Probabilistic Embedding through the Wasserstein Tangent Space

Maria Oprea

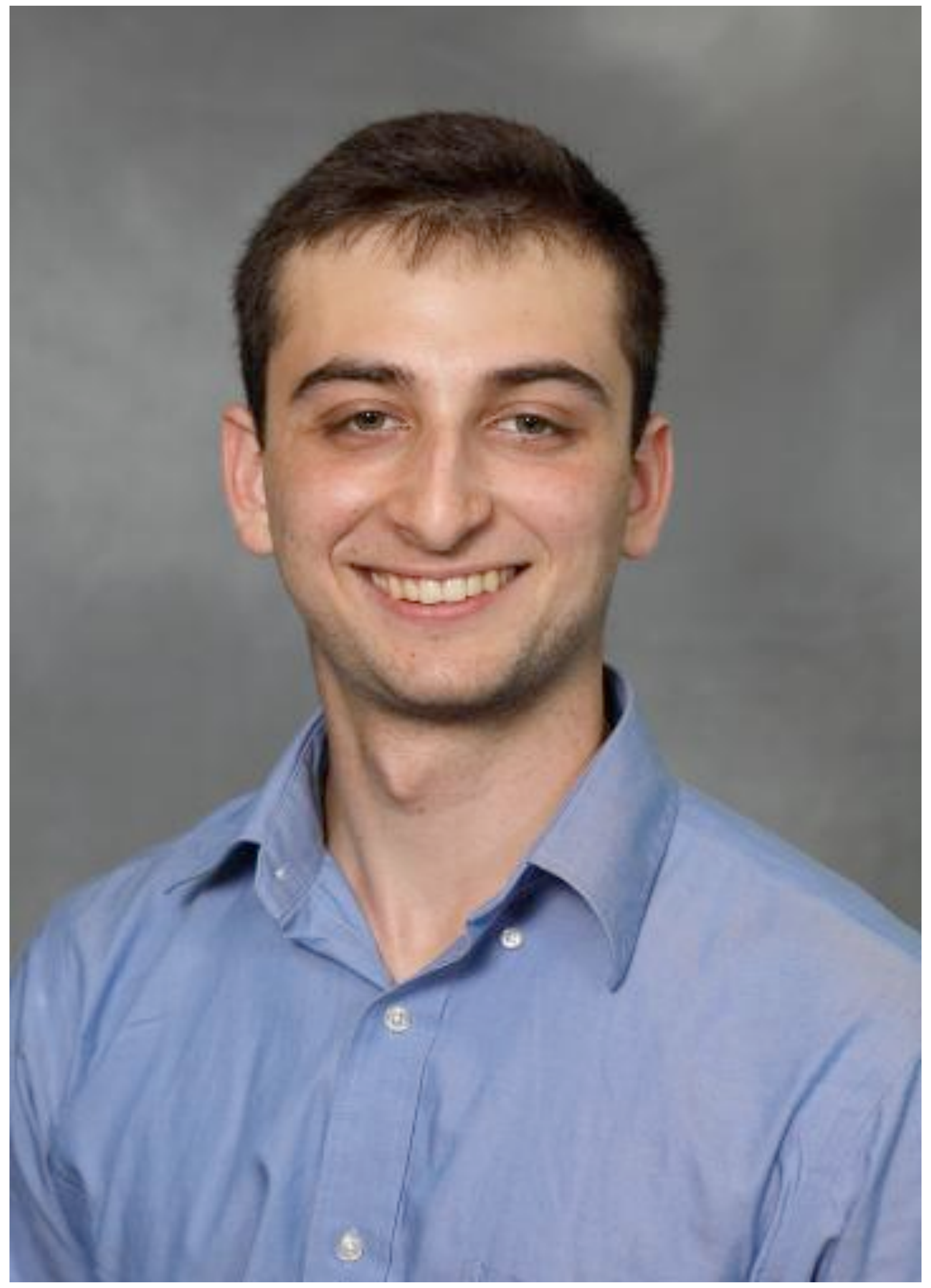
NANMAT 24

Joint work with:

Yunan Yang

Goenka Family Assistant Professor in Mathematics[1]

Jonah Botvinick-Greenhouse



PhD candidate at the Center for Applied Mathematics[1]

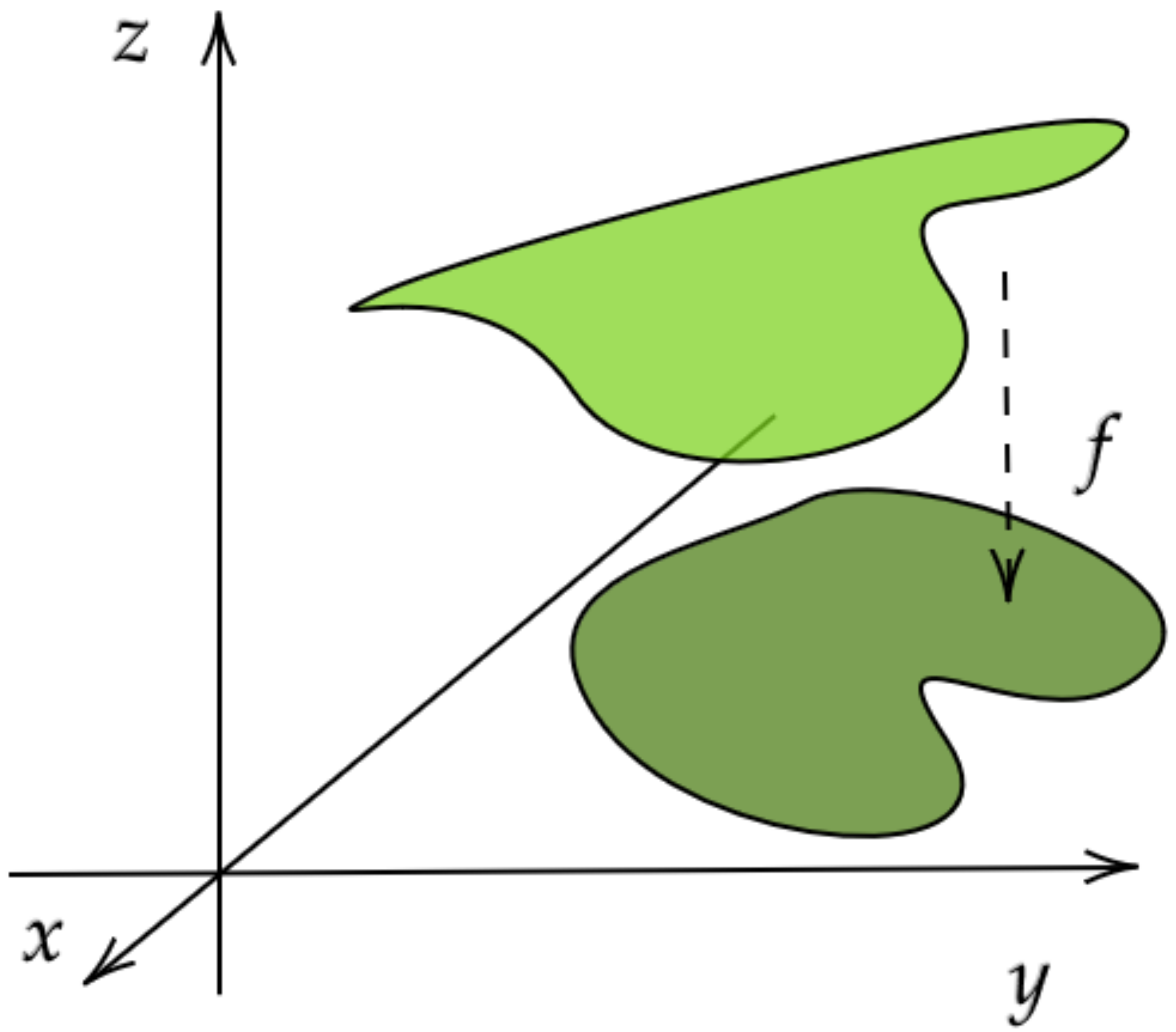
Romit Maulik

Assistant Professor at Penn State College of Information Sciences and Technologies

[1]Cornell University

Embeddings

- Notion of „sameness"
- Diffeomorphism that preserves the differential structure

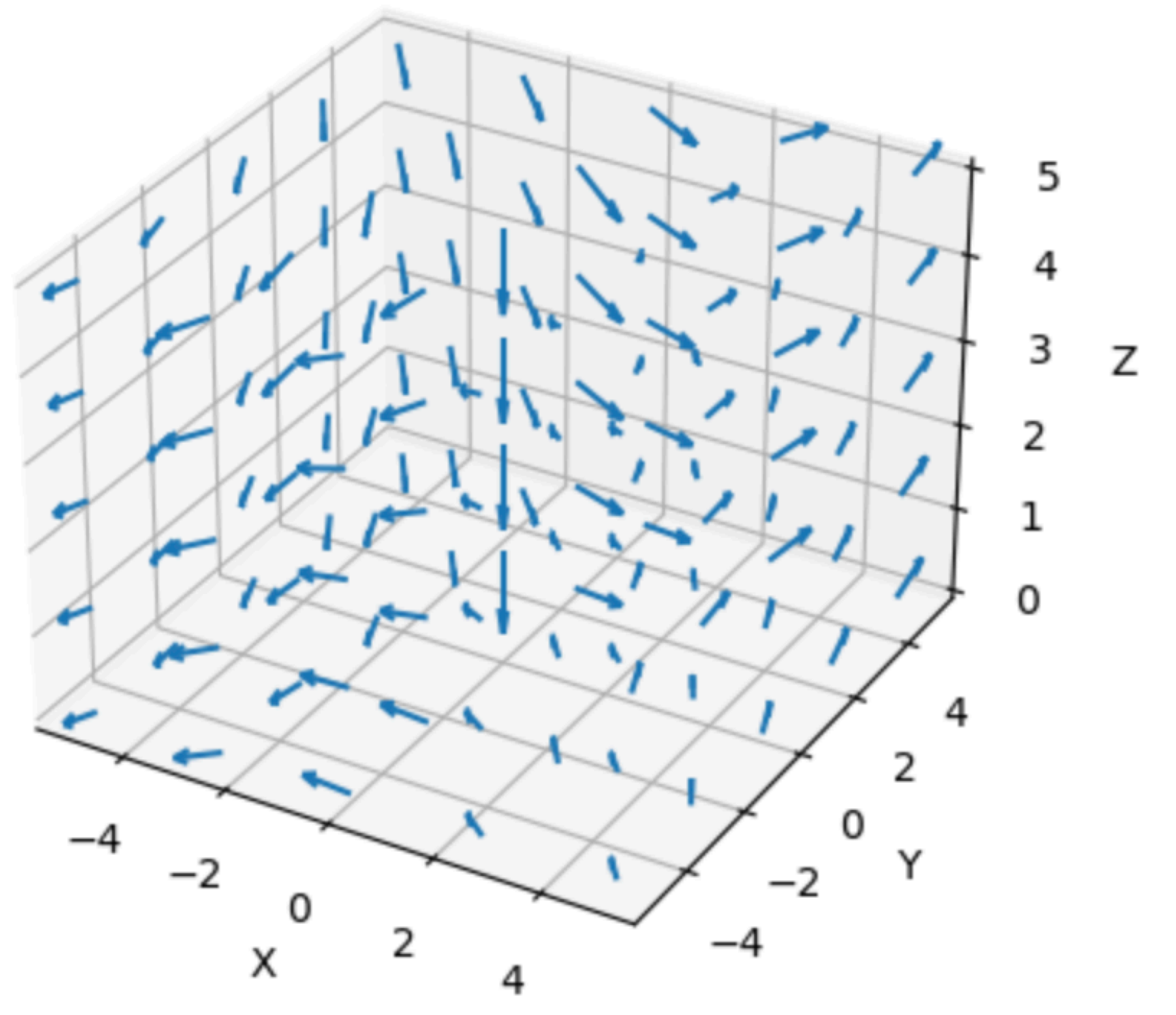


Definition: $f : M \to N$ is an embedding if:

- f is bijective onto $f(M)$
- f is differentiable
- Df_x is injective $\forall x \in M$

Takens' embedding

- Can we reconstruct the full state given partial information?
- Dynamical system $\dot{x} = f(x)$ with flow $\phi_t : M \to M, x(t) = \phi_t(x_0)$
- An observation: $h : M \to \mathbb{R}$



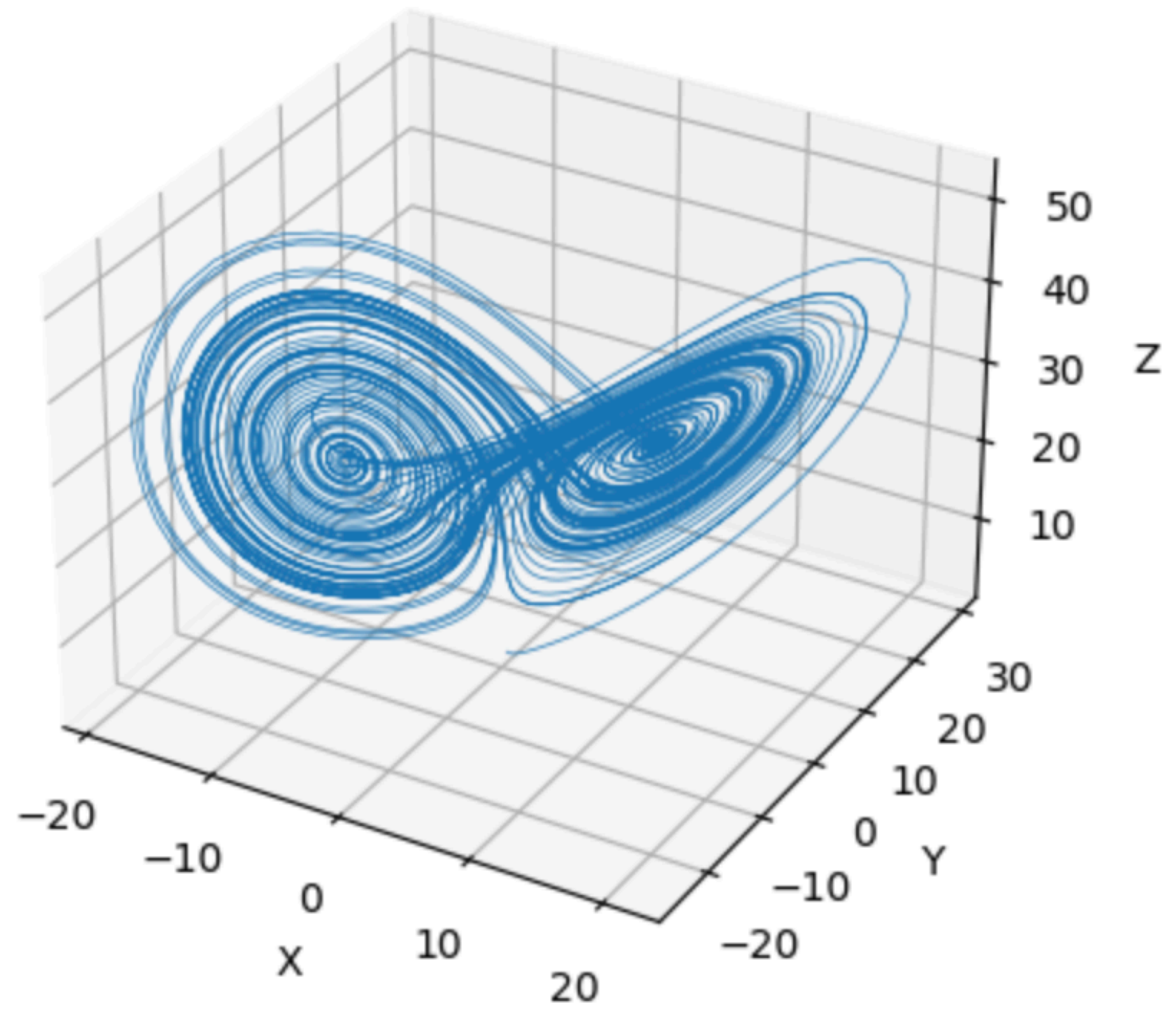
Theorem[1] If f satisfies:

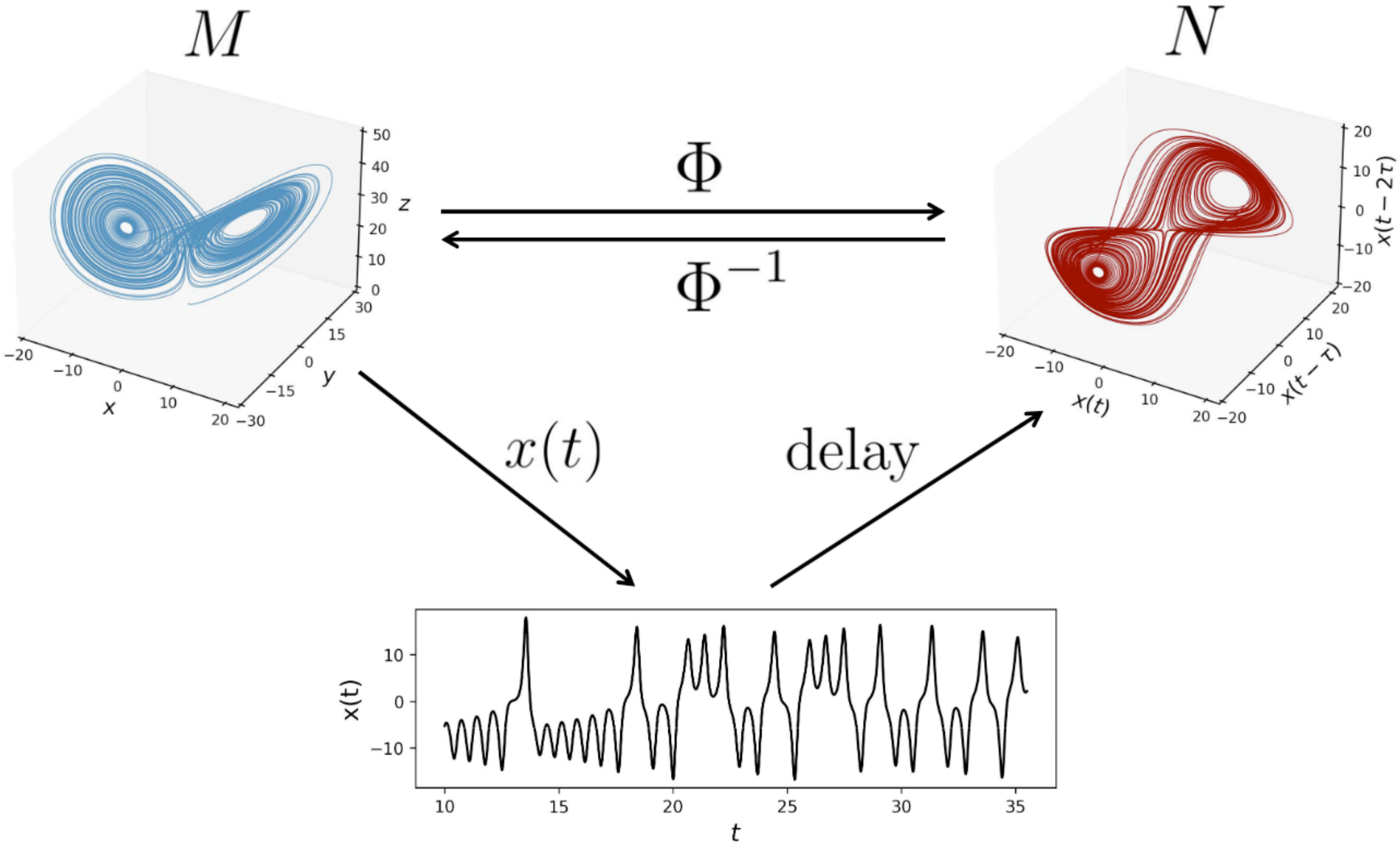
1. If $f(x) = 0$ then the eigenvalues of $d\phi_\tau|_x$ are all different and different to from 1,
2. No periodic integral curve of f has period less that $2d + 1$,

then it is a generic property that the delay coordinate map given by

$$\Phi_{h,\phi_t}(x) := (h(x), h(\phi_\tau(x)), \ldots h(\phi_{(d-1)\tau}(x))) \in \mathbb{R}^d$$

an embedding of M into $\mathbb{R}^d$ provided that $d \geq 2n + 1$.

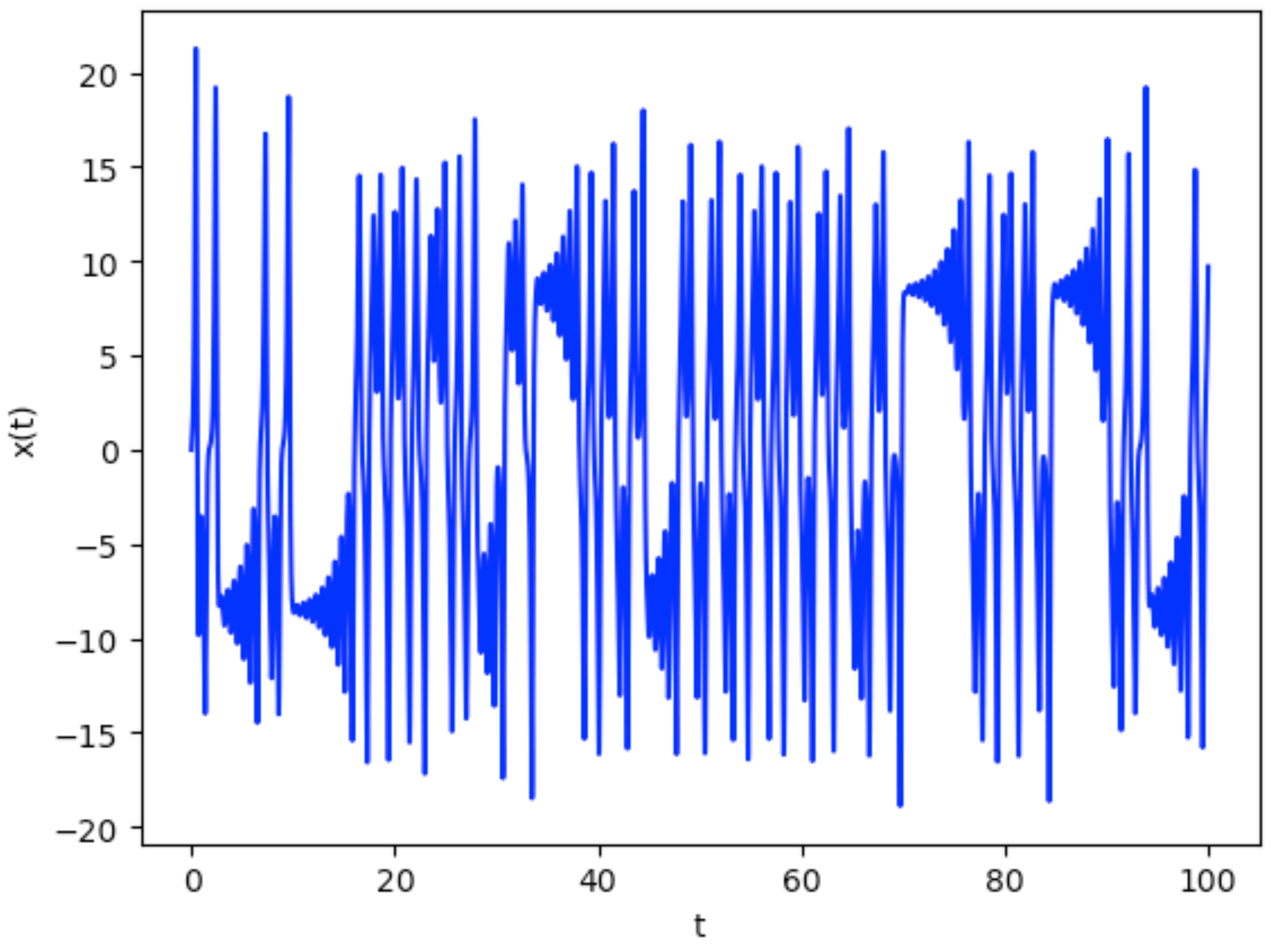




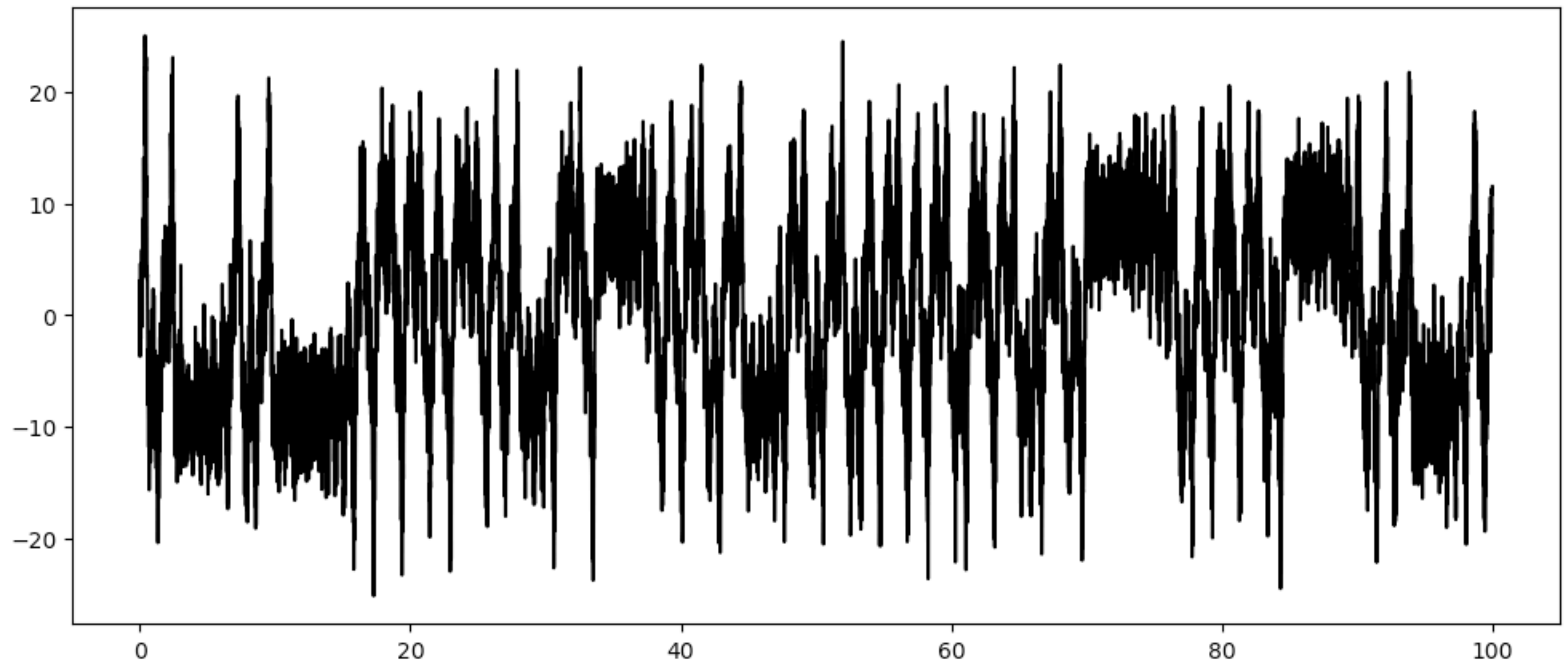
M
N
Φ
Φ^{-1}
$x(t)$
delay
x
y
z
x(t)
$x(t-\tau)$
$x(t-2\tau)$
t

Limitations

- Takens requires clean data, deterministic setup.
- Chaos + Uncertainty in the initial position $\rho_0 \implies \phi_t \# \rho_0$
- Stochastic dynamics $\rho_t = law(X_t)$
- Noisy data $\tilde{x}_t = x_t + \mathcal{N}(0,1)$, $\rho_t = \delta_{x_t} * \mathcal{N}(0,1)$



Clean

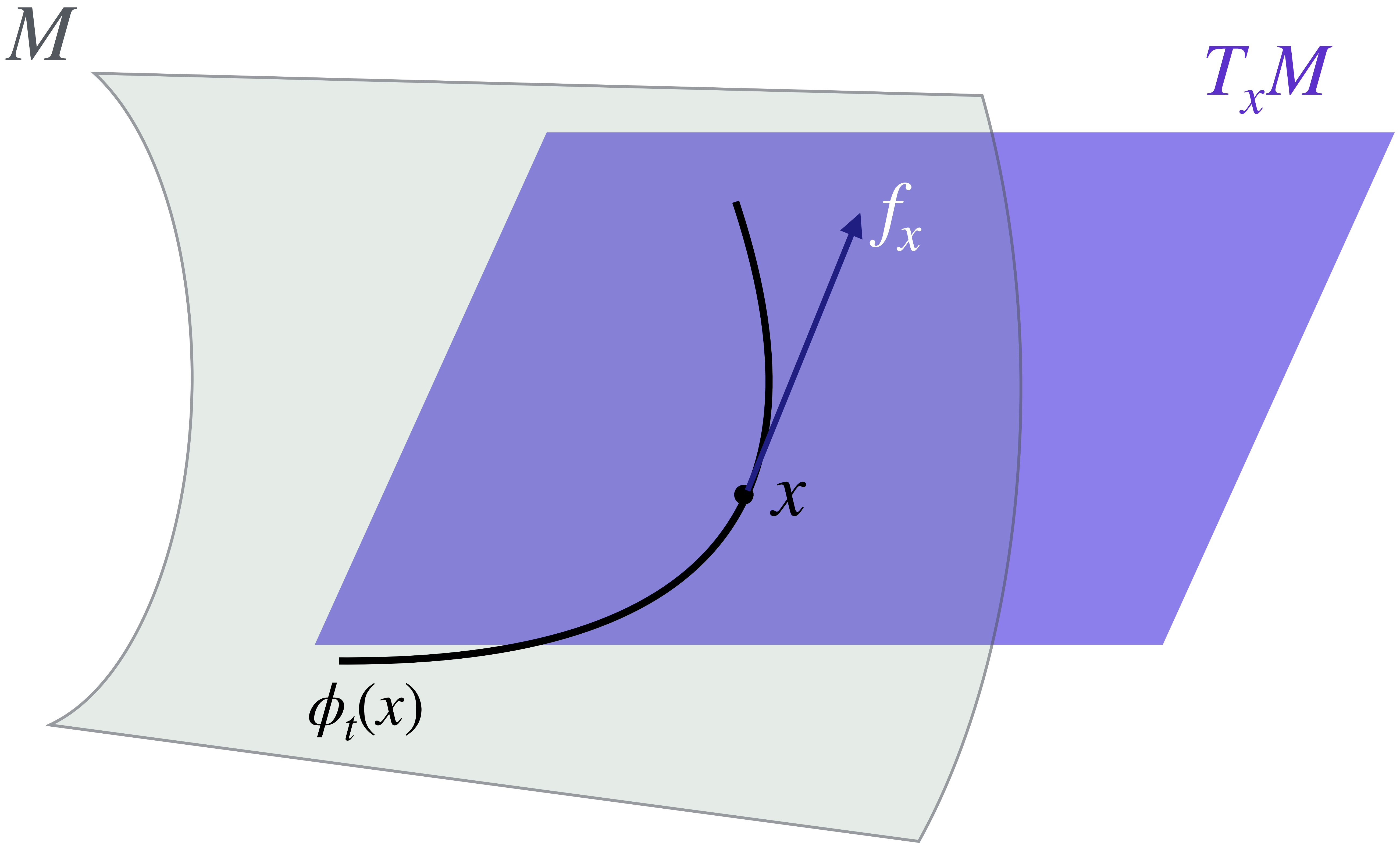


Noisy

Change of paradigm

Lagrangian	**Eulerian**
Manifold M	$\mathscr{P}_2(M) = \left\{ \rho, \int_M \rho = 1, \int_M \lvert x \rvert^2 \rho < \infty \right\}$
Distance $d(x, y)$	Wasserstein metric $W_2(\rho_x, \rho_y)$
Vector field $f_x \in T_x M$	Modulo divergence free square integrable functions $v(\rho) \in L^2(TM, \rho)/ \sim \lvert v \sim w \iff \nabla \cdot ((v - w)\rho) = 0$
Integral curve $\frac{d}{dt}\phi_t(x) := \dot{x} = f(x)$	Distributions solving the continuity equation $\partial_t \rho_t + \nabla \cdot (\rho_t v_t) = 0$

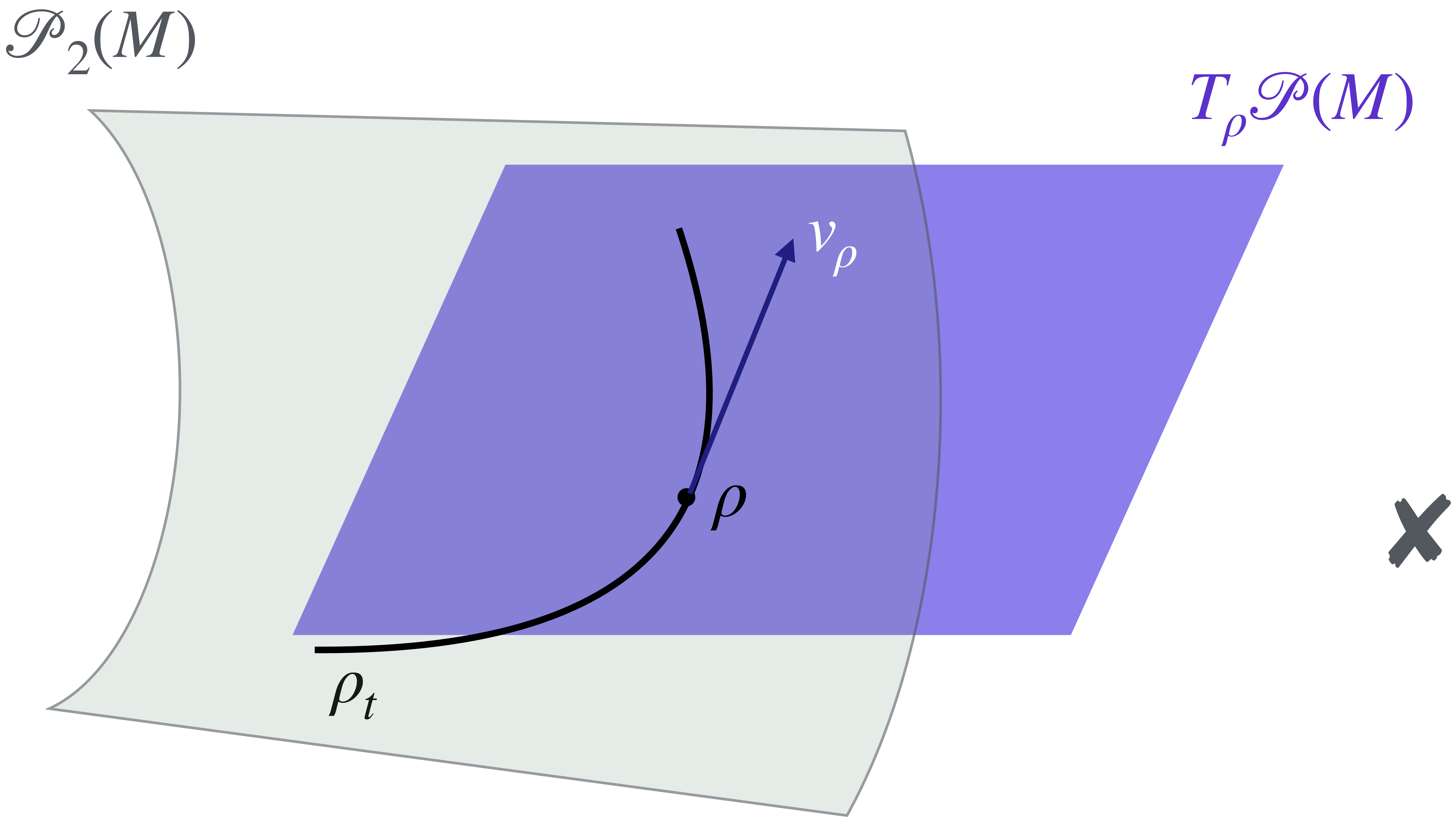
Schematic drawing



$$\frac{d}{dt}\phi_t(x) = \dot{x} = f_x$$

Schematic drawing

- If $X \sim \rho_0$ then $\Phi_t(X) = X_t \sim \rho_t$



✘ Infinite dimensional

v_ρ is a function for every ρ

$\partial_t \rho_t + \nabla \cdot (\rho_t v_t) = 0$ In the weak sense!!

- Many v_t satisfy the continuity equation. Project out the divergence free component![2]

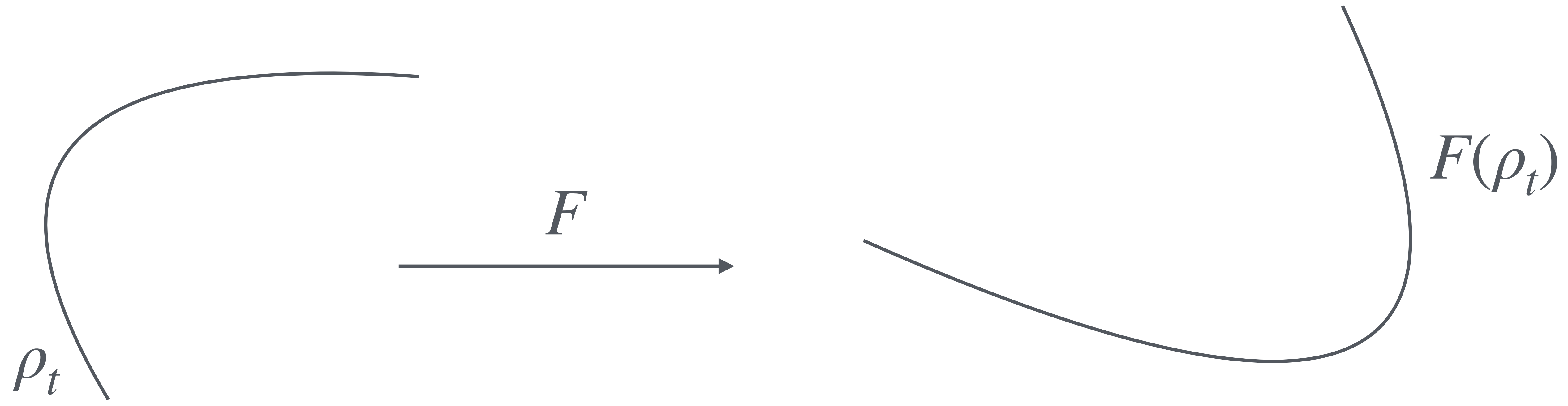
$$P_\rho : L^2(TM, \rho) \to T_\rho \mathscr{P}(M)$$

Embeddings in $\mathscr{P}_2(M)$

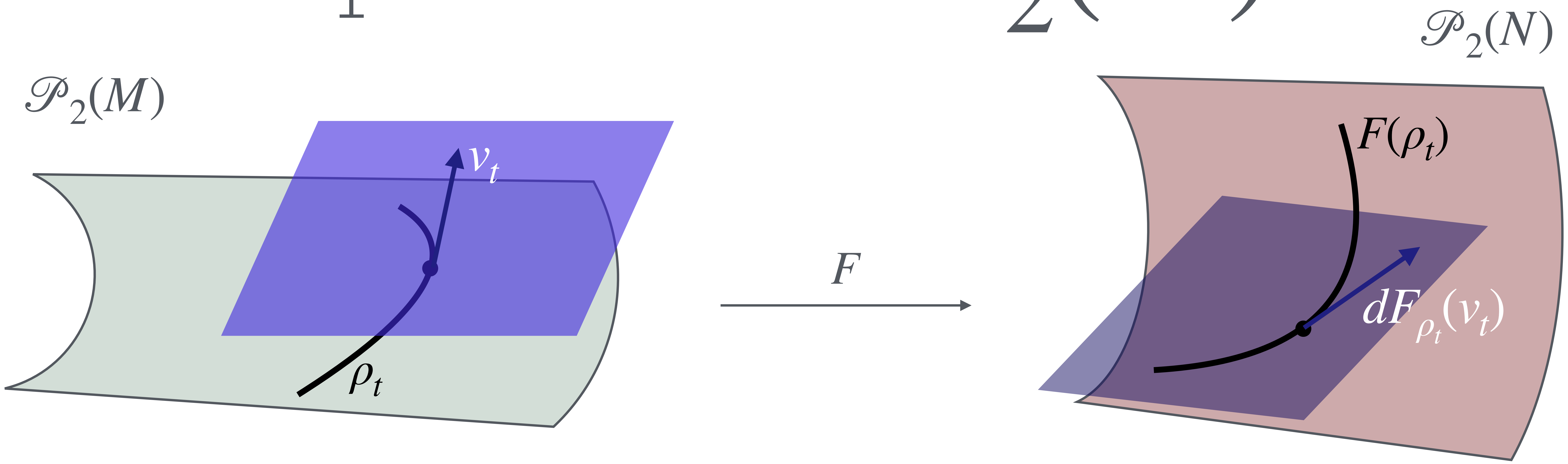
Definition: A map $F : \mathscr{P}(M) \to \mathscr{P}(N)$ is an embedding if

1. F is a bijection onto its image
2. F is differentiable
3. The derivative operator DF is injective i.e.
   $\forall v \neq w \in T_\rho \mathscr{P}(M), DF_\rho(v) \neq DF_\rho(w) \in T_{F(\rho)} \mathscr{P}(N)$

- A map is differentiable if it takes differentiable curves to differentiable curves
- Differentiable $\iff$ absolutely continuous[2][Ambrogio, Gigli, Savare, Thm 8.3.1]



Derivative operator in $\mathscr{P}_2(M)$

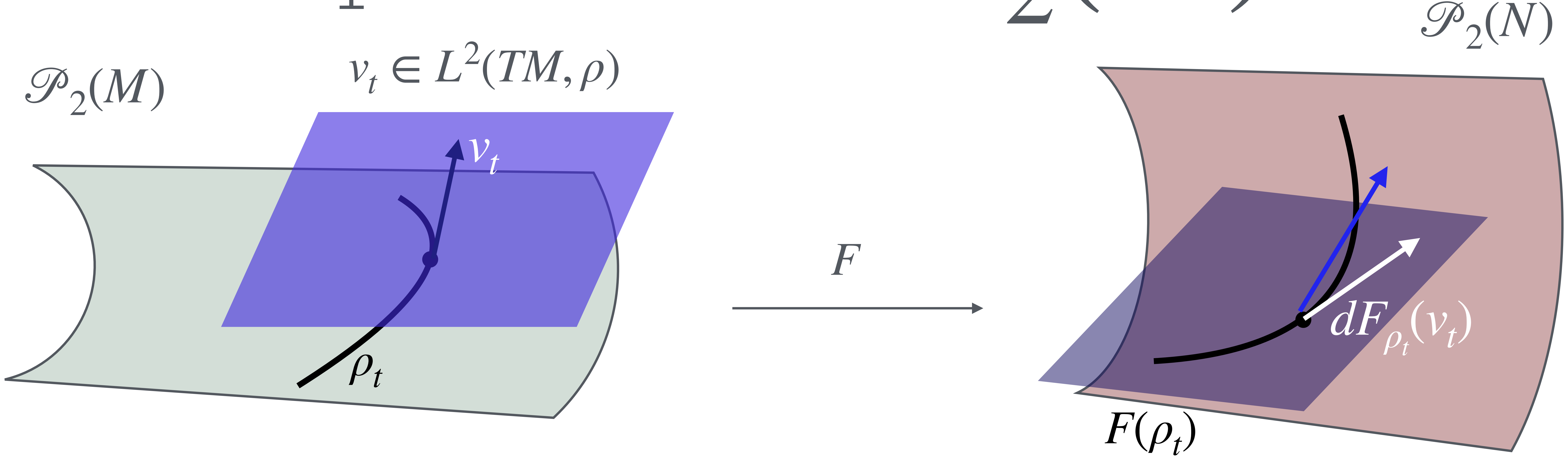


$\partial_t \rho_t + \nabla \cdot (\rho_t v_t) = 0$

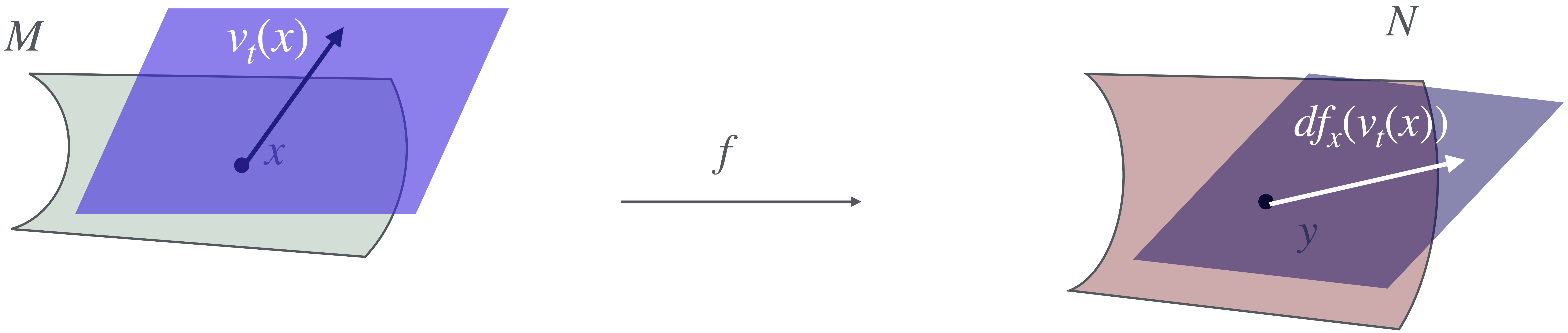
$\partial_t F(\rho_t) + \nabla \cdot (F(\rho_t) dF_{\rho_t}(v_t)) = 0$

- If $F = f\#$ then $y \to df_{f^{-1}(y)} v_t(f^{-1}(y))$ satisfies the continuity equation!
- Project to obtain a vector field in the tangent space $P_{F(\rho_t)} df_x v_x(x)$
- Define the derivative operator $dF_{\rho_t}(v_t)(y) = P_{F(\rho_t)} df_{f^{-1}(y)} v_t(f^{-1}(y))$

Derivative operator in $\mathscr{P}_2(M)$



$$dF_{\rho_t}(v_t) = P_{F(\rho_t)} df_{f^{-1}(\cdot)}(v_t(f^{-1}(\,\cdot\,))$$



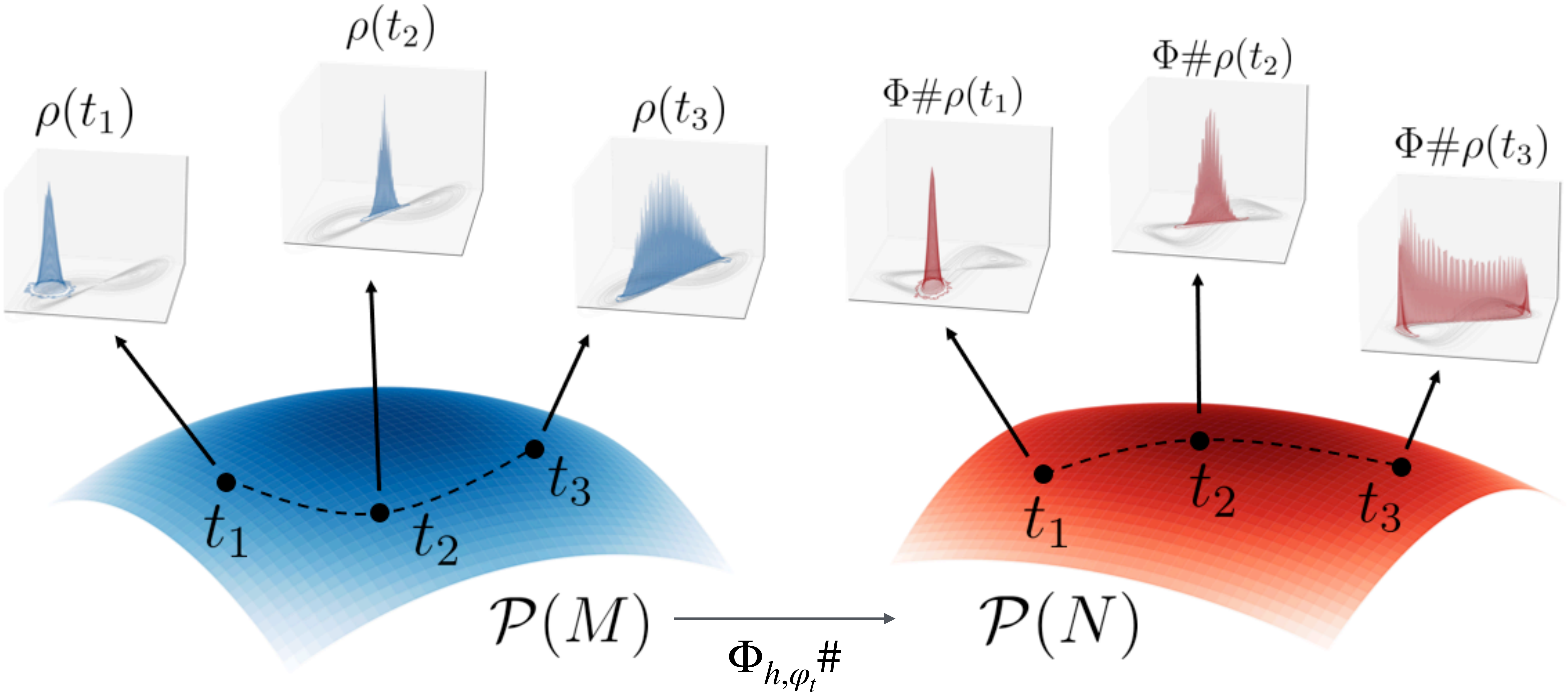
Main result

Theorem Let $f : M \to N$ be an embedding and denote by $F = f\# : \mathscr{P}_2(M) \to \mathscr{P}_2(N)$ its push-forward. Then F is an embedding with respect to the Wasserstein geometry.

- Explicit formula between df and $dF \implies$ Proprieties of f can be lifted to F

Lifted Takens Embedding

Corrolary: Let $\phi_t : M \to M$ be a dynamical system satisfying Takens conditions. Let $\Psi_{h,\phi_t} := \Phi_{h,\phi_t}\#$. Then if $d \geq n+1$ it is a generic propriety that Ψ_{h,ϕ_t} is an embedding of $\mathscr{P}(M)$ to $\mathscr{P}(\mathbb{R}^d)$.



Numerical implementation

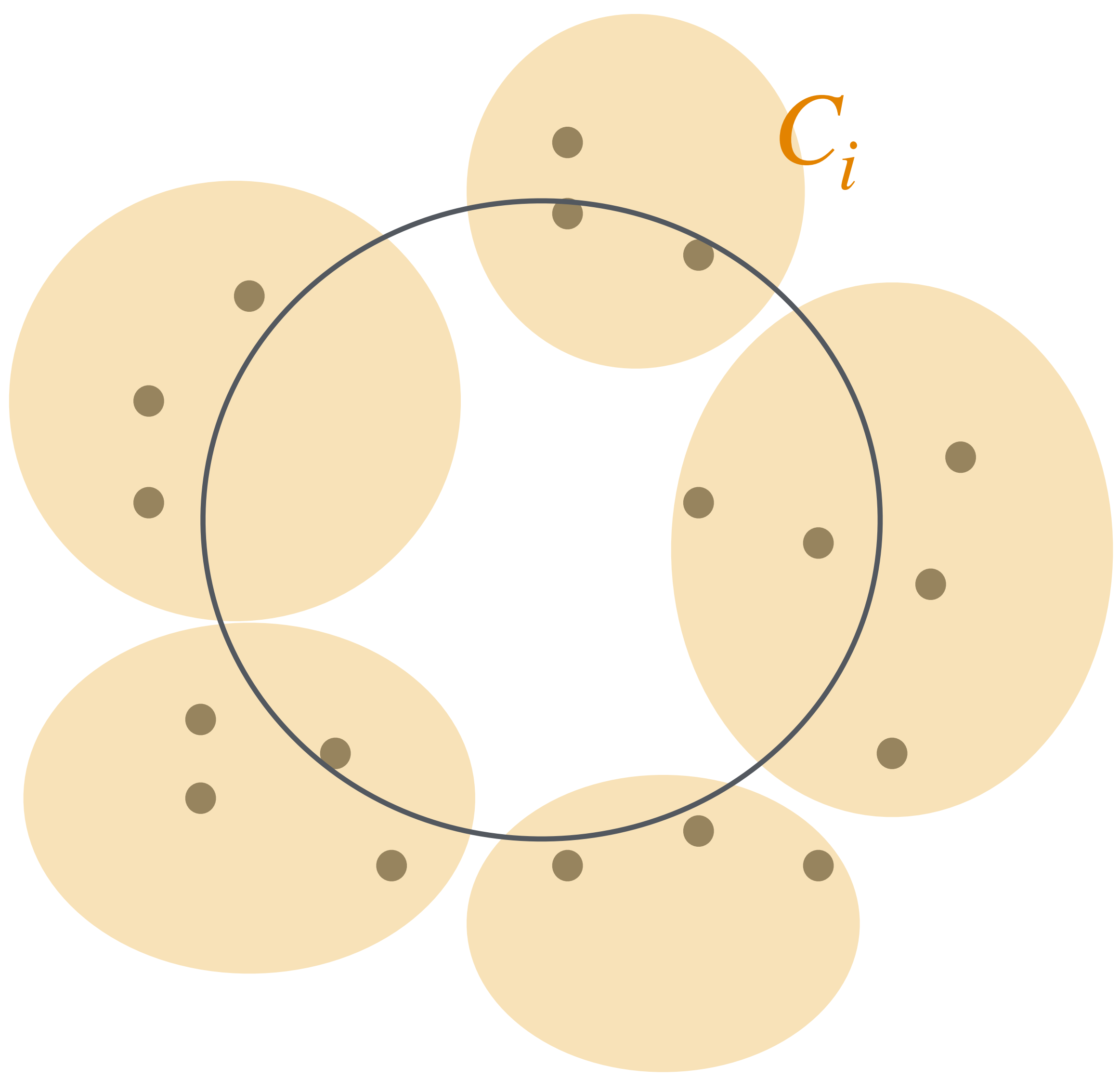
- Goal: Learn the reconstruction map i.e. Φ^{-1} from samples $\{x_i, \Phi(x_i)\}_{i=1}^N$
- Parametrize Φ^{-1} by a neural network R_θ
- Method 1: Point-wise matching $\displaystyle arg\min_\theta \sum_{i=1}^{N} \|x_i - R_\theta(\Phi(x_i)\|_2^2$
- Method 2: Measure based reconstruction $\displaystyle arg\min_\theta \sum_{i=1}^{M} \mathscr{D}(\mu_i, R_\theta \# \Phi \# \mu_i)$

  $\mathscr{D}$ = maximum mean discrepancy

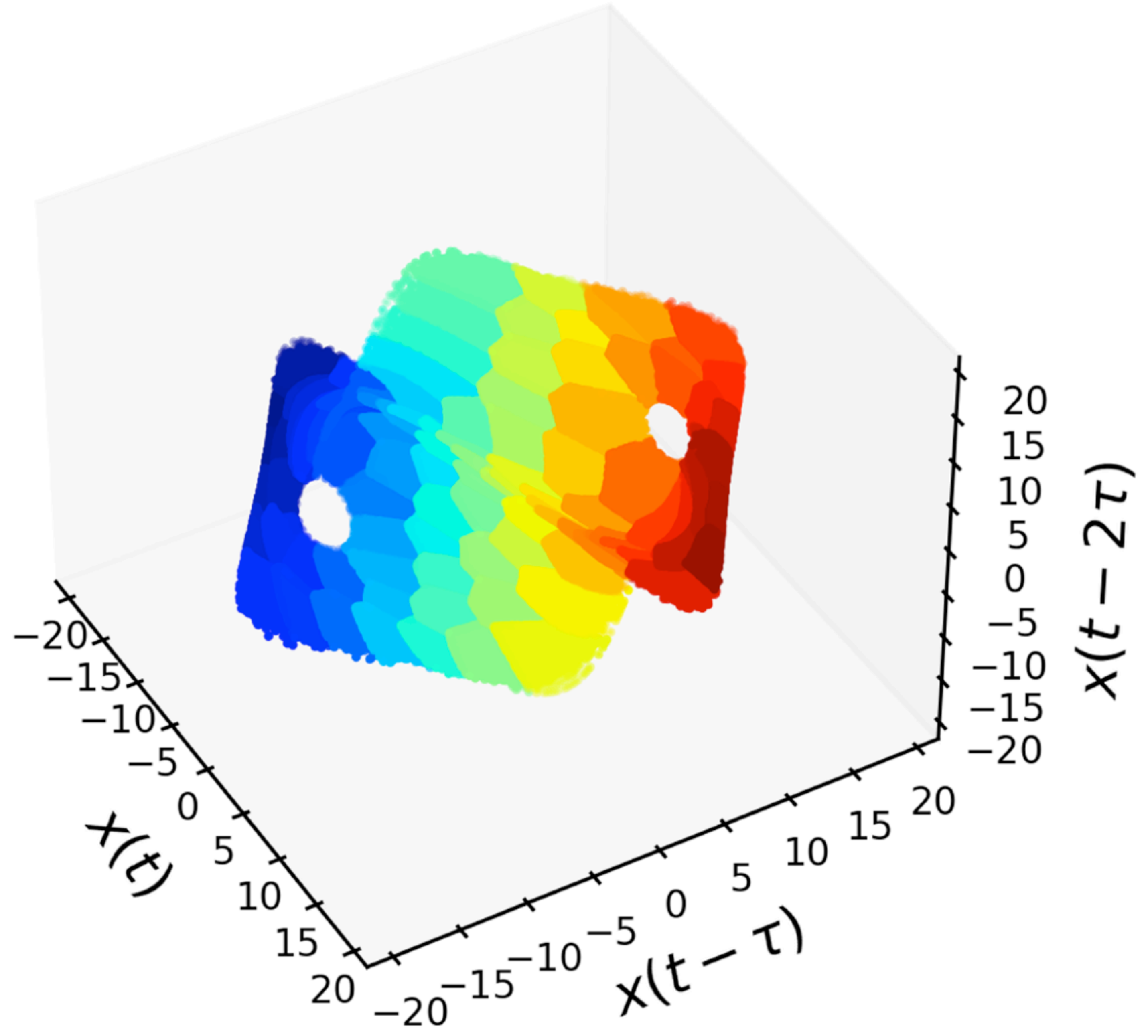
  The measures μ_i are obtained from a long trajectory by partitioning into Vornoi cells $\Longrightarrow \mu_i = \mu \,|\, C_i$

Numerical implementation

Method 2: Measure based reconstruction $\arg\min_{\theta} \sum_{i=1}^{M} \mathscr{D}(\mu_i, R_\theta \# \Phi \# \mu_i)$

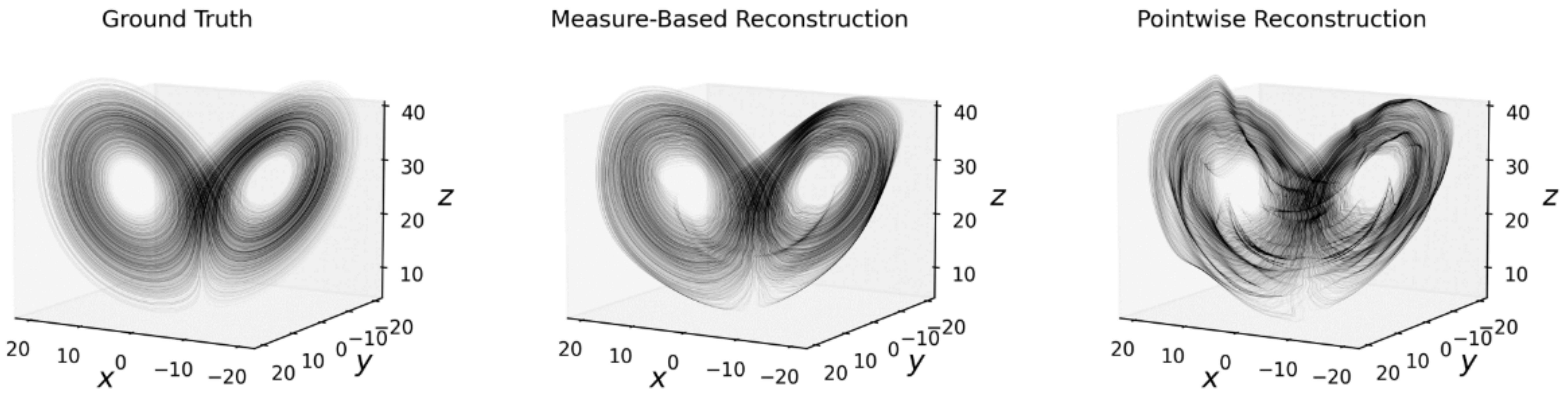


- Partition domain into $\{C_i\}_{i=1}^{M}$
- Empirical distribution conditioned on the cell $\mu_i \sim \sum_{x_j \in C_i} \delta_{x_j}$
- As diameter of $C_i \to 0$ we obtain Method 1
- Works with different data assumptions

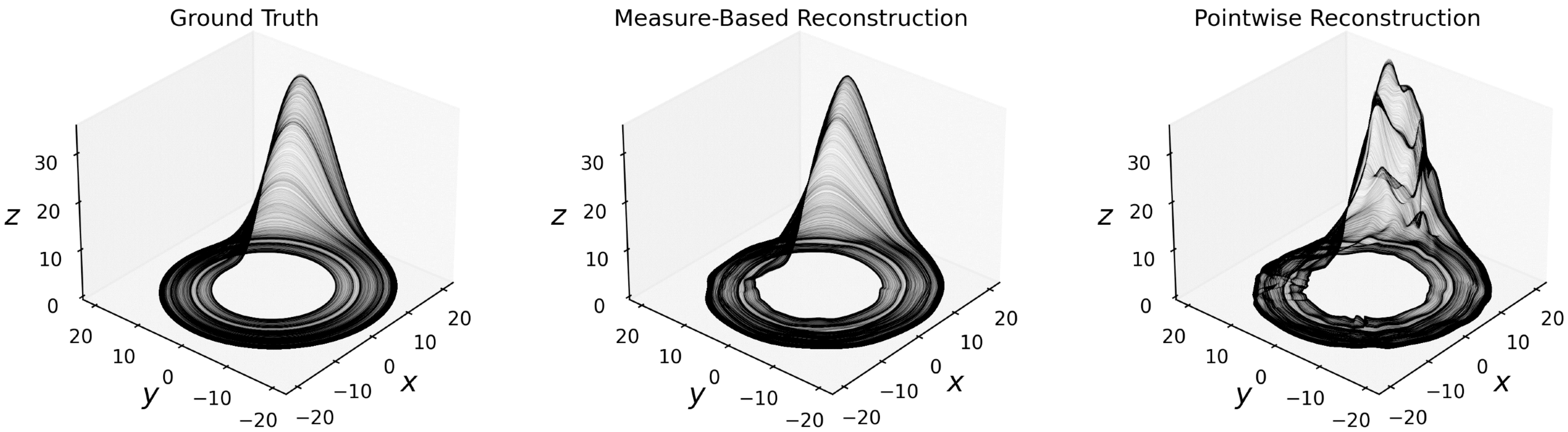


Numerical Results

Lorentz system



Roessler system



Thank you for your attention!

Contact me: mao237@cornell.edu

References: [1] Floris Takens. Detecting strange attractors in turbulence. In Dynamical Systems and Turbulence, Warwick 1980, pages 366–381. Springer, 1981

[2]Luigi Ambrosio, Nicola Gigli, and Giuseppe Savare. Gradient Flows in Metric Spaces and in the Space of Probability Measures. Birkhaeuser Verlag, 2005.

[3]Jonah Botvinick-Greenhouse, Maria Oprea, Yunan Yang and Romit Maulik, Measure-Theoretic Takens' Time-Delay Embedding:Analysis and Application, preprint, 2024

Supplimental slides/figures

- Error in the attractor reconstruction

System	Pointwise MSE	Measure MSE
Lorenz	7.15×10^{-1}	2.84×10^{-1}
Rössler	3.99×10^{-1}	8.34×10^{-2}
Lotka–Volterra	2.10×10^{-4}	9.45×10^{-5}

- Steps of the proof:

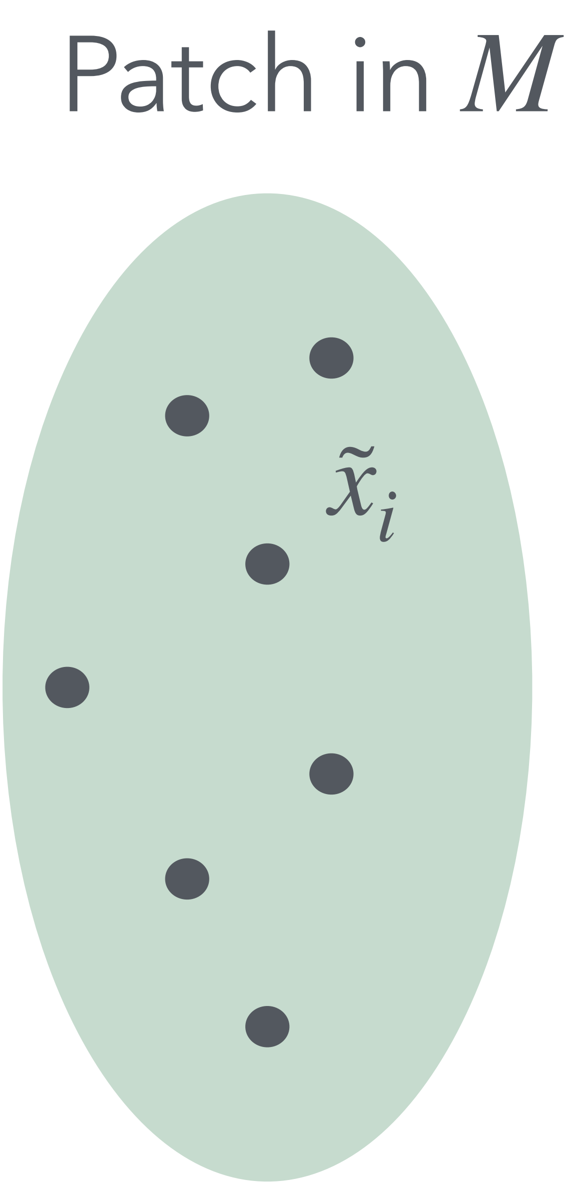
  - Show that $F(\rho_t)$ is absolutely continuous if ρ_t is

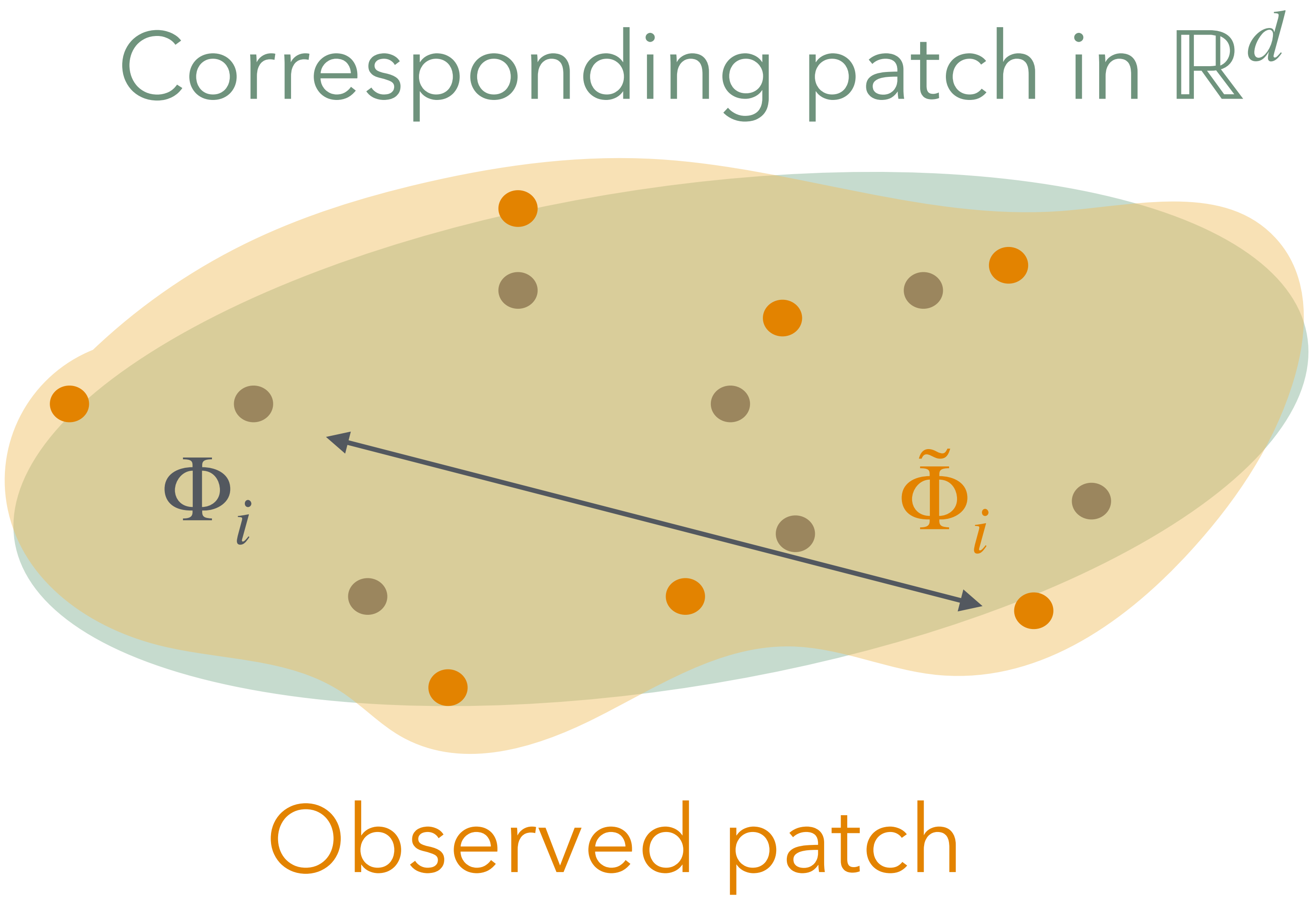
  - Show that $F(\rho_t)$ satisfies the continuity equation if ρ_t also satisfies I, with velocity field given by $DF(v_t)(y) = w_t(y) := df_{f^{-1}(y)} v_t(f^{-1}(y))$ projected onto the tangent space

  - Show that the derivative is injective i.e. $\exists \varphi$ s.t. $\int_M \langle \nabla \varphi(x), DF(v_t)(x) - DF(v_t')(x) \rangle d\rho(x) \neq 0.$

(They can only differ by a divergence free component)

- Observe $\tilde{x}_i = x_i + \epsilon_i$ then $\tilde{x}_i \sim \rho_i$
- Noisy data $\{\tilde{x}_i\} \implies$ smooth curve in $\rho_t \in \mathscr{P}_2(M)$





- Want to minimize the distance $\mathscr{D}(\mu_C, R\#\Phi\#\mu_C)$. Instead we minimize $\mathscr{D}(\mu_C, R\#\tilde{\Phi}_C)$.
  But the error is smaller than when minimizing the point wise distance