

# STATISTIQUES SUR DES VARIÉTÉS

(1)

## Introduction:

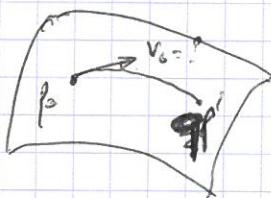
- $M$  une variété.  $\{y_i\}$  des données vues comme des points sur  $M$ .
- $D: M \times M \rightarrow \mathbb{R}_+$  une fonction de similitude différentiable sur  $M$ , facilement calculable et évidemment localement la proximité entre 2 points sur  $M$ .
  - on peut prendre la distance géodésique mais rarement calculable facile
  - $M$  un ouvert de  $\mathbb{R}^n$ : Dès la distance Euclidienne.
  - $M$  une sous-variété de  $\mathbb{R}^m$ : idem
  - $M$  une image  $\in L^2(\Omega, \mathbb{R})$ : idem
  - $M$  un ensemble de points  $\mathbb{L}^2$  ou connexes ou autres...

But: trouver des représentations bonnes [le géodésique est l'outil de choix pour "positionner" un pt par rapport à un autre].

## I - Le Recalage ou le calcul du lag Biromin.

[paramétrisé par un vecteur  $v \in T_p M$  enac linéaire]

1.



$$\text{Calculer } v_0 / \text{Eupp}_p(v_0) = q'$$

C'est un problème inverse.

C'est un problème de contrôle optimal.  $v_0$  est le contrôle (ou contrôlable négatif) pour l'état du système.

On va le résoudre par la résolution d'un problème d'optimisation.

$$\min_{v_0} D\mathcal{C} \text{Eupp}_p(v_0, q')$$

(2)

En utilisant la formulation hamiltonienne des cinématiques, on a:

$$\begin{aligned} \text{Exp}_{p_0}(v_0) = p(1) \quad \text{avec} \quad & \left\{ \begin{array}{l} \dot{p} = + \frac{\partial H}{\partial q} \\ \dot{q} = - \frac{\partial H}{\partial p} \end{array} \right. \quad \begin{array}{l} p(0) = p_0 \\ q(0) = G(p_0)^{-1} v_0 \end{array} \\ & \text{et } H(p, q) = \frac{1}{2} q^T G(p)^T q. \end{aligned}$$

En écrivant  $S(t) = (p(t), q(t))$  avec  $\dot{S}(t) = F(S(t))$   
 $S(0) = S_0 = (p_0, G(p_0)^{-1} v_0)$

on est amené à vouloir résoudre un problème du type:  $\min_{S(t)} f(S(t))$

où  $f$  est différentiable.

$$\begin{cases} \dot{S}(t) = f(S(t)) \\ S(0) = S_0 \end{cases}$$

## 2. Calcul du gradient

On calcule le dérivé de l'éikan de  $S(t)$  par une perturbation  $\delta S_0$  de sa condition initiale:

On calcule tout d'abord:  $\begin{cases} \dot{\delta S}(t) = \frac{dF}{dS} \delta S(t) \\ \delta S(0) = \delta S_0 \end{cases}$  D'où  $\delta S(t) = \exp\left(\int_0^t \frac{d}{dS(u)} F du\right) \delta S_0$

D'où  $\frac{d}{S_0} f \delta S_0 = \frac{d}{S(t)} f \delta S(t)$

$$= \frac{d}{S(t)} f \exp\left(\int_0^t \frac{d}{dS(u)} F du\right) \delta S_0$$

et:  $\nabla_{S_0} f = \underbrace{\exp\left(\int_0^t \frac{d}{dS(u)} F du\right)}_{= S(t)} \nabla_{S(t)} f$

En partant:  $v(t) = \text{solution de } \dot{v}(t) = - \frac{d}{dS(t)} F v(t)$

$$v(1) = \nabla_{S(1)} f$$

on a  $\boxed{v(0) = \nabla_{S_0} f}$

→ solution d'un op. linéaire rétrograde.  
 (backprop de  $\nabla_{S(1)} f$  à  $v(0) = \nabla_{S_0} f$ )

(3)

Rq: En prenant une variation de SCLI pour rapport à un autre temps de ref  $t_0$ , on peut montrer que  $\dot{Y}(t) = \nabla_{SCLI} f$  pour tout  $t \in [t_0, t]$ .

Algo:

$$V_0 \leftarrow V_0^{(0)} \text{ initialisation} \quad q_0^{(0)} \leftarrow G(p_0) \quad h=0 \quad \begin{cases} \text{Tout écrit dans une} \\ \text{carte locale.} \end{cases}$$

Jusqu'à (critère d'arrêt)

- ~~Initialiser  $q_0^{(0)} \leftarrow G(p_0)$~~
- Intégrer (Euler):  $(\dot{p}, \dot{q}) = F(p, q)$  avec  $p(0) = p_0$ ,  $q(0) = q_0^{(0)}$
- Calculer  $\nabla_{p^{(k)}} f$ ,  $\dot{Y}(1) \leftarrow \nabla_{p^{(k)}} f$
- Intégrer逆 (Euler):  $\dot{v}(t) = -d \frac{\partial}{\partial (q(t), p(t))} F^T \dot{Y}(t)$
- $q_0^{(k+1)} \leftarrow Y(0)$  (condition de  $Y(0)$  connue au moment  $q$ )
- $h \leftarrow k+1$ .

Rq:  $d_S f$  requiert de savoir calculer la différentielle de la co-matième.

Variante:

- Calcul de  $\nabla_S f$  par différentiation automatique
- Amélioration de la qualité de gradient (L-BFGS par exemple)
- ou méthodes d'ordre supérieur
- si dimension pas trop grande.

3. Exemple: Recalage différentiel par points de contrôle.

→ recalage d'image

→ recalage de feuilleau de fils

→ recalage de filtres.

Rq:  $H(v, s) = v^T F(s)$

$$\text{on a } \dot{s} = \frac{\partial H}{\partial v} \quad \dot{v} = -\frac{\partial H}{\partial s}$$

Rq: On a calculé  $y(0) = \nabla_{p_0} f$

si on ne le prend pas en compte

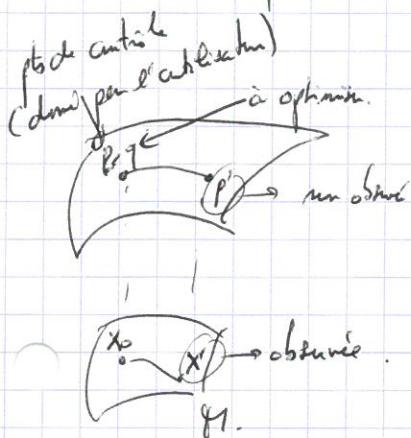
ce min  $f(s(1))$  n'est un optimum

évident:  $p_0 = p_1$  et  $v_0 = 0$  !!

3 b)

### 3. Cas d'une variété mythométrique.

Le schéma est un peu plus subtil puisque le gradient ne traîne pas directement l'objet  $x_0$  mais un ensemble de points de contrôle. On cherche à optimiser.



$$E(q_0) = \mathcal{F}(X(1)) \quad \text{où} \quad \begin{cases} \dot{S}(t) = F(S(t)) \\ \dot{X}(t) = G(S(t), X(t)) \end{cases} \quad \begin{matrix} S_0 = S \\ X_0 = X \end{matrix}$$

La même émanation donne :  $\delta S(t) = \exp\left(\int_0^t d_{S(u)} F du\right) \delta S_0$

$$\delta \dot{X}(t) = \partial_{S(t)} \delta S(t) + \partial_{X(t)} \delta X(t) \quad \underline{\delta X_0 = 0}.$$

Variation de la constante :  $\delta X(t) = \exp\left(\int_0^t \partial_{X(u)} G du\right) h(t)$

$$\text{avec } h'(t) = \exp\left(-\int_0^t \partial_{X(u)} G du\right) \partial_{S(t)} G \exp\left(\int_0^t d_{S(u)} F du\right) \delta S_0$$

$$h(t) = \int_0^t \exp\left(-\int_s^u \partial_{X(s)} G ds\right) \partial_{S(u)} G \exp\left(\int_s^u d_{S(s)} F ds\right) \delta S_0$$

$$\text{Finalement : } \delta X(1) = \int_0^1 \exp\left(\int_u^1 \partial_{X(s)} G ds\right) \partial_{S(u)} G \exp\left(\int_s^1 d_{S(s)} F ds\right) \delta S_0$$

$$\text{Et } \nabla_{S_0} f = \int_0^1 \exp\left(\int_0^t d_{S(u)} F^T du\right) \partial_{S(t)} G^T \exp\left(\int_t^1 \partial_{X(u)} G^T du\right) \nabla_{S(t)} f.$$

$$\theta(t) = \exp\left(\int_t^1 \partial_{X(u)} Q^\top du\right) V_{S(t)} f \quad : \quad \theta(1) = V_{S(1)} f \quad \dot{\theta}(t) = \partial_{X(t)} Q^\top \theta(t)$$

$$\xi(t) = \int_t^1 \underbrace{\exp\left(\int_s^1 d_{S(u)} F^\top du\right) \partial_{S(s)} Q^\top \theta(s)}_{g(t,s) \text{ avec } t \leq s} ds$$

$$\text{on a } \frac{\partial g}{\partial t}(t,s) = -d_{S(t)} F^\top g(t,s) \quad \text{et}$$

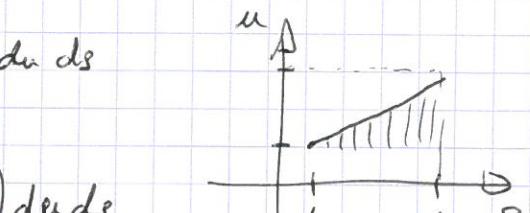
$$g(s,s) = \partial_{S(s)} Q^\top \theta(s)$$

$$\begin{aligned} \text{Donc } g(u,s) &= \int_s^u \frac{\partial g}{\partial u}(u,s) du + g(s,s) = \partial_{S(u)} Q^\top \theta(s) + \int_s^u d_{S(u)} F^\top g(u,s) du \\ &= \partial_{S(u)} Q^\top \theta(s) + \int_t^s d_{S(u)} F^\top g(u,s) du. \end{aligned}$$

$$\text{Donc } \xi(t) = \int_t^1 \left( \partial_{S(s)} Q^\top \theta(s) + \int_s^t d_{S(u)} F^\top g(u,s) \right) du ds$$

$$= \int_t^1 \left( \partial_{S(s)} Q^\top \theta(s) + \int_s^t d_{S(u)} F^\top g(s,u) \right) du ds$$

$$= \int_t^1 \left( \partial_{S(s)} Q^\top \theta(s) + d_{S(s)} F^\top \int_s^t \exp\left(\int_s^u d_{S(p)} F^\top dp\right) \cdot \right.$$



$$\begin{aligned} &\left. \partial_{S(u)} Q^\top \theta(u) du \right) = \int_t^1 \int_u^1 h(u,s) ds du \\ &\left. \right) ds = \int_t^1 \int_s^1 h(s,u) du ds \end{aligned}$$

$$= \int_t^1 \partial_{S(s)} Q^\top \theta(s) + d_{S(s)} F^\top \xi(s) ds$$

Et finalement  $\xi(t)$  solution de

$$\begin{cases} \dot{\xi}(t) = -\left[ \partial_{S(t)} Q^\top \theta(t) + d_{S(t)} F^\top \xi(t) \right] \\ \xi(1) = 0 \end{cases}$$

(4)

## II - Régression géodésique :

On a une série d'observations  $(y_i)_{i \in \mathbb{N}}$  associée à un co-facteur (typique un âge ou un point temporel)  $(t_i)_{i \in \mathbb{N}}$ .

On veut faire pour une géodésique aussi pris que possible de ces points en optimisant un critère des moindres carrés.

$$\min_{(p_0, v_0)} \sum_{i=1}^n D(E_{pp}(t_i, s_0), y_i).$$

On a donc un vecteur minimum par rapport au point de référence.

$$\text{Si } M = \mathbb{R}^2, \text{ mitigeur plate et } D = \| \cdot \|^2 \text{ alors } \min_{(p_0, v_0)} \sum_{i=1}^n \| p_0 + v_0 t_i - y_i \|^2$$

c'est une régression linéaire avec

$$\begin{cases} p_0 : \text{ordonnée à l'origine} \\ v_0 : \text{pente.} \end{cases}$$

Rq : Il est + commode de fixer un temps finit  $t = \bar{t} \neq 1 \Rightarrow \min_{p_0, v_0} \sum_{i=1}^n D(\gamma(t_i), y_i)$

Il convient donc de minimiser  $E(S_0) = \sum_{i=1}^n f(S(t_i))$  pour  $\begin{cases} S(t) = F(S(t)) \\ S(t) = S_0 \end{cases}$

~~Rq : ça peut compliquer les choses~~ pour un temps de référance le choix arbitrairement.

Pour une variation  $\delta S$  de la condition initiale, on a :

$$dE \delta S_0 = \sum_{i=1}^n d_{S(t_i)} f \delta S(t_i) = \sum_{i=1}^n d_{S(t_i)} f \exp\left(\int_{t_0}^{t_i} d_{S(u)} F du\right) \delta S$$

$$\text{D'où } \nabla_{S_0} E = \sum_{i=1}^n \exp\left(\int_{t_0}^{t_i} d_{S(t)} F^T dt\right) \nabla_{S(t)} f.$$

5

Prop:  $\nabla_{S_0} E$  est la solution à  $t=t_0$  du système :

$$\begin{cases} \dot{v}(t) = -d_{S(t)} F^T r(t) \rightarrow \sum_{i=1}^N \nabla_{S(t_i)} f \delta(t-t_i) \\ v(t) = 0 \end{cases}$$

Dém: Par la méthode de variation de la constante, la solution a pour forme

$$v(t) = \exp\left(\int_t^T d_{S(u)} F^T du\right) h(t)$$

$$\text{et } \dot{v}(t) = -d_{S(t)} F^T v(t) + \exp\left(\int_t^T d_{S(u)} F^T du\right) h'(t)$$

$$\text{d'où } h'(t) = -\sum_{i=1}^N \exp\left(-\int_t^{t_i} d_{S(u)} F^T du\right) \nabla_{S(t_i)} f \cdot \delta(t-t_i)$$

$$\text{et } h(t) = \sum_{i=1}^N \exp\left(-\int_{t_i}^T d_{S(u)} F^T du\right) \nabla_{S(t_i)} f \cdot \mathbb{1}_{\{t \leq t_i\}} + h_0$$

$$\text{et finalement } v(t) = \sum_{i=1}^N \exp\left(\int_t^{t_i} d_{S(u)} F^T du\right) \nabla_{S(t_i)} f \mathbb{1}_{\{t \leq t_i\}}$$

ce qui donne le résultat annoncé pour  $t=t_0$ .

Si  $T \geq t_i$   
pour toute  
 $i$

Dém: Notebook Jupyter pour la régression géodésique et sphère.

(6)

Pour une variante myométrique, on a :

$$\min_{S_0, X_0} \sum_i f(X(t_i))$$

$\left\{ \begin{array}{l} S(t) = f(S(t_i)) \\ X(t) = g(S(t), X(t_i)) \end{array} \right. \quad \begin{array}{l} S(0) = S_0 \\ X(0) = X_0 \end{array}$

moments initiaux  
 + pts de contrôle  
 initiaux

pour protéger  
 initial.

} → cas de meilleurs / p'ts  $X_i$  c'est-à-dire les points de tels que  
 → cas inverse : pts fins mais non à j'pas des initiales.

Donc, l'attention à l'origine  
 se concentre à la fois comme  
 pts de contrôle initiaux et finaux  
 initiaux.

$$\left\{ \begin{array}{l} \delta S(t) = d_{S(t)} F \delta S(t) \\ \delta S(0) = \delta S_0 \end{array} \right. \quad \delta S(t) = \exp \left( \int_0^t d_{S(u)} F du \right) \delta S_0$$

$$\left\{ \begin{array}{l} \delta X(t) = \partial_{S(t)} G \delta S(t) + \partial_{X(t)} G \delta X(t) \\ \delta X(0) = \delta X_0 \end{array} \right.$$

$$\delta X(t) = \int_0^t \exp \left( \int_u^t \partial_{X(s)} G ds \right) \partial_{S(u)} G \exp \left( \int_u^t d_{S(s)} F ds \right) \delta S du$$

avec  $(10 \text{ day})$   
~~avec  $(10 \text{ day})$~~

$$+ \exp \left( \int_0^t \partial_{X(u)} G du \right) \delta X_0.$$

$$\text{D'où } \nabla_{S_0} E = \sum_i \int_0^{t_i} \exp \left( \int_u^{t_i} d_{S(u)} F ds \right) \partial_{S(u)} G \exp \left( \int_u^{t_i} \partial_{X(s)} G ds \right) \nabla_{X(t_i)} f \mathbb{1}_{\{t \leq t_i\}} du$$

$$\nabla_{X_0} E = \sum_i \exp \left( \int_0^{t_i} \partial_{X(u)} G du \right) \nabla_{X(t_i)} f.$$

$$\Theta(t) = \sum_i \exp\left(\int_t^{t_i} \partial_{X(s)} G^i ds\right) \left( \sum_i \nabla_{X(s)} f \mathbb{1}_{\{t \leq t_i\}} \right)$$

Alors:  $\partial \Theta = \nabla_{X_0} E$  et au moment  $\xi(t) = \int_t^u \exp\left(\int_s^u \partial_{X(s)} F^i ds\right) \partial_{X(s)} G^i \partial u ds$

$$\text{on a } \xi(t) = \nabla_{S_0} E$$

$$\text{avec } \begin{cases} \dot{\xi}(t) = [\partial_{S(t)} G^i \partial u + \partial_{S(t)} F^i \xi(t)] \\ \xi(T) = 0 \end{cases}$$

par le même calcul que précédemment.

De plus  $\dot{\partial}(t) = 0$  pour  $T \geq t_i$  pour tout  $i$

$$\begin{aligned} \dot{\partial}(t) &= \sum_i -\partial_{X(t)} G^i \partial u \Leftrightarrow -\sum_i \exp\left(\int_t^{t_i} \partial_{X(s)} G^i ds\right) \nabla_{X(t)} \delta(t-t_i) \\ &= -\partial_{X(t)} G^i \partial u - \sum_i \nabla_{X(t)} f \delta(t-t_i) \end{aligned}$$

Demos:

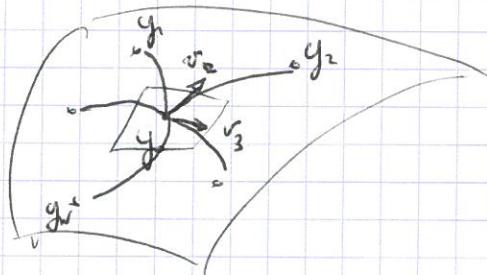
- profile de crêtes standing / Deformetrica / deformtrace - exemples - Crans - NonRidgeyf
- faisceau de fibres du article de Jans - GlobalRegime für PCCW32 / view results.m

## II - Construction de modèles anisotropiques.

(y)<sub>i=1..n</sub>: des observations d'individus / d'objets différents. Un ∈ des points d'une variété M.

But: Étendre à M des notions de moyen et variance.

↪ On résume la population par un point  $y_0$  et on relève les observations dans le plan tangent à  $y_0$  où on peut utiliser des opérations linéaires et définir des distributions.



Inconvénient: on résume en 1 point: ⇒ distribution multi-modale?

- relativement si plan tangent: modél adaptif si la dispersion des données est grande par rapport à la courbure.

### 1. Approche géométrique: le moyen de Fréchet.

- Dans  $\mathbb{R}^n$ , le moyenne empirique  $y_0$  d'un ensemble de points  $y_1, \dots, y_n$  peut être un cone:  $\min_{y_0} \sum_{i=1}^n \|y_i - y_0\|^2$ : de. le moyen minimal le plus courant.

- En remplaçant la norm par la distance géodésique, on obtient:

$$\min_{y_0} \sum_{i=1}^n d(y_0, y_i)^2 = \min_{y_0} \sum_{i=1}^n l_{y_0}^{y_i} \text{ tel que } y_0 = \text{Exp}_{y_0}(\omega_{0,i}).$$

Pour éviter un over-fitting (~~too much~~), on pénalise une variable absorbante:

$$E(y_0) = \sum_{i=1}^n l_{y_0}^{y_i} + \gamma D(\text{Exp}_{y_0}(\omega_{0,i}), y_i) \text{ où } \gamma \text{ est un multiplicateur de } \underline{\text{lambda}}$$

⑨.

- On a considérablement augmenté le nombre de variables à optimiser.
- Rq: Armande a prouvé l'existence de la moyenne de Fréchet  
En revanche, il n'y a pas en général unicité.

Optimisation: Ça n'est pas difficile car on fait calculer

$$\nabla E(S_{\alpha,i}, \xi_p) = \sum_i q_{\alpha,i}^T G(y_0)^{-1} q_{\alpha,i} + \gamma f(S_i(\alpha)) .$$

$$\left. \begin{aligned} \text{pour } S_{\alpha,i}(t) &= F(S_i(t)) \\ S_i(\alpha) &= S_{\alpha,i} = (y_0, q_{\alpha,i}) \\ &\quad q_i^T G(y_0)^{-1} q_{\alpha,i} \end{aligned} \right\}$$

$$\text{On sait que } \nabla_{S_{\alpha,i}} f(S_i(\alpha)) = v_i(\alpha) \quad \text{pour } \left. \begin{aligned} \dot{v}_i(t) &= -d_{S_i(t)} F^T v_i(t) \\ v_i(0) &= \nabla_{S_i(0)} f \end{aligned} \right\}$$

$$\text{Donc } \nabla_{q_{\alpha,i}} E = 2 G(y_0)^{-1} q_{\alpha,i} + \gamma v_i'(\alpha)$$

$$\nabla_{y_0} E = \sum_i \nabla_{y_0} (q_{\alpha,i}^T G(y_0)^{-1} q_{\alpha,i}) + \gamma v_i''(\alpha)$$

↳ on associe au contributif de chaque obs  $y_0$  son  $v_i$  partagé par tous les individus.

Cas d'une variable multivariée: cas des marchés -

$$E(S_{\alpha,i}, X_0) = \sum_i \alpha_{\alpha,i} K(c, c_i) \alpha_{\alpha,i} + \gamma f(X_i(\alpha))$$

$$\left. \begin{aligned} S_i(\alpha) &= f(S_i(\alpha)) \\ S_i(\alpha) &= S_{\alpha,i} = (x_0, \theta_i(\alpha)) \end{aligned} \right\} \left. \begin{aligned} \dot{x}_i(t) &= G(S_i(t), x_0, c_i) \\ x_i(0) &= x_0 \end{aligned} \right\}$$

$$\text{et } \nabla_{\alpha_{\alpha,i}} E = 2 K(c, c_i) \alpha_{\alpha,i} + \gamma \xi_i^X(\alpha)$$

$$\nabla_c E = \sum_i \nabla_c (K(c, c_i) \alpha_{\alpha,i}) + \gamma \xi_i^C(\alpha) \quad \text{avec } \left. \begin{aligned} \dot{\theta}_i(t) &= -\partial_{X_i(t)} G^T \theta_i(t) \\ \theta_i(0) &= \nabla_{X_i(0)} f \end{aligned} \right\}$$

$$\nabla_{x_0} E = \sum_i \theta_i(\alpha)$$

$$\left. \begin{aligned} \dot{\xi}_i^X(t) &= -(\partial_{S_i(t)} G^T \theta_i(t) + d_{S_i(t)} F^T \xi_i^X(t)) \\ \xi_i^X(0) &= \dots \end{aligned} \right\}$$

## 2. Discussion:

- Rep de variable à estimer
- Importance capitale de  $y_i$  dans our-fitting - relativement très différent  
pour  $y \neq$ : classe avec points inversés?
- Comment il estimer?
- Manque une approche statistique.

DERO : ANALYSE DE FEUILLES  
(+ pondérée ?)

## 2. Approche statistique:

Stanley / newey / Stanley / project Nodules Differ / leaves MA n.

On peut écrire le modèle binomialique suivant: Le stat pminmax = 3;

$$y_i = \text{Exp}(\beta_0 + v_{0,i}) + \varepsilon_i \quad y_i: \text{effet de pop.} \quad v_{0,i}: \text{effet individuel}$$

où on écrit:  $\ln \bar{Y}_0$  M est la normale où la matrice de covariance est donnée par la matrice  $G(y_0)^T$ .

$$p(v_{0,i}) \propto \exp\left(-\frac{1}{2} v_{0,i}^T G(y_0) v_{0,i}\right)$$

$$= \exp\left(-\frac{1}{2} q_{0,i} G(y_0)^T q_{0,i}\right).$$

et pour le bruit  $p(\varepsilon) \propto \exp\left(-\frac{D(\varepsilon)}{2\sigma_\varepsilon^2}\right)$  pour  $D(x,y) = D(x-y)$   
+ q.  $\int \exp\left(-\frac{D(x)}{\sigma_\varepsilon^2}\right) dx < \infty$

Dans ce cas on a:

$$p(\{y_i, v_{0,i}\}) = \prod_i p(y_i | v_{0,i}) p(v_{0,i})$$

$$\text{et } -\log p(\{y_i, v_{0,i}\}) = \sum_i^N \frac{1}{2\sigma_\varepsilon^2} D(y_i - \text{Exp}(v_{0,i}),) + \frac{1}{2} v_{0,i}^T G(y_0) v_{0,i}$$

C qui est le même critère que le moyenne de Fréchet.

Mais l'approche stat appelle de nouveaux outils.

où pmin min  $\bar{Y}_0^2$  (Et donc  $\hat{v}_{0,i}$ )  
apprendre une matrice de cov = co-matrice.

→ Postes itératifs.

dans un modèle hiérarchique, les variables latentes ne sont pas observées.

$$\text{on optimise plutôt: } p(y_{1:N}, y_0) = \prod_{i=1}^N p(y_i | \sigma_{y_i}) p(\sigma_{y_i}) d\sigma_{y_i}$$

où les paramètres  $\theta$  sont  $(\Gamma, \sigma_\varepsilon^2, y_0)$   
mais le  
modèle de  
corrélation  
des moments

→ l'intégration régularise l'estimation des  $\sigma_{y_i}$ .

- Mais rend l'algorithme + compliqué.

→ Il a été montré (Desilliers) que ~~l'algorithme~~

les log-likelihoods sont  $\neq$  au  $n+1$  ème.  $E(\min_{y_0} d(\text{Exp}_{y_0}, x)) > 0$

- Méthode PEMA : Une méthode générative pour estimer ces paramètres (avec preuve de convergence).

$$\text{Et: } \theta^{(k+1)} = \arg \max_{\theta} E \left( \log p(y, z; \theta) \right) = \max_{\theta} \int \log p(y, z; \theta) p(z|y, \theta^{(k)}) dz$$

$$\text{Ici: } \theta^{(k+1)} = \arg \max_{\theta} \sum_{i=1}^N \underbrace{\int \log(y_i | \sigma_{y_i}; \theta) p(\sigma_{y_i} | y_i, \theta^{(k)}) d\sigma_{y_i}}_{- \frac{1}{2\sigma_\varepsilon^2} D(\text{Exp}_{y_0}(\sigma_i), y_i) + \frac{1}{2} \sigma_i^\top \Gamma^{-1} \sigma_i}$$

On cherche famille exponentielle  $\log(y_i | \sigma_i; \theta) = \phi(\sigma_i)^\top S(y_i, \sigma) + C(\theta)$ .

ok pour  $\sigma_\varepsilon^2$  et  $\Gamma$  mais pas  $y_0$ .

on fait donc l'hypothèse suivante:  $y_0 \sim \text{exp}(\bar{y}_0, \bar{\sigma}_y^2 \text{Id})$

$$\text{d'où } \theta^{(k+1)} = \max_{\theta} \sum_{i=1}^N \underbrace{\int \log(y_i | \sigma_i, y_0; \theta) p(\sigma_i | y_i, \theta^{(k)}) d\sigma_i}_{- \frac{1}{2\sigma_\varepsilon^2} D(\text{Exp}_{y_0}(\sigma_i), y_i) - \frac{1}{2} \sigma_i^\top \Gamma^{-1} \sigma_i - \frac{1}{2} (\bar{y}_0 - \bar{y}_0)^\top \bar{y}_0 / \sigma_y^2} d\sigma_i$$

$$- \frac{1}{2} \log \sigma_\varepsilon^2 - \frac{1}{2} \log |\Gamma| - \frac{1}{2} \log \sigma_y^2$$

$$- \frac{1}{2} \bar{y}_0^\top \bar{y}_0 + \frac{1}{2} \bar{y}_0^\top \Gamma^{-1} \bar{y}_0$$

(12)

$$S_1(y_0, \sigma_i) = D(E_{\theta} p_{y_0}(\sigma_i), y_i)$$

$$S_2(y_0, \sigma_i) = \sigma_i^2 \bar{v}_i^2$$

$$S_3(y_0, \sigma_i) = |y_0|^2$$

$$S_4(y_0, \sigma_i) = y_0$$

$$\Phi_1(\bar{y}_0, \sigma_i, \Gamma) = -\frac{1}{2\sigma_i^2}$$

$$\Phi_2(\theta) = -\frac{1}{2} \Gamma^2$$

$$\Phi_3(\theta) = -\frac{1}{2\sigma_i^2}$$

$$\Phi_4(\theta) = -2 \frac{\partial \bar{y}_0}{\partial \theta^2}$$

$$C(\theta) = -\frac{1}{2} \log \sigma_e^2 - \frac{1}{2} \log \sigma_y^2 - \frac{1}{2} \log |\Gamma| \leq -\frac{1}{2} \frac{|\bar{y}_0|^2}{\sigma_y^2}$$

\* Pour  $\bar{S}_{\theta}^{(k)}(y_0, \sigma_i) = \sum_{i=1}^N E(S_p(y_0, \sigma_i) | y_i; \theta^{(k)})$

On a donc :  $\theta^{(k+1)} = \max_{\theta} \left\{ -\frac{\bar{S}_{1,k}}{2\sigma_e^2} - \frac{N}{2} \log \sigma_e^2 - \frac{1}{2} \langle \bar{S}_{2,k}, \Gamma^{-1} \rangle_F - \frac{N}{2} \log |\Gamma| \right.$   

$$\left. - \frac{1}{2} \frac{\bar{S}_{3,k}}{\sigma_y^2} - \frac{\bar{S}_{4,k}^T \bar{y}_0}{\sigma_y^2} - \frac{N}{2} \frac{|\bar{y}_0|^2}{\sigma_y^2} \right\} - \frac{N}{2} \log \sigma_y^2$$

~~et donc~~

$$(\bar{\sigma}_e^2)^{(k+1)} = \frac{1}{N} \bar{S}_{1,k}$$

$$\Gamma^{(k+1)} = \frac{1}{N} \bar{S}_{2,k}$$

$$\bar{y}_0^{(k+1)} = \frac{1}{N} \bar{S}_{4,k}$$

$$(\bar{\sigma}_y^2)^{(k+1)} = \frac{1}{N} (\bar{S}_{3,k} + \bar{S}_{4,k}^T \bar{y}_0 + N |\bar{y}_0|^2)$$

En général, on ne met pas à jour  
ce paramètre de variance

\* Comment calculer  $E(S_p(y_0, \sigma_i) | y_i; \theta^{(k)}) = \int S_p(y_0, \sigma_i) p(y_0, \sigma_i | y_i; \theta^{(k)}) dy d\sigma$

Comme c'est un facteur  
multiplicatif pris.

1/ On échantillonne  $p(y_0, \sigma_i)$  selon la loi conditionnelle

2/ on fait une somme de Monte-Carlo.

Dans 1/ on réalise une chaîne de Markov dont le loi ergodique est la loi conditionnelle

par ex: Metropolis-Hastings :  $y_0^{(p)} \sim \text{law prop. } q$  et  $\frac{p(y_0^{(p)}, \sigma_i^{(p)}; \theta^{(k)})}{q(y_0^{(p)}, \sigma_i^{(p)}; \theta^{(k)})}$

$$\frac{q(y_0^{(p+1)})}{q(y_0^{(p)})} / \frac{q(y_0^{(p)})}{q(y_0^{(p+1)})}$$

accepte avec proba min (1, u).

Pour 2/1, on réalise ~~une étape~~: l'approximation stochastique.

$$Q_{p_{k+1}}(c) = (1-\delta_h) Q_k(c) + \delta_h S_p(y_0^{(k)}, s_c^{(k)}) \quad \text{avec} \quad \sum \delta_h = +\infty \\ \sum \delta_h^2 < \infty$$

On Monte qui va faire 1 seul étape de simulation

- 1 \_\_\_\_\_ de l'approxim. stoch.
- 1 maximisation des probabilités

} un converge asymptotiquement

Déroulement, chose de l'échantillonnage. (MH within Gibbs par ex.)

- Phase de burn-in ( $\delta_h = 1$ )
- Choix de l'algorithme.
- Choisir le loi de prop (a priori, random walk, symétric random walk) etc..
- Mais permet de visiter le support de la distribution
- Majorité (de l'intégrale)  $\rightarrow$  stabilisation -
- Pas de calcul du gradient ( $+ \frac{\partial}{\partial x_1, \dots, \partial x_n} p(y_0, \dots, x_n)$ )

La machine fait l'échantillonage  $\rightarrow$  ~~échantillonage~~ loi de prop diffusion de Metropolis).

## Modèles longitudinaux:

- $y_{ij}$ :  $j^{\text{e}} \text{ observation du } i^{\text{e}} \text{ sujet.}$
- $t_{ij}$ : temps (ex-position) correspondant à cette observation.

But: Étudier les modèles linéaires à effets mixtes:  $y_{ij} = \alpha t_{ij} + b_i + \varepsilon_{ij}$

avec  $\alpha \sim N(\bar{\alpha}, \sigma_\alpha^2)$

$b_i \sim N(\bar{b}_i, \sigma_b^2)$

$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$

pour
 

- donner la variété
- de manière à prendre en compte l'estimation des décalages temporels / paramétrisations en temps.

→ tout en gardant l'identifiabilité du modèle.

### 1. Le modèle:

On se sent:  $y_{ij} = \bar{\alpha} t_{ij} + \bar{b} + \alpha_i t_{ij} + b_i + \varepsilon_{ij}$  avec  $\alpha \sim N(0, \sigma_\alpha^2)$  b  $\sim N(0, \sigma_b^2)$

$$= \gamma_i(t_{ij}) \quad \text{où } \gamma_i(t) = \alpha t + b_i \\ = T_i(\gamma_0)(t)$$

$$\text{ou } \gamma_0(t) = \bar{\alpha}t + \bar{b} \quad T_i: \gamma_{\bar{\alpha}, \bar{b}} \rightarrow \gamma_{\alpha_i, b_i}$$

transfert d'une const.

$$\hookrightarrow y_{ij} = T_i(\gamma_0)(t_{ij}) + \varepsilon_{ij}$$

Si on ajoute les u-paramétrisations simples:  $y_{ij} = T_i(\gamma_0)(\varphi_i(t_{ij})) + \varepsilon_{ij}$

$\varphi_i: R \rightarrow R$  C<sup>1</sup> monotone croissante.

~~param.~~

Où veut: identifiabilité du modèle: besoin de redéfinir  $T_i$ .

$$\text{dim avec } \varphi_i(t) = \alpha t + \beta: \text{on a } y_{ij} = \alpha_i (\alpha t_{ij} + \beta) + b_i + \varepsilon_{ij}$$

$$= \alpha_i \alpha t_{ij} + \alpha_i \beta + b_i + \varepsilon_{ij}$$

4 paramètres pour 2 seulement identifiés:  $\alpha_i \alpha$  et  $\alpha_i \beta + b_i$ .

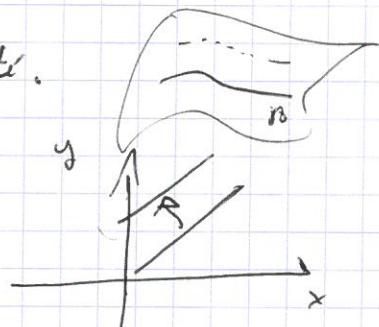
(13)

- Compatibilité du time-warp:  $T(\gamma_0) \circ \varphi = T(\gamma_0 \circ \varphi)$  pour ne pas avoir à chercher

- intrinsèque:  $T$  ne dépend pas de la paramétrisation de  $\gamma_0$ .

Pour cette raison, posons que  $\gamma_0$  soit d'un modèle linéaire:

- $\gamma_0$  est une géodésique
- $T$  est une "translation" de  $\gamma_0$  sur la variété.



### L'Exp-parallélisation:

$M$  une variété géodésiquement complète.

Def:  $\gamma: I \subset \mathbb{R} \rightarrow M$  une courbe différentiable sur  $M$ .

$$t_0 \in I \quad w \in \overline{T}_{\gamma(t_0)} M$$

L'Exp-parallèle à  $\gamma$  dans la direction  $w$  est:  $\gamma_{\gamma, w}(t) = \text{Exp}_{\gamma(t_0)}(P_{\gamma(t_0), t}(w))$

où  $P_{\gamma(t_0), t}(w)$  est le champ de vecteur parallèle à  $\gamma$ :  $\cdot \nabla_{\dot{\gamma}(t)} P_{\gamma(t_0), t}(w) = 0 \quad \forall t \in I$

$$\cdot P_{\gamma(t_0), t_0}(w) = w.$$

Def:  $*\gamma: I \rightarrow M$  une courbe différentiable sur  $M$ .

Donc  $p_0 = \gamma(0) \in M$ ,  $\overline{T}_{p_0}^w(\gamma_0)$  est une translation de  $\gamma_0$  ssi il définit une courbe différentiable sur  $M$  passant par tout  $t \in I$ .

\*  $\overline{T}_{p_0}^w$  est intrinsèque ssi:  $\overline{T}_{p_0}^w(\gamma_0) = \overline{T}_{\gamma_0(t_0)}^{P_{\gamma_0(t_0), t}(w)}(\gamma_0)$  pour  $p_0'$  un point sur  $\gamma_0$ .

\*  $\overline{T}_{p_0}^w$  est compatible avec time-warp ssi:  $\overline{T}_{p_0}^w(\gamma_0 \circ \varphi) = \overline{T}_{p_0}^w(\gamma_0) \circ \varphi$ .

Propriété: l'Exp-parallélisation est une translation intrinsèque et compatible avec le time-warp.

Dém: . translation car M est localement complet et différentiable par composition de fonctions différentiables.

- intrinsèque: car  $\mathcal{I}_{y_0, t_0, t} \circ \mathcal{I}_{y_0, t_0, t_0'} = \mathcal{I}_{y_0, t_0, t}$

Solution de l'équation différentielle

- compatible avec time-warp:  $T_p^w(y_0) \circ \psi(t) = y_{\psi(t), w}(t)$   $\Rightarrow \text{Exp}_{p \circ \psi(t)}(\mathcal{I}_{y_0, t_0, \psi(t)})$

ou  $\mathcal{I}_{y_0, t_0, \psi(t)} = \mathcal{I}_{y_0 \circ \psi, \psi(t), t}$  d'où  $= T_p^w(y_0 \circ \psi)(t)$ .

$\int \text{Int } w(t) = \mathcal{I}_{y_0, t_0, t}(w_0)$  on  $\frac{dw^k}{dt} = - \sum_{ij} \Gamma_{ij}^k w^i(t) \dot{y}^j(t) \Rightarrow f(w(t)) = \frac{dw^k}{dt}$

Et si  $\tilde{w}(t) = \mathcal{I}_{y_0 \circ \psi, \psi(t), t}(w_0)$  vérifie:  $\frac{d\tilde{w}^k}{dt} = - \sum_{ij} \Gamma_{ij}^k \tilde{w}^i(t) \dot{\psi}(t) \dot{y}^j(t)$

avec  $\tilde{w}(t \circ \psi(t)) = w_0 = w(t)$   $\Rightarrow \boxed{\tilde{w} = w \circ \psi}$   $\tilde{w}(t) = \psi(t) f(w(t), y(t))$

Prop: Si  $y_0(t) = \text{Exp}_p((t-t_0)\omega_0)$  géodésique passant par p au temps  $t_0$  avec la vitesse  $v_0$

Alors  $y_0 \circ \psi(t) = \text{Exp}_{\psi(t_0)}((t-t_0)\omega_0)$  avec  $t' = \psi(t)$  le geo. passant par  $\psi(t)$

par  $\psi(t) = \cancel{(t-t_0)} + t_0$  au temps  $t'$  avec vitesse  $v_0$

$$\alpha(t-t_0) + t_0 \quad (\psi(t_0) = t_0)$$

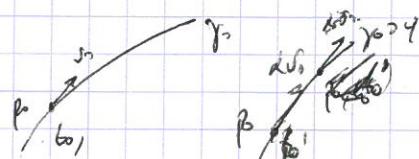
Dém: je veux  $\nabla_{\dot{y}_0} \dot{y}_0 = 0 \Rightarrow \ddot{y}_0^k(t) - \sum_{ij} \Gamma_{ij}^k \dot{y}_i^i(t) \dot{y}_j^j(t) = 0$

donc  $y_0 \circ \psi$  vérifie:  $\nabla_{\dot{y}_0} \dot{y}_0 = \frac{d^2}{dt^2} \psi(t) \nabla_{\dot{y}_0} \dot{y}_0 = 0$  c'est donc une géodésique.

Elle vérifie:  $y_0(\psi(t_0)) = y_0(t_0) = p$   $\xrightarrow{\text{Il faut que les su. indépendantes!}}$   $y_0 \circ \psi(t) = \text{Exp}_{\psi(t_0)}((t-t_0)\omega_0)$

$$\frac{d}{dt} y_0 \circ \psi \Big|_{t=t_0} = \dot{\psi}(t_0) \dot{y}_0(t_0) = \omega_0.$$

Dans  $y_0 \circ \psi = \text{Exp}_{\psi(t_0)}((t-t_0)\omega_0)$ .



[C'est jamais qu'une version étendue (affine et pas seulement linéaire) de la propriété d'homogénéité des géodésiques] -

• Si  $w_0 = \alpha v_0$  alors  $y_{\gamma_0, w_0}(t) = \underbrace{\text{Exp}_{\gamma_0(t)}(\alpha \underbrace{\mathcal{P}_{v_0, t}(v_0)}_{g(t)})}_{y_0(t)} = y_0(t)$  par homogénéité des fonctions.

Dans la composante de  $w_0$  sur  $v_0$  apparaît une exponentielle temporelle.

Pour assurer l'identifiabilité du modèle, on va supposer  $w_0 \perp v_0$ .  $g(v_0, w_0) = 0$ .

[Pas si clair car on n'a pas ce que vaut  $\text{Exp}(\mathcal{P}(w_0^{-1}) + \mathcal{P}(w_0))$ ]

Modèle à effets mixtes (variate simple).

$$y_{ij} = \underbrace{\gamma_{\gamma_0, w_{0,i}}(y_0(t_{ij}))}_{y_0} + \varepsilon_{ij}$$

$$\text{avec } y_0 = \text{Exp}_{\gamma_0}((t-t_0)v_0) \quad \gamma_0 \in \Gamma \quad v_0 \in \overline{\Gamma}_0(n)$$

$$g_{\gamma_0}(v_0, w_0) = 0 \quad w_{0,i} = \sum_k A_{ki} s_{ik} \quad g_f(A_k, w_0) = 0$$

$$y_i(t) = \alpha_i(t-t_0) + \varepsilon_i \quad (\text{le sujet } i \text{ passe en } y_0 \text{ à } t_0 \text{ avec initial } \alpha_i)$$

$$\alpha_i = e^{x_i}$$

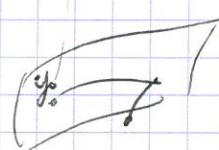
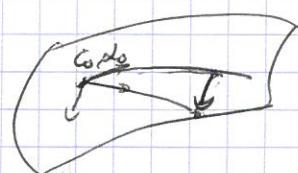
$$x_i \sim \mathcal{N}(0, \sigma_x^2) \quad \varepsilon_i \sim \mathcal{N}(0, \sigma_\varepsilon^2) \quad s_{ik} \sim \mathcal{N}(0, \sigma_s^2) \quad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma_{\varepsilon, ij}^2)$$

$$\text{et } p(x_i) \propto \exp\left(-\frac{|x_i|}{\sigma_x}\right)$$

↗ Invariance de la loi de  $w_i$  en temps !

Modèle à effets mixtes (variate moyennée par sujet):

$$y_{ij} = \underbrace{\gamma_{\gamma_0, w_{0,i}}(y_i(t_{ij})) * y_0}_{\text{en les pris de contre}} + \varepsilon_{ij}$$



• En venant de vérifier que  $\gamma_{\gamma_0, w_{0,i}}$  soit dans  $\mathcal{G}_{Y_i}$

•  $\text{Exp}_{\gamma_0}(E_{t_0,t}(w_0)) * y_0$  se calcule comme

$$\text{Exp}_{C(t)}(E_{t_0,t}(w_0)) * (\text{Exp}_{C_0}(t_{02}) * y_0)$$

## Exemple:

- 1D: ps de space-shift:  $y_{ij} = \gamma_0(\varphi_0(t_{ij})) + \varepsilon_j$

vante plate:  $y_{ij} = p_0 + v_0 (\alpha_0(t - t_0 - z_0)) + \varepsilon_j$        $\gamma_0(t) = p_0 + v_0(t - t_0)$

$$= p_0 + \alpha_0 v_0 (t - t_0 - z_0) + \varepsilon_j.$$

à comparer avec  $y_{ij} = (\alpha_0 + v_i)(t - t_0) + p_0 + \beta_i$

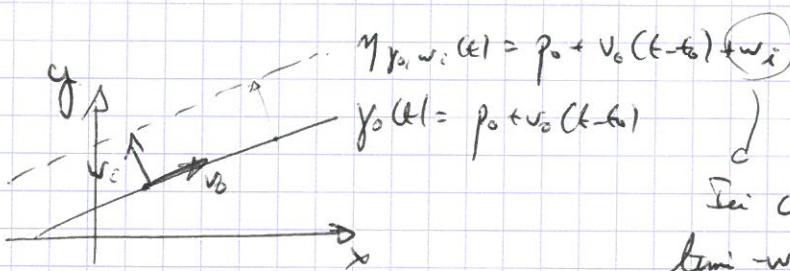
↳ Illustration ⊕ qd mêm modél non linéai

vante courb:  $g(t) = J_{0,1,2}$      $g(p) = \frac{1}{p^2(1-p)^2}$      $\gamma(t) = \frac{1}{1 + \frac{1-p_0}{p_0} \exp(-\frac{v_0(t-t_0)}{p_0(1-p_0)})}$

↳ Illustration (IPMI - ADAS-CO, gizi)

## Vante produit.

plate: 2D



Si on complète avec le  
term-way car  $w_i \perp v_0$ .

$$y_{ij} = p_0(w_i) + \alpha_0 v_0(t_{ij} - t_0 - z_0) + \varepsilon_j$$

$\left. \begin{matrix} 2 \text{ points} \\ 1 p_0 \end{matrix} \right\} 3 \text{ points} \quad \text{à part + addition à l'ajust.}$

↳ on a enlevé un degrés de liberté pour revenir à un modèle où  $w_i$  fournit l'effet aléatoire associé à  $p_0$ .

produit:  $M \rightsquigarrow M^N = M_1 M_2 \dots \rightarrow M$ .

$$g_{(p_1, \dots, p_N)}((x_1, u_1, \dots, x_N), (w_1, \dots, w_N)) = \sum_k g_{p_k}(x_k, w_k).$$

Et  $y_0$  gradéfin de  $M^N$  et de la form  $(y_1(t), \dots, y_N(t))$  avec  $y_k$  gradéfin de  $M$ .

(19)

L'Exp. parallelisation s'écrit :

$$\gamma_{p,w}(t) = \left( \gamma_1(t + \frac{w_1}{\tilde{\gamma}_1(t)}), \dots, \gamma_N(t + \frac{w_N}{\tilde{\gamma}_N(t)}) \right)$$

Modèle de progression

on cherche  $\tilde{\gamma}_0$  sous la forme paramétrique suivante :

$$\tilde{\gamma}_0(t) = \left( \tilde{\gamma}_0(t), \tilde{\gamma}_0(t - \delta_2), \dots, \tilde{\gamma}_0(t - \delta_N) \right)$$

$$\gamma_{p,w}(t) = \left( \tilde{\gamma}_0(t + \frac{w_1}{\tilde{\gamma}_0(t)}), \tilde{\gamma}_0(t - \delta_2 + \frac{w_2}{\tilde{\gamma}_0(t - \delta_2)}), \dots, \tilde{\gamma}_0(t - \delta_N + \frac{w_N}{\tilde{\gamma}_0(t - \delta_N)}) \right)$$

$$\parallel \quad \tilde{\delta}_k = \delta_k + \frac{w_k}{\tilde{\gamma}_0(t_0 + \delta_k)}$$

↳ Illustration.

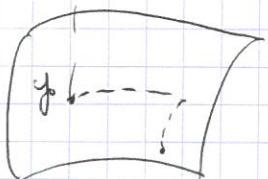
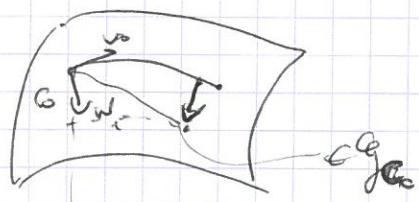
$$\text{Si } \tilde{\gamma}_0(t) = \frac{1}{1 + \frac{p_0(1-p_0)}{p_0} \exp(-\frac{v_0(t-t_0)}{p_0(1-p_0)})}$$

$$\text{Alors } \gamma_{ijk} = \frac{1}{1 + \frac{1-p_0}{p_0} \exp(-\frac{v_0(t_{ij}-t_0-\tau_i)}{p_0(1-p_0)} + \frac{w_k}{\tilde{\gamma}_0(t_0 + \tilde{\delta}_k)} / p_0(1-p_0))} + \epsilon_{ijk}$$

↳ Illustration.

Variétés mythomorphiques =

(9 bis)



$$\cancel{C_i(t_j) = \text{Exp}_G((\psi_i(t_{ij}) - t_0)v_0)} \quad \cancel{\text{for } j}$$

$$y_{ij} = \text{Exp}_{C_i(t_j)} \left( P_{c_i, t_j, \psi_i(t_j)} (w_i) \right) * y_0 + \varepsilon_{ij}.$$

paramètres:  $c, t_0, v_0$  et  $y_0$        $\sigma_c^2, \sigma_d^2, \sigma_\tau^2$       Ans

V.a :  $\alpha_i, \tau_i, w_i = \sum \alpha_k R_{ik}$

obtenir par intégration successive de 2 systèmes d'ODE.

↳ Besoin d'un schéma numérique pour le TP -

(on n'a pas l'équivalent des formulations variationnelles: Euler-Lagrange ou Hamilton).

## Transport parallèle : méthode numérique

- Pour le gradient ( $\nabla_{\gamma} \psi(t)$  le long de  $\gamma(t)$ ):  $\nabla_{\gamma} \psi =$   
 $\hookrightarrow$  l'approche variationnelle appelle des orbites numériques (Lagrange, Hamilton)
- Pour un TP quelconque, on n'a pas l'équivalent de  $\nabla_{\gamma} w = 0$ : comment intégrer  
 variationnellement? van ce que cela  
 est dérivable?

Champ de Jacob:  $J_{y(t)}^w(h) = \frac{d}{ds} \Big|_{s=0} \text{Exp}_{y(t)}(h(\dot{\gamma}(s) + \varepsilon w)) \in T_{\gamma(t+h)} M$ .

$$\text{On a } P_{0,t}(w) = \frac{J_{y(t)}^w(t)}{t} + O(t^2).$$

Def:  $\gamma_s(\cdot)$  une famille paramétrique de géodésiques différentiable en le paramètre

$$\text{on appelle champ de Jacob: } J(t) = \frac{d}{ds} \Big|_{s=0} \gamma_s(t)$$

Prop: Un champ de Jacob vérifie:  $J''(t) + R(J(t), \dot{\gamma}_s(t)) \dot{\gamma}_s(t) = 0$   
 avec  $R$  le tension de courbure.

Si on considère le champ de Jacob particulier:  $J(t) = \frac{d}{ds} \Big|_{s=0} \text{Exp}(t(\dot{\gamma}_0 + \varepsilon w))$

$$\text{qui vérifie donc: } J(0) = 0$$

$$J'(0) = w$$

$$J''(0) = -R(J(0), \dot{\gamma}_0) \dot{\gamma}_0 = 0.$$

D'où on voit que:  $J(t) = t w + o(t)$  mais  $J(t) \in T_{\gamma(t)} M$ .

Théorème: Soit  $V(t)$  un champ de vecteur le long de  $\gamma(t)$

on a  $\nabla_{\dot{\gamma}}^{(k)} V(t) = \frac{d}{dt} \Big|_{t=0}^k P_{0,t}^{-1}(V(t))$ .

Preuve: Soit  $E_i(t)$  un Basc de  $T_{\gamma(t)} M$ , étudier un autre Basc  $e_i(t) = P_{0,t}^{-1}(e_i(0))$  de  $T_{\gamma(t)} M$ .

D'où  $V(t) = \sum_i v_i(t) e_i(t)$

$$P_{0,t}^{-1}(V(t)) = \sum_i v_i(t) e_i(0)$$

$$\text{et } \frac{d}{dt} \Big|_{t=0}^k P_{0,t}^{-1}(V(t)) = \sum_i v_i^{(k)}(0) e_i(0)$$

$$\begin{aligned} \nabla_{\dot{\gamma}}^{(k)} V(t) &= \nabla_{\dot{\gamma}}^{(k)} \left( \sum_i v_i(t) e_i(t) \right) = \sum_i \nabla_{\dot{\gamma}}^{(k)} (v_i(t) e_i(t)) \quad \text{comme } \nabla_{\dot{\gamma}} e_i(t) = 0 \\ &= \sum_i v_i^{(k)}(t) e_i(t) \end{aligned}$$

D'où  $\nabla_{\dot{\gamma}}^{(k)} V(t) = \frac{d}{dt} \Big|_{t=0}^k P_{0,t}^{-1}(V(t))$

Finalement:  $P_{0,t}^{-1} J(t) = \underbrace{\nabla_{\dot{\gamma}}^{(0)} J(0)}_{=J(0)=0} + t \underbrace{\nabla_{\dot{\gamma}}^{(1)} J(0)}_w + t^2 \underbrace{\nabla_{\dot{\gamma}}^{(2)} J(0)}_0 + o(t^2)$

D'où  $P_{0,t}^{-1} J(t) = tw + o(t^2)$  et

$$\boxed{P_{0,t}^{-1}(w) = \frac{J(t)}{t} + o\left(\frac{t}{t}\right)}$$