# Most probable transition pathway of biological SDEs

Stanley Nicholson

Illinois Institute of Technology

March 12, 2021

# Why Stochastics?

- All physical systems are "noisy"

# Why Stochastics?

- All physical systems are "noisy"
- Noise: think of Brownian motion

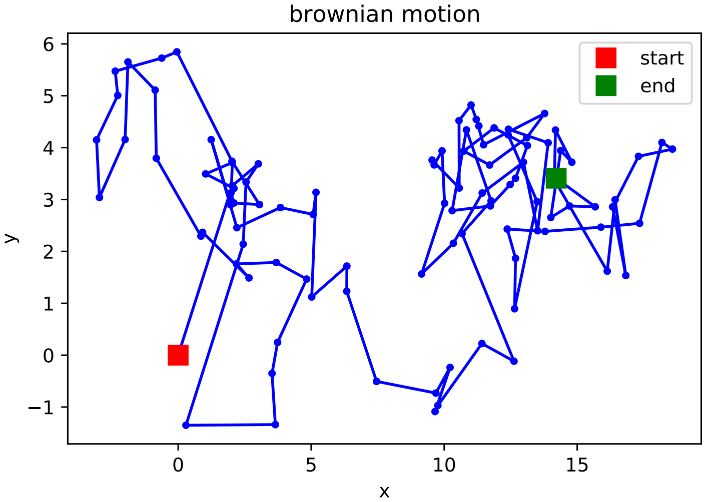
# Why Stochastics?

- All physical systems are "noisy"
- Noise: think of Brownian motion
- Accounts for sudden changes in system

# Why Stochastics?

- All physical systems are "noisy"
- Noise: think of Brownian motion
- Accounts for sudden changes in system
- Many biological processes are stochastic: mutation, gene regulation, etc.

# Brownian Motion



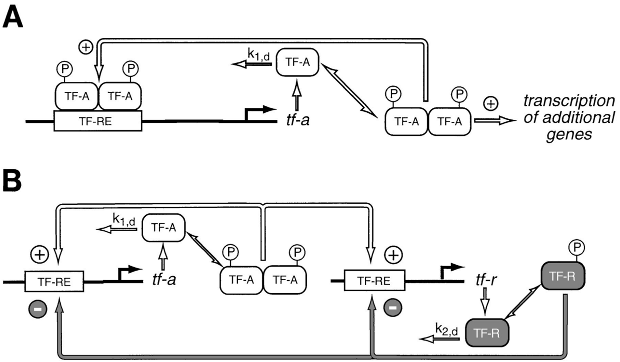
# What's an SDE?

The definition of a *stochastic differential equation* given a stochastic process $X_t$, drift function $f(X_t, t)$, and diffusion function $\sigma(X_t, t)$:

$$dX_t = f(X_t, t)dt + \sigma(X_t, t)dB_t \tag{1}$$

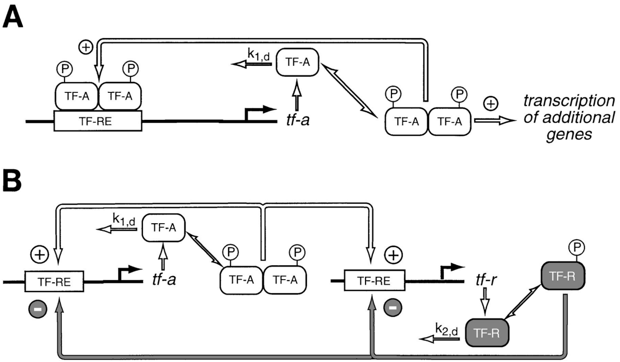
Think: $f$ determines direction, $\sigma$ denotes noise intensity

# The Biology



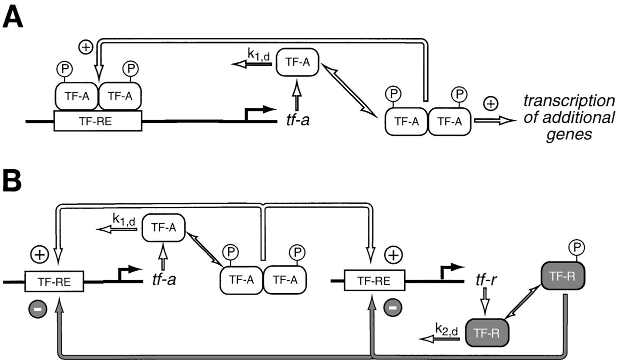
A
TF-A
TF-A
TF-RE
$k_{1,d}$
TF-A
tf-a
TF-A
TF-A
transcription of additional genes
B
$k_{1,d}$
TF-A
TF-RE
tf-a
TF-A
TF-A
TF-RE
tf-r
TF-R
$k_{2,d}$
TF-R

# The Biology




$$\frac{d[\text{TF A}]}{dt} = \frac{k_{1,f}[\text{TF A}]^2}{[\text{TF A}]^2 + K_{1,d}} - k_{1,d}[\text{TF A}] + r_{1,\text{bas}} \qquad (2)$$

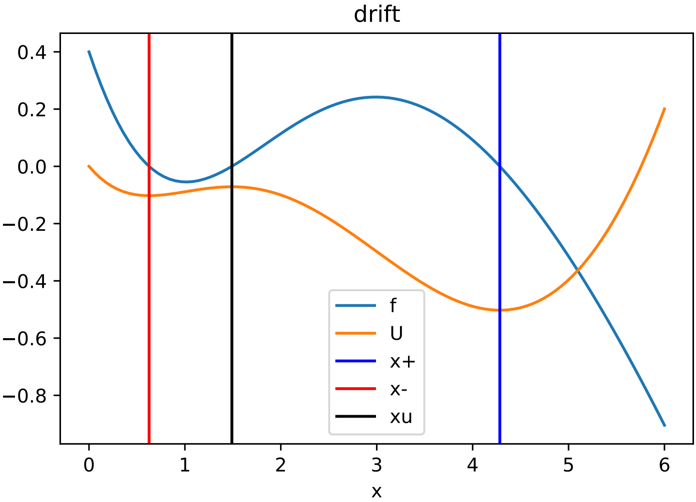
# The Biology




$$\frac{d[\text{TF A}]}{dt} = \frac{k_{1,f}[\text{TF A}]^2}{[\text{TF A}]^2 + K_{1,d}} - k_{1,d}[\text{TF A}] + r_{1,\text{bas}} \tag{2}$$

Let [TF A] be written as $x$.

# Our Model

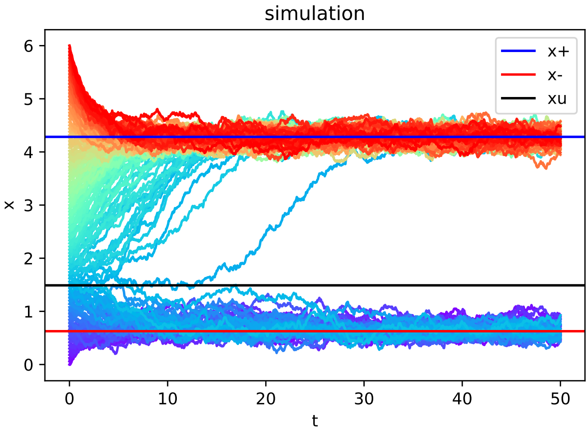


# Simulation

Given $f$ and $\sigma$, we simulate the SDE with the Euler-Maruyama method.

# Simulation

Given $f$ and $\sigma$, we simulate the SDE with the Euler-Maruyama method.



# Learn our Model

Given this simulated data, we wish to find learn the underlying drift and diffusion

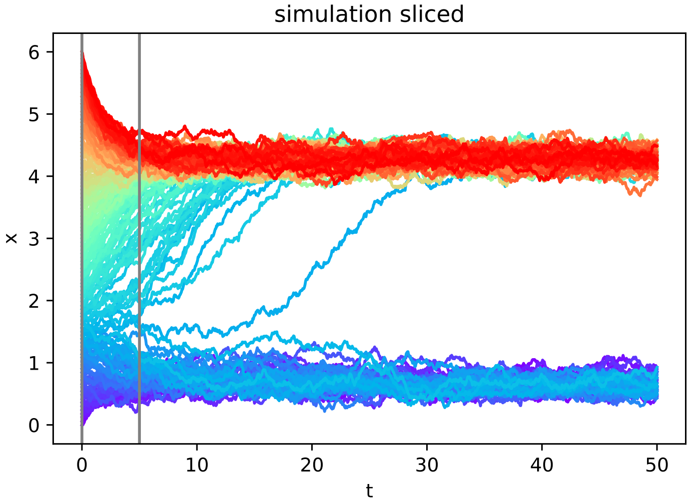
# Learn our Model

Given this simulated data, we wish to find learn the underlying drift and diffusion

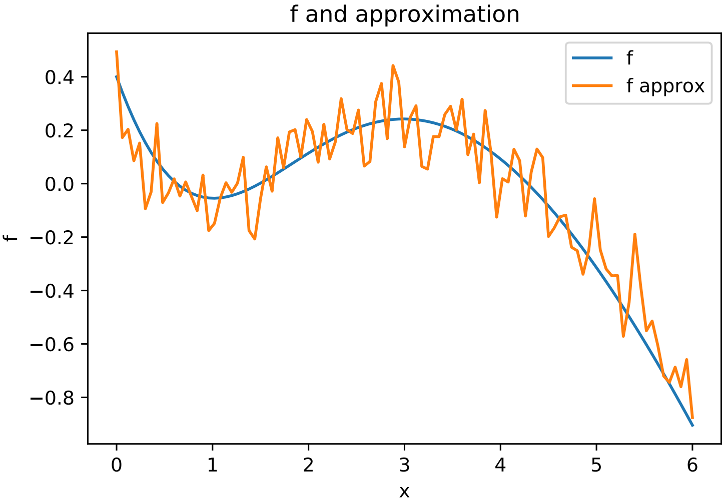
$$f(x) = \lim_{\Delta t \to 0} \mathbb{E}\left( \frac{X_{\Delta t} - X_0}{\Delta t} \middle| X_0 = x \right) \tag{3}$$

$$\sigma^2(x) = \lim_{\Delta t \to 0} \mathbb{E}\left( \frac{(X_{\Delta t} - X_0)^2}{\Delta t} \middle| X_0 = x \right) \tag{4}$$

# Simulation Zoom-In



# Learned $f$



# Most Probable Pathway

Starting at the stable concentration $x_-$ of transcription factor, can we jump to $x_+$?

# Most Probable Pathway

Starting at the stable concentration $x_-$ of transcription factor, can we jump to $x_+$?
If so, what is the most probable path that will be taken?

# Most Probable Pathway

Starting at the stable concentration $x_-$ of transcription factor, can we jump to $x_+$?
If so, what is the most probable path that will be taken?

$$\ddot{z} = \frac{\sigma(z)^2}{2} f''(z) + f'(z) f(z). \tag{5}$$

with the condition that $z(t_0) = x_-$ and $z(t_f) = x_+$.

# Most Probable Pathway

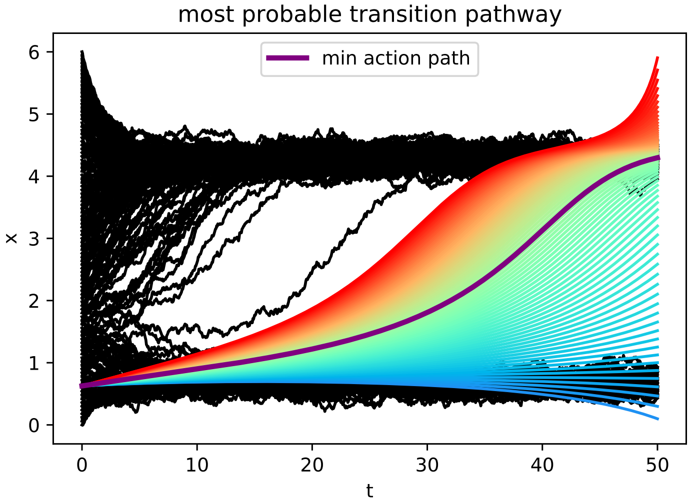
Starting at the stable concentration $x_-$ of transcription factor, can we jump to $x_+$?
If so, what is the most probable path that will be taken?

$$\ddot{z} = \frac{\sigma(z)^2}{2} f''(z) + f'(z) f(z). \tag{5}$$

with the condition that $z(t_0) = x_-$ and $z(t_f) = x_+$.

- Solving analytically is very difficult
- The shooting method lets us find the "velocity" that minimizes the "loss"

# Learned Probable Pathway

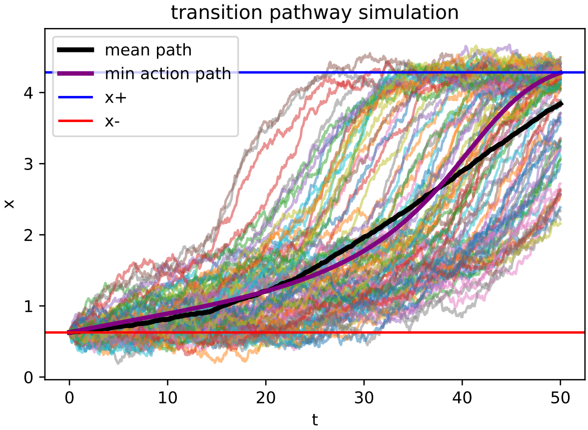


# Mean Path vs Learned Path

Running the simulation over $50,000$ times we obtain 93 runs that jump from $x_-$ to $x_+$. That is only $0.186\%$ of the runs.

# Mean Path vs Learned Path

Running the simulation over $50,000$ times we obtain 93 runs that jump from $x_{-}$ to $x_{+}$. That is only 0.186% of the runs.



# Future Work

- Fine tune/optimize specific data extraction
- Applying machine learning techniques to pathway distribution
- Consider higher dimensional data: consider the second transcription factor TF-R

# 3D Pathway Distribution