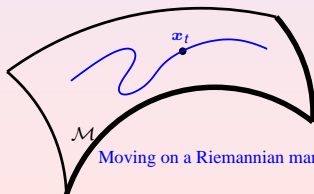


Géométrie de l'information et tracking sur les variétés riemanniennes

Hichem Snoussi

Université de Technologie de Troyes
Laboratoire Modélisation et Sécurité des Systèmes

10 octobre 2009



Moving on a Riemannian manifold

Plan

- 1 Quelques notions de géométrie différentielle
- 2 Filtrage bayésien sur les variétés riemanniennes
- 3 Cas des matrices définies positives
- 4 Perspectives
- 5 Problèmes à étudier ...

Plan

- 1 Quelques notions de géométrie différentielle
- 2 Filtrage bayésien sur les variétés riemanniennes
- 3 Cas des matrices définies positives
- 4 Perspectives
- 5 Problèmes à étudier ...

Plan

- 1 Quelques notions de géométrie différentielle
- 2 Filtrage bayésien sur les variétés riemanniennes
- 3 Cas des matrices définies positives
- 4 Perspectives
- 5 Problèmes à étudier ...

Plan

- 1 Quelques notions de géométrie différentielle
- 2 Filtrage bayésien sur les variétés riemanniennes
- 3 Cas des matrices définies positives
- 4 Perspectives
- 5 Problèmes à étudier ...

Plan

- 1 Quelques notions de géométrie différentielle
- 2 Filtrage bayésien sur les variétés riemanniennes
- 3 Cas des matrices définies positives
- 4 Perspectives
- 5 Problèmes à étudier ...

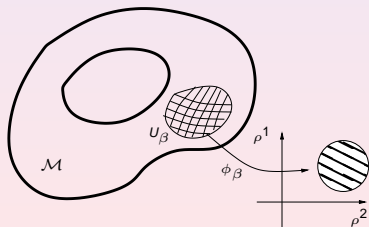
Plan

- 1 Quelques notions de géométrie différentielle
- 2 Filtrage bayésien sur les variétés riemanniennes
- 3 Cas des matrices définies positives
- 4 Perspectives
- 5 Problèmes à étudier ...

Notion de variété topologique

Définition

Une variété topologique est un **ensemble de points** \mathcal{S} muni d'un système de coordonnées $[\xi_i]_{i=1}^n$ ($\dim(\mathcal{S}) = n$). Chaque point $p \in \mathcal{S}$ possède un voisinage U et un homéomorphisme de U vers un voisinage ouvert de \mathbb{R}^n .



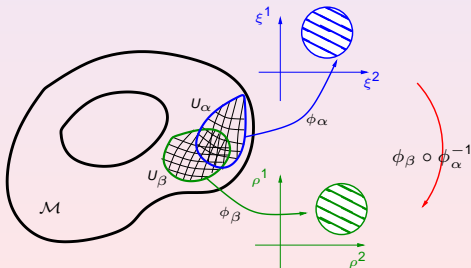
Exemples:

- Distributions paramétriques.
- Matrices définies positives.
- Sous-espaces de \mathbb{R}^n .

Notion de variété différentiable

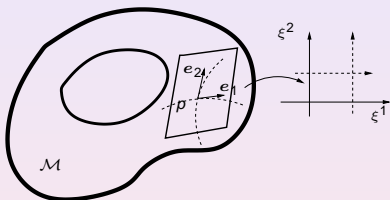
Définition

Une variété différentiable \mathcal{M} est une variété topologique munie d'une famille $\mathcal{U} = \{U_\alpha, \phi_\alpha\}$ de systèmes de voisinages tels que $\phi_\alpha \circ \phi_\beta^{-1}$ et $\phi_\beta \circ \phi_\alpha^{-1}$ sont des difféomorphismes sur les ouverts $\phi_\beta(U_\alpha \cap U_\beta)$ et $\phi_\alpha(U_\alpha \cap U_\beta)$ de \mathbb{R}^n .



Espace tangent

- Intuitivement, l'espace tangent, en un point p , est la **linéarisation** de \mathcal{S} autour de p .



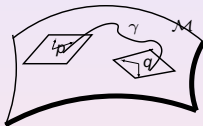
- $f : \mathcal{S} \rightarrow \mathbb{R}$ une fonction différentiable et $\gamma : I \rightarrow \mathcal{S}$ une courbe dans \mathcal{S} , la **dérivée directionnelle** de f le long de la courbe γ s'écrit:

$$\frac{d}{dt}f(\gamma(t)) = \sum \frac{\partial f}{\partial \xi_i} \frac{d\gamma_i}{dt} = \left[\sum \frac{d\gamma_i}{dt} \frac{\partial}{\partial \xi_i} \right] f$$

- Le vecteur $e_i = \frac{\partial}{\partial \xi_i}$ est un **opérateur de dérivation**.

Métrie riemannienne

- Définition: une **métrie riemannienne** est un tenseur g qui à chaque point $p \in \mathcal{S}$ associe un produit scalaire \langle, \rangle_p défini sur l'espace tangent T_p .



- Dans un système de coordonnées $\{\xi\}$, g est entièrement définie par les **matrices définies positives** G_p :

$$G_{ij}(p) = \left\langle \frac{\partial}{\partial \xi_i}, \frac{\partial}{\partial \xi_j} \right\rangle_p$$

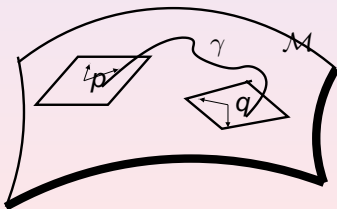
- Sur une variété \mathcal{S} , il existe une **infinité** de métriques riemanniennes \implies la métrique n'est pas une propriété intrinsèque.

Métrie riemannienne

- Soit une courbe $\gamma : [a, b] \rightarrow (\mathcal{S}, g)$, on définit sa longueur $\|\gamma\|$:

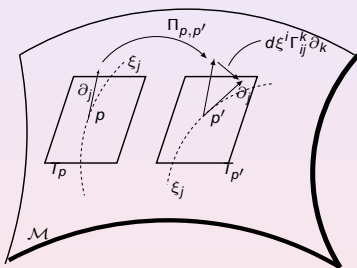
$$\|\gamma\| = \int_a^b \left\| \frac{d\gamma}{dt} \right\| dt = \int_a^b \sqrt{g_{ij} \dot{\gamma}^i \dot{\gamma}^j} dt$$

- Les **géodésiques** coïncident avec le chemin le plus **court** entre deux points.



Connexions affines

- Une **connexion affine** est une relation linéaire $\Pi_{p,p'}$ infinitésimale entre les **espaces tangents** de deux points voisins p et p' .



- Une connexion affine est définie par ses n^3 coefficients Γ_{ij}^k :

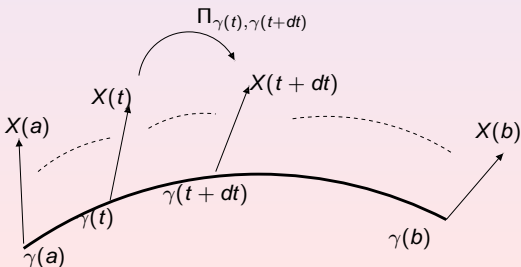
$$\Pi_{p,p'}((\partial_j)_p) = (\partial_j)_{p'} - d\xi^i (\Gamma_{ij}^k)_p (\partial_k)_{p'}$$

Connexions affines

- Soient p et q deux points de \mathcal{S} et γ une courbe reliant p et q . Si les vecteurs tangents $X(t)$ le long de la courbe γ vérifient la relation suivante:

$$X(t + dt) = \Pi_{\gamma(t), \gamma(t+dt)}(X(t))$$

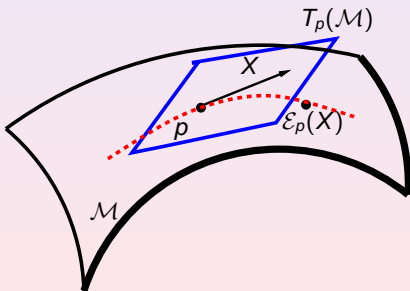
alors X est **parallèle** le long de γ et Π_γ est la **translation parallèle sur γ** .



Mapping exponentiel

Définition

Pour un point p et un vecteur tangent $X \in T_p(\mathcal{M})$, soit $\gamma : t \Rightarrow \gamma(t)$ la géodésique tel que $\gamma(0) = p$ et $\frac{d\gamma}{dt}(0) = X$. Le mapping exponentiel de X est défini par $\mathcal{E}_p(X) = \gamma(1)$.



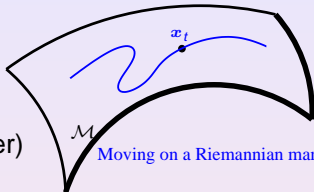
\Rightarrow Le **mapping exponentiel** représente un lien intéressant entre les techniques géométriques et les techniques euclidiennes usuelles.

Plan

- 1 Quelques notions de géométrie différentielle
- 2 Filtrage bayésien sur les variétés riemanniennes**
- 3 Cas des matrices définies positives
- 4 Perspectives
- 5 Problèmes à étudier ...

Filtrage bayésien sur les variétés riemanniennes

$$\begin{cases} \mathbf{x}_t \sim p_x(\mathbf{x}_t | \mathbf{x}_{t-1}), \mathbf{x}_t \in \mathcal{M} \\ \mathbf{y}_t \sim p_y(\mathbf{y}_t | \mathbf{x}_t), \text{ (ou critère à optimiser)} \end{cases}$$



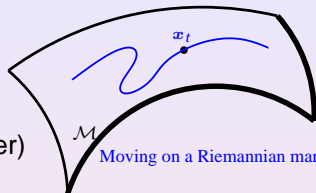
Moving on a Riemannian manifold

Contraintes

- Système dynamique non-linéaire et/ou non gaussien.
- L'état caché du système appartient à une variété riemannienne \mathcal{M} .
- Tracking dans l'espace euclidien $\mathcal{E} \supseteq \mathcal{M}$ et projection sur \mathcal{M} est **inefficace !**
- L'espace euclidien \mathcal{E} ne peut pas être défini (\mathcal{M} est définie d'une manière abstraite).

Filtrage bayésien sur les variétés riemanniennes

$$\begin{cases} \mathbf{x}_t \sim p_{\mathbf{x}}(\mathbf{x}_t | \mathbf{x}_{t-1}), \mathbf{x}_t \in \mathcal{M} \\ \mathbf{y}_t \sim p_{\mathbf{y}}(\mathbf{y}_t | \mathbf{x}_t), \text{ (ou critère à optimiser)} \end{cases}$$



Moving on a Riemannian manifold

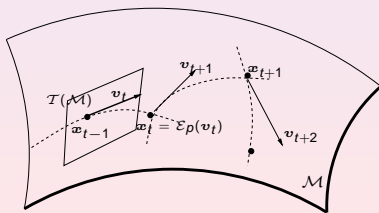
Objectifs

- Généraliser le filtrage bayésien sur les variétés riemanniennes en se basant sur les outils de la **géométrie différentielle**.
- Utilisation des propriétés géométriques **intrinsèques** (invariantes par rapport au choix de la paramétrisation).
- Application pour le tracking des matrices de Gram **variables dans le temps**.

Filtrage bayésien sur les variétés riemanniennes

Modèle de transition

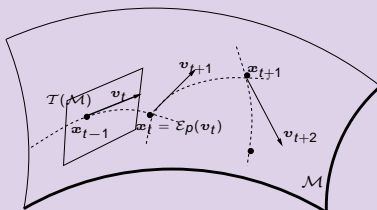
- Une chaîne de Markov $p_x(x_t | x_{t-1}, u_t)$ sur \mathcal{M} peut être définie par:
 - 1 Générer un échantillon v_t sur l'**espace tangent** $\mathcal{T}_{x_{t-1}}\mathcal{M}$ selon $p_v(\cdot)$.
 - 2 x est obtenu par **le mapping exponentiel** de v_t selon la connexion affine ∇ .



Filtrage bayésien sur les variétés riemanniennes

Filtrage particulaire:

- 1 Propager les trajectoires sur \mathcal{M} selon le processus markovien:



- 2 Pondérer les particules selon leurs vraisemblances.

Filtrage bayésien sur les variétés riemanniennes

Estimation:

- Le calcul de la moyenne empirique des particules n'est plus une opération valide sur la variété \mathcal{M} .
- L'estimation ponctuelle peut être définie par la **moyenne intrinsèque** (barycentre riemannien [Fréchet 1948]):

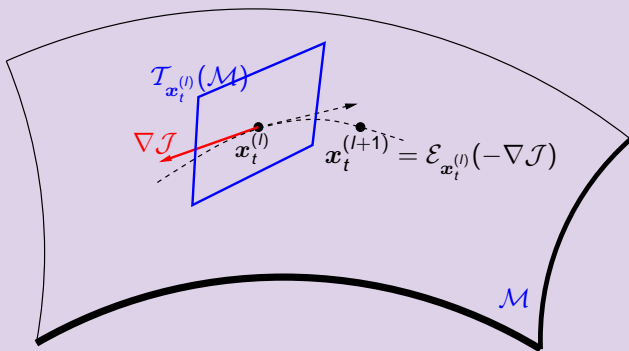
$$\begin{aligned}\hat{x}_t &= \arg \min_{x_t \in \mathcal{M}} \mathbb{E}[(\mathcal{D}(x_t, s_t))^2] \\ &\approx \arg \min_{x_t \in \mathcal{M}} \sum_{i=1}^N w_t^{(i)} (\mathcal{D}(x_t, s_t^{(i)}))^2\end{aligned}$$

où $s_t^{(i)}$ et $w_t^{(i)}$ représentent les particules et leurs poids respectifs et $\mathcal{D}(\cdot, \cdot)$ est **la distance géodésique**.

Filtrage bayésien sur les variétés riemanniennes

Estimation:

- La minimisation du critère peut être implémentée par un algorithme de gradient en se basant sur le mapping exponentiel:



Plan

- 1 Quelques notions de géométrie différentielle
- 2 Filtrage bayésien sur les variétés riemanniennes
- 3 Cas des matrices définies positives**
- 4 Perspectives
- 5 Problèmes à étudier ...

Exemple: Tracking de la covariance de bruit

$$\left\{ \begin{array}{l} \mathbf{x}_t \sim p_x(\mathbf{x}_t | \mathbf{x}_{t-1}, \boldsymbol{\alpha}) \\ \boldsymbol{\Sigma}_t \sim \mathcal{GN}(\boldsymbol{\Sigma}_t | \boldsymbol{\Sigma}_{t-1}, \boldsymbol{\Lambda}) \longrightarrow \text{covariance du bruit} \\ \mathbf{y}_t \sim \mathcal{N}(\mathbf{y}_t | f(\mathbf{x}_t), \boldsymbol{\Sigma}_t), \end{array} \right.$$

où $\boldsymbol{\Sigma}_t \sim \mathcal{GN}(\boldsymbol{\Sigma}_t | \boldsymbol{\Sigma}_{t-1}, \boldsymbol{\Lambda})$ est une marche aléatoire sur la variété riemannienne \mathcal{S}_+ des **matrices définies positives** :

- 1 Echantillonnage d'une matrice symétrique gaussienne $B \in \mathcal{S}$ avec une précision $\boldsymbol{\Lambda} \left(\frac{n_x(n_x+1)}{2} \times \frac{n_x(n_x+1)}{2} \right)$: $B \sim \mathcal{N}(0; \boldsymbol{\Lambda})$
- 2 La matrice suivante $\boldsymbol{\Sigma}_t$ est obtenue par le mapping exponentiel:

$$\boldsymbol{\Sigma}_t = \mathcal{E}_{\boldsymbol{\Sigma}_{t-1}}(B) = \boldsymbol{\Sigma}_{t-1}^{1/2} \exp \left[\boldsymbol{\Sigma}_{t-1}^{-1/2} B \boldsymbol{\Sigma}_{t-1}^{-1/2} \right] \boldsymbol{\Sigma}_{t-1}^{1/2}$$

Exemple: Tracking de la covariance de bruit

- L'expression explicite de la distance géodésique permet le calcul du barycentre riemannien.
- La variance *a posteriori* de l'erreur peut être approchée par:

$$\bar{\Sigma}_t = \arg \min_{\Sigma \in \mathcal{S}_+} \sum_{i=1}^N w_t^{(i)} \frac{1}{2} \text{tr} \text{Ln}^2(\Sigma^{-1/2} \Sigma_t^{(i)} \Sigma^{-1/2})$$

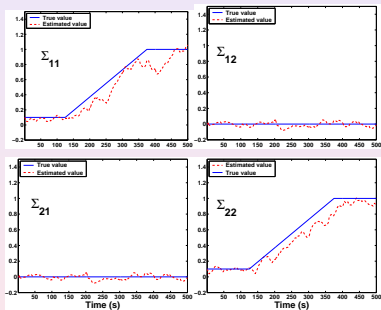
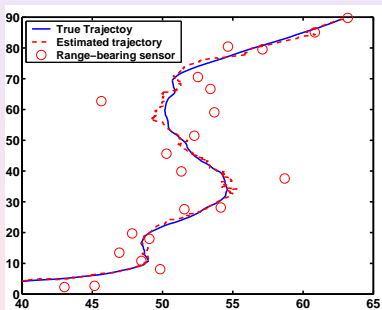
- Le gradient, appartenant à l'espace tangent $\mathcal{T}_{\Sigma} \mathcal{S}_+$, possède l'expression suivante:

$$\nabla \mathcal{J}(\Sigma) = \frac{\Sigma}{N} \sum_{i=1}^N \text{Ln}((\Sigma_t^{(i)})^{-1} \Sigma)$$

- Etant données les expressions du gradient et du mapping exponentiel, l'algorithme de gradient se met sous la forme:

$$\begin{aligned} \bar{\Sigma}^{(l+1)} &= \mathcal{E}_{\bar{\Sigma}^{(l)}}(-\nabla \mathcal{J}(\bar{\Sigma}^{(l)})) \\ &= (\bar{\Sigma}^{(l)})^{1/2} \exp \left[(-\bar{\Sigma}^{(l)})^{-1/2} \nabla \mathcal{J}(\bar{\Sigma}^{(l)})^{-1/2} \right] (\bar{\Sigma}^{(l)})^{1/2} \end{aligned}$$

Exemple: Tracking de la covariance de bruit



Plan

- 1 Quelques notions de géométrie différentielle
- 2 Filtrage bayésien sur les variétés riemanniennes
- 3 Cas des matrices définies positives
- 4 Perspectives**
- 5 Problèmes à étudier ...

Perspectives

- Estimer en ligne des lois normales multivariées non centrées (exemple: biais du bruit inconnu et variant dans le temps).
- Difficulté: calcul des géodésiques (nécessaire pour simuler la chaîne de Markov):
 - 1 Adopter des schémas de calcul numérique.
 - 2 Changer la géométrie et ne pas se restreindre aux géodésiques riemanniennes.
 - 3 Approcher les géodésiques par des inégalités triangulaires.
- Application en machine learning: estimer la matrice de Gram / comparer les métriques pour des données matricielles.

Perspectives

- Estimer en ligne des lois normales multivariées non centrées (exemple: biais du bruit inconnu et variant dans le temps).
- Difficulté: calcul des géodésiques (nécessaire pour simuler la chaîne de Markov):
 - 1 Adopter des schémas de calcul numérique.
 - 2 Changer la géométrie et ne pas se restreindre aux géodésiques riemanniennes.
 - 3 Approcher les géodésiques par des inégalités triangulaires.
- Application en machine learning: estimer la matrice de Gram / comparer les métriques pour des données matricielles.

Perspectives

- Estimer en ligne des lois normales multivariées non centrées (exemple: biais du bruit inconnu et variant dans le temps).
- Difficulté: calcul des géodésiques (nécessaire pour simuler la chaîne de Markov):
 - 1 Adopter des schémas de calcul numérique.
 - 2 Changer la géométrie et ne pas se restreindre aux géodésiques riemanniennes.
 - 3 Approcher les géodésiques par des inégalités triangulaires.
- Application en machine learning: estimer la matrice de Gram / comparer les métriques pour des données matricielles.

Plan

- 1 Quelques notions de géométrie différentielle
- 2 Filtrage bayésien sur les variétés riemanniennes
- 3 Cas des matrices définies positives
- 4 Perspectives
- 5 Problèmes à étudier ...

Sélection de modèles

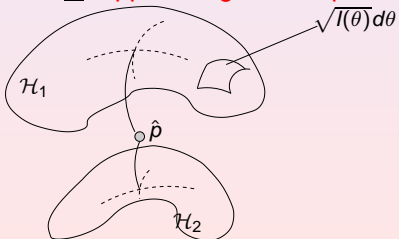
- Données $\mathbf{y} = (y_1, y_2, \dots, y_N)$ et plusieurs modèles $\mathcal{H}_i = \{f_i(\mathbf{y} | \theta^i)\}$: \mathbf{y} sont distribuées selon quel modèle ?
- Approches classiques: compromis entre attache aux données et complexité,

$$\text{AIC} = -2 \log f(\mathbf{y} | \hat{\theta}) + 2k$$

$$\text{BIC} = -2 \log f(\mathbf{y} | \hat{\theta}) + k \log N$$

$$\text{SC} = -\log f(\mathbf{y} | \hat{\theta}) + \frac{k}{2} \log N$$

- Approche géométrique:



- Avec un critère bayésien:

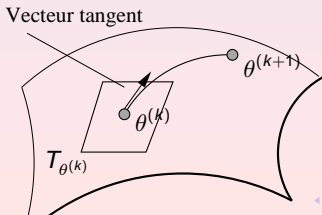
$$-\log \text{Pr}(\mathcal{H} | \mathbf{y}) = -\log f(\mathbf{y} | \hat{\theta}) + \frac{k}{2} \log\left(\frac{N}{2}\right) + \log \int d\theta \sqrt{\det I(\theta)} + \frac{1}{2} \log\left(\frac{\det J(\hat{\theta})}{\det I(\hat{\theta})}\right)$$

Algorithmes d'apprentissage sur les variétés

- Objectif: **estimer** un paramètre $\theta \in \mathcal{S}$ (variété riemannienne) en minimisant une fonction coût $J(\theta)$.
 → Algorithmes de type **gradient**: construire une suite $\theta^{(k)}$ convergent vers la solution:

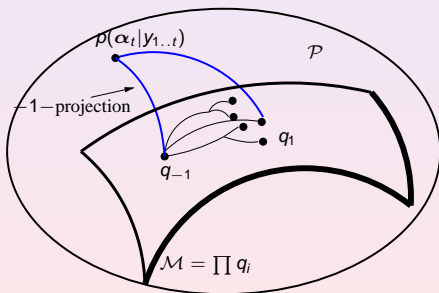
$$\theta^{(k+1)} = \theta^{(k)} + \alpha_k L(\theta^{(k)}), \quad k = 1, \dots, M, \dots$$

- **Non adapté** à la structure géométrique de l'espace des paramètres \mathcal{S} .
- Approche géométrique: **Apprentissage géodésique** sur \mathcal{S} →
Se déplacer sur des géodésiques de vecteur tangent (à l'origine) le gradient de la fonction coût



Bifurcations

- Unicité des solutions (projections) en fonction de la géométrie choisie: étude de la courbure de la sous-variété \mathcal{M} et du hessien du critère: **effet de bifurcation !**



Merci pour votre attention !