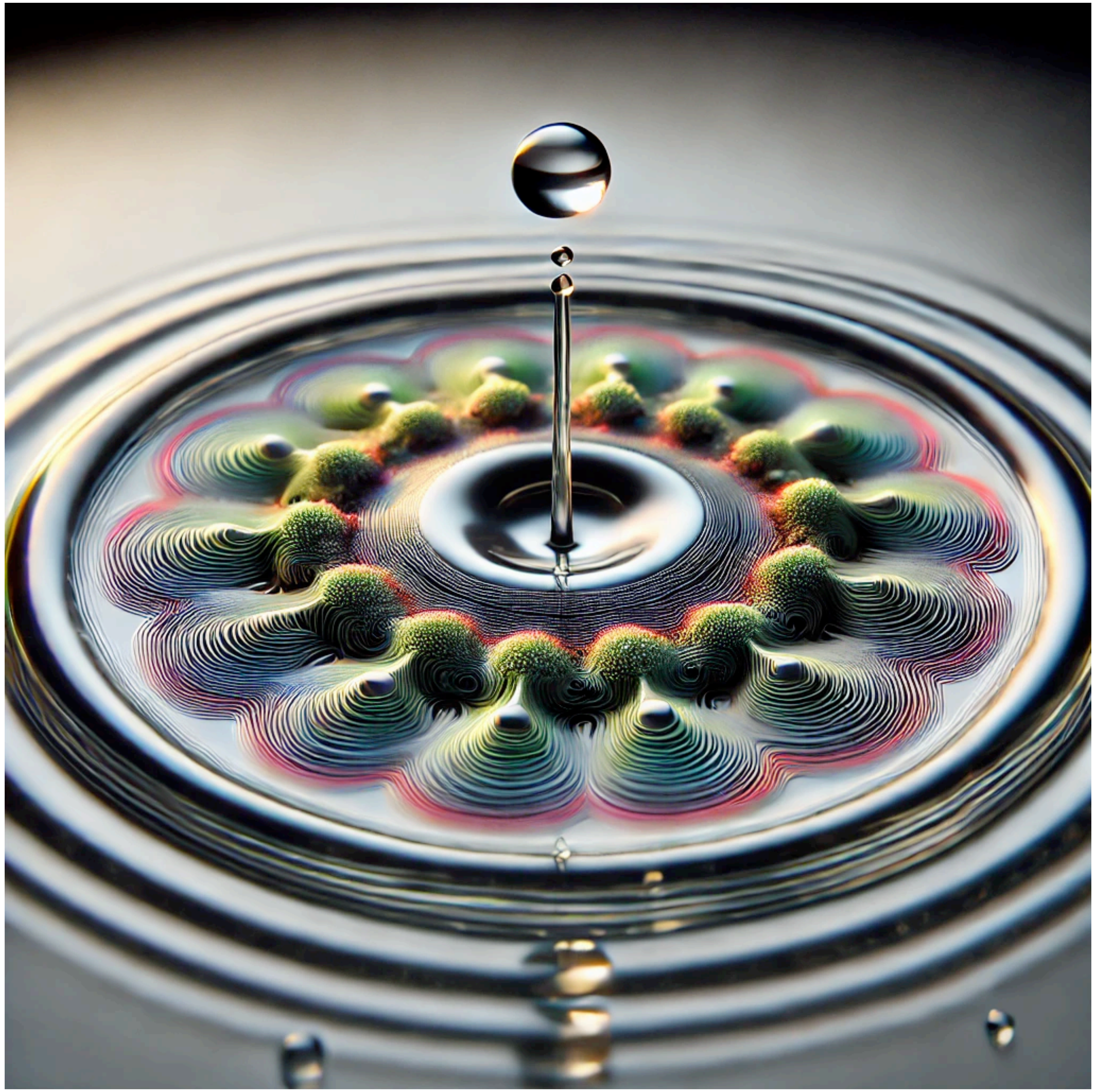
School of Data Analysis

Diffusion models via SDE and ODE

Mikhail Romanov

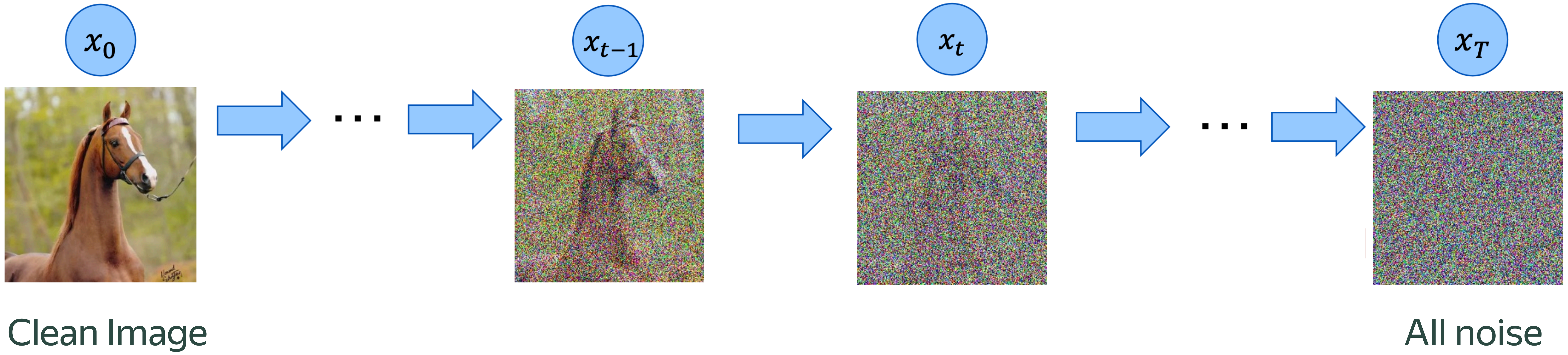
Senior Engineer at YandexART

Discrete Diffusion Process

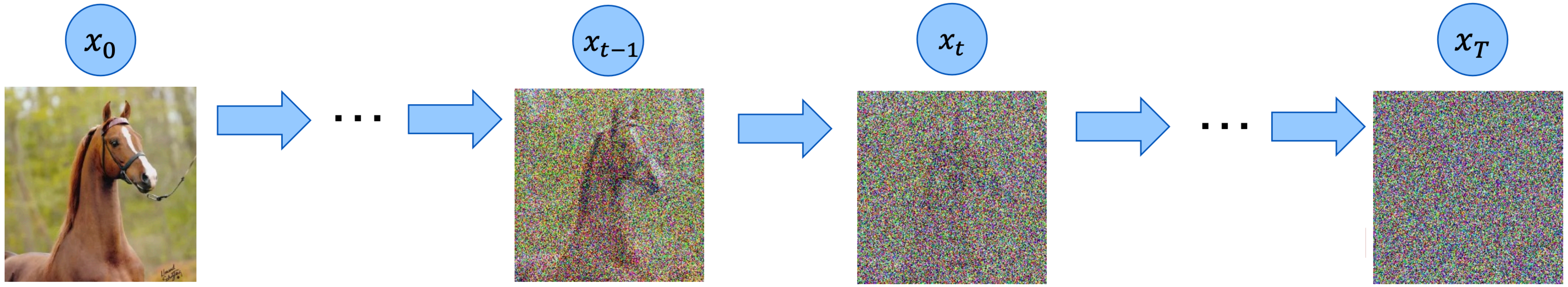


Where we stopped last time...

Discrete Forward Diffusion Process



Discrete Forward Diffusion Process



Clean Image

All noise

$$\mathbf{x}_t = \sqrt{1-\bar{\beta}_t} \cdot \mathbf{x}_0 + \sqrt{\bar{\beta}_t} \cdot \boldsymbol{\epsilon}_t$$

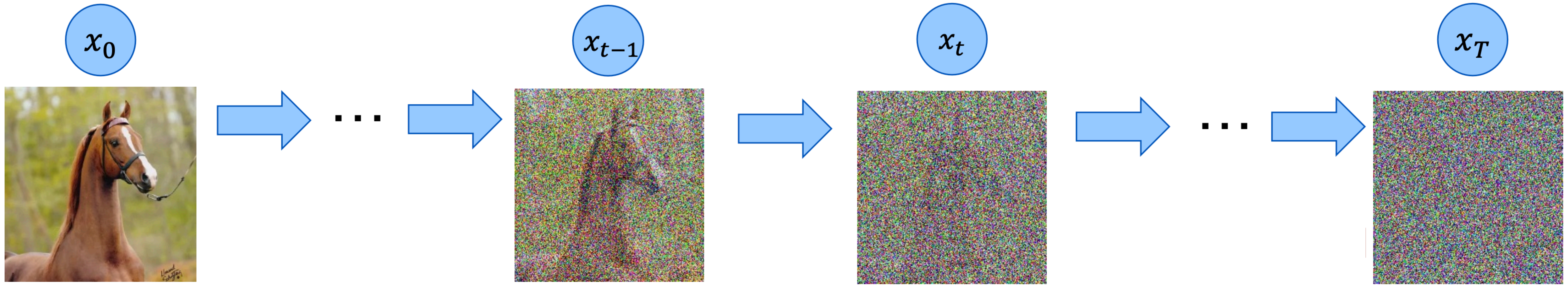
$$\boldsymbol{\epsilon}_t \sim \mathcal{N}(0, I)$$

$\bar{\beta}_t = 1$ — Noise
$\bar{\beta}_t = 0$ — Clean Image

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

$$d\bar{\beta} = \beta dt$$

Discrete Forward Diffusion Process



Clean Image

All noise

$$\mathbf{x}_t = \sqrt{1-\beta_t} \cdot \mathbf{x}_0 + \sqrt{\beta_t} \cdot \boldsymbol{\epsilon}_t$$

$$\boldsymbol{\epsilon}_t \sim \mathcal{N}(0, I)$$

$\beta_t = 1$ — Noise
$\beta_t = 0$ — Clean Image

$$1 - \beta_t = \bar{\alpha}_t = 1 - \sigma_t^2$$

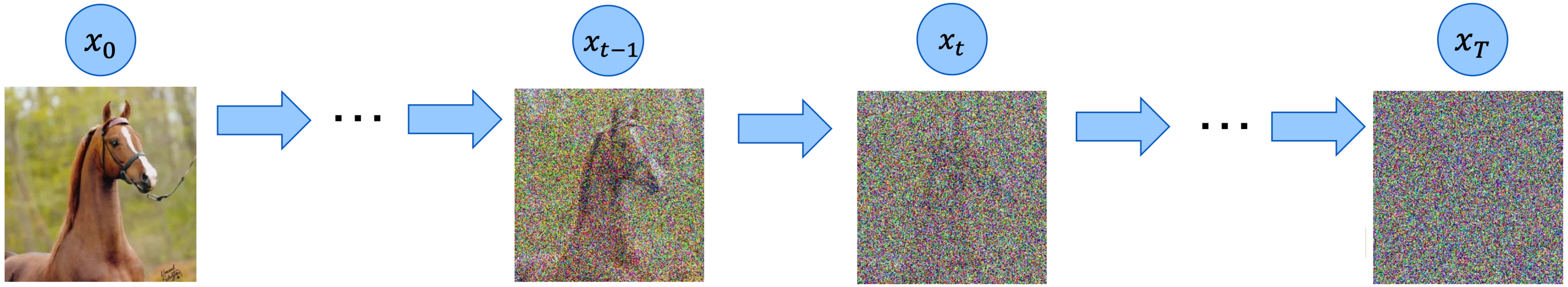
Network extracts noise:

$$\mathbf{x}_t \rightarrow \text{NNet} \rightarrow \boldsymbol{\epsilon}(\mathbf{x}_t, t)$$

Input: noisy image $\mathbf{x}_t$
Output: noise $\boldsymbol{\epsilon}(\mathbf{x}_t, t)$

$$\mathbf{x}_0 = \frac{\mathbf{x}_t - \boldsymbol{\epsilon}_t \sqrt{\beta_t}}{\sqrt{1-\beta_t}}$$

Discrete Forward Diffusion Process



Clean Image

All noise

$$\mathbf{x}_t = \sqrt{1-\beta_t} \cdot \mathbf{x}_0 + \sqrt{\beta_t} \cdot \boldsymbol{\epsilon}_t$$

$$\boldsymbol{\epsilon}_t \sim \mathcal{N}(0, I)$$

$\beta_t = 1$ — Noise
$\beta_t = 0$ — Clean Image

$$1 - \beta_t = \bar{\alpha}_t = 1 - \sigma_t^2$$

Network extracts noise:

$$\mathbf{x}_t \rightarrow \text{NNet} \rightarrow \boldsymbol{\epsilon}(\mathbf{x}_t, t)$$

Input: noisy image $\mathbf{x}_t$
Output: noise $\boldsymbol{\epsilon}(\mathbf{x}_t, t)$

$$\mathbf{x}_0 = \frac{\mathbf{x}_t - \boldsymbol{\epsilon}_t\sqrt{\beta_t}}{\sqrt{1-\beta_t}}$$

Having $\boldsymbol{\epsilon}_t$, we can estimate $\mathbf{x}_0$:

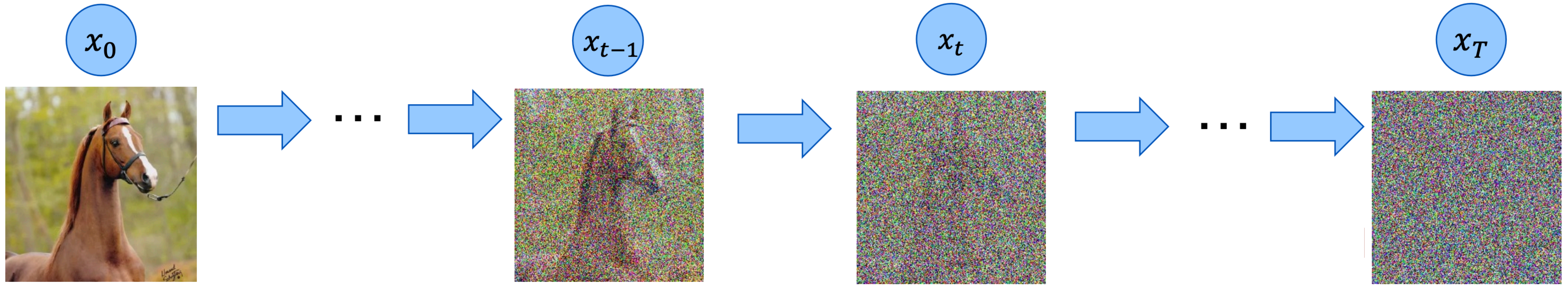
$$\mathbf{x}_0 = \frac{\mathbf{x}_t - \boldsymbol{\epsilon}_t\sqrt{\beta_t}}{\sqrt{1-\beta_t}}$$

From pure noise we can predict the image (?!?!)

No, we can't!

BUT we will have a hallucination of $\mathbf{x}_0$ (convolution with large Gaussian Kernel)

Discrete Forward Diffusion Process



Clean Image

All noise

$$\mathbf{x}_t = \sqrt{1-\beta_t} \cdot \mathbf{x}_0 + \sqrt{\beta_t} \cdot \boldsymbol{\epsilon}_t$$

$$\boldsymbol{\epsilon}_t \sim \mathcal{N}(0, I)$$

$\beta_t = 1$ — Noise
$\beta_t = 0$ — Clean Image

$$1 - \beta_t = \bar{\alpha}_t = 1 - \sigma_t^2$$

<u>Network</u> extracts noise:

$$\mathbf{x}_t \rightarrow \text{NNet} \rightarrow \boldsymbol{\epsilon}(\mathbf{x}_t, t)$$

Input: noisy image $\mathbf{x}_t$
Output: noise $\boldsymbol{\epsilon}(\mathbf{x}_t, t)$

$$\mathbf{x}_0 = \frac{\mathbf{x}_t - \boldsymbol{\epsilon}_t \sqrt{\beta_t}}{\sqrt{1-\beta_t}}$$

Having $\boldsymbol{\epsilon}_t$, we can estimate $\mathbf{x}_0$:

$$\mathbf{x}_0 = \frac{\mathbf{x}_t - \boldsymbol{\epsilon}_t \sqrt{\beta_t}}{\sqrt{1-\beta_t}}$$

From pure noise we can predict the image (?!?!)

No, we can't!

BUT we will have a hallucination of $\mathbf{x}_0$ (convolution with large Gaussian Kernel)

Several problems here:

- Noise means $\beta_t = 1$, division by zero!
- Can solve it: take $\beta_t = 0.99$

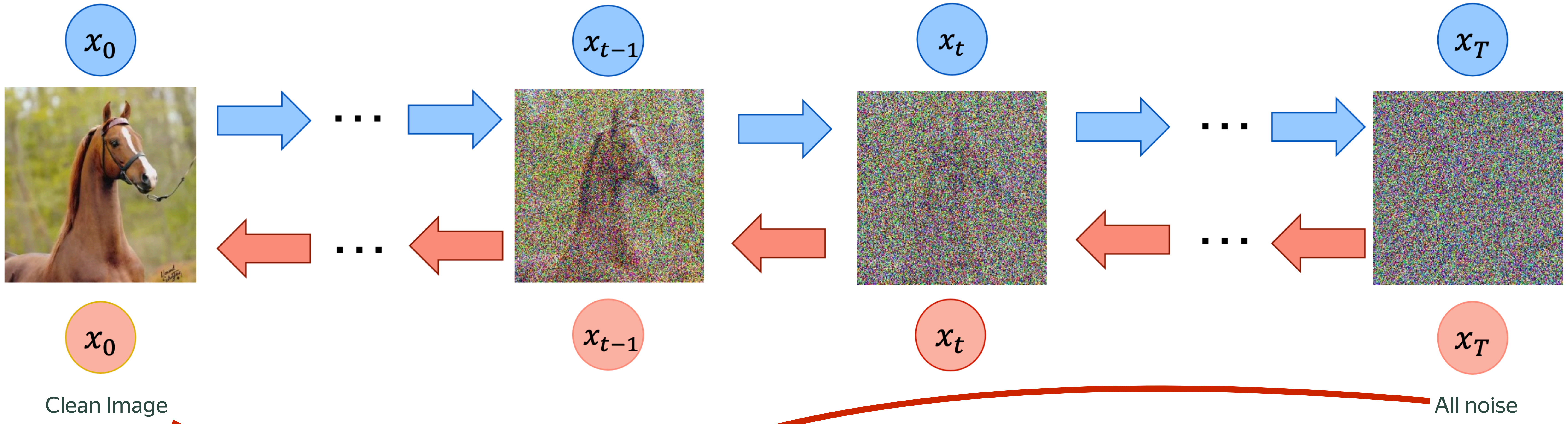
In $\boldsymbol{\epsilon}_t$, there is 1% of image $\mathbf{x}_0$ information

Quality of $\mathbf{x}_0$ depends on the quality of ϵ_t estimate

Surely, $\mathbf{x}_0$ will not be a high quality image

But it will be a good start

DDIM: Denoising Diffusion Implicit Model



Further approach:

$$\mathbf{x}_t \to \tilde{\mathbf{x}}_{t-1} \to \tilde{\mathbf{x}}_{t-2} \to \dots \to \tilde{\mathbf{x}}_0$$

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t} \cdot \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \cdot \epsilon_t$$

$$\epsilon_t \sim \mathcal{N}(0, I)$$

$\bar{\alpha}_t = 0$ — Noise

$\bar{\alpha}_t = 1$ — Clean Image

$$\mathbf{x}_t \to \text{NNet} \to \tilde{\epsilon}_t$$

Estimate of Image:

$$\tilde{\mathbf{x}}_0 = \frac{\mathbf{x}_t - \tilde{\epsilon}_t \sqrt{1 - \bar{\alpha}_t}}{\sqrt{\bar{\alpha}_t}}$$

Surely, $\mathbf{x}_0$ will not be a high quality image

But it will be a good start

Let us estimate $\mathbf{x}_{t-1}$

$$\tilde{\mathbf{x}}_0 = \frac{\mathbf{x}_t - \tilde{\epsilon}_t \sqrt{1 - \bar{\alpha}_t}}{\sqrt{\bar{\alpha}_t}}$$

$$\tilde{\mathbf{x}}_{t-1} = \tilde{\mathbf{x}}_0 \sqrt{\bar{\alpha}_{t-1}} + \tilde{\epsilon}_t \sqrt{1 - \bar{\alpha}_{t-1}}$$

$$\tilde{\mathbf{x}}_{t-1} = \underbrace{\frac{\mathbf{x}_t - \tilde{\epsilon}_t \sqrt{1 - \bar{\alpha}_t}}{\sqrt{\bar{\alpha}_t}}}_{\tilde{\mathbf{x}}_0} \sqrt{\bar{\alpha}_{t-1}} + \tilde{\epsilon}_t \sqrt{1 - \bar{\alpha}_{t-1}}$$

DDIM

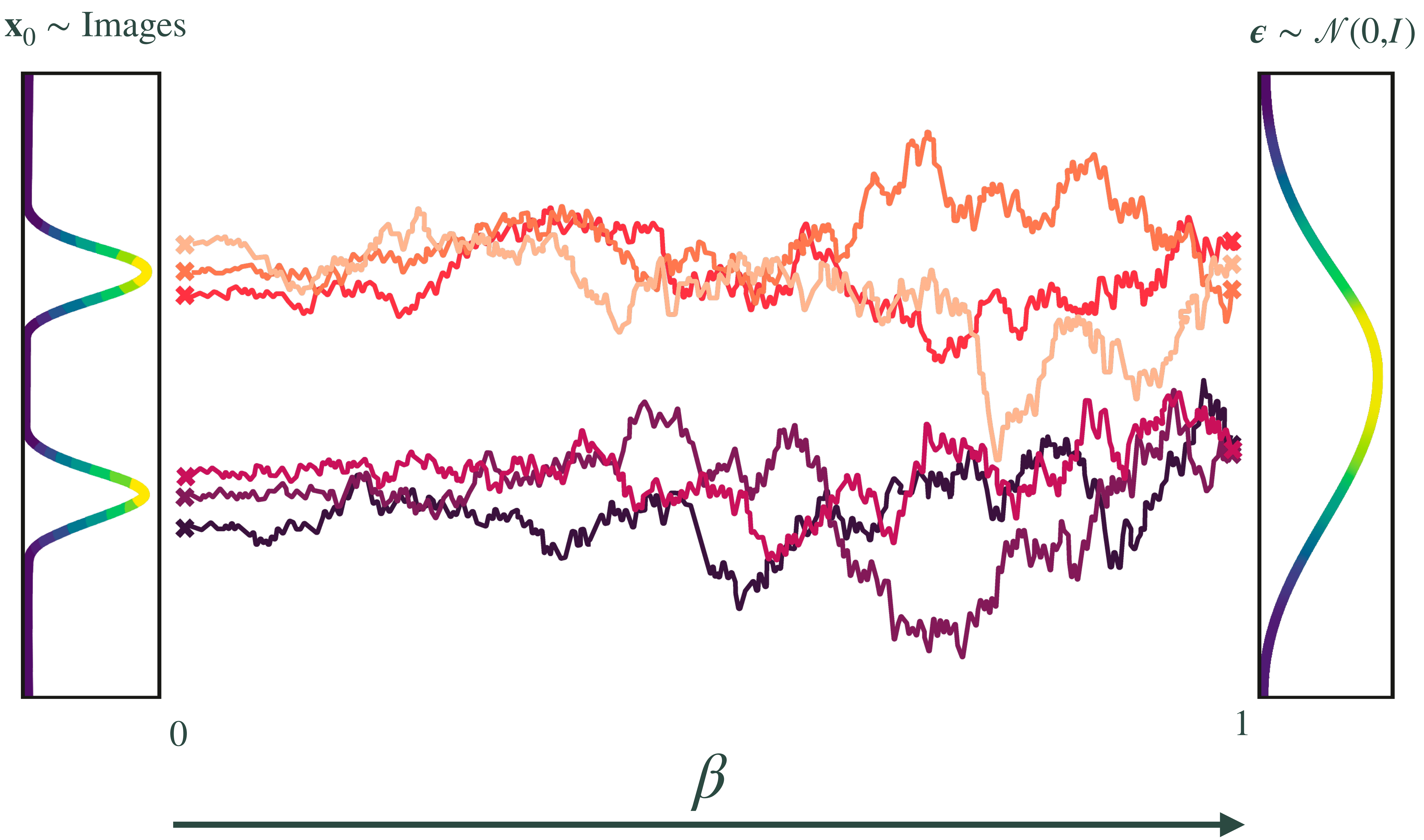
Now we have a slightly denoised Image!

The process that we have:

$$\mathbf{x}_t = \sqrt{1-\beta_t} \cdot \mathbf{x}_0 + \sqrt{\beta_t} \cdot \boldsymbol{\epsilon}_t$$

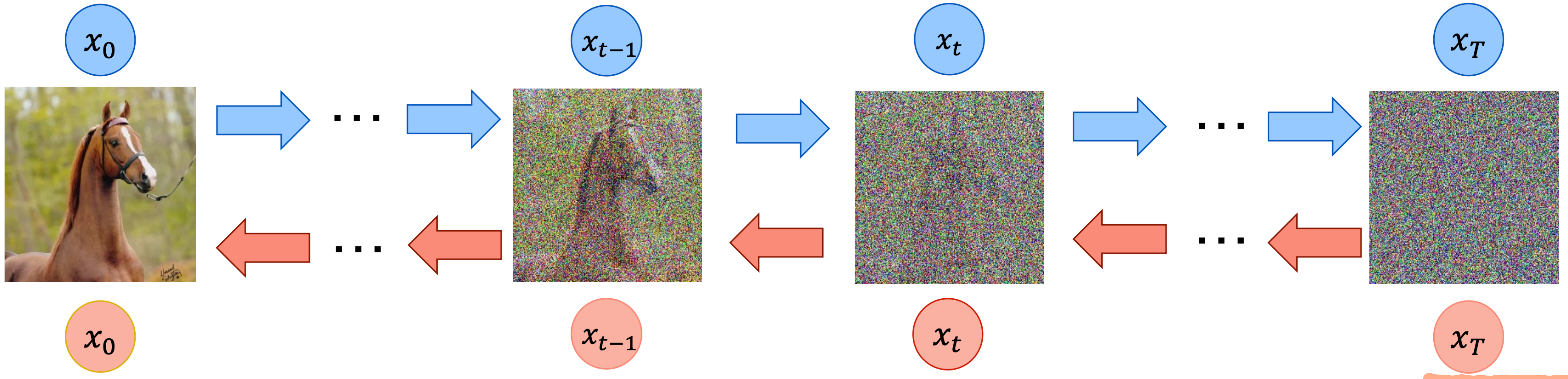
Example:

$\mathbf{x}_0$ built from one pixel



Continuous Diffusion Process

Discrete -> Continuous Forward Diffusion



$$\mathbf{x}_t = \sqrt{1-\beta_t} \cdot \mathbf{x}_0 + \sqrt{\beta_t} \cdot \boldsymbol{\epsilon}_t$$

$$\mathbf{x}_t = \sqrt{1-d\beta} \cdot \mathbf{x}_{t-1} + \sqrt{d\beta} \cdot \boldsymbol{\epsilon}_t$$

Forward diffusion, 1 step

$$\mathbf{x}_t = \sqrt{1-d\beta} \cdot \mathbf{x}_{t-1} + \sqrt{d\beta} \cdot \boldsymbol{\epsilon}_t$$

Let us try to rewrite it in the form

$\mathbf{x}_t - \mathbf{x}_{t-1} = f(\,\cdot\,)$

Receipt to cook $\mathbf{x}_t$:

- Take $\mathbf{x}_{t-1}$, make space for noise $\sqrt{1-d\beta}\ \mathbf{x}_{t-1}$
- Add a little noise $\sqrt{d\beta}\ \boldsymbol{\epsilon}_{t-1}$
- Enjoy your spoiled $\mathbf{x}_t$

Actually!

Here there is a mistake!

To fix it, perform the replacement:

$$d\beta \leftarrow \frac{d\beta}{1-\beta}$$

Nevertheless, for every fixed β_t, this term is infinitesimally small, so we can just continue keeping in mind that in the end we will need to replace $d\beta$

How to derive this fix:

$$\bar{\alpha}_t = \alpha_1 \alpha_2 \ldots \alpha_t$$

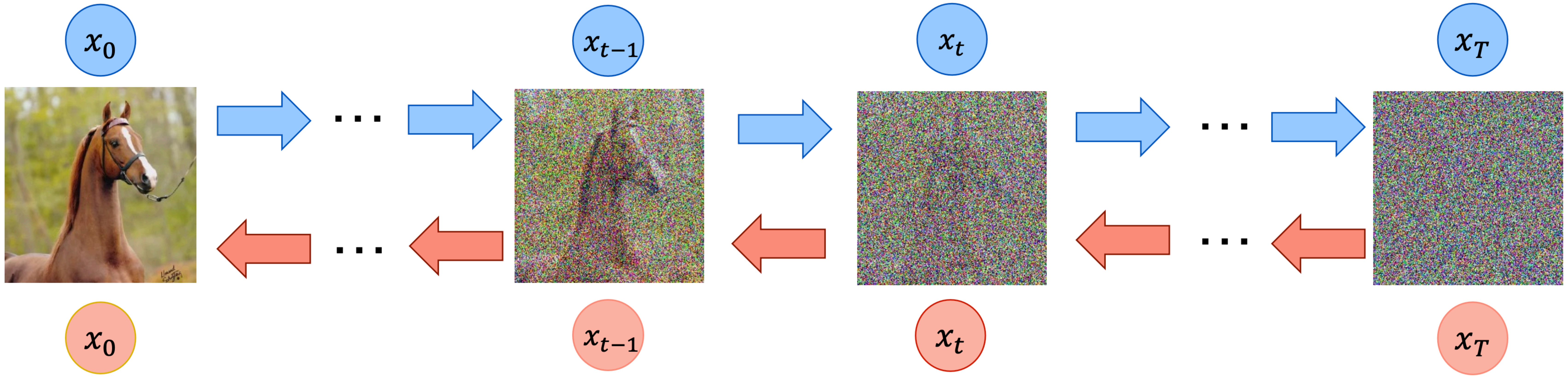
$$\bar{\beta}_t = 1 - \alpha_1 \alpha_2 \ldots \alpha_t$$

$$d\bar{\beta}_t = \bar{\beta}_t - \bar{\beta}_{t-1} =$$
$$= (1-\alpha_t)\alpha_1 \ldots \alpha_{t-1} =$$
$$= \beta_t (1 - \bar{\beta}_{t-1})$$

$$\beta_t = \frac{d\bar{\beta}_t}{1-\bar{\beta}_{t-1}}$$

Diffusion step: $\mathbf{x}_t = \sqrt{1-\beta_t}\mathbf{x}_{t-1} + \sqrt{\beta_t}\boldsymbol{\epsilon}_t$

Discrete -> Continuous Forward Diffusion



$$\mathbf{x}_t = \sqrt{1-\beta_t} \cdot \mathbf{x}_0 + \sqrt{\beta_t} \cdot \boldsymbol{\epsilon}_t$$

$$\mathbf{x}_t = \sqrt{1-d\beta} \cdot \mathbf{x}_{t-1} + \sqrt{d\beta} \cdot \boldsymbol{\epsilon}_t$$

Forward diffusion, 1 step

$$\mathbf{x}_t = \sqrt{1-d\beta} \cdot \mathbf{x}_{t-1} + \sqrt{d\beta} \cdot \boldsymbol{\epsilon}_t$$

Let us try to rewrite it in the form

$$\mathbf{x}_t - \mathbf{x}_{t-1} = f(\,\cdot\,)$$

Receipt to cook $\mathbf{x}_t$:

- Take $\mathbf{x}_{t-1}$, make space for noise $\sqrt{1-d\beta}\ \mathbf{x}_{t-1}$
- Add a little noise $\sqrt{d\beta}\ \boldsymbol{\epsilon}_{t-1}$
- Enjoy your spoiled $\mathbf{x}_t$

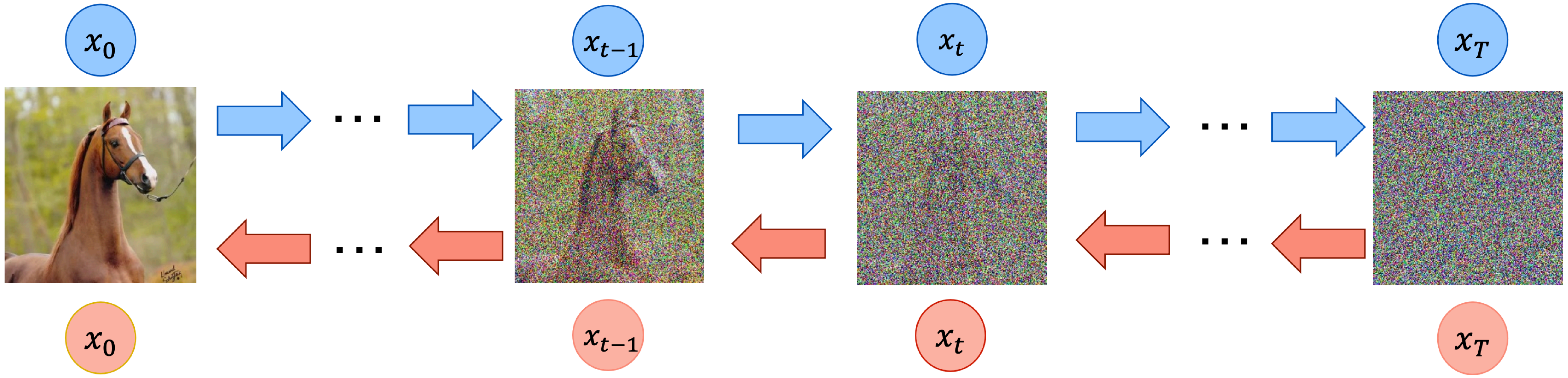
Assuming that $d\beta$ is really small

$$\sqrt{1-\beta} = 1 + \frac{\partial\sqrt{1-\beta}}{\partial\beta}\bigg|_{\beta=0} d\beta + o(d\beta) =$$

$$= 1 - \frac{1}{2}d\beta + o(d\beta)$$

(McLaurin formula)

Discrete -> Continuous Forward Diffusion



$$\mathbf{x}_t = \sqrt{1-\beta_t} \cdot \mathbf{x}_0 + \sqrt{\beta_t} \cdot \boldsymbol{\epsilon}_t$$

$$\mathbf{x}_t = \sqrt{1-d\beta} \cdot \mathbf{x}_{t-1} + \sqrt{d\beta} \cdot \boldsymbol{\epsilon}_t$$

Forward diffusion, 1 step

$$\mathbf{x}_t = \sqrt{1-d\beta} \cdot \mathbf{x}_{t-1} + \sqrt{d\beta} \cdot \boldsymbol{\epsilon}_t$$

Let us try to rewrite it in the form

$\mathbf{x}_t - \mathbf{x}_{t-1} = f(\,\cdot\,)$

Receipt to cook $\mathbf{x}_t$:

- Take $\mathbf{x}_{t-1}$, make space for noise $\sqrt{1-d\beta}\ \mathbf{x}_{t-1}$
- Add a little noise $\sqrt{d\beta}\ \boldsymbol{\epsilon}_{t-1}$
- Enjoy your spoiled $\mathbf{x}_t$

Assuming that $d\beta$ is really small

$$\sqrt{1-\beta} = 1 + \frac{\partial\sqrt{1-\beta}}{\partial\beta}\bigg|_{\beta=0} d\beta + o(d\beta) =$$

$$= 1 - \frac{1}{2}d\beta + o(d\beta)$$

(McLaurin formula)

$$\mathbf{x}_t = \left(1 - \frac{1}{2}d\beta\right)\mathbf{x}_{t-1} + \boldsymbol{\epsilon}_\beta\sqrt{d\beta}$$

$$\mathbf{x}_t - \mathbf{x}_{t-1} = -\frac{\mathbf{x}_{t-1}}{2}d\beta + \boldsymbol{\epsilon}_\beta\sqrt{d\beta}$$

$$d\mathbf{x} = -\frac{\mathbf{x}}{2}d\beta + \boldsymbol{\epsilon}_\beta\sqrt{d\beta}$$

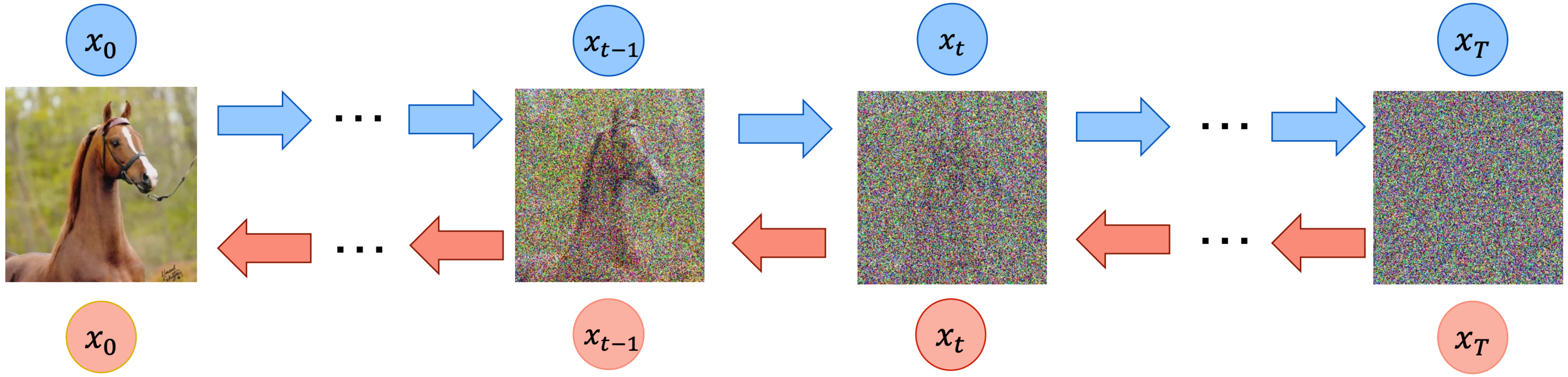
Differential equation :))))

But… $\boldsymbol{\epsilon}_\beta \propto \mathcal{N}(0,I)$:((((

General case:

$d\mathbf{x} = f(\mathbf{x},\beta)d\beta + g(\mathbf{x},\beta)\epsilon_\beta\sqrt{d\beta}$

Continuous Forward Diffusion Process



Discrete forward diffusion:

$$\mathbf{x}_t = \sqrt{1 - d\beta} \cdot \mathbf{x}_{t-1} + \sqrt{d\beta} \cdot \boldsymbol{\epsilon}_t$$

Continuous forward diffusion:

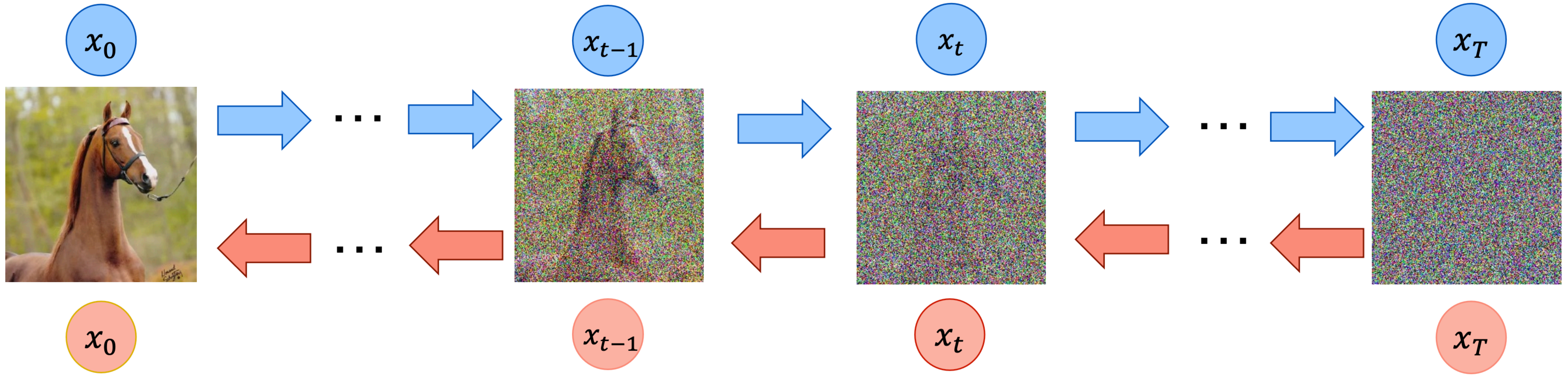
$$d\mathbf{x} = -\frac{\mathbf{x}}{2} d\beta + \boldsymbol{\epsilon}_\beta \sqrt{d\beta}$$

Deterministic part Wiener process

General case:

$$d\mathbf{x} = f(\mathbf{x}, \beta) d\beta + g(\mathbf{x}, \beta) \boldsymbol{\epsilon}_\beta \sqrt{d\beta}$$

Continuous Forward Diffusion Process



Discrete forward diffusion:

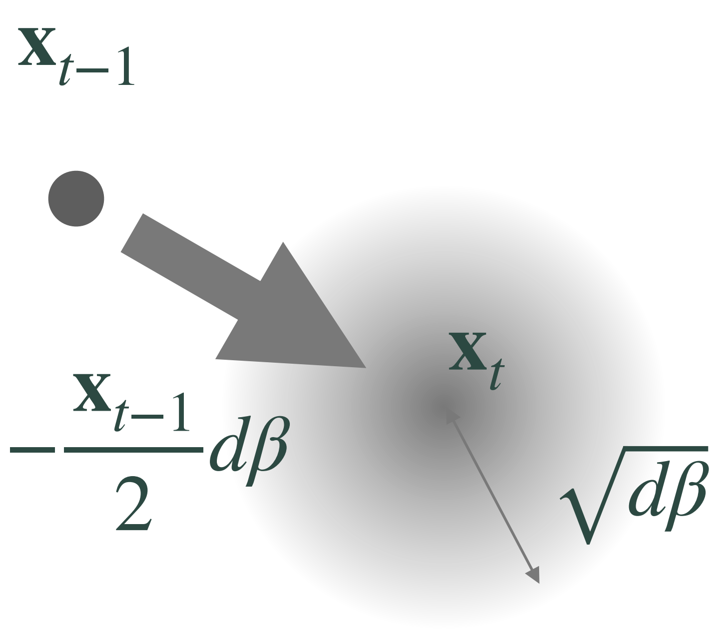
$$\mathbf{x}_t = \sqrt{1 - d\beta} \cdot \mathbf{x}_{t-1} + \sqrt{d\beta} \cdot \boldsymbol{\epsilon}_t$$

$$d\mathbf{x} = \underbrace{-\frac{\mathbf{x}}{2}d\beta}_{\mu(\mathbf{x},\beta)} + \epsilon_\beta \underbrace{\sqrt{d\beta}}_{\sigma(\mathbf{x},\beta)}$$

Continuous forward diffusion:

$$d\mathbf{x} = -\frac{\mathbf{x}}{2}d\beta + \boldsymbol{\epsilon}_\beta\sqrt{d\beta}$$

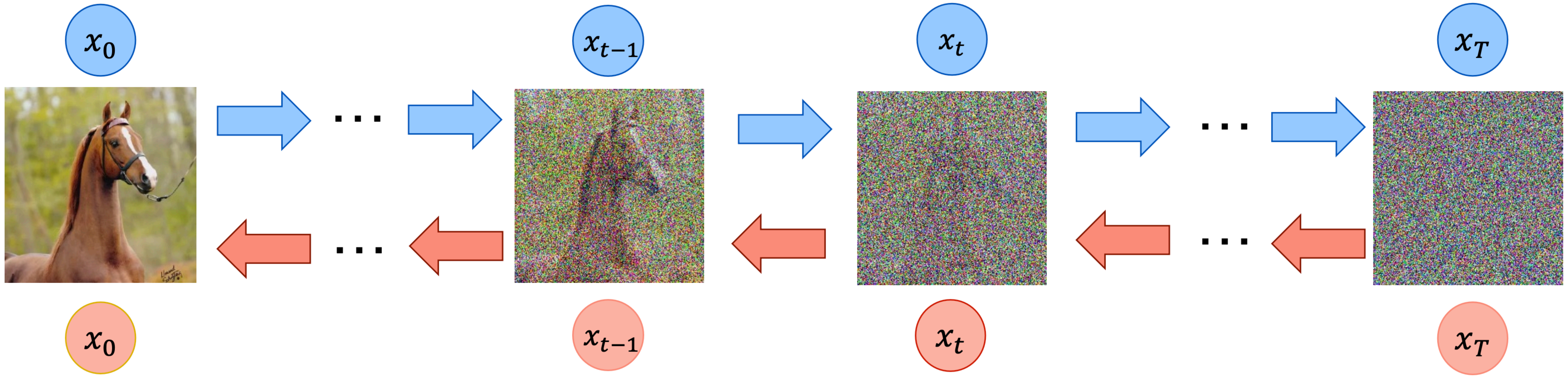
Deterministic part Wiener process



General case:

$$d\mathbf{x} = f(\mathbf{x}, \beta)d\beta + g(\mathbf{x}, \beta)\boldsymbol{\epsilon}_\beta\sqrt{d\beta}$$

Continuous Forward Diffusion Process



Discrete forward diffusion:

$$\mathbf{x}_t = \sqrt{1 - d\beta} \cdot \mathbf{x}_{t-1} + \sqrt{d\beta} \cdot \boldsymbol{\epsilon}_t$$

Continuous forward diffusion:

$$d\mathbf{x} = -\frac{\mathbf{x}}{2}d\beta + \boldsymbol{\epsilon}_\beta\sqrt{d\beta}$$

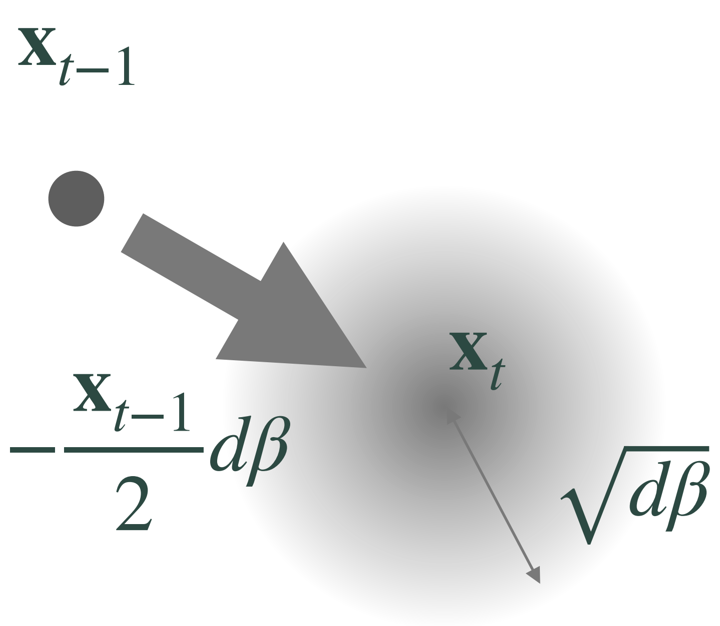
Deterministic part

Wiener process

General case:

$$d\mathbf{x} = f(\mathbf{x}, \beta)d\beta + g(\mathbf{x}, \beta)\boldsymbol{\epsilon}_\beta\sqrt{d\beta}$$

$$d\mathbf{x} = \underbrace{-\frac{\mathbf{x}}{2}d\beta}_{\mu(\mathbf{x},\beta)} + \epsilon_\beta \underbrace{\sqrt{d\beta}}_{\sigma(\mathbf{x},\beta)}$$



Cannot work with material point dynamics :'(

Move to distribution dynamics!

At moment β we have image distribution $p_\beta(\mathbf{x})$

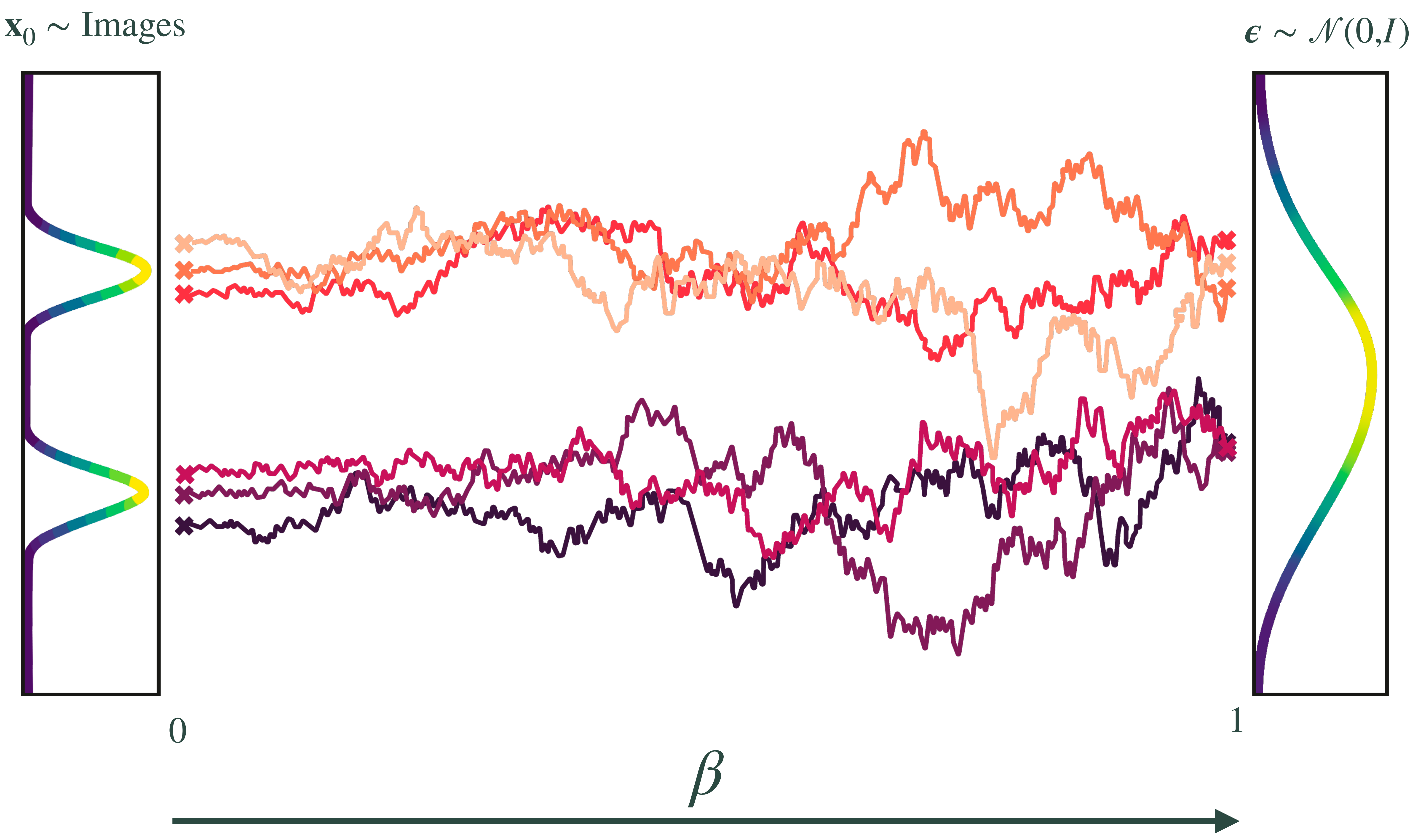
Dynamics consists of:
1. Small mean displacement
2. Small diffusion

The process that we have:

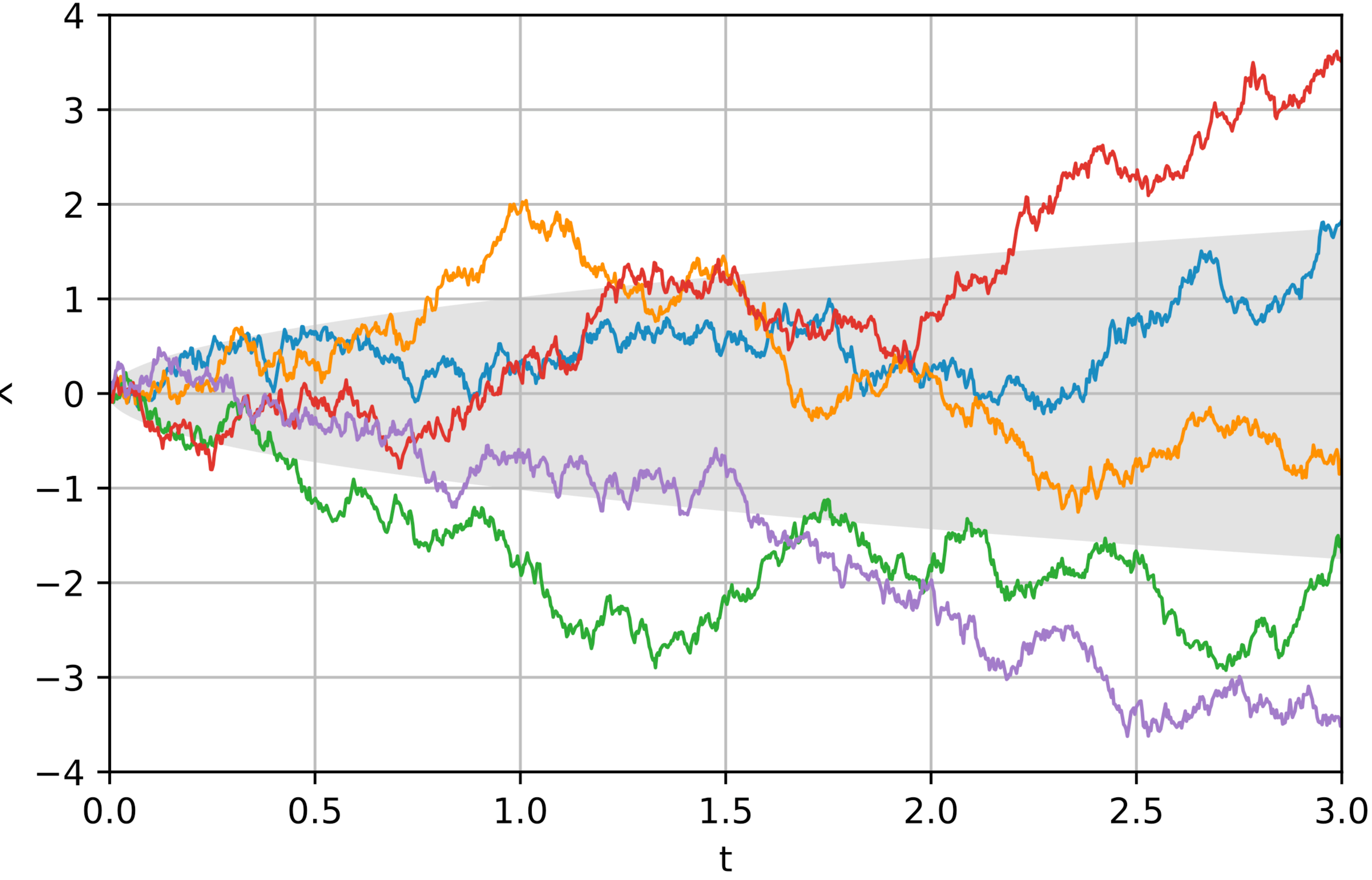
$$d\mathbf{x} = -\frac{\mathbf{x}}{2}d\beta + \boldsymbol{\epsilon}_\beta\sqrt{d\beta}$$

1D Example:

$\mathbf{x}_0$ is a pixel



Wiener's Process (Brownian Motion)



Point dynamics

Every moment the coordinate changes:

$dx = \epsilon_t \sqrt{dt}, \; \epsilon_t \sim \mathcal{N}(0,1)$

$\mu = 0$
$\sigma = \sqrt{dt}$

Point drift for time T:

$$\Delta x = \sum_{t=0}^{T} \epsilon_t \sqrt{dt}$$

$$\mathbb{E} \sum_{k=0}^{K} \xi_k = \sum_{k=0}^{K} \mathbb{E} \xi_k$$

$$\sigma^2 = \mathbb{V} \sum_{k=0}^{K} \xi_k = \sum_{k=0}^{K} \mathbb{V} \xi_k = \sum_{k=0}^{K} \sigma_k^2$$

$$\Delta x = \sqrt{\int_0^T \sqrt{dt}^2} = \sqrt{T}$$

Probabilistic Diffusion Process

Because we can!

What we will need

Fourier Transform:
(Same as Basis Expansion)

$$\phi_f(\lambda) = \left\langle e^{-i\lambda x}, f(x) \right\rangle = \int_{\mathbb{R}} e^{i\lambda x} f(x) dx$$

$$f(x) = \sum_{\lambda} \phi(\lambda) e^{-i\lambda x} = \frac{1}{2\pi} \int_{\mathbb{R}} \phi(\lambda) e^{-i\lambda x} d\lambda$$

Delta Function:

$$\phi_\delta(\lambda) = \int_{\mathbb{R}} \delta(x) e^{i\lambda x} dx = e^{i\lambda 0} = 1$$

$$\delta(x) = \frac{1}{2\pi} \int_{\mathbb{R}} 1 e^{-i\lambda x} d\lambda = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-i\lambda x} d\lambda$$

What we will need

Fourier Transform:
(Same as Basis Expansion)

$$\phi_f(\lambda) = \left\langle e^{-i\lambda x}, f(x) \right\rangle = \int_{\mathbb{R}} e^{i\lambda x} f(x) dx$$

$$f(x) = \sum_{\lambda} \phi(\lambda) e^{-i\lambda x} = \frac{1}{2\pi} \int_{\mathbb{R}} \phi(\lambda) e^{-i\lambda x} d\lambda$$

Delta Function:

$$\phi_\delta(\lambda) = \int_{\mathbb{R}} \delta(x) e^{i\lambda x} dx = e^{i\lambda 0} = 1$$

$$\delta(x) = \frac{1}{2\pi} \int_{\mathbb{R}} 1 e^{-i\lambda x} d\lambda = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-i\lambda x} d\lambda$$

Characteristic Function:

$$\phi(\lambda) = \int_{\mathbb{R}} e^{i\lambda x} p(x) dx = \mathbb{E}_p e^{i\lambda \xi}$$

$$\frac{\partial}{\partial \lambda} \phi(\lambda) \Big|_{\lambda=0} = \frac{\partial}{\partial \lambda} \int_{\mathbb{R}} e^{i\lambda x} p(x) dx \Big|_{\lambda=0} =$$

$$= \int_{\mathbb{R}} ixe^{i\lambda x} p(x) dx \Big|_{\lambda=0} =$$

$$= i \int_{\mathbb{R}} xp(x) dx = i\mathbb{E}\xi$$

$$\frac{\partial^k}{\partial \lambda^k} \phi(\lambda) \Big|_{\lambda=0} = i^k \mathbb{E}\xi^k$$

What we will need

Fourier Transform:
(Same as Basis Expansion)

$$\phi_f(\lambda) = \left\langle e^{-i\lambda x}, f(x) \right\rangle = \int_{\mathbb{R}} e^{i\lambda x} f(x) dx$$

$$f(x) = \sum_{\lambda} \phi(\lambda) e^{-i\lambda x} = \frac{1}{2\pi} \int_{\mathbb{R}} \phi(\lambda) e^{-i\lambda x} d\lambda$$

Delta Function:

$$\phi_\delta(\lambda) = \int_{\mathbb{R}} \delta(x) e^{i\lambda x} dx = e^{i\lambda 0} = 1$$

$$\delta(x) = \frac{1}{2\pi} \int_{\mathbb{R}} 1 e^{-i\lambda x} d\lambda = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-i\lambda x} d\lambda$$

Characteristic Function:

$$\phi(\lambda) = \int_{\mathbb{R}} e^{i\lambda x} p(x) dx = \mathbb{E}_p e^{i\lambda \xi}$$

$$\frac{\partial}{\partial \lambda} \phi(\lambda) \Big|_{\lambda=0} = \frac{\partial}{\partial \lambda} \int_{\mathbb{R}} e^{i\lambda x} p(x) dx \Big|_{\lambda=0} =$$

$$= \int_{\mathbb{R}} i x e^{i\lambda x} p(x) dx \Big|_{\lambda=0} =$$

$$= i \int_{\mathbb{R}} x p(x) dx = i \mathbb{E} \xi$$

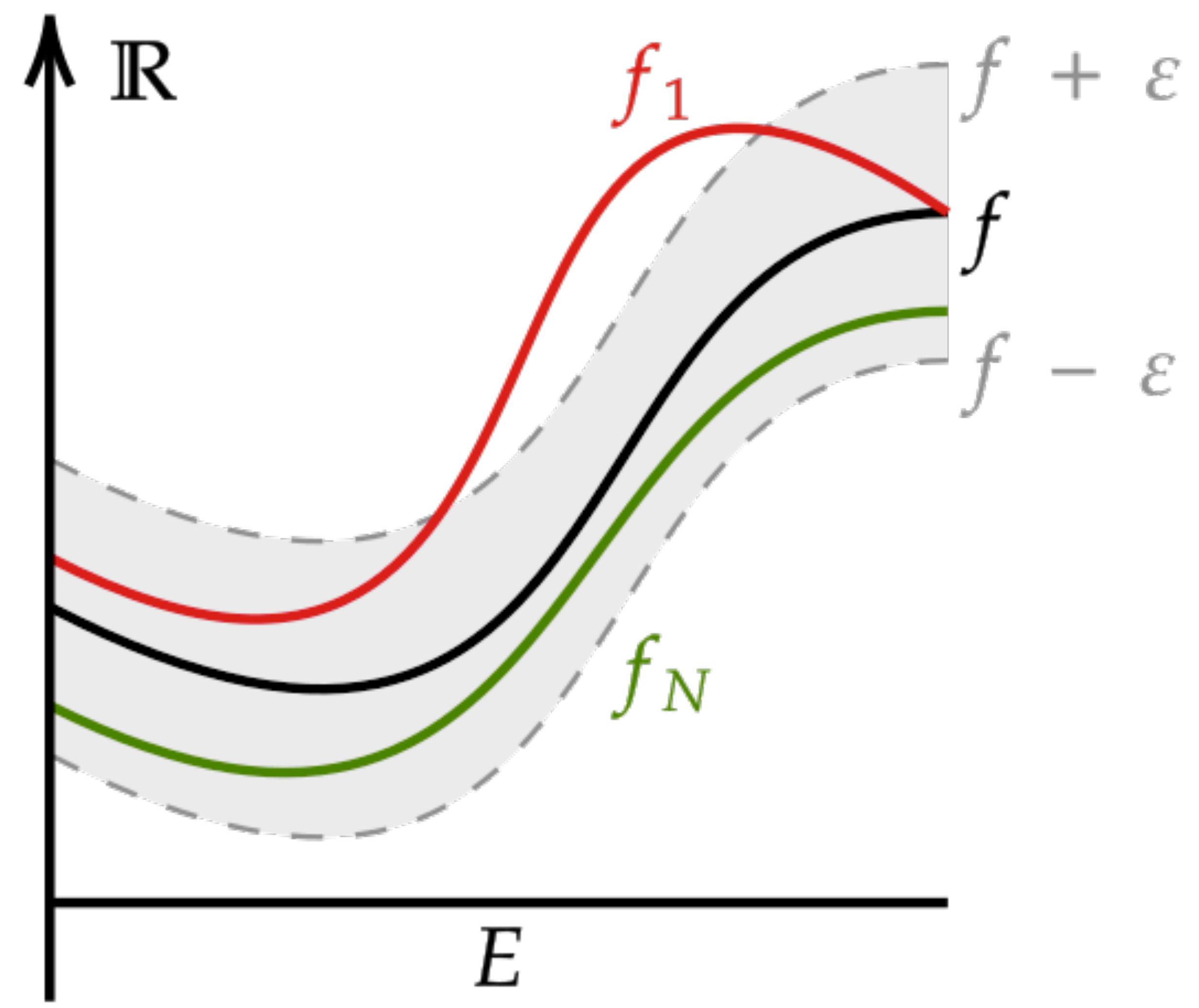
$$\frac{\partial^k}{\partial \lambda^k} \phi(\lambda) \Big|_{\lambda=0} = i^k \mathbb{E} \xi^k$$

Taylor's Series:

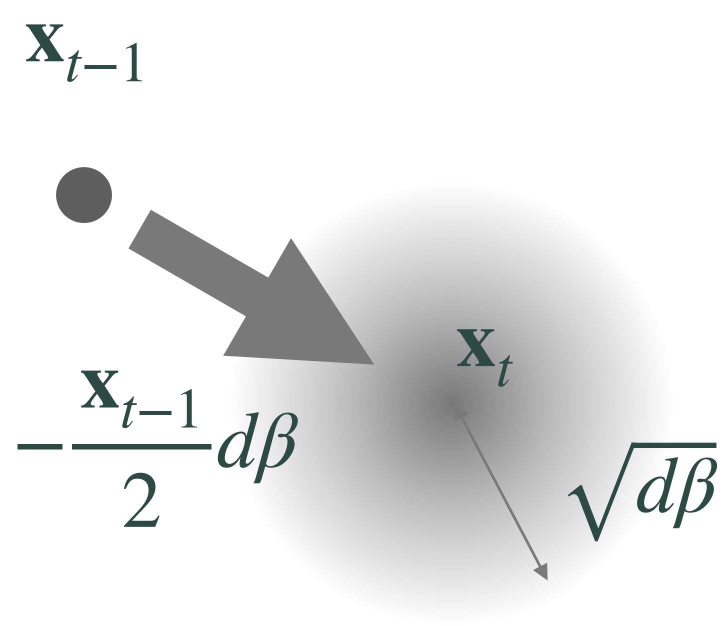
$$f(x) = \sum_{k=0}^{\infty} \frac{1}{k!} \frac{\partial^k f}{\partial x^k} (x - x_0)^k$$

Uniform Convergence

$$\forall \varepsilon > 0\ \exists N :\ \forall n > N\ \forall x \in [a, b]\ \ |f(x) - f_n(x)| < \varepsilon$$



Probabilistic Diffusion



We have $p(\mathbf{x}, \beta)$

Dynamics consists of:

1. Small mean displacement

   (by $\mu(\mathbf{x}, \beta) = f(\mathbf{x}, \beta)d\beta$)

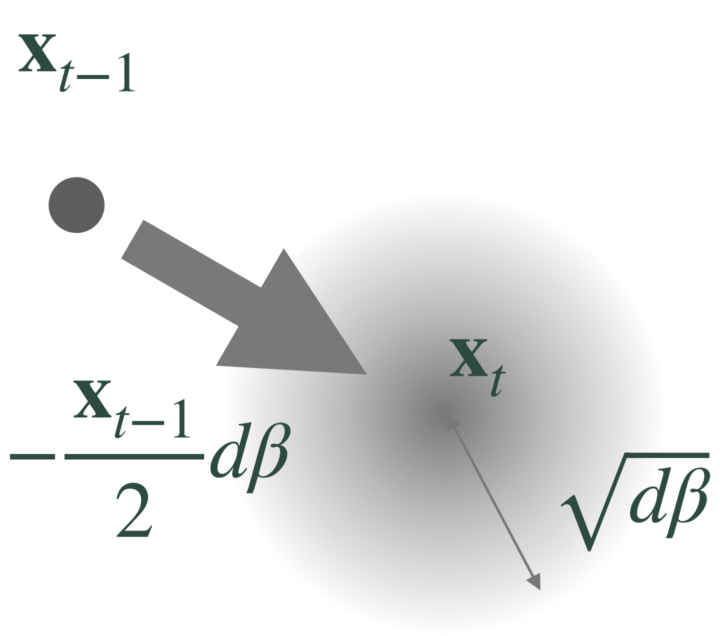
2. Small diffusion
   (by $\sigma(\mathbf{x}, \beta) = g(\mathbf{x}, \beta)\sqrt{d\beta}$)

Transition :

$p(x, \beta + d\beta) =$

$$= \int_{\mathbb{R}} p_{diff}(x, \beta + d\beta \,|\, x', \beta) p(x', \beta) dx'$$

Probabilistic Diffusion



We have $p(\mathbf{x}, \beta)$

Dynamics consists of:

1. Small mean displacement

   (by $\mu(\mathbf{x}, \beta) = f(\mathbf{x}, \beta)d\beta$)

2. Small diffusion
   (by $\sigma(\mathbf{x}, \beta) = g(\mathbf{x}, \beta)\sqrt{d\beta}$)

Transition :
$p(x, \beta + d\beta) =$

$$= \int_{\mathbb{R}} p_{diff}(x, \beta + d\beta \,|\, x', \beta) p(x', \beta) dx'$$

Transition: $p(x, \beta + d\beta) = \int_{\mathbb{R}} p_{diff}(x, \beta + d\beta \,|\, x', \beta) p(x', \beta) dx'$

$\phi(\lambda)$ — characteristic function (Fourier Transform)

$$\phi_{diff}(\lambda) = \mathbb{E} e^{i\lambda(x - x')} = \int_{\mathbb{R}} e^{i\lambda(x-x')} p_{diff}(x, \beta + d\beta \,|\, x', \beta) dx$$

$$p_{diff}(x, \beta + d\beta \,|\, x', \beta) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-i\lambda(x-x')} \phi_{diff}(\lambda, \beta + d\beta \,|\, x', \beta) d\lambda$$

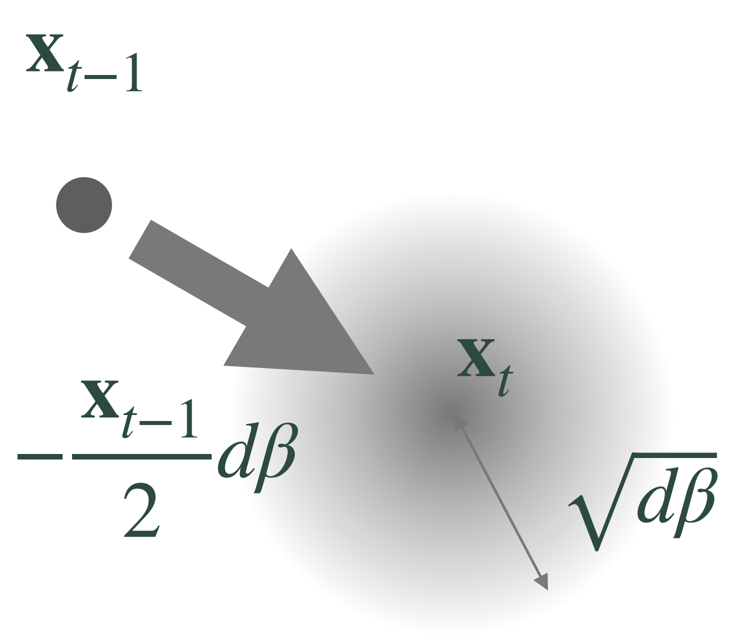
$$\phi_{diff}(\lambda) = \sum_{k=0}^{\infty} \frac{1}{k!} \frac{\partial^k \phi}{\partial \lambda^k} \lambda^k = \sum_{k=0}^{\infty} \frac{\mathbb{E}\xi^k}{k!} (i)^k \lambda^k$$

$$\mathbb{E}\xi^k = \begin{cases} \mu(x', \beta) & \text{if } k = 1 \\ \mu^2(x', \beta) + \sigma^2(x', \beta) & \text{if } k = 2 \\ O(\mu^3, \sigma^3) & \text{if } k = 3 \\ \dots & \dots \end{cases}$$

$$p_{t+1} = \int_{\mathbb{R}} \frac{1}{2\pi} \int_{\mathbb{R}} e^{-i\lambda(x-x')} \sum_{k=0}^{\infty} \frac{(i\lambda)^k \mathbb{E}\xi^k}{k!} p(x', \beta) d\lambda dx' =$$

$$= \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \frac{\partial^k}{\partial x^k} \left[\int_{\mathbb{R}} \underbrace{\left[\int_{\mathbb{R}} \frac{1}{2\pi} e^{-i\lambda(x-x')} d\lambda \right]}_{\delta(x-x')} \mathbb{E}\xi^k p(x', \beta) dx' \right] =$$

Probabilistic Diffusion



We have $p(\mathbf{x}, \beta)$

Dynamics consists of:

1. Small mean displacement

   (by $\mu(\mathbf{x}, \beta) = f(\mathbf{x}, \beta)d\beta$)

2. Small diffusion
   (by $\sigma(\mathbf{x}, \beta) = g(\mathbf{x}, \beta)\sqrt{d\beta}$)

Transition :

$p(x, \beta + d\beta) =$

$$= \int_{\mathbb{R}} p_{diff}(x, \beta + d\beta \,|\, x', \beta) p(x', \beta) dx'$$

<u>Transition</u>: $p(x, \beta + d\beta) = \int_{\mathbb{R}} p_{diff}(x, \beta + d\beta \,|\, x', \beta) p(x', \beta) dx'$

$\phi(\lambda)$ — characteristic function (Fourier Transform)

$$\phi_{diff}(\lambda) = \mathbb{E}e^{i\lambda(x-x')} = \int_{\mathbb{R}} e^{i\lambda(x-x')} p_{diff}(x, \beta + d\beta \,|\, x', \beta) dx$$

$$p_{diff}(x, \beta + d\beta \,|\, x', \beta) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-i\lambda(x-x')} \phi_{diff}(\lambda, \beta + d\beta \,|\, x', \beta) d\lambda$$

$$\phi_{diff}(\lambda) = \sum_{k=0}^{\infty} \frac{1}{k!} \frac{\partial^k \phi}{\partial \lambda^k} \lambda^k = \sum_{k=0}^{\infty} \frac{\mathbb{E}\xi^k}{k!} (i)^k \lambda^k$$

$$\mathbb{E}\xi^k = \begin{cases} \mu(x', \beta) & \text{if } k = 1 \\ \mu^2(x', \beta) + \sigma^2(x', \beta) & \text{if } k = 2 \\ O(\mu^3, \sigma^3) & \text{if } k = 3 \\ \dots & \dots \end{cases}$$

$$p_{t+1} = \int_{\mathbb{R}} \frac{1}{2\pi} \int_{\mathbb{R}} e^{-i\lambda(x-x')} \sum_{k=0}^{\infty} \frac{(i\lambda)^k \mathbb{E}\xi^k}{k!} p(x', \beta) d\lambda dx' =$$

$$= \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \frac{\partial^k}{\partial x^k} \left[\int_{\mathbb{R}} \underbrace{\left[\int_{\mathbb{R}} \frac{1}{2\pi} e^{-i\lambda(x-x')} d\lambda \right]}_{\delta(x-x')} \mathbb{E}\xi^k p(x', \beta) dx' \right] =$$

$$= \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \frac{\partial^k}{\partial x^k} \mathbb{E}\xi^k p(x, \beta) =$$

$$= p(x, \beta) - \frac{\partial}{\partial x} \mu(x, \beta) p(x, \beta) +$$

$$+ \frac{1}{2} \frac{\partial^2}{\partial x^2} \left(\mu^2(x, \beta) + \sigma^2(x, \beta) \right) p(x, \beta) + o(\mu^2, \sigma^2) =$$

$$= p(x, \beta) - d\beta \frac{\partial}{\partial x} f(x, \beta) p(x, \beta) +$$

$$+ d\beta \frac{1}{2} \frac{\partial^2}{\partial x^2} g^2(x, \beta) p(x, \beta) + o(d\beta)$$

<u>**Finally:**</u>

$$\frac{dp}{d\beta} = -\frac{\partial}{\partial x} f(x, \beta) p(x, \beta) + \frac{1}{2} \frac{\partial^2}{\partial x^2} g^2(x, \beta) p(x, \beta)$$

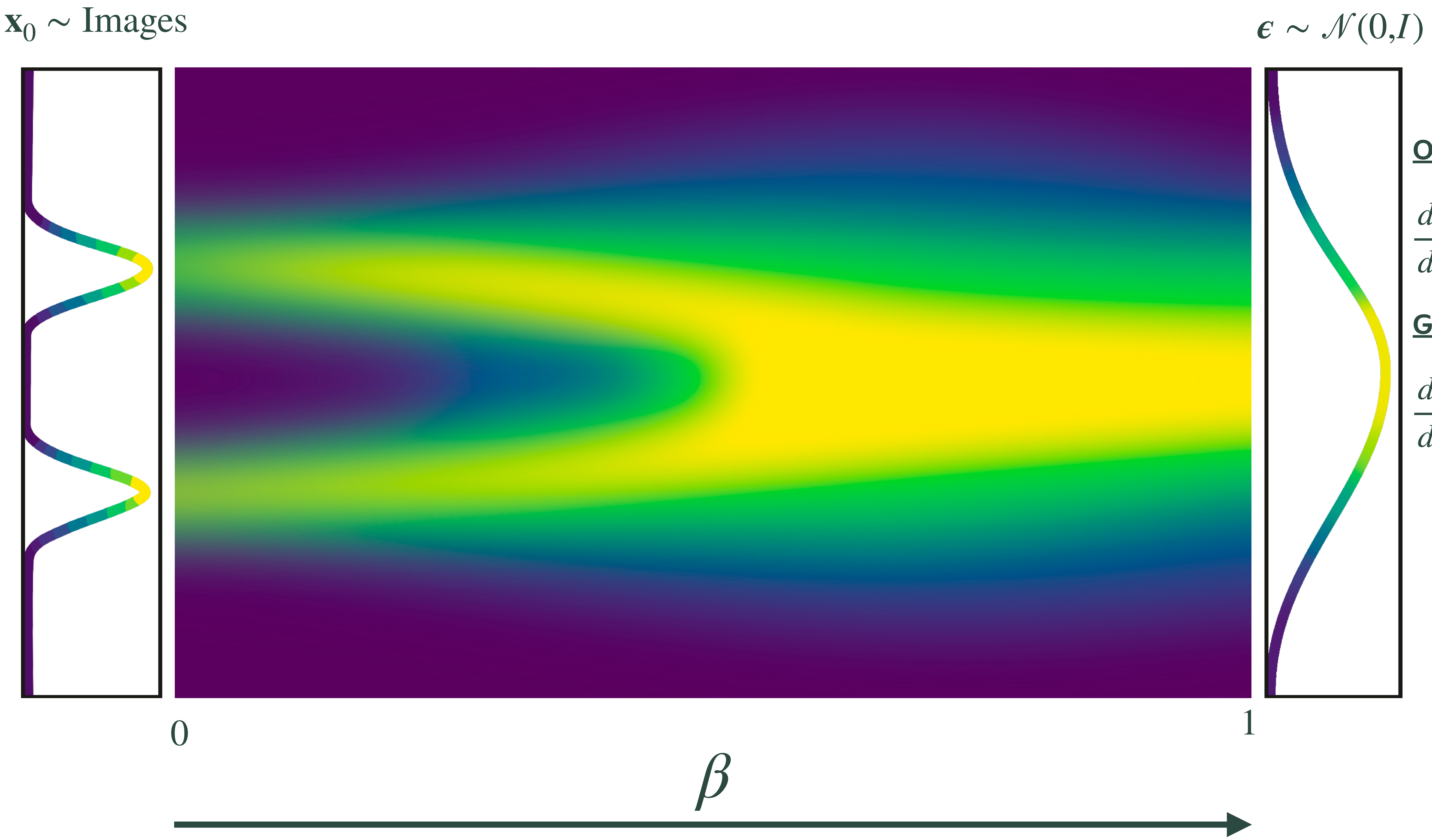
<u>**In our case:**</u>

$$\frac{dp}{d\beta} = \frac{\partial}{\partial x} \frac{x}{2} p(x, \beta) + \frac{1}{2} \frac{\partial^2}{\partial x^2} p(x, \beta)$$

The process that we have:

1D Example:

$\mathbf{x}_0$ is a pixel



Our Case:

$$\frac{dp}{d\beta} = \frac{\partial}{\partial x}\frac{x}{2}p + \frac{1}{2}\frac{\partial^2}{\partial x^2}p$$

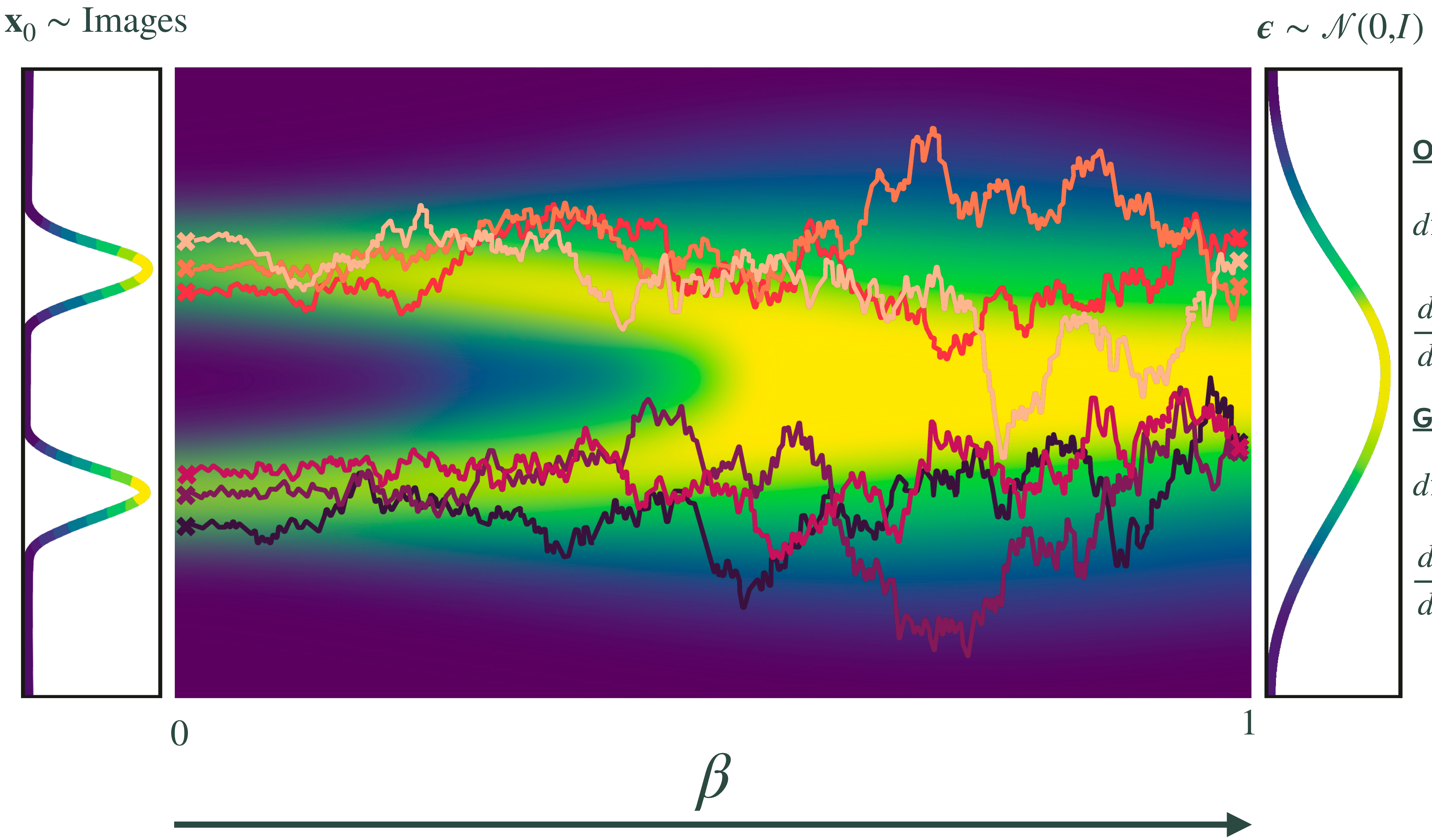
General Case:

$$\frac{dp}{d\beta} = -\frac{\partial}{\partial x}f(x, \beta)p + \frac{1}{2}\frac{\partial^2}{\partial x^2}g^2(x, \beta)p$$

The process that we have:

1D Example:

$\mathbf{x}_0$ is a pixel



Our Case:

$$d\mathbf{x} = -\frac{\mathbf{x}}{2} d\beta + \boldsymbol{\epsilon}_\beta \sqrt{d\beta}$$

$$\frac{dp}{d\beta} = \frac{\partial}{\partial x}\frac{x}{2}p + \frac{1}{2}\frac{\partial^2}{\partial x^2}p$$

General Case:

$$d\mathbf{x} = f(\mathbf{x}, \beta)d\beta + \mathbf{g}(\mathbf{x}, \beta)\epsilon_\beta \sqrt{d\beta}$$

$$\frac{dp}{d\beta} = -\frac{\partial}{\partial x}f(x, \beta)p + \frac{1}{2}\frac{\partial^2}{\partial x^2}g^2(x, \beta)p$$

Inverting Time

Now we can do that!

Inverting Time

Discrete forward diffusion:

$$\mathbf{x}_t = \sqrt{1 - d\beta} \cdot \mathbf{x}_{t-1} + \sqrt{d\beta} \cdot \epsilon_t$$

Continuous forward diffusion:

$$d\mathbf{x} = \underbrace{-\frac{\mathbf{x}}{2}d\beta}_{\mu(\mathbf{x},\beta)} + \epsilon_\beta \underbrace{\sqrt{d\beta}}_{\sigma(\mathbf{x},\beta)}$$

Deterministic part

Stochastic part

Or: $dx = f(x, \beta)d\beta + g(x, \beta)dW_\beta$

Our case: $dx = -\frac{x}{2}d\beta + dW_\beta$

Probabilistic forward diffusion (1D):

$$\frac{dp(x, \beta)}{d\beta} = \frac{\partial}{\partial x}\frac{x}{2}p(x, \beta) + \frac{\partial^2}{\partial x^2}\frac{1}{2}p(x, \beta)$$

Inverting Time

Discrete forward diffusion:

$$\mathbf{x}_t = \sqrt{1 - d\beta} \cdot \mathbf{x}_{t-1} + \sqrt{d\beta} \cdot \epsilon_t$$

Continuous forward diffusion:

$$d\mathbf{x} = \underbrace{-\frac{\mathbf{x}}{2}d\beta}_{\mu(\mathbf{x},\beta)} + \epsilon_\beta \underbrace{\sqrt{d\beta}}_{\sigma(\mathbf{x},\beta)}$$

Deterministic part

Stochastic part

Or: $dx = f(x,\beta)d\beta + g(x,\beta)dW_\beta$

Our case: $dx = -\frac{x}{2}d\beta + dW_\beta$

Probabilistic forward diffusion (1D):

$$\frac{dp(x,\beta)}{d\beta} = \frac{\partial}{\partial x}\frac{x}{2}p(x,\beta) + \frac{\partial^2}{\partial x^2}\frac{1}{2}p(x,\beta)$$

Plan:

1. We invert the time flow $\tilde{\beta} = 1 - \beta$
2. We try to build the SDE of the same form

$$\frac{dp(x)}{d(1-\tilde{\beta})} = \frac{\partial}{\partial x}\frac{x}{2}p(x,1-\tilde{\beta}) + \frac{\partial^2}{\partial x^2}\frac{1}{2}p(x,1-\tilde{\beta})$$

$$\frac{dp(x)}{d\tilde{\beta}} = -\frac{\partial}{\partial x}\frac{x}{2}p(x,\tilde{\beta}) - \frac{\partial^2}{\partial x^2}\frac{1}{2}p(x,\tilde{\beta})$$

$$\frac{dp(x)}{d\beta} = -\frac{\partial}{\partial x}\frac{x}{2}p(x,\beta) - \frac{\partial^2}{\partial x^2}p(x,\beta) + \frac{\partial^2}{\partial x^2}\frac{1}{2}p(x,\beta)$$

$$\frac{dp(x)}{d\beta} = \frac{\partial}{\partial x}\left[-\frac{x}{2}p(x,\beta) - \frac{\partial}{\partial x}p(x,\beta)\right] + \frac{\partial^2}{\partial x^2}\frac{1}{2}p(x,\beta)$$

$$\frac{dp(x)}{d\beta} = \frac{\partial}{\partial x}\left[-\frac{x}{2}p(x,\beta) - \frac{p(x,\beta)}{p(x,\beta)}\frac{\partial}{\partial x}p(x,\beta)\right] + \frac{\partial^2}{\partial x^2}\frac{1}{2}p(x,\beta)$$

$$\frac{dp(x)}{d\beta} = \frac{\partial}{\partial x}\left[-\frac{x}{2}p(x,\beta) - p(x,\beta)\frac{\partial}{\partial x}\ln p(x,\beta)\right] + \frac{\partial^2}{\partial x^2}\frac{1}{2}p(x,\beta)$$

Inverting Time

Discrete forward diffusion:

$$\mathbf{x}_t = \sqrt{1 - d\beta} \cdot \mathbf{x}_{t-1} + \sqrt{d\beta} \cdot \epsilon_t$$

Continuous forward diffusion:

$$d\mathbf{x} = \underbrace{-\frac{\mathbf{x}}{2} d\beta}_{\mu(\mathbf{x},\beta)} + \epsilon_\beta \underbrace{\sqrt{d\beta}}_{\sigma(\mathbf{x},\beta)}$$

Deterministic part

Stochastic part

Or: $dx = f(x, \beta)d\beta + g(x, \beta)dW_\beta$

Our case: $dx = -\dfrac{x}{2}d\beta + dW_\beta$

Probabilistic forward diffusion (1D):

$$\frac{dp(x, \beta)}{d\beta} = \frac{\partial}{\partial x}\frac{x}{2}p(x, \beta) + \frac{\partial^2}{\partial x^2}\frac{1}{2}p(x, \beta)$$

Plan:

1. We invert the time flow $\tilde{\beta} = 1 - \beta$
2. We try to build the SDE of the same form

$$\frac{dp(x)}{d(1 - \tilde{\beta})} = \frac{\partial}{\partial x}\frac{x}{2}p(x, 1 - \tilde{\beta}) + \frac{\partial^2}{\partial x^2}\frac{1}{2}p(x, 1 - \tilde{\beta})$$

$$\frac{dp(x)}{d\tilde{\beta}} = -\frac{\partial}{\partial x}\frac{x}{2}p(x, \tilde{\beta}) - \frac{\partial^2}{\partial x^2}\frac{1}{2}p(x, \tilde{\beta})$$

$$\frac{dp(x)}{d\beta} = -\frac{\partial}{\partial x}\frac{x}{2}p(x, \beta) - \frac{\partial^2}{\partial x^2}p(x, \beta) + \frac{\partial^2}{\partial x^2}\frac{1}{2}p(x, \beta)$$

$$\frac{dp(x)}{d\beta} = \frac{\partial}{\partial x}\left[-\frac{x}{2}p(x, \beta) - \frac{\partial}{\partial x}p(x, \beta)\right] + \frac{\partial^2}{\partial x^2}\frac{1}{2}p(x, \beta)$$

$$\frac{dp(x)}{d\beta} = \frac{\partial}{\partial x}\left[-\frac{x}{2}p(x, \beta) - \frac{p(x, \beta)}{p(x, \beta)}\frac{\partial}{\partial x}p(x, \beta)\right] + \frac{\partial^2}{\partial x^2}\frac{1}{2}p(x, \beta)$$

$$\frac{dp(x)}{d\beta} = \frac{\partial}{\partial x}\left[-\frac{x}{2}p(x, \beta) - p(x, \beta)\frac{\partial}{\partial x}\ln p(x, \beta)\right] + \frac{\partial^2}{\partial x^2}\frac{1}{2}p(x, \beta)$$

Resulting Probabilistic DE:

$$\frac{dp}{d\beta} = \frac{\partial}{\partial x}\left[-\frac{x}{2} - \frac{\partial}{\partial x}\ln p(x, \beta)\right]p(x, \beta) + \frac{\partial^2}{\partial x^2}\frac{1}{2}p(x, \beta)$$

Corresponding Reverse SDE:

$$dx = \left[-\frac{x}{2} - \frac{\partial}{\partial x}\ln p(x, \beta)\right]d\beta + \epsilon_\beta\sqrt{-d\beta}$$

One Special Case:

$$dx = \left[f(\beta)x - g^2(\beta)\frac{\partial}{\partial x}\ln p\right]d\beta + g(\beta)\epsilon_\beta\sqrt{-d\beta}$$

Annealed Langevin Dynamics

$$d\mathbf{x} = \left[\frac{\mathbf{x}}{2} + \frac{\partial}{\partial \mathbf{x}} \ln p(\mathbf{x}, \beta)\right](-d\beta) + \boldsymbol{\epsilon}_\beta \sqrt{-d\beta}$$

Pushes away from zero

Pushes towards higher score
(our model's output)

Injects noise

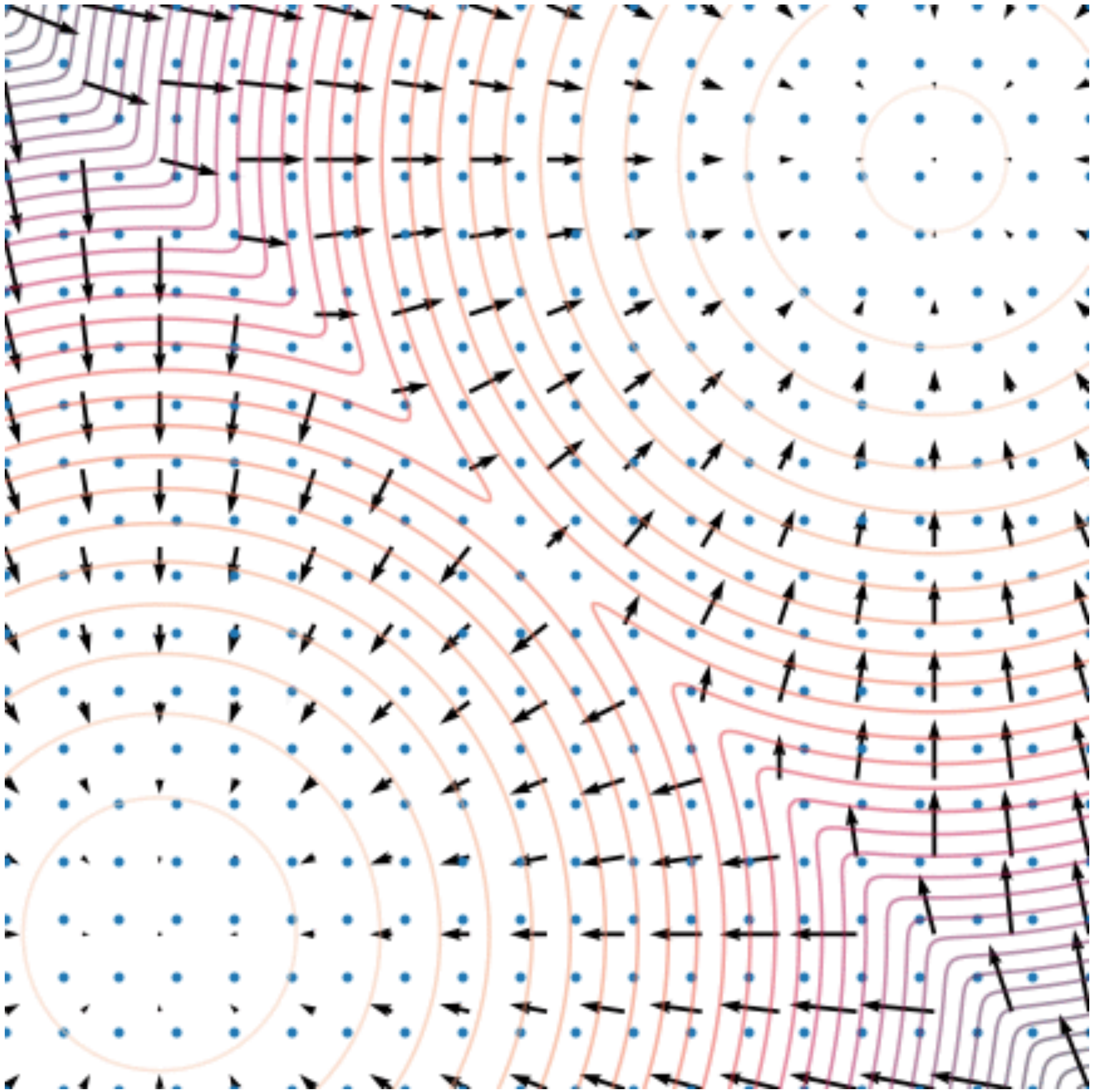
$$\ln p(\mathbf{x}, \beta) = \ln \mathcal{N}\left(\mathbf{x}_0\sqrt{1-\beta}, \beta\right) =$$

$$= \ln \frac{1}{Z} \exp\left(-\frac{(\mathbf{x}_t - \boldsymbol{\mu})^2}{2\sigma^2}\right) =$$

$$= -\frac{1}{\sigma^2}(\mathbf{x}_\beta - \boldsymbol{\mu}) + C = -\frac{1}{\beta}\underbrace{(\mathbf{x}_\beta - \sqrt{1-\beta}\mathbf{x}_0)}_{\sqrt{\beta}\boldsymbol{\epsilon}} + C =$$

$$= -\frac{1}{\sqrt{\beta}}\boldsymbol{\epsilon} + C = -\frac{1}{\sigma_\beta}\boldsymbol{\epsilon} + C$$

$$d\mathbf{x} = \left[\frac{\mathbf{x}}{2} - \frac{\boldsymbol{\epsilon}(\mathbf{x}, \beta)}{\sigma_\beta}\right](-d\beta) + \boldsymbol{\epsilon}_\beta \sqrt{-d\beta}$$

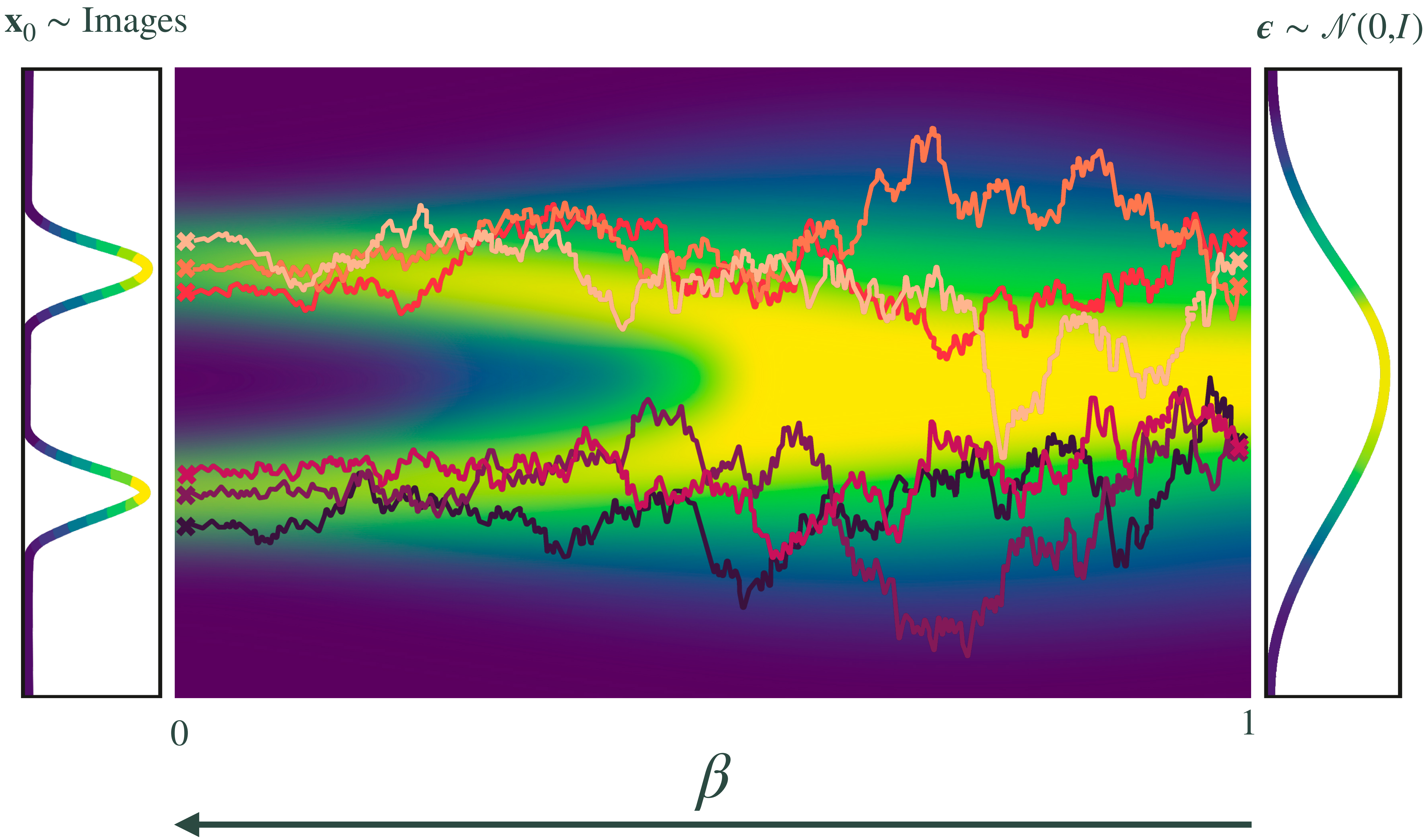


The process that we have:

$$d\mathbf{x} = \left[\frac{\mathbf{x}}{2} - \frac{\boldsymbol{\epsilon}(\mathbf{x},\beta)}{\sigma_\beta}\right](-d\beta) + \boldsymbol{\epsilon}_\beta\sqrt{-d\beta}$$

<u>**1D Example:**</u>

$\mathbf{x}_0$ is a pixel



No more Randomnicity!

Because we don't need it

SDE -> ODE 🤯

Discrete forward diffusion:

$$\mathbf{x}_t = \sqrt{1 - d\beta} \cdot \mathbf{x}_{t-1} + \sqrt{d\beta} \cdot \epsilon_t$$

Continuous forward diffusion:

$$d\mathbf{x} = \underbrace{-\frac{\mathbf{x}}{2} d\beta}_{\mu(\mathbf{x},\beta)} + \epsilon_\beta \underbrace{\sqrt{d\beta}}_{\sigma(\mathbf{x},\beta)}$$

Deterministic part

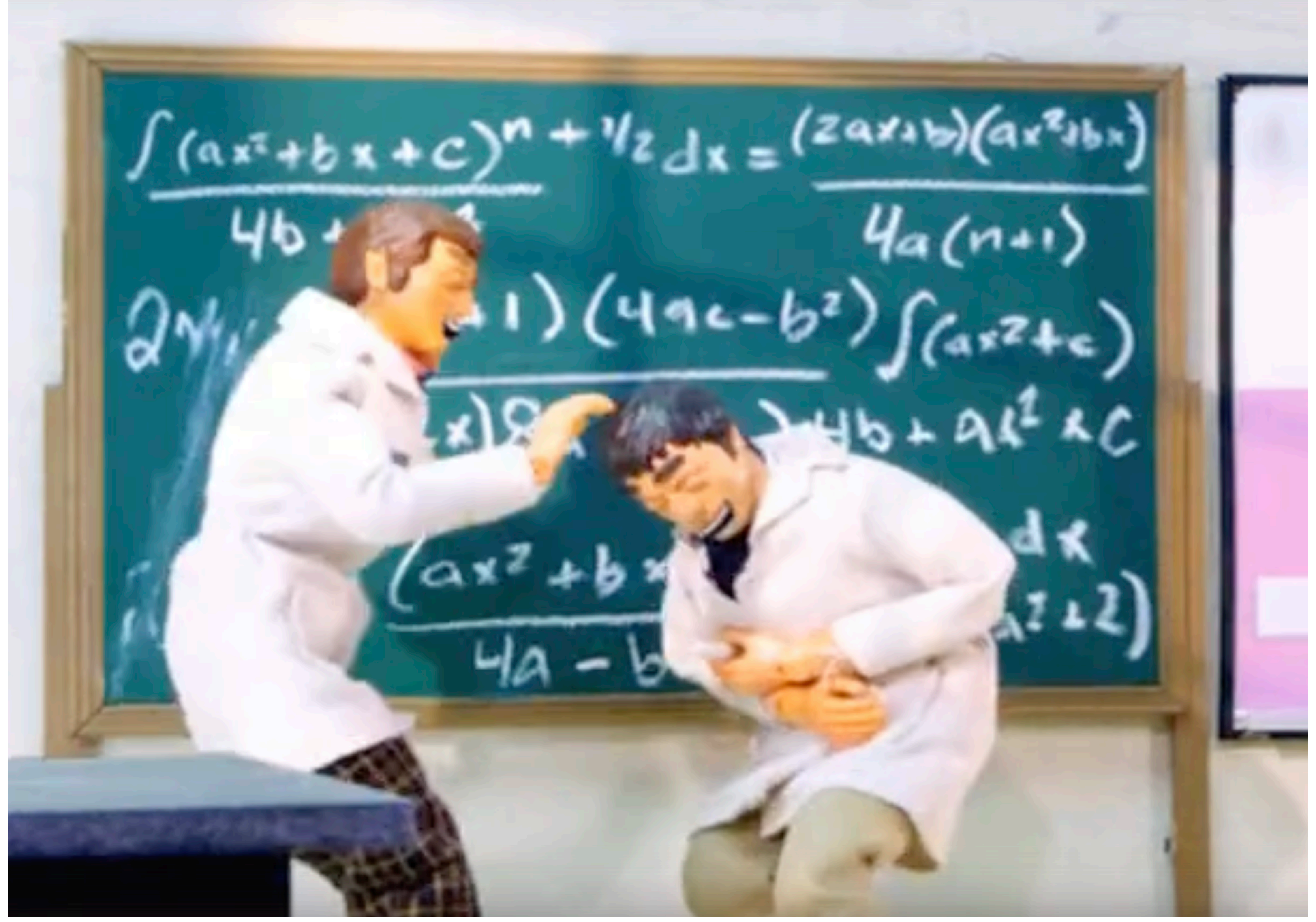
Stochastic part

Or: $dx = f(x, \beta)d\beta + g(x, \beta)dW_\beta$

Our case: $dx = -\frac{x}{2}d\beta + dW_\beta$

Probabilistic forward diffusion (1D):

$$\frac{dp(x, \beta)}{d\beta} = \frac{\partial}{\partial x}\frac{x}{2}p(x, \beta) + \frac{\partial^2}{\partial x^2}\frac{1}{2}p(x, \beta)$$



SDE -> ODE 🤯

Discrete forward diffusion:

$$\mathbf{x}_t = \sqrt{1-d\beta}\cdot\mathbf{x}_{t-1} + \sqrt{d\beta}\cdot\epsilon_t$$

Continuous forward diffusion:

$$d\mathbf{x} = \underbrace{-\frac{\mathbf{x}}{2}d\beta}_{\mu(\mathbf{x},\beta)} + \epsilon_\beta\underbrace{\sqrt{d\beta}}_{\sigma(\mathbf{x},\beta)}$$

Deterministic part

Stochastic part

Or: $dx = f(x,\beta)d\beta + g(x,\beta)dW_\beta$

Our case: $dx = -\frac{x}{2}d\beta + dW_\beta$

Probabilistic forward diffusion (1D):

$$\frac{dp(x,\beta)}{d\beta} = \frac{\partial}{\partial x}\frac{x}{2}p(x,\beta) + \frac{\partial^2}{\partial x^2}\frac{1}{2}p(x,\beta)$$

Plan:

1. We invert the time flow $\tilde{\beta} = 1-\beta$
2. We try to build the SDE of the same form

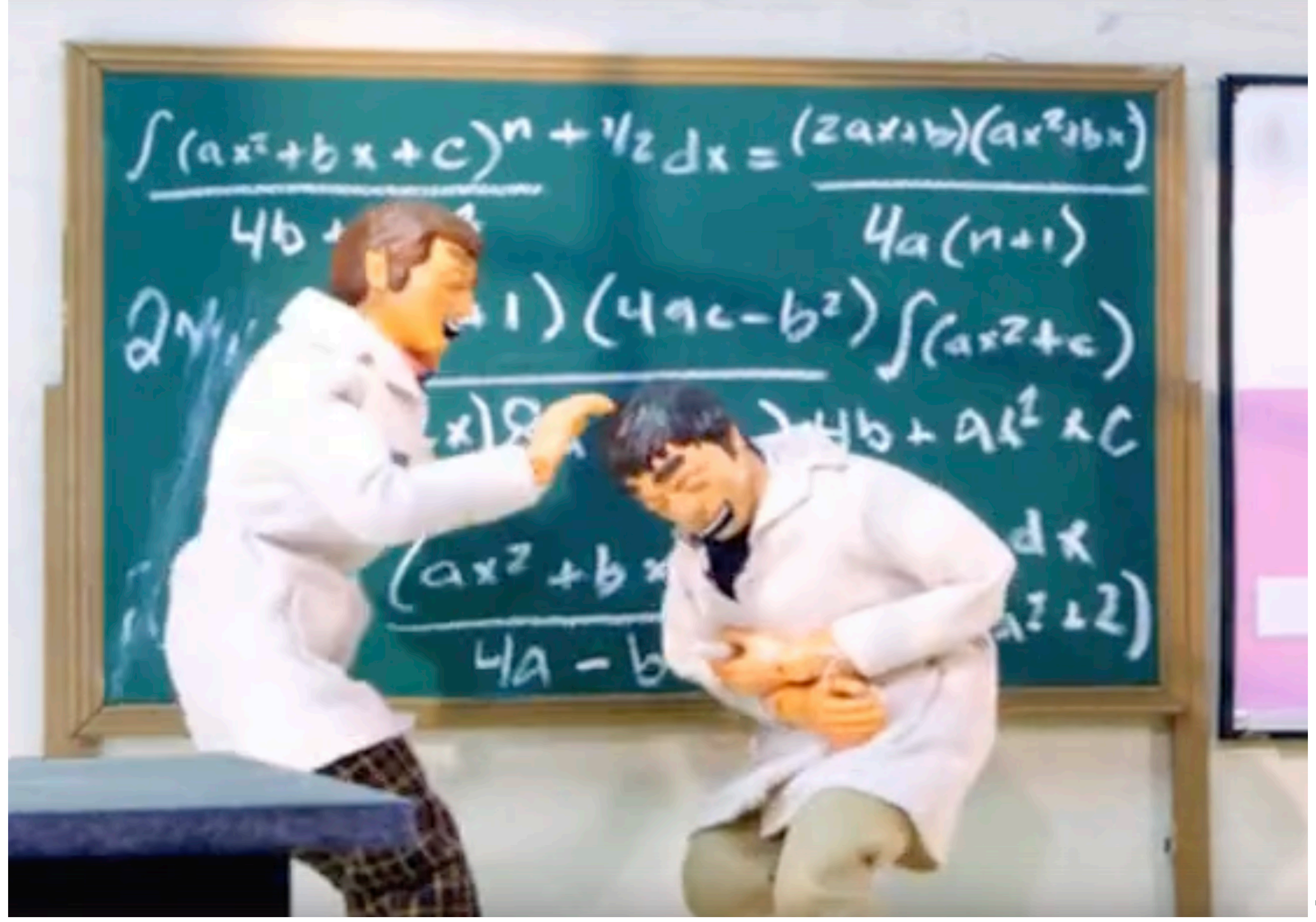
$$\frac{dp(x)}{d(1-\tilde{\beta})} = \frac{\partial}{\partial x}\frac{x}{2}p(x,1-\tilde{\beta}) + \frac{\partial^2}{\partial x^2}\frac{1}{2}p(x,1-\tilde{\beta})$$

$$\frac{dp(x)}{d\tilde{\beta}} = -\frac{\partial}{\partial x}\frac{x}{2}p(x,\tilde{\beta}) - \frac{\partial^2}{\partial x^2}\frac{1}{2}p(x,\tilde{\beta})$$

$$\frac{dp(x)}{d\beta} = \frac{\partial}{\partial x}\left[-\frac{x}{2}p(x,\beta) - \frac{1}{2}\frac{\partial}{\partial x}p(x,\beta)\right]$$

$$\frac{dp(x)}{d\beta} = \frac{\partial}{\partial x}\left[-\frac{x}{2}p(x,\beta) - \frac{1}{2}\frac{p(x,\beta)}{p(x,\beta)}\frac{\partial}{\partial x}p(x,\beta)\right]$$

$$\frac{dp(x)}{d\beta} = \frac{\partial}{\partial x}\left[-\frac{x}{2}p(x,\beta) - \frac{1}{2}p(x,\beta)\frac{\partial}{\partial x}\ln p(x,\beta)\right]$$



SDE -> ODE 🤯

Discrete forward diffusion:

$$\mathbf{x}_t = \sqrt{1 - d\beta} \cdot \mathbf{x}_{t-1} + \sqrt{d\beta} \cdot \epsilon_t$$

Continuous forward diffusion:

$$d\mathbf{x} = \underbrace{-\frac{\mathbf{x}}{2} d\beta}_{\mu(\mathbf{x},\beta)} + \epsilon_\beta \underbrace{\sqrt{d\beta}}_{\sigma(\mathbf{x},\beta)}$$

Deterministic part

Stochastic part

Or: $dx = f(x, \beta)d\beta + g(x, \beta)dW_\beta$

Our case: $dx = -\frac{x}{2}d\beta + dW_\beta$

Probabilistic forward diffusion (1D):

$$\frac{dp(x,\beta)}{d\beta} = \frac{\partial}{\partial x}\frac{x}{2}p(x,\beta) + \frac{\partial^2}{\partial x^2}\frac{1}{2}p(x,\beta)$$

Plan:

1. We invert the time flow $\tilde{\beta} = 1 - \beta$
2. We try to build the SDE of the same form

$$\frac{dp(x)}{d(1-\tilde{\beta})} = \frac{\partial}{\partial x}\frac{x}{2}p(x,1-\tilde{\beta}) + \frac{\partial^2}{\partial x^2}\frac{1}{2}p(x,1-\tilde{\beta})$$

$$\frac{dp(x)}{d\tilde{\beta}} = -\frac{\partial}{\partial x}\frac{x}{2}p(x,\tilde{\beta}) - \frac{\partial^2}{\partial x^2}\frac{1}{2}p(x,\tilde{\beta})$$

$$\frac{dp(x)}{d\beta} = \frac{\partial}{\partial x}\left[-\frac{x}{2}p(x,\beta) - \frac{1}{2}\frac{\partial}{\partial x}p(x,\beta)\right]$$

$$\frac{dp(x)}{d\beta} = \frac{\partial}{\partial x}\left[-\frac{x}{2}p(x,\beta) - \frac{1}{2}\frac{p(x,\beta)}{p(x,\beta)}\frac{\partial}{\partial x}p(x,\beta)\right]$$

$$\frac{dp(x)}{d\beta} = \frac{\partial}{\partial x}\left[-\frac{x}{2}p(x,\beta) - \frac{1}{2}p(x,\beta)\frac{\partial}{\partial x}\ln p(x,\beta)\right]$$

Resulting Probabilistic ODE:

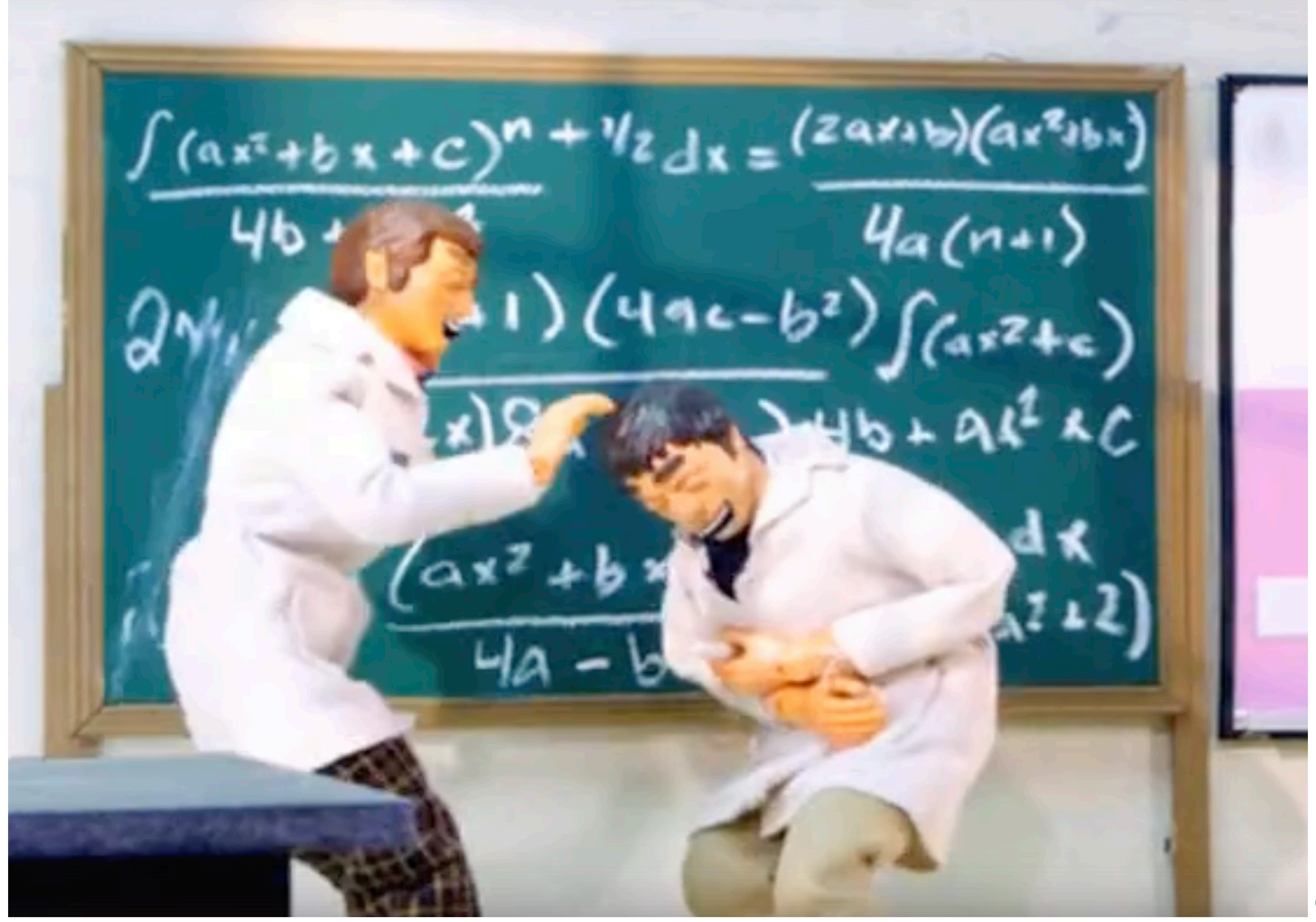
$$\frac{dp}{d\beta} = \frac{\partial}{\partial x}\left[-\frac{x}{2} - \frac{1}{2}\frac{\partial}{\partial x}\ln p(x,\beta)\right]p(x,\beta)$$

Corresponding ODE:

$$dx = \frac{1}{2}\left[-x - \frac{\partial}{\partial x}\ln p(x,\beta)\right]d\beta$$

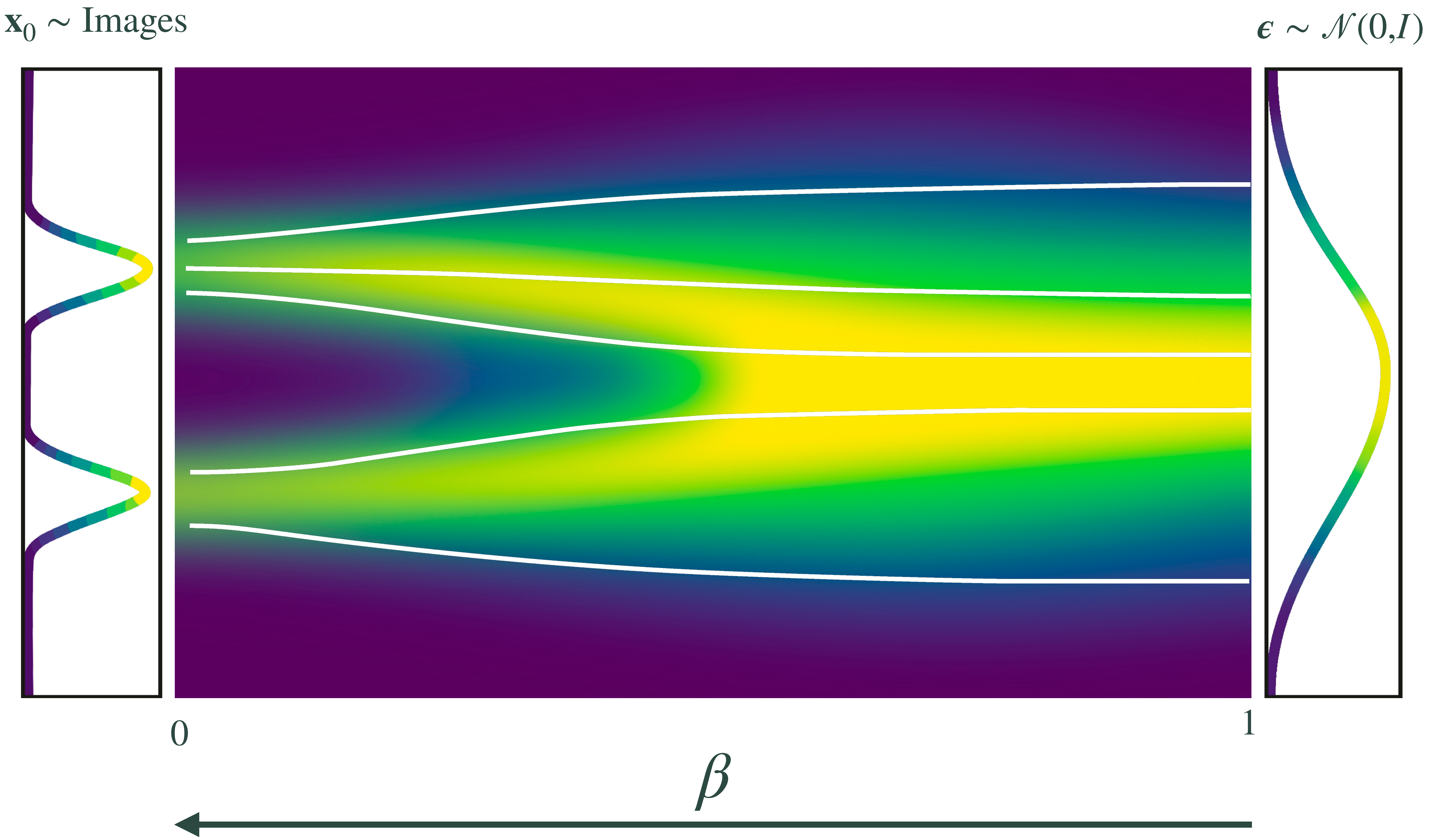
One Special Case:

$$dx = \left[f(\beta)x - \frac{1}{2}g^2(\beta)\frac{\partial}{\partial x}\ln p(x,\beta)\right]d\beta$$



The process that we have:

$$d\mathbf{x} = \frac{1}{2}\left[\mathbf{x} - \frac{\boldsymbol{\epsilon}(\mathbf{x},\beta)}{\sigma_\beta}\right](-d\beta)$$



ODE Solvers

ODE Solvers

We have ODE:

$$d\mathbf{x} = \frac{1}{2}\left[\mathbf{x} - \frac{\epsilon(\mathbf{x}, \beta)}{\sigma_\beta}\right](-d\beta)$$

General ODE:

$$\frac{d\mathbf{x}}{dt} = f(\mathbf{x}, t)$$

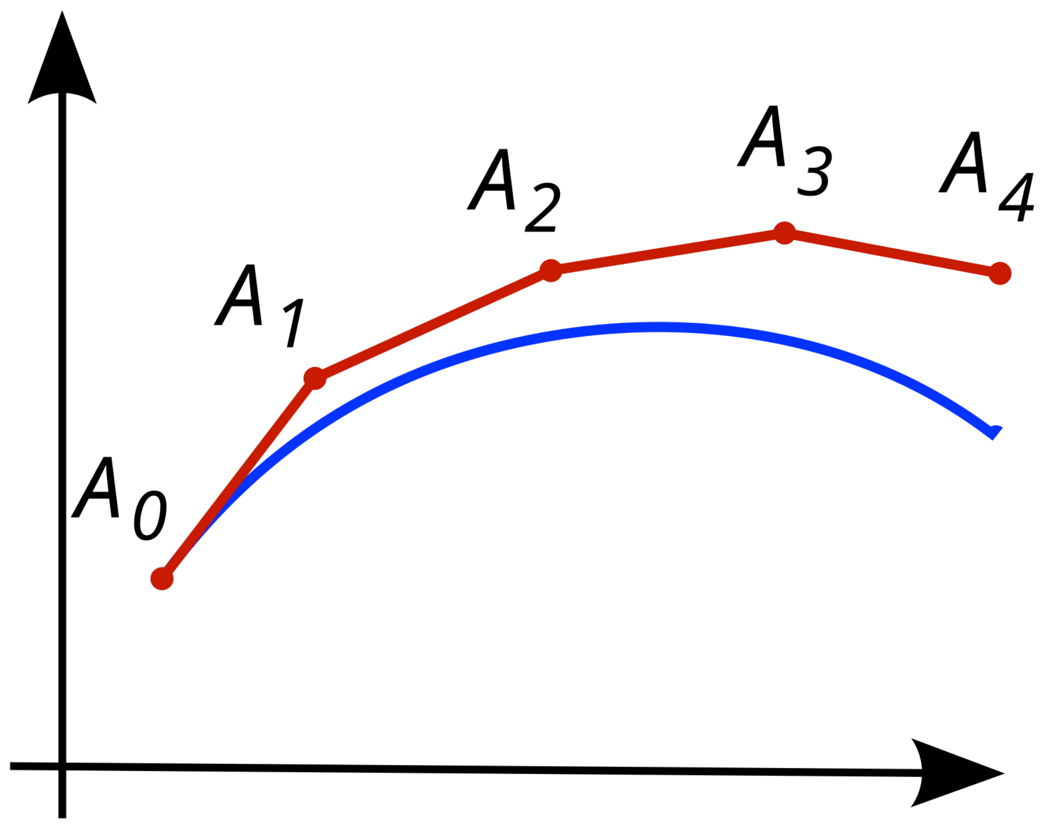
Basic Solver
(Euler's Forward Solver):

$$\mathbf{x}_{t+\Delta t} = \mathbf{x}_t + f(\mathbf{x}_t, t)\Delta t$$

This solver is 1st order solver

Local Truncation Error $o(\Delta t)$
Global Truncation Error $o(1)$

1 Network evaluation



ODE Solvers

We have ODE:

$$d\mathbf{x} = \frac{1}{2}\left[\mathbf{x} - \frac{\epsilon(\mathbf{x}, \beta)}{\sigma_\beta}\right](-d\beta)$$

General ODE:

$$\frac{d\mathbf{x}}{dt} = f(\mathbf{x}, t)$$

Basic Solver (Euler's Forward Solver):

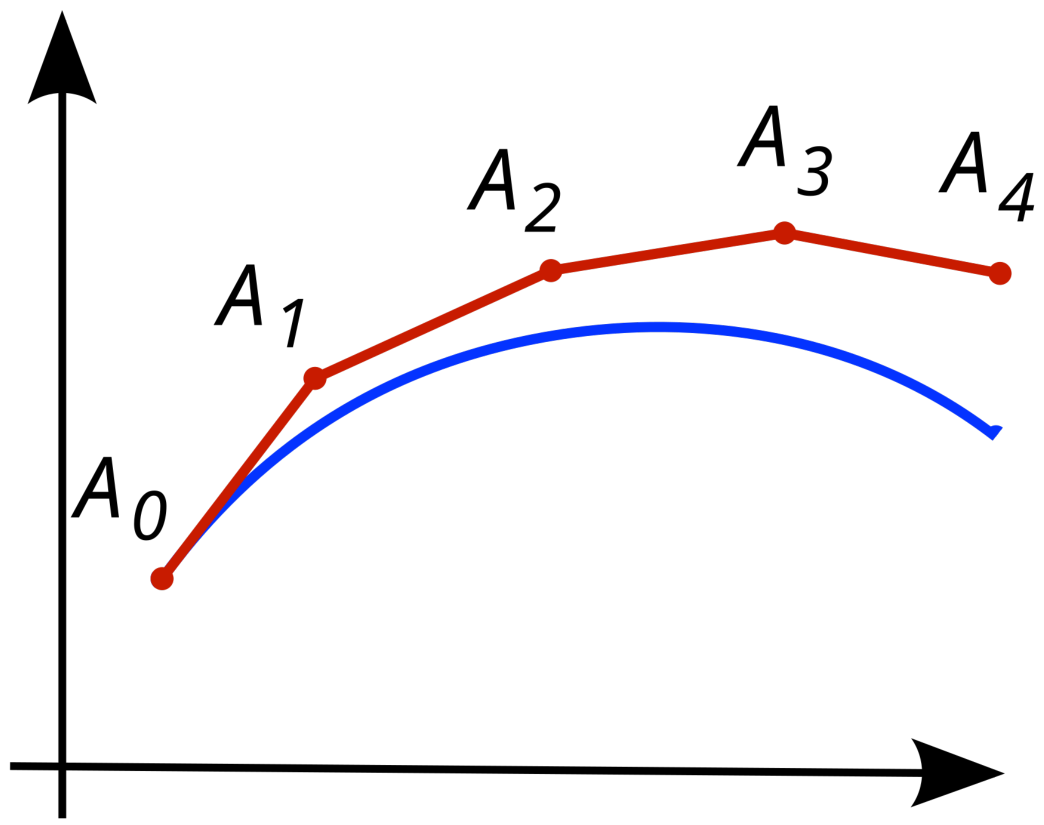
$$\mathbf{x}_{t+\Delta t} = \mathbf{x}_t + f(\mathbf{x}_t, t)\Delta t$$

This solver is 1st order solver

Local Truncation Error $o(\Delta t)$

Global Truncation Error $o(1)$

1 Network evaluation



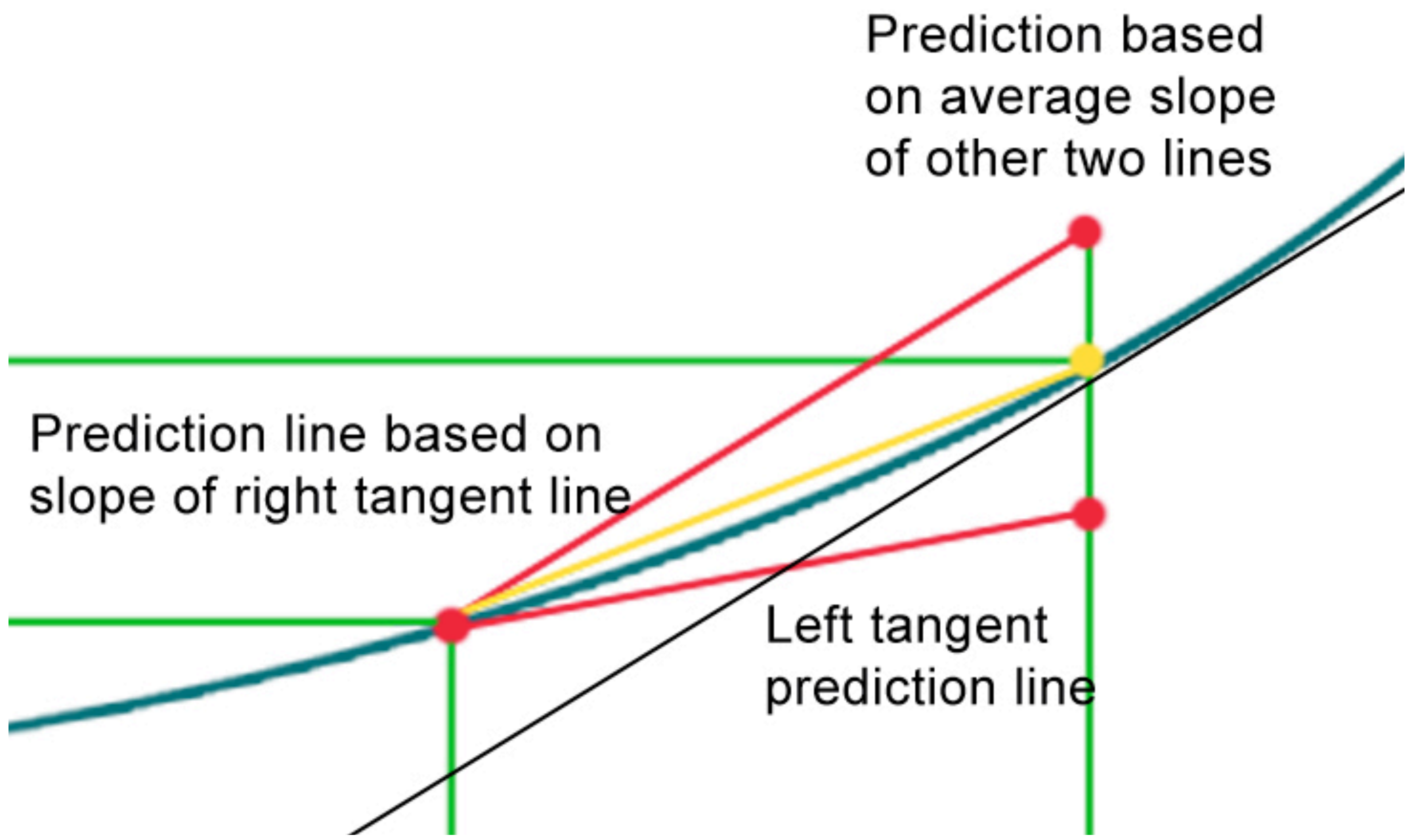
Heun's solver:

$$\tilde{\mathbf{x}}_{t+\Delta t} = \mathbf{x}_t + f(\mathbf{x}_t, t)\Delta t$$

$$\mathbf{x}_{t+\Delta t} = \mathbf{x}_t + \left[f(\mathbf{x}_t, t) + f(\tilde{\mathbf{x}}_{t+\Delta t}, t + \Delta t)\right]\frac{\Delta t}{2}$$

2nd order solver

2 Network evaluations



ODE Solvers

We have ODE:

$$d\mathbf{x} = \frac{1}{2}\left[\mathbf{x} - \frac{\epsilon(\mathbf{x},\beta)}{\sigma_\beta}\right](-d\beta)$$

General ODE:

$$\frac{d\mathbf{x}}{dt} = f(\mathbf{x}, t)$$

Basic Solver (Euler's Forward Solver):

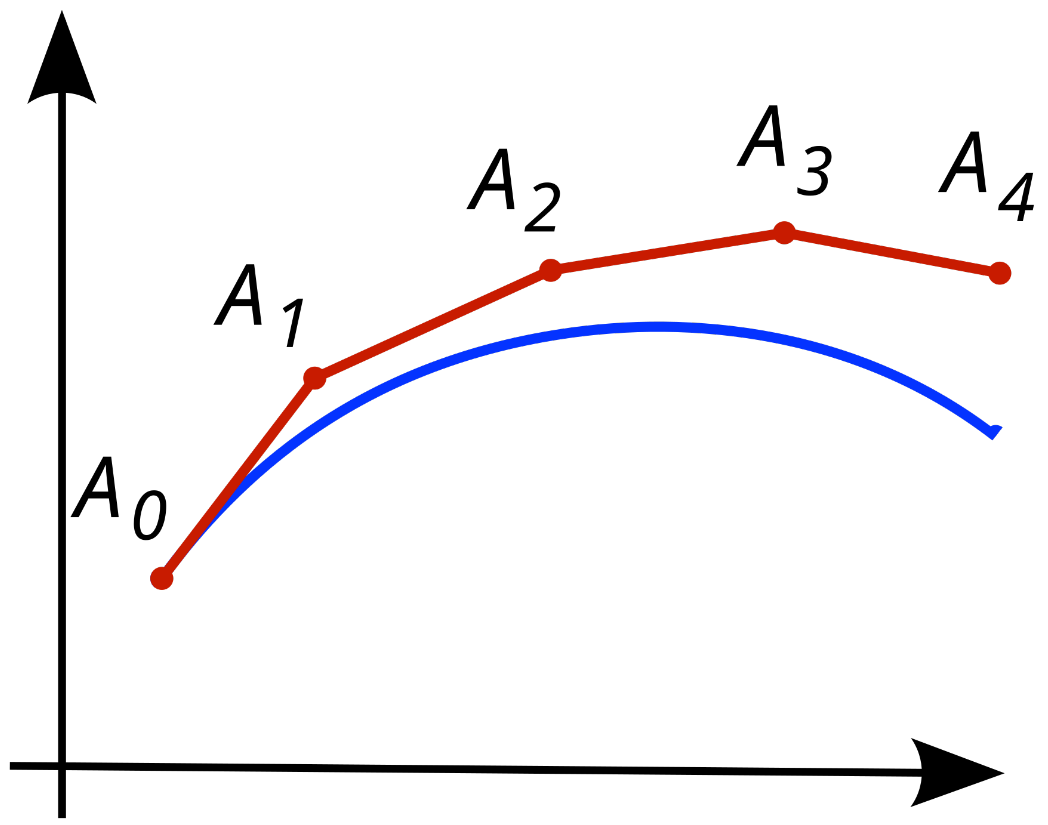
$$\mathbf{x}_{t+\Delta t} = \mathbf{x}_t + f(\mathbf{x}_t, t)\Delta t$$

This solver is 1st order solver

Local Truncation Error $o(\Delta t)$

Global Truncation Error $o(1)$

1 Network evaluation



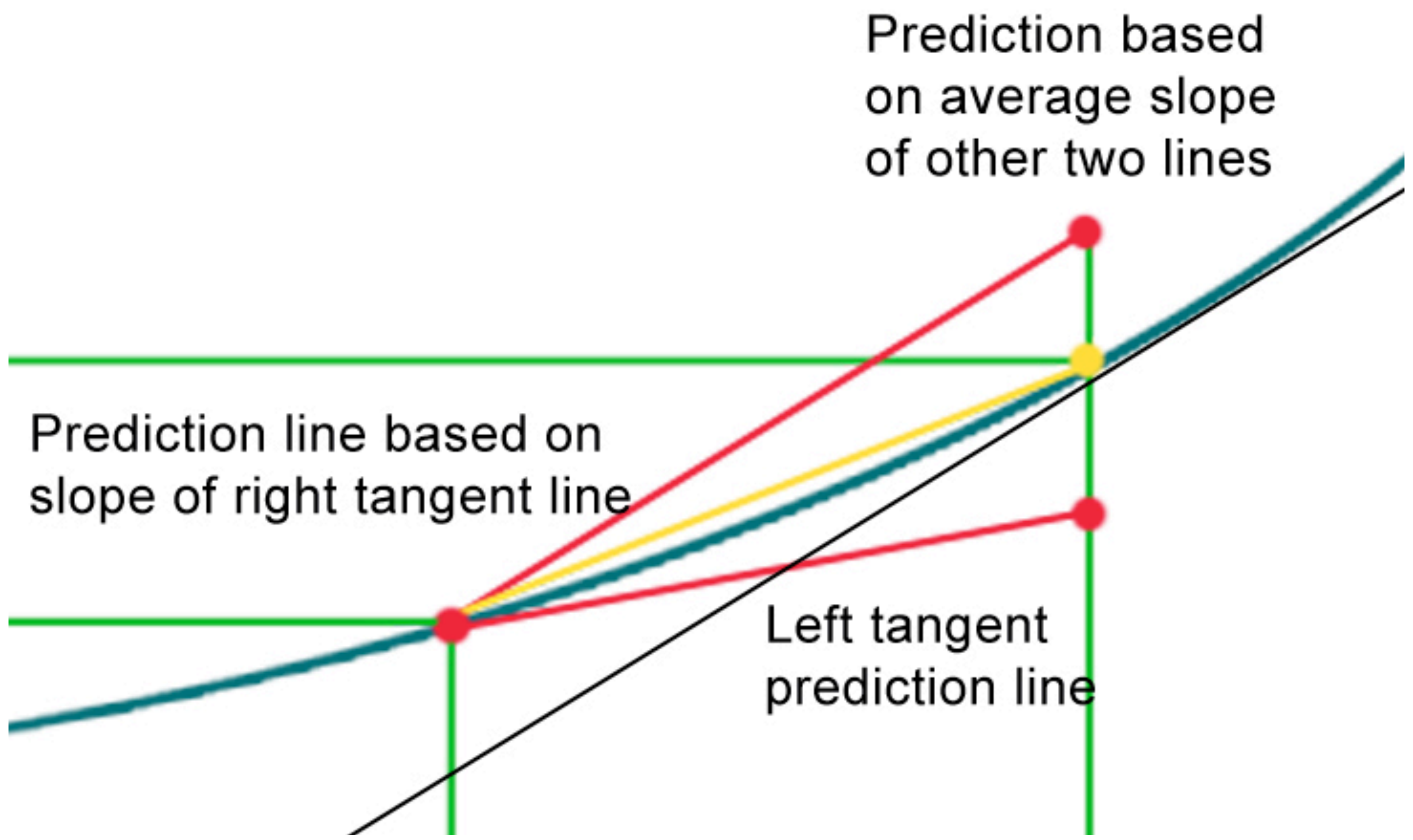
Heun's solver:

$$\tilde{\mathbf{x}}_{t+\Delta t} = \mathbf{x}_t + f(\mathbf{x}_t, t)\Delta t$$

$$\mathbf{x}_{t+\Delta t} = \mathbf{x}_t + \left[f(\mathbf{x}_t, t) + f(\tilde{\mathbf{x}}_{t+\Delta t}, t+\Delta t)\right]\frac{\Delta t}{2}$$

2nd order solver

2 Network evaluations



Runge-Kutta solver (4th order):

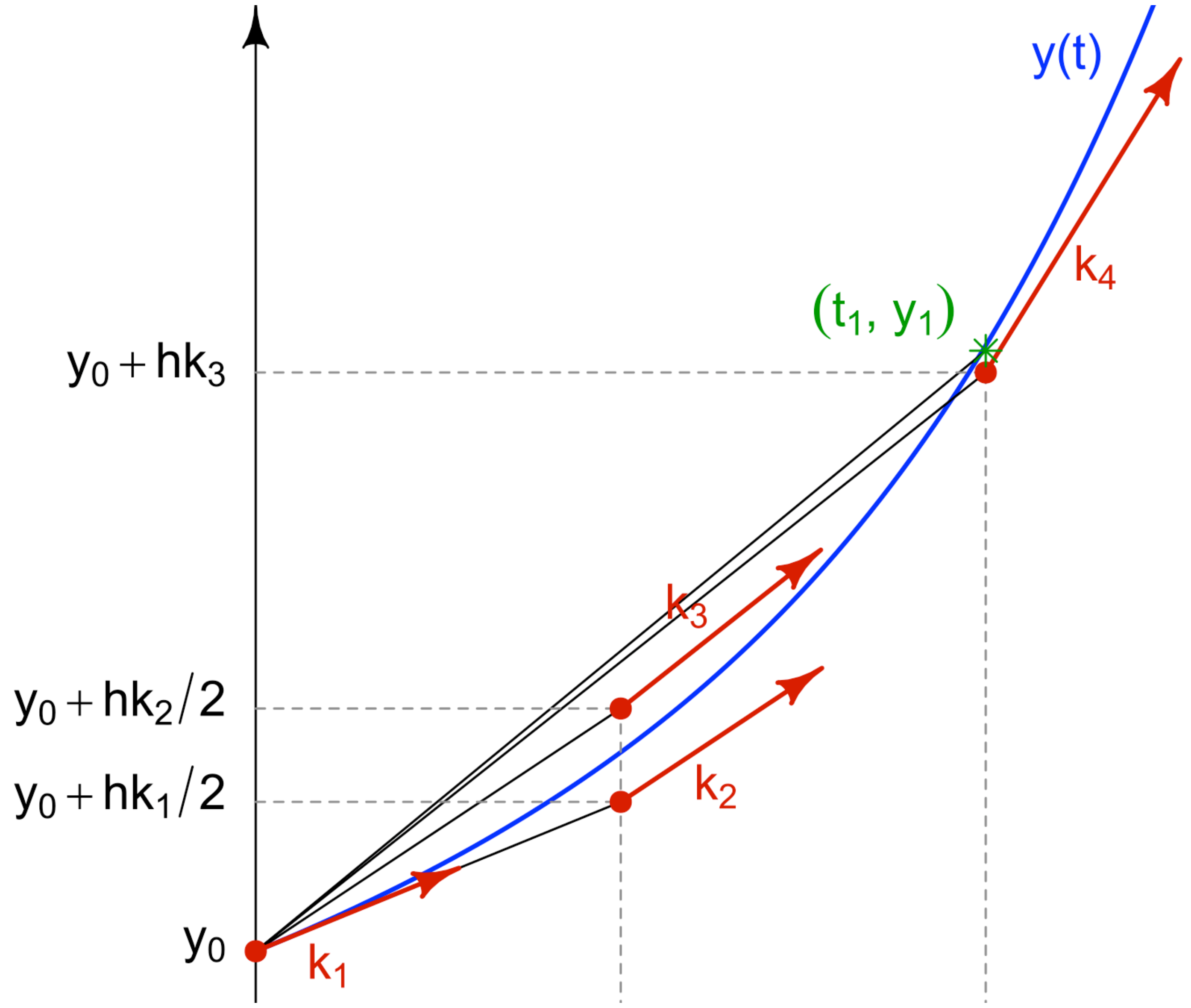
$$\mathbf{k}_1 = f(\mathbf{x}_t, t)$$

$$\mathbf{k}_2 = f\left(\mathbf{x}_t + \frac{\Delta t}{2}\mathbf{k}_1, t + \frac{\Delta t}{2}\right)$$

$$\mathbf{k}_3 = f\left(\mathbf{x}_t + \frac{\Delta t}{2}\mathbf{k}_2, t + \frac{\Delta t}{2}\right)$$

$$\mathbf{k}_4 = f\left(\mathbf{x}_t + \mathbf{k}_3\Delta t, t + \Delta t\right)$$

$$\mathbf{x}_{t+\Delta t} = \mathbf{x}_t + \frac{\Delta t}{6}(\mathbf{k}_1 + 2\mathbf{k}_2 + 2\mathbf{k}_3 + \mathbf{k}_4)$$



Adams methods

General ODE:

$$\frac{d\mathbf{x}}{dt} = f(\mathbf{x}, t)$$

We have ODE:

$$d\mathbf{x} = \frac{1}{2}\left[\mathbf{x} - \frac{\epsilon(\mathbf{x}, \beta)}{\sigma_\beta}\right](-d\beta)$$

Basic Solver (Euler's Forward Solver):

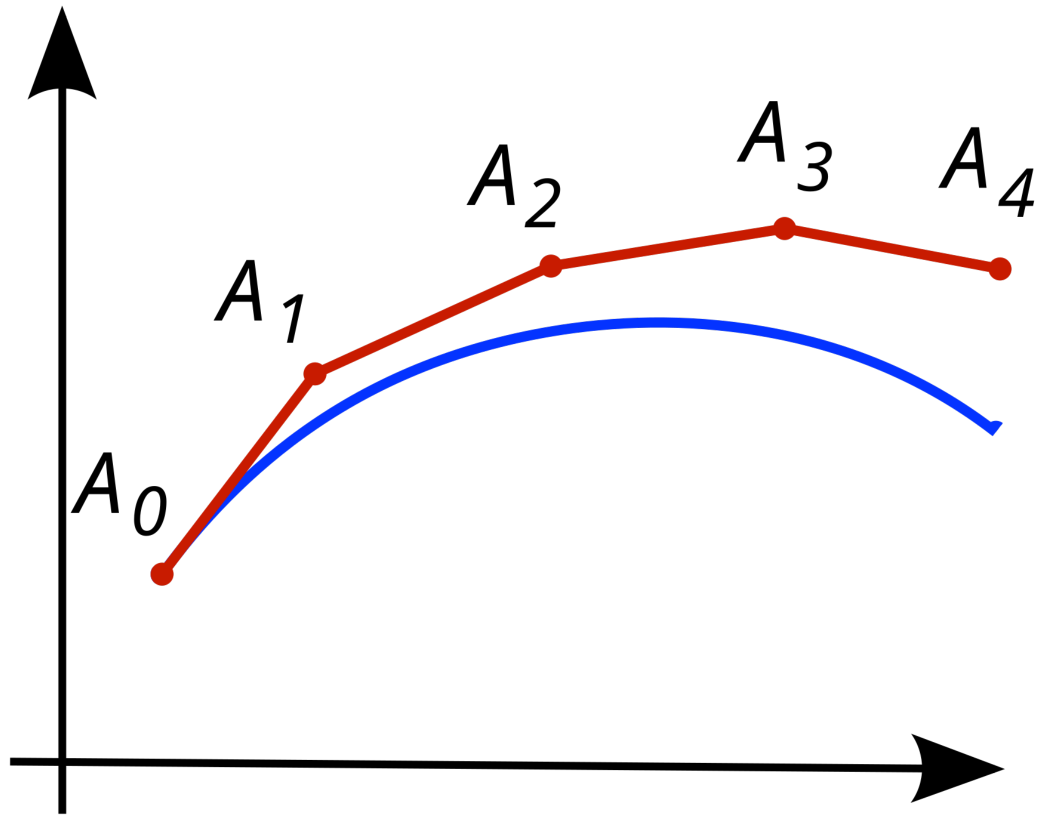
$$\mathbf{x}_{t+\Delta t} = \mathbf{x}_t + f(\mathbf{x}_t, t)\Delta t$$

This solver is 1st order solver

Local Truncation Error $o(\Delta t)$

Global Truncation Error $o(1)$

1 Network evaluation



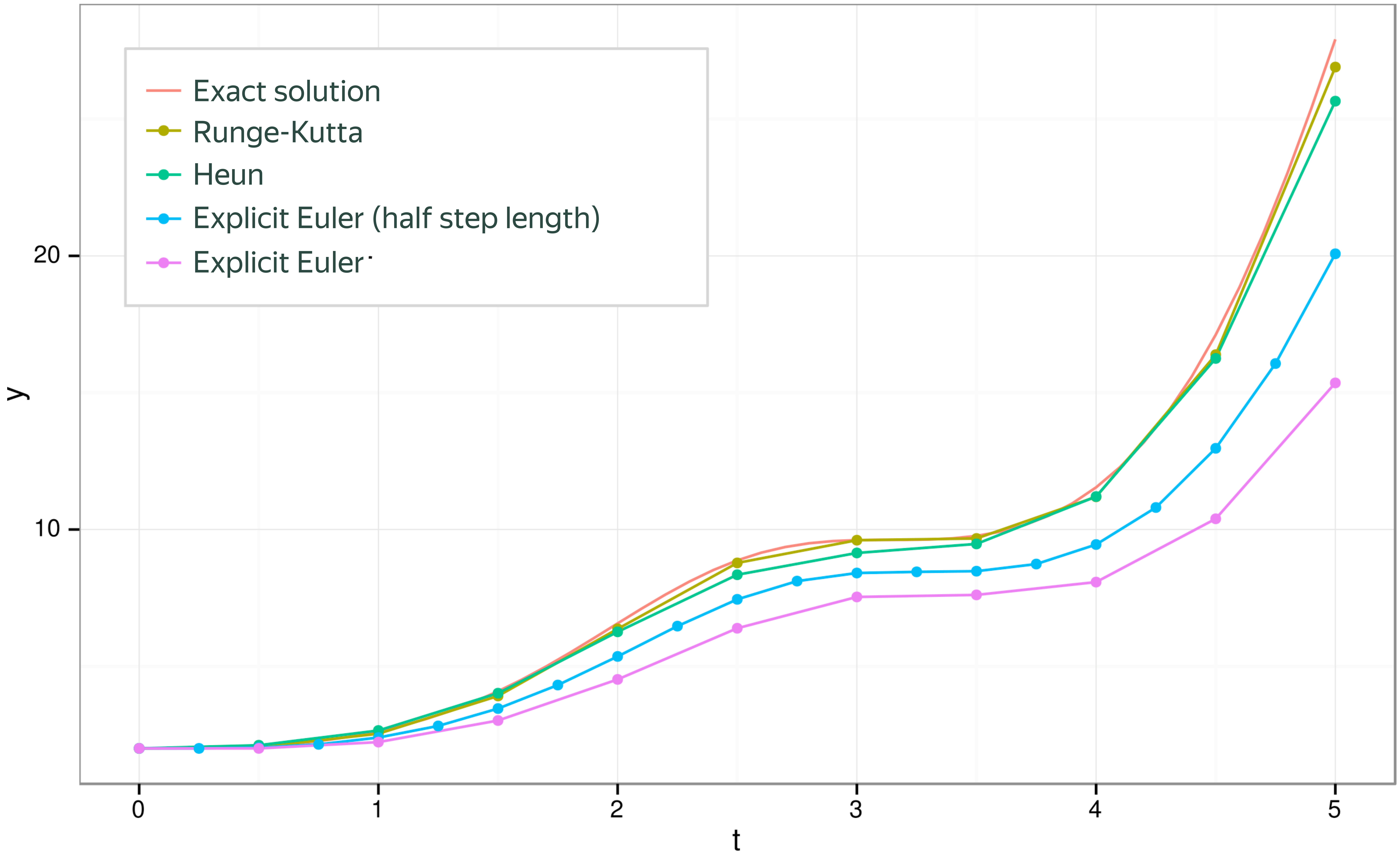
Adams-Bashforth Solver:

$$\mathbf{x}_{t+\Delta t} = \mathbf{x}_t + \left[\frac{3}{2}f(\mathbf{x}_t, t) - \frac{1}{2}f(\mathbf{x}_{t-\Delta t}, t - \Delta t)\right]\Delta t$$

2nd order solver

1 Network evaluation per iteration
(because the previous we already know)

ODE Solvers Comparison



Truncation Error

Equation:

$$\frac{dx}{dt} = f(x, t)$$

Basic Solver
(Euler's Forward Solver):

$$x_{t+\Delta t} = A(t, x_t, \Delta t, f)x_t + f(x_t, t)\Delta t$$

Local Truncation Error:

$$\tau_t = x(t + \Delta t) - x_{t+\Delta t} = x(t + \Delta t) - x_t + f(x_t, t)\Delta t =$$

$$= x(t + \Delta t) - x_t + \left.\frac{dx}{dt}\right|_t \Delta t =$$

$$= x(t) + \left.\frac{dx}{dt}\right|_t \Delta t + \frac{\partial^2 x}{\partial t^2}\Delta t^2 + o(\Delta t^2) - x_t + \left.\frac{dx}{dt}\right|_t \Delta t =$$

$$= \frac{d^2 x}{dt^2}\Delta t^2 + o(\Delta t^2) = o(\Delta t)$$

Truncation Error

Equation:

$$\frac{dx}{dt} = f(x, t)$$

Basic Solver
(Euler's Forward Solver):

$$x_{t+\Delta t} = A(t, x_t, \Delta t, f)x_t + f(x_t, t)\Delta t$$

Local Truncation Error:

$$\tau_t = x(t+\Delta t) - x_{t+\Delta t} = x(t+\Delta t) - x_t + f(x_t, t)\Delta t =$$

$$= x(t+\Delta t) - x_t + \left.\frac{dx}{dt}\right|_t \Delta t =$$

$$= x(t) + \left.\frac{dx}{dt}\right|_t \Delta t + \frac{\partial^2 x}{\partial t^2}\Delta t^2 + o(\Delta t^2) - x_t + \left.\frac{dx}{dt}\right|_t \Delta t =$$

$$= \frac{d^2 x}{dt^2}\Delta t^2 + o(\Delta t^2) = o(\Delta t)$$

Total Accumulated Error:

$$e_{n+1} = e_n + \big(A(t_n, x(t_n), \Delta t, f) - A(t_n, x_n, \Delta t, f)\big)\,\Delta t + \tau_{n+1}$$

Not tractable, so a lot of interesting math here...

Spectral Stability Criterion...

Total Accumulated Error for our methods:

* **Euler:** $O(\Delta t)$
* **Heun:** $O(\Delta t^2)$
* **Runge-Kutta:** $O(\Delta t^4)$

Different (De)Noising Processes

Commonly Used Diffusion Processes

Forward Diffusion:

$$dx = f(\beta)x d\beta + g(\beta)\boldsymbol{\epsilon}_\beta \sqrt{d\beta}$$

ODE:

$$dx = \left[f(\beta)x - \frac{1}{2} g^2(\beta) \frac{\partial}{\partial x} \ln p(x, \beta) \right] d\beta$$

Variance preserving:

$$f(t) = \frac{1}{2}\frac{\mathrm{d}\log\alpha_t}{\mathrm{d}t},\ g(t) = \sqrt{\left(-\frac{\mathrm{d}\log\alpha_t}{\mathrm{d}t}\right)},$$

where α_t — bounded decreasing sequence from 1 to 0

Variance exploding:

$$f(t) = 0,\ g(t) = \sqrt{\frac{\mathrm{d}\sigma_t^2}{\mathrm{d}t}}$$

where σ_t is an unbounded increasing sequence, from 0 to C

DPM Solver

One Special Case:

$$\frac{dx}{d\beta} = \left[f(\beta)x - \frac{1}{2}g(\beta)^2 \frac{\partial}{\partial x} \ln p(x, \beta) \right]$$

$$f(\beta) = \frac{d \log \alpha_t}{dt}$$

$$g^2(\beta) = \frac{d\sigma_t^2}{dt} - 2\frac{d \log \alpha_t}{dt}\sigma_t^2$$

Equation:

$$\frac{dx}{d\beta} - f(\beta)x = -\frac{1}{2}g(\beta)\ln p(x, \beta)$$

Uniform equation:

$$\frac{dx}{d\beta} - f(\beta)x = 0$$

Uniform Soltuion:

$$x = Ce^{\int f(\beta)d\beta}$$

Variation of constants:

$$x = C(\beta)e^{\int f(\beta)d\beta}$$

Total solution:

$$\mathbf{x}_t = \mathbf{x}_s e^{\int_s^t f(\beta)d\beta} + \int_s^t e^{\int_\tau^t f(r)dr} \frac{g^2(\tau)}{2\sigma_\tau} \boldsymbol{\epsilon}(\mathbf{x}_\tau, \tau)d\tau$$

$$\lambda = \log \frac{\alpha_\beta}{\sigma_\beta}$$

$$\mathbf{x}_t = \frac{\alpha_t}{\alpha_s}\mathbf{x}_s + \alpha_t \int_s^t e^{-\lambda}\boldsymbol{\epsilon}(\mathbf{x}_s, \lambda)d\lambda$$

Summary

01

Started from diffusion formulation (Variance Preserving)

02

Derived simple sampler

03

Derived forward SDE

04

Derived forward Probability ODE

05

Inverted time for probability ODE

06

Derived backward SDE

07

Reduced it to ODE

08

Proposed how to solve it

09

DPM solver (partly analytical)

10

EDM solver