On the Accuracy of Diffusion Models in Bayesian Image Inverse Problems: A Gaussian Case Study

Émile Pierret[a], supervised by Bruno Galerne[a,b]

Journées MISTIC, Lyon, 26 mai

[a] Institut Denis Poisson – Université d'Orléans, Université de Tours, CNRS
[b] Institut universitaire de France (IUF)

Introduction

$$\boldsymbol{v} = \boldsymbol{A}\boldsymbol{x} + \sigma \boldsymbol{n}, \quad \boldsymbol{n} \sim \mathcal{N}_0 \tag{1}$$

where $\boldsymbol{A}$ is an inpainting, super-resolution, or blur operator.

Clean image | Super-resolution $(\times 4)$ | Inpainting | Gaussian blur | Motion blur

Inverse problems

$$\boldsymbol{v} = \boldsymbol{A}\boldsymbol{x} + \sigma\boldsymbol{n}, \quad \boldsymbol{n} \sim \mathcal{N}_0 \tag{1}$$

where $\boldsymbol{A}$ is an inpainting, super-resolution, or blur operator.

Clean image | Super-resolution $(\times 4)$ | Inpainting | Gaussian blur | Motion blur

$\Longrightarrow$ One popular solution: diffusion models.

Ground truth	Degraded image	DDRM [Kawar et al., 2022]	DPS [Chung et al., 2023]	ΠGDM [Song et al., 2023]

Ground truth	Degraded image	DDRM [Kawar et al., 2022]	DPS [Chung et al., 2023]	ΠGDM [Song et al., 2023]

Ground truth	Degraded image	DDRM [Kawar et al., 2022]	DPS [Chung et al., 2023]	ΠGDM [Song et al., 2023]

Ground truth	Degraded image	DDRM [Kawar et al., 2022]	DPS [Chung et al., 2023]	ΠGDM [Song et al., 2023]

One pending question

$$\boldsymbol{v} = \boldsymbol{A}\boldsymbol{x} + \sigma \boldsymbol{n}, \quad \boldsymbol{n} \sim \mathcal{N}_0 \tag{2}$$

where $\boldsymbol{A}$ inpainting, SR or blur operator.

One pending question

$$\boldsymbol{v} = \boldsymbol{A}\boldsymbol{x} + \sigma \boldsymbol{n}, \quad \boldsymbol{n} \sim \mathcal{N}_0 \tag{2}$$

where $\boldsymbol{A}$ inpainting, SR or blur operator.

Aim: Sample from $p(\boldsymbol{x} \mid \boldsymbol{v})$

One pending question

$$v = Ax + \sigma n, \quad n \sim \mathcal{N}_0 \tag{2}$$

where A inpainting, SR or blur operator.

Aim: Sample from $p(x \mid v)$

Question: To what extent do diffusion models allow sampling from the target posterior distribution?

One pending question

$$\boldsymbol{v} = \boldsymbol{A}\boldsymbol{x} + \sigma \boldsymbol{n}, \quad \boldsymbol{n} \sim \mathcal{N}_0 \tag{2}$$

where $\boldsymbol{A}$ inpainting, SR or blur operator.

Aim: Sample from $p(\boldsymbol{x} \mid \boldsymbol{v})$

Question: To what extent do diffusion models allow sampling from the target posterior distribution?

Idea: Observe what happens for Gaussian image distributions, for which calculations are tractable.

Diffusion models for image generation

Discrete DDPM [Ho et al., 2020][1]

Forward process

$$\boldsymbol{x}_t = \sqrt{1-\beta_t}\boldsymbol{x}_{t-1} + \sqrt{\beta_t}\boldsymbol{z}_t, \quad 1 \leqslant t \leqslant T, \quad \boldsymbol{z}_t \sim \mathcal{N}_0, \quad \boldsymbol{x}_0 \sim p_{\mathrm{data}}, \tag{3}$$

Ones can write

$$\boldsymbol{x}_t = \sqrt{\overline{\alpha}_t}\boldsymbol{x}_0 + \sqrt{1-\overline{\alpha}_t}\varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}_0 \tag{4}$$

[1] Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020*

Discrete DDPM [Ho et al., 2020][1]

Forward process

$$\boldsymbol{x}_t = \sqrt{1-\beta_t}\boldsymbol{x}_{t-1} + \sqrt{\beta_t}\boldsymbol{z}_t, \quad 1 \leqslant t \leqslant T, \quad \boldsymbol{z}_t \sim \mathcal{N}_0, \quad \boldsymbol{x}_0 \sim p_{\text{data}}, \tag{3}$$

Ones can write

$$\boldsymbol{x}_t = \sqrt{\overline{\alpha}_t}\boldsymbol{x}_0 + \sqrt{1-\overline{\alpha}_t}\varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}_0 \tag{4}$$

Backward process

By learning $\boldsymbol{\varepsilon}_\theta$ such that $\boldsymbol{\varepsilon}_\theta(\boldsymbol{x}_t, t) \approx \varepsilon_t$,

$$\boldsymbol{y}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{y}_t - \frac{\beta_t}{\sqrt{1-\overline{\alpha}_t}}\boldsymbol{\varepsilon}_\theta(\boldsymbol{y}_t, t)\right) + \sqrt{\tilde{\beta}_t}\boldsymbol{z}_t, \quad \boldsymbol{z}_t \sim \mathcal{N}_0. \tag{5}$$

[1] Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020*

Forward process

$$\boldsymbol{x}_t = \sqrt{1-\beta_t}\boldsymbol{x}_{t-1} + \sqrt{\beta_t}\boldsymbol{z}_t, \quad 1 \leqslant t \leqslant T, \quad \boldsymbol{z}_t \sim \mathcal{N}_0, \quad \boldsymbol{x}_0 \sim p_{\text{data}}, \tag{3}$$

Ones can write

$$\boldsymbol{x}_t = \sqrt{\overline{\alpha}_t}\boldsymbol{x}_0 + \sqrt{1-\overline{\alpha}_t}\varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}_0 \tag{4}$$

Backward process

By learning $\boldsymbol{\varepsilon}_\theta$ such that $\boldsymbol{\varepsilon}_\theta(\boldsymbol{x}_t, t) \approx \varepsilon_t$,

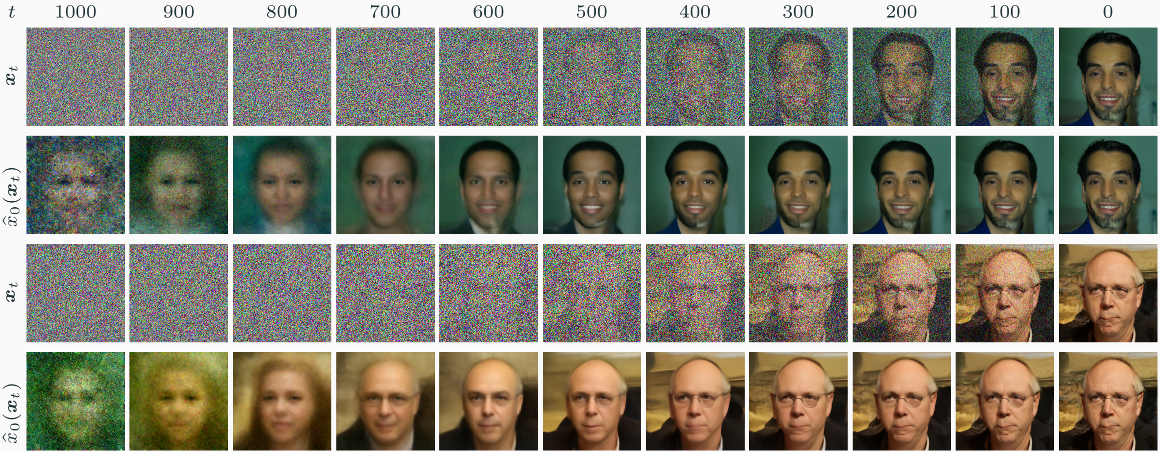
$$\boldsymbol{y}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{y}_t - \frac{\beta_t}{\sqrt{1-\overline{\alpha}_t}}\boldsymbol{\varepsilon}_\theta(\boldsymbol{y}_t, t)\right) + \sqrt{\tilde{\beta}_t}\boldsymbol{z}_t, \quad \boldsymbol{z}_t \sim \mathcal{N}_0. \tag{5}$$

By denoting p_t the marginals of the forward process and learning $\boldsymbol{s}_\theta(\boldsymbol{x}, t) \approx \nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x}_t)$,

$$\boldsymbol{y}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{y}_t + \beta_t s_{\theta^\star}(\boldsymbol{y}_t, t)\right) + \sigma_t \boldsymbol{z}_t, \boldsymbol{z}_t \sim \mathcal{N}_0, 1 \leqslant t \leqslant T, \boldsymbol{z}_t \sim \mathcal{N}_0, \boldsymbol{y}_T \sim \mathcal{N}_0. \tag{6}$$

[1] Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020*

Generation examples



Diffusion models for inverse problems

And for solving image problems ?

$$\boldsymbol{v} = \boldsymbol{A}\boldsymbol{x}_0 + \sigma\boldsymbol{n}, \quad \boldsymbol{x}_0 \sim p_0, \boldsymbol{n} \sim \mathcal{N}_0 \tag{7}$$

Aim: Sampling $p_0(\cdot \mid \boldsymbol{v})$

And for solving image problems ?

$$\boldsymbol{v} = \boldsymbol{A}\boldsymbol{x}_0 + \sigma \boldsymbol{n}, \quad \boldsymbol{x}_0 \sim p_0, \boldsymbol{n} \sim \mathcal{N}_0 \tag{7}$$

Aim: Sampling $p_0(\cdot \mid \boldsymbol{v})$

$\Longrightarrow$ **conditional forward process**

$$\tilde{\boldsymbol{x}}_t = \sqrt{1-\beta_t}\tilde{\boldsymbol{x}}_{t-1} + \sqrt{\beta_t}\boldsymbol{z}_t, \quad 1 \leqslant t \leqslant T, \quad \boldsymbol{z}_t \sim \mathcal{N}_0, \quad \tilde{\boldsymbol{x}}_0 \sim p_{\text{data}}(\cdot \mid \boldsymbol{v}), \tag{8}$$

And for solving image problems ?

$$\boldsymbol{v} = \boldsymbol{A}\boldsymbol{x}_0 + \sigma \boldsymbol{n}, \quad \boldsymbol{x}_0 \sim p_0, \boldsymbol{n} \sim \mathcal{N}_0 \tag{7}$$

Aim: Sampling $p_0(\cdot \mid \boldsymbol{v})$

$\Longrightarrow$ **conditional forward process**

$$\tilde{\boldsymbol{x}}_t = \sqrt{1-\beta_t}\tilde{\boldsymbol{x}}_{t-1} + \sqrt{\beta_t}\boldsymbol{z}_t, \quad 1 \leqslant t \leqslant T, \quad \boldsymbol{z}_t \sim \mathcal{N}_0, \quad \tilde{\boldsymbol{x}}_0 \sim p_{\text{data}}(\cdot \mid \boldsymbol{v}), \tag{8}$$

$\Longrightarrow$ **conditional backward process**

$$\tilde{\boldsymbol{y}}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\tilde{\boldsymbol{y}}_t + \beta_t \nabla \log \tilde{p}_t(\tilde{\boldsymbol{y}}_t)\right) + \sigma_t \boldsymbol{z}_t, \quad \boldsymbol{z}_t \sim \mathcal{N}_0, 1 \leqslant t \leqslant T, \boldsymbol{z}_t \sim \mathcal{N}_0, \tilde{\boldsymbol{y}}_T \sim \mathcal{N}_0 \tag{9}$$

Bayes theorem

Our point of interest is $\nabla \log \tilde{p}_t(\boldsymbol{x}_t) = \nabla \log p_t(\boldsymbol{x}_t \mid \boldsymbol{v})$.

Bayes theorem

Our point of interest is $\nabla \log \tilde{p}_t(\boldsymbol{x}_t) = \nabla \log p_t(\boldsymbol{x}_t \mid \boldsymbol{v})$. By Bayes' formula,

$$\nabla_{\boldsymbol{x}} \log \tilde{p}_t(\boldsymbol{x}_t) = \nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x}_t) + \nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{v} \mid \boldsymbol{x}_t), \tag{10}$$

Bayes theorem

Our point of interest is $\nabla \log \tilde{p}_t(\boldsymbol{x}_t) = \nabla \log p_t(\boldsymbol{x}_t \mid \boldsymbol{v})$. By Bayes' formula,

$$\nabla_{\boldsymbol{x}} \log \tilde{p}_t(\boldsymbol{x}_t) = \nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x}_t) + \nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{v} \mid \boldsymbol{x}_t), \tag{10}$$

Algorithm 5
Unconditional DDPM backward process

$\boldsymbol{y}_T \sim \mathcal{N}_0$
for t=T **to** 1 **do**
    $\boldsymbol{z}_t \sim \mathcal{N}_0$
    $\boldsymbol{y}_{t-1} = \frac{1}{\sqrt{\alpha_t}}(\boldsymbol{y}_t + \beta_t s_{\theta^\star}(\boldsymbol{y}_t, t)) + \sigma_t \boldsymbol{z}_t$
end for

Algorithm 6 Conditional DDPM backward process

$\boldsymbol{y}_T \sim \mathcal{N}_0$
for $t = T$ **to** 1 **do**
    $\widehat{\boldsymbol{x}}_0(\boldsymbol{x}_t) = \frac{1}{\sqrt{\overline{\alpha}_t}}(\boldsymbol{x}_t + (1-\overline{\alpha}_t) s_{\theta^\star}(\boldsymbol{y}_t, t))$
    $\tilde{s}_\theta(\boldsymbol{y}_t, t) = s_{\theta^\star}(\boldsymbol{y}_t, t) + \nabla \log p_t(\boldsymbol{x} \mid \boldsymbol{v})$
    $\boldsymbol{z}_t \sim \mathcal{N}_0$
    $\boldsymbol{y}_{t-1} = \frac{1}{\sqrt{\alpha_t}}(\boldsymbol{y}_t + \beta_t \tilde{s}_{\theta^\star}(\boldsymbol{y}_t, t)) + \sigma_t \boldsymbol{z}_t$
end for

Description of two algorithms from the literature

Mean of $p_t(v \mid x_t)$

$$\boldsymbol{v} = \boldsymbol{A}\boldsymbol{x}_0 + \sigma \boldsymbol{n}, \quad \boldsymbol{x}_0 \sim p_0, \boldsymbol{n} \sim \mathcal{N}_0 \tag{11}$$

$$\nabla_{\boldsymbol{x}} \log \tilde{p}_t(\boldsymbol{x}_t) = \nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x}_t) + \nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{v} \mid \boldsymbol{x}_t), \tag{12}$$

Some assumptions lead to "$p_t(\boldsymbol{v} \mid \boldsymbol{x}_t)$ is Gaussian".

Mean of $p_t(\boldsymbol{v} \mid \boldsymbol{x}_t)$

$$\boldsymbol{v} = \boldsymbol{A}\boldsymbol{x}_0 + \sigma\boldsymbol{n}, \quad \boldsymbol{x}_0 \sim p_0, \boldsymbol{n} \sim \mathcal{N}_0 \tag{11}$$

$$\nabla_{\boldsymbol{x}} \log \tilde{p}_t(\boldsymbol{x}_t) = \nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x}_t) + \nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{v} \mid \boldsymbol{x}_t), \tag{12}$$

Some assumptions lead to "$p_t(\boldsymbol{v} \mid \boldsymbol{x}_t)$ is Gaussian". Let note that an approximation of $\mathbb{E}(\boldsymbol{v} \mid \boldsymbol{x}_t)$ is known.

By Tweedie's formula, that is,

$$\widehat{\boldsymbol{x}}_0(\boldsymbol{x}_t) := \mathbb{E}\left[\boldsymbol{x}_0 \mid \boldsymbol{x}_t\right] = \frac{1}{\sqrt{\overline{\alpha}_t}}\left(\boldsymbol{x}_t + (1 - \overline{\alpha}_t)\nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x}_t)\right). \tag{13}$$

Mean of $p_t(\boldsymbol{v} \mid \boldsymbol{x}_t)$

$$\boldsymbol{v} = \boldsymbol{A}\boldsymbol{x}_0 + \sigma\boldsymbol{n}, \quad \boldsymbol{x}_0 \sim p_0, \boldsymbol{n} \sim \mathcal{N}_0 \tag{11}$$

$$\nabla_{\boldsymbol{x}} \log \tilde{p}_t(\boldsymbol{x}_t) = \nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x}_t) + \nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{v} \mid \boldsymbol{x}_t), \tag{12}$$

Some assumptions lead to "$p_t(\boldsymbol{v} \mid \boldsymbol{x}_t)$ is Gaussian". Let note that an approximation of $\mathbb{E}(\boldsymbol{v} \mid \boldsymbol{x}_t)$ is known.

By Tweedie's formula, that is,

$$\widehat{\boldsymbol{x}}_0(\boldsymbol{x}_t) := \mathbb{E}\left[\boldsymbol{x}_0 \mid \boldsymbol{x}_t\right] = \frac{1}{\sqrt{\overline{\alpha}_t}}\left(\boldsymbol{x}_t + (1-\overline{\alpha}_t)\nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x}_t)\right). \tag{13}$$

As a consequence, $\mathbb{E}\left[\boldsymbol{v} \mid \boldsymbol{x}_t\right]$ is given by

$$\mathbb{E}\left[\boldsymbol{v} \mid \boldsymbol{x}_t\right] = \boldsymbol{A}\widehat{\boldsymbol{x}}_0(\boldsymbol{x}_t) = \frac{1}{\sqrt{\overline{\alpha}_t}}\boldsymbol{A}\left(\boldsymbol{x}_t + (1-\overline{\alpha}_t)\nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x}_t)\right). \tag{14}$$

Mean of $p_t(\boldsymbol{v} \mid \boldsymbol{x}_t)$

$$\boldsymbol{v} = \boldsymbol{A}\boldsymbol{x}_0 + \sigma \boldsymbol{n}, \quad \boldsymbol{x}_0 \sim p_0, \boldsymbol{n} \sim \mathcal{N}_0 \tag{11}$$

$$\nabla_{\boldsymbol{x}} \log \tilde{p}_t(\boldsymbol{x}_t) = \nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x}_t) + \nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{v} \mid \boldsymbol{x}_t), \tag{12}$$

Some assumptions lead to "$p_t(\boldsymbol{v} \mid \boldsymbol{x}_t)$ is Gaussian". Let note that an approximation of $\mathbb{E}(\boldsymbol{v} \mid \boldsymbol{x}_t)$ is known.

By Tweedie's formula, that is,

$$\widehat{\boldsymbol{x}}_0(\boldsymbol{x}_t) := \mathbb{E}\left[\boldsymbol{x}_0 \mid \boldsymbol{x}_t\right] = \frac{1}{\sqrt{\overline{\alpha}_t}}\left(\boldsymbol{x}_t + (1 - \overline{\alpha}_t)\nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x}_t)\right). \tag{13}$$

As a consequence, $\mathbb{E}\left[\boldsymbol{v} \mid \boldsymbol{x}_t\right]$ is given by

$$\mathbb{E}\left[\boldsymbol{v} \mid \boldsymbol{x}_t\right] = \boldsymbol{A}\widehat{\boldsymbol{x}}_0(\boldsymbol{x}_t) = \frac{1}{\sqrt{\overline{\alpha}_t}}\boldsymbol{A}\left(\boldsymbol{x}_t + (1 - \overline{\alpha}_t)\nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x}_t)\right). \tag{14}$$

Finally,

$$p(\boldsymbol{x}_t \mid \boldsymbol{v}) = \mathcal{N}(\boldsymbol{A}\widehat{\boldsymbol{x}}_0(\boldsymbol{x}_t), \boldsymbol{C}_{\boldsymbol{v}|t}) \tag{15}$$

with $\boldsymbol{C}_{\boldsymbol{v}|t}$ to fix.

Covariance of $p_t(v \mid x)$

We denote it $\boldsymbol{C}_{\boldsymbol{v}|t}$.

- Denoising Posterior Sampling (DPS) algorithm [Chung et al., 2023][2]

$$\log p(\boldsymbol{v} \mid \boldsymbol{x}_t) \approx \log p(\boldsymbol{v} \mid \boldsymbol{x}_0 = \widehat{x}_0(\boldsymbol{x}_t)) \tag{16}$$

$$p(\boldsymbol{v} \mid \boldsymbol{x}_0) = \mathcal{N}\left(\boldsymbol{A}\boldsymbol{x}_0, \sigma^2 \boldsymbol{I}\right). \tag{17}$$

$$\nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{v} \mid \boldsymbol{x}_0 = \widehat{x}_0(\boldsymbol{x}_t)) = -\frac{1}{2\sigma^2} \nabla_{\boldsymbol{x}_t} \left\| \boldsymbol{v} - \boldsymbol{A}\widehat{x}_0(\boldsymbol{x}_t) \right\|^2. \tag{18}$$

[2] Chung, H., Kim, J., Mccann, M. T., Klasky, M. L., & Ye, J. C. (2023). Diffusion posterior sampling for general noisy inverse problems. *The Eleventh International Conference on Learning Representations*

Covariance of $p_t(v \mid x)$

We denote it $\boldsymbol{C}_{\boldsymbol{v}|t}$.

- Denoising Posterior Sampling (DPS) algorithm [Chung et al., 2023][2]

$$\log p(\boldsymbol{v} \mid \boldsymbol{x}_t) \approx \log p(\boldsymbol{v} \mid \boldsymbol{x}_0 = \widehat{x}_0(\boldsymbol{x}_t)) \tag{16}$$

$$p(\boldsymbol{v} \mid \boldsymbol{x}_0) = \mathcal{N}\left(\boldsymbol{A}\boldsymbol{x}_0, \sigma^2 \boldsymbol{I}\right). \tag{17}$$

$$\nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{v} \mid \boldsymbol{x}_0 = \widehat{x}_0(\boldsymbol{x}_t)) = -\frac{1}{2\sigma^2} \nabla_{\boldsymbol{x}_t} \left\| \boldsymbol{v} - \boldsymbol{A}\widehat{x}_0(\boldsymbol{x}_t) \right\|^2. \tag{18}$$

In practice,

$$\nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{v} \mid \boldsymbol{x}_0 = \widehat{x}_0(\boldsymbol{x}_t)) \approx -\frac{\alpha_{\text{DPS}}}{2\sigma^2} \nabla_{\boldsymbol{x}_t} \left\| \boldsymbol{v} - \boldsymbol{A}\widehat{x}_0(\boldsymbol{x}_t) \right\|^2. \tag{19}$$

$\Longrightarrow \boldsymbol{C}_{\boldsymbol{v}|t}^{\text{DPS}} = \sigma^2 \boldsymbol{I}$

[2] Chung, H., Kim, J., Mccann, M. T., Klasky, M. L., & Ye, J. C. (2023). Diffusion posterior sampling for general noisy inverse problems. *The Eleventh International Conference on Learning Representations*

Covariance of $p_t(v \mid x)$

We denote it $\boldsymbol{C}_{\boldsymbol{v}|t}$.

- Pseudo-Guided Diffusion model (ΠGDM) algorithm [Song et al., 2023][3]

$$p(\boldsymbol{x}_0 \mid \boldsymbol{x}_t) \approx \mathcal{N}\left(\widehat{x}_0(\boldsymbol{x}_t), r_t^2 \boldsymbol{I}\right). \tag{20}$$

  Consequently,

$$p(\boldsymbol{v} \mid \boldsymbol{x}_t) \approx \mathcal{N}\left(\boldsymbol{A}\widehat{x}_0(\boldsymbol{x}_t), r_t^2 \boldsymbol{A}\boldsymbol{A}^T + \sigma^2 \boldsymbol{I}\right) \tag{21}$$

[3] Song, J., Vahdat, A., Mardani, M., & Kautz, J. (2023). Pseudoinverse-guided diffusion models for inverse problems. *International Conference on Learning Representations*

Covariance of $p_t(v \mid x)$

We denote it $\boldsymbol{C}_{\boldsymbol{v}|t}$.

- Pseudo-Guided Diffusion model (ΠGDM) algorithm [Song et al., 2023][3]

$$p(\boldsymbol{x}_0 \mid \boldsymbol{x}_t) \approx \mathcal{N}\left(\widehat{x}_0(\boldsymbol{x}_t), r_t^2 \boldsymbol{I}\right). \tag{20}$$

Consequently,

$$p(\boldsymbol{v} \mid \boldsymbol{x}_t) \approx \mathcal{N}\left(\boldsymbol{A}\widehat{x}_0(\boldsymbol{x}_t), r_t^2 \boldsymbol{A}\boldsymbol{A}^T + \sigma^2 \boldsymbol{I}\right) \tag{21}$$

$\Longrightarrow \boldsymbol{C}_{\boldsymbol{v}|t}^{\Pi\text{GDM}} = r_t^2 \boldsymbol{A}\boldsymbol{A}^T + \sigma^2 \boldsymbol{I}$

[3] Song, J., Vahdat, A., Mardani, M., & Kautz, J. (2023). Pseudoinverse-guided diffusion models for inverse problems. *International Conference on Learning Representations*

Under Gaussian assumption

By adding the assumption: $\boldsymbol{x}_0$ follows a Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Under Gaussian assumption

By adding the assumption: $\boldsymbol{x}_0$ follows a Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

$$p(\boldsymbol{v} \mid \boldsymbol{x}_t) = \mathcal{N}\left(\boldsymbol{A}\widehat{\boldsymbol{x}}_0(\boldsymbol{x}_t), (1-\overline{\alpha}_t)\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{\Sigma}_t^{-1}\boldsymbol{A}^T + \sigma^2\boldsymbol{I}\right), \tag{22}$$

$$\text{with } \widehat{\boldsymbol{x}}_0(\boldsymbol{x}_t) = \boldsymbol{\mu} + \sqrt{\overline{\alpha}_t}\boldsymbol{\Sigma}\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{x}_t - \sqrt{\overline{\alpha}_t}\boldsymbol{\mu}). \tag{23}$$

We call this setting "Conditional Gaussian Diffusion Model" (CGDM) $\Longrightarrow \boldsymbol{C}_{\boldsymbol{v}|t}^{\mathsf{CGDM}} = (1-\overline{\alpha}_t)\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{\Sigma}_t^{-1}\boldsymbol{A}^T + \sigma^2\boldsymbol{I}$

Comparison of the choice made by different algorithms

	$\boldsymbol{C}_{\boldsymbol{v}\|t}$
DPS Chung et al., 2023	$\frac{\sigma^2}{\alpha_{\text{DPS}}}\boldsymbol{I}$
ΠGDM Song et al., 2023	$(1-\overline{\alpha}_t)\boldsymbol{A}\boldsymbol{A}^T + \sigma^2\boldsymbol{I}$
CGDM	$(1-\overline{\alpha}_t)\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{\Sigma}_t^{-1}\boldsymbol{A}^T + \sigma^2\boldsymbol{I}, \quad \boldsymbol{\Sigma}_t = \overline{\alpha}_t\boldsymbol{\Sigma} + (1-\overline{\alpha}_t)\boldsymbol{I}$

Comparison of the choice made by different algorithms

	$\boldsymbol{C}_{\boldsymbol{v}\|t}$
DPS Chung et al., 2023	$\frac{\sigma^2}{\alpha_{\text{DPS}}}\boldsymbol{I}$
ΠGDM Song et al., 2023	$(1-\overline{\alpha}_t)\boldsymbol{A}\boldsymbol{A}^T + \sigma^2\boldsymbol{I}$
CGDM	$(1-\overline{\alpha}_t)\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{\Sigma}_t^{-1}\boldsymbol{A}^T + \sigma^2\boldsymbol{I}, \quad \boldsymbol{\Sigma}_t = \overline{\alpha}_t\boldsymbol{\Sigma} + (1-\overline{\alpha}_t)\boldsymbol{I}$

Algorithm 7
Unconditional DDPM backward process

$\boldsymbol{y}_T \sim \mathcal{N}_0$
for t=T **to** 1 **do**
  $\boldsymbol{z}_t \sim \mathcal{N}_0$
  $\boldsymbol{y}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{y}_t + \beta_t s_{\theta^\star}(\boldsymbol{y}_t, t)\right) + \sigma_t \boldsymbol{z}_t$
end for

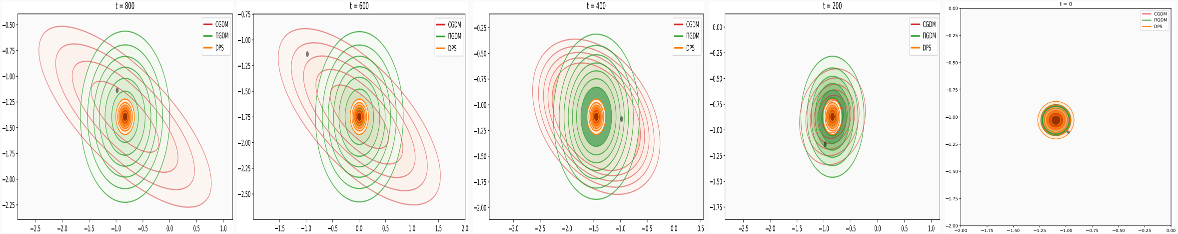
Algorithm 8 Conditional DDPM backward process

$\boldsymbol{y}_T \sim \mathcal{N}_0$
for $t = T$ **to** 1 **do**
  $\widehat{\boldsymbol{x}}_0(\boldsymbol{x}_t) = \frac{1}{\sqrt{\overline{\alpha}_t}}\left(\boldsymbol{x}_t + (1-\overline{\alpha}_t)s_{\theta^\star}(\boldsymbol{y}_t, t)\right)$
  $\tilde{s}_\theta(\boldsymbol{y}_t, t) = s_{\theta^\star}(\boldsymbol{y}_t, t) - \frac{1}{2}\nabla_{\boldsymbol{x}_t}\left\|\boldsymbol{v} - \boldsymbol{A}\widehat{\boldsymbol{x}}_0(\boldsymbol{x}_t)\right\|^2_{\boldsymbol{C}^{-1}_{\boldsymbol{v}|t}}$
  $\boldsymbol{z}_t \sim \mathcal{N}_0$
  $\boldsymbol{y}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{y}_t + \beta_t \tilde{s}_{\theta^\star}(\boldsymbol{y}_t, t)\right) + \sigma_t \boldsymbol{z}_t$
end for

Comparison of the choice made by different algorithms

	$\boldsymbol{C}_{\boldsymbol{v}\|t}$
DPS Chung et al., 2023	$\frac{\sigma^2}{\alpha_{\text{DPS}}}\boldsymbol{I}$
ΠGDM Song et al., 2023	$(1-\overline{\alpha}_t)\boldsymbol{A}\boldsymbol{A}^T + \sigma^2\boldsymbol{I}$
CGDM	$(1-\overline{\alpha}_t)\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{\Sigma}_t^{-1}\boldsymbol{A}^T + \sigma^2\boldsymbol{I}, \quad \boldsymbol{\Sigma}_t = \overline{\alpha}_t\boldsymbol{\Sigma} + (1-\overline{\alpha}_t)\boldsymbol{I}$

$t=$ 800	600	400	200	0



Comparison of the algorithms via 2-Wasserstein distance

The Asymptotic Discrete Spot Noise (ADSN) model [Galerne et al., 2011b] [5]

Let $\boldsymbol{u} \in \mathbb{R}^{\Omega_{M,N}}$ be a grayscale image, m its grayscale mean and $\boldsymbol{t} = \frac{1}{\sqrt{MN}}(\boldsymbol{u} - m)$ its associated texton. Let $\boldsymbol{w}$ be a white Gaussian noise,

$$\boldsymbol{X} = \boldsymbol{t} \star \boldsymbol{w} \sim \mathrm{ADSN}(\boldsymbol{u}) = \mathscr{N}(\boldsymbol{0}, \boldsymbol{\Gamma})$$ **which is a stationary law**

$\boldsymbol{u}$ | Samples of $\mathrm{ADSN}(\boldsymbol{u})$

Image extracted from [Galerne et al., 2011a][4]

[4] Galerne, B., Gousseau, Y., & Morel, J.-M. (2011a). Micro-texture synthesis by phase randomization. *Image Processing On Line*

[5] Galerne, B., Gousseau, Y., & Morel, J.-M. (2011b). Random Phase Textures: Theory and Synthesis. *IEEE Transactions on Image Processing*, *20*(1), 257–267

The Asymptotic Discrete Spot Noise (ADSN) model [Galerne et al., 2011b] [5]

Let $\boldsymbol{u} \in \mathbb{R}^{\Omega_{M,N}}$ be a grayscale image, m its grayscale mean and $\boldsymbol{t} = \frac{1}{\sqrt{MN}}(\boldsymbol{u} - m)$ its associated texton. Let $\boldsymbol{w}$ be a white Gaussian noise,

$$\boldsymbol{X} = \boldsymbol{t} \star \boldsymbol{w} \sim \mathrm{ADSN}(\boldsymbol{u}) = \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Gamma}) \quad \textbf{which is a stationary law}$$

$\boldsymbol{\Gamma}$ represents the convolution by the kernel $\gamma = \boldsymbol{t} \star \check{\boldsymbol{t}}$.

$\boldsymbol{u}$ Samples of $\mathrm{ADSN}(\boldsymbol{u})$

Image extracted from [Galerne et al., 2011a] [4]

[4] Galerne, B., Gousseau, Y., & Morel, J.-M. (2011a). Micro-texture synthesis by phase randomization. *Image Processing On Line*

[5] Galerne, B., Gousseau, Y., & Morel, J.-M. (2011b). Random Phase Textures: Theory and Synthesis. *IEEE Transactions on Image Processing*, *20*(1), 257–267

The Asymptotic Discrete Spot Noise (ADSN) model [Galerne et al., 2011b] [5]

Let $\boldsymbol{u} \in \mathbb{R}^{\Omega_{M,N}}$ be a grayscale image, m its grayscale mean and $\boldsymbol{t} = \frac{1}{\sqrt{MN}}(\boldsymbol{u} - m)$ its associated texton. Let $\boldsymbol{w}$ be a white Gaussian noise,

$$\boldsymbol{X} = \boldsymbol{t} \star \boldsymbol{w} \sim \mathrm{ADSN}(\boldsymbol{u}) = \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma})$$ **which is a stationary law**

$\boldsymbol{\Gamma}$ represents the convolution by the kernel $\gamma = \boldsymbol{t} \star \check{\boldsymbol{t}}$.

$\boldsymbol{u}$ Samples of $\mathrm{ADSN}(\boldsymbol{u})$

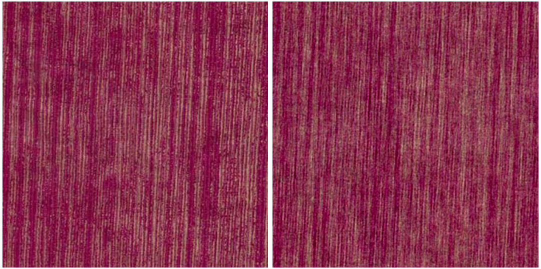


Image extracted from [Galerne et al., 2011a][4]

[4] Galerne, B., Gousseau, Y., & Morel, J.-M. (2011a). Micro-texture synthesis by phase randomization. *Image Processing On Line*

[5] Galerne, B., Gousseau, Y., & Morel, J.-M. (2011b). Random Phase Textures: Theory and Synthesis. *IEEE Transactions on Image Processing*, *20*(1), 257–267

The Asymptotic Discrete Spot Noise (ADSN) model [Galerne et al., 2011b] [5]

Let $\boldsymbol{u} \in \mathbb{R}^{\Omega_{M,N}}$ be a grayscale image, m its grayscale mean and $\boldsymbol{t} = \frac{1}{\sqrt{MN}}(\boldsymbol{u} - m)$ its associated texton. Let $\boldsymbol{w}$ be a white Gaussian noise,

$$\boldsymbol{X} = \boldsymbol{t} \star \boldsymbol{w} \sim \mathrm{ADSN}(\boldsymbol{u}) = \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Gamma}) \quad \textbf{which is a stationary law}$$

$\boldsymbol{\Gamma}$ represents the convolution by the kernel $\gamma = \boldsymbol{t} \star \check{\boldsymbol{t}}$.

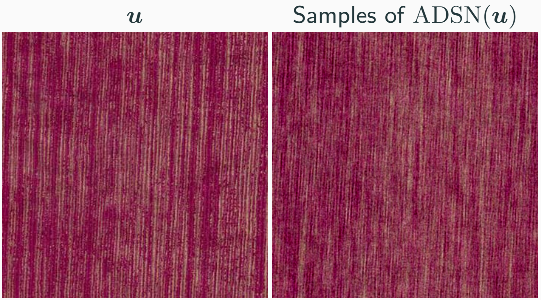


Image extracted from [Galerne et al., 2011a][4]

[4] Galerne, B., Gousseau, Y., & Morel, J.-M. (2011a). Micro-texture synthesis by phase randomization. *Image Processing On Line*

[5] Galerne, B., Gousseau, Y., & Morel, J.-M. (2011b). Random Phase Textures: Theory and Synthesis. *IEEE Transactions on Image Processing*, *20*(1), 257–267

The Asymptotic Discrete Spot Noise (ADSN) model [Galerne et al., 2011b] [5]

Let $\boldsymbol{u} \in \mathbb{R}^{\Omega_{M,N}}$ be a grayscale image, m its grayscale mean and $\boldsymbol{t} = \frac{1}{\sqrt{MN}}(\boldsymbol{u} - m)$ its associated texton. Let $\boldsymbol{w}$ be a white Gaussian noise,

$$\boldsymbol{X} = \boldsymbol{t} \star \boldsymbol{w} \sim \mathrm{ADSN}(\boldsymbol{u}) = \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Gamma}) \quad \textbf{which is a stationary law}$$

$\boldsymbol{\Gamma}$ represents the convolution by the kernel $\boldsymbol{\gamma} = \boldsymbol{t} \star \check{\boldsymbol{t}}$.

$\boldsymbol{u}$

Samples of $\mathrm{ADSN}(\boldsymbol{u})$

Image extracted from [Galerne et al., 2011a][4]

[4] Galerne, B., Gousseau, Y., & Morel, J.-M. (2011a). Micro-texture synthesis by phase randomization. *Image Processing On Line*

[5] Galerne, B., Gousseau, Y., & Morel, J.-M. (2011b). Random Phase Textures: Theory and Synthesis. *IEEE Transactions on Image Processing*, *20*(1), 257–267

The Asymptotic Discrete Spot Noise (ADSN) model [Galerne et al., 2011b] [5]

Let $\boldsymbol{u} \in \mathbb{R}^{\Omega_{M,N}}$ be a grayscale image, m its grayscale mean and $\boldsymbol{t} = \frac{1}{\sqrt{MN}}(\boldsymbol{u} - m)$ its associated texton. Let $\boldsymbol{w}$ be a white Gaussian noise,

$$\boldsymbol{X} = \boldsymbol{t} \star \boldsymbol{w} \sim \mathrm{ADSN}(\boldsymbol{u}) = \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}) \quad \textbf{which is a stationary law}$$

$\boldsymbol{\Gamma}$ represents the convolution by the kernel $\gamma = \boldsymbol{t} \star \check{\boldsymbol{t}}$.

$\boldsymbol{u}$ | Samples of $\mathrm{ADSN}(\boldsymbol{u})$

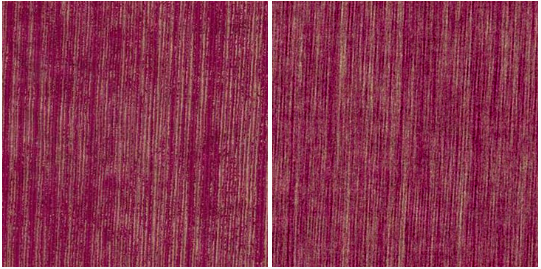


Image extracted from [Galerne et al., 2011a] [4]

[4] Galerne, B., Gousseau, Y., & Morel, J.-M. (2011a). Micro-texture synthesis by phase randomization. *Image Processing On Line*

[5] Galerne, B., Gousseau, Y., & Morel, J.-M. (2011b). Random Phase Textures: Theory and Synthesis. *IEEE Transactions on Image Processing*, *20*(1), 257–267

Let $\boldsymbol{u} \in \mathbb{R}^{\Omega_{M,N}}$ be a grayscale image, m its grayscale mean and $\boldsymbol{t} = \frac{1}{\sqrt{MN}}(\boldsymbol{u} - m)$ its associated texton. Let $\boldsymbol{w}$ be a white Gaussian noise,

$$\boldsymbol{X} = \boldsymbol{t} \star \boldsymbol{w} \sim \mathrm{ADSN}(\boldsymbol{u}) = \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}) \qquad \textbf{which is a stationary law}$$

$\boldsymbol{\Gamma}$ represents the convolution by the kernel $\boldsymbol{\gamma} = \boldsymbol{t} \star \check{\boldsymbol{t}}$.

$\boldsymbol{u}$ | Samples of $\mathrm{ADSN}(\boldsymbol{u})$

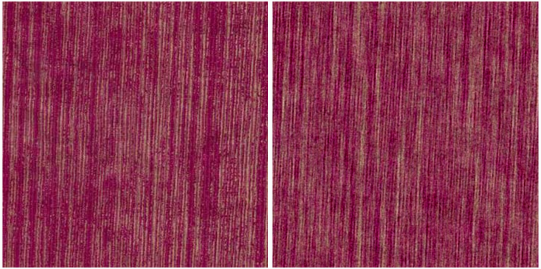


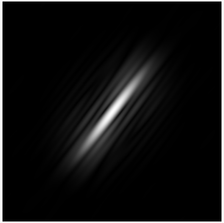
Image extracted from [Galerne et al., 2011a][4]

[4] Galerne, B., Gousseau, Y., & Morel, J.-M. (2011a). Micro-texture synthesis by phase randomization. *Image Processing On Line*

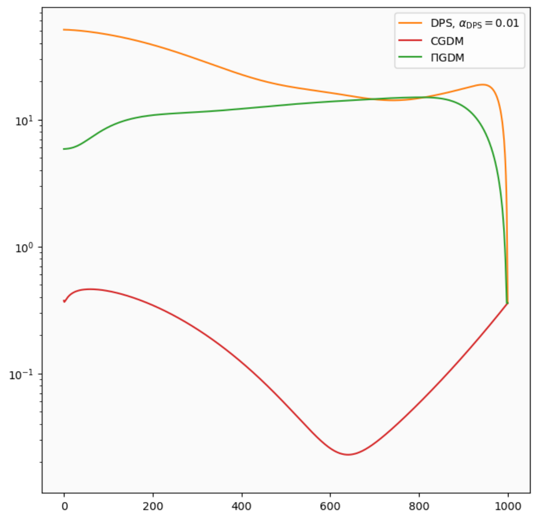
[5] Galerne, B., Gousseau, Y., & Morel, J.-M. (2011b). Random Phase Textures: Theory and Synthesis. *IEEE Transactions on Image Processing*, *20*(1), 257–267

Exact Wasserstein error for deblurring

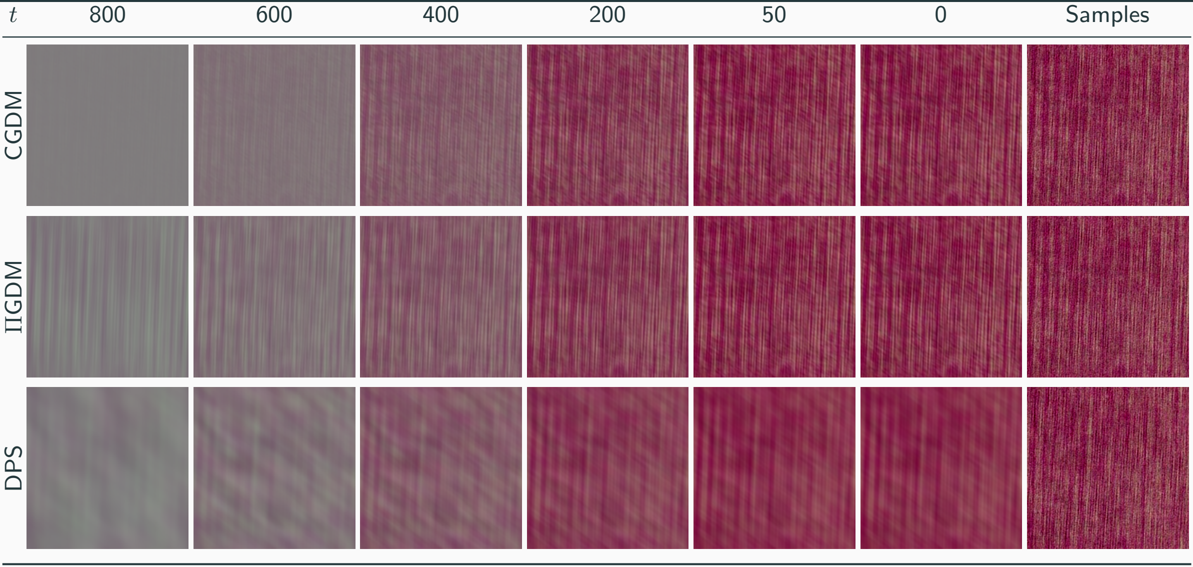
Blur kernel



Blurred image v



Study of the bias



Reverse Bayes rule

For $t \approx T$, $\boldsymbol{\Sigma}_t \approx \boldsymbol{I}$ and

$$\boldsymbol{C}_{\boldsymbol{v}|t}^{\text{DPS}} = \sigma^2 \boldsymbol{I}, \tag{24}$$

$$\boldsymbol{C}_{\boldsymbol{v}|t}^{\Pi\text{GDM}} \approx (1 - \overline{\alpha}_t) \boldsymbol{A}\boldsymbol{A}^T + \sigma^2 \boldsymbol{I}, \tag{25}$$

$$\boldsymbol{C}_{\boldsymbol{v}|t}^{\text{CGDM}} \approx (1 - \overline{\alpha}_t) \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^T + \sigma^2 \boldsymbol{I}. \tag{26}$$

Reverse Bayes rule

For $t \approx T$, $\boldsymbol{\Sigma}_t \approx \boldsymbol{I}$ and

$$\boldsymbol{C}_{\boldsymbol{v}|t}^{\mathsf{DPS}} = \sigma^2 \boldsymbol{I}, \tag{24}$$

$$\boldsymbol{C}_{\boldsymbol{v}|t}^{\Pi\mathsf{GDM}} \approx (1 - \overline{\alpha}_t) \boldsymbol{A}\boldsymbol{A}^T + \sigma^2 \boldsymbol{I}, \tag{25}$$

$$\boldsymbol{C}_{\boldsymbol{v}|t}^{\mathsf{CGDM}} \approx (1 - \overline{\alpha}_t) \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^T + \sigma^2 \boldsymbol{I}. \tag{26}$$

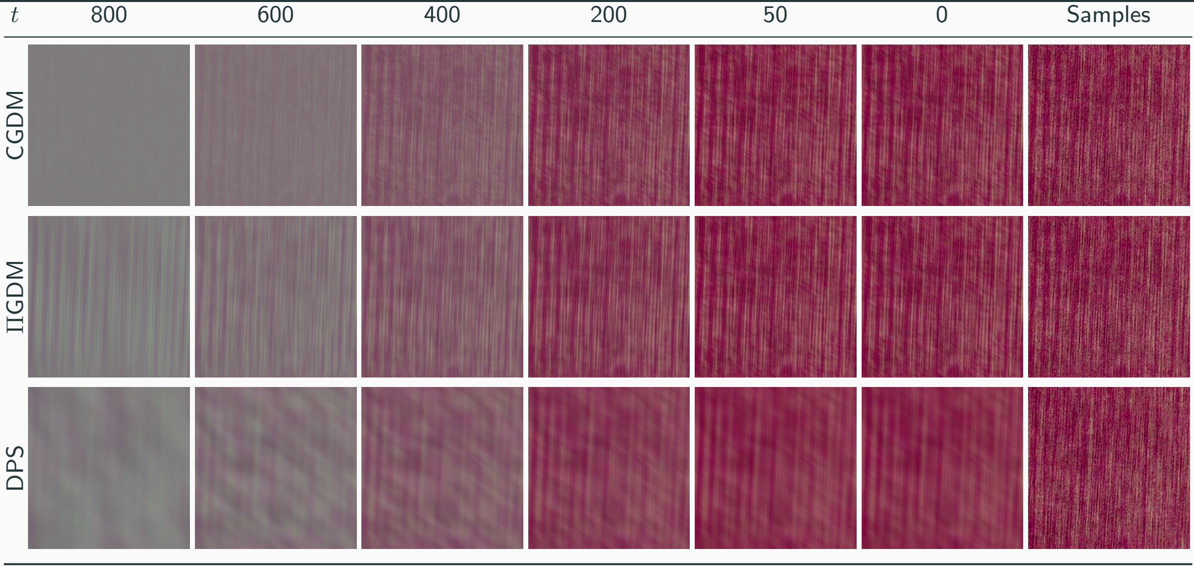
For $t \approx 0$, $\boldsymbol{\Sigma}_t \approx \boldsymbol{\Sigma}$ and

$$\boldsymbol{C}_{\boldsymbol{v}|t}^{\mathsf{DPS}} = \sigma^2 \boldsymbol{I}, \tag{27}$$

$$\boldsymbol{C}_{\boldsymbol{v}|t}^{\Pi\mathsf{GDM}} \approx (1 - \overline{\alpha}_t) \boldsymbol{A}\boldsymbol{A}^T + \sigma^2 \boldsymbol{I}, \tag{28}$$

$$\boldsymbol{C}_{\boldsymbol{v}|t}^{\mathsf{CGDM}} \approx (1 - \overline{\alpha}_t) \boldsymbol{A}\boldsymbol{A}^T + \sigma^2 \boldsymbol{I}. \tag{29}$$

Study of the bias



Conclusion

Conclusion

- Generalization to other inverse problems.
- Generalization to multimodal distributions.
- An important direction of research [Rozet et al., 2024][6]:

$$\mathrm{Cov}(\boldsymbol{x} \mid \boldsymbol{x}_t) = \sigma_t^2 + \sigma_t^4 \nabla_{\boldsymbol{x}}^2 \log p_t(\boldsymbol{x}_t), \tag{30}$$

[6] Rozet, F., Andry, G., Lanusse, F., & Louppe, G. (2024). Learning diffusion priors from observations by expectation maximization. *The Thirty-eighth Annual Conference on Neural Information Processing Systems*

Conclusion

- Generalization to other inverse problems.
- Generalization to multimodal distributions.
- An important direction of research [Rozet et al., 2024][6]:

$$\mathrm{Cov}(\boldsymbol{x} \mid \boldsymbol{x}_t) = \sigma_t^2 + \sigma_t^4 \nabla_{\boldsymbol{x}}^2 \log p_t(\boldsymbol{x}_t), \tag{30}$$

Thank you for your attention !

[6] Rozet, F., Andry, G., Lanusse, F., & Louppe, G. (2024). Learning diffusion priors from observations by expectation maximization. *The Thirty-eighth Annual Conference on Neural Information Processing Systems*

References

Chung, H., Kim, J., Mccann, M. T., Klasky, M. L., & Ye, J. C. (2023). Diffusion posterior sampling for general noisy inverse problems. *The Eleventh International Conference on Learning Representations*.

Galerne, B., Gousseau, Y., & Morel, J.-M. (2011a). Micro-texture synthesis by phase randomization. *Image Processing On Line*.

Galerne, B., Gousseau, Y., & Morel, J.-M. (2011b). Random Phase Textures: Theory and Synthesis. *IEEE Transactions on Image Processing*, *20*(1), 257–267.

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020*.

Kawar, B., Elad, M., Ermon, S., & Song, J. (2022). Denoising diffusion restoration models. *ICLR Workshop on Deep Generative Models for Highly Structured Data*.

Rozet, F., Andry, G., Lanusse, F., & Louppe, G. (2024). Learning diffusion priors from observations by expectation maximization. *The Thirty-eighth Annual Conference on Neural Information Processing Systems.*

Song, J., Vahdat, A., Mardani, M., & Kautz, J. (2023). Pseudoinverse-guided diffusion models for inverse problems. *International Conference on Learning Representations.*