Coupling, hypoellipticity and gradient estimates

Sayan Banerjee

(Joint works with Maria Gordina, Wilfrid S. Kendall and Phanuel Mariano)

University of North Carolina at Chapel Hill

July 28, 2017

Couplings

- Let $(\Omega_1, \mathcal{F}_1, \mu_1)$ and $(\Omega_2, \mathcal{F}_2, \mu_2)$ be two probability spaces. A coupling of μ_1 and μ_2 is a measure μ on $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2)$ with marginals μ_1 and μ_2.
- We will consider coupling of (the laws of) Markov processes X and Y.
- **Coupling Time:** $\tau = \inf\{s > 0 : X_t = Y_t \text{ for all } t > s\}$.

Coupling and 'closeness' of laws of Markov processes

- **Aldous' Inequality**: For any $t > 0$,

$$||\mu_{1,t} - \mu_{2,t}||_{TV} \leq P(\tau > t),$$

where

  - $\mu_{1,t}$ and $\mu_{2,t}$ are distributions of X_t and Y_t respectively.
  - $||\cdot||_{TV}$ is the total variation distance between measures given by
$$||\mu_{1,t} - \mu_{2,t}||_{TV} = \sup_{A \text{ Borel set}} |\mu_{1,t}(A) - \mu_{2,t}(A)|$$

- **Aldous' Inequality**: For any $t > 0$,

$$||\mu_{1,t} - \mu_{2,t}||_{TV} \leq P(\tau > t),$$

where

  - $\mu_{1,t}$ and $\mu_{2,t}$ are distributions of X_t and Y_t respectively.
  - $||\cdot||_{TV}$ is the total variation distance between measures given by
$$||\mu_{1,t} - \mu_{2,t}||_{TV} = \sup_{A \text{ Borel set}} |\mu_{1,t}(A) - \mu_{2,t}(A)|$$

- Aldous' inequality can be used to estimate how 'close' the laws of X and Y are after time t. If stationary distribution exists, this gives mixing time estimates.

Coupling and 'closeness' of laws of Markov processes

- **Aldous' Inequality:** For any $t > 0$,

$$||\mu_{1,t} - \mu_{2,t}||_{TV} \leq P(\tau > t),$$

  where

  - $\mu_{1,t}$ and $\mu_{2,t}$ are distributions of X_t and Y_t respectively.
  - $||\cdot||_{TV}$ is the total variation distance between measures given by

$$||\mu_{1,t} - \mu_{2,t}||_{TV} = \sup_{A \text{ Borel set}} |\mu_{1,t}(A) - \mu_{2,t}(A)|$$

- Aldous' inequality can be used to estimate how 'close' the laws of X and Y are after time t. If stationary distribution exists, this gives mixing time estimates.

- Maximal coupling: Equality above for all t. (Exists under regularity assumptions, but usually hard to describe.)

Markovian couplings

- A common feature of typically used couplings is that the coupled processes are co-adapted to the same filtration.

Markovian couplings

- A common feature of typically used couplings is that the coupled processes are co-adapted to the same filtration.
- Intuitively, the "next move" of each of the coupled processes depends *only on the past history of both the processes*.

- A coupling of Markov processes X and Y starting from x_0 and y_0 is called **Markovian** if

$$(X_{t+s}, Y_{t+s})_{t\geq 0} \mid \mathcal{F}_s$$

is again a coupling of the laws of X and Y starting from (X_s, Y_s). Here $\mathcal{F}_s = \sigma\{(X_{s'}, Y_{s'}) : s' \leq s\}$.

Example: *Reflection coupling* of simple random walks / Brownian motions.

Markovian couplings (contd.)

- A coupling of Markov processes X and Y starting from x_0 and y_0 is called **Markovian** if

$$(X_{t+s}, Y_{t+s})_{t\geq 0} \mid \mathcal{F}_s$$

is again a coupling of the laws of X and Y starting from (X_s, Y_s). Here $\mathcal{F}_s = \sigma\{(X_{s'}, Y_{s'}) : s' \leq s\}$.

- The coupling is not allowed to look into the future.

Example: *Reflection coupling* of simple random walks / Brownian motions.

- A coupling of Markov processes X and Y starting from x_0 and y_0 is called **Markovian** if

$$(X_{t+s}, Y_{t+s})_{t\geq 0} \mid \mathcal{F}_s$$

is again a coupling of the laws of X and Y starting from (X_s, Y_s). Here $\mathcal{F}_s = \sigma\{(X_{s'}, Y_{s'}) : s' \leq s\}$.

- The coupling is not allowed to look into the future.
- Usually easier to describe and analyze explicitly.

Example: *Reflection coupling* of simple random walks / Brownian motions.

Questions

- Can we get close to the maximal rate using Markovian couplings?

Questions

- Can we get close to the maximal rate using Markovian couplings?
- If so, what class of Markov processes admit Markovian couplings that are "near maximal"?

Questions

- Can we get close to the maximal rate using Markovian couplings?
- If so, what class of Markov processes admit Markovian couplings that are "near maximal"?
- When Markovian couplings fail, can we construct general explicit ways to construct non-Markovian couplings that are "near maximal"?

We will investigate these questions for **diffusion processes**.

Coupling and diffusions

Diffusions are Markov process in $\mathbb{R}^d$ $(d \geq k)$ given by

$$X(t) = x + \int_0^t V_0(X(s))ds + \sum_{i=1}^{k} \int_0^t V_i(X(s)) \circ dW_i(s)$$

where $x \in \mathbb{R}^d$ and $(W_1, \ldots, W_k)$ is a standard Brownian motion on $\mathbb{R}^k$.

- When $k = d$ and $(V_1(x), \dots, V_k(x))$ span $\mathbb{R}^d$ for each $x \in \mathbb{R}^d$, X is called an elliptic diffusion.

  In this case, $\mathbb{R}^d$ furnished with the Riemannian metric $G(x) = (\sigma(x)\sigma(x)^T)^{-1}$ where $\sigma(x) = [V_1(x), \dots, V_d(x)]$ becomes a Riemannian manifold and $X(t)$ becomes a Brownian motion with drift on this new space.

Coupling and diffusions (contd.)

- When $k = d$ and $(V_1(x), \ldots, V_k(x))$ span $\mathbb{R}^d$ for each $x \in \mathbb{R}^d$, X is called an elliptic diffusion.

  In this case, $\mathbb{R}^d$ furnished with the Riemannian metric $G(x) = (\sigma(x)\sigma(x)^T)^{-1}$ where $\sigma(x) = [V_1(x), \ldots, V_d(x)]$ becomes a Riemannian manifold and $X(t)$ becomes a Brownian motion with drift on this new space.

- When $k < d$, the *driving Brownian motion has dimension lower than the diffusion itself*. Nevertheless, it might have a smooth density if $V_0, V_1, \ldots, V_k$ satisfy the Hörmander condition (iterated Lie brackets span the whole tangent space. Then X is called a hypoelliptic diffusion.

Markovian maximal couplings for elliptic diffusions

Markovian maximal couplings are indeed rare. In fact, one can completely characterize the elliptic diffusions which admit such couplings.

Theorem (B. and Kendall, 2014)

If a Markovian maximal coupling exists for two copies of an elliptic diffusion started from sufficiently many pairs of starting points, then the Riemannian manifold obtained via the intrinsic metric must be a sphere, Euclidean space or a hyperbolic space.

Moreover, the drift vector fields are in one-one correspondence with generators of flows of isometries.

Efficient couplings tend to be non-Markovian

- Typically efficient couplings (coupling rates comparable to TV distance) tend to be non-Markovian.

Efficient couplings tend to be non-Markovian

- Typically efficient couplings (coupling rates comparable to TV distance) tend to be non-Markovian.
- In the hypoelliptic setup, there are examples where *Markovian couplings do not even come close to attaining rates comparable to TV distance*.

Efficient couplings tend to be non-Markovian

- Typically efficient couplings (coupling rates comparable to TV distance) tend to be non-Markovian.
- In the hypoelliptic setup, there are examples where *Markovian couplings do not even come close to attaining rates comparable to TV distance*.
- In the existing literature, applicable couplings are usually Markovian and the few examples of non-Markovian couplings are either highly abstract and hence unusable, or highly specialised to particular cases.

Efficient couplings tend to be non-Markovian

- Typically efficient couplings (coupling rates comparable to TV distance) tend to be non-Markovian.
- In the hypoelliptic setup, there are examples where *Markovian couplings do not even come close to attaining rates comparable to TV distance*.
- In the existing literature, applicable couplings are usually Markovian and the few examples of non-Markovian couplings are either highly abstract and hence unusable, or highly specialised to particular cases.
- A schematic approach to constructing explicit non-Markovian couplings is of utmost importance.

Towards a general efficient non-Markovian coupling strategy

We will outline an explicit efficient non-Markovian coupling strategy for the Kolmogorov diffusion and see how the technique extends to the Brownian motion on the Heisenberg group, yielding sharp total variation bounds and also providing important geometric information (gradient estimates).

- Consider the iterated Kolmogorov diffusion of order n given by

$$X_t = \left(B_t, \int_0^t B_s ds, \ldots, \int_0^t \int_0^{s_{n-1}} \cdots \int_0^{s_2} B_{s_1} ds_2 \ldots ds_{n-1}\right)$$

- Consider the iterated Kolmogorov diffusion of order n given by

$$X_t = \left(B_t, \int_0^t B_s ds, \ldots, \int_0^t \int_0^{s_{n-1}} \cdots \int_0^{s_2} B_{s_1} ds_2 \ldots ds_{n-1}\right)$$

- This is a Gaussian process, so we can explicitly compute total variation distances. If the process is started from distinct points in $\mathbb{R}^n$ such that the first k co-ordinates agree, then TV distance at time t is $\sim t^{-k-\frac{1}{2}}$.

Some efficient non-Markovian couplings and applications

- Consider the iterated Kolmogorov diffusion of order n given by

$$X_t = \left(B_t, \int_0^t B_s ds, \ldots, \int_0^t \int_0^{s_{n-1}} \cdots \int_0^{s_2} B_{s_1} ds_2 \ldots ds_{n-1} \right)$$

- This is a Gaussian process, so we can explicitly compute total variation distances. If the process is started from distinct points in $\mathbb{R}^n$ such that the first k co-ordinates agree, then TV distance at time t is $\sim t^{-k-\frac{1}{2}}$.
- Markovian couplings couple at rate at best $t^{-1/2}$ as the 'Brownian motions have to separate before coupling' and the coupling time stochastically dominates the Brownian coupling time. Thus, Markovian couplings can never be efficient.

Outline of the efficient non-Markovian coupling [B.-Kendall, 2015]

- The coupling is based on the Karhunen-Loeve expansion of Brownian motion on $[0, T]$:

$$B(t) = \sqrt{T} \sum_{i=1}^{\infty} Z_k \frac{\sqrt{2} \sin\left(\left(k - \frac{1}{2}\right) \pi t / T\right)}{\left(k - \frac{1}{2}\right) \pi}, \quad Z_k \text{ i.i.d } N(0, 1).$$

Outline of the efficient non-Markovian coupling [B.-Kendall, 2015]

- The coupling is based on the Karhunen-Loeve expansion of Brownian motion on $[0, T]$:

$$B(t) = \sqrt{T}\sum_{i=1}^{\infty} Z_k \frac{\sqrt{2}\sin\left(\left(k-\frac{1}{2}\right)\pi t/T\right)}{\left(k-\frac{1}{2}\right)\pi}, \quad Z_k \text{ i.i.d } N(0,1).$$

- Write $Z_k = W_k(1)$ where W_k are i.i.d. Brownian motions.

Outline of the efficient non-Markovian coupling [B.-Kendall, 2015]

- The coupling is based on the Karhunen-Loeve expansion of Brownian motion on $[0, T]$:

$$B(t) = \sqrt{T} \sum_{i=1}^{\infty} Z_k \frac{\sqrt{2} \sin\left(\left(k - \frac{1}{2}\right)\pi t/T\right)}{\left(k - \frac{1}{2}\right)\pi}, \quad Z_k \text{ i.i.d } N(0,1).$$

- Write $Z_k = W_k(1)$ where W_k are i.i.d. Brownian motions.

- Appropriate Markovian couplings of the infinite dimensional Brownian motions $\{W_k(t) : t \in [0, T]\}_{k \geq 1}$ and $\{\tilde{W}_k(t) : t \in [0, T]\}_{k \geq 1}$ produce non-Markovian couplings of the respective Brownian paths $\{B(t) : t \in [0, T]\}$ and $\{\tilde{B}(t) : t \in [0, T]\}$.

Outline of the efficient non-Markovian coupling [B.-Kendall, 2015]

- The coupling is based on the Karhunen-Loeve expansion of Brownian motion on $[0, T]$:

$$B(t) = \sqrt{T} \sum_{i=1}^{\infty} Z_k \frac{\sqrt{2} \sin \left(\left(k - \frac{1}{2} \right) \pi t / T \right)}{\left(k - \frac{1}{2} \right) \pi}, \quad Z_k \text{ i.i.d } N(0,1).$$

- Write $Z_k = W_k(1)$ where W_k are i.i.d. Brownian motions.
- Appropriate Markovian couplings of the infinite dimensional Brownian motions $\{W_k(t) : t \in [0, T]\}_{k \geq 1}$ and $\{\tilde{W}_k(t) : t \in [0, T]\}_{k \geq 1}$ produce non-Markovian couplings of the respective Brownian paths $\{B(t) : t \in [0, T]\}$ and $\{\tilde{B}(t) : t \in [0, T]\}$.
- Iterating this construction on successive intervals $[2^j, 2^{j+1}]$ yield an efficient non-Markovian coupling.

Brownian Motion on Heisenberg group

- Heisenberg group $\mathbb{H}^3$: $\mathbb{R}^3$ furnished with the group structure

$$(x_1, y_1, z_1) \star (x_2, y_2, z_2) = (x_1+x_2, y_1+y_2, z_1+z_2+(y_2x_1-x_2y_1)).$$

Brownian Motion on Heisenberg group

- Heisenberg group $\mathbb{H}^3$: $\mathbb{R}^3$ furnished with the group structure

  $$(x_1, y_1, z_1) \star (x_2, y_2, z_2) = (x_1+x_2, y_1+y_2, z_1+z_2+(y_2x_1-x_2y_1)).$$

- Canonical example of a sub-Riemannian space with applications in physics, harmonic analysis, geometry and rough paths theory (Neuenschwander, Elderidge, Baudoin, Bakry, Bonnefont, Chafai, Lyons, Hairer, etc.)

Brownian Motion on Heisenberg group

- Heisenberg group $\mathbb{H}^3$: $\mathbb{R}^3$ furnished with the group structure

$$(x_1, y_1, z_1) \star (x_2, y_2, z_2) = (x_1+x_2, y_1+y_2, z_1+z_2+(y_2x_1-x_2y_1)).$$

- Canonical example of a sub-Riemannian space with applications in physics, harmonic analysis, geometry and rough paths theory (Neuenschwander, Elderidge, Baudoin, Bakry, Bonnefont, Chafai, Lyons, Hairer, etc.)

- Brownian motion on the Heisenberg group is two-dimensional standard BM along with Levy stochastic area:

$$X(t) = \left(B_1(t), B_2(t), \int_0^t B_1(s)dB_2(s) - \int_0^t B_2(s)dB_1(s)\right).$$

Total variation bounds

Theorem (B., Gordina and Mariano, 2016)

The total variation distance between the laws $\mathcal{L}(\mathbf{X}_t), \mathcal{L}\left(\widetilde{\mathbf{X}}_t\right)$ *of two Brownian motions on the Heisenberg group started from* (b_1, b_2, a) *and* $\left(\widetilde{b}_1, \widetilde{b}_2, \widetilde{a}\right)$ *respectively satisfy*

$$d_{TV}\left(\mathcal{L}(\mathbf{X}_t), \mathcal{L}\left(\widetilde{\mathbf{X}}_t\right)\right) \leq C_1 \left(\frac{\left|\mathbf{b}-\widetilde{\mathbf{b}}\right|}{\sqrt{t}} + \frac{\left|a-\widetilde{a}+b_1\widetilde{b}_2-b_2\widetilde{b}_1\right|}{t}\right)$$

$$d_{TV}\left(\mathcal{L}(\mathbf{X}_t), \mathcal{L}\left(\widetilde{\mathbf{X}}_t\right)\right) \geq C_2 \left(\frac{\left|\mathbf{b}-\widetilde{\mathbf{b}}\right|}{\sqrt{t}}\mathbb{I}(\mathbf{b}\neq\widetilde{\mathbf{b}}) + \frac{|a-\widetilde{a}|}{t}\mathbb{I}(\mathbf{b}=\widetilde{\mathbf{b}})\right)$$

for $t \geq \max\left\{\left|\mathbf{b}-\widetilde{\mathbf{b}}\right|^2, 2\left|a-\widetilde{a}+b_1\widetilde{b}_2-b_2\widetilde{b}_1\right|\right\}$.

Gradient estimates

The sub-Laplacian on the Heisenberg group is given by

$$\Delta_{\mathcal{H}} = \mathcal{X}^2 + \mathcal{Y}^2$$

where

$$\mathcal{X} = \partial_x - y\partial_z$$
$$\mathcal{Y} = \partial_y + x\partial_z$$

are the left-invariant vector fields.

u is said to be harmonic in a domain D if $\Delta_{\mathcal{H}} u = 0$ on D and u is continuous on $\overline{D}$.

Gradient estimates (contd.)

Theorem (B., Gordina and Mariano, 2016)

Suppose u is non-negative and harmonic on a domain D. There exists a constant $C > 0$ that does not depend on u such that for each $x \in D$,

$$|\nabla_{\mathcal{H}} u(x)| \leqslant C \left(1 + \frac{1}{\delta_x} + \frac{1}{\delta_x^4} + \frac{(1+\delta_x)^3}{\delta_x^4}\right) u(x).$$

where $\delta_x = d_{CC}(x, D^c)$ (Carnot-Caretheodory distance).

Similar theorems were obtained by Cranston ('91, '92) for some elliptic diffusions using Markovian couplings (which fail in our case).

Conclusions and Remarks (Markovian couplings)

- Unlike the elliptic setup, very little is known about existence of Markovian couplings for general hypoelliptic diffusions.

Conclusions and Remarks (Markovian couplings)

- Unlike the elliptic setup, very little is known about existence of Markovian couplings for general hypoelliptic diffusions.
- Challenging problem as we have to successfully couple Brownian motions along with a collection of their path functionals simultaneously.

Conclusions and Remarks (Markovian couplings)

- Unlike the elliptic setup, very little is known about existence of Markovian couplings for general hypoelliptic diffusions.
- Challenging problem as we have to successfully couple Brownian motions along with a collection of their path functionals simultaneously.
- [B.-Kendall, 2017] construct successful Markovian couplings for hypoelliptic diffusions driven by a two-dimensional Brownian motion (W_1, W_2) and polynomial vector fields. Coupling achieved by simultaneously coupling (W_1, W_2) along with the set of Brownian integrals $\{\int W_1^i W_2^j \circ dW_2\}_{i+j \le n}$ using a multi-scale technique.

Conclusions and Remarks (Markovian couplings)

- Unlike the elliptic setup, very little is known about existence of Markovian couplings for general hypoelliptic diffusions.
- Challenging problem as we have to successfully couple Brownian motions along with a collection of their path functionals simultaneously.
- [B.-Kendall, 2017] construct successful Markovian couplings for hypoelliptic diffusions driven by a two-dimensional Brownian motion (W_1, W_2) and polynomial vector fields. Coupling achieved by simultaneously coupling (W_1, W_2) along with the set of Brownian integrals $\{\int W_1^i W_2^j \circ dW_2\}_{i+j \leq n}$ using a multi-scale technique.
- The technique is extendable to nilpotent diffusions.

Conclusions and Remarks (non-Markovian couplings)

- How much of the existing analytic results in sub-Riemannian geometry can we recover via couplings (Poincare inequalities, gradient bounds on heat kernel, etc.)?

  An important ingredient in this direction is the Kuwada duality, which establishes the equivalence of L^p-heat kernel gradient bounds and L^q-Wasserstein distance bounds $(p^{-1} + q^{-1} = 1)$.

Conclusions and Remarks (non-Markovian couplings)

- How much of the existing analytic results in sub-Riemannian geometry can we recover via couplings (Poincare inequalities, gradient bounds on heat kernel, etc.)?

  An important ingredient in this direction is the Kuwada duality, which establishes the equivalence of L^p-heat kernel gradient bounds and L^q-Wasserstein distance bounds $(p^{-1} + q^{-1} = 1)$.

- Total variation distance for Kolmogorov diffusion and BM on Heisenberg group decays faster if the driving Brownian motions start from the same point. Is this phenomenon more general?

Conclusions and Remarks (non-Markovian couplings)

- How much of the existing analytic results in sub-Riemannian geometry can we recover via couplings (Poincare inequalities, gradient bounds on heat kernel, etc.)?

  An important ingredient in this direction is the Kuwada duality, which establishes the equivalence of L^p-heat kernel gradient bounds and L^q-Wasserstein distance bounds $(p^{-1} + q^{-1} = 1)$.

- Total variation distance for Kolmogorov diffusion and BM on Heisenberg group decays faster if the driving Brownian motions start from the same point. Is this phenomenon more general?

- How robust is the developed non-Markovian coupling scheme? Can similar schemes be applied to other hypoelliptic diffusions?

Thank You!