Guarantees for the Langevin Monte Carlo sampling algorithm in terms of Wasserstein distance

The problem

Let μ be a probability measure of $\mathbb{R}^p$ with $d\mu = e^{-u}dx$. How to sample from μ? (i.e. obtaining a random variable with measure μ).

The problem

Let μ be a probability measure of $\mathbb{R}^p$ with $d\mu = e^{-u}dx$. How to sample from μ? (i.e. obtaining a random variable with measure μ).

Applications in

- Bayesian statistics
- Volume computation
- ...

(One of) The Solution

The diffusion process solution of

$$dX_t = -\nabla u dt + \sqrt{2} dW_t$$

has stationary measure μ.

(One of) The Solution

The diffusion process solution of

$$dX_t = -\nabla u dt + \sqrt{2} dW_t$$

has stationary measure μ.

$\Rightarrow$ Approximate X_t long enough for it to converge to μ, using the scheme

$$M^{n+1} = M^n - h(n)\nabla u(M^n) + \sqrt{2h(n)}\mathcal{N}_n,$$

with the $\mathcal{N}_n$ i.i.d. normal random variables.

(One of) The Solution

The diffusion process solution of

$$dX_t = -\nabla u dt + \sqrt{2} dW_t$$

has stationary measure μ.

$\Rightarrow$ Approximate X_t long enough for it to converge to μ, using the scheme

$$M^{n+1} = M^n - h(n)\nabla u(M^n) + \sqrt{2h(n)}\mathcal{N}_n,$$

with the $\mathcal{N}_n$ i.i.d. normal random variables.

Problem: good approximation $\Rightarrow$ small h $\Rightarrow$ many iterations to approximate X_t up to time T.
How to choose h? How many iterations do we need to achieve a given accuracy ϵ.

Total variation distance and log-concave μ

Let μ, ν be two measures. The total variation distance between them is given by

$$TV(\nu, \mu) = \max_{\mathcal{B}} |P_\mu(\mathcal{B}) - P_\nu(\mathcal{B})|,$$

where $\mathcal{B}$ is a Borel set.

μ is log-concave if it has density e^{-u} with u convex. This property traditionaly ensures exponential convergence for many quantities quantifying the distance between X_t and μ (Entropy, Fisher information, total variation, Wasserstein distance, ...).

Total variation distance and log-concave μ

Theorem [Dalalyan 2015, fixed h]
Suppose μ is log-concave and ∇u is bounded. Then, one can choose h and T in order to reach an ϵ accuracy in total variation in no more than $O(\epsilon^{-2} p(p^2 + \log^2(1/\epsilon))$ steps.

Total variation distance and log-concave μ

Theorem [Dalalyan 2015, fixed h]
Suppose μ is log-concave and ∇u is bounded. Then, one can choose h and T in order to reach an ϵ accuracy in total variation in no more than $O(\epsilon^{-2} p(p^2 + \log^2(1/\epsilon))$ steps.

Sketch of Proof

$$TV(M^n, \mu) \leq TV(M^n, X_{nh}) + TV(X_{nh}, \mu).$$

Total variation distance and log-concave μ

Theorem [Dalalyan 2015, fixed h]
Suppose μ is log-concave and ∇u is bounded. Then, one can choose h and T in order to reach an ϵ accuracy in total variation in no more than $O(\epsilon^{-2} p(p^2 + \log^2(1/\epsilon))$ steps.

Sketch of Proof

$$TV(M^n, \mu) \leq TV(M^n, X_{nh}) + TV(X_{nh}, \mu).$$

Approximation term.

$$\leq (CpTh)^{1/2}$$

Exponential convergence from log-concavity.

$$\leq \frac{1}{2} e^{C_1 p - C_2 T}$$

Total variation distance and log-concave μ

Theorem [Dalalyan 2015, fixed h]
Suppose μ is log-concave and ∇u is bounded. Then, one can choose h and T in order to reach an ϵ accuracy in total variation in no more than $O(\epsilon^{-2} p(p^2 + \log^2(1/\epsilon))$ steps.

Problems:

- In this analysis, taking too large of a T is harmful.
- In practice, by the discrete nature of the computation, the total variation distance between μ and our approximation is infinite.
- Better rates for diffusion approximation are known for other distances. [Alfonsi, Jourdain, Kohatsu-Higa 2015]

Total variation distance and log-concave μ

Theorem [Dalalyan 2015, fixed h]
Suppose μ is log-concave and ∇u is bounded. Then, one can choose h and T in order to reach an ϵ accuracy in total variation in no more than $O(\epsilon^{-2}p(p^2 + \log^2(1/\epsilon))$ steps.

Disappears if one has access to a *warm start*.

Problems:

- In this analysis, taking too large of a T is harmful.
- In practice, by the discrete nature of the computation, the total variation distance between μ and our approximation is infinite.
- Better rates for diffusion approximation are known for other distances. [Alfonsi, Jourdain, Kohatsu-Higa 2015]

Wasserstein Distance

Let μ and ν be two measures over a metric space (E, d). The Wasserstein disance between these two measures is given by :

$$W_p^p(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \int_{E \times E} d(x, y)^p d\pi(x, y)$$

Where $\Pi(\mu, \nu)$ is the set of measures of $E \times E$ with marginals μ and ν.

Wasserstein Distance

Let μ and ν be two measures over a metric space (E, d). The Wasserstein disance between these two measures is given by :

$$W_p^p(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \int_{E \times E} d(x, y)^p d\pi(x, y)$$

Where $\Pi(\mu, \nu)$ is the set of measures of $E \times E$ with marginals μ and ν.

Theorem [Durmus and Moulines 2015]
Suppose μ is log-concave and ∇u is bounded. Then, one can choose $h(n)$ in order to reach an ϵ accuracy in total variation/W_2 distance in no more than $O(\epsilon^{-2} p)$ steps.

Our assumptions

- μ has a finite second moment.
- The first 2 derivatives of u are Lipschitz continuous with respective Lipschitz constants L_1, L_2, i.e.

$$\forall 1 \leq k \leq 2, \forall x, y \in \mathbb{R}^p, \|\nabla^k u(x) - \nabla^k u(y)\| \leq L_k \|x - y\|.$$

- For any n, $\mathbb{E}[\|M^n\|^2] \leq Cp$.
- There exists $\rho > 0$ such that, if X_0 has finite second moment,

$$W_2(X_t, \mu) \leq e^{-\rho t} W_2(X_0, \mu).$$

Our assumptions

- μ has a finite second moment.
- The first 2 derivatives of u are Lipschitz continuous with respective Lipschitz constants L_1, L_2, i.e.

$$\forall 1 \leq k \leq 2, \forall x, y \in \mathbb{R}^p, \|\nabla^k u(x) - \nabla^k u(y)\| \leq L_k \|x - y\|.$$

- For any n, $\mathbb{E}[\|M^n\|^2] \leq Cp$.
- There exists $\rho > 0$ such that, if X_0 has finite second moment,

$$W_2(X_t, \mu) \leq e^{-\rho t} W_2(X_0, \mu).$$

Last assumption is true if μ is strictly log-concave, but also holds in more general cases, for instance it holds if μ is strictly log-concave outside of a bounded set **Theorem**[Bolley, Gentil, Guilli, 2012].

Main Result

Theorem
For any $n \geq 0$,

$$W_2(M^{n+1}, \mu) \leq Ch(n)^{3/2}\sqrt{p} + e^{-\rho h(n)} W_2(M^n, \mu).$$

For any n sufficiently large,

$$W_2(M^{n+1}, \mu) \leq Ch(n)^2 p + e^{-\rho h(n)} W_2(M^n, \mu).$$

Main Result

Theorem
For any $n \geq 0$,

$$W_2(M^{n+1}, \mu) \leq Ch(n)^{3/2}\sqrt{p} + e^{-\rho h(n)} W_2(M^n, \mu).$$

For any n sufficiently large,

$$W_2(M^{n+1}, \mu) \leq Ch(n)^2 p + e^{-\rho h(n)} W_2(M^n, \mu).$$

Sketch of proof Let X_t continuous process started in M^n.

$$W_2(M^{n+1}, \mu) \leq W_2(M^{n+1}, X_{h(n)}) + W_2(X_{h(n)}, \mu).$$

Main Result

Theorem
For any $n \geq 0$,

$$W_2(M^{n+1}, \mu) \leq Ch(n)^{3/2}\sqrt{p} + e^{-\rho h(n)} W_2(M^n, \mu).$$

For any n sufficiently large,

$$W_2(M^{n+1}, \mu) \leq Ch(n)^2 p + e^{-\rho h(n)} W_2(M^n, \mu).$$

Sketch of proof Let X_t continuous process started in M^n.

$$W_2(M^{n+1}, \mu) \leq W_2(M^{n+1}, X_{h(n)}) + W_2(X_{h(n)}, \mu).$$

Approximation

Exponential convergence to μ

$\leq max(C_1 h(n)^{3/2}\sqrt{p}, C_2(h)h(n)^2 p).$

$\leq e^{-\rho h(n)} W_2(M^n, \mu).$

Main Result

Theorem
For any $n \geq 0$,

$$W_2(M^{n+1}, \mu) \leq Ch(n)^{3/2}\sqrt{p} + e^{-\rho h(n)} W_2(M^n, \mu).$$

For any n sufficiently large,

$$W_2(M^{n+1}, \mu) \leq Ch(n)^2 p + e^{-\rho h(n)} W_2(M^n, \mu).$$

Corollary
For any $\nu > 0$, there exists C such that for $h(n) = (2+\nu)/(n+1)\rho$,

$$W_2(M^n, \mu) \leq Cp/(n+1).$$

Main Result

Theorem
For any $n \geq 0$,

$$W_2(M^{n+1}, \mu) \leq Ch(n)^{3/2}\sqrt{p} + e^{-\rho h(n)} W_2(M^n, \mu).$$

For any n sufficiently large,

$$W_2(M^{n+1}, \mu) \leq Ch(n)^2 p + e^{-\rho h(n)} W_2(M^n, \mu).$$

Corollary
For any $\nu > 0$, there exists C such that for $h(n) = (2+\nu)/(n+1)\rho$,

$$W_2(M^n, \mu) \leq Cp/(n+1).$$

No more than $O(\epsilon^{-1}p)$ steps required to reach an ϵ accuracy.

Sketch of proof

Sketch of proof

For $t < h(n)$, $\bar{X}$ is the continuous Euler's scheme associated to M.

$$\frac{dW_2(X_t, \bar{X}_t)}{dt} \leq L_1 W_2(X_t, \bar{X}_t) + \mathbb{E}[|b(\bar{X}_s) - \mathbb{E}[b(\bar{X}_0)|\bar{X}_s]|^2]^{1/2}$$

Taylor expansion:

$$\begin{aligned} b(\bar{X}_s) - b(\bar{X}_0) &= \nabla b(\bar{X}_s).(\bar{X}_s - \bar{X}_0) \\ &\quad + \left[\int_0^1 \nabla b(v\bar{X}_s + (1-v)\bar{X}_0) - \nabla b(\bar{X}_s) dv\right](\bar{X}_s - \bar{X}_0) \end{aligned}$$

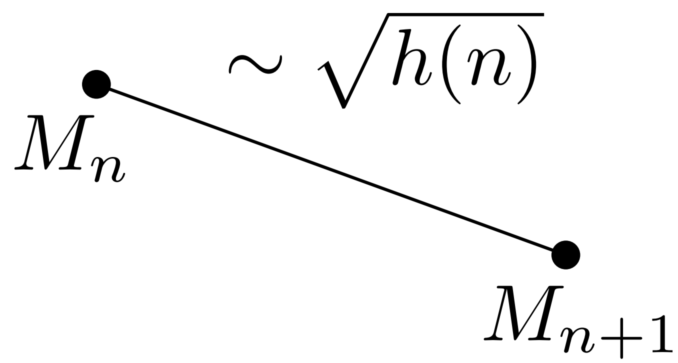
Sketch of proof

Strong error: $\mathbb{E}[|b(\bar{X}(s)) - b(\bar{X}_0)|^2]$.

M_{n+1}

Sketch of proof

Strong error: $\mathbb{E}[|b(\bar{X}(s)) - b(\bar{X}_0)|^2]$.



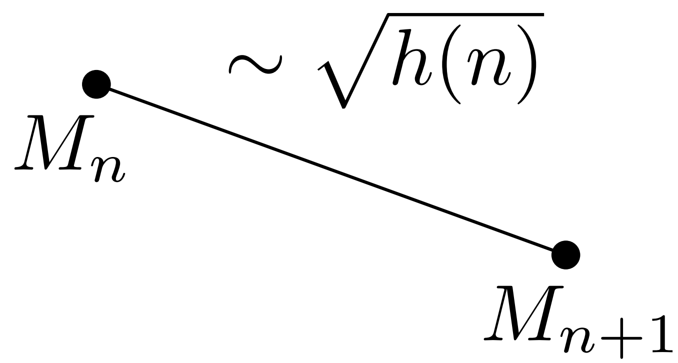
Sketch of proof

Strong error: $\mathbb{E}[|b(\bar{X}(s)) - \mathbb{E}[b(\bar{X}_0)|\bar{X}_s]|^2]$.

M_{n+1}

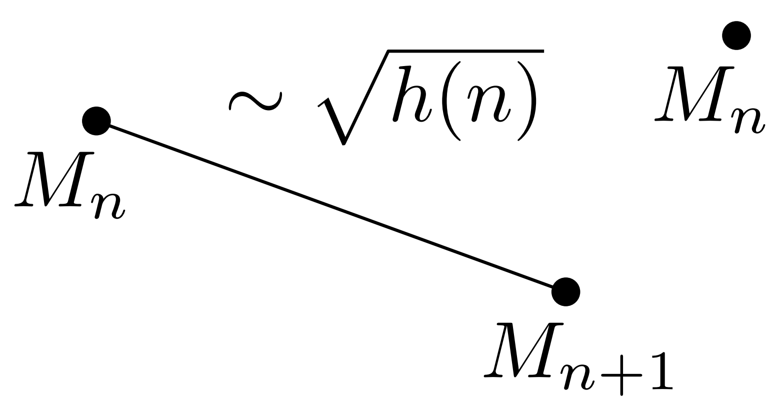
Sketch of proof

Strong error: $\mathbb{E}[|b(\bar{X}(s)) - \mathbb{E}[b(\bar{X}_0)|\bar{X}_s]|^2]$.



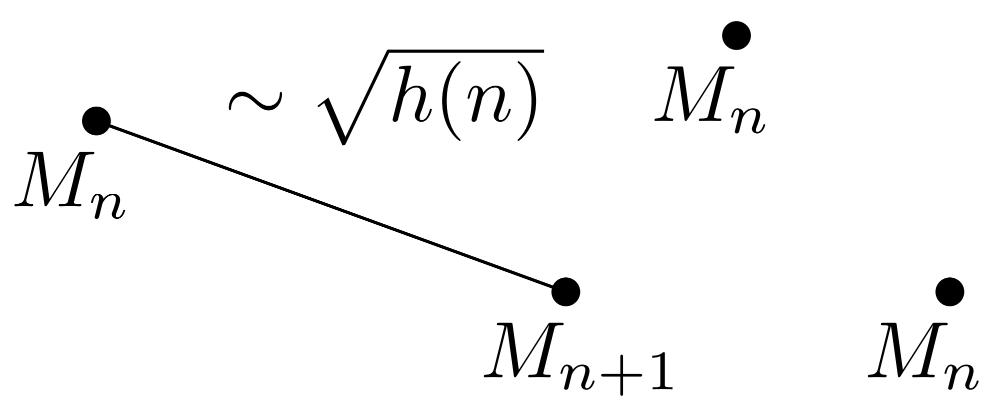
Sketch of proof

Strong error: $\mathbb{E}[|b(\bar{X}(s)) - \mathbb{E}[b(\bar{X}_0)|\bar{X}_s]|^2]$.



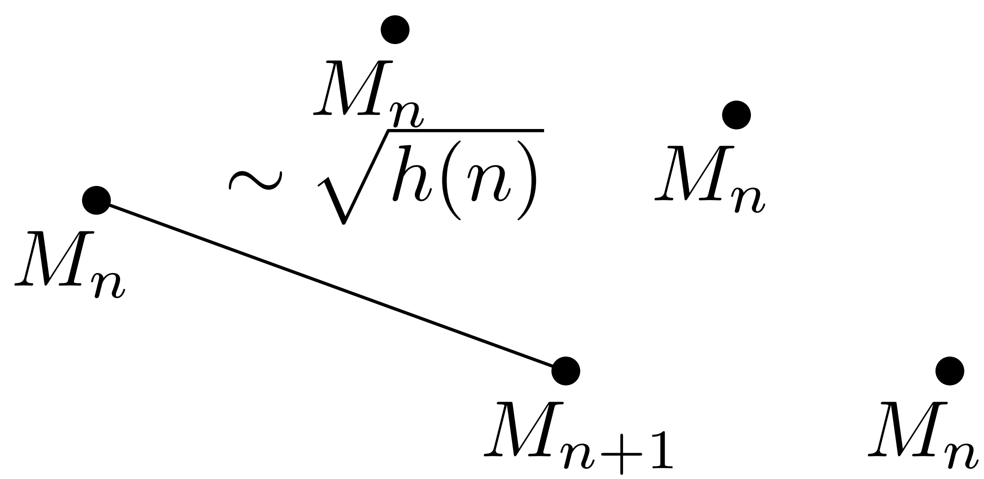
Sketch of proof

Strong error: $\mathbb{E}[|b(\bar{X}(s)) - \mathbb{E}[b(\bar{X}_0)|\bar{X}_s]|^2]$.



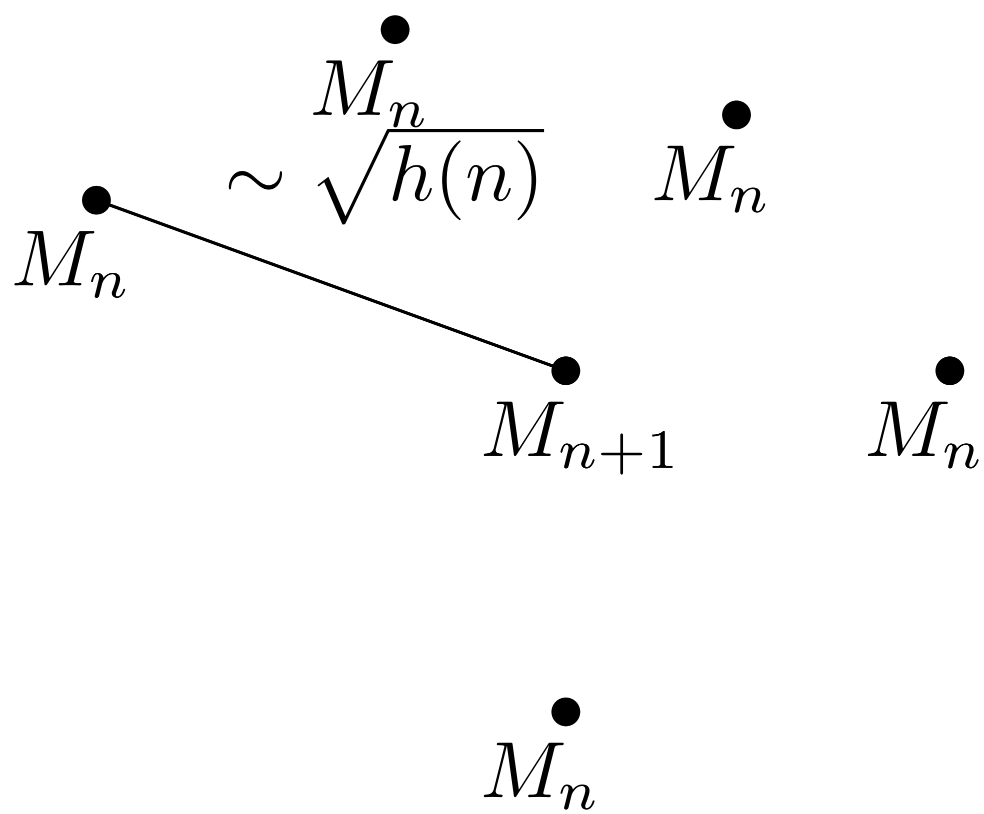
Sketch of proof

Strong error: $\mathbb{E}[|b(\bar{X}(s)) - \mathbb{E}[b(\bar{X}_0)|\bar{X}_s]|^2]$.



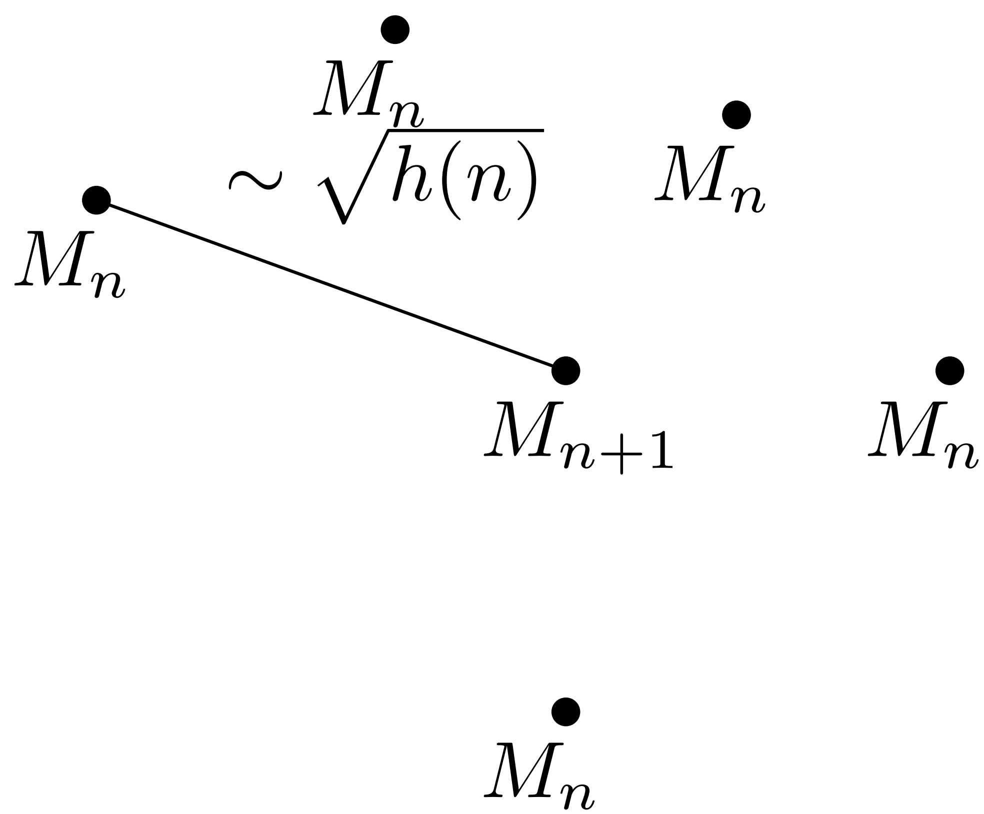
Sketch of proof

Strong error: $\mathbb{E}[|b(\bar{X}(s)) - \mathbb{E}[b(\bar{X}_0)|\bar{X}_s]|^2]$.



Sketch of proof

Strong error: $\mathbb{E}[|b(\bar{X}(s)) - \mathbb{E}[b(\bar{X}_0)|\bar{X}_s]|^2]$.



If the measure of M^n is *smooth enough*, the first term of the Taylor expansion is of order h.

Other Schemes

Other Schemes

Ozaki's discretization scheme: approximates X_t by $\bar{X}_t$ with drift a diffusion process $\bar{X}_t$ with drift

$$b_t(\bar{X}_t) = -\nabla u(\bar{X}_{\tau_t}) - \nabla^2 u(\bar{X}_{\tau_t})(\bar{X}_t - \bar{X}_{\tau_t}),$$

Other Schemes

Ozaki's discretization scheme: approximates X_t by $\bar{X}_t$ with drift a diffusion process $\bar{X}_t$ with drift

$$b_t(\bar{X}_t) = -\nabla u(\bar{X}_{\tau_t}) - \nabla^2 u(\bar{X}_{\tau_t})(\bar{X}_t - \bar{X}_{\tau_t}),$$

Closed form solution: let $B = (I - e^{-h(n)\nabla b})(\nabla b)^{-1} b$ and $\Sigma = (I - e^{-2h(n)\nabla b})(\nabla b)^{-1}$ then $\bar{X}_{X_{\sum_{i=0}^{n-1} h(i)}} = M^n$ with

$$M^{n+1} = M^n + B + \Sigma^{1/2}\mathcal{N}_n.$$

Other Schemes

Ozaki's discretization scheme: approximates X_t by $\bar{X}_t$ with drift a diffusion process $\bar{X}_t$ with drift

$$b_t(\bar{X}_t) = -\nabla u(\bar{X}_{\tau_t}) - \nabla^2 u(\bar{X}_{\tau_t})(\bar{X}_t - \bar{X}_{\tau_t}),$$

Accuracy of ϵ for total variation can be reached in $O^*(\epsilon^{-1}p^{5/2})$ [Dalalyan 2015]

Other Schemes

Ozaki's discretization scheme: approximates X_t by $\bar{X}_t$ with drift a diffusion process $\bar{X}_t$ with drift

$$b_t(\bar{X}_t) = -\nabla u(\bar{X}_{\tau_t}) - \nabla^2 u(\bar{X}_{\tau_t})(\bar{X}_t - \bar{X}_{\tau_t}),$$

Accuracy of ϵ for total variation can be reached in $O^*(\epsilon^{-1}p^{5/2})$ [Dalalyan 2015]

Theorem
For any $n \geq 0$,

$$W_2(M^{n+1}, \mu) \leq e^{h(n)L_1} \frac{L_2}{\sqrt{2}} \left[\frac{h(n)^3 \mathbb{E}[\|M^n\|_2^2]}{3} + \frac{h(n)^2 p}{2} \right] + e^{-\rho h(n)} W_2(M^n, \mu).$$

Other Schemes

Ozaki's discretization scheme: approximates X_t by $\bar{X}_t$ with drift a diffusion process $\bar{X}_t$ with drift

$$b_t(\bar{X}_t) = -\nabla u(\bar{X}_{\tau_t}) - \nabla^2 u(\bar{X}_{\tau_t})(\bar{X}_t - \bar{X}_{\tau_t}),$$

Accuracy of ϵ for total variation can be reached in $O^*(\epsilon^{-1} p^{5/2})$ [Dalalyan 2015]

Theorem
For any $n \geq 0$,

$$W_2(M^{n+1}, \mu) \leq e^{h(n)L_1} \frac{L_2}{\sqrt{2}} \left[\frac{h(n)^3 \mathbb{E}[\|M^n\|_2^2]}{3} + \frac{h(n)^2 p}{2} \right] + e^{-\rho h(n)} W_2(M^n, \mu).$$

$\Rightarrow$ No major improvement on the rate, each iteration is more costly (computing Hessian matrix of u).

Other Schemes

General scheme:

$$M^{n+1} = M^n + \xi_n(M^n)$$

Other Schemes

General scheme:

$$M^{n+1} = M^n + \xi_n(M^n)$$

Rough idea (maybe works?): use Central Limit Theorem for W_2 distance (Bobkov 2013, B. 2015) $\Rightarrow$ the addition of multiple noise converges to a Gaussian $\Rightarrow$ fallback to the Euler's scheme.

Other Schemes

General scheme:

$$M^{n+1} = M^n + \xi_n(M^n)$$

Let π be the stationary measure of M_n, we have

$$W_2(M^n, \mu) \leq W_2(\pi, \mu) + W_2(M^n, \pi)$$

Other Schemes

General scheme:

$$M^{n+1} = M^n + \xi_n(M^n)$$

Let π be the stationary measure of M_n, we have

$$W_2(M^n, \mu) \leq W_2(\pi, \mu) + W_2(M^n, \pi)$$

- μ has a finite second moment.
- The first i derivatives of u are Lipschitz continuous with respective Lipschitz constants $L_1, \dots, L_i$, i.e.

$$\forall 1 \leq k \leq 2, \forall x, y \in \mathbb{R}^p, \|\nabla^k u(x) - \nabla^k u(y)\| \leq L_k \|x - y\|.$$

- $\mathbb{E}_\pi[\|X\|^2] < \infty$.
- μ is strictly ρ log-concave

Other Schemes

General scheme:

$$M^{n+1} = M^n + \xi_n(M^n)$$

Theorem[B. 2015] If μ is log-concave then, for any $T > 0$

$$\begin{aligned} W_2(\pi,\mu) \leq & T\mathbb{E}_\pi[\|\nabla u\|^2]^{1/2} + C(L_2)\sqrt{T}p^{1/2} + \frac{1}{\rho}\mathbb{E}_\pi[\|\mathbb{E}[\frac{\xi}{h} - b]\|^2]^{1/2} \\ & + C(\rho, L_2)\mathbb{E}_\pi[\|\mathbb{E}[\frac{\xi^{\otimes 2}}{2h} - I_d]\|^2]^{1/2} \\ & + \log(T)C(\rho, L_2, L_3)\mathbb{E}_\pi[\|\mathbb{E}[\frac{(\xi)^{\otimes 3}}{6h}]\|^2]^{1/2} \\ & + \sum_{4<k<i} \frac{C(\rho, L_2, \ldots, L_k)}{T^{(k-3)/2}}\mathbb{E}_\pi[\|\mathbb{E}[\frac{(\xi)^{\otimes k}}{k!h}]\|^2]^{1/2} \\ & + \frac{C(\rho, L_2, \ldots, L_i)}{T^{(i-3)/2}}\mathbb{E}_{\pi,\xi}[\frac{\|\xi\|^i}{i!h}]\|^2]^{1/2}. \end{aligned}$$

Other Schemes

General scheme:

$$M^{n+1} = M^n + \xi_n(M^n)$$

Theorem[B. 2015] If μ is log-concave then, for any $T > 0$

$$\begin{aligned} W_2(\pi,\mu) \leq & \boxed{T\mathbb{E}_\pi[\|\nabla u\|^2]^{1/2} + C(L_2)\sqrt{T}p^{1/2}} + \frac{1}{\rho}\mathbb{E}_\pi[\|\mathbb{E}[\frac{\xi}{h} - b]\|^2]^{1/2} \\ & + C(\rho, L_2)\mathbb{E}_\pi[\|\mathbb{E}[\frac{\xi^{\otimes 2}}{2h} - I_d]\|^2]^{1/2} \\ & + \log(T)C(\rho, L_2, L_3)\mathbb{E}_\pi[\|\mathbb{E}[\frac{(\xi)^{\otimes 3}}{6h}]\|^2]^{1/2} \\ & + \sum_{4<k<i} \frac{C(\rho, L_2, \dots, L_k)}{T^{(k-3)/2}}\mathbb{E}_\pi[\|\mathbb{E}[\frac{(\xi)^{\otimes k}}{k!h}]\|^2]^{1/2} \\ & + \frac{C(\rho, L_2, \dots, L_i)}{T^{(i-3)/2}}\mathbb{E}_{\pi,\xi}[\frac{\|\xi\|^i}{i!h}]\|^2]^{1/2}. \end{aligned}$$

Other Schemes

General scheme:

$$M^{n+1} = M^n + \xi_n(M^n)$$

Theorem[B. 2015] If μ is log-concave then, for any $T > 0$

$$\begin{aligned} W_2(\pi,\mu) \leq & T\mathbb{E}_\pi[\|\nabla u\|^2]^{1/2} + C(L_2)\sqrt{T}p^{1/2} + \boxed{\frac{1}{\rho}\mathbb{E}_\pi[\|\mathbb{E}[\frac{\xi}{h} - b]\|^2]^{1/2}} \\ & + C(\rho, L_2)\mathbb{E}_\pi[\|\mathbb{E}[\frac{\xi^{\otimes 2}}{2h} - I_d]\|^2]^{1/2} \\ & + \log(T)C(\rho, L_2, L_3)\mathbb{E}_\pi[\|\mathbb{E}[\frac{(\xi)^{\otimes 3}}{6h}]\|^2]^{1/2} \\ & + \sum_{4<k<i} \frac{C(\rho, L_2, \ldots, L_k)}{T^{(k-3)/2}}\mathbb{E}_\pi[\|\mathbb{E}[\frac{(\xi)^{\otimes k}}{k!h}]\|^2]^{1/2} \\ & + \frac{C(\rho, L_2, \ldots, L_i)}{T^{(i-3)/2}}\mathbb{E}_{\pi,\xi}[\frac{\|\xi\|^i}{i!h}]\|^2]^{1/2}. \end{aligned}$$

Other Schemes

General scheme:

$$M^{n+1} = M^n + \xi_n(M^n)$$

Theorem[B. 2015] If μ is log-concave then, for any $T > 0$

$$\begin{aligned}
W_2(\pi,\mu) \leq & T\mathbb{E}_\pi[\|\nabla u\|^2]^{1/2} + C(L_2)\sqrt{T}p^{1/2} + \frac{1}{\rho}\mathbb{E}_\pi[\|\mathbb{E}[\frac{\xi}{h} - b]\|^2]^{1/2} \\
& + \boxed{C(\rho, L_2)\mathbb{E}_\pi[\|\mathbb{E}[\frac{\xi^{\otimes 2}}{2h} - I_d]\|^2]^{1/2}} \\
& + \log(T)C(\rho, L_2, L_3)\mathbb{E}_\pi[\|\mathbb{E}[\frac{(\xi)^{\otimes 3}}{6h}]\|^2]^{1/2} \\
& + \sum_{4<k<i} \frac{C(\rho, L_2, \ldots, L_k)}{T^{(k-3)/2}}\mathbb{E}_\pi[\|\mathbb{E}[\frac{(\xi)^{\otimes k}}{k!h}]\|^2]^{1/2} \\
& + \frac{C(\rho, L_2, \ldots, L_i)}{T^{(i-3)/2}}\mathbb{E}_{\pi,\xi}[\frac{\|\xi\|^i}{i!h}]\|^2]^{1/2}.
\end{aligned}$$

Other Schemes

General scheme:

$$M^{n+1} = M^n + \xi_n(M^n)$$

Theorem[B. 2015] If μ is log-concave then, for any $T > 0$

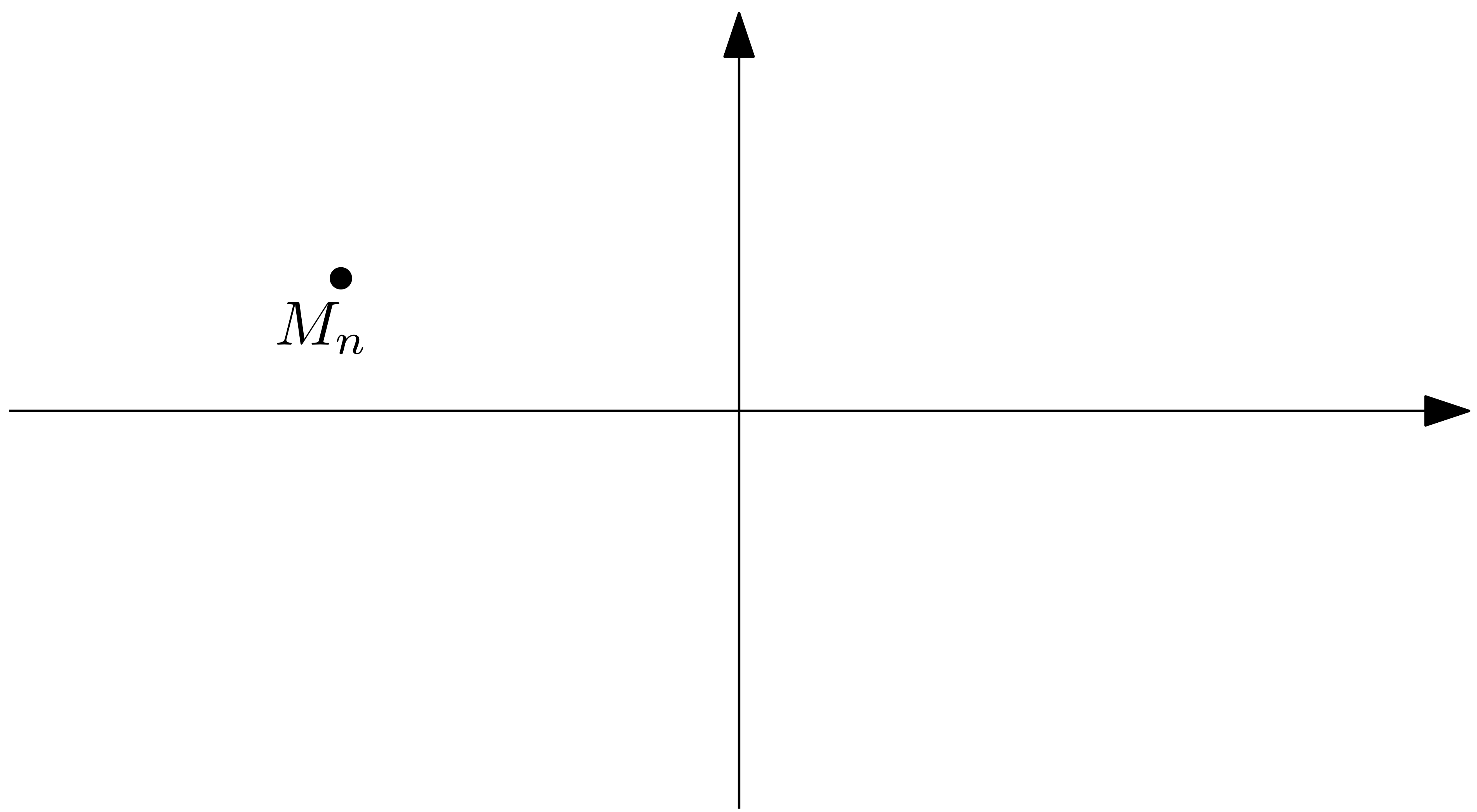
$$\begin{aligned}
W_2(\pi,\mu) \leq & T\mathbb{E}_\pi[\|\nabla u\|^2]^{1/2} + C(L_2)\sqrt{T}p^{1/2} + \frac{1}{\rho}\mathbb{E}_\pi[\|\mathbb{E}[\frac{\xi}{h} - b]\|^2]^{1/2} \\
& + C(\rho, L_2)\mathbb{E}_\pi[\|\mathbb{E}[\frac{\xi^{\otimes 2}}{2h} - I_d]\|^2]^{1/2} \\
& + \log(T)C(\rho, L_2, L_3)\mathbb{E}_\pi[\|\mathbb{E}[\frac{(\xi)^{\otimes 3}}{6h}]\|^2]^{1/2} \\
& + \sum_{4<k<i} \frac{C(\rho, L_2, \ldots, L_k)}{T^{(k-3)/2}}\mathbb{E}_\pi[\|\mathbb{E}[\frac{(\xi)^{\otimes k}}{k!h}]\|^2]^{1/2} \\
& + \frac{C(\rho, L_2, \ldots, L_i)}{T^{(i-3)/2}}\mathbb{E}_{\pi,\xi}[\frac{\|\xi\|^i}{i!h}]\|^2]^{1/2}.
\end{aligned}$$

Other Schemes: an example

$\xi_n = e_{I_n}.(-hp\nabla u + \sqrt{2hp}B_n)$ where I_n are independent uniform random variable on $[|1, p|]$, B_n are independent Bernoulli random variable and the e_i are the canonical basis of $\mathbb{R}^p$.
Computations on a single direction $\Rightarrow$ cost of an iteration is $O(1)$ compared to $O(p)$ for the Euler's scheme.

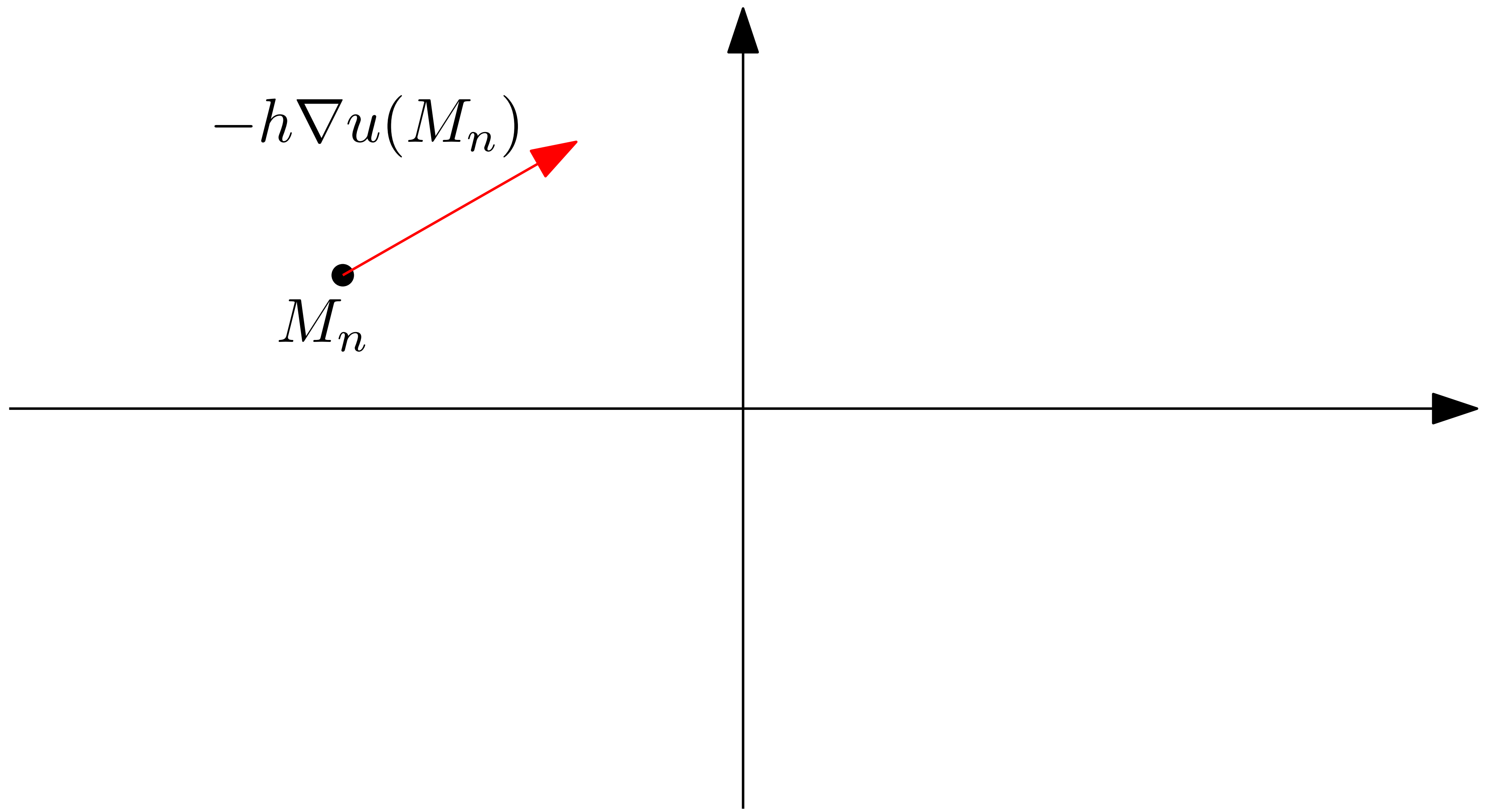
Other Schemes: an example

$\xi_n = e_{I_n}.(-hp\nabla u + \sqrt{2hp}B_n)$ where I_n are independent uniform random variable on $[|1,p|]$, B_n are independent Bernoulli random variable and the e_i are the canonical basis of $\mathbb{R}^p$.
Computations on a single direction $\Rightarrow$ cost of an iteration is $O(1)$ compared to $O(p)$ for the Euler's scheme.



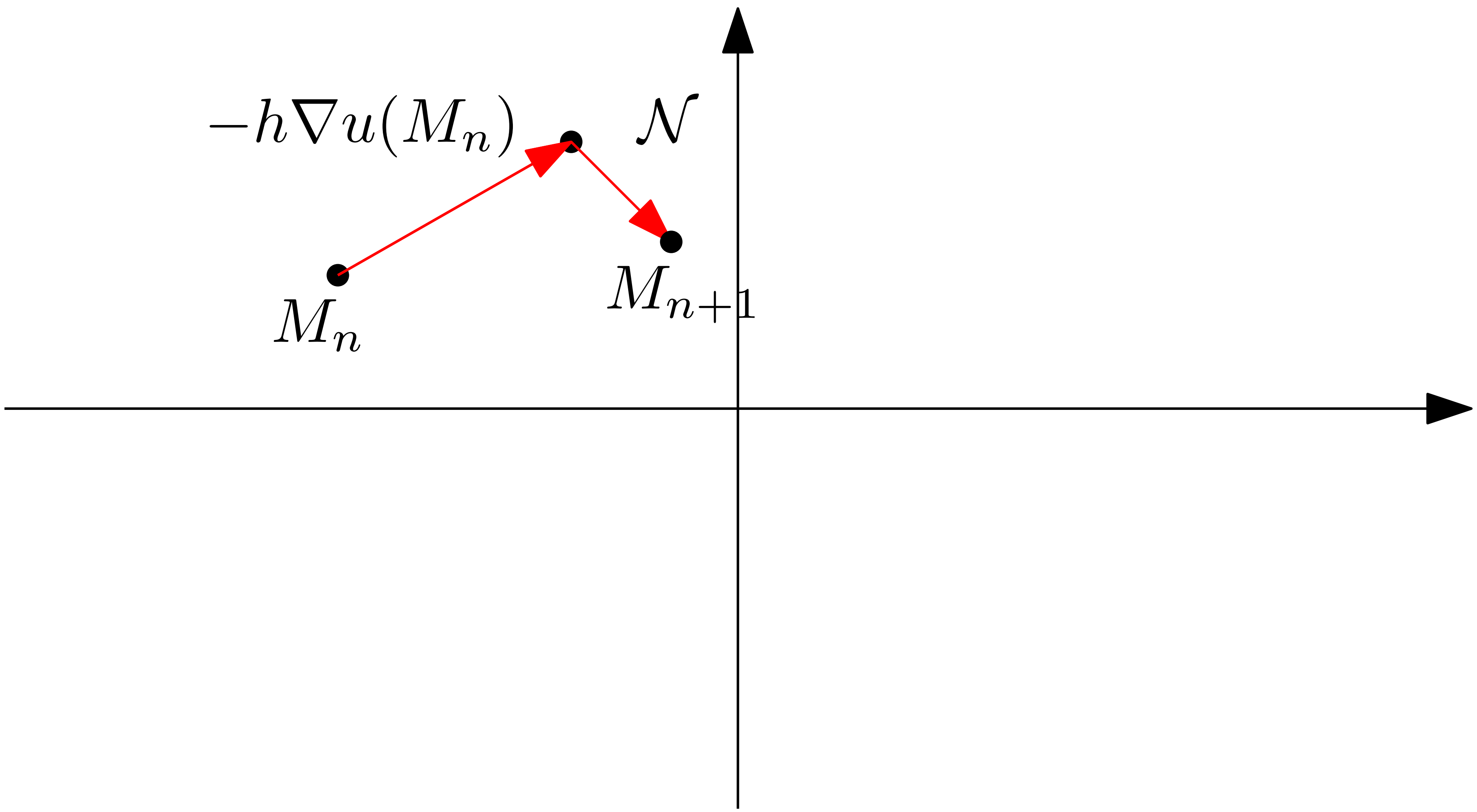
Other Schemes: an example

$\xi_n = e_{I_n}.(-hp\nabla u + \sqrt{2hp}B_n)$ where I_n are independent uniform random variable on $[|1,p|]$, B_n are independent Bernoulli random variable and the e_i are the canonical basis of $\mathbb{R}^p$.
Computations on a single direction $\Rightarrow$ cost of an iteration is $O(1)$ compared to $O(p)$ for the Euler's scheme.



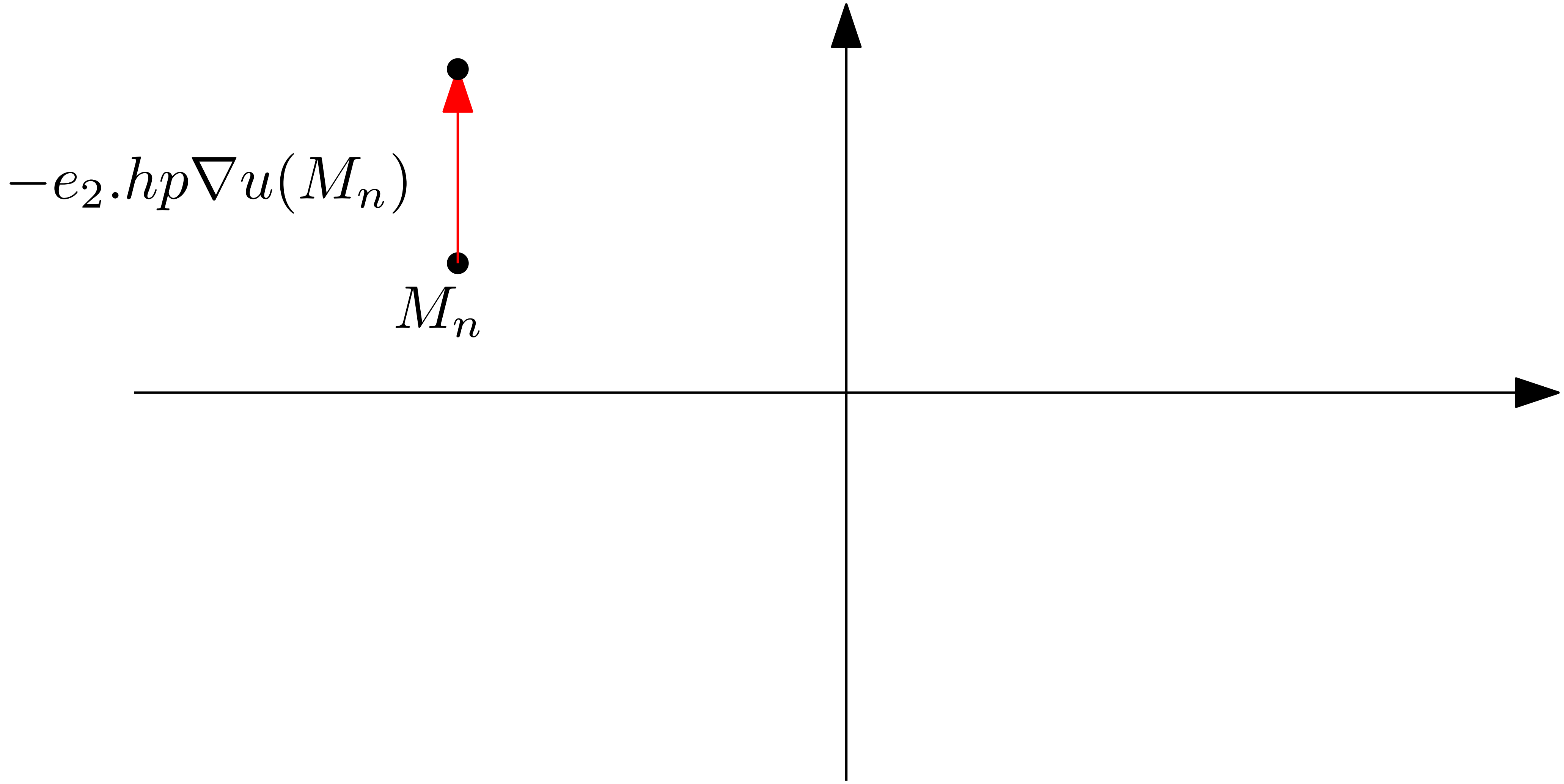
Other Schemes: an example

$\xi_n = e_{I_n}.(-hp\nabla u + \sqrt{2hp}B_n)$ where I_n are independent uniform random variable on $[|1,p|]$, B_n are independent Bernoulli random variable and the e_i are the canonical basis of $\mathbb{R}^p$.
Computations on a single direction $\Rightarrow$ cost of an iteration is $O(1)$ compared to $O(p)$ for the Euler's scheme.



Other Schemes: an example

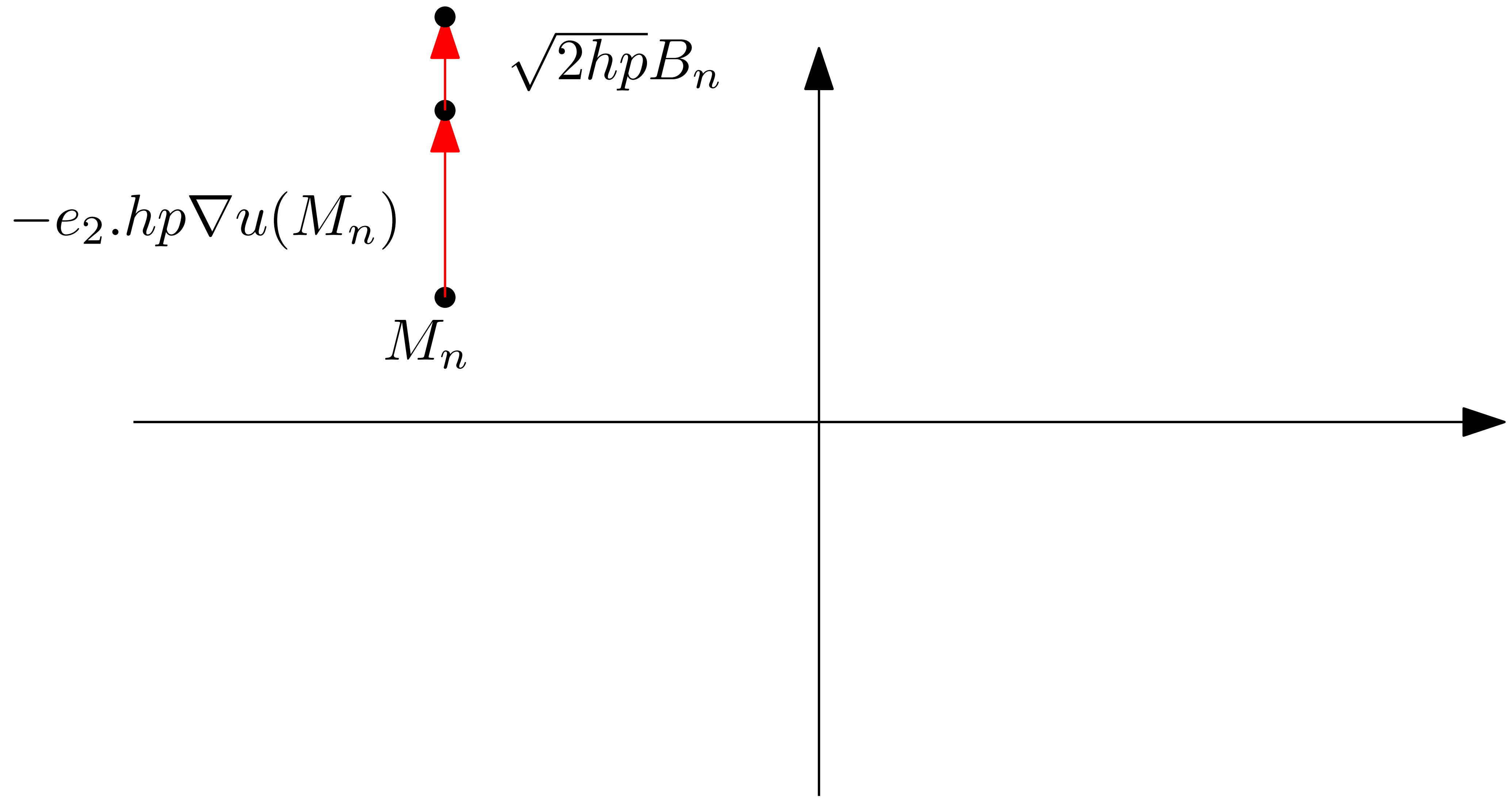
$\xi_n = e_{I_n}.(-hp\nabla u+\sqrt{2hp}B_n)$ where I_n are independent uniform random variable on $[|1,p|]$, B_n are independent Bernoulli random variable and the e_i are the canonical basis of $\mathbb{R}^p$.
Computations on a single direction $\Rightarrow$ cost of an iteration is $O(1)$ compared to $O(p)$ for the Euler's scheme.



Other Schemes: an example

$\xi_n = e_{I_n}.(-hp\nabla u + \sqrt{2hp}B_n)$ where I_n are independent uniform random variable on $[|1,p|]$, B_n are independent Bernoulli random variable and the e_i are the canonical basis of $\mathbb{R}^p$.

Computations on a single direction $\Rightarrow$ cost of an iteration is $O(1)$ compared to $O(p)$ for the Euler's scheme.



Other Schemes: an example

$\xi_n = e_{I_n}.(-hp\nabla u + \sqrt{2hp}B_n)$ where I_n are independent uniform random variable on $[|1,p|]$, B_n are independent Bernoulli random variable and the e_i are the canonical basis of $\mathbb{R}^p$.
Computations on a single direction $\Rightarrow$ cost of an iteration is $O(1)$ compared to $O(p)$ for the Euler's scheme.

Theorem Assume $h \leq \frac{\sqrt{\rho}}{\sqrt{p}L_1}$. Then, there exists a constant C such that

$$W_2(M^n,\mu) \leq (1-2h\rho+h^2L_1^2p)^{n/2}\frac{2hp}{2\rho h - pL_1^2h^2} + Ch^{1/2}p^{1+1/(4(i-2))}.$$

Other Schemes: an example

$\xi_n = e_{I_n}.(-hp\nabla u + \sqrt{2hp}B_n)$ where I_n are independent uniform random variable on $[|1,p|]$, B_n are independent Bernoulli random variable and the e_i are the canonical basis of $\mathbb{R}^p$.
Computations on a single direction $\Rightarrow$ cost of an iteration is $O(1)$ compared to $O(p)$ for the Euler's scheme.

Theorem Assume $h \leq \frac{\sqrt{\rho}}{\sqrt{p}L_1}$. Then, there exists a constant C such that

$$W_2(M^n, \mu) \leq (1 - 2h\rho + h^2 L_1^2 p)^{n/2} \frac{2hp}{2\rho h - pL_1^2 h^2} + Ch^{1/2} p^{1+1/(4(i-2))}.$$

$\Rightarrow$ needs no more than $O^*(\epsilon^{-2} p^{2+1/2(i-1)})$ (we conjecture $O^*(\epsilon^{-1} p^3)$) steps to reach an ϵ-accuracy. Still worse than Euler's Scheme.

Conclusion

- Linear rates (accuracy/dimension) for the LMC algorithm.
- Heuristic to choose h.
- Ozaki's discretization not interesting in higher dimension.
- Able to cope with general schemes.

- Sampling from a manifold.
- Dealing with measures on a convex set [Bubeck, Lehec, Eldan 2015].
- General schemes are not tight yet.
- Applications to Stochastic Gradient Descent.

Bibliography

- Alfonsi, Jourdain, Kohatsu-Higa 2015, Optimal transport bounds between the time-marginals of a multidimensional diffusion and its Euler scheme.
- Bubeck, Lehec, Eldan 2015, Sampling from a log-concave distribution with Projected Langevin Monte Carlo.
- Bobkov 2013, Entropic approach to E. Rio' central limit theorem for W_2 transport distance.
- Bolley, Gentil, Guillin 2012, Convergence to equilibrium in Wasserstein distance for Fokker-Planck equations.
- Dalalyan 2015, Theoretical guarantees for approximate sampling from smooth and log-concave densities.
- Durmus and Moulines 2015, Non-asymptotic convergence analysis for the Unadjusted Langevin Algorithm.
- Bonis 2015, Stable measures and Stein's method: rates in the Central Limit Theorem and diffusion approximation.