

Stabilité des chaînes de Markov par la technique des limites fluides. Applications MCMC

Gersende FORT

LTCI
CNRS - TELECOM ParisTech

En collaboration avec Sean MEYN (Univ. Illinois), Eric MOULINES (TELECOM ParisTech) et Pierre PRIOURET (Univ. Paris 6).

Nous introduisons

- ▶ une transformation d'une Chaîne de Markov \longrightarrow famille de processus à temps continu \longrightarrow un processus (limite) à temps continu
- ▶ tq *la stabilité* de ce processus, est liée à *l'ergodicité* de la chaîne de Markov.
 - \Rightarrow caractériser l'ergodicité;
 - \Rightarrow identifier les facteurs responsables de la dynamique de la Chaîne de Markov.

Nous introduisons

- ▶ une transformation d'une Chaîne de Markov \longrightarrow famille de processus à temps continu \longrightarrow un processus (limite) à temps continu
- ▶ tq *la stabilité* de ce processus, est liée à *l'ergodicité* de la chaîne de Markov.
 - \Rightarrow caractériser l'ergodicité;
 - \Rightarrow identifier les facteurs responsables de la dynamique de la Chaîne de Markov.

Les algorithmes *Markov Chain Monte Carlo* (MCMC)

- ▶ sont des algorithmes itératifs qui produisent une chaîne de Markov de loi stationnaire donnée;
- ▶ et dont les performances sont liées (*entre autre*) à des paramètres de mise en oeuvre.
- ▶ \Rightarrow identifier le rôle de ces paramètres dans la définition des *limites fluides* et proposer un "choix optimal" de ces paramètres.

↔ Plan de l'exposé

I. Méthodes MCMC

II. Limites fluides

III. Echantillonneur Metropolis-within-Gibbs adaptatif

IV. Conclusion

Usage des méthodes MCMC :

Pour une loi donnée π , produire une Chaîne de Markov $\{\Phi_n, n \geq 0\}$ admettant π pour unique loi invariante.

↔ Permettent

- ▶ d'explorer la densité cible π .
- ▶ d'approcher des espérances $\mathbb{E}_\pi[g(\Phi)]$ dès lors que la chaîne est suffisamment régulière pour justifier LGN (et autres théos limites).

↔ Nécessaires dès que

- (i) π est trop complexe et tirages i.i.d. de loi π impossibles.
- (ii) (ne nécessitent que) π connu à une constante de normalisation près.

↔ Algorithmes : Hastings-Metropolis, Gibbs, ...

Echantillonneurs de Hastings-Metropolis

- ▶ Choisir un noyau de transition $q(x,y)$ ex. $q(x,y) = \mathcal{N}(x,\Sigma)[y]$
- ▶ Itérativement :
 - proposer une nouvelle valeur $Y \sim q(\Phi_n, \cdot)$.
 - accepter / rejeter le candidat :

$$\Phi_{n+1} = \begin{cases} Y & \text{avec proba} \\ \Phi_n & \text{sinon.} \end{cases} \quad \alpha(\Phi_n, Y) = 1 \wedge \frac{\pi(Y)}{\pi(\Phi_n)} \frac{q(Y, \Phi_n)}{q(\Phi_n, Y)}$$

Exemple :

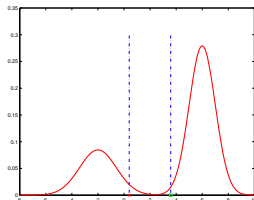
HM à marche aléatoire symétrique

$$Y = \Phi_n + \bullet \quad \bullet \sim \text{loi symétrique} \quad \text{donc } \alpha(\Phi_n, Y) = \frac{\pi(Y)}{\pi(\Phi_n)} \wedge 1.$$

- Explorer la loi sur \mathbb{R}^2

$$\pi = \alpha \mathcal{N}(\mu_1, \sigma_1^2) + (1 - \alpha) \mathcal{N}(\mu_2, \sigma_2^2)$$

en proposant des accroissements $\mathcal{N}(0, \sigma^2)$.



Exemple :

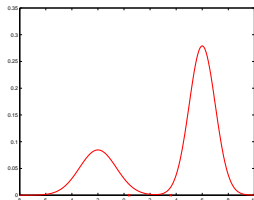
HM à marche aléatoire symétrique

$$Y = \Phi_n + \bullet \quad \bullet \sim \text{loi symétrique} \quad \text{donc } \alpha(\Phi_n, Y) = \frac{\pi(Y)}{\pi(\Phi_n)} \wedge 1.$$

- Explorer la loi sur \mathbb{R}^2

$$\pi = \alpha \mathcal{N}(\mu_1, \sigma_1^2) + (1 - \alpha) \mathcal{N}(\mu_2, \sigma_2^2)$$

en proposant des accroissements $\mathcal{N}(0, \sigma^2)$.



Exemple :

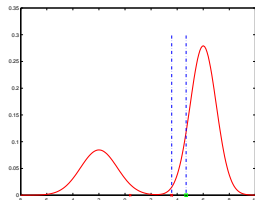
HM à marche aléatoire symétrique

$$Y = \Phi_n + \bullet \quad \bullet \sim \text{loi symétrique} \quad \text{donc } \alpha(\Phi_n, Y) = \frac{\pi(Y)}{\pi(\Phi_n)} \wedge 1.$$

- Explorer la loi sur \mathbb{R}^2

$$\pi = \alpha \mathcal{N}(\mu_1, \sigma_1^2) + (1 - \alpha) \mathcal{N}(\mu_2, \sigma_2^2)$$

en proposant des accroissements $\mathcal{N}(0, \sigma^2)$.



Exemple :

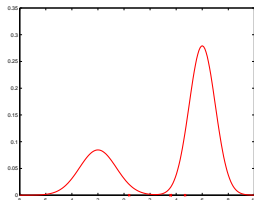
HM à marche aléatoire symétrique

$$Y = \Phi_n + \bullet \quad \bullet \sim \text{loi symétrique} \quad \text{donc } \alpha(\Phi_n, Y) = \frac{\pi(Y)}{\pi(\Phi_n)} \wedge 1.$$

- Explorer la loi sur \mathbb{R}^2

$$\pi = \alpha \mathcal{N}(\mu_1, \sigma_1^2) + (1 - \alpha) \mathcal{N}(\mu_2, \sigma_2^2)$$

en proposant des accroissements $\mathcal{N}(0, \sigma^2)$.



Exemple :

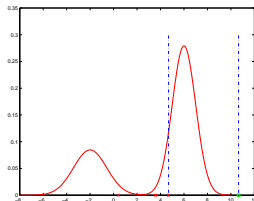
HM à marche aléatoire symétrique

$$Y = \Phi_n + \bullet \quad \bullet \sim \text{loi symétrique} \quad \text{donc } \alpha(\Phi_n, Y) = \frac{\pi(Y)}{\pi(\Phi_n)} \wedge 1.$$

- Explorer la loi sur \mathbb{R}^2

$$\pi = \alpha \mathcal{N}(\mu_1, \sigma_1^2) + (1 - \alpha) \mathcal{N}(\mu_2, \sigma_2^2)$$

en proposant des accroissements $\mathcal{N}(0, \sigma^2)$.



Exemple :

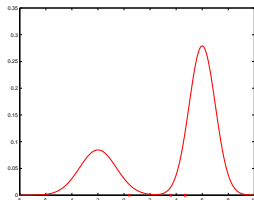
HM à marche aléatoire symétrique

$$Y = \Phi_n + \bullet \quad \bullet \sim \text{loi symétrique} \quad \text{donc } \alpha(\Phi_n, Y) = \frac{\pi(Y)}{\pi(\Phi_n)} \wedge 1.$$

- Explorer la loi sur \mathbb{R}^2

$$\pi = \alpha \mathcal{N}(\mu_1, \sigma_1^2) + (1 - \alpha) \mathcal{N}(\mu_2, \sigma_2^2)$$

en proposant des accroissements $\mathcal{N}(0, \sigma^2)$.



Exemple :

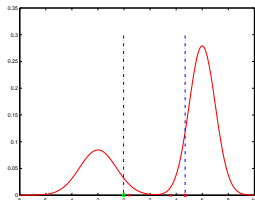
HM à marche aléatoire symétrique

$$Y = \Phi_n + \bullet \quad \bullet \sim \text{loi symétrique} \quad \text{donc } \alpha(\Phi_n, Y) = \frac{\pi(Y)}{\pi(\Phi_n)} \wedge 1.$$

- Explorer la loi sur \mathbb{R}^2

$$\pi = \alpha \mathcal{N}(\mu_1, \sigma_1^2) + (1 - \alpha) \mathcal{N}(\mu_2, \sigma_2^2)$$

en proposant des accroissements $\mathcal{N}(0, \sigma^2)$.



Exemple :

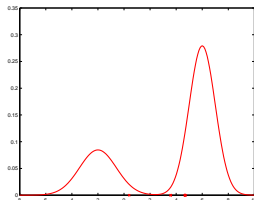
HM à marche aléatoire symétrique

$$Y = \Phi_n + \bullet \quad \bullet \sim \text{loi symétrique} \quad \text{donc } \alpha(\Phi_n, Y) = \frac{\pi(Y)}{\pi(\Phi_n)} \wedge 1.$$

- Explorer la loi sur \mathbb{R}^2

$$\pi = \alpha \mathcal{N}(\mu_1, \sigma_1^2) + (1 - \alpha) \mathcal{N}(\mu_2, \sigma_2^2)$$

en proposant des accroissements $\mathcal{N}(0, \sigma^2)$.



Exemple :

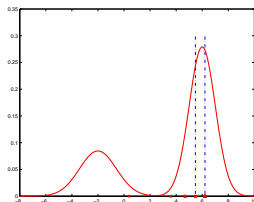
HM à marche aléatoire symétrique

$$Y = \Phi_n + \bullet \quad \bullet \sim \text{loi symétrique} \quad \text{donc } \alpha(\Phi_n, Y) = \frac{\pi(Y)}{\pi(\Phi_n)} \wedge 1.$$

- Explorer la loi sur \mathbb{R}^2

$$\pi = \alpha \mathcal{N}(\mu_1, \sigma_1^2) + (1 - \alpha) \mathcal{N}(\mu_2, \sigma_2^2)$$

en proposant des accroissements $\mathcal{N}(0, \sigma^2)$.



Exemple :

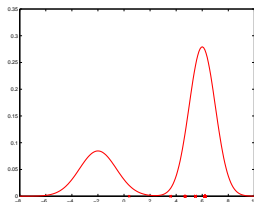
HM à marche aléatoire symétrique

$$Y = \Phi_n + \bullet \quad \bullet \sim \text{loi symétrique} \quad \text{donc } \alpha(\Phi_n, Y) = \frac{\pi(Y)}{\pi(\Phi_n)} \wedge 1.$$

- Explorer la loi sur \mathbb{R}^2

$$\pi = \alpha \mathcal{N}(\mu_1, \sigma_1^2) + (1 - \alpha) \mathcal{N}(\mu_2, \sigma_2^2)$$

en proposant des accroissements $\mathcal{N}(0, \sigma^2)$.



Exemple :

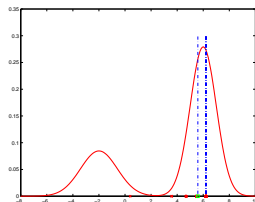
HM à marche aléatoire symétrique

$$Y = \Phi_n + \bullet \quad \bullet \sim \text{loi symétrique} \quad \text{donc } \alpha(\Phi_n, Y) = \frac{\pi(Y)}{\pi(\Phi_n)} \wedge 1.$$

- Explorer la loi sur \mathbb{R}^2

$$\pi = \alpha \mathcal{N}(\mu_1, \sigma_1^2) + (1 - \alpha) \mathcal{N}(\mu_2, \sigma_2^2)$$

en proposant des accroissements $\mathcal{N}(0, \sigma^2)$.



Exemple :

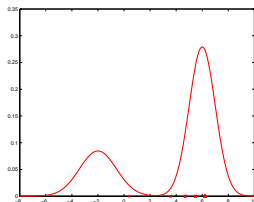
HM à marche aléatoire symétrique

$$Y = \Phi_n + \bullet \quad \bullet \sim \text{loi symétrique} \quad \text{donc } \alpha(\Phi_n, Y) = \frac{\pi(Y)}{\pi(\Phi_n)} \wedge 1.$$

- Explorer la loi sur \mathbb{R}^2

$$\pi = \alpha \mathcal{N}(\mu_1, \sigma_1^2) + (1 - \alpha) \mathcal{N}(\mu_2, \sigma_2^2)$$

en proposant des accroissements $\mathcal{N}(0, \sigma^2)$.



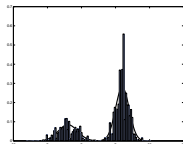
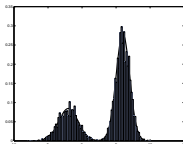
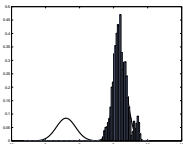
Autres échantillonneurs :

- ▶ **Echantillonneur de Gibbs** : Mettre à jour composante par composante la chaîne en proposant
$$\Phi_{n+1,k} \sim \pi(\cdot | (\Phi_{n,1:(k-1)}, \Phi_{n+1,(k+1):d})).$$
- ▶ **Echantillonneur Metropolis-within-Gibbs** :
 - ▶ Se donner
 - une procédure de sélection de la direction de mise à jour $\{\omega_i, i \in \{1, \dots, d\}\}$.
 - des lois de proposition sur \mathbb{R} dans chaque direction.
 - ▶ A chaque itération :
 - choisir aléatoirement LA composante mise à jour $\mathbb{P}(I = k) = \omega_k$.
 - faire une mise à jour type Hastings-Metropolis dans la direction sélectionnée.

Choix des paramètres de mise en oeuvre

l'efficacité des échantillonneurs dépend des paramètres de mise en oeuvre. par ex.

- ▶ pour HM : choix des lois de proposition (ex. : choix de la structure de cov quand $q(x, \cdot) \sim \mathcal{N}(x, \gamma)$).
- ▶ pour MwG : procédure de sélection des composantes $\{\omega_i, i \leq d\}$, et lois de proposition dans chaque direction.



↪ Questions ouvertes

- ▶ Valeur optimale de ces paramètres.
- ▶ Méthodes adaptatives : correction en ligne de ces paramètres basée sur le comportement passé de l'algorithme.

↪ Dans la suite

- ▶ identifier le rôle de ces paramètres de mise en oeuvre dans la dynamique de la chaîne;
- ▶ en déduire une définition de ces paramètres.

II. Limites fluides

Processus renormalisés

Soit $\{\Phi_k, k \geq 0\}$ une chaîne de Markov à valeur X ($X = \mathbb{R}^d$).

Famille de transformations : processus renormalisés η_r , pour $r > 0$

(i) en le point initial :

$$\eta_r(0; x) = \frac{1}{r} \Phi_0 = x, \quad \Phi_0 = rx$$

(ii) en temps et en espace :

$$\eta_r(t; x) = \frac{1}{r} \Phi_{\lfloor tr \rfloor}.$$

Processus renormalisés

Soit $\{\Phi_k, k \geq 0\}$ une chaîne de Markov à valeur X ($X = \mathbb{R}^d$).

Famille de transformations : processus renormalisés η_r , pour $r > 0$

(i) en le point initial :

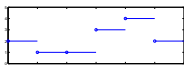
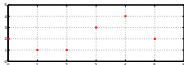
$$\eta_r(0; x) = \frac{1}{r} \Phi_0 = x, \quad \Phi_0 = rx$$

(ii) en temps et en espace :

$$\eta_r(t; x) = \frac{1}{r} \Phi_{\lfloor tr \rfloor}.$$

Donc $\eta_r(\cdot; x) = \frac{1}{r} \Phi_k$ sur l'intervalle de temps $\left[\frac{k}{r}; \frac{(k+1)}{r} \right)$.

Par construction, trajectoires cad-lag.



Définition

↔ Distributions

- \mathbb{P}_x : distribution de la chaîne $\{\Phi_k, k \geq 0\}$ de loi initiale δ_x .
- $\mathbb{Q}_{r;x}$: prob. image de \mathbb{P}_{rx} par $\eta_r(\cdot; x)$,
probabilité sur l'espace $\mathbb{D}(\mathbb{R}^+, \mathbf{X})$ des fonctions cad-lag $\mathbb{R}^+ \rightarrow \mathbf{X}$

Définition

↔ Distributions

- \mathbb{P}_x : distribution de la chaîne $\{\Phi_k, k \geq 0\}$ de loi initiale δ_x .
- $\mathbb{Q}_{r;x}$: prob. image de \mathbb{P}_{rx} par $\eta_r(\cdot; x)$,
probabilité sur l'espace $\mathbb{D}(\mathbb{R}^+, \mathbf{X})$ des fonctions cad-lag $\mathbb{R}^+ \rightarrow \mathbf{X}$

↔ **Définition Limite Fluide.** \mathbb{Q} probabilité sur $\mathbb{D}(\mathbb{R}^+, \mathbf{X})$ est une **limite fluide** si il existe une famille de facteurs d'échelles $r_n \rightarrow +\infty$ tq

$$\mathbb{Q}_{r_n;x} \implies \mathbb{Q}.$$

Notée \mathbb{Q}_x ci-après.

Définition

↔ Distributions

- \mathbb{P}_x : distribution de la chaîne $\{\Phi_k, k \geq 0\}$ de loi initiale δ_x .
- $\mathbb{Q}_{r;x}$: prob. image de \mathbb{P}_{rx} par $\eta_r(\cdot; x)$,
probabilité sur l'espace $\mathbb{D}(\mathbb{R}^+, X)$ des fonctions cad-lag $\mathbb{R}^+ \rightarrow X$

↔ **Définition Limite Fluide.** \mathbb{Q} probabilité sur $\mathbb{D}(\mathbb{R}^+, X)$ est une **limite fluide** si il existe une famille de facteurs d'échelles $r_n \rightarrow +\infty$ tq

$$\mathbb{Q}_{r_n;x} \implies \mathbb{Q}.$$

Notée \mathbb{Q}_x ci-après.

↔ **Rque :** limite fluide $\leftrightarrow \lim_r \mathbb{Q}_{r,x}$ et $\mathbb{Q}_{r,x}$ image de $\mathbb{P}_{rx} \leftrightarrow$ comportement des chaînes initialisées loin dans les queues de π .

Exemple

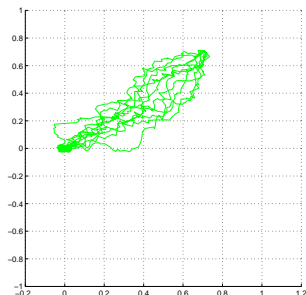
$\{\Phi_n, n \geq 0\}$ chaîne de Hastings-Metropolis lorsque la loi cible sur \mathbb{R}^2 est donnée par

$$\pi(x_1, x_2) \propto (1 + x_1^2 + x_2^2 + x_1^8 x_2^2) \exp(-(x_1^2 + x_2^2))$$

et la loi de proposition est $4 \mathcal{N}_2(x, \mathbb{I})$.

Figures : Différentes réalisations du processus normalisé $\eta_r(\cdot, x)$ sur $[0, T]$; puis pour différents points initiaux x ; puis pour différents facteurs d'échelle r .

un point initial



Exemple

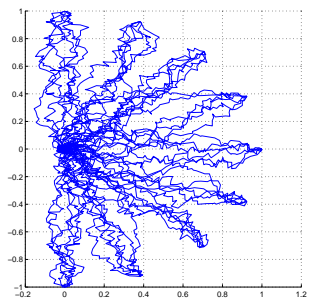
$\{\Phi_n, n \geq 0\}$ chaîne de Hastings-Metropolis lorsque la loi cible sur \mathbb{R}^2 est donnée par

$$\pi(x_1, x_2) \propto (1 + x_1^2 + x_2^2 + x_1^8 x_2^2) \exp(-(x_1^2 + x_2^2))$$

et la loi de proposition est $4 \mathcal{N}_2(x, \mathbb{I})$.

Figures : Différentes réalisations du processus normalisé $\eta_r(\cdot, x)$ sur $[0, T]$; puis pour différents points initiaux x ; puis pour différents facteurs d'échelle r .

un point initial, différents points initiaux ($r = 100$)



Exemple

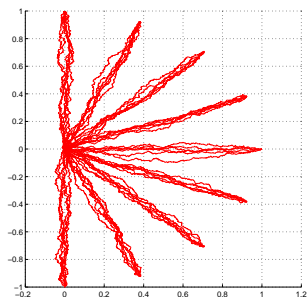
$\{\Phi_n, n \geq 0\}$ chaîne de Hastings-Metropolis lorsque la loi cible sur \mathbb{R}^2 est donnée par

$$\pi(x_1, x_2) \propto (1 + x_1^2 + x_2^2 + x_1^8 x_2^2) \exp(-(x_1^2 + x_2^2))$$

et la loi de proposition est $4 \mathcal{N}_2(x, \mathbb{I})$.

Figures : Différentes réalisations du processus normalisé $\eta_r(\cdot, x)$ sur $[0, T]$; puis pour différents points initiaux x ; puis pour différents facteurs d'échelle r .

un point initial, différents points initiaux ($r = 100$), différentes valeurs du scaling r ($r = 1000$)



Exemple

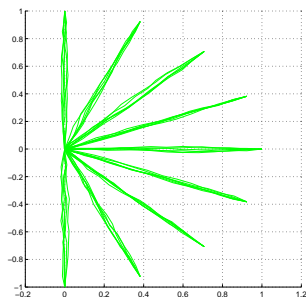
$\{\Phi_n, n \geq 0\}$ chaîne de Hastings-Metropolis lorsque la loi cible sur \mathbb{R}^2 est donnée par

$$\pi(x_1, x_2) \propto (1 + x_1^2 + x_2^2 + x_1^8 x_2^2) \exp(-(x_1^2 + x_2^2))$$

et la loi de proposition est $4 \mathcal{N}_2(x, \mathbb{I})$.

Figures : Différentes réalisations du processus normalisé $\eta_r(\cdot, x)$ sur $[0, T]$; puis pour différents points initiaux x ; puis pour différents facteurs d'échelle r .

un point initial, différents points initiaux ($r = 100$), différentes valeurs du scaling r ($r = 1000$) ($r = 5000$)



Exemple

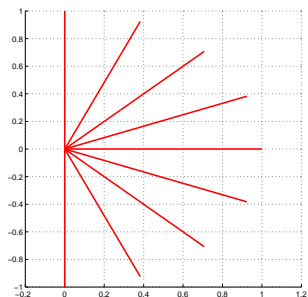
$\{\Phi_n, n \geq 0\}$ chaîne de Hastings-Metropolis lorsque la loi cible sur \mathbb{R}^2 est donnée par

$$\pi(x_1, x_2) \propto (1 + x_1^2 + x_2^2 + x_1^8 x_2^2) \exp(-(x_1^2 + x_2^2))$$

et la loi de proposition est $4 \mathcal{N}_2(x, \mathbb{I})$.

Figures : Différentes réalisations du processus normalisé $\eta_r(\cdot, x)$ sur $[0, T]$; puis pour différents points initiaux x ; puis pour différents facteurs d'échelle r .

un point initial, différents points initiaux ($r = 100$), différentes valeurs du scaling r ($r = 1000$) ($r = 5000$) (Limite Fluide)



CS d'Existence

$$\begin{aligned}\Phi_{k+1} &= \Phi_k + \mathbb{E}[\Phi_{k+1}|\mathcal{F}_k] - \Phi_k + \Phi_{k+1} - \mathbb{E}[\Phi_{k+1}|\mathcal{F}_k] \\ &= \Phi_k + \underbrace{\mathbb{E}_x[\Phi_{k+1} - \Phi_k|\mathcal{F}_k]}_{\Delta(\Phi_k)} + \underbrace{(\Phi_{k+1} - \mathbb{E}_x[\Phi_{k+1}|\mathcal{F}_k])}_{\epsilon_{k+1} \quad \text{incrément de martingale}}.\end{aligned}$$

CS d'Existence

$$\begin{aligned}
 \Phi_{k+1} &= \Phi_k + \mathbb{E}[\Phi_{k+1} | \mathcal{F}_k] - \Phi_k + \Phi_{k+1} - \mathbb{E}[\Phi_{k+1} | \mathcal{F}_k] \\
 &= \Phi_k + \underbrace{\mathbb{E}_x[\Phi_{k+1} - \Phi_k | \mathcal{F}_k]}_{\Delta(\Phi_k)} + \underbrace{(\Phi_{k+1} - \mathbb{E}_x[\Phi_{k+1} | \mathcal{F}_k])}_{\epsilon_{k+1} \text{ incrément de martingale}}.
 \end{aligned}$$

► Résultat (Fort et al, 2007)

Si

- $\exists p > 1, \quad \lim_{K \rightarrow +\infty} \sup_{x \in X} \mathbb{E}_x [|\epsilon_1|^p \mathbf{1}_{|\epsilon_1| > K}] \rightarrow 0.$
- $0 < \sup_{x \in X} |\Delta(x)| < \infty.$

Alors $\forall x$

- $\forall r_n \rightarrow +\infty, \exists$ sous-suite $\{r_{n_j}, j \geq 1\}$ tq $\mathbb{Q}_{r_{n_j}; x} \Rightarrow \mathbb{Q}_x$
- \mathbb{Q}_x prob. sur l'espace des fonctions continues de $\mathbb{R}^+ \rightarrow X.$

Stabilité des Limites Fluides

↔ *Définition Modèle Fluide stable*: il existe $T > 0$ et $\rho < 1$ tq pour tout x sur la sphère unité,

$$\mathbb{Q}_x \left(\eta \in \mathbb{D}(\mathbb{R}^+, \mathbf{X}), \inf_{[0, T]} |\eta(t)| \leq \rho \right) = 1.$$

Stabilité des Limites Fluides

↔ *Définition Modèle Fluide stable*: il existe $T > 0$ et $\rho < 1$ tq pour tout x sur la sphère unité,

$$\mathbb{Q}_x \left(\eta \in \mathbb{D}(\mathbb{R}^+, \mathbf{X}), \inf_{[0, T]} |\eta(t)| \leq \rho \right) = 1.$$

↔ *Stabilité du modèle fluide et stabilité de la chaîne initiale*: on montre que la stabilité du modèle fluide entraîne une condition de drift pour la chaîne de noyau $\bar{P}(x, \cdot) := \mathbb{P}_x(\Phi_{\tau_{\Phi_0}} \in \cdot)$

$$\tau_{\Phi_0} := \lfloor T|\Phi_0| \rfloor \wedge \sigma, \quad \sigma := \inf \{k, |\Phi_k| \leq \rho|\Phi_0|\}.$$

On en déduit l'ergodicité de la chaîne initiale.

Résultat (★★★★) (Fort et al, 2007)

Si

- $\{\Phi_k, k \geq 0\}$ est phi-irréductible apériodique; et que les ensembles compact sont petites.
- le modèle fluide existe et est stable

Alors la chaîne de Markov est (f, r) -ergodique

$$(n+1)^{q-1} \sup_{\{f, |f| \leq 1+|x|^{p-q}\}} |\mathbb{E}_x[f(\Phi_n)] - \pi(f)| \xrightarrow{n \rightarrow +\infty} 0, \quad 1 \leq q \leq p.$$

Résultat (★★★★) (Fort et al, 2007)

Si

- $\{\Phi_k, k \geq 0\}$ est phi-irréductible apériodique; et que les ensembles compact sont petites.
- le modèle fluide existe et est stable

Alors la chaîne de Markov est (f, r) -ergodique

$$(n+1)^{q-1} \sup_{\{f, |f| \leq 1+|x|^{p-q}\}} |\mathbb{E}_x[f(\Phi_n)] - \pi(f)| \xrightarrow{n \rightarrow +\infty} 0, \quad 1 \leq q \leq p.$$

p : contrôle du terme martingale dans la décomposition de la chaîne

$$\Phi_{n+1} - \Phi_n = \Delta(\Phi_n) + \text{incrément de martingale.}$$

Le temps T d'atteinte de la boule de rayon ρ par le modèle fluide joue un rôle sur le contrôle de la convergence de $P^n(x, \cdot)$ vers π . (rôle du retour vers le "centre")

Limite fluide = Squelette de la chaîne

$$\Phi_{k+1} = \Phi_k + \underbrace{(\mathbb{E}_x [\Phi_{k+1} | \mathcal{F}_k] - \Phi_k)}_{\Delta(\Phi_k)} + \underbrace{(\Phi_{k+1} - \mathbb{E}_x [\Phi_{k+1} | \mathcal{F}_k])}_{\epsilon_{k+1} \text{ incrément de martingale}}$$

► Pour le processus normalisé (constant par morceaux, saut aux instants k/r):

$$\begin{aligned} \eta_r \left[\frac{k+1}{r}, x \right] &= \frac{1}{r} \Phi_{k+1} = \eta_r \left[\frac{k}{r}, x \right] + \frac{1}{r} \Delta \left(r \eta_r \left[\frac{k}{r}, x \right] \right) + \frac{1}{r} \epsilon_{k+1} \\ &= \eta_r \left[\frac{k}{r}, x \right] + \frac{1}{r} h \left(\eta_r \left[\frac{k}{r}, x \right] \right) + \frac{1}{r} (\xi_k + \epsilon_{k+1}) \end{aligned}$$

en ayant posé

$$h(x) = \lim_{r \rightarrow +\infty} \Delta(r x).$$

Limite fluide = Squelette de la chaîne

$$\Phi_{k+1} = \Phi_k + \underbrace{(\mathbb{E}_x [\Phi_{k+1} | \mathcal{F}_k] - \Phi_k)}_{\Delta(\Phi_k)} + \underbrace{(\Phi_{k+1} - \mathbb{E}_x [\Phi_{k+1} | \mathcal{F}_k])}_{\epsilon_{k+1} \text{ incrément de martingale}}$$

► Pour le processus normalisé (constant par morceaux, saut aux instants k/r):

$$\begin{aligned} \eta_r \left[\frac{k+1}{r}, x \right] &= \frac{1}{r} \Phi_{k+1} = \eta_r \left[\frac{k}{r}, x \right] + \frac{1}{r} \Delta \left(r \eta_r \left[\frac{k}{r}, x \right] \right) + \frac{1}{r} \epsilon_{k+1} \\ &= \eta_r \left[\frac{k}{r}, x \right] + \frac{1}{r} h \left(\eta_r \left[\frac{k}{r}, x \right] \right) + \frac{1}{r} (\xi_k + \epsilon_{k+1}) \end{aligned}$$

en ayant posé

$$h(x) = \lim_{r \rightarrow +\infty} \Delta(r x).$$

► Ainsi **version bruitée de**

$$\mu \left(\frac{k+1}{r} \right) = \mu \left(\frac{k}{r} \right) + \frac{1}{r} h \left(\mu \left(\frac{k}{r} \right) \right) \longleftrightarrow \text{ODE : } \dot{\mu}(t) = h(\mu(t))$$

En fait, limites fluides caractérisées par

$$\lim_{r \rightarrow +\infty} \sup_{x \in H} |\Delta(rx) - h(x)| = 0,$$

pour tout compact $H \subset \mathbb{R}^d$?

En fait, limites fluides caractérisées par

$$\lim_{r \rightarrow +\infty} \sup_{x \in H} |\Delta(rx) - h(x)| = 0,$$

pour tout compact $H \subset ?$

- ▶ Dans les cas “favorables” ($? = X$), les limites fluides sont des distributions dégénérées : masses de Dirac en une fonction μ solution de l'ODE $\dot{\mu} = h(\mu)$.
Stabilité du modèle fluide \longleftrightarrow Stabilité de l'ODE.
- ▶ Sinon, situations plus complexes, pas de résultats généraux.

Caractérisation : cas 1

► Si

- $\exists h$ continue tq pour tout compact H de $X \setminus \{0\}$,

$$\lim_{r \rightarrow +\infty} \sup_{x \in H} |\Delta(rx) - h(x)| = 0.$$

- l' ODE $\dot{\mu} = h(\mu)$ est stable pour tout point initial x .

Alors Le modèle fluide est stable.

Caractérisation : cas 1

► Si

- $\exists h$ continue tq pour tout compact H de $X \setminus \{0\}$,

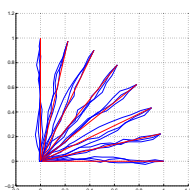
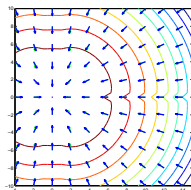
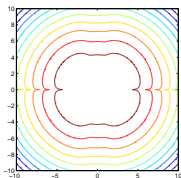
$$\lim_{r \rightarrow +\infty} \sup_{x \in H} |\Delta(rx) - h(x)| = 0.$$

- l'ODE $\dot{\mu} = h(\mu)$ est stable pour tout point initial x .

Alors Le modèle fluide est stable.

► Exemple : Hastings-Metropolis

$$\pi(x_1, x_2) \propto (1 + x_1^2 + x_2^2 + x_1^8 x_2^2) \exp(-(x_1^2 + x_2^2))$$



Courbes de niveau de π et champs Δ, h et réalisations de la limite fluide

Caractérisation : cas 2

► Si

- $\exists h$ continue tq pour tout compact H d' un cône de $X \setminus \{0\}$,

$$\lim_{r \rightarrow +\infty} \sup_{x \in H} |\Delta(rx) - h(x)| = 0.$$

- les ODE $\dot{\mu} = h(\mu)$ issues d'un point du cône sont stables.
- le cône est " attractif".

Alors Le modèle fluide est stable.

Caractérisation : cas 2

► Si

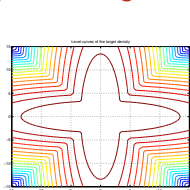
- $\exists h$ continue tq pour tout compact H d' un cône de $X \setminus \{0\}$,

$$\lim_{r \rightarrow +\infty} \sup_{x \in H} |\Delta(rx) - h(x)| = 0.$$

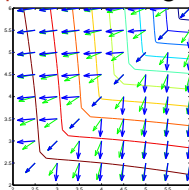
- les ODE $\dot{\mu} = h(\mu)$ issues d'un point du cône sont stables.
- le cône est " attractif".

Alors Le modèle fluide est stable.

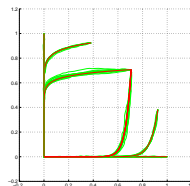
► Exemple : Hastings-Metropolis. π mélange de gaussiennes



Courbes de niveau de π et



champs Δ, h et



réalisations de limites fluides

Caractérisation : cas 3

$(X = \mathbb{R}^2)$

▶ Si

- $X = \bigcup_{\alpha=1}^a O_{\alpha} \cup \bigcup_{\beta=1}^b \{x, f'_{\beta}x = 0\}$.
- $\exists \Sigma_{\alpha}$ tq pour tout compact H de O_{α} ,

$$\lim_{r \rightarrow +\infty} \sup_{x \in H} |\Delta(rx) - \Sigma_{\alpha}| = 0.$$

- hyperplans sont “attractifs” et “stables”.

Alors le modèle fluide est stable.

Caractérisation : cas 3 ($X = \mathbb{R}^2$)

► Si

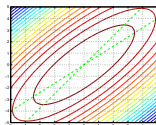
- $X = \bigcup_{\alpha=1}^a O_{\alpha} \cup \bigcup_{\beta=1}^b \{x, f'_{\beta}x = 0\}$.
- $\exists \Sigma_{\alpha}$ tq pour tout compact H de O_{α} ,

$$\lim_{r \rightarrow +\infty} \sup_{x \in H} |\Delta(rx) - \Sigma_{\alpha}| = 0.$$

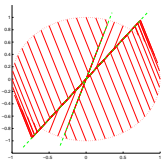
- hyperplans sont “attractifs” et “stables”.

Alors le modèle fluide est stable.

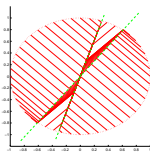
► Exemple : Metropolis within Gibbs



Courbes de niveau de π et



limites fluides quand $\omega_1 = 0.25$



et limites fluides quand $\omega_1 = 0.5$

Conclusion

- ▶ Par renormalisation de la chaîne :
 - ▶ le modèle fluide caractérise le comportement de la chaîne initialisée “loin dans les queues” $\Phi_0 = rx$ et $r \rightarrow +\infty$.
 - ▶ on extrait le comportement déterministe sous-jacent en éliminant les perturbations stochastiques.

Conclusion

- ▶ Par renormalisation de la chaîne :
 - ▶ le modèle fluide caractérise le comportement de la chaîne initialisée “loin dans les queues” $\Phi_0 = rx$ et $r \rightarrow +\infty$.
 - ▶ on extrait le comportement déterministe sous-jacent en éliminant les perturbations stochastiques.
- ▶ L’ergodicité de la chaîne initiale est liée à la stabilité du modèle fluide.
- ▶ Modèle fluide caractérisé (presque partout) par une ODE.

Conclusion

- ▶ Par renormalisation de la chaîne :
 - ▶ le modèle fluide caractérise le comportement de la chaîne initialisée “loin dans les queues” $\Phi_0 = rx$ et $r \rightarrow +\infty$.
 - ▶ on extrait le comportement déterministe sous-jacent en éliminant les perturbations stochastiques.
- ▶ L’ergodicité de la chaîne initiale est liée à la stabilité du modèle fluide.
- ▶ Modèle fluide caractérisé (presque partout) par une ODE.
- ▶ La limite fluide apporte des informations sur la dynamique de la chaîne dans le régime **transitoire** (i.e. avant régime stationnaire).
- ▶ Mais au prix d’un travail (parfois très) fastidieux pour obtenir une caractérisation explicite par une ODE.

Extensions non présentées

- ▶ Lorsque $\sup_{x \in X} |x|^\beta |\Delta(x)| < +\infty$ pour $0 \leq \beta < 1$.

Extensions non présentées

- ▶ Lorsque $\sup_{x \in X} |x|^\beta |\Delta(x)| < +\infty$ pour $0 \leq \beta < 1$.
 - ▶ la chaîne a une dynamique plus lente.
 - ▶ limite fluide triviale: $\mathbb{Q}_x = \delta_\mu$ avec $\mu(t) = x$.
 - ▶ changer la définition du processus normalisé

$$\eta_r(t, x) = \frac{1}{r} \Phi_{\lceil tr^{1+\beta} \rceil} \quad \Phi_0 = rx.$$

- ▶ ergodicité plus lente.

Extensions non présentées

- ▶ Lorsque $\sup_{x \in X} |x|^\beta |\Delta(x)| < +\infty$ pour $0 \leq \beta < 1$.
 - ▶ la chaîne a une dynamique plus lente.
 - ▶ limite fluide triviale: $\mathbb{Q}_x = \delta_\mu$ avec $\mu(t) = x$.
 - ▶ changer la définition du processus normalisé

$$\eta_r(t, x) = \frac{1}{r} \Phi_{\lceil tr^{1+\beta} \rceil} \quad \Phi_0 = rx.$$

- ▶ ergodicité plus lente.
-
- ▶ Espace d'état : pas forcément $X = \mathbb{R}^d$.

III. Metropolis-within-Gibbs

Pour un MCMC particulier,

- (a) voir l'influence des paramètres de mise en oeuvre sur la limite fluide.
- (b) en déduire de "bons" paramètres.

Metropolis-within-Gibbs

↔ Algorithme

- choix aléatoire d'une direction de mise à jour :
 $\mathbb{P}(I = k) = \omega_k$.
- mise à jour de la composante sélectionnée, selon une étape de Hastings-Metropolis.

↔ Paramètres de mise en oeuvre

- Probabilités ω_k , $k \in \{1, \dots, d\}$.
- Lois de proposition : ici $q_k(x, \cdot) = \mathcal{N}(x, \sigma_k^2)$ donc variances σ_k^2 , $k \in \{1, \dots, d\}$.

Limite radiale de $\Delta(x) = \mathbb{E}_x[\Phi_1 - \Phi_0]$

Pour tout $i \in \{1, \dots, d\}$, $q_i = \mathcal{N}(0, \sigma_i^2)$

$$\Delta_i(x) = \omega_i \int_{\{y \in \mathbb{R}, \pi(x + ye_i) < \pi(x)\}} y \left(\frac{\pi(x + ye_i)}{\pi(x)} - 1 \right) q_i(y) dy.$$

Limite radiale de $\Delta(x) = \mathbb{E}_x[\Phi_1 - \Phi_0]$

Pour tout $i \in \{1, \dots, d\}$, $q_i = \mathcal{N}(0, \sigma_i^2)$

$$\Delta_i(x) = \omega_i \int_{\{y \in \mathbb{R}, \pi(x + ye_i) < \pi(x)\}} y \left(\frac{\pi(x + ye_i)}{\pi(x)} - 1 \right) q_i(y) dy.$$

↔ Pour toute densité cible π tq

- $\lim_{r \rightarrow +\infty} |\nabla \ln \pi(rx)| = +\infty$.
- ℓ donnée par $\lim_{r \rightarrow +\infty} \frac{\nabla \ln \pi(rx)}{|\nabla \ln \pi(rx)|} = \ell(x)$ est continue⁽⁻⁾.

Limite radiale de $\Delta(x) = \mathbb{E}_x[\Phi_1 - \Phi_0]$

Pour tout $i \in \{1, \dots, d\}$, $q_i = \mathcal{N}(0, \sigma_i^2)$

$$\Delta_i(x) = \omega_i \int_{\{y \in \mathbb{R}, \pi(x + ye_i) < \pi(x)\}} y \left(\frac{\pi(x + ye_i)}{\pi(x)} - 1 \right) q_i(y) dy.$$

↔ Pour toute densité cible π tq

- $\lim_{r \rightarrow +\infty} |\nabla \ln \pi(rx)| = +\infty$.
- ℓ donnée par $\lim_{r \rightarrow +\infty} \frac{\nabla \ln \pi(rx)}{|\nabla \ln \pi(rx)|} = \ell(x)$ est continue⁽⁻⁾.

↔ Lorsque $r \rightarrow +\infty$

$$\lim_r \Delta_i(rx) = \text{sign}(\ell_i(x)) \omega_i \int_{\mathbb{R}^+} y q_i(y) dy = \text{sign}(\ell_i(x)) \frac{\omega_i \sigma_i}{\sqrt{2\pi}}.$$

$$\lim_r \Delta_i(rx) = \text{sign}(\ell_i(x)) \frac{\omega_i \sigma_i}{\sqrt{2\pi}}, \quad \ell(x) := \lim_{r \rightarrow +\infty} \frac{\nabla \ln \pi(rx)}{|\nabla \ln \pi(rx)|}.$$

↔ Cela entraîne que

- ▶ La limite radiale de Δ dépend de la loi cible π via le “gradient normalisé limite”.
- ▶ La limite radiale de Δ dépend des paramètres de mise en oeuvre via le produit $\omega_i \sigma_i$.
- ▶ La limite radiale est constante sur les ensembles

$$O_\alpha = \{x, \text{sign}(\ell(x)) = \gamma_\alpha\}$$

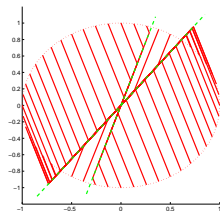
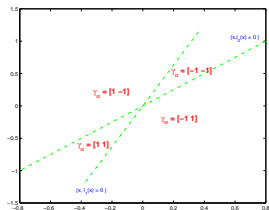
où $\gamma_\alpha \in \{-1, 1\}^d$.

- ▶ La valeur du champ limite sur O_α est donnée par les paramètres de mise en oeuvre.

Limite fluide linéaire par morceaux

↪ Exemple: MwG, $\pi \sim \mathcal{N}_2(0, \Gamma)$

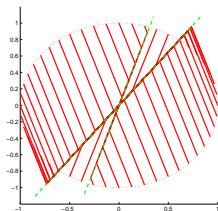
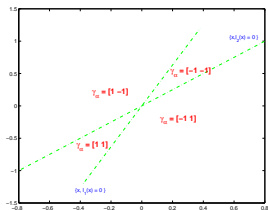
$$\implies \ell(x) = -\frac{\Gamma^{-1}x}{|\Gamma^{-1}x|}.$$



Limite fluide linéaire par morceaux

↔ Exemple: MwG, $\pi \sim \mathcal{N}_2(0, \Gamma)$

$$\implies \ell(x) = -\frac{\Gamma^{-1}x}{|\Gamma^{-1}x|}.$$



↔ Linéaire jusqu'au temps d'atteinte de $\partial[\cup_{\alpha=1}^a \mathcal{O}_\alpha]$ On peut mq :
 $\forall x \in \mathcal{O}_\alpha,$

$$\forall t \leq T(x) \quad \eta(t) = x + t \gamma_\alpha \circ \omega \circ \sigma, \quad \mathbb{Q}_x - \text{p.s.}$$

où $T(x)$: temps d'atteinte de $\partial\mathcal{O}_\alpha$.

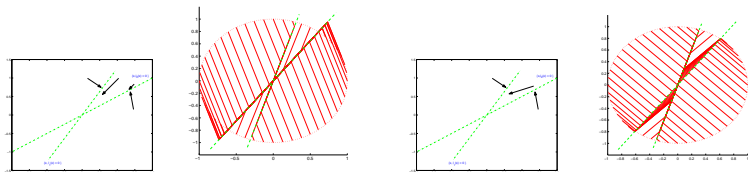
Stabilité du modèle fluide

(Résultats lorsque : $d = 2$ et les frontières sont des hyperplans)

► Réulte

- de l'existence d'une frontière "absorbante".
- du fait que toute frontière absorbante est "stable" (pour ces limites fluides associées au MwG).

► Exemple : MwG, $\pi \sim \mathcal{N}_2(0, \Gamma)$



Stratégie adaptative :

Puisque la limite fluide dépend des paramètres de mise en oeuvre à travers le produit $\omega_i \sigma_i$,

Stratégie 1. Fixer $\omega_i = 1/d$ et choisir $\sigma_i(x)$.

Stratégie 2. Fixer $\sigma_i = c$ et choisir $\omega_i(x)$.

Stratégie adaptative :

Puisque la limite fluide dépend des paramètres de mise en oeuvre à travers le produit $\omega_i \sigma_i$,

Stratégie 1. Fixer $\omega_i = 1/d$ et choisir $\sigma_i(x)$.

Stratégie 2. Fixer $\sigma_i = c$ et choisir $\omega_i(x)$.

alors la limite fluide \longleftrightarrow solution de l'ODE $\dot{\mu} = h(\mu)$ avec

$$h(x) = \frac{1}{\sqrt{2\pi}} \text{sign}(\ell_i(x)) [\omega_i \sigma_i](x)$$

Stratégie adaptative :

Puisque la limite fluide dépend des paramètres de mise en oeuvre à travers le produit $\omega_i \sigma_i$,

Stratégie 1. Fixer $\omega_i = 1/d$ et choisir $\sigma_i(x)$.

Stratégie 2. Fixer $\sigma_i = c$ et choisir $\omega_i(x)$.

alors la limite fluide \longleftrightarrow solution de l'ODE $\dot{\mu} = h(\mu)$ avec

$$h(x) = \frac{1}{\sqrt{2\pi}} \text{sign}(\ell_i(x)) [\omega_i \sigma_i](x)$$

Par exemple, en prenant

$$[\omega_i \sigma_i](x) = c |\ell_i(x)| \qquad \ell_i(x) = \lim_r \frac{\nabla_i \ln \pi(rx)}{|\nabla \ln \pi(rx)|}.$$

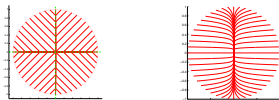
on a

$$h(x) = \frac{c}{\sqrt{2\pi}} \left(\lim_r \frac{\nabla \ln \pi(rx)}{|\nabla \ln \pi(rx)|} \right)$$

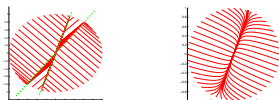
Algorithme de gradient pour ramener la chaîne vers les modes de π

Ex. : Limites fluides du MwG [gauche] non-adaptatif [droite] adaptatif

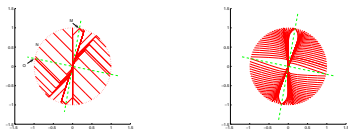
► Quand $\pi \sim \mathcal{N}_2(0, \Gamma_1)$ Γ_1 diagonale



► Quand $\pi \sim \mathcal{N}_2(0, \Gamma_2)$ Γ_2 non-diagonale



► Quand $\pi \sim \mathcal{N}_2(0, \Gamma_1) + \mathcal{N}_2(0, \Gamma_2)$



Comparaison des procédures

↔ **Critère 1**: Basé sur la **Limite fluide** et sur le temps d'atteinte d'une sphère de rayon $\rho \in]0,1[$ par la limite fluide issue de la sphère unité.

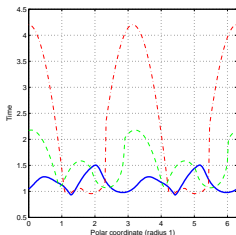
Comparaison des procédures

↪ **Critère 1** : Basé sur la **Limite fluide** et sur le temps d'atteinte d'une sphère de rayon $\rho \in]0,1[$ par la limite fluide issue de la sphère unité.

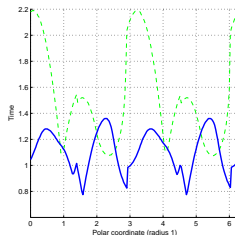
x-axes : coordonnée polaire de la valeur initiale.

y-axes : temps d'atteinte.

pour les trois algorithmes **Stratégie adaptative** **Non-Adaptative, $\omega_1 = 0.25$** **Non-Adaptative, $\omega_1 = 0.5$**



$$\pi \sim \mathcal{N}_2(0, \Gamma_2) \quad \Gamma_2 \text{ non diagonale}$$



$$\pi \sim \mathcal{N}_2(0, \Gamma_1) + \mathcal{N}_2(0, \Gamma_2)$$

↔ Critère 2 : Basé sur la chaîne de Markov et le temps d'atteinte du "centre de l'espace" quand chaîne initialisée "loin" du centre.

↪ **Critère 2** : Basé sur la **chaîne de Markov** et le temps d'atteinte du "centre de l'espace" quand chaîne initialisée "loin" du centre.

► **Exemple : comparaison des deux procédures adaptatives**

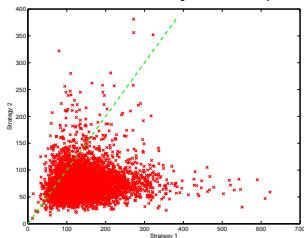
$$\pi \sim \mathcal{N}_8(0, \Gamma) \quad d = 8$$

Γ : diagonale, avec $\Gamma_{i,i} \sim \mathcal{E}(1)$.

5000 chaînes adaptatives, issues de $x \in \{z' \Gamma^{-1} z = d\}$.

x -axes : temps d'atteinte de la boule de rayon \sqrt{d} pour la Strat 1 (σ adapté)

y -axes : temps d'atteinte de la boule de rayon \sqrt{d} pour la Strat 2 (ω adapté)



► Exemple : intérêt de l'adaptation

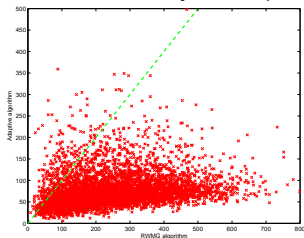
$$\pi \sim \mathcal{N}_8(0, \Gamma) \quad d = 8$$

Γ : diagonale, avec $\Gamma_{i,i} \sim \mathcal{E}(1)$.

5000 chaînes adaptatives, issues de $x \in \{z' \Gamma^{-1} z = d\}$

x -axes : temps d'atteinte de la boule de rayon \sqrt{d} pour l'algorithme standard

y -axes : temps d'atteinte de la boule de rayon \sqrt{d} pour la Stratégie 2 (ω adapté)



Conclusion

Pour une famille de densités cibles, nous avons

- ▶ établi l'existence d'un modèle fluide et la stabilité du modèle.
- ▶ montré que ce modèle dépendait des paramètres de mise en oeuvre de l'algorithme.
- ▶ exploité la forme de dépendance pour proposer des procédures adaptatives de choix des paramètres.
- ▶ vérifié l'intérêt de ces stratégies d'adaptation en regardant le comportement des chaînes dans la phase "transitoire".

Conclusion

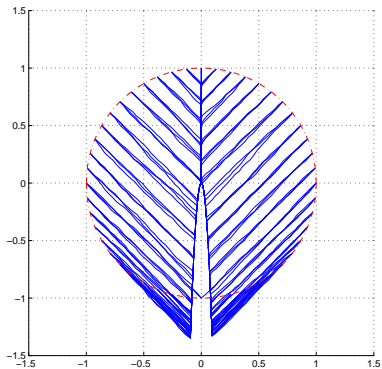
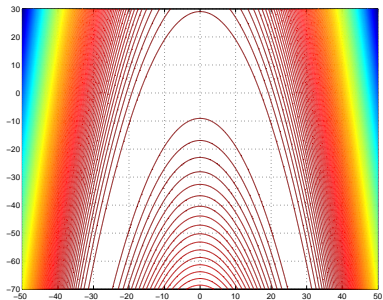
Pour une famille de densités cibles, nous avons

- ▶ établi l'existence d'un modèle fluide et la stabilité du modèle.
- ▶ montré que ce modèle dépendait des paramètres de mise en oeuvre de l'algorithme.
- ▶ exploité la forme de dépendance pour proposer des procédures adaptatives de choix des paramètres.
- ▶ vérifié l'intérêt de ces stratégies d'adaptation en regardant le comportement des chaînes dans la phase "transitoire".

Néanmoins,

- ▶ résultats théoriques en faible dimension (conditions complexes et peu générales).
- ▶ informations portées par les fluctuations autour de la dérive moyenne (étude à l'ordre 2) non exploitées (travail en cours).

Autre Exemple (MwG)



En conclusion de cette présentation,

- ▶ Hist. : les limites fluides sont utilisées pour l'étude des proc. de Markov à temps continu modélisant les files d'attente. Extension à l'étude de certaines Chaînes de Markov.
- ▶ Limites fluides ou conditions de drifts pour établir l'ergodicité des chaînes?
- ▶ Pour l'étude des méthodes MCMC : étude du régime transitoire de la chaîne (avant "stationnarité").

Résultats disponibles

- G. Fort, S. Meyn, E. Moulines and P. Priouret. *The ODE method for the stability of skip-free Markov Chains with applications to MCMC*. To be published, Ann. Appl. Probab. (2007)
- G. Fort. *Fluid limit-based tuning of some hybrid MCMC samplers*. Submitted (2007).