NEURAL INFORMATION PROCESSING SYSTEMS

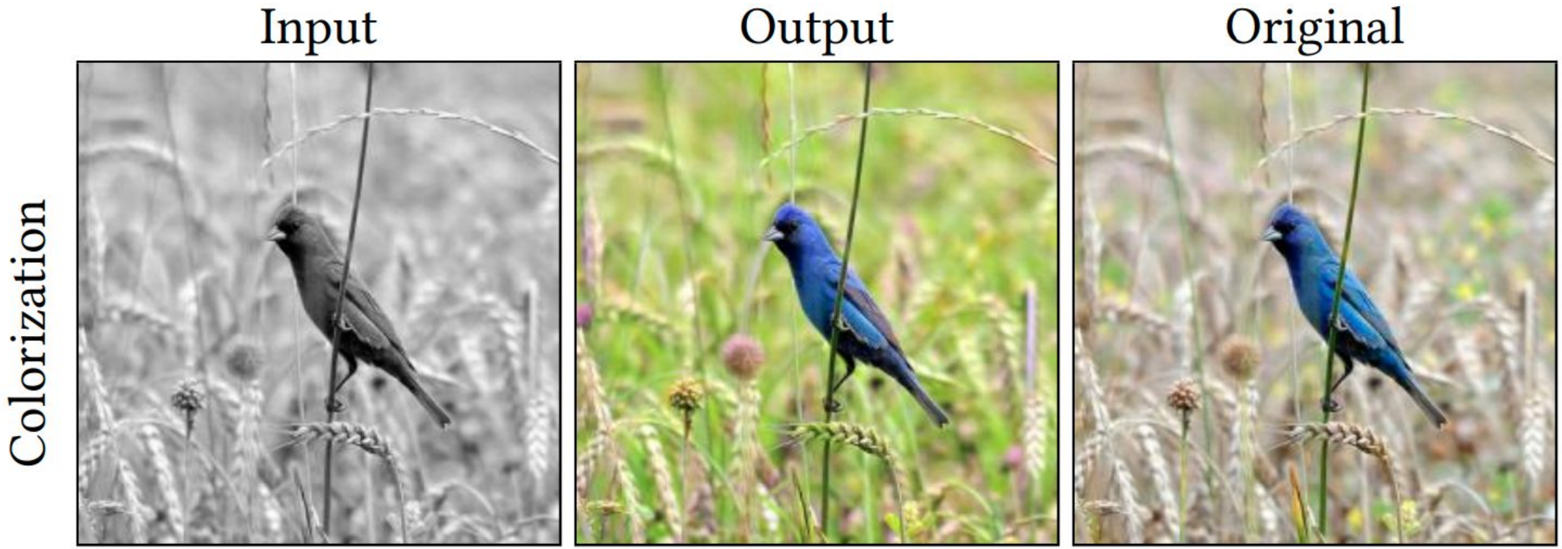
UCI University of California, Irvine

Fast Samplers for Inverse Problems in Iterative Refinement Models

Kushagra Pandey*, Ruihan Yang*, Stephan Mandt
(Slide Made By: Kushagra Pandey, Poster Presenter: Ruihan Yang)

Background - Inverse Problems



Source: Palette: Image to Image Diffusion Models, Saharia et al.

Background - Inverse Problems

Inpainting

Source: Palette: Image to Image Diffusion Models, Saharia et al.

Background - Inverse Problems

Source: Palette: Image to Image Diffusion Models, Saharia et al.

Background - Inverse Problems

Main Idea: Given a degradation model $h(.)$ and observations y infer x_0

$$y = h(x_0) + n$$

Background - Inverse Problems

Main Idea: Given a degradation model $h(.)$ and observations y infer x_0

$$y = h(x_0) + n$$

⬇

$$p(x_0|y) \propto p(y|x_0)p(x_0)$$

$$p(y|x_0) = \mathcal{N}(h(x_0), \sigma_y^2)$$

Background - Inverse Problems

Main Idea: Given a degradation model $h(.)$ and observations y infer x_0

$$y = h(x_0) + n$$

$$p(x_0|y) \propto p(y|x_0)p(x_0)$$

$$p(y|x_0) = \mathcal{N}(h(x_0), \sigma_y^2)$$

Idea: Use pretrained generative models

Background - Inverse Problems with Diffusion Models

Main Idea: Condition the reverse diffusion process on the degradations y

$$\frac{dx_t}{dt} = F_t x_t - \frac{1}{2} G_t G_t^\top \nabla_{x_t} \log p(x_t)$$

Background - Inverse Problems with Diffusion Models

Main Idea: Condition the reverse diffusion process on the degradations y

$$\frac{dx_t}{dt} = F_t x_t - \frac{1}{2} G_t G_t^\top \nabla_{x_t} \log p(x_t)$$

↓

$$\frac{dx_t}{dt} = F_t x_t - \frac{1}{2} G_t G_t^\top \nabla_{x_t} \log p(x_t | y)$$

Background - Inverse Problems with Diffusion Models

Main Idea: Condition the reverse diffusion process on the degradations y

$$\frac{dx_t}{dt} = F_t x_t - \frac{1}{2} G_t G_t^\top \nabla_{x_t} \log p(x_t)$$

↓

$$\frac{dx_t}{dt} = F_t x_t - \frac{1}{2} G_t G_t^\top \nabla_{x_t} \log p(x_t | y)$$

↓

$$\nabla_{x_t} \log p(x_t|y) = \nabla_{x_t} \log p(x_t) + w_t \nabla_{x_t} \log p(y|x_t)$$

$$\approx s_\theta(x_t, t) + w_t \nabla_{x_t} \log p(y|x_t)$$

Background - Inverse Problems with Diffusion Models

Main Idea: Condition the reverse diffusion process on the degradations y

$$\frac{dx_t}{dt} = F_t x_t - \frac{1}{2} G_t G_t^\top \nabla_{x_t} \log p(x_t)$$

↓

$$\frac{dx_t}{dt} = F_t x_t - \frac{1}{2} G_t G_t^\top \nabla_{x_t} \log p(x_t | y)$$

↓

$$\frac{dx_t}{dt} = F_t - \frac{1}{2} G_t G_t^\top \left[s_\theta(x_t, t) + w_t \nabla_{x_t} \log p(y | x_t) \right]$$

Background - Inverse Problems with Diffusion Models

$$\frac{dx_t}{dt} = F_t - \frac{1}{2} G_t G_t^\top \left[s_\theta(x_t, t) + w_t \nabla_{x_t} \log p(y|x_t) \right]$$

Caveat 1: Can use a pretrained score model $s_\theta(x_t, t)$

Background - Inverse Problems with Diffusion Models

$$\frac{dx_t}{dt} = F_t - \frac{1}{2} G_t G_t^\top \left[s_\theta(x_t, t) + w_t \nabla_{x_t} \log p(y|x_t) \right]$$

Caveat 1: Can use a pretrained score model $s_\theta(x_t, t)$

Caveat 2: Need to approximate $\nabla_{x_t} \log p(y|x_t)$

Background - Inverse Problems with Diffusion Models

$$\frac{dx_t}{dt} = F_t - \frac{1}{2} G_t G_t^\top \left[s_\theta(x_t, t) + w_t \nabla_{x_t} \log p(y|x_t) \right]$$

Caveat 1: Can use a pretrained score model $s_\theta(x_t, t)$

Caveat 2: Need to approximate $\nabla_{x_t} \log p(y|x_t)$

$$p(y|x_t) = \int p(y|x_0) p(x_0|x_t) dx_0$$

Degradation Model

$\mathcal{N}(h(x_0), \sigma_y^2)$

Background - Inverse Problems with Diffusion Models

$$\frac{dx_t}{dt} = F_t - \frac{1}{2} G_t G_t^\top \left[s_\theta(x_t, t) + w_t \nabla_{x_t} \log p(y|x_t) \right]$$

Caveat 1: Can use a pretrained score model $s_\theta(x_t, t)$

Caveat 2: Need to approximate $\nabla_{x_t} \log p(y|x_t)$

$$p(y|x_t) = \int p(y|x_0) p(x_0|x_t) dx_0$$

Diffusion Posterior

DPS: $p(x_0|x_t) \approx \delta(\mathbb{E}[x_0|x_t])$

ΠGDM: $p(x_0|x_t) \approx \mathcal{N}(\mathbb{E}[x_0|x_t], \Sigma_t)$

Setup - Inverse Problems

Degradation Model (Linear): $y = Hx_0 + \sigma_y z, \qquad p(y|x_0) = \mathcal{N}(Hx_0, \sigma_y^2 I_d)$

Setup - Inverse Problems

Degradation Model (Linear): $y = Hx_0 + \sigma_y z, \qquad p(y|x_0) = \mathcal{N}(Hx_0, \sigma_y^2 I_d)$

Posterior Approximation(πGDM) $p(x_0|x_t) \approx \mathcal{N}(\bar{x}_t, r_t^2 I_d)$

$$\bar{x}_t = \mathbb{E}[x_0|x_t] = x_t + \sigma_t^2 s_\theta(x_t, t)$$

Setup - Inverse Problems

Degradation Model (Linear): $y = Hx_0 + \sigma_y z, \qquad p(y|x_0) = \mathcal{N}(Hx_0, \sigma_y^2 I_d)$

Posterior Approximation(πGDM) $p(x_0|x_t) \approx \mathcal{N}(\bar{x}_t, r_t^2 I_d)$

$$\bar{x}_t = \mathbb{E}[x_0|x_t] = x_t + \sigma_t^2 s_\theta(x_t, t)$$

Noisy Likelihood $p(y|x_t) = \int p(x_0|x_t) p(y|x_0) dx_0$

$$= \mathcal{N}(H\bar{x}_t, r_t^2 H H^\top + \sigma_y^2 I_d)$$

Setup - Inverse Problems

Degradation Model (Linear):

$$y = Hx_0 + \sigma_y z, \qquad p(y|x_0) = \mathcal{N}(Hx_0, \sigma_y^2 I_d)$$

Posterior Approximation(πGDM)

$$p(x_0|x_t) \approx \mathcal{N}(\bar{x}_t, r_t^2 I_d)$$

$$\bar{x}_t = \mathbb{E}[x_0|x_t] = x_t + \sigma_t^2 s_\theta(x_t, t)$$

Noisy Likelihood Score

$\nabla_{x_t} \log p(y|x_t)$

$$(y - H\bar{x}_t)^\top (r_t^2 HH^\top + \sigma_y^2 I_d)^{-1} H \frac{\partial \bar{x}_t}{\partial x_t}$$

$\sigma_y = 0$ [Noiseless inverse problems]

$$r_t^{-2} \left[H^\dagger (y - H\bar{x}_t) \frac{\partial \bar{x}_t}{\partial x_t} \right]$$

Setup - Inverse Problems

Degradation Model (Linear): $y = Hx_0 + \sigma_y z, \qquad p(y|x_0) = \mathcal{N}(Hx_0, \sigma_y^2 I_d)$

Posterior Approximation(πGDM) $p(x_0|x_t) \approx \mathcal{N}(\bar{x}_t, r_t^2 I_d)$

$$\bar{x}_t = \mathbb{E}[x_0|x_t] = x_t + \sigma_t^2 s_\theta(x_t, t)$$

Noisy Likelihood Score

$\nabla_{x_t} \log p(y|x_t)$

$$r_t^{-2}\left[H^\dagger (y - H\bar{x}_t) \frac{\partial \bar{x}_t}{\partial x_t}\right]$$

[Noiseless inverse problems]

$\sigma_y = 0$

Setup - Inverse Problems

Degradation Model (Linear): $y = Hx_0 + \sigma_y z, \qquad p(y|x_0) = \mathcal{N}(Hx_0, \sigma_y^2 I_d)$

Posterior Approximation(πGDM) $p(x_0|x_t) \approx \mathcal{N}(\bar{x}_t, r_t^2 I_d)$

$$\bar{x}_t = \mathbb{E}[x_0|x_t] = x_t + \sigma_t^2 s_\theta(x_t, t)$$

Noisy Likelihood Score

$\nabla_{x_t} \log p(y|x_t)$

$$r_t^{-2}\left[H^\dagger(y - H\bar{x}_t)\frac{\partial \bar{x}_t}{\partial x_t}\right]$$

[Noiseless inverse problems]

$\sigma_y = 0$

$H^\dagger = H^\top(HH^\top)^{-1}$ denotes the "Pseudo-Inverse" of the degradation operator H

Setup - Inverse Problems

Degradation Model (Linear): $y = Hx_0 + \sigma_y z, \qquad p(y|x_0) = \mathcal{N}(Hx_0, \sigma_y^2 I_d)$

Posterior Approximation(πGDM) $p(x_0|x_t) \approx \mathcal{N}(\bar{x}_t, r_t^2 I_d)$

$$\bar{x}_t = \mathbb{E}[x_0|x_t] = x_t + \sigma_t^2 s_\theta(x_t, t)$$

Noisy Likelihood Score

$\nabla_{x_t} \log p(y|x_t)$

$$r_t^{-2}\left[H^\dagger (y - H\bar{x}_t)\frac{\partial \bar{x}_t}{\partial x_t}\right]$$

[Noiseless inverse problems]

$\sigma_y = 0$

$$\frac{dx_t}{dt} = F_t - \frac{1}{2} G_t G_t^\top \left[s_\theta(x_t, t) + w_t \nabla_{x_t} \log p(y|x_t)\right]$$

Motivation

Noisy Likelihood Score

$$\nabla_{x_t} \log p(y|x_t)$$

$$r_t^{-2}\left[H^\dagger\left(y - H\hat{x}_t\right)\frac{\partial \hat{x}_t}{\partial x_t}\right]^\top$$

Vector

Jacobian

Motivation

Noisy Likelihood Score

$\nabla_{x_t} \log p(y|x_t)$

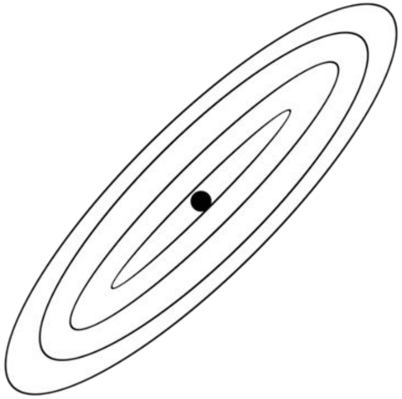
$$r_t^{-2}\left[H^{\dagger}\left(y - H\hat{x}_t\right)\frac{\partial \hat{x}_t}{\partial x_t}\right]^{\top}$$

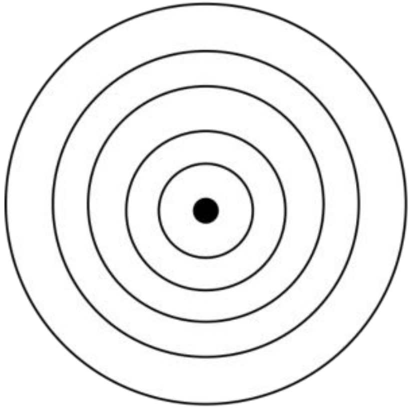
$$\frac{dx_t}{dt} = F_t - \frac{1}{2}G_t G_t^{\top}\left[s_\theta(x_t, t) + w_t \nabla_{x_t} \log p(y|x_t)\right]$$

Score Function Evaluation + Jacobian-Vector Product

Solving Inverse Problems with pretrained models is very slow!

Conditional Conjugate Integrators - Overview

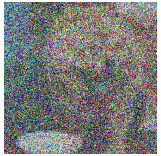


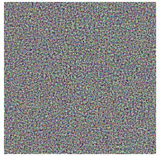


Project to another space

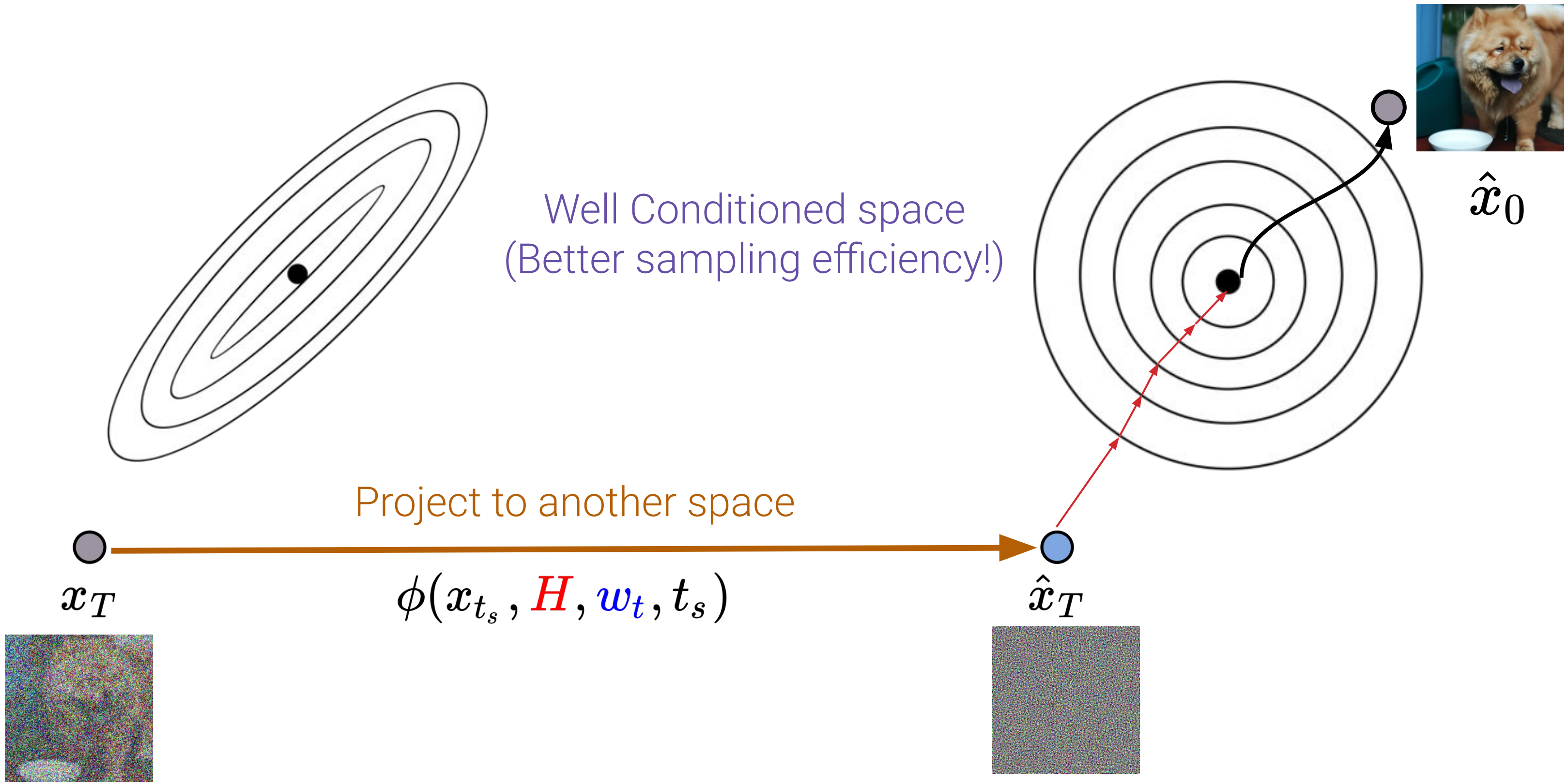
x_T $\longrightarrow$ $\hat{x}_T$

$\phi(x_{t_s}, H, w_t, t_s)$

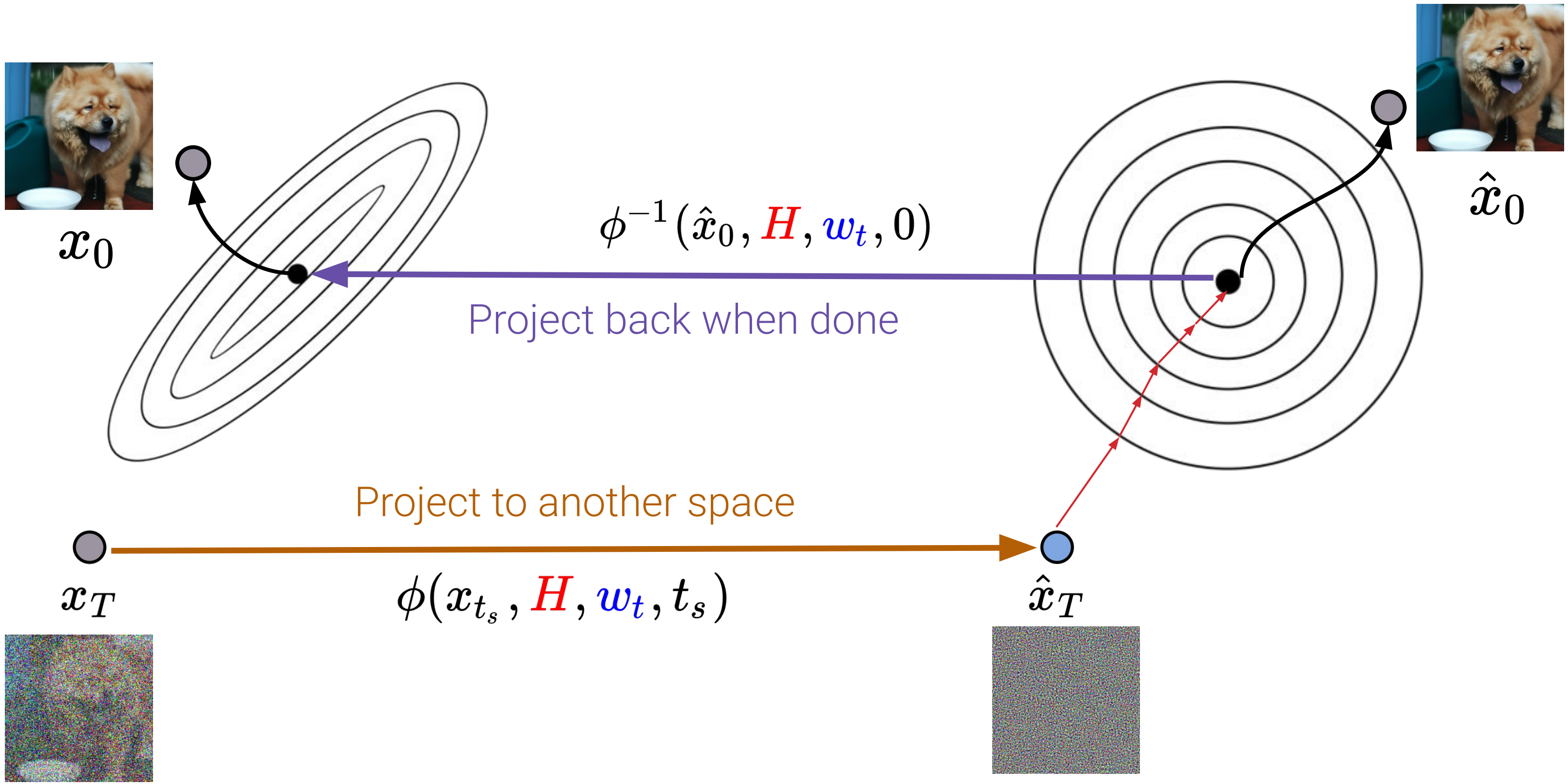




Conditional Conjugate Integrators - Overview



Conditional Conjugate Integrators - Overview



Conditional Conjugate Integrators - Formulation

Original Space: $$\frac{dx_t}{dt} = F_t - \frac{1}{2} G_t G_t^\top \left[s_\theta(x_t, t) + w_t \nabla_{x_t} \log p(y|x_t) \right]$$

$$\hat{x}_t = A_t x_t$$

Projected Space: $$\frac{d\hat{x}_t}{dt} = A_t B_t A_t^{-1} \hat{x}_t + d\Phi_t \epsilon_\theta(x_t, t) + d\Phi_y y + d\Phi_j \left[\partial_{x_t} \epsilon_\theta(x_t, t)(H^\dagger y - P\hat{x}_0) \right]$$

$$A_t = \exp\left[\int_0^t B_s - \left(F_s + \frac{w_s r_s^{-2}}{2\mu_s^2} G_s G_s^\top P \right) \right]$$

Conditional Conjugate Integrators - Formulation

Original Space: $$\frac{dx_t}{dt} = F_t - \frac{1}{2}G_t G_t^\top \left[s_\theta(x_t, t) + w_t \nabla_{x_t} \log p(y|x_t)\right]$$

$$\hat{x}_t = A_t x_t$$

Projected Space: $$\frac{d\hat{x}_t}{dt} = A_t B_t A_t^{-1}\hat{x}_t + d\Phi_t \epsilon_\theta(x_t, t) + d\Phi_y y + d\Phi_j \left[\partial_{x_t} \epsilon_\theta(x_t, t)(H^\dagger y - P\hat{x}_0)\right]$$

$$A_t = \exp\left[\int_0^t B_s - \left(F_s + \frac{w_s r_s^{-2}}{2\mu_s^2} G_s G_s^\top P\right)\right]$$

$P = H^\dagger H$ denotes the Orthogonal Projector operator

Conditional Conjugate Integrators - Design Space

Choice of Diffusion (VP-SDE): $F_t = -\frac{1}{2}\beta_t I_d$ $\quad G_t = \sqrt{\beta_t} I_d$

Score Parameterization: $C_{\text{skip}}(t) = 0$ $\quad s_\theta(x_t, t) = C_{\text{out}}(t)\epsilon_\theta(x_t, t)$

Conditional Conjugate Integrators - Design Space

Choice of Diffusion (VP-SDE): $F_t = -\frac{1}{2}\beta_t I_d$ $G_t = \sqrt{\beta_t} I_d$

Score Parameterization: $C_{\text{skip}}(t) = 0$ $s_\theta(x_t, t) = C_{\text{out}}(t)\epsilon_\theta(x_t, t)$

Projection: $\hat{x}_t = A_t x_t$

Design Choice: $B_t = \lambda I_d$

Conditional Conjugate Integrators - Design Space

Choice of Diffusion (VP-SDE): $F_t = -\frac{1}{2}\beta_t I_d \qquad G_t = \sqrt{\beta_t} I_d$

Score Parameterization: $C_{\text{skip}}(t) = 0 \qquad s_\theta(x_t, t) = C_{\text{out}}(t)\epsilon_\theta(x_t, t)$

Projection: $\hat{x}_t = A_t x_t$

Design Choice: $B_t = \lambda I_d$

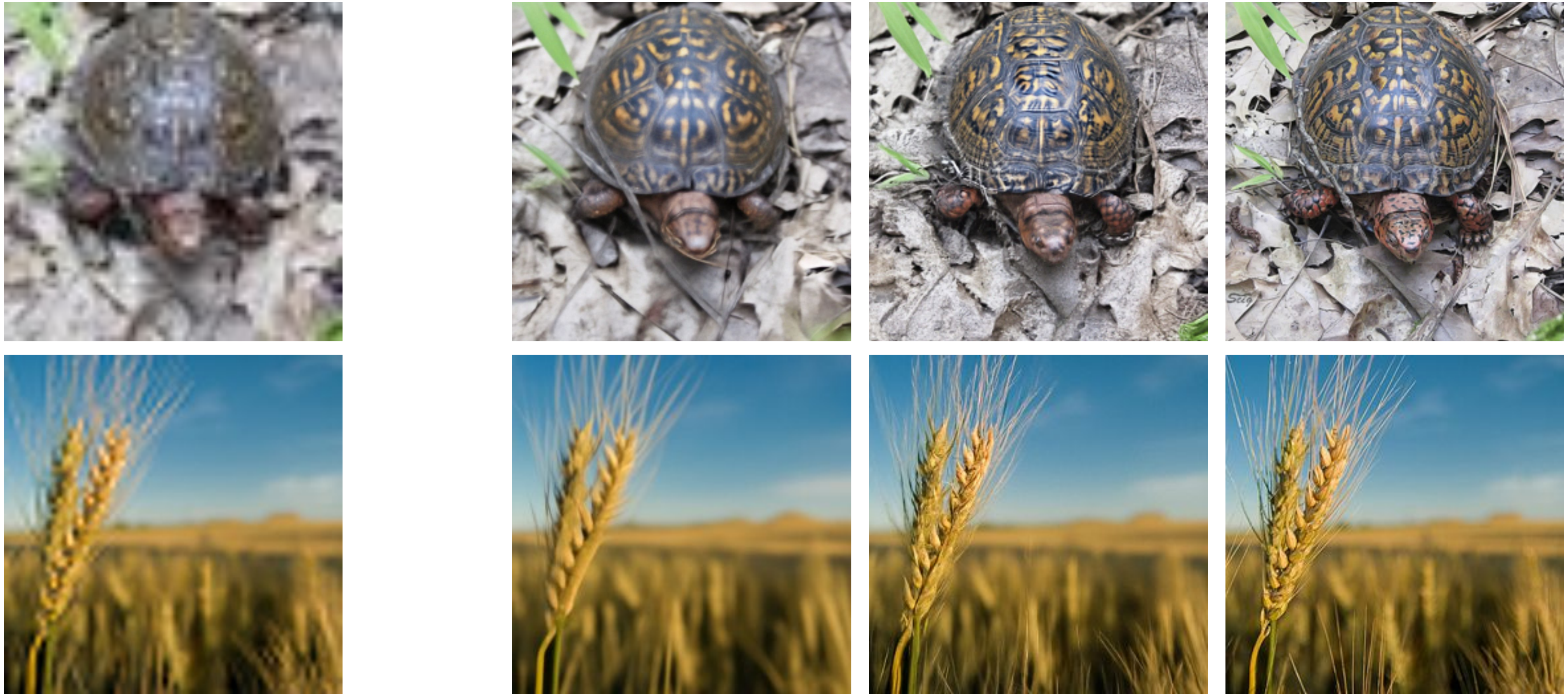
$$A_t = \exp\left[\int_0^t B_s - \left(F_s + \frac{w_s r_s^{-2}}{2\mu_s^2} G_s G_s^\top P\right)\right]$$

$$\Downarrow$$

$$A_t = \kappa_t^1(\lambda, w_t)\Big[I_d + \kappa_t^2(\lambda, w_t) P\Big]$$

Conditional Conjugate Integrators - Results (4x SR)

Degraded Input	ΠGDM (5 steps)	Conjugate-ΠGDM (5 steps)	Reference



Conditional Conjugate Integrators - Results (4x SR)

Diffusion Results		C-ΠGDM	ΠGDM	DPS	DDRM	C-ΠGDM	ΠGDM	DPS	DDRM	C-ΠGDM	ΠGDM	DPS	DDRM
Super-Resolution	5	**0.220**	0.306			**2.7**	6.3			**37.31**	49.06		
	10	**0.206**	0.252	0.252	0.318	**1.6**	4.8	5.8	14.1	**34.22**	44.30	38.18	51.64
	20	**0.207**	0.222			**1.7**	2.5			**34.28**	37.36		
Deblurring	5	**0.272**	0.349			**3.89**	14.1			**44.42**	63.94		
	10	**0.272**	0.294	0.619	0.336	**3.6**	5.3	59.5	12.3	**43.37**	47.80	139.58	62.53
	20	0.268	**0.259**			**3.5**	4.2			**43.70**	44.20		

4x improvement in Speed-vs-Quality Tradeoffs over vanilla ΠGDM

Conditional Conjugate Integrators - Results (Noisy 4xSR)

Degraded Input | Conjugate-ΠGDM (5 steps) | Reference

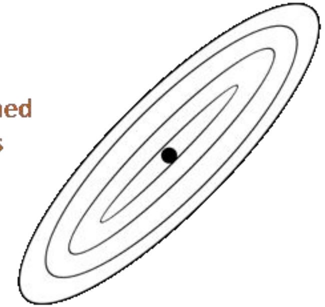
$$\sigma_y = 0.05$$

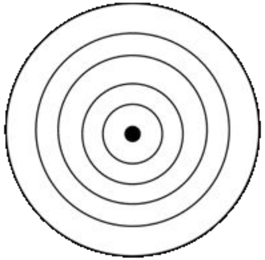
Extends to Non-linear inverse problems as well

$$A_t^{\sigma_y} = A_t + \kappa_t H^{\dagger}(H^{\dagger})^{\top} + \mathcal{O}(\sigma_y^4)$$

TLDR

Ill-Conditioned
Dynamics





Better-Conditioned
Dynamics