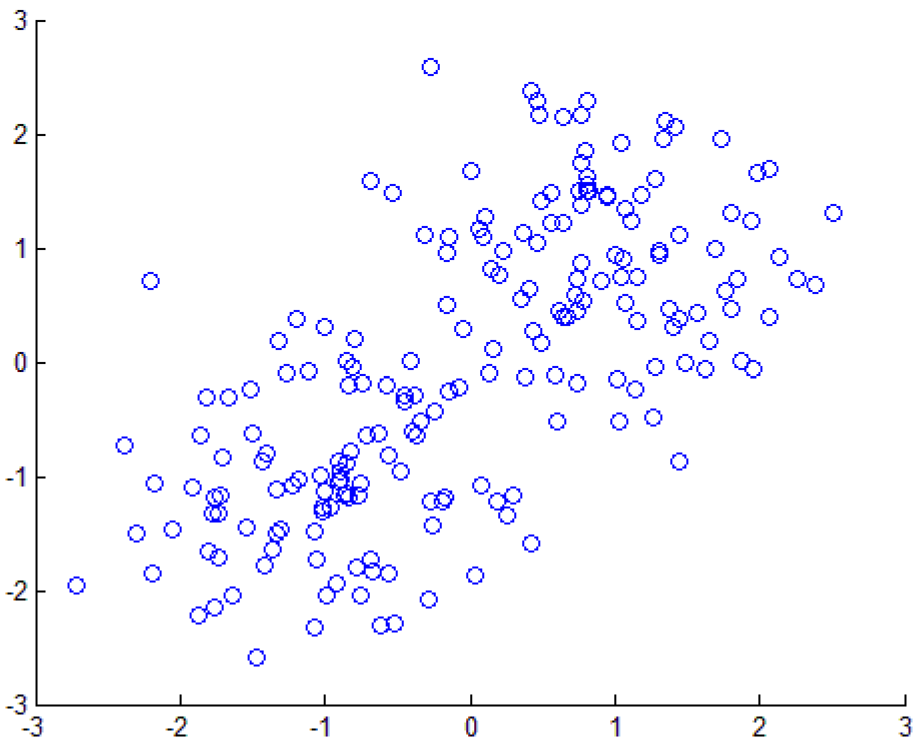
Stein's method for diffusion approximation

Thomas Bonis

DataShape team, Inria

K-nearest-neighbor graph

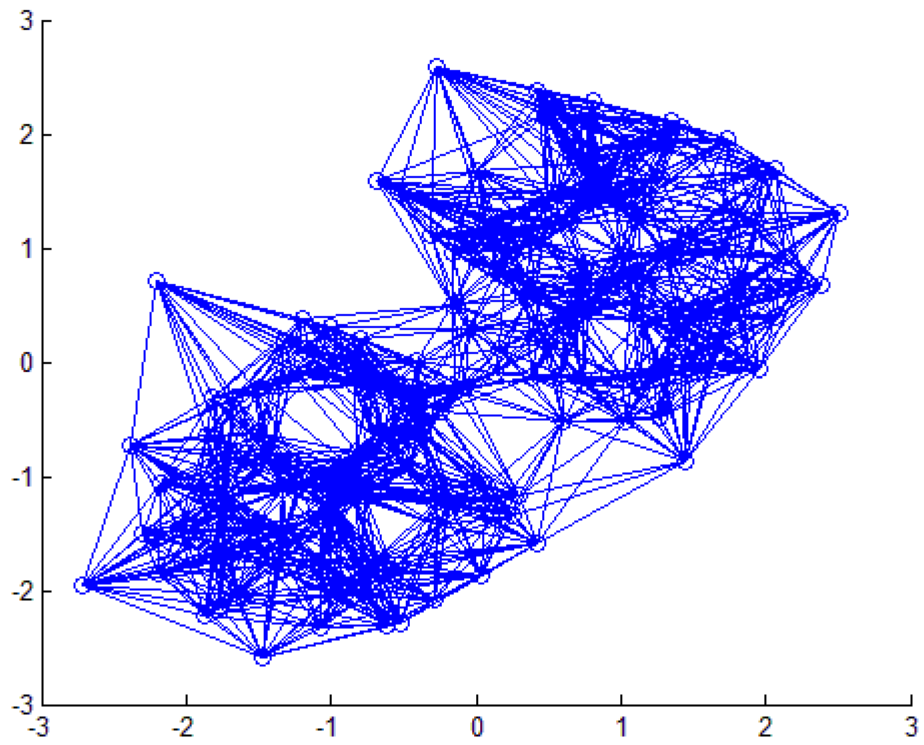
We draw n points in $\mathbb{R}^d$, $X_1, \ldots, X_n \sim d\nu = f d\lambda$



K-nearest-neighbor graph

We draw n points in $\mathbb{R}^d$, $X_1, \ldots, X_n \sim d\nu = f d\lambda$

We add an edge between each point and its K-nearest-neighbors

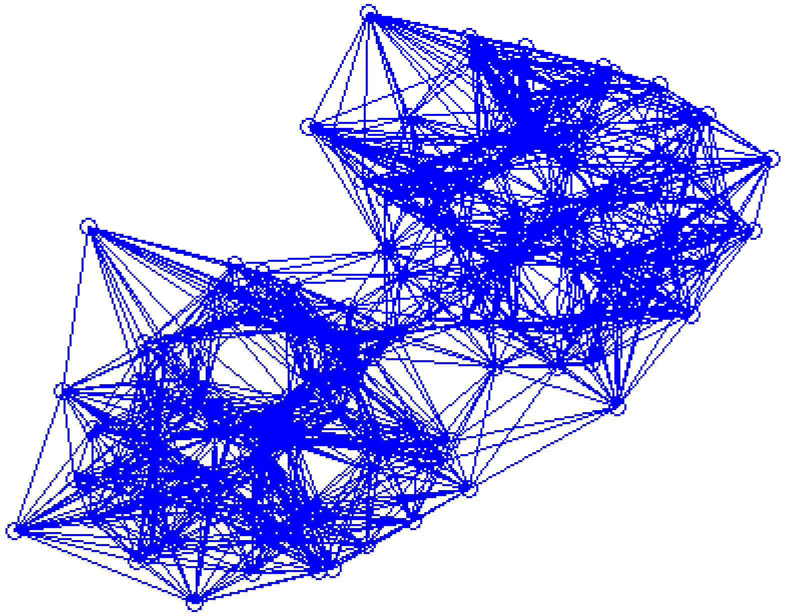


K-nearest-neighbor graph

We draw n points in $\mathbb{R}^d$, $X_1, \ldots, X_n \sim d\nu = f d\lambda$

We add an edge between each point and its K-nearest-neighbors

When $K, n \to \infty, K/n \to 0$ (+other condition on K), can we recover f using only the graph structure?



Random walk on K-nearest-neighbor graph

A random walk on the graph captures information on ν.

Random walk on K-nearest-neighbor graph

A random walk on the graph captures information on ν.

The random walk approximates the diffusion process with generator

$$f^{-2/d}(\nabla(\log f).\nabla + \frac{1}{2}\Delta).$$

Random walk on K-nearest-neighbor graph

A random walk on the graph captures information on ν.

The random walk approximates the diffusion process with generator

$$f^{-2/d}(\nabla(\log f).\nabla + \frac{1}{2}\Delta).$$

Diffusion has invariant measure μ with density proportional to $f^{2+2/d}$.

Random walk on K-nearest-neighbor graph

A random walk on the graph captures information on ν.

The random walk approximates the diffusion process with generator

$$f^{-2/d}(\nabla(\log f).\nabla + \frac{1}{2}\Delta).$$

Diffusion has invariant measure μ with density proportional to $f^{2+2/d}$.

Does π, the invariant measure of the random walk, converge to μ ?

Random walk on ϵ-graph

Edge between X_i and X_j if $\|X_i - X_j\| \leq \epsilon$.

The random walk approximates the diffusion process with generator

$$\nabla(\log f).\nabla + \frac{1}{2}\Delta.$$

Diffusion has invariant measure μ with density proportional to f^2.

$\pi(X_i)$ proportional to the degree of X_i (the ball density estimator $\to f$). Since we have more points where f is large, π converges to a measure with density proportional to f^2.

Stein discrepancy

Let γ be the gaussian measure Z be drawn from γ. Then, $\forall \phi$,

$$\mathbb{E}[-Z.\nabla\phi(Z) + \Delta\phi(Z)] = 0.$$

Stein discrepancy

Let γ be the gaussian measure Z be drawn from γ. Then, $\forall \phi$,

$$\mathbb{E}[-Z.\nabla\phi(Z) + \Delta\phi(Z)] = 0.$$

Let X be drawn from ν. We say that a measure ν admits a Stein kernel τ_ν with respect to γ if there exists τ_ν such tha, $\forall \phi$,

$$\mathbb{E}[-X.\nabla\phi(X) + < \tau_\nu(X), Hess(\phi)(X)) >_{HS}] = 0.$$

Stein discrepancy

Let γ be the gaussian measure Z be drawn from γ. Then, $\forall \phi$,

$$\mathbb{E}[-Z.\nabla\phi(Z) + \Delta\phi(Z)] = 0.$$

Let X be drawn from ν. We say that a measure ν admits a Stein kernel τ_ν with respect to γ if there exists τ_ν such tha, $\forall \phi$,

$$\mathbb{E}[-X.\nabla\phi(X) + < \tau_\nu(X), Hess(\phi)(X)) >_{HS}] = 0.$$

Intuitively, if τ_ν is close to I_d then ν is close to γ. The distance between τ_ν and I_d is quantified by:

$$S(\nu,\mu)^2 = \mathbb{E}[\|\tau_\nu(X) - I_d\|^2].$$

Bounding the Wasserstein distance with S

Bounding the Wasserstein distance with S

Theorem [Ledoux, Nourdin, Peccati 2015]

Let ν be a measure admitting a Stein kernel τ_ν and let $S(\nu)$ be the associated Stein discrepancy. We have:

$$W_2(\nu, \gamma) \leq S(\nu)$$

Bounding the Wasserstein distance with S

Theorem [Ledoux, Nourdin, Peccati 2015]

Let ν be a measure admitting a Stein kernel τ_ν and let $S(\nu)$ be the associated Stein discrepancy. We have:

$$W_2(\nu, \gamma) \leq S(\nu)$$

Problem: in the general case, discrete measures do not admit a Stein kernel.
Example: if the Rademacher measure admited a Stein kernel, there would exist τ such that for any smooth function ϕ

$$\phi'(-1) - \phi'(1) + \tau(1)\phi''(1) + \tau(-1)\phi''(-1) = 0.$$

Can be dealt with using a smoothing procedure (relying on the zero-bias distribution), but not practical in high dimensions.

Generalizing the Stein kernel

$X \sim \nu$. There exists an operator $\mathcal{L}$ such that

$$\forall \phi, \mathbb{E}[\mathcal{L}\phi(X)] = 0,$$

compare $\mathcal{L}$ with $-x.\nabla + \Delta$.

Generalizing the Stein kernel

$X \sim \nu$. There exists an operator $\mathcal{L}$ such that

$$\forall \phi, \mathbb{E}[\mathcal{L}\phi(X)] = 0,$$

compare $\mathcal{L}$ with $-x.\nabla + \Delta$.

Let X and X' be drawn from ν, then $\forall \phi$,

$$\mathbb{E}[\phi(X) - \phi(X')] = 0,$$

and by a Taylor-Expansion

$$\mathbb{E}[\sum_{k=0}^{\infty} \frac{\mathbb{E}[X'^k]}{k!} \phi^{(k)}] = 0.$$

Another bound on W_2

Theorem[Dimension 1]
Let ν be a measure of $\mathbb{R}$ and r.v. X, $(X_t)_{t\geq 0}$ drawn from ν. Let $Y_t = X_t - X$, for any $h > 0$,

$$\begin{aligned} W_2(\nu,\gamma) \leq & \int_0^\infty e^{-t}\mathbb{E}[(\frac{1}{h}\mathbb{E}[Y_t|X] + X)^2]^{1/2}dt \\ & + \int_0^\infty \frac{e^{-2t}}{\sqrt{1-e^{-2t}}}\mathbb{E}[(\frac{1}{h}\frac{\mathbb{E}[Y_t^2|X]}{2} - 1)^2]^{1/2}dt \\ & + \sum_{k>2}\int_0^\infty \frac{e^{-kt}}{h\sqrt{k!}(1-e^{-2t})^{(k-1)/2}}\mathbb{E}[\mathbb{E}[Y_t^k|X]^2]^{1/2}dt \end{aligned}$$

Another bound on W_2

Theorem[Dimension 1]
Let ν be a measure of $\mathbb{R}$ and r.v. X, $(X_t)_{t\geq 0}$ drawn from ν. Let $Y_t = X_t - X$, for any $h > 0$,

Rescaling

$$W_2(\nu,\gamma) \leq \int_0^\infty e^{-t}\mathbb{E}[(\frac{1}{h}\mathbb{E}[Y_t|X] + X)^2]^{1/2}dt$$
$$+ \int_0^\infty \frac{e^{-2t}}{\sqrt{1-e^{-2t}}}\mathbb{E}[(\frac{1}{h}\frac{\mathbb{E}[Y_t^2|X]}{2} - 1)^2]^{1/2}dt$$
$$+ \sum_{k>2}\int_0^\infty \frac{e^{-kt}}{h\sqrt{k!}(1-e^{-2t})^{(k-1)/2}}\mathbb{E}[\mathbb{E}[Y_t^k|X]^2]^{1/2}dt$$

Another bound on W_2

Theorem[Dimension 1]
Let ν be a measure of $\mathbb{R}$ and r.v. X, $(X_t)_{t\geq 0}$ drawn from ν. Let $Y_t = X_t - X$, for any $h > 0$,

First moment close to $-X$

Rescaling

$$W_2(\nu,\gamma) \leq \int_0^\infty e^{-t}\mathbb{E}[(\frac{1}{h}\mathbb{E}[Y_t|X] + X)^2]^{1/2}dt$$
$$+ \int_0^\infty \frac{e^{-2t}}{\sqrt{1-e^{-2t}}}\mathbb{E}[(\frac{1}{h}\frac{\mathbb{E}[Y_t^2|X]}{2} - 1)^2]^{1/2}dt$$
$$+ \sum_{k>2}\int_0^\infty \frac{e^{-kt}}{h\sqrt{k!}(1-e^{-2t})^{(k-1)/2}}\mathbb{E}[\mathbb{E}[Y_t^k|X]^2]^{1/2}dt$$

Another bound on W_2

Theorem[Dimension 1]
Let ν be a measure of $\mathbb{R}$ and r.v. X, $(X_t)_{t\geq 0}$ drawn from ν. Let $Y_t = X_t - X$, for any $h > 0$,

First moment close to $-X$

Rescaling

$$W_2(\nu,\gamma) \leq \int_0^\infty e^{-t}\mathbb{E}[(\frac{1}{h}\mathbb{E}[Y_t|X] + X)^2]^{1/2}dt$$

Second moment close to 1.

$$+ \int_0^\infty \frac{e^{-2t}}{\sqrt{1-e^{-2t}}}\mathbb{E}[(\frac{1}{h}\frac{\mathbb{E}[Y_t^2|X]}{2} - 1)^2]^{1/2}dt$$

$$+ \sum_{k>2}\int_0^\infty \frac{e^{-kt}}{h\sqrt{k!}(1-e^{-2t})^{(k-1)/2}}\mathbb{E}[\mathbb{E}[Y_t^k|X]^2]^{1/2}dt$$

Another bound on W_2

Theorem[Dimension 1]
Let ν be a measure of $\mathbb{R}$ and r.v. X, $(X_t)_{t\geq 0}$ drawn from ν. Let $Y_t = X_t - X$, for any $h > 0$,

First moment close to $-X$

Rescaling

$$W_2(\nu,\gamma) \leq \int_0^\infty e^{-t}\mathbb{E}[(\frac{1}{h}\mathbb{E}[Y_t|X] + X)^2]^{1/2}dt$$

Second moment close to 1.

$$+ \int_0^\infty \frac{e^{-2t}}{\sqrt{1-e^{-2t}}}\mathbb{E}[(\frac{1}{h}\frac{\mathbb{E}[Y_t^2|X]}{2} - 1)^2]^{1/2}dt$$

$$+ \sum_{k>2}\int_0^\infty \frac{e^{-kt}}{h\sqrt{k!}(1-e^{-2t})^{(k-1)/2}}\mathbb{E}[\mathbb{E}[Y_t^k|X]^2]^{1/2}dt$$

Higher moments small, start at 0 and grow as t increases.

Roughly, Y_t has to be bounded by $\sqrt{t}$.

Another bound on W_2

Theorem[Dimension 1]
Let ν be a measure of $\mathbb{R}$ and r.v. X, $(X_t)_{t\geq 0}$ drawn from ν. Let $Y_t = X_t - X$, for any $h > 0$,

$$
\begin{aligned}
W_2(\nu,\gamma) \leq & \int_0^\infty e^{-t}\mathbb{E}[(\frac{1}{h}\mathbb{E}[Y_t|X] + X)^2]^{1/2} dt \\
& + \int_0^\infty \frac{e^{-2t}}{\sqrt{1-e^{-2t}}}\mathbb{E}[(\frac{1}{h}\frac{\mathbb{E}[Y_t^2|X]}{2} - 1)^2]^{1/2} dt \\
& + \sum_{k>2}\int_0^\infty \frac{e^{-kt}}{h\sqrt{k!}(1-e^{-2t})^{(k-1)/2}}\mathbb{E}[\mathbb{E}[Y_t^k|X]^2]^{1/2} dt
\end{aligned}
$$

Similar result (in dimension 1 only !) for $W_p, p \geq 1$.

Another bound on W_2

Let μ be the invariant measure of an operator

$$\mathcal{L} = b.\nabla + < a, \nabla^2 > .$$

Under technical conditions on $\mathcal{L}$ (a curvature dimension inequality), a similar result holds.

Another bound on W_2

Let μ be the invariant measure of an operator

$$\mathcal{L} = b.\nabla + < a, \nabla^2 > .$$

Under technical conditions on $\mathcal{L}$ (a curvature dimension inequality), a similar result holds.

A similar result holds under technical conditions on $\mathcal{L}$ (for example under a curvature dimension inequality). If

- $\mathbb{E}[Y_t|X]$ is close to $b(X)$.
- $\mathbb{E}[Y_t^2|X]$ is close to $a(X)$.
- $\mathbb{E}[Y_t^k]$ are small for $k > 2$.

then $W_2(\nu, \mu)$ is small.

Convergence rates in the Central Limit Theorem

Consider i.i.d random variables $X_1, \ldots, X_n$ with measure ν and $\mathbb{E}[X_1] = 0, \mathbb{E}[X_1^2] = 1$. The Central Limit Theorem gives

$$S_n = n^{-1/2} \sum_{i=1}^{n} \to \mathcal{N}(0,1).$$

How fast does it converge?

Convergence rates in the Central Limit Theorem

Consider i.i.d random variables $X_1, \ldots, X_n$ with measure ν and $\mathbb{E}[X_1] = 0, \mathbb{E}[X_1^2] = 1$. The Central Limit Theorem gives

$$S_n = n^{-1/2} \sum_{i=1}^{n} \to \mathcal{N}(0, 1).$$

How fast does it converge?

Let $\tilde{X}_1, \ldots, \tilde{X}_n$ i.i.d. copies of $X_1, \ldots, X_n$ and I a uniform r.v. on $1, \ldots, n$. We pose

$$(S_n)_t = S_n + n^{-1/2}(\tilde{X}_I - X_I).$$

Convergence rates in the Central Limit Theorem

Consider i.i.d random variables $X_1, \ldots, X_n$ with measure ν and $\mathbb{E}[X_1] = 0, \mathbb{E}[X_1^2] = 1$. The Central Limit Theorem gives

$$S_n = n^{-1/2} \sum_{i=1}^{n} \to \mathcal{N}(0,1).$$

How fast does it converge?

Let $\tilde{X}_1, \ldots, \tilde{X}_n$ i.i.d. copies of $X_1, \ldots, X_n$ and I a uniform r.v. on $1, \ldots, n$. We pose

$$(S_n)_t = S_n + n^{-1/2}(\tilde{X}_I - X_I) 1_{X_i, \tilde{X}_i \in [-\sqrt{tn}, \sqrt{tn}]}.$$

Convergence rates in the Central Limit Theorem

Theorem

Consider $X_1, ..., X_n$ i.i.d random variables in $\mathbb{R}^d$ and let ν_n be the measure of $n^{-1/2}\sum_{i=1}^n X_i$. If X_1 admits a moment of order $p+m$ (that is $\mathbb{E}[\|X_1\|_2^{p+m} < \infty$) for some $m \in [0,2]$ then

$$W_2(\nu_n, \gamma) = O(n^{-1/2+(2-m)/4}).$$

Moreover, if $d = 1$ then for any $p \geq 2$, if X_1 admits a finite moment of order $p+m$ for some $m \in [0,2]$ then

$$W_p(\nu_n, \gamma) = O(n^{-1/2+(2-m)/2p}).$$

Convergence rates in the Central Limit Theorem

Theorem

Consider $X_1, ..., X_n$ i.i.d random variables in $\mathbb{R}^d$ and let ν_n be the measure of $n^{-1/2}\sum_{i=1}^n X_i$. If X_1 admits a moment of order $p+m$ (that is $\mathbb{E}[\|X_1\|_2^{p+m} < \infty)$ for some $m \in [0,2]$ then

$$W_2(\nu_n, \gamma) = O(n^{-1/2+(2-m)/4}).$$

Moreover, if $d=1$ then for any $p \geq 2$, if X_1 admits a finite moment of order $p+m$ for some $m \in [0,2]$ then

$$W_p(\nu_n, \gamma) = O(n^{-1/2+(2-m)/2p}).$$

- $p > 2, m = 0$ proved by Sakhanenko (1985).
- $p \in [1,2], m = 2$ proved by Rio (2009).
- $p = 2, m = 2$ proved by Bobkov (2013) by other means, can be extended to higher dimensions at the cost of stronger assumptions.
- $p > 2$ also proved by Bobkov (2016).

Convergence of Markov chains towards diffusion processes

Consider a Markov chain (X_n) with invariant measure π and transition kernel K approximating a diffusion process with operator

$$\mathcal{L} = b.\nabla + < a, \nabla^2 >$$

and invariant measure μ.

Convergence of Markov chains towards diffusion processes

Consider a Markov chain (X_n) with invariant measure π and transition kernel K approximating a diffusion process with operator

$$\mathcal{L} = b.\nabla + < a, \nabla^2 >$$

and invariant measure μ.
X is drawn from π, ξ is a jump from X.

$$X_t = 1_{t \geq T} \xi + X.$$

Convergence of Markov chains towards diffusion processes

Consider a Markov chain (X_n) with invariant measure π and transition kernel K approximating a diffusion process with operator

$$\mathcal{L} = b.\nabla + < a, \nabla^2 >$$

and invariant measure μ.
X is drawn from π, ξ is a jump from X.

$$X_t = 1_{t \geq T} \xi + X.$$

h is the time step of the Markov Chain. If

- Conditions on $\mathcal{L}$.
- $\mathbb{E}[(\int \frac{(y-X)}{h} K(x,dy) - b(X))^2]$;
- $\mathbb{E}[(\int \frac{(y-X)^2}{2h} K(x,dy) - b(X))^2]$;
- higher moments of the jumps

are small then $W_2(\pi, \mu)$ is small.

Convergence of Markov chains towards diffusion processes

Theorem [Strook Varadhan]

Consider a family of discrete Markov chains M^h defined on $S_h \subset \mathbb{R}^d$ with transition kernel K_h. Then, if

- $\lim_{h\to 0} \sup_{x\in S_h} \frac{1}{h}\int_{y\in S_h} (y-x)K_h(x,dy) = b$,
- $\lim_{h\to 0} \sup_{x\in S_h} \frac{1}{h}\int_{y\in S_h} \frac{(y-x)^2}{2} K_h(x,dy) = a$
- $\forall r > 0, \lim_{h\to 0} \sup_{x\in S_h} \int_{y\in S_h, |y-x|>r} K_h(x,y) = 0$ Then, for any $T > 0$, M^h converges weakly on $[0,T]$ toward the diffusion process with infinitesimal generator $\mathcal{L}$.

Langevin Monte-Carlo algorithm

How to draw points from a smooth log concave measure μ with density f?

Langevin Monte-Carlo algorithm

How to draw points from a smooth log concave measure μ with density f?

μ is the stationary distribution of the diffusion process with generator

$$\mathcal{L}f = \nabla \log f . \nabla + \Delta f,$$

thus we can draw from f by simulating the diffusion operator.

Langevin Monte-Carlo algorithm

How to draw points from a smooth log concave measure μ with density f?

μ is the stationary distribution of the diffusion process with generator

$$\mathcal{L}f = \nabla \log f . \nabla + \Delta f,$$

thus we can draw from f by simulating the diffusion operator.

Sample from a discretization of the form

$$X_{t+h} = X_t + hb(X_t) + \sqrt{h}\xi.$$

Langevin Monte-Carlo algorithm

How to draw points from a smooth log concave measure μ with density f?

μ is the stationary distribution of the diffusion process with generator

$$\mathcal{L}f = \nabla \log f.\nabla + \Delta f,$$

thus we can draw from f by simulating the diffusion operator.

Sample from a discretization of the form

$$X_{t+h} = X_t + hb(X_t) + \sqrt{h}\xi.$$

Complexity to achieve an accuracy of ϵ between the sampling and the target in W_2 is

- $O^*(d\epsilon^{-1})$ if $\xi = \mathcal{N}(0,1)$ (direct approach).
- $O^*(d\epsilon^{-2})$ for more general ξ (Stein's method).

Conclusion

Rates in Wasserstein Distance for the muldim CLT.

Quantitative convergence result for the invariant measure of an approximation scheme of a diffusion process.

Conclusion

Rates in Wasserstein Distance for the muldim CLT.

Quantitative convergence result for the invariant measure of an approximation scheme of a diffusion process.

Density estimation for K-nn graphs.

Rates for Langevin Monte-Carlo algorithm.

Conclusion

Rates in Wasserstein Distance for the muldim CLT.

Quantitative convergence result for the invariant measure of an approximation scheme of a diffusion process.

Density estimation for K-nn graphs.

Too strong assumptions on k as $n \to \infty$, requires new stochastic homogenization results.

Rates for Langevin Monte-Carlo algorithm.