Stability of Stein kernels, moment maps and invariant measures

Dan Mikulincer

Weizmann Institute of Science
Joint work with Max Fathi

What This Talk Is About

Let X_t, Y_t be two diffusions in $\mathbb{R}^d$ which satisfy

$$dX_t = a(X_t)dt + \tau(X_t)dB_t,$$
$$dY_t = b(Y_t)dt + \sigma(Y_t)dB_t.$$

Assume ν, μ to be their respective (unique) invariant measures.

Question (Stability of invariant measures)

Suppose that $\|a - b\|_{\cdot} + \|\tau - \sigma\|_{\cdot}$ is small, is μ close to ν?

What This Talk Is About

Let X_t, Y_t be two diffusions in $\mathbb{R}^d$ which satisfy

$$dX_t = a(X_t)dt + \tau(X_t)dB_t,$$
$$dY_t = b(Y_t)dt + \sigma(Y_t)dB_t.$$

Assume ν, μ to be their respective (unique) invariant measures.

Question (Stability of invariant measures)

Suppose that $\|a - b\|_{\cdot} + \|\tau - \sigma\|_{\cdot}$ is small, is μ close to ν?

The Motivation (What This Talk is Really About)

If μ is a measure on $\mathbb{R}^d$ we will associate to it the following objects:

- A matrix valued map $\tau_\mu : \mathbb{R}^d \to M_d(\mathbb{R})$, called a Stein kernel.
- A convex function $\varphi_\mu : \mathbb{R}^d \to \mathbb{R}$, called the moment map.

Question (Stability of Stein kernels)

Suppose that $\|\tau_\mu - \tau_\nu\|_.$ is small, is μ close to ν?

Question (Stability of moment maps)

Suppose that $\|\varphi_\mu - \varphi_\nu\|_.$ is small, is μ close to ν?

The Motivation (What This Talk is Really About)

If μ is a measure on $\mathbb{R}^d$ we will associate to it the following objects:

- A matrix valued map $\tau_\mu : \mathbb{R}^d \to M_d(\mathbb{R})$, called a Stein kernel.
- A convex function $\varphi_\mu : \mathbb{R}^d \to \mathbb{R}$, called the moment map.

Question (Stability of Stein kernels)

Suppose that $\|\tau_\mu - \tau_\nu\|_\cdot$ is small, is μ close to ν?

Question (Stability of moment maps)

Suppose that $\|\varphi_\mu - \varphi_\nu\|_\cdot$ is small, is μ close to ν?

The Motivation (What This Talk is Really About)

If μ is a measure on $\mathbb{R}^d$ we will associate to it the following objects:

- A matrix valued map $\tau_\mu : \mathbb{R}^d \to M_d(\mathbb{R})$, called a Stein kernel.
- A convex function $\varphi_\mu : \mathbb{R}^d \to \mathbb{R}$, called the moment map.

Question (Stability of Stein kernels)

Suppose that $\|\tau_\mu - \tau_\nu\|_\cdot$ is small, is μ close to ν?

Question (Stability of moment maps)

Suppose that $\|\varphi_\mu - \varphi_\nu\|_\cdot$ is small, is μ close to ν?

The Motivation (What This Talk is Really About)

If μ is a measure on $\mathbb{R}^d$ we will associate to it the following objects:

- A matrix valued map $\tau_\mu : \mathbb{R}^d \to M_d(\mathbb{R})$, called a Stein kernel.
- A convex function $\varphi_\mu : \mathbb{R}^d \to \mathbb{R}$, called the moment map.

Question (Stability of Stein kernels)

Suppose that $\|\tau_\mu - \tau_\nu\|_.$ is small, is μ close to ν?

Question (Stability of moment maps)

Suppose that $\|\varphi_\mu - \varphi_\nu\|_.$ is small, is μ close to ν?

The Motivation (What This Talk is Really About)

If μ is a measure on $\mathbb{R}^d$ we will associate to it the following objects:

- A matrix valued map $\tau_\mu : \mathbb{R}^d \to M_d(\mathbb{R})$, called a Stein kernel.
- A convex function $\varphi_\mu : \mathbb{R}^d \to \mathbb{R}$, called the moment map.

Question (Stability of Stein kernels)

Suppose that $\|\tau_\mu - \tau_\nu\|_.$ is small, is μ close to ν?

Question (Stability of moment maps)

Suppose that $\|\varphi_\mu - \varphi_\nu\|_.$ is small, is μ close to ν?

Stein's method

Basic observation: If $G \sim \gamma$ is the standard Gaussian on $\mathbb{R}^d$. Then,

$$\mathbb{E}\left[\langle G, \nabla f(G)\rangle\right] = \mathbb{E}\left[\Delta f(G)\right],$$

for any test function $f : \mathbb{R}^d \to \mathbb{R}$. Moreover, the Gaussian is the only measure which satisfies this relation.

Stein's idea: This property is stable. If X is any other random vector in $\mathbb{R}^d$.

$$\mathbb{E}\left[\langle X, \nabla f(X)\rangle\right] \simeq \mathbb{E}\left[\Delta f(G)\right] \implies X \simeq G,$$

Stein's method

Basic observation: If $G \sim \gamma$ is the standard Gaussian on $\mathbb{R}^d$. Then,

$$\mathbb{E}\left[\langle G, \nabla f(G)\rangle\right] = \mathbb{E}\left[\Delta f(G)\right],$$

for any test function $f : \mathbb{R}^d \to \mathbb{R}$. Moreover, the Gaussian is the only measure which satisfies this relation.

Stein's idea: This property is stable. If X is any other random vector in $\mathbb{R}^d$.

$$\mathbb{E}\left[\langle X, \nabla f(X)\rangle\right] \simeq \mathbb{E}\left[\Delta f(G)\right] \implies X \simeq G,$$

Stein Kernels

A Stein kernel of $X \sim \mu$ is a matrix valued map $\tau : \mathbb{R}^d \to M_d(\mathbb{R})$, such that

$$\mathbb{E}\left[\langle X, \nabla f(X)\rangle\right] \simeq \mathbb{E}\left[\langle \tau, \nabla^2 f(G)\rangle\right].$$

We have that $\tau \equiv \mathrm{Id}$ iff $\mu = \gamma$. The discrepancy is then defined as

$$S^2(\mu||\gamma) = \mathbb{E}_\mu\left[\|\tau - \mathrm{Id}\|_{HS}^2\right].$$

Stein Kernels

A Stein kernel of $X \sim \mu$ is a matrix valued map $\tau : \mathbb{R}^d \to M_d(\mathbb{R})$, such that

$$\mathbb{E}\left[\langle X, \nabla f(X)\rangle\right] \simeq \mathbb{E}\left[\langle \tau, \nabla^2 f(G)\rangle\right].$$

We have that $\tau \equiv \mathrm{Id}$ iff $\mu = \gamma$. The discrepancy is then defined as

$$S^2(\mu||\gamma) = \mathbb{E}_\mu\left[\|\tau - \mathrm{Id}\|_{HS}^2\right].$$

Stein Kernels - Example

If $X \sim \mu$ is a 'nice' centered random variable on $\mathbb{R}$, with density ρ its unique Stein kernel is given by

$$\tau(x) := \frac{\int_x^\infty y\rho(y)dy}{\rho(x)}.$$

Indeed, we can integrate by parts,

$$\mathbb{E}\left[Xf'(X)\right] = \int_{-\infty}^{\infty} f'(x)\rho(x)dx = \int_{-\infty}^{\infty} f''(x)\left(\int_x^\infty y\rho(y)dy\right)dx$$

$$= \int_{-\infty}^{\infty} f''(x)\frac{\left(\int_x^\infty y\rho(y)dy\right)}{\rho(x)}\rho(x)dx = \mathbb{E}\left[\tau(X)f''(X)\right].$$

Stein Kernels - Example

If $X \sim \mu$ is a 'nice' centered random variable on $\mathbb{R}$, with density ρ its unique Stein kernel is given by

$$\tau(x) := \frac{\int\limits_x^\infty y\rho(y)dy}{\rho(x)}.$$

Indeed, we can integrate by parts,

$$\mathbb{E}\left[Xf'(X)\right] = \int\limits_{-\infty}^{\infty} f'(x)\rho(x)dx = \int\limits_{-\infty}^{\infty} f''(x)\left(\int\limits_x^\infty y\rho(y)dy\right)dx$$

$$= \int\limits_{-\infty}^{\infty} f''(x)\frac{\left(\int\limits_x^\infty y\rho(y)dy\right)}{\rho(x)}\rho(x)dx = \mathbb{E}\left[\tau(X)f''(X)\right].$$

Stein Kernels

Suppose now that $|\tau(x) - 1|$ is small. So, $\rho(x) \simeq \int_x^\infty y\rho(y)dy$. In this case, one can use Gronwall's inequality to show $\rho(x) \simeq e^{-x^2/2}$.

yo

yo

In higher dimension, many different constructions for Stein kernels are known. The known constructions do not have explicit tractable expressions in general.

Stein Kernels

Suppose now that $|\tau(x) - 1|$ is small. So, $\rho(x) \simeq \int_x^\infty y\rho(y)dy$. In this case, one can use Gronwall's inequality to show $\rho(x) \simeq e^{-x^2/2}$.

yo

yo

In higher dimension, many different constructions for Stein kernels are known. The known constructions do not have explicit tractable expressions in general.

Stein Discrepancy

Recall that $S^2(\mu||\gamma) = \mathbb{E}_\mu\left[\|\tau - \mathrm{Id}\|_{HS}^2\right]$. It's an exercise to show,

$$W_1(\mu,\gamma) \leq S(\mu||\gamma).$$

What is more impressive is that

$$W_2(\mu,\gamma) \leq S(\mu||\gamma)$$

as well, as shown in (Ledoux, Nourdin, Pecatti 14').

In fact,

$$\mathrm{Ent}(\mu,\gamma)\frac{1}{2} \leq S^2(\mu||\gamma) \ln\left(1 + \frac{\mathrm{I}(\mu||\gamma)}{S^2(\mu||\gamma)}\right).$$

Stein Discrepancy

Recall that $S^2(\mu||\gamma) = \mathbb{E}_\mu\left[\|\tau - \mathrm{Id}\|_{HS}^2\right]$. It's an exercise to show,

$$W_1(\mu,\gamma) \leq S(\mu||\gamma).$$

What is more impressive is that

$$W_2(\mu,\gamma) \leq S(\mu||\gamma)$$

as well, as shown in (Ledoux, Nourdin, Pecatti 14').

In fact,

$$\mathrm{Ent}(\mu,\gamma)\frac{1}{2} \leq S^2(\mu||\gamma)\ln\left(1 + \frac{\mathrm{I}(\mu||\gamma)}{S^2(\mu||\gamma)}\right).$$

Stein Discrepancy

Recall that $S^2(\mu||\gamma) = \mathbb{E}_\mu\left[\|\tau - \mathrm{Id}\|_{HS}^2\right]$. It's an exercise to show,

$$W_1(\mu, \gamma) \leq S(\mu||\gamma).$$

What is more impressive is that

$$W_2(\mu, \gamma) \leq S(\mu||\gamma)$$

as well, as shown in (Ledoux, Nourdin, Pecatti 14').
In fact,

$$\mathrm{Ent}(\mu, \gamma)\frac{1}{2} \leq S^2(\mu||\gamma) \ln\left(1 + \frac{\mathrm{I}\left(\mu||\gamma\right)}{S^2(\mu||\gamma)}\right).$$

Stein Discrepancy - Rough Sketch

Consider the OU process $dX_t = -X_t dt + \sqrt{2} dB_t$, with $X_0 \sim \mu$. γ is the unique invariant measure of the process and we wish to bound:

$$W_2(X_0, X_\infty) = \int_0^\infty \frac{d}{dt} W_2(X_0, X_t) dt.$$

A result of Otto-Villani allows to bound $\frac{d}{dt} W_2(X_0, X_t)$ by $\mathrm{I}(X_t || \gamma)$.

Integration by parts is then used to bound $\mathrm{I}(X_t || \gamma)$ by $S^2(\mu || \gamma)$.

Stein Discrepancy - Rough Sketch

Consider the OU process $dX_t = -X_t dt + \sqrt{2} dB_t$, with $X_0 \sim \mu$.
γ is the unique invariant measure of the process and we wish to bound:

$$W_2(X_0, X_\infty) = \int_0^\infty \frac{d}{dt} W_2(X_0, X_t) dt.$$

A result of Otto-Villani allows to bound $\frac{d}{dt} W_2(X_0, X_t)$ by $\mathrm{I}(X_t || \gamma)$.

Integration by parts is then used to bound $\mathrm{I}(X_t || \gamma)$ by $S^2(\mu || \gamma)$.

Stein Discrepancy - Rough Sketch

Consider the OU process $dX_t = -X_t dt + \sqrt{2} dB_t$, with $X_0 \sim \mu$.
γ is the unique invariant measure of the process and we wish to bound:

$$W_2(X_0, X_\infty) = \int_0^\infty \frac{d}{dt} W_2(X_0, X_t) dt.$$

A result of Otto-Villani allows to bound $\frac{d}{dt} W_2(X_0, X_t)$ by $\mathrm{I}(X_t || \gamma)$.

yo

Integration by parts is then used to bound $\mathrm{I}(X_t || \gamma)$ by $S^2(\mu || \gamma)$.

Stein Discrepancy with Respect to Other Measures

Stein kernels and discrepancy have found numerous applications for normal approximations:

- Central limit theorems
- Stability of functional inequalities
- Second order Poincaré inequalities

Can we extend the theory by bounding $\mathrm{dist}(\mu, \nu)$ with $\|\tau_\mu - \tau_\nu\|$?

Stein Discrepancy with Respect to Other Measures

Stein kernels and discrepancy have found numerous applications for normal approximations:

- Central limit theorems
- Stability of functional inequalities
- Second order Poincaré inequalities

Can we extend the theory by bounding $\mathrm{dist}(\mu, \nu)$ with $\|\tau_\mu - \tau_\nu\|$?

Moment Maps

For a measure $\mu = e^{-\psi(x)}dx$ on $\mathbb{R}^d$ we define its moment map by:

Definition (Moment map)

A moment map of μ, is a convex function $\varphi_\mu : \mathbb{R}^d \to \mathbb{R}$ such that $e^{-\varphi_\mu}$ is a centered probability density whose push-forward by $\nabla\varphi_\mu$ is μ. The measure $e^{-\varphi_\mu}dx$ is called the moment measure.

Remark: convexity of φ_μ implies that $\nabla\varphi_\mu$ is the optimal transport map between $e^{-\varphi_\mu}dx$ and μ and in particular it satisfies the following Monge–Ampère equation:

$$e^{-\varphi_\mu(x)} = e^{-\psi(\nabla\varphi_\mu(x))}\det(\nabla^2\varphi_\mu(x)).$$

Moment Maps - Examples

Some examples:

- If γ is the standard Gaussian, then $\varphi_\gamma(x) = \frac{\|x\|^2}{2}$.
- For $\mu \sim \mathrm{Uniform}(\mathbb{S}^{d-1})$, $\varphi_\mu(x) = \|x\|$.
- For $\mu \sim \mathrm{Uniform}([-1,1]^d)$, $\varphi_\mu(x) = \sum_{i=1}^{d} 2 \log\cosh\left(\frac{x_i}{2}\right) + C$.

The last example can be seen as special case of the following relation, which can be derived in the one-dimensional case:

$$(\psi^{-1})' \left(-\log \left| \int_x^1 t d\mu(t) \right| \right) = \frac{1}{x}.$$

Moment Maps - Examples

Some examples:

- If γ is the standard Gaussian, then $\varphi_\gamma(x) = \frac{\|x\|^2}{2}$.
- For $\mu \sim \mathrm{Uniform}(\mathbb{S}^{d-1})$, $\varphi_\mu(x) = \|x\|$.
- For $\mu \sim \mathrm{Uniform}([-1,1]^d)$, $\varphi_\mu(x) = \sum_{i=1}^{d} 2 \log \cosh\left(\frac{x_i}{2}\right) + C$.

The last example can be seen as special case of the following relation, which can be derived in the one-dimensional case:

$$(\psi^{-1})'\left(-\log\left|\int_x^1 t d\mu(t)\right|\right) = \frac{1}{x}.$$

Moment Maps - Examples

Some examples:

- If γ is the standard Gaussian, then $\varphi_\gamma(x) = \frac{\|x\|^2}{2}$.
- For $\mu \sim \mathrm{Uniform}(\mathbb{S}^{d-1})$, $\varphi_\mu(x) = \|x\|$.
- For $\mu \sim \mathrm{Uniform}([-1,1]^d)$, $\varphi_\mu(x) = \sum_{i=1}^{d} 2 \log \cosh \left(\frac{x_i}{2}\right) + C$.

The last example can be seen as special case of the following relation, which can be derived in the one-dimensional case:

$$(\psi^{-1})' \left(-\log \left| \int_x^1 t d\mu(t) \right| \right) = \frac{1}{x}.$$

Moment Maps - Examples

Some examples:

- If γ is the standard Gaussian, then $\varphi_\gamma(x) = \frac{\|x\|^2}{2}$.
- For $\mu \sim \mathrm{Uniform}(\mathbb{S}^{d-1})$, $\varphi_\mu(x) = \|x\|$.
- For $\mu \sim \mathrm{Uniform}([-1,1]^d)$, $\varphi_\mu(x) = \sum_{i=1}^{d} 2 \log\cosh\left(\frac{x_i}{2}\right) + C$.

The last example can be seen as special case of the following relation, which can be derived in the one-dimensional case:

$$(\psi^{-1})'\left(-\log\left|\int_x^1 t d\mu(t)\right|\right) = \frac{1}{x}.$$

Moment Maps - Examples

Some examples:

- If γ is the standard Gaussian, then $\varphi_\gamma(x) = \frac{\|x\|^2}{2}$.
- For $\mu \sim \mathrm{Uniform}(\mathbb{S}^{d-1})$, $\varphi_\mu(x) = \|x\|$.
- For $\mu \sim \mathrm{Uniform}([-1,1]^d)$, $\varphi_\mu(x) = \sum_{i=1}^{d} 2 \log \cosh\left(\frac{x_i}{2}\right) + C$.

The last example can be seen as special case of the following relation, which can be derived in the one-dimensional case:

$$(\psi^{-1})' \left(-\log \left| \int_x^1 t d\mu(t) \right| \right) = \frac{1}{x}.$$

Moment Maps - Existence

In general, it is hard to give explicit expressions for φ_μ.

Theorem (Cordero-Erausquin, Klartag '15)

Under some mild regularity assumptions, if μ is a centered measure on $\mathbb{R}^d$. Then, the moment map exists and is unique.

It is somewhat clear that if $\varphi_\mu(x) \simeq \frac{x^2}{2}$, then $\mu \simeq \gamma$.

As before, if ν and μ are not Gaussians , what can we say when $\|\varphi_\mu - \varphi_\nu\|$ is small?

It turns out that this is very much related to the previous question about Stein kernels.

Moment Maps - Existence

In general, it is hard to give explicit expressions for φ_μ.

Theorem (Cordero-Erausquin, Klartag '15)

Under some mild regularity assumptions, if μ is a centered measure on $\mathbb{R}^d$. Then, the moment map exists and is unique.

It is somewhat clear that if $\varphi_\mu(x) \simeq \frac{x^2}{2}$, then $\mu \simeq \gamma$.

As before, if ν and μ are not Gaussians , what can we say when $\|\varphi_\mu - \varphi_\nu\|$ is small?

It turns out that this is very much related to the previous question about Stein kernels.

Moment Maps - Existence

In general, it is hard to give explicit expressions for φ_μ.

Theorem (Cordero-Erausquin, Klartag '15)

Under some mild regularity assumptions, if μ is a centered measure on $\mathbb{R}^d$. Then, the moment map exists and is unique.

It is somewhat clear that if $\varphi_\mu(x) \simeq \frac{x^2}{2}$, then $\mu \simeq \gamma$.
As before, if ν and μ are not Gaussians , what can we say when $\|\varphi_\mu - \varphi_\nu\|$ is small?
It turns out that this is very much related to the previous question about Stein kernels.

From Moment Maps to Stein Kernels

Theorem (Fathi 18')

Let μ be a measure on $\mathbb{R}^d$ with moment map $\varphi := \varphi_\mu$. Then, the matrix valued map

$$\tau_\mu(x) = \nabla^2\varphi(\nabla\varphi^{-1}(x)),$$

is a Stein kernel for μ.

Proof.

$$\begin{aligned}\int \langle \nabla f(x), x\rangle d\mu(x) &= \int \langle \nabla f(\nabla\varphi(y)), \nabla\varphi(y)\rangle e^{-\varphi(y)} dy \\ &= \int \langle \nabla^2 f(\nabla\varphi(y)), \nabla^2\varphi(y)\rangle_{HS} e^{-\varphi(y)} dy \\ &= \int \langle \nabla^2 f(x), \nabla^2\varphi(\nabla\varphi^{-1}(x))\rangle_{HS} d\mu(x)\end{aligned}$$

□

From Moment Maps to Stein Kernels

Theorem (Fathi 18')

Let μ be a measure on $\mathbb{R}^d$ with moment map $\varphi := \varphi_\mu$. Then, the matrix valued map

$$\tau_\mu(x) = \nabla^2 \varphi(\nabla \varphi^{-1}(x)),$$

is a Stein kernel for μ.

Proof.

$$\begin{aligned}
\int \langle \nabla f(x), x \rangle d\mu(x) &= \int \langle \nabla f(\nabla \varphi(y)), \nabla \varphi(y) \rangle e^{-\varphi(y)} dy \\
&= \int \langle \nabla^2 f(\nabla \varphi(y)), \nabla^2 \varphi(y) \rangle_{HS} e^{-\varphi(y)} dy \\
&= \int \langle \nabla^2 f(x), \nabla^2 \varphi(\nabla \varphi^{-1}(x)) \rangle_{HS} d\mu(x)
\end{aligned}$$

□

Stability of Moment Maps

We can now use the Stein discrepancy to deduce some stability bounds on the moment map.

$$W_2^2(\mu||\gamma) \leq S^2(\mu||\gamma) = \int \|\nabla^2\varphi(\nabla\varphi^{-1}(x)) - \mathrm{Id}\|_{HS}\, d\mu(x)$$
$$= \int \|\nabla^2\varphi(y) - \mathrm{Id}\|_{HS}\, e^{-\varphi(y)} dy.$$

From Stein Kernels to Stochastic Processes

Now, let μ be a measure and τ_μ its (moment) Stein kernel. We define a stochastic process

$$dX_t = -X_t dt + \sqrt{2\tau_\mu(X_t)} dB_t.$$

Remark: compare this to the OU process:

$$dY_t = -Y_t dt + \sqrt{2}\mathrm{Id} dB_t.$$

Lemma

μ is an invariant measure of X_t.

From Stein Kernels to Stochastic Processes

Now, let μ be a measure and τ_μ its (moment) Stein kernel. We define a stochastic process

$$dX_t = -X_t dt + \sqrt{2\tau_\mu(X_t)} dB_t.$$

Remark: compare this to the OU process:

$$dY_t = -Y_t dt + \sqrt{2}\mathrm{Id} dB_t.$$

Lemma

μ is an invariant measure of X_t.

From Stein Kernels to Stochastic Processes

Now, let μ be a measure and τ_μ its (moment) Stein kernel. We define a stochastic process

$$dX_t = -X_t dt + \sqrt{2\tau_\mu(X_t)} dB_t.$$

Remark: compare this to the OU process:

$$dY_t = -Y_t dt + \sqrt{2}\mathrm{Id} dB_t.$$

Lemma

μ is an invariant measure of X_t.

Invariant Measures

Proof.

The infinitesimal generator of X_t is given by:

$$Lf(x) = -\langle x, \nabla f(x) \rangle + \langle \tau_\mu(x), \nabla^2 f(x) \rangle_{HS}.$$

μ is an invariant measure of X_t, if and only if,

$$\mathbb{E}_\mu \left[Lf(x) \right] = 0.$$

Or, in other words,

$$\mathbb{E}_\mu \left[\langle x, \nabla f(x) \rangle \right] = \mathbb{E}_\mu \left[\langle \tau_\mu(x), \nabla^2 f(x) \rangle_{HS} \right],$$

which is the Stein relation. □

Invariant Measures

Proof.

The infinitesimal generator of X_t is given by:

$$Lf(x) = -\langle x, \nabla f(x) \rangle + \langle \tau_\mu(x), \nabla^2 f(x) \rangle_{HS}.$$

μ is an invariant measure of X_t, if and only if,

$$\mathbb{E}_\mu \left[Lf(x) \right] = 0.$$

Or, in other words,

$$\mathbb{E}_\mu \left[\langle x, \nabla f(x) \rangle \right] = \mathbb{E}_\mu \left[\langle \tau_\mu(x), \nabla^2 f(x) \rangle_{HS} \right],$$

which is the Stein relation. □

Invariant Measures

Proof.

The infinitesimal generator of X_t is given by:

$$Lf(x) = -\langle x, \nabla f(x) \rangle + \langle \tau_\mu(x), \nabla^2 f(x) \rangle_{HS}.$$

μ is an invariant measure of X_t, if and only if,

$$\mathbb{E}_\mu [Lf(x)] = 0.$$

Or, in other words,

$$\mathbb{E}_\mu [\langle x, \nabla f(x) \rangle] = \mathbb{E}_\mu \left[\langle \tau_\mu(x), \nabla^2 f(x) \rangle_{HS} \right],$$

which is the Stein relation. □

Stochastic Process - Properties

This process was studied before, in different settings:

- The Dirichlet form is $\mathbb{E}_{\mu}\left[fLf\right] = \mathbb{E}_{\mu}\left[\nabla f^{T}\tau_{\mu}\nabla f\right]$. Moreover

$$\mathrm{Var}_{\mu}(f) \leq \mathbb{E}_{\mu}\left[\nabla f^{T}\tau_{\mu}\nabla f\right].$$

- It has an exponential convergence to equilibrium. If $X_t \sim \mu_t$,

$$W_{\cdot}(\mu_t, \mu) \leq e^{-\frac{t}{2}} W_{\cdot}(\mu_0, \mu).$$

Those properties make it tempting to use the processes in order to sample from μ. The problem is that τ_{μ} is not tractable, in general.

Stochastic Process - Properties

This process was studied before, in different settings:

- The Dirichlet form is $\mathbb{E}_\mu [fLf] = \mathbb{E}_\mu \left[\nabla f^T \tau_\mu \nabla f\right]$. Moreover

$$\mathrm{Var}_\mu (f) \leq \mathbb{E}_\mu \left[\nabla f^T \tau_\mu \nabla f\right].$$

- It has an exponential convergence to equilibrium. If $X_t \sim \mu_t$,

$$W_\cdot(\mu_t, \mu) \leq e^{-\frac{t}{2}} W_\cdot(\mu_0, \mu).$$

Those properties make it tempting to use the processes in order to sample from μ. The problem is that τ_μ is not tractable, in general.

Stochastic Process - Properties

This process was studied before, in different settings:

- The Dirichlet form is $\mathbb{E}_\mu [fLf] = \mathbb{E}_\mu \left[\nabla f^T \tau_\mu \nabla f\right]$. Moreover

$$\mathrm{Var}_\mu (f) \le \mathbb{E}_\mu \left[\nabla f^T \tau_\mu \nabla f\right].$$

- It has an exponential convergence to equilibrium. If $X_t \sim \mu_t$,

$$W_\cdot(\mu_t, \mu) \le e^{-\frac{t}{2}} W_\cdot(\mu_0, \mu).$$

Those properties make it tempting to use the processes in order to sample from μ. The problem is that τ_μ is not tractable, in general.

Stochastic Process - Properties

This process was studied before, in different settings:

- The Dirichlet form is $\mathbb{E}_\mu\left[fLf\right] = \mathbb{E}_\mu\left[\nabla f^T \tau_\mu \nabla f\right]$. Moreover

$$\mathrm{Var}_\mu(f) \leq \mathbb{E}_\mu\left[\nabla f^T \tau_\mu \nabla f\right].$$

- It has an exponential convergence to equilibrium. If $X_t \sim \mu_t$,

$$W_\cdot(\mu_t, \mu) \leq e^{-\frac{t}{2}} W_\cdot(\mu_0, \mu).$$

Those properties make it tempting to use the processes in order to sample from μ. The problem is that τ_μ is not tractable, in general.

Summary Up to Now

We have a nice measure $\mu = e^{-\psi(x)}dx$ on $\mathbb{R}^d$. To this measure we associate the moment map φ_μ,

$$e^{-\varphi_\mu(x)} = e^{-\psi(\nabla\varphi_\mu(x))}\det(\nabla^2\varphi_\mu(x)).$$

We use the moment map to construct a positive-definite Stein kernel τ_μ:

$$\tau_\mu(x) := \nabla^2\varphi(\nabla\varphi^{-1}(x)).$$

From the kernel we build a stochastic process which has μ as an invariant measure.

$$dX_t = -X_t dt + \sqrt{2\tau_\mu(X_t)}dB_t.$$

Summary Up to Now

We have a nice measure $\mu = e^{-\psi(x)}dx$ on $\mathbb{R}^d$. To this measure we associate the moment map φ_μ,

$$e^{-\varphi_\mu(x)} = e^{-\psi(\nabla\varphi_\mu(x))}\det(\nabla^2\varphi_\mu(x)).$$

We use the moment map to construct a positive-definite Stein kernel τ_μ:

$$\tau_\mu(x) := \nabla^2\varphi(\nabla\varphi^{-1}(x)).$$

From the kernel we build a stochastic process which has μ as an invariant measure.

$$dX_t = -X_t dt + \sqrt{2\tau_\mu(X_t)}dB_t.$$

Summary Up to Now

We have a nice measure $\mu = e^{-\psi(x)}dx$ on $\mathbb{R}^d$. To this measure we associate the moment map φ_μ,

$$e^{-\varphi_\mu(x)} = e^{-\psi(\nabla\varphi_\mu(x))}\det(\nabla^2\varphi_\mu(x)).$$

We use the moment map to construct a positive-definite Stein kernel τ_μ:

$$\tau_\mu(x) := \nabla^2\varphi(\nabla\varphi^{-1}(x)).$$

From the kernel we build a stochastic process which has μ as an invariant measure.

$$dX_t = -X_t dt + \sqrt{2\tau_\mu(X_t)}dB_t.$$

What This Talk Is About

Let X_t, Y_t be two diffusions in $\mathbb{R}^d$ which satisfy

$$dX_t = a(X_t)dt + \tau(X_t)dB_t,$$
$$dY_t = b(Y_t)dt + \sigma(Y_t)dB_t.$$

Assume ν, μ to be their respective (unique) invariant measures.

Question (Stability of invariant measures)

Suppose that $\|a - b\|_{\cdot} + \|\tau - \sigma\|_{\cdot}$ is small, is μ close to ν?

Easy Case

Suppose that,

$$dX_t = a(X_t)dt + dB_t,$$
$$dY_t = b(Y_t)dt + dB_t.$$

Then, the processes are equivalent in the Wiener space, and one can use Girsanov's theorem to write their relative densities.

This allows a bound of the form

$$\mathrm{Ent}\left(X_t || Y_t\right) \leq \int_0^t \mathbb{E}\left[\|a(X_t) - b(X_t)\|^2\right] dt.$$

Another Easy Case - Lipschitz Coefficients

Suppose that $\|a(x) - a(y)\|, \|\tau(x) - \tau(y)\|_{HS} \leq \|x - y\|$.
Fix $X_0 = Y_0 \sim \mu$ and apply Itô's formula to $\|X_t - Y_t\|^2$ and obtain

$$\begin{aligned} \frac{d}{dt}\mathbb{E}\left[\|X_t - Y_t\|^2\right] &= 2\mathbb{E}\left[\langle X_t - Y_t, a(X_t) - b(Y_t)\rangle\right] + \mathbb{E}\left[\|\sigma(X_t) - \tau(Y_t)\|_{HS}^2\right] \\ &\leq 2\mathbb{E}\left[\|X_t - Y_t\|^2\right] + 2\left[\|a(X_t) - b(Y_t)\|^2\right] + \mathbb{E}\left[\|\sigma(X_t) - \tau(Y_t)\|_{HS}^2\right]. \end{aligned}$$

Then,

$$\begin{aligned} \left[\|a(X_t) - b(Y_t)\|^2\right] &\leq 2\mathbb{E}\left[\|a(X_t) - a(Y_t)\|^2\right] + \left[\|a(Y_t) - b(Y_t)\|^2\right] \\ &\leq 2\mathbb{E}\left[\|X_t - Y_t\|^2\right] + \mathbb{E}_\mu\left[\|a - b\|^2\right]. \end{aligned}$$

Another Easy Case - Lipschitz Coefficients

Suppose that $\|a(x) - a(y)\|, \|\tau(x) - \tau(y)\|_{HS} \leq \|x - y\|$.
Fix $X_0 = Y_0 \sim \mu$ and apply Itô's formula to $\|X_t - Y_t\|^2$ and obtain

$$\frac{d}{dt}\mathbb{E}\left[\|X_t - Y_t\|^2\right] = 2\mathbb{E}\left[\langle X_t - Y_t, a(X_t) - b(Y_t)\rangle\right] + \mathbb{E}\left[\|\sigma(X_t) - \tau(Y_t)\|_{HS}^2\right]$$
$$\leq 2\mathbb{E}\left[\|X_t - Y_t\|^2\right] + 2\left[\|a(X_t) - b(Y_t)\|^2\right] + \mathbb{E}\left[\|\sigma(X_t) - \tau(Y_t)\|_{HS}^2\right].$$

Then,

$$\left[\|a(X_t) - b(Y_t)\|^2\right] \leq 2\mathbb{E}\left[\|a(X_t) - a(Y_t)\|^2\right] + \left[\|a(Y_t) - b(Y_t)\|^2\right]$$
$$\leq 2\mathbb{E}\left[\|X_t - Y_t\|^2\right] + \mathbb{E}_\mu\left[\|a - b\|^2\right].$$

Another Easy Case - Lipschitz Coefficients

Suppose that $\|a(x) - a(y)\|, \|\tau(x) - \tau(y)\|_{HS} \leq \|x - y\|$.
Fix $X_0 = Y_0 \sim \mu$ and apply Itô's formula to $\|X_t - Y_t\|^2$ and obtain

$$\begin{aligned} \frac{d}{dt}\mathbb{E}\left[\|X_t - Y_t\|^2\right] &= 2\mathbb{E}\left[\langle X_t - Y_t, a(X_t) - b(Y_t)\rangle\right] + \mathbb{E}\left[\|\sigma(X_t) - \tau(Y_t)\|_{HS}^2\right] \\ &\leq 2\mathbb{E}\left[\|X_t - Y_t\|^2\right] + 2\left[\|a(X_t) - b(Y_t)\|^2\right] + \mathbb{E}\left[\|\sigma(X_t) - \tau(Y_t)\|_{HS}^2\right]. \end{aligned}$$

Then,

$$\begin{aligned} \left[\|a(X_t) - b(Y_t)\|^2\right] &\leq 2\mathbb{E}\left[\|a(X_t) - a(Y_t)\|^2\right] + \left[\|a(Y_t) - b(Y_t)\|^2\right] \\ &\leq 2\mathbb{E}\left[\|X_t - Y_t\|^2\right] + \mathbb{E}_\mu\left[\|a - b\|^2\right]. \end{aligned}$$

Another Easy Case - Lipschitz Coefficients

We conclude:

$$\frac{d}{dt}\mathbb{E}\left[\|X_t - Y_t\|^2\right] \leq 8\mathbb{E}\left[\|X_t - Y_t\|^2\right] + 4\mathbb{E}_\mu\left[\|a - b\|^2\right] + 2\mathbb{E}_\mu\left[\|\tau - \sigma\|^2\right].$$

Gronwall's inequality yields

$$\begin{aligned} W_2^2(\mu, \nu_t) = W_2^2(Y_t, X_t) &\leq \mathbb{E}\left[\|X_t - Y_t\|_2^2\right] \\ &\leq \left(4\mathbb{E}_\mu\left[\|a - b\|^2\right] + 2\mathbb{E}_\mu\left[\|\tau - \sigma\|^2\right]\right)\frac{e^{8t} - 1}{8}. \end{aligned}$$

Another Easy Case - Lipschitz Coefficients

We conclude:

$$\frac{d}{dt}\mathbb{E}\left[\|X_t - Y_t\|^2\right] \leq 8\mathbb{E}\left[\|X_t - Y_t\|^2\right] + 4\mathbb{E}_\mu\left[\|a-b\|^2\right] + 2\mathbb{E}_\mu\left[\|\tau-\sigma\|^2\right].$$

Gronwall's inequality yields

$$\begin{aligned} W_2^2(\mu,\nu_t) = W_2^2(Y_t, X_t) &\leq \mathbb{E}\left[\|X_t - Y_t\|_2^2\right] \\ &\leq \left(4\mathbb{E}_\mu\left[\|a-b\|^2\right] + 2\mathbb{E}_\mu\left[\|\tau-\sigma\|^2\right]\right)\frac{e^{8t}-1}{8}. \end{aligned}$$

Another Easy Case - Lipschitz Coefficients

Assume that X_t converges to equilibrium exponentially fast.

$$W_2(\nu_t, \nu) \leq e^{-t} W_2(\nu_0, \nu).$$

By optimizing over t, we have proven

Theorem

Suppose that a, τ are Lipschitz and that X_t has exponential convergence to equilibrium. Then

$$W_2^2(\mu, \nu) \leq C(\mathbb{E}_\mu \left[\|a - b\|^2\right] + \mathbb{E}_\mu \left[\|\tau - \sigma\|^2\right]).$$

Another Easy Case - Lipschitz Coefficients

Assume that X_t converges to equilibrium exponentially fast.

$$W_2(\nu_t, \nu) \leq e^{-t} W_2(\nu_0, \nu).$$

By optimizing over t, we have proven

Theorem

Suppose that a, τ are Lipschitz and that X_t has exponential convergence to equilibrium. Then

$$W_2^2(\mu, \nu) \leq C(\mathbb{E}_\mu \left[\|a - b\|^2\right] + \mathbb{E}_\mu \left[\|\tau - \sigma\|^2\right]).$$

The General Case

In general, there is no reason to assume that the coefficients will be Lipschitz. In particular, the Stein kernel τ_μ is typically not Globally Lipschitz.

However, in many interesting cases, we can find a proxy for the Lipschitz condition.

Theorem (Ambrosio, Brué, Trevisan - 2017)

If μ is log-concave and $f \in W^{1,p}(\mu)$. Then, there exists a function g, such that

$$\|f(x) - f(y)\| \leq (g(x) + g(y)) \|x - y\|,$$

and

$$\mathbb{E}_\mu \left[\|g\|^p\right] \leq \mathbb{E}_\mu \left[\|\nabla f\|^p\right].$$

The General Case

In general, there is no reason to assume that the coefficients will be Lipschitz. In particular, the Stein kernel τ_μ is typically not Globally Lipschitz.

However, in many interesting cases, we can find a proxy for the Lipschitz condition.

Theorem (Ambrosio, Brué, Trevisan - 2017)

If μ is log-concave and $f \in W^{1,p}(\mu)$. Then, there exists a function g, such that

$$\|f(x) - f(y)\| \leq (g(x) + g(y))\,\|x - y\|,$$

and

$$\mathbb{E}_\mu\left[\|g\|^p\right] \leq \mathbb{E}_\mu\left[\|\nabla f\|^p\right].$$

The General Case

In general, there is no reason to assume that the coefficients will be Lipschitz. In particular, the Stein kernel τ_μ is typically not Globally Lipschitz.

However, in many interesting cases, we can find a proxy for the Lipschitz condition.

Theorem (Ambrosio, Brué, Trevisan - 2017)

If μ is log-concave and $f \in W^{1,p}(\mu)$. Then, there exists a function g, such that

$$\|f(x) - f(y)\| \leq (g(x) + g(y)) \|x - y\|,$$

and

$$\mathbb{E}_\mu [\|g\|^p] \leq \mathbb{E}_\mu [\|\nabla f\|^p].$$

The General Case - Challenges

In the Lipschitz case, we had

$$\mathbb{E}\left[\|a(X_t) - b(Y_t)\|^2\right] \leq \mathbb{E}\left[\|X_t - Y_t\|^2\right].$$

Now, we will get

$$\mathbb{E}\left[\|a(X_t) - b(Y_t)\|^2\right] \leq \mathbb{E}\left[(g(X_t) + g(Y_t)\|X_t - Y_t\|^2\right],$$

which isn't comparable to $\mathbb{E}\left[\|X_t - Y_t\|^2\right]$.

Idea: use another distance which will be more tractable with Itô's formula:

$$D_\delta(X, Y) = \inf \mathbb{E}\left[\ln\left(1 + \frac{\|X - Y\|^2}{\delta^2}\right)\right].$$

The General Case - Challenges

In the Lipschitz case, we had

$$\mathbb{E}\left[\|a(X_t) - b(Y_t)\|^2\right] \leq \mathbb{E}\left[\|X_t - Y_t\|^2\right].$$

Now, we will get

$$\mathbb{E}\left[\|a(X_t) - b(Y_t)\|^2\right] \leq \mathbb{E}\left[(g(X_t) + g(Y_t)\|X_t - Y_t\|^2\right],$$

which isn't comparable to $\mathbb{E}\left[\|X_t - Y_t\|^2\right]$.

yo

Idea: use another distance which will be more tractable with Itô's formula:

$$D_\delta(X, Y) = \inf \mathbb{E}\left[\ln\left(1 + \frac{\|X - Y\|^2}{\delta^2}\right)\right].$$

The General Case - Challenges

In the Lipschitz case, we had

$$\mathbb{E}\left[\|a(X_t) - b(Y_t)\|^2\right] \leq \mathbb{E}\left[\|X_t - Y_t\|^2\right].$$

Now, we will get

$$\mathbb{E}\left[\|a(X_t) - b(Y_t)\|^2\right] \leq \mathbb{E}\left[(g(X_t) + g(Y_t)\|X_t - Y_t\|^2\right],$$

which isn't comparable to $\mathbb{E}\left[\|X_t - Y_t\|^2\right]$.

Idea: use another distance which will be more tractable with Itô's formula:

$$D_\delta(X, Y) = \inf \mathbb{E}\left[\ln\left(1 + \frac{\|X - Y\|^2}{\delta^2}\right)\right].$$

The General Case - Challenges

In the Lipschitz case, we had

$$\mathbb{E}\left[\|a(X_t) - b(Y_t)\|^2\right] \leq \mathbb{E}\left[\|X_t - Y_t\|^2\right].$$

Now, we will get

$$\mathbb{E}\left[\|a(X_t) - b(Y_t)\|^2\right] \leq \mathbb{E}\left[(g(X_t) + g(Y_t)\|X_t - Y_t\|^2\right],$$

which isn't comparable to $\mathbb{E}\left[\|X_t - Y_t\|^2\right]$.

yo

Idea: use another distance which will be more tractable with Itô's formula:

$$D_\delta(X, Y) = \inf \mathbb{E}\left[\ln\left(1 + \frac{\|X - Y\|^2}{\delta^2}\right)\right].$$

The General Case

We now make the following assumptions:

- $\|\tau(x) - \tau(y)\|_{HS}, \|a(x) - a(y)\| \leq (g(x) + g(y))\|x - y\|$.
- $\frac{d\mu}{d\nu}$ is in $L^p(\nu)$ for some p.
- X_t as an exponential convergence to equilibrium.

Theorem

Set $r := \mathbb{E}_\mu [\|a - b\|] + \mathbb{E}_\mu \left[\|\tau - \sigma\|^2\right]$. *With the above assumptions,*

$$W_\cdot^2(\mu, \nu) \lesssim \ln\left(1 + \frac{1}{r}\right)^{-1}.$$

The General Case - Stein Kernels

If ν is a well-conditioned log-concave measure, and φ is its moment map, then we can show $\nabla^2\varphi \in W^{1,2}(e^{-\varphi}dx)$. Which yields

Theorem

Suppose ν is a well-conditioned log-concave measure and let μ be a measure with $\frac{d\mu}{d\nu}$ bounded. Then, τ_ν, τ_μ are their respective (moment) Stein kernels.

$$W_2^2(\mu,\nu) \lesssim \ln\left(1 + \frac{1}{\mathbb{E}_\mu\left[\|\tau_\mu - \tau_\nu\|^2\right]}\right)^{-1}.$$

Thank you!