Wasserstein 距離にまつわる確率解析

桒田 和正

(お茶の水女子大学)

確率論シンポジウム (2011 年 12 月 19–22 日 関西大学)

1. What is Wasserstein distance?

Framework

(M, d): Polish space

$\mathcal{P}(M)$: probability measures on M

Wasserstein distance $\cdots$ distance on $\mathcal{P}(M)$

Framework

(M, d): Polish space

$\mathcal{P}(M)$: probability measures on M

Wasserstein distance $\cdots$ distance on $\mathcal{P}(M)$

For $\mu, \nu \in \mathcal{P}(M)$,

$\Pi(\mu, \nu)$: set of couplings between μ & ν i.e.

$$\Pi(\mu, \nu) \subset \mathcal{P}(X \times X),$$

$$\Pi(\mu, \nu) := \left\{ \pi \;\middle|\; \begin{array}{l} \pi(A \times X) = \mu(A), \\ \pi(X \times A) = \nu(A) \end{array} \right\}$$

L^p-Wasserstein distance

For $p \in [1, \infty]$, $\mu, \nu \in \mathcal{P}(X)$

$$W_p(\mu, \nu) := \inf_{\pi \in \Pi(\mu,\nu)} \|d\|_{L^p(\pi)} \in [0, \infty]$$

L^p-Wasserstein space

$\mathcal{P}_p(M) \subset \mathcal{P}(M)$,

$$\mathcal{P}_p(M) := \left\{ \mu \;\middle|\; \int_M d(^\exists x, y)^p \mu(dy) < \infty \right\}$$

- $W_p(\mu, \nu) < \infty$ for $\mu, \nu \in \mathcal{P}_p(M)$

Alternative definition of W_p

For $\mu, \nu \in \mathcal{P}(M)$,

$$W_p(\mu, \nu) = \inf_{(X,Y)} \mathbb{E}[d(X, Y)^p]^{1/p},$$

where the infimum runs over

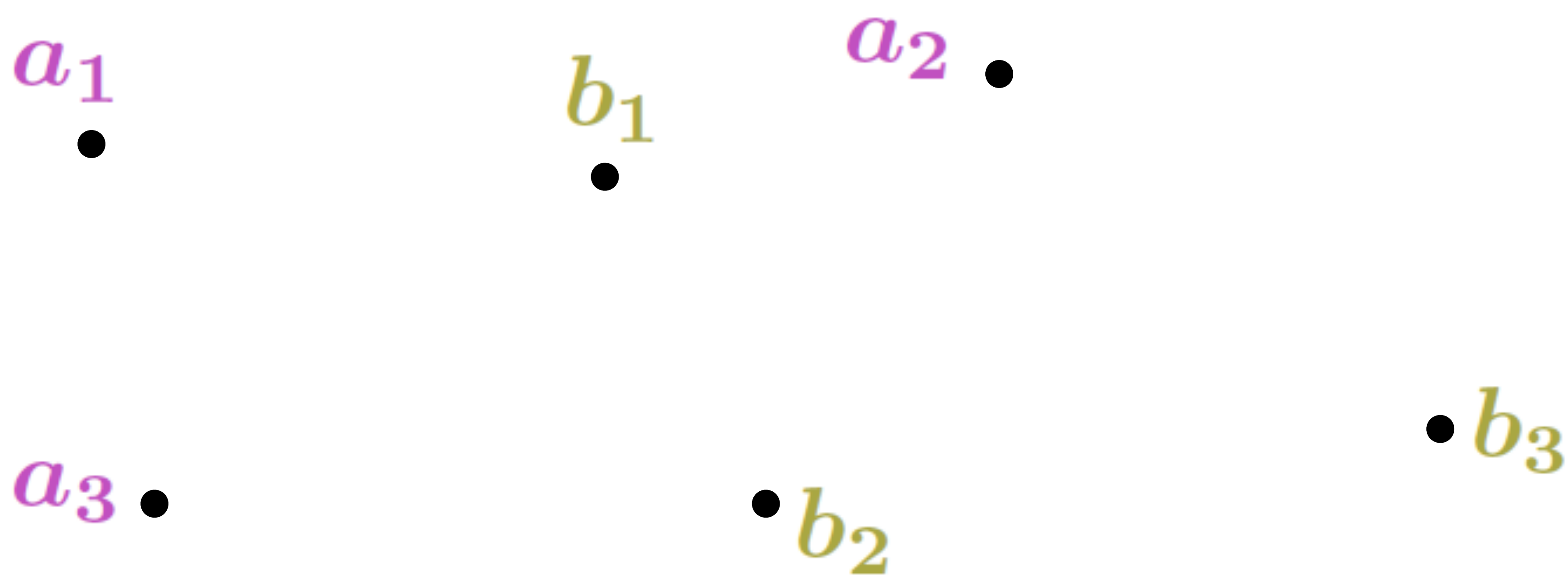
- (X, Y): $M \times M$-valued r.v.s,
- $X \overset{\mathcal{L}}{\sim} \mu$, $Y \overset{\mathcal{L}}{\sim} \nu$

Example

$$\operatorname{supp} \mu = \{a_1, a_2, a_3\}, \operatorname{supp} \nu = \{b_1, b_2, b_3\}$$

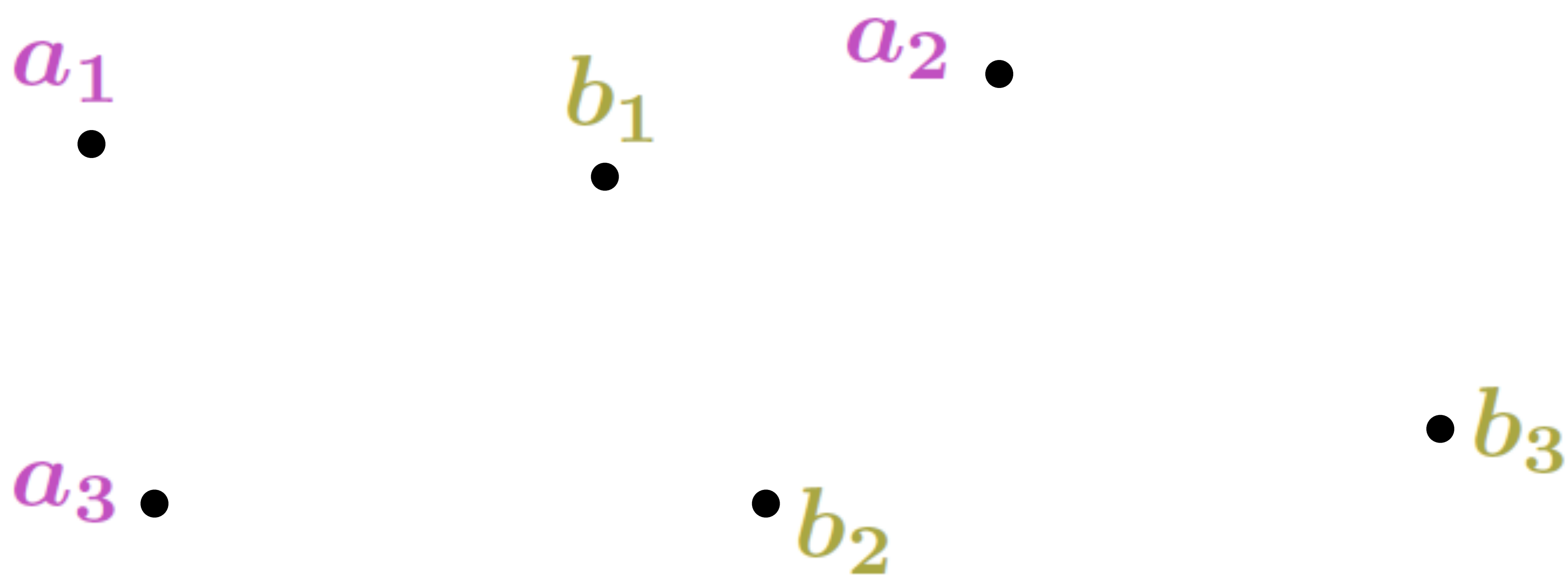
Example

supp $\mu = \{a_1, a_2, a_3\}$, supp $\nu = \{b_1, b_2, b_3\}$



Example

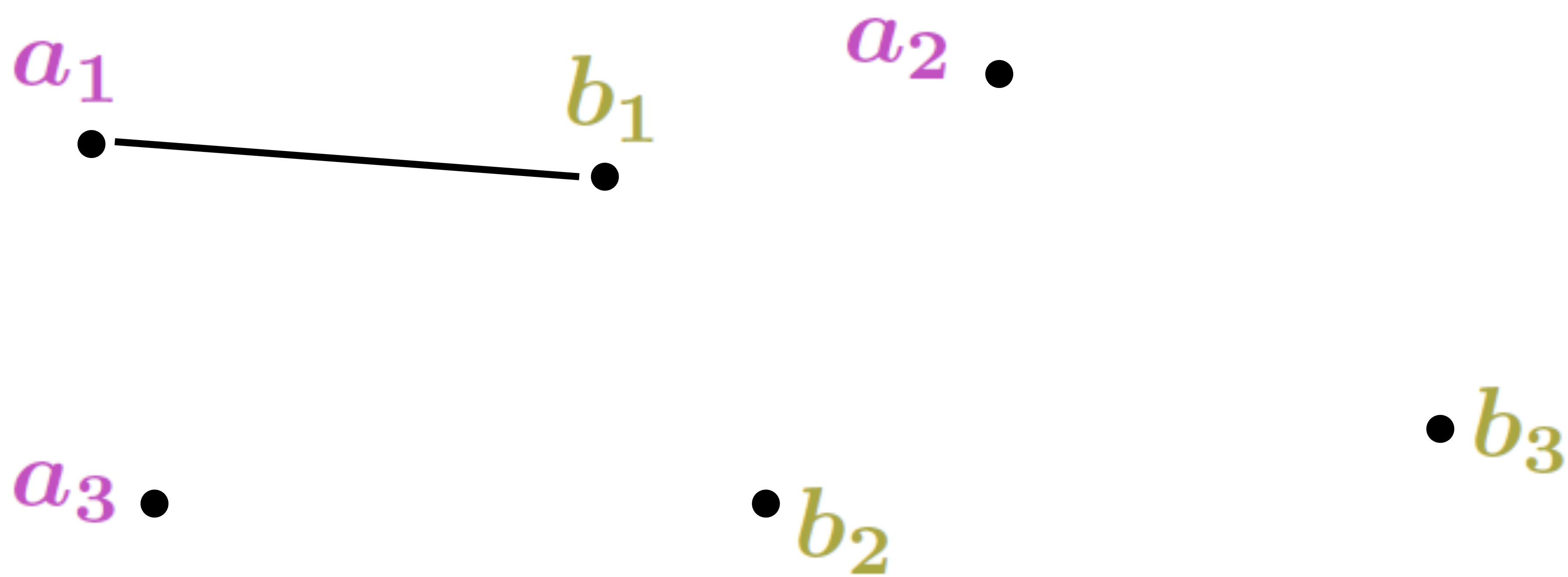
$\mathrm{supp}\ \mu = \{a_1, a_2, a_3\}$, $\mathrm{supp}\ \nu = \{b_1, b_2, b_3\}$



$\pi \in \Pi(\mu, \nu)$: a prob. meas.
supported on $\{(a_i, b_j) \mid i, j \in \{1, 2, 3\}\}$

Example

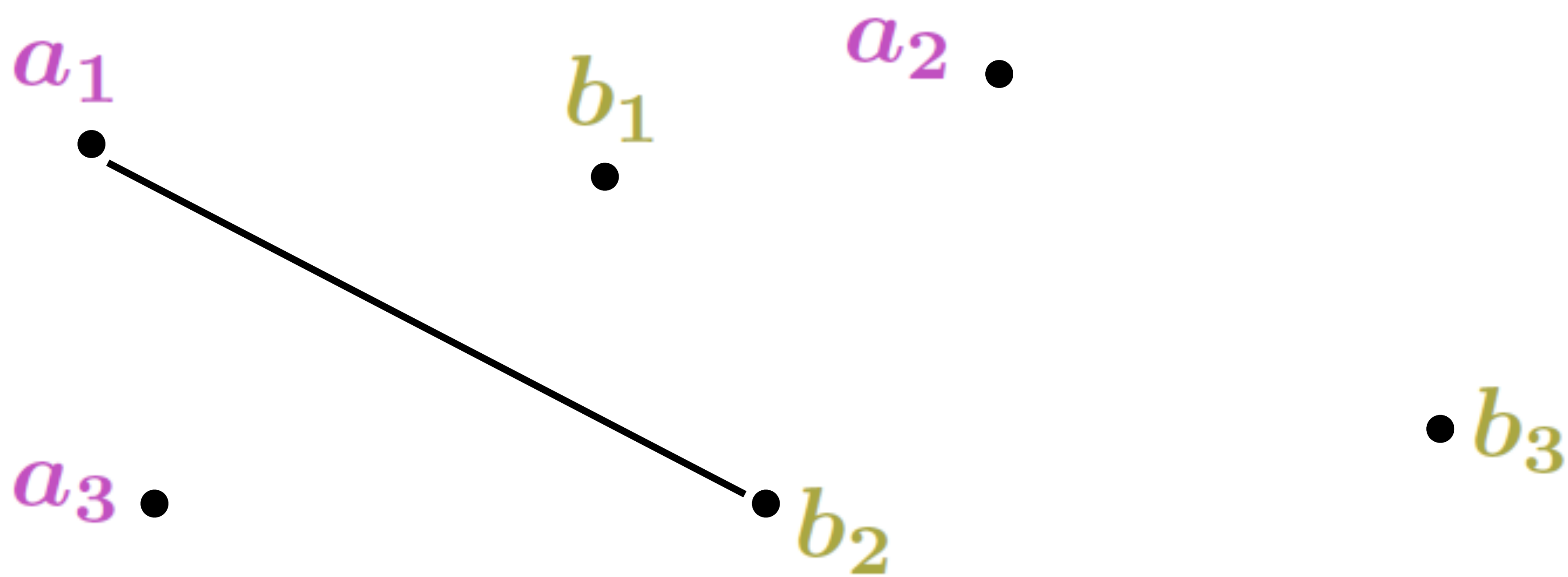
$\operatorname{supp}\mu = \{a_1, a_2, a_3\}$, $\operatorname{supp}\nu = \{b_1, b_2, b_3\}$



$\pi \in \Pi(\mu, \nu)$: a prob. meas.
supported on $\{(a_i, b_j) \mid i, j \in \{1, 2, 3\}\}$

Example

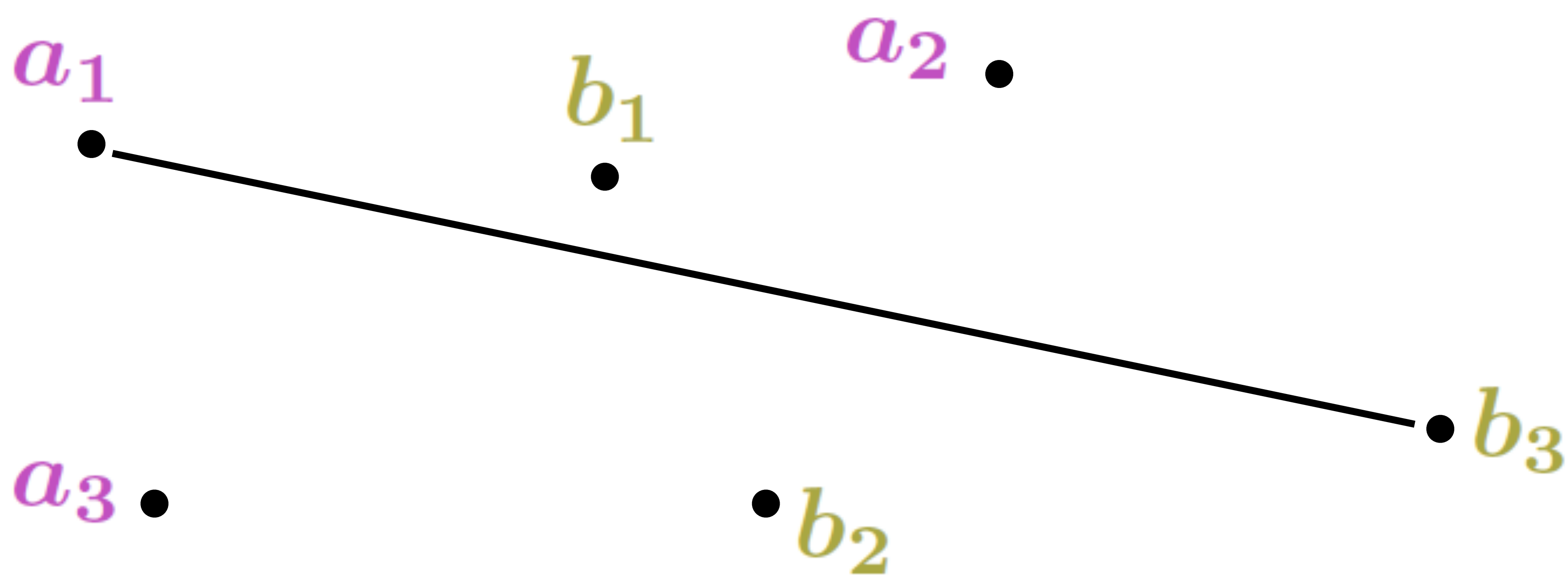
$\mathrm{supp}\ \mu = \{a_1, a_2, a_3\}$, $\mathrm{supp}\ \nu = \{b_1, b_2, b_3\}$



$\pi \in \Pi(\mu, \nu)$: a prob. meas.
supported on $\{(a_i, b_j) \mid i, j \in \{1, 2, 3\}\}$

Example

$\text{supp}\ \mu = \{a_1, a_2, a_3\}$, $\text{supp}\ \nu = \{b_1, b_2, b_3\}$

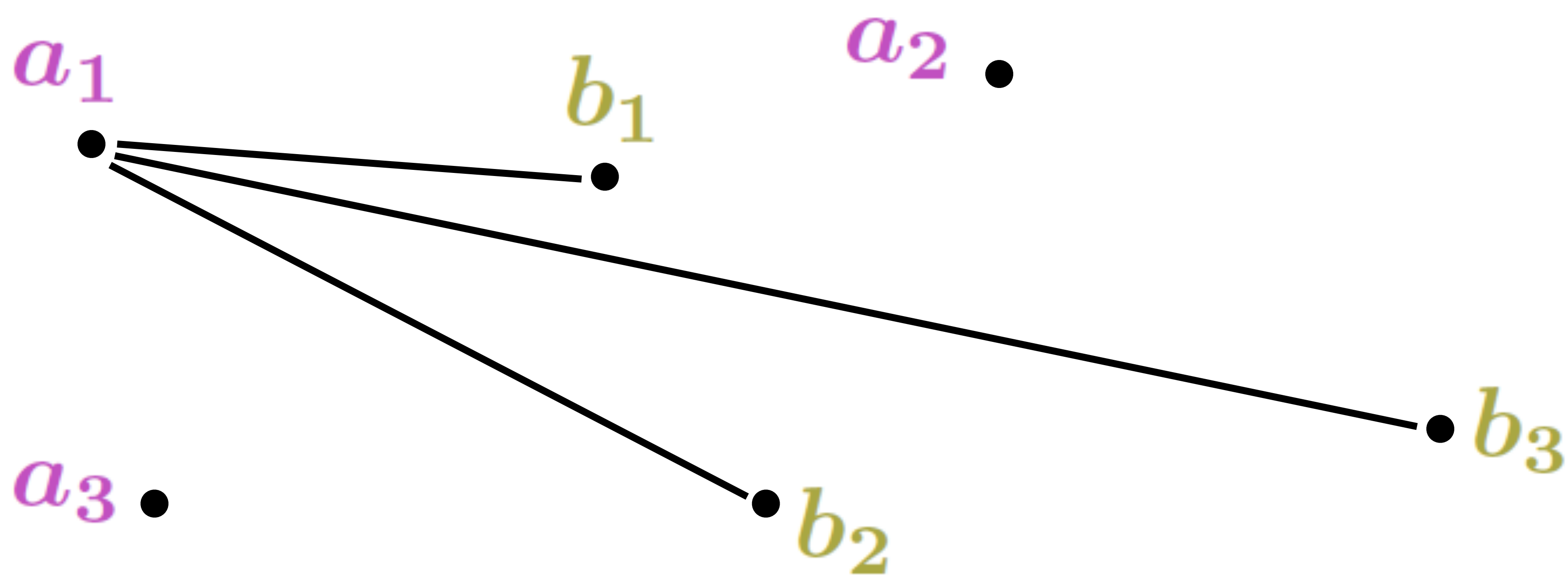


$\pi \in \Pi(\mu, \nu)$: a prob. meas.

supported on $\{(a_i, b_j) \mid i, j \in \{1, 2, 3\}\}$

Example

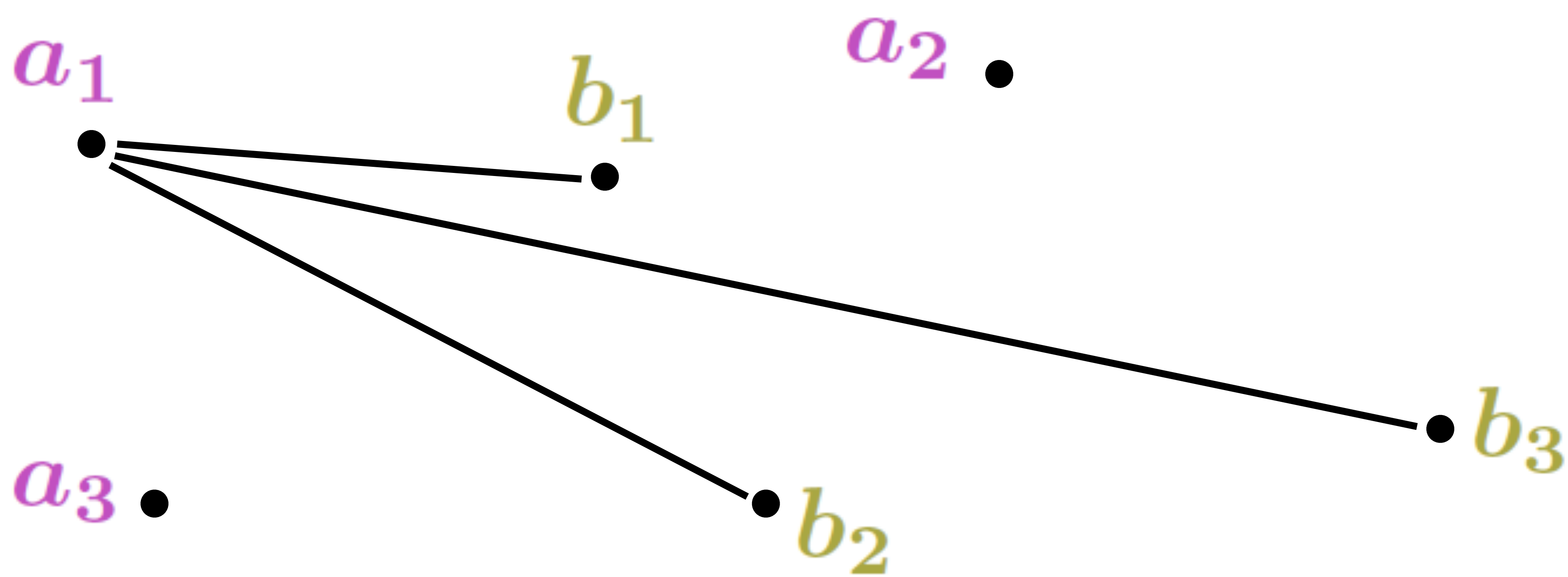
$\operatorname{supp}\mu = \{a_1, a_2, a_3\}$, $\operatorname{supp}\nu = \{b_1, b_2, b_3\}$



$\pi \in \Pi(\mu, \nu)$: a prob. meas.
supported on $\{(a_i, b_j) \mid i, j \in \{1, 2, 3\}\}$

Example

$\operatorname{supp}\mu = \{a_1, a_2, a_3\}$, $\operatorname{supp}\nu = \{b_1, b_2, b_3\}$



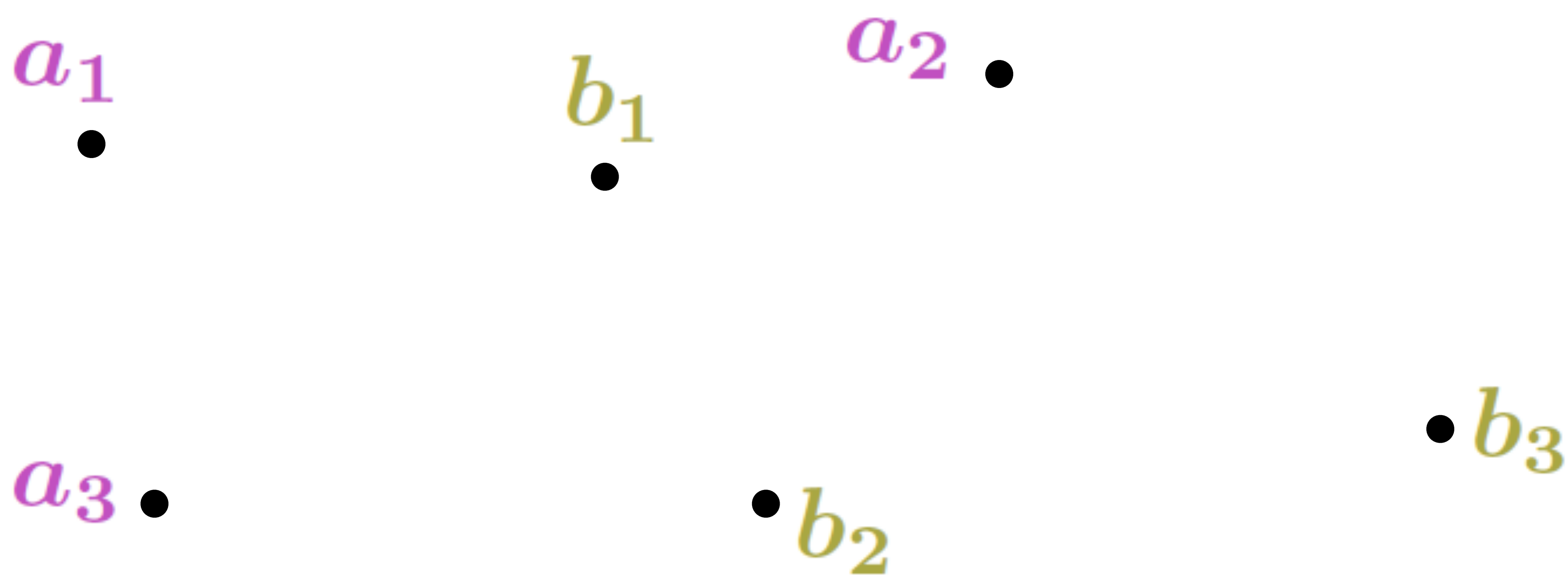
$\pi \in \Pi(\mu, \nu)$: a prob. meas.
supported on $\{(a_i, b_j) \mid i, j \in \{1, 2, 3\}\}$

- μ, ν: unif. $\Rightarrow {}^{\exists}\varphi : \{1, 2, 3\} \to \{1, 2, 3\}$ bij.
  s.t. "unif. meas. on $\{(a_i, b_{\varphi(i)})\}$" is a minimizer

[Birkhoff's thm]

Example

$\mathbf{supp}\ \mu = \{a_1, a_2, a_3\}$, $\mathbf{supp}\ \nu = \{b_1, b_2, b_3\}$



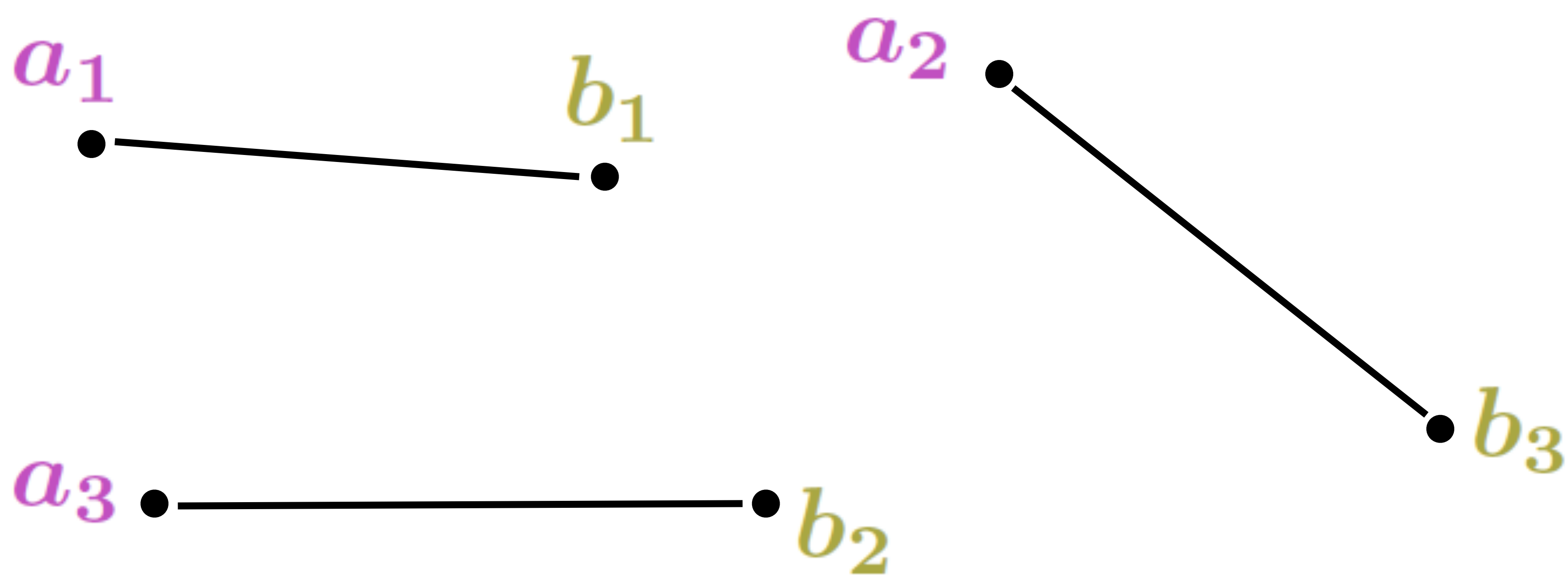
$\pi \in \Pi(\mu, \nu)$: a prob. meas.
supported on $\{(a_i, b_j) \mid i, j \in \{1, 2, 3\}\}$

- μ, ν: unif. $\Rightarrow {}^{\exists}\varphi : \{1, 2, 3\} \to \{1, 2, 3\}$ bij.
  s.t. "unif. meas. on $\{(a_i, b_{\varphi(i)})\}$" is a minimizer

[Birkhoff's thm]

Example

$\mathrm{supp}\,\mu = \{a_1, a_2, a_3\}$, $\mathrm{supp}\,\nu = \{b_1, b_2, b_3\}$



$\pi \in \Pi(\mu, \nu)$: a prob. meas.
supported on $\{(a_i, b_j) \mid i, j \in \{1, 2, 3\}\}$

- μ, ν: unif. $\Rightarrow {}^{\exists}\varphi : \{1, 2, 3\} \to \{1, 2, 3\}$ bij.
  s.t. "unif. meas. on $\{(a_i, b_{\varphi(i)})\}$" is a minimizer

[Birkhoff's thm]

Properties of W_p

- $\lim_n W_p(\mu_n, \mu) = 0$ for $\mu \in \mathcal{P}_p(M)$
  iff $\mu_n \to \mu$ & $\sup_n \int_M d(x,y)^p \mu_n(dy) < \infty$
- W_p: distance on $\mathcal{P}_p(M)$
- Another variational formula (Kantorovich duality)
- $W_p(\mu, \nu) \le \mathbb{E}[d(X,Y)^p]^{1/p}$
  for $^\forall$couplings (X, Y)

More geometric properties

- W_p is stable under perturbation of (M, d)
- W_p reflects the geometry of (M, d) well.
  For instance,
  - (M, d): complete $\Rightarrow$ so is $(\mathcal{P}_p(M), W_p)$
  - (M, d): geodesic sp. $\Rightarrow$ so is $(\mathcal{P}_p(M), W_p)$

Proof of some properties

Lemma 1 ($^\exists$minimizer)

$$W_p(\mu,\nu) = \min_{\pi\in\Pi(\mu,\nu)} \|d\|_{L^p(\pi)}$$

Proof

$(\pi_n)_n$: minimizing sequence

$\star$ $\Pi(\mu,\nu)$: compact. $\therefore$ $\pi_n \to {}^\exists\pi$ w.l.o.g.

$$\begin{aligned}
\|d\|_{L^p(\pi)} &= \lim_{R\to\infty} \|d\wedge R\|_{L^p(\pi)} \\
&= \lim_{R\to\infty} \left(\lim_n \|d\wedge R\|_{L^p(\pi_n)} \right) \\
&\le \lim_n \|d\|_{L^p(\pi_n)}
\end{aligned}$$

$\square$

M^Λ: product of M with index set Λ

$p_{i,j}$: proj. to i-th and j-th components

Lemma 2 (gluing)

$\pi_1 \in \Pi(\mu_1, \mu_2)$, $\pi_2 \in \Pi(\mu_2, \mu_3)$

$\Rightarrow {}^\exists \tilde{\pi} \in \mathcal{P}(M^3)$ s.t.

$$p^\#_{1,2}\tilde{\pi} = \pi_1,\ p^\#_{2,3}\tilde{\pi} = \pi_2$$

Proof

$$(X_1, Y_1) \overset{\mathcal{L}}{\sim} \pi_1,\ (Y_2, Z_2) \overset{\mathcal{L}}{\sim} \pi_2$$

$$\mathbb{P}[X_1 \in dx | Y_1 = y]\mathbb{P}[Z_2 \in dz | Y_2 = y]\mu_2(dy) =: \tilde{\pi}(dxdydz) \quad \square$$

Pf: W_p: dist

- $\pi_1 \in \Pi(\mu_1, \mu_2)$, $\pi_2 \in \Pi(\mu_2, \mu_3)$: minimizer
  $\tilde{\pi}$: gluing of π_1 and π_2

$$\Rightarrow W_p(\mu_1, \mu_3) \le \|d\|_{L^p(p_{1,3}^{\#}\tilde{\pi})}$$
$$\le \|d\|_{L^p(\pi_1)} + \|d\|_{L^p(\pi_2)} \quad //$$

- $W_p(\mu, \nu) = 0$, π: minimizer

$$\Rightarrow \mu(A) = \pi(A \times X)$$
$$= \pi(\{(x, x) \mid x \in A\}) = \nu(A) \quad \square$$

Pf: $\lim_n W_p(\mu_n, \mu) = 0 \Rightarrow \mu_n \to \mu$

$\pi_n \in \Pi(\mu, \mu_n)$: minimizer

$\star\ {}^{\exists}\hat{\pi} \in \mathcal{P}(M^{\mathbb{N}_0})$, $p_{0,n}^{\#}\hat{\pi} = \pi_n$

(via gluing & Kolmogorov's extension thm)

$\Rightarrow p_n \to p_0$ in $L^p(M^{\mathbb{N}_0}, \hat{\pi}; M)$

$$\Rightarrow \int_M f\, d\mu_n = \int_{M^{\mathbb{N}_0}} f \circ p_n\, d\hat{\pi}$$

$$\to \int_{M^{\mathbb{N}_0}} f \circ p_0\, d\hat{\pi}$$

for ${}^{\forall} f \in C_b(M)$ $\square$

Kantorovich duality

&

displacement interpolation

Theorem 1 (Kantorovich duality)

$$W_p(\mu,\nu)^p = \sup_{g,f} \left[\int_M g \, d\mu + \int_M f \, d\nu \right],$$

$$= \sup_{f} \left[\int_M \hat{f} \, d\mu + \int_M f \, d\nu \right],$$

where $f, g \in C_b(M)$,

$$g(x) + f(y) \leq d(x,y)^p,$$

$$\hat{f}(x) := \inf_{y \in M} [d(x,y)^p - f(y)]$$

Constraint: $g(x) + f(y) \leq d(x,y)^p$

$\Rightarrow$ For $\pi \in \Pi(\mu, \nu)$,

$$\int_M g\,d\mu + \int_M f\,d\nu$$
$$= \int_M (g(x) + f(y))\pi(dxdy)$$
$$\leq \|d\|_{L^p(\pi)}^p$$

$$\Rightarrow W_p(\mu,\nu)^p \geq \sup_{g,f}\left[\int_M g\,d\mu + \int_M f\,d\nu\right]$$

If $\pi \in \Pi(\mu, \nu)$: minimizer, (f_0, g_0): maximizer

- $g_0(x) + f_0(y) = d(x, y)^p$ π-a.e. (x, y)

If $\pi \in \Pi(\mu, \nu)$: minimizer, (f_0, g_0): maximizer

- $\underline{g_0(x) + f_0(y) = d(x,y)^p}$ π-a.e. (x, y)
- $g_0 = \hat{f}_0$ & $f_0 = \hat{g}_0$

$$\Rightarrow W_1(\mu, \nu) = \sup_{f\text{: 1-Lip}} \int_M f \, d(\mu - \nu)$$

If $\pi \in \Pi(\mu, \nu)$: minimizer, (f_0, g_0): maximizer

- $\underline{g_0(x) + f_0(y) = d(x, y)^p}$ π-a.e. (x, y)
- $g_0 = \hat{f}_0$ & $f_0 = \hat{g}_0$

$$\Rightarrow W_1(\mu, \nu) = \sup_{f\text{: 1-Lip}} \int_M f \, d(\mu - \nu)$$

- When $M = \mathbb{R}^m$, $d(x, y)^2 = \frac{1}{2}|x - y|^2$,

$$g(x) + f(y) \le d(x, y)^2$$
$$\Leftrightarrow \tilde{g}(x) + \tilde{f}(y) \ge x \cdot y$$
$$\left(\tilde{g}(x) := \frac{1}{2}|x|^2 - g(x)\right)$$

$\Rightarrow \tilde{g}_0 =$ Legendre conj. of $\tilde{f}_0$

Property: $(\mathcal{P}_p(M), W_p)$: geodesic sp.

(M, d): geodesic sp.

iff $^\forall x, y \in M$, $^\exists \gamma : [0,1] \to M$ s.t.

$\gamma(0) = x$, $\gamma(1) = y$,

$d(\gamma(s), \gamma(t)) = |s - t| d(x, y)$

(γ: constant speed minimal geodesic)

Example

- ○ $(\mathbb{R}^m, \|\cdot\|_p)$, $p \in [1, \infty]$
- ○ M: Riemannian mfd with the Riem. distance
- × $A \subset \mathbb{R}^m$: not convex, with $\|\cdot\|_2$

$\Gamma := \{\gamma : [0,1] \to M$ const. speed min. geod.$\}$

$e_t : \Gamma \to M,\ e_t(\gamma) := \gamma(t)$

Theorem 2 (displacement interpolation)

Suppose (M, d): geodesic sp.

For $\mu_0, \mu_1 \in \mathcal{P}_p(M)$, ${}^{\exists}\Xi \in \mathcal{P}(\Gamma)$ s.t.

- $e_0^{\#}\Xi = \mu_0,\ e_1^{\#}\Xi = \mu_1$
- $W_p(e_t^{\#}\Xi, e_s^{\#}\Xi) = \|d\|_{L^p((e_t, e_s)^{\#}\Xi)}$
  $= |t - s| W_p(\mu_0, \mu_1)$

$(\Rightarrow (e_t^{\#}\Xi)_{t \in [0,1]}$: min. geod. in $(\mathcal{P}_p(X), W_p))$

2. Wasserstein contraction and equivalent conditions

Example

$$\mathcal{L} := \Delta - \nabla V \cdot \nabla \text{ on } \mathbb{R}^m$$

$$\updownarrow$$

$$dX_t^x = \sqrt{2}dB_t - \nabla V(X_t^x)dt, \quad X_0^x = x$$

Assumption:

$$(\nabla V(x) - \nabla V(y)) \cdot (x - y) \geq K|x - y|^2$$

$$(\Uparrow \text{Hess}\, V \geq K)$$

$$dX_t^x = \sqrt{2}\,dB_t - \nabla V(X_t^x)\,dt,$$

$$\Downarrow$$

$$\begin{aligned} W_p(\mathbb{P}^{X_t^x}, \mathbb{P}^{X_t^y}) &\leq \mathrm{e}^{-Kt}|x - y| \\ &= \mathrm{e}^{-Kt} W_p(\delta_x, \delta_y) \end{aligned}$$

$$dX_t^x = \sqrt{2}dB_t - \nabla V(X_t^x)dt,$$

X_t^x, X_t^y: str. sol. of the SDE with a common B_t

$$\Downarrow$$

$$\begin{aligned} W_p(\mathbb{P}^{X_t^x}, \mathbb{P}^{X_t^y}) &\le \mathrm{e}^{-Kt}|x - y| \\ &= \mathrm{e}^{-Kt} W_p(\delta_x, \delta_y) \end{aligned}$$

$$dX_t^x = \sqrt{2}dB_t - \nabla V(X_t^x)dt,$$

X_t^x, X_t^y: str. sol. of the SDE with a common B_t

$$\Downarrow$$

$$d(X_t^x - X_t^y) = -\left(\nabla V(X_t^x) - \nabla V(X_t^y)\right) dt$$

$$\Downarrow$$

$$d|X_t^x - X_t^y|^2 \leq -2K|X_t^x - X_t^y|^2 dt$$

$$\Downarrow$$

$$|X_t^x - X_t^y|^2 \leq \mathrm{e}^{-2Kt}|x - y|^2$$

$$\Downarrow$$

$$W_p(\mathbb{P}^{X_t^x}, \mathbb{P}^{X_t^y}) \leq \mathrm{e}^{-Kt}|x - y| = \mathrm{e}^{-Kt} W_p(\delta_x, \delta_y)$$

M: complete Riemannian manifold

$P_t = \mathrm{e}^{t\Delta}$: heat semigroup

M: complete Riemannian manifold

$P_t = \mathrm{e}^{t\Delta}$: heat semigroup

Theorem 3 [von Renesse & Sturm '05]

For $K \in \mathbb{R}$, the following are equivalent:

(a) $W_2(P_t^* \mu_0, P_t^* \mu_1) \leq \mathrm{e}^{-Kt} W_2(\mu_0, \mu_1)$

(b) $\mathrm{Ric} \geq K$

(c) $|\nabla P_t f|(x)^2 \leq \mathrm{e}^{-2Kt} P_t(|\nabla f|^2)(x)$

(d) $\mathrm{CD}(K, \infty)$

Relative entropy: $\mathrm{Ent}(\mu) := \int_M \rho \log \rho \, dv$

(if $d\mu = \rho \, dv$ & $[\rho \log \rho]_- \in L^1$)

<u>The condition CD(K, ∞):</u>

For $^\forall W_2$-geod. $(\mu_t)_{t \in [0,1]}$,

$$\mathrm{Ent}(\mu_t) \leq (1-t)\,\mathrm{Ent}(\mu_0) + t\,\mathrm{Ent}(\mu_1) - \frac{K}{2} t(1-t) W_2(\mu_0, \mu_1)^2$$

($\Leftrightarrow$ "$\mathrm{Hess}\,\mathrm{Ent} \geq K$")

Significance of Theorem 3

- Each condition has rich applications
- (a) & (d) are stable under the measured Gromov-Hausdorff convergence
- Source of several trials to extend the existing theory, once we obtain a variant of them

  $\rightsquigarrow$ the latter part of the talk

3. Coupling by parallel transport

In Theorem 3

(M: cpl. Riem. mfd, P_t: heat semigroup),

(b) $\mathrm{Ric} \geq K$

$\Downarrow$

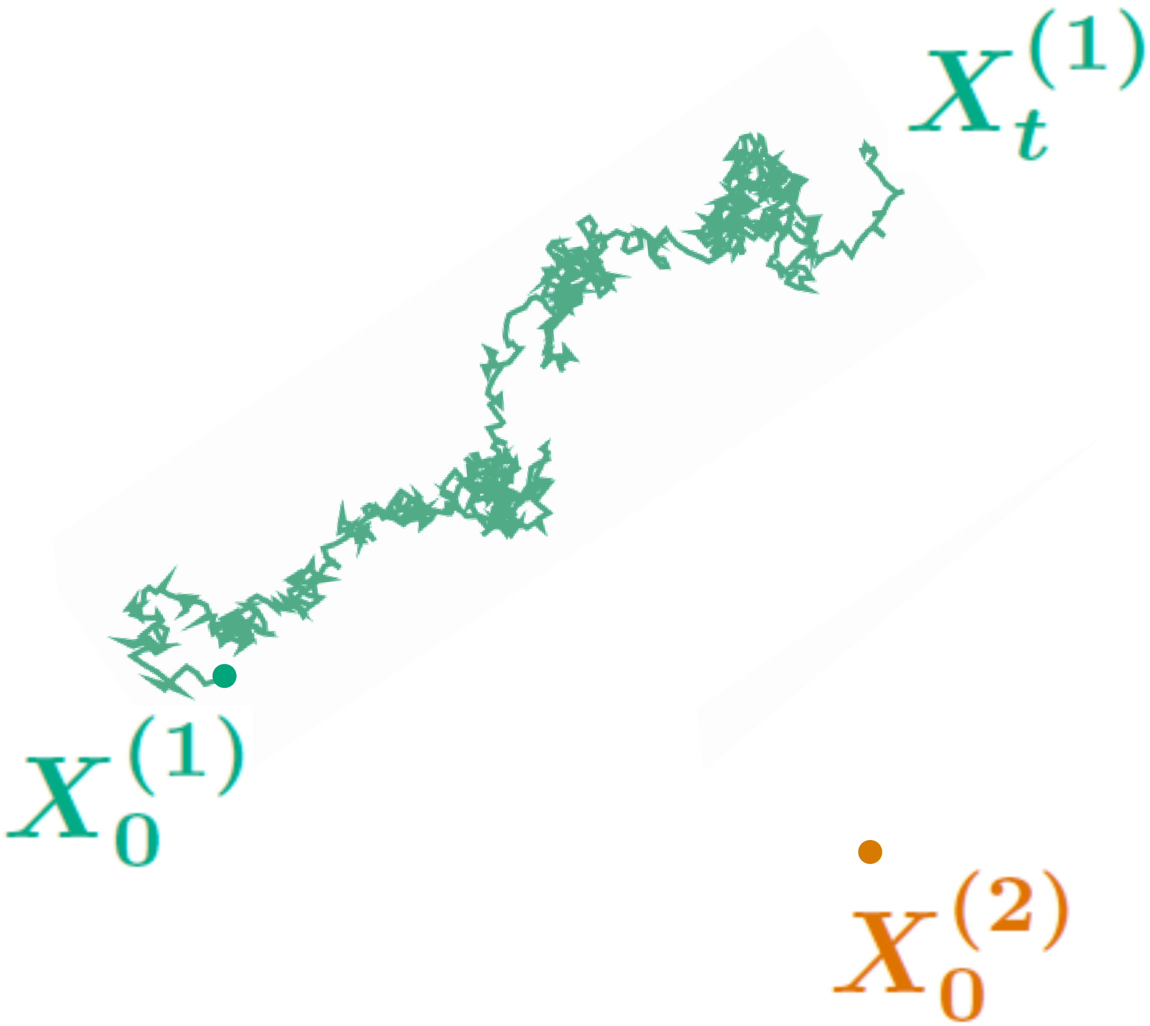
(a) $W_2(P_t^* \mu_0, P_t^* \mu_1) \leq \mathrm{e}^{-Kt} W_2(\mu_0, \mu_1)$

by studying a coupling by parallel transport of BMs

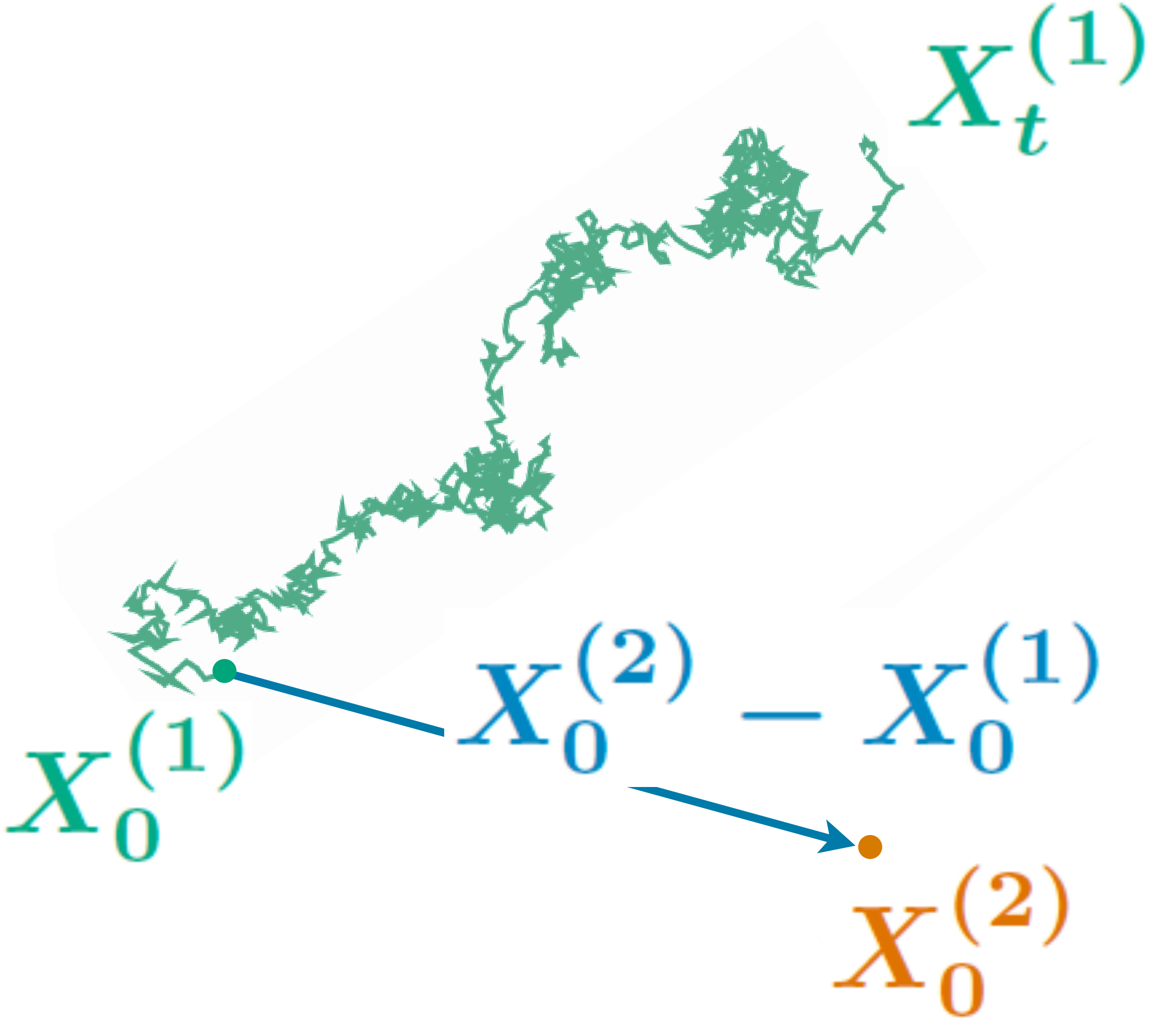
[F.-Y. Wang '05, K. '10, etc.]

Example (BM on $\mathbb{R}^m$)

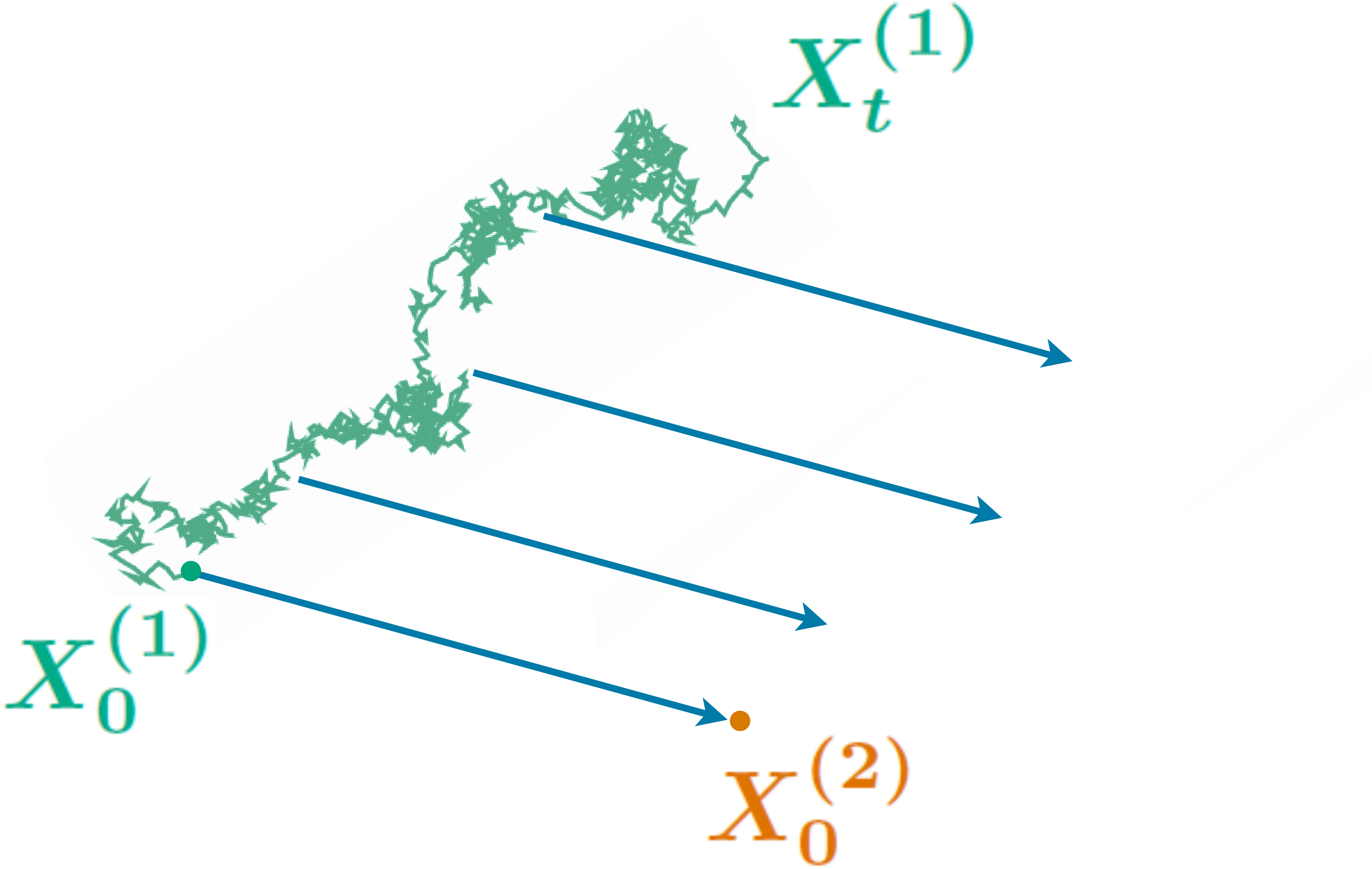
Example (BM on $\mathbb{R}^m$)



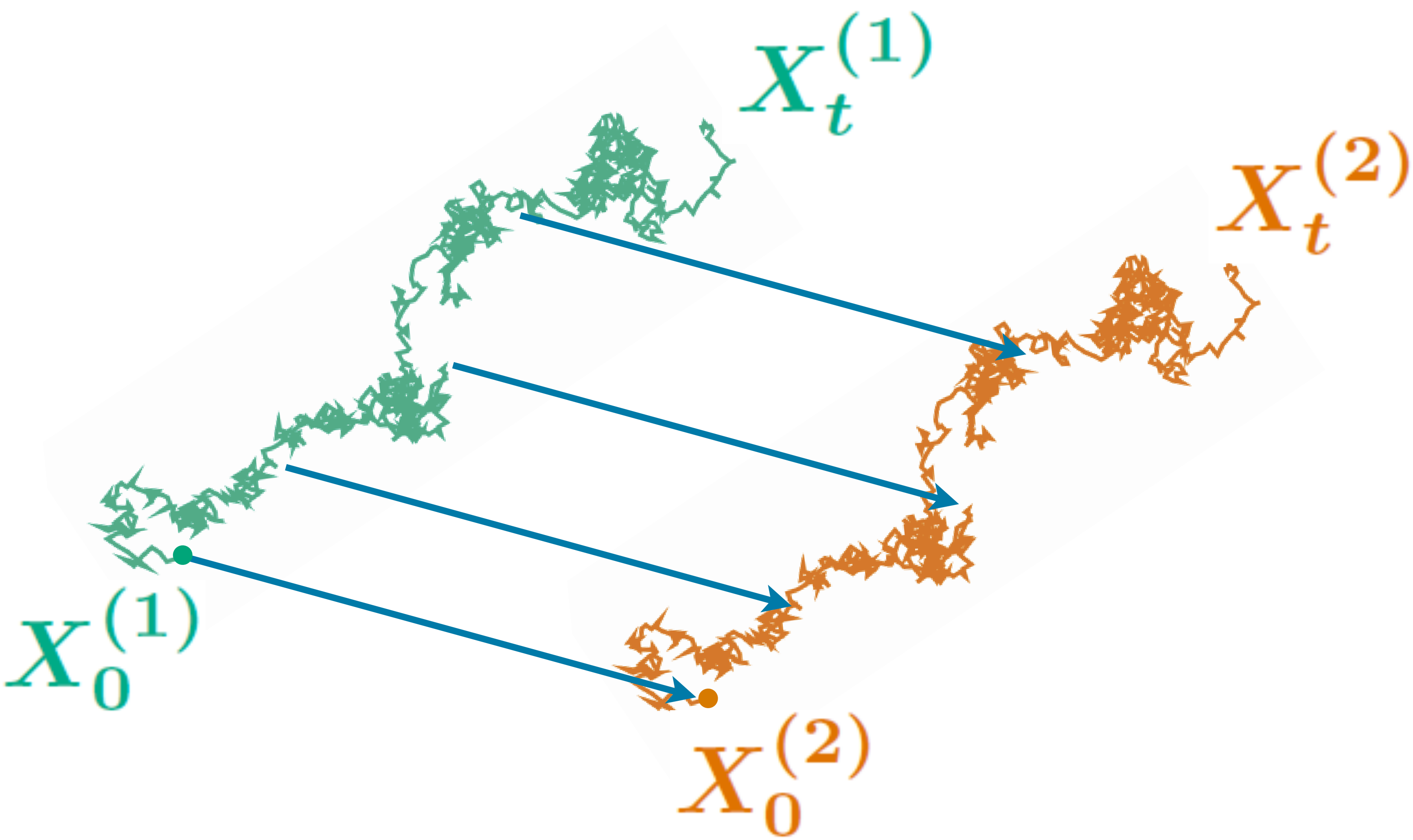
Example (BM on $\mathbb{R}^m$)



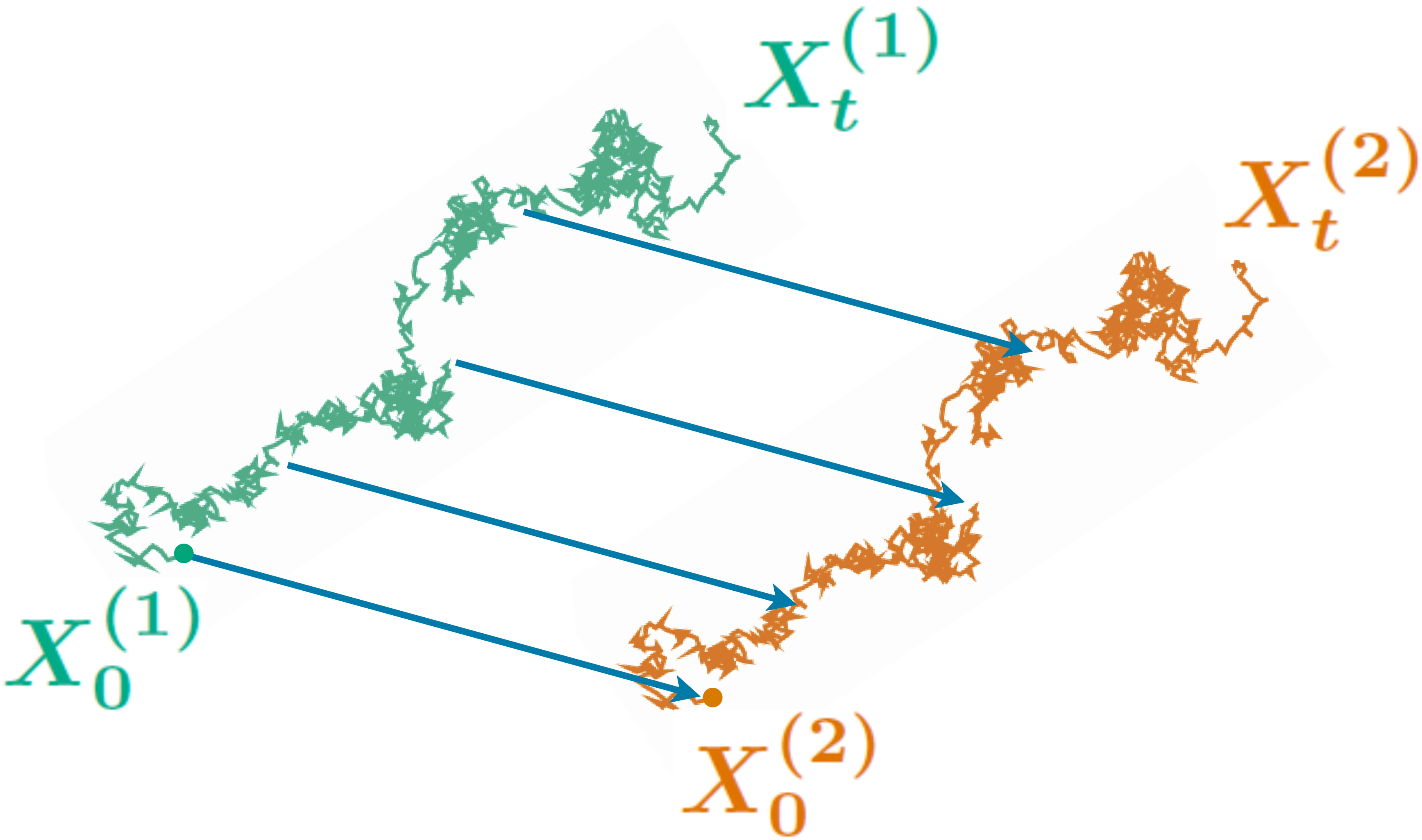
Example (BM on $\mathbb{R}^m$)



Example (BM on $\mathbb{R}^m$)

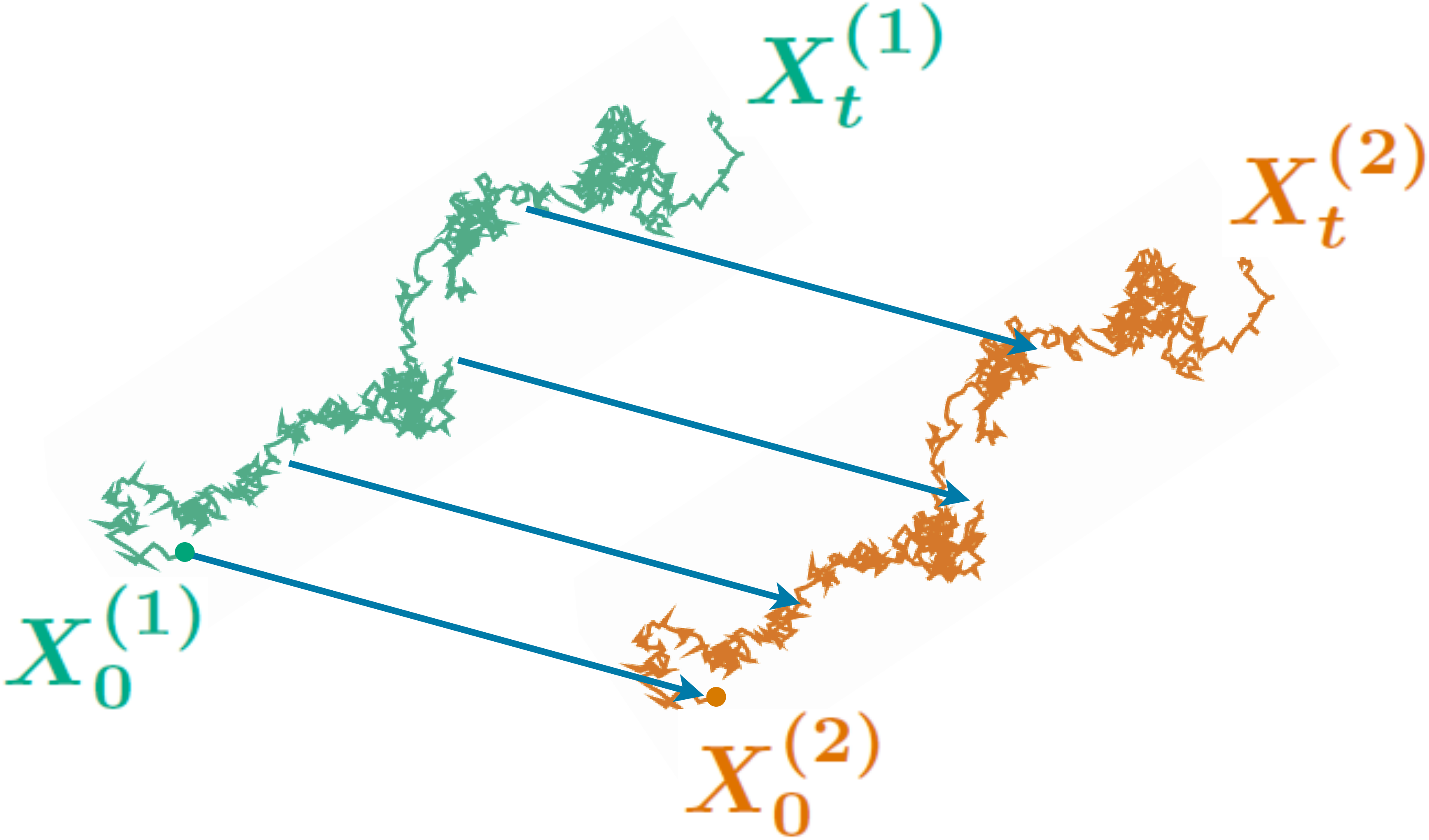


Example (BM on $\mathbb{R}^m$)



$$\Rightarrow |X_t^{(1)} - X_t^{(2)}| = |X_0^{(1)} - X_0^{(2)}|$$

Example (BM on $\mathbb{R}^m$)



$$\Rightarrow |X_t^{(1)} - X_t^{(2)}| = |X_0^{(1)} - X_0^{(2)}|$$

$$\Rightarrow W_2(P_t^* \mu_1, P_t^* \mu_2) \leq W_2(\mu_1, \mu_2)$$

Coupling by parallel transport on a Riem. mfd M

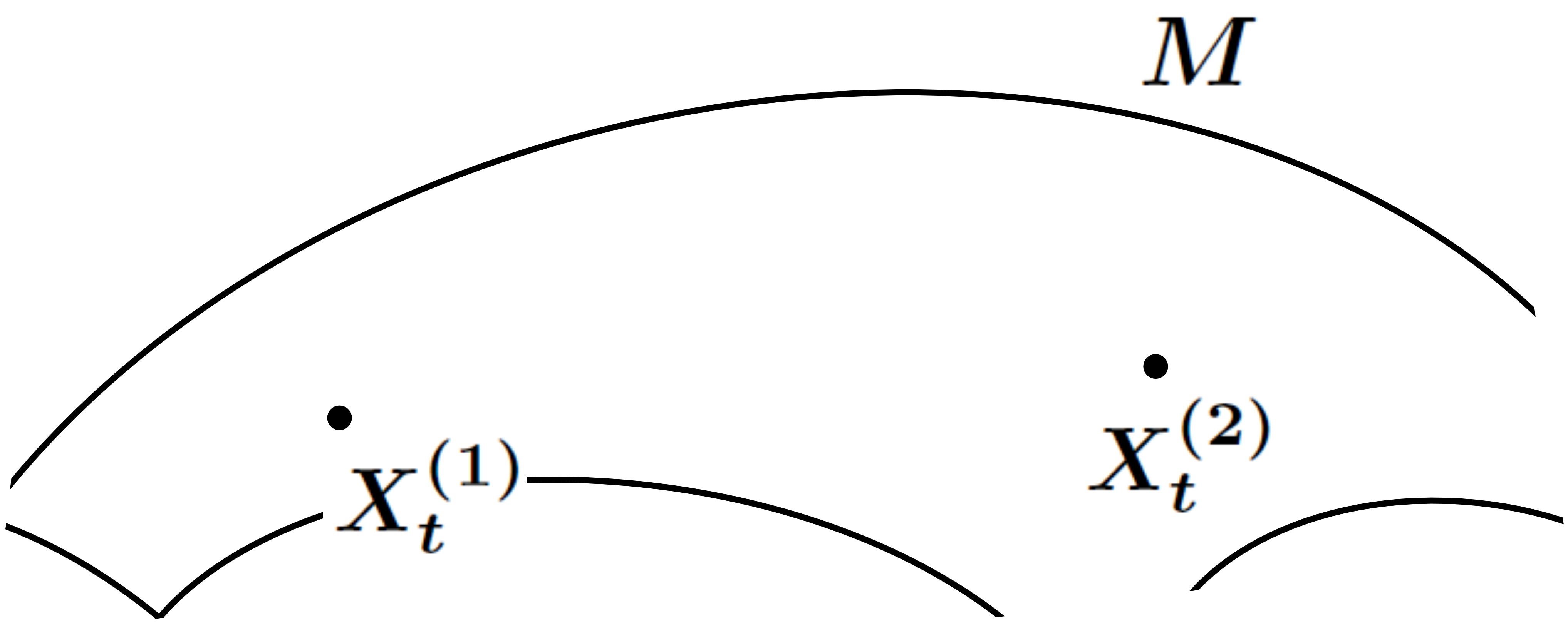
$(X_t^{(1)}, X_t^{(2)})$: coupled BMs on M,

driving noise $dB_t^{(2)}$ of $X_t^{(2)}$
$= $ parallel transport of $dB_t^{(1)}$ along a geod.

Coupling by parallel transport on a Riem. mfd M

$(X_t^{(1)}, X_t^{(2)})$: coupled BMs on M,

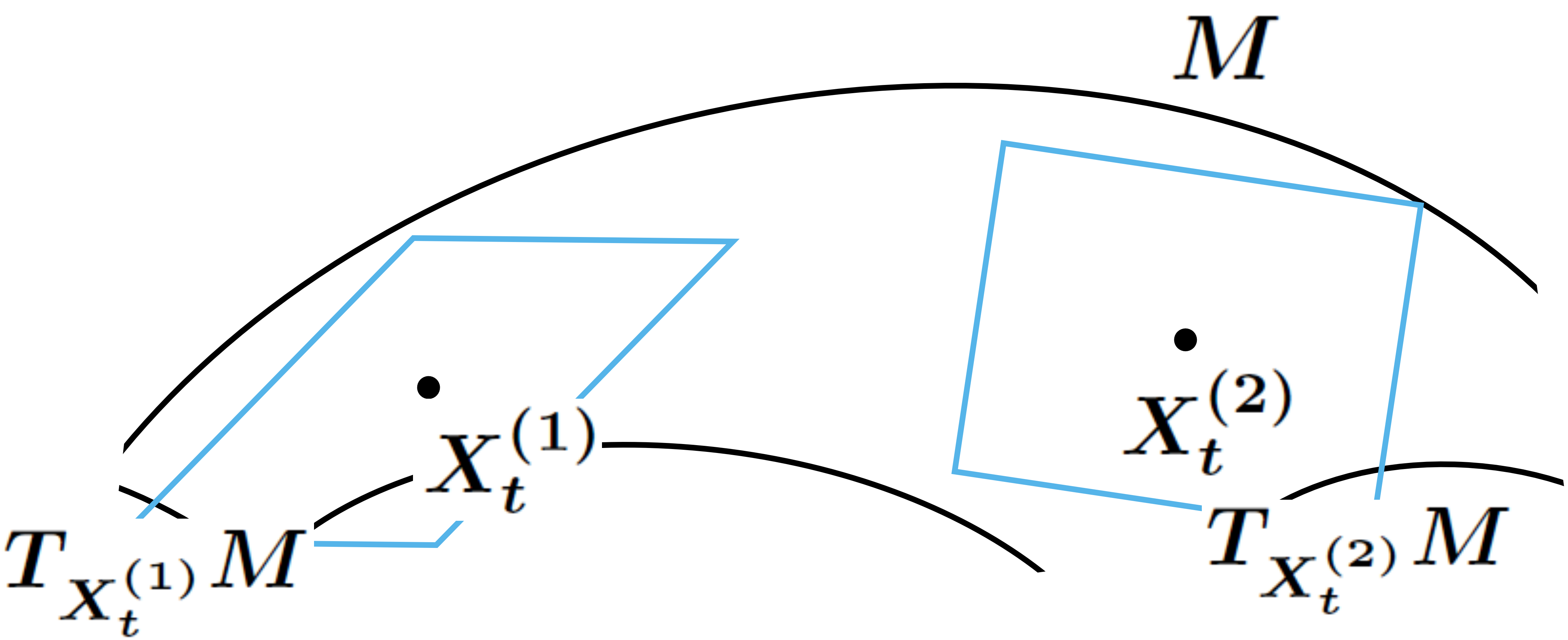
driving noise $dB_t^{(2)}$ of $X_t^{(2)}$
$= $ parallel transport of $dB_t^{(1)}$ along a geod.



Coupling by parallel transport on a Riem. mfd M

$(X_t^{(1)}, X_t^{(2)})$: coupled BMs on M,

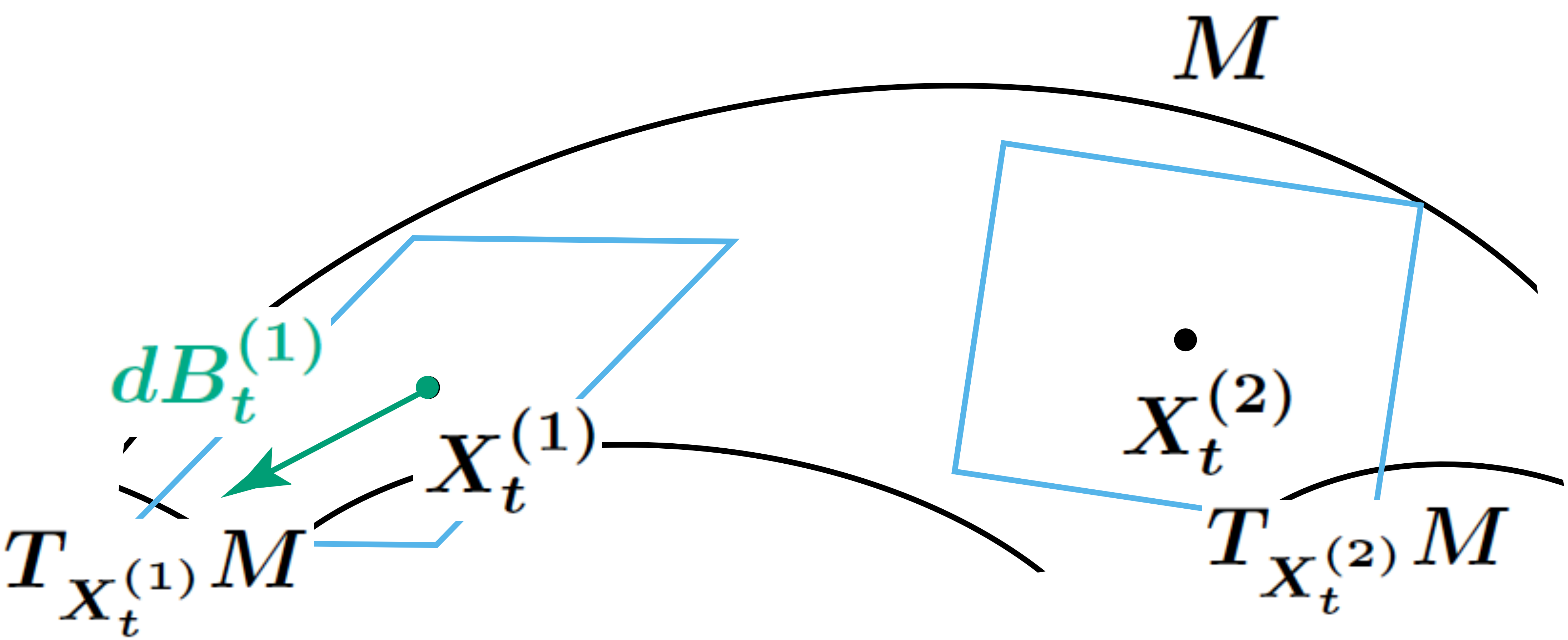
driving noise $dB_t^{(2)}$ of $X_t^{(2)}$
$= $ parallel transport of $dB_t^{(1)}$ along a geod.



Coupling by parallel transport on a Riem. mfd M

$(X_t^{(1)}, X_t^{(2)})$: coupled BMs on M,

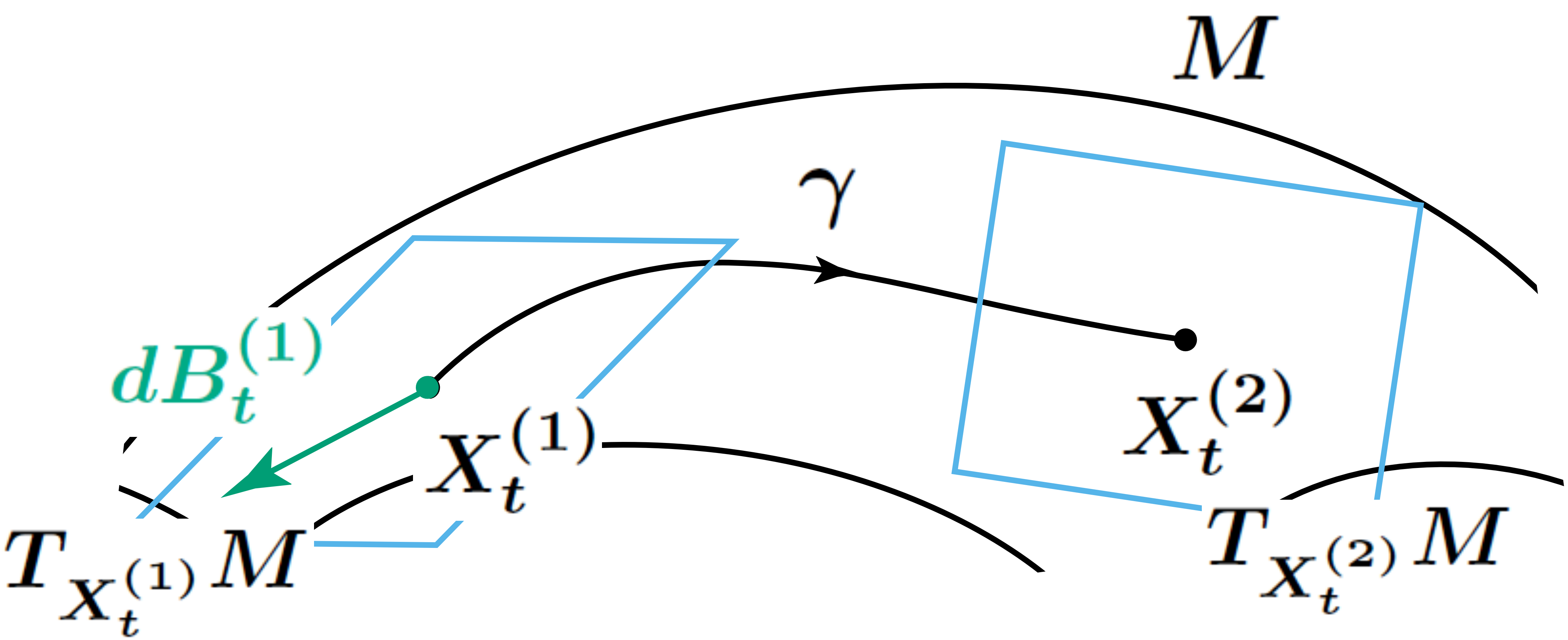
driving noise $dB_t^{(2)}$ of $X_t^{(2)}$
$=$ parallel transport of $dB_t^{(1)}$ along a geod.



Coupling by parallel transport on a Riem. mfd M

$(X_t^{(1)}, X_t^{(2)})$: coupled BMs on M,

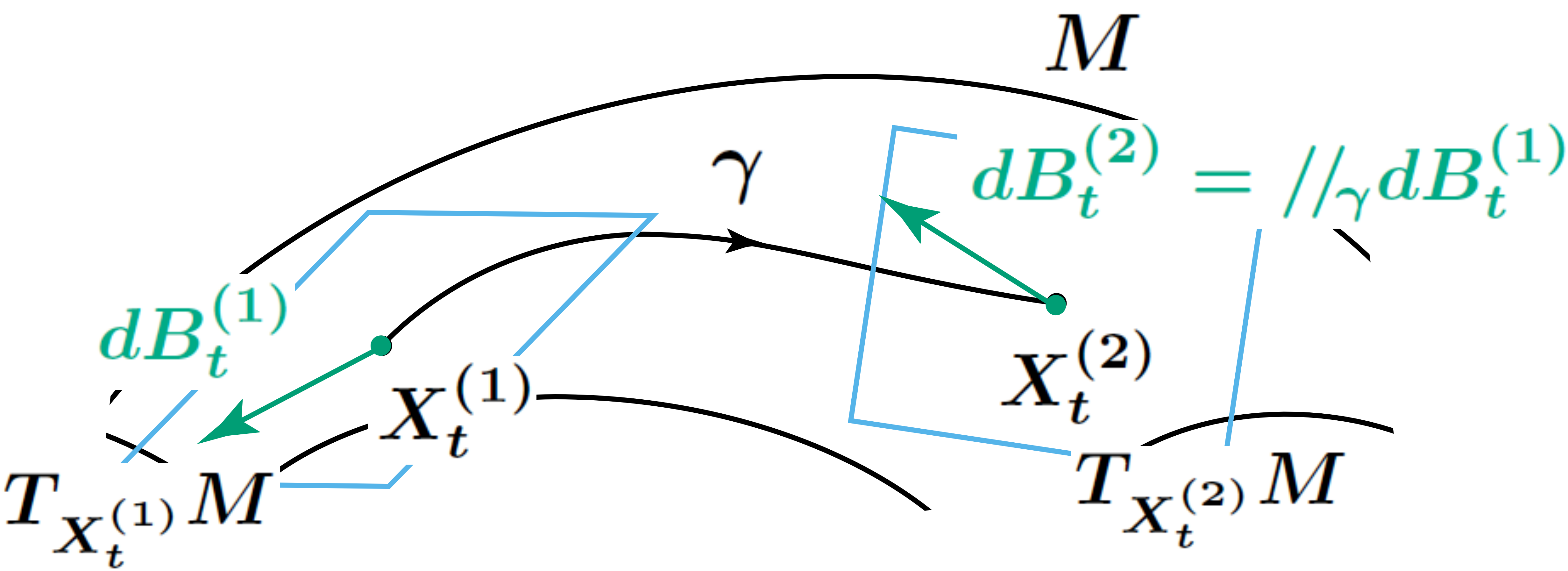
driving noise $dB_t^{(2)}$ of $X_t^{(2)}$
$=$ parallel transport of $dB_t^{(1)}$ along a geod.



Coupling by parallel transport on a Riem. mfd M

$(X_t^{(1)}, X_t^{(2)})$: coupled BMs on M,

> driving noise $dB_t^{(2)}$ of $X_t^{(2)}$
> $=$ parallel transport of $dB_t^{(1)}$ along a geod.



Idea of Pf: (b) $\Rightarrow$ (a)

- "Itô formula" for $\rho_t := d(X_t^{(1)}, X_t^{(2)})$

$$\Rightarrow d\rho_t = 0 + \sum_i (\nabla_{(e_i, e_i)})^2 d(X_t^{(1)}, X_t^{(2)}) dt$$

- Ric $\geq K$

$\Rightarrow$ (2nd) $\leq -K \rho_t dt$

$$\Rightarrow d(X_t^{(1)}, X_t^{(2)}) \leq \mathrm{e}^{-Kt} d(X_0^{(1)}, X_0^{(2)}) \qquad \square$$

Idea of Pf: (b) $\Rightarrow$ (a)

- "Itô formula" for $\rho_t := d(X_t^{(1)}, X_t^{(2)})$

$$\Rightarrow d\rho_t = 0 + \sum_i (\nabla_{(e_i, e_i)})^2 d(X_t^{(1)}, X_t^{(2)}) dt$$

- Ric $\geq K$

$\Rightarrow$ (2nd) $\leq -K \rho_t dt$

$$\Rightarrow d(X_t^{(1)}, X_t^{(2)}) \leq \mathrm{e}^{-Kt} d(X_0^{(1)}, X_0^{(2)}) \qquad \square$$

Remark

d is NOT differentiable at some off-diagonal points

$\Rightarrow$ technical difficulties

Time-dependent metric

$(g(t))_t$: complete Riem. metrics on M

- BM on $(M, g(t))_t$: diffusion $\leftrightarrow$ $\mathbf{\Delta}_{g(t)}$

Time-dependent metric

$(g(t))_t$: complete Riem. metrics on M

- BM on $(M, g(t))_t$: diffusion $\leftrightarrow \Delta_{g(t)}$

★ If $\partial_t g(t) \le 2\,\mathrm{Ric}_{g(t)} - 2K g(t)$, then

$$W_2^{d_t}(P_t^* \mu_1, P_t^* \mu_2) \le \mathrm{e}^{-Kt} W_2^{d_0}(\mu_1, \mu_2)$$

[McCann & Topping '10,
Arnaudon, Coulibaly & Thalmaier '09, K.]

(backward) Ricci flow: $\partial_t g(t) = 2\,\mathrm{Ric}_{g(t)}$

(backward) Ricci flow: $\partial_t g(t) = 2\,\mathrm{Ric}_{g(t)}$

- Perelman's $\mathcal{L}$-distance: $L(\tau_1, x; \tau_2, y)$

  $(\tau_1 < \tau_2,\ x, y \in M)$

- Given $0 \leq \bar{\tau}_1 < \bar{\tau}_2$,

  $\Theta_t(x, y)$: a normalization of $L(\bar{\tau}_1 t, x; \bar{\tau}_2 t, y)$

(backward) Ricci flow: $\partial_t g(t) = 2\,\mathrm{Ric}_{g(t)}$

- Perelman's $\mathcal{L}$-distance: $L(\tau_1, x; \tau_2, y)$
  $(\tau_1 < \tau_2,\ x, y \in M)$
- Given $0 \le \bar{\tau}_1 < \bar{\tau}_2$,
  $\Theta_t(x, y)$: a normalization of $L(\bar{\tau}_1 t, x; \bar{\tau}_2 t, y)$

Theorem 4 [Topping '10, K. & Philipowski '11]

Under t-unif. lower bound of $\mathrm{Ric}_{g(t)}$,

$^\exists (X_t^{(1)}, X_t^{(2)})$: coupling of $g(t)$-BMs s.t.

$$\Theta_t(X^{(1)}_{\bar{\tau}_1 t}, X^{(2)}_{\bar{\tau}_2 t}) \searrow \text{ a.s.}$$

(coupling by space-time parallel transport)

4. Dual approach

In Theorem 3

(M: cpl. Riem. mfd, P_t: heat semigroup),

(c) $|\nabla P_t f|(x)^2 \le \mathrm{e}^{-2Kt} P_t(|\nabla f|^2)(x)$

$\Downarrow$

(a) $W_2(P_t^* \mu_0, P_t^* \mu_1) \le \mathrm{e}^{-Kt} W_2(\mu_0, \mu_1)$

by passing through (b) $\mathrm{Ric} \ge K$

In Theorem 3

(M: cpl. Riem. mfd, P_t: heat semigroup),

(c) $|\nabla P_t f|(x)^2 \le \mathrm{e}^{-2Kt} P_t(|\nabla f|^2)(x)$

$\Downarrow$

(a) $W_2(P_t^* \mu_0, P_t^* \mu_1) \le \mathrm{e}^{-Kt} W_2(\mu_0, \mu_1)$

by passing through (b) $\mathrm{Ric} \ge K$

★ Direct proof is possible in a generalized framework

Framework

(M, d): Polish metric sp.

Framework

(M, d): Polish metric sp.

$p(x, dy)$: Markov kernel on M

Framework

(M, d): Polish metric sp.

$p(x, dy)$: Markov kernel on M

$$\Rightarrow P : C_b(M) \to C_b(M),$$

$$P^* : \mathcal{P}(M) \to \mathcal{P}(M)$$

Framework

(M, d): Polish metric sp.

$p(x, dy)$: Markov kernel on M

$$\Rightarrow P : C_b(M) \to C_b(M),$$

$$P^* : \mathcal{P}(M) \to \mathcal{P}(M)$$

(When M: Riem. mfd, $P = P_t$ for instance)

Framework

(M, d): Polish metric sp.

$p(x, dy)$: Markov kernel on M

$\Rightarrow P : C_b(M) \to C_b(M)$,

$P^* : \mathcal{P}(M) \to \mathcal{P}(M)$

(When M: Riem. mfd, $P = P_t$ for instance)

For $f : M \to M$,

$$|\nabla f|(x) := \limsup_{y \to x} \frac{|f(x) - f(y)|}{d(x, y)}$$

Given $C > 0$,

$$W_p(P^*\mu, P^*\nu) \le C W_p(\mu, \nu) \qquad (W_p)$$

$$|\nabla P f|(x) \le C P(|\nabla f|^q)(x)^{1/q} \qquad (G_q)$$

$$(\mu, \nu \in \mathcal{P}(X),\ f \in C_b^{\mathrm{Lip}}(X))$$

Given $C > 0$,

$$W_p(P^*\mu, P^*\nu) \le C W_p(\mu, \nu) \qquad (W_p)$$

$$|\nabla P f|(x) \le C P(|\nabla f|^q)(x)^{1/q} \qquad (G_q)$$

$$(\mu, \nu \in \mathcal{P}(X),\ f \in C_b^{\mathrm{Lip}}(X))$$

Theorem 5 [K. cf. K. '10]

For $p, q \in [1, \infty]$ with $\frac{1}{p} + \frac{1}{q} = 1$,

$$(W_p) \Leftrightarrow (G_q)$$

Idea of Pf: $(G_q) \Rightarrow (W_p)$ for $p \in (1, \infty)$

By the Kantorovich duality,

$$\frac{W_p(P^*\delta_x, P^*\delta_y)^p}{p} = \sup_f [PQ_1 f(x) - Pf(y)]$$

($Q_t f$: Hopf-Lax/Hamilton-Jacobi semigroup)

$\star\ \gamma : [0,1] \to M$, $\gamma(0) = y$, $\gamma(1) = x$

$$\Rightarrow PQ_1 f(x) - Pf(y) = \int_0^1 \partial_t (Q_t f(\gamma_t)) dt$$

$\therefore$ [calc. of $\partial_t (Q_t f(\gamma_t))$ & (G_q)] $\Rightarrow (W_p)$ $\square$

Previous argument: (iii) $\Rightarrow$ (ii)

$$|\nabla P_t f|(x)^2 \leq \mathrm{e}^{-2Kt} P_t(|\nabla f|^2)(x)$$

$\Downarrow$ ∂_t

Γ_2-condition:

$$\frac{1}{2}\Delta(|\nabla f|^2) - \langle \nabla f, \nabla \Delta f \rangle \geq K|\nabla f|^2$$

$\Downarrow$ Bochner formula

$$\mathrm{Ric} \geq K$$

Previous argument: (iii) $\Rightarrow$ (ii)

$$|\nabla P_t f|(x)^2 \le \mathrm{e}^{-2Kt} P_t(|\nabla f|^2)(x)$$

$\Downarrow$ ∂_t

Γ_2-condition:

$$\frac{1}{2}\Delta(|\nabla f|^2) - \langle \nabla f, \nabla \Delta f \rangle \ge K |\nabla f|^2$$

$\Downarrow$ Bochner formula

$$\mathrm{Ric} \ge K$$

Remark

- e^{-2Kt}: conti. at $t = 0$ is essential

Gradient estimates for hypoelliptic diffusions

[Driver, Melcher, Bakry, Baudoin, . . .]

$$|\nabla P_t f(x)|^p \leq C_p(t) P_t(|\nabla f|)(x)^p$$

with $|\nabla f|^2$: carré du champ $\leftrightarrow P_t$

Gradient estimates for hypoelliptic diffusions

[Driver, Melcher, Bakry, Baudoin, . . .]

$$|\nabla P_t f(x)|^p \leq C_p(t) P_t(|\nabla f|)(x)^p$$

with $|\nabla f|^2$: carré du champ $\leftrightarrow P_t$

- $C_p(t)$: NOT conti. at $t = 0$

Gradient estimates for hypoelliptic diffusions

[Driver, Melcher, Bakry, Baudoin, . . .]

$$|\nabla P_t f(x)|^p \le C_p(t) P_t(|\nabla f|)(x)^p$$

with $|\nabla f|^2$: carré du champ $\leftrightarrow P_t$

- $C_p(t)$: NOT conti. at $t = 0$
- Theorem 5 is applicable

Gradient estimates for hypoelliptic diffusions

[Driver, Melcher, Bakry, Baudoin, . . .]

$$|\nabla P_t f(x)|^p \le C_p(t) P_t(|\nabla f|)(x)^p$$

with $|\nabla f|^2$: carré du champ $\leftrightarrow P_t$

- $C_p(t)$: NOT conti. at $t = 0$
- Theorem 5 is applicable
- Coupling methods seems to be hard

(e.g. [Kendall '07])

5. Heat distribution as a gradient flow on $\mathcal{P}_2(M)$

In Theorem 3,

(M: cpl. Riem. mfd, P_t: heat semigroup),

(d) $\mathrm{CD}(K, \infty)$

$\Downarrow$

(a) $W_2(P_t^* \mu_0, P_t^* \mu_1) \le \mathrm{e}^{-Kt} W_2(\mu_0, \mu_1)$

([Erbar '10] for a direct proof)

In Theorem 3,

(M: cpl. Riem. mfd, P_t: heat semigroup),

(d) CD(K, ∞): For $^\forall W_2$-geod. $(\mu_t)_{t\in[0,1]}$,

$$\mathrm{Ent}(\mu_t) \le (1-t)\,\mathrm{Ent}(\mu_0) + t\,\mathrm{Ent}(\mu_1) - \frac{K}{2} t(1-t) W_2(\mu_0, \mu_1)^2$$

$\Downarrow$

(a) $W_2(P_t^* \mu_0, P_t^* \mu_1) \le \mathrm{e}^{-Kt} W_2(\mu_0, \mu_1)$

([Erbar '10] for a direct proof)

Heuristics [Otto '01]

$(\mu_t)_t$: heat distribution

$=$ a gradient curve of $-\mathrm{Ent}$ on $(\mathcal{P}_2(M), W_2)$

i.e.

$$\frac{d}{dt}\mu_t = -\nabla \mathrm{Ent}(\mu_t)$$

Heuristics [Otto '01]

$(\mu_t)_t$: heat distribution

$=$ a gradient curve of $-\mathrm{Ent}$ on $(\mathcal{P}_2(M), W_2)$

i.e.

$$\frac{d}{dt}\mu_t = -\nabla\,\mathrm{Ent}(\mu_t)$$

$$\Downarrow$$

$\mathrm{Hess}\,\mathrm{Ent} \geq K$

$\Rightarrow {}^\forall \mu_t^{(1)}, \mu_t^{(2)}$: gradient curves,

$$W_2(\mu_t^{(1)}, \mu_t^{(2)}) \leq \mathrm{e}^{-Kt} W_2(\mu_0^{(1)}, \mu_0^{(2)})$$

Heuristics [Otto '01]

$(\mu_t)_t$: heat distribution

$=$ a gradient curve of $-\mathrm{Ent}$ on $(\mathcal{P}_2(M), W_2)$

i.e.

$$\frac{d}{dt}\mu_t = -\nabla\,\mathrm{Ent}(\mu_t)$$

$$\Downarrow$$

$\mathrm{Hess}\,\mathrm{Ent} \geq K$

$\Rightarrow {}^\forall \mu_t^{(1)}, \mu_t^{(2)}$: gradient curves,

$$W_2(\mu_t^{(1)}, \mu_t^{(2)}) \leq \mathrm{e}^{-Kt} W_2(\mu_0^{(1)}, \mu_0^{(2)})$$

★ It works even when M is NOT smooth

Theorem 6

M: cpt. m-dim. Alexandrov sp., curvature $\geq k$

Theorem 6

M: cpt. m-dim. Alexandrov sp., curvature $\geq k$

(i) [Petrunin '09] CD(K, ∞), $K = (m-1)k$

Theorem 6

M: cpt. m-dim. Alexandrov sp., curvature $\geq k$

(i) [Petrunin '09] $\mathrm{CD}(K,\infty)$, $K=(m-1)k$

(ii) [Ohta '09] $\mu_t^{(i)}$: grad. curves of $-\,\mathrm{Ent}$

$$\Rightarrow W_2(\mu_t^{(1)},\mu_t^{(2)}) \leq \mathrm{e}^{-Kt} W_2(\mu_0^{(1)},\mu_0^{(2)})$$

Theorem 6

M: cpt. m-dim. Alexandrov sp., curvature $\geq k$

(i) [Petrunin '09] $\mathrm{CD}(K, \infty)$, $K = (m-1)k$

(ii) [Ohta '09] $\mu_t^{(i)}$: grad. curves of $-\mathrm{Ent}$

$$\Rightarrow W_2(\mu_t^{(1)}, \mu_t^{(2)}) \leq \mathrm{e}^{-Kt} W_2(\mu_0^{(1)}, \mu_0^{(2)})$$

(iii) [Gigli, K. & Ohta '10] $^\forall \mu_t$: grad. curve,

$$\mu_t = P_t^* \mu_0,$$

where P_t: the heat semigroup

$\leftrightarrow$ the canonical Dirichlet form on $L^2(M)$

The “metric” on $(\mathcal{P}_2(M), W_2)$
&
Gradient curve of $-\,\mathrm{Ent}$

One more property of W_2

One more property of W_2

For simplicity, suppose $M = \mathbb{R}^m$

One more property of W_2

For simplicity, suppose $M = \mathbb{R}^m$

Proposition [Brenier '91, McCann '95]

$\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^m)$, abs. conti. w.r.t. Leb. meas.

$^\exists \tilde{\varphi} : \mathbb{R}^m \to \mathbb{R}$: convex s.t.

- $(\nabla \tilde{\varphi})^\# \mu_0 = \mu_1$,
- $\dfrac{W_2(\mu_0, \mu_1)^2}{2} = \displaystyle\int_{\mathbb{R}^m} |x - \nabla \tilde{\varphi}(x)|^2 \mu_0(dx)$
- $\mu_t := ((1-t)I + t\nabla\tilde{\varphi})^\# \mu_0$

  $\Rightarrow (\mu_t)_t$: W_2-min. geod.

$$\varphi := \frac{1}{2}|\cdot|^2 - \tilde{\varphi}$$

$$\Rightarrow (1-t)x + t\nabla\varphi(x) = \exp_x(-t\nabla\varphi)$$

$$\varphi := \frac{1}{2}|\cdot|^2 - \tilde{\varphi}$$

$$\Rightarrow (1-t)x + t\nabla\varphi(x) = \exp_x(-t\nabla\varphi)$$

$$\Rightarrow \text{“}|\dot{\mu}_0|^2\text{”} = \int_{\mathbb{R}^m} |\nabla\varphi|^2 \, d\mu_0$$

$$\varphi := \frac{1}{2}|\cdot|^2 - \tilde{\varphi}$$

$$\Rightarrow (1-t)x + t\nabla\varphi(x) = \exp_x(-t\nabla\varphi)$$

$$\Rightarrow \text{“}|\dot{\mu}_0|^2\text{”} = \int_{\mathbb{R}^m} |\nabla\varphi|^2 \, d\mu_0$$

“Natural” definition

- Tangent space at $\mu \in \mathcal{P}_2(\mathbb{R}^m)$:

$$T_\mu\mathcal{P}_2(\mathbb{R}^m) := \overline{\{\nabla\varphi \mid \varphi \in C^\infty(\mathbb{R}^m)\}}^{L^2(\mu)}$$

- Riem. metric on $T_\mu\mathcal{P}_2(\mathbb{R}^m)$:

$$\sigma(\nabla\varphi, \nabla\psi)(\mu) := \int_{\mathbb{R}^m} \langle\nabla\varphi, \nabla\psi\rangle d\mu$$

"Regular" curve in $\mathcal{P}_2(\mathbb{R}^m)$

$\varphi_t \in C_0^\infty(\mathbb{R}^m)$, Φ_t: grad. flow of φ_t on $\mathbb{R}^m$

$\mu_t := \Phi_t^{\#}\mu$ ($\Rightarrow \nabla\varphi_t \in T_{\mu_t}\mathcal{P}_2(\mathbb{R}^m)$)

"Regular" curve in $\mathcal{P}_2(\mathbb{R}^m)$

$\varphi_t \in C_0^\infty(\mathbb{R}^m)$, Φ_t: grad. flow of φ_t on $\mathbb{R}^m$

$\mu_t := \Phi_t^\# \mu$ ($\Rightarrow \nabla\varphi_t \in T_{\mu_t}\mathcal{P}_2(\mathbb{R}^m)$)

$$\Rightarrow \frac{d}{dt}\int_{\mathbb{R}^m} f\, d\mu_t = \frac{d}{dt}\int_{\mathbb{R}^m} f \circ \Phi_t\, d\mu$$

$$= \int_{\mathbb{R}^m} \langle (\nabla f)\circ \Phi_t, \partial_t \Phi_t \rangle d\mu$$

$$= \int_{\mathbb{R}^m} \langle \nabla f, \nabla \varphi_t \rangle d\mu_t$$

$$\Rightarrow \frac{d}{dt}\mu_t = \mathrm{div}_{\mu_t}(\nabla \varphi_t)\mu_t \text{ (weakly)}$$

Gradient of Ent

For $\mu_t = \Phi_t^{\#}\mu = \rho_t v$ (v: Leb. meas.),

Gradient of Ent

For $\mu_t = \Phi_t^{\#}\mu = \rho_t v$ (v: Leb. meas.),

$$\frac{d}{dt}\,\mathrm{Ent}(\mu_t)|_{t=0} = \frac{d}{dt}\int_{\mathbb{R}^m} \log\rho_t\, d\mu_t\Big|_{t=0}$$

$$= \int_{\mathbb{R}^m} \partial_t\rho_0\, dv + \int_{\mathbb{R}^m} \langle \frac{\nabla\rho_0}{\rho_0}, \nabla\varphi_0\rangle d\mu$$

$$= \sigma(\frac{\nabla\rho_0}{\rho_0}, \nabla\varphi_0)(\mu)$$

Gradient of Ent

For $\mu_t = \Phi_t^{\#}\mu = \rho_t v$ (v: Leb. meas.),

$$\frac{d}{dt}\mathrm{Ent}(\mu_t)|_{t=0} = \frac{d}{dt}\int_{\mathbb{R}^m} \log \rho_t \, d\mu_t \Big|_{t=0}$$

$$= \int_{\mathbb{R}^m} \partial_t \rho_0 \, dv + \int_{\mathbb{R}^m} \langle \frac{\nabla \rho_0}{\rho_0}, \nabla \varphi_0 \rangle d\mu$$

$$= \sigma(\frac{\nabla \rho_0}{\rho_0}, \nabla \varphi_0)(\mu)$$

$$\Rightarrow \boxed{\frac{d}{dt}\mu_t = -\nabla \mathrm{Ent}(\mu_t) \text{ iff } \nabla \varphi_t = -\frac{\nabla \rho_t}{\rho_t}}$$

- $$\frac{d}{dt}\mu_t = \mathrm{div}_{\mu_t}(\nabla\varphi_t)\mu_t$$
- $$\frac{d}{dt}\mu_t = -\nabla\,\mathrm{Ent}(\mu_t) \text{ iff } \nabla\varphi_t = -\frac{\nabla\rho_t}{\rho_t}$$

- $\frac{d}{dt}\mu_t = \mathrm{div}_{\mu_t}(\nabla \varphi_t)\mu_t$
- $\frac{d}{dt}\mu_t = -\nabla \,\mathrm{Ent}(\mu_t)$ iff $\nabla \varphi_t = -\frac{\nabla \rho_t}{\rho_t}$

$\Rightarrow$ When μ_t: grad. curve of $-\,\mathrm{Ent}$,

$$\frac{d}{dt}\int_{\mathbb{R}^m} f \, d\mu_t = -\int_{\mathbb{R}^m} \langle \nabla f, \nabla \rho_t \rangle dv$$

$$= \int_{\mathbb{R}^m} \Delta f \, d\mu_t$$

$\therefore$ μ_t solves the heat equation (weakly)

6. Curvature-dimension conditions

Framework: M: cpl. Riem. mfd, P_t: heat semigroup

Framework: M: cpl. Riem. mfd, P_t: heat semigroup

$$\mathrm{Ric} \geq K$$

$$\Leftrightarrow \frac{1}{2}\Delta(|\nabla f|^2) - \langle \nabla f, \nabla \Delta f \rangle \geq K|\nabla f|^2$$

[Bakry & Émery '84]

$$\Leftrightarrow |\nabla P_t f|^2 \leq \mathrm{e}^{-2Kt} P_t(|\nabla f|^2)$$

Framework: M: cpl. Riem. mfd, P_t: heat semigroup

$$\mathrm{Ric} \geq K \;\&\; \dim M \leq N$$

$$\Leftrightarrow \frac{1}{2}\Delta(|\nabla f|^2) - \langle \nabla f, \nabla \Delta f \rangle \geq K|\nabla f|^2 + \frac{1}{N}(\Delta f)^2$$

[Bakry & Émery '84]

Framework: M: cpl. Riem. mfd, P_t: heat semigroup

$$\mathrm{Ric} \geq K \ \& \ \dim M \leq N$$

$$\Leftrightarrow \frac{1}{2}\Delta(|\nabla f|^2) - \langle \nabla f, \nabla \Delta f \rangle \geq K|\nabla f|^2 + \frac{1}{N}(\Delta f)^2$$

[Bakry & Émery '84]

$$\Leftrightarrow |\nabla P_t f|^2 \leq \mathrm{e}^{-2Kt} P_t(|\nabla f|^2) + \frac{1 - \mathrm{e}^{-2Kt}}{NK}(\Delta P_t f)^2$$

[F.-Y. Wang '10]

Theorem 7 [K.]

The last inequality is equivalent to the following:

$$W_2(P_s^* \mu_1, P_t^* \mu_2)^2 \leq \frac{e^{-2Kt} - e^{-2Ks}}{2K(s-t)} W_2(\mu_0, \mu_1)^2 + (s-t) \int_t^s \frac{NK}{e^{2Ku} - 1} du$$