Fluid Limits for some MCMC samplers

- Gersende FORT, CNRS, Paris, France.

Joint work with

- Sean MEYN (Univ. of Illinois, Urbana, USA),
- Eric MOULINES (GET, France),
- Pierre PRIOURET (University Paris VI, France).

Outline of the talk

We are interested in

- the existence + stability of the fluid limits for `skip free` Markov Chains.
- their use in the study of (some) MCMC samplers.

Outline of the talk

We are interested in

- the existence + stability of the fluid limits for `skip free` Markov Chains.
- their use in the study of (some) MCMC samplers.

We will discuss

1. *Fluid Limits* for *skip-free* Markov Chains.
   - the existence of *fluid limits*
   - their characterization
   - their stability and the stability of the Markov Chain.
2. Applications to Metropolis-Hastings Markov Chains
   - Convergence of the samplers
   - How to tune the parameters ?

MCMC samplers / Hastings-Metropolis

Sample from a (complex, unnormalized) distribution π on $\mathbb{R}^d$ when exact sampling is not possible :

Define a Markov Chain $(\Phi_n, n \geq 0)$,
with unique stationary distribution $\propto \pi$
and *ergodic*.

MCMC samplers / Hastings-Metropolis

Sample from a (complex, unnormalized) distribution π on $\mathbb{R}^d$ when exact sampling is not possible :

Define a Markov Chain $(\Phi_n, n \geq 0)$,
with unique stationary distribution $\propto \pi$
and *ergodic*.

Ex. Hastings-Metropolis algorithm
Given Φ_t, define Φ_{t+1} by

- $\Phi_{t+1/2} \sim Q(\Phi_t, \cdot)$.
- $\Phi_{t+1} = \begin{cases} \Phi_{t+1/2} & \text{with prob. } \alpha(\Phi_t, \Phi_{t+1/2}) \\ \Phi_t & \text{with prob. } 1 - \alpha(\Phi_t, \Phi_{t+1/2}) \end{cases}$,

where $\alpha(x, z) = 1 \wedge \frac{\pi(z) Q(z,x)}{\pi(x) Q(x,z)}$.

MCMC samplers / Hastings-Metropolis

Problems :

- ($\star$) Convergence ? (ergodicity)

$$\kappa(n)\ \left|\mathbb{E}_x\left[g(\Phi_n)\right] - \pi(g)\right| \to 0 \qquad \forall x, \qquad g \in ?$$

- Limit Theorems

$$n^{-1}\sum_{k=1}^{n} g(\Phi_k) \to_{\text{a.s.}} \pi(g) \qquad \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\{g(\Phi_k) - \pi(g)\} \to_{\text{d}} \mathcal{N}(0, \sigma_g^2).$$

- ($\star$) How to tune the parameters i.e. (here) the proposal kernel $Q(x, y)$

MCMC samplers / Hastings-Metropolis

Problems :

- $(\star)$ Convergence ? (ergodicity)

$$\kappa(n) \; |\mathbb{E}_x\left[g(\Phi_n)\right] - \pi(g)| \to 0 \qquad \forall x, \qquad g \in ?$$

- Limit Theorems

$$n^{-1}\sum_{k=1}^{n} g(\Phi_k) \to_{\text{a.s.}} \pi(g) \qquad \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\{g(\Phi_k) - \pi(g)\} \to_{\text{d}} \mathcal{N}(0,\sigma_g^2).$$

- $(\star)$ How to tune the parameters i.e. (here) the proposal kernel $Q(x,y)$

Hereafter, illustrations in the case

- symmetric HM : $Q(x,y) = q(|x-y|)$
- $q(z) \sim \sigma\, \mathcal{N}_d(0,\mathbb{I})[z]$

Existence of fluid limits (a)

$\hookrightarrow$ Define a normalized process

(i) in the initial point

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

(ii) in time and space

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

$$\eta_r(t;x) = \frac{1}{r}\Phi_k \qquad \text{on } \left[\frac{k}{r}; \frac{(k+1)}{r}\right).$$

Existence of fluid limits (a)

↪ Define a normalized process

(i) in the initial point

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

(ii) in time and space

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

$$\eta_r(t;x) = \frac{1}{r}\Phi_k \qquad \text{on } \left[\frac{k}{r}; \frac{(k+1)}{r}\right).$$

- Distributions
  - $\mathbb{P}_x$: distribution of the Markov Chain with initial distribution δ_x.
  - $\mathbb{Q}_{r;x}$: image prob. of $\mathbb{P}_x$ by $\eta_r(\cdot;x)$ prob. on the space of càd-làg functions $\mathbb{R}^+ \to \mathsf{X}$.

Existence of fluid limits (b)

- Définition : $\mathbb{Q}_x$ is a fluid limitif there exists $\{r_n\}_n \to +\infty$, $\{x_n\}_n \to x$ s.t.

$$\mathbb{Q}_{r_n;x_n} \Longrightarrow \mathbb{Q}_x \qquad \text{on the space of the càd-làg functions } \mathbb{R}^+ \to \mathsf{X}.$$

Existence of fluid limits (b)

- Définition : $\mathbb{Q}_x$ is a fluid limitif there exists $\{r_n\}_n \to +\infty$, $\{x_n\}_n \to x$ s.t.

$$\mathbb{Q}_{r_n;x_n} \Longrightarrow \mathbb{Q}_x \qquad \text{on the space of the càd-làg functions } \mathbb{R}^+ \to \mathsf{X}.$$

$$\Phi_{k+1} = \Phi_k + \mathbb{E}\left[\Phi_{k+1}|\mathcal{F}_k\right] - \Phi_k + \Phi_{k+1} - \mathbb{E}\left[\Phi_{k+1}|\mathcal{F}_k\right]$$

Existence of fluid limits (b)

- Définition : $\mathbb{Q}_x$ is a fluid limitif there exists $\{r_n\}_n \to +\infty$, $\{x_n\}_n \to x$ s.t.

$$\mathbb{Q}_{r_n;x_n} \Longrightarrow \mathbb{Q}_x \qquad \text{on the space of the càd-làg functions } \mathbb{R}^+ \to \text{X}.$$

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

Existence of fluid limits (b)

- Définition : $\mathbb{Q}_x$ is a fluid limitif there exists $\{r_n\}_n \to +\infty$, $\{x_n\}_n \to x$ s.t.

$$\mathbb{Q}_{r_n;x_n} \Longrightarrow \mathbb{Q}_x$$

on the space of the càd-làg functions $\mathbb{R}^+ \to \mathsf{X}$.

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

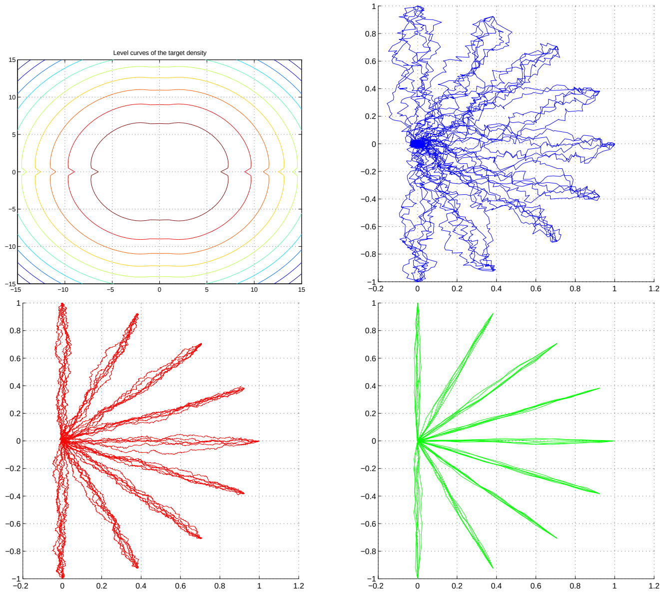
- Result

if

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

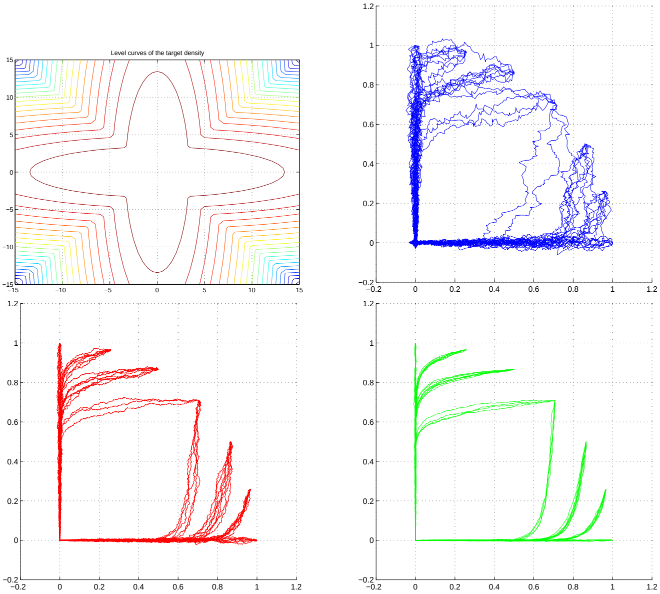
Then fluid limits exist, prob. on the space of continuous functions (whatever the initial point on the unit sphere)

Example 1 : (regular case)



$\pi(x, y) \propto (1 + x^2 + y^2 + x^8 y^2) \exp(-(x^2 + y^2)), \quad q \sim \mathcal{N}(0, 4),$ r=100, r=1000, r=5000

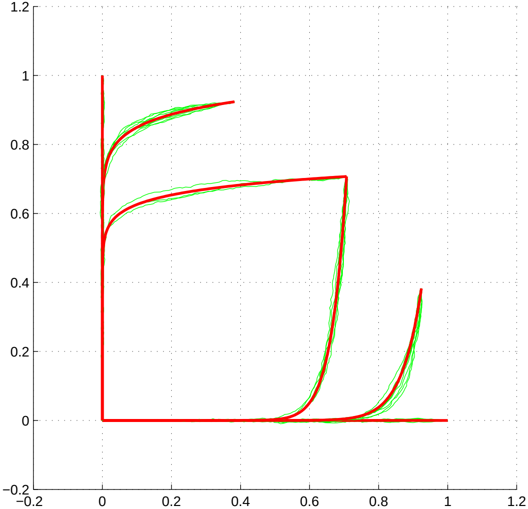
Example 2 : (irregular case)



$\pi(x, y) \propto \mathcal{N}(0, \Gamma_1^{-1}) + \mathcal{N}(0, \Gamma_2^{-1}), \quad q \sim \mathcal{N}(0, 1),$ r=100, r=1000, r=5000

Characterisation of the fluid limits

$\hookrightarrow$ Can we describe the distributions $\mathbb{Q}_x$?



$\pi(x, y) \propto$ mixture of Gaussian,
$q \sim \mathcal{N}(0, \mathbb{I})$, r=5000 T=5

Characterization (b)

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

- For the normalized process

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

where

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

Characterization (b)

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

- For the normalized process

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

where

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

- Thus the dynamic

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

in an additive noise.

Characterisation (c)

- Theorem

If

- Existence of the fluid limit.
- there exists an open cone O de X $\setminus \{0\}$,
- h : O $\rightarrow$ X s.t.

$$\sup_{x \in \mathsf{H}} \left| r^{\beta} \Delta(rx) - |x|^{-\beta} h(x) \right| \rightarrow 0, \qquad r \rightarrow +\infty,$$

for any compact H $\subseteq$ O

Characterisation (c)

- Theorem

If

- Existence of the fluid limit.
- there exists an open cone O de X $\setminus \{0\}$,
- h : O $\rightarrow$ X s.t.

$$\sup_{x \in \mathsf{H}} \left| r^{\beta} \Delta(rx) - |x|^{-\beta} h(x) \right| \rightarrow 0, \qquad r \rightarrow +\infty,$$

for any compact H $\subseteq$ O

Then for all $0 \leq s \leq t$, on $\{\eta, \eta(u) \in \mathsf{O}, s \leq u \leq t\}$,

$$\sup_{s \leq u \leq t} \left| \eta(u) - \eta(s) - \int_s^u h \circ \eta(v) \; dv \right| = 0, \quad \mathbb{Q}_x^{\beta} - a.s.$$

Characterisation (c)

▶ Theorem

If

- Existence of the fluid limit.
- there exists an open cone O de X $\setminus \{0\}$,
- h : O $\to$ X s.t.

$$\sup_{x \in \mathsf{H}} \left| r^{\beta} \Delta(rx) - |x|^{-\beta} h(x) \right| \to 0, \qquad r \to +\infty,$$

for any compact H $\subseteq$ O

Then for all $0 \leq s \leq t$, on $\{\eta, \eta(u) \in \mathsf{O}, s \leq u \leq t\}$,

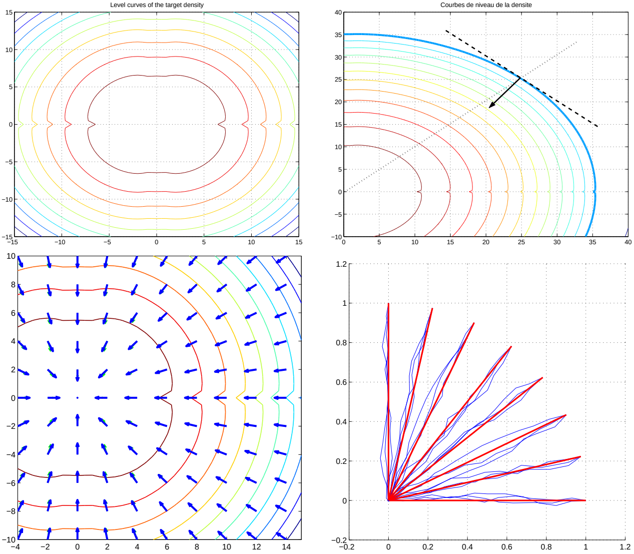
$$\sup_{s \leq u \leq t} \left| \eta(u) - \eta(s) - \int_s^u h \circ \eta(v)\ dv \right| = 0, \quad \mathbb{Q}_x^{\beta} - a.s.$$

▶ i.e. the fluid limit $\mathbb{Q}_x^{\beta}$ is a Dirac mass at the point η satisfying

$$\eta(u) = \eta(s) + \int_s^u h \circ \eta(v)\ dv, \qquad s \leq u \leq t,$$

whenever $\eta([s,t]) \subset \mathsf{O}$.

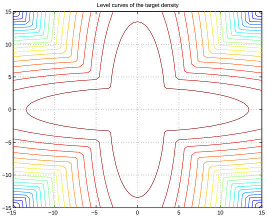
Example 3 : Super-exponential case, $O = X \setminus \{0\}$



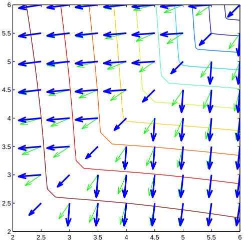
UpperLeft- Level curves of π UpperRight- Rejection area

LowerLeft- Level curves, $\triangle$ and h LowerRight- Process η_x^{β} and flow of the ODE.

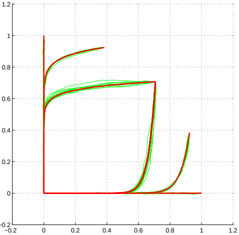
Example 4 : Super-exponential case, $O \subsetneq X \setminus \{0\}$



Level curves of π

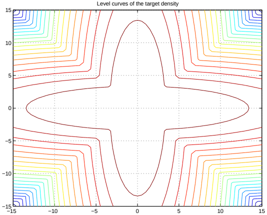


Level curves, $\triangle$ and h

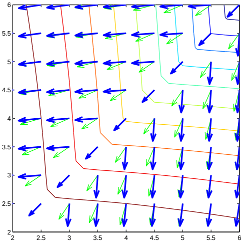


Process η_x^β and flow of the ODE.

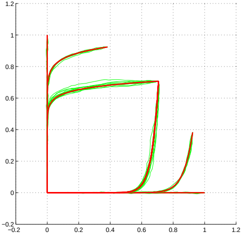
Example 4 : Super-exponential case, $\mathsf{O} \subsetneq \mathsf{X} \setminus \{0\}$



Level curves of π



Level curves, $\triangle$ and h



Process η_x^β and flow of the ODE.

$\hookrightarrow$ There exists $T_0 < \infty$ s.t. for all $x \in \mathsf{X}$, $|x| = 1$, and any fluid limit $\mathbb{Q}_x$,

$$\mathbb{Q}_x \left(\eta, \eta([0, T_0]) \cap \mathsf{O} \neq \emptyset \right) = 1.$$

[Appl 1] Stability : fluid limit → Markov Chain

- A fluid limit is stable if $\exists\, T > 0$ and $0 < \rho < 1$, s.t.

$$\forall\, x, |x| = 1, \quad \mathbb{Q}_x \left(\eta, \inf_{[0,T]} |\eta(\cdot)| \leq \rho \right) = 1.$$

[Appl 1] Stability : fluid limit → Markov Chain

- A fluid limit is stable if $\exists\, T > 0$ and $0 < \rho < 1$, s.t.

$$\forall\, x, |x| = 1, \quad \mathbb{Q}_x\left(\eta, \inf_{[0,T]} |\eta(\cdot)| \le \rho\right) = 1.$$

- If
  - irreducible, aperiodic, compact sets are petite.
  - Existence of the fluid limits.
  - Stability of the fluid limits.

Then polynomial ergodicity,

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

[Appl 1] Stability : fluid limit → Markov Chain

- A fluid limit is stable if $\exists\, T > 0$ and $0 < \rho < 1$, s.t.

$$\forall\, x, |x| = 1, \quad \mathbb{Q}_x\left(\eta, \inf_{[0,T]} |\eta(\cdot)| \le \rho\right) = 1.$$

- If
  - · irreducible, aperiodic, compact sets are petite.
  - · Existence of the fluid limits.
  - · Stability of the fluid limits.

Then polynomial ergodicity,

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

- Stability ODE / Stability fluid limit

(i) Case 1 : $\mathsf{O} = \mathsf{X} \setminus \{0\}$

$$\exists\, T, \rho \qquad \forall\, x, |x| = 1, \qquad \inf_{[0,T]} |\mu(\cdot; x)| \le \rho < 1.$$

(ii) Case 2 : $\mathsf{O} \neq \mathsf{X} \setminus \{0\}$ and $\mathbb{Q}_x\left(\eta, \eta([0, T_0]) \cap \mathsf{O} \neq \emptyset\right) = 1$.

$$\forall K > 0, \ \exists\, T_K, \rho_K \qquad \forall\, x \in \mathsf{O}, |x| \le K, \qquad \inf_{[0, T_K \wedge T_x]} |\mu(\cdot; x)| \le \rho_K < 1.$$

[Appl 2] How to choose the parameters of the algorithms ?

- Hybrid Hastings-Metropolis :

$$P(x, dy) = \sum_{k=1}^{d} \omega_k \; P_k(x, dy) \qquad \sum_{k=1}^{d} \omega_k = 1.$$

  - choose a direction $i \in \{1, \cdots, d\}$ with prob. ω_i.
  - update the component i-th with a $\mathbb{R}$-valued HM (proposal $\mathcal{N}(0, \sigma_i^2)$).

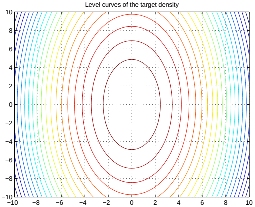
- Under `conditions` $\cdots$

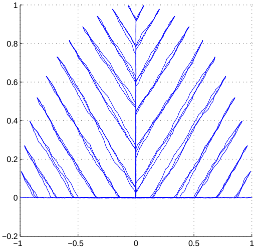
$$h_i(x) = \frac{1}{\sqrt{2\pi}} \; \omega_i \; \sigma_i \; \text{sign}\left(\lim_{r \to \infty} \frac{\nabla_i \ln \pi(rx)}{|\nabla \ln \pi(rx)|} \right).$$

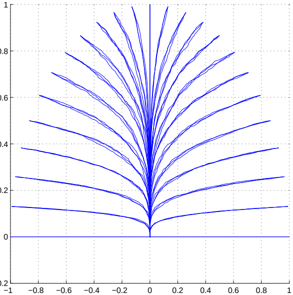
- "Parameters" : $(\omega_k, \sigma_k)_{1 \leq k \leq d}$, for ex.

$$\omega_i = \frac{1}{d} \qquad\qquad \sigma_i = c_i \;\; \text{or} \;\; \sigma_i = \ell \; \lim_{r \to \infty} \frac{|\nabla_i \ln \pi(rx)|}{|\nabla \ln \pi(rx)|}.$$

Example 5 : Gaussian $\mathbb{R}^2$, diagonal dispersion matrix

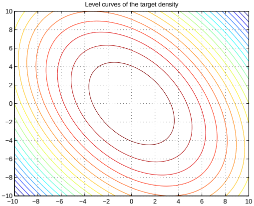


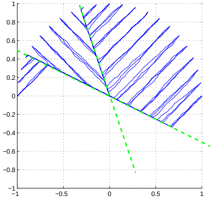


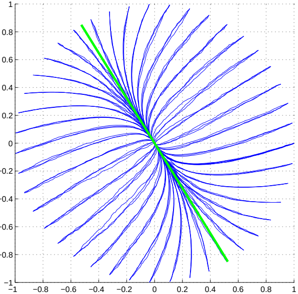


$$\Gamma = \mathrm{diag}(1, 4)$$

Example 6 : Gaussian $\mathbb{R}^2$, non-diagonal dispersion matrix

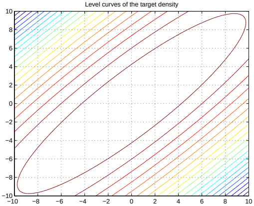


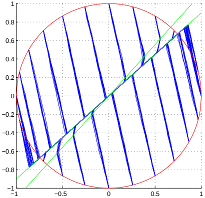


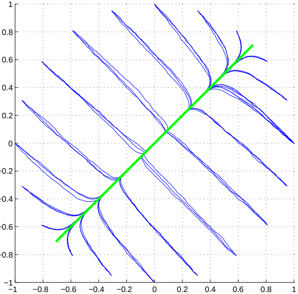


$$\Gamma^{-1} = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$$

Example 7 : Gaussian $\mathbb{R}^2$, non-diagonal dispersion matrix _







$$\Gamma^{-1} \propto \begin{bmatrix} 1 & -0.9 \\ -0.9 & 1 \end{bmatrix} \qquad \sigma_1^2 = 0.5 \qquad \sigma_2^2 = 2/0.9$$

Conclusion

- Existence of fluid limits for `skip free` Markov Chains.
- **[Not Detailed]** Case when for some $0 < \beta < 1$,

$$\eta_r(t;x) = \frac{1}{r}\Phi_{\lfloor tr^{1+\beta} \rfloor},$$

  $\hookrightarrow$ ergodicity at a lower rate.
- Characterization of the limit fluid

- Stable fluid limits $\rightarrow$ Ergodic Markov Chains, but $\cdots$
- more information on the Markov Chain $\cdots$ other normalization (diffusion) ?