Essentially non-branchingness of Riemannian/Lorentzian mfds w. lower regularity metrics

Sabrina Lin
University of Toronto

Essentially non-branchingness of

Riemannian/Lorentzian mfds w. lower regularity metrics

$(+,+,\dots,+)$ $(-,+,+,\dots,+)$

Sabrina Lin

University of Toronto

Riemannian mfd

hard

$$(\overset{\in C^\infty}{M}, \overset{\in C^\infty}{g_{ij}})$$

Riemannian mfd

hard

$(M^{\in C^\infty}, g_{ij}^{\in C^\infty})$

Riemannian mfd

hard

Metric sp

$$(M^{\in C^\infty}, g_{ij}^{\in C^\infty})$$

Riemannian mfd

hard

Metric sp

easy

$$(M^{\in C^\infty}, g_{ij}^{\in C^\infty})$$

Riemannian mfd

hard

$$(X, d)$$

Metric sp

easy

Curvature dimension bounds on (X, d, m)

Curvature dimension bounds on (X, d, m)

Ex: $CD^e(K, N)$

Curvature dimension bounds on (X, d, m)

Ex: $CD^e(K,N)$

$\Downarrow$:

Curvature dimension bounds on (X, d, m)

Ex: $CD^e(K,N)$

$\in C^\infty$

$\Downarrow$: Φ $\quad (M^n, g_{ij}, e^{-V} d\mathrm{vol}_g)$

Curvature dimension bounds on (X, d, m)

Ex: $CD^e(K,N)$

$\Updownarrow$: $\exists$ $(M^n, g_{ij}, e^{-V} dvol_g)$ is $CD^e(K,N)$

$\in C^\infty$

Curvature dimension bounds on (X,d,m)

Ex: $CD^e(K,N)$

$\Downarrow$: ① $(M^n, g_{ij}, e^{-V} dvol_g)$ is $CD^e(K,N)$ $\quad V \in C^\infty$

$\Leftrightarrow$ $Ric^{(N,V)} \geq K$ and $n \leq N$

Curvature dimension bounds on (X,d,m)

Ex: $CD^e(K,N)$

$\Downarrow$: ① $(M^n, g_{ij}, e^{-V} d\mathrm{vol}_g)$ is $CD^e(K,N)$, $V \in C^\infty$

$\Leftrightarrow \mathrm{Ric}^{(N,V)} \geq K$ and $n \leq N$

② Stable under GH

Curvature dimension bounds on (X, d, m)

Ex: $CD^e(K,N)$

$\Downarrow$: ① $(M^n, g_{ij}, e^{-V} dvol_g)$ is $CD^e(K,N)$, $V \in C^\infty$

$\Leftrightarrow Ric^{(N,V)} \geq K$ and $n \leq N$

② Stable under GH . . . ..

Curvature dimension bounds on (X, d, m)

Ex: $CD^e(K,N)$

$\Downarrow$: ① $(M^n, g_{ij}, e^{-V} d\mathrm{vol}_g)$ is $CD^e(K,N)$ ($g_{ij} \in C^\infty$)

$\Leftrightarrow Ric^{(N,V)} \geq K$ and $n \leq N$

② Stable under GH . . . . .

Curvature dimension bounds on (X,d,m)

Ex: $CD^e(K,N)$

$\in C^\infty$

☺ : ① $(M^n, g_{ij}, e^{-V} d\mathrm{vol}_g)$ is $CD^e(K,N)$

$\Leftrightarrow \mathrm{Ric}^{(N,V)} \geq K$ and $n \leq N$

② Stable under GH $\cdot \quad \cdot \quad \cdot \quad \cdot\cdot\cdot$

$CD^e(K,N)$

Curvature dimension bounds on (X,d,m)

Ex: $CD^e(K,N)$

① $(M^n, g_{ij}, e^{-V} d\mathrm{vol}_g)$ is $CD^e(K,N)$, $V \in C^\infty$

$\Leftrightarrow$ $Ric^{(N,V)} \geq K$ and $n \leq N$

② Stable under GH

$CD^e(K,N)$

$CD^e(K,N)$

Curvature dimension bounds on (X, d, m)

Ex: $CD^e(K,N)$

① $\underbrace{(M^n, g_{ij}, e^{-V} dvol_g)}_{\circledast}$ is $CD^e(K,N)$, $\quad V \in C^\infty$

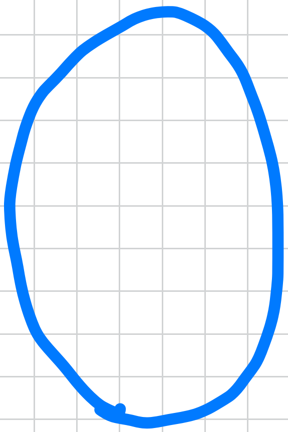
$\Longleftrightarrow \quad Ric^{(N,V)} \geq K$ and $n \leq N$

② Stable under GH $\quad \cdot \quad \cdot \quad \cdot \quad \cdot\cdot\; \to$ $CD^e(K,N)$ (red); $CD^e(K,N)$ (blue)

③ $\circledast$ is dense in $CD^e(K,N)$

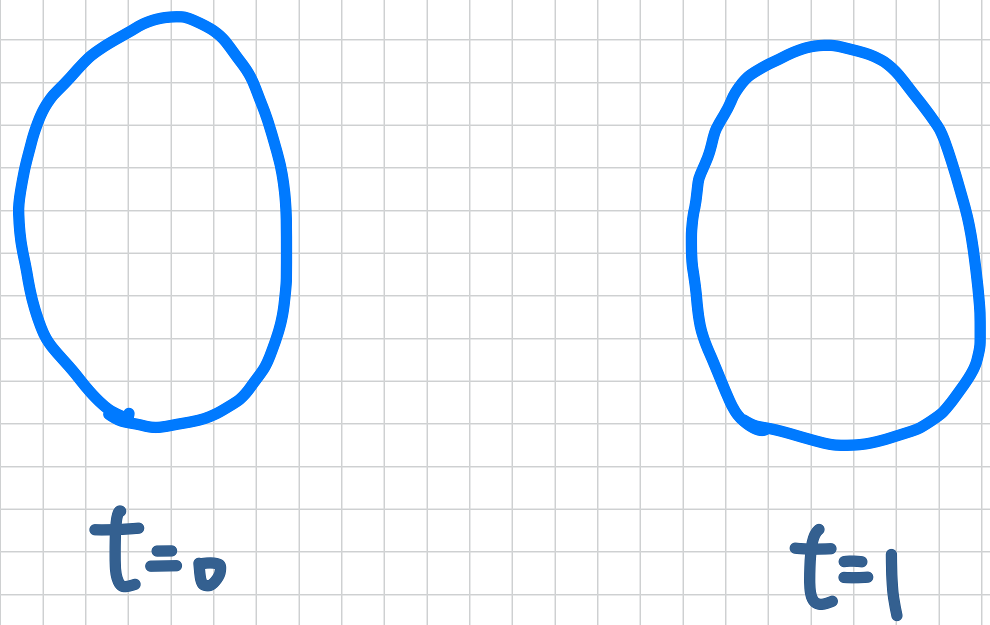
Motivation: Lazy Gas Experiment

Motivation: Lazy Gas Experiment

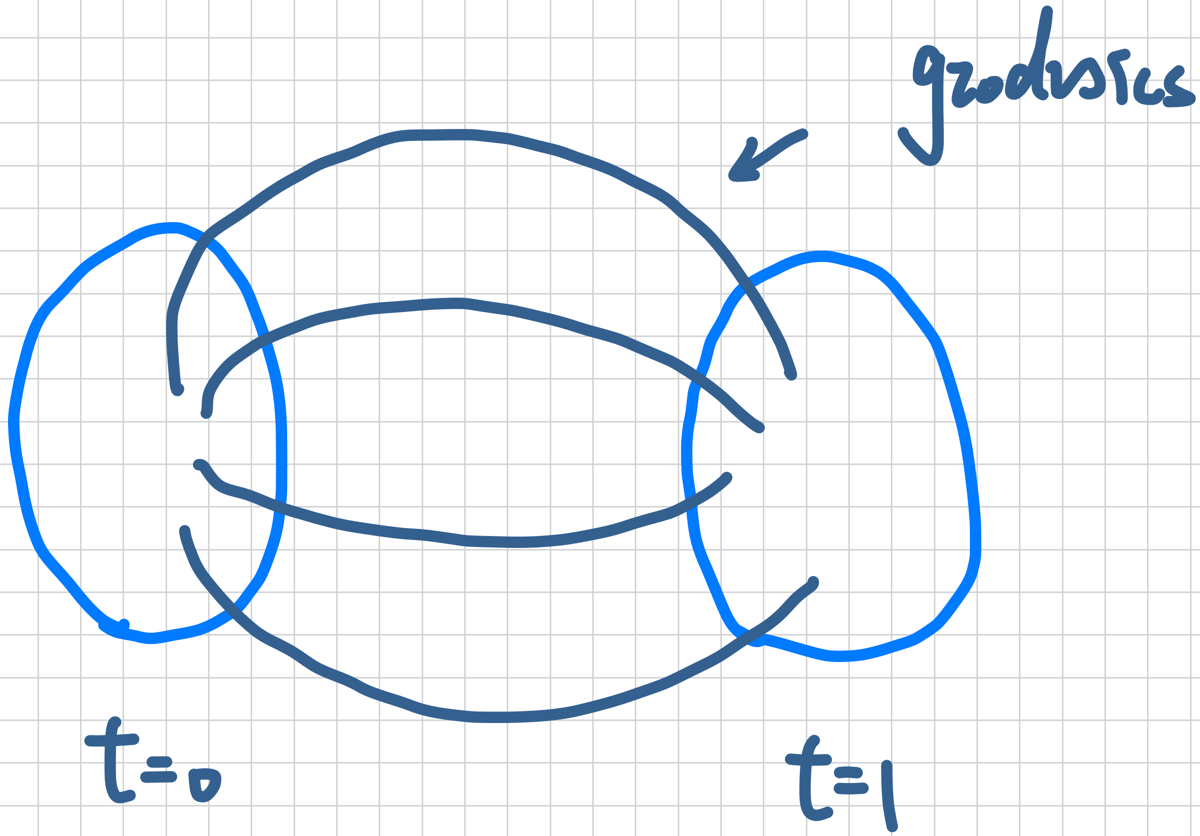


$t=0$

Motivation: Lazy Gas Experiment



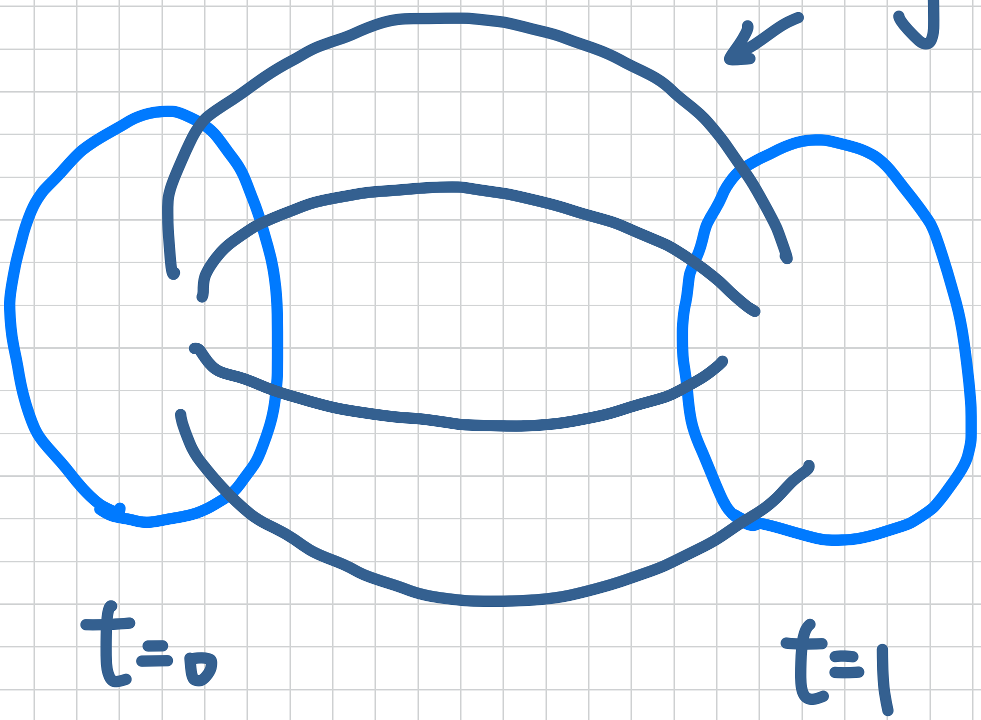
Motivation: Lazy Gas Experiment



Motivation: Lazy Gas Experiment

On smooth Riemannian mfd

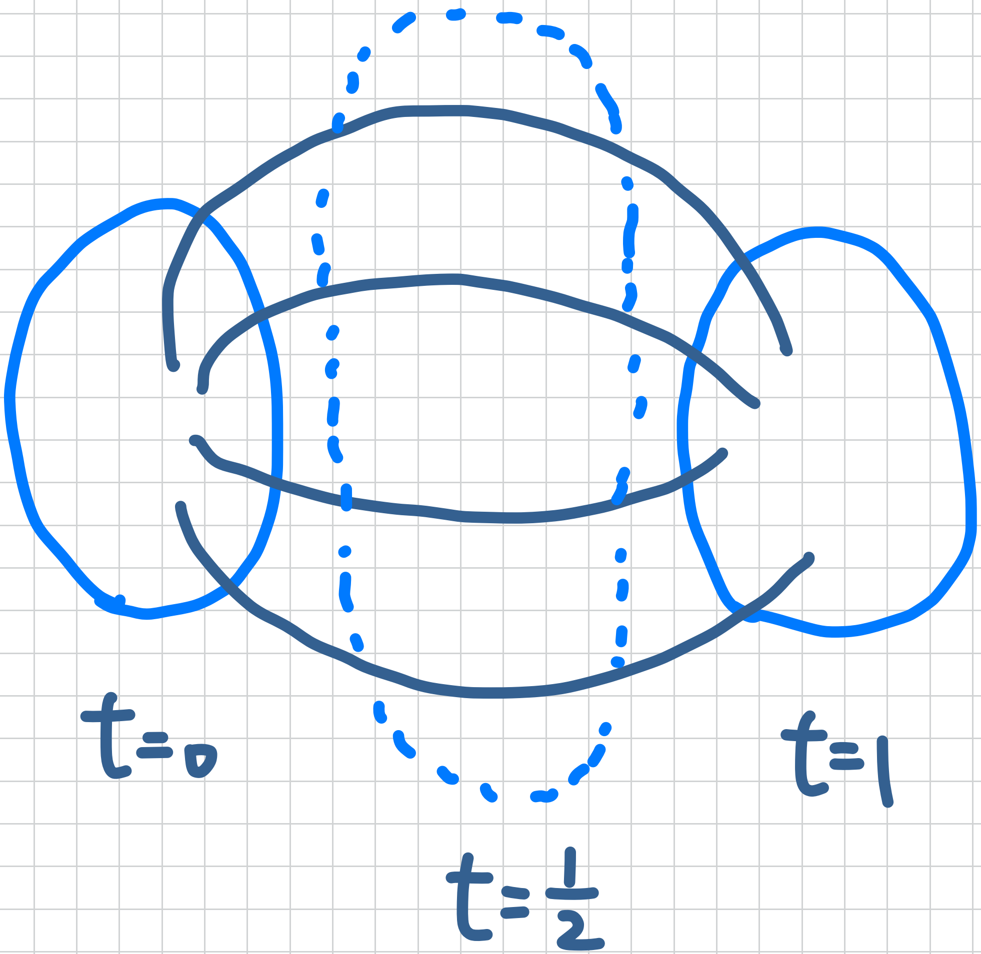
$\mathrm{Ric} \geq 0$.



Motivation: Lazy Gas Experiment

On smooth Riemannian mfd

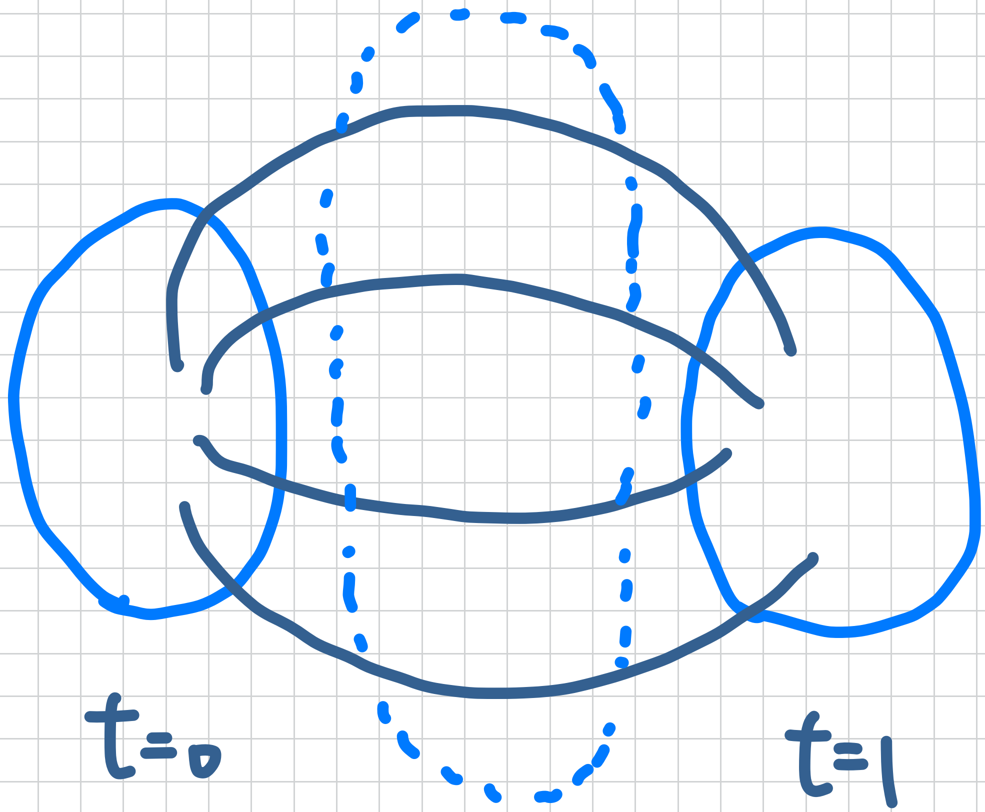
$Ric \geq 0$.

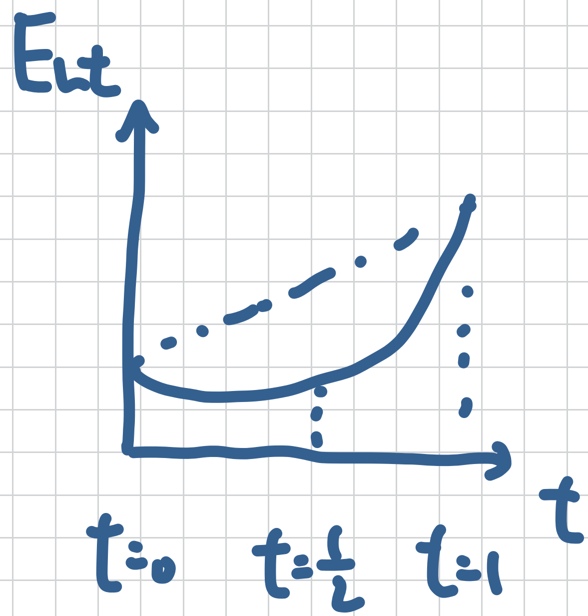


Motivation: Lazy Gas Experiment

On smooth Riemannian mfd

$Ric \geq 0$.

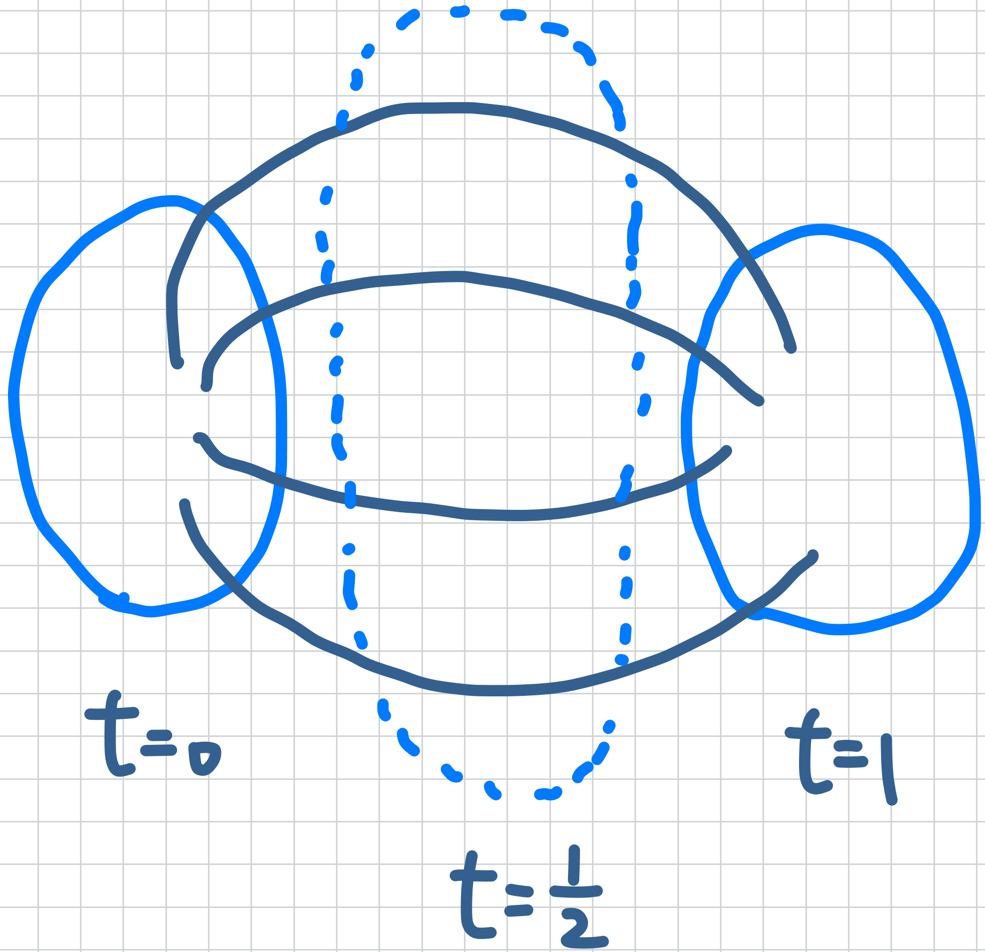


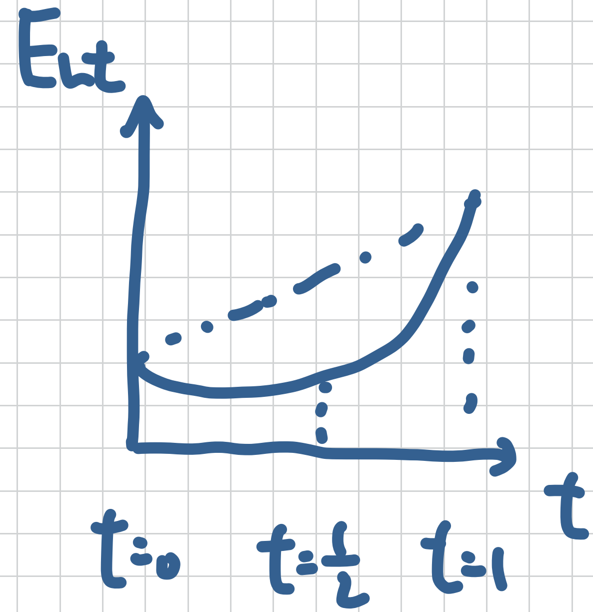


Motivation: Lazy Gas Experiment

On smooth Riemannian mfd

$Ric \geq 0$.



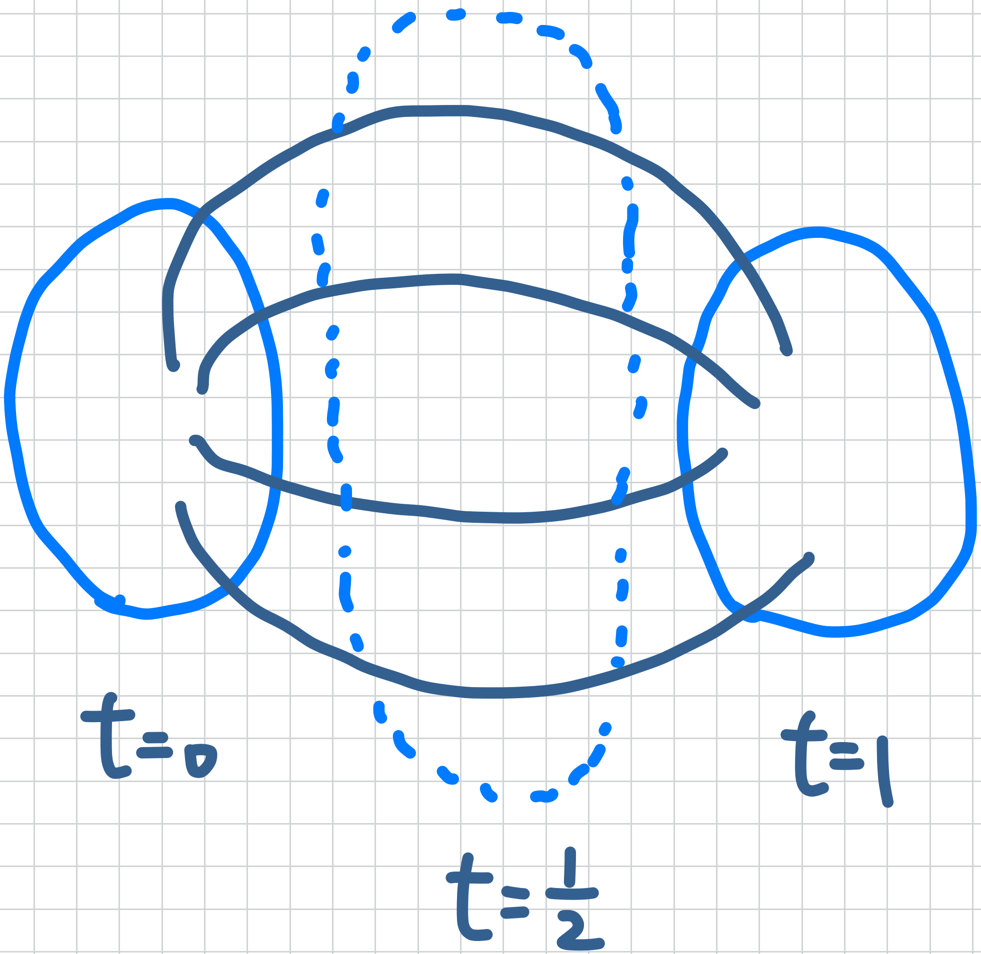


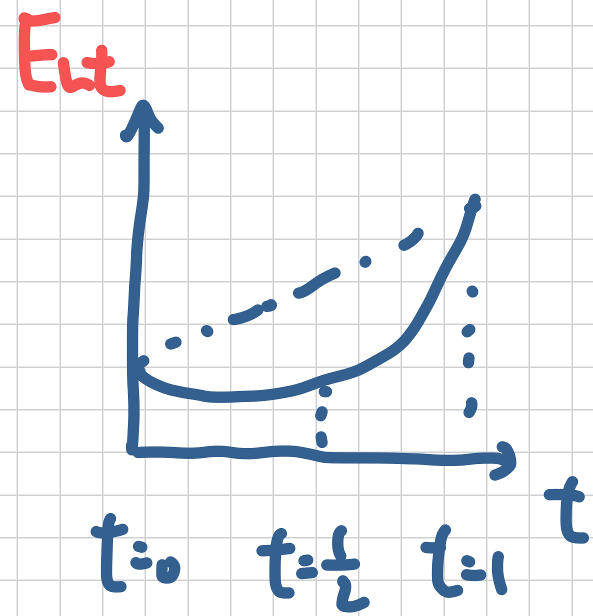
Convex

Motivation: Lazy Gas Experiment

On smooth Riemannian mfd

$Ric \geq 0$.





Convex

Curvature dimension bounds on (X,d,m)

Ex: $CD^e(K,N)$

$\in C^\infty$

① $\underbrace{(M^n, g_{ij}, e^{-V}\, d\mathrm{vol}_g)}_{\circledast}$ is $CD^e(K,N)$

$\Leftrightarrow \mathrm{Ric}^{(N,V)} \geq K$ and $n \leq N$

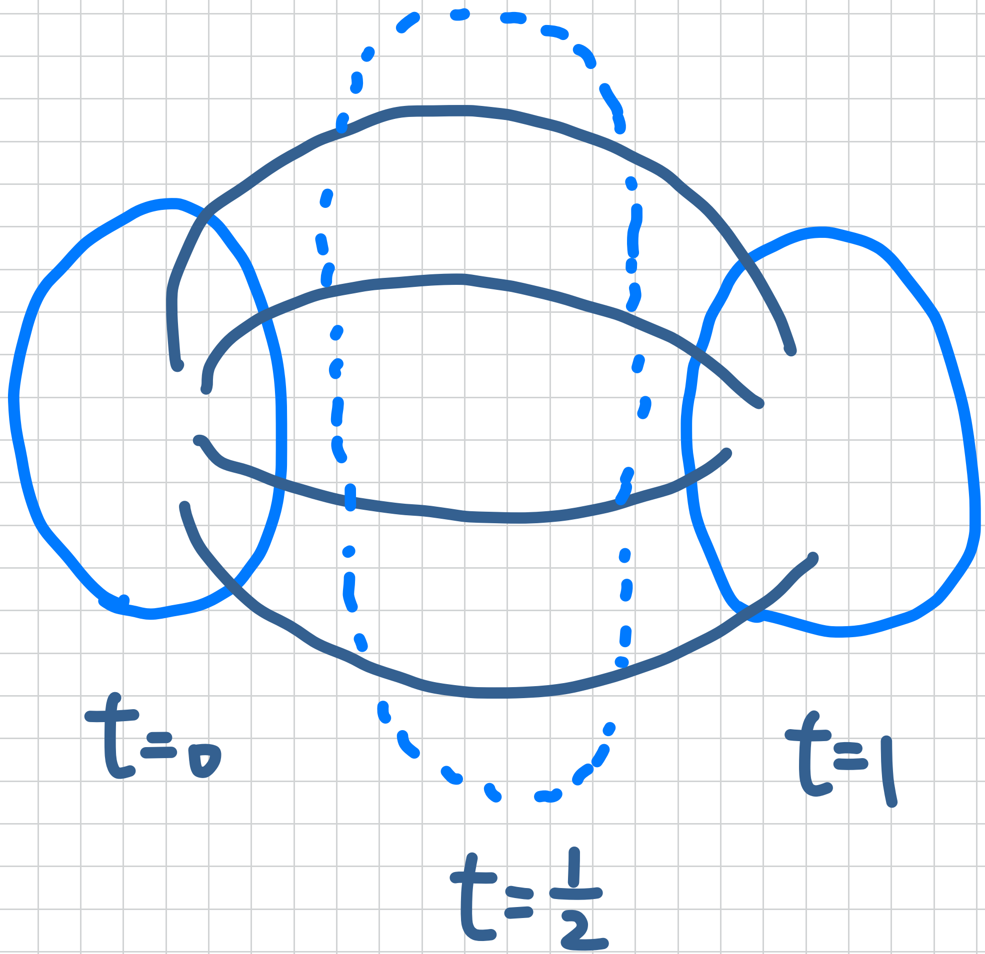
$CD^e(K,N)$

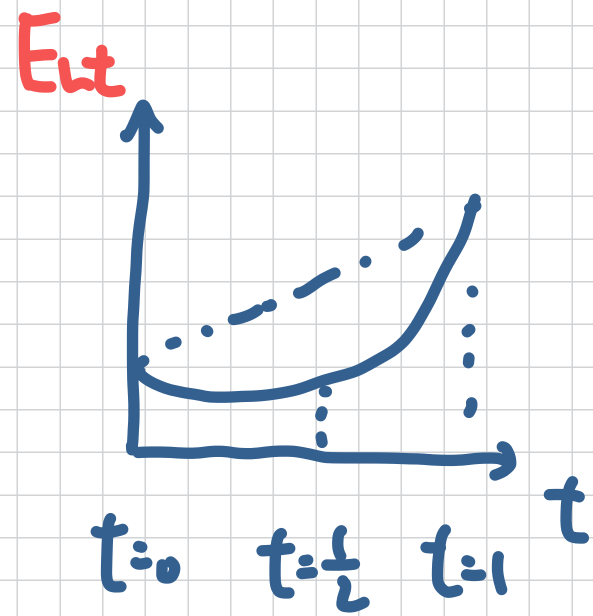
② Stable under GH

$CD^e(K,N)$

③ $\circledast$ is dense in $CD^e(K,N)$

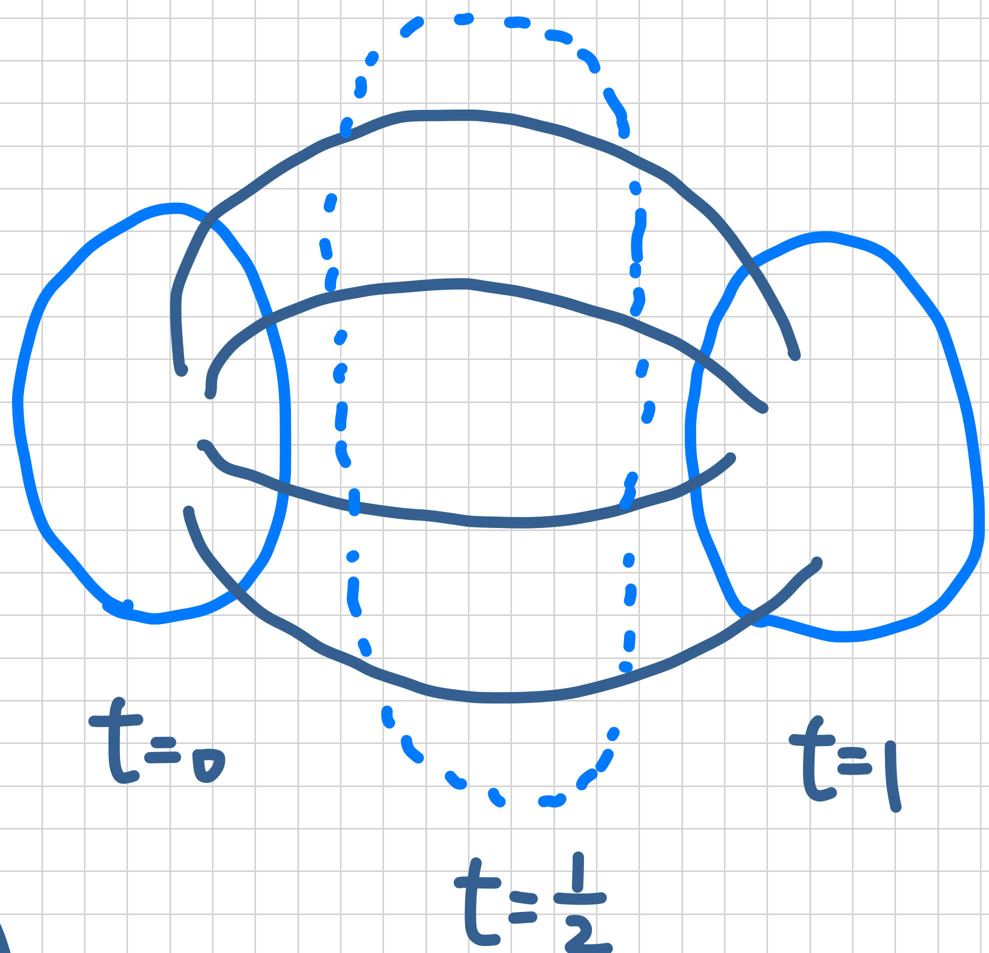
Motivation: Lazy Gas Experiment

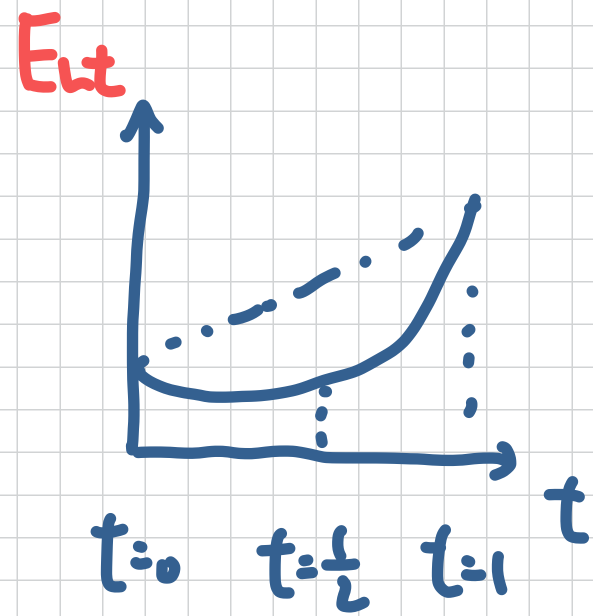




Convex

Motivation: Lazy Gas Experiment



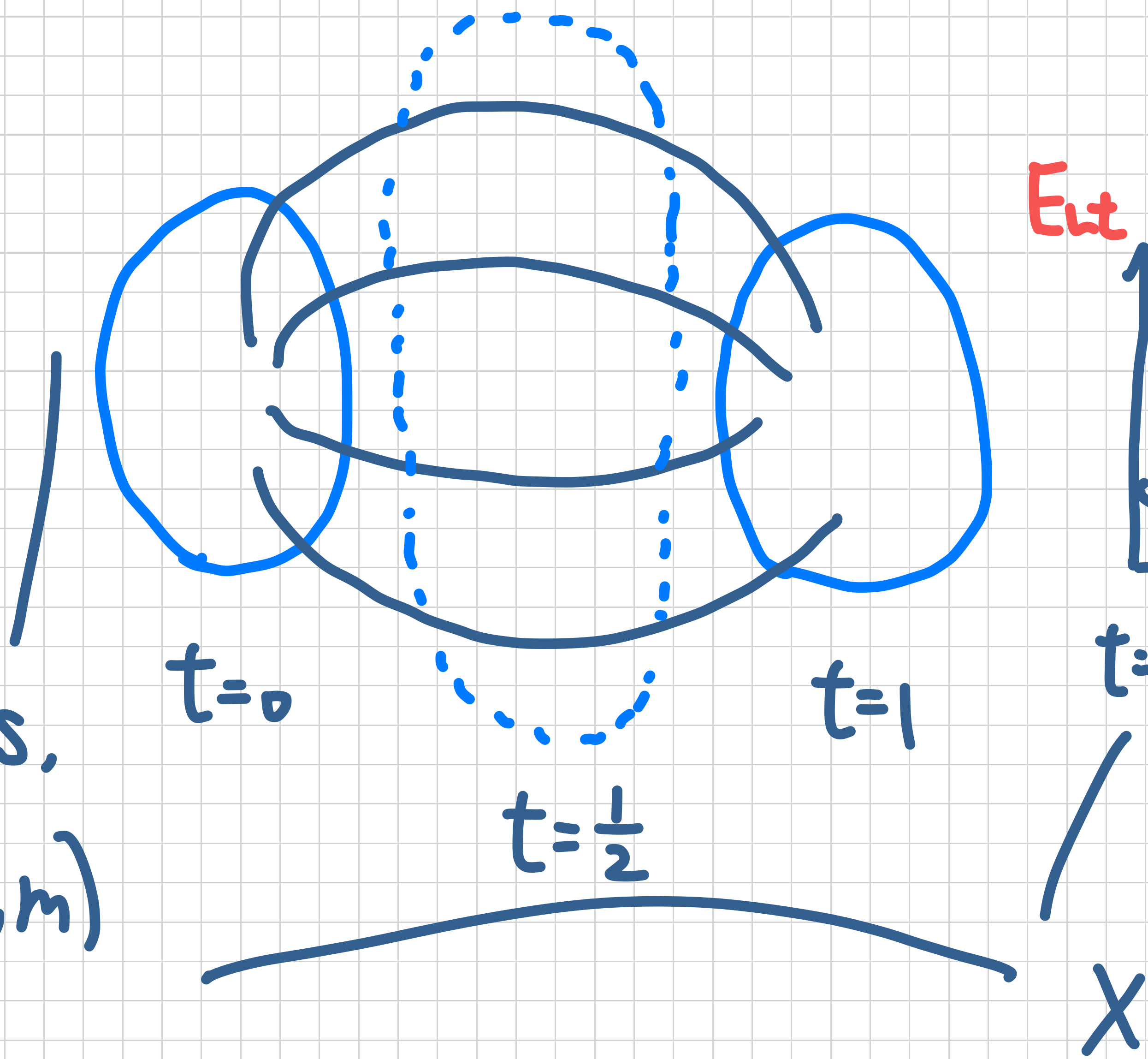


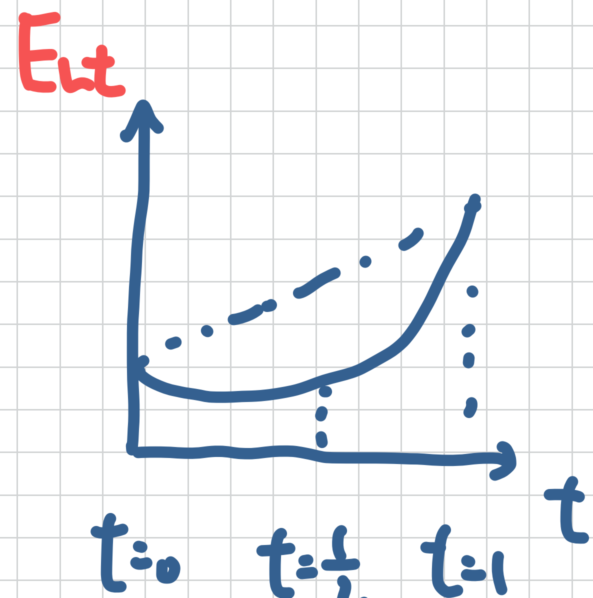
Convex

m.m.s, (X,d,m)

Motivation: Lazy Gas Experiment

m.m.s,
(X, d, m)

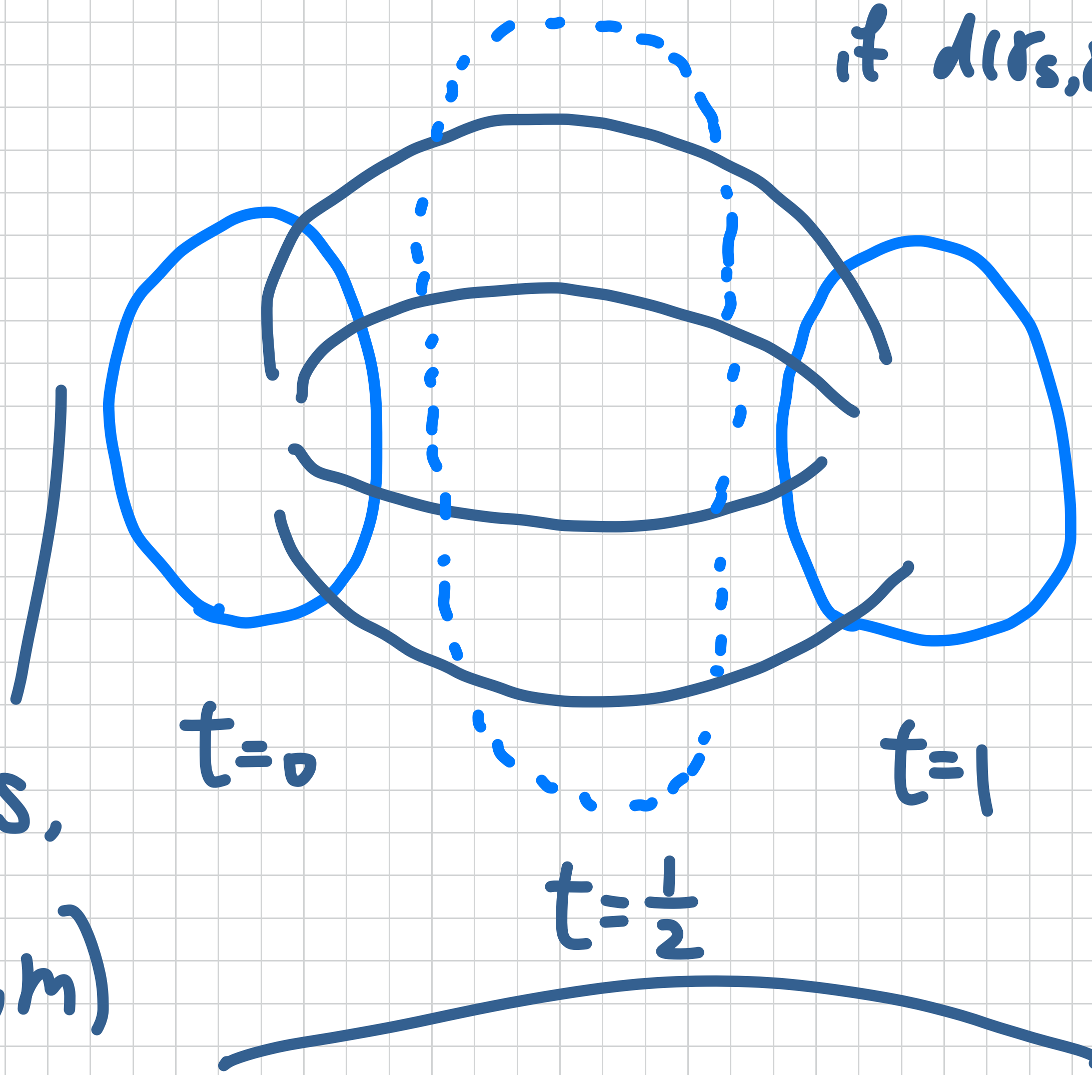


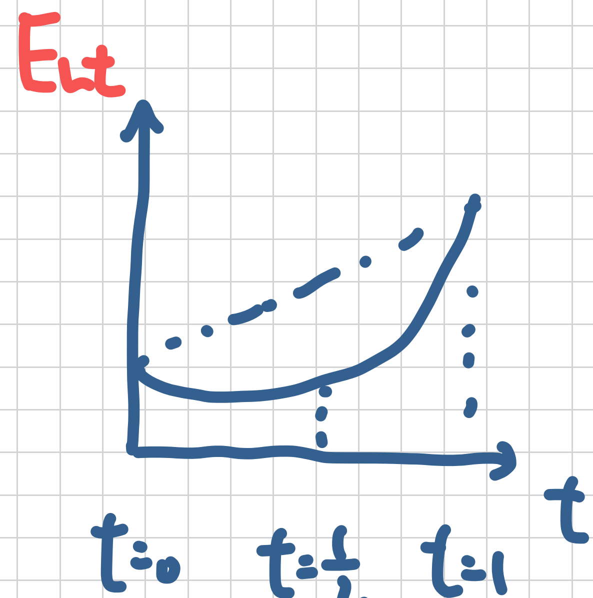


Convex

Motivation: Lazy Gas Experiment

$\gamma:[0,1]\to X$ is a geodesic X
if $d(\gamma_s,\gamma_t)=|s-t|\,d(\gamma_0,\gamma_1)\ \forall s,t$





m.m.s,
(X,d,m)

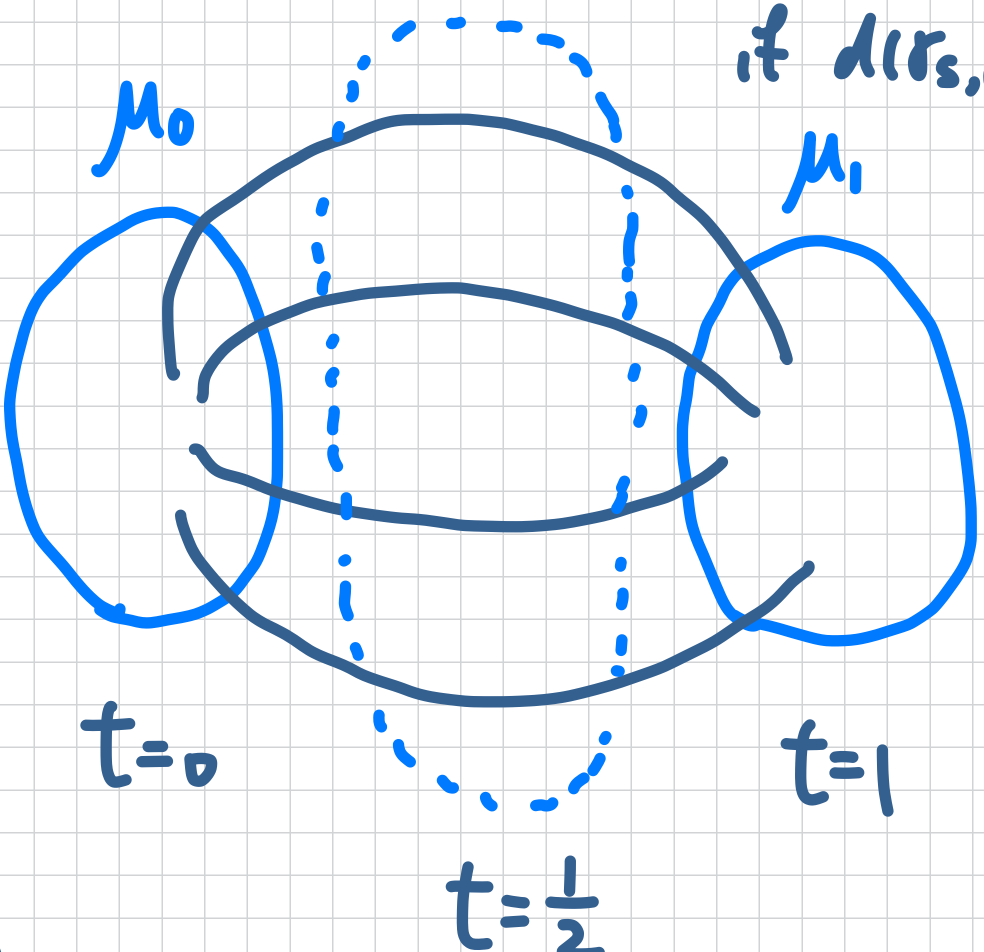
Convex

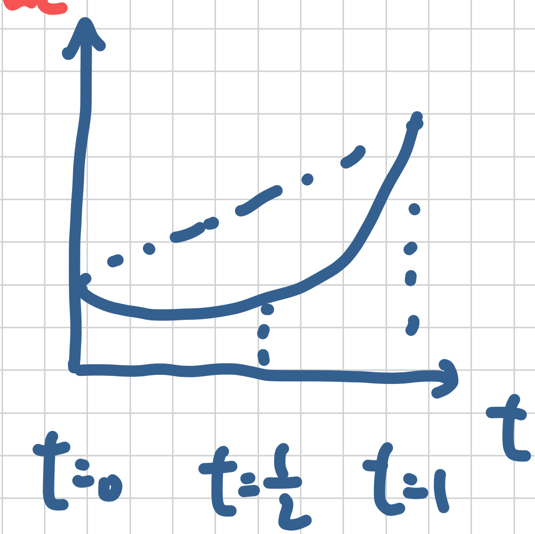
X

Motivation: Lazy Gas Experiment

$\gamma:[0,1]\to X$ is a geodesic X
if $d(\gamma_s,\gamma_t)=|s-t|\,d(\gamma_0,\gamma_1)\ \forall s,t$

$P_2(X)$





m.m.s,
(X,d,m)

Convex

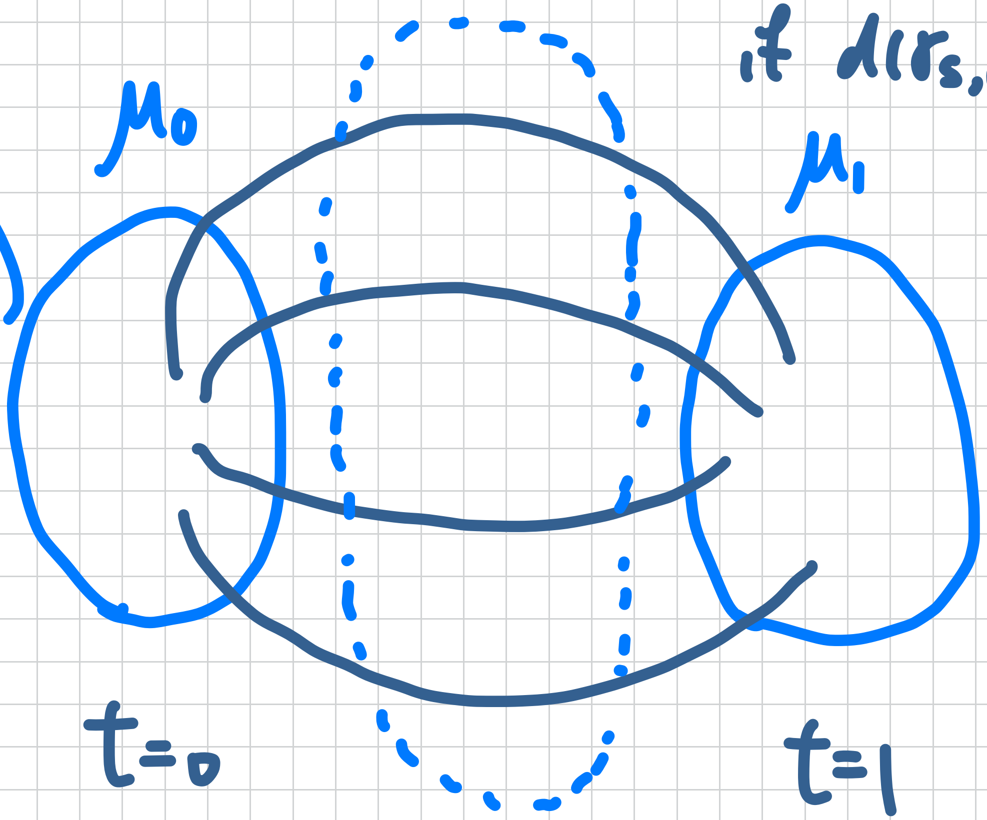
X

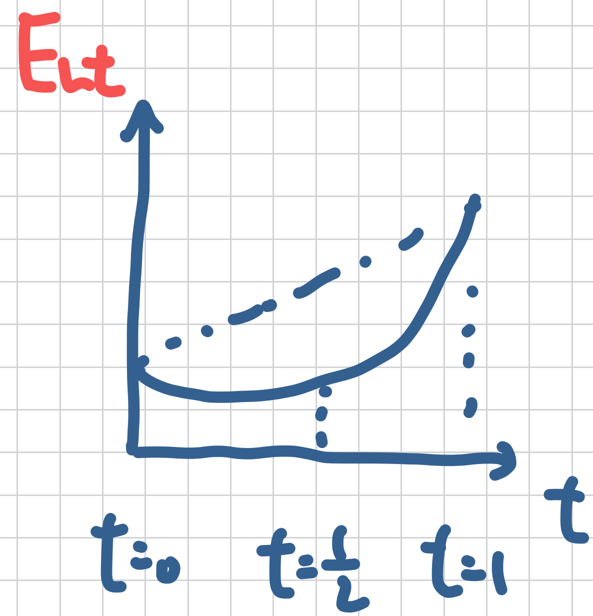
Motivation: Lazy Gas Experiment

$\gamma : [0,1] \to X$ is a geodesic X
if $d(\gamma_s, \gamma_t) = |s-t|\, d(\gamma_0, \gamma_1)$ $\forall s,t$

$(\mathcal{P}_2(X), d_W)$

$c = \dfrac{d_g^2}{2}$

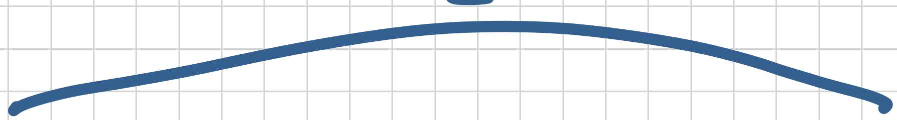


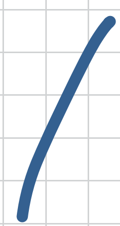


m.m.s,

(X, d, m)

$t = \frac{1}{2}$





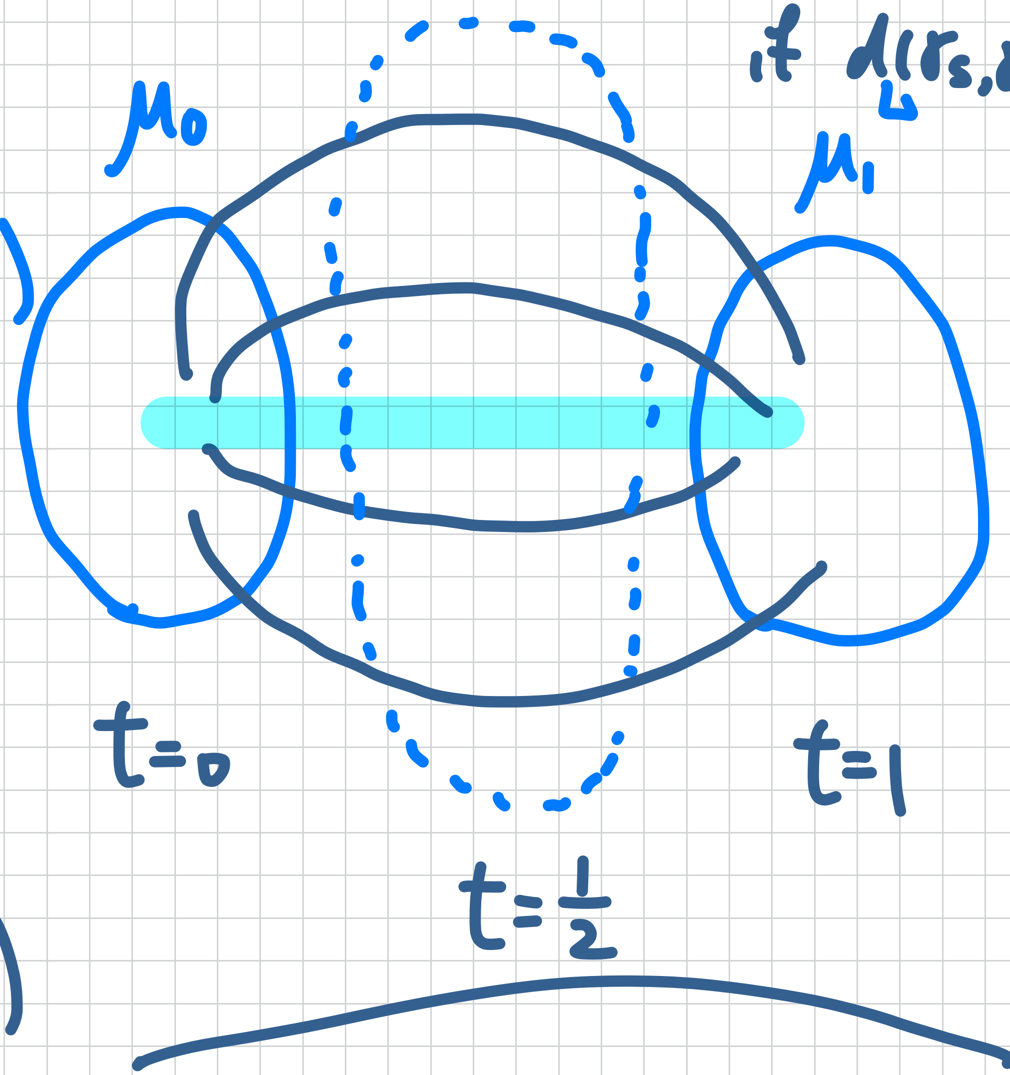
Convex

X

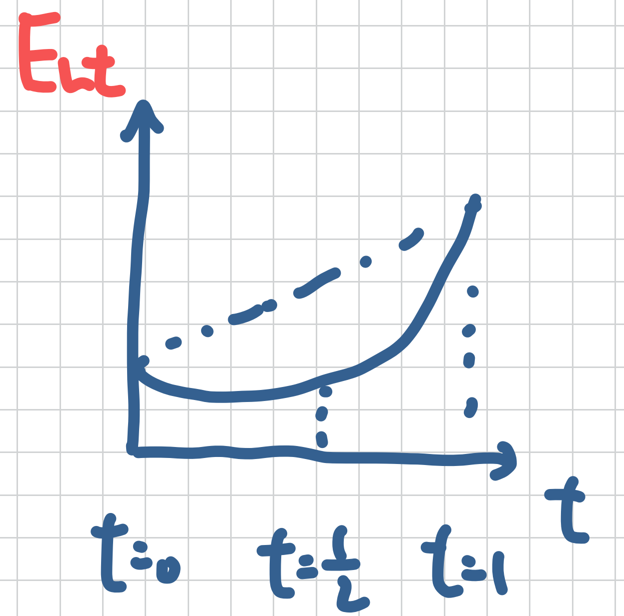
Motivation: Lazy Gas Experiment

$\gamma:[0,1]\to P_2(X)$ is a geodesic $P_2(X)$
if $d_W(\gamma_s,\gamma_t)=|s-t|\, d_W(\gamma_0,\gamma_1)$ $\forall s,t$

$(P_2(X), d_W)$



m.m.s.
(X,d,m)

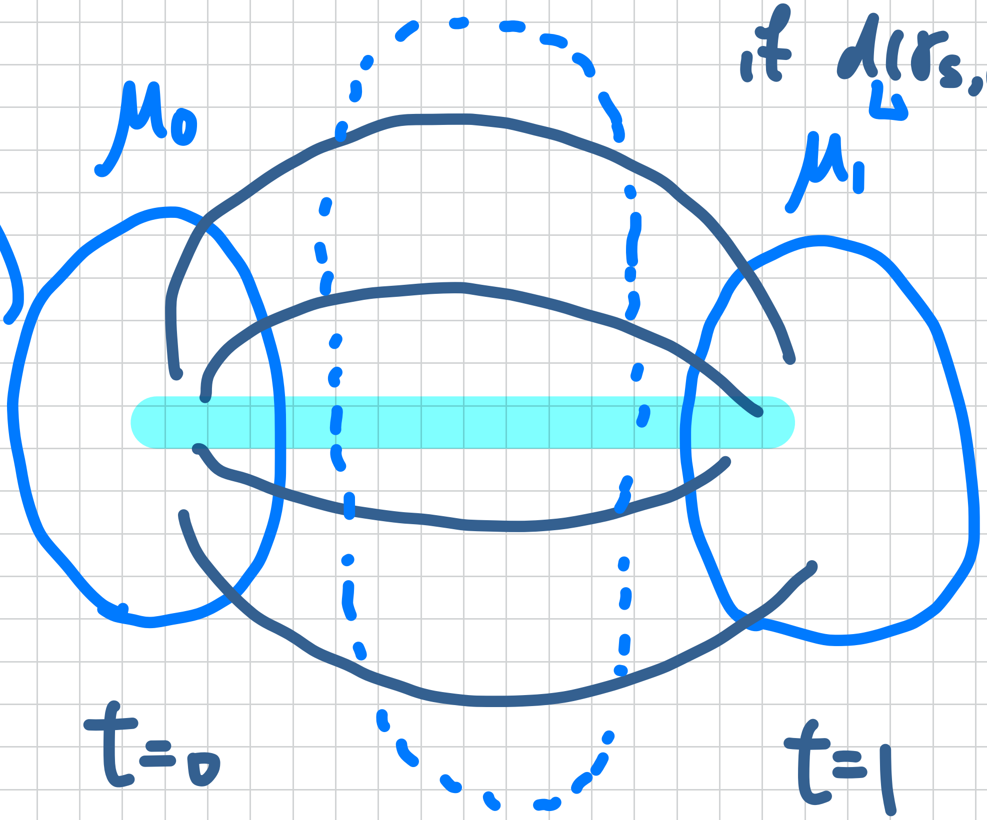


Convex

Motivation: Lazy Gas Experiment

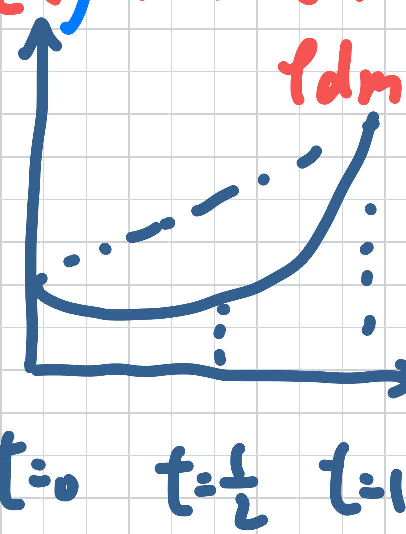
$\gamma:[0,1]\to P_2(X)$ is a geodesic $P_2(X)$ if $d_W(\gamma_s,\gamma_t)=|s-t|\,d_W(\gamma_0,\gamma_1)\ \forall s,t$

$(P_2(X), d_W)$



$\mathrm{Ent}(\mu|m)=\int \rho\log\rho\,dm$

$\rho\,dm=\mu$



m.m.s.

(X,d,m)

X

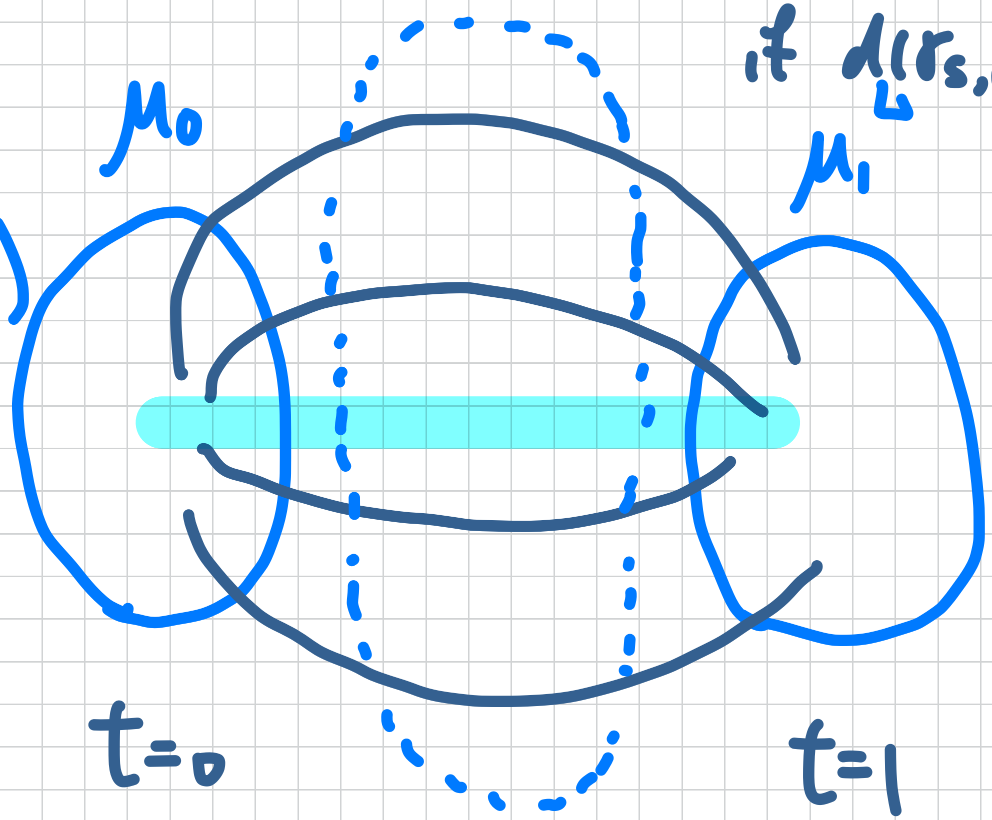
Convex

Motivation: Lazy Gas Experiment

$\gamma:[0,1]\to P_2(X)$ is a geodesic $P_2(X)$

if $d_W(\gamma_s,\gamma_t)=|s-t|\,d_W(\gamma_0,\gamma_1)$ $\forall s,t$

$(P_2(X), d_W)$



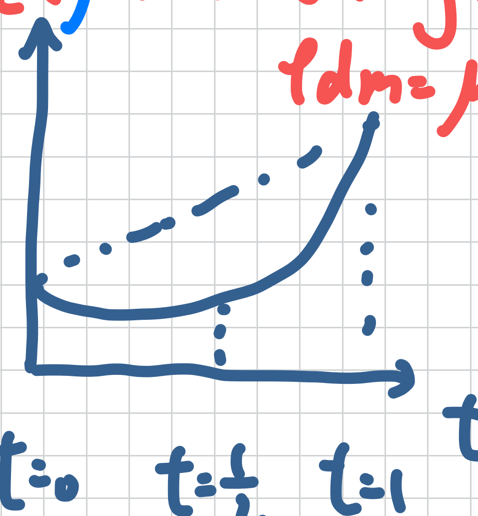
$t=\frac{1}{2}$

m.m.s,

(X,d,m)

$\mathrm{Ent}(\mu|m)=\int \rho\log\rho\,dm$

$\rho\,dm=\mu$



u is K-N convex if

X $\quad u''\ge K+\frac{1}{N}(u')^2$

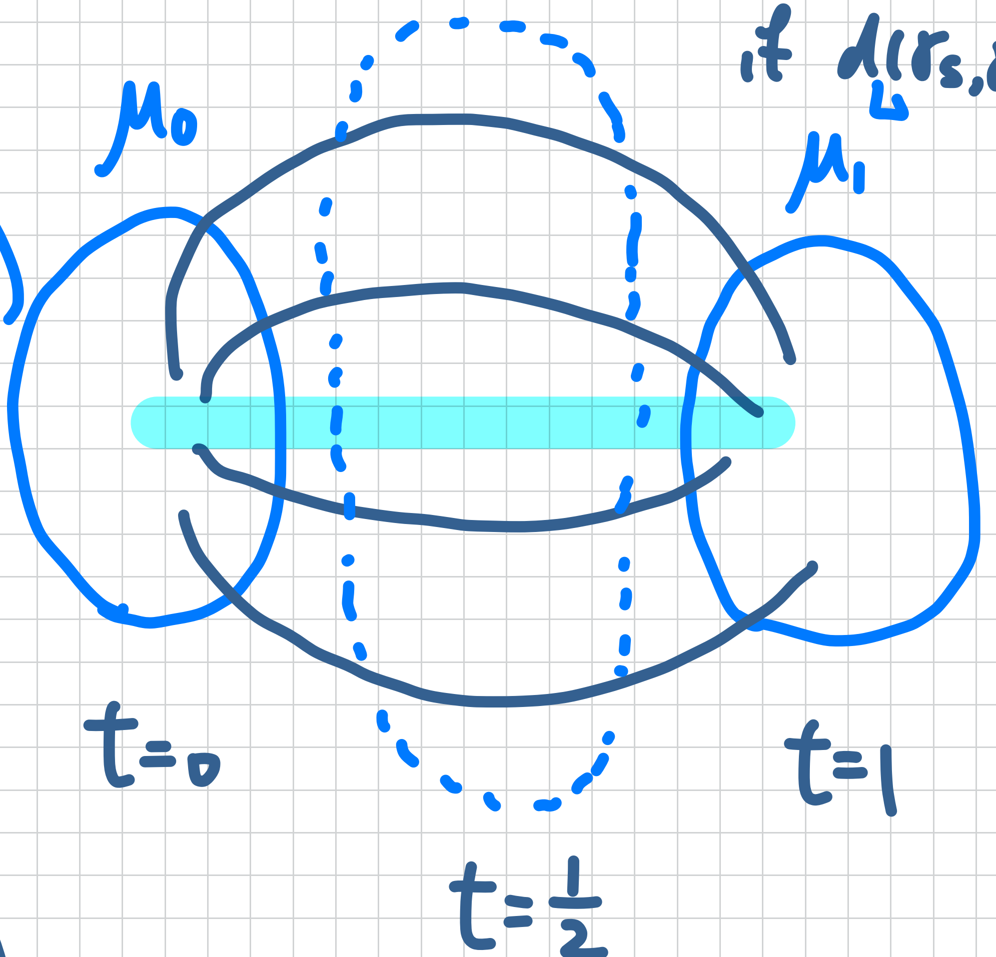
Similarly, we can define

$CD(K,N)$, $CD^*(K,N)$, $CD_1(K,N)$, ...

$\gamma : [0,1] \to P_2(X)$ is a geodesic $P_2(X)$

if $d_W(\gamma_s, \gamma_t) = |s-t| \, d_W(\gamma_0, \gamma_1)$ $\forall s,t$

$(P_2(X), d_W)$

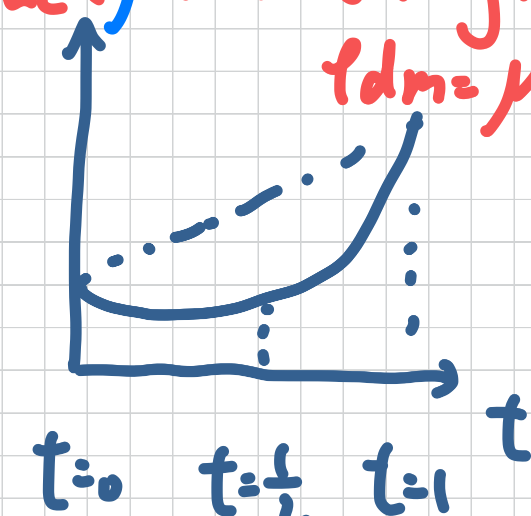


m.m.s,

(X, d, m)

$Ent(\mu | m) = \int \rho \log \rho \, dm$

$\rho \, dm = \mu$



u is K-N convex if

$u'' \geq K + \frac{1}{N}(u')^2$

Similarly, we can define

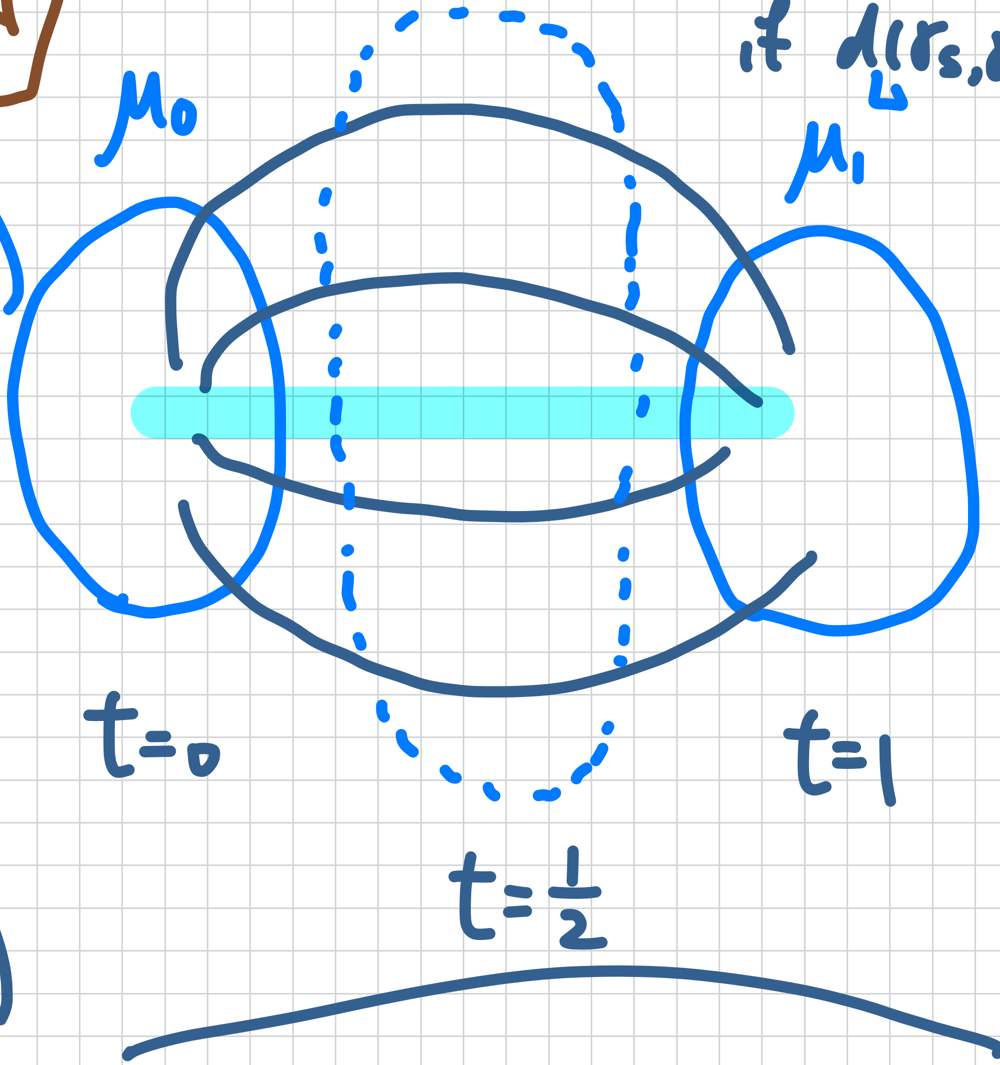
$$\boxed{CD(K,N) = CD^*(K,N) = CD_1(K,N) = \dots = CD^e(K,N)}$$

ENB

$\gamma : [0,1] \to P_2(X)$ is a geodesic $P_2(X)$

if $d_W(\gamma_s, \gamma_t) = |s-t|\, d_W(\gamma_0, \gamma_1)$ $\forall s,t$

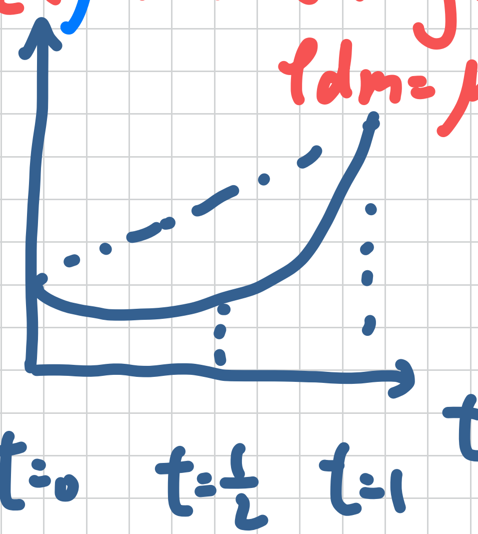
$(P_2(X), d_W)$



m.m.s. (X, d, m)

$Ent(\mu | m) = \int \rho \log \rho \, dm$

$\rho \, dm = \mu$



u is K-N convex if

$u'' \geq K + \frac{1}{N}(u')^2$

ENB

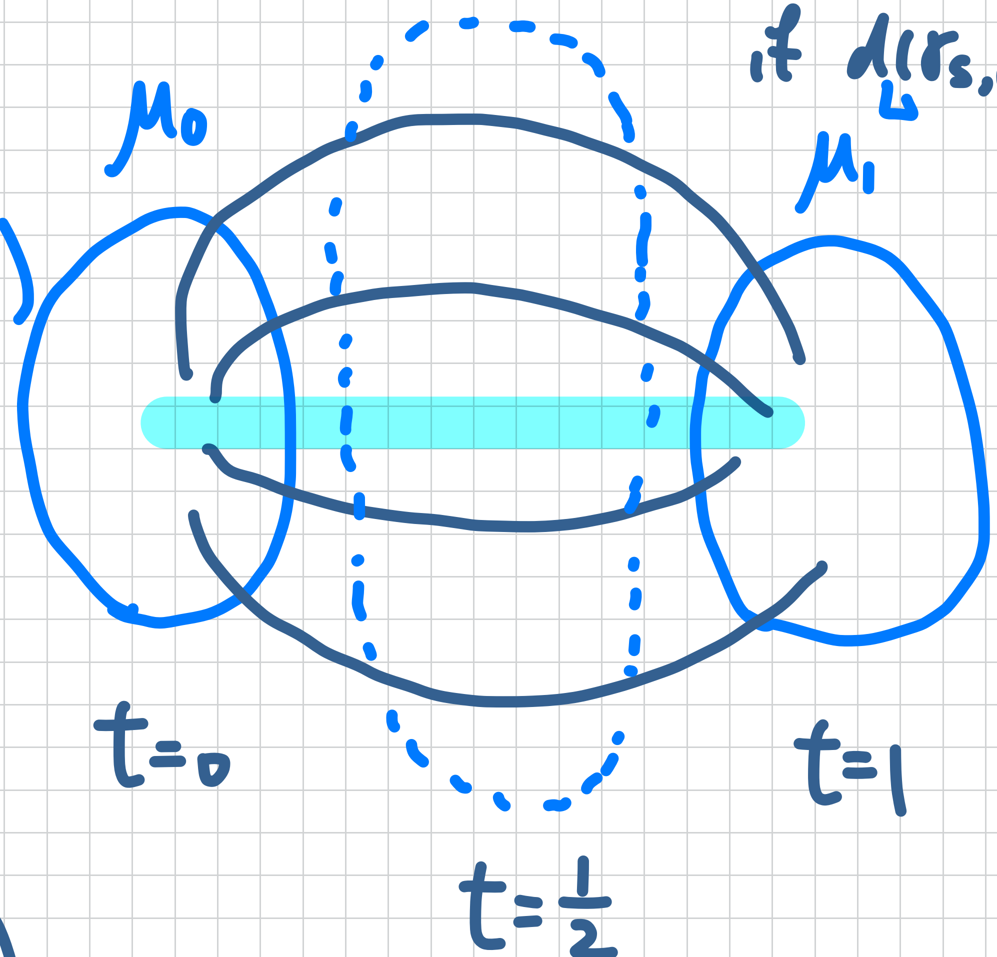
On smooth Riemannian mfd

$\mathrm{Ric} \geq 0$ ·

$(P_2(X), d_W)$

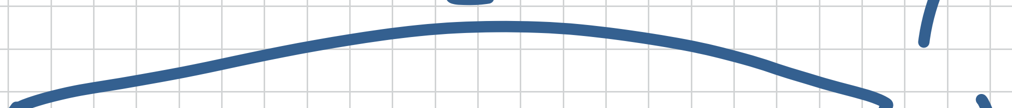
$\gamma : [0,1] \to P_2(X)$ is a geodesic $P_2(X)$

if $d_W(\gamma_s, \gamma_t) = |s-t| \, d_W(\gamma_0, \gamma_1)$ $\forall s,t$

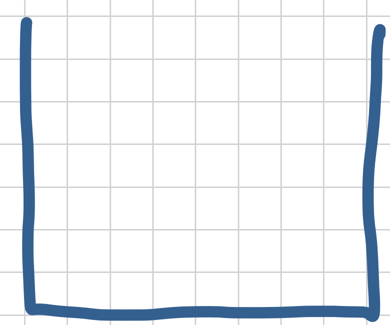


m.m.s.

(X, d, m)



X



$\mathrm{Geo}(X)$

ENB

On smooth Riemannian mfd

$Ric \geq 0.$

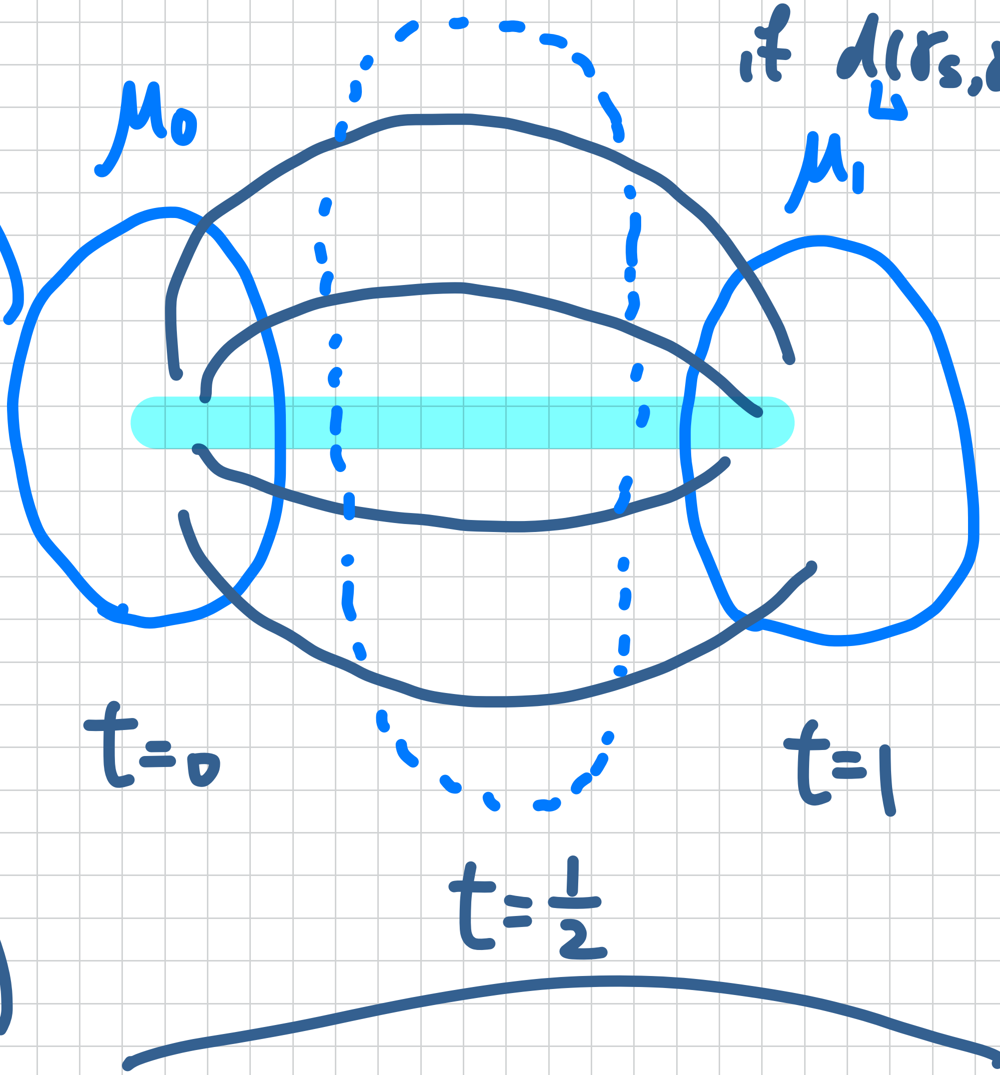
$(\mathcal{P}_2(X), d_W)$

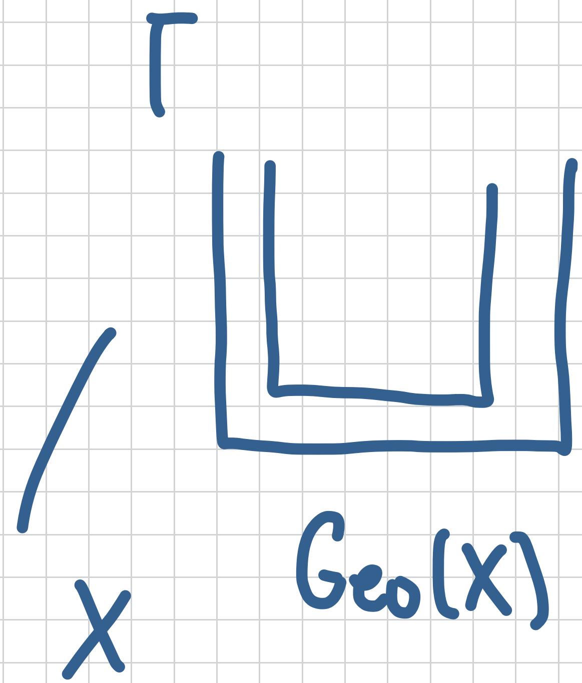
m.m.s.

(X, d, m)

$\gamma : [0,1] \to \mathcal{P}_2(X)$ is a geodesic $\mathcal{P}_2(X)$

if $d_W(\gamma_s, \gamma_t) = |s-t| \, d_W(\gamma_0, \gamma_1)$ $\forall s,t$





ENB

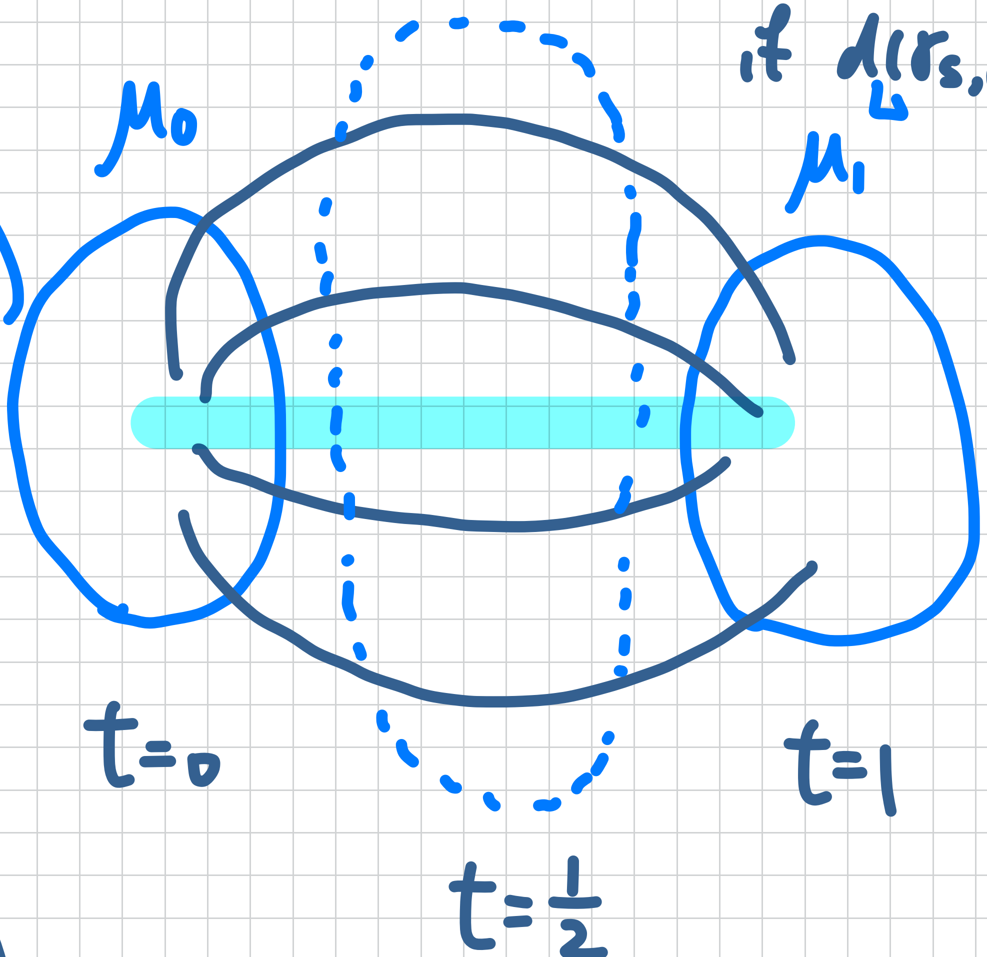
On smooth Riemannian mfd

$Ric \geq 0$ ·

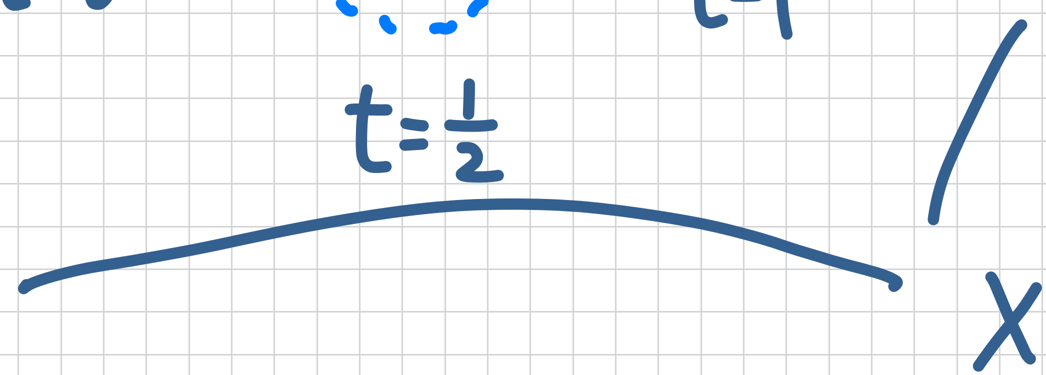
$(\mathcal{P}_2(X), d_W)$

$\gamma : [0,1] \to \mathcal{P}_2(X)$ is a geodesic $\mathcal{P}_2(X)$

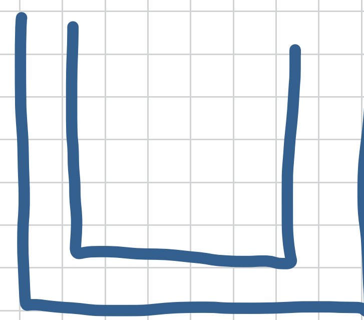
if $d_W(\gamma_s, \gamma_t) = |s-t|\, d_W(\gamma_0, \gamma_1)$ $\forall s,t$



m.m.s, (X, d, m)



Γ is non branching



Geo(X)

ENB

On smooth Riemannian mfd

Ric ≥ 0 ·

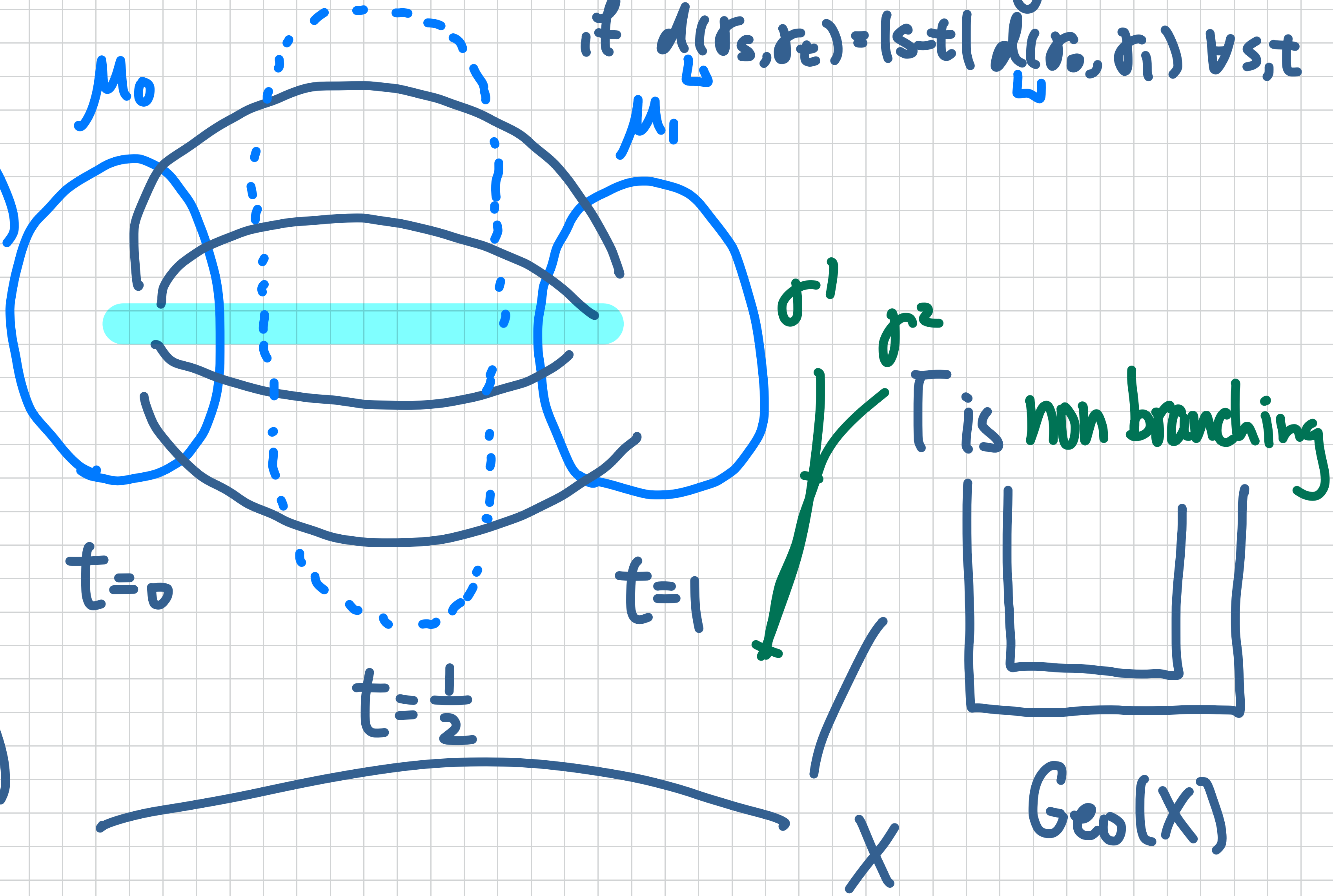
$(\mathcal{P}_2(X), d_W)$

m.m.s,

(X, d, m)

$\gamma : [0,1] \to \mathcal{P}_2(X)$ is a geodesic $\mathcal{P}_2(X)$

if $d_W(\gamma_s, \gamma_t) = |s-t| \, d_W(\gamma_0, \gamma_1) \ \forall s,t$



ENB

On smooth Riemannian mfd

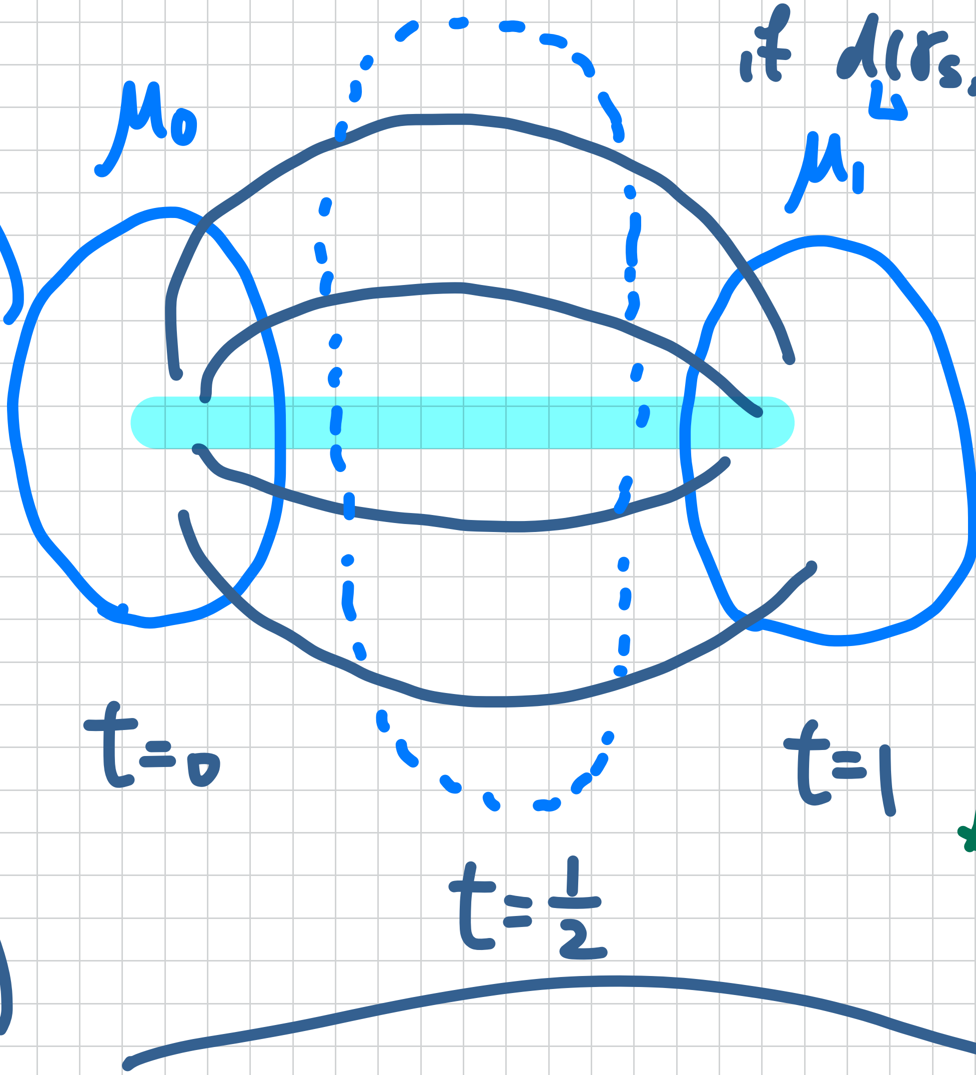
Ric ≥ 0 ·

$(\mathcal{P}_2(X), d_W)$

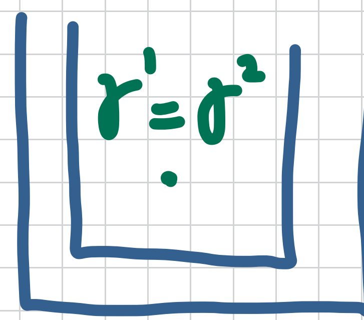
m.m.s,

(X, d, m)

$\gamma : [0,1] \to \mathcal{P}_2(X)$ is a geodesic $\mathcal{P}_2(X)$
if $d_W(\gamma_s, \gamma_t) = |s-t| \, d_W(\gamma_0, \gamma_1)$ $\forall s,t$



Γ is non branching



Geo(X)

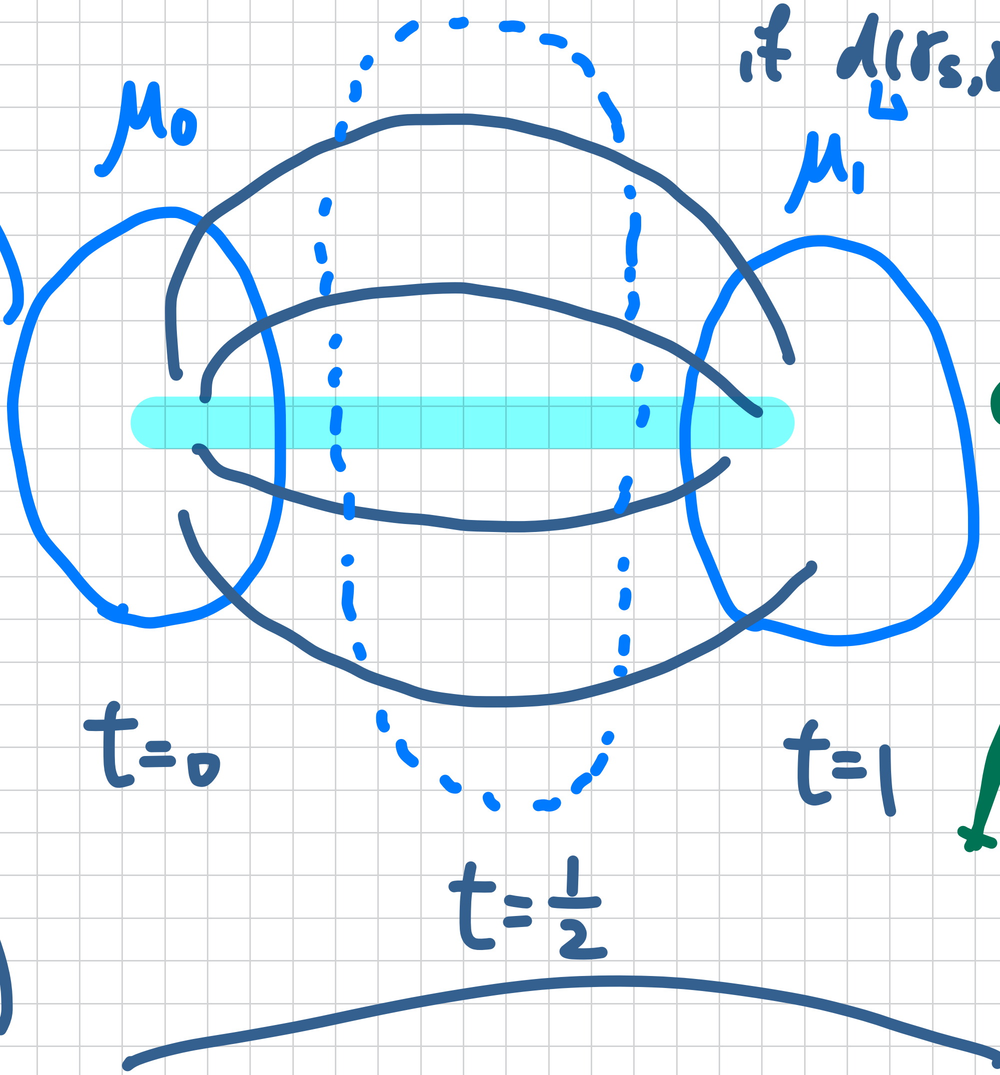
ENB

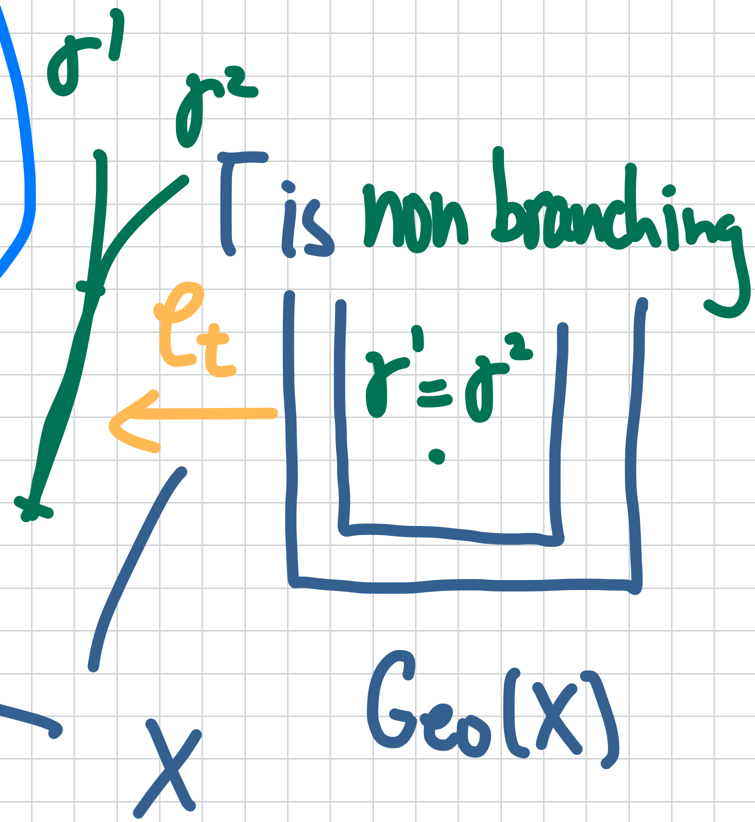
On smooth Riemannian mfd $\mathrm{Ric} \geq 0$.

$(\mathcal{P}_2(X), d_W)$

m.m.s, (X, d, m)

$\gamma : [0,1] \to \mathcal{P}_2(X)$ is a geodesic $\mathcal{P}_2(X)$ if $d_W(\gamma_s, \gamma_t) = |s-t|\, d_W(\gamma_0, \gamma_1)$ $\forall s,t$





ENB

On smooth Riemannian mfd

$Ric \geq 0$.

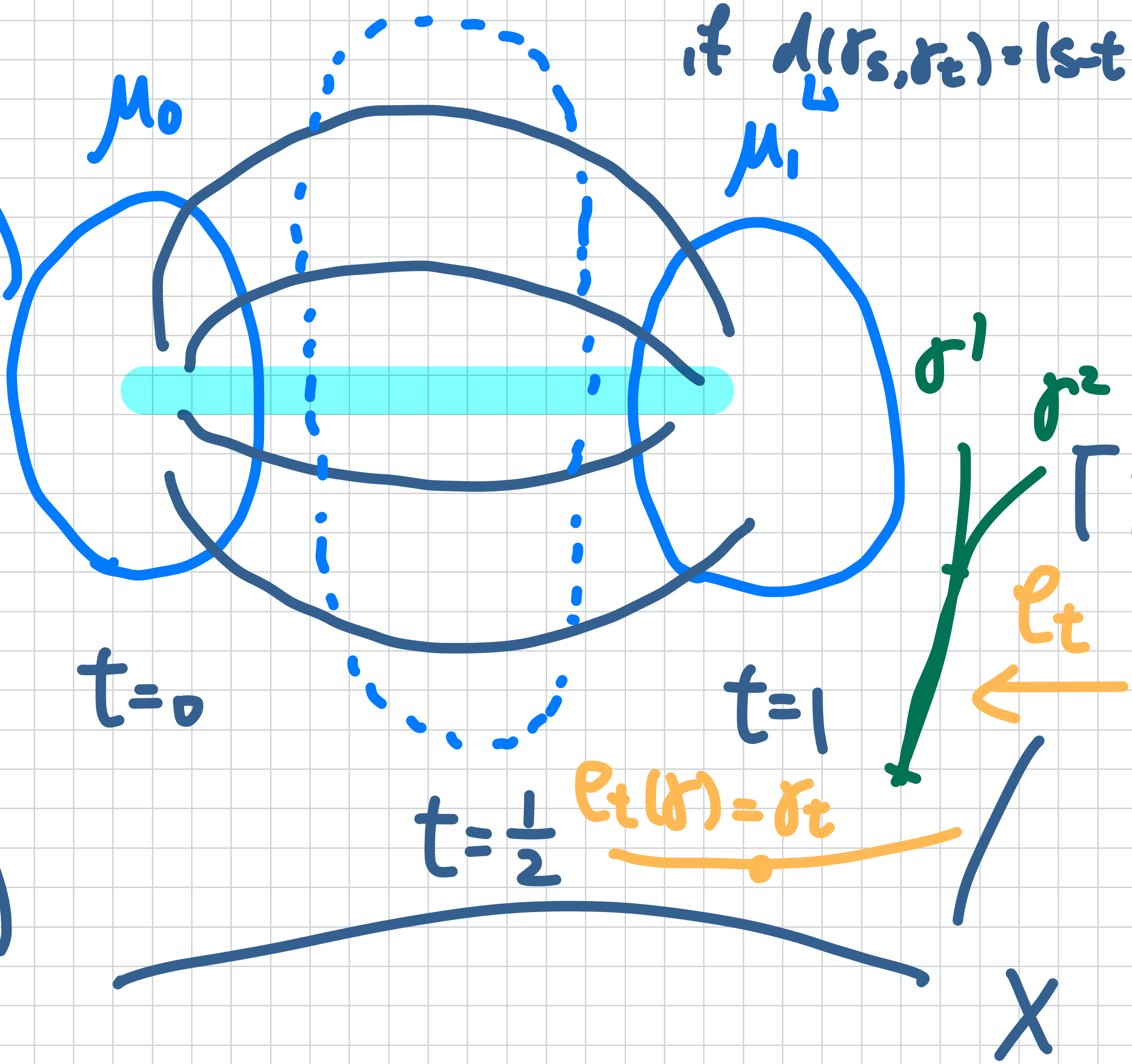
$(P_2(X), d_W)$

m.m.s.

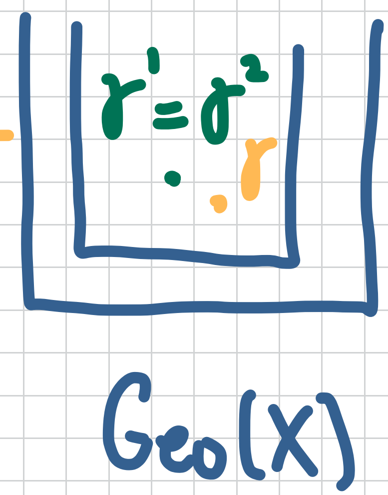
(X, d, m)

$\gamma : [0,1] \to P_2(X)$ is a geodesic $P_2(X)$

if $d_W(\gamma_s, \gamma_t) = |s-t| \, d_W(\gamma_0, \gamma_1) \ \forall s,t$



Γ is non branching



ENB

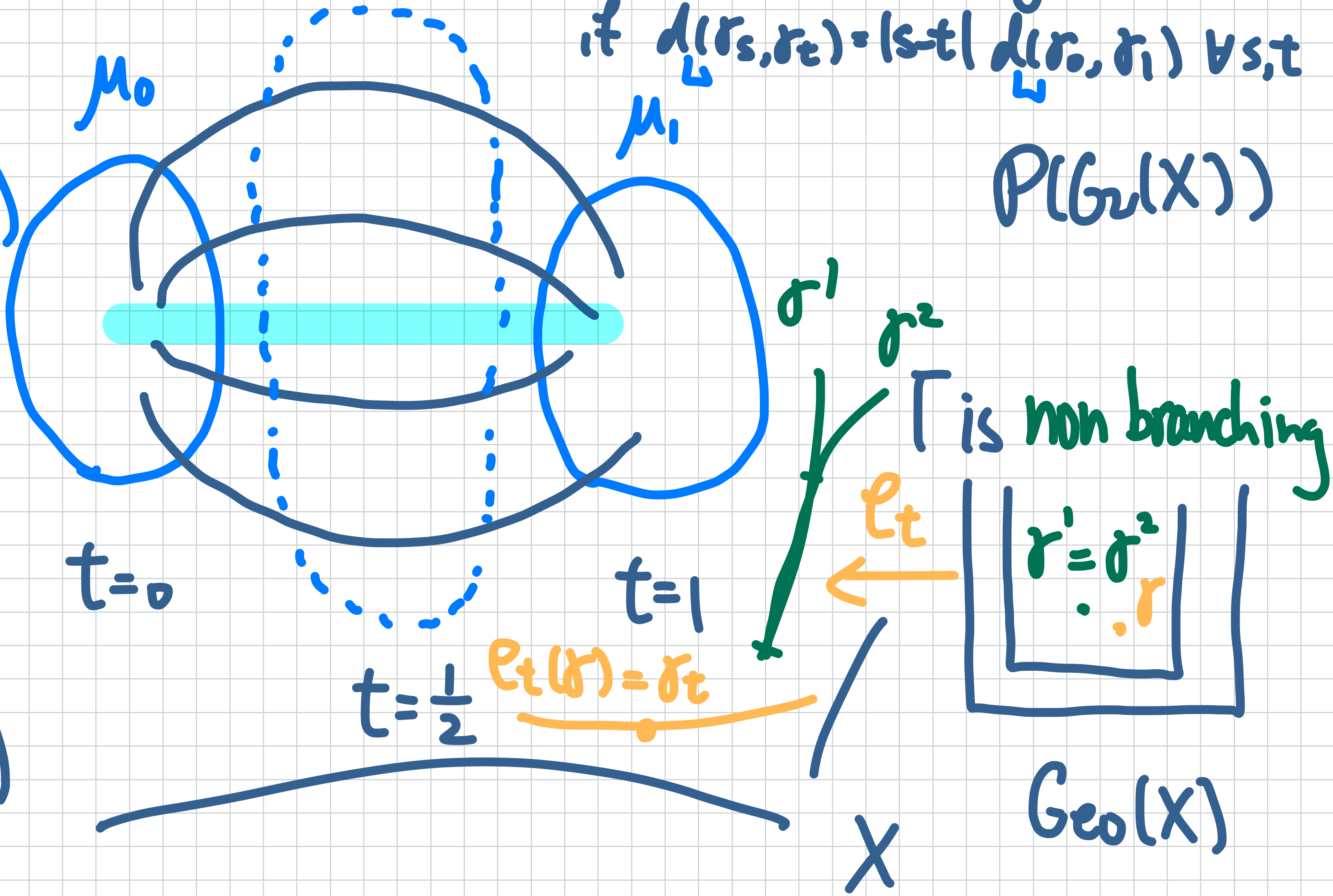
On smooth Riemannian mfd

$Ric \geq 0$.

$(\mathcal{P}_2(X), d_W)$

$\gamma : [0,1] \to \mathcal{P}_2(X)$ is a geodesic $\mathcal{P}_2(X)$

if $d_W(\gamma_s, \gamma_t) = |s-t| \, d_W(\gamma_0, \gamma_1)$ $\forall s,t$



m.m.s.

(X, d, m)

ENB

On smooth Riemannian mfd

$Ric \geq 0$.

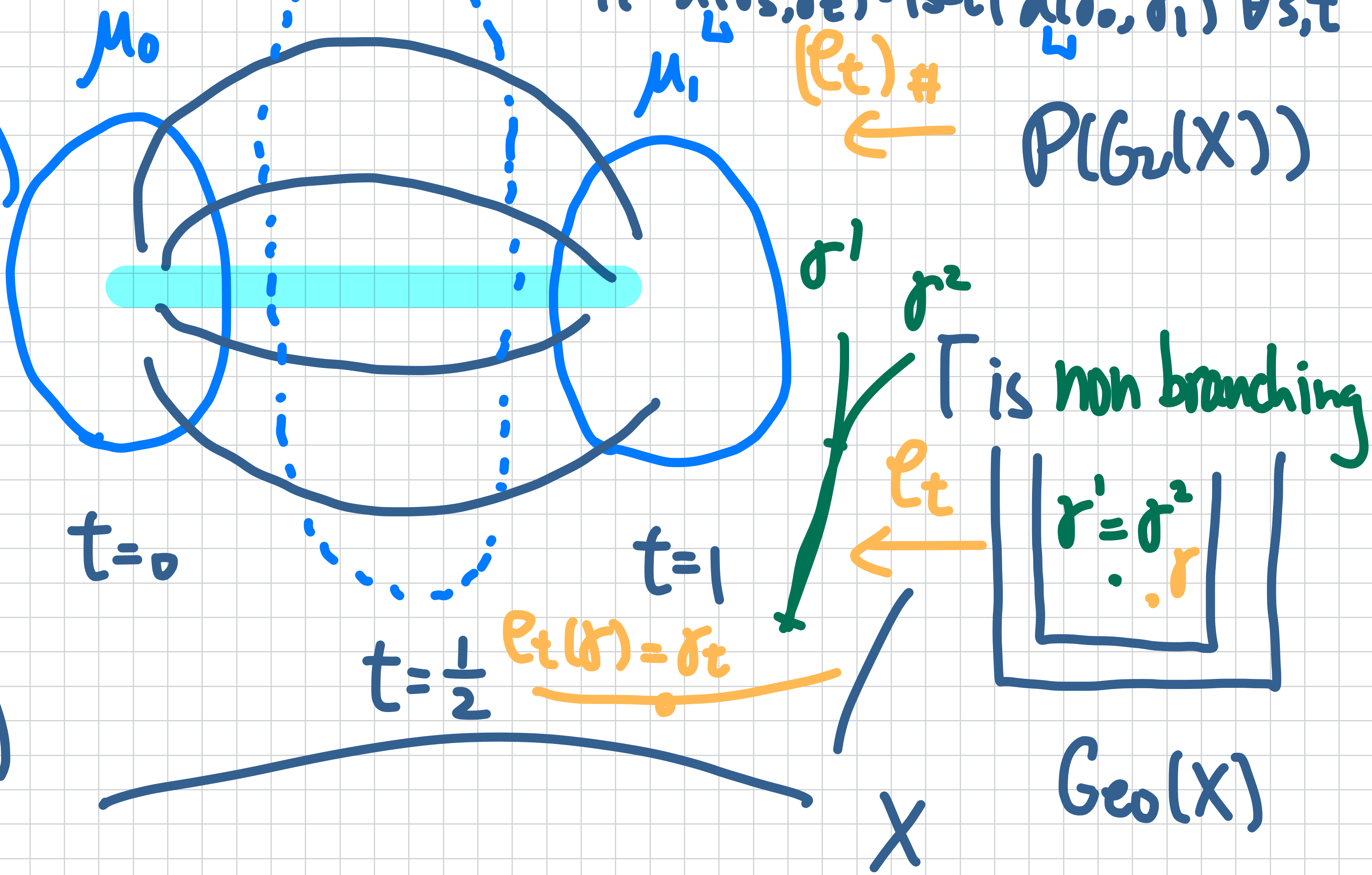
$(\mathcal{P}_2(X), d_W)$

m.m.s.

(X, d, m)

$\gamma : [0,1] \to \mathcal{P}_2(X)$ is a geodesic $\mathcal{P}_2(X)$

if $d_W(\gamma_s, \gamma_t) = |s-t| \, d_W(\gamma_0, \gamma_1)$ $\forall s,t$

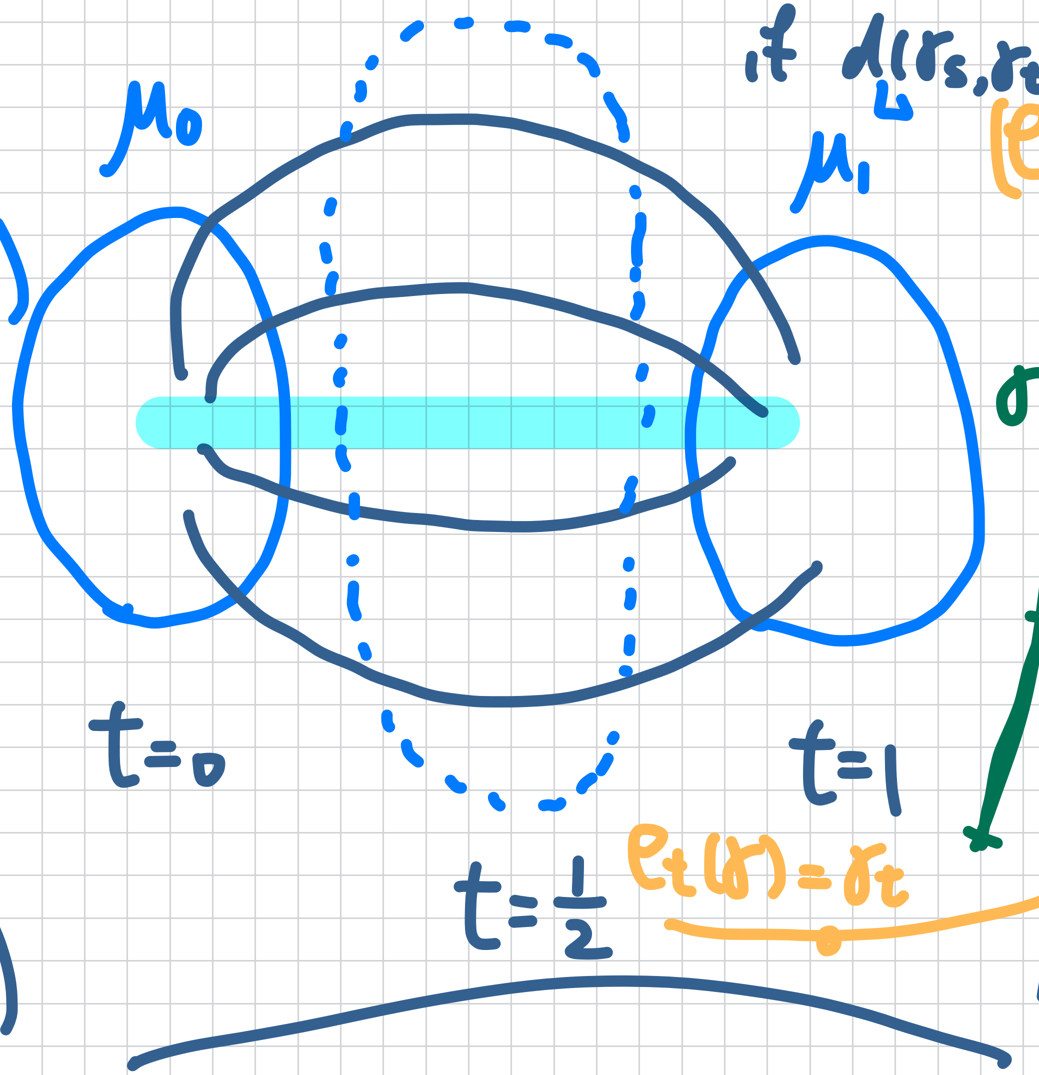


ENB

On smooth Riemannian mfd

Ric ≥ 0.

$(\mathcal{P}_2(X), d_W)$



m.m.s,

(X, d, m)

$\gamma : [0,1] \to \mathcal{P}_2(X)$ is a geodesic $\mathcal{P}_2(X)$

if $d_W(\gamma_s, \gamma_t) = |s-t| \, d_W(\gamma_0, \gamma_1) \ \forall s,t$

$\mathcal{P}(\mathrm{Geo}(X)) \xrightarrow{(e_t)_\#}$

$\gamma^1, \gamma^2 \in \mathrm{OptGeo}(\mu_0, \mu_1)$

Γ is non branching

$\xleftarrow{e_t}$

$\gamma^1 = \gamma^2$

γ

$\mathrm{Geo}(X)$

ENB

On smooth Riemannian mfd

Ric ≥ 0.

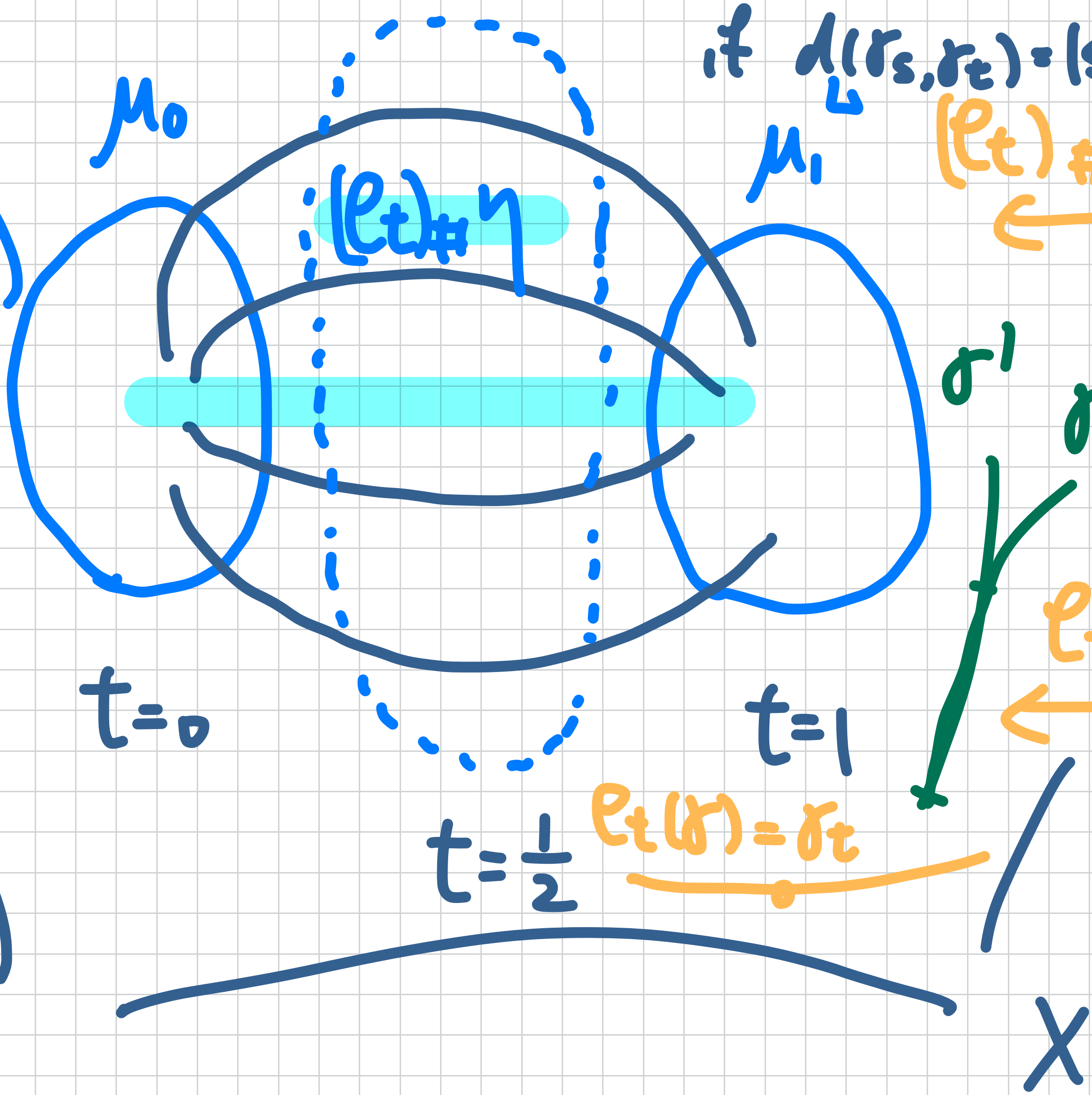
$(P_2(X), d_W)$

m.m.s.

(X, d, m)

$\gamma : [0,1] \to P_2(X)$ is a geodesic $P_2(X)$

if $d_W(\gamma_s, \gamma_t) = |s-t|\, d_W(\gamma_0, \gamma_1)$ $\forall s,t$



$(e_t)_\#$ $\leftarrow$ $P(\mathrm{Geo}(X))$

γ^1 γ^2 $\eta \in \mathrm{OptGeo}(\mu_0, \mu_1)$

Γ is non branching

e_t $\leftarrow$

$\gamma^1 = \gamma^2$ γ

Geo(X)

(X, d, m) is ENB

On smooth Riemannian mfd

$\mathrm{Ric} \geq 0$.

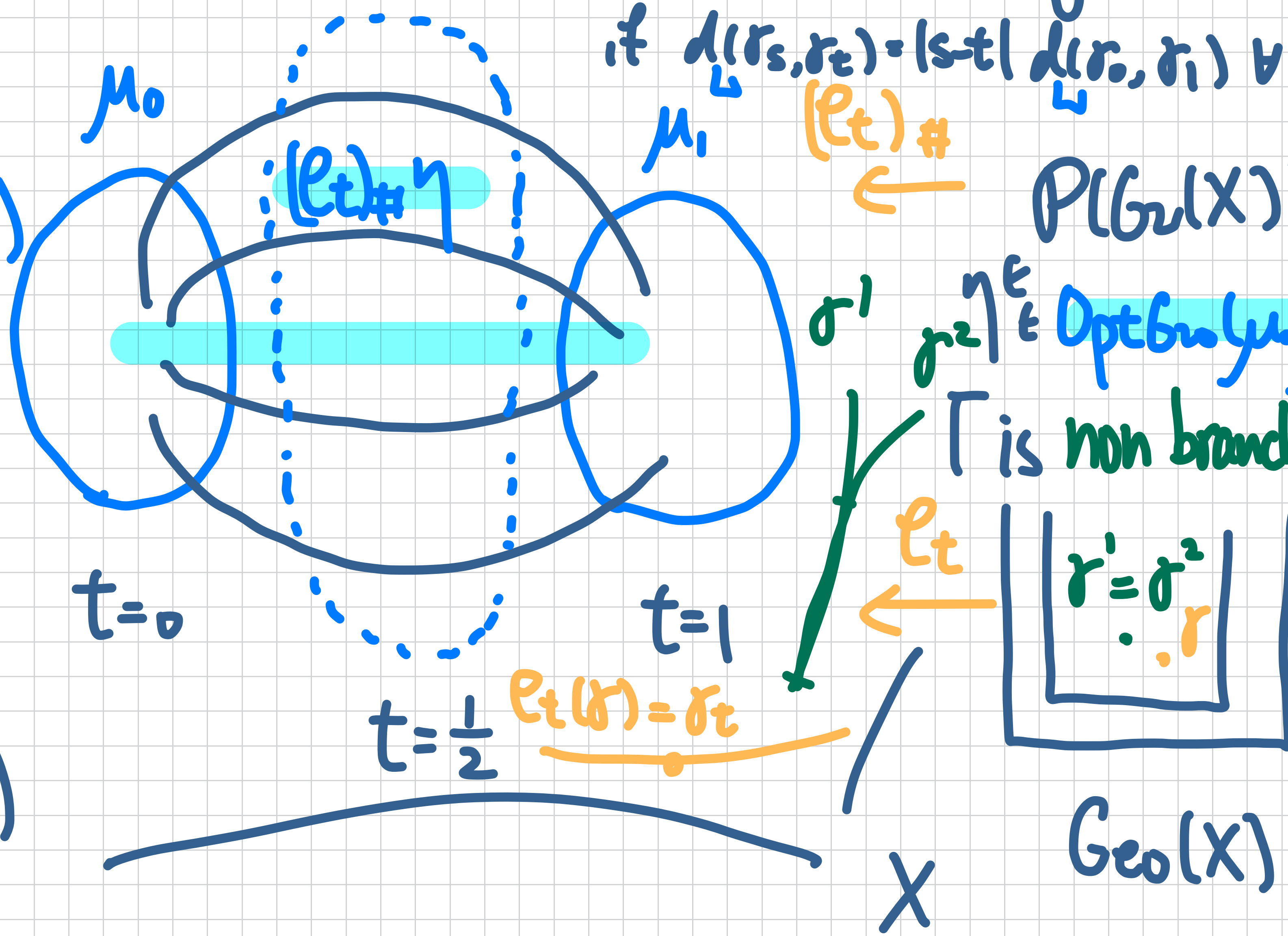
$(\mathcal{P}_2(X), d_W)$

m.m.s.

(X, d, m)

$\gamma : [0,1] \to \mathcal{P}_2(X)$ is a geodesic $\mathcal{P}_2(X)$

if $d_W(\gamma_s, \gamma_t) = |s-t|\, d_W(\gamma_0, \gamma_1)$ $\forall s,t$



(X, d, m) is ENB

On smooth Riemannian mfd

$Ric \geq 0$.

$(\mathcal{P}_2(X), d_W)$

m.m.s.

(X, d, m)

$\gamma : [0,1] \to \mathcal{P}_2(X)$ is a geodesic $\mathcal{P}_2(X)$

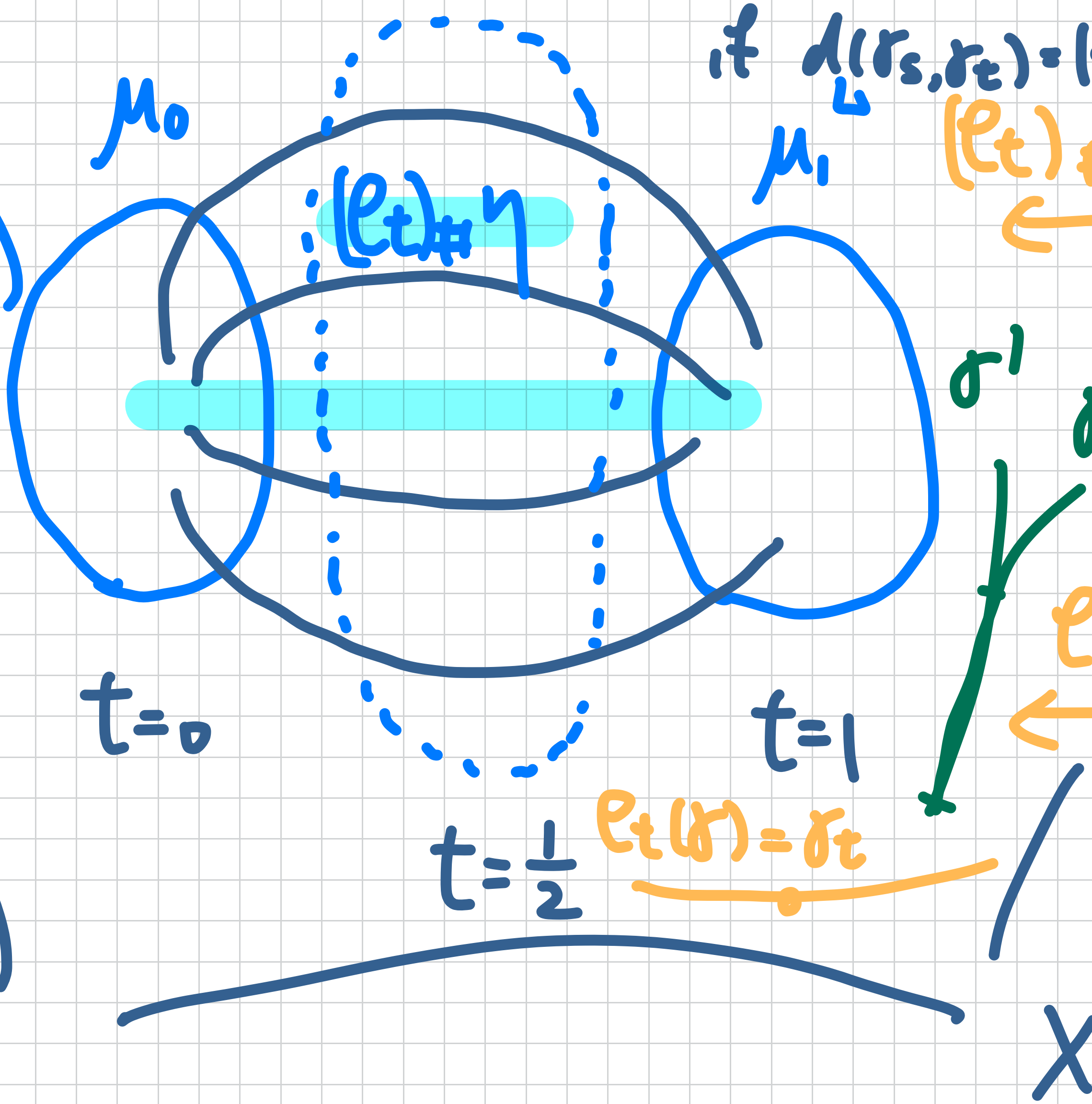
if $d_W(\gamma_s, \gamma_t) = |s-t|\, d_W(\gamma_0, \gamma_1)$ $\forall s,t$

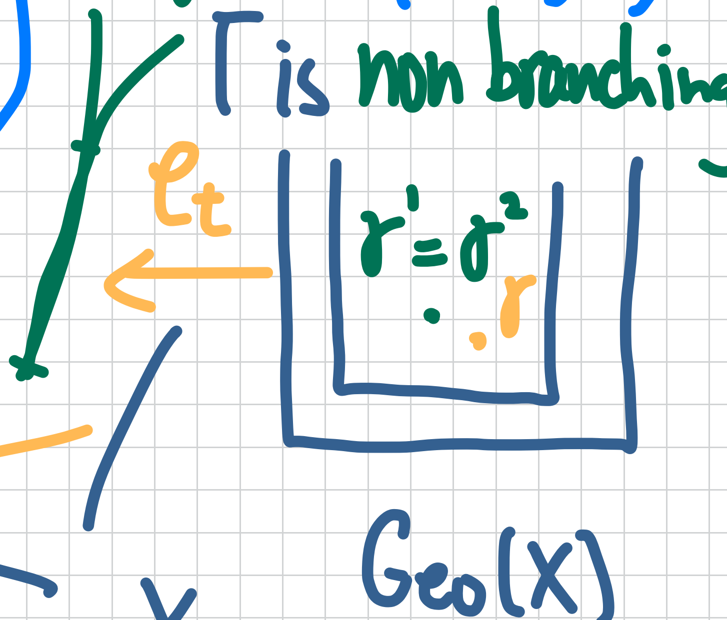
$(e_t)_\# \leftarrow \mathcal{P}(G_{eo}(X))$

$\forall$

$\gamma^1, \gamma^2 \; \eta \in \mathrm{OptGeo}(\mu_0, \mu_1)$

Γ is non branching





(X, d, m) is ENB

On smooth Riemannian mfd

$Ric \geq 0$.

$(P_2(X), d_W)$

m.m.s.

(X, d, m)

$\gamma : [0,1] \to P_2(X)$ is a geodesic $P_2(X)$

if $d_W(\gamma_s, \gamma_t) = |s-t| \, d_W(\gamma_0, \gamma_1)$ $\forall s,t$

$(e_t)_\#$ $\longleftarrow$ $P(Geo(X))$

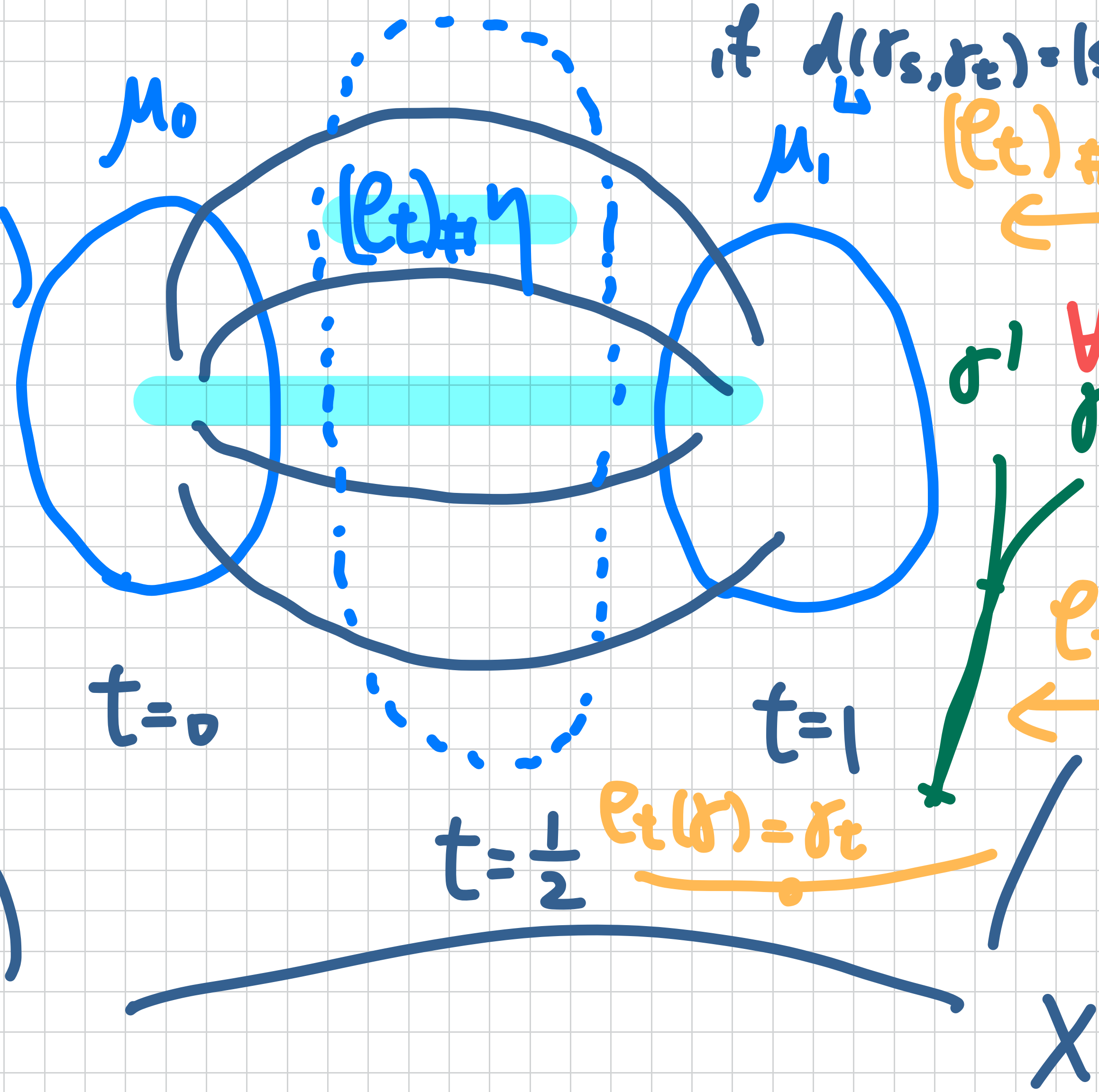
γ^1 $\forall \gamma^2$ $\forall \eta \in OptGeo(\mu_0, \mu_1)$

Γ is non branching

e_t $\longleftarrow$ $Geo(X)$

$\gamma^1 = \gamma^2$

γ



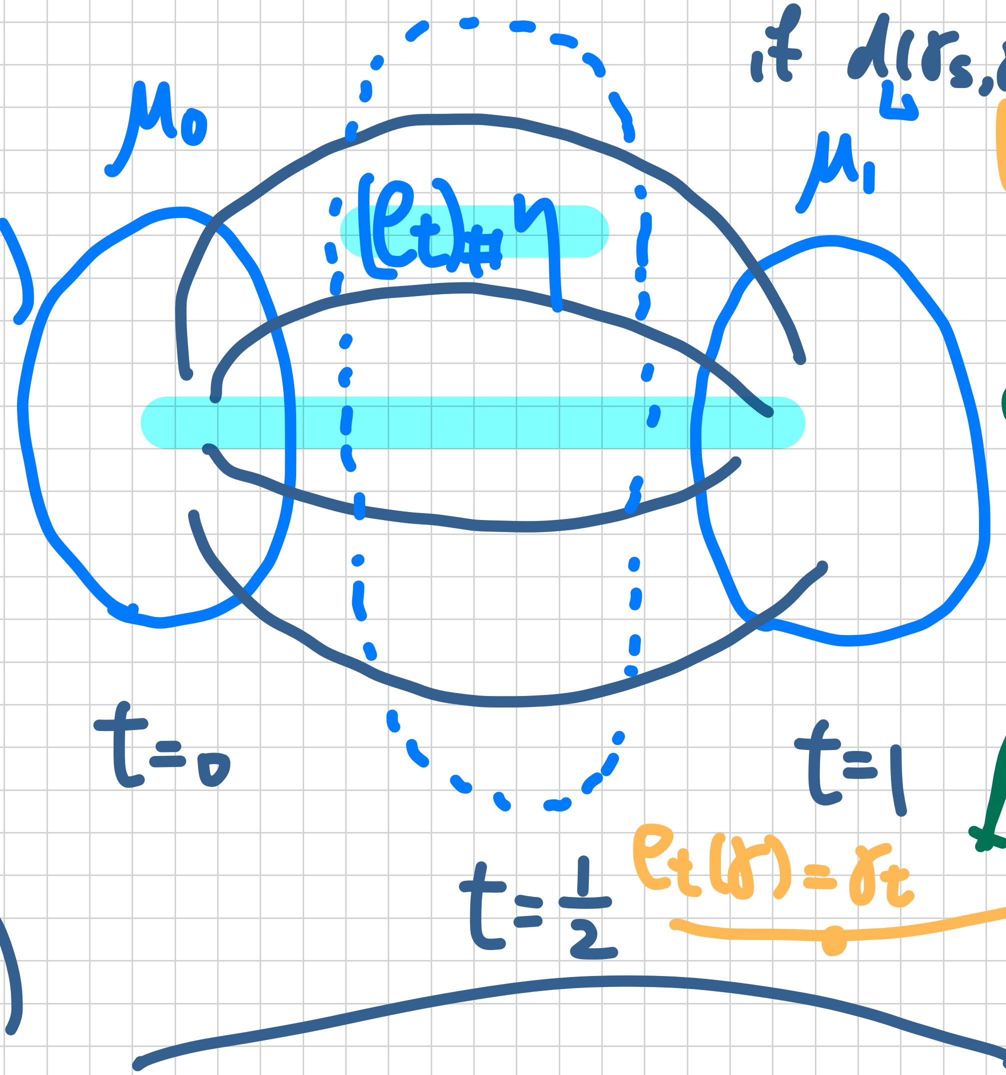
(X, d, m) is ENB

On smooth Riemannian mfd

$\mathrm{Ric} \geq 0$.

$(\mathcal{P}_2(X), d_W)$

m.m.s, (X, d, m)



$\gamma : [0,1] \to \mathcal{P}_2(X)$ is a geodesic $\mathcal{P}_2(X)$

if $d_W(\gamma_s, \gamma_t) = |s-t| \, d_W(\gamma_0, \gamma_1)$ $\forall s,t$

$(e_t)_\#$ $\leftarrow$ $\mathcal{P}(\mathrm{Geo}(X))$

γ^1 $\forall$ $\cancel{\gamma^2}$ $\eta \in \mathrm{OptGeo}(\mu_0, \mu_1)$ $\forall$

$\exists \Gamma$ is non branching

e_t $\leftarrow$ $\gamma^1 = \gamma^2$, γ

X

$\mathrm{Geo}(X)$

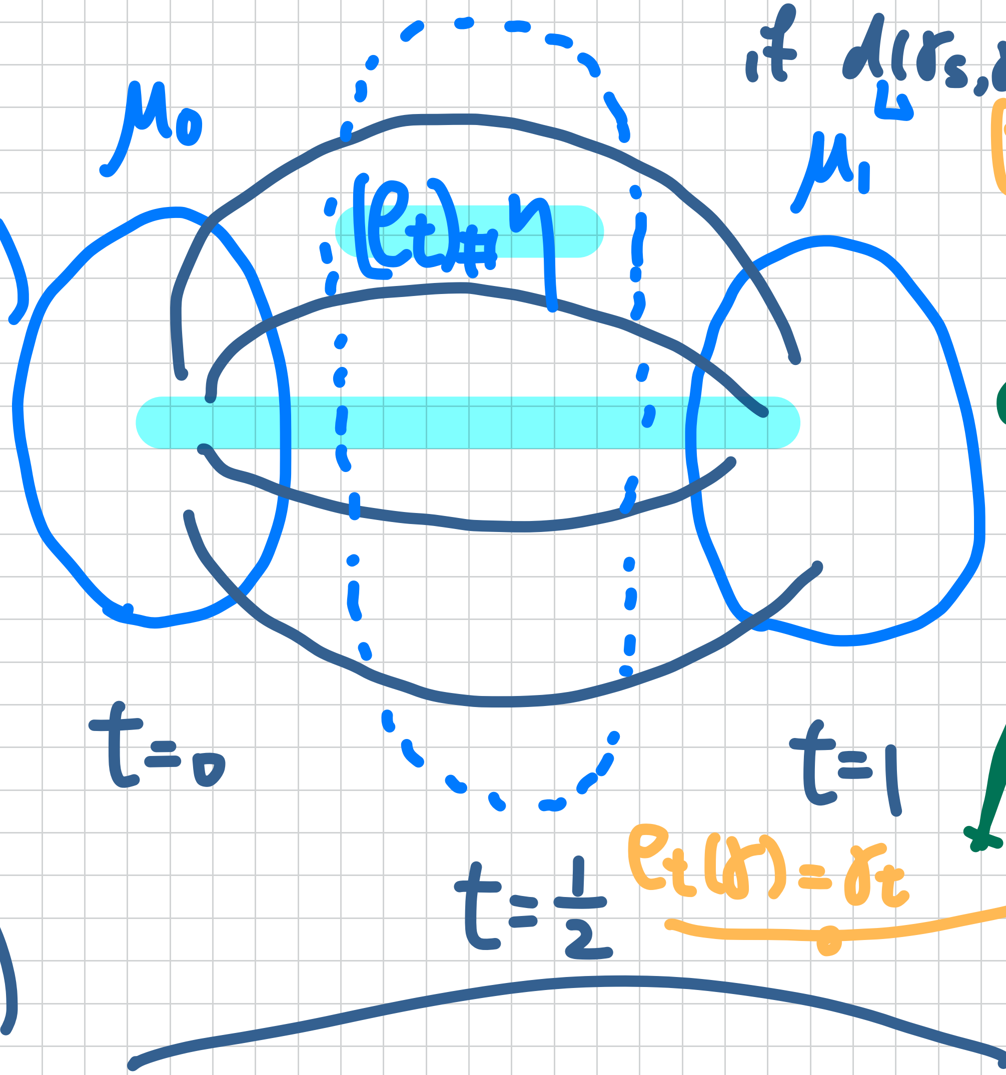
(X, d, m) is ENB

On smooth Riemannian mfd

$\mathrm{Ric} \geq 0$.

$(\mathcal{P}_2(X), d_W)$

m.m.s, (X, d, m)



$\gamma : [0,1] \to \mathcal{P}_2(X)$ is a geodesic $\mathcal{P}_2(X)$
if $d_W(\gamma_s, \gamma_t) = |s-t| \, d_W(\gamma_0, \gamma_1)$ $\forall s,t$

$(e_t)_\#$

$\mathcal{P}(\mathrm{Geo}(X))$

$\forall \eta \in \mathrm{OptGeo}(\mu_0, \mu_1)$ γ^1 γ^2

$\exists \Gamma$ $\eta(\Gamma) = 1$

Γ is non branching

e_t

$\gamma^1 = \gamma^2$

γ

$e_t(\gamma) = \gamma_t$

X

$\mathrm{Geo}(X)$

$$(M, \underset{\in C^{1,1}}{g_{ij}}, m) \text{ is ENB}$$

On smooth Riemannian mfd

Ric ≥ 0 .

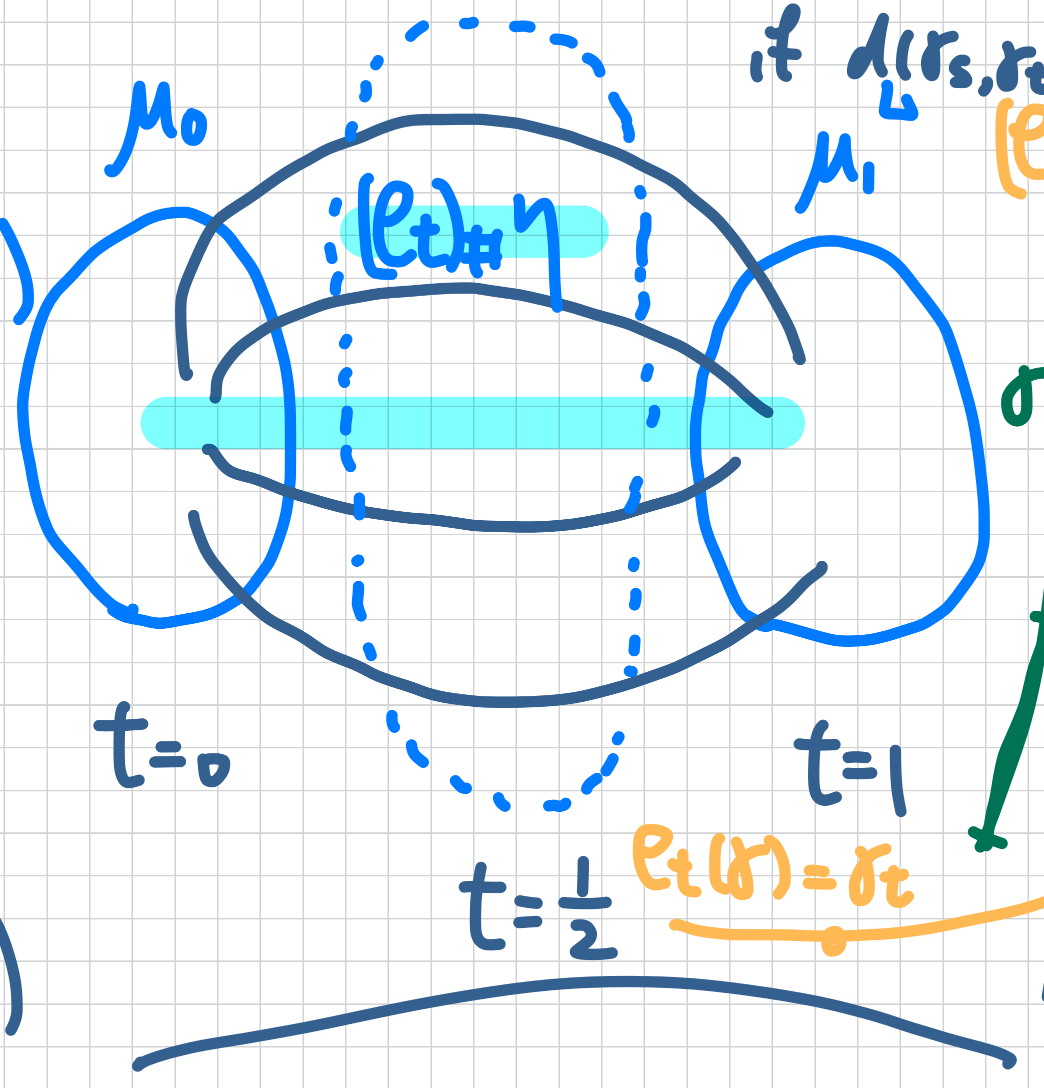
$(P_2(X), d_W)$

m.m.s,

(X, d, m)

$\gamma : [0,1] \to P_2(X)$ is a geodesic $P_2(X)$

if $d_W(\gamma_s, \gamma_t) = |s-t|\, d_W(\gamma_0, \gamma_1)$ $\forall s,t$

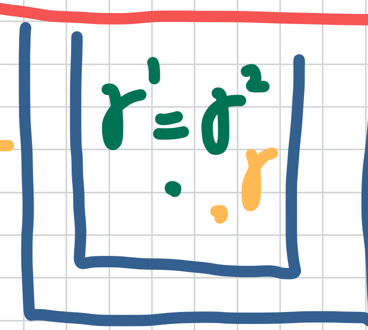


$(e_t)_\#$ $\leftarrow$ $P(Geo(X))$

$\forall \eta \in OptGeo(\mu_0, \mu_1)$ $\exists \Gamma$, $\eta(\Gamma) = 1$, Γ is non branching

γ^1 γ^2

e_t $\leftarrow$



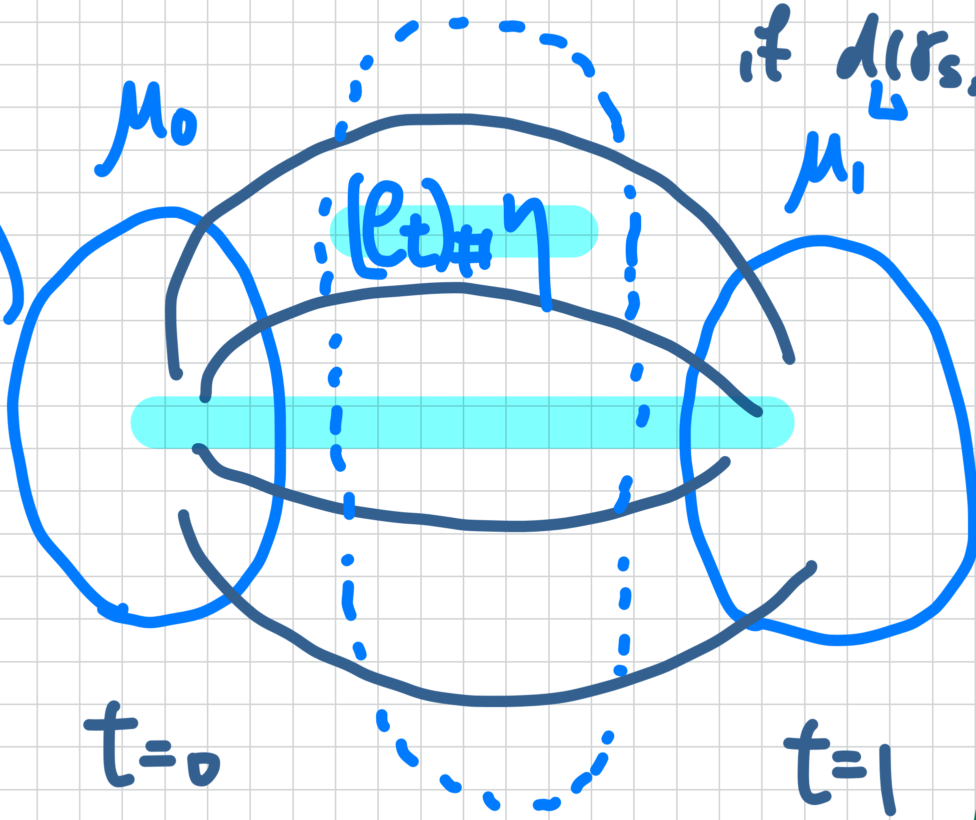
Geo(X)

(M, g_{ij}, m) $\in \cancel{C^{1,1}}$ $C^{0,1}$ is ENB

On smooth Riemannian mfd

$Ric \geq 0.$

$(\mathcal{P}_2(X), d_W)$



m.m.s,

(X, d, m)

$\gamma : [0,1] \to \mathcal{P}_2(X)$ is a geodesic $\mathcal{P}_2(X)$

if $d_W(\gamma_s, \gamma_t) = |s-t|\, d_W(\gamma_0, \gamma_1)$ $\forall s,t$

$(e_t)_\#$

$\longleftarrow$ $P(Geo(X))$

γ^1 η γ^2 $\forall \eta \in$ $\forall$ $OptGeo(\mu_0, \mu_1)$

$\exists \Gamma$, $\eta(\Gamma) = 1$, Γ is non branching

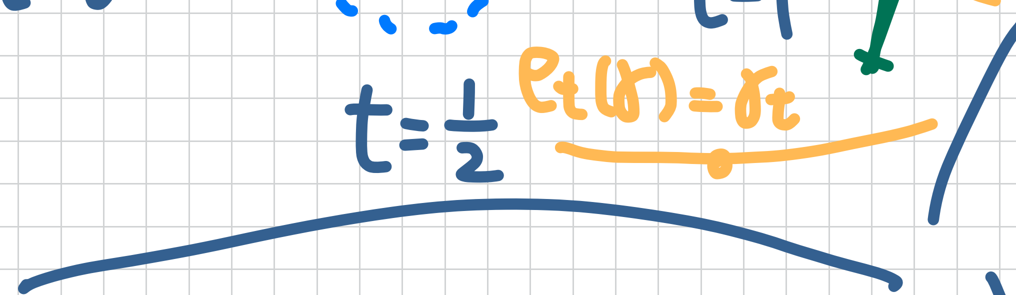
e_t $\longleftarrow$

$\gamma^1 = \gamma^2$

γ

Geo(X)

$e_t(\gamma) = \gamma_t$



X

Dimension Reduction of Distributionally Robust Optimization Problems

Optimal Transport Summer School 2025

Brandon Tam[1]
Silvana Pesenti[1]

[1]Department of Statistical Sciences
University of Toronto

July 2025

Problem Introduction

- We study an application of optimal transport to distributionally robust optimization (DRO) problems.

Problem Introduction

- We study an application of optimal transport to distributionally robust optimization (DRO) problems.
- Our problem takes the following form:

$$\sup_{\boldsymbol{X} \in \mathcal{U}} \rho(g(\boldsymbol{X})).$$

UNIVERSITY OF TORONTO

Problem Introduction

- We study an application of optimal transport to distributionally robust optimization (DRO) problems.
- Our problem takes the following form:

$$\sup_{\boldsymbol{X} \in \mathcal{U}} \rho(g(\boldsymbol{X})).$$

- Consider the worst-case over the set of plausible distributions for the random vector $\boldsymbol{X}$.
- This optimization addresses the issue that the exact distribution of $\boldsymbol{X}$ is often unknown.
- For this talk, I will focus on uncertainty sets $\mathcal{U}$ defined by the Wasserstein distance.

Notation

- We use the following notation:

$$W^{n}(F,G) := \inf_{F_{\boldsymbol{X}}=F, F_{\boldsymbol{Y}}=G} (\mathbb{E}[\|\boldsymbol{X} - \boldsymbol{Y}\|_a^p])^{\frac{1}{p}}.$$

- As the dimension plays an important role in our problem, we will denote it with a superscript n for $n > 1$.
- The dependence on a and p are not important to our problem, so we omit them in the notation whenever there is no confusion.

Wasserstein Uncertainty Sets for Random Variables (rvs)

For $\varepsilon \geq 0$, we define the following uncertainty sets:

1. The univariate Wasserstein uncertainty set (for rvs) around the rv $X \in \mathcal{L}^p$ is given by

$$\mathcal{U}_\varepsilon(X) := \{Z \in \mathcal{L}^p \mid W(F_Z, F_X) \leq \varepsilon\}.$$

UNIVERSITY OF TORONTO

Wasserstein Uncertainty Sets for Random Variables (rvs)

For $\varepsilon \geq 0$, we define the following uncertainty sets:

1. The univariate Wasserstein uncertainty set (for rvs) around the rv $X \in \mathcal{L}^p$ is given by

$$\mathcal{U}_\varepsilon(X) := \{Z \in \mathcal{L}^p \mid W(F_Z, F_X) \leq \varepsilon\}.$$

2. The multivariate Wasserstein uncertainty set (for random vectors) around the random vector $\boldsymbol{X} \in \mathcal{L}^p_n$ is given by

$$\mathcal{U}^n_\varepsilon(\boldsymbol{X}) := \{\boldsymbol{Z} \in \mathcal{L}^p_n \mid W^n(F_{\boldsymbol{Z}}, F_{\boldsymbol{X}}) \leq \varepsilon\}.$$

Two Approaches for Introducing Uncertainty in $g(\boldsymbol{X})$

1. Uncertainty in the Risk Factors:

$$g\big(\mathcal{U}_\varepsilon^n(\boldsymbol{X})\big) := \{g(\boldsymbol{Z}) \mid \boldsymbol{Z} \in \mathcal{U}_\varepsilon^n(\boldsymbol{X})\}$$

2. Uncertainty in the Aggregate:

$$\mathcal{U}_\varepsilon(g(\boldsymbol{X}))$$

UNIVERSITY OF TORONTO

Two Approaches for Introducing Uncertainty in $g(\boldsymbol{X})$

1. Uncertainty in the Risk Factors:

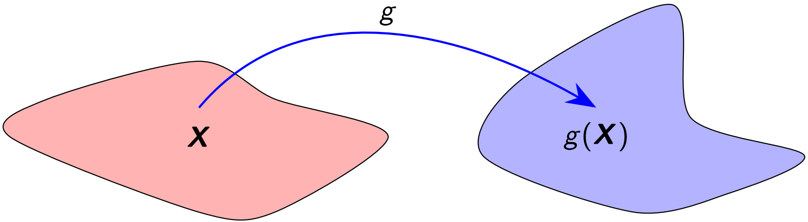
$$g\big(\boldsymbol{\mathcal{U}}_\varepsilon^n(\boldsymbol{X})\big) := \{g(\boldsymbol{Z}) \mid \boldsymbol{Z} \in \boldsymbol{\mathcal{U}}_\varepsilon^n(\boldsymbol{X})\}$$

2. Uncertainty in the Aggregate:

$$\mathcal{U}_\varepsilon(g(\boldsymbol{X}))$$

- In general, these two approaches for introducing uncertainty are not the same, i.e., $g\big(\boldsymbol{\mathcal{U}}_\varepsilon^n(\boldsymbol{X})\big) \neq \mathcal{U}_\varepsilon(g(\boldsymbol{X}))$.

Visualization of $g\left(\mathcal{U}^n_\varepsilon(\boldsymbol{X})\right)$



$$\mathcal{U}^n_\varepsilon(\boldsymbol{X}) = \{\boldsymbol{Z} \mid W^n(F_{\boldsymbol{Z}}, F_{\boldsymbol{X}}) \le \varepsilon\} \qquad g\left(\mathcal{U}^n_\varepsilon(\boldsymbol{X})\right)$$

Figure: Visualization of $g\left(\mathcal{U}^n_\varepsilon(\boldsymbol{X})\right)$

UNIVERSITY OF TORONTO

First Major Result

- Idea: We want g to be sufficiently nice so that the image of the multivariate uncertainty set under g is still a Wasserstein ball.

Theorem 1

Let $\mathfrak{g} : \mathbb{R}^n \to \mathbb{R}$ be L-Lipschitz. Assume $\boldsymbol{X} \in \mathcal{L}_n^p$ and $\mathfrak{g}(\boldsymbol{X}) \in \mathcal{L}^p$. Then, for any $\varepsilon \geq 0$,

$$\mathfrak{g}\big(\boldsymbol{\mathcal{U}}_\varepsilon^n(\boldsymbol{X})\big) \subseteq \mathcal{U}_{L\varepsilon}(\mathfrak{g}(\boldsymbol{X})).$$

Proposition 1

Let ρ be a law invariant risk functional. Under the conditions of Theorem 1,

$$\sup_{\boldsymbol{Y} \in \boldsymbol{\mathcal{U}}_\varepsilon^n(\boldsymbol{X})} \rho\big(\mathfrak{g}(\boldsymbol{Y})\big) \leq \sup_{Y \in \mathcal{U}_{L\varepsilon}(\mathfrak{g}(\boldsymbol{X}))} \rho(Y)\,.$$

- Under slightly stronger conditions on $\mathfrak{g}$, the inequality becomes an equality.

Signed Choquet Integrals

Signed Choquet Integral

A signed Choquet integral is given by

$$I_h(X) = \int_{-\infty}^{0} [h(\mathbb{P}(X \geq x)) - h(1)]dx + \int_{0}^{\infty} h(\mathbb{P}(X \geq x))dx,$$

where $h : [0,1] \to \mathbb{R}$ such that h has bounded variation and $h(0) = 0$.

Signed Choquet Integral

A signed Choquet integral is given by

$$I_h(X) = \int_{-\infty}^{0} [h(\mathbb{P}(X \geq x)) - h(1)]dx + \int_{0}^{\infty} h(\mathbb{P}(X \geq x))dx,$$

where $h : [0,1] \to \mathbb{R}$ such that h has bounded variation and $h(0) = 0$.

- If h is absolutely continuous, then there exists $\gamma : [0,1] \to \mathbb{R}$ such that

$$I_h(X) = \int_0^1 \gamma(u) F_X^{-1}(u) du.$$

- The function γ is called a distortion weight function.

Explicit Bound for Worst-Case Signed Choquet Integrals

Proposition 7

Let I_h be a signed Choquet integral with square integrable, non-decreasing distortion weight function γ satisfying $\|\gamma\|_2 > 0$.

Under the conditions of Theorem 1,

$$\sup_{\boldsymbol{Y} \in \mathcal{U}^n_\varepsilon(\boldsymbol{X})} I_h\big(\mathfrak{g}(\boldsymbol{Y})\big) \leq I_h(\mathfrak{g}(\boldsymbol{X})) + L\varepsilon \|\gamma\|_2 .$$

UNIVERSITY OF TORONTO

Numerical Example - Aggregation Function

- The aggregation function is given by

$$g(\boldsymbol{x}) = -(x_1 + 2\max\{x_2 - 5, 0\} + 3\max\{35 - x_3, 0\} + 4x_4).$$

Numerical Example - Aggregation Function

- The aggregation function is given by

$$g(\boldsymbol{x}) = -(x_1 + 2\max\{x_2 - 5, 0\} + 3\max\{35 - x_3, 0\} + 4x_4).$$

- This aggregation function corresponds to the negative payoff of a portfolio consisting of:
  - 1 unit of a risky asset X_1
  - 2 units of a call option on X_2 with strike price 5
  - 3 units of a put option on X_3 with strike price 35
  - 4 units of a risky asset X_4.
- Clearly, g is Lipschitz with Lipschitz constant $L = 4$.

Numerical Example - Worst-Case Expected Shortfall (ES)

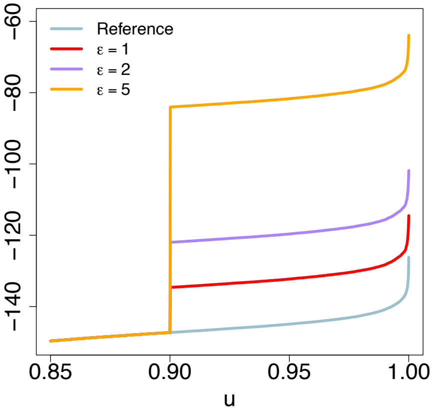


Figure: Worst-case quantile functions for the ES.

Questions?

Thank you for your attention! Any questions?

Figure: QR code for my preprint.

UNIVERSITY OF TORONTO

Solution to Optimization Problem

- If γ is non-decreasing and $\|\gamma\|_2 > 0$, then the supremum in the bound of Proposition 1 is attained by the rv with quantile function

$$F^{-1}_{\mathfrak{g}(\boldsymbol{X})}(u) + \frac{L\varepsilon}{\|\gamma\|_2}\gamma(u).$$

- The assumptions on γ ensures that the proposed quantile function satisfies the properties of a quantile function.
- The non-decreasing assumption can be relaxed, but the form of the solution is more complicated.

References I

Artzner, P., Delbaen, F., Jean-Marc, E., and Heath, D. D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3):203–228.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M., and Brunk, H. (1972). *Statistical Inference Under Order Restrictions: The Theory and Application of Isotonic Regression*. Wiley, London.

Barlow, R. E. and Brunk, H. (1972). The isotonic regression problem and its dual. *Journal of the American Statistical Association*, 67(337):140–147.

Ben-Tal, A. and Nemirovski, A. (1998). Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805.

Ben-Tal, A. and Nemirovski, A. (1999). Robust solutions of uncertain linear programs. *Operations Research Letters*, 25(1):1–13.

Bernard, C., Pesenti, S. M., and Vanduffel, S. (2024). Robust distortion risk measures. *Mathematical Finance*, 34(3):774–818.

Blanchet, J. and Murphy, K. (2019). Quantifying distributional model risk via optimal transport. *Mathematics of Opertations Research*, 44(2):565–600.

Brighi, B. and Chipot, M. (1994). Approximated convex envelope of a function. *SIAM Journal on Numerical Analysis*, 31(1):128–148.

Brunk, H. (1965). Conditional expectation given a sigma lattice and applications. *The Annals of Mathematical Statistics*, 36(5):1339–1350.

Carlier, G. and Jimenez, C. (2007). On monge's problem for bregman-like cost functions. *Journal of Convex Analysis*, 14(3):647–655.

Choquet, G. (1954). Theory of capacities. *Annales de l'institut Fourier*, 5:131–295.

Cont, R., Deguest, R., and Scandolo, G. (2010). Robustness and sensitivity analysis of risk measurement procedures. *Quantitative Finance*, 10(6):593–606.

References III

Cuturi, M. (2013). Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in Neural Information Processing Systems*, 26.

Dantzig, G. B. (1955). Linear programming under uncertainty. *Management Science*, 1(3-4):197–206.

de Leeuw, J., Hornik, K., and Mair, P. (2009). Isotone optimization in r: Pool-adjacent-violators algorithm (pava) and active set methods. *Journal of Statistical Software*, 32(5):1–24.

Deshpande, I., Hu, Y.-T., Sun, R., Pyrros, A., Siddiqui, N., Koyejo, S., Zhao, Z., Forsyth, D., and Schwing, A. (2019). Max-sliced wasserstein distance and its use for gans. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Dobrushin, R. (1970). Prescribing a system of random variables by conditional distributions. *Theory of Probability and Its Applications*, 15(3):458–486.

References IV

Dowson, D. and Landau, B. (1982). The fréchet distance between multivariate normal distributions. *Journal of Multivariate Analysis*, 12:450–455.

Embrechts, P., Wang, B., and Wang, R. (2015). Aggregation-robustness and model uncertainty of regulatory risk measures. *Finance and Stochastics*, 19(4):763–790.

Föllmer, H. and Shied, A. (2002). Convex measures of risk and trading constraints. *Finance and Stochastics*, 6(4):429–447.

Gao, R. and Kleywegt, A. (2023). Distributionally robust stochastic optimization with wasserstein distance. *Mathematics of Opertations Research*, 48(2):603–655.

Guo, X., Hong, J., and Yang, N. (2017). Ambiguity set and learning via bregman and wasserstein. *arXiv Preprint*, arXiv:1705.08056.

Kolouri, S., Nadjahi, K., Simsekli, U., Badeau, R., and Rohde, G. (2019). Generalized sliced wasserstein distances. *Advances in Neural Information Processing Systems*, 32.

Kuhn, D., Esfahani, P. M., Nguyen, V. A., and Shafieezadeh-Abadeh, S. (2019). Wasserstein distributionally robust optimization: Theory and applications in machine learning. *INFORMS TutORials in Operations Research*, pages 130–166.

Kusuoka, S. (2001). On law invariant coherent risk measures. *Advances in Mathematical Economics*, 3:83–95.

Mao, T., Wang, R., and Wu, Q. (2022). Model aggregation for risk evaluation and robust optimization. *arXiv Preprint*, arXiv:2201.06370.

McLachlan, G. J. (1999). Mahalanobis distance. *Resnonance*, 4(6):20–26.

Michaud, R. O. (1989). The markowitz optimization enigma: Is 'optimized' optimal? *Financial Analysts Journal*, 45(1):31–42.

References VI

Miller, B. K., Federici, M., Weniger, C., and Forré, P. (2023). Simulation based inference with the generalized kullbeck-liebler divergence. *arXiv Preprint*, arXiv:2310.01808.

Munk, A. and Czado, C. (1998). Nonparametric validation of similar distributions and assessment of goodness of fit. *Journal of the Royal Statistical Society, Series B*, 60(1):223–241.

Olea, J. L. M., Rush, C., Velez, A., and Wiesel, J. (2024). The out-of-sample prediction error of the lasso and related estimators. *arXiv Preprint*, arXiv:2211.07608.

Panaretos, V. M. and Zemel, Y. (2019). Statistical aspects of wasserstein distances. *Annual Review of Statistics and Its Applications*, 6(1):405–431.

Paty, F.-P. and Cuturi, M. (2019). Subspace robust wasserstein distances. *International Conference on Machine Learning*.

Pesenti, S. M. (2022). Reverse sensitivity analysis for risk modelling. *Risks*, 10(7).

Pesenti, S. M., Millossovich, P., and Tsanakas, A. (2016). Robustness regions for measures of risk aggregation. *Dependence Modelling*, 4:348–367.

Pesenti, S. M. and Vanduffel, S. (2024). Optimal transport divergences induced by scoring functions. *Operations Research Letters*, 57.

Pesenti, S. M., Vanduffel, S., Yang, Y., and Yao, J. (2024). Optimal payoff under bregman-wasserstein divergence constraints. *arXiv Preprint*, arXiv:2411.18397.

Rachev, S. T. and Rüshendorf, L. (1998). *Mass Transportation Problems: Volume I: Theory*. Springer New York, New York, NY.

Rankin, C. and Wong, T.-K. L. (2023). Bregman-wasserstein divergence: Geometry and applications. *arXiv Preprint*, arXiv:2302.05833.

Rockafellar, R. T., Uryasev, S., and Zabarankin, M. (2006). Generalized deviations in risk analysis. *Finance and Stochastics*, 10:51–74.

Scanlon, C. H. (1970). Rings of functions with certain lipschitz properties. *Pacific Journal of Mathematics*, 32(1):197–201.

Smith, J. E. and Winkler, R. L. (2006). The optimizer's curse: Skepticism and postdecision surprise in decision analysis. *Management Science*, 52(3):311–322.

Soyster, A. L. (1973). Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research*, 21(5):1154–1157.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5:297–323.

Villani, C. (1973). *Topics in Optimal Transportation*. American Mathematical Society.

UNIVERSITY OF TORONTO

References IX

Wang, R., Wei, Y., and Willmot, G. E. (2020). Characterization, robustness and aggregation of signed choquet integrals. *Mathematics of Operations Research*, 45(3):797–1192.

UNIVERSITY OF TORONTO

Optimal transport-based denoising and applications to collider physics

Benjamin Faktor, UCLA

In collaboration with:
Katy Craig, UCSB
Benjamin Nachman, SLAC/CMS

UCSB Summer School on Optimal Transport and Applications
22 July 2025

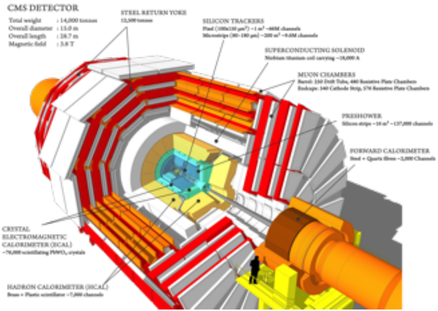
Outline

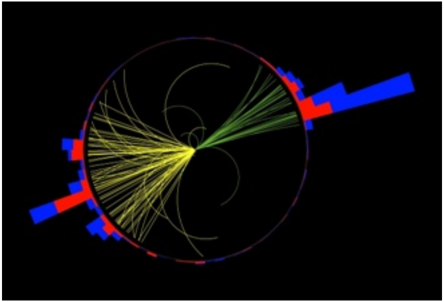
1. Large Hadron Collider and Motivation

2. Formulation

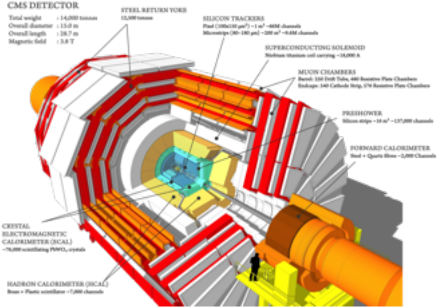
3. Numerical Solution

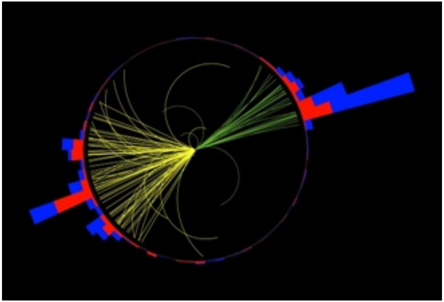
Large Hadron Collider





Large Hadron Collider



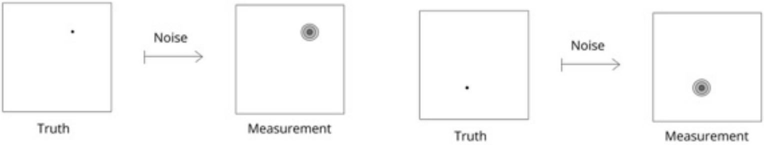


Moral: Collider measurement is a probability measure on $\mathbb{R}^2$

Motivation

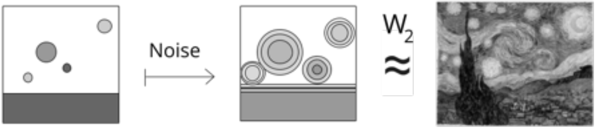
Given are a fixed measurement

and a noise model:



etc.

Question: Which image's noisification is closest?



Formulation

Given: fixed measurement $\nu \in \mathscr{P}_2(\mathbb{R}^d)$ and noise model $\rho : \mathbb{R}^d \to \mathscr{P}(\mathbb{R}^d)$

Formulation

Given: fixed measurement $\nu \in \mathscr{P}_2(\mathbb{R}^d)$ and noise model $\rho : \mathbb{R}^d \to \mathscr{P}(\mathbb{R}^d)$

ρ defines the *noisification* $\nu_\sigma \in \mathscr{P}(\mathbb{R}^d)$ of $\sigma \in \mathscr{P}(\mathbb{R}^d)$ via

$$\int f(y)\, d\nu_\sigma(y) = \iint f(y)\, d\rho_x(y)\, d\sigma(x), \quad \forall f \in C_b(\mathbb{R}^d)$$

Formulation

Given: fixed measurement $\nu \in \mathscr{P}_2(\mathbb{R}^d)$ and noise model $\rho : \mathbb{R}^d \to \mathscr{P}(\mathbb{R}^d)$

ρ defines the *noisification* $\nu_\sigma \in \mathscr{P}(\mathbb{R}^d)$ of $\sigma \in \mathscr{P}(\mathbb{R}^d)$ via

$$\int f(y)\, d\nu_\sigma(y) = \iint f(y)\, d\rho_x(y)\, d\sigma(x), \quad \forall f \in C_b(\mathbb{R}^d)$$

Remark:

$$\nu_{\delta_a} = \rho_a \text{ for } a \in \mathbb{R}^d$$

Formulation

Given: fixed measurement $\nu \in \mathscr{P}_2(\mathbb{R}^d)$ and noise model $\rho : \mathbb{R}^d \to \mathscr{P}(\mathbb{R}^d)$

ρ defines the *noisification* $\nu_\sigma \in \mathscr{P}(\mathbb{R}^d)$ of $\sigma \in \mathscr{P}(\mathbb{R}^d)$ via

$$\int f(y)\, d\nu_\sigma(y) = \iint f(y)\, d\rho_x(y)\, d\sigma(x), \quad \forall f \in C_b(\mathbb{R}^d)$$

Remark:

$\nu_{\delta_a} = \rho_a$ for $a \in \mathbb{R}^d$

If $d\rho_x(y) = \phi(y - x)\, dy$ then noisification is convolution with ϕ

Formulation

Given: fixed measurement $\nu \in \mathscr{P}_2(\mathbb{R}^d)$ and noise model $\rho : \mathbb{R}^d \to \mathscr{P}(\mathbb{R}^d)$

ρ defines the *noisification* $\nu_\sigma \in \mathscr{P}(\mathbb{R}^d)$ of $\sigma \in \mathscr{P}(\mathbb{R}^d)$ via

$$\int f(y)\, d\nu_\sigma(y) = \iint f(y)\, d\rho_x(y)\, d\sigma(x), \quad \forall f \in C_b(\mathbb{R}^d)$$

Remark:

$\nu_{\delta_a} = \rho_a$ for $a \in \mathbb{R}^d$

If $d\rho_x(y) = \phi(y-x)\, dy$ then noisification is convolution with ϕ

We seek to study the *denoising problem*

$$\sigma_* := \arg\min_\sigma \; W_2(\nu_\sigma, \nu)$$

Formulation

Given: fixed measurement $\nu \in \mathscr{P}_2(\mathbb{R}^d)$ and noise model $\rho : \mathbb{R}^d \to \mathscr{P}(\mathbb{R}^d)$

ρ defines the *noisification* $\nu_\sigma \in \mathscr{P}(\mathbb{R}^d)$ of $\sigma \in \mathscr{P}(\mathbb{R}^d)$ via

$$\int f(y)\, d\nu_\sigma(y) = \iint f(y)\, d\rho_x(y)\, d\sigma(x), \quad \forall f \in C_b(\mathbb{R}^d)$$

Remark:

$\nu_{\delta_a} = \rho_a$ for $a \in \mathbb{R}^d$

If $d\rho_x(y) = \phi(y - x)\, dy$ then noisification is convolution with ϕ

We seek to study the *denoising problem*

$$\sigma_* := \arg\min_\sigma\ W_2(\nu_\sigma, \nu)$$

Thm [Craig, F. '23] If ρ is W_2-continuous and $x \mapsto M_2(\rho_x)$ has compact sublevel sets then a solution exists. Moreover, if $\sigma \mapsto \nu_\sigma$ is injective and $\nu \ll \mathscr{L}^d$ then the solution is unique.

Numerical Solution

The denoising problem has poor stability;

e.g. if $\nu = \nu_{\sigma^*}$ there may be $\bar{\sigma}$ with $W_2(\nu_{\bar{\sigma}}, \nu_{\sigma^*})$ small but $W_2(\bar{\sigma}, \sigma^*)$ large.

Numerical Solution

The denoising problem has poor stability;

e.g. if $\nu = \nu_{\sigma^*}$ there may be $\bar{\sigma}$ with $W_2(\nu_{\bar{\sigma}}, \nu_{\sigma^*})$ small but $W_2(\bar{\sigma}, \sigma^*)$ large.

Instead, we add entropic regularization:

$$\sigma^* := \underset{(\sigma,\Gamma):\, \Gamma \in \Pi(\nu_\sigma, \nu)}{\arg\min} \left\{ \iint |y - y'|^2 d\Gamma + \varepsilon \int \sigma \log \sigma + \varepsilon \iint \Gamma \log \Gamma \right\}$$

Numerical Solution

The denoising problem has poor stability;

e.g. if $\nu = \nu_{\sigma^*}$ there may be $\bar{\sigma}$ with $W_2(\nu_{\bar{\sigma}}, \nu_{\sigma^*})$ small but $W_2(\bar{\sigma}, \sigma^*)$ large.

Instead, we add entropic regularization:

$$\sigma^* := \operatorname*{arg\,min}_{(\sigma,\Gamma):\,\Gamma\in\Pi(\nu_\sigma,\nu)} \left\{ \iint |y-y'|^2 d\Gamma + \varepsilon \int \sigma \log \sigma + \varepsilon \iint \Gamma \log \Gamma \right\}$$

At the discrete level,

Numerical Solution

The denoising problem has poor stability;

e.g. if $\nu = \nu_{\sigma^*}$ there may be $\bar{\sigma}$ with $W_2(\nu_{\bar{\sigma}}, \nu_{\sigma^*})$ small but $W_2(\bar{\sigma}, \sigma^*)$ large.

Instead, we add entropic regularization:

$$\sigma^* := \underset{(\sigma,\Gamma):\, \Gamma \in \Pi(\nu_\sigma,\nu)}{\arg\min} \left\{ \iint |y-y'|^2 d\Gamma + \varepsilon \int \sigma \log \sigma + \varepsilon \iint \Gamma \log \Gamma \right\}$$

At the discrete level, this is equivalent to

$$\underset{A \in \mathcal{A}_1 \cap \mathcal{A}_2}{\arg\min}\ KL(A|B), \qquad \mathcal{A}_1, \mathcal{A}_2 \text{ affine, } B \text{ fixed}$$

and is solvable by alternating KL-projections.

Numerical Solution

The denoising problem has poor stability;

e.g. if $\nu = \nu_{\sigma^*}$ there may be $\bar{\sigma}$ with $W_2(\nu_{\bar{\sigma}}, \nu_{\sigma^*})$ small but $W_2(\bar{\sigma}, \sigma^*)$ large.

Instead, we add entropic regularization:

$$\sigma^* := \operatorname*{arg\,min}_{(\sigma,\Gamma):\, \Gamma \in \Pi(\nu_\sigma, \nu)} \left\{ \iint |y - y'|^2 d\Gamma + \varepsilon \int \sigma \log \sigma + \varepsilon \iint \Gamma \log \Gamma \right\}$$

At the discrete level, this is equivalent to

$$\operatorname*{arg\,min}_{A \in \mathcal{A}_1 \cap \mathcal{A}_2} KL(A|B), \qquad \mathcal{A}_1, \mathcal{A}_2 \text{ affine, } B \text{ fixed}$$

and is solvable by alternating KL-projections.

E-L equations allow us to find equivalent Sinkhorn-type iterations

Thank you!

Why OT?

– In the LHC data, the supports of ν and the ρ_{x_i} are not the same

– To compare ν_σ and ν using e.g. KL, ν must be "binned" to fit the support of the ρ_{x_i}, losing information

– An OT cost circumvents this and better quantifies horizontal shifts

Why OT?

– In the LHC data, the supports of ν and the ρ_{x_i} are not the same

– To compare ν_σ and ν using e.g. KL, ν must be "binned" to fit the support of the ρ_{x_i}, losing information

– An OT cost circumvents this and better quantifies horizontal shifts