

# Notes

## [11/2021] Concernant la preuve du théorème de plongement de Takens présentée dans [1]

- Voir le théorème de Kupka-Smale. Voici un énoncé approximatif.

### Theorem 1: Kupka-Smale

Soit  $M$  une variété compacte, et  $n$  un entier positif.

Pour  $\phi \in \text{Diff}^1(M)$  générique, le nombre de points périodiques de période inférieure ou égale à  $n$  est fini.

De plus, la propriété est toujours vraie si on suppose que les points périodiques sont *hyperboliques*. D'où, pour des  $\phi$  génériques, les points fixes ont des valeurs propres distinctes, et les points périodiques ont des valeurs propres distinctes lorsqu'ils sont vus comme des points fixes de  $\phi^k$ .

Un point fixe  $x$  de  $\phi$  est hyperbolique si  $D\phi_x$  n'a pas de valeurs propres de module unité. Un point périodique  $x$  de période  $k$  est hyperbolique si c'est un point fixe hyperbolique de  $\phi^k$ .

- Pourquoi un plongement est-il ici défini comme une application injective et immersive ? Ne doit-il pas aussi être un homéomorphisme dans son image munie de la topologie induite par celle de l'espace d'arrivée (et pas celle induite par l'application elle-même). Sinon, l'image ne serait-elle pas une sous-variété immergée et pas une sous-variété régulière ?  
Réponse : Si la variété de départ est compacte alors une application injective et immersive est automatiquement un homéomorphisme, et on a bien une sous-variété régulière.

## [12/2021] Concernant la preuve du théorème de plongement de Takens présentée dans [2]

- Voir la preuve du théorème de plongement de Whitney. En effet, ce résultat constitue la raison d'être du "2m + 1". Le théorème de plongement de Whitney nous dit qu'il existe en plongement entre une variété  $M$  de dimension  $m$  et  $\mathbb{R}^{2m+1}$ , et même de façon encore plus forte que les plongements de  $M$  dans  $\mathbb{R}^{2m+1}$  sont denses dans les fonctions lisses de  $M$  dans  $\mathbb{R}^{2m+1}$ .

### Theorem 2: Plongement de Whitney

$\text{Emb}(M, \mathbb{R}^{2m+1})$  est dense dans  $C^1(M, \mathbb{R}^{2m+1})$ .

Remarque : Il semble qu'avec très peu d'hypothèses supplémentaires (la compacité en fait-elle partie ?), il soit possible de montrer le même résultat mais avec  $2m$  au lieu de  $2m + 1$ . Cela se fait par un "argument de chirurgie" de Dehn. Certaines variétés très pathologiques (avec "beaucoup" de trous par exemple) ne pourront pas être plongées dans  $\mathbb{R}^{2m}$  (mais bien dans  $\mathbb{R}^{2m+1}$ ).

- Rappel : Une valeur critique est l'image d'un point critique, c'est à dire d'un point où  $f$  n'est pas submersive (sa matrice jacobienne n'y est pas surjective). Une valeur régulière est une valeur non critique (y compris si elle n'est l'image d'aucun point).

### Theorem 3: Densité de Sard (ou de Sard-Brown)

Soit  $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$  de classe  $C^r$ . Si  $r > \max(0, n - m)$ , alors l'ensemble des valeurs critiques est négligeable (pour la mesure de Lebesgue). Autrement dit, l'ensemble des valeurs régulières est dense dans  $\mathbb{R}^m$ .

- Voici une généralisation du théorème précédent.

### Theorem 4: Densité de Smale

Si  $f : N^n \rightarrow M^m$  est  $C^r$  avec  $r > \max(0, n - m)$  alors l'ensemble des valeurs régulières est comaigne (donc dense).

- Voir la preuve du résultat suivant.

### Lemma 1

$C^2(M, \mathbb{R})$  est dense dans  $C^1(M, \mathbb{R})$  (avec la topologie  $C^1$ ).  
De même,  $\text{Diff}^2(M)$  est dense dans  $\text{Diff}^1(M)$ .  
On en déduit que  $\text{Diff}^2(M) \times C^2(M, \mathbb{R})$  est dense dans  $\text{Diff}^1(M) \times C^1(M, \mathbb{R})$ .

La topologie  $C^1$  sur  $C^2(M, \mathbb{R})$  est simplement la topologie induite lorsque  $C^2(M, \mathbb{R})$  est vu comme un sous-ensemble de  $C^1(M, \mathbb{R})$  (de même pour  $\text{Diff}^2(M)$  et  $\text{Diff}^1(M) \times C^2(M, \mathbb{R})$ ).

- Voir le *principe de transversalité*. En particulier, voir le “multijet transversality theorem”. Commencer par la preuve du théorème de transversalité de Thom.
- Voir la preuve du résultat affirmant que les immersions (ou les plongements) forment un ensemble ouvert des fonctions lisses.  
Remarque : Ce n'est pas trivial. Bien sûr, le résultat dépend de la topologie. On souhaite montrer cela dans la topologie  $C^1$  qui est une topologie de Whitney. Il semblerait qu'il s'agissent de la même topologie que la topologie dite compacte-ouverte. La topologie compacte-ouverte est définie via une base dont les éléments sont les applications partant d'un compact et allant dans un ouvert. Cette topologie n'est pas métrisable (ce qui veut sûrement dire qu'il n'existe aucune distance dont les boules ouvertes engendrent cette topologie).
- Remarque : On travaillera toujours (ou presque) sur des variétés compactes. En effet, une variété quelconque (i.e., non compacte, on parle alors de variété ouverte) peut rapidement présenter des dégénérescences très délicates à traiter. Considérer par exemple la sphère privée d'un point (qui n'est pas compacte car pas fermée). Alors certains points ne sont reliés par aucune géodésique. En effet, les géodésiques (grands cercles) passant auparavant par le point qui a été retiré n'existent plus. Et toute autre courbe reliant deux tels points n'est pas minimisante (on peut toujours trouver une courbe plus proche du “trou” qui sera plus courte).

## [14/12/2021] Questions sur les distributions

- Distinction entre distribution intégrable (ou intégrale ?) et distribution complètement intégrable ?  
Réponse : Examinons tout d'abord la notion de distribution intégrable. Si la distribution est de dimension 1, elle est toujours intégrable. C'est en substance ce que nous dit le théorème de Cauchy pour l'existence d'une solution à un problème de Cauchy. En effet, étant donnée une condition initiale  $x_0 \in M$  et un champ de vecteurs lisse  $X$  sur  $M$ , on cherche une courbe  $\gamma$  telle que  $\gamma(0) = x_0$  et  $\dot{\gamma}(t) = X_{\gamma(t)}$  pour  $t$  variant dans un certain intervalle ouvert de  $\mathbb{R}$  contenant 0. Le théorème de Cauchy

nous dit qu'il existe un  $\epsilon > 0$  une courbe  $\gamma$  définie sur  $(-\epsilon, \epsilon)$ . Pour une distribution  $P$  de dimension  $> 1$ , on cherche une sous-variété  $N$  telle que pour tout  $x \in N$ ,  $T_x N = P_x$ .

Le théorème de Frobenius généralise le théorème de Cauchy en disant qu'une distribution est intégrable si et seulement si pour tout couple de champ de vecteurs tangents à la distribution, le crochet de Lie des deux champ de vecteurs est aussi tangent à la distribution. Pour comprendre le sens de ce résultat, il faut voir le crochet de Lie comme la dérivée de Lie, c'est à dire que  $[X, Y] = \mathcal{L}_X Y$  est la dérivée de  $Y$  par rapport à  $X$ .

Ce résultat est à mettre en parallèle avec la relation liant dérivées partielles et différentiabilité. Dans cette analogie, le théorème de Cauchy nous dit que si une fonction de la variable réelle admet une dérivée partielle alors elle est dérivable ce qui est toujours trivialement vrai. En revanche, pour des fonctions de  $\mathbb{R}^n \rightarrow \mathbb{R}$  pour  $n > 1$ , l'existence des dérivées partielles n'implique pas la différentiabilité. De même, l'existence (d'après Cauchy) de courbes intégrables dans toutes les directions d'une base de champs de vecteurs de notre distribution n'implique pas que la distribution soit intégrable. Des contre-exemples existent. Tout comme les contre-exemples de fonctions non différentiables mais admettant des dérivées partielles, il s'agit de trouver une distribution dont les éléments de contact varient "trop vite" par exemple proche de l'origine (sous-entendu, les éléments de contact ont des variations visibles à l'ordre 1 alors que pour que la distribution soit intégrable, il faut que les variations soient nulles à l'ordre 1 près du point considéré, ici l'origine, mais peuvent être non nulles à des ordres supérieures). Considérer par exemple [3] la distribution de dimension 2 dans  $\mathbb{R}^3$  engendrée par  $\partial_1$  et  $\partial_2 + x^1 \partial_3$ . Dans ce cas, l'élément de contact à l'origine est le plan horizontal (i.e., engendré par  $\partial_1$  et  $\partial_2$ ). Si on s'éloigne de l'origine en restant dans le plan orthogonal à  $x^1$ , l'élément de contact reste horizontal. Mais si on sort de ce plan, l'élément de contact va pivoter (dans un sens pour les  $x^1$  positifs et dans l'autre pour les  $x^1$  négatifs). Le pivotement sera le plus rapide si on s'éloigne le long de  $x^1$ .

Concernant la distinction entre intégrable et complètement intégrable, on peut reprendre l'analogie avec les équations différentielles. Le théorème de Cauchy est un résultat local, puisqu'il ne garantit l'existence que dans un intervalle  $(-\epsilon, \epsilon)$ . L'existence d'une courbe maximale, ou sur tout  $\mathbb{R}$ , ou à temps infini, nécessite d'autres arguments. De même, le théorème de Frobenius est un résultat local, et ne donne des conditions pour l'intégrabilité d'une distribution que localement (les champs de vecteurs  $X$  et  $Y$  peuvent n'être définis que localement). Pour que la distribution soit complètement intégrable (i.e., sur tout  $M$ ), il faut et il suffit de montrer que pour tout champs de vecteurs  $X$  et  $Y$  définis sur tout  $M$  et tangents à la distribution, alors  $[X, Y]$  est tangent à la distribution sur tout  $M$ .

- Lien entre distributions et sous-modules de champs de vecteurs tangents à la distribution ?  
Réponse : Il y a deux grands types de modules. Les modules libres sont très proches des espaces vectoriels, la seule différence étant que les scalaires forment un anneau au lieu d'un corps. A partir d'un module libre, on peut définir une structure moins forte appelée module projectif. Un module projectif est défini comme le quotient d'un module libre par un idéal. Cela revient à créer des "relations". Cela signifie qu'il existe des familles de vecteurs telles que  $\sum \alpha_i x_i = 0 \Rightarrow \alpha_i = 0$  mais la famille est quand même liée. Il y a apparemment un théorème dû à Whitney (encore lui) à propos des modules projectifs admettant une famille génératrice finie mais je ne saurais pas en dire plus ...

## [14/12/2021] Vers un théorème style Takens pour les distributions

- Soit  $P$  une distribution inconnue sur une variété compacte  $M$  de dimension  $m$ . On suppose  $P$  intégrable. A partir d'observations de  $M \rightarrow \mathbb{R}$ , on souhaite reconstruire  $P$ . Le théorème de Takens nous dit que si  $P$  est de dimension 1, il suffit d'observer au moins  $2m + 1$  points le long d'une courbe intégrale afin de pouvoir reconstruire  $P$ . Si  $P$  a une dimension strictement supérieure à 1, on sent bien qu'observer des retards le long d'une seule courbe intégrale ne sera pas suffisant. Il faudra aussi observer des courbes partant dans d'autres directions. Cela revient à "faire varier les conditions initiales".
- On va s'intéresser à un champ de vecteurs contrôlé. Il s'agit d'un champ de vecteur  $X$  défini sur la variété  $M \times U$  où  $U$  est un sous-ensemble d'un espace vectoriel, généralement un ouvert ou un compact. Dans un premier temps, on considérera  $U = [0, 1]$  un intervalle compact de  $\mathbb{R}$ . Pour  $p \in M$  et  $u \in U$ , on a  $X_{p,u}$  est un vecteur tangent de  $T_{p,u}(M \times U)$ . On peut aussi le voir comme un vecteur tangent de  $T_p M$ . Dans ce cas,  $X$  est une famille de champs de vecteurs de  $M$  paramétrée par  $U$ .

- Le champ de vecteurs est dit *contrôlé* car il s'agit d'une notion de *contrôle optimal*. Cela signifie que l'on peut agir sur  $u$ . Partant d'une condition initiale  $p_0$ , on choisit une commande  $u : I \subset \mathbb{R} \rightarrow U$  de telle sorte à ce que notre courbe intégrale  $\gamma$ , telle que  $\gamma(0) = p_0$  (condition initiale) et  $\dot{\gamma}(t) = X_{\gamma(t), u(t)}$ , minimise un certain critère. On choisit  $u$  de telle sorte à ce que toutes petites perturbations de  $u$  fasse augmenter le critère, il s'agit donc de trouver un minimum local.
- Dans notre cas, il ne s'agit pas d'un problème de contrôle mais d'un problème d'identification. On suppose donc que  $u$  est donnée, et on se demande à quelles conditions (sur  $u$  notamment) on va pouvoir reconstruire notre distribution inconnue  $P$  sur  $M$ , ou plutôt à quelles conditions on va pouvoir déterminer une dynamique conjuguée comme celle obtenue dans le théorème de Takens pour un seul champ de vecteurs.
- A minima,  $u$  sera mesurable bornée. Dans un premier temps, on pourra prendre des hypothèses très fortes sur  $u$  afin d'obtenir une première preuve, puis on essaiera de retirer des hypothèses au fur et à mesure. On supposera donc que  $u$  est  $C^\infty$  (mais pas analytique, car cette hypothèse est trop forte et risque de rendre le théorème indémontrable). Ensuite, on pourra très certainement passer à des fonctions  $C^1$  et peut-être  $C^0$ . En géométrie, le passage de  $C^\infty$  à  $C^1$  est souvent facile, alors que le passage de  $C^1$  à  $C^0$  est plus délicat. Une fois le  $C^0$  établi, le passage de  $C^0$  à mesurable est aisé (si le théorème reste vrai pour des fonctions mesurables bien sûr). Dans d'autres domaines (lesquels ?), c'est le passage de  $C^2$  à  $C^1$  qui est le plus difficile.

**[15/12/2021] Théorème de Takens pour les champs de vecteurs et dynamique conjuguée (en partie incorrect)**

Soit  $M$  une variété compacte de dimension  $m$  et  $X$  un champ de vecteurs lisse défini sur  $M$ . Ici, lisse signifie  $C^1$ . Soit  $y : M \rightarrow \mathbb{R}$  une fonction lisse.

Soit  $x_0 \in M$ ,  $U$  un voisinage de  $x_0$ , et  $\epsilon > 0$ . On appelle *flot local* de  $X$  au voisinage de  $x_0$ , une application  $\phi : (-\epsilon, \epsilon) \times U \rightarrow M$  telle que :

- $\forall x \in U, \phi_0(x) = x$
- $\forall x \in U, \forall t \in (-\epsilon, \epsilon), \dot{\phi}_t(x) = X_{\phi_t(x)}$
- $\forall x \in U, \forall s, t \in (-\epsilon, \epsilon), \phi_s(\phi_t(x)) = \phi_{s+t}(x)$

Pour tout  $t \in (-\epsilon, \epsilon)$ ,  $\phi_t$  est un difféomorphisme dans son image. Ainsi,  $\phi$  est une famille paramétrée de difféomorphismes (avec un seul paramètre). On dit que  $\phi$  est un *flot global* de  $X$  s'il est défini sur  $\mathbb{R} \times M$ .

Un champ de vecteur est dit *complet* s'il admet un flot global.

Sur une variété compacte sans bord, tous les champs de vecteurs lisses sont complets.

Le théorème de Takens nous dit que pour des  $X$  et  $y$  génériques, et/ou pour  $X$  fixé et  $t$  générique (notamment il faut  $t \neq 0$ ), l'application  $\Phi_t$  définie par  $\Phi_t : x \mapsto (y(x), y(\phi_t(x)), \dots, y(\phi_{2mt}(x)))$  est un plongement de  $M$  dans  $\mathbb{R}^{2m+1}$ . Plus généralement, on pourrait prendre des pas de temps non réguliers (ils n'ont même pas besoin d'être croissants) : pour tout  $T = (t_0, \dots, t_{2m}) \in \mathbb{R}^{2m+1}$  tel que toutes les composantes soient distinctes, l'application  $\Phi_T : x \mapsto (y(\phi_{t_0}(x)), y(\phi_{t_1}(x)), \dots, y(\phi_{t_{2m}}(x)))$  est un plongement.

**Question :** *Que signifie que  $X$  soit générique, pour quelle topologie, faut-il réadapter la preuve du théorème de Takens ? A-t-on un plongement pour tout  $t \neq 0$  ?*

On souhaiterait maintenant éclaircir la notion de dynamique conjuguée.

On définit  $\pi_i : \mathbb{R}^{2m+1} \rightarrow \mathbb{R}^{2m}$  par  $(r_0, \dots, r_{i-1}, r_i, r_{i+1}, \dots, r_{2m}) \mapsto (r_0, \dots, r_{i-1}, r_{i+1}, \dots, r_{2m})$ .

Soit  $t \in \mathbb{R}$ . Le plongement  $\Phi_t$  induit une sous-variété  $\Phi_t(M)$  de  $\mathbb{R}^{2m+1}$  de dimension  $m$ . Ainsi, si  $r \in \Phi_t(M)$  alors il existe un unique  $r' \in \Phi_t(M)$  tel que  $\pi_{2m}(r') = \pi_0(r)$ . Comme  $\Phi_t(M)$  est de dimension  $m$ , la restriction  $\pi_{i,t} : \Phi_t(M) \rightarrow \pi_i(\Phi_t(M))$  est un difféomorphisme. Ainsi, on a  $r' = (\pi_{2m,t}^{-1} \circ \pi_{0,t})(r)$ . On pose  $\psi_t = \pi_{2m,t}^{-1} \circ \pi_{0,t}$ .

Ainsi,  $\psi$  est un flot local d'un champ de vecteurs  $Y$  défini sur un certain ouvert de  $\mathbb{R}^{2m+1}$  de telle sorte que, pour tout  $t$  tel que  $\Phi_t$  soit un plongement et pour tout  $r$  dans un certain ouvert de  $\mathbb{R}^{2m+1}$ , on ait  $Y_{\psi_t(r)} = \dot{\psi}_t(r)$ . Le champ de vecteurs  $Y$  est donc une *dynamique (topologiquement) conjuguée* de  $X$ . En effet, on a pour tout  $t$ ,  $\psi_t \circ \Phi_t = \Phi_t \circ \phi_t$ .

**Question :** *Est-ce la bonne définition d'une dynamique conjuguée (sûrement pas). Peut-on démontrer toutes ces affirmations ? Sur quel ouvert  $Y$  est-il défini ? Et sur quel autre ouvert  $\psi$  est-il défini ?*

En pratique, supposons que l'on dispose d'une série de mesures  $r \in \mathbb{R}^{2m}$  de notre système dynamique inconnu. On suppose que les mesures sont régulièrement espacées d'un temps  $t$ . On souhaiterait prédire la mesure suivante. On utilise ce que je suppose être la *méthode des retards*. Pour cela, il faut d'abord estimer notre sous-variété  $\Phi_t(M)$ . On suppose que l'on connaît la dimension  $m$  de  $M$  (si ce n'est pas le cas, il faut à minima estimer une borne supérieure de  $m$ ). On échantillonne un grand nombre de conditions initiales  $x_i \in M$  de telle sorte à bien recouvrir  $M$ . A partir de chacun des  $x_i$ , on laisse évoluer le système et on effectue une mesure à chaque intervalle de temps  $t$ . Une fois qu'on a obtenu  $2m+1$  mesures, on dispose d'un point de  $\mathbb{R}^{2m+1}$  qui n'est autre que  $\Phi_t(x_i)$ . Si on a suffisamment de  $x_i$ , on pourra sans doute interpoler d'une façon ou d'une autre la sous-variété  $\Phi_t(M)$  (ou de façon équivalente interpoler notre plongement  $\Phi_t$ ). Soit  $r$  notre série de  $2m$  mesures dont on souhaiterait prédire la mesure suivante (il s'agit en fait de notre  $\pi_{0,t}(r)$ ). On cherche l'unique point de notre sous-variété interpolée dont les  $2m$  premières composantes sont égales à  $r$  (ou on cherche le point de notre sous-variété dont les  $2m$  premières composantes sont les plus proches de  $r$  en utilisant la distance euclidienne (?)). La  $2m+1$ -ième composante de ce point est notre prédiction.

**Question :** *Est-ce bien cela que l'on appelle la méthode des retards ? Est-ce bien la bonne façon d'utiliser le théorème de Takens pour faire de la prédiction ?*

Maintenant, on aimerait réécrire la relation de conjugaison entre  $X$  et  $Y$  en une relation de conjugaison entre des feuilletages (ou entre des distributions). Soit  $N_{x_0} = \{x \in M : \exists t \in \mathbb{R}, x = \phi_t(x_0)\}$  l'image de la courbe maximale passant par  $x_0$ . Il s'agit d'une sous-variété de dimension 1. Les  $N_{x_0}$  forment une partition de  $M$ . Il s'agit d'un feuilletage  $F$  associé à une distribution  $P$  de dimension 1. Le champ de vecteurs  $Y$  engendre une distribution  $Q$  (et donc un feuilletage  $G$ ) sur  $\mathbb{R}^{2m+1}$ .  $F$  et  $G$  sont conjugués. En passant des champs de vecteurs aux distributions, on perd l'information "dynamique", à savoir la dépendance en  $t$ . Cela signifie que, quel que soit  $t$ ,  $\Phi_t$  envoie les feuilles de  $F$  sur les feuilles de  $G$ . On peut écrire :  $G = \Phi(F)$  où  $\Phi$  désigne la famille des plongements  $\Phi_t$ .

#### [16/12/2021] Sur l'identifiabilité des distributions intégrables (*incorrect*)

Soit  $M$  une variété compacte de dimension  $m$ .

Soit  $P$  une distribution de  $M$  de dimension  $p$ , supposée  $C^\infty$  et intégrable. On note  $q = m - p$ .

Soit  $y : M \rightarrow \mathbb{R}$  supposée  $C^\infty$ .

Soit  $X$  un champ de vecteurs contrôlé tangent à  $P$ . Soit  $\phi$  le flot de  $X$ .

$X$  est un champ de vecteurs de  $M \times [0, 1]$  tangent à  $P$ . On a donc :

- $\forall x \in M, \forall u \in [0, 1], \phi_0(x, u) = (x, u)$ .
- $\forall x \in M, \forall u \in [0, 1], \forall t \in \mathbb{R}, \dot{\phi}_t(x, u) = X_{\phi_t(x, u)}$ .
- $\forall x \in M, \forall u \in [0, 1], \forall s, t \in \mathbb{R}, \phi_{s+t}(x, u) = \phi_s(\phi_t(x, u))$ .

Soit une fonction  $u : \mathbb{R} \rightarrow [0, 1], t \mapsto u(t)$  supposée  $C^\infty$ .

Sous quelles conditions l'application définie par  $\Phi_t : x \mapsto (y(x), y(\phi_{t,u(t)}(x)), \dots, y(\phi_{Nt,u(Nt)}(x)))$  est-elle un plongement ? On cherche des conditions sur  $u(t)$  et sur  $N$ .

#### [17/12/2021] Théorème de Takens pour les distributions et généralisation

Soit  $M$  une variété  $C^\infty$  compacte de dimension  $m$ .

Soit  $X$  un champ de vecteur  $C^\infty$  défini sur  $M$ . Afin d'éviter les singularités, on suppose que  $X$  ne s'annule pas sur  $M$ .

Soit  $h : M \rightarrow \mathbb{R}$  une fonction  $C^\infty$  appelée *fonction de mesure*.

Soit  $0 \leq t_1 < \dots < t_{2m} < +\infty$  des instants de mesures.

Comme  $M$  est compacte,  $X$  admet un flot global noté  $\phi$ . Pour comprendre cela, il faut voir les raisons qui peuvent empêcher un champ de vecteurs d'admettre un flot global. Si  $X$  n'est pas défini en certains points de  $M$ , alors les courbes intégrales qui devraient passer par ces points ne peuvent pas être globales. Le flot de  $X$  peut aussi exploser (en temps infini ou en temps fini), mais cela nécessite que  $X$  soit défini sur un ouvert. Le théorème de Cauchy nous dit que  $X$  admet un flot local au voisinage de tout point. Si  $M$  est compacte, on peut recouvrir  $M$  par un nombre fini de tels voisinages. Il sera donc possible de recoller les flots locaux

pour obtenir un flot global.  
Le flot  $\phi$  vérifie :

- $\forall x \in M, \phi_0(x) = x$
- $\forall x \in M, \forall s \in \mathbb{R}, \dot{\phi}_s(x) = (\phi(x))_* (\frac{d}{dt}|_s) = X_{\phi_t(x)}$
- $\forall x \in M, \forall s, t \in \mathbb{R}, \phi_s(\phi_t(x)) = \phi_{s+t}(x)$  (propriété de pseudo-groupe)

D'après le théorème de Takens, l'application  $\Phi : M \rightarrow \mathbb{R}^{2m+1}$  définie par  $x \mapsto (h(x), h(\phi_{t_1}(x)), \dots, h(\phi_{t_{2m}}(x)))$  est un plongement de  $M$  dans  $\mathbb{R}^{2m+1}$ . Autrement dit,  $M$  et la sous-variété  $\Phi(M)$  de dimension  $m$  dans  $\mathbb{R}^{2m+1}$  sont difféomorphes. Ainsi, sur  $\mathbb{R}^{2m+1}$ , on peut définir une dynamique  $\psi$  “différentiellement” conjuguée à  $\phi$  : pour tout  $t$ ,  $\psi_t \circ \Phi = \Phi \circ \phi_t$ .  $\psi$  est le flot d'un champ de vecteurs  $Y$  défini sur la sous-variété  $\Phi(M)$ . Le flot  $\phi$  engendre un feuilletage  $F$  de  $M$  associée à la distribution  $P$  de dimension 1 engendrée par  $X$ . Sur  $\Phi(M)$ , nous avons également un feuilletage  $G$  engendré par  $\psi$ .

### Lemma 2: Premier Jour

Le plongement  $\Phi$  envoie les feuilles de  $F$  sur les feuilles de  $G$ .  
Autrement dit,  $\forall x \in M, G_{\Phi(x)} = \Phi(F_x)$ .  
En termes de feuilletages, on notera  $G = \Phi(F)$ .

### Proof: Premier Jour

Prenons  $x_0 \in M$  et notons  $F_{x_0}$  la feuille contenant  $x_0$ . Soit  $x \in F_{x_0}$ . Il existe un  $t_x$  tel que  $x = \phi_{t_x}(x_0)$ . On a la relation  $\psi_{t_x} \circ \Phi = \Phi \circ \phi_{t_x}$ . En appliquant la relation au point  $x_0$ , on obtient  $\psi_{t_x}(\Phi(x_0)) = \Phi(\phi_{t_x}(x_0)) = \Phi(x)$ . En notant  $y_0 = \Phi(x_0)$  et  $y = \Phi(x)$ , on voit que  $y = \psi_{t_x}(y_0)$  donc  $y \in G_{y_0}$ . Ainsi  $G_{y_0} = \Phi(F_{x_0})$ . Plus encore, on voit que deux points d'une même feuille séparés par un temps donné ont leurs images séparées par le même temps sur la feuille image. Cette conjugaison “temporelle” ne sera pas généralisée.  $\square$

### Lemma 3: Deuxième Jour

Pour tout  $x \in M$  et pour tout  $t \in \mathbb{R}$ , on a  $\Phi_* X_{\phi_t(x)} = Y_{\psi_t(\Phi(x))}$ .  
Autrement dit,  $\Phi_* \dot{\phi}_t(x) = \dot{\psi}_t(\Phi(x))$ .  
Cela implique le résultat suivant, plus faible, sur les éléments de contact :  $\forall x \in M, \Phi_* P_x = Q_{\Phi(x)}$ .  
En termes de distributions, on notera  $Q = \Phi_* P$ .

### Proof: Deuxième Jour

$$\begin{aligned} \Phi_* \dot{\phi}_t(x) &= \Phi_*(\phi_*(x)_*(\frac{d}{dr}|_t)), \\ &= \Phi_{*,\phi_t(x)} \circ \phi_*(x)_*(\frac{d}{dr}|_t), \\ &= (\Phi \circ \phi_*(x))_* (\frac{d}{dr}|_t) \text{ (dérivation en chaîne)}. \end{aligned}$$

Soit  $g \in C^\infty(\mathbb{R}^{2m+1}, \mathbb{R})$ .

$$\begin{aligned}
(\Phi \circ \phi.(x))_* \left( \frac{d}{dr} \Big|_t g \right) &= \frac{d}{dr} \Big|_t (g \circ \Phi \circ \phi.(x)), \\
&= \frac{d}{dr} (g \circ \Phi(\phi_t(x))), \\
&= \frac{d}{dr} (g \circ \psi_t(\Phi(x))) \text{ (relation de conjugaison)}, \\
&= \psi.(\Phi(x))_{*,t} \frac{d}{dr} \Big|_t g, \\
&= \dot{\psi}_t(\Phi(x))g.
\end{aligned}$$

□

Plaçons-nous maintenant dans le cas général d'une distribution  $P$  de dimension  $p$  sur  $M$ . On suppose que  $P$  est intégrable, c'est à dire qu'il existe un feuilletage  $F$  sur  $M$  tel qu'en tout point de  $x \in M$ , l'espace tangent en  $x$  de la feuille  $F_x$  contenant  $x$  n'est autre que l'élément de contact  $P_x$ . On souhaite trouver des conditions sur le nombre de retards  $N$  et sur le nombre et la direction des champs de vecteurs nécessaires pour que la distribution "conjuguée" sur  $\mathbb{R}^N$  soit intégrable.

#### [12/01/2022] Encore une reformulation de notre généralisation

Let  $P$  be a (smooth regular) distribution of dimension  $p$  ( $1 \leq p \leq m$ ) on  $M$ .

Assume that  $P$  is integrable.

Let  $X$  be a controlled vector field tangent to  $P$ . We see  $X$  as a vector field of  $M \times [0, 1]$  whose flow  $\phi$  verifies:

- $\forall x \in M, \forall u \in [0, 1], \phi_0(x, u) = (x, u)$ .
- $\forall x \in M, \forall u \in [0, 1], \forall t \in \mathbb{R}, \dot{\phi}_t(x, u) = X_{\phi_t(x, u)} \in P_{(\pi \circ \phi_t)(x, u)}$ .
- $\forall x \in M, \forall u \in [0, 1], \forall s, t \in \mathbb{R}, \phi_{s+t}(x, u) = \phi_s(\phi_t(x, u))$ .

Let  $u \in C^\infty(\mathbb{R}, [0, 1])$  and  $T > 0$ .

Let  $f_{y, \phi, u} : x \mapsto (y(x), y(\phi_T(x, u(T))), \dots, y(\phi_{NT}(x, u(NT))))$ .

Denote  $f = f_{y, \phi, u}$  and define the distribution  $Q$  on  $f(M) \subset \mathbb{R}^N$  by  $Q_{f(x)} = f_*(P_x)$ .

By Frobenius theorem:  $Q$  is integrable if and only if for every  $Y$  and  $Z$  tangent to  $Q$  we have  $[Y, Z]$  is also tangent to  $Q$ .

We are looking for conditions on  $u$  and  $N$  such that, for generic  $(\phi, y)$ , the distribution  $Q$  is integrable.

#### [27/01/2022] Des précisions sur l'étude géométrique des réseaux de neurones récurrents

L'objectif est de conduire une analyse de sensibilité des réseaux de neurones récurrents.

Un réseau de neurones récurrent est une fonction  $f(e, \theta, \mathcal{W})$  avec  $e \in \mathbb{R}^N$  l'entrée du réseau,  $\theta \in \mathbb{R}^K$  l'état du réseau, et  $\mathcal{W}$  les poids du réseau. Dans la suite, on suppose que les poids  $\mathcal{W}$  sont fixés donc on aura  $f(e, \theta, \mathcal{W}) = f(e, \theta)$ .

A l'itération  $k$ , le réseau produit une sortie  $s_k = f(e_k, \theta_k) \in \mathbb{R}^M$ . On définit alors la relation de récurrence  $\theta_{k+1} = \phi(s_k) = \phi(f(e_k, \theta_k))$ . La fonction  $\phi$  sert deux objectifs :

- Supprimer les composantes de la sortie  $s_k$  n'intervenant pas dans l'état  $\theta_k$  (par exemple si le réseau produit une prédiction qui n'est pas utilisée dans l'état). Sans perte de généralité, on supposera dans la suite que toutes les composantes de la sortie interviennent dans l'état.
- Construire  $\theta_{k+1}$  en concaténant plusieurs sorties successives:  $\theta_{k+1} = (s_k^T \ s_{k-1}^T \ \dots \ s_{k-p}^T)^T$ . Ainsi  $\theta_k$  et  $\theta_{k+1}$  partagent les  $p \times M$  composantes  $s_{k-1}, \dots, s_{k-p}$ . On a  $K = (p+1) \times M$ . Afin d'alléger les notations, on supposera dans la suite que  $\theta_{k+1} = f(e_k, \theta_k)$ . Cela revient à dire qu'une fois la sortie  $s_k$  produite, le réseau va lui-même la concaténer avec les entrées précédentes afin de construire  $\theta_{k+1}$ .

Si le réseau est parfaitement entraîné, on a  $\theta_k = (\psi(X_{k-1}) \ \psi(X_{k-2}) \ \dots \ \psi(X_{k-p-1}))$  où  $X$  est un système dynamique vivant sur une variété  $\mathcal{M}$  supposée non observable et  $\psi$  est une fonction de mesure. Dans le cas d'un avion, on a  $\mathcal{M} = \mathbb{R}^{12}$  puisque que le modèle dynamique complet de l'avion comporte trois positions, trois angles, et leurs dérivées temporelles. Par exemple,  $\psi$  pourra être la mesure de l'altitude (ou de la masse ?). Si le réseau est parfaitement entraîné, le théorème de Takens nous dit que le réseau est capable de prédire  $s_k = \psi(X_k)$  à partir de  $\theta_k$ , à condition que  $p \geq 2\dim(\mathcal{M})$ . Dans ce cas, à quoi sert l'entrée externe  $e_k$  ? Je suppose qu'elle sert à corriger le réseau (puisque en pratique, celui-ci ne sera jamais parfait), ainsi qu'à introduire les  $p+1$  premières mesures  $\psi(X_k)$ .

On munit l'espace de sortie de la métrique de l'information de Fisher  $g$  (en supposant que le réseau définit une loi gaussienne en chaque point de l'espace de sortie). L'espace d'entrée est alors munit de la métrique pullback  $f^*g$ .

Notre objectif est de construire petit à petit une analyse du cas général en réalisant une succession d'étapes simples. On décrit ici le cas le plus simple via les hypothèses suivantes :

- $f$  est purement récurrent, c'est à dire qu'il ne dépend pas d'une entrée externe  $e$ . Dans ce cas, l'espace d'entrée est égal à l'espace de sortie.
- $\theta \in \mathbb{R}$  i.e.,  $K = 1$  et  $p = 0$ . Dans ce cas, on a  $\theta = s$  et on travaille en fait directement sur  $s$ .
- $f$  est linéaire.
- On se place dans le cas continu i.e.,  $\theta$  obéit à l'équation différentielle  $\dot{\theta}(t) = f(\theta(t))$  plutôt qu'à l'équation aux différences  $\theta_{k+1} - \theta_k = f(\theta_k) - \theta_k$ .

Afin de pouvoir étudier la sensibilité du réseau (aux conditions initiales ?), on suppose que  $\theta$  est en fait un processus stochastique obéissant à l'équation différentielle stochastique suivante :

$$d\theta(t, \omega) = f(\theta(t, \omega))dt + dw(t, \omega)$$

où  $w$  est un processus de Wiener (i.e., mouvement Brownien) défini comme suit :

- $w$  est à accroissements indépendants i.e., pour tout  $t_1 < t_2 < t_3 < t_4$ ,  $w(t_2) - w(t_1)$  est indépendant de  $w(t_4) - w(t_3)$ .
- Pour tout  $n$  et pour tout  $t_1, \dots, t_n$ ,  $(w(t_1) \ \dots \ w(t_n))$  est un vecteur gaussien.

Sous ces conditions, il y a un résultat d'unicité pour  $w$  (à voir).

On peut montrer que  $\theta(t) \sim \mathcal{N}(\mu(t), \sigma(t))$ . L'objectif de cette première étape est d'étudier les propriétés métriques des trajectoires de  $\theta$  dans l'espace  $(\mu, \sigma)$  en fonction des conditions initiales  $\mu(0), \sigma(0)$ . Comme  $\theta$  suit une loi normale, l'espace  $(\mu, \sigma)$  est un demi-plan de Poincaré (hyperbolique). Plus précisément, il s'agit de déterminer :

- La longueur des trajectoires  $\int_0^t |\dot{\theta}(\tau)| d\tau$  en fonction de  $t$ .
- La courbure en tout point de la trajectoire.
- Le comportement asymptotique des trajectoires.  
 $f$  étant linéaire, les trajectoires peuvent être décrites explicitement. On s'attend à avoir deux types de comportements en fonction des valeurs propres de  $f$ . Si  $f$  est contractante (valeur propre négative ?, de module inférieur à 1 ?), les trajectoires vont exponentiellement converger vers une valeur limite  $\sigma_*$  égale à l'écart-type de  $w$ . Cela revient à dire que les bruits précédents sont atténués et que seul le bruit courant intervient dans l'écart-type de  $\theta$ . Si  $f$  est expansive, les trajectoires vont diverger exponentiellement selon  $\sigma$ . Il est alors impossible de prédire  $\theta$  car son écart-type diverge.

Dans ce cas, l'espace d'entrée est de même dimension que l'espace de sortie (puisque les deux espaces sont égaux). La métrique pullback n'est donc pas dégénérée et aucun feuilletage n'apparaît dans l'espace d'entrée.

Dans l'étape suivante, on relaxe l'hypothèse  $K = 1$  et on s'intéresse au cas  $\theta \in \mathbb{R}^K$  pour  $K$  et  $p$  quelconques. Dans ce cas, les trajectoires évoluent dans un produit de demi-plans de Poincaré. Il est de nouveau possible d'exprimer explicitement les trajectoires de  $\theta$ . La dimension de l'espace d'entrée est strictement supérieure à celle de l'espace de sortie, la métrique pullback est donc dégénérée (c'est une pseudométrique) ce qui permet de définir plusieurs feuilletages :



- Un feuilletage dont les feuilles sont les noyaux de la métrique. Cela signifie qu'en se déplaçant le long de ces feuilles dans l'espace d'entrée, on reste immobile dans l'espace de sortie.
- Un feuilletage transverse (normal) aux feuilles précédentes. Un déplacement le long de ces feuilles correspond à un déplacement maximal dans l'espace de sortie.

Notre objectif est alors d'étudier l'interaction entre ces feuilletages et les trajectoires de  $\theta$ .

#### [27/01/2022] Petite remarque concernant le lien entre les connexions duales et les probabilités

L'intérêt de la géométrie de l'information repose sur un résultat pour les familles exponentielles (e.g., gaussiennes, gammas, bêtas, exponentielles, de Poisson ...). Une famille exponentielle est une famille de distributions de probabilités dont les densités par rapport à une mesure (généralement la mesure de Lebesgue mais pas que, puisqu'avec la mesure de comptage, on obtient des lois discrètes) sont de la forme :

$$\exp(\langle x, \theta \rangle - \psi(\theta))$$

où  $\psi$  est une fonction de normalisation permettant d'avoir une probabilité égale à 1 lorsqu'on intègre sur le domaine de  $x$ . Les paramètres  $\theta$  sont appelés "paramètres canoniques". L'espace des  $\theta$  noté  $\Theta$  peut être muni de la métrique de l'information de Fisher  $g$  et d'une connexion euclidienne (plate)  $\nabla_m$  qu'on appelle connexion de mélange car  $\langle x, \theta \rangle$  est un mélange dans le vocabulaire probabiliste. Le triplet  $(\Theta, g, \nabla_m)$  est appelé un espace de jauge.

On note  $X$  la variable aléatoire de paramètre  $\theta$  et on considère  $\eta(\theta) = \partial_\theta \mathbb{E}[X]$ . Un résultat fondamental de la géométrie de l'information nous dit que  $\eta$  est un difféomorphisme. On obtient alors de nouvelles coordonnées appelées "paramètres naturels". Par exemple, si la famille est celle des gaussiennes unidimensionnelles, on a  $\eta = (\mu, \sigma^2)$ . Si on considère  $\nabla_e$  la connexion euclidienne sur les  $\eta$  alors  $\nabla_m$  et  $\nabla_e$  sont des connexions conjuguées ou duales.  $\nabla_e$  est appelé la connexion d'espérance car elle est obtenue par un calcul d'espérance. La théorie des connexions duales nous dit que la connexion de Levi-Civita s'obtient par  $\nabla^{LC} = (\nabla_m + \nabla_e)/2$ .

Note à soi-même : il faut regarder les filtres particuliers.

#### [04/02/2022] Objectifs de l'étude géométrique d'une équation différentielle stochastique linéaire

On considère l'équation différentielle stochastique (EDS) scalaire, linéaire, autonome, et à bruit additif suivante :

$$\begin{aligned} d\theta(t) &= (a.\theta(t) + b)dt + c.dW(t), \\ \theta(0) &= \theta_0 \text{ p.s.} \end{aligned}$$

où  $\theta : \mathbb{R}^+ \times (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  est un processus stochastique,  $W$  est un processus de Wiener standard, et  $a, b, c, \theta_0 \in \mathbb{R}$ . Soit  ${}_N g$  le champ de tenseurs métrique d'information de Fisher associé à la famille des lois normales unidimensionnelles paramétrées par leur espérance  $\mu$  et leur variance  $\sigma^2$ .

Objectifs de compréhension :

1. Définition, construction, et propriétés du processus de Wiener standard.
2. Montrer que la loi de  $\theta(t)|\theta_0$  est normale pour tout  $t$ , i.e.,  $\theta(t)|\theta_0 \sim \mathcal{N}(\mu(t), \sigma^2(t))$ .
3. Montrer que  $(\mathbb{R} \times \mathbb{R}_*^+, {}_N g)$  est un demi-plan de Poincaré.

Objectifs d'exercice :

4. Expressions de  $\mu(t)$  et  $\sigma^2(t)$ .
5. Calculer la longueur des trajectoires  $L(\theta_0, t_0, t) = \int_{t_0}^t |\dot{\theta}(s)| ds$ .
6. Calculer la courbure des trajectoires en tout point.
7. Décrire le comportement asymptotique des trajectoires.

## [08/02/2022] Précisions et prochaines étapes

Il faut étudier les propriétés asymptotiques de la longueur et de la courbure des trajectoires.

Pour pouvoir appliquer la géométrie de l'information aux réseaux de neurones récurrents, il faut introduire de l'aléatoire. Dans cette optique, il y a trois points de vue différents :

1. Le réseau définit une équation différentielle déterministe. La condition initiale est une variable aléatoire. En sortie, on obtient un processus stochastique. On peut alors étudier comment les moments de ce processus évoluent au cours du temps. Dans ce cas, l'espace de sortie est muni de la métrique de Fisher. Il n'y a pas à proprement parler d'espace d'entrée (la sortie est directement réinjectée dans le réseau).
2. Le réseau définit une équation différentielle stochastique. La condition initiale est une variable aléatoire. De plus, à chaque instant  $t$ , le réseau prend une variable aléatoire en entrée dont la loi est connue, par exemple un processus de Wiener (qui modélise un bruit purement aléatoire). On étudie alors la trajectoires du processus de sortie. Les espaces d'entrée et de sortie sont identiques et tout deux munis de la métrique de Fisher.
3. C'est le cas le plus complet et le plus cohérent. Le réseau définit une équation différentielle stochastique. La condition initiale est une variable aléatoire. A chaque instant  $t$ , le réseau prend une variable aléatoire en entrée dont la loi est inconnue. On observe les trajectoires du processus de sortie dans l'espace de sortie. L'espace de sortie est muni de la métrique de Fisher. L'espace d'entrée est muni de la métrique pullback.

On pourra étudier le cas scalaire et linéaire avec le point de vue 1. Dans un second temps, on étudiera le cas multidimensionnel (linéaire) selon les points de vue 1 et 2. Il ne sera peut-être plus possible de calculer explicitement les longueurs et la courbure. Dans le cas général (non-linéaire), ce sera de toute façon impossible. Le but sera alors d'encadrer la longueur et la courbure du processus de sortie par des processus linéaires connus. Cela se fait généralement en supposant que le réseau est Lipschitz. Sous cette hypothèse, on peut encadrer la sortie du réseau par deux processus linéaires dont les 'pentes' sont  $\pm$  une constante de Lipschitz. C'est une méthode classique dans l'étude des réseaux de neurones (cf. Nicolas Couellan). Dans le cadre des équations différentielles déterministes, le lemme de Grönwall permet d'obtenir des encadrements similaires.

Toutes ces études ne prennent pas en compte la corrélation entre les instants successifs du processus d'entrée. Par la suite, il s'agira d'étudier les trajectoires d'un réseau prenant en entrée  $p$  instants successifs, dans l'esprit du théorème de Takens. Les trajectoires évolueront alors dans un produit d'espaces hyperboliques.

## [22/02/2022] Un point et quelques remarques plus générales

- Dans le cas du processus de Ornstein-Uhlenbeck, il est possible que la courbure diverge vers l'infini alors que la trajectoire converge vers un point. C'est par exemple le cas si la trajectoire s'enroule autour d'un point en formant une spirale, ou simplement un "début de spirale". De plus, dans le demi-plan de Poincaré, les coordonnées sont euclidiennes alors que l'espace est hyperbolique, ce qui rend ce genre de phénomène d'autant plus difficile à voir. Pour avoir une autre représentation, on peut se placer dans le disque de Poincaré. Si on voit le demi-plan de Poincaré comme le demi-plan complexe, la transformation  $z \mapsto \frac{z-i}{z+i}$  envoie le demi-plan de Poincaré sur le disque de Poincaré. Cette transformation est appelée "projection gnomonique". Dans le cas du globe terrestre, elle consiste à projeter depuis le centre de la Terre sur un plan tangent à un pôle. Dans ce cas, les grands cercles deviennent des droites (car les grands cercles sont l'intersection de la sphère avec un plan passant par le centre de la Terre et deviennent donc l'intersection entre ce même plan et le plan de projection, donc des droites). Cette projection est cependant non-conforme. Une représentation encore meilleure serait celle du disque de Klein, dans lequel toutes les géodésiques sont des droites.
- Il faut trouver des bornes, notamment des bornes inférieures à la courbure. En effet, la courbure ne peut pas être inférieure à la courbure de l'espace ( $-1$  dans le cas d'un espace hyperbolique de rayon 1), car les géodésiques sont les courbes de courbure minimale égale à la courbure de l'espace.

- Il faudra se pencher sur les équations de Jacobi. Elles permettent d'accéder à la courbure dans certains cas, ou d'accéder à un encadrement de celle-ci. Les équations de Jacobi permettent d'étudier comment les géodésiques varient lorsqu'on change légèrement les conditions initiales.
- Dans un second temps, on s'intéressera à la loi jointe de points successifs, et on étudiera la trajectoire de cette loi jointe au cours du temps, dans un esprit théorème de Takens. L'idée est ensuite d'étudier la sensibilité des trajectoires aux conditions initiales. Plus précisément, on peut regarder comment la courbure change quand la moyenne initiale change, i.e., on regarde la dérivée de la courbure par rapport à la moyenne initiale. L'idée générale de notre étude est de transformer un problème statistique en un problème géométrique. L'intérêt d'étudier la courbure (ou un encadrement de la courbure) est de pouvoir décrire des comportements asymptotiques sans connaître de formes closes pour les trajectoires. Intuitivement, la courbure décrit à quel point la courbe s'éloigne d'une géodésique. Par exemple, si la courbure tend vers 0, la courbe doit tendre vers une géodésique. En particulier, dans le cas hyperbolique, on s'attendrait alors à ce que la courbe tende vers une droite verticale (qui est une géodésique), donc une moyenne constante et un écart-type qui diverge.
- Concernant l'interprétation à donner à l'information de Fisher, elle correspond au premier terme non nul (d'ordre deux) dans le développement de la divergence de Kullback-Leibler (KL) aussi appelée entropie relative. La divergence KL entre  $p$  et  $q$  s'interprète comme l'information supplémentaire nécessaire, partant de la connaissance complète de  $q$ , pour connaître entièrement  $p$ . Quel lien entre l'entropie relative et l'entropie conditionnelle ? Il existe une autre métrique entre distributions, à savoir la distance de Wasserstein définie comme  $W(p, q) = \min_{\pi} \int \pi(x, y) D(x, y) dx dy$  où le min est pris sur l'ensemble des mesures de probabilités sous les contraintes  $\int_y \pi(x, y) dy = p(x)$  et  $\int_x \pi(x, y) dx = q(y)$ . La distance  $D$  pourra typiquement être la distance euclidienne. Ainsi, cela revient à chercher une loi jointe pour  $p$  et  $q$  telle que la distance moyenne entre un échantillon de  $p$  et un échantillon de  $q$  tirés selon cette loi jointe soit minimale (selon la distance  $D$ ). Il s'agit d'une approche de type transport optimal. La distance de Wasserstein souffre de deux défauts qui font qu'on lui préfère la divergence KL. Tout d'abord, dans le cas de la distance entre deux lois gaussiennes, la distance de Wasserstein se réduit à la distance euclidienne entre les paramètres des lois (si  $D$  est la distance euclidienne). Ce comportement n'est pas souhaitable car, par exemple, si  $p$  et  $q$  convergent vers deux diracs disjoints, la distance va converger vers une valeur finie, alors qu'on s'attendrait plutôt à ce qu'elle diverge vers l'infini (ce qui est le cas avec la divergence KL). Le second défaut est que le min étant pris sur l'ensemble des mesures de probabilités, la loi jointe correspondant à la distance peut être en dehors de la famille paramétrée que l'on avait choisie. Dans le cas de la distance entre gaussiennes, la loi jointe optimale sera aussi gaussienne, mais ce n'est pas vrai en général.
- Existe-il des mesures de probabilité qui ne sont ni discrètes, ni absolument continues par rapport à la mesure de Lebesgue, ni un mélange des deux ? La réponse est oui. Construisons l'escalier du diable. Pour cela, on part de l'ensemble de Cantor. Il s'agit de l'ensemble des réels de l'intervalle  $[0, 1]$  dont l'écriture en base 3 ne contient que des 0 et des 2 (pas de 1). Cet ensemble est de mesure de Lebesgue nulle, pourtant il est en bijection avec  $[0, 1]$ . Pour cela, il suffit d'associer à chaque réel de l'ensemble de Cantor, le réel de  $[0, 1]$  s'écrivant en base 2 de la même manière que notre réel de départ en base 3 mais avec les 2 remplacés par des 1. L'escalier de diable n'est autre que la fonction de répartition d'une mesure de probabilité croissant linéairement sur l'ensemble de Cantor et constante ailleurs. On peut montrer que cette mesure ne charge aucun singleton (donc elle n'est pas discrète) et n'est pas absolument continue par rapport à la mesure de Lebesgue. On dit qu'elle est étrangère à la mesure de Lebesgue. En effet, d'après le théorème de décomposition de Lebesgue, étant donné une mesure  $\nu$  (par exemple la mesure de Lebesgue), on peut décomposer toute mesure  $\mu$  de la façon suivante  $\mu = \mu_c + \mu_s$  avec  $\mu_c \ll \nu$  et  $\mu_s = \mu_d + \mu_e$  se décompose à son tour en une partie discrète (et étrangère) et une partie étrangère (non discrète).  $\mu_e$  est étrangère par rapport à  $\nu$  signifie qu'il existe un  $D$  tel que  $\mu_s(D) = 0$  et  $\nu(D^C) = 0$ . Un théorème montre que la partie étrangère  $\mu_e$  prend toujours une forme similaire à l'escalier du diable. En ce sens, l'escalier du diable est déjà le cas général (et pas un simple exemple). L'escalier du diable est aussi un contre-exemple du théorème fondamental de l'analyse. En effet, l'escalier du diable est une fonction continue, intégrable, et presque partout dérivable de dérivée nulle. Cependant il n'est égale à l'intégrale de sa dérivée (par contre il est toujours vrai que la dérivée

de l'intégrale d'une fonction mesurable est la fonction elle-même).

- Question géométrique : étant donnée une métrique riemannienne, est-il toujours possible que cette métrique soit la dérivée seconde d'une fonction. C'est toujours vrai localement. Pour que ce soit vrai globalement, il faut que la métrique soit "hessienne", et il faut pour cela que la variété soit affine, c'est à dire qu'il doit exister un atlas tel que les fonctions de changement de coordonnées soient affines (de sorte que leur dérivée seconde s'annulent et ne viennent pas s'ajouter à la métrique). Ma question était surtout de savoir à quelles conditions la fonction en question avait les propriétés d'une distance ou d'une divergence ... Une autre question intéressante est la suivante : étant donnée une métrique riemannienne, est-il toujours possible de voir cette métrique comme l'information de Fisher d'une certaine famille de distributions de probabilité. Il y a des cas où il a été montré que c'était impossible. Néanmoins, il n'y a pas à l'heure actuelle de résultat donnant des conditions nécessaires et suffisantes pour qu'une métrique soit une information de Fisher. Une condition suffisante est d'avoir un *recouvrement universel*. Un recouvrement universel est un fibré de notre variété de départ qui est simplement connexe (tout lacet peut se ramener au lacet trivial) et qui est de type  $\mathbb{R}^n$ . Par exemple, le recouvrement universel du cercle est l'axe des réels  $\mathbb{R}$  qu'on aurait "enroulé" de façon hélicoïdale, ou autrement dit le cercle n'est autre que  $\mathbb{R}/2\pi\mathbb{Z}$ . Autre exemple, le tore n'est autre que  $\mathbb{R}^2/\mathbb{Z}^2$ . Le recouvrement universel du tore est donc  $\mathbb{R}^2$ . Si on dispose du recouvrement universel, on peut retrouver la variété de base en quotientant par un groupe discret. S'il existe, le recouvrement universel est unique. L'avantage du recouvrement universel est de pouvoir travailler dans un espace bien connu dont la simple connexité évite de nombreux comportements pathologiques. Avec cette méthode, Stéphane a montré dans un papier à GSI qu'une métrique hyperbolique sur une surface de Riemann peut toujours être vu comme l'information de Fisher d'une famille de lois normales. Pour cela, l'idée est de définir la famille de distributions sur le recouvrement universel puis de projeter (en quotientant) sur la variété étudiée. Qu'est-ce qu'une surface de Riemann ? Il s'agit d'une courbe de  $\mathbb{C}^2$  de la forme  $\gamma = \{(z, \omega) \in \mathbb{C}^2 : z = \omega^2\}$ . Comme c'est une courbe de complexes, on peut la voir comme une surface de réels. Riemann avait introduit ces surfaces afin de pouvoir étudier  $\sqrt{z}$  qui n'est pas une fonction puisque tout complexe a deux racines (sauf 0) et qu'il est impossible d'en "choisir" une comme dans  $\mathbb{R}$  puisqu'il est possible de boucler et donc de devoir changer de choix (j'ai pas très bien compris ce passage), par exemple considérez le parcours du cercle unité par  $\theta \mapsto \exp i(\theta + 2k\pi)$ . On "choisit" la racine carrée "positive" (pas de moins devant l'exponentielle) :  $\theta \mapsto \exp i(\theta/2 + k\pi)$ . Par exemple, avec  $\theta = 2\pi$ , on obtient  $-1$  ou  $+1$ , il y a ambiguïté (?).
- Enfin, concernant le cas de l'ODE aléatoire, j'obtenais une courbure constante. Cela semble tout à fait cohérent. Une courbure constante correspond à des cercles (hyperboliques), ou plutôt des arcs de cercle, c'est à dire des points situés à égale distance hyperbolique d'un certain point. Dans ce cas, l'accélération est orthogonale à la vitesse (et c'est bien ce que donne le calcul).

#### Point du 15/03/2022. Objectifs pour l'étude géométrique de l'EDS linéaire de dimension $n$

- Concernant le cas scalaire, on peut interpréter les résultats de la façon suivante. Lorsque  $a > 0$  (trajectoires divergentes), la courbure converge vers une valeur finie. Cela signifie que la trajectoire s'approche d'un cercle hyperbolique, que l'on peut voir comme une courbe s'éloignant linéairement d'une géodésique. Lorsque  $a < 0$ , la courbure croît exponentiellement. Dans ce cas, la trajectoire s'éloigne très rapidement d'une géodésique.
- On s'intéresse maintenant à l'équation  $d\theta(t) = (A\theta(t) + b)dt + cdw(t)$  avec  $\theta(t) \in \mathbb{R}^n$ . La solution suit une loi normale  $\theta(t) \sim \mathcal{N}(\mu_t, \Sigma_t)$ . Une fois munie de la métrique de l'information de Fisher, la solution vit sur la variété  $\mathbb{R}^n \times \mathcal{S}_n^+$  où  $\mathcal{S}_n^+$  est l'ensemble des matrices symétriques définies positives de taille  $n \times n$ . L'espace tangent est  $\mathbb{R}^n \times \mathcal{S}_n \sim \mathbb{R}^n \times \mathbb{R}^{n(n+1)/2}$  où  $\mathcal{S}_n$  est l'ensemble des matrices symétriques de taille  $n \times n$ .
- Dans un premier temps, il s'agit de voir si on peut obtenir une expression pour les moments de la solution, puis pour l'information de Fisher. On souhaite ensuite étudier la courbure.

- On peut s'intéresser à la courbure sectionnelle, qui est définie de la façon suivante. Soit  $p$  un point d'une variété  $M$  et  $T_p M$  l'espace tangent à  $p$ . Soit  $\Pi$  un plan de  $T_p M$  engendrés par deux vecteurs non colinéaires  $(X, Y)$ . On peut définir la section plane déterminée par  $\Pi$ , notée  $S_\Pi$ , par  $S_\Pi = \exp_p \Pi$  (en toute rigueur, il faudrait restreindre  $\Pi$  à un voisinage de l'origine sur lequel  $\exp_p$  est un difféomorphisme). La section plane  $S_\Pi$  est une sous-variété de dimension 2 de  $M$  contenant  $p$  (il s'agit de l'ensemble des points atteints par les géodésiques partant de  $p$  avec un vecteur tangent initial dans  $\Pi$ ). La courbure sectionnelle (de  $M$  associée à  $\Pi$ ) n'est autre que la courbure de Gauss de la surface  $S_\Pi$  au point  $p$ , munie de la métrique induite. Elle est notée  $K(X, Y)$  ou  $K(\Pi)$ . La courbure sectionnelle vérifie  $K(X, Y) = \frac{Rm(X, Y, Y, X)}{|X|^2|Y|^2 - \langle X, Y \rangle^2}$  où  $Rm$  est le tenseur de Riemann.

La courbure de Gauss est définie de la façon suivante. Si  $II$  est la seconde forme fondamentale, on peut définir l'opérateur de forme  $s$  par : pour tous champs de vecteurs  $X$  et  $Y$ ,  $\langle X, sY \rangle = \langle II(X, Y), N \rangle$  où  $N$  est un champ de vecteurs normal unitaire à notre surface  $S_\Pi$  (dans ce cas, on voit  $S_\Pi$  comme une (hyper)surface plongée dans une variété de dimension 3 (?), le choix d'une orientation définissant un unique champ de vecteurs normal unitaire). La courbure de Gauss (=la courbure sectionnelle) est  $K(\Pi) = \det s = \kappa_1 \times \dots \times \kappa_n$  avec  $\kappa_1, \dots, \kappa_n$  les valeurs propres de  $s$  appelées courbures principales.

- On peut également s'intéresser à la courbure géodésique de la courbe  $t \mapsto (\mu_t, \Sigma_t)$ . Si la courbe a une vitesse de norme constante égale à 1, la courbure géodésique est définie par  $\kappa(t) = |D_t \dot{\gamma}(t)|$ . Pour une courbe quelconque, il faut d'abord reparamétriser la courbe pour obtenir une vitesse unitaire, ou utiliser cette formule :  $\kappa(t) = \frac{|D_t \dot{\gamma}(t)|}{|\dot{\gamma}(t)|^2} - \frac{\langle D_t \dot{\gamma}(t), \dot{\gamma}(t) \rangle}{|\dot{\gamma}(t)|^3}$ .
- Un calcul explicite (via les symboles de Christoffel etc.) sera sans doute difficile. Il faudra alors réfléchir à d'autres méthodes : soit renoncer à calculer la courbure et tenter de trouver des encadrements, soit passer par d'autres "astuces". Une telle astuce consiste à utiliser les champs de Jacobi ... mais je ne vois pas encore très bien comment. Cela nécessiterait d'avoir une forme close pour les géodésiques. Concernant les encadrements, on peut utiliser le lemme de Grönwall.
- Faire une recherche bibliographique sur l'étude géométrique des solutions de l'EDS linéaire afin de voir si des résultats n'ont pas déjà été obtenus sur ce sujet. Voir en particulier le *demi espace de Siegel* (par exemple [4]) qui pourrait avoir la même structure que la variété  $\mathcal{S}_n^+$  munie de l'information de Fisher (?).
- Voir aussi les matrices de Toeplitz symétriques qui pourraient correspondre aux matrices de covariance d'un processus avec des retards (???)
- Remarque. L'équation  $d\theta(t) = (A\theta(t) + b)dt + cdw(t)$  peut être vue comme la linéarisation d'une dynamique quelconque. Par exemple, il pourrait s'agir de la linéarisation de la dynamique conjuguée dont l'existence est donnée par le théorème de Takens (même si dans ce cas, les éléments de  $\theta$  ne seront pas indépendants). L'étude de cette équation est donc un premier pas vers l'étude de cette dynamique conjuguée.
- Aparté sur les distances. Si une variété est complète (ce qui est le cas pour les lois normales munies de l'information de Fisher), la métrique induit une distance  $d(p, q) = \min_\gamma l(\gamma)$  où  $\gamma$  est une courbe lisse quelconque telle que  $\gamma(0) = p$  et  $\gamma(1) = q$ . Le min est alors atteint lorsque  $\gamma$  est une géodésique.

Je me demandais si il y avait un lien privilégié entre cette distance et la divergence de Kullback-Leibler (KL). Notez que la divergence KL n'est pas une distance ; même si on la symétrise, elle ne vérifie pas l'inégalité triangulaire. On peut effectivement dire que la divergence KL correspond localement au carré de la distance  $d$ . En effet,  $KL(p||p + \epsilon) = g_p(\epsilon, \epsilon) + o(\epsilon^2)$  alors que  $d(p, p + \epsilon) = \sqrt{g_p(v, v)}$  où  $v$  est un vecteur tangent à  $p$  tel que la géodésique partant de  $p$  avec le vecteur tangent initial  $v$  atteigne  $p + \epsilon$  pour  $t = 1$ . Néanmoins, on pourra généralement (toujours ?) trouver des liens (i.e., des encadrements) entre diverses distances (ou "divergences") et la distance  $d$ .

Voici d'autres exemples de distances entre distributions de probabilités. La distance en variation totale  $d_{VT}(\mu, \nu) = \sup_{A \in \mathcal{F}} |\mu(A) - \nu(A)|$ . On voit bien que la distance VT s'annule ssi  $\mu = \nu$ . De plus, la distance VT est majorée par 1 (e.g., dans le cas de deux mesures étrangères, il existe un  $A$  tel que  $\mu(A) = 1$  et  $\nu(A^C) = 1$ , donc  $\nu(A) = 0$ ). Autre exemple, les distances  $L_p$  pour  $p \geq 1$  :

$d_p(q_1, q_2) = (\int_{\Omega} |q_1 - q_2|^p d\mu)^{1/p}$  avec  $q_1, q_2$  absolument continues par rapport à  $\mu$ . On peut encadrer ces distances les unes par rapport aux autres.

## References

- [1] J. P. Huke, “Embedding nonlinear dynamical systems: A guide to takens’ theorem,” tech. rep., Manchester Institute for Mathematical Sciences, 1993.
- [2] L. Noakes, “The takens embedding theorem,” *International Journal of Bifurcation and Chaos*, vol. 1, no. 4, pp. 867–872, 1991.
- [3] P. Molino, *Riemannian Foliations*. Birkhäuser, 1988.
- [4] F. Nielsen, “The Siegel–Klein Disk: Hilbert Geometry of the Siegel Disk Domain,” *Entropy*, vol. 22, Sept. 2020.