

Exposé d'Habilitation à Diriger les Recherches

S. Gadat

Institut de Mathématiques de Toulouse
Université Paul Sabatier

22 Novembre 2012

I - Modèles déformables en statistique

I - 1 Motivations

I - 2 Modèle statistique

I - 3 Fiabilité des méthodes de recalage

I - 4 Estimation par approche Bayésienne

I - 5 Conclusion sur le modèle SIM

II Diffusion sur une descente de gradient à mémoire

II - 1 Introduction au modèle de gradient moyenné

II - 2 Hypo-ellipticité

II - 3 Comportement asymptotique

II - 4 Convergence lorsque $\sigma \rightarrow 0$

II - 5 Formulation variationnelle

II - 6 Conclusion

I - 1 Motivations

Traitement du signal et statistiques : Comparer des objets qui partagent des caractéristiques de **formes communes** et extraire des informations sur la distribution de ces objets

Outils classiques :

- ▶ Statistiques du premier ordre : moyenne empirique
- ▶ Statistiques du second ordre : Analyse en Composantes Principales

Pour n variables aléatoires $(Y_m)_{m=1,\dots,n}$ i.i.d.

$$\bar{Y}^n = \frac{1}{n} \sum_{m=1}^n Y_m$$

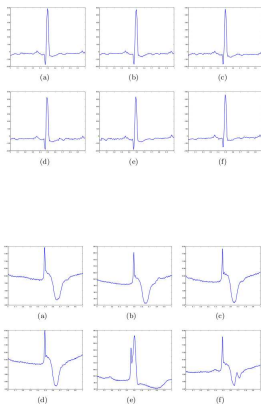
Puis, l'ACP consiste à diagonaliser la covariance empirique

$$S = \frac{1}{n} \sum_{m=1}^n (Y_m - \bar{Y}^n)(Y_m - \bar{Y}^n)',$$

et choisir les premiers vect. propres comme modes de variations principaux.

I - 1 Motivations : Applications en biologie

Problématique : On s'intéresse à des données qui partagent des caractéristiques de **forme commune**.



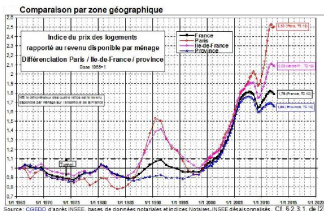
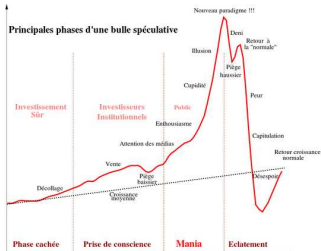
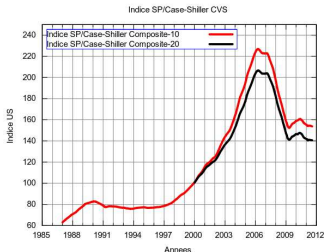
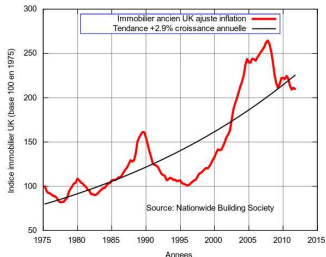
Source : J. Bigot 2012, Preprint, Fréchet means of curves for signal averaging and application to ECG data analysis

I - 1 Motivations : Applications en économie

Bulles spéculatives

Sources : (b) Standard & Pears Case Shiller, (c) J.-P. Rodrigue, Hofstra University (d) J. Friggit,

Conseil Général de l'Environnement et du Gouvernement Durable.



I - 1 Motivations : Applications en traitement de l'image

Données : Images de $N \times N$ pixels



Problème : Approches standard non adaptées pour les données Y_m déformées puisque \bar{Y}^n n'est pas une bonne estimation de la forme moyenne.

I - Modèles déformables en statistique

I - 1 Motivations

I - 2 Modèle statistique

I - 3 Fiabilité des méthodes de recalage

I - 4 Estimation par approche Bayésienne

I - 5 Conclusion sur le modèle SIM

II Diffusion sur une descente de gradient à mémoire

II - 1 Introduction au modèle de gradient moyenné

II - 2 Hypo-ellipticité

II - 3 Comportement asymptotique

II - 4 Convergence lorsque $\sigma \rightarrow 0$

II - 5 Formulation variationnelle

II - 6 Conclusion

I - 2 Modèle statistique

Objectifs idéaux : Proposer un estimateur « consistant » de la forme moyenne d'un ensemble de données et estimer leurs modes de variations.

Dans cet exposé : Décrire ce qui est possible ou non d'un point de vue statistique, dans un cadre simplifié pour la **forme moyenne**.

Modèle déformable : l'observation $Y_m : \Omega \rightarrow \mathbb{R}$, $m = 1, \dots, n$ vérifie

$$Y_m(x) = \mathbf{f} \circ \phi_m(x) + W_m(x), \quad x \in \Omega, \text{ où}$$

- ▶ $\mathbf{f} : \mathbb{R}^d \rightarrow \mathbb{R}$ est la forme commune inconnue
- ▶ ϕ_m : déformations aléatoires qui modélisent des variations de forme de \mathbf{f} .
- ▶ W_m est un bruit additif indépendant de moyenne nulle

Dans cet exposé : **Comment estimer \mathbf{f} et ϕ_m lorsque $n \rightarrow +\infty$?**

I - 2 Modèle statistique

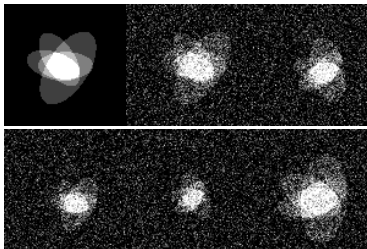
Déformations rigides

- ▶ Translation : $\phi(x) = x - \tau$ où $b \in \mathbb{R}^d$
- ▶ Rotation + scaling (dans \mathbb{R}^2) : $\phi(x) = \frac{1}{a}A_\theta x$ avec $a \in \mathbb{R}^+$ et

$$A_\theta = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}$$

- ▶ Affine (Translation + rotation + scaling) : $\phi(x) = \frac{1}{a}A_\theta(x - \tau)$, soit 2D ou 3D
- ▶ Exemple sur le « Fantôme de Shepp-Logan » Source : Jain, Anil K., Fundamentals of Digital

Image Processing, Englewood Cliffs, NJ, Prentice Hall, 1989



Déformations élastiques : Un autre monde...

I - 2 Shape Invariant Model

Modèle le plus simple : Observation de courbes shiftées aléatoirement

$Y_m : [0, 1] \rightarrow \mathbb{R}$, $m = 1, \dots, n$ telles que

$$Y_m(x) = f(x - \tau_m) + W_m(x), \text{ pour } x \in [0, 1], \text{ où} \quad (1)$$

- ▶ $f : [0, 1] \rightarrow \mathbb{R}$ est 1-périodique
- ▶ les τ_m sont des shifts aléatoires i.i.d. de loi de densité g
- ▶ les W_m sont des bruits additifs centrés indépendants des shifts

La moyenne empirique est inconsistante : si $n \rightarrow +\infty$

$$\bar{Y}^n(x) = \frac{1}{n} \sum_{m=1}^n Y_m(x) \rightarrow \mathbb{E}f(x - \tau_1) = \int f(x - \tau)g(\tau)d\tau = f \star g(x) \neq \mathbf{f}(x)$$

La suite de cet exposé se cantonnera à des estimations dans le modèle de shift aléatoire lorsque $n \mapsto +\infty$ pour f , $(\tau_m)_{m=1\dots n}$ et g .

I - Modèles déformables en statistique

I - 1 Motivations

I - 2 Modèle statistique

I - 3 Fiabilité des méthodes de recalage

I - 4 Estimation par approche Bayésienne

I - 5 Conclusion sur le modèle SIM

II Diffusion sur une descente de gradient à mémoire

II - 1 Introduction au modèle de gradient moyenné

II - 2 Hypo-ellipticité

II - 3 Comportement asymptotique

II - 4 Convergence lorsque $\sigma \rightarrow 0$

II - 5 Formulation variationnelle

II - 6 Conclusion

I - 3 Fiabilité des méthodes de recalage

a) Idée Naïve

- ▶ On estime puis inverse tous les paramètres de déformation $\tau_j \mapsto -\hat{\tau}_j$.
- ▶ On obtient alors une estimation en « recalant » les observations :

$$\bar{Y}^n = \frac{1}{n} \sum_{j=1}^n Y_j(x + \hat{\tau}_j).$$

- ▶ Est-ce envisageable d'estimer correctement les τ_j ? Sous quelle asymptotique (nombre de répétitions, niveau de bruit, ...) ?

b) Moyenne de Fréchet **Quelle distance ?**

- ▶ **Distance euclidienne** : $d(y, y') = \|y - y'\|_{\mathcal{H}} \rightsquigarrow$ Moyenne Empirique

$$\bar{Y}_n = \frac{1}{n} \sum_{j=1}^n Y_j = \arg \min_{y \in \mathcal{H}} \sum_{j=1}^n \|Y_j - y\|_{\mathcal{H}}^2$$

- ▶ **Distance associée aux déformations** : \rightsquigarrow **Moyenne au sens de Fréchet**

$$d_G^2(Y, h) := \inf_{\tau \in [0,1]} \int_0^1 |Y(x) - h(x - \tau)|^2 dx$$

$$\tilde{Y}^n \in \arg \min_{h \in \mathcal{H}} \sum_j d_G^2(Y_j, h).$$

I - 3 c) Estimation par Moyenne de Fréchet

Moyenne de Fréchet empirique $\tilde{Y}^n \in \arg \min_{y \in \mathcal{H}} \frac{1}{n} \sum_{m=1}^n d_G^2(y, Y_m)$.

Moyenne de Fréchet théorique $Y^F := \arg \min_{y \in \mathcal{H}} \mathbb{E}_{W,g} d_G^2(y, Y_{W,g})$

- ▶ Consistance et convergence de \tilde{Y}^n vers Y^F lorsque $n \rightarrow +\infty$ démontrée dans le cas où \mathcal{H} est une variété Riemannienne de **dimension finie**.
- ▶ **Problème** : Résultats non adaptés à l'étude des moyennes de Fréchet pour les courbes ou images paramétrées (dimension "infinie").

Calcul de la moyenne de Fréchet Empirique

Étape 1 - Registration/warping des données

$$(\hat{\tau}_1, \dots, \hat{\tau}_n) = \arg \min_{(\tau_1, \dots, \tau_n) \in [0,1]^n} \left\{ \frac{1}{n} \sum_{j=1}^n \int_{\Omega} \left| Y_j(t - \tau_j) - \frac{1}{n} \sum_{m=1}^n Y_m(t - \tau_m) \right|^2 dt \right.$$

Étape 2 - Alignement et moyennisation des données

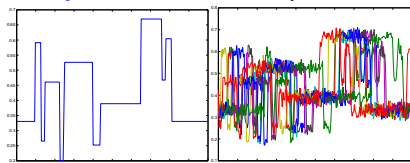
$$\tilde{Y}_n = \frac{1}{n} \sum_{j=1}^n Y_j(\cdot - \hat{\tau}_j)$$

I - 3 - d) Application de la Moyenne de Fréchet

Données : courbes translattées $f \in L^2_{per}([0, 1])$ + modèle de bruit blanc

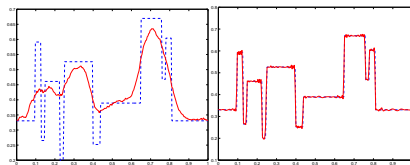
$$Y_j(t) = f(t - \tau_j) + W_j(t), \text{ for } t \in [0, 1], \text{ et } \tau_j \text{ i.i.d. translations aléatoires,}$$

Courbe f / Échantillon de 10 parmi $n = 200$



Moyenne euclidienne

Moyenne de Fréchet



I - 3 e) Fiabilité des processus de recalage

On suppose

- ▶ (H_g) : g à support compact dans $\mathcal{T} = [-\frac{1}{4}, \frac{1}{4}]$, de moyenne nulle.
- ▶ (H_f) : la fonction f est telle que $c_1(f) \neq 0$

Theorem (Bigot & G. '10)

Sous (H_f) et (H_g) , le modèle en f est identifiable. ☺

Mais : On ne peut pas estimer les paramètres de recalage correctement :

Theorem (Bigot & G. '10, Bigot & Charlier '11, B.G. Klein Marteau '12)

Sous (H_f) et (H_g) , **quels que soient des estimateurs** $(\hat{\tau}_1, \dots, \hat{\tau}_n)$ les vrais paramètres de recalage $(\tau_1^*, \dots, \tau_n^*)$, on a

$$\mathbb{E} \left(\frac{1}{n} \sum_{j=1}^n (\hat{\tau}_j - \tau_j^*)^2 \right) \geq \frac{\epsilon^2}{\sum_{k \in \mathbb{Z}} (2\pi k)^2 |c_k(f)|^2 + \epsilon^2 \int_{\mathcal{T}} \left(\frac{\partial}{\partial \theta} \log g(\tau) \right)^2 g(\tau) d\tau} > 0.$$

Sous (H_f) et (H_g) , **quels que soient des estimateurs** $(\hat{\tau}_1, \dots, \hat{\tau}_n)$ les vrais paramètres de recalage $(\tau_1^*, \dots, \tau_n^*)$, on a

$$\mathbb{E} \left\| \frac{1}{n} \sum_{j=1}^n f(\cdot - \hat{\tau}_j + \tau_j^*) - f \right\|_2^2 \geq C \epsilon^2 \frac{c_1(f)^2}{\|f'\|_2 + \epsilon^2 I(g)} > 0.$$



I - Modèles déformables en statistique

I - 1 Motivations

I - 2 Modèle statistique

I - 3 Fiabilité des méthodes de recalage

I - 4 Estimation par approche Bayésienne

I - 5 Conclusion sur le modèle SIM

II Diffusion sur une descente de gradient à mémoire

II - 1 Introduction au modèle de gradient moyenné

II - 2 Hypo-ellipticité

II - 3 Comportement asymptotique

II - 4 Convergence lorsque $\sigma \rightarrow 0$

II - 5 Formulation variationnelle

II - 6 Conclusion

I - 4 Estimation par approche Bayésienne

Cadre :

$$Y_j(x) = f^0(x - \tau_j) + \epsilon W_j(x),$$

où (τ_j) i.i.d. $\sim g^0$ et f^0 et g^0 inconnus dans une famille non paramétrique.

Méthode : Poser une loi *a priori* π_1 sur f^0 et π_2 sur g^0 , puis utiliser la loi *a posteriori* pour fabriquer des estimations de f^0 et g^0 .

$$\pi := (\pi_1 \otimes \pi_2) \text{ et } \pi_n := \pi[\cdot \mid Y_1^n]$$

Questions "fréquentistes" :

Question 1 : Lorsque $n \mapsto +\infty$, a-t-on $\pi_n \mapsto \delta_{f^0, g^0}$?

Question 2 : Quelle est la vitesse de contraction autour de δ_{f^0, g^0} ?

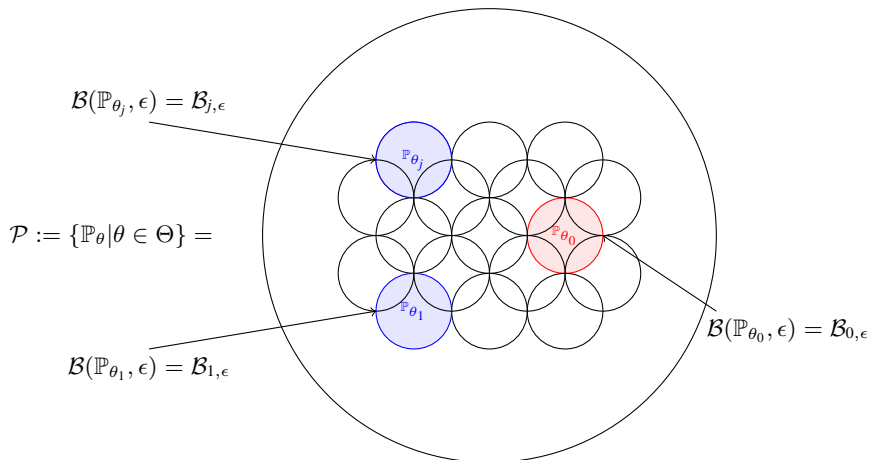
Dynamique markovienne de l'*a posteriori* :

$$\pi_{n+1}(\theta) = \frac{\pi_n(\theta) \mathbb{P}_\theta(Y_{n+1})}{\int_{\Theta} \pi_n(u) \mathbb{P}_u(Y_{n+1}) du}$$

I - 4 Approche Bayésienne : recouvrements

Cadre :

$$Y_j(x) = f^0(x - \tau_j) + \epsilon W_j(x),$$



- ▶ f et g sont paramétrées par $\theta \in \Theta$, espace des paramètres.
- ▶ Les observations $(Y_j)_{j=1\dots n}$ sont issues d'un θ^0 inconnu, ici $\theta^0 = (\theta_f^0, \theta_g^0)$

I - 4 Approche Bayésienne : contraction de l'a posteriori

Par des inégalités de convexité, on démontre que

$$\mathbb{E} \left[\sqrt{\frac{\int_{\mathcal{B}_{j,\epsilon}} \pi_{n+1}(u) du}{\int_{\mathcal{B}_{0,\epsilon}} \pi_{n+1}(u) du}} \middle| \pi_n \right] \leq e^{-c\epsilon^2} \sqrt{\frac{\int_{\mathcal{B}_{j,\epsilon}} \pi_n(u) du}{\int_{\mathcal{B}_{0,\epsilon}} \pi_n(u) du}}. \quad (2)$$

Si $\mathcal{B}_{j,\epsilon}$ est un ϵ -recouvrement de Θ , alors

$$\pi_n(\mathcal{B}_{0,\epsilon}) \geq 1 - \pi_n(\mathcal{B}_{0,\epsilon}) \sum_{j \geq 1}^{N_\epsilon} \frac{\pi_n(\mathcal{B}_{j,\epsilon})}{\pi_n(\mathcal{B}_{0,\epsilon})} \geq 1 - C \pi_n(\mathcal{B}_{0,\epsilon}) N_\epsilon e^{-2c\epsilon^2 n} \quad (3)$$

Bilan : contraction de π_n sur $\mathcal{B}_{0,\epsilon}$ (sur $\mathcal{B}_{0,\epsilon_n}$) si

- ▶ $\log N_\epsilon \lesssim n\epsilon^2 \quad \rightsquigarrow \quad \log N_{\epsilon_n} \lesssim n\epsilon_n^2$
- ▶ $\pi(\mathcal{B}_{0,\epsilon}) \geq c(\epsilon, n) > 0 \quad \rightsquigarrow \quad \pi(\mathcal{B}_{0,\epsilon_n}) \geq c(\epsilon_n, n) > 0.$

Theorem (Bontemps & G. '12)

Dans le modèle où $f \in \mathcal{H}_s^\ell(A)$, $g \in \mathfrak{M}([0, 1])$ et $\mathcal{P}_{\ell,A}$ décrit par

$$Y_j(x) = f^0(x - \tau_j) + \epsilon W_j(x), \tau_j \text{ i.i.d } \sim g^0$$

alors $\epsilon > 0$ assez petit et si $A_\epsilon \lesssim \sqrt{\ell_\epsilon}$

$$\log N(\epsilon, \mathcal{P}_{\ell_\epsilon, A_\epsilon}, d_{VT}) \lesssim \ell_\epsilon^2 \left[\log \frac{1}{\epsilon} + \log \ell_\epsilon \right].$$

I - 4 Approche Bayésienne

Construction de l'*a priori* π :

- ▶ On choisit un entier ℓ_{max} suivant une loi vérifiant $p(\ell_{max} = k) \propto e^{-ck^2 \log^{3/2} k}$.
- ▶ Chaque coefficient "allumé" suit une $\mathcal{N}_{\mathbb{C}}(0, n^{-1/4} \log^{-3/2} n)$.
- ▶ Processus de Dirichlet $D(\alpha)$ comme *a priori* sur g .

Theorem (Consistance Bayésienne, Bontemps & G. '12)

Si $f^0 \in \mathcal{H}_s$ avec $s \geq 1$, alors pour $\epsilon_n = n^{-\lceil s/(2s+2) \wedge 3/8 \rceil} \log n$:

$$\pi_n \left\{ \mathbb{P}_{f,g} \text{ s.t. } d_H(\mathbb{P}_{f,g}, \mathbb{P}_{f^0,g^0}) \leq M\epsilon_n \right\} = 1 + O_p(1)$$

Remarques :

- ▶ On obtient une vitesse polynomiale. L'*a priori* est adaptatif pour $s \in [1, 3]$.
- ▶ On paye le $\ell_\epsilon^2 \log \frac{1}{\epsilon}$ en obtenant une vitesse $2s/(2s+2)$ au lieu de $2s/(2s+1)$: il faut estimer f^0 et g^0 .

I - 4 Approche Bayésienne : revenir aux objets dans \mathcal{H}_s et $\mathfrak{M}_\nu([0, 1])$?

Classe d'identifiabilité : $\mathcal{F}_s := \{f \in H_s \text{ et } c_1(f) > 0\}$ et

$$\mathfrak{M}_\nu([0, 1]) := \{g \in \mathfrak{M}([0, 1]) \mid \exists \forall k \in \mathbb{Z} \quad c|k|^{-\nu} < |c_k(g)| < C|k|^{-\nu}\}$$

Proposition (Bontemps & G. '12)

Si $(f, g) \in \mathcal{F}_s \times \mathfrak{M}_\nu([0, 1])$, alors le modèle est identifiable.

Idée de la preuve :

- ▶ Établir que $d_{VT}(\mathbb{P}_{f,g}^1, \mathbb{P}_{\tilde{f},\tilde{g}}^1) > 0 \Rightarrow c_1(f) = c_1(\tilde{f})$.
- ▶ En déduire que nécessairement $g = \tilde{g}$ par méthode Hilbertienne ou transformée de Laplace.
- ▶ Conclure sur les autres coefficients de f en utilisant les autres marginales de $\mathbb{P}_{f,g}$.

I - 4 Approche Bayésienne : revenir aux objets dans \mathcal{H}_s et $\mathfrak{M}_\nu([0, 1])$?

- ▶ Implicitement, cela pose la question de la robustesse de l'identifiabilité précédente.
- ▶ Pour g , il faut utiliser une minoration du type

$$d_{VT}(\mathbb{P}_{f,g}, \mathbb{P}_{f,g^0}) \geq C(\theta_1^0) \sum_{n \in \mathbb{Z}} |c_n(g - g^0)|^2 \omega_n^2,$$

où ω_n est à relier à la fonction de Bessel modifiée (densité d'une CTRW).
Malheureusement, $\omega_n^2 \sim e^{-n \log n}$!

- ▶ Très forte dégénérescence du problème inverse et ...

Theorem (Consistance Bayésienne, Bontemps & G. '12)

Si $(f^0, g^0) \in \mathcal{F}_s(A) \times \mathfrak{M}_\nu([0, 1])$ inconnus, alors

$$\Pi_n \left(g : \|g - g^0\|_2^2 > M \log^{-2\nu}(n) |Y_1, \dots, Y_n \right) \mapsto 0$$

$$\Pi_n \left(f : \|f - f^0\|_2^2 > M (\log n)^{-2s \times \frac{2\nu}{2s+2\nu+1}} |Y_1, \dots, Y_n \right) \longrightarrow 0$$

III - 7 Approche Bayésienne : bornes inférieures

Optimalité des bornes précédentes ?

- ▶ Il s'agit de se placer dans les cas où l'injectivité de l'application $(f, g) \mapsto \mathbb{P}_{f,g}$ est peu robuste.
- ▶ Dans la preuve de l'identifiabilité, on constate que les difficultés arrivent lorsque les modules des coeffs. de Fourier sont égaux : $|\theta_\ell^0| = |\theta_\ell|$.
- ▶ On construit des fonctions tests $f_j(x) = e^{i2\pi x} + \theta_j e^{i2\pi \nu x}$ avec $|\theta_j| = |\theta_{j'}|, \forall j \neq j'$ et g_j telle que $f_j \star g_j \simeq f_{j'} \star g_{j'}$ (on matche les premières fréquences).

Le lemme de Fano permet d'obtenir :

Theorem (Bontemps & G. '12)

Si $(f^0, g^0) \in \mathcal{F}_s(A) \times \mathfrak{M}_\nu([0, 1])$, on peut trouver c tel que le risque minimax sur $\mathcal{F}_s(A) \times \mathfrak{M}_\nu([0, 1])$ est minoré par

$$\liminf_{n \rightarrow +\infty} (\log n)^{2s+2} \inf_{\hat{f} \in \mathcal{F}_s(A)} \sup_{(f^0, g^0) \in \mathcal{F}_s(A) \times \mathfrak{M}_\nu([0, 1])} \|\hat{f} - f\|_2^2 \geq c,$$

et

$$\liminf_{n \rightarrow +\infty} (\log n)^{2\nu+1} \inf_{\hat{g} \in \mathcal{F}_s(A)} \sup_{(f^0, g^0) \in \mathcal{F}_s(A) \times \mathfrak{M}_\nu([0, 1])} \|\hat{g} - g\|_2^2 \geq c.$$

I - Modèles déformables en statistique

I - 1 Motivations

I - 2 Modèle statistique

I - 3 Fiabilité des méthodes de recalage

I - 4 Estimation par approche Bayésienne

I - 5 Conclusion sur le modèle SIM

II Diffusion sur une descente de gradient à mémoire

II - 1 Introduction au modèle de gradient moyenné

II - 2 Hypo-ellipticité

II - 3 Comportement asymptotique

II - 4 Convergence lorsque $\sigma \rightarrow 0$

II - 5 Formulation variationnelle

II - 6 Conclusion

I - 5 Conclusion sur le modèle SIM

Sur la partie Bayésienne :

- ▶ Le problème statistique est du type "problème inverse" et devient très mal posé lorsque la loi des déformations est inconnue.
- ▶ Les bornes supérieures et inférieures ne sont pas toujours collées.
- ▶ Question ouverte pour des **algorithmes efficaces de simulation de loi a posteriori**...

Et plus généralement (non présenté aujourd'hui) :

- ▶ Lorsque g est connue, on peut mettre en œuvre des stratégies de déconvolution *statistiquement* quasi-optimales.
- ▶ La régularité de la loi de déformation g est fondamentale et dégrade la qualité d'estimation. Vitesse en $(\log(n)/n)^{-2s/(2s+2\nu+1)}$.
- ▶ Avant tout procédé de recalage, il faut supprimer le bruit sur les observations (statistiquement, il y a consistance lorsque $\epsilon \rightarrow 0$).
- ▶ La pertinence des méthodes de recalage lorsque les fonctions sont non régulières reste à étudier en asymptotique $n \rightarrow +\infty$.

I - Modèles déformables en statistique

I - 1 Motivations

I - 2 Modèle statistique

I - 3 Fiabilité des méthodes de recalage

I - 4 Estimation par approche Bayésienne

I - 5 Conclusion sur le modèle SIM

II Diffusion sur une descente de gradient à mémoire

II - 1 Introduction au modèle de gradient moyenné

II - 2 Hypo-ellipticité

II - 3 Comportement asymptotique

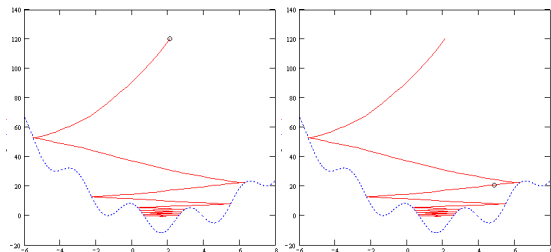
II - 4 Convergence lorsque $\sigma \rightarrow 0$

II - 5 Formulation variationnelle

II - 6 Conclusion

II - 1 Introduction au modèle de gradient moyenné

Comment permettre à une descente de gradient ceci ?



On s'intéresse au comportement en temps long de :



$$dx_t = - \left[\frac{1}{k(t)} \int_0^t k'(s) \nabla U(x_s) ds \right] dt, \quad (GM)$$

où U est un potentiel coercif strictement positif, $(x_t)_{t \geq 0} \in \mathbb{R}^d$.

▶ Ainsi qu'à :

$$dX_t = - \left[\frac{1}{k(t)} \int_0^t k'(s) \nabla U(X_s) ds \right] dt + \sigma dB_t. \quad (DM)$$

II - 1 Introduction au modèle de gradient moyenné

L'origine du modèle :

- ▶ La méthode de la boule pesante (cas particulier de (GM)) est utilisée en optimisation, elle est décrite par une O.D.E.

$$\ddot{x}_t + \lambda \dot{x}_t + \nabla U(x_t) = 0 \rightsquigarrow \dot{x}_t = -\lambda e^{-\lambda t} \int_0^t e^{\lambda s} \nabla(\lambda U)(x_s) ds$$

- ▶ Aménagement par hasard de Robbins-Monro : $(\mathbb{E}d_k(X_k) = \nabla U(X_k))$

$$X_{k+1} = X_k + \alpha_k d_k(X_k) + \sqrt{\alpha_k} \xi_k \rightsquigarrow dX_t = -\nabla U(X_t) dt + dB_t$$

en

$$X_{k+1} = X_k + \alpha_k \frac{\sum_{j \leq k} \omega_j d_j(X_j)}{\sum_{j \leq k} \omega_j} + \sqrt{\alpha_k} \xi_k \rightsquigarrow dX_t = - \left(\int_0^t \frac{k'(s)}{k(t)} \nabla U(X_s) ds \right) dt + dB_t$$

Les questions abordées pour cet exposé :

- ▶ Comportement en temps long de l'E.D.S. (stabilité, explosion, stationarité)
- ▶ Comportement lorsque $\sigma \rightarrow 0$

Les questions qui sont laissées de côté : principalement la situation déterministe (du moins en apparence).

II - 1 Introduction au modèle : propriétés d'existence

- ▶ U potentiel « coercif » positif et régulier.
- ▶ k croissante, positive, de classe \mathcal{C}^2 .
- ▶ On définit $(B_t)_{t \geq 0}$ un Brownien standard, et $(X_t)_{t \geq 0}$ solution de :

$$dX_t = - \underbrace{\left[\frac{1}{k(t)} \int_0^t k'(s) \nabla U(X_s) ds \right]}_{:= Y_t} dt + \sigma dB_t. \quad (DM)$$

- ▶ $(X_t)_{t \geq 0}$ n'a aucune chance d'être Markovien mais :

$$dX_t = -Y_t dt + \sigma dB_t \quad \text{et} \quad dY_t = \frac{k'(t)}{k(t)} [\nabla U(X_t) - Y_t] dt.$$

- ▶ Une utilisation de $\mathcal{L}_1(x, y, t) = U(x) + \frac{y^2}{2r(t)}$ avec $r = k'/k$ implique

Proposition

Si $|D^2(U)| = \mathcal{O}_{+\infty}(U)$, alors il y a existence forte du processus $(Z_t)_{t \geq 0} := (X_t, Y_t)_{t \geq 0}$ et non explosion en temps fini.

I - Modèles déformables en statistique

I - 1 Motivations

I - 2 Modèle statistique

I - 3 Fiabilité des méthodes de recalage

I - 4 Estimation par approche Bayésienne

I - 5 Conclusion sur le modèle SIM

II Diffusion sur une descente de gradient à mémoire

II - 1 Introduction au modèle de gradient moyenné

II - 2 Hypo-ellipticité

II - 3 Comportement asymptotique

II - 4 Convergence lorsque $\sigma \rightarrow 0$

II - 5 Formulation variationnelle

II - 6 Conclusion

II - 2 Hypo-ellipticité

La régularité du semi-groupe n'est pas évidente (dégénérescence de la partie Brownienne sur Y). On définit

$$\mathcal{E}_U := \left\{ x, \det(D^2 U(x)) = 0 \right\}$$

Proposition (G. & Panloup '12)

Si \mathcal{E}_U n'a que des points isolés, alors $(Z_t)_{t>0}$ a une densité $p_t \in C^\infty$ par rapport à $\mu_{\mathcal{L}, \mathbb{R}^d \times \mathbb{R}^d}$. De plus, si $U(x)/x \rightarrow_{+\infty} +\infty$, alors $\forall z, \text{Supp}(p_t(z, \cdot)) = \mathbb{R}^d \times \mathbb{R}^d$.

- ▶ La régularité provient de propriétés sur les crochets de Lie des champs de vecteur.
- ▶ La propriété de support provient principalement d'une propriété de contrôlabilité du système : $z_\varphi := (x_\varphi, y_\varphi)$ solution de

$$\dot{x}_\varphi(t) = -y_\varphi(t) + \varphi(t) \quad \text{et} \quad \dot{y}_\varphi(t) = r(t)[\nabla U(x_\varphi(t)) - y_\varphi(t)]$$

I - Modèles déformables en statistique

I - 1 Motivations

I - 2 Modèle statistique

I - 3 Fiabilité des méthodes de recalage

I - 4 Estimation par approche Bayésienne

I - 5 Conclusion sur le modèle SIM

II Diffusion sur une descente de gradient à mémoire

II - 1 Introduction au modèle de gradient moyenné

II - 2 Hypo-ellipticité

II - 3 Comportement asymptotique

II - 4 Convergence lorsque $\sigma \rightarrow 0$

II - 5 Formulation variationnelle

II - 6 Conclusion

II - 3 a) Comportement asymptotique, cas instable

Situation instable : $k'(t)/k(t) \rightarrow 0$ lorsque $t \rightarrow +\infty$.

- ▶ Ce contexte correspond aux situations du type $k(t) = t^n$. Mémoire à **long** terme.
- ▶ Il n'y a pas de propriété de compacité et même :

Theorem (G. & Panloup '12)

(i) Si $|\nabla U|^2 \leq C(1 + U)$, et $k(t) = t^n$, alors

$$\limsup_{t \rightarrow +\infty} \frac{|X_t|^2}{t} = +\infty.$$

(ii) Dans le cas Gaussien $U(x) = x^2/2$, on a même

$$\left(\sqrt{2n+1}t^{-1/2}X_t, \sqrt{(2n+1)/n}Y_t \right) \rightarrow \mathcal{N}(0, I_2).$$

- ▶ Pas de mesure stationnaire dans le cas gaussien. Processus à ne considérer qu'avec $\sigma(t) \mapsto 0$.

II - 3 b) Comportement asymptotique, cas stable

Situation stable : $k'(t)/k(t) \rightarrow \lambda > 0$ lorsque $t \rightarrow +\infty$.

- ▶ Ce contexte correspond aux situations du type $k(t) = e^{\lambda t}$. Mémoire à **court** terme.
- ▶ La fonction \mathcal{L}_1 n'est pas satisfaisante pour contrôler le système dynamique puisque $\mathcal{A}\mathcal{L}_1 \leq -cy^2 + \frac{1}{2}\text{Tr}(D^2U(x))$.
- ▶ Une modification pertinente pour les modèles cinétiques fait intervenir

$$\mathcal{L}(x, y, t) = \mathcal{L}_1(x, y, t) + m \left[x^2/2 - \frac{\langle x, y \rangle}{\rho(t)} \right]$$

puisque l'on obtient une "vraie" fonction de Lyapunov

$$\mathcal{A}\mathcal{L}^p \leq \beta - \alpha\mathcal{L}^q.$$

- ▶ On déduit alors l'existence, l'unicité d'une mesure invariante μ_σ .
- ▶ Vitesse de convergence pour W_2 ou VT exponentielle/polynomiale selon U . (pas de résultat en L_2 ou H .)

I - Modèles déformables en statistique

I - 1 Motivations

I - 2 Modèle statistique

I - 3 Fiabilité des méthodes de recalage

I - 4 Estimation par approche Bayésienne

I - 5 Conclusion sur le modèle SIM

II Diffusion sur une descente de gradient à mémoire

II - 1 Introduction au modèle de gradient moyenné

II - 2 Hypo-ellipticité

II - 3 Comportement asymptotique

II - 4 Convergence lorsque $\sigma \rightarrow 0$

II - 5 Formulation variationnelle

II - 6 Conclusion

II- 4 Convergence lorsque $\sigma \rightarrow 0$: potentiel non convexe « intéressant »

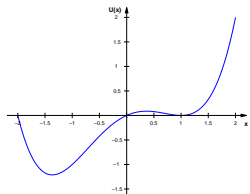
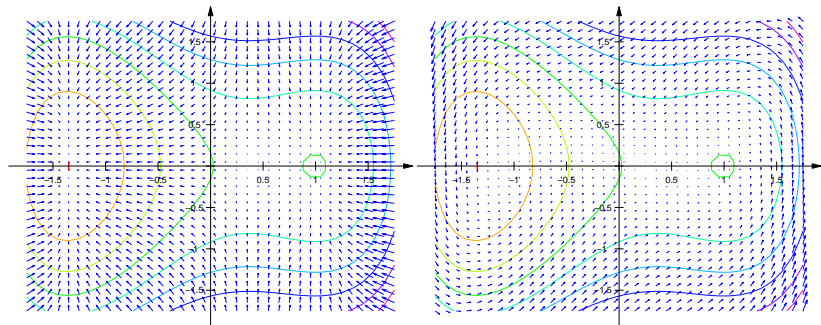


FIGURE: Potentiel $U = x^4/4 - 3x^2/2 + x/5$, Lignes de champ de $U(x) + y^2/2$ pour le cas standard et le cas à inertie.



II- 4 Convergence lorsque $\sigma \rightarrow 0$

- ▶ Dans le cas « standard » :

$$\mathcal{L}_K = -\nabla U(x) \partial_x + \sigma^2 \partial_x^2,$$

la distribution invariante est le champ de Gibbs $\nu_\sigma \propto e^{-U/(2\sigma^2)}$.

Méthode de Laplace \Rightarrow P.G.D. + $\nu_\sigma(\arg \min U) \rightarrow_{\sigma \rightarrow 0} 1$.

- ▶ Pour nous, (hormis le cas gaussien) pas de forme explicite pour μ_σ ☹.

On peut démontrer le théorème

Theorem (G. & Panloup & Pellegrini '12)

La suite de mesures (μ_σ) est exponentiellement tendue et lorsque les équilibres $(z_i)_i$ du champ sont discrets avec $D^2 U(z_i)$ inversible, (μ_σ) satisfait un P.G.D.

$$\lim_{\sigma \rightarrow 0} \sigma^2 \log[\mu_\sigma(x, y)] = -W(x, y)$$

où W est solution de l'équation d'Hamilton Jacobi

$$H(W, DW) := -\langle y, \partial_x W \rangle + \langle (U'(x) - y), \partial_y W \rangle + \frac{\{\partial_x W\}^2}{2} = 0.$$

I - Modèles déformables en statistique

I - 1 Motivations

I - 2 Modèle statistique

I - 3 Fiabilité des méthodes de recalage

I - 4 Estimation par approche Bayésienne

I - 5 Conclusion sur le modèle SIM

II Diffusion sur une descente de gradient à mémoire

II - 1 Introduction au modèle de gradient moyenné

II - 2 Hypo-ellipticité

II - 3 Comportement asymptotique

II - 4 Convergence lorsque $\sigma \rightarrow 0$

II - 5 Formulation variationnelle

II - 6 Conclusion

II- 5 Formulation variationnelle et conclusion

En fait, $(\mu_\sigma)_\sigma$ se concentre sur le minimum de la fonction de taux W . Ce minimum peut être écrit au travers du quasi-potentiel V :

$$V(z_i, z_j) = \inf_{\begin{cases} \varphi \in \mathbb{H}^1 \\ z_\varphi(0) = z_i, z_\varphi(\infty) = z_j \end{cases}} \left[\frac{1}{2} \int_0^{+\infty} \dot{\varphi}^2 \right],$$

Theorem (G. & Panloup & Pellegrini '12)

Sous les conditions précédentes, W est continue au voisinage des $(z_i)_i$, et

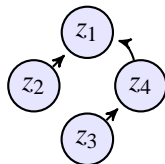
$$W(z_i) := \min_{g \in G(i)} \sum_{(z_k \rightarrow z_l) \in g} V(z_k, z_l) \quad \text{vérifie} \quad W(z_i) = \mathcal{W}(z_i) - \min_i \mathcal{W}(z_i).$$

Un z_1 -graphe de $G(z_1)$

z_1 est le nœud terminal

chaque nœud possède

une unique arête sortante



I - Modèles déformables en statistique

I - 1 Motivations

I - 2 Modèle statistique

I - 3 Fiabilité des méthodes de recalage

I - 4 Estimation par approche Bayésienne

I - 5 Conclusion sur le modèle SIM

II Diffusion sur une descente de gradient à mémoire

II - 1 Introduction au modèle de gradient moyenné

II - 2 Hypo-ellipticité

II - 3 Comportement asymptotique

II - 4 Convergence lorsque $\sigma \rightarrow 0$

II - 5 Formulation variationnelle

II - 6 Conclusion

II - 6 Conclusion

- ▶ Bonne connaissance du système déterministe (GM).
- ▶ Pour la diffusion (DM)
 - ▶ $k(t) = e^{\lambda t}$ Comportement stable mais non standard.
 - ▶ Si $k(t) = e^{e^t}$ (non exposé aujourd'hui), comportement comparable à la diffusion

$$dS_t = -\nabla U(S_t)dt + \sigma(S_t)dB_t.$$

- ▶ Diffusion à petits paramètres
 - ▶ Existence d'un PGD
 - ▶ Description précise des minimiseurs (non exposé aujourd'hui).
- ▶ Élargissements :
- ▶ Coupler $\sigma_t \rightarrow 0$ avec la vitesse L^2 , application dédiée à l'optimisation.
- ▶ Généraliser à des mémoires différentes :

$$dX_t = -\frac{1}{Z_t} \left[\int_{a(t)}^{b(t)} \nabla U(X_s) ds \right] dt + \sigma(X_s)dB_s.$$

- ▶ Borne inférieure de vitesses inconnues.
- ▶ Du point de vue de la modélisation, étudier des systèmes à mémoire ayant une écriture comparable (Bulles financières, dynamique des populations, ...)

Merci de votre attention!