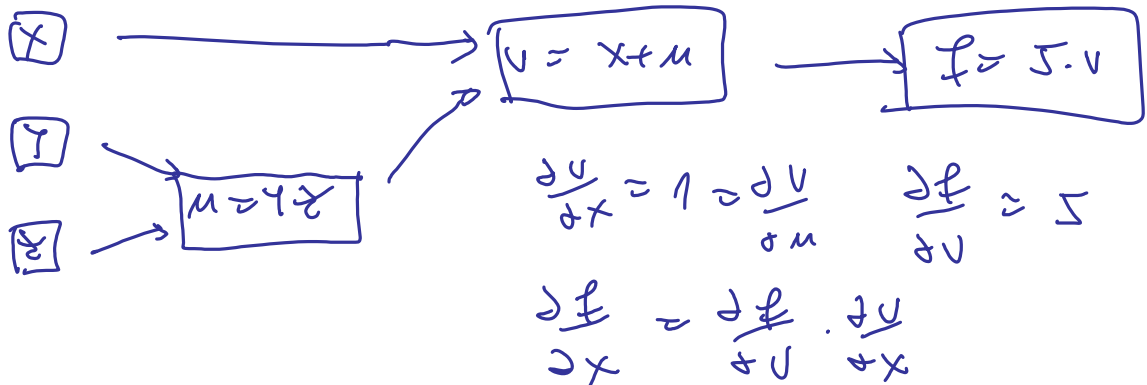
Méthodes numériques: problèmes dépendant du temps

cours de M1 2021/22 Université Paris Dauphine -

→ www

→ livre

$$f(x, y, z) = 5(x + y \cdot z)$$



$$f(t, x) = r x (1 - x) \quad \dot{x} = r x (1 - x) \dots$$

$$|f(t, x_1) - f(t, x_2)| = |r [x_1(1 - x_1) - x_2(1 - x_2)]|$$

$$= |r| |(x_1 - x_2)(1 - x_1 - x_2)| = |r| |1 - x_1 - x_2| |x_1 - x_2|$$

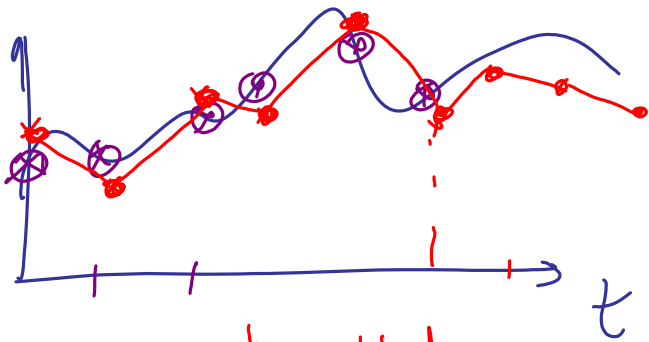
OK pour thm ≥ 1 avec $L = |r| (|1 - 2x_0| + \eta)$

car pour $x_1, x_2 \in B(x_0, \eta/2)$ on a

$$|1 - x_1 - x_2| = |1 - x_0 + x_0 - x_1 - x_0 + x_0 - x_2| \leq$$

$$|1 - 2x_0| + |x_0 - x_1| + |x_0 - x_2| \leq |1 - 2x_0| + \eta/2 + \eta/2 = |1 - 2x_0| + \eta$$

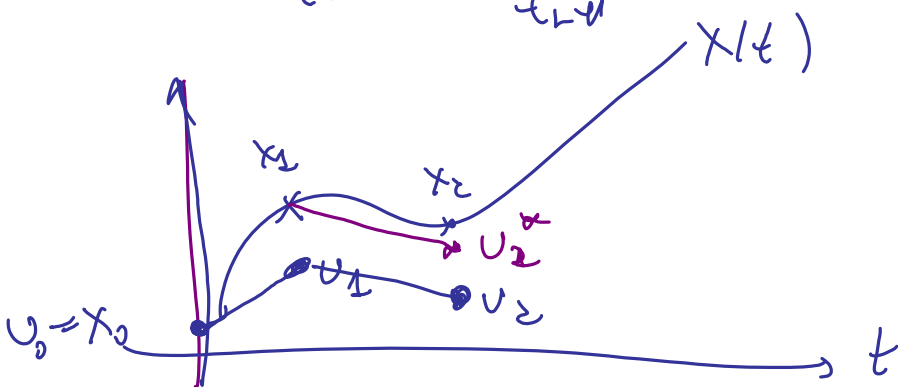
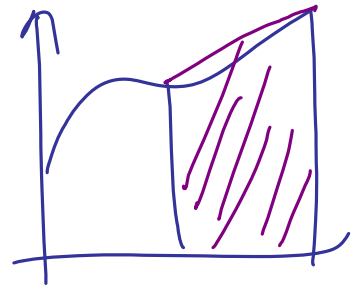
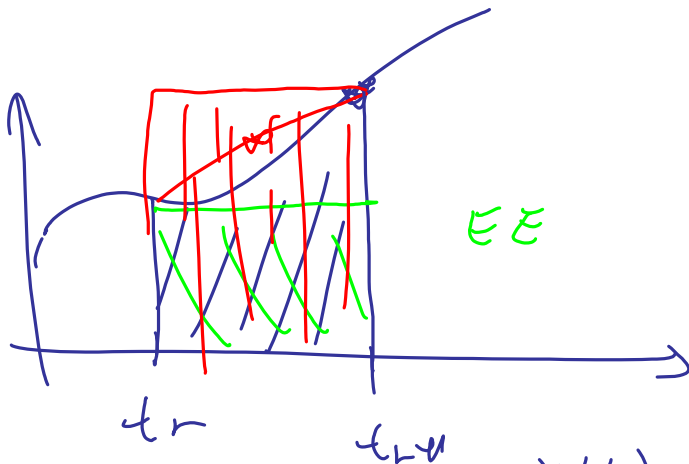
Pour contourner pour le Thm 2.2 on peut pas l'appliquer directement ss informations sur la taille des "X".



$$x' = rx^2 \quad \frac{x'}{x} = rx$$

$$f(t, x) = r \cdot x$$

pour $f(t_n) = f(t_n, u_n) = \underbrace{r u_n}_{= \phi} = \phi$ pour $\epsilon \bar{\epsilon}$



Calcul de l'erreur de troncature pour $\epsilon \bar{\epsilon}$

$$\tau_{n+1}(h) = \frac{X(t_{n+1}) - X(t_n) - h f(t_n, X_n)}{h} =$$

$$= \frac{x(t_n+h) - x(t_n) - h x'(t_n)}{h} \quad \frac{\text{Taylor exacte}}{\text{d'ordre 2}} \quad \frac{\frac{h^2}{2} x''(\xi_n)}{h}$$

(avec $\xi_n \in [t_n, t_n+h]$) = $h \cdot \frac{x''(\xi_n)}{2} = \tau_{n+1}(h)$

Donc $\tau(h) = O(h)$ si x'' bornée / admissible.

Rappel $\mathcal{O}(h^p)$ si $\lim_{h \rightarrow 0} \frac{\mathcal{O}}{h^p} = \text{borné}$

$\mathcal{O}(h^p) = o(h^p)$ si $\lim_{h \rightarrow 0} \frac{\mathcal{O}}{h^p} = 0$.

si $\mathcal{O}(h) = O(h^2)$

TD Heun (troncature)

τ pour EI : $\tau = \frac{x_{n+1} - x_n - h f(t_n, x_n)}{h} = \frac{x_{n+1} - x_n - h x'(t_n)}{h}$

$$= \frac{h x'(t_n) + \frac{h^2}{2} x''(\xi_n) - h x'(t_n)}{h} = \frac{h}{2} x''(\xi_n) - h \cdot \frac{x'(t_{n+1}) - x'(t_n)}{h}$$

$$= \frac{h}{2} x''(\xi_n) - h x''(\xi_n) = O(h)$$

$$\frac{g(b) - g(a)}{b-a} = g'(u) \quad (m/l)$$

$$w_{n+1} \leq w_n + \dots$$

$$w_n \leq w_{n-1} + \dots$$

$$w_1 \leq w_0 + \dots$$

$$w_{n+1} \leq w_0 + \dots$$

$$k_s = hA \quad p_s = h\delta_s \quad \rho_n = |w_n|$$

(Gronwall)

Verif hyp thm zero-stabilité pour $\in \mathbb{I}$

$$U_{n+1} = U_n + h f(t_n, U_{n+1})$$

$$U_{n+1} = U_n + h \underbrace{\phi(t_n, U_{n+1})}_{\phi(t_n, U_{n+1})}$$

On veut q $U_n \rightarrow f(t_n, U_{n+1})$ soit lipschitz. Comme f est lipschitz il suffit que $U_n \rightarrow U_{n+1}$ soit lipschitz.

$$|U_{n+1} - V_{n+1}| \leq |U_n - V_n| + hL |U_{n+1} - V_{n+1}| \quad \text{Pour } hL < 1$$

$$\text{on obtient } (1 - hL) |U_{n+1} - V_{n+1}| \leq |U_n - V_n|$$

$$\text{donc } |U_{n+1} - V_{n+1}| \leq \frac{1}{1 - hL} |U_n - V_n| \quad \text{e.t.c.}$$

Dern Thm 2.12.

$$U_{n+1} = U_n + h \phi(\cdot, U_n)$$

$$X_{n+1} = X_n + h \phi(\cdot, X_n) + h z_n(h) \quad (\text{eq. avant } 2.5)$$

ds thm 2.10

$$z_{n+1} = z_n + h \phi(\cdot, z_n) + h \delta_{n+1} \quad (\text{eq. (2.7)})$$

Donc si $\delta_{n+1} := z_n(h)$ alors $z_k = X_k \quad \forall k$.

Par thm 2.10 $|U_n - X_n| = |U_n - z_n| \leq C \varepsilon$ Thm 2.10.

avec ε qui majore tous les f_k par ex $O(h^p)$.

Donc $|U_n - x_n| = O(h^p)$) $\left(\begin{array}{l} f_0 \text{ par hyp. est} \\ O(h^p) \end{array} \right)$

Si ~~est~~ $h \approx 10^{-3}$ \rightarrow EE, EI : erreur 10^{-3}
 \leftarrow CM, M : erreur 10^{-6}
 $h = 10^{-8}$ \rightarrow 10^{-8}
 \rightarrow 10^{-16}

$$U_{n+1} = U_n + \frac{h}{2} (\lambda U_n + \lambda U_{n+1})$$

$$\left(1 - \frac{h\lambda}{2}\right) U_{n+1} = \left(1 + \frac{h\lambda}{2}\right) U_n \Rightarrow U_{n+1} = \frac{1+h\lambda/2}{1-h\lambda/2} U_n$$

$$\left|1 + \frac{h\lambda}{2}\right| < \left|1 - \frac{h\lambda}{2}\right|$$

$$\downarrow$$

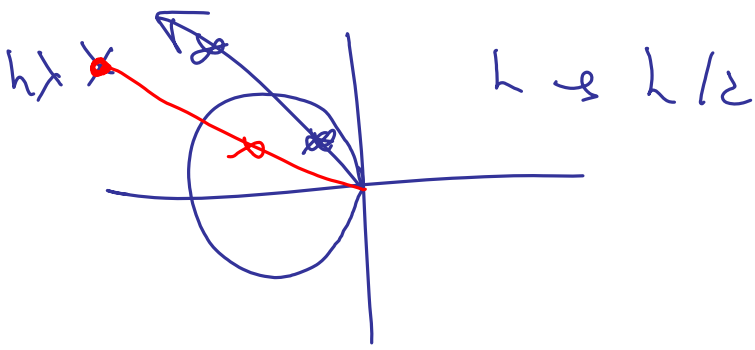
$$\text{dist}\left(\frac{h\lambda}{2}, -1\right) < \text{dist}\left(\frac{h\lambda}{2}, 1\right)$$

$$U_{n+1} = U_n + \frac{h}{2} \left[f(t_n, U_n + h \frac{h\lambda}{2} U_n) + f(t_n, U_n) \right]$$

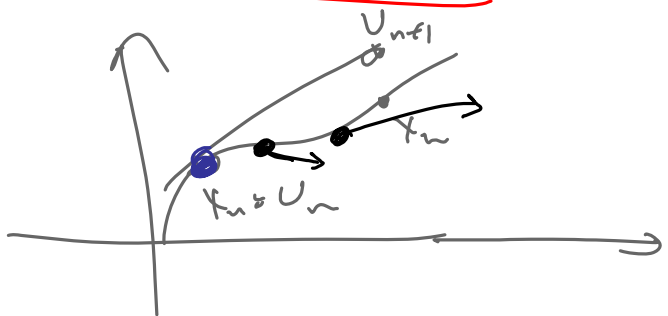
$$U_{n+1} = U_n + \frac{h}{2} \left[\lambda (U_n + h \lambda U_n) + \lambda U_n \right]$$

$$= U_n \left(1 + \frac{h}{2} + \frac{h^2 \lambda^2}{2} + \frac{h}{2} \right) = U_n \left(1 + h + \frac{h^2 \lambda^2}{2} \right)$$

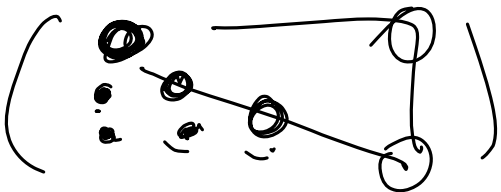




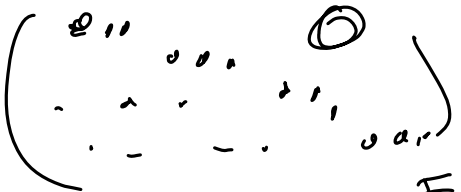
\mathbb{R}^k



$$X_{n+1} = X_n + h \int_{t_n}^{t_{n+1}} f(t, X(t)) dt$$



" Δ inf" (stuck)

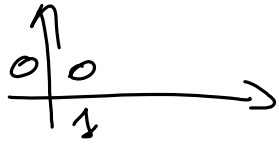


" Δ inf" (pas stuck)

$\exists \epsilon$

$$U_{n+1} = U_n + h \int_{t_n}^{t_{n+1}} f(t, U_n) dt$$

$c_1 = 0, a_1, \dots, 0$



Dem Prop 2.24

Consistance si $\lim_{h \rightarrow 0} \tau(h) = 0$

Pour les schémas R-K : $\tau(h) = \frac{|X(t_{n+1}, h) - U_{n+1}|}{h}$
(explicit)

avec $U_{n+1}^{\alpha} = X_n + h \sum_{i=1}^s b_i k_i$ k_i : calculés à partir de X_n

$$\tau(h) = \frac{X(t_n+h) - X(t_n) - h \sum_{i=1}^s b_i k_i}{h}$$

$$\underline{\text{Taylor}} \quad \frac{h X'(t_n) + \frac{h^2}{2} X''(\xi) - h \sum_{i=1}^s b_i (f(t_n, X_n) + \mathcal{O}(h))}{h}$$

car $k_i = f(t_n + c_i h, X_n + h \sum_{j=1}^s c_{ij} k_j)$

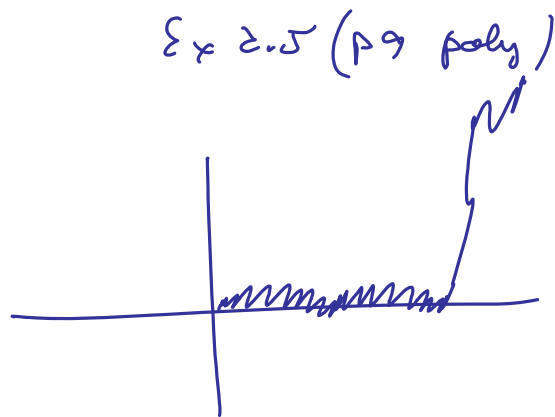
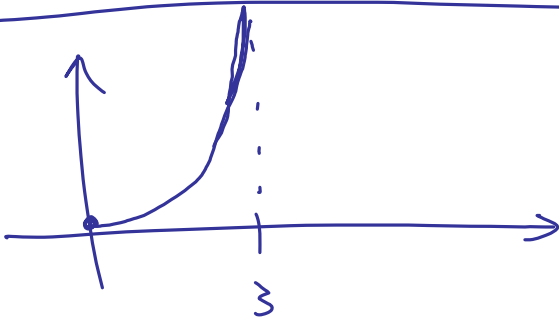
Taylor

$$f(t_n, X_n) + f_t(t_n, X_n) \cdot c_i h + f_x(t_n, X_n) \cdot h (\sum_{j=1}^s c_{ij} k_j) + \mathcal{O}(h^2) = f(t_n, X_n) + \mathcal{O}(h)$$

$$= \left[X'(t_n) - \left(\sum_{i=1}^s b_i \right) f(t_n, X_n) \right] + \mathcal{O}(h)$$

$$= f(t_n, X_n) [1 - \sum b_i] + \mathcal{O}(h). \quad \text{Pour que } \tau(h) \xrightarrow{h \rightarrow 0} 0$$

il faut et il suffit que $\sum_{i=1}^s b_i$ soit égale à 1.



$\tau'(t) = -1000 \tau$ est stable pour quel h ?

$$U_{n+1} = U_n + h f_n = U_n + h (-1000 U_n) = (1 - 1000h) U_n$$

$$U_{n+1} = (1 - 1000h) U_n = \dots = (1 - 1000h)^{n+1} U_0.$$

Pour stabilité il faut $\underbrace{|1 - 1000h| < 1}_{-1 < 1 - 1000h < 1}$

$$\underbrace{1000h < 2}_{\Rightarrow h < 2/1000}$$

R-K implémente : il faut résoudre un système d'eq. (peut être non-linéaires).

$$\begin{cases} k_1 = f(t_n + c_1 h, U_n + h \sum_{e=1}^s a_{1e} k_e) \\ k_2 = \dots \\ k_s = f(t_n + c_s h, U_n + h \sum_{e=1}^s a_{se} k_e) \\ \dots \end{cases}$$

On utilise des itérations de type Picard.

Rappel pour trouver un point fixe de $g(\cdot)$:

c'est à dire une solution de $g(x) = x$, on définit la suite $x_0 = g(x_0)$, $x_{n+1} = g(x_n)$.

Si $g =$ contraction alors $x_n \xrightarrow{n \rightarrow \infty} x^*$ avec $x^* =$ pt fixe.

→ def: $|g(x) - g(y)| < L |x - y| \quad \forall x, y$
avec $L < 1$.

RK CG marche aussi en dim > 1 .

thm de Picard

Méthodes multi-pas

$$y_{n+s} + a_{s-1} y_{n+s-1} + \dots = h \sum_{k=0}^s b_k f(t_{n+k}, y_{n+k})$$

inconnue
déjà calculés

Coût : 1 calcul ~~à~~ nouveau de $f(\cdot, \cdot)$ par pas de tps.

Montrons la consistance de BDF

BDF satisfait $s=2$ $a_0 = \frac{1}{3}$ $a_1 = -\frac{4}{3}$ $a_2 = 1$
 $b_0 = 0$, $b_1 = 0$, $b_2 = \frac{2}{3}$

Alors $\sum a_n = \frac{1}{3} - \frac{4}{3} + 1 = 0$

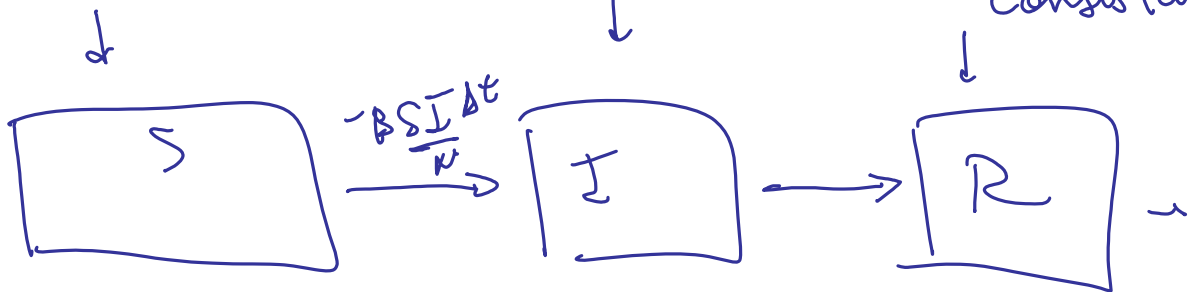
$\sum_{k=0}^s b_k = b_0 + b_1 + b_2 = \frac{2}{3}$

$\sum_{k=0}^s k a_k = 0 \cdot a_0 + 1 \cdot a_1 + 2 \cdot a_2 = 0 - \frac{4}{3} + 2 = \frac{2}{3}$

$\Rightarrow \sum b_k = \sum a_n \cdot k$ Thm 2.35

Comme $\sum a_n = 0$

\Rightarrow BDF est consistant.



~~$I(t) - I(t+\Delta t)$~~

$R(t+\Delta t) - R(t)$

$\approx I \cdot \Delta t \cdot \gamma$

\approx

$R'(t) \stackrel{\Delta t \rightarrow 0}{\leftarrow} \frac{R(t+\Delta t) - R(t)}{\Delta t} \approx \gamma I$

$$\frac{S(t) - S(t+\Delta t)}{\Delta t} \xrightarrow{\Delta t \rightarrow 0} -S'(t)$$

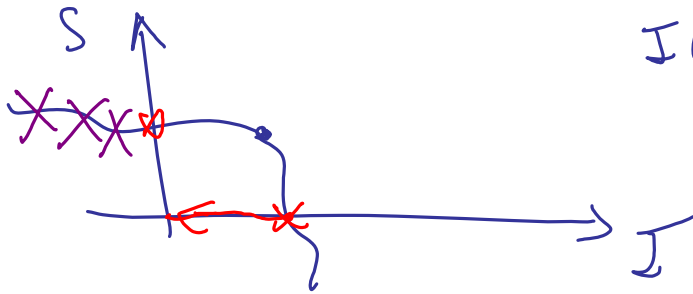
$$\parallel$$

$$\beta S I$$

$$S' = -\beta S I + \mu \text{ tx naissances / ...}$$

Question S, I, R sont-ils positifs ?

$$s = \frac{S}{N}, i = \frac{I}{N}, r = \frac{R}{N} \text{ (proportions)}$$



$$I(t^*) = 0 \Rightarrow S'(t^*), R'(t^*) = 0$$

$$\downarrow t > t^*$$

$$S(t^*) = 0 \quad \begin{cases} S' = 0 \\ I' = -\alpha I \end{cases}$$

~~$$S' = -\beta S I - \mu S$$~~

$$\left\{ \begin{array}{l} S(t^*) = \dots \\ I(t^*) = 0 \\ R(t^*) = \dots \end{array} \right\} \begin{cases} S(t) = S(t^*) \\ I(t) = 0 \\ R(t) = R(t^*) \end{cases}$$

par unicité

$$s' = -\beta s i$$

$$i' = (\beta s - \alpha) i$$

$$R_0 = \beta / \alpha$$

$$\beta s > \alpha \quad s > \frac{\alpha}{\beta} ? \quad \left(\frac{1}{R_0} \right)$$

$$s = \frac{S}{N} > \frac{1}{R_0}$$

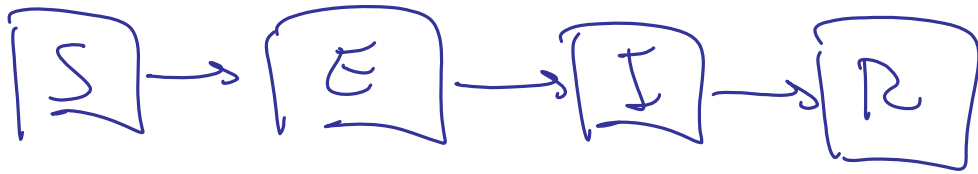
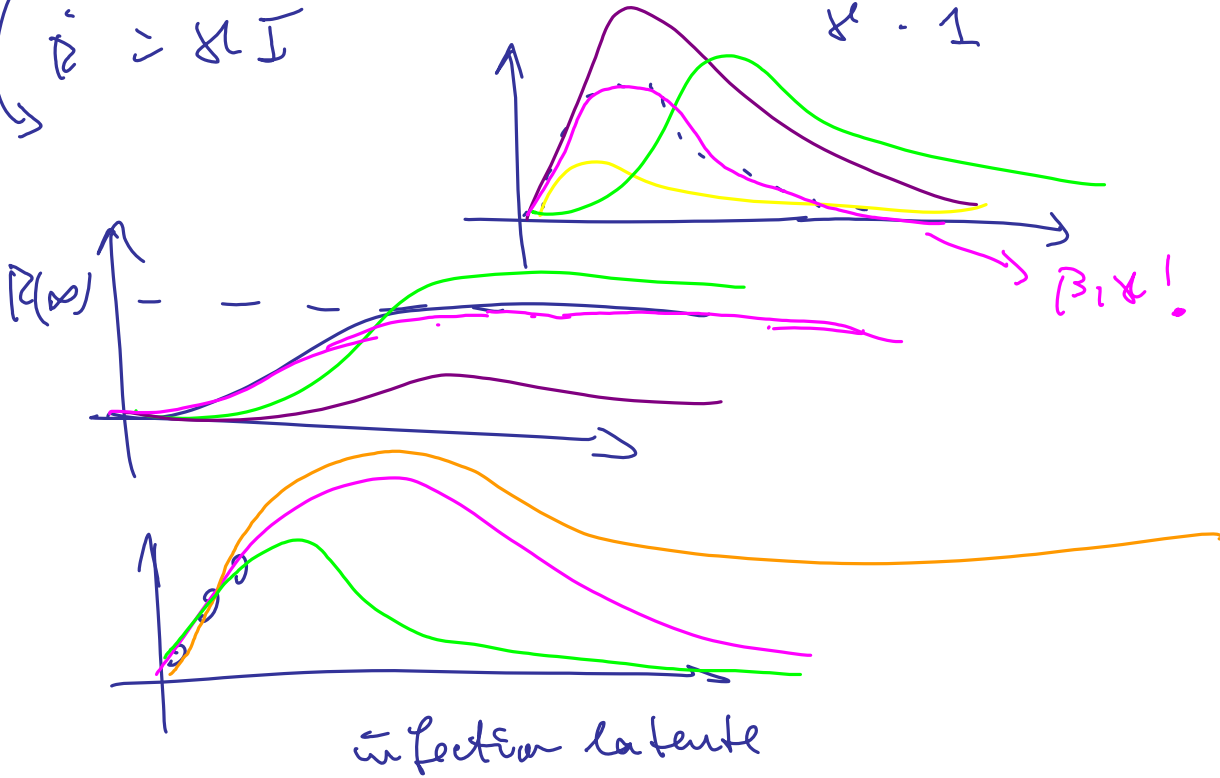
$$R_0 < 1$$

$$\frac{1}{R_0} > 1$$

~~$$s > \frac{1}{R_0}$$~~

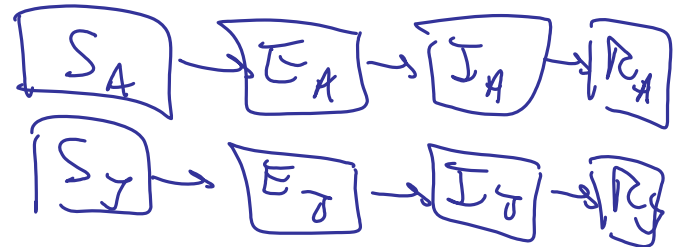
$$R(t) \rightarrow I(t) \approx \frac{R(\text{max} - R(t))}{\mu - 1}$$

$\dot{R} = \mu I$



$$\begin{cases} \dot{S} = -\beta SI \\ \dot{E} = \beta SI - \mu_E E \\ \dot{I} = \mu_E E - \mu_I I \\ \dot{R} = \mu I \end{cases} \quad 1/\mu_E$$

"Strukturmatrix" (par age)



$$\dot{S}_A = -\beta S_A \left(\frac{I_Y}{N} c_1 + \frac{I_A}{N} c_2 \right)$$

...

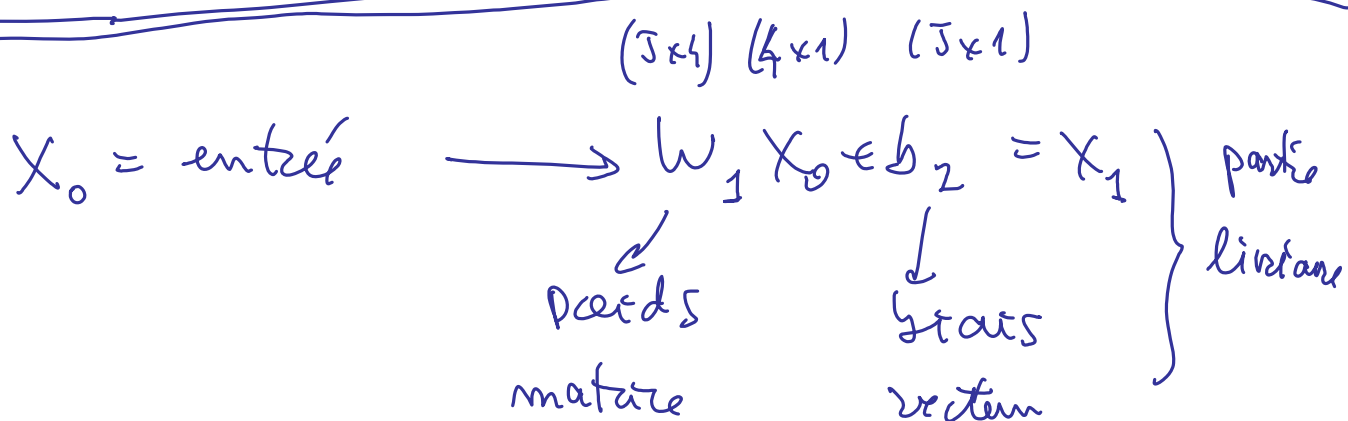
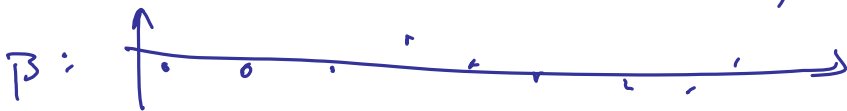
§3. Backward = $\Sigma(\beta)$

$$\frac{d}{d\beta(t)} \int_0^T c(\beta(t)) dt = c'(\beta(t))$$

$$\left[\begin{aligned} c(\beta) &= \alpha(t) \cdot \beta(t) & \int_0^T c(\beta(t)) dt &= \int_0^T \alpha(t) \beta(t) dt \\ & & & \\ &= \langle \alpha, \beta \rangle_{L^2(\Sigma_{0,T})} & \nabla \left[\int_0^T c(\beta(t)) dt \right] &= \left(\alpha(t) \right)_{t \in \Sigma_{0,T}} \end{aligned} \right]$$

fonction

$$\Sigma(\beta) = \Sigma(\beta_0) + \langle c'(\beta(\cdot)), \beta - \beta_0 \rangle \in \mathcal{O}(\|\beta - \beta_0\|)$$



Partie "activation" : $X_1^a = \text{fct d'activation}(X_1)$

Ex : $\text{actio} = \text{ReLU} = (-)_+$



$$X_2 = W_2 X_1^a + b_2 \quad X_2^a = \text{ReLU}(X_2)$$

La dernière couche activation = e.g. sigmoïde.
 $\mathbb{R} \rightarrow \Sigma_{0,1}$

CIFAR 10 / MNIST : 10 categ.

CIFAR 100 : 100 = K

Activation = "softmax" $(x_1, \dots, x_k) \rightarrow \left(\frac{e^{x_k}}{\sum_{l=1}^k e^{x_l}} \right)_{k=1}^k$
 $= P_k$

Taille de X = $\left(\begin{matrix} 5 \times 4 & + & 5 \\ \underbrace{\hspace{1.5cm}}_{w_1} & & b_1 \end{matrix} \right) + \left(\begin{matrix} 7 \times 5 & + & 7 \\ \underbrace{\hspace{1.5cm}}_{w_2} & & b_2 \end{matrix} \right) + \left(\begin{matrix} 7 \times 1 \\ \underbrace{\hspace{1.5cm}}_{w_3} \end{matrix} \right)$

$$f'(x) = \frac{f(x+h) - f(x)}{h} + \dots$$

$$\frac{f(x+h) - f(x)}{h} \stackrel{\text{Taylor}}{=} \frac{f'(x)h + f''(\xi_1) \frac{h^2}{2}}{h} = f'(x) + O(h)$$

$$\frac{f(x+h) - f(x-h)}{2h} \stackrel{\text{Taylor}}{=} \frac{\cancel{f(x)} + f'(x)h + \cancel{f''(\xi_1) \frac{h^2}{2}} + f^{(3)}(\xi_1) \frac{h^3}{6}}{2h}$$

$$= \frac{\cancel{f(x)} - f'(x)h + \cancel{f''(\xi_2) \frac{h^2}{2}} - f^{(3)}(\xi_2) \frac{h^3}{6}}{2h}$$

$$= f'(x) + \frac{f^{(3)}(\xi_1) + f^{(3)}(\xi_2)}{2 \cdot 6} h^2 = f'(x) + O(h^2)$$

$f(x)$: $\mathcal{I} \rightarrow \mathcal{O}(Y_1, \dots, Y_{k-1}, Y_{k+h}, Y_{k+1}, \dots, Y_N)$

A calculer et $\mathcal{O}(Y_1, \dots, Y_{k-1}, Y_{k-h}, Y_{k+1}, \dots, Y_N)$

Cost : \geq calculs de G par dimension de Y .

$= \geq \cdot N$ calculs (en tout cas $> N$) de $G(\cdot)$

Si un calcul coûte $O(N)$ unités de tps, coût total $O(N^2)$.

$$\frac{10^{-16}}{2h} \approx h^2$$

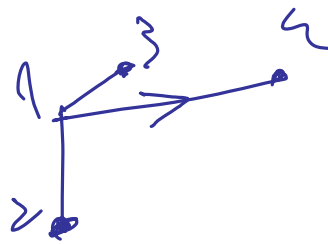
$$h^3 \approx 10^{-16} \quad h \approx 10^{-16/3}$$

$$h^2 \approx 10^{-16 \cdot \frac{2}{3}} \quad 10^{-10.66}$$

$$\rightarrow \frac{10^{-16}}{h} \approx h$$

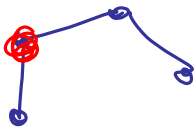
$$h \approx 10^{-8} \quad \text{En } 10^{-8}$$

Graph



J

$V \in \{1, 2, 3, 4, 5\}$
(1,4)



$$\mathbb{R}^N \xrightarrow{G} \mathbb{R}^L \xrightarrow{f} \mathbb{R}$$

$$F = f \circ G$$

$$\delta U_6 \approx \frac{\delta f}{\delta U_6} = \frac{\delta U_7}{\delta U_6} = \frac{\delta (U_6 - U_{114})}{\delta U_6} = 1$$

$$\delta U_{113} \approx \frac{\delta f}{\delta U_{113}} = \frac{\delta f(U_{114}, U_6)}{\delta U_{113}} = \frac{\delta f}{\delta U_{114}} \cdot \frac{\delta U_{114}}{\delta U_{113}} + \dots$$

$$\frac{\partial f}{\partial U_6} \cdot \frac{\partial U_6}{\partial U_{13}} = \delta U_{14} \cdot \frac{\partial U_{14}}{\partial U_{13}} + \delta U_6 \cdot \frac{\partial U_6}{\partial U_{13}}$$

$$= (-1) \cdot 1 + 1 \cdot 0 = -1.$$

$$\delta U_5 = \frac{\partial f}{\partial U_5} = \frac{\partial f(U_{14}, U_6)}{\partial U_5}$$

$$= \frac{\partial f(U_{14}(U_{13}, U_5, U_{32}), U_6(U_{13}, U_5, U_{32}))}{\partial U_5}$$

$$= \frac{\partial f}{\partial U_{14}} \frac{\partial U_{14}}{\partial U_5} + \frac{\partial f}{\partial U_6} \frac{\partial U_6}{\partial U_5} = \delta U_{14} \cdot 0 + \delta U_6 \cdot U_{32}$$

$$= U_{32}.$$

$$\delta U_{32} = \delta U_{14} \frac{\partial U_{14}}{\partial U_{32}} + \delta U_6 \frac{\partial U_6}{\partial U_{32}} = 1 \cdot U_5 = U_5$$

$$\delta U_{13} = \delta U_{14} \frac{\partial U_{14}}{\partial U_{13}} + \delta U_6 \frac{\partial U_6}{\partial U_{13}} = (-1) \cdot 1 = -1$$

$$\delta U_{12} = 0 + \delta U_{13} \frac{\partial U_{13}}{\partial U_{12}} = (-1) \cdot 1 = -1$$

$$\delta U_4 = 0 + \delta U_5 \frac{\partial U_5}{\partial U_4} = U_{32} \cdot 1 = U_{32}$$

$$\delta U_{212} = 0 + \delta U_5 \frac{\partial U_5}{\partial U_{212}} = U_{32} \cdot 1 = U_{32}$$

$$\delta U_{312} = \delta U_{32} \frac{\partial U_{32}}{\partial U_{312}} = \delta U_{32} = U_5$$

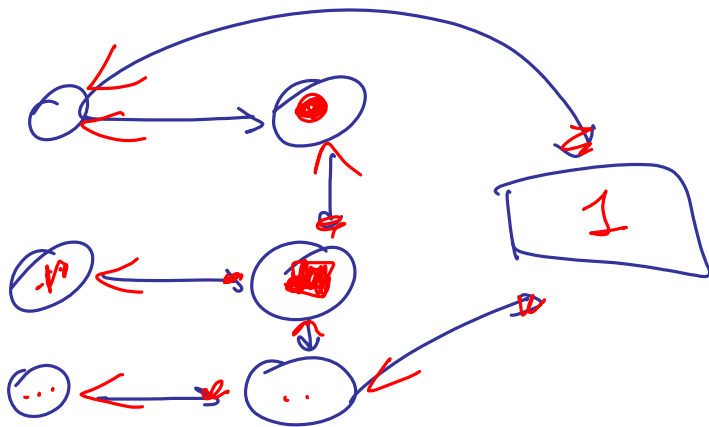
$$\delta U_1 = \delta U_{12} \frac{\partial U_{12}}{\partial U_1} + \delta U_4 \frac{\partial U_4}{\partial U_1} = (-1) \cdot 1 + U_{32} \cdot \frac{\partial U_1}{\partial U_1}$$

il devrait y avoir toutes les dérivées des U_{13}, U_5, U_{32}

$$\delta y = \delta U_2 = \delta U_{z_{12}} \cdot \frac{\partial U_{z_{12}}}{\partial U_2} = U_{z_{13}} - 1 = U_{z_{13}}$$

$$\delta z = \delta U_3 = \delta U_{z_{12}} \cdot \frac{\partial U_{z_{12}}}{\partial U_3} = U_5 \cdot 1 = U_5$$

$$\text{Donc } (\delta x, \delta y, \delta z) = \nabla f = (6U_1 \cdot U_{z_{13}} - 1, U_{z_{13}}, U_5) \\ = (6x^2 - 1, z, 3x^2 + y)$$



$\nabla_x [\mathbb{E}_w [L(x, w)]] = \text{très long à calculer!}$
 $|\Omega|$ # propagations directes et backward

On remplace par une approximation bruitée :

$$\nabla_x L(x, w) = \nabla_x [\mathbb{E}(L)] + (\text{bruit})$$

ou sinon $\nabla_x \mathbb{E}(L) \approx \frac{\sum_{b=1}^B \nabla_x L(x, w_b)}{B}$ $B = \text{"batch"}$
 $\approx \dots \approx 32/64$
 128

$w_b = \text{tirés au sort ds } \Omega!$

Ceci est couplé avec un algo de type descente de gradient

$$x_{k+1} = x_k - \epsilon_k \cdot \nabla_x L(x_k, w)$$

dit SGD gradient ptr à x_k

$$\nabla_x \mathbb{E}(L) = \frac{\sum_{w \in \Omega} \nabla_x L(x, w)}{|\Omega|}$$

Algo similaires : "Adam", "Nesterov", "momentum"
 "AdaGrad", "AdaMax", ..., SGD-G2.

$$\delta w_2 = \frac{\partial g}{\partial y_2} \cdot \frac{\partial y_2}{\partial \tilde{y}_2} \cdot \frac{\partial \tilde{y}_2}{\partial w_2} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} \in \mathbb{R}^{7 \times 5}$$

$w_1: 5 \times 4$
 $b_1: 5 \times 1$
 $w_2: 7 \times 5$ $b_2: 7 \times 1$

$$\tilde{y}_2 = w_2 y_1 + b_2$$

$$\frac{\partial \tilde{y}_2}{\partial w_2} = y_1 \otimes \text{Id}_{7 \times 7}$$

matrice $7 \times (7 \times 5)$

produit tensoriel

$w_2 \in \mathbb{R}^{7 \times 5}$
 $\tilde{y}_2 \in \mathbb{R}^7$

$\varepsilon_x = (2, 1) \otimes \text{Id}_{3 \times 3} = (2 \text{ Id}_1, \text{ Id}_2)$

$(2 \times 3) \quad \begin{pmatrix} 3 & \sqrt{2} & \sqrt{7} \\ 12 & 11 & 41 \end{pmatrix} \otimes \text{Id}_{2 \times 2} = \begin{pmatrix} 3 \text{ Id}_2 & \sqrt{2} \text{ Id}_2 & \sqrt{7} \text{ Id}_2 \\ 12 \text{ Id}_2 & 11 \text{ Id}_2 & 41 \text{ Id}_2 \end{pmatrix}$

$$\begin{pmatrix} 3 & 0 & \sqrt{2} & 0 & \sqrt{7} & 0 \\ 0 & 3 & 0 & \sqrt{2} & 0 & \sqrt{7} \\ 12 & 0 & 11 & 0 & 41 & 0 \\ 0 & 12 & 0 & 11 & 0 & 41 \end{pmatrix}$$

$2 \times 3 \otimes 2 \times 2 \rightarrow 4 \times 6$

Exemples de fonctions d'activation



$\rightarrow \text{ReLU}(x_1, \dots, x_N) = ((x_1)_+, \dots, (x_N)_+)$

$\rightarrow \text{Softmax } \mathcal{S}(x_1, \dots, x_N) = \left(\frac{e^{x_k}}{\sum_{e=1}^N e^{x_e}} \right)_{k=1}^N$

résultat = loi de proba discrète sur un ensemble à N éléments

Exemples de fonction "loss"

\rightarrow moindres carrés $\mathcal{L}(x) = \sum_{i=1}^N (x_i - y_i)^2$ ($Y = (y_1, \dots, y_N)$)
 = données

\rightarrow cross-entropy $\mathcal{L}(q_1, \dots, q_N) = - \sum_{k=1}^N p_k \ln(q_k)$

(p_k) = loi de proba données, "divergence Kullback-Libler"

TD Dériver "cross-entropy" o softmax.

