Identifying Latent Stochastic Differential Equations with Variational Auto-Encoders

Ali Hasan,[1 *] João M. Pereira,[2 *] Sina Farsiu,[12] Vahid Tarokh[2]
[1]Department of Biomedical Engineering, Duke University
[2]Department of Electrical and Computer Engineering, Duke University
{ali.hasan, joao.pereira, sina.farsiu, vahid.tarokh}@duke.edu

Presented By:
Achint
21st April, 2021

Plan for the talk

- Introduction to SDE and VAE
- Problem Statement
- Three Theorems and their Proofs
- Conclusion

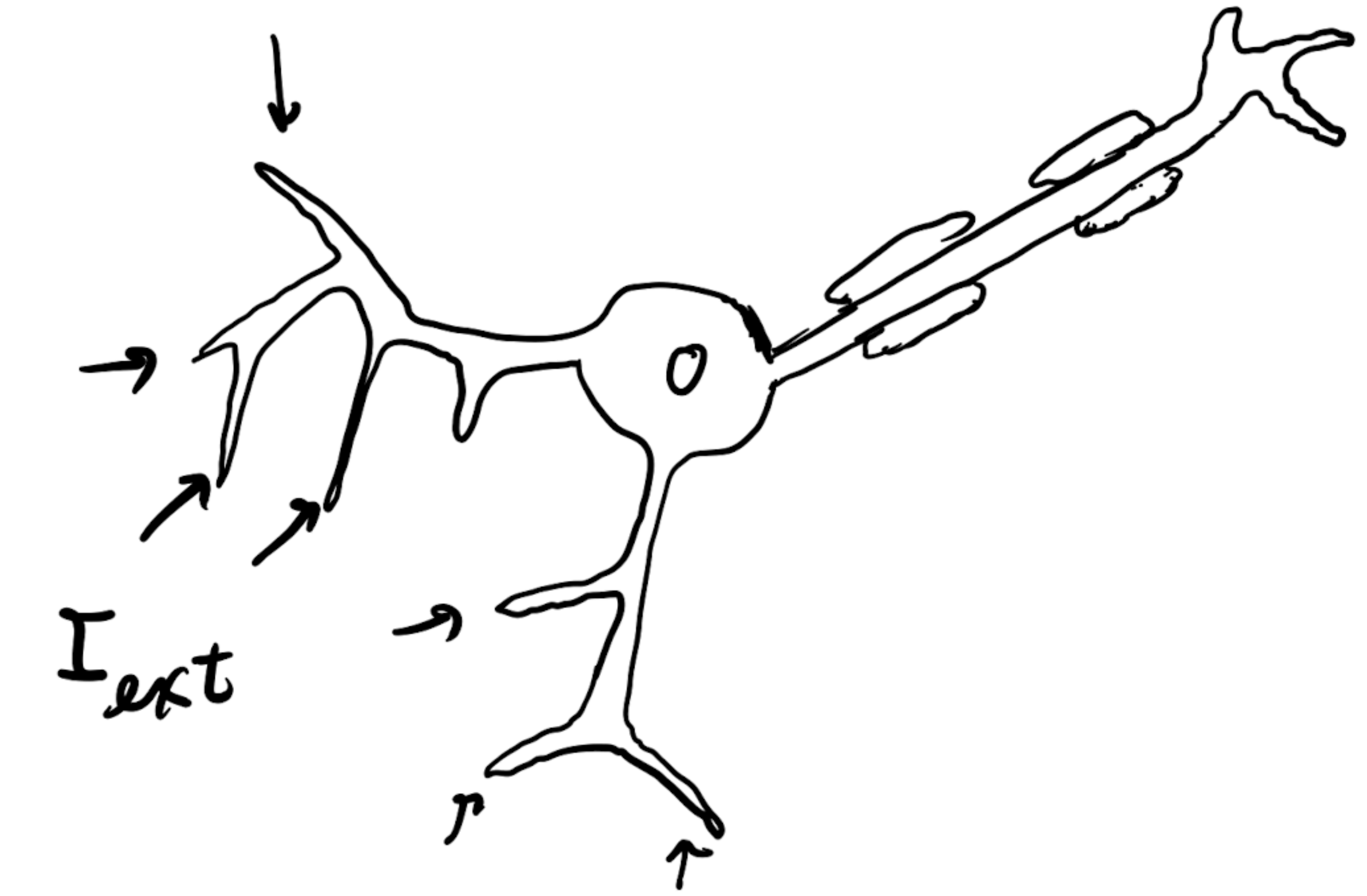
Plan for the talk

- Introduction to SDE and VAE
- Problem Statement
- Three Theorems and their Proofs
- Conclusion

> Never underestimate the joy people derive from hearing something they already know.
>
> — Enrico Fermi —

Stochastic Differential Equations

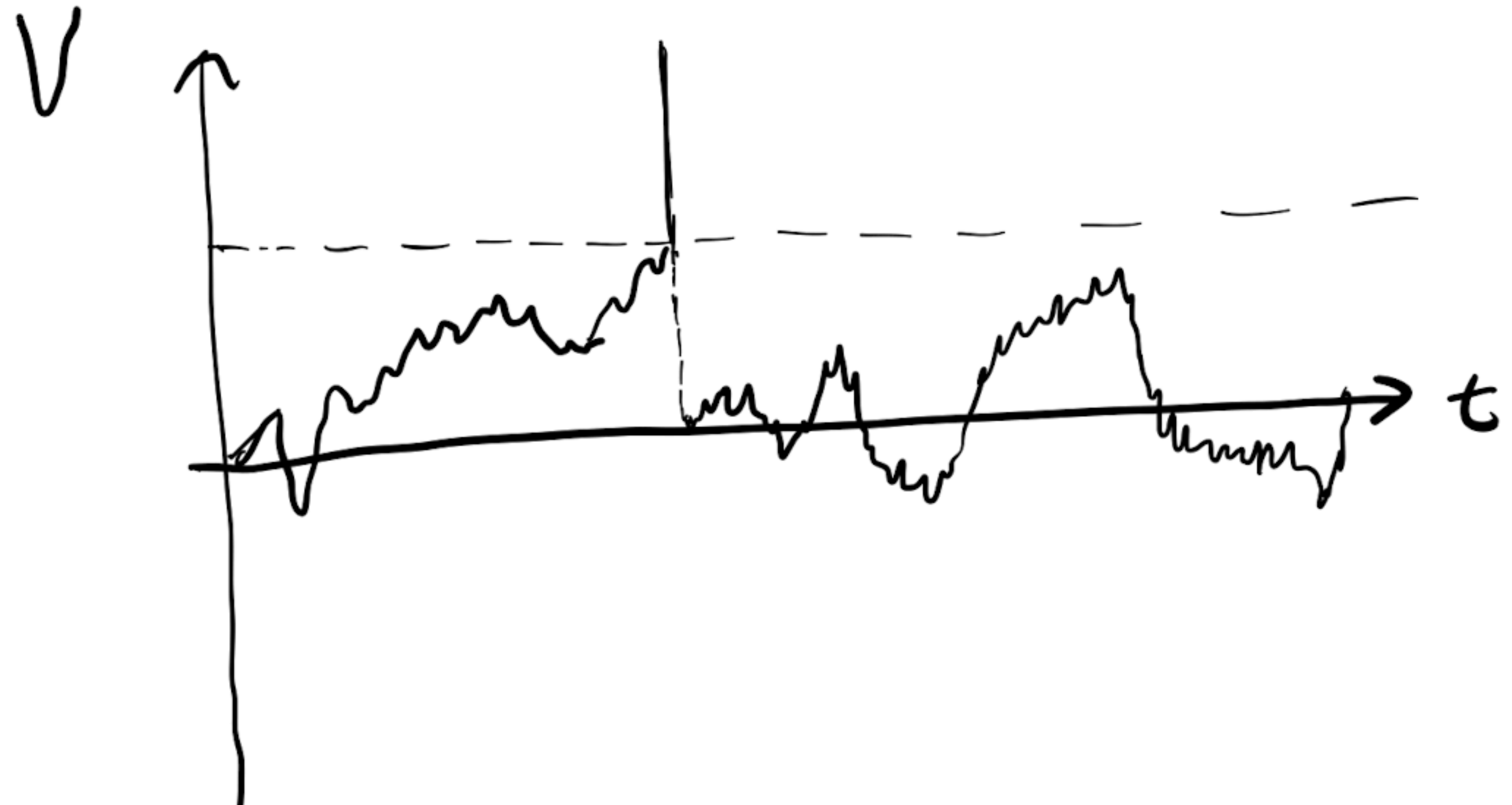
- Equations with derivatives and noise (‘stochasticity’)
- Consider Lapicque neuron,



$$\tau \frac{dV}{dt} = -V + I_{\text{ext}}$$

$$\Rightarrow \tau \frac{dV}{dt} = -V + \eta(t)$$

Gaussian White noise



General form of SDE

$$\tau \frac{dV(t)}{dt} = -V(t) + \eta(t)$$

$$\Rightarrow dV_t = -\frac{1}{\tau} V_t dt + \frac{1}{\tau} dW_t$$

$$\downarrow$$

$$\boxed{dZ_t = \mu(Z_t)dt + \sigma(Z_t)dW_t}$$

- Z_t could represent voltage, position, etc
- Sneak peak: The goal of this paper would be to infer the drift and diffusion terms given some data

General form of SDE

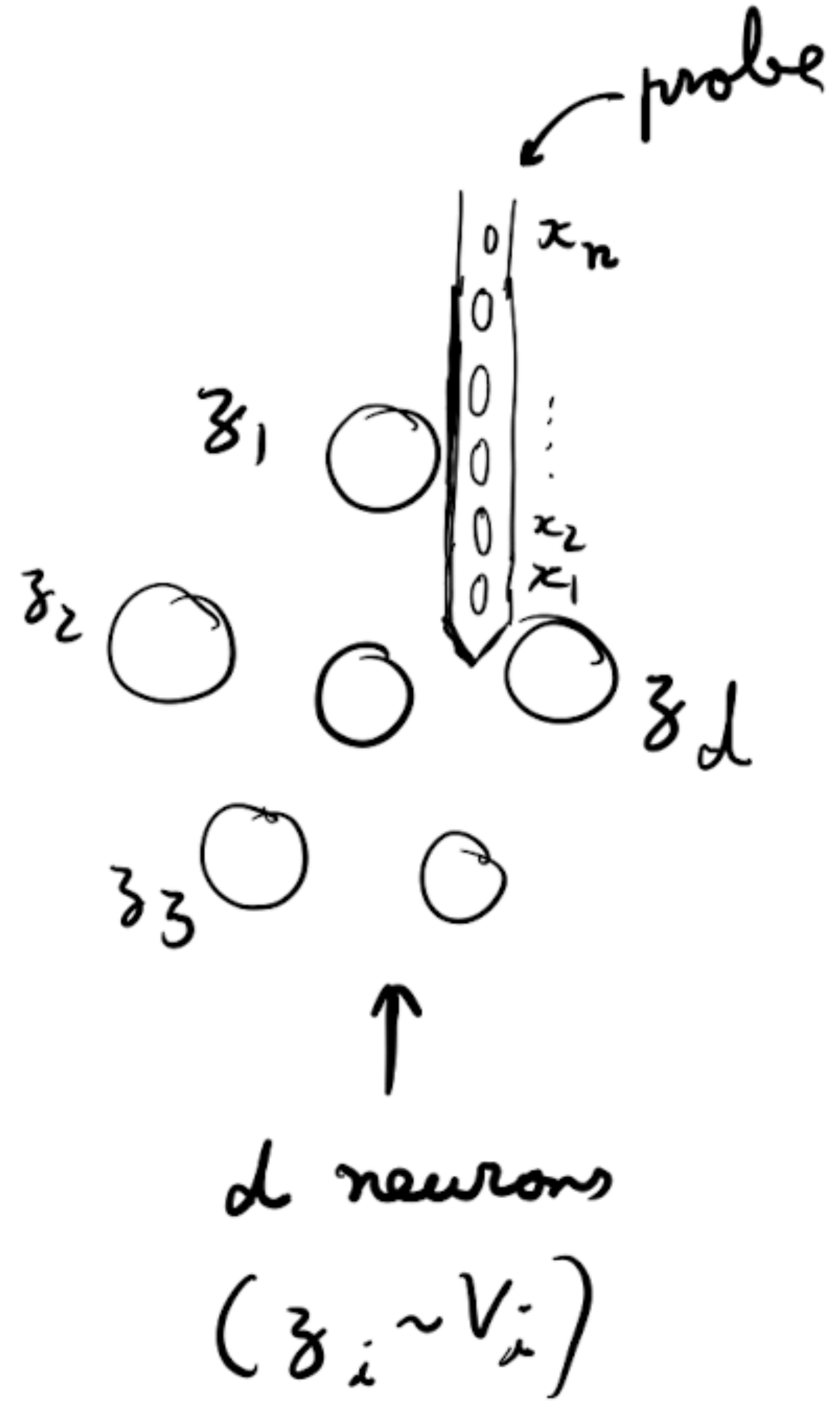
$$dZ_t = \mu(Z_t)dt + \sigma(Z_t)dW_t$$

- If $\mu(Z_t) = -\alpha Z_t$ and $\sigma(Z_t) = \text{constant}$, then it is called an Ornstein-Uhlenbeck Process

Gathering data

$$dZ_t = \mu(Z_t)dt + \sigma(Z_t)dW_t$$

$$X_t = f(Z_t)$$



Another Example: Stochastic Yellow Ball

(a) Yellow ball moving according to a 2D Ornstein-Uhlenbeck process;

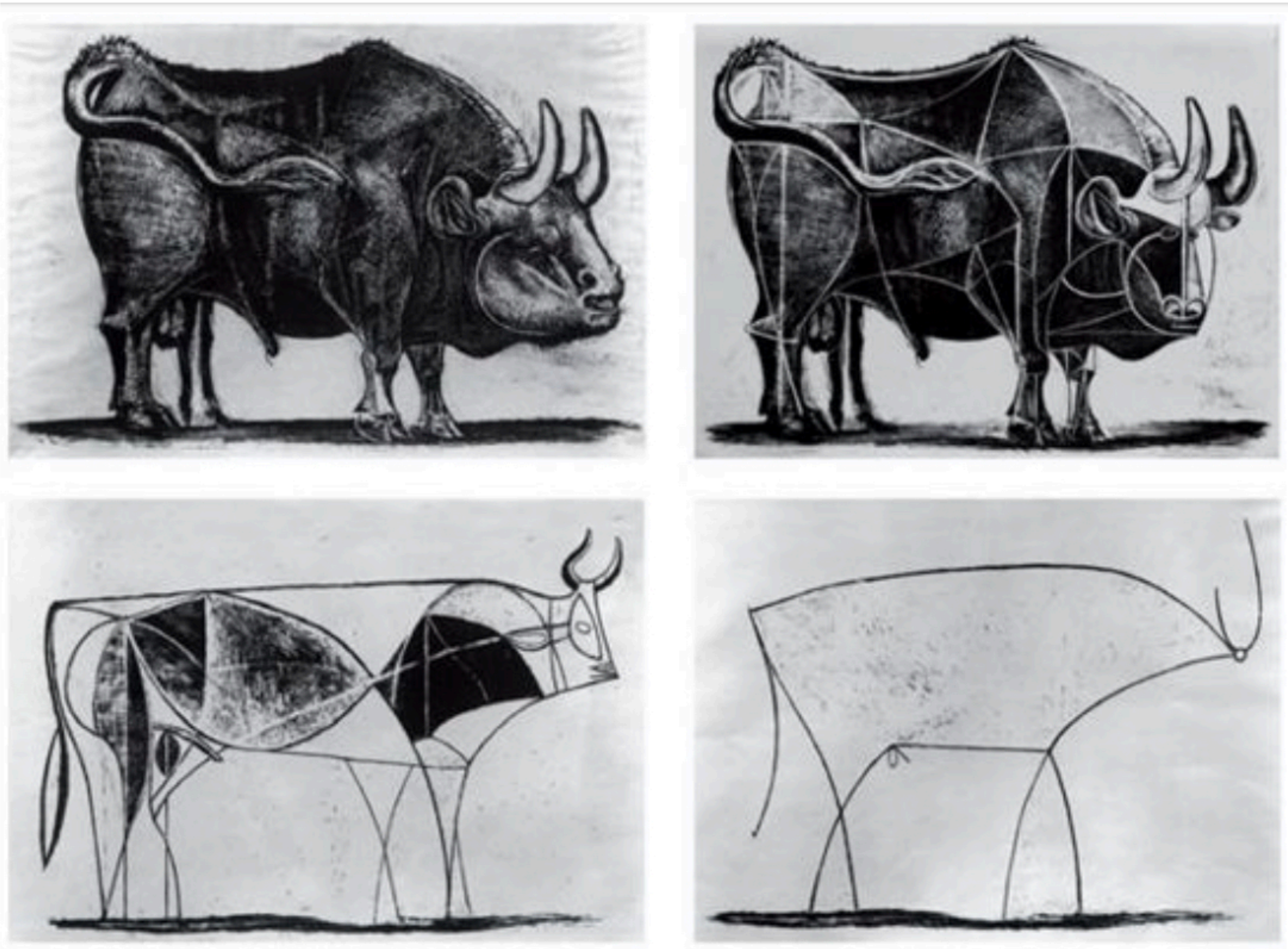
$$dx_t = -4x_t dt + dW_t$$

$$dy_t = -4y_t dt + dW_t$$

- Observable: X_t = 64x64x3 dimensional Images representing one realization of SDE for 1000 time steps

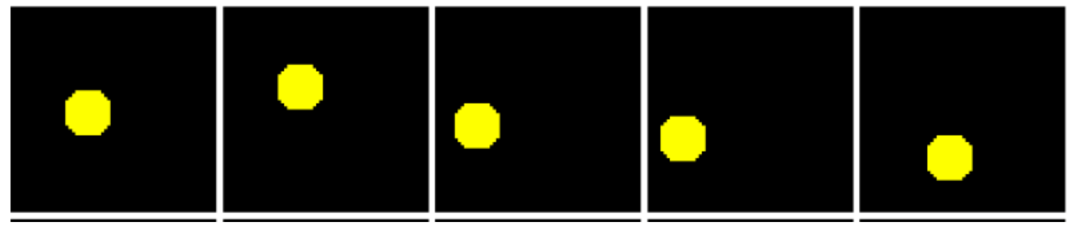
- Goal: How do we infer f, drift and diffusion coefficient of the underlying SDE

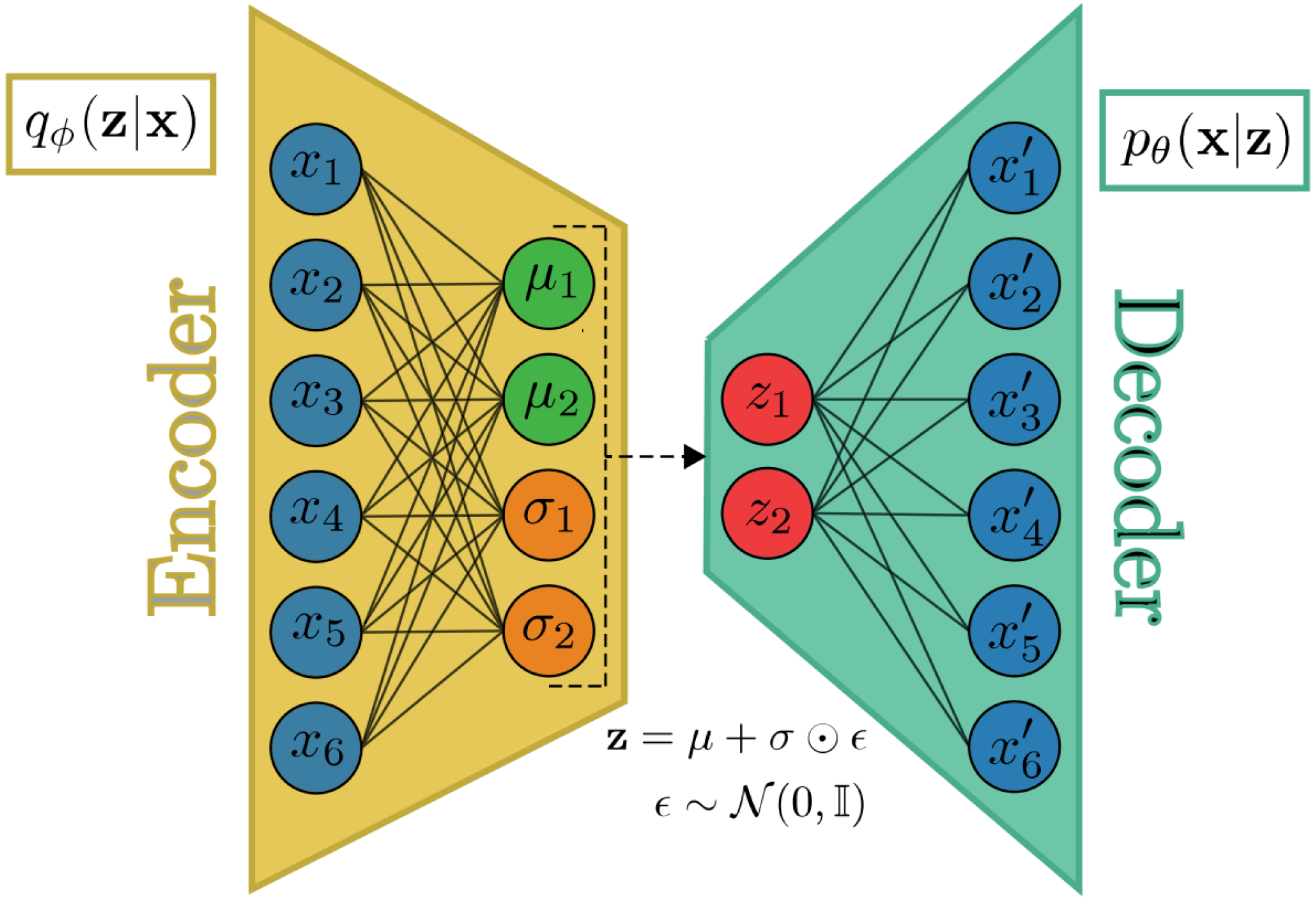
Autoencoders: Capturing the essence of the substance

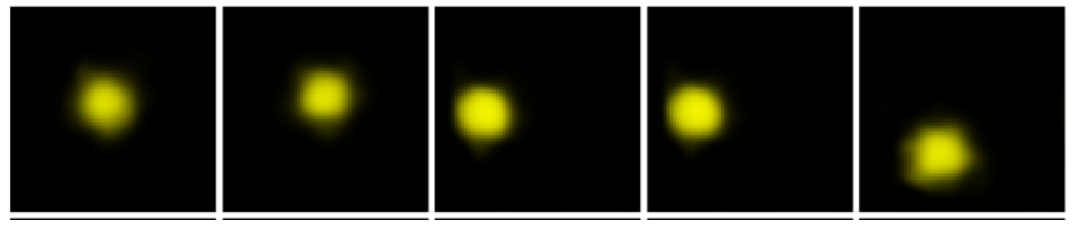


The Bull by Picasso

Variational Autoencoders







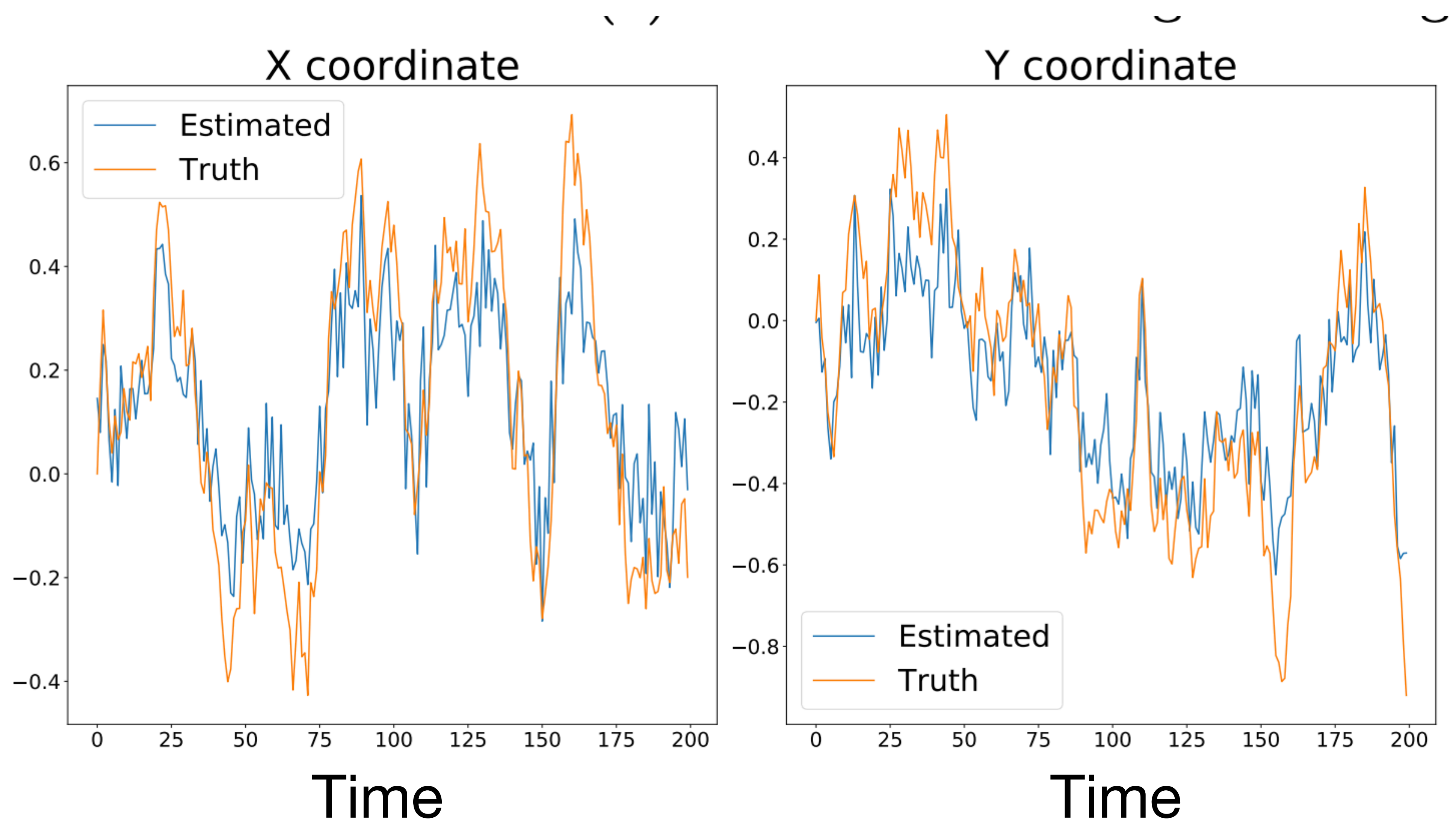
Angelo, PRL 2020

Wish list

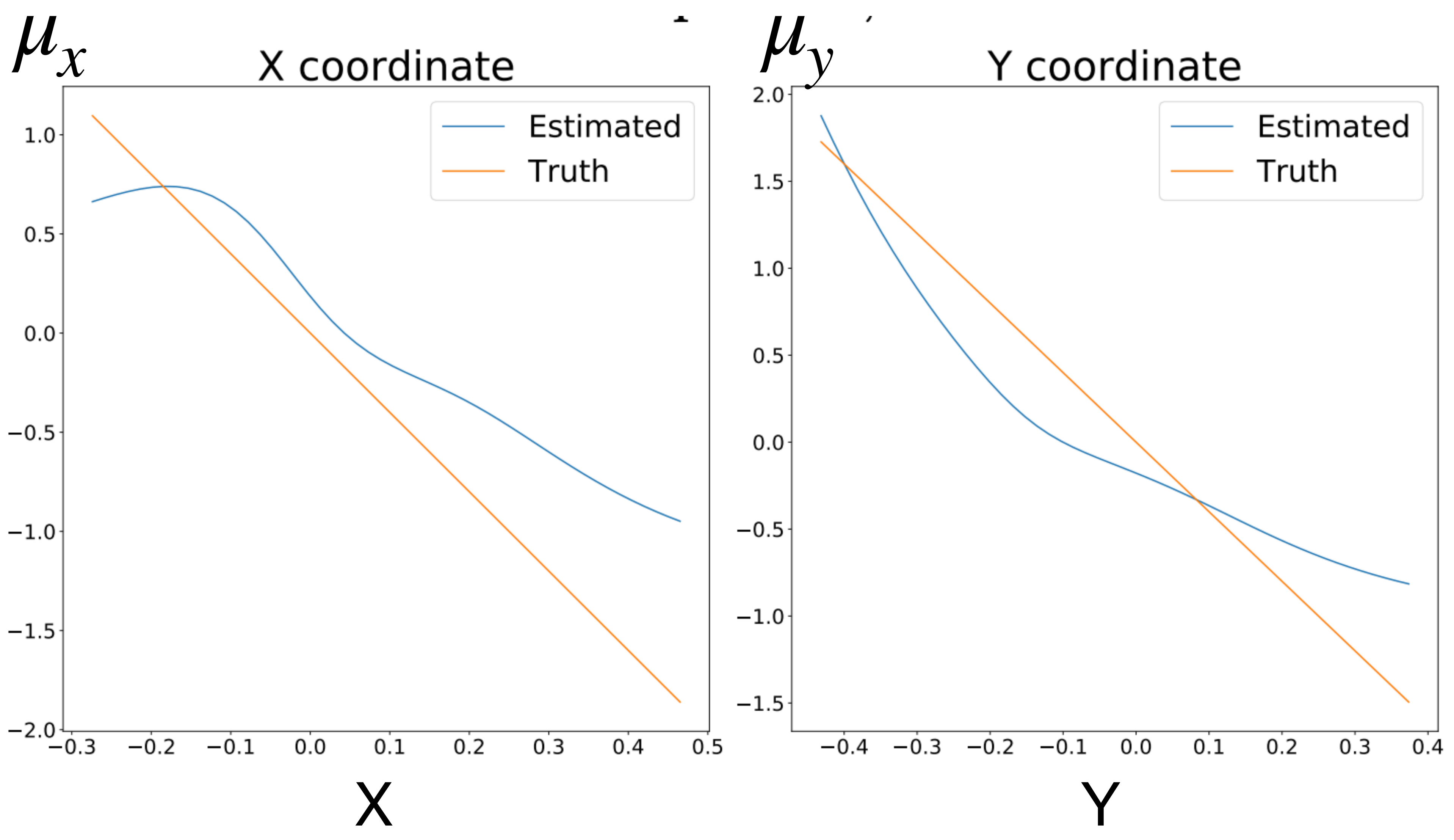
- It would be great if we could interpret the latent dimensions as the x and y coordinate of center of the yellow ball
- Using these latent dimensions we would like to infer the drift and diffusion coefficient

Example: Stochastic Yellow Ball

(a) Yellow ball moving according to a 2D Ornstein-Uhlenbeck process;



(b) Comparison between the true centers of the ball and the latent representation learned by the VAE at different frames of the video;



(c) Comparison between the true drift coefficient and the drift coefficient estimated by the VAE;

Plan for the talk

- Introduction to SDE and VAE
- Problem Statement: Find (f, μ, σ) given $X_t|_{t=1:\mathbf{T}}$
- Three Theorems and their Proofs
- Conclusion

Plan for the talk

- Introduction to SDE and VAE
- Problem Statement
- Three Theorems and their Proofs
- Conclusion

Theorem 1: (f, μ, σ) cannot be uniquely determined

If $Z \xrightarrow[(f,\mu,\sigma)]{} X$ then $Y = g(Z) \xrightarrow[(f',\mu',\sigma')]{} X$

Proof: We are given,

$$dZ_t = \mu\, dt + \sigma\, dW_t$$

Let $Y_t = g(Z_t)$, then using Taylor expansion (assuming $d=1$)

$$dY_t = \frac{\partial Y_t}{\partial t} dt + \frac{\partial Y_t}{\partial Z_t} dZ_t + \frac{1}{2} \frac{\partial^2 Y_t}{\partial Z_t^2} (dZ_t)^2$$

$$= \frac{\partial Y_t}{\partial t} dt + \frac{\partial Y_t}{\partial Z_t} (\mu\, dt + \sigma\, dW_t) + \frac{1}{2} \frac{\partial^2 Y_t}{\partial Z_t^2} (\mu\, dt + \sigma\, dW_t)^2$$

Using the fact, $dW_t^2 = dt$ $\quad \left(\because \langle W_t^2 \rangle - \underbrace{\langle W_t \rangle}_{0}^2 = t\right)$

We get,

$$dY_t = \underbrace{\left(\frac{\partial Y_t}{\partial t} + \mu \frac{\partial Y_t}{\partial Z_t} + \frac{\sigma^2}{2} \frac{\partial^2 Y_t}{\partial Z_t^2} \right)}_{\mu'} dt + \sigma \underbrace{\frac{\partial Y_t}{\partial Z_t}}_{\sigma'} dW_t$$

This equation is called Ito's lemma

Theorem 1: (f, μ, σ) cannot be uniquely determined

$$dY_t = \mu' dt + \sigma' dW_t$$

Now, we need to find f'.

Ansatz: $f'(Y_t) = f(g^{-1}(Y_t))$

Proof: $Y_t = g(Z_t)$, so

$$f(g^{-1}(Y_t)) = f(g^{-1}(g(Z_t)))$$

$$= f(Z_t) = X_t$$

So, (f', μ', σ') also leads to the same X_t

Theorem 2: We can make $\sigma = \mathrm{I}_d$ by carefully choosing g

Proof:

Recall,

$$dY_t = \underbrace{\left(\frac{\partial Y_t}{\partial t} + \mu \frac{\partial Y_t}{\partial Z_t} + \frac{\sigma^2}{2}\frac{\partial^2 Y_t}{\partial Z_t^2}\right)}_{\mu'} dt + \underbrace{\sigma \frac{\partial Y_t}{\partial Z_t}}_{\sigma'} dW_t$$

$$\sigma' = \sigma \frac{\partial Y_t}{\partial Z_t} = 1 \quad (\text{WANT})$$

$$\Rightarrow Y_t = \int_0^{Z_t} \frac{1}{\sigma(Z)} dZ \quad \text{(Lamperti Transform)}$$

So, if $g(Z_t) = \int_0^{Z_t} \frac{1}{\sigma(Z)} dZ$ then $\sigma' = 1$

Theorem 3: We can recover the true (f^*, μ^*) up to an isometry

$$f(z) = f^*(Qz + b) \qquad \mu(z) = Q^T \mu^*(Qz + b)$$

- Isometry = (rotation or reflection) + translation. You can also think of it as distance preserving transformation

Recall, if $Y_t = g(Z_t)$ then,

$(\sigma = 1)$

$$dY_t = \left(\frac{\partial Y_t}{\partial t} + \mu \frac{\partial Y_t}{\partial Z_t} + \frac{\partial^2 Y_t}{\partial Z_t^2} \right) dt + \frac{\partial Y_t}{\partial Z_t} dW_t$$

Since, we want $\frac{\partial Y_t}{\partial Z_t} = 1 \Rightarrow \frac{\partial g(Z_t)}{\partial Z_t} = 1$

$g(Z)$ must be a linear function

$$g(Z) = QZ + b$$

When, $d = 1$, $Q = \pm 1$

For arbitrary d, Q must be orthogonal matrix, since $QW_t = W_t$. This is because W_t is rotationally-invariant.

Plan for the talk

- Introduction to SDE and VAE
- Problem Statement
- Content of three Theorem
- High level discussion of the proofs
- Conclusion

Conclusion

- VAE can allow us infer parameters of stochastic differential equations
- The inferred latent space is interpretable
- They also show how the dimensionality of the latent space can be inferred

Some reflections

- Can we use the VAE framework to infer parameters of ODEs from the dynamics?
- What happens if the SDE evolves according to non-Gaussian noise?
- How does one calibrate uncertainty in drift coefficient?