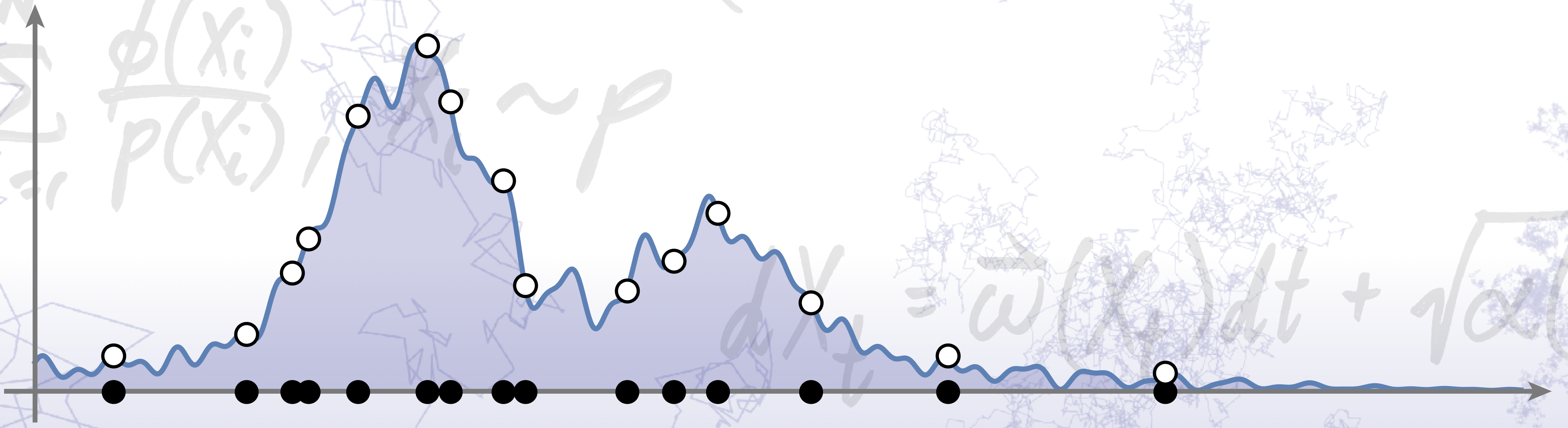
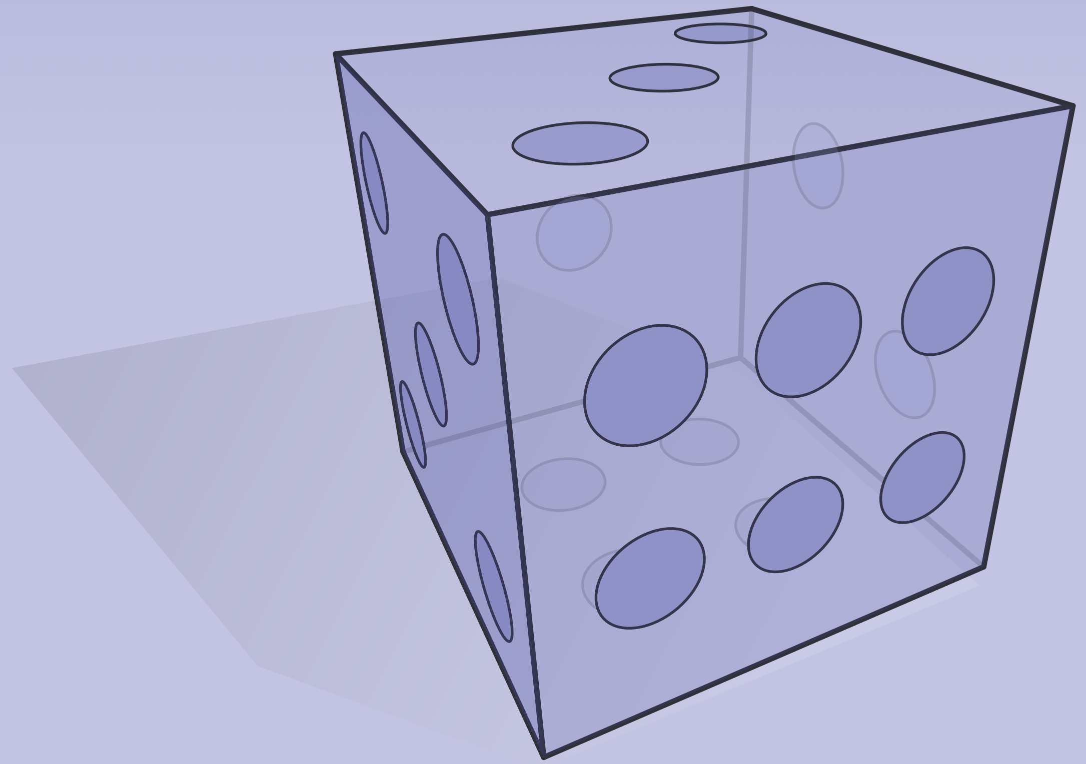
Monte Carlo Methods and Applications



Carnegie Mellon University | 21-387 / 15-327 / 15-627 / 15-860 | Fall 2023

Lecture 11
Stochastic Differential Equations



Monte Carlo Methods
and Applications

Carnegie Mellon University | 21-387 | 15-327 | 15-627 | 15-860 | Fall 2023

Deterministic Motion

depends on history

Brownian Motion

independent of history

Overview—SDEs & PDEs

- **Ordinary & Stochastic Differential Equations** (this lecture)
  - how do we describe systems evolving *over time*? **(ODEs)**
  - how do we incorporate randomness? **(SDEs)**
  - how do we simulate motion numerically?
- **Partial Differential Equations** (next lecture)
  - how do we describe systems evolving over *time & space?* **(PDEs)**
  - how do we simulate these systems numerically?
- SDE $\longleftrightarrow$ PDE connection
  - Somewhat surprising perspective: can use **stochastic** ODEs to understand—*and simulate*—**deterministic** PDEs
  - …and vice-versa!

analogy: *trajectory of rock (+wind)*

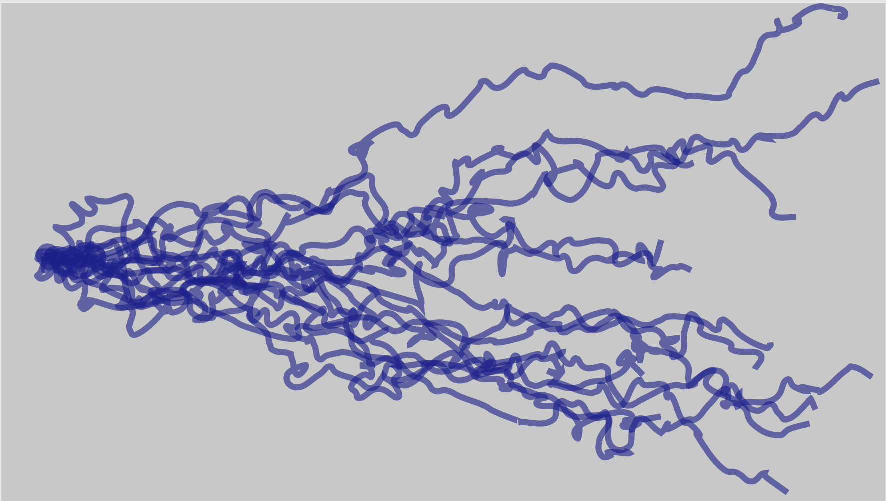
analogy: *ripples on pond*

Goal: Connect "microscopic" & "macroscopic"

Understand statements of two major **concepts**
and see how they can be used for **computation**.

Feynman-Kac formula

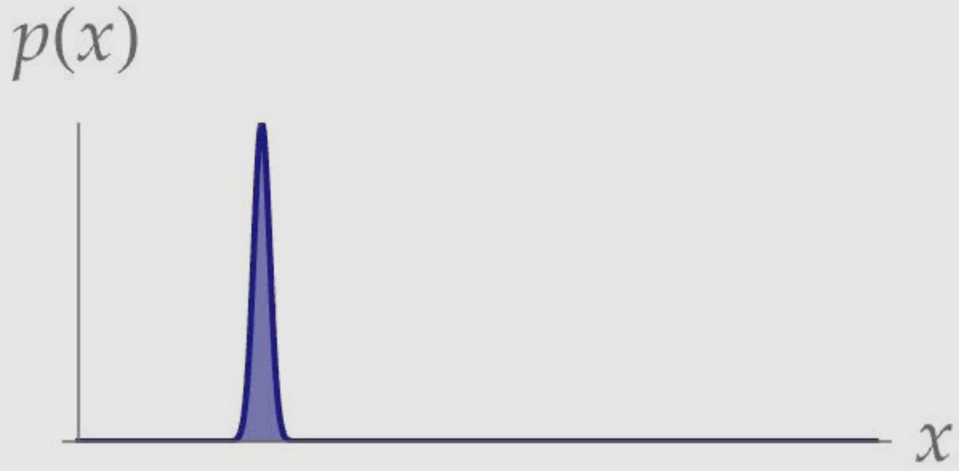
$$u(x) = \mathbb{E}\left[\int_0^T e^{-\int_0^t \sigma(X_s)ds} f(X_t)dt + e^{-\int_0^T \sigma(X_t)dt} g(X_t)\right]$$



use random walks to solve PDEs

Fokker-Planck equation

$$\frac{\partial p}{\partial t} = \alpha \Delta p - \nabla \cdot (p\omega)$$



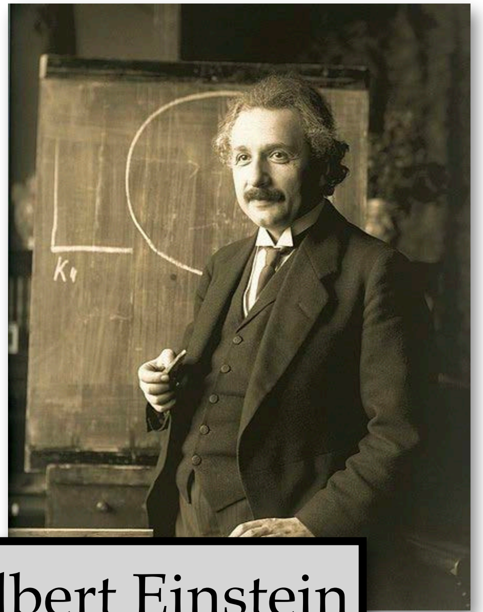
solve PDEs to model random walks

(Proofs that they're *true* will come later.)

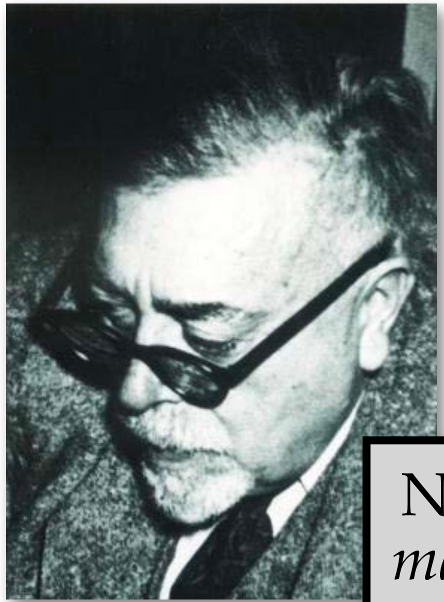
History of Brownian Motion

- **Brown's "life force"**
  - "spontaneous" motion of organic particles
  - …but also inorganic particles
- **Einstein's mystery:** how does random motion arise?
  - random "kicks" from water molecules are both too small, and too frequent
  - but occasionally random events "conspire" to give big kick in same direction
  - foundation of *statistical physics*
- **Wiener process**
  - formalize Brownian motion as a "non differentiable curve" (*Wiener process*)

Robert Brown
botanist

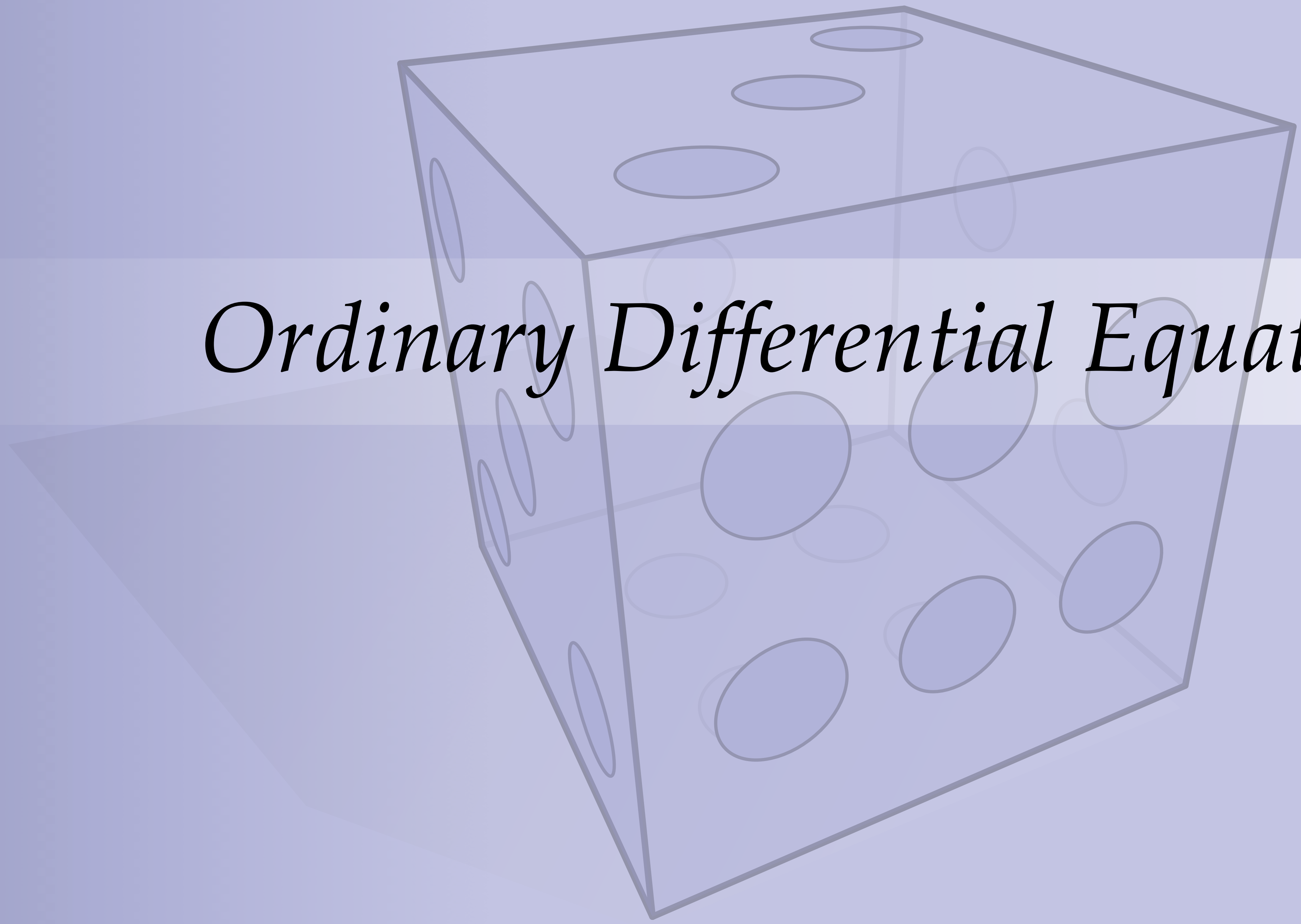


Albert Einstein
physicist



Norbert Wiener
mathematician &
computer scientist

Ordinary Differential Equations



Ordinary Differential Equations—Overview

- Differential equations *"lingua franca"* for phenomena appearing throughout nature, technology, & society
- Give an **implicit** description of quantities in terms of relative rates of change
  - *"if I change quantity A by a little bit, how much does quantity B change?"*
- Very different from an **explicit** description
  - *"what are the actual values of A or B?"*
- Basic task in mathematics & computation is therefore to solve for **explicit** values, given **implicit** description

You've probably already done this in your intro physics class! (Solve *"F=ma"*)

Ordinary Differential Equation

An **ordinary differential equation** is any equation of the form

$$F\left(t, x, x', \ldots, x^{(n)}\right) = 0$$

where F is any function of the (unknown) function $x(t)$ and its first n derivatives in time.

We say this ODE is:

– *nth order in time* (or simply *nth order*)

– *linear* (or *nonlinear*) if F is a linear (or nonlinear) function of its inputs

Example — 1st-order Linear ODEs

Simple but important example:

some constant

$$\frac{d}{dt}x(t) = ax(t)$$

"the function is proportional to its derivative"

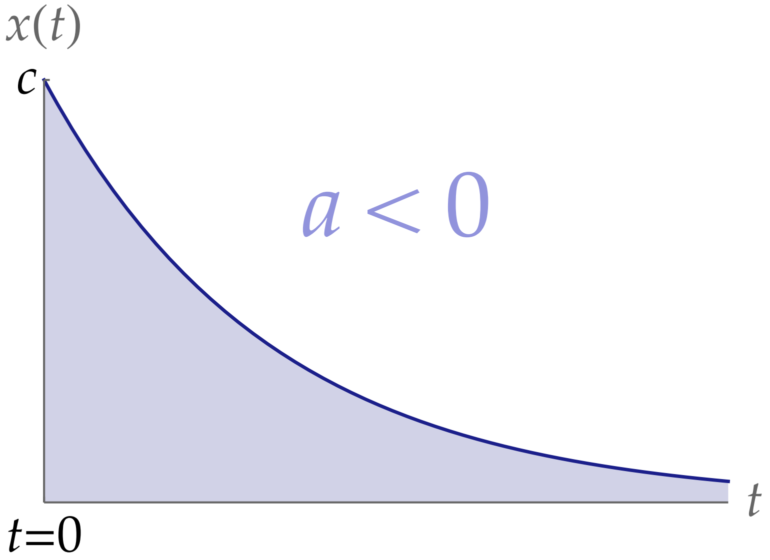
Q: Solution?

initial value

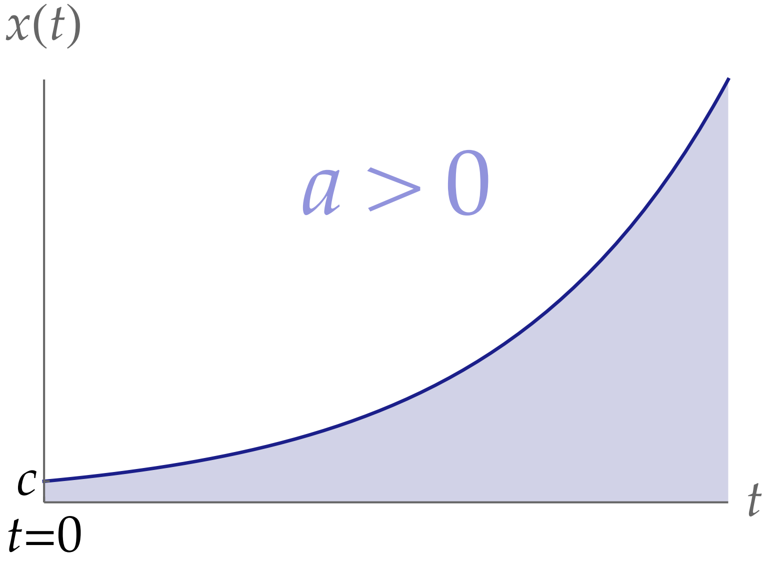
$$x(t) = \boxed{c}e^{at}$$

Check:

$$\frac{d}{dt}ce^{at} = ace^{at} = ax(t) \checkmark$$



exponential decay *(e.g., caffeine in blood)*



exponential growth *(e.g., bacteria on food)*

1st-order Linear ODEs—General Solution

More generally, 1st-order linear ODE has the form

$$ax'(t) + bx(t) + ct + d = 0,$$

highest-order derivative

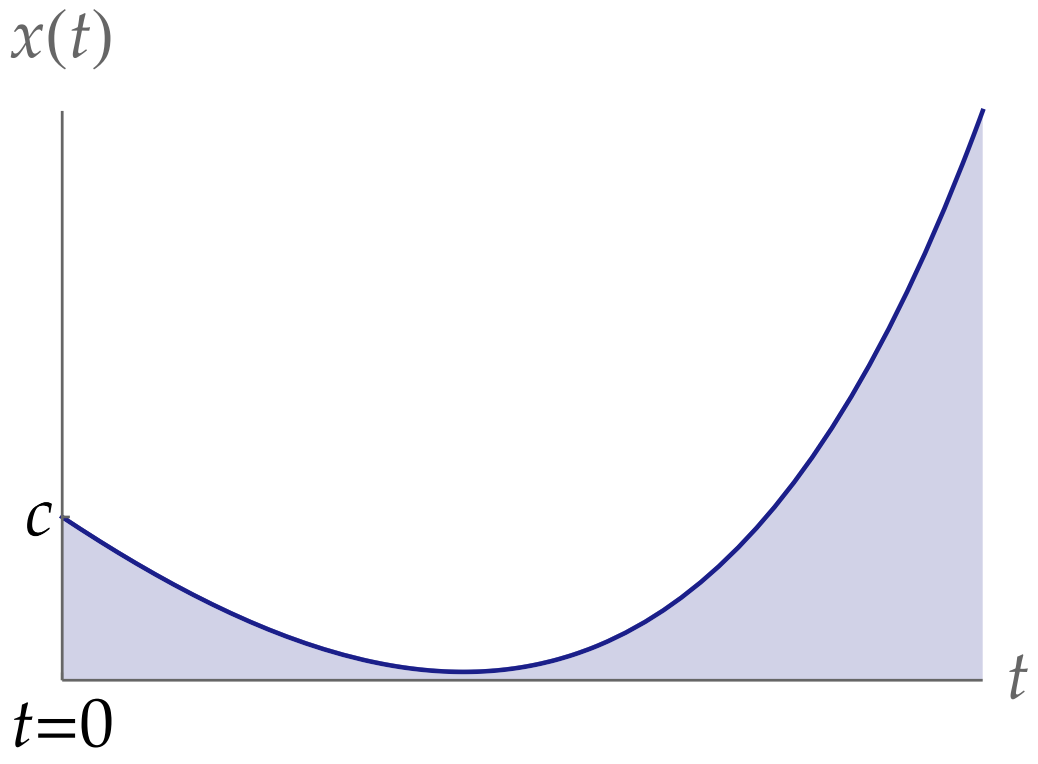
$$a, b, c, d \in \mathbb{R}, a \neq 0$$

still dominated by exponential growth (for $b > 1$)

Solution:

$$se^{-bt/a} + \frac{ac - b(d+ct)}{b^2},$$

$$s \in \mathbb{R}$$



Trivial Example—0th Order ODEs

Q: By the way, why didn't we start with 0th-order ODEs? :-)

Example.

$$x(t)^2 = bt + c \qquad \Longrightarrow \qquad x(t) = \pm\sqrt{bt+c}$$

Example.

$$\sin(x(t)) = at \quad \Longrightarrow \quad x(t) = \arcsin(at)$$

A: Because 0th-order "differential" equations are just *equations!*
(No relationship between different moments in time…)

Example—Projectile Motion

Quite famous ODE: **Newton's 2nd law of motion** (*"F=ma"*)

$$x''(t) = F/m$$

assuming force, mass are constant

2nd-order linear ODE

Q: Solution?

initial position *initial velocity*

$$x(t) = \boxed{x_0} + t\boxed{v_0} + \frac{F}{2m}t^2$$

in reality: a lot more complicated
(aerodynamic drag, spin of ball, wind, ...)

Systems of ODEs

One way to solve Newton's 2nd law: split into <u>system</u> of *1st-order* equations:

$$\boxed{x''(t) = F/m}$$

original ODE (2nd-order)

$$v(t) := x'(t)$$

think of velocity v as <u>independent</u> quantity

$$\begin{cases} x'(t) = v(t) \\ v'(t) = F/m \end{cases}$$

"couple" position x and velocity v into a <u>system</u> of ODEs

Now solve each linear equation in sequence:

$$v(t) = \frac{F}{m}t + c$$

determined by initial velocity: $v(0) = c$

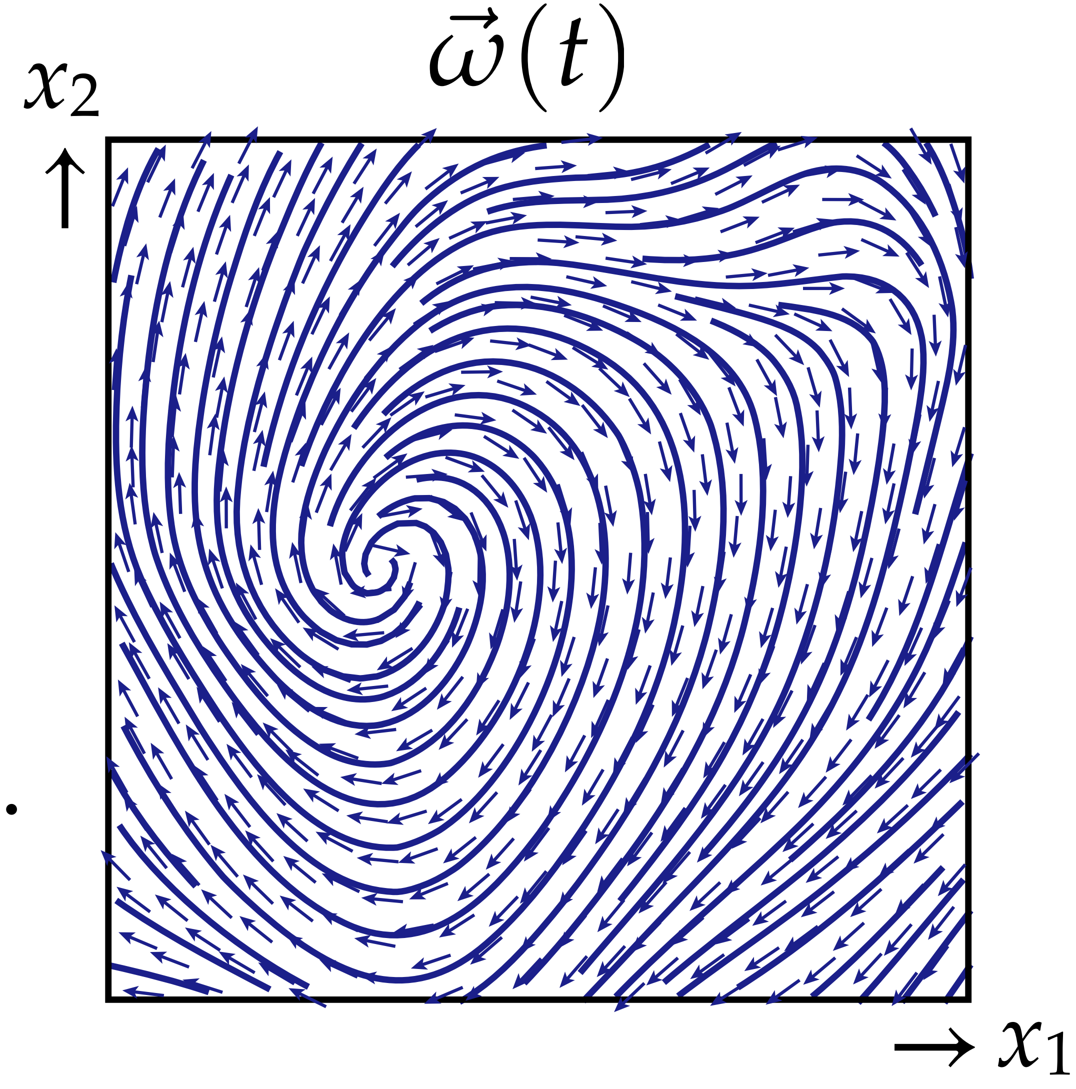
$$x(t) = \frac{F}{2m}t^2 + tv_0 + d$$

determined by initial position: $x(0) = d$

ODEs—Vector Field Perspective

In general, ODE in several variables $x(t) = (x_1(t), \ldots, x_n(t))$ can be viewed as "flow" along a vector field $\vec{\omega}$.

$$\underbrace{x'(t)}_{\text{change in position}} = \underbrace{\vec{\omega}(x(t))}_{\text{velocity vector field}}$$

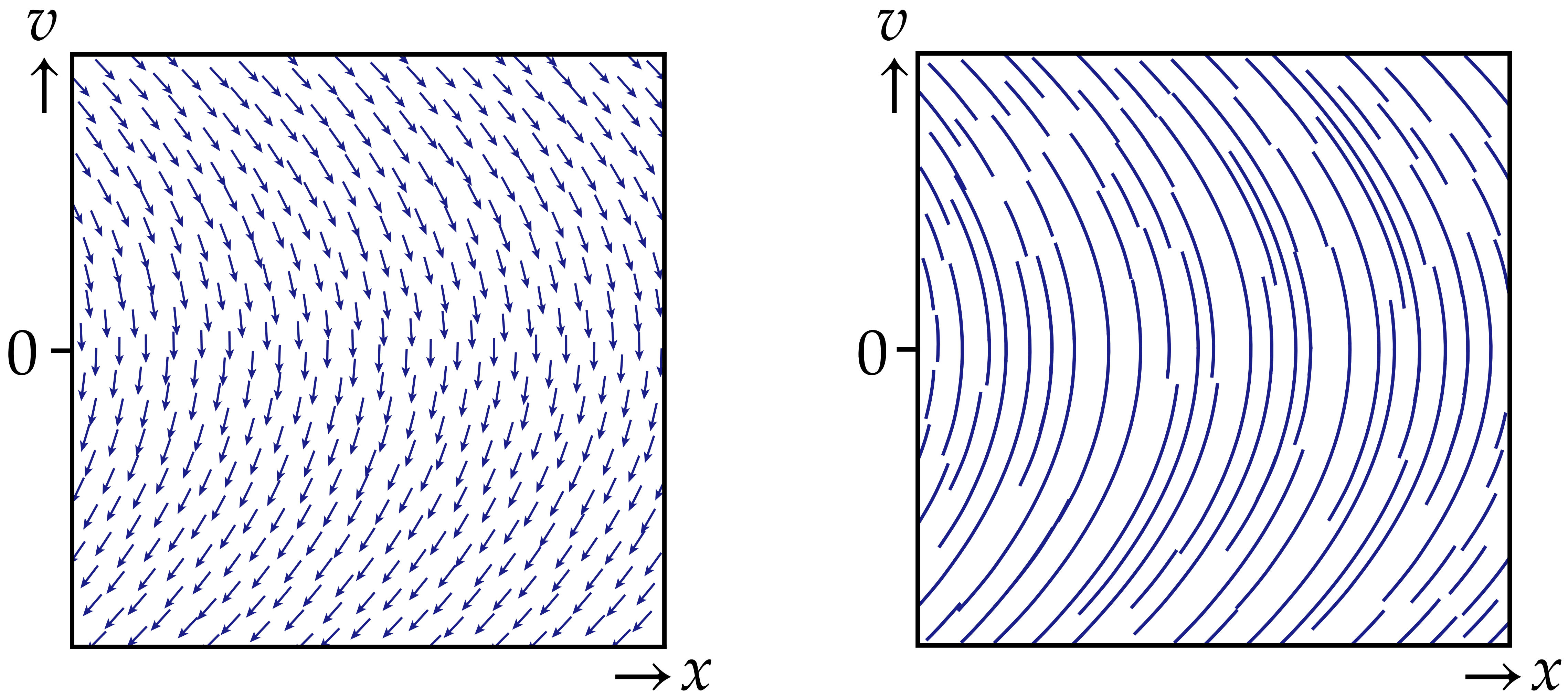


Solution corresponds to streamline of vector field, starting from initial conditions.

We'll use this visualization later to develop an understanding of SDEs…

Example — Projectile Motion

$$(x'(t), v'(t)) = \underbrace{(v(t), F/m)}_{\vec{\omega}}$$



Solving a System of Linear ODEs

Consider the system of linear 1st-order ODEs

$$\begin{aligned} x_1'(t) &= ax_1(t) + bx_2(t) \\ x_2'(t) &= cx_1(t) + dx_2(t) \end{aligned}$$

Can write in matrix form as

$$\underbrace{\begin{bmatrix} x_1'(t) \\ x_2'(t) \end{bmatrix}}_{x'(t)} = \underbrace{\begin{bmatrix} a & b \\ c & d \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}}_{x(t)}$$

$$x'(t) = Ax(t)$$

Helpful for understanding *infinitesimal generator* of stochastic process...

Q: What do you think the solution should be?

For a single ODE we had

$$x'(t) = ax(t) \implies x(t) = e^{at}x_0$$

So, perhaps unsurprisingly,

$$x(t) = e^{tA}x_0$$

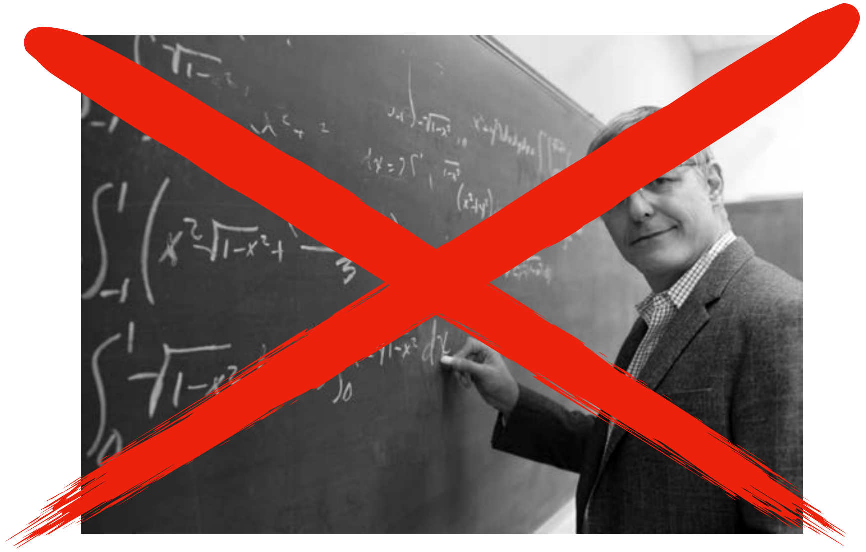
matrix exponential

$$e^A = \sum_{k=0}^{\infty} \frac{1}{k!} A^k$$

Numerical Integration of ODEs

Numerical Integration

- As usual, can't integrate most equations in closed form



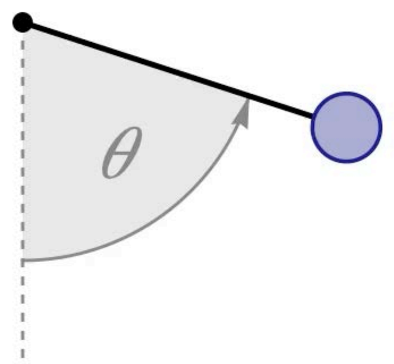
- Instead, use numerical "time stepping" to approximate solution
- General strategy:
  - replace derivatives with differences
  - solve for the unknowns!

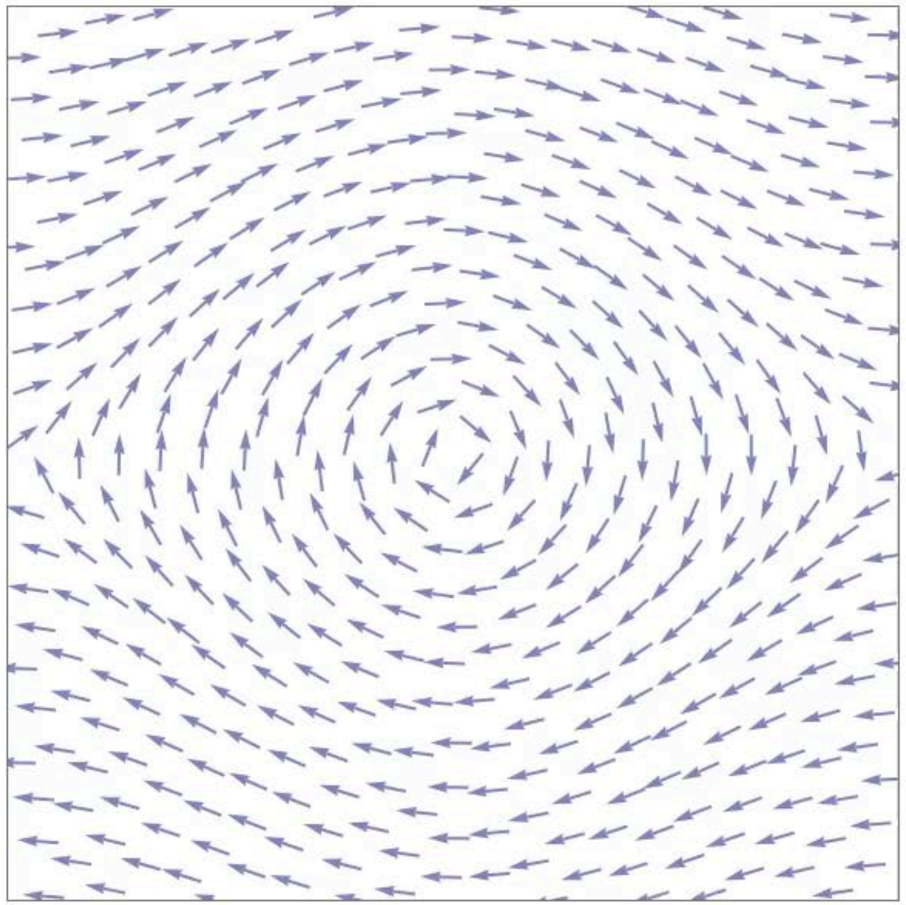
$$\boxed{\frac{dx}{dt} \Longrightarrow \frac{x_{t+\varepsilon} - x_t}{\varepsilon}}$$

- (This will also be the basis of the finite difference method for PDEs...)

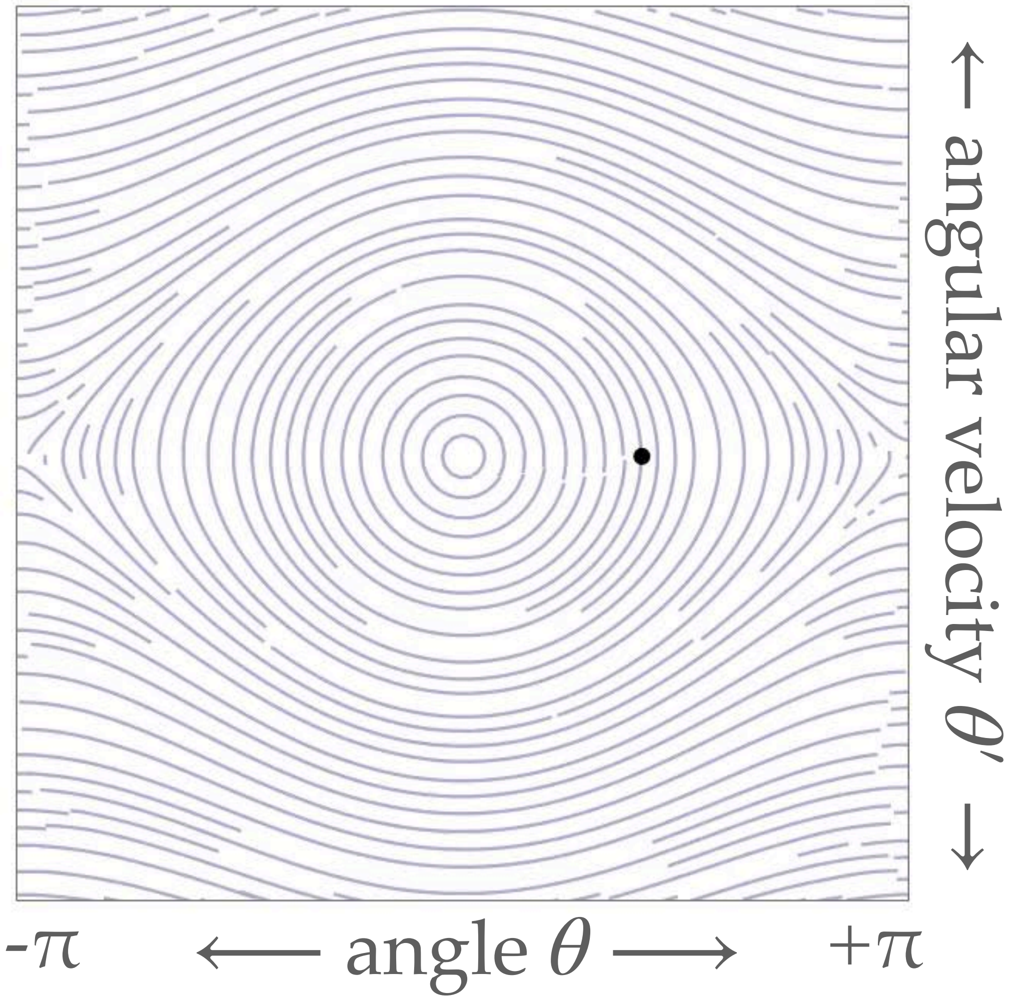
Running Example—Frictionless Pendulum

$$\theta''(t) = -\sin(\theta(t))$$





$\overrightarrow{\omega}(\theta, \theta') = (\theta', -\sin(\theta))$



Forward Euler

Consider any ODE of the form

$$\frac{d}{dt}x(t) = \vec{\omega}(x(t))$$

where $x(t)$ is an $\mathbb{R}^n$-valued function of time t, and the velocity $\vec{\omega}$ is a vector field on $\mathbb{R}^n$.

We can approximate the time derivative dx/dt by a difference

$$\frac{x(t+\varepsilon) - x(t)}{\varepsilon}, \quad \varepsilon > 0$$

Question: at which of the two points should we evaluate the velocity?

Forward Euler: assuming current point $x(t)$ is <u>known</u>, and next point x(t) is <u>unknown</u>, probably easiest to evaluate $\vec{\omega}$ at the known point.

Forward Euler (continued)

$$\tfrac{d}{dt}x(t) = \vec{\omega}(x(t))$$

$$\frac{x(t+\varepsilon) - x(t)}{\varepsilon} \approx \vec{\omega}(x(t))$$

Forward Euler (continued)

$$\frac{d}{dt}x(t) = \vec{\omega}(x(t)) \qquad\qquad x(t+\varepsilon) \approx x(t) + \varepsilon\,\vec{\omega}(x(t))$$

Suppose we have initial conditions $x(0) = x_0$.
Then we can repeatedly apply this approximation to get a sequence

forward Euler

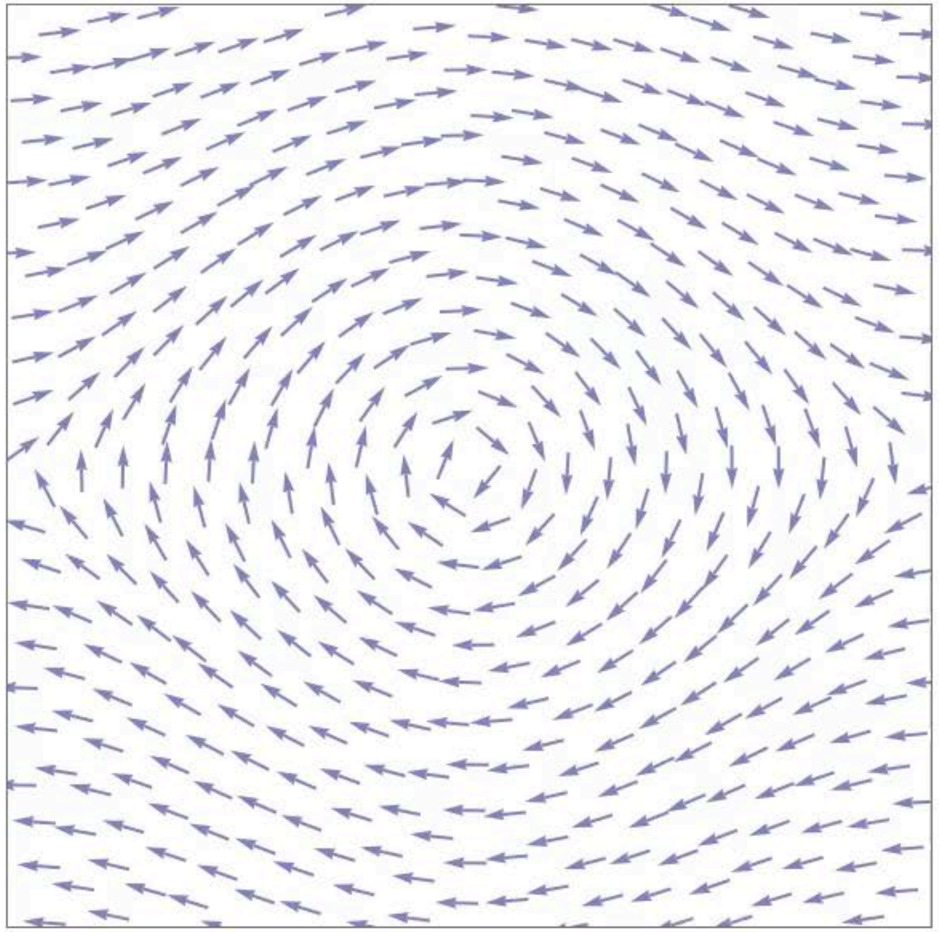
$$x_{k+1} = x_k + \varepsilon\vec{\omega}(x_k)$$

Intuition: to get the next state, just step a little along the direction of velocity…

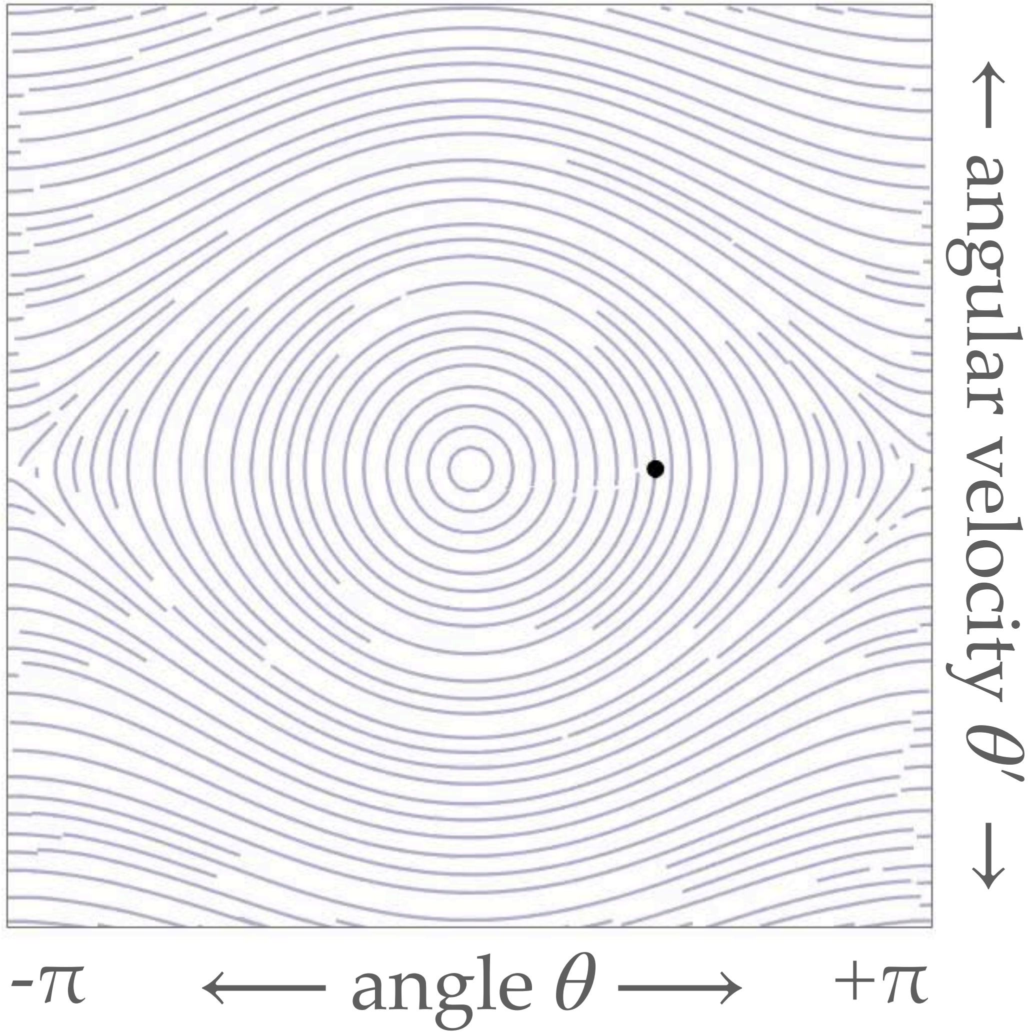
Pendulum — Forward Euler

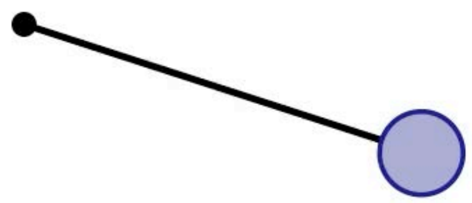
$$\boxed{\theta_{k+1} = \theta_k - \varepsilon \sin(\theta_k)}$$

Why does this happen?



velocity $\overrightarrow{\omega}$





Forward Euler—Stability Analysis

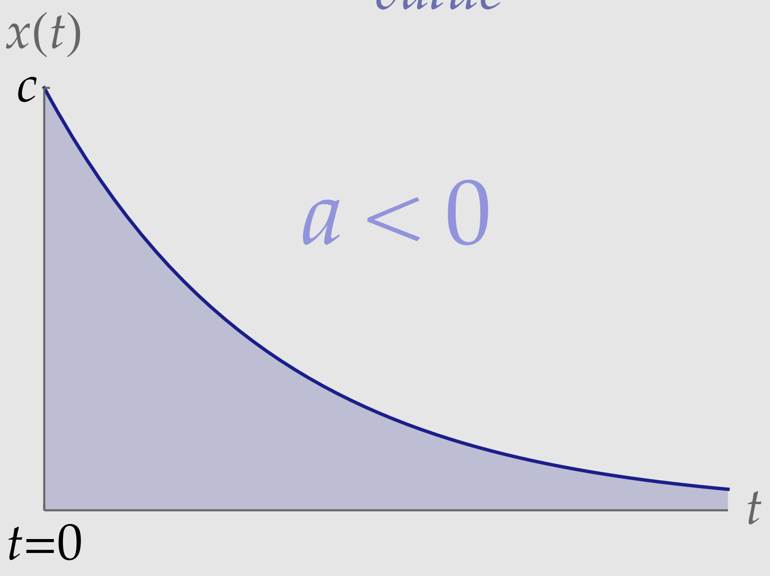
Consider a simpler (linear) problem:

exponential decay

$$\frac{d}{dt}x(t) = ax(t)$$

$$x(t) = \boxed{c}e^{at}$$

initial value



<u>forward Euler</u>

$$\begin{aligned} x_{k+1} &= x_k + \varepsilon a x_k \\ &= (1+\varepsilon a)x_k \\ &= (1+\varepsilon a)^{k+1} x_0 \end{aligned}$$

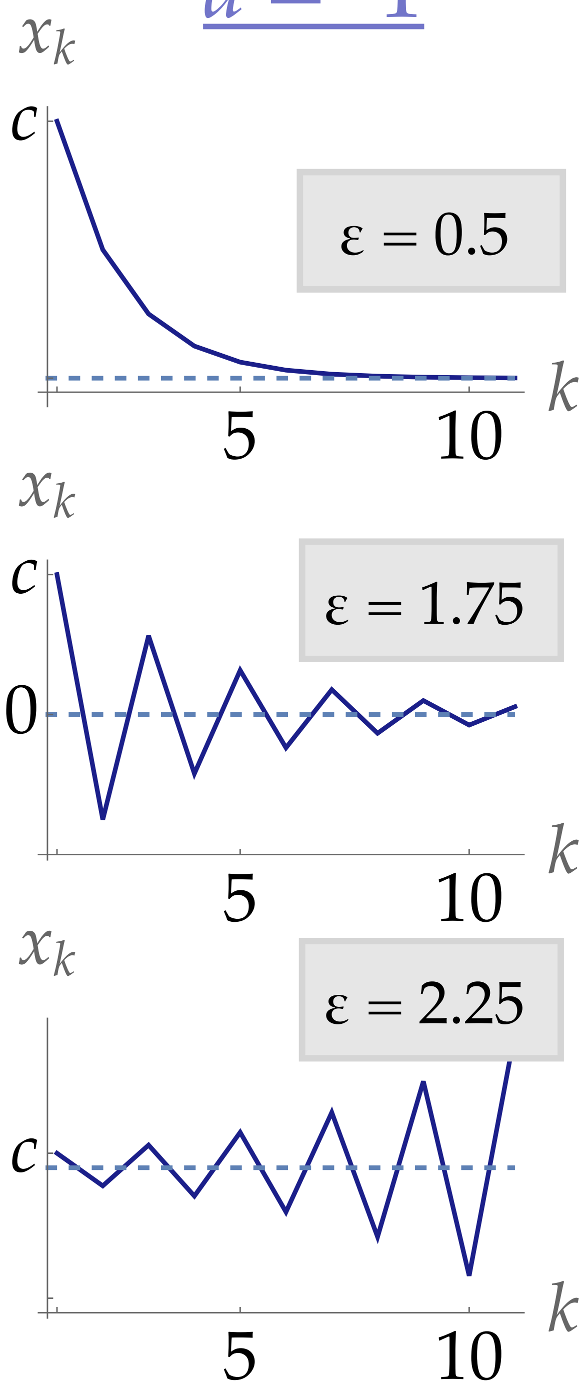
Q: will we always get <u>decay</u>?

A: No—must have $|1+\varepsilon a| < 1$.

Stay *monotonic*: $\varepsilon < 1/|a|$.

For general (nonlinear) ODE: bound ε in terms of eigenvalues of Jacobian at every point

<u>a = -1</u>



Backward Euler

Consider again any ODE

$$\frac{d}{dt}x(t) = \vec{\omega}(x(t))$$

where $x(t)$ is an $\mathbb{R}^n$-valued function of time t, and the velocity $\vec{\omega}$ is a vector field on $\mathbb{R}^n$.

Approximation of time derivative involves two points:

$$\frac{x(t+\varepsilon) - x(t)}{\varepsilon}, \ \varepsilon > 0$$

Question: what if we evaluate the velocity at $x(t+\varepsilon)$ instead of $x(t)$?

Backward Euler: even though next point $x(t+\varepsilon)$ is not known, we can still evaluate velocity "implicitly," *i.e.*, solve for a point $x(t+\varepsilon)$ such that the finite difference in time equals the velocity at $x(t+\varepsilon)$.

Backward Euler (continued)

$$\frac{d}{dt}x(t) = \vec{\omega}(x(t))$$

$$\frac{x(t+\varepsilon) - x(t)}{\varepsilon} \approx \vec{\omega}(x(t+\varepsilon))$$

Backward Euler (continued)

$$\frac{d}{dt}x(t) = \vec{\omega}(x(t)) \qquad\qquad x(t+\varepsilon) - \varepsilon\,\vec{\omega}(x(t+\varepsilon)) \approx x(t)$$

Suppose we have <u>initial conditions</u> $x(0) = x_0$.
Then we can repeatedly apply this approximation to get a sequence

backward Euler

$$x_{k+1} - \varepsilon\vec{\omega}(x_{k+1}) = x_k$$

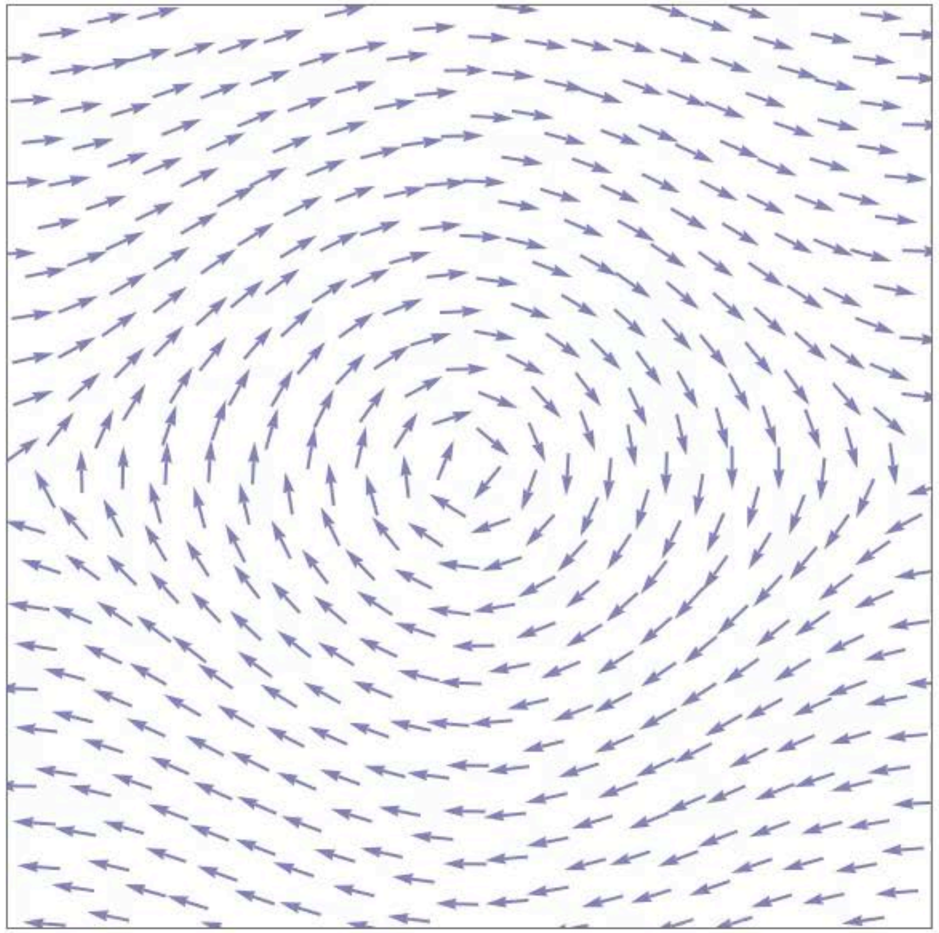
Summary: solve a (possibly nonlinear) equation for the next state.

Backward Euler

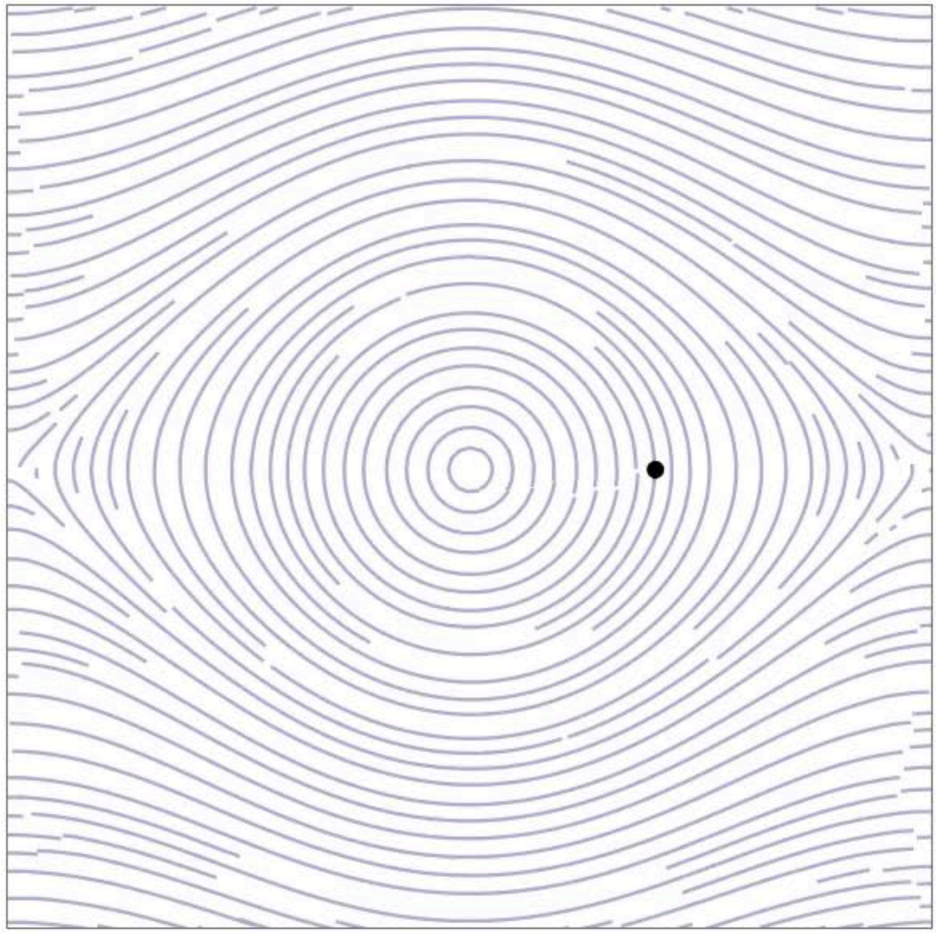
$$\boxed{\theta_{k+1} - \varepsilon \sin(\theta_{k+1}) = \theta_k}$$

solve via, e.g., Newton's method

Why does this happen?

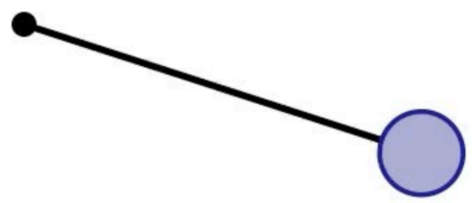


velocity $\overrightarrow{\omega}$



-π ⟵ angle θ ⟶ +π

↑ angular velocity θ' ↓



Backward Euler—Stability Analysis

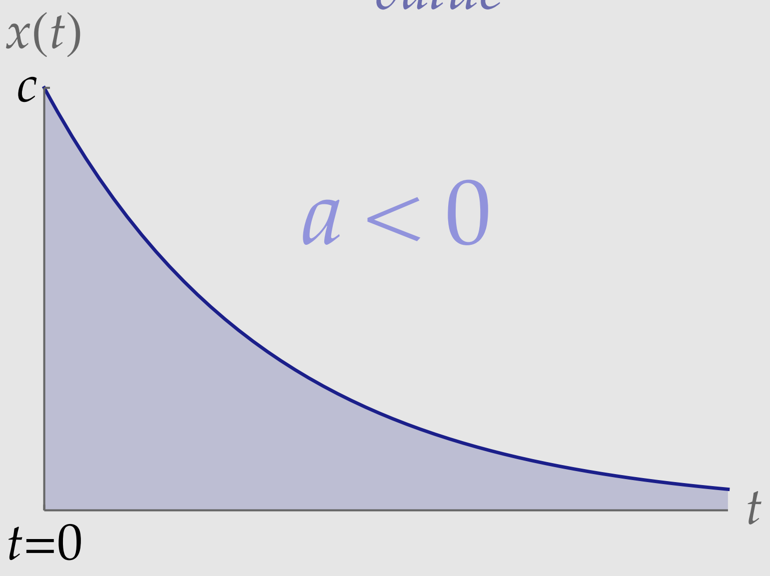
Consider a simpler (linear) problem:

exponential decay

$$\frac{d}{dt}x(t) = ax(t)$$

$$x(t) = \boxed{c}e^{at}$$

initial value



backward Euler

$$x_{k+1} - \varepsilon a x_{k+1} = x_k$$

$$\iff (1-\varepsilon a)x_{k+1} = x_k$$

$$\iff x_{k+1} = \frac{1}{(1-\varepsilon a)}x_k$$

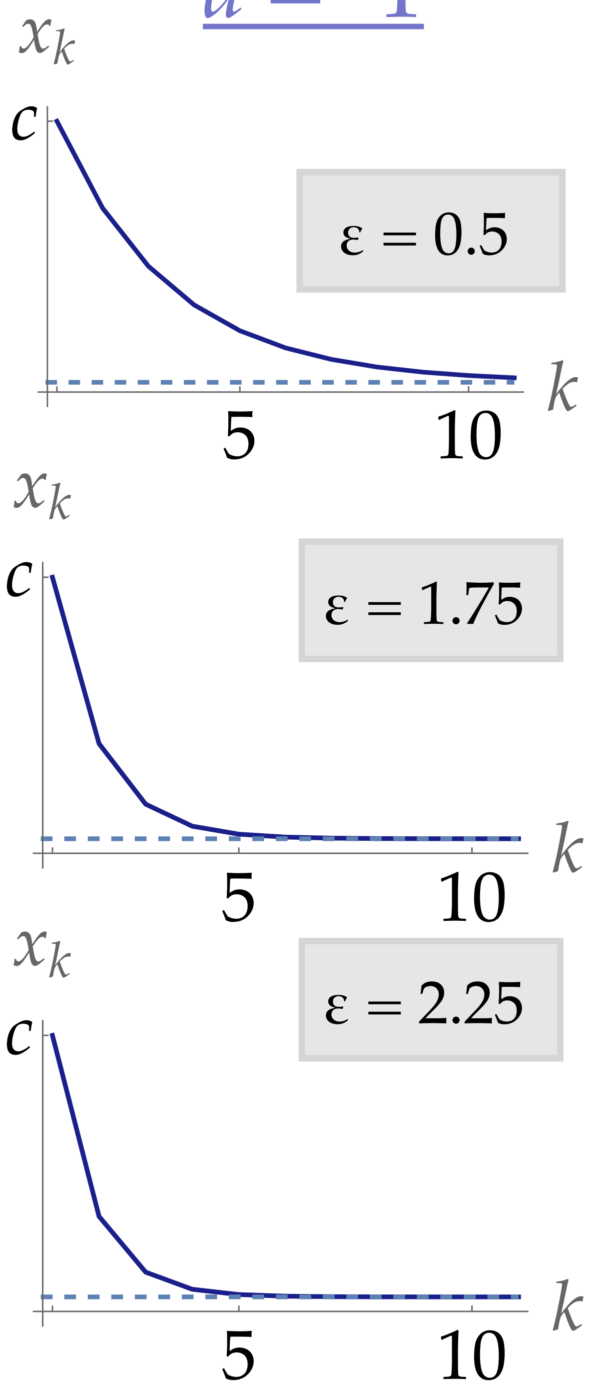
$$\Rightarrow x_{k+1} = \left(\frac{1}{1-\varepsilon a}\right)^{k+1} x_0$$

Q: will we always get decay?

A: Yes—since $a < 0$, $\varepsilon > 0$, factor always less than 1 (*"unconditionally stable"*)

But may be "over-damped!"

a = -1



Symplectic Euler

For ODEs arising from <u>dynamical systems</u> (e.g., Newton's 2nd law), another option:

– first, update velocity from **<u>old</u>** position

–then, update positions from **<u>new</u>** velocity

- For conservative systems (no friction, etc.) energy, momentum, etc., will not "drift" significantly up or down even over very long time scales

–exactly preserve *symplectic form* (sum of 2D phase-space areas in each dimension)

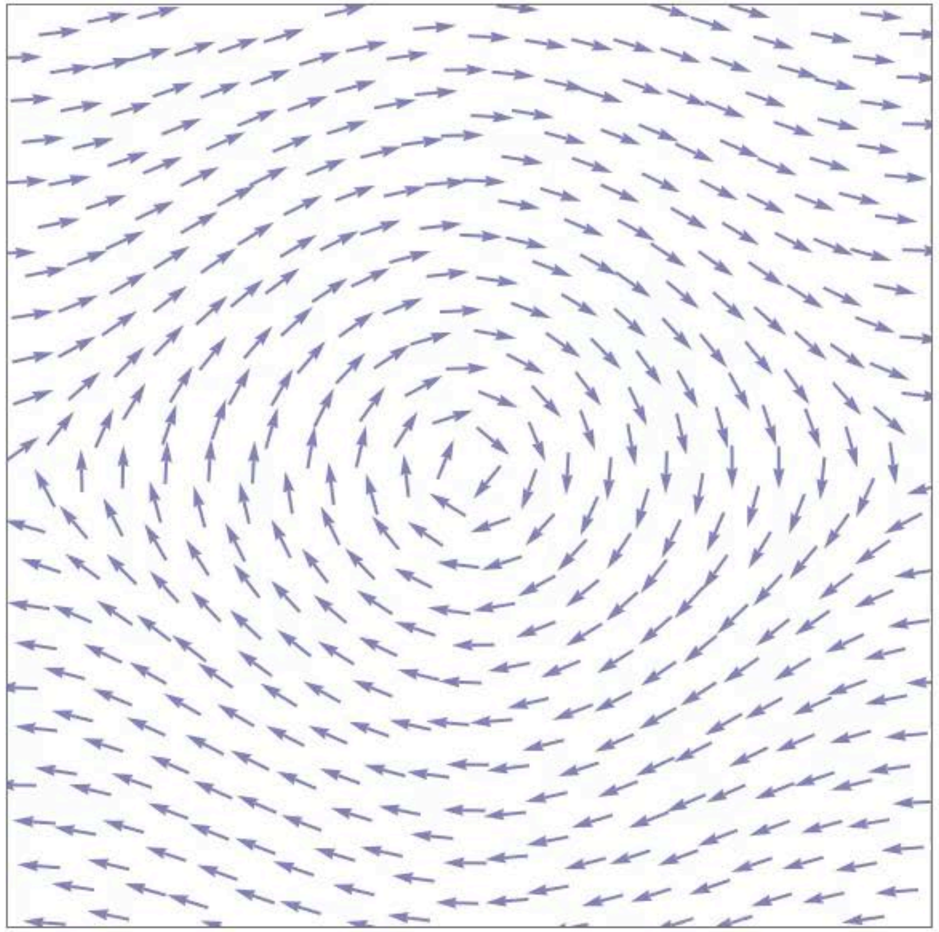
forward Euler **backward Euler** **symplectic Euler**

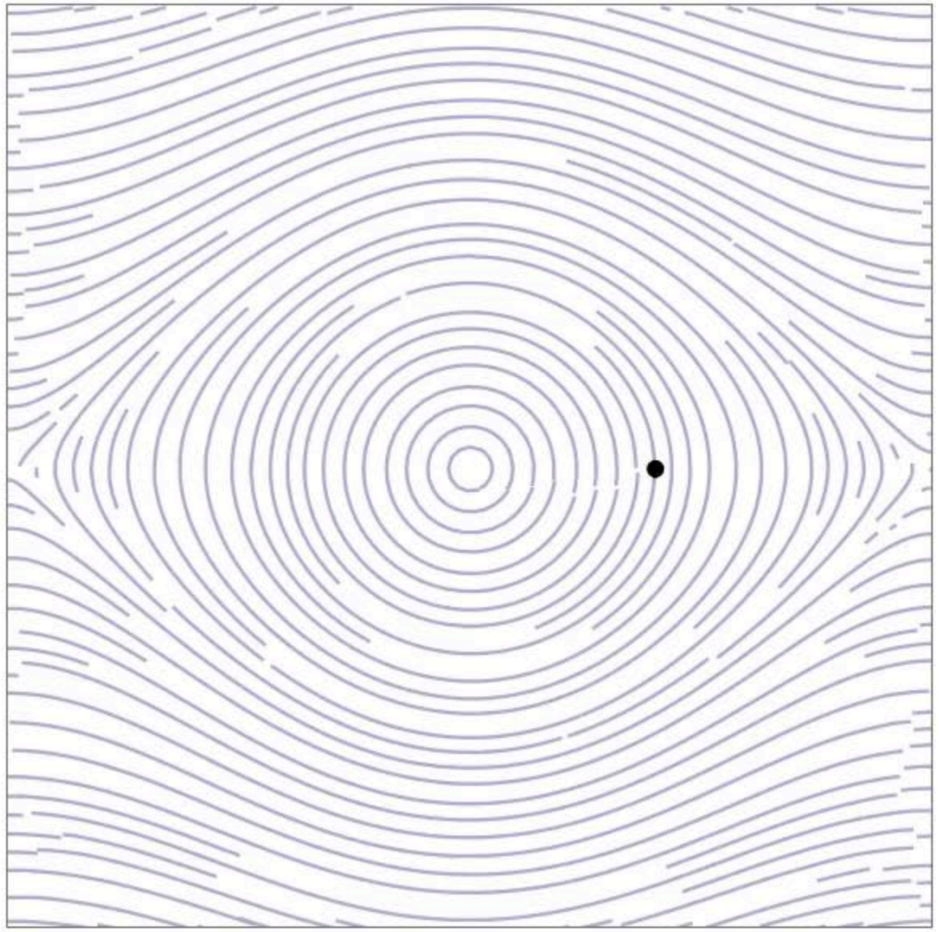
Symplectic Euler

$$\begin{aligned} \theta'_{k+1} &= \theta'_k - \varepsilon \sin(\theta_k) \\ \theta_{k+1} &= \theta_k + \varepsilon\theta'_{k+1} \end{aligned}$$

use new velocity to update old position

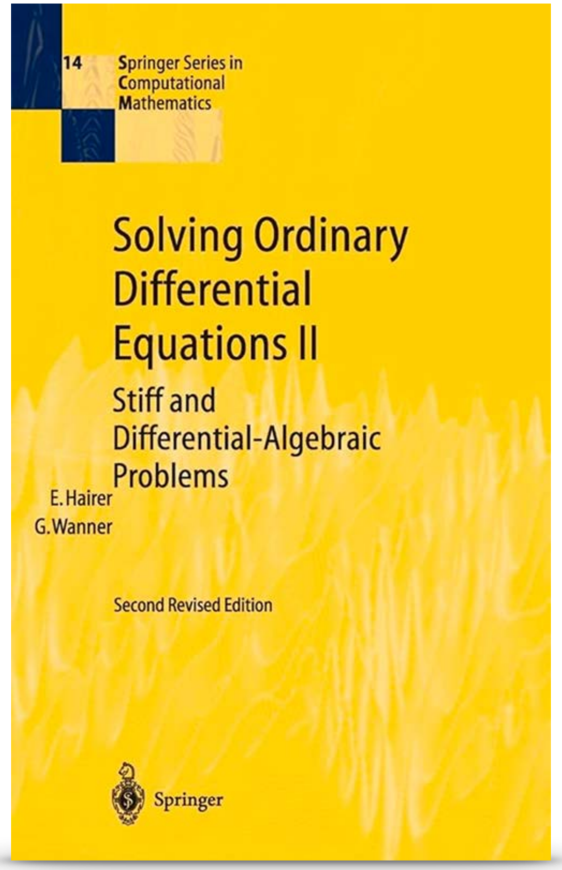
This will (provably) continue forever.

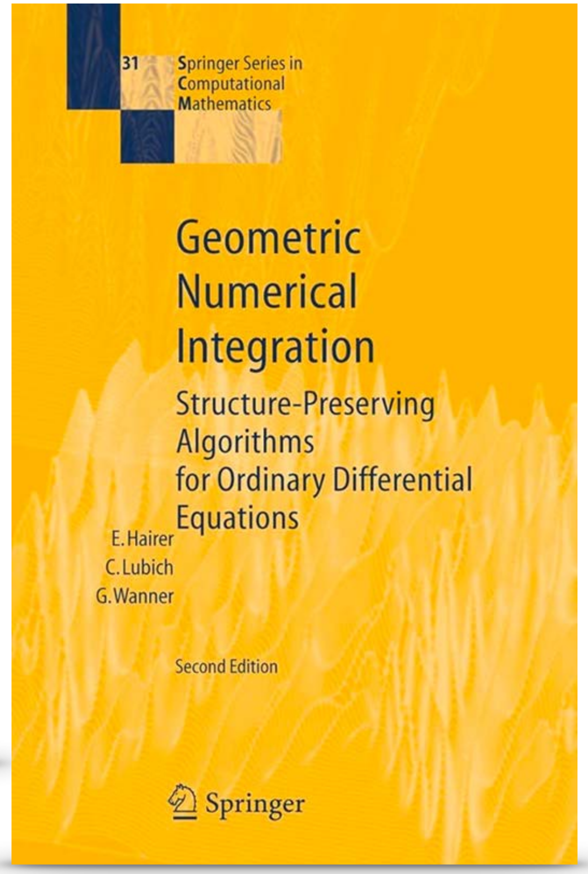




ODE Integration—Beyond the Basics

- A <u>lot</u> to say about numerical integrators beyond forward/backward Euler
- E.g., can we get the "right" behavior for systems more complex than pendulum?
  - Yes! can use **geometric** integrators like *symplectic Euler* to get good long-term behavior for many systems (dissipative, non-conservative forces, …)
- More generally, can improve integrator **accuracy**
  - Adams-Bashforth, Runge Kutta, …
  - less error per step, but error can still *accumulate* over long times

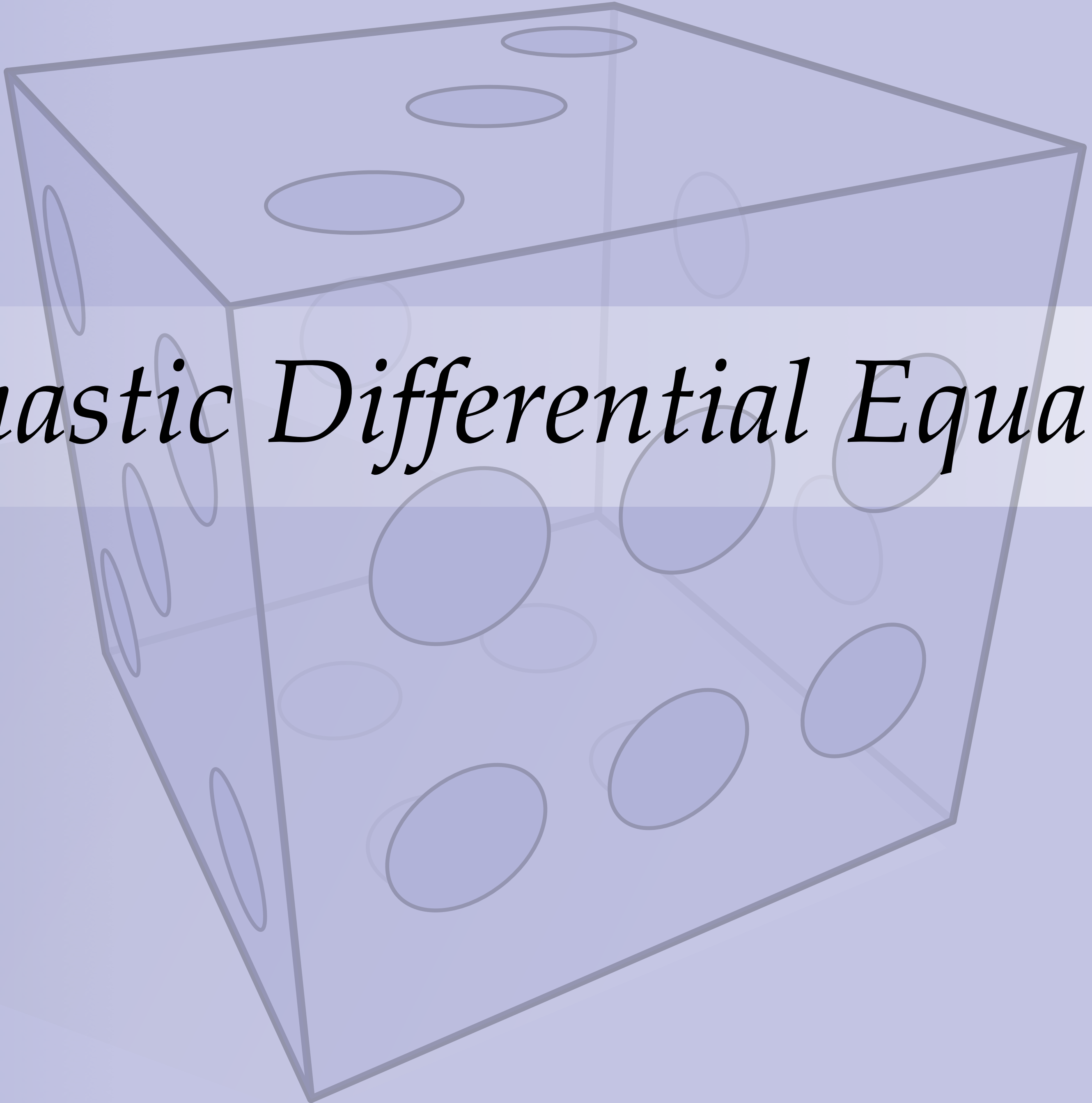




Numerical solution may not reflect reality!

`scipy.integrate`

Can often just invoke library functions *(but <u>please</u> understand what they do!)*



Stochastic Differential Equations

Stochastic Differential Equations—Overview

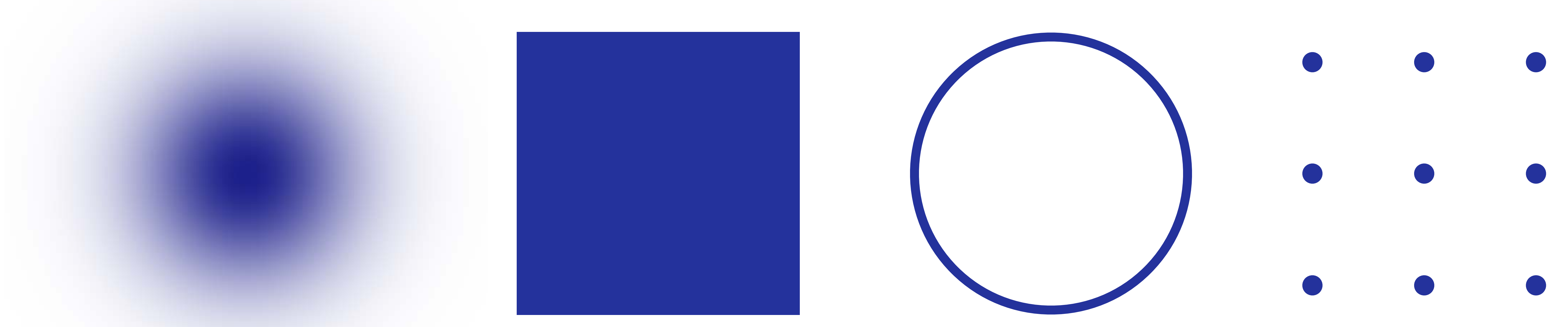
- Now that we understand how to describe functions in terms of their derivatives, can add randomness to the picture
- A few key pieces:
  - **Brownian motion** — basic notion of randomness for continuous functions
  - **Diffusion process** — more general class of "random functions" that connect to broader applications & algorithms
  - **Ito calculus**
    - **Ito's lemma** — basic notion of "stochastic differentiation"
    - **Ito integral** — basic notion of "stochastic integration"
  - **Numerical integrators** for SDEs

Stochastic Differential Equations—Motivation

- Consider particles jiggling in a water. What would it take to simulate this system using an ODE integrator?
- The issue is not merely that there are a lot of particles: to capture the "jiggling" motion, we'd also have to integrate ODEs for trajectories of a <u>huge</u> number of water molecules ($\sim 10^{23}$).
- If mass of particles is large—or fluid is very cold—motion due to thermal fluctuation is negligible, and we can just simulate projectile motion plus a linear drag force (linear ODE!)
- Otherwise, we have to actually model & simulate the forces that induce jiggling (*"Langevin force"*)

Brownian motion — Motivation

- Processes found in nature, finance, etc., have very different physical/dynamical origins
- Each one "jiggles around" according to a very different distribution $P(x_{k+1} \mid x_k)$
- For fun, let's simulate random walks using a few distributions p (centered at 0):

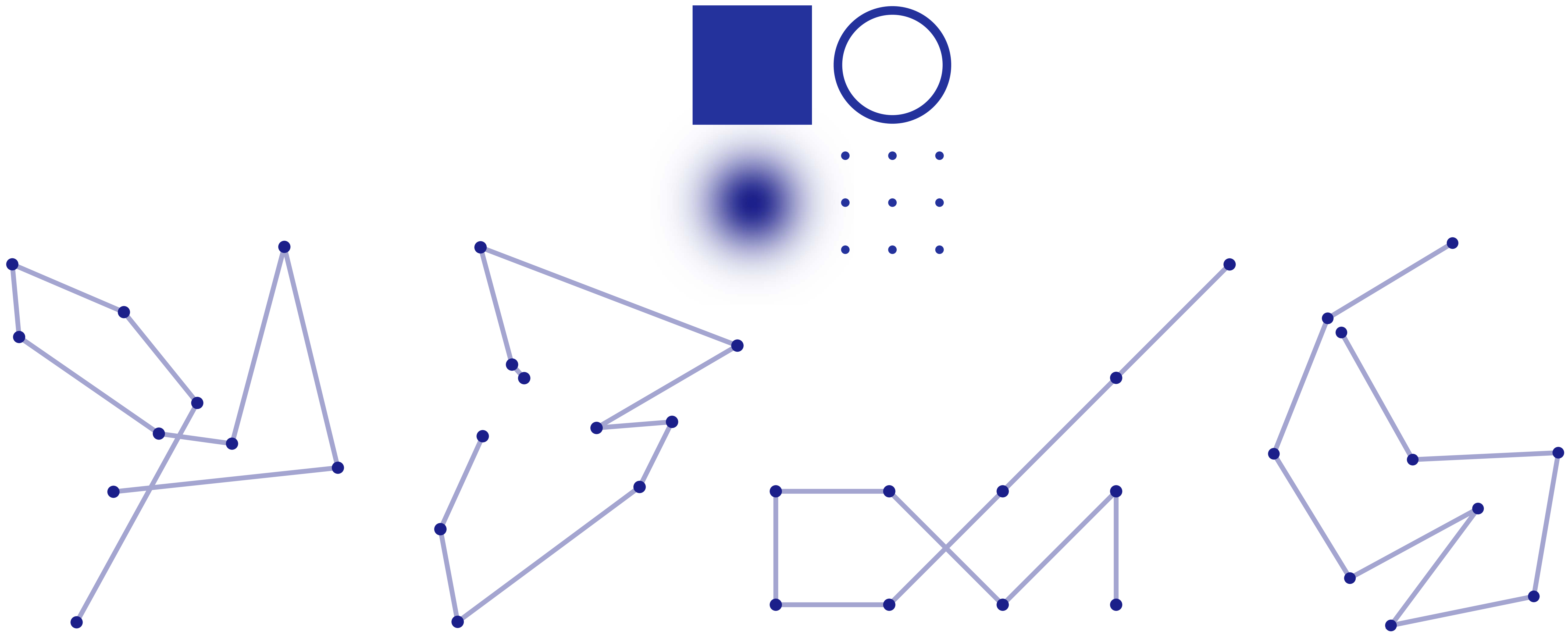


normal **square** **circle** **points**

ALGORITHM: $x_{k+1} \leftarrow x_k + \xi_k, \quad \xi_k \sim p \ \text{(i.i.d.)}$

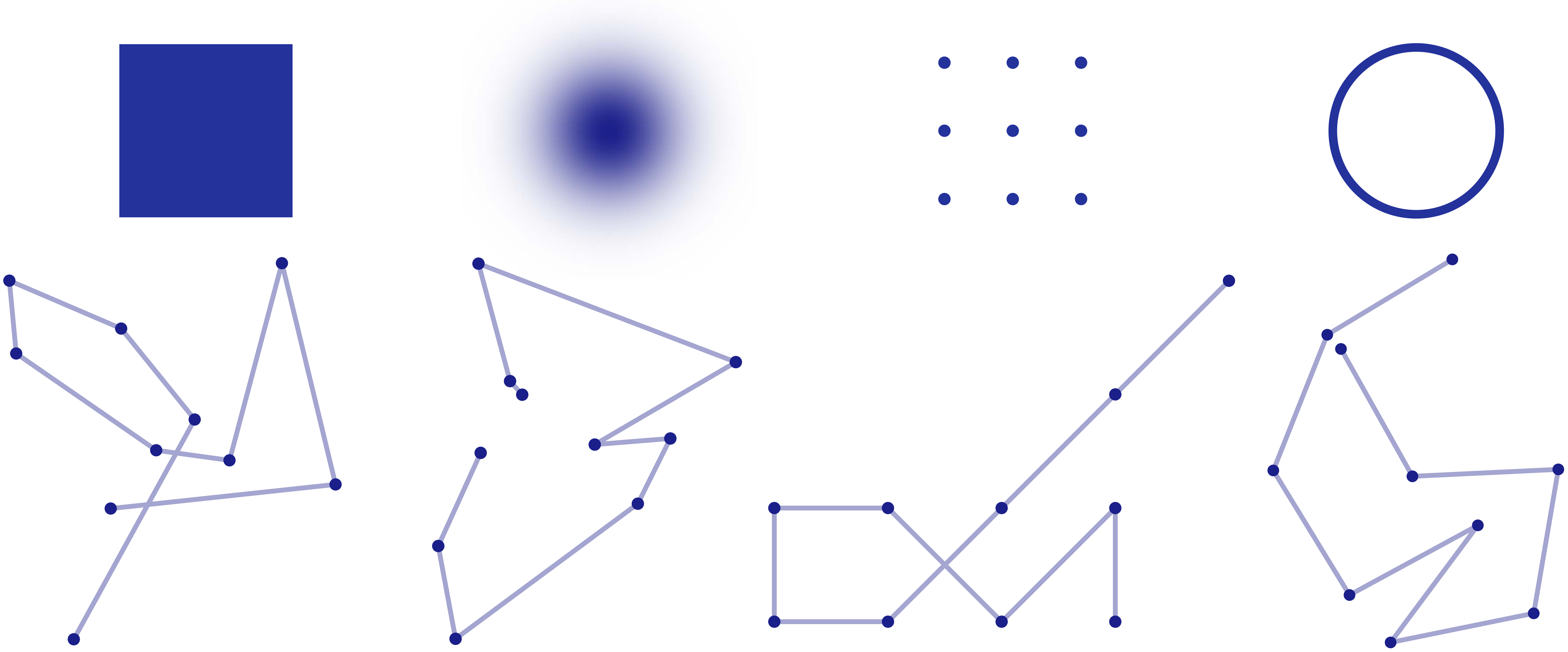
Random Walks — A Few Steps

Suppose we take 10 steps. Can you tell which walk comes from which distribution?



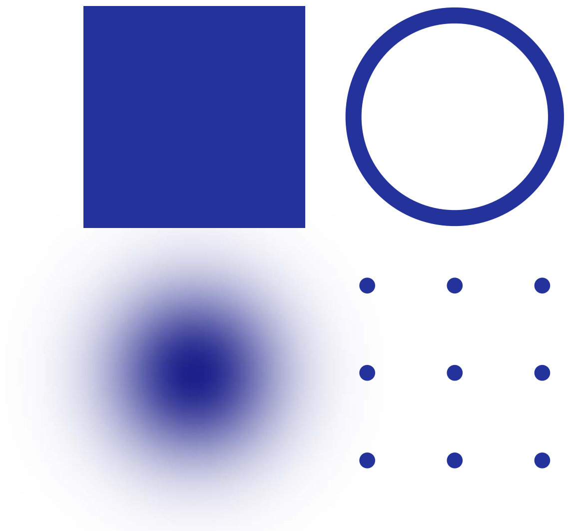
Random Walks — A Few Steps

Suppose we take 10 steps. Can you tell which walk comes from which distribution?



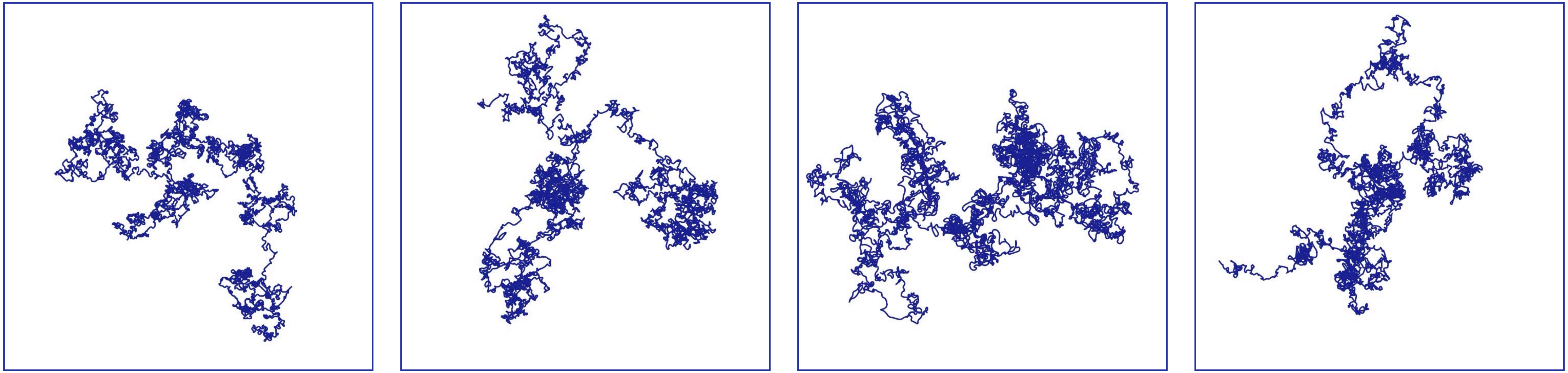
Random Walks—Many Steps

Suppose we now take 10,000 steps. Can you still tell which walk is which?



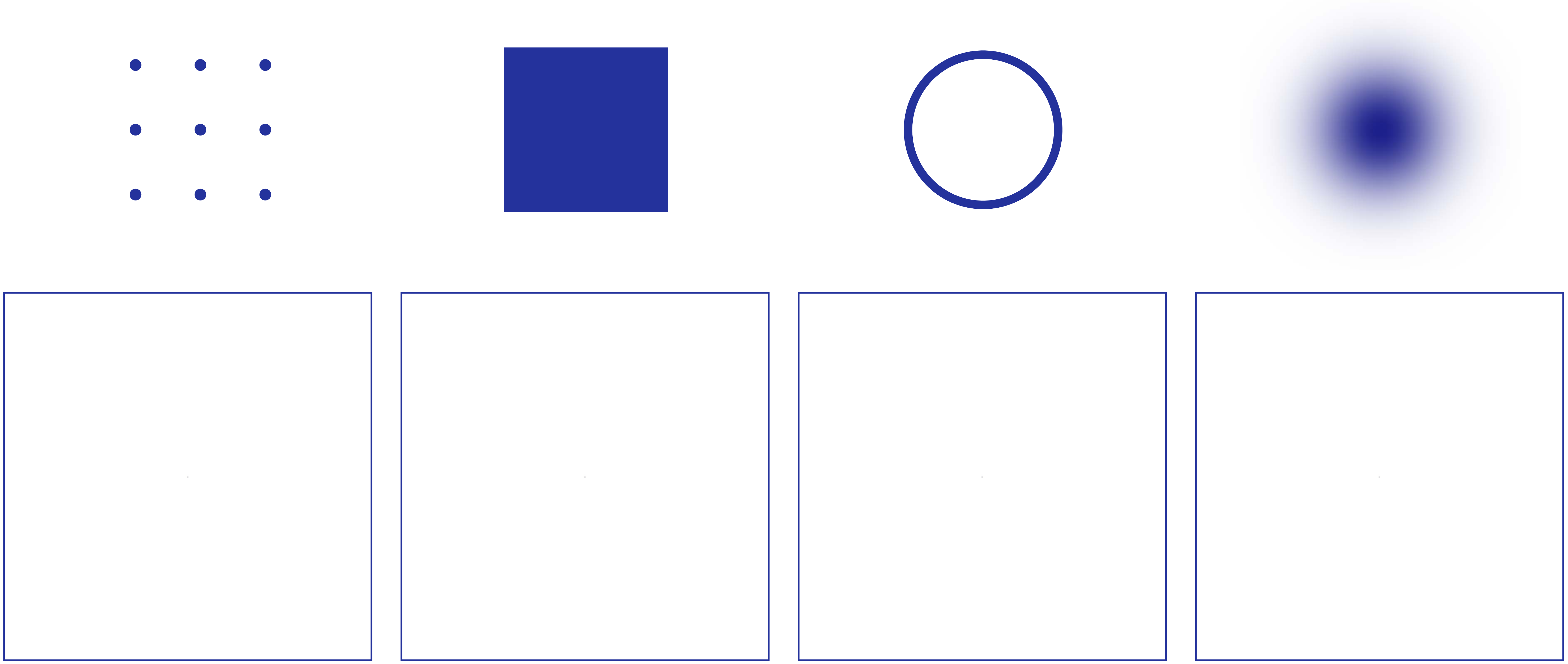
Random Walks—Many Steps

Suppose we now take 10,000 steps. Can you still tell which walk is which?



Random Walks—Zooming Out

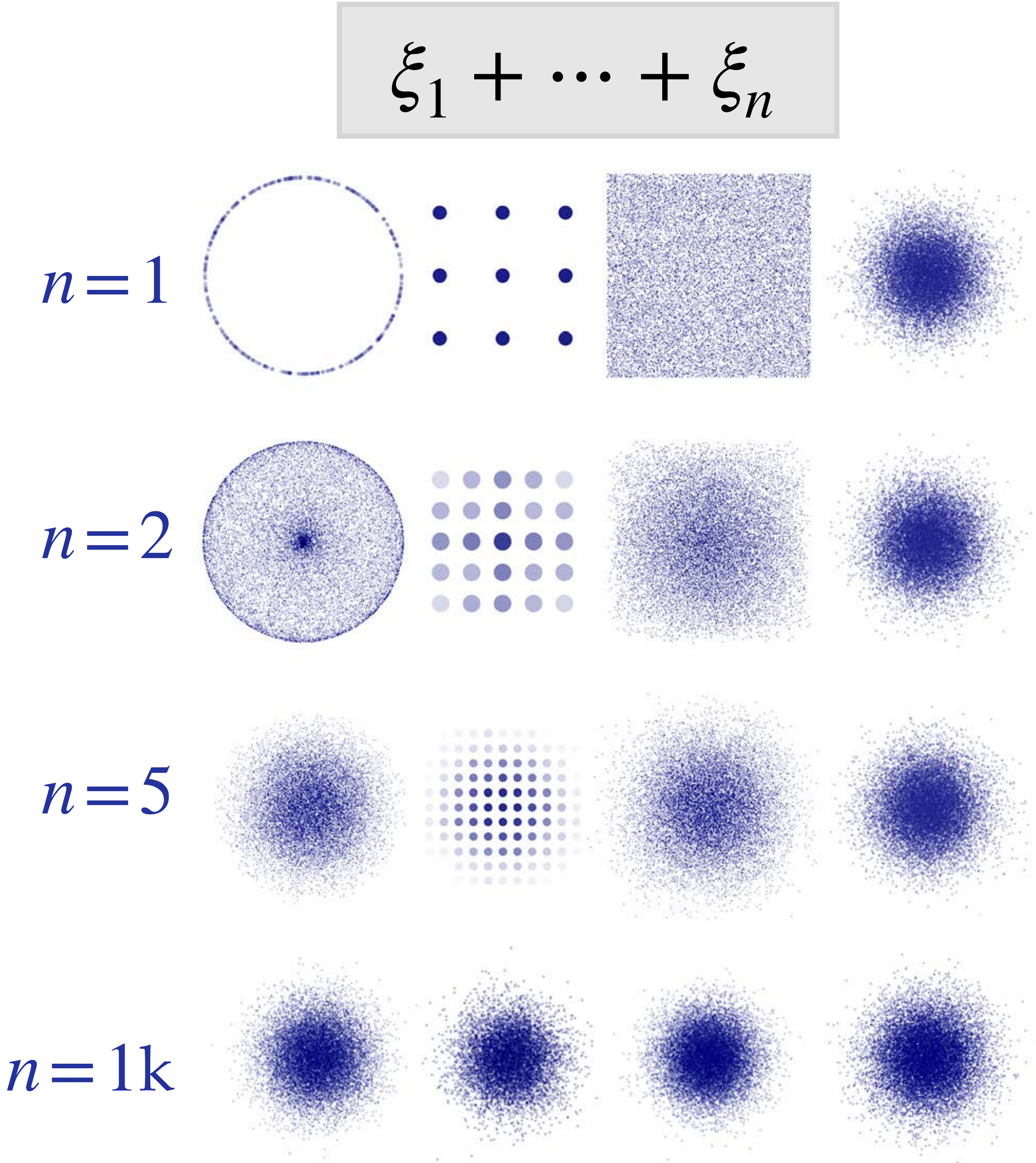
Let's watch what happens as we gradually zoom out:



Q: Why do these walks all look so similar “from a distance”—
even though they look very different “up close?”

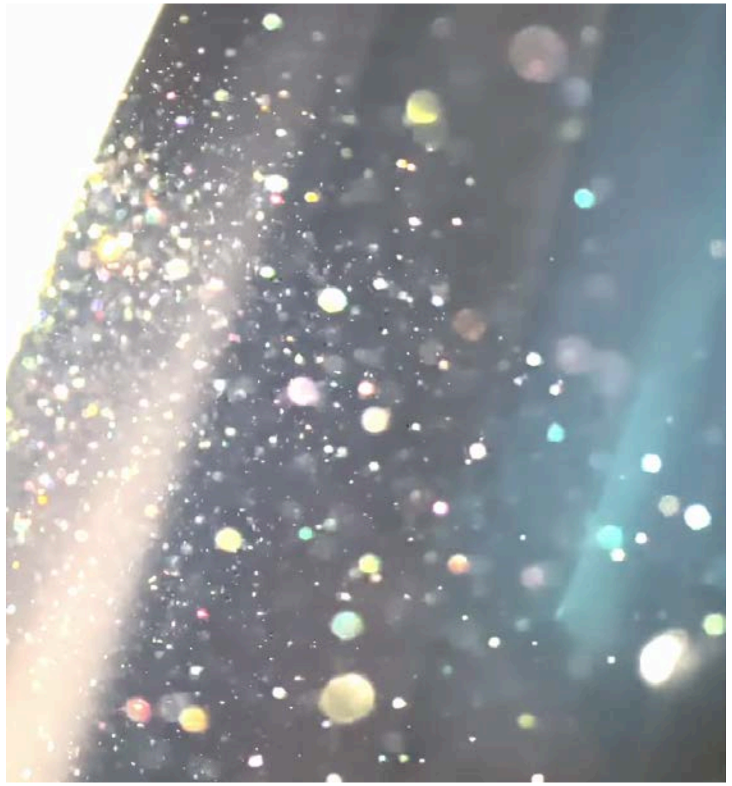
Brownian motion — Big Picture

- **A:** Because of the central limit theorem!
- The distribution describing the location of the nth step is the sum of n copies of single-step distribution p
- Central limit theorem tells us that this distribution approaches a normal distribution as $n \to \infty$, *no matter what p looks like*
  - when we zoom out, can't see individual steps—only the effect of n steps, for fairly large n



Universality of Brownian Motion

Takeaway: Even though random processes found in nature, science, technology, etc., all have **very** different origins, their aggregate behavior is in many* cases extremely well-predicted by one universal model.



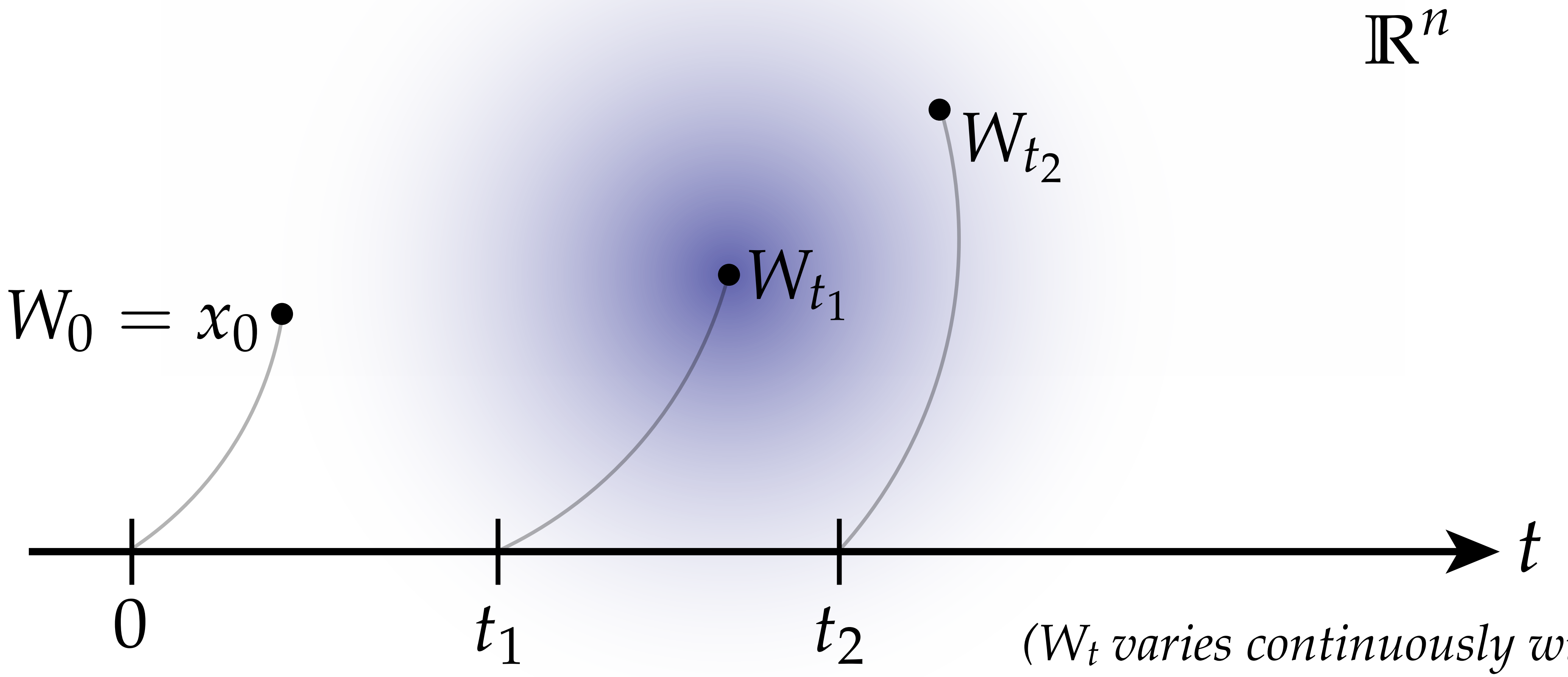
*Though other stochastic processes *do* arise in nature!

Brownian Motion / Wiener Process

Brownian motion or *Wiener process* assigns random variable W_t to each time t:

$$\boxed{W_{t_2} - W_{t_1} \sim \mathcal{N}(0, t_2 - t_1)}$$

independent Gaussian increments



(W_t varies continuously with respect to t)

Wiener Process — Definition

More formally, a **Wiener process** is a time-parameterized family of random variables W_t (i.e., one random variable for each $t \in \mathbb{R}_{>0}$) such that:

Brownian motion exhibits <u>Markov</u> property!

(continuity) W_t is continuous in t *almost surely* *i.e., with probability 1*

(independent increments) The "random increment" $W_{t_2} - W_{t_1}$ is independent of any past state W_{t_0} for all $0 \leq t_0 < t_1 \leq t_2$

(Gaussian increments) Each increment $W_{t_2} - W_{t_1}$ follows a normal distribution $\mathcal{N}(0, t_2 - t_1)$

Often, "Gaussian increments" condition given without any motivation

- e.g., why not consider other kinds of increments?
- hopefully you now understand why! ;-)

Donsker's Theorem

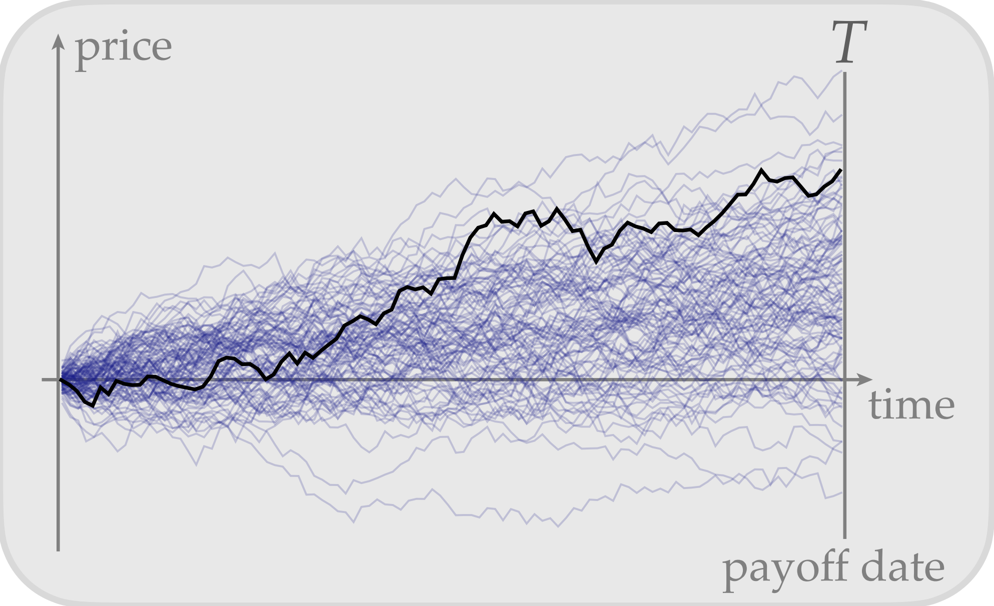
- Consider a sequence of i.i.d. random variables $X_1, \ldots, X_n$
- Can associate these discrete steps with a time-continuous function

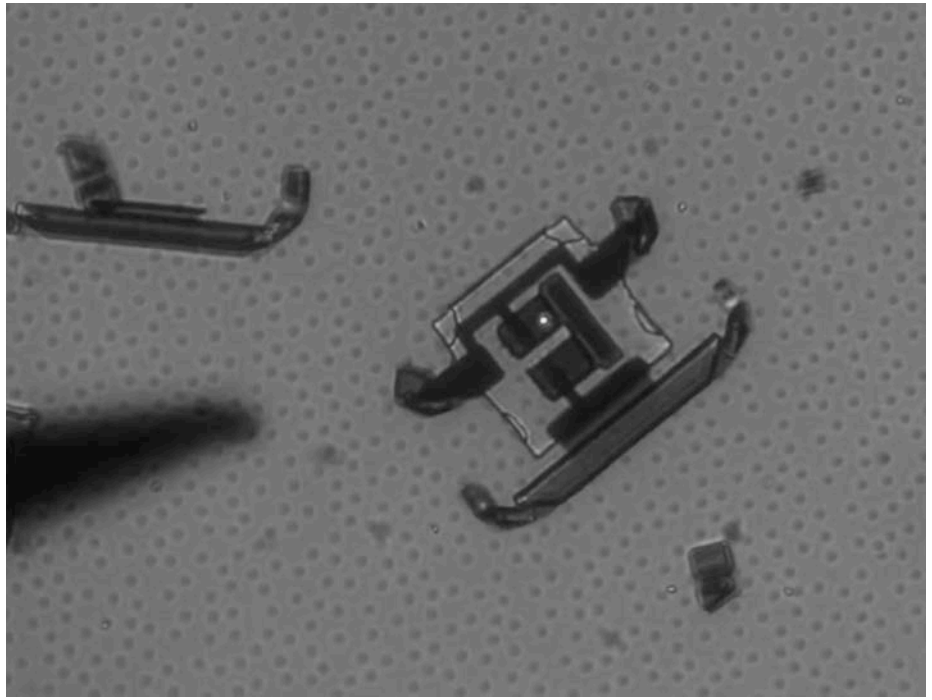
$$\widehat{W}_n(t) := \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor tn \rfloor} X_i, \quad t \in [0,1]$$

- **Donsker's theorem.** As $n \to \infty$, $\widehat{W}_n(t)$ converges* to a standard Brownian motion W_t over $t \in [0,1]$
  - *in an appropriate sense (*Skorokhod topology*)

Stopping Time

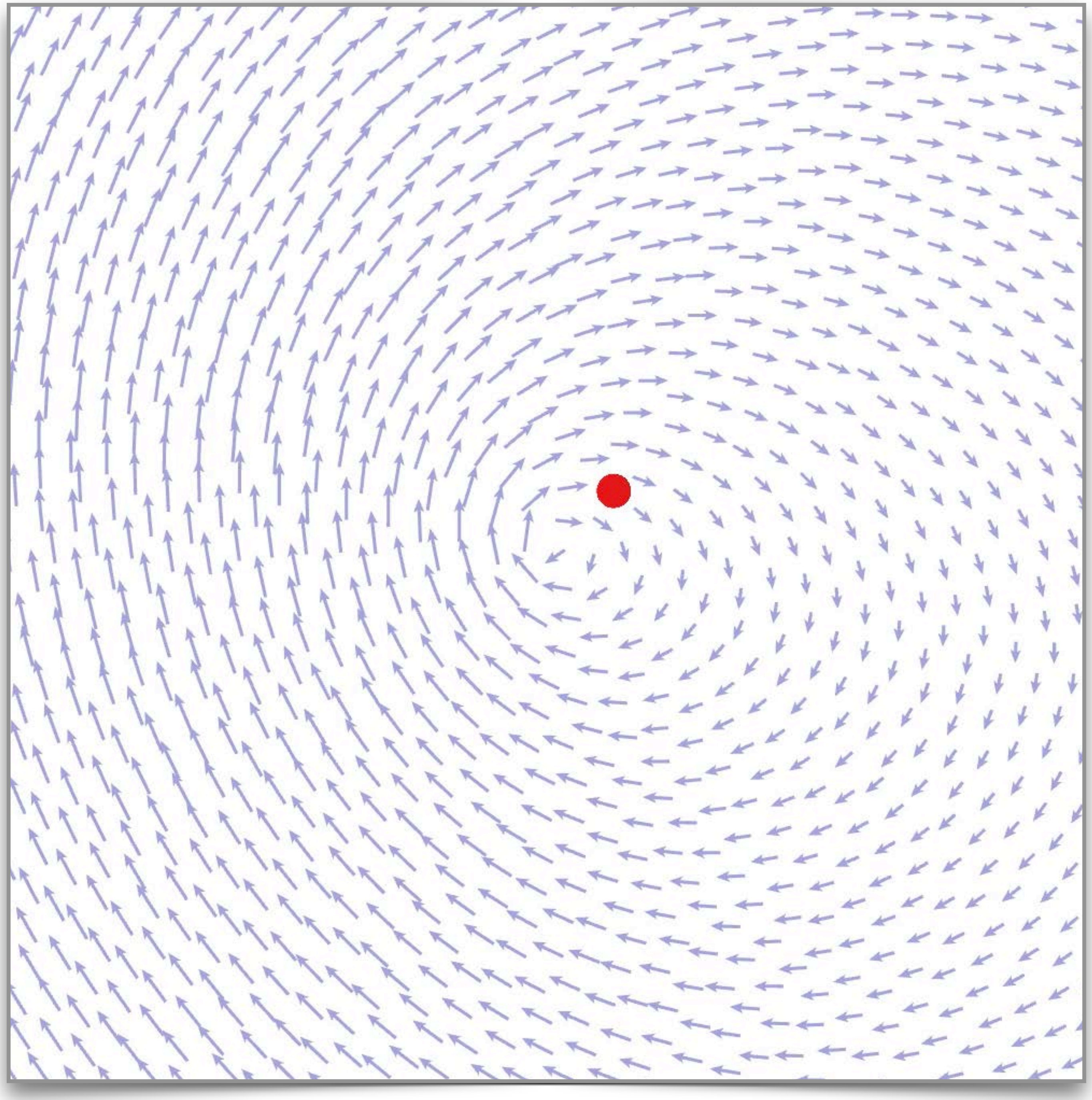
- Although many random processes could continue indefinitely, there is often a natural **stopping time**
  - For process X_t, often denote stopping time by capital T
  - e.g., *stock options*: we purchase the option to purchase an asset at an alternative price <u>at a fixed time T</u>
  - e.g., *control theory*: need to “steer” noisy process toward a goal <u>over a fixed time T</u>
- Stopping time can itself be a random variable
  - *e.g.*, gamble until you run out of money!





video: Marc Miskin

Deterministic Process



ordinary differential equation (ODE)

CHANGE IN POSITION

CHANGE IN TIME

$$dX_t = \vec{\omega}(X_t)dt$$

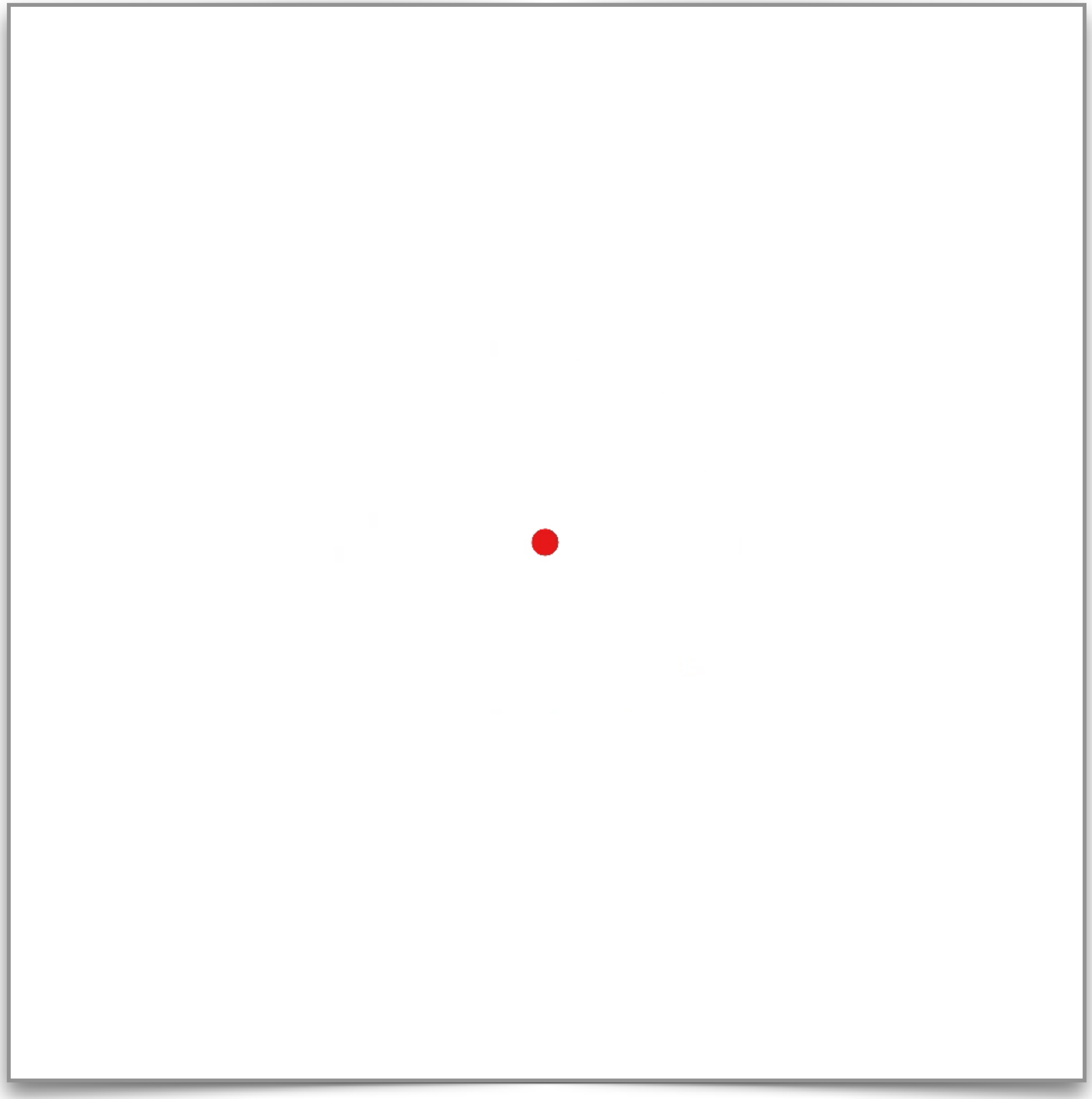
VELOCITY

Note: if we "divide by dt", get usual ODE $dx/dt = \omega(x(t))$

trajectory (X_t)

drift direction $(\vec{\omega})$

Brownian Process



stochastic differential equation (SDE)

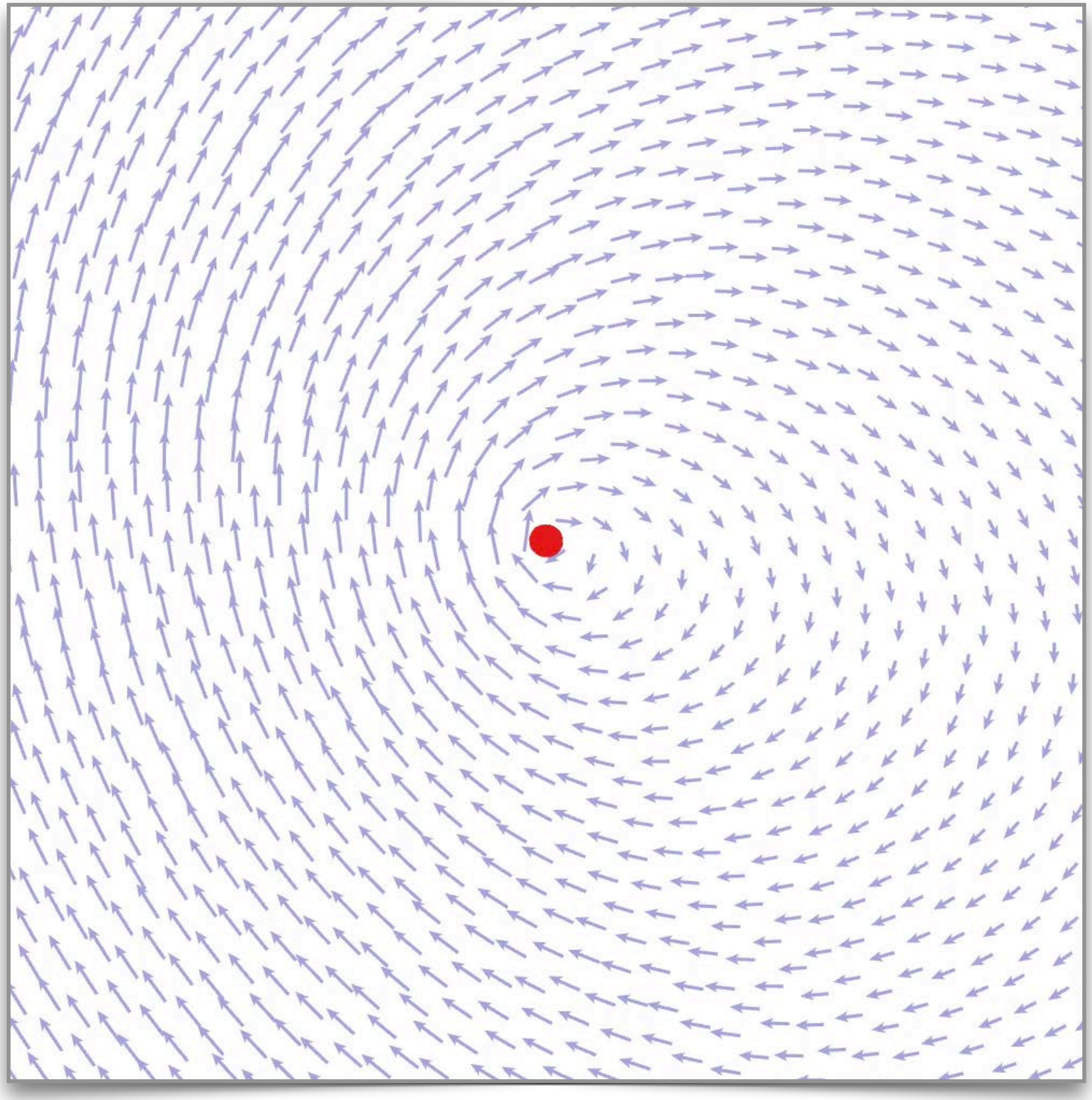
CHANGE IN POSITION

$$\boxed{dX_t} = \boxed{dW_t}$$

"NOISE"

● trajectory (X_t)

Brownian Process with Drift



deterministic motion + "noise"

–or–

random motion + "drift"

CHANGE IN POSITION: dX_t

VELOCITY: $\vec{\omega}(X_t)$

CHANGE IN TIME: dt

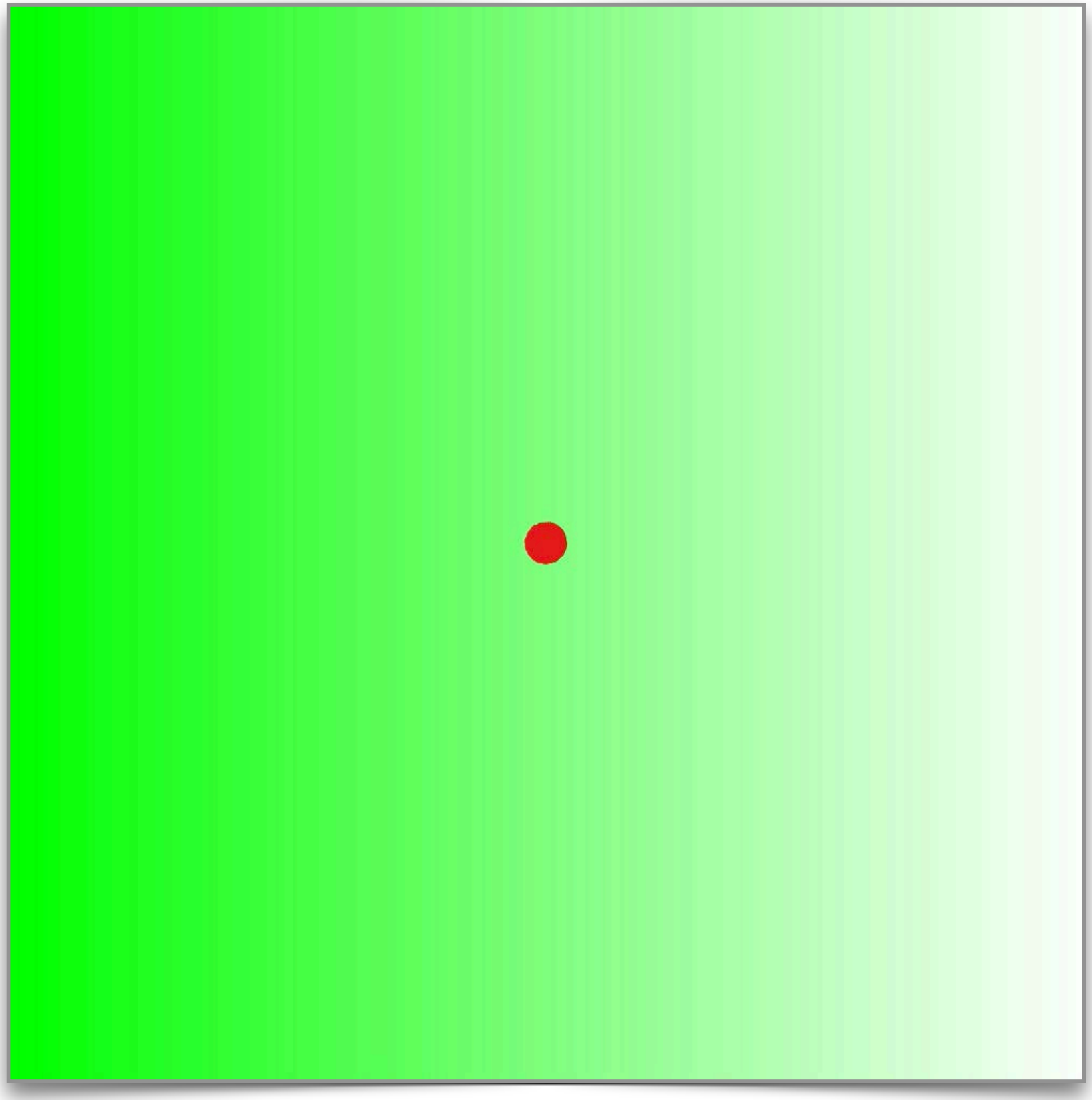
BROWNIAN MOTION: dW_t

$$dX_t = \vec{\omega}(X_t)dt + dW_t$$

● trajectory (X_t)

→ drift direction $(\vec{\omega})$

Brownian Process with Variable Diffusion



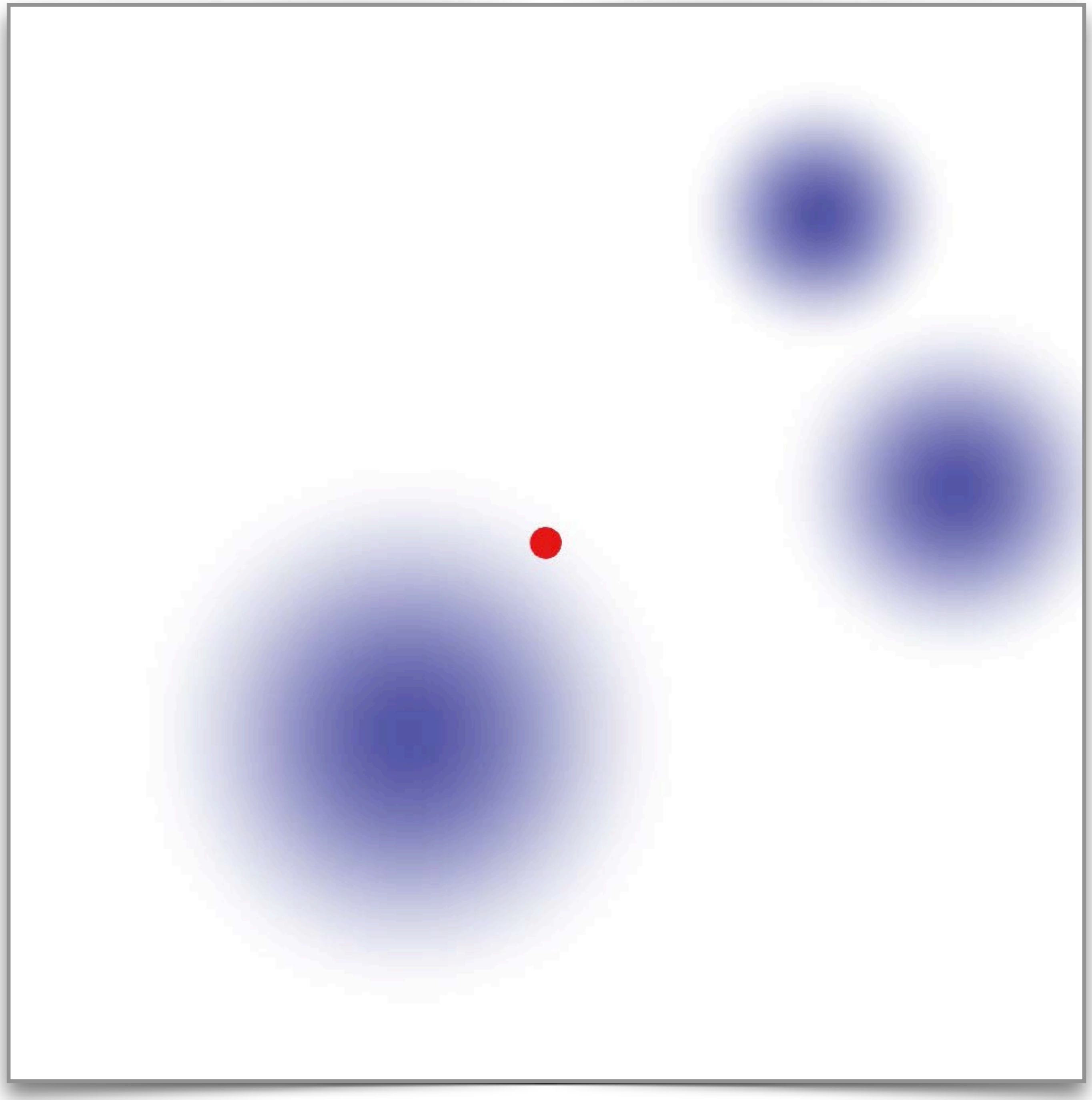
$$dX_t = \boxed{\alpha(X_t)}dW_t$$

RATE OF
DIFFUSION

trajectory (X_t)

diffusivity (α)

Brownian Process in Absorbing Medium



In general, may need to talk about random walk getting *"killed"* or *"absorbed"*—even though absorption does not appear in the SDE itself.

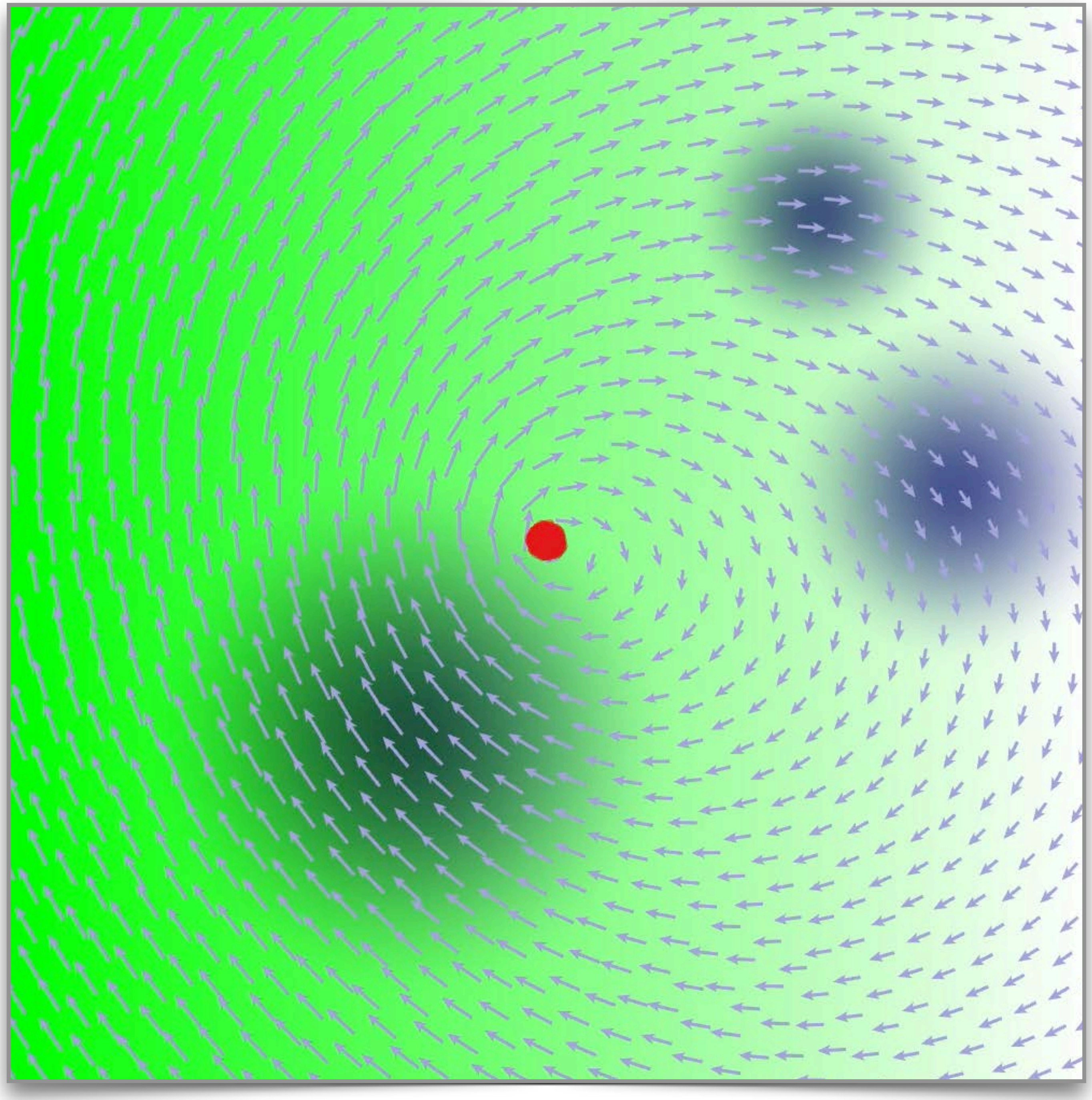
$$dX_t = dW_t$$

Roughly: integrating absorption over time determines (random) stopping time.

trajectory (X_t)

absorption (σ)

Diffusion Process



$$\underbrace{\overbrace{dX_t}^{\text{CHANGE IN POSITION}}}_{} = \underbrace{\vec{\omega}(X_t)}_{\text{VELOCITY}}\overbrace{dt}^{\text{CHANGE IN TIME}} + \underbrace{\alpha(X_t)}_{\text{RATE OF DIFFUSION}}\overbrace{dW_t}^{\text{"NOISE"}}$$

trajectory (X_t)

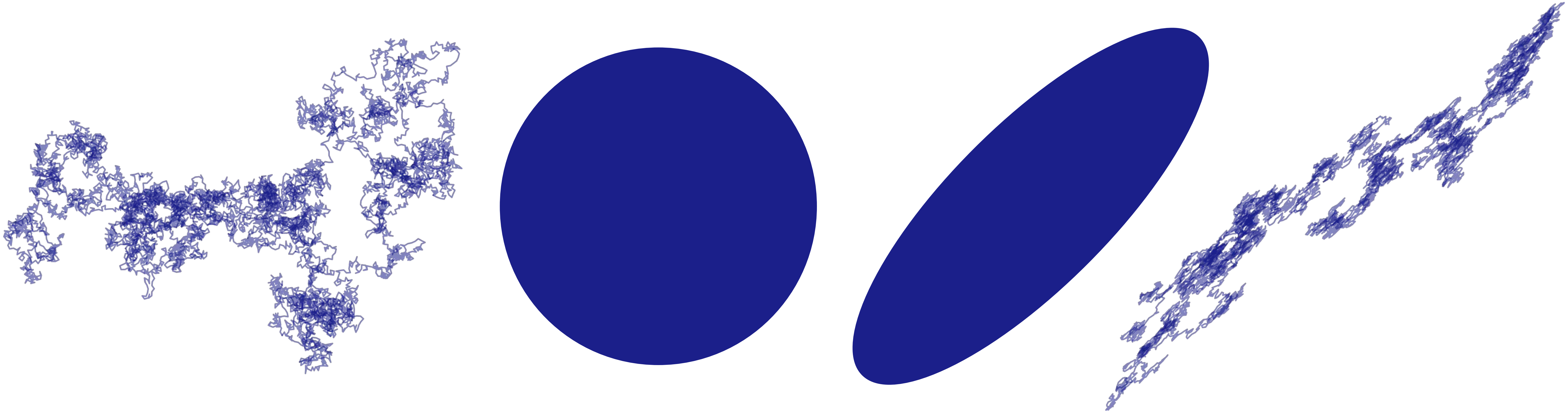
drift direction $(\vec{\omega})$

absorption (σ)

diffusivity (α)

Anisotropic Diffusion

Q: Do you think our random walk will look the same (as $n \to \infty$) if we sample our step direction from these two distributions?



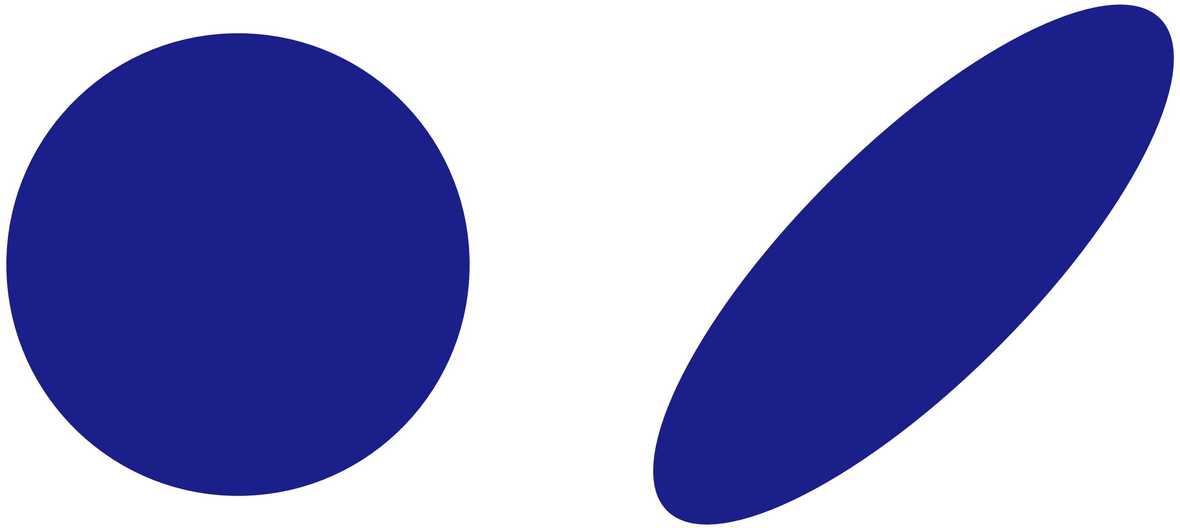
A: No! If our distribution is *anisotropic* (i.e., lacks rotational symmetry), our random walk will likewise be anisotropic.

Anisotropic Diffusion & Central Limit Theorem

- In multiple dimensions, the central limit theorem says that a sum of i.i.d. samples X_i from any distribution converges to a normal distribution with the same mean μ and covariance matrix Σ
  - in general, Σ can look very different from a constant multiple of the identity!

covariance matrix

$$\Sigma_{i,j} = \mathbb{E}\left[(X_i - \mu_i)(X_j - \mu_j)\right]$$

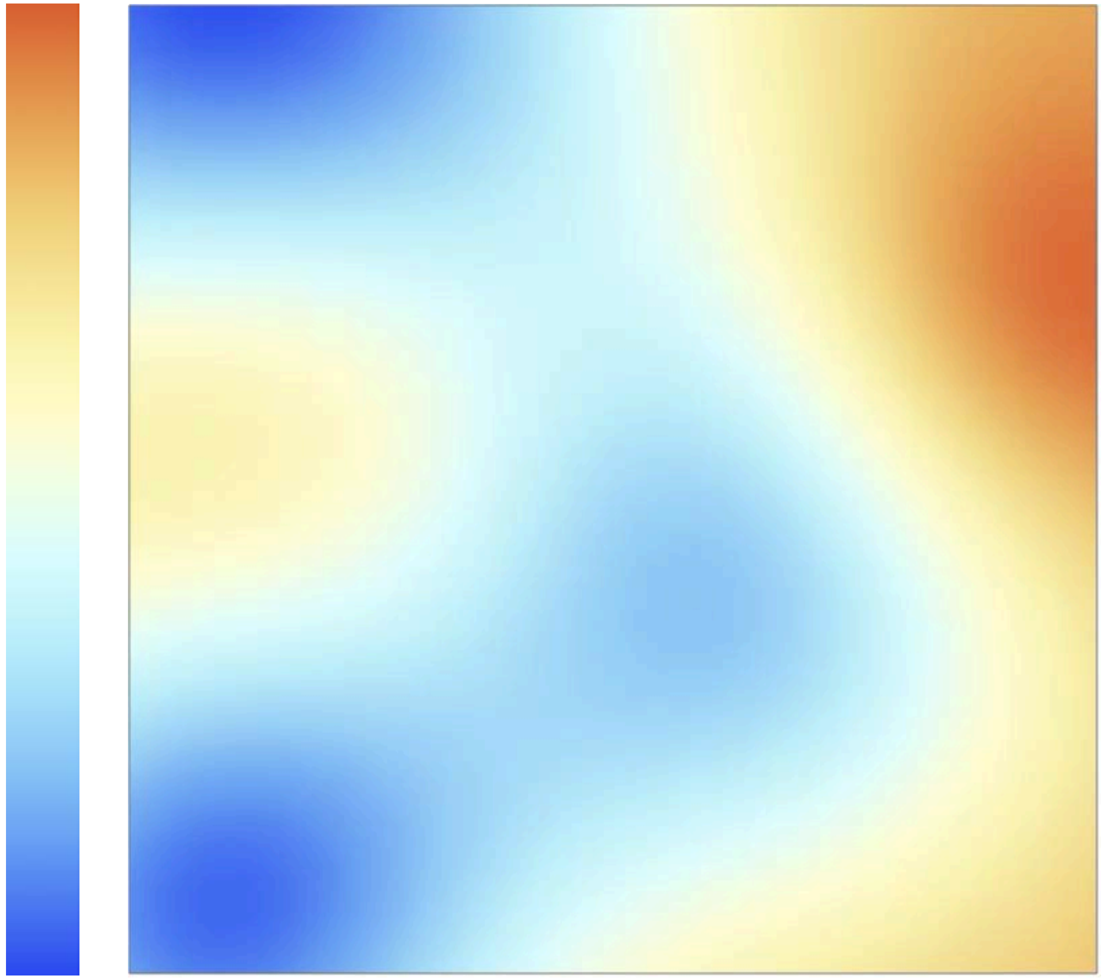


$$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1 \end{bmatrix}$$

Function of a Stochastic Processes

- **Recall:** a function of a random variable is a random variable
- Likewise, a function of stochastic process is a stochastic process

$f : \mathbb{R}^n \to \mathbb{R}$



$$dX_t = \vec{\omega}(X_t)dt + \sigma dW_t$$

Natural question: what is the SDE governing Y_t?

$$Y_t := f(X_t)$$

Itô's Formula

Itô's lemma provides "chain rule" for stochastic processes.

Rough intuition.

Deterministic derivatives ask: how much does output change if we vary the input along a given direction?

Stochastic derivatives ask: what distribution of change do we get by varying the input along a distribution of directions?

Kyoshi Itō

Itô's Formula—Ordinary Differential Equation

Itô's lemma provides "chain rule" for stochastic processes.

Example. For *deterministic* ODE, just the usual chain rule:

ordinary differential equation

$$dX_t = \vec{\omega}(X_t)dt$$

time-varying function

$$f : \mathbb{R}^n \times \mathbb{R}_{\geq 0} \to \mathbb{R}$$

derived differential equation

$$Y_t := f(X_t)$$

change in derived value over time

temporal change in f itself

spatial change in f due to motion along trajectory

$$\implies \quad dY_t = \left(\frac{\partial f}{\partial t}\bigg|_{X_t} + \vec{\omega}|_{X_t} \cdot \nabla f|_{X_t} \right) dt$$

Itô's Formula — Brownian Motion

Itô's lemma provides "chain rule" for stochastic processes.

Example. Most essential question: what about Brownian motion?

pure Brownian motion

$$dX_t = dW_t$$

time-independent function

$$f : \mathbb{R}^n \to \mathbb{R}$$

derived stochastic process

$$Y_t := f(X_t)$$

change in derived value over time — *Laplacian of function f* — *spatial change in f along random trajectory*

$$\Longrightarrow \quad dY_t = \tfrac{1}{2}\Delta f|_{X_t}\, dt + \nabla f|_{X_t} \cdot dW_t$$

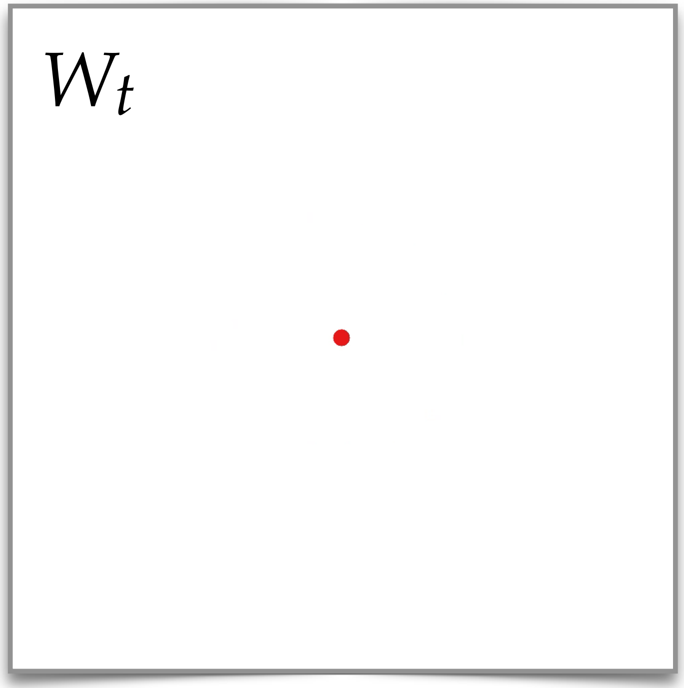
$$\Delta f = \sum_{k=1}^{n} \frac{\partial^2 f}{\partial x_i^2}$$

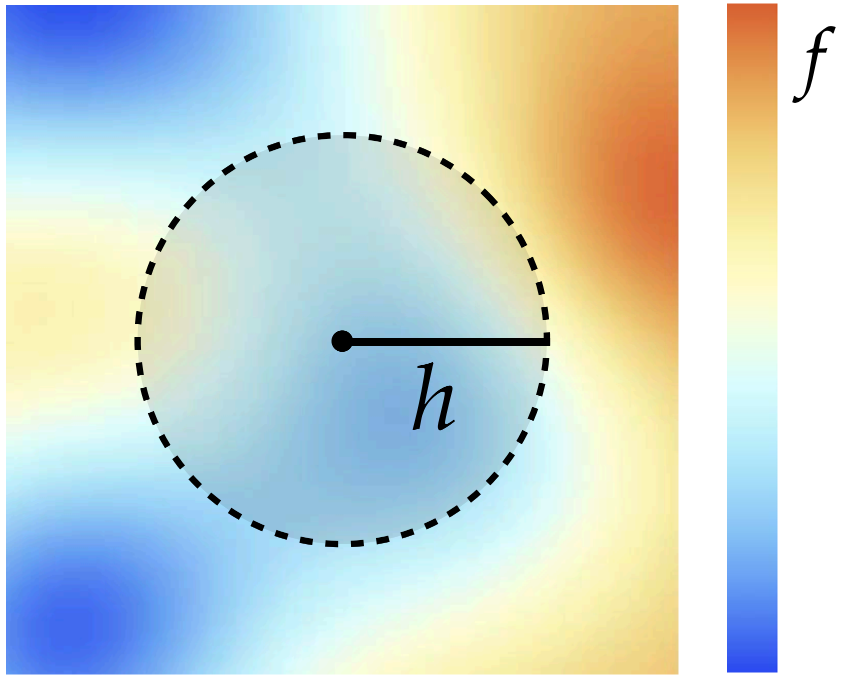
Really strange: we only took one derivative (d). How did we end up with 2nd derivatives?

Itô's Lemma & Laplacian

Intuition.

- Over small time t, Brownian motion W_t explores a **small neighborhood** of X_0.
- At any point x, the Laplacian $\Delta f(x)$ gives difference between the value at x and value in **small neighborhood**.
- Hence, 1st-order change in observed value over time involves 2nd-order derivative in space.
- (Formal treatment: Øksendal §4.2)





$$\bar{f}_h(x) := \frac{1}{|B_h(x)|} \int_{B_h(x)} f(y)\, dy$$

$$\Delta f(x) \propto \frac{\bar{f}_h(x) - f(x)}{h^2} + O(h)^3$$

Itô's Formula — Diffusion Process

Itô's lemma provides "chain rule" for stochastic processes.

Example. Overall we get a formula for general diffusion processes:

diffusion process	time-varying function	"derived" process
$dX_t = \vec{\omega}dt + \alpha dW_t$	$f : \mathbb{R}^n \times \mathbb{R}_{\geq 0} \to \mathbb{R}$	$Y_t := f(X_t)$

change in derived value over time

temporal change in f itself

spatial change in f due to "exploration" of local neighborhood

spatial change in f due to motion along random path

$$\underbrace{dY_t}_{\text{change in derived value over time}} = \left(\underbrace{\frac{\partial f}{\partial t}}_{\text{temporal change in } f \text{ itself}} + \underbrace{\vec{\omega} \cdot \nabla f}_{\text{spatial change in } f \text{ due to deterministic motion}} + \underbrace{\tfrac{1}{2}\alpha^2 \Delta f}_{\text{spatial change in } f \text{ due to "exploration" of local neighborhood}}\right) dt + \underbrace{(\alpha \nabla f) \cdot dW_t}_{\text{spatial change in } f \text{ due to motion along random path}}$$

spatial change in f due to deterministic motion

Note: since all directions are equally likely & f is locally linear, $\mathbb{E}[\nabla f \cdot dW_t] = 0$

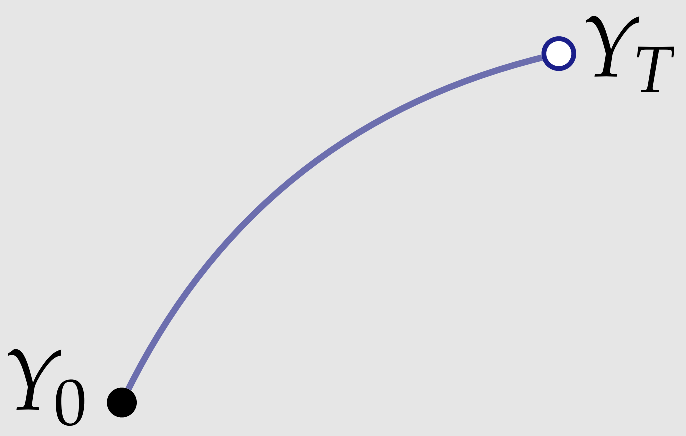
Even more general form: Øksendal [2013, §4.2]

Ito Integration

Deterministic Integral

"start at an initial point and add total change due to a deterministic function / vector field"

$$Y = \int_0^T \vec{\omega}\, dt$$

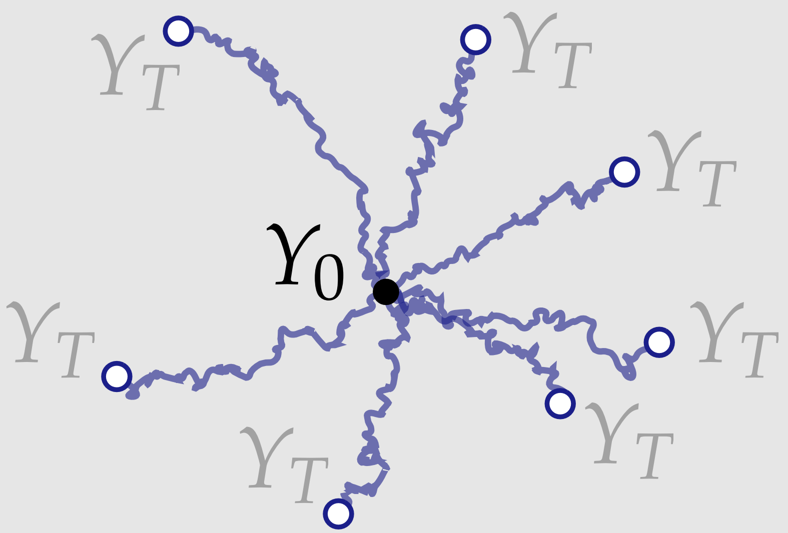


result is a point in space

Stochastic (Itô) Integral

"start at an initial point and add total change due to a stochastic function / random walk"

$$Y = \int_0^T dW_t$$



result is a random variable

Ito Integration (continued)

Perhaps easiest to understand in terms of numerical integration:

$$dX_t = \vec{\omega}(X_t)dt + \alpha(X_t)dW_t$$

Euler-Maruyama

$$\underbrace{x_{k+1}}_{\text{next state}} = \underbrace{x_k}_{\text{last state}} + \underbrace{\varepsilon}_{\text{time step}}(\overbrace{\vec{\omega}(x_k)}^{\text{velocity at last state}} + \overbrace{\alpha(x_k)}^{\text{diffusivity}}\overbrace{W_k}^{\text{noise}})$$

$$W_k \sim \mathcal{N}(0,1)$$

Get better & better approximation of <u>one trajectory</u> by taking more steps n:

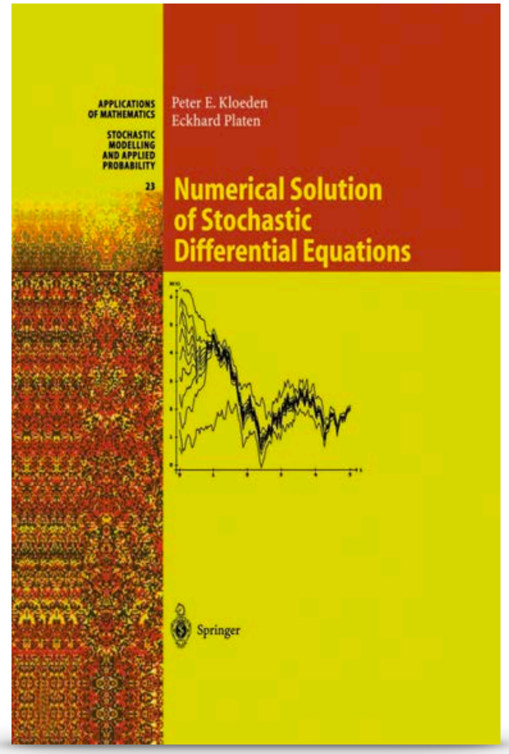
$$x_T^n := x_0 + \sum_{k=0}^{n-1} \varepsilon\vec{\omega}(x_k) + \varepsilon\alpha(x_k)W_k, \quad \varepsilon := \frac{T}{n}$$

As $n \to \infty$, the <u>distribution</u> of points x_T^n essentially describes result of Ito integral.

More formal treatment: Øksendal [2013, Ch. 3]

Numerical Integration of SDEs

- Numerically integrating SDEs not much different from ODEs
  - roughly speaking: take a step and "add noise"
  - amount of noise should be proportional to time step ε



[Kloeden & Platen]

diffusion process

$$dX_t = \vec{\omega}(X_t)dt + \alpha(X_t)dW_t$$

<u>**Euler-Maruyama** (forward)</u>

$$x_{k+1} = x_k + \varepsilon(\vec{\omega}(x_k) + \alpha(x_k)W_k)$$

explicit

<u>**Euler-Maruyama** (backward)</u>

$$x_{k+1} - \varepsilon\vec{\omega}(x_{k+1}) = x_k + \varepsilon\alpha(x_k)W_k$$

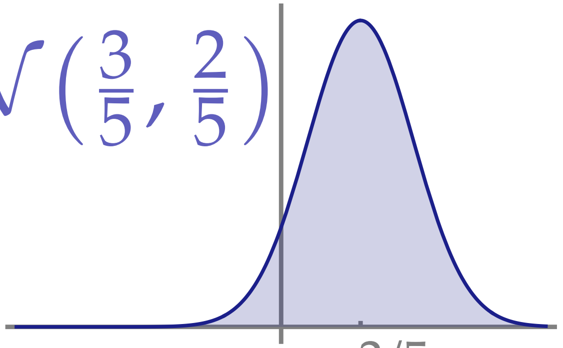
implicit

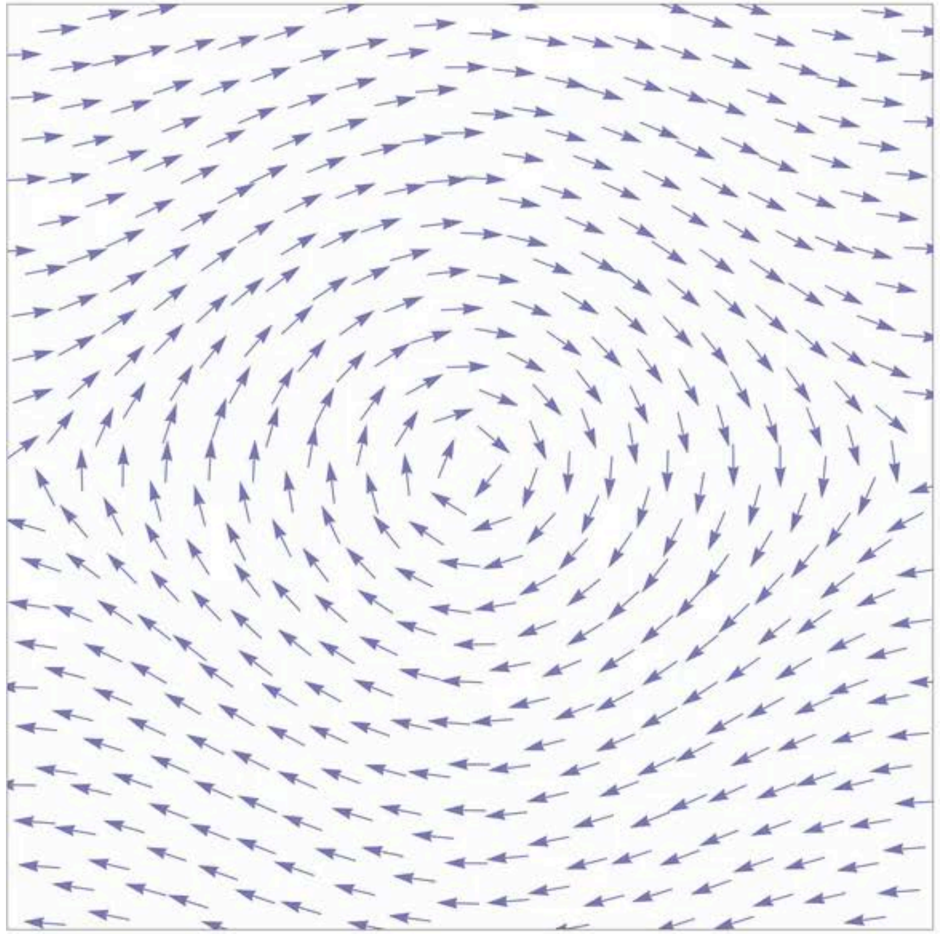
$$W_k \sim \mathcal{N}(0,1) \quad (i.i.d.)$$

Pendulum in the Wind — Forward Euler-Maruyama

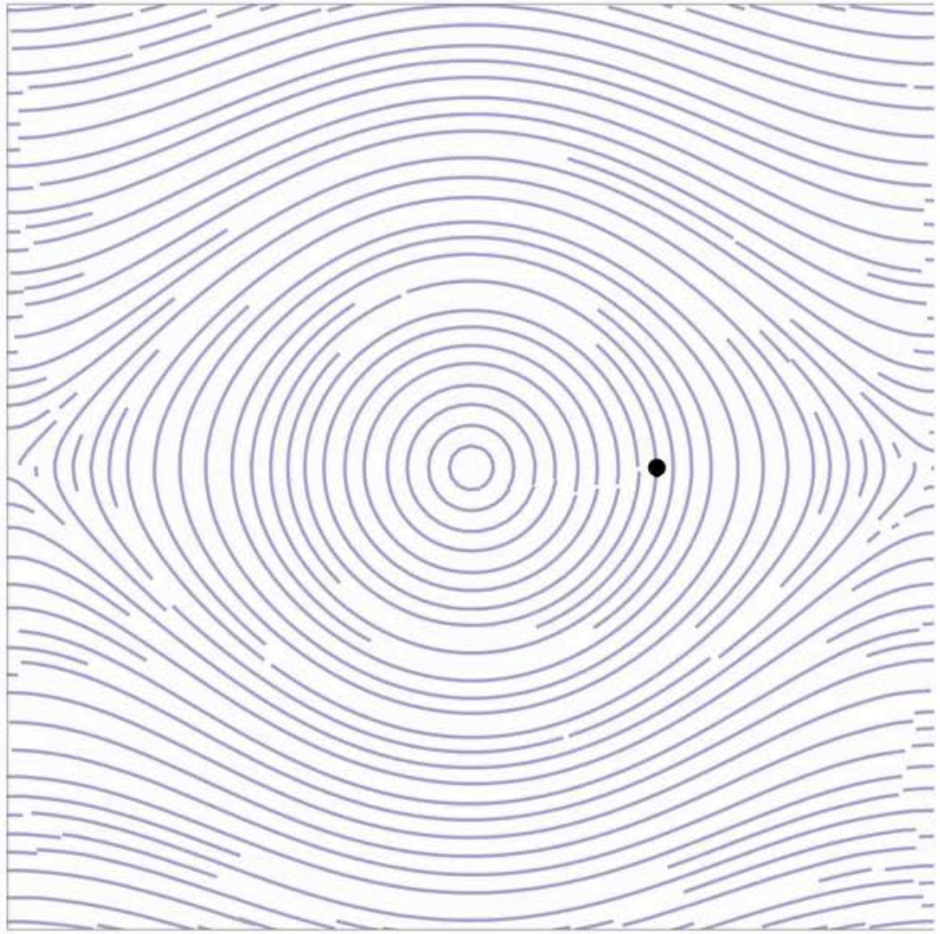
$$(d\theta_t, d\theta'_t) = (\theta'_t, -\sin(\theta_t))dt + \begin{bmatrix} 0 & 0 \\ 0 & \sigma_{\text{wind}} \end{bmatrix} dW_t$$

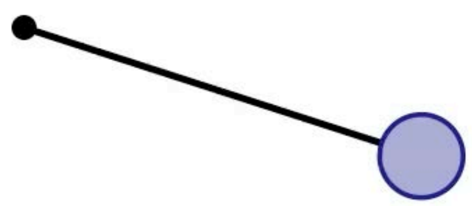
$\sigma_{\text{wind}} \sim \mathcal{N}\left(\frac{3}{5}, \frac{2}{5}\right)$





drift $\overrightarrow{\omega}(\theta, \theta')$

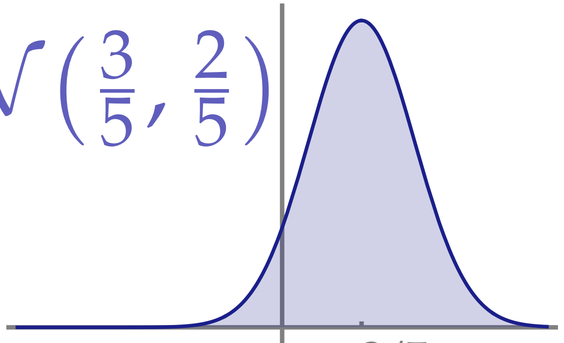




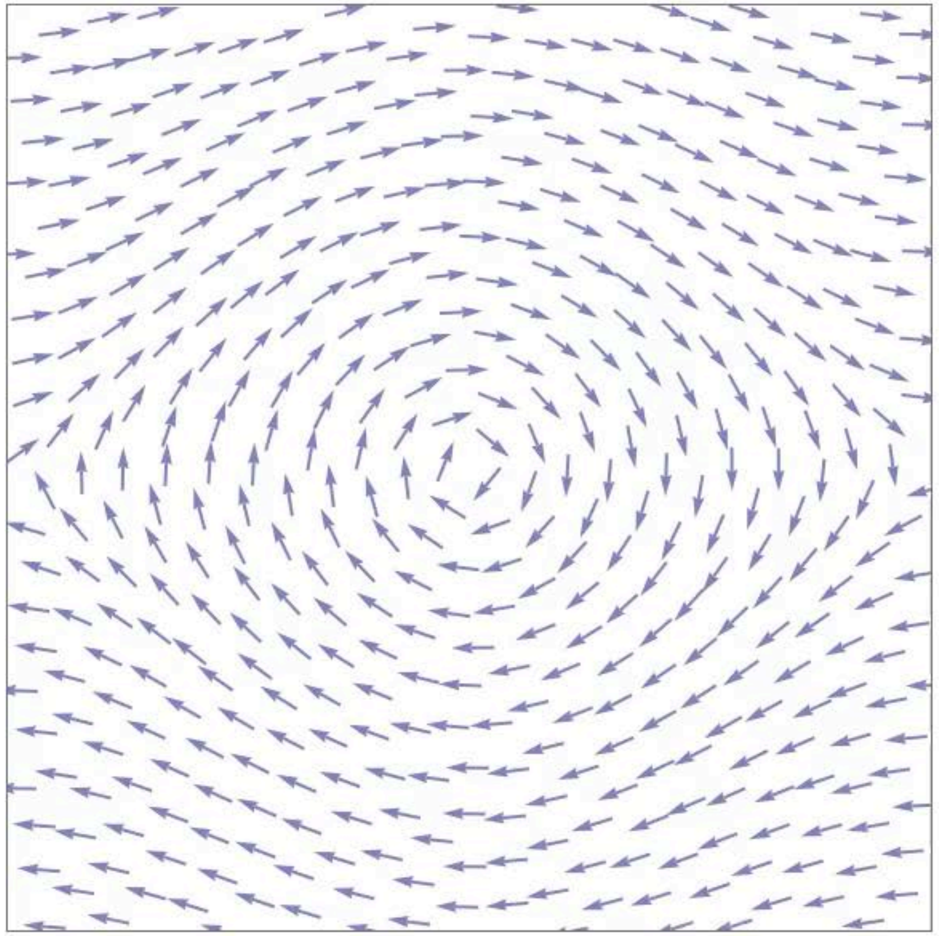
Pendulum in the Wind — Backward Euler-Maruyama

$$(d\theta_t, d\theta'_t) = (\theta'_t, -\sin(\theta_t))dt + \begin{bmatrix} 0 & 0 \\ 0 & \sigma_{\text{wind}} \end{bmatrix} dW_t$$

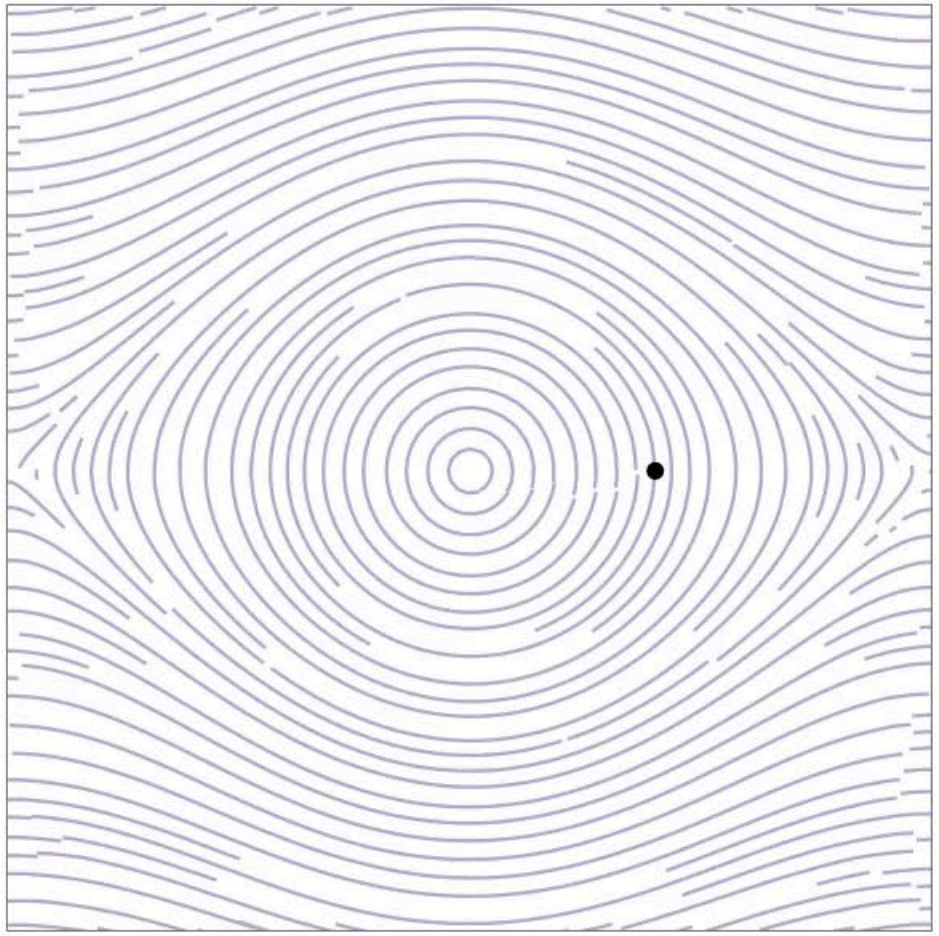
$\sigma_{\text{wind}} \sim \mathcal{N}\left(\frac{3}{5}, \frac{2}{5}\right)$



wind velocity

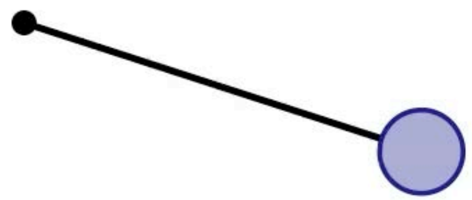


drift $\overrightarrow{\omega}(\theta, \theta')$



-π ⟵ angle θ ⟶ +π

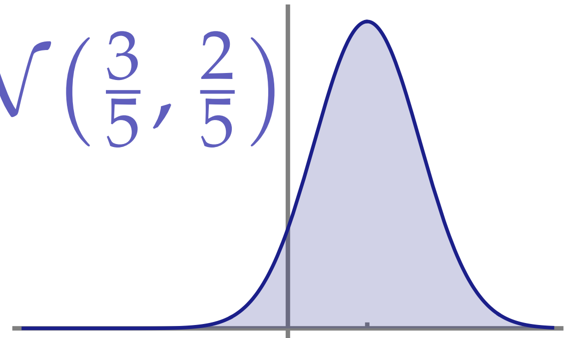
↑ angular velocity θ' ↓



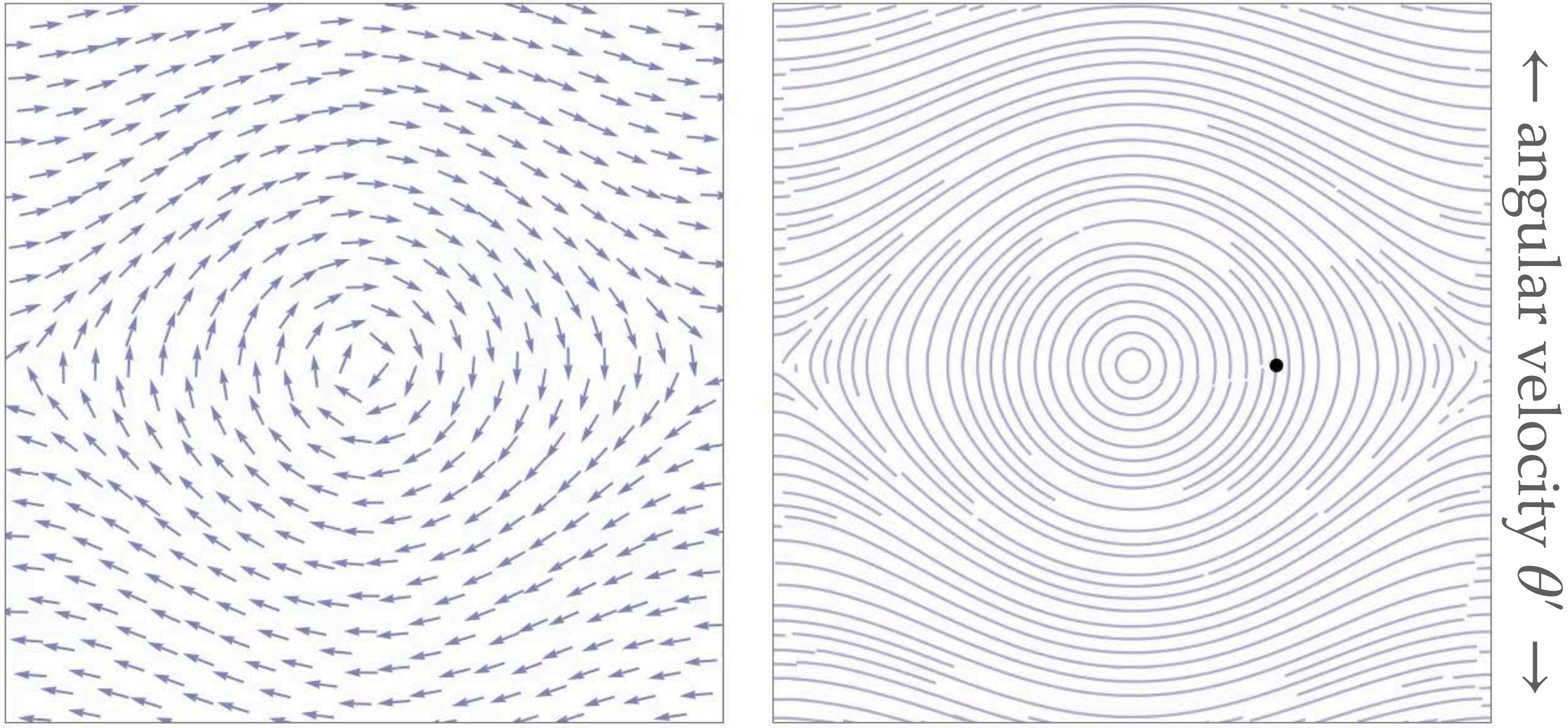
Pendulum in the Wind—Symplectic Euler-Maruyama

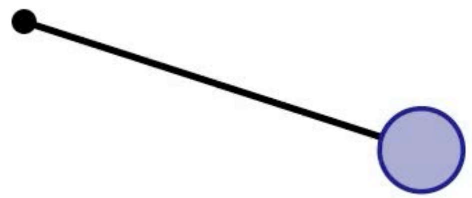
$$\begin{aligned} \theta'_{k+1} &= \theta'_k - \varepsilon(\sin(\theta_k) - \sigma_{\text{wind}}) \\ \theta_{k+1} &= \theta_k + \varepsilon\theta'_{k+1} \end{aligned}$$

$\sigma_{\text{wind}} \sim \mathcal{N}\left(\frac{3}{5}, \frac{2}{5}\right)$



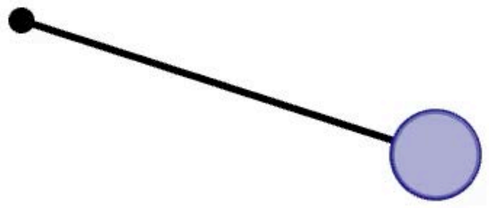
wind velocity





Pendulum as Random Variable

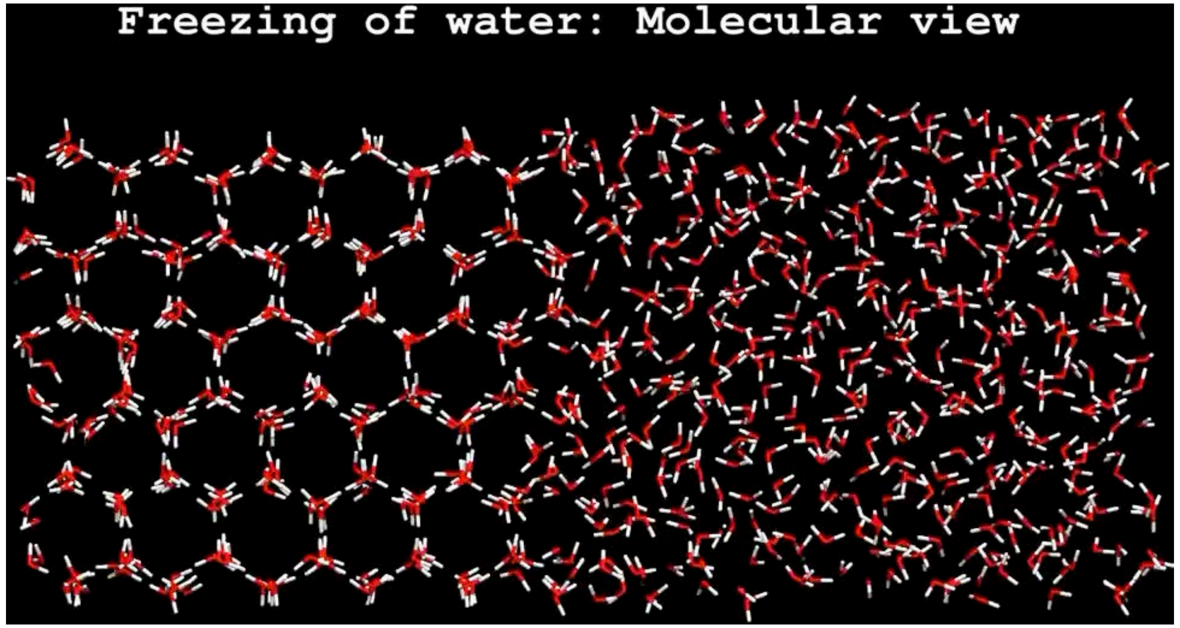
But wait a minute—didn't we say the result of Itô integration is a <u>random variable</u>?
(Not just one "noisy" trajectory.)

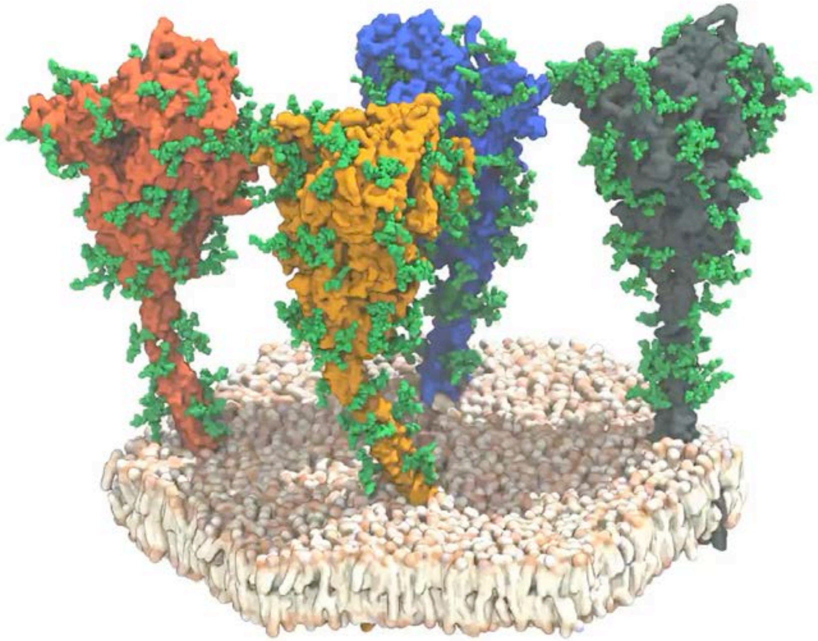


Can think of our SDE integrator as a tool for approximating a *distribution*, rather than finding just <u>one</u> trajectory.

Application: Molecular Dynamics

- In fact, this is often the goal in **molecular dynamics**
  - use SDE integrator to simulate trajectory of molecules in "noisy" environment
  - perform many trials to understand typical/ average behavior of large *ensemble* of molecules
  - use information to predict behavior of diseases, response to drugs, build new materials, …
  - alternative perspective: simulation is strategy for sampling states of system according to their probability
  - **later:** *Langevin dynamics* ↔ *Langevin Monte Carlo*





video credits: Bohemian chemists, Max Planck society

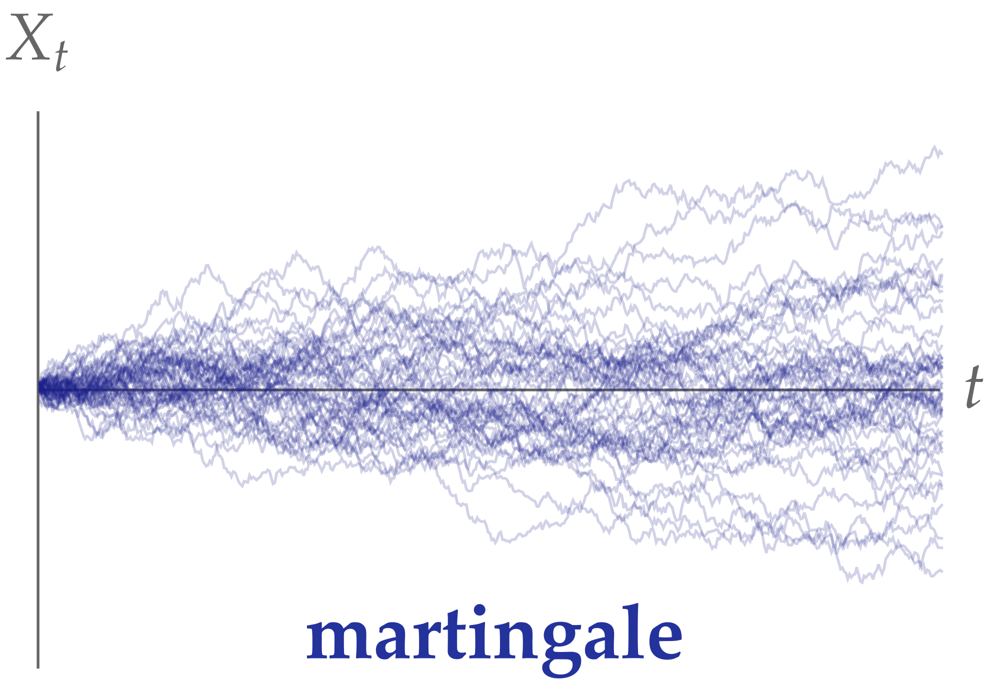
Beyond Brownian Motion — Martingales

- In general, **martingale** is stochastic process where:
  - average value doesn't change
  - average value is independent of history
- Discrete sequence of random variables $X_1, \ldots, X_k$ is a martingale if $\mathbb{E}[X_{k+1} \mid X_1, \ldots, X_k] = X_k$
  - (X_i need not be independent)
- Brownian motion is a model example in the continuous case

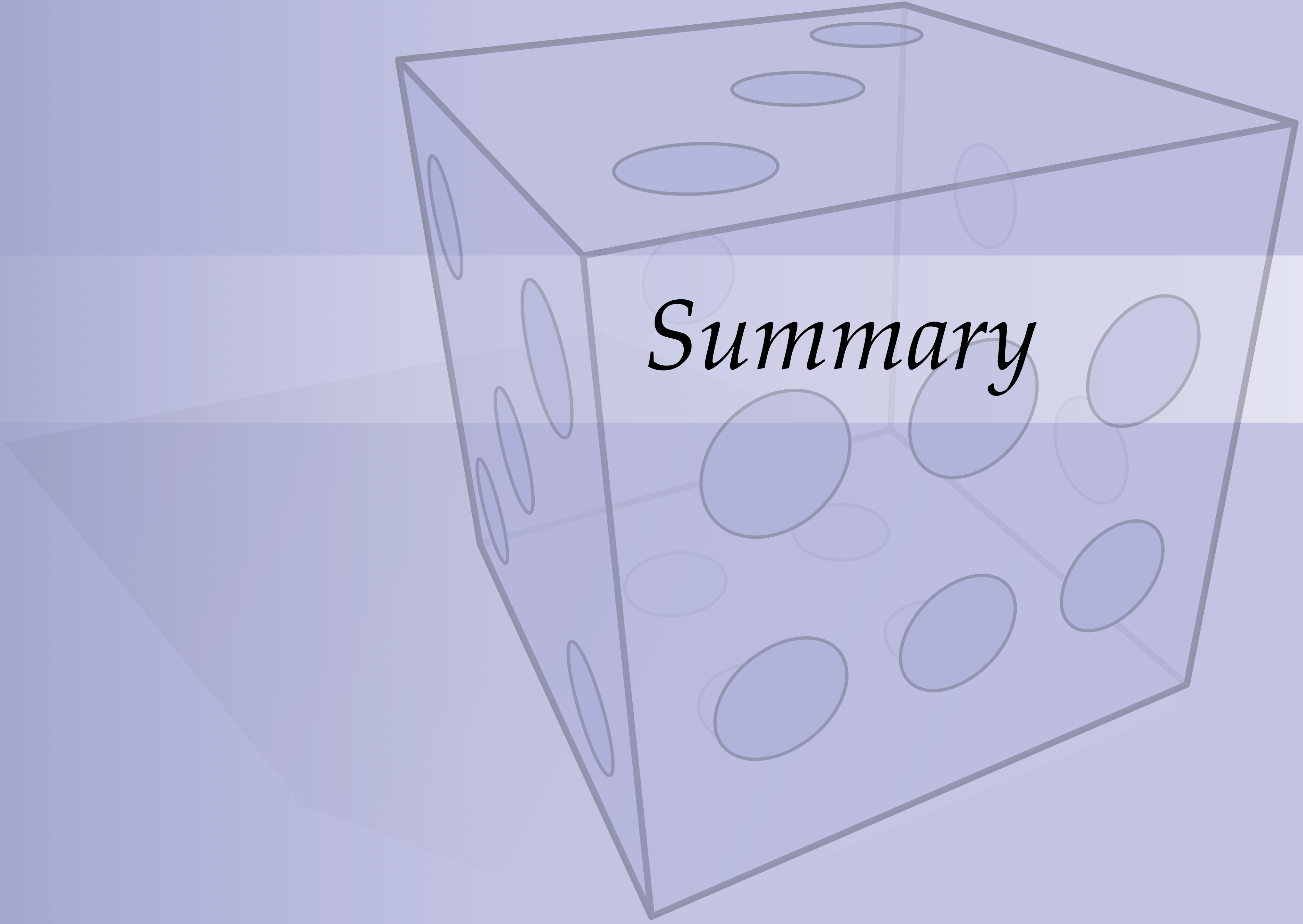
Basic **regularity** condition for stochastic processes

Makes it possible to generalize Brownian motion (and still say useful things...)

nightingale



martingale

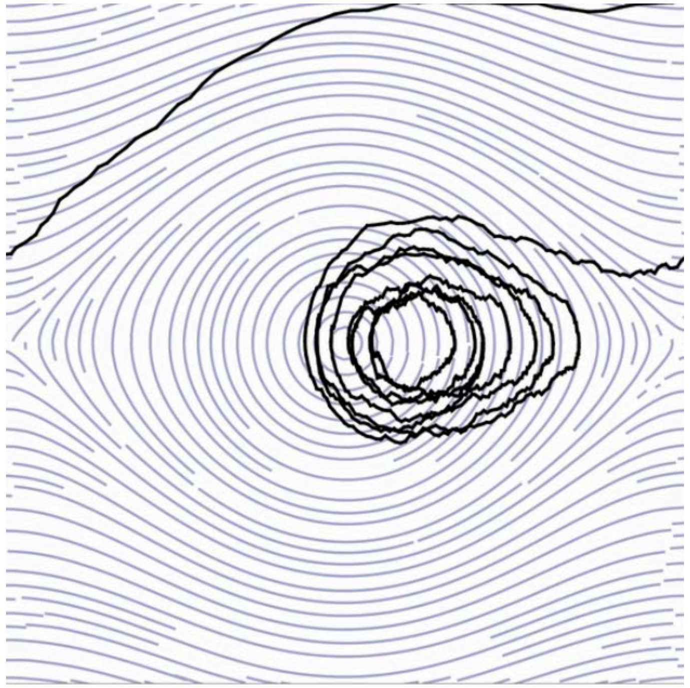


Summary

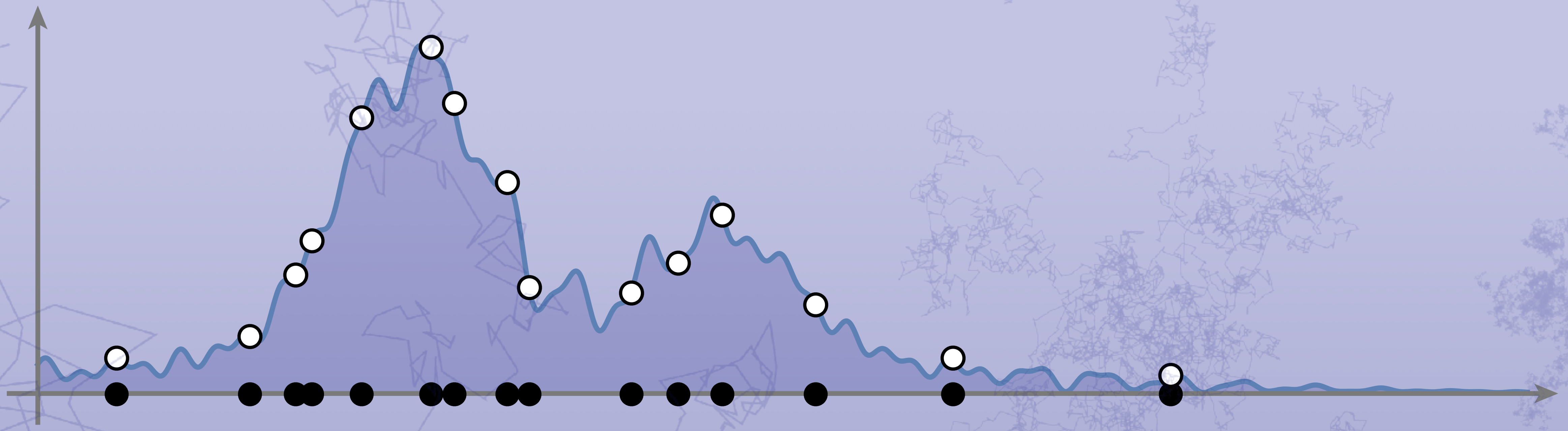
Overview—Stochastic Differential Equations

- **ODEs** implicitly describe systems evolving *over time*
- **SDEs** add *randomness* to this picture
- use **numerical integration** to recover explicit function from implicit description
  - **forward Euler** — simple/cheap but unstable
  - **backward Euler** — trickier/more expensive but stable
  - **Euler-Maruyama** — "just add noise" to simulate SDEs
- **Ito calculus** lets us analyze SDEs
  - **Ito's lemma** — basic analogue of differentiation
  - **Ito integration** — basic analogue of integration
  - unlike ordinary calculus, get *distributions* (not definite values)

analogy: *trajectory of rock (+wind)*



Thanks!



MONTE CARLO METHODS
AND APPLICATIONS