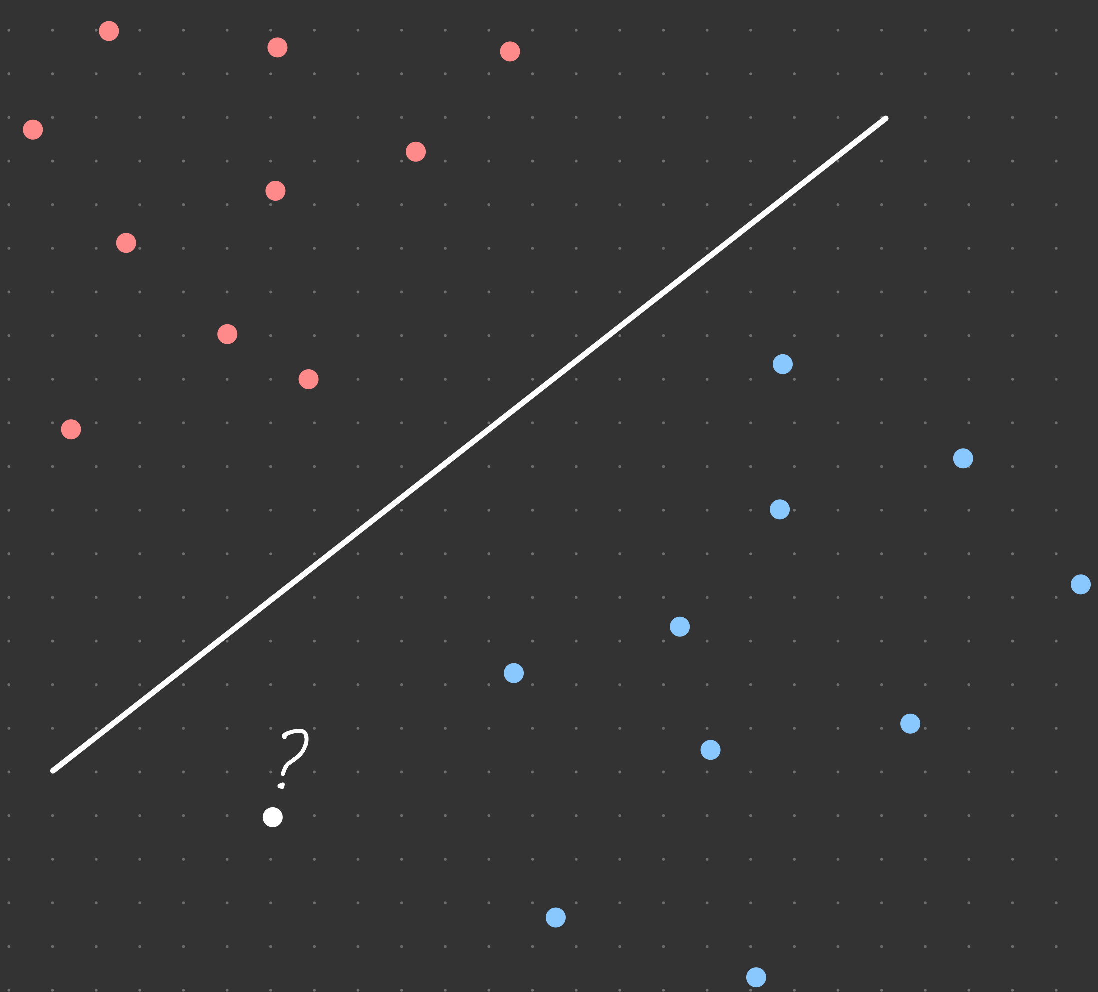
Embedding Non-linear Systems Data into a Reproducing Kernel Hilbert Space

Ben Russo

Binary Classification: Support Vector Machines are a solution to the binary classification problem with geometric origins.

We assume there exists a $w \in \mathbb{R}^d$ with $\|w\| = 1$ and $b \in \mathbb{R}$ such that

$$\langle w, x_i \rangle + b > 0 \quad \forall i \text{ with } y_i = +1$$
$$\langle w, x_i \rangle + b < 0 \quad \forall i \text{ with } y_i = -1.$$



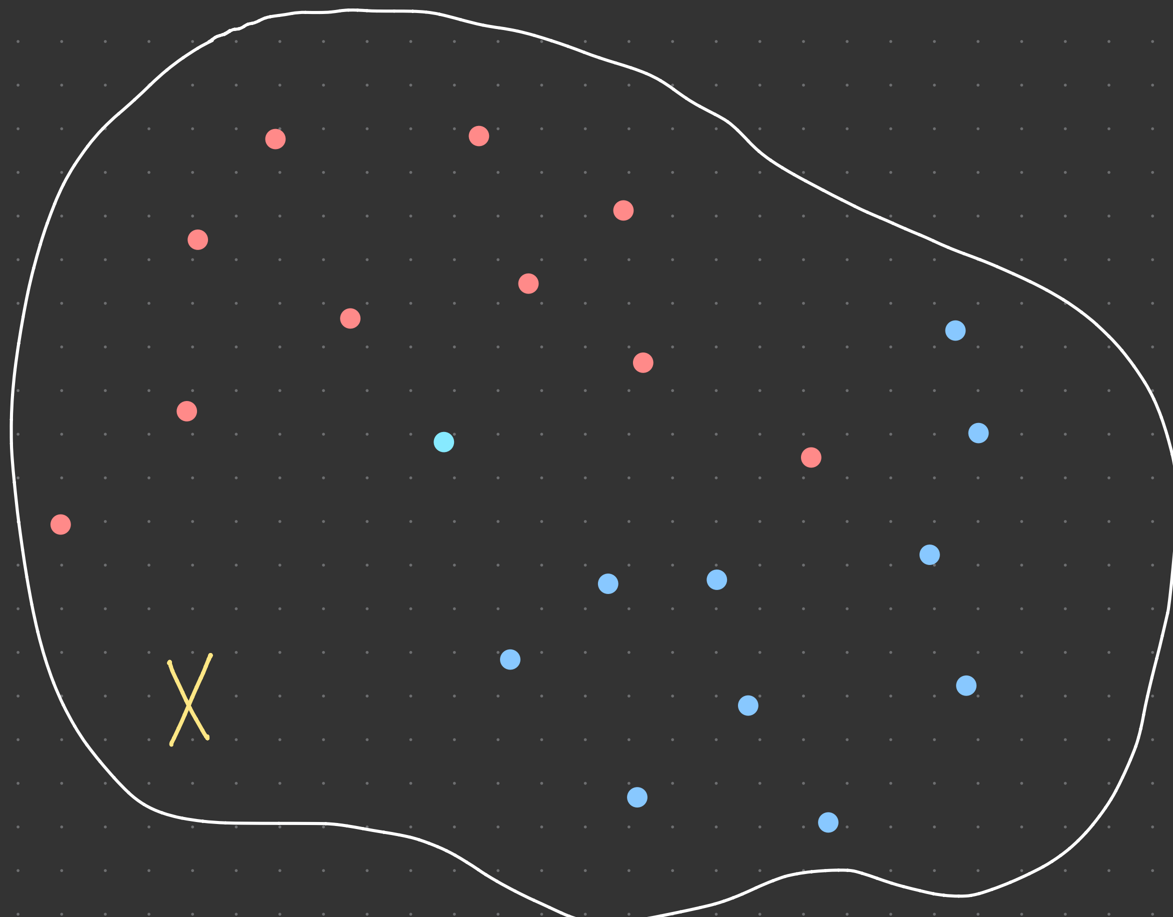
Then one can use the decision function

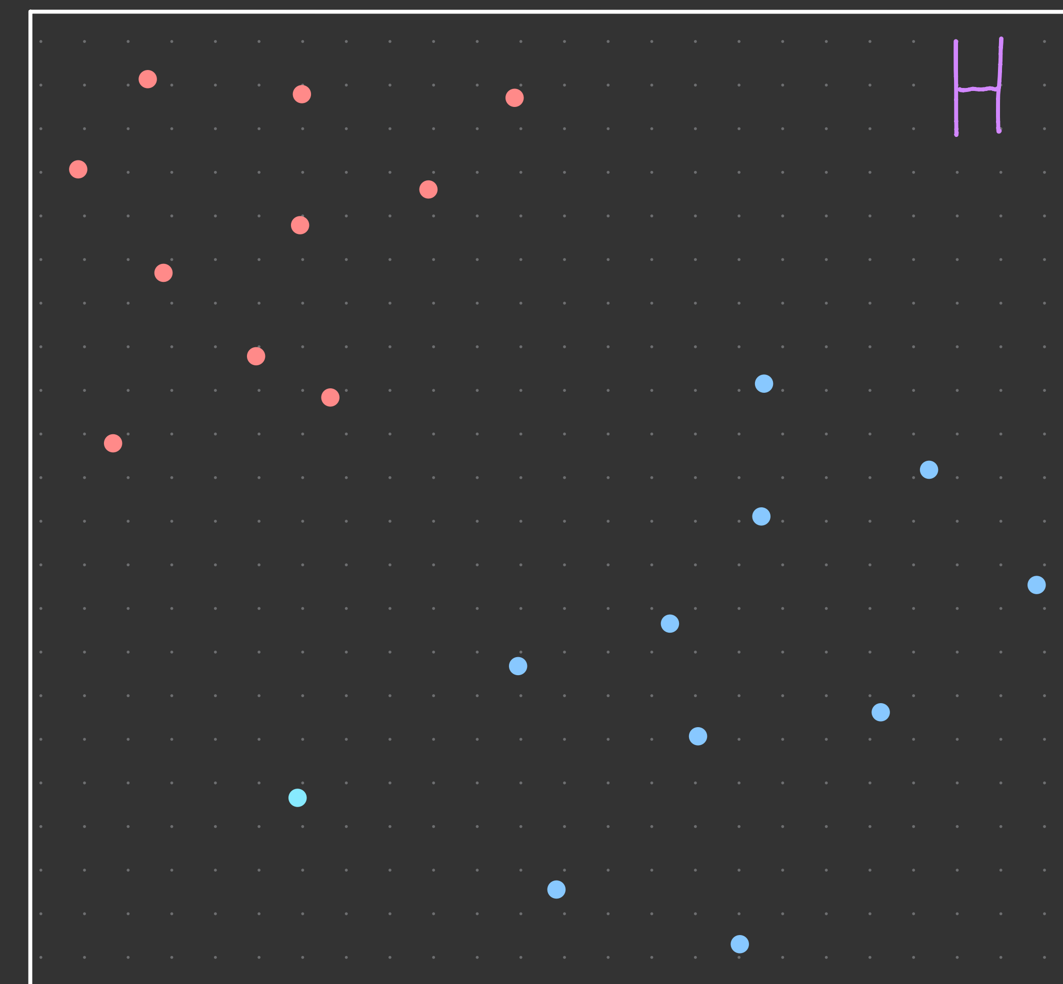
$$f_D(\cdot) = \operatorname{sign}(\langle w, \cdot \rangle + b)$$

to classify future points.

$$D = \{(x_1, y_1), \ldots, (x_n, y_n)\}$$

Binary Classification: Use a feature map $\Phi : X \longrightarrow H$ such that $\langle \Phi(x), \Phi(y) \rangle = K(x,y)$ where K is a Kernel function.





$$\min \ \langle w,w \rangle$$
subject to $y_i(\langle w, x_i \rangle + b) \geq 1.$
over $w \in \mathbb{R}^g$, $b \in \mathbb{R}$, $i = 1, \ldots, n.$

$\Phi : X \longrightarrow H$

$$\min \ \langle w,w \rangle$$
subject to $y_i(\langle w, \Phi(x_i) \rangle + b) \geq 1.$
over $w \in H$, $b \in \mathbb{R}$, $i = 1, \ldots, n.$

Definition:

A Reproducing Kernel Hilbert Space (RKHS) over a set X is a Hilbert space of $\mathbb{R}(\mathbb{C})$-valued functions such that $e_x(g) = g(x)$ is a continuous linear functional for all $x \in X$.

As such, Riesz representation guarantees for all $x \in X$ a function K_x such that $f(x) = \langle f, K_x \rangle$ for all $f \in H$.
The Kernel function at x will be denoted $K_x(y) = K(y,x)$.

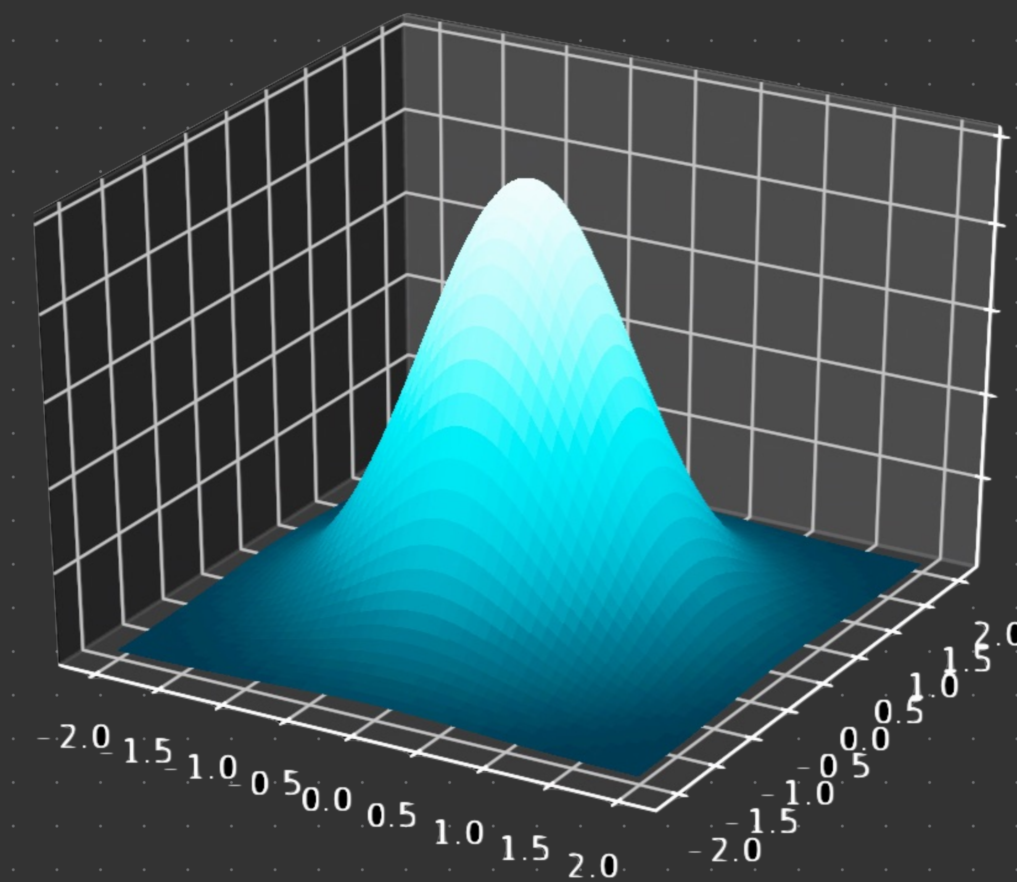
Theorem: Suppose K is a symmetric, positive definite Kernel on a set X. Then there is a unique Hilbert space with K as its Kernel.

For a given holomorphic function $f:\mathbb{C}^d \longrightarrow \mathbb{C}$ define

$$\|f\|_\gamma := \left(\frac{2^d}{\pi^d\gamma^{2d}} \int_{\mathbb{C}} |f(z)|^2 \exp\left(\gamma^{-2}\sum_{j=1}^d (z_j - \bar{z}_j)^2\right) dz\right)^{1/2}$$

$$H_\gamma = \{ f:\mathbb{C}^d \longrightarrow \mathbb{C} : f \text{ holomorphic } \|f\|_\gamma < \infty \}$$

is the RKHS with $K(z,z') = \exp\left(-\gamma^{-2}\sum_{j=1}^d (z_j - \bar{z}_j')^2\right)$.

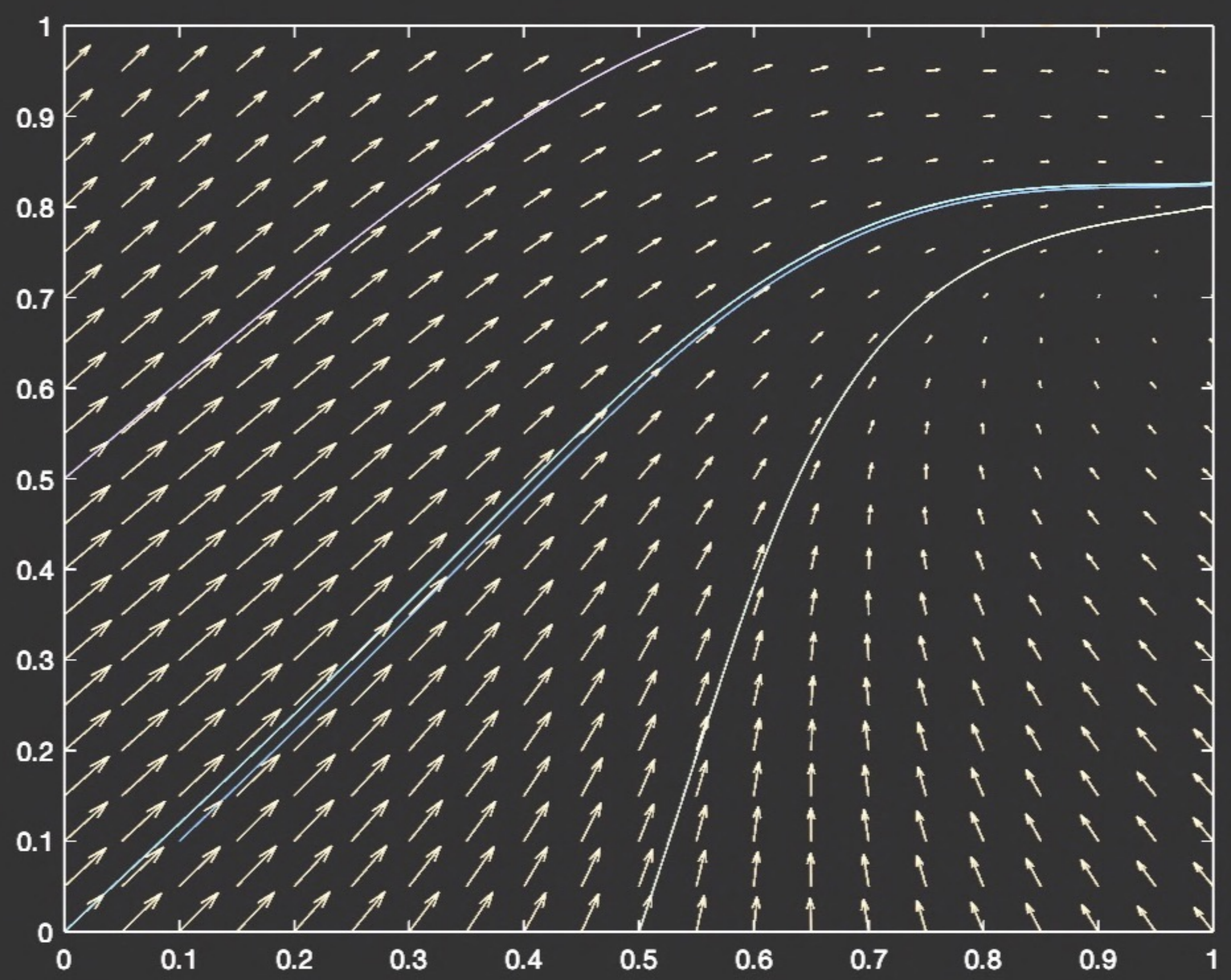


For the real version we take

$$H_\gamma^{\mathbb{R}} := \{ f: X \subseteq \mathbb{R}^d \longrightarrow \mathbb{R}^d \mid \exists\, g \in H_\gamma \text{ with } \operatorname{Re} g|_X = f \}$$

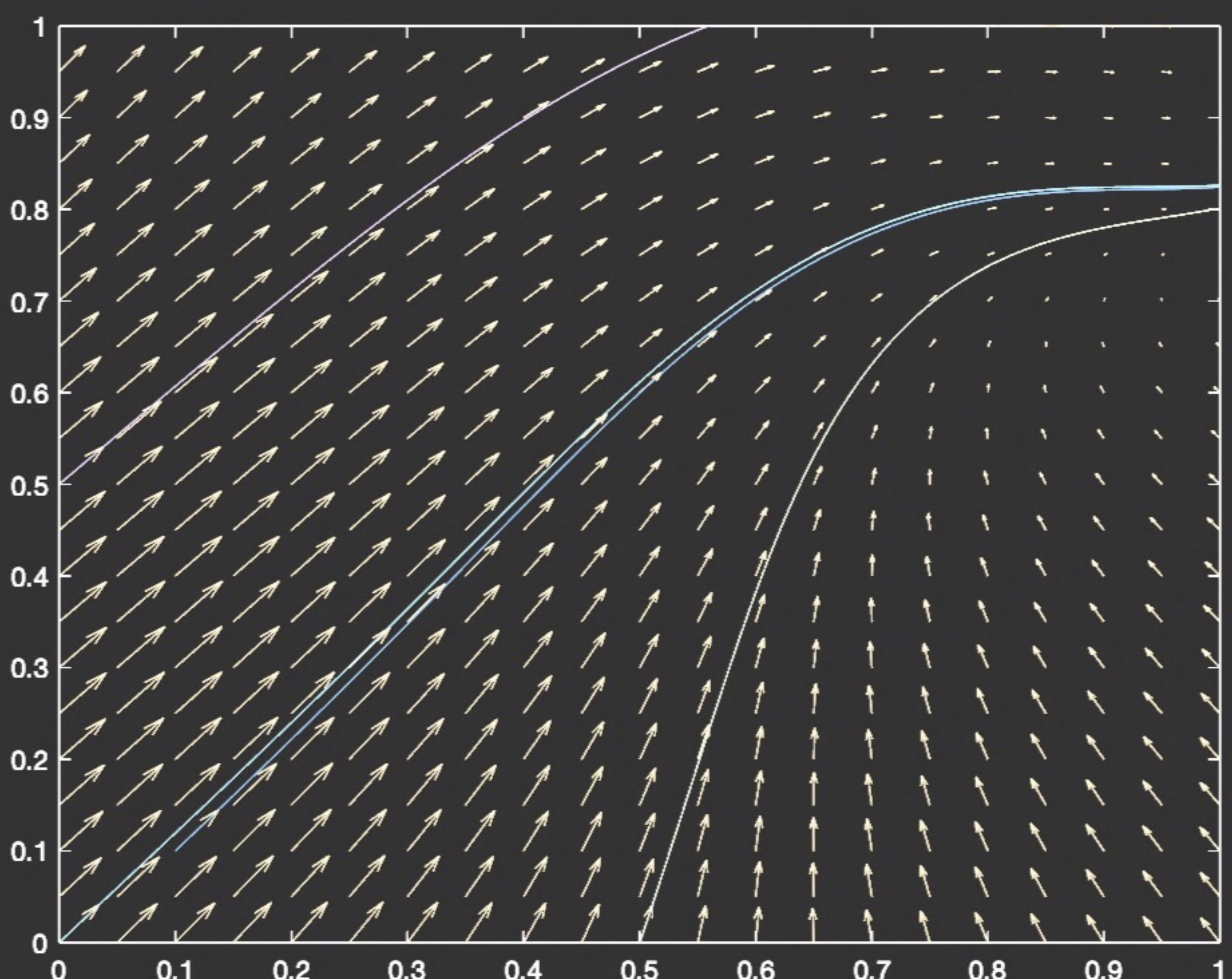
$$\|f\|_{H_\gamma^{\mathbb{R}}} = \inf\{ \|g\|_\gamma : g \in H_\gamma \text{ with } \operatorname{Re} g|_X = f \}$$

Dynamical System: $\dot{x}(t) = f(x)$; $f: \mathbb{R}^n \longrightarrow \mathbb{R}^n$

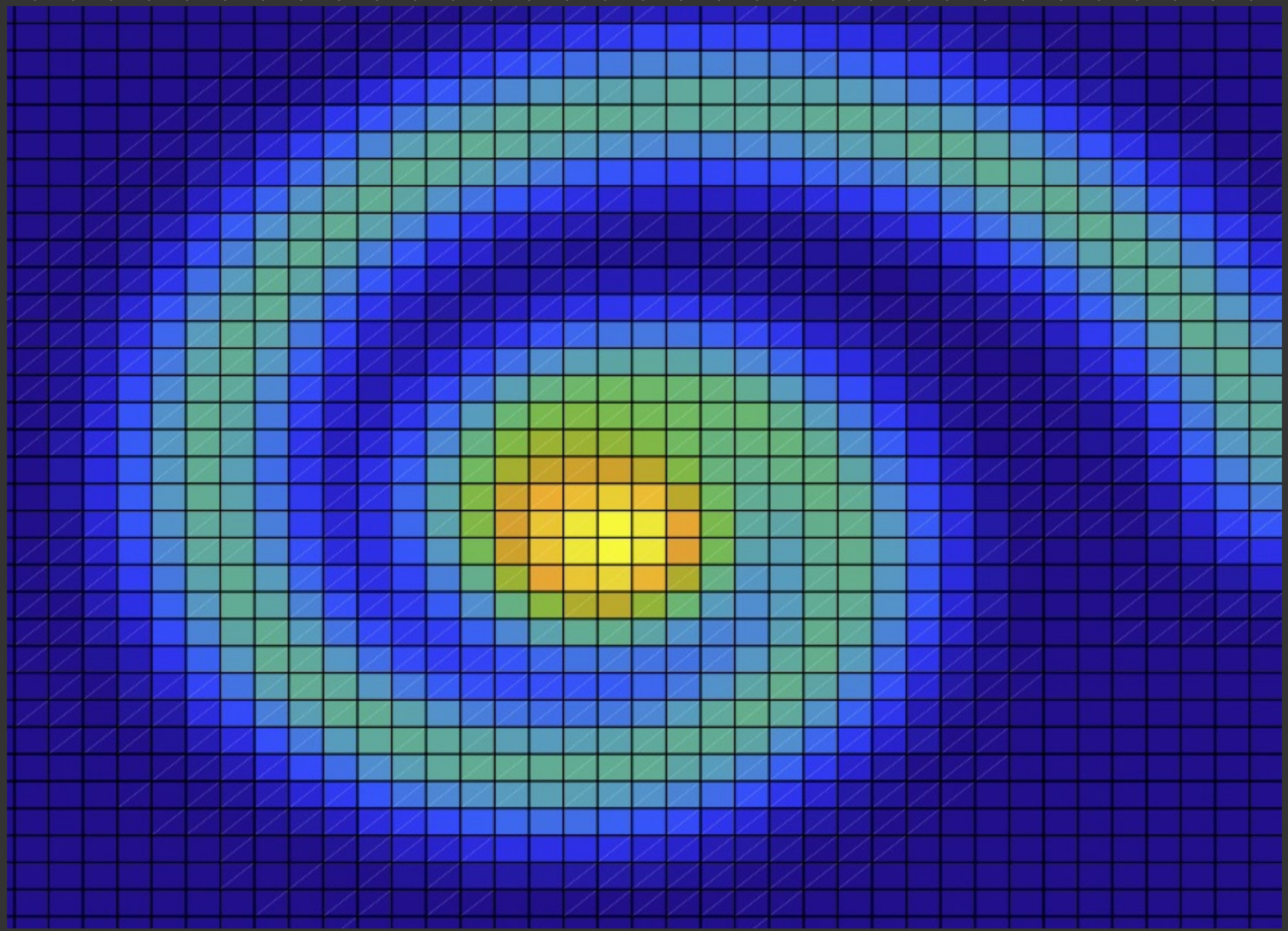


We encode trajectories with occupation kernels and the dynamics via Liouville operators.

Dynamical System: $\dot{x}(t) = f(x)$; $f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$



Definition: Given a function $f : X \longrightarrow \mathbb{R}^n$ and H a RKHS over the set X we define the Liouville operator with symbol f as $A_f : \mathcal{D}(A_f) \longrightarrow H$ given by $A_f(g) := \nabla_x g \cdot f$, where $\mathcal{D}(A_f) := \{ g \in H : \nabla_x g \cdot f \in H \}$.



heat map of an occupation kernel for a spiral trajectory.

Definition: Let $X \subset \mathbb{R}^n$ be compact, H be a RKHS of continuous functions over X, and $\gamma : [0,T] \longrightarrow X$ be a continuous trajectory. The functional

$$g \longmapsto \int_0^T g(\gamma(t))\,dt$$

is bounded and may be represented as $\int_0^T g(\gamma(t))\,dt = \langle g, \Gamma_\gamma \rangle_H$.

Here we call the function, $\Gamma_\gamma \in H$, the occupation kernel for $\gamma(t)$.

Occupation Kernels:

- Occupation Kernels have representation in terms of the Kernel

  Given $\gamma:[0,T]\longrightarrow\mathbb{R}^n$, $\Gamma_\gamma(x)=\int_0^T K(x,\gamma(t))\,dt$.

- If we define $\Gamma_{\gamma,\tau}:=\int_0^\tau K(x,\gamma(t))dt$ for $\tau\in[0,T]$ then the mapping $t\mapsto\Gamma_{\gamma,t}$ is continuous in the Hilbert space norm.

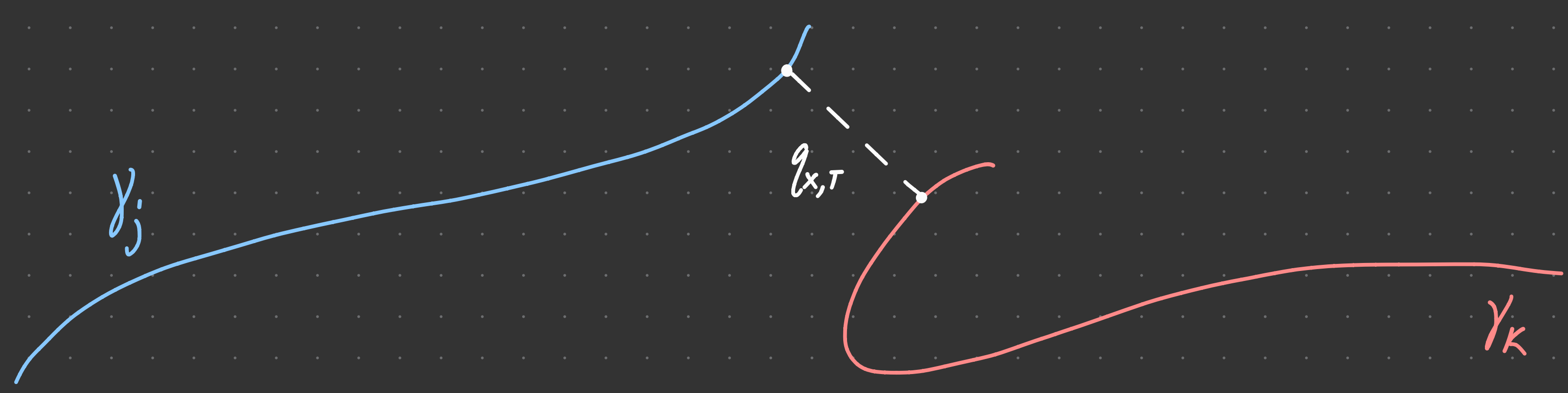
- If there exists a homotopy $\{\gamma_s\}$ between two paths then the mapping $s\mapsto\Gamma_{\gamma_s}$ is continuous in the Hilbert space norm.

Occupation Kernels:

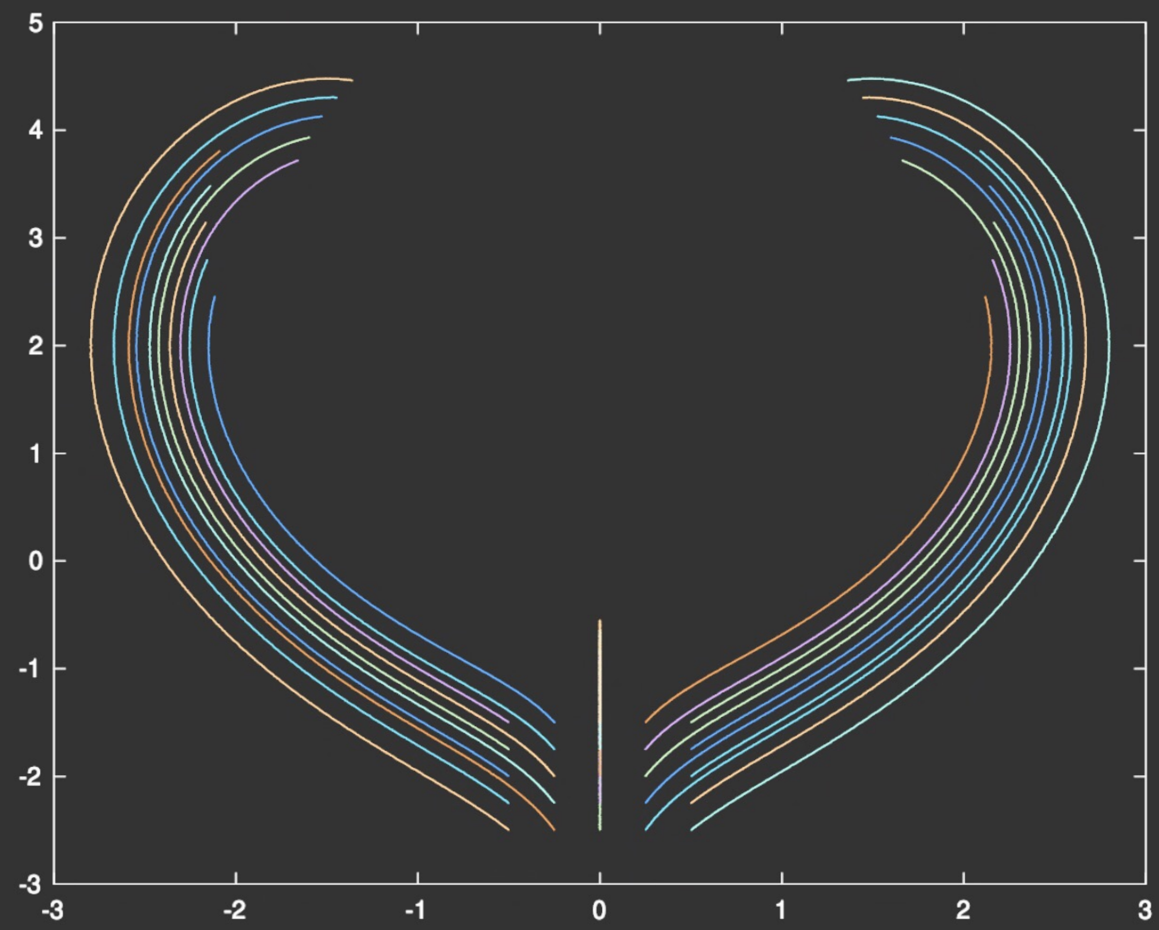
- Since linear combinations of Kernel functions are dense in the RKHS then linear combinations of occupation Kernels are as well. In fact, take $\gamma:[0,T]\to\mathbb{R}^n$; $\gamma(t):=x$ then $\Gamma_\gamma = \int_0^T K(y,x)\,dt = T\cdot K_x(y)$.

- Letting $G_\Gamma = (\langle \Gamma_{\gamma_i}, \Gamma_{\gamma_j} \rangle)_{i,j=1,1}^{N,N}$ be the occupation Kernel Gramian for a set of paths $\{\gamma_j\}_{j=1}^N$, if $K(x,x') = \exp(-\mu\|x-x'\|_2^2)$ then the condition number $:= \frac{\lambda_{max}}{\lambda_{min}}$ estimates for G_Γ are comparable to the condition number estimates for Gramians for Gaussian RBFs.

- The λ_{min} is dependent on the path separation distance $q_{x,T} := \frac{1}{2} \min_{\substack{j\neq k \\ t,\tau\in[0,T]}} \|\gamma_j(t) - \gamma_k(\tau)\|_2$



System Identification:

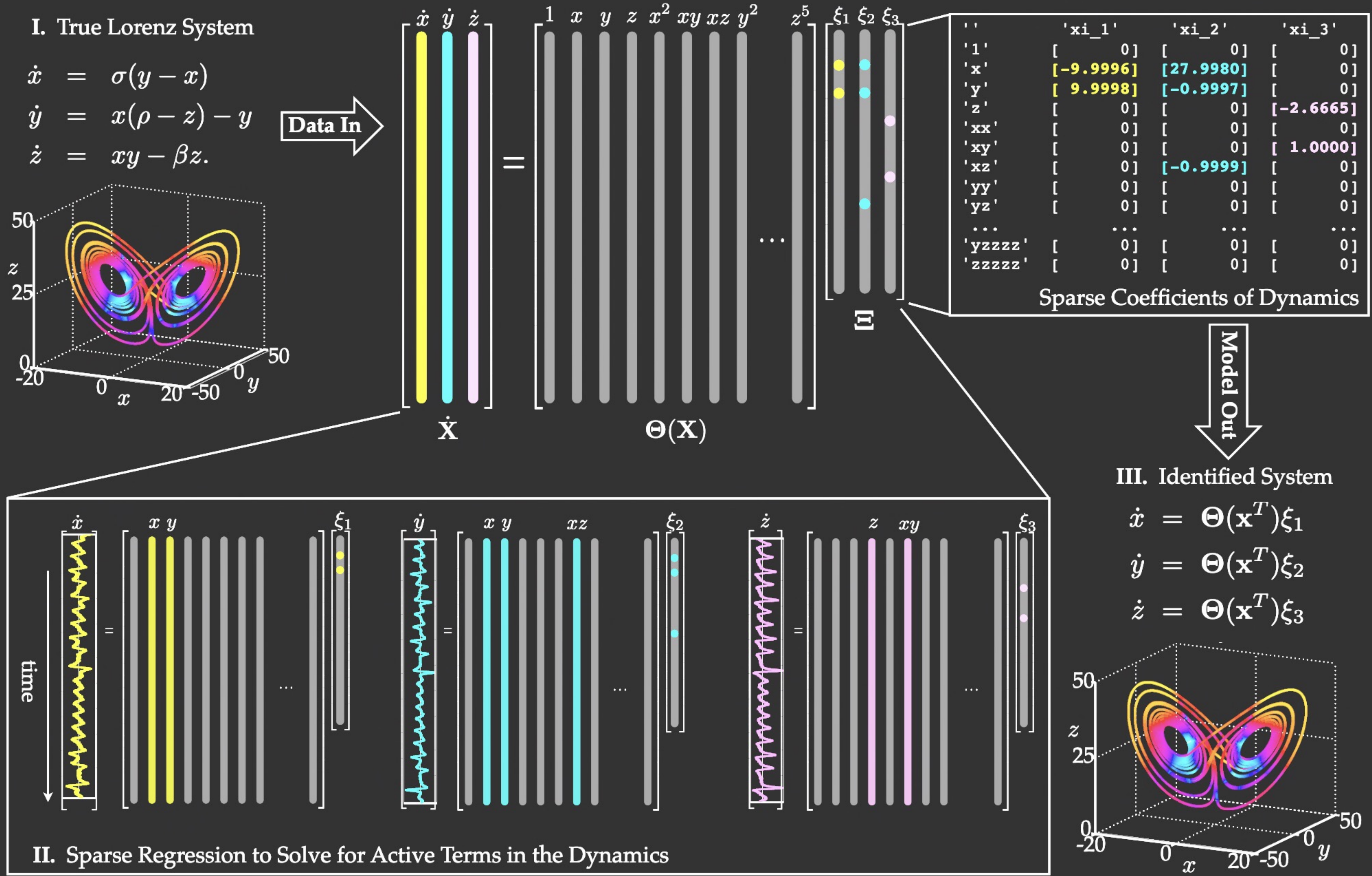


Observed data

System id techniques

$$\dot{X}(t) = \begin{pmatrix} 2X_1 - X_1 X_2 \\ 2X_1^2 - X_2 \end{pmatrix}$$

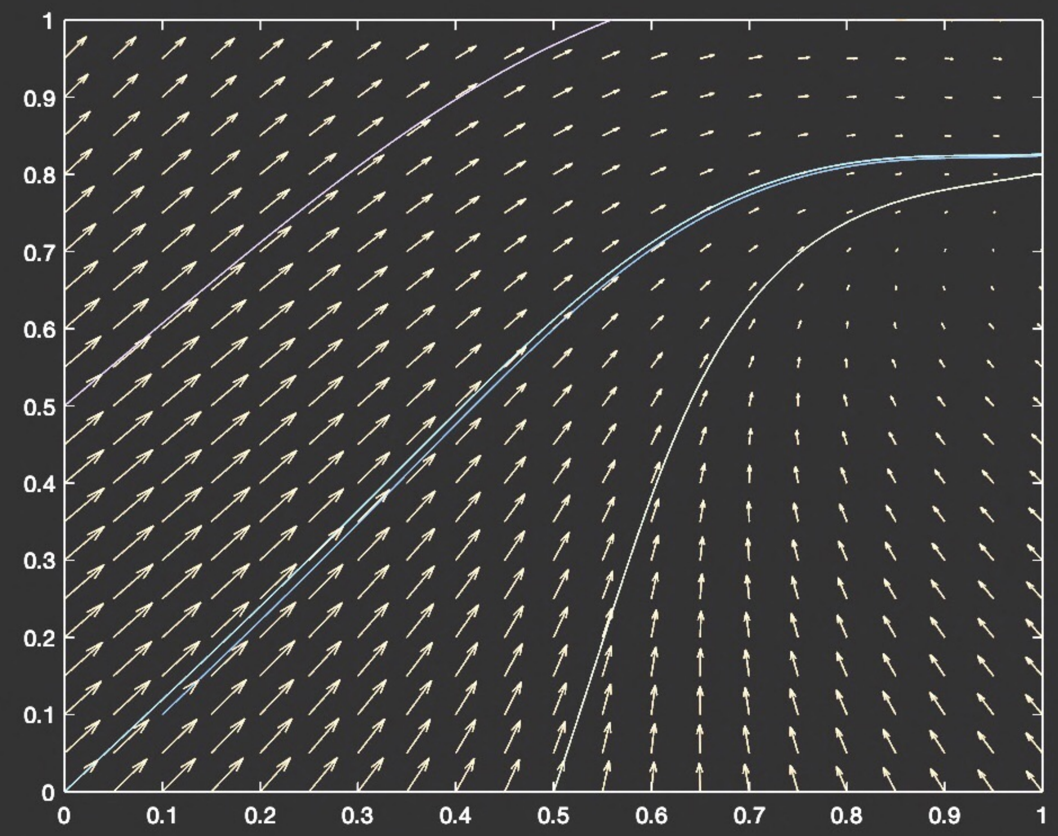
The SINDy Algorithm



Brunton et.al.

Integrating Non-linear systems data via Occupation Kernels

Suppose $\gamma(t)$ is a trajectory satisfying the dynamics $\dot{\gamma}(t) = f(\gamma(t))$



$$\int_0^T \nabla g(\gamma(t)) f(\gamma(t))\, dt = g(\gamma(t)) - g(\gamma(0))$$

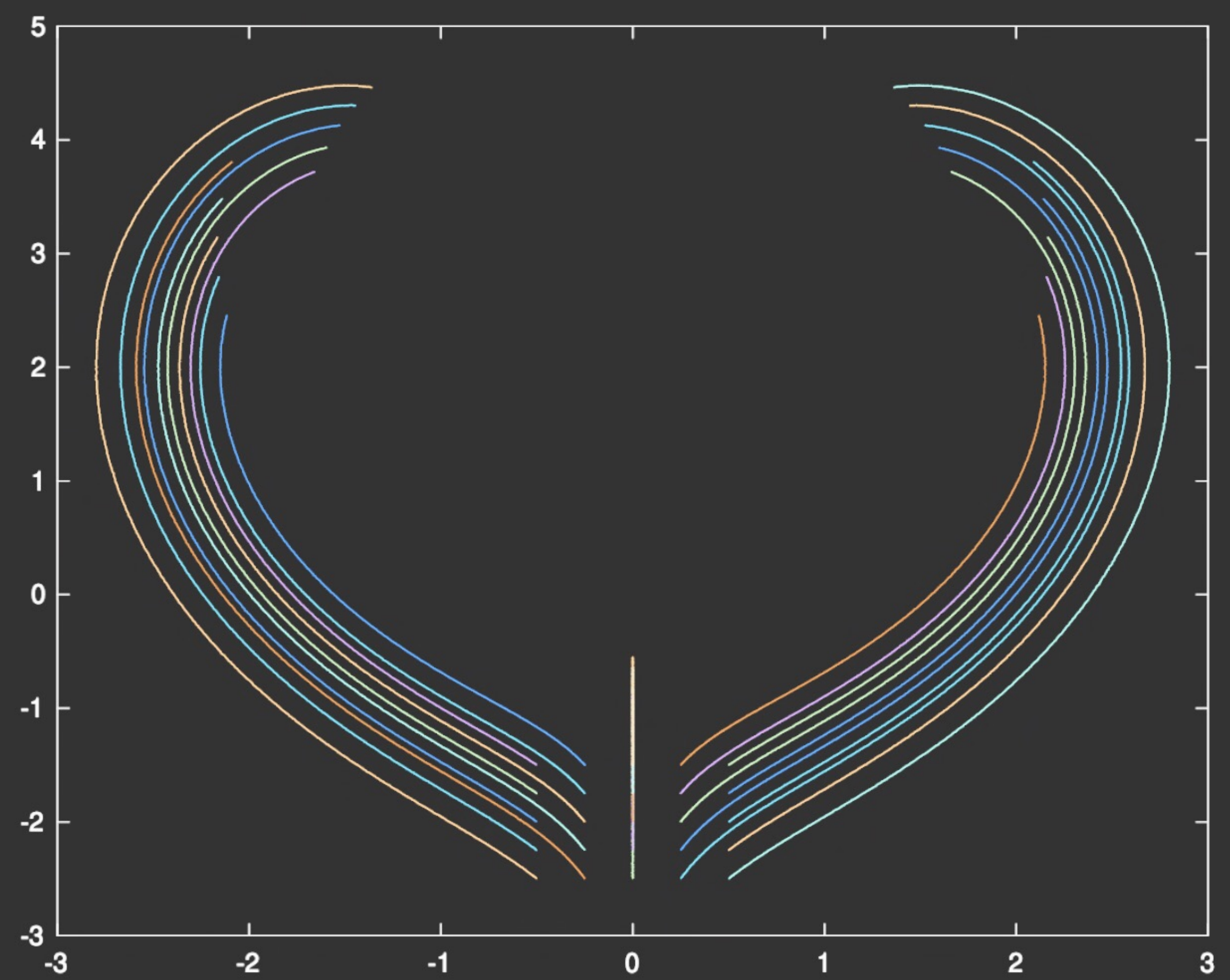
$$\langle A_f g, \Gamma_\gamma \rangle = g(\gamma(t)) - g(\gamma(0))$$

For a compact set $X \subset \mathbb{R}^n$, let $\{\gamma_j : [0,T] \longrightarrow X\}_{j=1}^{N}$ be a collection of trajectories satisfying the dynamics $\dot{x} = f(x) = \sum_{i=1}^{M} \theta_i Y_i(x)$ and let Γ_{γ_j} be the corresponding occupation kernels inside a RKHS, H, of continuously differentiable functions over X. Suppose that $\{c_s\}_{s=1}^{\infty} \subset X$ is dense. Constraints on θ_i are established as

$$\langle A_f K(\cdot, c_s), \Gamma_{\gamma_j} \rangle_H = \sum_{i=1}^{M} \theta_i \langle A_{Y_i} K(\cdot, c_s), \Gamma_{\gamma_j} \rangle_H = K(\gamma_j(T), c_s) - K(\gamma_j(0), c_s)$$

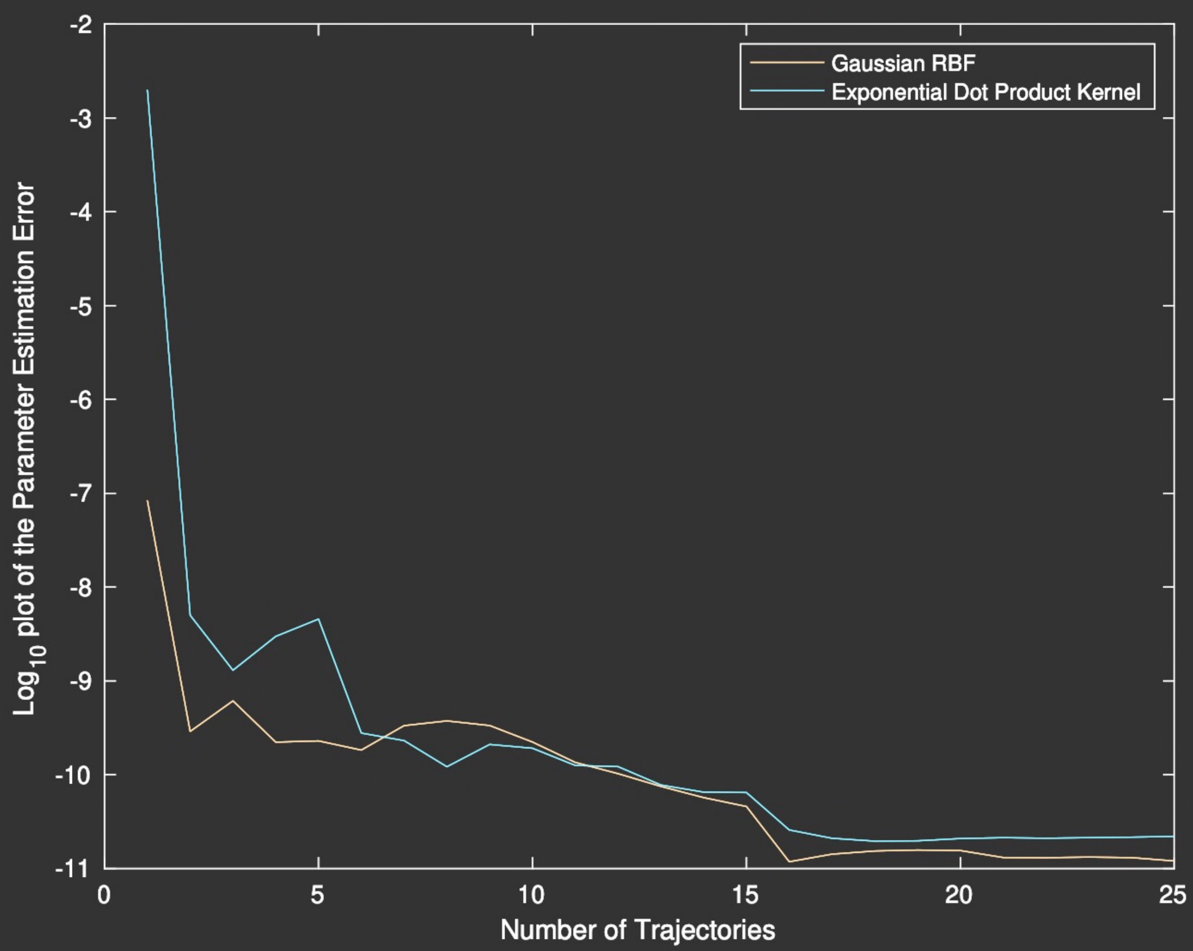
for each $s = 1, 2, 3, \ldots,\ j = 1, 2, \ldots N$.

$$\langle A_{Y_\ell} K(\cdot, c_s), \Gamma_{\gamma_j} \rangle_H = \int_0^T \nabla K(\gamma_j(t), c_s)\, Y_\ell(\gamma_j(t))\, dt$$



$$\dot{X}(t) = \begin{pmatrix} 2X_1 - X_1 X_2 \\ 2X_1^2 - X_2 \end{pmatrix}$$

#Trajectories v.s. Error



Robustness to Noise

Methods like SINDy require error prone numerical derivative estimates.

However under mild smoothness conditions the error introduced by a bounded zero mean disturbance $\varepsilon \in \mathcal{L}^2([0,T],\mathbb{R}^n)$ is $O(T \cdot \sigma(\varepsilon))$ for the occupation kernel method.

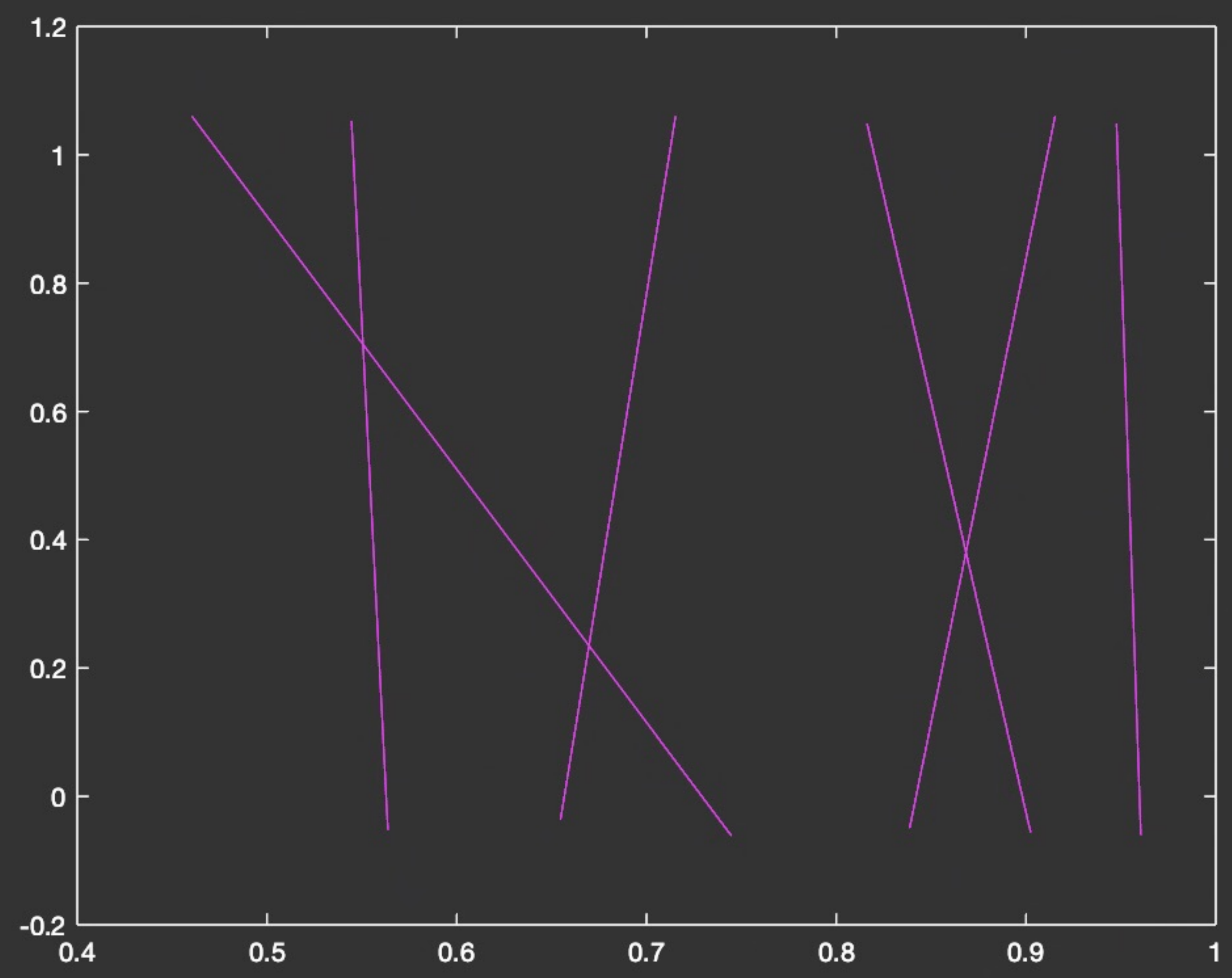
$$\left| \langle A_{Y_\ell} K(\cdot, c_s), \Gamma_{\gamma_j} \rangle_H - \langle A_{Y_\ell} K(\cdot, c_s), \Gamma_{\gamma_j+\varepsilon} \rangle_H \right| \leq M \cdot \int_0^T \| \gamma_j(t) + \varepsilon(t) - \gamma_j(t) \| \, dt$$

$$\leq M \cdot T \cdot \sigma(\varepsilon)$$

Motion Tomography

Goal: Reconstruct a vector field using its accumulated effects on mobile sensing units as they travel through the field.

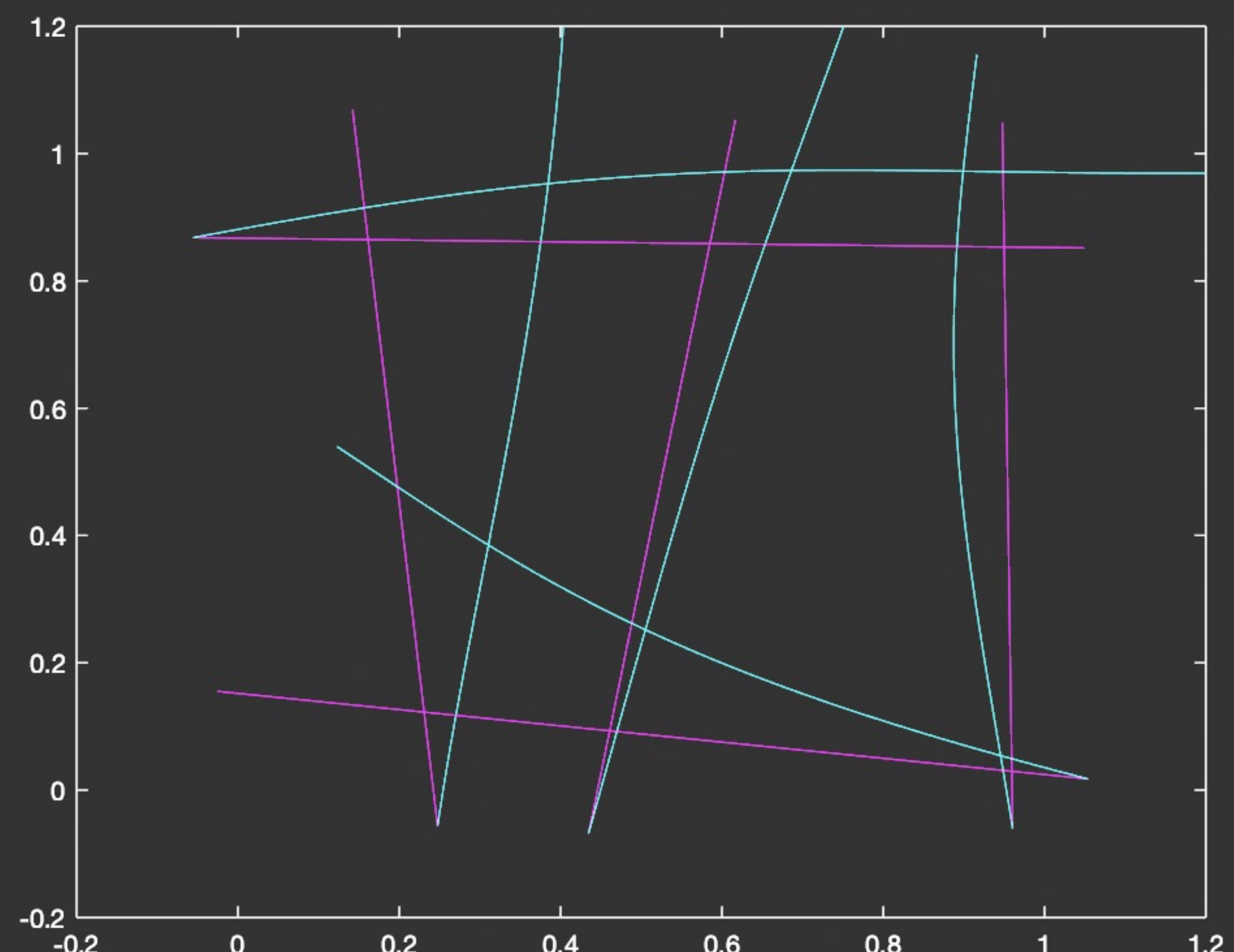
We have anticipated straight line paths.



Motion Tomography

Goal: Reconstruct a vector field using its accumulated effects on mobile sensing units as they travel through the field.

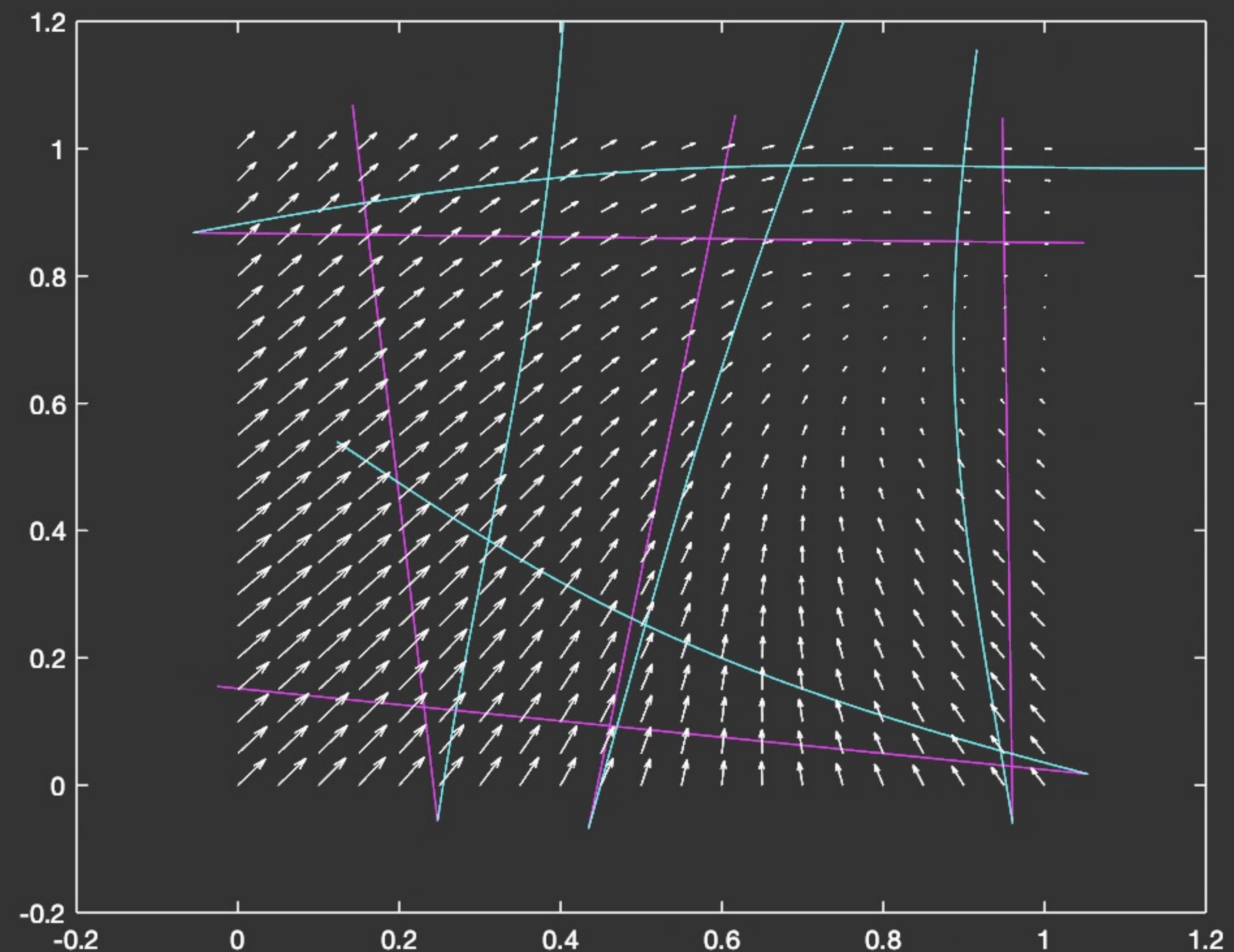
The vehicles are then influenced by the flow-field and stray from the anticipated paths.



Motion Tomography

Goal: Reconstruct a vector field using its accumulated effects on mobile sensing units as they travel through the field.

From this infomation we aim to construct the vector field.



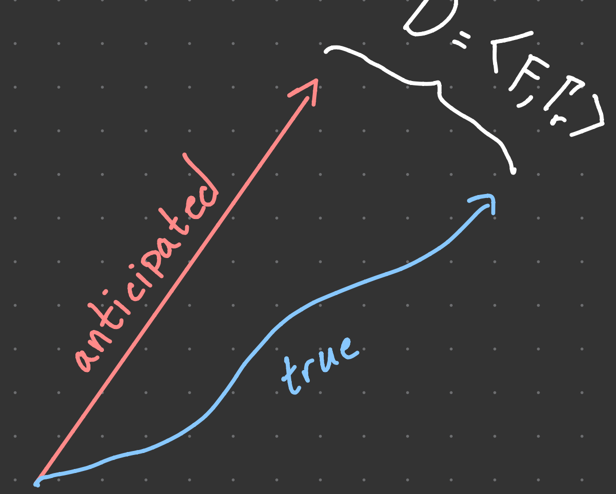
Let $r:[0,T]\longrightarrow\mathbb{R}^2$ represent the continuous path for a mobile sensor attempting to travel in a straight line but subject to an unknown flow-field $F:R\subseteq\mathbb{R}^2\longrightarrow\mathbb{R}^2$, where R is a compact subset.

Let $\dot{r}(t)=s(\cos(\theta),\sin(\theta))^T+F(r)$, with $s>0$ and F locally Lipschitz, represent the true dynamics.

As the flow field is unknown the anticipated dynamics are given as $\dot{\tilde{r}}=s(\cos(\theta),\sin(\theta))^T$

After a vehicle has traveled through the flow-field over a time period $[0,T]$, during which there is little to no knowledge of the vehicles position the difference between the actual location, $r(T)$, and anticipated location, $\tilde{r}(T)$, is given as

$$D=r(T)-\tilde{r}(T)=\int_0^T(\dot{r}(t)-\dot{\tilde{r}}(t))\,dt=\int_0^T F(r(t))\,dt=\langle F,\Gamma_r\rangle_H$$



Let $\{s_i\}_{i=1}^{M}$, $\{\theta_i\}_{i=1}^{M}$ be a collection of speeds and angles used to generate the anticipated trajectories.

After the initial data collection no further experiments are needed to approximate the flowfield.

Since, representation of the flow field is in terms of occupation Kernels we have access to the approximation powers of RKHS's.

Algorithm 1: Iterative Algorithm

Define *N as the number of iterates*
Input *Samples* $r_i(T) \quad i \in \{1, \dots, M\}$
begin

 Generate *via a numerical method* $\tilde{r}_{i,0} : [0,1] \to R$, *the unique solution to*

$$\dot{p} = s_i \begin{pmatrix} \cos(\theta_i) & \sin(\theta_i) \end{pmatrix}^\top, \quad p(0) = p_i$$

 for $i \in \{1, \dots, M\}$
 Set $D_{i,0} = r_i(T) - \tilde{r}_{i,0}(T) \quad i \in \{1, \dots, M\}$
 Set $\hat{F}_{-1} = 0$

for *n in $\{0, \dots, N\}$* **do**

 Input $\hat{F}_n = \sum_{i=1}^{M} w_{i,n} \Gamma_{\tilde{r}_{i,n}}$
 Generate *via a numerical method* $\tilde{r}_{i,n+1} : [0,1] \to R$ *the unique solution to*

$$\dot{p} = s_i \begin{pmatrix} \cos(\theta_i) & \sin(\theta_i) \end{pmatrix}^\top + \hat{F}_n(p), \quad p(0) = p_i$$

 for $i \in \{1, \dots, M\}$.
 Set $D_{i,n+1} = r_i(T) - \tilde{r}_{i,n+1}(T) \quad i \in \{1, \dots, M\}$
 Compute $\hat{F}_{n+1} = \sum_{i=1}^{M} w_{i,n+1} \Gamma_{\tilde{r}_{i,n+1}}$ *by solving*

$$\left(\langle \Gamma_{\tilde{r}_{i,n+1}}, \Gamma_{\tilde{r}_{j,n+1}} \rangle\right)_{i,j=1}^{M,M} \begin{pmatrix} w_{1,n+1} \\ \vdots \\ w_{M,n+1} \end{pmatrix} = \begin{pmatrix} D_{1,n+1} + \langle \hat{F}_n, \Gamma_{\tilde{r}_{1,n+1}} \rangle \\ \vdots \\ D_{M,n+1} + \langle \hat{F}_n, \Gamma_{\tilde{r}_{M,n+1}} \rangle \end{pmatrix}.$$

 for all $i \in \{1, \dots, M\}$
 Output $\hat{F}_{n+1} = \sum_{i=1}^{M} w_{i,n+1} \Gamma_{\tilde{r}_{i,n+1}}$

Theorem: With $\Phi(x) = K(x,x) = \exp(-\mu \|x\|_2^2)$ the minimal eigen-value of $G_\Gamma = (\langle \Gamma_{\gamma_i}, \Gamma_{\gamma_j} \rangle)_{i,j=1,1}^{N,N}$ is bounded by

$$\lambda_{min} \geq C/2\mu \, \frac{\exp(-M_2^2/q_{xT}^2 \cdot \mu)}{q_{xT}^2} \cdot T^2$$

with $M_2 = 12\sqrt[3]{\pi/9}$, $C = M_2^2/16$ and $q_{x,T} := \frac{1}{2} \min_{\substack{j \neq k \\ t,\tau \in [0,T]}} \|\gamma_j(t) - \gamma_k\|_2$.

Theorem: Under the same conditions as the above $\lambda_{max} \leq N T^2 \Phi(0)$.

$$\lambda_{max} \leq N \int_0^T \int_0^T \max_{i,j=1,\dots,N} |\Phi(r_i(t) - r_j(\tau))| \, dt \, d\tau$$

To prove convergence we found sufficient conditions to apply the contraction mapping theorem.

We make the assumptions that our process starts reasonably close to the true solution and that the true flow field F is not unreasonably strong.

We also assume $\|\Gamma_{r_\varphi} - \Gamma_s\|_H$ is small where Γ_s is the occupation Kernel for a straight line trajectory and $\|\Gamma_{r_\varphi} - \Gamma_{r_\psi}\|_H \leq CT\|\varphi - \psi\|_H$ i.e. trajectories do not differ wildly from straight paths and that close flow fields produce similar trajectories.

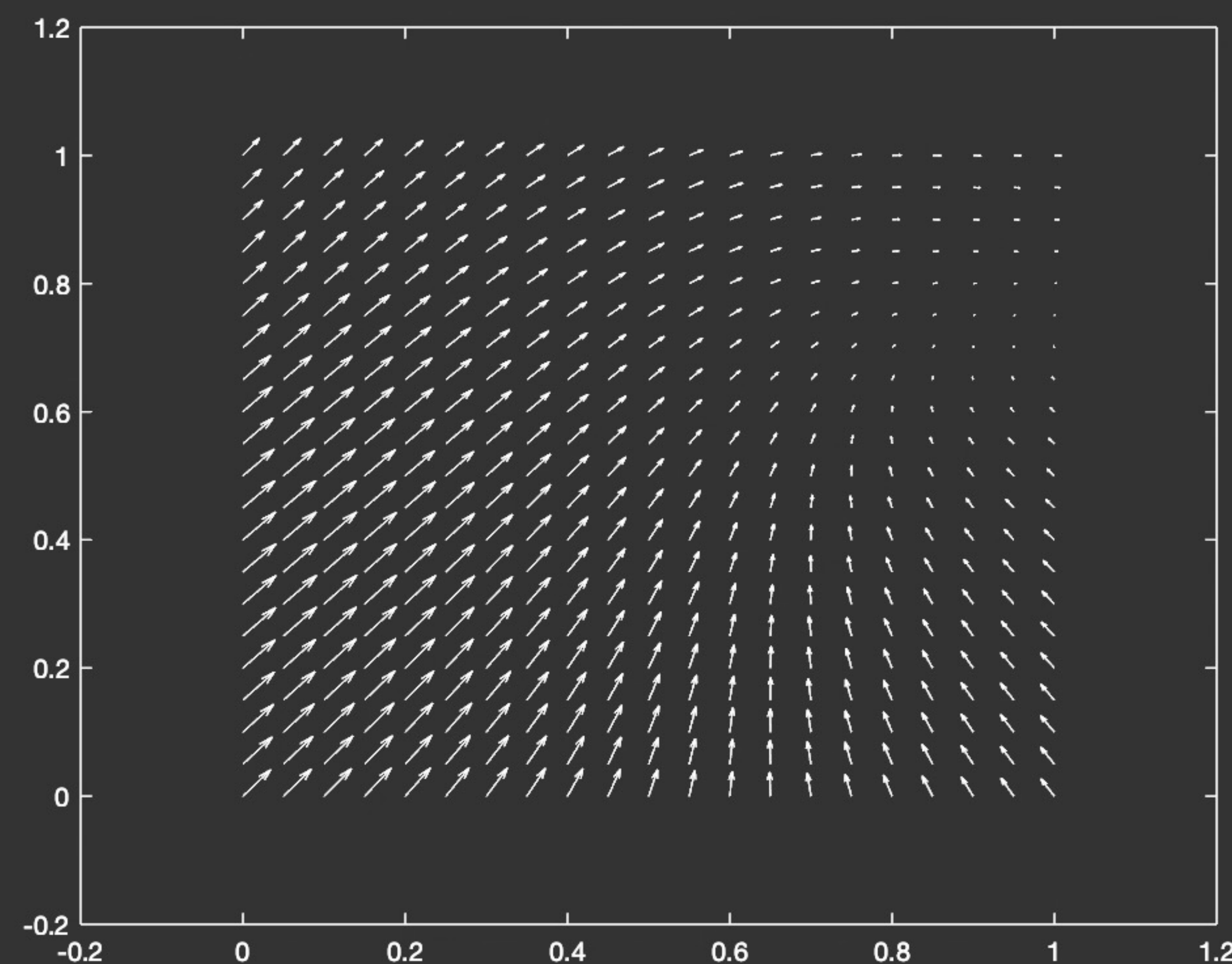
Experiments:

For an artificial experiment we used the flow-field given by

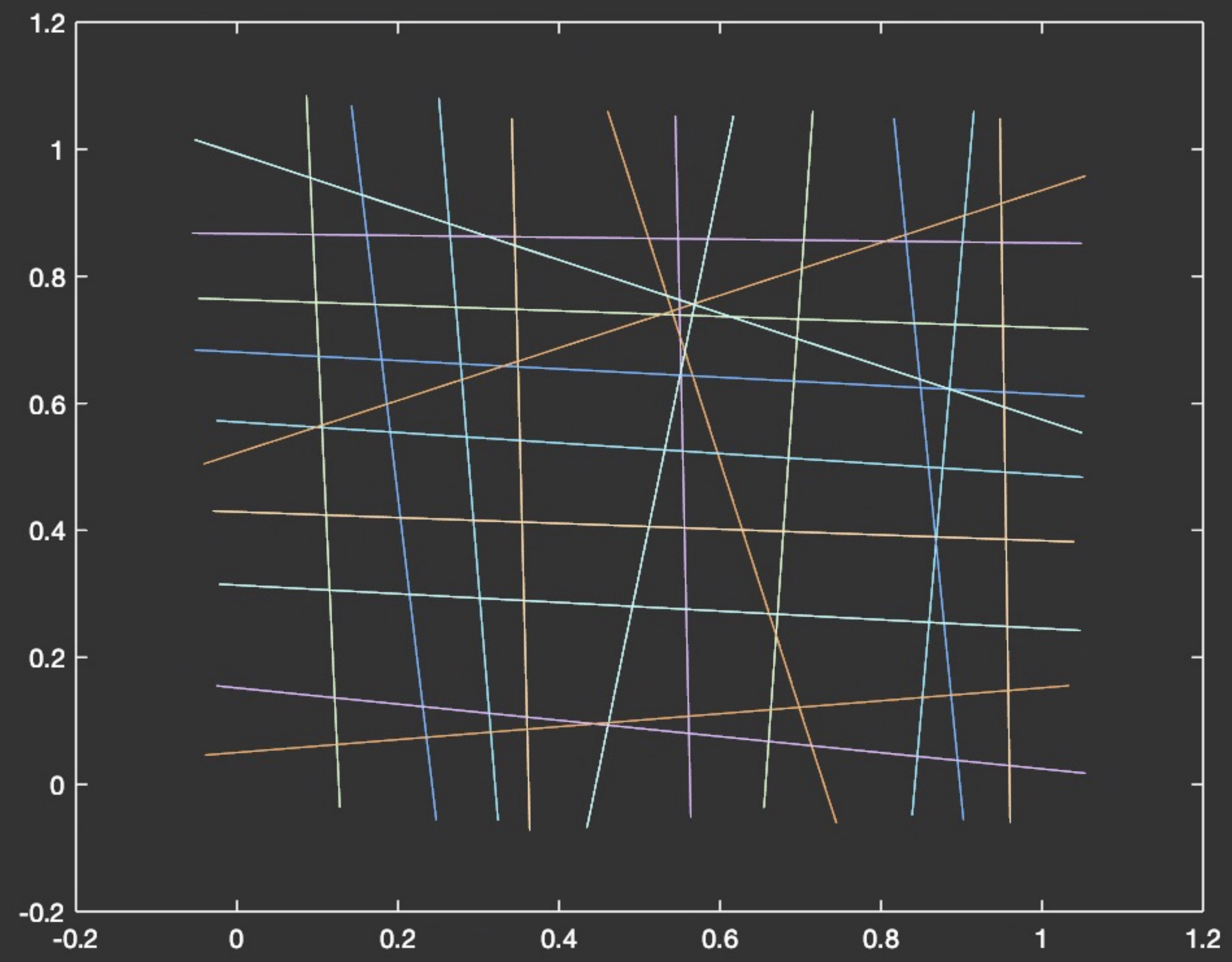
$$F(x) = \tfrac{1}{8}\left(f_1(x) \;\; f_2(x)\right)^T$$

$$f_1(x) = 5\exp\left(-2\left\|x - \begin{bmatrix}1/4\\1/4\end{bmatrix}\right\|^2\right) - .2\exp\left(-\left\|x - \begin{bmatrix}1/4\\3/4\end{bmatrix}\right\|^2\right) + 2\exp\left(-\left\|x - \begin{bmatrix}3/4\\3/4\end{bmatrix}\right\|^2\right) - 5\exp\left(-2\left\|x - \begin{bmatrix}3/4\\1/4\end{bmatrix}\right\|^2\right)$$

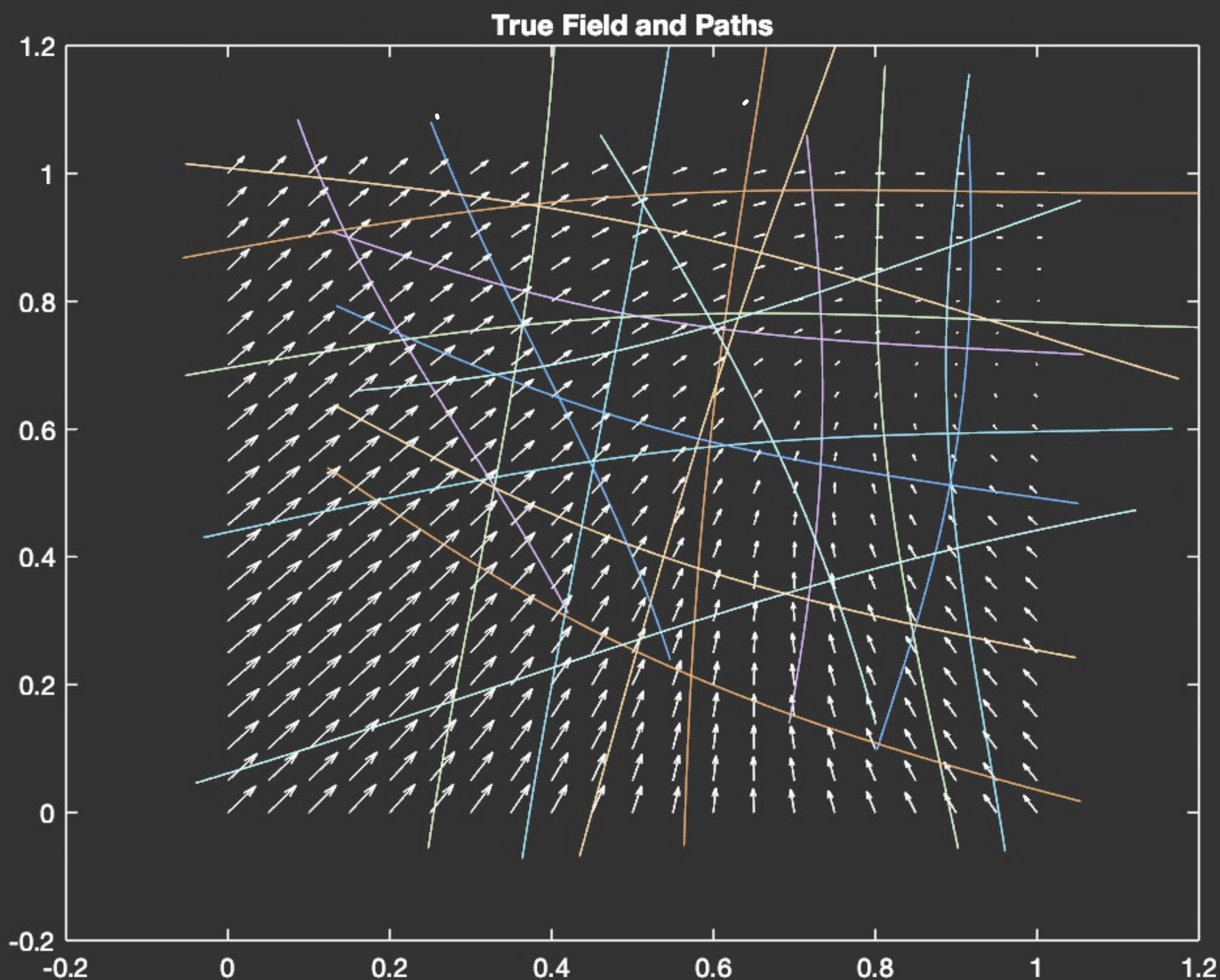
$$f_2(x) = 3\exp\left(-\left\|x - \begin{bmatrix}1/4\\1/4\end{bmatrix}\right\|^2\right) - \exp\left(-\left\|x - \begin{bmatrix}1/4\\3/4\end{bmatrix}\right\|^2\right) - 3\exp\left(-3\left\|x - \begin{bmatrix}3/4\\3/4\end{bmatrix}\right\|^2\right) + \exp\left(-\left\|x - \begin{bmatrix}3/4\\1/4\end{bmatrix}\right\|^2\right)$$

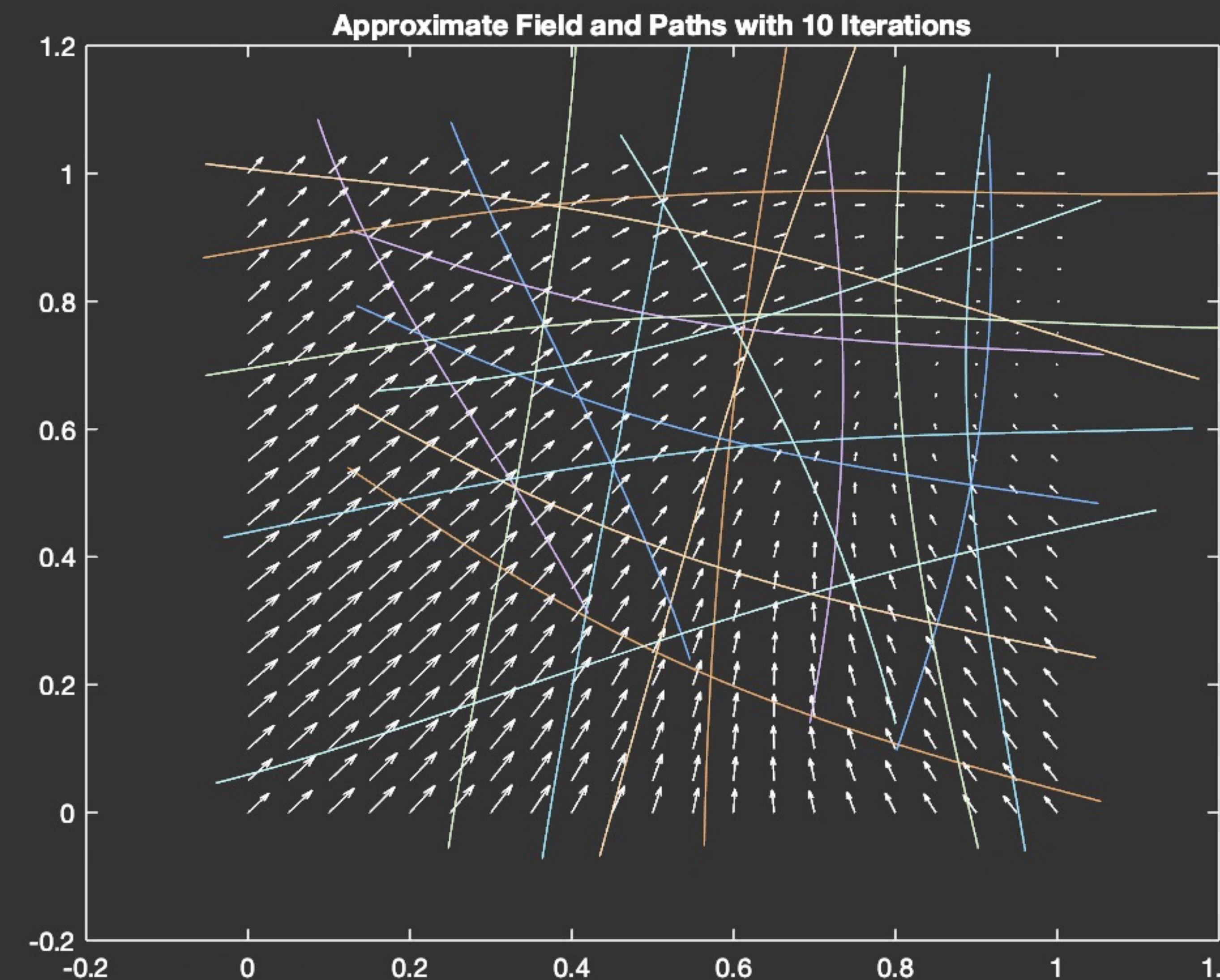


We then generated a set of points and angles to serve as our anticipated trajectories.



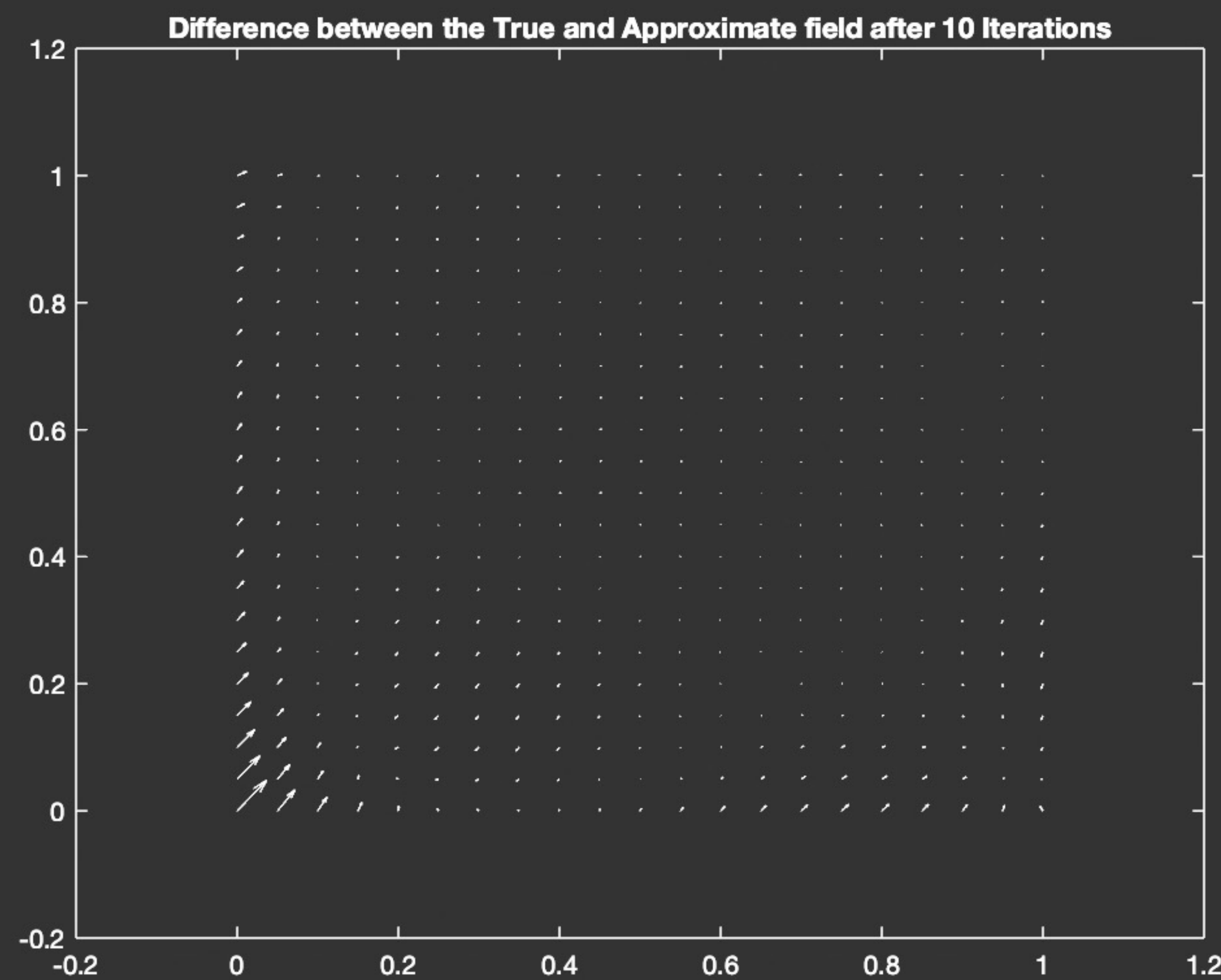
These are our results after only 10 iterations.

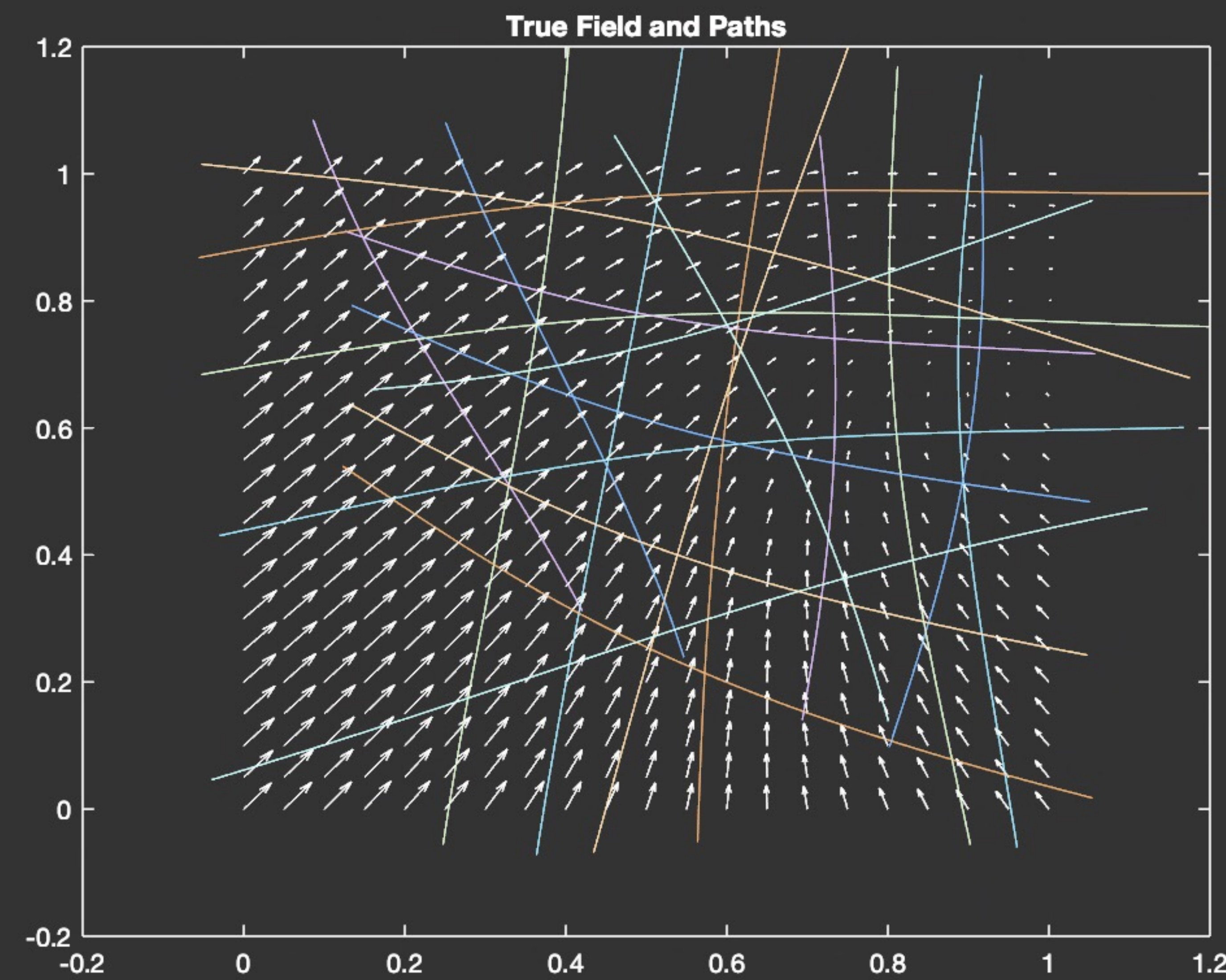




This was done using Gaussian RBF's with a Kernel width of $\mu = 1/50$.

We can visually see the difference.



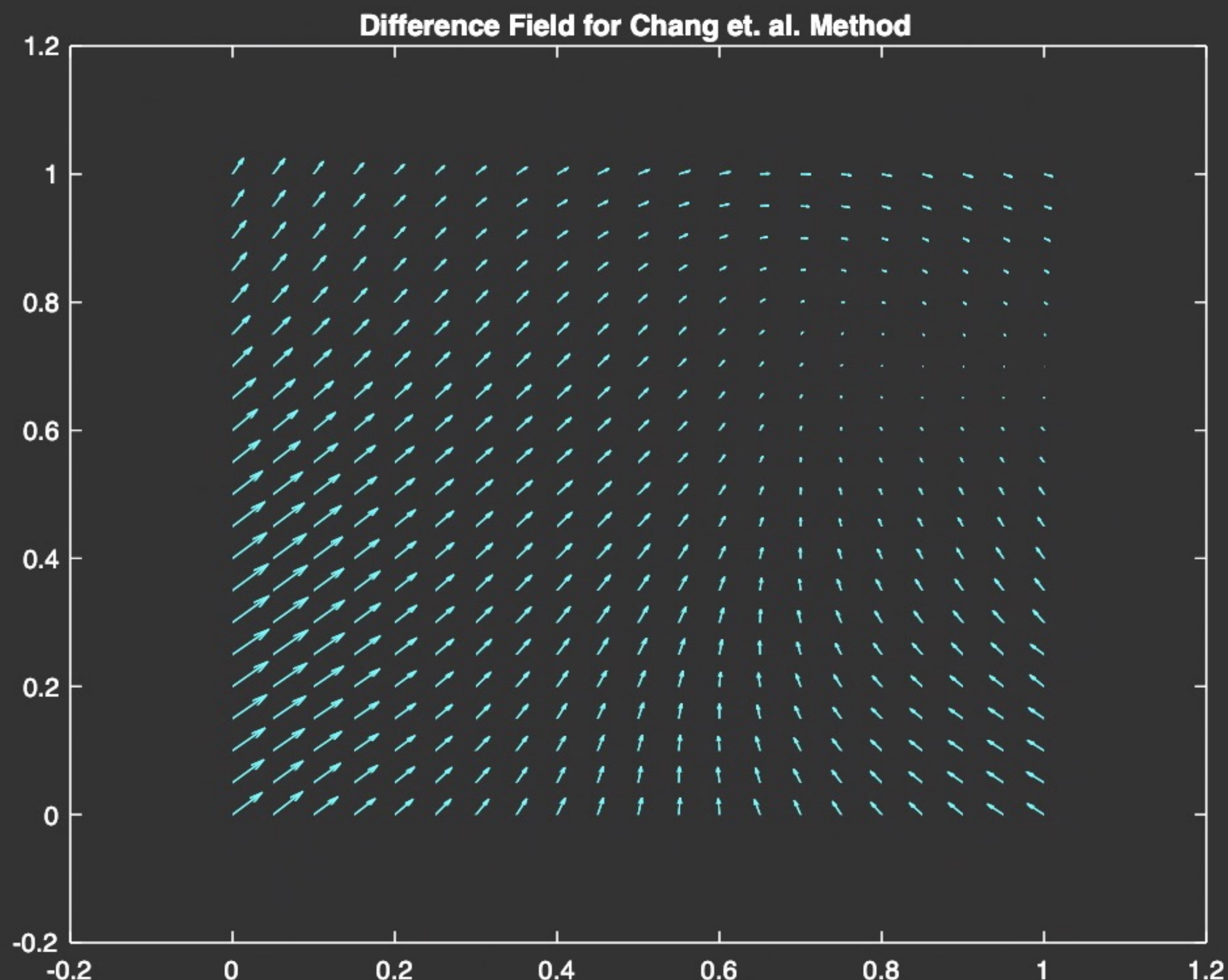


Some additional information:

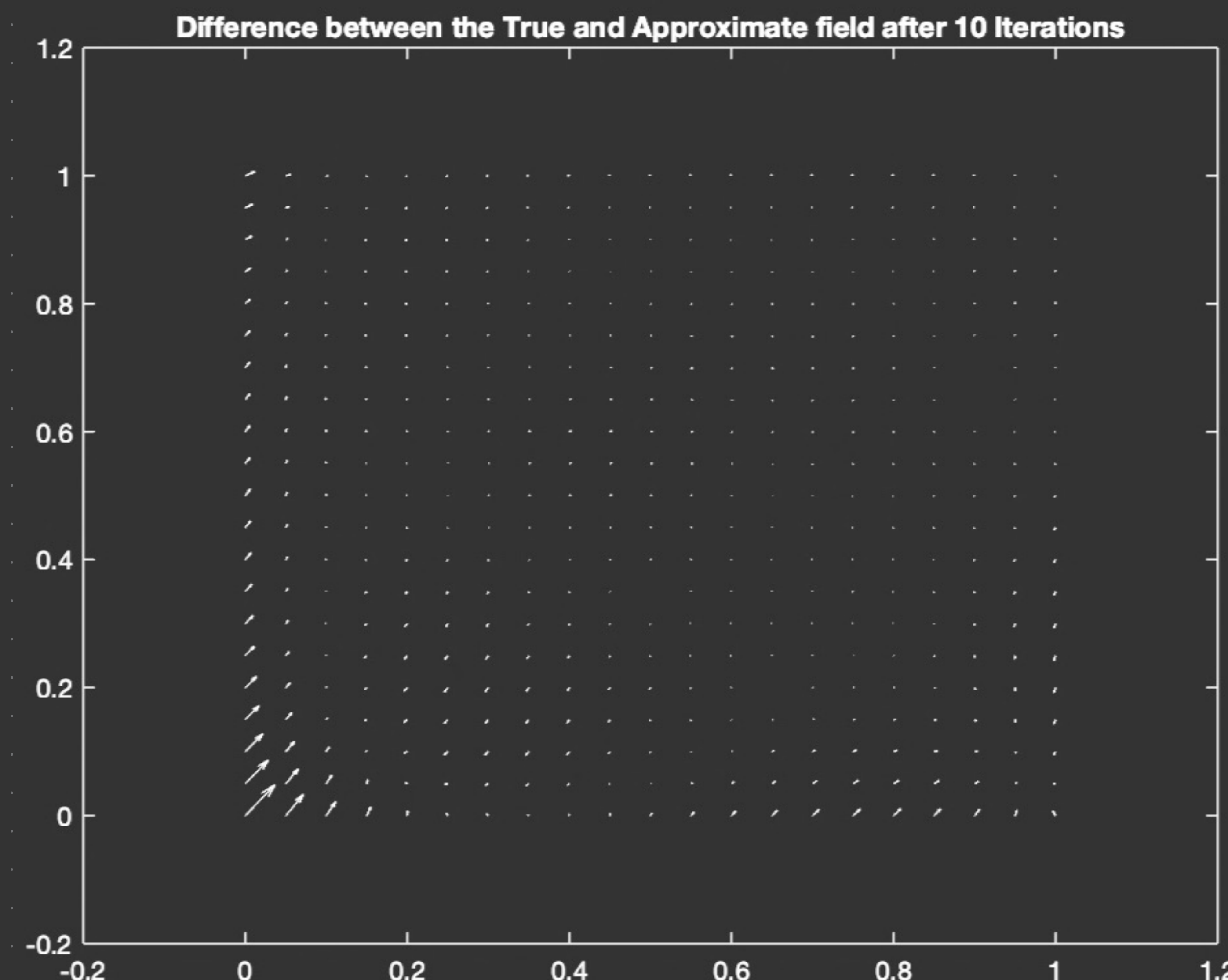
To generate the paths we used RK4 and computed integrals with Simpson's Rule.

We used $[0,1]$ as our time frame, a constant speed of 1, and a stepsize of $h=0.01$.

Comparison of methods

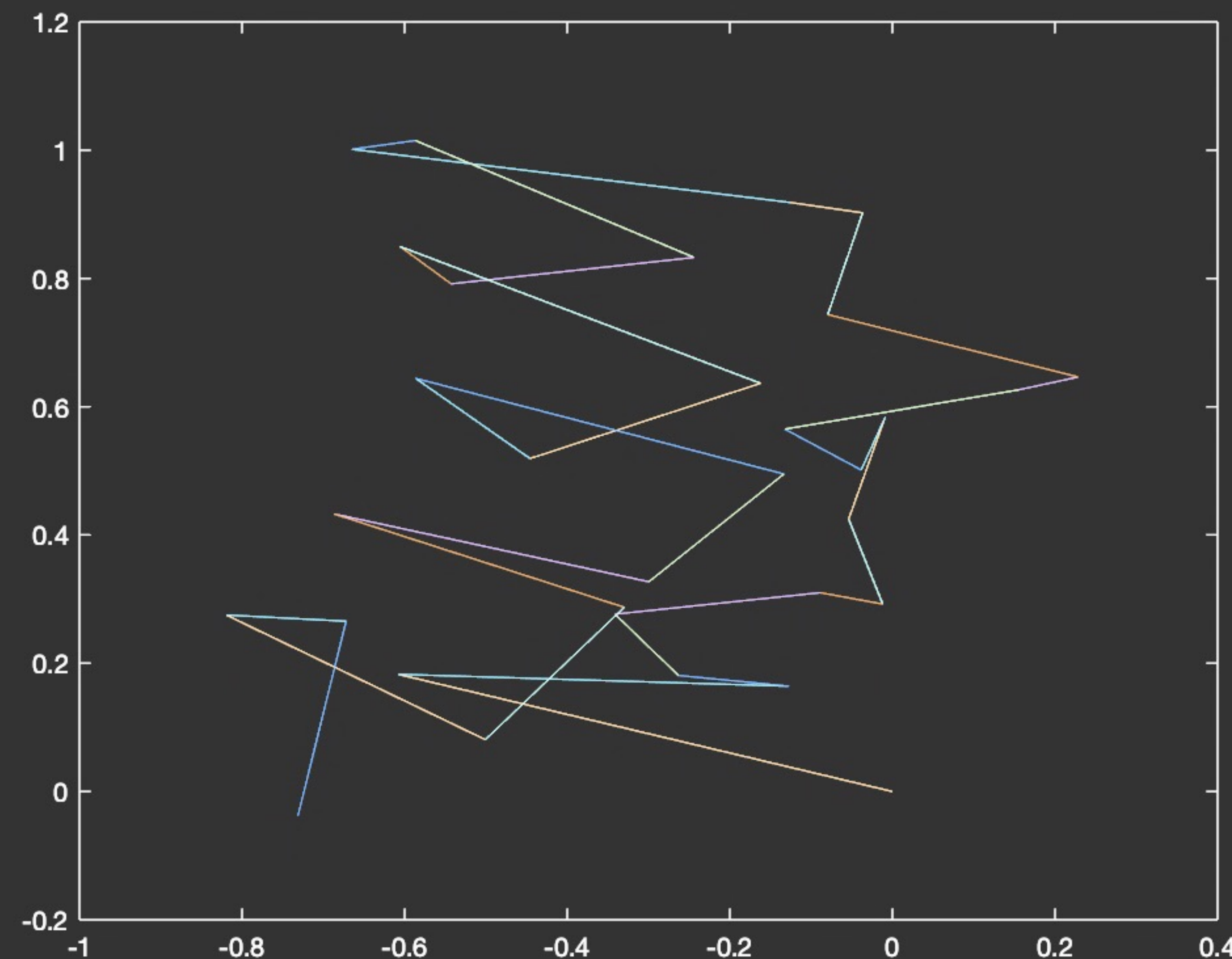


Chang et. al

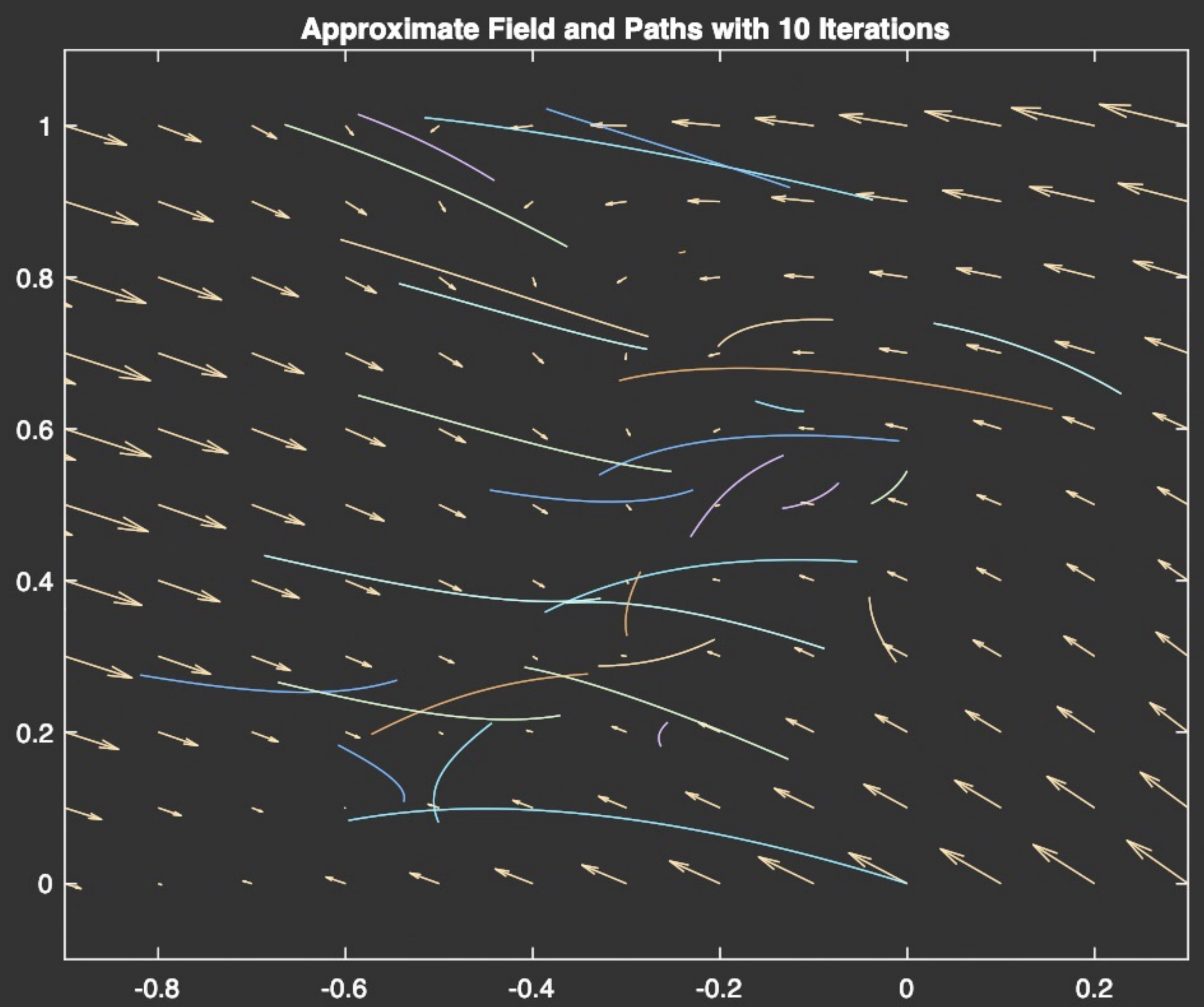


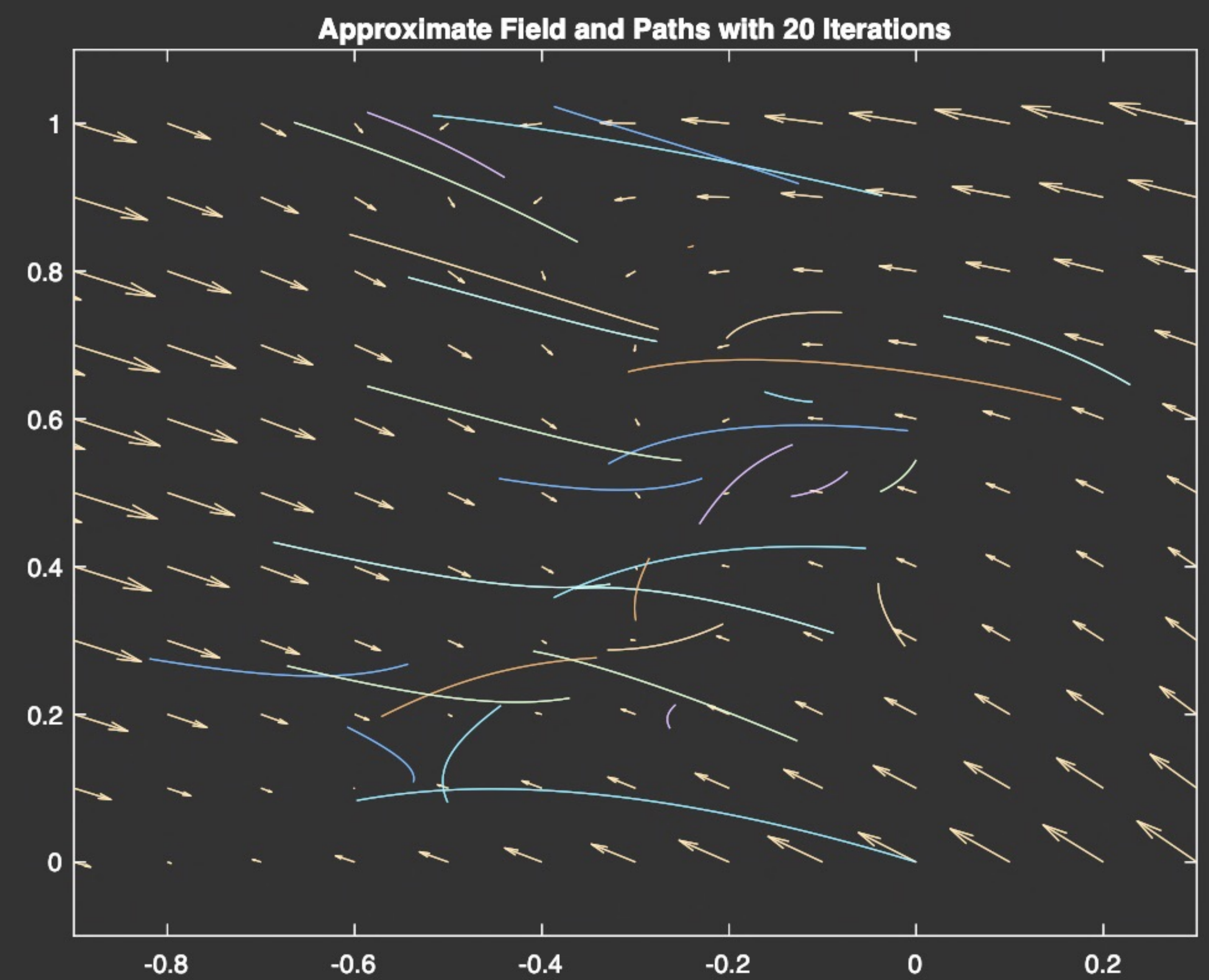
Kernel method

Real Example: For a real experiment we used 2013 Gliderpalooza data graciously shared by Chang et.al containing initial, final, and anticipated positions for 31 sequential trajectory segments each lasting a fixed time.

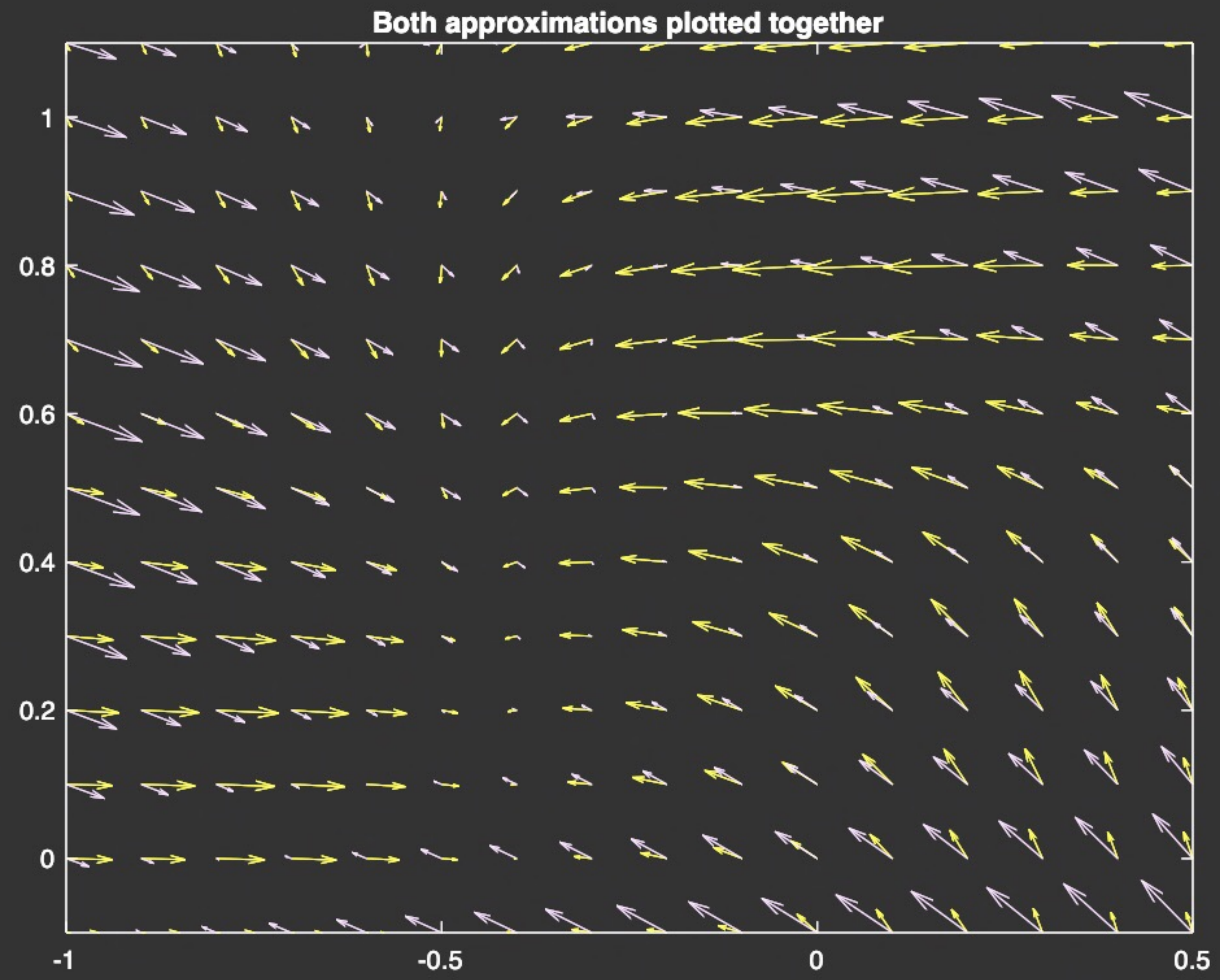


Since our algorithm is designed to run on several trajectories we broke the single trajectory into 31 separate segments.





In the absence of ground truth, we will simply remark that the algorithm seems to converge. In this example we used exponential Kernels, $K(x,y) = \exp\left(\frac{x^T \cdot y}{\mu}\right)$, with a parameter value of $\mu = 10{,}000$.

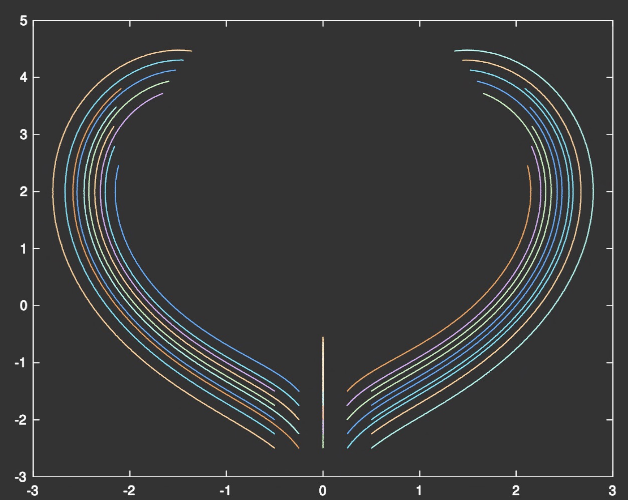


Kernel Method

Chang et al

Current Frontiers: Dynamic mode decomposition is a dimension reduction technique centered on finding the eigenmodes of a dynamical system.

Overview:



Cast the finite dimensional non-linear system as a comp. op K_F

$\Longrightarrow$

Use data to find eigenfunctions

$$\varphi_i(X_{j+1}) = K_F \varphi_i(X_j) = \lambda_i \varphi_i(X_j)$$

$\Longrightarrow$

Project the identity onto span$\{\phi_i\}$

$$X_j = \sum_{i=1}^{\infty} \xi_i \lambda_i^j \varphi_i(X_0)$$

Here, $X_{i+1} = F(X_i)$ where F is called a flow map.

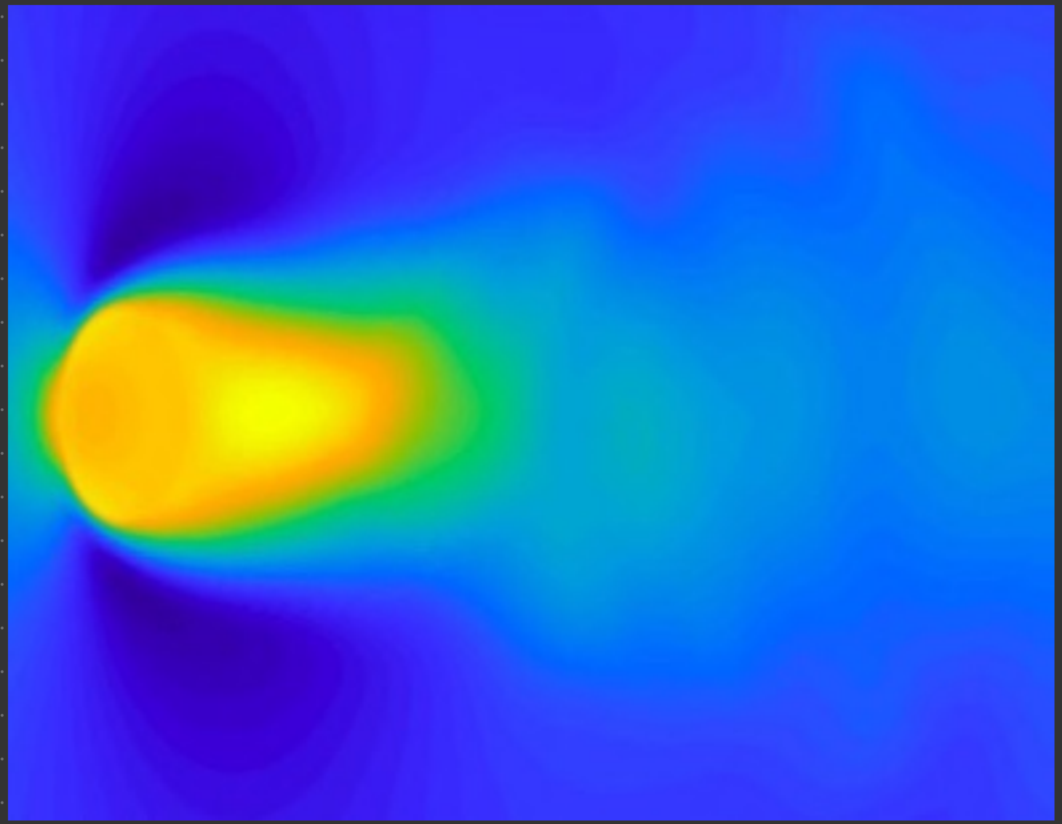
$$K_F g = g \circ F \quad \text{for } g \in H$$

Occupation Kernels interfaces well with DMD

Dynamic Mode Decomposition for Continuous Time Systems with the Liouville Operator

Joel A. Rosenfeld · Rushikesh Kamalapurkar · L. Forest Gruss · Taylor T. Johnson

arXiv:1910.03977v1 [math.DS] 9 Oct 2019



"Liouville" mode

Higher order systems:

$$\ddot{x} = f(x) \implies \left.\begin{aligned} z &:= (x \;\; \dot{x})^T \\ \dot{z} &= (z_2,\; f(z_1))^T \end{aligned}\right\} \text{Requires numerical estimation of derivatives}$$

We'd like to find a way to gain the same benefits as before.

Recall some Key elements

We have a RKHS of observables $g \in H$.

$\int_0^T g(\gamma(t))\,dt$ gets encoded as a linear functional $\langle g, \Gamma_\gamma \rangle_H$.

$\dot{\gamma}(t) = f(\gamma(t)) \Rightarrow \frac{d}{dt}(g \circ \gamma(t)) = \nabla g(\gamma(t)) f(\gamma(t))$ so the dynamics are encoded as an operator A_f.

We built methods around

$$\langle A_f g, \Gamma_\gamma \rangle = \int_0^T \nabla g(\gamma(t)) f(\gamma(t))\,dt = g(\gamma(t)) - g(\gamma(0)).$$

Redux

We have a vvRKHS of observables $\psi_g[\gamma] = g(\gamma(t))$.

$\frac{1}{(m-1)!}\int_0^T (T-t)^{m-1} g(\gamma(t))\,dt$ gets encoded as a linear functional $\langle g, \Gamma_\gamma^m \rangle_H$.

$\ddot{\gamma}(t) = f(\gamma(t)) \Rightarrow \frac{d^2}{dt^2}(g \circ \gamma(t)) = \left(\dot{\gamma}(0) + \int_0^T f(\gamma(t))dt\right)^T \mathcal{H}[g](\gamma(t))\left(\dot{\gamma}(0) + \int_0^T f(\gamma(t))dt\right)$ so the dynamics are encoded as an operator B_f.

We built methods around

$$\langle B_f \psi_g, \Gamma_\gamma^{(2)} \rangle_H = \int_0^T (T-t)\, B_f \psi_g[\gamma](t)\, dt = \psi_g[\gamma](T) - \psi_g[\gamma](0) - T\nabla\psi[\gamma](0)\dot{\gamma}(0)$$

Thanks!

Collaborators:

Joel Rosenfeld

University of South Florida

Rushikesh Kamalapurkar

Oklahoma State University

Taylor Johnson

Vanderbilt University