

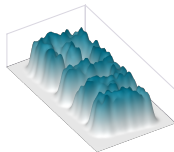


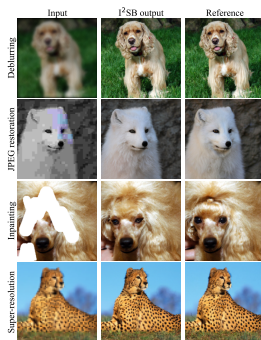
I²SB: Image-to-Image Schrödinger Bridge

From generative modeling to interpolation

Johannes Herter

Mathematical Foundations of Computer Graphics and Vision · 30 May 2023





Outputs of I²SB on ImageNet 256×256: super-resolution, deblurring, JPEG restoration, inpainting. Liu, Vahdat, Huang, Theodorou, Nie, Anandkumar, ICML 2023, Figure 1.

Start from noise, or start from the image?

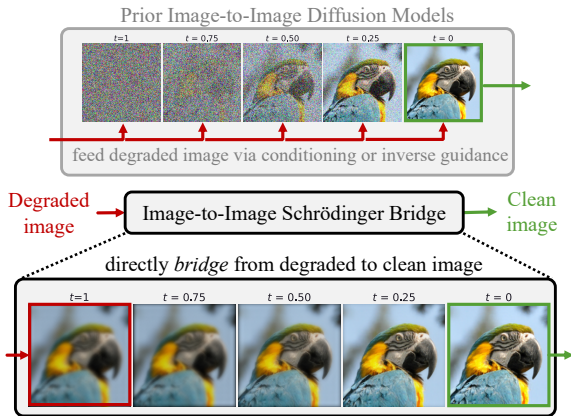


Figure 3 of the paper.

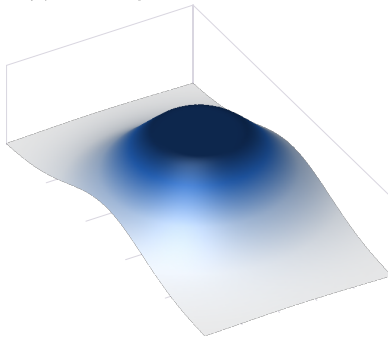
- Diffusion models generate from **white noise**, which carries no information about the image.
- In restoration, the degraded image is already **close** to the answer.
- I²SB: learn a diffusion **bridge** directly from degraded to clean images.

Plan Diffusion models → Schrödinger bridges → a tractable bridge that trains exactly like a diffusion model.

1 *Diffusion models*

forward diffusion: the density of X_t as a landscape

$t = 1.00$



The density of the ETH logo, $x_0 \sim p_0$, smoothed by noise until only a Gaussian hill is left. Generation means running this film backwards.

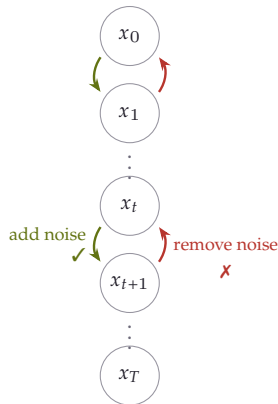
- Forward: $x_0 \sim p_0$, then $x_{t+1} \sim p_{t+1|t}(\cdot | x_t)$, a kernel we **choose**: add Gaussian noise. ✓

$$p(x_{0:T}) = p_0(x_0) \prod_{t=0}^{T-1} p_{t+1|t}(x_{t+1} | x_t)$$

- Backward: the same law, factorised the other way,

$$p(x_{0:T}) = p_T(x_T) \prod_{t=0}^{T-1} p_{t|t+1}(x_t | x_{t+1})$$

- Ancestral sampling: $x_T \sim p_T$, then $x_t \sim p_{t|t+1}(\cdot | x_{t+1})$, $t = T - 1, \dots, 0$.
- But the backward kernel depends on the unknown data. ✗



[Sohl-Dickstein et al. 2015; Ho et al. 2020]

- Choose a discretised Ornstein–Uhlenbeck step, Gaussian and stationary: ✓

$$p_{t+1|t}(x_{t+1} | x_t) = \mathcal{N}(x_{t+1}; \alpha x_t, (1 - \alpha^2)I)$$

- Bayes' rule turns it around:

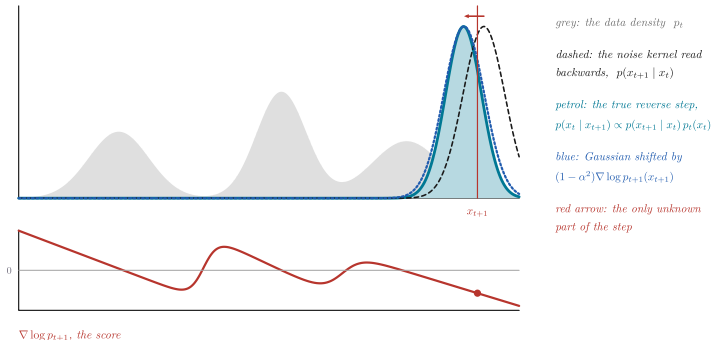
$$p_{t|t+1}(x_t | x_{t+1}) = p_{t+1|t}(x_{t+1} | x_t) \exp [\log p_t(x_t) - \log p_{t+1}(x_{t+1})]$$

- For small steps ($\alpha \rightarrow 1$), a first-order Taylor expansion of the exponent gives a Gaussian again:

$$p_{t|t+1}(x_t | x_{t+1}) \approx \mathcal{N}(x_t; (2 - \alpha) x_{t+1} + (1 - \alpha^2) \nabla \log p_{t+1}(x_{t+1}), (1 - \alpha^2)I)$$

The only unknown The shape is Gaussian, the variance is known and most of the mean is known. One vector field is missing: the **score** $\nabla \log p_t$. Learn it, and everything else is analytic. Derivation: A1.

one reverse step: where can x_t have come from?



The true backward step is the noise kernel tilted towards higher density. The tilt is the score.

- The marginal score $\nabla \log p_t(x_t)$ is unknown. ✗ The conditional score given x_0 is Gaussian: ✓

$$\nabla \log p_{t|0}(x_t | x_0) = -\frac{x_t - a_t x_0}{\sigma_t^2}, \quad a_t = \alpha^t, \quad \sigma_t^2 = 1 - \alpha^{2t}$$

- The marginal score is the conditional expectation of the conditional one, so it solves a **regression problem**:

$$s_t = \operatorname{argmin}_s \mathbb{E}_{p_{0,t}} \|s(x_t) - \nabla \log p_{t|0}(x_t | x_0)\|^2$$

- One network $s_\theta(t, x_t) \approx \nabla \log p_t(x_t)$ for all times:

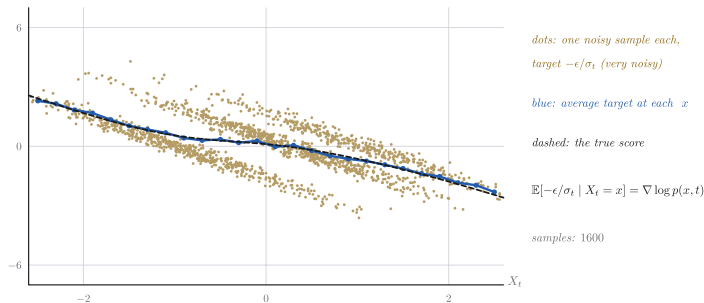
$$\theta^\star = \operatorname{argmin}_\theta \mathbb{E}_{t \sim \mathcal{U}} \mathbb{E}_{p_0(x_0)} \mathbb{E}_{p_{t|0}(x_t | x_0)} \|s_\theta(t, x_t) - \nabla \log p_{t|0}(x_t | x_0)\|^2$$

- Sample with $x_T \sim \mathcal{N}(0, I)$ and the recursion

$$x_t = (2 - \alpha) x_{t+1} + (1 - \alpha^2) s_{\theta^\star}(t+1, x_{t+1}) + \sqrt{1 - \alpha^2} z_{t+1}, \quad z_{t+1} \sim \mathcal{N}(0, I)$$

[Song and Ermon 2019; Vincent 2011]

denoising score matching: regress $-\epsilon/\sigma_t$ on $X_t = \alpha_t X_0 + \sigma_t \epsilon$



Eq. (3). $\|\epsilon(X_t, t; \theta) - \sigma_t \nabla \log p(X_t, t \mid X_0)\|$: every sample brings its own known target; their average at each X_t is the true score. Proof: A3.

Analytic ✓

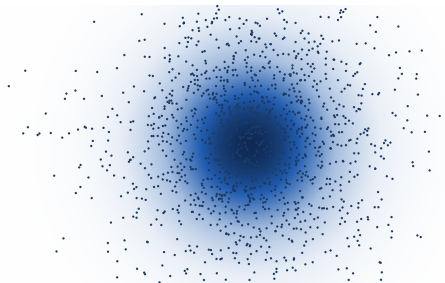
- the forward kernel $p_{t+1|t}$ and $p_{t|0}$
- the regression target $\nabla \log p_{t|0}(x_t | x_0)$
- the Gaussian form and variance of every backward step
- drawing x_t given x_0 , without simulating anything

Learned

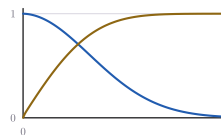
- one network $s_\theta(t, x) \approx \nabla \log p_t(x)$

The same split carries over to I²SB: a Gaussian bridge in closed form, and a single learned score.

forward: add noise, $dX_t = f_t(X_t) dt + \sqrt{\beta_t} dW_t$



$t = 1.00$



α_t : signal

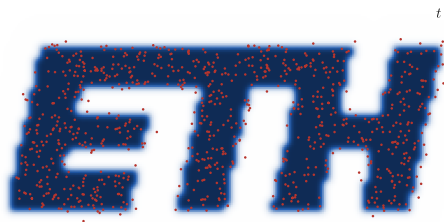
σ_t : noise

$$X_t = \alpha_t X_0 + \sigma_t \epsilon$$

at $t = 1$: $X_1 \approx \mathcal{N}(0, I)$

Eq. (1), forward. $dX_t = f_t(X_t) dt + \sqrt{\beta_t} dW_t$, $X_0 \sim p_{\mathcal{A}}$. With $\beta_t \in \mathbb{R}$ and f_t linear, $X_1 \approx \mathcal{N}(0, I)$ and, given X_0 , every X_t is Gaussian. [Song et al. 2021]

backward: $dX_t = [f_t - \beta_t \nabla \log p(X_t, t)] dt + \sqrt{\beta_t} d\bar{W}_t$



$t = 0.00$

arrows: the score

$\nabla \log p(\cdot, t)$, the direction
in which density grows

red: samples started
from $\mathcal{N}(0, I)$ at $t = 1$

Anderson (1982): this
SDE has the same
marginals, run backward

Eq. (2), backward (Anderson 1982). $dX_t = [f_t - \beta_t \nabla \log p(X_t, t)] dt + \sqrt{\beta_t} d\bar{W}_t$, $X_1 \sim \mathcal{N}(0, I)$: the recursion in continuous time; sample with s_θ and Euler–Maruyama. A2.

- Every parameterisation predicts the clean image: since $X_t = a_t X_0 + b_t \epsilon$,

$$X_0^\epsilon := (X_t - b_t \epsilon(X_t, t; \theta)) / a_t.$$

- DDPM then samples the next state as if X_0^ϵ were the truth.

Eq. (4): DDPM

$$X_n \sim p(X_n \mid X_0^\epsilon, X_{n+1}), \quad X_N \sim \mathcal{N}(0, I)$$

In words Guess the clean image, then take a small, exactly Gaussian step towards it. Repeat. I²SB keeps this sampler and only changes where it starts.

too few steps: Euler-Maruyama with N steps from $\mathcal{N}(0, I)$



$N = 2$



$N = 5$



$N = 20$



$N = 200$

too short a forward process: start at $\mathcal{N}(0, I)$ although $p_T \neq \mathcal{N}(0, I)$



$T = 0.15$



$T = 0.3$



$T = 0.5$



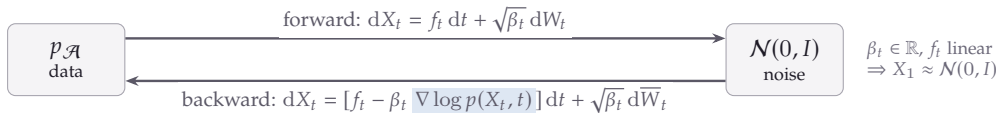
$T = 1.0$

- Too few steps: a poor approximation of the backward process.
- A forward process that has not mixed: $p_1 \neq \mathcal{N}(0, I)$, so the backward process starts in the wrong place.

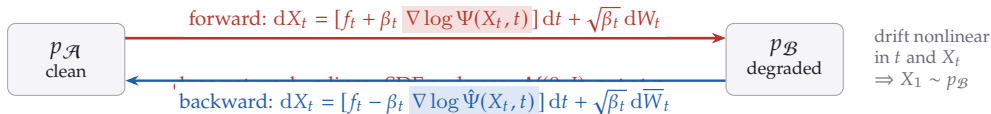
- Long sampling times. For restoration, why start from noise at all, when the degraded image is already close?

[De Bortoli et al. 2021]

2 *Schrödinger bridges*

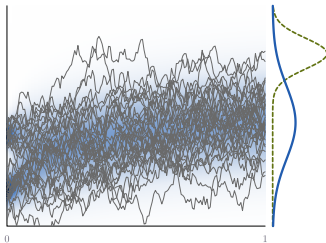


...what if we want to interpolate between **two complicated** distributions?



Two new drifts The forward drift $\nabla \log \Psi$ steers the process into $p_{\mathcal{B}}$; the backward drift $\nabla \log \hat{\Psi}$ is no longer a known score.

linear drift (as in diffusion models)

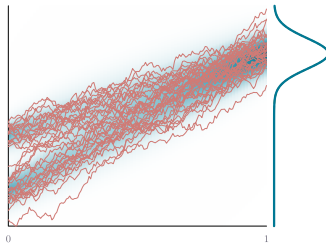


$$dX_t = f_t(X_t) dt + \sqrt{\beta_t} dW_t$$

reached at $t=1$: $\neq p_{\mathcal{B}}$: does not work

start: $p_{\mathcal{A}}$ at $t=0$; dashed green: the target $p_{\mathcal{B}}$ at $t=1$

nonlinear drift (Schrödinger bridge)

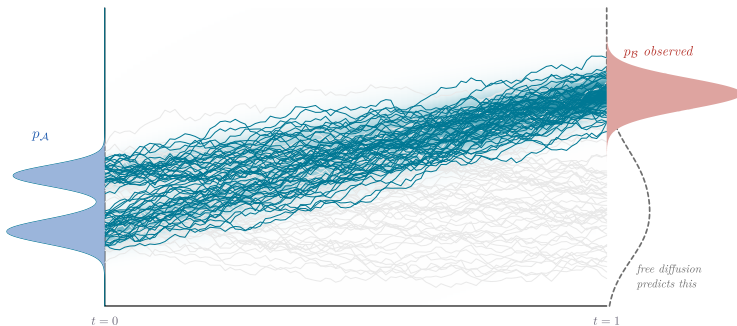


$$dX_t = [f_t + \beta_t \nabla \log \Psi(X_t, t)] dt + \sqrt{\beta_t} dW_t$$

reached at $t=1$: $= p_{\mathcal{B}}$

Left: the process forgets $p_{\mathcal{A}}$ and ends at a Gaussian. Right: the extra drift $\beta_t \nabla \log \Psi$ lands on $p_{\mathcal{B}}$.

Schrödinger (1931): *what is the most likely evolution between the two?*



petrol: the bridge, the path measure closest to free diffusion (in KL) that matches both observations

Dynamic Schrödinger bridge

$$\min_{\mathbb{Q}: \mathbb{Q}_0=p_{\mathcal{A}}, \mathbb{Q}_1=p_{\mathcal{B}}} \text{KL}(\mathbb{Q} \parallel \mathbb{P})$$

Reference $\mathbb{P}$: the path measure of the forward SDE (1).

Eq. (21): as optimal control

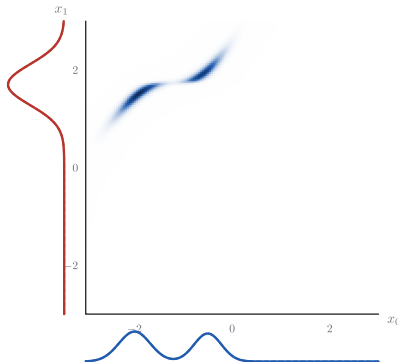
$$\min_u \mathbb{E} \int_0^1 \frac{1}{2} \|u(X_t, t)\|^2 dt$$

subject to $dX_t = [f_t + u] dt + \sqrt{\beta_t} dW_t$, $X_0 \sim p_{\mathcal{A}}$,
 $X_1 \sim p_{\mathcal{B}}$.

In words Among all ways to move $p_{\mathcal{A}}$ into $p_{\mathcal{B}}$, take the one that disturbs the reference diffusion least: the minimum-energy steering.

Its static version (the joint law of (X_0, X_1)) is **entropy-regularised optimal transport**; as the noise vanishes it becomes optimal transport [Léonard 2013; Mikami 2004].

less noise: $\sigma^2 = 0.012$, the coupling sharpens to the optimal transport map



$$\pi(x_0, x_1) = \hat{\Psi}_0(x_0) K_\sigma(x_0, x_1) \Psi_1(x_1)$$

solid: the marginals of π , dashed: targets

marginal error 0.0007

$$\min_{\pi \in \Pi(p_A, p_B)} \mathbb{E}_\pi \left[\frac{|x_0 - x_1|^2}{2\sigma^2} \right] - H(\pi)$$

*entropic optimal transport: the
static Schrödinger bridge; the same
IPF as for survey weights*

The path measure of the forward and backward SDEs, Eq. (5),

$$dX_t = [f_t + \beta_t \nabla \log \Psi(X_t, t)] dt + \sqrt{\beta_t} dW_t \quad \text{forward}$$

$$dX_t = [f_t - \beta_t \nabla \log \hat{\Psi}(X_t, t)] dt + \sqrt{\beta_t} d\bar{W}_t \quad \text{backward}$$

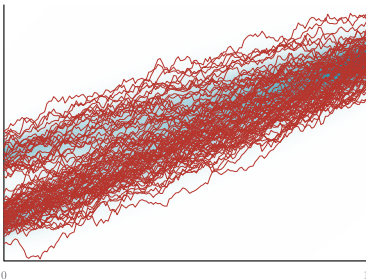
solves the Schrödinger bridge problem if Ψ and $\hat{\Psi}$ solve the **Schrödinger system**, Eq. (6):

$$\partial_t \Psi = -\nabla \Psi^\top f - \frac{1}{2} \beta \Delta \Psi, \quad \partial_t \hat{\Psi} = -\nabla \cdot (\hat{\Psi} f) + \frac{1}{2} \beta \Delta \hat{\Psi},$$

$$\begin{aligned} \Psi(\cdot, 0) \hat{\Psi}(\cdot, 0) &= p_{\mathcal{A}} \\ \Psi(\cdot, 1) \hat{\Psi}(\cdot, 1) &= p_{\mathcal{B}} \end{aligned}$$

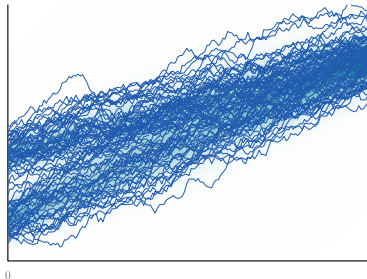
Why this is hard Two linear PDEs, one forward in time and one backward. Each alone is easy; they are coupled **nonlinearly through their boundary values**, so one must iterate (IPF / Sinkhorn over path measures). Optimal control $\rightarrow$ (6): A4.

forward from p_A



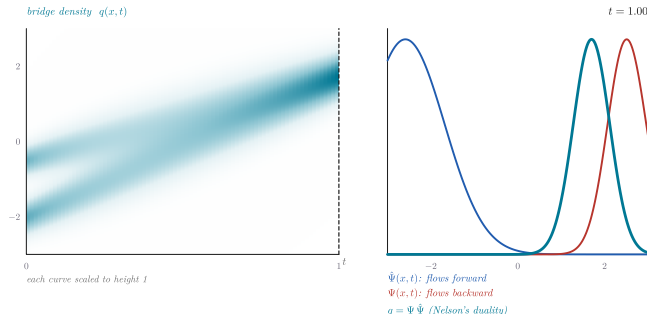
$$dX_t = [f_t + \beta_t \nabla \log \Psi] dt + \sqrt{\beta_t} dW_t$$

backward from p_B



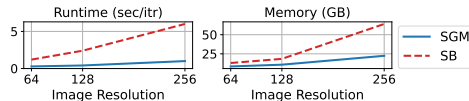
$$dX_t = [f_t - \beta_t \nabla \log \hat{\Psi}] dt + \sqrt{\beta_t} d\bar{W}_t$$

both follow the same marginals $q(\cdot, t)$ (petrol); each reverses the other



Eq. (7) and (8). $\Psi \hat{\Psi} = q$, so $\nabla \log \Psi - \nabla \log q = -\nabla \log \hat{\Psi}$: reversing the forward SDE with Anderson's formula gives exactly the backward SDE. A5.

- Set $\Psi \equiv 1$: the forward drift vanishes and $\hat{\Psi} = q$.
- Then (5) is exactly the pair (1), (2), and $\nabla \log \hat{\Psi} = \nabla \log p$ is the score.
- SB adds one nonlinear drift, $\nabla \log \Psi$, which lets the process end anywhere, not only at a Gaussian.



But solving (6) needs iterative projections over whole path measures: at 256×256, 6× slower and 3× the memory of a diffusion model. Figure 2 of the paper.

3 I^2SB

We want to learn the backward drift $\nabla \log \hat{\Psi}(X_t, t)$ of the nonlinear SDEs

$$\begin{aligned} dX_t &= [f_t + \beta_t \nabla \log \Psi(X_t, t)] dt + \sqrt{\beta_t} dW_t & X_0 &\sim p_{\mathcal{A}} \\ dX_t &= [f_t - \beta_t \nabla \log \hat{\Psi}(X_t, t)] dt + \sqrt{\beta_t} d\bar{W}_t & X_1 &\sim p_{\mathcal{B}} \end{aligned}$$

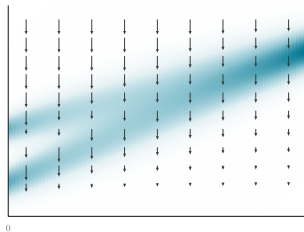
- ...but we cannot use score matching: it needs a conditional score in closed form, and here the forward process is nonlinear and unknown. ✗
- ...however (Theorem 3.1), $\nabla \log \hat{\Psi}$ and $\nabla \log \Psi$ are the score functions of **linear** SDEs:

$$dX_t = f_t dt + \sqrt{\beta_t} dW_t, \quad X_0 \sim \hat{\Psi}(\cdot, 0) \quad dX_t = f_t dt + \sqrt{\beta_t} d\bar{W}_t, \quad X_1 \sim \Psi(\cdot, 1)$$

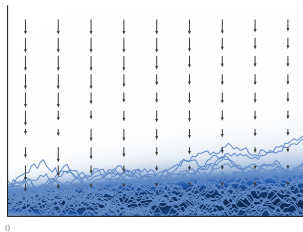
So close If we could sample $\hat{\Psi}(\cdot, 0)$, score matching on a linear SDE would learn $s_\theta \approx \nabla \log \hat{\Psi}$. But the boundary densities are fixed only through the couplings in (6): intractable.

Theorem 3.1: the drifts are scores

SB: bridge from p_A to p_B (nonlinear)



linear SDE from $\hat{\Psi}(\cdot, 0)$: density $\hat{\Psi}(\cdot, t)$



arrows on both: $\nabla \log \hat{\Psi}(x, t)$, the SB backward drift (left) is the score of the plain diffusion (right)

Theorem 3.1: the nonlinearity is absorbed into the starting density $\hat{\Psi}(\cdot, 0)$

Theorem 3.1, Eq. (9). $\nabla \log \hat{\Psi}$ is the score of $dX_t = f_t dt + \sqrt{\beta_t} dW_t$ from $X_0 \sim \hat{\Psi}(\cdot, 0)$; $\nabla \log \Psi$ that of the same SDE run back from $X_1 \sim \Psi(\cdot, 1)$. Proof: A6.

- Look at what score matching does, Eq. (3):

$$\theta^* = \operatorname{argmin}_{\theta} \mathbb{E}_t \mathbb{E}_{p_{\mathcal{A}}(x_0)} \mathbb{E}_{p_{t|0}(x_t|x_0)} \|s_{\theta}(t, x_t) - \nabla \log p_{t|0}(x_t | x_0)\|^2$$

For each x_0 : the score between δ_{x_0} and a Gaussian, then the average over $x_0 \sim p_{\mathcal{A}}$.

- So set $p_{\mathcal{A}} := \delta_{x_0}$. By Corollary 3.2 this **eliminates one of the couplings**:

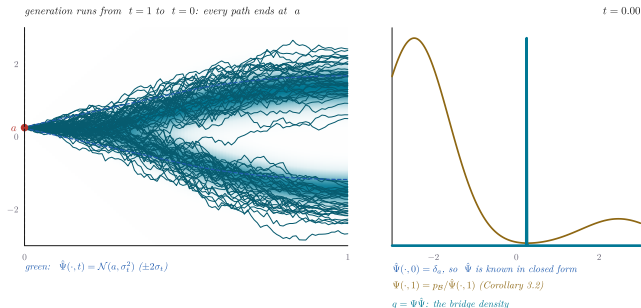
$$\hat{\Psi}(\cdot, 0) = \delta_{x_0},$$

$$\Psi(\cdot, 1) = \frac{p_{\mathcal{B}}}{\hat{\Psi}(\cdot, 1)}$$

- $\hat{\Psi}$ is now a known Gaussian, $\hat{\Psi}(\cdot, t) = \mathcal{N}(x_0, \sigma_t^2 I)$ for $f = 0$, and the conditional score

$$\nabla \log \hat{\Psi}(X_t, t | x_0) = -\frac{X_t - x_0}{\sigma_t^2} \text{ is an analytic regression target. } \checkmark$$

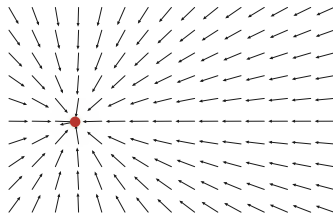
The trick The move that makes denoising score matching work: condition on the clean image, regress on a closed-form target. Proof: A7.



Corollary 3.2, Eq. (10). If $p_{\mathcal{A}} = \delta_a$, then $\hat{\Psi}(\cdot, 0) = \delta_a$ and $\Psi(\cdot, 1) = p_{\mathcal{B}} / \hat{\Psi}(\cdot, 1)$: the coupling at $t = 0$ is gone and $\hat{\Psi}$ is a known Gaussian.

$$\nabla \log \hat{\Psi}(x, t \mid X_0 = a) = \frac{a - x}{\sigma_t^2} \text{ and } \mathcal{N}(a, \sigma_t^2 I)$$

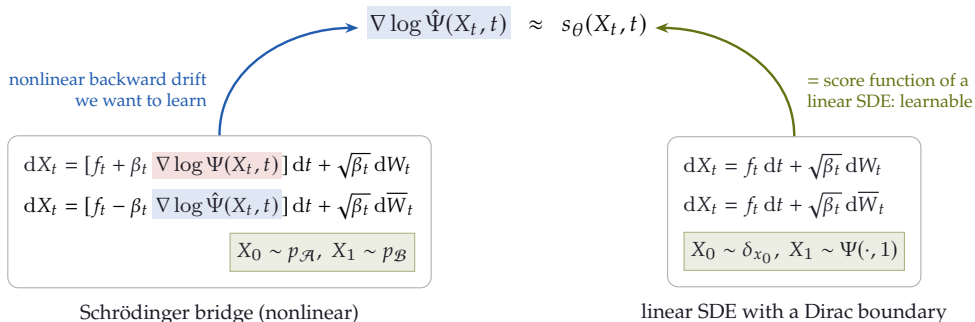
the bridge density $q(\cdot, t)$ between p_S and δ_a



$t = 0.040$

whatever p_S is, the drift always points at a : the backward process collapses onto it
(the paper's Figure 4, recomputed exactly)

With $p_{\mathcal{A}} = \delta_a$, the backward drift is $\nabla \log \hat{\Psi} = -(x - a)/\sigma_t^2$ everywhere: every degraded sample is pulled straight towards a .



The model is an ordinary score network, trained with the ordinary score-matching machinery. Only the data it sees change.

- Training data come in pairs:
 $p(X_0, X_1) = p_{\mathcal{A}}(X_0) p_{\mathcal{B}}(X_1 | X_0).$
- One tractable bridge per clean image X_0 , ending in its own degraded distribution $p_{\mathcal{B}}(\cdot | X_0).$
- The endpoint is no longer Gaussian, so the base drift is unnecessary: $f := 0.$

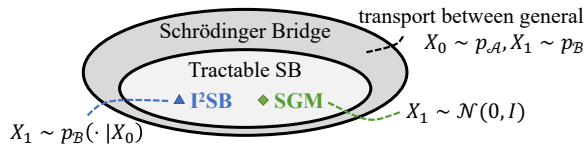
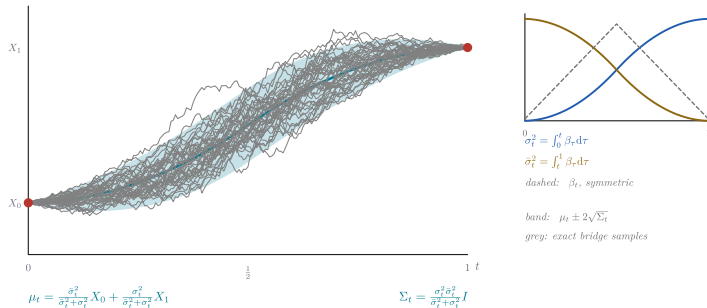


Figure 3 of the paper (Venn diagram).

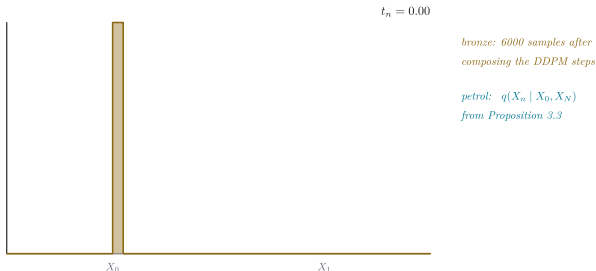
Where I^2SB sits A tractable class of SB that trains with the computational framework of diffusion models.

Proposition 3.3: given a pair (X_0, X_1) , the SB posterior is Gaussian



Eq. (11). Given the pair, every intermediate state is Gaussian in closed form: training needs no simulation of any SDE. Proof: A8, A9.

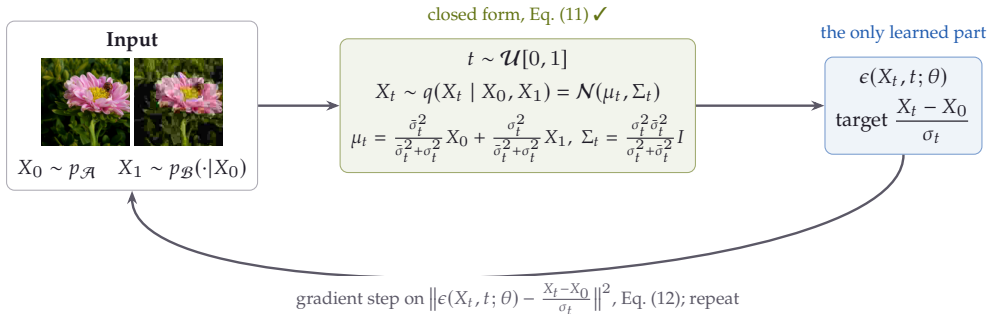
DDPM: $X_n \sim p(X_n | X_0, X_{n+1})$, step 12 of 12



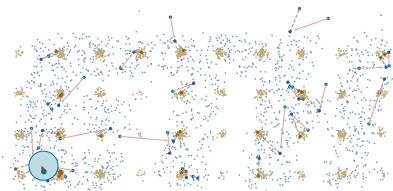
$$q(X_n | X_0, X_N) = \int \prod_{k=n}^{N-1} p(X_k | X_0, X_{k+1}) dX_{k+1}$$

they agree at every step: DDPM run from $X_N = X_1$ traces the SB bridge

Run DDPM from $X_N = X_1$ with the true X_0 , and its marginals are exactly those of (11). With a predicted $X_0^\epsilon \approx X_0$ the ordinary DDPM sampler follows the bridge. Induction: A9.



Algorithm 1. Identical to training a diffusion model, except that X_t interpolates between a clean and a degraded image instead of between a clean image and noise. A10.



no SDE is simulated: every X_t is drawn in closed form

Eq. (12). $\left\| \epsilon(X_t, t; \theta) - \frac{X_t - X_0}{\sigma_t} \right\|$ with $X_t \sim q(X_t \mid X_0, X_1)$: each training pair brings its own Gaussian bridge and its own closed-form target.

Algorithm 1 (training)

3: $t \sim \mathcal{U}([0, 1])$, $X_0 \sim p_A$, $X_1 \sim p_B(\cdot \mid X_0)$

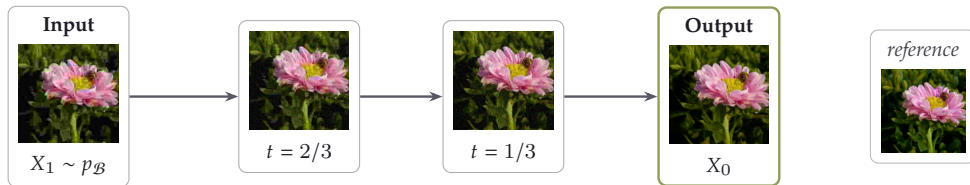
4: $X_t \sim q(X_t \mid X_0, X_1)$ by (11)

5: step on $\left\| \epsilon(X_t, t; \theta) - \frac{X_t - X_0}{\sigma_t} \right\|$ (12)

blue: clean X_0 , bronze: degraded X_1

petrol: X_t on the bridge (circle: $\sqrt{\Sigma_t}$)

red: the regression target points to X_0



each arrow: $X_0^\epsilon = X_n - \sigma_n \epsilon(X_n, t_n; \theta), \quad X_{n-1} \sim p(X_{n-1} \mid X_0^\epsilon, X_n)$

simulate $dX_t = -\beta_t s_\theta(X_t, t) dt + \sqrt{\beta_t} d\bar{W}_t$ with DDPM steps, Eq. (4)

Algorithm 2. Start at the degraded image, not at noise: each intermediate is a less degraded image. Images:
Figure 7 of the paper (JPEG).

$t = 0.000$, step 100 of 100



generation is a gradual, interpretable restoration of the degraded input

Algorithm 2 (generation)

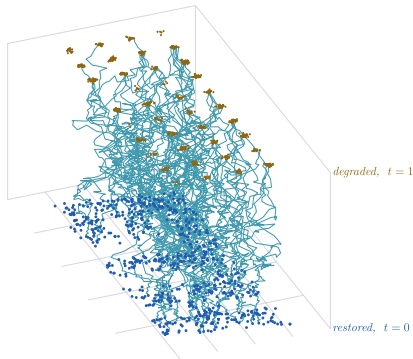
start: $X_N \sim p_S$ (the degraded grid, not noise)

each step: predict X_0^ϵ ,
then $X_{n-1} \sim p(X_{n-1} | X_0^\epsilon, X_n)$

bronze: the network's
current guess X_0^ϵ

blue: the bridge state X_t

Toy model: clean data are the ETH letters, degraded data snap each point to a coarse grid. The network is replaced by the exact minimiser of (12) over 6000 training pairs.



Algorithm 2 on the toy model, time running downwards: from degraded points at $t = 1$ to the ETH letters at $t = 0$.

Analytic ✓

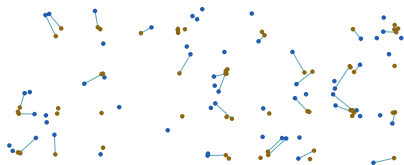
- the bridge posterior $q(X_t | X_0, X_1)$, Eq. (11)
- $\hat{\Psi}(\cdot, t | X_0) = \mathcal{N}(X_0, \sigma_t^2 I)$ and the target $\frac{X_t - X_0}{\sigma_t}$
- every sampling step $p(X_{n-1} | X_0^\epsilon, X_n)$

Learned

one network $\epsilon(X_t, t; \theta) \approx \sigma_t \nabla \log \hat{\Psi}(X_t, t)$ (up to sign)

Never needed the forward drift $\nabla \log \Psi$, iterative fitting over path measures, or simulating an SDE during training

As for diffusion models: the Gaussian bridge replaces the noising kernel, and $\nabla \log \hat{\Psi}$ replaces the score.



blue: clean X_0 , bronze: degraded X_1 ; the paths straighten into the optimal transport displacement

diffusion scaled by $\lambda = 0.000$

Proposition 3.4: as $\beta_t \rightarrow 0$

$$dX_t = \frac{\beta_t}{\sigma_t^2} (X_t - X_0) dt$$

the OT-ODE; its solution is μ_t

Corollary 3.5: constant β

$$\mu_t = (1 - t)X_0 + tX_1,$$

straight lines at constant speed

Proposition 3.4, Corollary 3.5, Eq. (13). As $\beta_t \rightarrow 0$: $dX_t = \frac{\beta_t}{\sigma_t^2} (X_t - X_0) dt$, solved by μ_t ; for constant β , $\mu_t = (1 - t)X_0 + tX_1$, the OT displacement. A11.

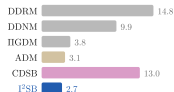
Model	$p(X_0)$	$p(X_1)$	$\nabla \log \Psi$	$\nabla \log \hat{\Psi}$
(C)SGM	$p_{\mathcal{A}}$	$\mathcal{N}(0, I)$	0	tractable
I²SB	$p_{\mathcal{A}}$	$p_{\mathcal{B}}(\cdot X_0)$	intractable	tractable
SB	$p_{\mathcal{A}}$	$p_{\mathcal{B}}$	intractable	intractable

Table 1 of the paper. I²SB needs paired data, but never needs the forward drift: only $\nabla \log \hat{\Psi}$ is learned.

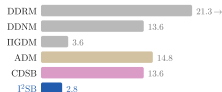
4 *Results*

FID on ImageNet 256×256 (lower is better; Tables 2–5 of the paper; axis cut at 20)

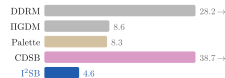
4x super-resolution, Pool



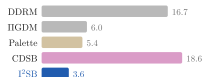
4x super-resolution, Bicubic



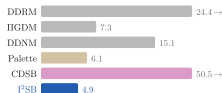
JPEG restoration, $QF = 5$



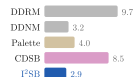
JPEG restoration, $QF = 10$



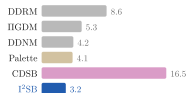
Inpainting, Center 128×128



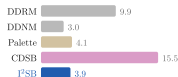
Inpainting, Freeform 10-20%



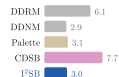
Inpainting, Freeform 20-30%



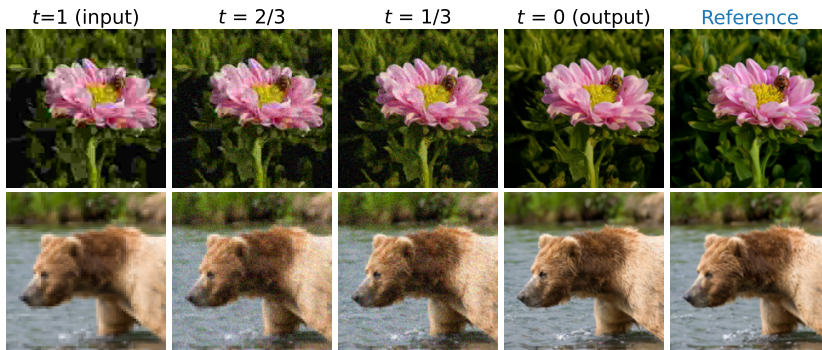
Deblurring, Uniform



Deblurring, Gaussian



■ I²SB
 ■ conditional diffusion (Palette, ADM)
 ■ inverse models, need the corruption operator
 ■ standard SB (CDSB)



JPEG restoration (top) and 4× super-resolution (bottom): each intermediate image is a less degraded version of the input. Figure 7 of the paper.

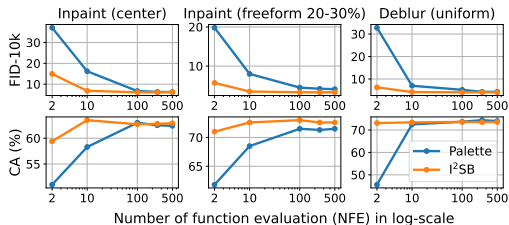


Figure 8 of the paper.

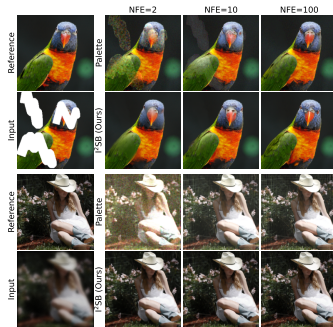
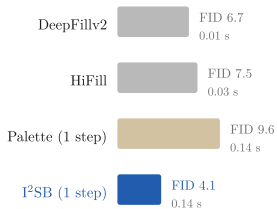


Figure 9 of the paper.

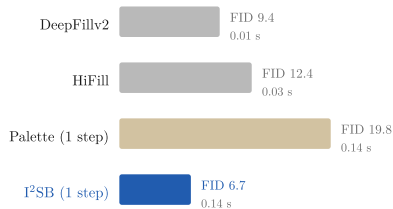
Starting from the degraded image, I²SB loses little as the number of network evaluations drops; on inpainting 2–10 suffice where Palette needs about 100.

one network call: I^2SB against inpainting GANs and Palette (Table 8 of the paper; V100 time per image)

Inpainting, Freeform 10-20%

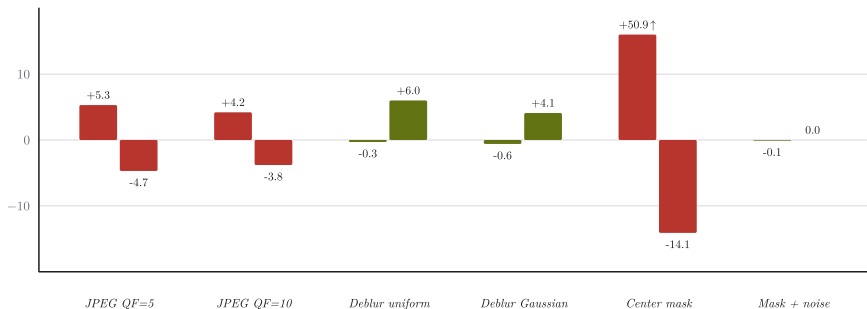


Inpainting, Freeform 20-30%



starting from the degraded image, a single step of I^2SB already beats GANs built for one-shot inpainting

OT-ODE instead of the bridge: change in FID (left bars) and CA (right bars); green = better



deterministic paths help where the inverse is nearly unique (deblurring) and hurt where much is lost (JPEG, empty masks)

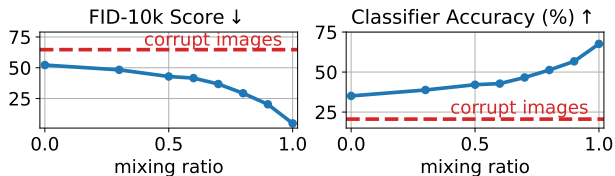
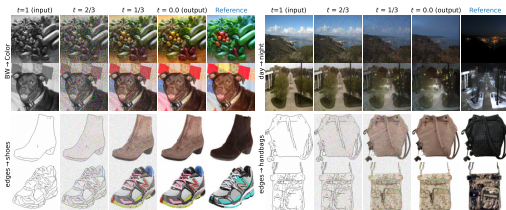


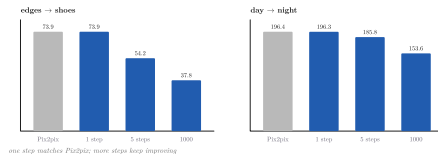
Figure 10 of the paper (JPEG, QF = 5).

- Mixing the training proposal between the linear SDE (9a) and the bridge posterior (11).
- Both metrics improve steadily as the proposal approaches (11): train where generation actually goes.



Edges $\rightarrow$ shoes, day $\rightarrow$ night: generation from $t = 1$ to $t = 0$.

general image-to-image translation: FID (Table 7 of the paper)



5 *Summary*

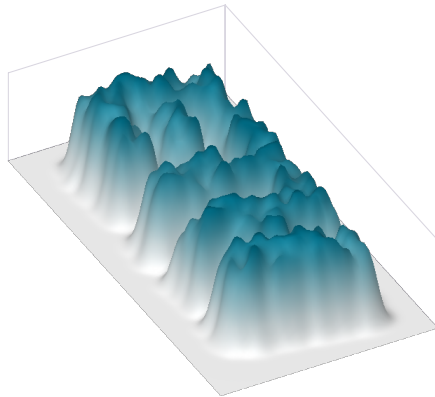
- Diffusion models: everything is analytic except the score, learned by regression on closed-form conditional scores.
- Schrödinger bridges reach any p_B with a nonlinear drift, at the price of two potentials coupled through their boundaries.
- **Theorem 3.1:** the bridge drifts are scores of linear SDEs.
- **Corollary 3.2:** a Dirac delta at the clean image removes one coupling: the score-matching trick.
- **Proposition 3.3:** the bridge between a pair is Gaussian, and DDPM follows it.
- **Proposition 3.4:** without noise, optimal transport.

In practice Train like a diffusion model, start from the degraded image instead of noise, sample with DDPM. More interpretable, and far fewer steps.

Open question from the original talk: what does this offer over bilinear upsampling, CNNs, ResNets and GANs?

I²SB: the bridge density $q(\cdot, t)$, from the degraded data at $t = 1$ to the clean data at $t = 0$

$t = 0.00$



Thank you.

Johannes Herter

6 *Appendix*

A1. The backward kernel by Taylor expansion

