Stability of Markov Chains based on fluid limit techniques. Applications to MCMC

Gersende FORT

LTCI
CNRS - TELECOM ParisTech

In collaboration with Sean MEYN (Univ. Illinois), Eric MOULINES (TELECOM ParisTech) and Pierre PRIOURET (Univ. Paris 6).

We introduce

- a transformation of the Markov Chain $\longrightarrow$ family of time-continuous processes $\longrightarrow$ a limiting time-continuous process
- such that *the stability* of this process, is related to *the ergodicity* of the Markov chain.

$\Rightarrow$ characterization of the ergodicity ;

$\Rightarrow$ identification of the factors that play a role in the dynamic of the Markov chain.

We introduce

- a transformation of the Markov Chain $\longrightarrow$ family of time-continuous processes $\longrightarrow$ a limiting time-continuous process
- such that *the stability* of this process, is related to *the ergodicity* of the Markov chain.

  $\Rightarrow$ characterization of the ergodicity ;

  $\Rightarrow$ identification of the factors that play a role in the dynamic of the Markov chain.

The *Markov Chain Monte Carlo* (MCMC) algorithms

- are iterative algorithms that draw path of a Markov chain with given stationary distribution ;
- the performances of which are related (among other factors) to some parameters of implementation *(design parameters)*.
- $\Rightarrow$ find the role of the parameters in the definition of the *fluid limit* and propose an "optimal choice" of these parameters.

$\hookrightarrow$ Outline of the talk

I. A MCMC sampler : the Metropolis-within-Gibbs (MwG), and its design parameters.

II. Fluid limits.

III. Applications : guidelines on the choice of the design parameters for the MwG.

MCMC samplers :

Given a probability π, sample a Markov chain $\{\Phi_n, n \geq 0\}$ with unique stationary distribution π.

$\hookrightarrow$ Allow

- to explore the target density π.
- to approximate quantities of the form $\mathbb{E}_\pi[g(\Phi)]$ as soon as a LLN exists (and other limit theorems).

$\hookrightarrow$ Algorithms : Hastings-Metropolis, Gibbs, Metropolis-within-Gibbs, $\cdots$

Metropolis-within-Gibbs samplers in $\mathbb{R}^d$

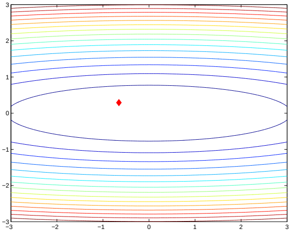
- Choose a selection probability : $\underline{\omega} = \{\omega_i, i \in \{1, \cdots, d\}\}$
- Choose a family of transition kernels on $\mathbb{R}$, $q_i(x, y)$ ex. $q_i(x, y) = \mathcal{N}(x, \sigma_i^2)[y]$
- Repeat :
  - select a direction I with prob. $\mathbb{P}(I = k) = \omega_k$.
  - draw a candidate $Y \sim q_I(\Phi_{n,I}, \cdot)$.
  - accept or reject the candidate : all the components are unchanged except the I-th

$$\Phi_{n+1,I} = \begin{cases} Y & \text{with proba} \\ \Phi_{n,I} & \text{otherwise.} \end{cases} \qquad \alpha(\Phi_n, Y) = 1 \wedge \frac{\pi(Y, \Phi_{n,-I})}{\pi(\Phi_n)} \frac{q_I(Y, \Phi_{n,I})}{q_I(\Phi_{n,I}, Y)}$$

Example : Metropolis-within-Gibbs (MwG)

- Explore on $\mathbb{R}^2$ a Gaussian distribution π with diagonal dispersion matrix
- and in each direction, the move is Gaussian.

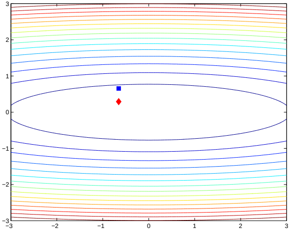
Initial value (and level curves of π)



Example : Metropolis-within-Gibbs (MwG)

- Explore on $\mathbb{R}^2$ a Gaussian distribution π with diagonal dispersion matrix
- and in each direction, the move is Gaussian.

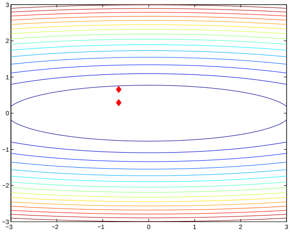
Initial value (and level curves of π), Propose



Example : Metropolis-within-Gibbs (MwG)

- Explore on $\mathbb{R}^2$ a Gaussian distribution π with diagonal dispersion matrix
- and in each direction, the move is Gaussian.

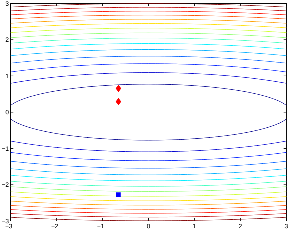
Initial value (and level curves of π), Propose , Accepted



Example : Metropolis-within-Gibbs (MwG)

- Explore on $\mathbb{R}^2$ a Gaussian distribution π with diagonal dispersion matrix
- and in each direction, the move is Gaussian.

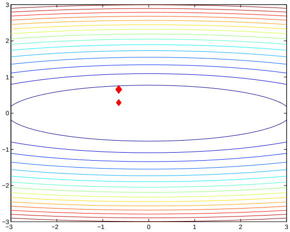
Initial value (and level curves of π), Propose , Accepted , Propose



Example : Metropolis-within-Gibbs (MwG)

- Explore on $\mathbb{R}^2$ a Gaussian distribution π with diagonal dispersion matrix
- and in each direction, the move is Gaussian.

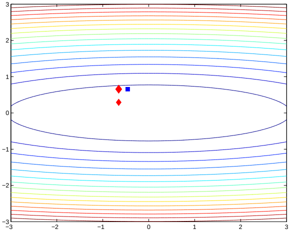
Initial value (and level curves of π), Propose , Accepted , Propose , Rejected



Example : Metropolis-within-Gibbs (MwG)

- Explore on $\mathbb{R}^2$ a Gaussian distribution π with diagonal dispersion matrix
- and in each direction, the move is Gaussian.

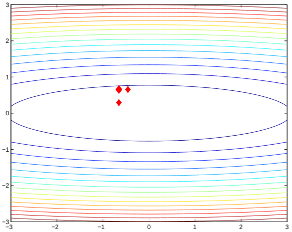
Initial value (and level curves of π), Propose , Accepted , Propose , Rejected , Propose



Example : Metropolis-within-Gibbs (MwG)

- Explore on $\mathbb{R}^2$ a Gaussian distribution π with diagonal dispersion matrix
- and in each direction, the move is Gaussian.

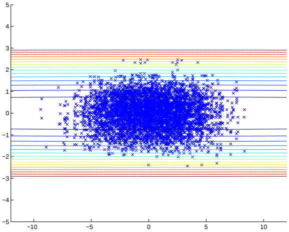
Initial value (and level curves of π), Propose , Accepted , Propose , Rejected , Propose , Accepted



Example : Metropolis-within-Gibbs (MwG)

- Explore on $\mathbb{R}^2$ a Gaussian distribution π with diagonal dispersion matrix
- and in each direction, the move is Gaussian.

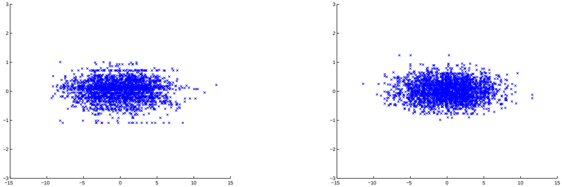
Initial value (and level curves of π), Propose , Accepted , Propose , Rejected , Propose , Accepted , After 10000 iterations.



Design parameters for the MwG

- Selection $\{\omega_i, i \leq d\}$,
- Gaussian proposal distributions in each direction, with std σ_i.

$\hookrightarrow$ Efficiency of the algorithm $\pi \sim \mathcal{N}_2(0, \Delta)$ with diagonal dispersion matrix Δ such that $\Delta_{1,1} >> \Delta_{2,2}$,



(left) $\omega_1 = \omega_2$, $\sigma_1 = \sigma_2$ (right) $\omega_1 = \omega_2$, $\sigma_1 >> \sigma_2$.

$\hookrightarrow$ Questions

- Optimal value of the design parameters.
- Adaptive methods : modify "on line" these parameters based on the past behavior of the algorithm.

$\hookrightarrow$ Hereafter,

- characterization of the role of these parameters on the dynamic of the chain.
- guidelines to fix / adapt the value of these parameters.

II. Fluid Limits

Normalized processes

Let $\{\Phi_k, k \geq 0\}$ be a Markov chain on X (X $= \mathbb{R}^d$).

A set of transformations : normalized process η_r, for $r > 0$

(i) in the initial value :

$$\eta_r(0; x) = \frac{1}{r}\Phi_0 = x \in \mathbb{R}^d, \qquad \Phi_0 = rx$$

(ii) in time and space :

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

Normalized processes

Let $\{\Phi_k, k \geq 0\}$ be a Markov chain on X (X $= \mathbb{R}^d$).

A set of transformations : normalized process η_r, for $r > 0$

(i) in the initial value :

$$\eta_r(0; x) = \frac{1}{r}\Phi_0 = x \in \mathbb{R}^d, \qquad \Phi_0 = rx$$

(ii) in time and space :

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

Hence $\eta_r(\cdot; x) = \frac{1}{r}\Phi_k$ on the time interval $\left[\frac{k}{r}; \frac{(k+1)}{r}\right)$.

By definition, cad-lag paths.

Definition

$\hookrightarrow$ Distributions

- $\mathbb{P}_x$: law of the canonical chain $\{\Phi_k, k \geq 0\}$ with initial value δ_x.
- $\mathbb{Q}_{r;x}$: distribution image of $\mathbb{P}_{rx}$ by $\eta_r(\cdot; x)$,

  distribution on $\mathbb{D}(\mathbb{R}^+, \mathsf{X})$ of cadlag functions $\mathbb{R}^+ \to \mathsf{X}$

Definition

↪ Distributions

- $\mathbb{P}_x$: law of the canonical chain $\{\Phi_k, k \geq 0\}$ with initial value δ_x.
- $\mathbb{Q}_{r;x}$: distribution image of $\mathbb{P}_{rx}$ by $\eta_r(\cdot; x)$,

  distribution on $\mathbb{D}(\mathbb{R}^+, \mathsf{X})$ of cadlag functions $\mathbb{R}^+ \to \mathsf{X}$

↪ Definition *Fluid Limit.* $\mathbb{Q}$ distribution on $\mathbb{D}(\mathbb{R}^+, \mathsf{X})$ is a fluid limit is there exists a family of scaling factors $r_n \to +\infty$ such that

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

Denoted by $\mathbb{Q}_x$ hereafter.

Definition

$\hookrightarrow$ Distributions

- $\mathbb{P}_x$: law of the canonical chain $\{\Phi_k, k \geq 0\}$ with initial value δ_x.
- $\mathbb{Q}_{r;x}$: distribution image of $\mathbb{P}_{rx}$ by $\eta_r(\cdot; x)$,
  distribution on $\mathbb{D}(\mathbb{R}^+, \mathsf{X})$ of cadlag functions $\mathbb{R}^+ \to \mathsf{X}$

$\hookrightarrow$ Definition *Fluid Limit*. $\mathbb{Q}$ distribution on $\mathbb{D}(\mathbb{R}^+, \mathsf{X})$ is a fluid limit is there exists a family of scaling factors $r_n \to +\infty$ such that

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

Denoted by $\mathbb{Q}_x$ hereafter.

$\hookrightarrow$ Rmk : fluid limit $\leftrightarrow \lim_r \mathbb{Q}_{r,x}$ and $\mathbb{Q}_{r,x}$ image of $\mathbb{P}_{rx}$ $\leftrightarrow$ behavior of the chains when started in the tails of π.

Example

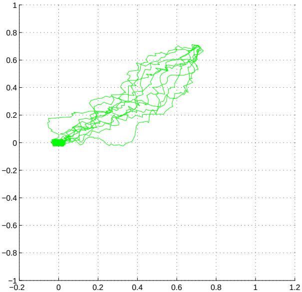
$\{\Phi_n, n \geq 0\}$ Hastings-Metropolis chain with target distribution on $\mathbb{R}^2$ given by

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

and Gaussian proposal distribution $4\,\mathcal{N}_2(x, \mathbb{I})$.

Figures : Different draws of the normalized process $\eta_r(\cdot, x)$ on $[0, T]$; for different initial values x ; and different scaling factors r.

One initial value



Example

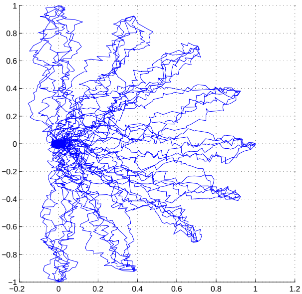
$\{\Phi_n, n \geq 0\}$ Hastings-Metropolis chain with target distribution on $\mathbb{R}^2$ given by

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

and Gaussian proposal distribution $4\,\mathcal{N}_2(x, \mathbb{I})$.

Figures : Different draws of the normalized process $\eta_r(\cdot, x)$ on $[0, T]$; for different initial values x ; and different scaling factors r.

One initial value, different initial values ($r = 100$)



Example

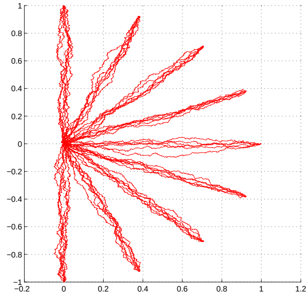
$\{\Phi_n, n \geq 0\}$ Hastings-Metropolis chain with target distribution on $\mathbb{R}^2$ given by

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

and Gaussian proposal distribution $4\,\mathcal{N}_2(x, \mathbb{I})$.

Figures : Different draws of the normalized process $\eta_r(\cdot, x)$ on $[0, T]$; for different initial values x ; and different scaling factors r.

One initial value, different initial values ($r = 100$), different scaling factors r ($r = 1000$)



Example

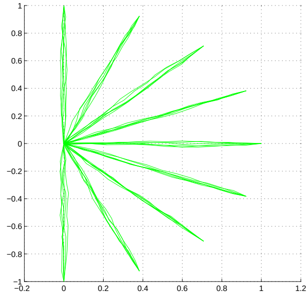
$\{\Phi_n, n \geq 0\}$ Hastings-Metropolis chain with target distribution on $\mathbb{R}^2$ given by

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

and Gaussian proposal distribution $4\,\mathcal{N}_2(x, \mathbb{I})$.

Figures : Different draws of the normalized process $\eta_r(\cdot, x)$ on $[0, T]$; for different initial values x ; and different scaling factors r.

One initial value, different initial values ($r = 100$), different scaling factors r ($r = 1000$) ($r = 5000$)



Example

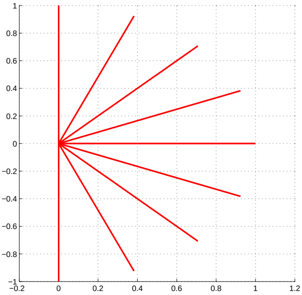
$\{\Phi_n, n \geq 0\}$ Hastings-Metropolis chain with target distribution on $\mathbb{R}^2$ given by

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

and Gaussian proposal distribution $4\,\mathcal{N}_2(x, \mathbb{I})$.

Figures : Different draws of the normalized process $\eta_r(\cdot, x)$ on $[0, T]$; for different initial values x ; and different scaling factors r.

One initial value, different initial values ($r = 100$), different scaling factors r ($r = 1000$) ($r = 5000$) (Fluid Limit)



Suff Cond for existence

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

Suff Cond for existence

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

▶ Theorem (Fort et al, 2007)

If

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

Then $\forall x$

- $\forall r_n \to +\infty$, $\exists$ sub-sequence $\{r_{n_j}, j \geq 1\}$ such that $\mathbb{Q}_{r_{n_j};x} \Rightarrow \mathbb{Q}_x$
- $\mathbb{Q}_x$ prob. on the space of the continuous functions from $\mathbb{R}^+$ to X.

Stability of the fluid limits

$\hookrightarrow$ Definition *Stable Fluid model* : there exist $T > 0$ and $\rho < 1$ such that for any x on the unit sphere,

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

Theorem $(\star\star\star\star)$ (Fort et al, 2007)

If

- $\{\Phi_k, k \geq 0\}$ is phi-irreducible, aperiodic ; and compact sets are petite.
- the fluid model exists and is stable.

Then the Markov chain is (f, r)-ergodic

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

Theorem $(\star\star\star\star)$ (Fort et al, 2007)

If

- $\{\Phi_k, k \geq 0\}$ is phi-irreducible, aperiodic ; and compact sets are petite.
- the fluid model exists and is stable.

Then the Markov chain is (f, r)-ergodic

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

p : control of the martingale increment in the decomposition

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

The hitting-time T of the ball of radius ρ by the fluid model plays a role in the control of convergence of $P^n(x, \cdot)$ to π. (control of the returns to the "center")

Fluid Limit = Skeleton of the chain

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

- For the normalized process (piecewise constant, jumps at time k/r) :

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

where we set

$$h\left(x\right) = \lim_{r\to+\infty} \Delta(r\ x).$$

Fluid Limit = Skeleton of the chain

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

- For the normalized process (piecewise constant, jumps at time k/r) :

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

where we set

$$h\left(x\right) = \lim_{r\to+\infty} \Delta(r\ x).$$

- Hence, noisy 'observation' of

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

To be more precise, fluid limit are characterized by

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

for any compact $\mathsf{H} \subset$?

To be more precise, fluid limit are characterized by

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

for any compact $\mathsf{H} \subset$?

- In the easiest cases (? $=$ X), fluid limits are Dirac mass at a function μ that solves the ODE $\dot{\mu} = h(\mu)$.
  Stability of the fluid model $\longleftrightarrow$ Stability of the ODE.
- Otherwise, more technical results, no general conditions.

Characterization : case 1

- If

  · $\exists \;\; h$ continuous such that $\mathsf{H} \subsetneq \mathsf{X} \setminus \{0\}$,

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

  · the ODE $\dot{\mu} = h(\mu)$ is stable for any initial value x.

Then the fluid model is stable.

Characterization : case 1

- If
  - $\exists \;\; h$ continuous such that $\mathsf{H} \subsetneq \mathsf{X} \setminus \{0\}$,

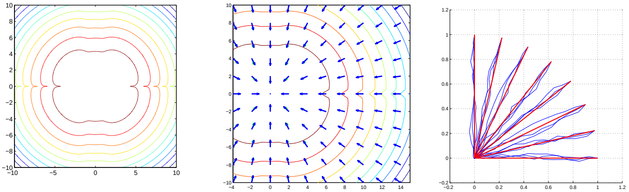
$$\lim_{r \to +\infty} \sup_{x \in \mathsf{H}} |\Delta(rx) - h(x)| = 0.$$

  - the ODE $\quad \dot{\mu} = h(\mu) \quad$ is stable for any initial value x.

Then the fluid model is stable.

- Example : Hastings-Metropolis

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



Level curves of π and fields Δ, h and draws of the fluid limit

Characterization : case 2

- If
  - $\exists \ h$ continuous such that for any compact H in a cone of $\mathsf{X} \setminus \{0\}$,
  $$\lim_{r \to +\infty} \sup_{x \in \mathsf{H}} |\Delta(rx) - h(x)| = 0.$$
  - the ODE $\dot{\mu} = h(\mu)$ started from a point in the cone are stable
  - the cone is “ attractive”.

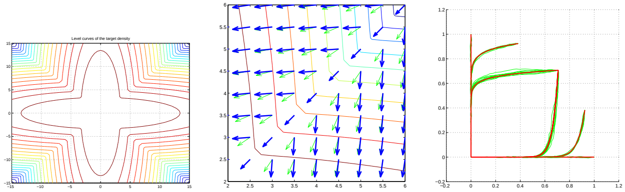
Then the fluid model is stable.

Characterization : case 2

- If
  - $\exists \ h$ continuous such that for any compact H in a cone of $\mathsf{X} \setminus \{0\}$,
  $$\lim_{r \to +\infty} \sup_{x \in \mathsf{H}} |\Delta(rx) - h(x)| = 0.$$
  - the ODE $\dot{\mu} = h(\mu)$ started from a point in the cone are stable
  - the cone is “ attractive”.

Then the fluid model is stable.

- Example : Hastings-Metropolis. π mixture of Gaussian distributions



Level curves of π and fields Δ, h and realizations of the fluid limits

Characterization : case 3 $(\mathsf{X} = \mathbb{R}^2)$

- If
  - $\mathsf{X} = \bigcup_{\alpha=1}^{a} \mathsf{O}_\alpha \cup \bigcup_{\beta=1}^{b} \{x, f_\beta' x = 0\}$.
  - $\exists \ \Sigma_\alpha$ such that for any compact H of O_α,

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

  - hyperplanes are "attractive" and "stable".

Then the fluid model is stable.

Characterization : case 3 $(\mathsf{X} = \mathbb{R}^2)$

- If

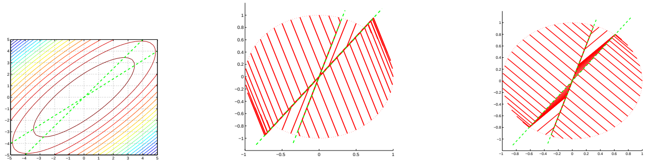
  - $\mathsf{X} = \bigcup_{\alpha=1}^{a} \mathsf{O}_\alpha \cup \bigcup_{\beta=1}^{b} \{x, f'_\beta x = 0\}$.
  - $\exists \ \Sigma_\alpha$ such that for any compact H of O_α,

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

  - hyperplanes are "attractive" and "stable".

  Then the fluid model is stable.

- Example : Metropolis within Gibbs



Level curves of π and fluid limits when $\omega_1 = 0.25$ and fluid limits when $\omega_1 = 0.5$

Conclusion (II)

- By renormalization of the chain,
  - the fluid model characterizes the behavior of the chain started “far in the tails” $\Phi_0 = rx$ and $r \to +\infty$.
  - the deterministic 'hidden' behavior is obtained by removing the stochastic perturbations.

Conclusion (II)

- By renormalization of the chain,
  - the fluid model characterizes the behavior of the chain started "far in the tails" $\Phi_0 = rx$ and $r \to +\infty$.
  - the deterministic 'hidden' behavior is obtained by removing the stochastic perturbations.
- Ergodicity of the initial chain is related to the stability of the fluid model.
- Fluid model characterized (almost everywhere) by an ODE.

Conclusion (II)

- By renormalization of the chain,
  - the fluid model characterizes the behavior of the chain started "far in the tails" $\Phi_0 = rx$ and $r \to +\infty$.
  - the deterministic 'hidden' behavior is obtained by removing the stochastic perturbations.
- Ergodicity of the initial chain is related to the stability of the fluid model.
- Fluid model characterized (almost everywhere) by an ODE.

- The fluid limit gives informations on the dynamic of the chain in the transient phase (i.e. before the stationary phase).
- But- in some cases - with quite cumbersome and fastidious computations in order to obtain an explicit characterization by an ODE.

Other results not discussed here

- When $\sup_{x \in \mathsf{X}} |x|^{\beta} |\Delta(x)| < +\infty$ for some $0 \leq \beta < 1$.

Other results not discussed here

- When $\sup_{x \in \mathsf{X}} |x|^\beta |\Delta(x)| < +\infty$ for some $0 \leq \beta < 1$.
  - the chain has a slower dynamic.
  - trivial fluid limit : $\mathbb{Q}_x = \delta_\mu$ with $\mu(t) = x$.
  - modify the definition of the normalized process

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

  - weaker ergodicity.

Other results not discussed here

- When $\sup_{x \in \mathsf{X}} |x|^{\beta} |\Delta(x)| < +\infty$ for some $0 \leq \beta < 1$.
  - the chain has a slower dynamic.
  - trivial fluid limit : $\mathbb{Q}_x = \delta_\mu$ with $\mu(t) = x$.
  - modify the definition of the normalized process

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

  - weaker ergodicity.
- State space : not necessarily $\mathsf{X} = \mathbb{R}^d$.

III. Metropolis-within-Gibbs

↪ Design parameters

(a) the selection probability $\underline{\omega} = \{\omega_i, i \leq d\}$.

(b) the size of the moves in each direction (e.g. the variances σ_i^2 when the proposal is Gaussian in each direction i).

↪ Which approach ?

(a) try to optimize the choice of $\underline{\omega}$ and fix the variances $\sigma_i^2 = c$.

(b) try to optimize the choice of the variances σ_i^2 and fix the probability $\omega_i = 1/d$.

(c) try to optimize both σ_i^2 and ω_i, $i \leq d$.

Expression of $\Delta(x) = \mathbb{E}_x[\Phi_1 - \Phi_0]$

For any $i \in \{1, \cdots, d\}$, $q_i = \mathcal{N}(0, \sigma_i^2)$

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

where $\{e_i, i \leq d\}$ is the canonical basis.

Expression of $\Delta(x) = \mathbb{E}_x[\Phi_1 - \Phi_0]$

For any $i \in \{1, \cdots, d\}$, $q_i = \mathcal{N}(0, \sigma_i^2)$

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

where $\{e_i, i \leq d\}$ is the canonical basis.

In order to characterize fluid limit, the radial limit

$$h(x) = \lim_{r \to +\infty} \Delta(r\, x)$$

is required. To that goal, assumptions on

- the limit of the rejection area $\{y \in \mathbb{R}, \pi(r\, x + ye_i) < \pi(r\, x)\}$ when $r \to +\infty$,
- the behavior of the gradient $\nabla \ln \pi(r\, x)$

are needed.

$\hookrightarrow$ For any target density π such that

- $\lim_{r \to +\infty} |\nabla \ln \pi(rx)| = +\infty$.
- ℓ given by $\lim_{r \to +\infty} \frac{\nabla \ln \pi(rx)}{|\nabla \ln \pi(rx)|} = \ell(x)$ is continuous$^{(-)}$.

$\hookrightarrow$ For any target density π such that

- $\lim_{r \to +\infty} |\nabla \ln \pi(rx)| = +\infty$.
- ℓ given by $\lim_{r \to +\infty} \frac{\nabla \ln \pi(rx)}{|\nabla \ln \pi(rx)|} = \ell(x)$ is continuous$^{(-)}$.

$\hookrightarrow$ the field h is given by

$$h_i(x) = \mathrm{sign}(\ell_i(x)) \, \frac{\omega_i \; \sigma_i}{\sqrt{2\pi}} \qquad \qquad \ell(x) := \lim_{r \to +\infty} \frac{\nabla \ln \pi(rx)}{|\nabla \ln \pi(rx)|}.$$

$$h_i(x) = \text{sign}(\ell_i(x)) \; \frac{\omega_i \; \sigma_i}{\sqrt{2\pi}} \qquad\qquad \ell(x) := \lim_{r \to +\infty} \frac{\nabla \ln \pi(rx)}{|\nabla \ln \pi(rx)|}.$$

$\hookrightarrow$ This implies that

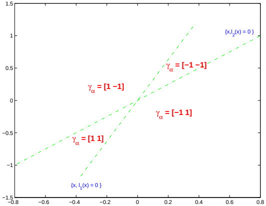
- h (and thus, the fluid limit) depends upon π through the "normalized limiting gradient".
- The fluid limit depends upon the design parameters through the products $\{\omega_i \sigma_i, i \leq d\}$.
- The field h is constant (and thus, the ODE is linear) on the sets

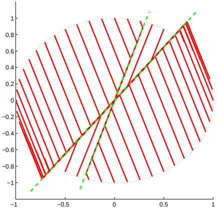
$$\mathsf{O}_\alpha = \{x, \text{sign}(\ell(x)) = \gamma_\alpha\}$$

where $\gamma_\alpha \in \{-1, 1\}^d$.

Piecewise linear fluid limits

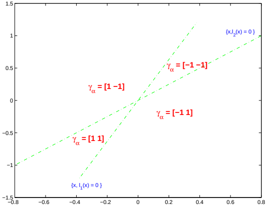
$\hookrightarrow$ Example : MwG, $\pi \sim \mathcal{N}_2(0, \Gamma)$ $\Longrightarrow \ell(x) = -\frac{\Gamma^{-1}x}{|\Gamma^{-1}x|}$.

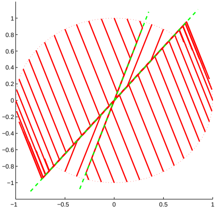




Piecewise linear fluid limits

↪ Example : MwG, $\pi \sim \mathcal{N}_2(0, \Gamma)$ $\Longrightarrow \ell(x) = -\frac{\Gamma^{-1}x}{|\Gamma^{-1}x|}$.





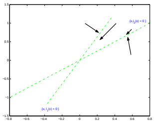
↪ The fluid limit is

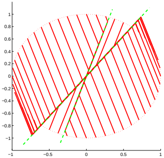
- linear till the first time it enters one of the sets $\{x, \ell_i(x) = 0\}$, $i \leq d$ - which in the above example - are the hyperplanes in green.
- then, the behavior depends on the field h in a neighborhood of these sets.

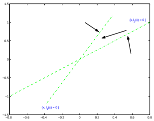
In any cases,

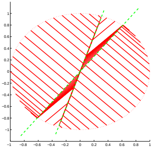
- there exists at least one “absorbing” set.
- this set is “stable” i.e. the fluid limits - when trapped in these sets - move towards the origin.

Two situations, obtained with different values of the design parameters









Strategy :

- Since the fluid limit depends on the design parameters through the product $\omega_i \sigma_i$,

  Strategy 1. Fix $\omega_i = 1/d$ and choose the std of the form $\sigma_i(x)$.

  Strategy 2. Fix $\sigma_i = c$ and choose the selection of the form $\omega_i(x)$.

Strategy :

- Since the fluid limit depends on the design parameters through the product $\omega_i \sigma_i$,

  Strategy 1. Fix $\omega_i = 1/d$ and choose the std of the form $\sigma_i(x)$.

  Strategy 2. Fix $\sigma_i = c$ and choose the selection of the form $\omega_i(x)$.

then, the fluid limit $\longleftrightarrow$ solves the ODE $\dot{\mu} = h(\mu)$ with

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

Strategy :

- Since the fluid limit depends on the design parameters through the product $\omega_i \sigma_i$,

  Strategy 1. Fix $\omega_i = 1/d$ and choose the std of the form $\sigma_i(x)$.

  Strategy 2. Fix $\sigma_i = c$ and choose the selection of the form $\omega_i(x)$.

then, the fluid limit $\longleftrightarrow$ solves the ODE $\dot{\mu} = h(\mu)$ with

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

- We propose

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

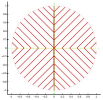
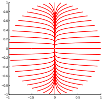
so that

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

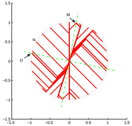
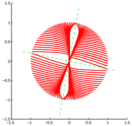
A gradient algorithm so that the chain - started far in the tails - is attracted towards the mode of π (i.e. the "center" of the space)

Ex. : Fluid limits of the MwG [left] non-adaptive [right] adaptive

- When $\pi \sim \mathcal{N}_2(0, \Gamma_1)$ Γ_1 diagonal

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

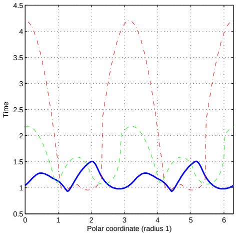
Comparison of the strategies

$\hookrightarrow$ Criterion 1 : Based on the Fluid Limit and on its hitting-time of a sphere of radius $\rho \in]0, 1[$ when initialized on the unit sphere.

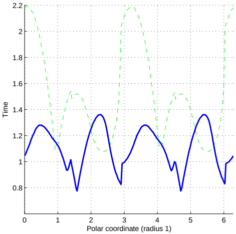
Comparison of the strategies

$\hookrightarrow$ Criterion 1 : Based on the Fluid Limit and on its hitting-time of a sphere of radius $\rho \in]0, 1[$ when initialized on the unit sphere.

x-axes : polar coordinate of the initial value.
y-axes : hitting-time.
for the three algorithms Adaptive strategy Non-Adaptive, $\omega_1 = 0.25$ Non-Adaptive, $\omega_1 = 0.5$



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



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

$\hookrightarrow$ Criterion 2 : Based on the Markov chain and its hitting-time of the "center of the space " when chain started "far" from the center.

$\hookrightarrow$ Criterion 2 : Based on the Markov chain and its hitting-time of the "center of the space" when chain started "far" from the center.

▶ Example : comparison of the two adaptive procedures

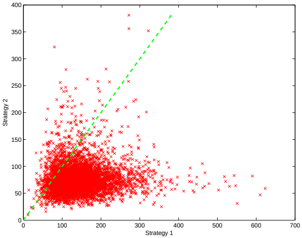
$\pi \sim \mathcal{N}_8(0, \Gamma) \qquad d = 8$

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

5000 adaptive chains, started from $x \in \{z'\Gamma^{-1}z = d\}$.

x-axes : hitting-time of the ball of radius $\sqrt{d}$ for Strat 1 (adapt σ)

y-axes : hitting-time of the ball of radius $\sqrt{d}$ for Strat 2 (adapt ω)



▶ Example : adaptive vs non-adaptive

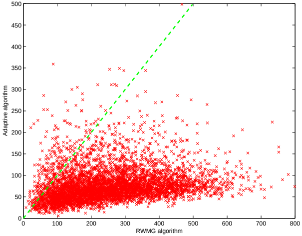
$\pi \sim \mathcal{N}_8(0, \Gamma)$ $\qquad d = 8$

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

5000 adaptive chains, started from $x \in \{z'\Gamma^{-1}z = d\}$

x-axes : hitting-time of the ball of radius $\sqrt{d}$ for the classical algorithm

y-axes : hitting-time of the ball of radius $\sqrt{d}$ for the Strat 2 (adapt ω)



To conclude,

- Hist. : fluid limits are common tools in queuing theory (continuous-time Markov process)
  We provided an extension of this theory to the study of some (discrete-time) Markov chains.
- Fluid limits or drift conditions to prove the ergodicity of the chain ?
- provide an analysis of the chain in its transient phase (before "stationnarity")

Available results

- 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. (2008)
- G. Fort. *Fluid limit-based tuning of some hybrid MCMC samplers*. Submitted (2007).