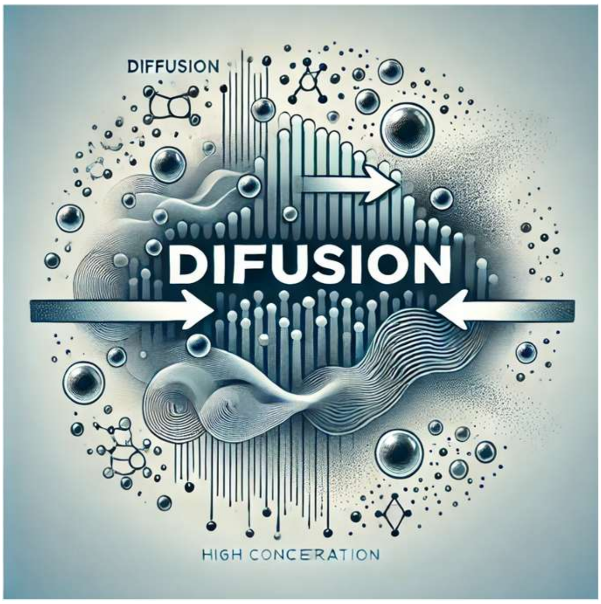
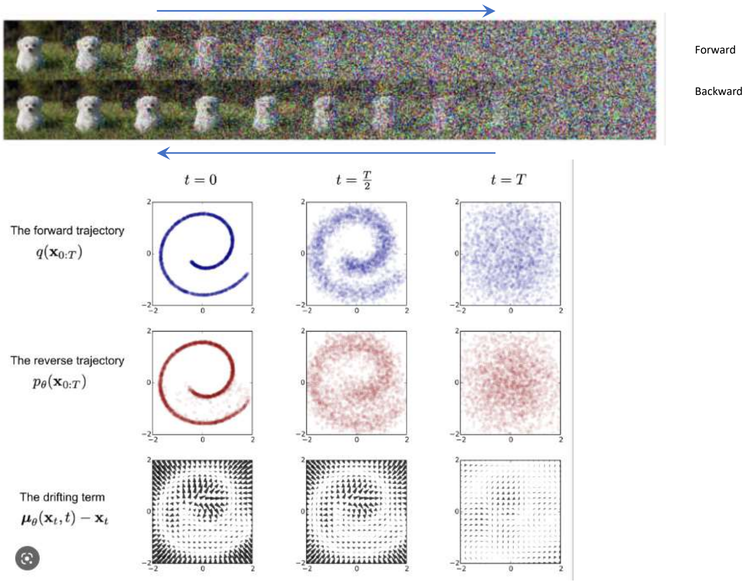


Mathematics of Diffusion

CVPR 2025

1



https://lilianweng.github.io/posts/2021-07-11-diffusion-models/

2

Ornstein Uhlenbeck process

the transition from $\mathbf{x}_{t-1}$ to $\mathbf{x}_t$ may be written as

$$\boxed{\mathbf{x}_t \leftarrow \alpha_t \mathbf{x}_{t-1} \oplus \sqrt{\beta_t}\epsilon}$$

$$X_t = \alpha_t X_{t-1} \oplus \sqrt{\beta_t} G$$

Conditional probabilities are

$$p(\mathbf{x}_t \,|\, \mathbf{x}_{t-1}) = G_{\beta_t}(\mathbf{x}_t - \alpha_t \mathbf{x}_{t-1})$$

One can also compute a sample from $\mathbf{x}_t$ directly according to the formula

$$\boxed{X_t = \bar{\alpha}_t X_0 + \sqrt{\bar{\beta}_t} G\ .}$$

$$\boxed{\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_t\ .}$$

$$\boxed{\bar{\beta}_t = 1 - \bar{\alpha}_t^2\ .}$$

(assuming $\beta_t + \alpha_t^2 = 1$).

3

Diffusion - backward and forward processes

- Diffusion consists of a forward and a backward stochastic process, starting with a sample x_0.
- Forward process (noising)

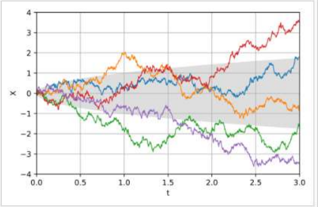
$$\boxed{\mathbf{x}_t \quad = \alpha_t \mathbf{x}_{t-1} + \beta_t \epsilon}$$

where $\alpha_t < 1$ and β_t is small.

- Backward process (denoising)

$$\boxed{\mathbf{x}_{t-1} = \mu_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) + \beta_t \epsilon}$$

- Can be made continuous as a so-called Wiener process.
- Reverse process worked out by Brian Anderson (ANU) in 1982 paper. Solution involves solving a Stochastic Differential Equation.



Five sampled processes, with expected standard deviation in gray.

From Wikipedia – Wiener process

4

5

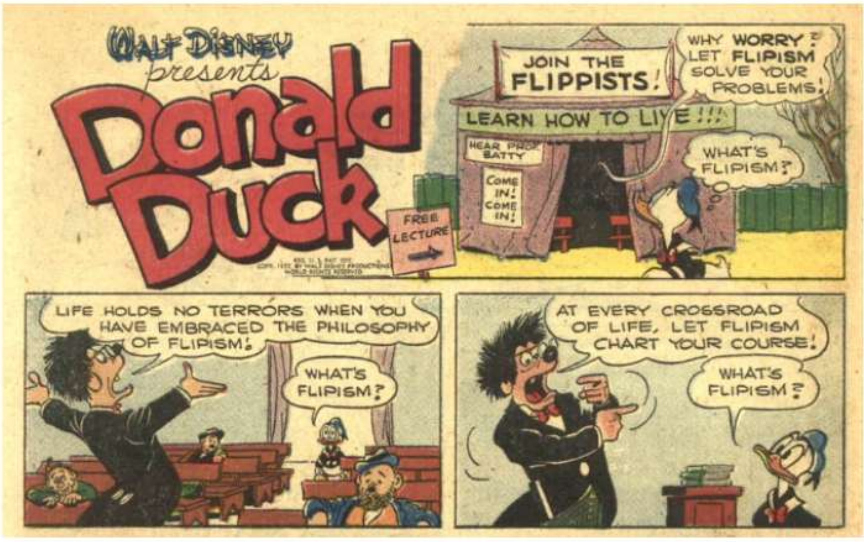


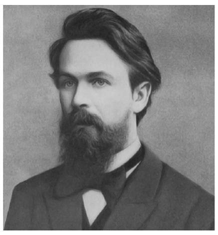
Figure 2.5: *When in doubt, flip a coin: life as a stochastic process. Donald Duck in "Flip Decision" — Carl Barks, 1953.*

6

Given any set of joint random variables $X_0, X_1, \ldots, X_T$, with probability distribution p, we can write

$$\begin{aligned} p(\mathbf{x}_1, \mathbf{x}_0) &= p(\mathbf{x}_1 \mid \mathbf{x}_0)\, p(\mathbf{x}_0) \\ p(\mathbf{x}_2, \mathbf{x}_1, \mathbf{x}_0) &= p(\mathbf{x}_2 \mid \mathbf{x}_1, \mathbf{x}_0)\, p(\mathbf{x}_1, \mathbf{x}_0) \\ &= p(\mathbf{x}_2 \mid \mathbf{x}_1, \mathbf{x}_0) p(\mathbf{x}_1 \mid \mathbf{x}_0) p(\mathbf{x}_0) \\ &\vdots \\ p(\mathbf{x}_{0:T}) &= p(\mathbf{x}_T \mid \mathbf{x}_{0:T-1})\, p(\mathbf{x}_{T-1} \mid \mathbf{x}_{0:T-2}) \ldots p(\mathbf{x}_1 \mid \mathbf{x}_0)\, p(\mathbf{x}_0) \qquad (2.1) \end{aligned}$$

7



Andrej Markov: Image from Wikimedia commons

Markov Property: $p(\mathbf{x}_t \mid \mathbf{x}_{0:t-1}) = p(\mathbf{x}_t \mid \mathbf{x}_{t-1})$.

This allows (2.1) to be written as

$$\begin{aligned} p(\mathbf{x}_{0:T}) &= p(\mathbf{x}_T \mid \mathbf{x}_{T-1})\, p(\mathbf{x}_{T-1} \mid \mathbf{x}_{T-2}) \ldots p(\mathbf{x}_1 \mid \mathbf{x}_0)\, p(\mathbf{x}_0) \\ &= p(\mathbf{x}_0) \prod_{t=1}^{T} p(\mathbf{x}_t \mid \mathbf{x}_{t-1}) \ . \end{aligned}$$

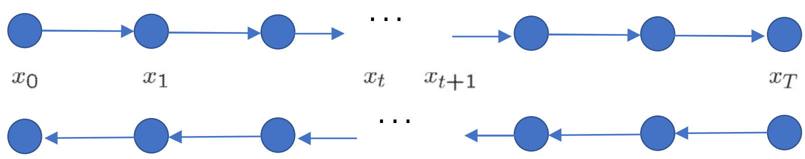
8

Markov Chains

Given any set of random variables $X_0, \ldots, X_T$ associated with a Markov Chain, the joint probability distribution q can be written

$$\boxed{q(\mathbf{x}_{0:T}) = q(\mathbf{x}_T|\mathbf{x}_{T-1})q(\mathbf{x}_{T-1}|\mathbf{x}_{T-2})\ldots q(\mathbf{x}_1|\mathbf{x}_0)q(\mathbf{x}_0)}$$

Does a Markov Chain have a reverse Markov Chain



9

Inversion of a Markov Chain

Joint probability of a Markov chain can be written as

$$\begin{aligned} q(\mathbf{x}_{0:T}) &= q(\mathbf{x}_T|\mathbf{x}_{T-1})\, q(\mathbf{x}_{T-1}|\mathbf{x}_{T-2}) \ldots q(\mathbf{x}_1|\mathbf{x}_0)\, q(\mathbf{x}_0) \\ &= q(\mathbf{x}_0) \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1}) \ . \end{aligned} \tag{3}$$

Inversion of a Markov chain. The formula (3) defines a joint probability distribution q on all the random variables $\mathbf{x}_t$ for $t = 0, \ldots, T$. Using Bayes' Law, we can also write (3) as

$$\begin{aligned} q(\mathbf{x}_{0:T}) &= q(\mathbf{x}_0) \prod_{t=1}^{T} \frac{q(\mathbf{x}_{t-1}|\mathbf{x}_t) q(\mathbf{x}_t)}{q(\mathbf{x}_{t-1})} \\ &= q(\mathbf{x}_0) \ \frac{q(\mathbf{x}_0|\mathbf{x}_1)q(\mathbf{x}_1)}{q(\mathbf{x}_0)} \ \frac{q(\mathbf{x}_1|\mathbf{x}_2)q(\mathbf{x}_2)}{q(\mathbf{x}_1)} \ \ldots \ \frac{q(\mathbf{x}_{T-1}|\mathbf{x}_T)q(\mathbf{x}_T)}{q(\mathbf{x}_{T-1})} \\ &= \cancel{q(\mathbf{x}_0)} \ \frac{q(\mathbf{x}_0|\mathbf{x}_1)\cancel{q(\mathbf{x}_1)}}{\cancel{q(\mathbf{x}_0)}} \ \frac{q(\mathbf{x}_1|\mathbf{x}_2)q(\mathbf{x}_2)}{\cancel{q(\mathbf{x}_1)}} \ \ldots \ \frac{q(\mathbf{x}_{T-1}|\mathbf{x}_T)q(\mathbf{x}_T)}{\cancel{q(\mathbf{x}_{T-1})}} \end{aligned}$$

This product telescopes by cancellation of the factors $q(\mathbf{x}_t)$, resulting in

$$\boxed{q(\mathbf{x}_{0:T}) = q(\mathbf{x}_T) \prod_{t=1}^{T} q(\mathbf{x}_{t-1}|\mathbf{x}_t) \ .}$$

10

Theorem 2.2.5 Invertibility of Markov Chains. *If a joint probability distribution $p_{0:T}$ is defined by a Markov chain with initial probability distribution $p(\mathbf{x}_0)$ and forward conditional probabilities $p(\mathbf{x}_t \mid \mathbf{x}_{t-1})$, then it may also be expressed as a "backwards" Markov chain with initial probability $p(\mathbf{x}_T)$ and conditional probabilities $p(\mathbf{x}_{t-1} \mid \mathbf{x}_t)$.* $\triangle$

$$Y = X \oplus G_\beta$$

$$\boxed{\mu(X \mid y) \approx y + \beta \, \nabla \log p_X(y)}$$

11

Computing probabilities of samples, after training

Write probabilities in two ways

$$\begin{aligned} q(\mathbf{x}_{0:T}) &= q(\mathbf{x}_T) \prod_{t=1}^{T} q(\mathbf{x}_{t-1}|\mathbf{x}_t) \\ &= q(\mathbf{x}_0) \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1}) \ . \end{aligned}$$

So

$$q(\mathbf{x}_0) = q(\mathbf{x}_T) \prod_{t=1}^{T} \frac{q(\mathbf{x}_{t-1}|\mathbf{x}_t)}{q(\mathbf{x}_t|\mathbf{x}_{t-1})} \ .$$

More exactly

$$\boxed{q(\mathbf{x}_0) = q(\mathbf{x}_T) \prod_{t=1}^{T} \frac{p_\theta(\mathbf{x}'_{t-1}|\mathbf{x}_t)}{q(\mathbf{x}_t|\mathbf{x}'_{t-1})}} \ .$$

12

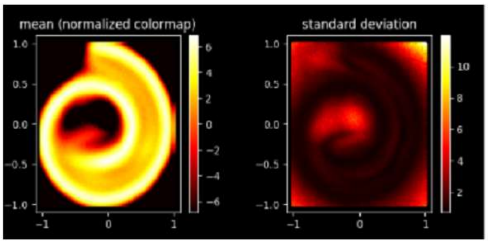
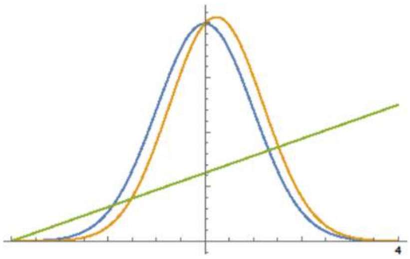


Figure 1: *Computed probabilities for points on a swiss roll. The left-hand figure shows graphically the probabilities of samples (2-d coordinates) from the swiss-roll dataset. For each point, the log-10 probability of each point (relative to a reference value) is encoded as a colour value. The graphic shows that points lying on the swiss-roll have probabilities a few orders of magnitude higher than all out-of-distribution points, and up to 10^6 times higher than more distant data (coloured black and dark red). The right hand graphic shows the standard-deviation of different estimates of each point. Overall, the standard-deviation of the probability estimate was around 0.01 of the probability value for each point.*

13

Form of inverse conditional probabilities



14

Form of inverse conditional probabilities

This is the approximation to the probabilities when $q(\mathbf{x}_t)$ is linear in the region of support of $q(\mathbf{x}_t|\mathbf{x}_{t-1})$.

$$\boxed{\mu(X_{t-1}|\mathbf{x}_t) = (\mathbf{x}_t + \beta_t^2 \mathbf{m})/\alpha_t}$$

Mean learnt by training

and covariance matrix

$$\boxed{\Sigma = \beta_t^2(\mathrm{I} - \beta^2 \mathbf{m}\mathbf{m}^\top)/\alpha_t^2 \ .}$$

Variance approximately same as forward variance.

where

$$\mathbf{m} = \nabla \log p(\mathbf{x}_{t-1}|\mathbf{x}_t) .$$

Bonus:

If we compute $\mu(X_{t-1}|\mathbf{x}_t)$, then we know

$$\nabla \log p(\mathbf{x}_{t-1}|\mathbf{x}_t) \equiv \mu(X_{t-1}|\mathbf{x}_t) - \mathbf{x}_t/\alpha_t \ .$$

15

How do we find the backward conditional mean?
(learning)

Theorem 2.2.8. *Let X, Y be a pair of joint random variables and let $f_*(\mathbf{y})$ be the function that minimizes the loss*

$$L(f_\theta) = \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim(X,Y)}\left[\|f_\theta(\mathbf{y}) - \mathbf{x}\|^2\right]$$

then

$$f_*(\mathbf{y}) = \mathbb{E}_{\mathbf{x}\sim X|\mathbf{y}}[\mathbf{x}] = \mu(X \mid \mathbf{y}) \ .$$

16

Theorem 2.2.6. *Let X and Y be two vector-valued random variables such that*

$$Y = X \oplus G_\beta \ .$$

Assume that p_X is well approximated by a linear function in the support region of the Gaussian $G_\beta(\mathbf{x}-\mathbf{y})$, a Gaussian with variance β centred at $\mathbf{y}$. Then

$$\mu(X \,|\, \mathbf{y}) \approx \mathbf{y} + \beta \nabla \log p_X(\mathbf{y}) \tag{2.16}$$

$$\mathrm{Covar}(X \,|\, \mathbf{y}) \approx \beta\, \mathtt{I} - (\mu - \mathbf{y})(\mu - \mathbf{y})^T \tag{2.17}$$

$$= \beta \mathtt{I} + O(\beta^2) \ , \tag{2.18}$$

abbreviating $\mu(X \,|\, \mathbf{y}$ as μ. △

Theorem 2.2.7. *Let X and Y be two vector-valued random variables such that*

$$Y = \alpha X \oplus G_\beta \ .$$

$$\mu(X \,|\, \mathbf{y}) \approx (\mathbf{y} + \beta \nabla \log p_X(\mathbf{y}/\alpha))\, /\alpha$$

$$\mathrm{Covar}(X \,|\, \mathbf{y}) \approx \beta/\alpha^2\, \mathtt{I} - (\mu - \mathbf{y}/\alpha)(\mu - \mathbf{y}/\alpha)^T$$

17

If $Y = X \oplus G_\beta$ for β small, then $\nabla \log p_X(\mathbf{x}) \approx \dfrac{\mu(X \,|\, Y = \mathbf{x}) - \mathbf{x}}{\beta}$

$$\boxed{\nabla \log p_X(\mathbf{x}) \approx \frac{\mu_\theta(\mathbf{x}) - \mathbf{x}}{\beta} \ .}$$

18

Minimizing least-squared error computes the conditional

Theorem 2.2. *Let X, Y be a pair of joint random variables. Let $f_*(\mathbf{y})$ be the function that minimizes the functional*

$$\boxed{L(f_\theta) = \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim(X,Y)}\left[\|f_\theta(\mathbf{y}) - \mathbf{x}\|^2\right]}$$

then

$$\boxed{f_*(\mathbf{y}) = \mathbb{E}_{\mathbf{x}\sim X|\mathbf{y}}[\mathbf{x}] = \mu(X|\mathbf{y})\ .}$$

□

Proof. We summarize the proof given above. By definition,

$$\begin{aligned} f_* &= \arg\min_{f_\theta} \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim(X,Y)}\left[\|f_\theta(\mathbf{y}) - \mathbf{x}\|^2\right] \\ &= \arg\min_{f_\theta} \mathbb{E}_{\mathbf{y}\sim Y}\left[\mathbb{E}_{\mathbf{x}\sim X|\mathbf{y}}\left[\|f_\theta(\mathbf{y}) - \mathbf{x}\|^2\right]\right] \end{aligned}$$

This minimum is attained by $f_*(\mathbf{y})$ defined independently for each $\mathbf{y}$ by

$$\begin{aligned} f_*(\mathbf{x}) &= \arg\min_{\hat{\mathbf{x}}} \mathbb{E}_{\mathbf{x}\sim X|\mathbf{y}}\left[\|\hat{\mathbf{x}} - \mathbf{x}\|^2\right] \\ &= \mathbb{E}_{x\sim X|\mathbf{y}}[\mathbf{x}]\ . \end{aligned}$$

□

19

Algorithm

1. A value of t is chosen, at random or by some process.
2. Samples $\mathbf{x}_i$ are sampled from X_{t-1}.
3. A sample $\mathbf{y}_i$ is chosen from $X_t = \alpha_t X_{t-1} \oplus G_{\beta_t}$ according to $\mathbf{y}_i = \alpha_t \mathbf{x}_i = \sqrt{\beta_t}\,\epsilon$.
4. Applying the function μ_θ one obtains $\hat{\mathbf{x}}_i = \mu_\theta(\mathbf{y}_i, t)$.
5. The cost function[3] is given as $\|\mu_\theta(\mathbf{y}_i, t) - \mathbf{x}_i\|^2$.
6. Therefore, over several samples, the total cost is
$$L(\theta, t, \{\mathbf{x}_i, \mathbf{y}_i\}) = \sum_i \|\mu_\theta(\mathbf{y}_i, t) - \mathbf{x}_i\|^2$$
which is minimized over the parameters θ.
7. The backward conditional $p(X_{t-1} \,|\, \mathbf{x}_t)$ is defined as a Gaussian with
$$\begin{aligned} \mu(X_{t-1} \,|\, \mathbf{x}_t) &= \mu_\theta(\mathbf{x}_t, t) \\ \mathrm{Covar}(X_{t-1} \,|\, \mathbf{x}_t) &= \Sigma_{\theta,t}(\mathbf{x}_t) = \beta_t \mathtt{I} - (\mu_\theta(\mathbf{x}_t, t) - \mathbf{x}_t)(\mu_\theta(\mathbf{x}_t, t) - \mathbf{x}_t)^T \end{aligned}$$
or more simply $\mathrm{Covar}(X_{t-1} \,|\, \mathbf{x}_t) = \beta_t/\alpha_t^2\,\mathtt{I}$.

20

Finding the backward conditionals by training.

As long as the variances β_t^2 are small, a reasonable approximation is that the backward conditionals are normal distributions, and that they have the same variance as the forward conditionals. In this case, one can assume a form

$$\boxed{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \mu_{\theta,t}(\mathbf{x}_t), \beta_t^2 \mathtt{I})} \tag{18}$$

and the goal of training is to compute the means

$$\mu_{\theta,t}(\mathbf{x}_t) = \mathbb{E}_{\mathbf{x}_{t-1} \sim (X_{t-1}|\mathbf{x}_t)}[\mathbf{x}_{t-1}] \ .$$

21

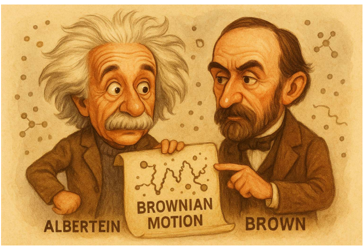
The solution

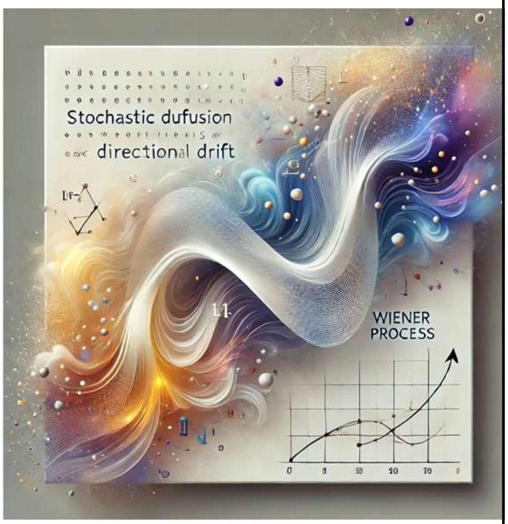
22

Computing the way back.

23

Continuous Stochastic Processes





Ito process

24

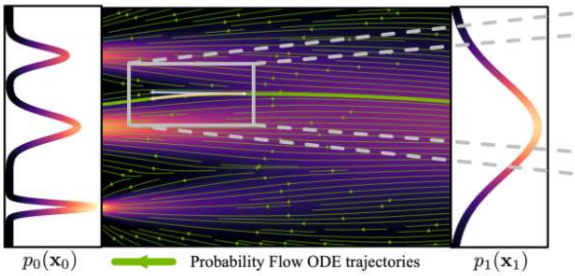
4 Continuous diffusion

The formula for $\mathbf{x}_t$

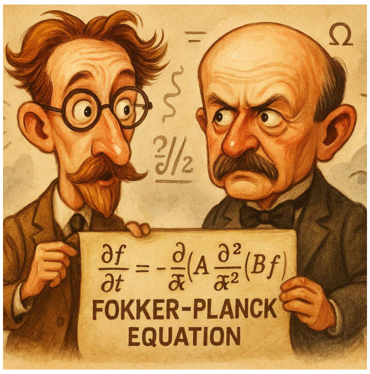
$$\boxed{\mathbf{x}_t = \bar{\alpha}_t \mathbf{x}_0 + \bar{\beta}_t \epsilon} \tag{20}$$

can be made continuous, as follows:

$$\boxed{\mathbf{x}(t) = \bar{\alpha}(t) \mathbf{x}_0 + \bar{\beta}(t) \epsilon} \tag{21}$$



25



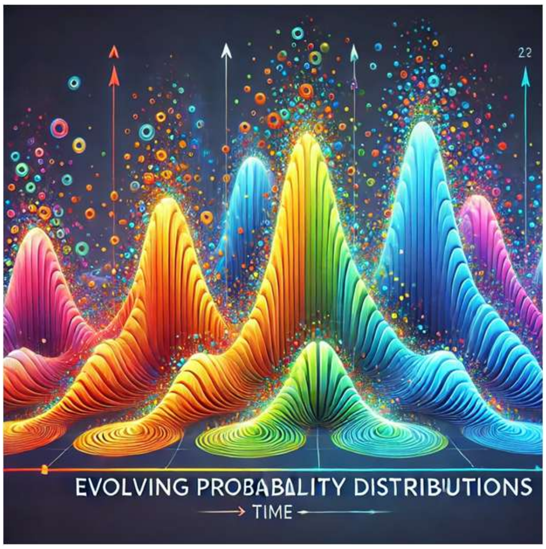
26

The term "stochastic" originates from the Ancient Greek word στοχαστικός (*stokhastikós*), which means "skillful in aiming" or "pertaining to guessing." This, in turn, is derived from στοχάζομαι (*stokházomai*), meaning "to aim at a target" or "to guess," and στόχος (*stókhos*), meaning "an aim" or "a target." EN.WIKTIONARY.ORG

In English, "stochastic" was first recorded in 1662, used as an adjective meaning "pertaining to conjecturing." Over time, its usage evolved, and by the 20th century, it came to describe processes involving randomness or probability. For instance, in 1934, mathematician Joseph L. Doob used the term "stochastic process" to refer to a collection of random variables indexed by time or space. EN.WIKIPEDIA.ORG

Today, "stochastic" is commonly used in fields such as mathematics, statistics, and physics to describe systems or processes that are inherently random or probabilistic in nature.

While "stochastic" is the primary English word directly connected to στόχος, other terms like "stich" and "stichos" are derived from related Greek roots such as στίχος (*stíkhos*), meaning "row" or "line," which is associated with the concept of alignment or order. EN.WIKTIONARY.ORG EN.WIKTIONARY.ORG

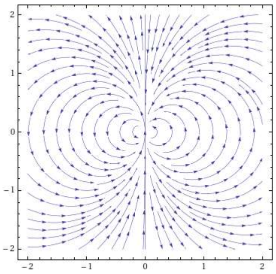


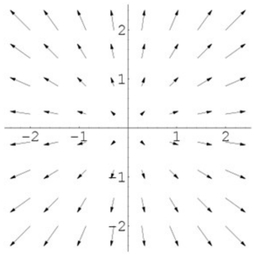
How does a probability distribution evolve with time?

Drift but no diffusion

. If a family of probability density functions $\{p_t\}$ are associated with drift $\alpha(\mathbf{x},t)$

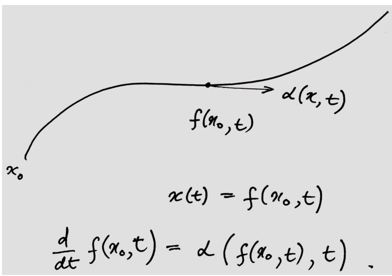
$$\frac{d}{dt}p_t(\mathbf{x}) = -\langle \nabla p_t(\mathbf{x}), \alpha(\mathbf{x},t)\rangle - p_t(\mathbf{x})\,\mathbf{div}\,\alpha(\mathbf{x},t)$$





29

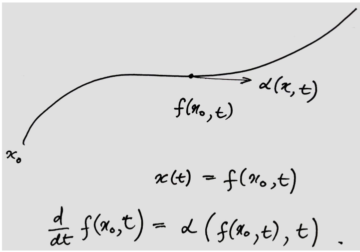
Oil floating on a slowly moving stream. How does the density of oil at a given point in the stream change?



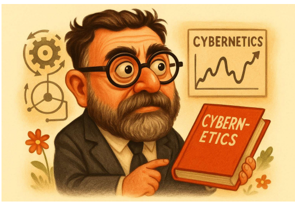
30

Lemma 3.2.11. *Given a deterministic process defined by the ODE* $\mathbf{x}'(t) = \alpha(\mathbf{x}, t)$*, then*

$$\begin{aligned}\frac{\partial p(\mathbf{x},t)}{\partial t} &= -\big\langle \nabla p(\mathbf{x},t), \alpha(\mathbf{x},t)\big\rangle - p(\mathbf{x},t)\,\mathbf{div}\,\alpha(\mathbf{x},t)\\ &= -\mathbf{div}\,(p\,\alpha)\end{aligned}$$



31



Wiener Process: no drift

$$dX = \sqrt{\beta}\; dW$$

Lemma 3.2.10. *If* $\{p_t\}$ *is a family of distributions defined by an Itô process* $\mathcal{I}(0, \beta, p)$, , *defined on* $\mathbb{R}^n$*, then*

$$\begin{aligned}\frac{\partial}{\partial t} p(\mathbf{x},t) &= \frac{\beta(t)}{2}\nabla^2 p(\mathbf{x},t)\\ &= \frac{\beta}{2}\;\mathbf{div}\,(p\,\nabla \log p)\ .\end{aligned} \tag{3.6}$$

Lemma 3.2.11. *Given a deterministic process defined by the ODE* $\mathbf{x}'(t) = \alpha(\mathbf{x}, t)$*, then*

$$\frac{\partial p(\mathbf{x},t)}{\partial t} = -\mathbf{div}\,(p\,\alpha)\ .$$

32

Focker-Planck Equations

Describe the evolution of the probability p(x, t) of a point at time t, undergoing a diffusion process.

Pure diffusion

Lemma 3.3.14. *If* $\{p_t\}$ *are a Gaussian family of distributions,* $p_t = G_{\beta(t)} * p_0$*, defined on* $\mathbb{R}^n$*, then*

$$\frac{d}{dt}p_t(\mathbf{x}) = \beta'(t)\frac{\nabla^2 p_t(\mathbf{x})}{2} \ .$$

33

33

Theorem 3.2.12 (Fokker-Planck equation). *For a general Itô process* $\mathcal{I}(\alpha, \beta, p_0)$ *with vector field* $\alpha(\mathbf{x}, t)$*, noise-schedule* $\beta(t)$ *and starting distribution* $p_0(\mathbf{x})$*,*

$$\boxed{\frac{\partial p}{\partial t} = -\mathbf{div}\,(p\,\mathbf{v}) \ .}$$

where

$$\boxed{\mathbf{v}(\mathbf{x}, t) = \alpha(\mathbf{x}, t) - \frac{\beta(t)}{2}\nabla \log p(\mathbf{x}, t) \ .}$$

34

Lemma 3.3.13. *The process $\mathcal{I}(\alpha, \beta, p_0)$ is equivalent in distribution to the non-stochastic process*[3]

$$\mathcal{I}(\alpha - \beta/2\, \nabla \log p, 0, p_0)$$

$\triangle$

In fact, to any process

$$\mathcal{I}(\alpha - (\lambda/2)\, \nabla \log p, \beta - \lambda, p_0)\ ,$$

where $\beta(t) - \lambda(t) > 0$ for all t.

Proposition 3.3.14. *Two processes $\mathcal{I}(\alpha_1, \beta_1, p_0)$ and $\mathcal{I}(\alpha_2, \beta_2, q_0)$ are equivalent in distribution if and only if $p_0 = q_0$ and*

$$\alpha_1 - \beta_1/2\, \nabla \log p = \alpha_2 - \beta_2/2\, \nabla \log p\ .$$

$\triangle$

35

We now define the reverse of an Itô process. Two processes $\mathcal{I}(\alpha_1, \beta_1, p_0)$ and $\mathcal{I}(\alpha_2, \beta_1, q_0)$ defined for a time interval $t \in [0, 1)$, are called reverse of each other provided $p_t = q_{1-t}$. An easy case where two processes are reverses is available in the deterministic case.

Lemma 3.3.15. *Non-stochastic processes $\mathcal{I}(\alpha, 0, p_0)$ and $\mathcal{I}(-\tilde{\alpha}, 0, p_1)$ defined for $t \in [0, 1]$ are reverses.* $\triangle$

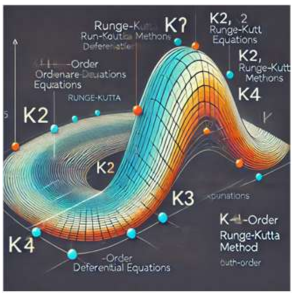
Theorem 3.3.16 (Reversibility of an Itô process). *Itô processes $\mathcal{I}(\alpha, \beta, p_0)$ and $\mathcal{I}(-\tilde{\alpha} + \tilde{\beta}\, \nabla \log \tilde{p}, \tilde{\beta}, p_1)$ are reverses:*

$$\boxed{\mathcal{I}(\alpha, \beta, p_0)^{-1} \sim \mathcal{I}(-\tilde{\alpha} + \tilde{\beta}\, \nabla \log \tilde{p}, \tilde{\beta}, p_1)}\ .$$

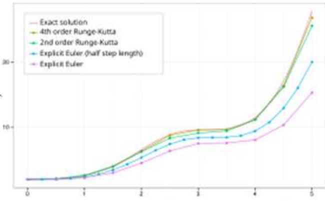
Proof. We compute

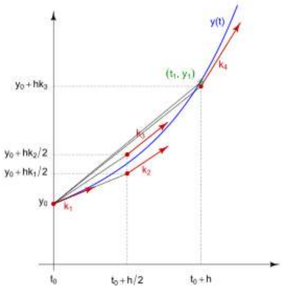
$$\begin{aligned}
\mathcal{I}(\alpha, \beta, p_0)^{-1} &\sim \mathcal{I}(\alpha - \beta/2\, \nabla \log p, 0, p_0)^{-1} \\
&\sim \mathcal{I}(-\tilde{\alpha} + \tilde{\beta}/2\, \nabla \log \tilde{p}, 0, p_1) \\
&\sim \mathcal{I}(-\tilde{\alpha} + \tilde{\beta} \nabla \log \tilde{p} - \tilde{\beta}/2\, \nabla \log \tilde{p}, 0, p_1) \\
&\sim \mathcal{I}(-\tilde{\alpha} + \tilde{\beta} \nabla \log \tilde{p}, \tilde{\beta}, p_1)\ .
\end{aligned}$$

36



Runge-Kutta method





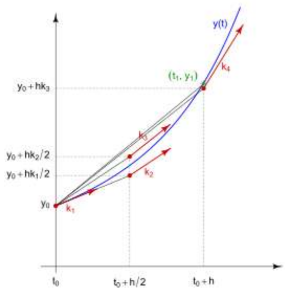
37

Now we pick a step-size $h > 0$ and define:

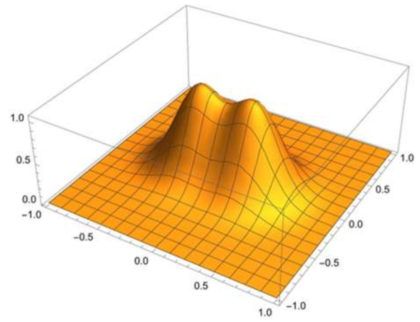
$$y_{n+1} = y_n + \frac{h}{6}\left(k_1 + 2k_2 + 2k_3 + k_4\right),$$
$$t_{n+1} = t_n + h$$

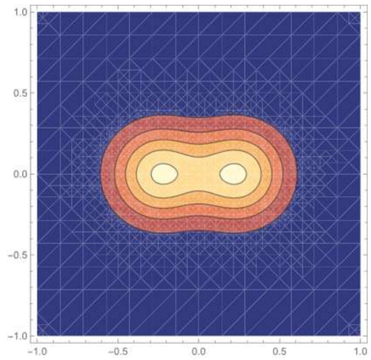
for n = 0, 1, 2, 3, ..., using[3]

$$k_1 = f(t_n, y_n),$$
$$k_2 = f\left(t_n + \frac{h}{2}, y_n + h\frac{k_1}{2}\right),$$
$$k_3 = f\left(t_n + \frac{h}{2}, y_n + h\frac{k_2}{2}\right),$$
$$k_4 = f(t_n + h, y_n + hk_3).$$

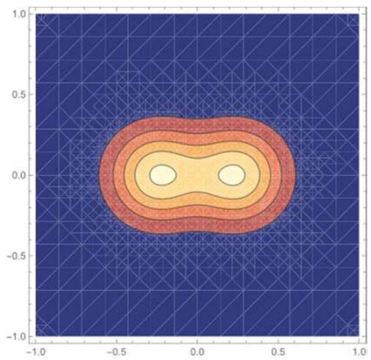


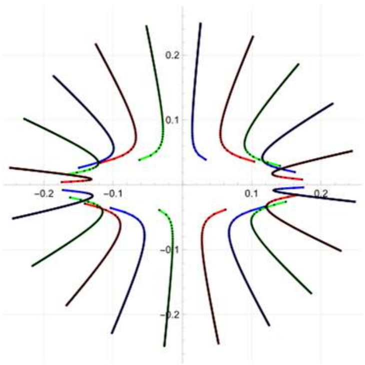
38


1.0
0.5
0.0
−1.0
−0.5
0.0
0.5
1.0
1.0
0.5
0.0
−0.5
−1.0

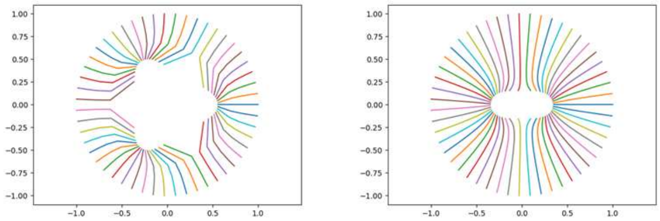

1.0
0.5
0.0
−0.5
−1.0
−1.0
−0.5
0.0
0.5
1.0

39


1.0
0.5
0.0
−0.5
−1.0
−1.0
−0.5
0.0
0.5
1.0


0.2
0.1
−0.2
−0.1
0.1
0.2
−0.1
−0.2

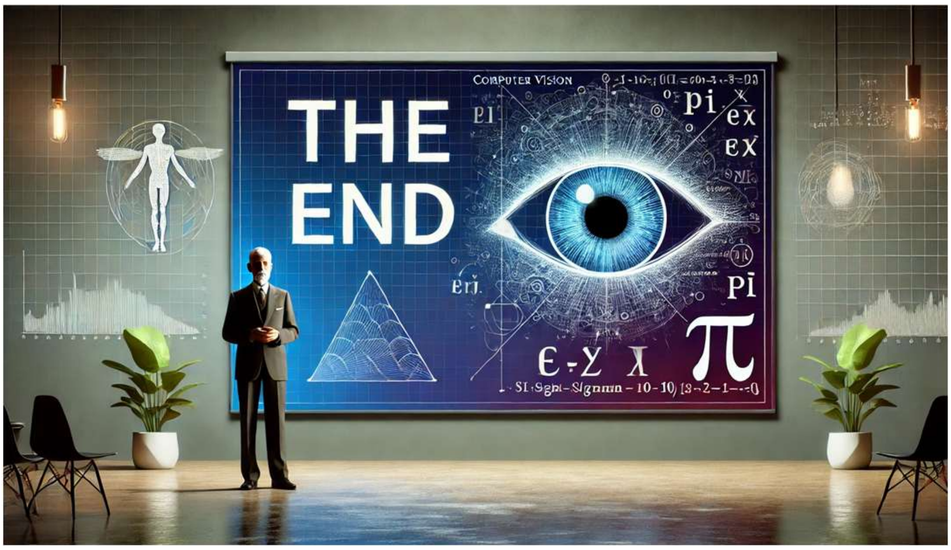
40



Exercise: Verify that the reverse of the reverse is the original process. In other words, show that

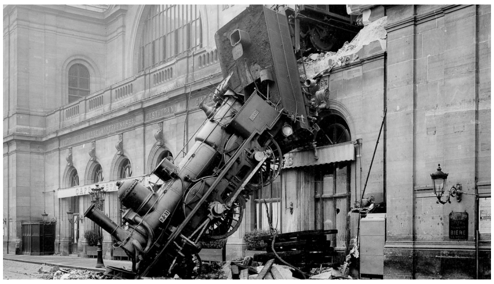
$$\mathcal{I}(-\tilde{\alpha} + \tilde{\beta}\nabla \log \tilde{p}, \tilde{\beta}, p_1)^{-1} = \mathcal{I}(\alpha, \beta, p_0)$$

Exercise: Let p_0 be an initial distribution, and define a (time independent) vector field by $\alpha(\mathbf{x}, t) = \beta(t)/2\,\nabla \log p_0$. What can be said about the Itô process $\mathcal{I}[\alpha, \beta, p_0)$?



43

Failure cases – always something that will go wrong.



44