

802 Appendix for "System-Embedded Diffusion Bridge Models"

803	A Proofs and derivations	21
804	A.1 Derivation of drift and diffusion coefficients for SDB	21
805	A.2 Proof of theorem 2	22
806	B Extended related works	23
807	B.1 Diffusion models for inverse problems	23
808	B.2 Diffusion and Schrödinger bridges	24
809	B.3 Baselines	24
810	C Extended methodology	26
811	C.1 Network parameterization	26
812	C.2 Nonlinear measurement systems	26
813	C.3 Pseudocode	26
814	D Experimental setup	26
815	D.1 Training hyperparameters	26
816	D.2 Computational requirements	26
817	E Additional experimental results	27
818	E.1 Evaluation under a misspecified model (CT reconstruction)	27
819	E.2 Hyperparameter analysis	27
820	F Broader impact	29
821	G Additional visual results	29

822 A Proofs and derivations

823 A.1 Derivation of drift and diffusion coefficients for SDB

824 We begin with deriving the drift and diffusion coefficients $\mathbf{F}_t, \mathbf{G}_t$ of SDB (eqs. (10a) and (10b)) by
825 using theorem 1 with the mean and covariance matrices \mathbf{H}_t, Σ_t (eqs. (8a) and (8b)) of its forward
826 process. Recall that $\mathbf{H}_t = \mathbf{A}^+ \mathbf{A} + \alpha_t (\mathbf{I} - \mathbf{A}^+ \mathbf{A}), \Sigma_t = \gamma_t \mathbf{A}^+ \Sigma \mathbf{A}^{+\top} + \beta_t (\mathbf{I} - \mathbf{A}^+ \mathbf{A})$. Following
827 theorem 1, we obtain

$$\mathbf{A}^+ \mathbf{A} \mathbf{F}_t = \frac{d}{dt} \mathbf{A}^+ \mathbf{A} \log \mathbf{H}_t = \frac{d}{dt} \mathbf{A}^+ \mathbf{A} = \mathbf{0}, \quad (12a)$$

$$(\mathbf{I} - \mathbf{A}^+ \mathbf{A}) \mathbf{F}_t = \frac{d}{dt} \log \alpha_t (\mathbf{I} - \mathbf{A}^+ \mathbf{A}), \quad (12b)$$

$$\frac{d}{dt} \mathbf{A}^+ \mathbf{A} \Sigma_t = \frac{d\gamma_t}{dt} \mathbf{A}^+ \Sigma \mathbf{A}^{+\top}, \quad (12c)$$

$$\frac{d}{dt} (\mathbf{I} - \mathbf{A}^+ \mathbf{A}) \Sigma_t = \frac{d\beta_t}{dt} (\mathbf{I} - \mathbf{A}^+ \mathbf{A}). \quad (12d)$$

828 Moreover, note that $\mathbf{F}_t^\top = \mathbf{F}_t, \Sigma_t^\top = \Sigma_t$ due to the symmetry of the range and null space projections,
829 and $\mathbf{F}_t \Sigma_t = \Sigma_t \mathbf{F}_t$ since \mathbf{F}_t only affects the null space and $(\mathbf{I} - \mathbf{A}^+ \mathbf{A})^2 = (\mathbf{I} - \mathbf{A}^+ \mathbf{A})$ (idempotent).

830 Following theorem [1](#)

$$\mathbf{F}_t = \frac{d}{dt} \log \alpha_t (\mathbf{I} - \mathbf{A}^+ \mathbf{A}), \quad (13a)$$

$$\mathbf{G}_t \mathbf{G}_t = \frac{d\boldsymbol{\Sigma}_t}{dt} - \mathbf{F}_t \boldsymbol{\Sigma}_t - \boldsymbol{\Sigma}_t \mathbf{F}_t^\top \quad (13b)$$

$$= \frac{d\boldsymbol{\Sigma}_t}{dt} - 2\mathbf{F}_t \boldsymbol{\Sigma}_t \quad (13c)$$

$$= \frac{d\gamma_t}{dt} \mathbf{A}^+ \boldsymbol{\Sigma} \mathbf{A}^{+\top} + \left(\frac{d\beta_t}{dt} - 2\beta_t \frac{d}{dt} \log \alpha_t \right) (\mathbf{I} - \mathbf{A}^+ \mathbf{A}). \quad (13d)$$

831 A.2 Proof of theorem [2](#)

832 We begin by recalling two propositions of [Liu et al. \[2023\]](#) using our notation, which will serve as
833 the basis of the proof.

834 **Proposition 1 (Analytic posterior given boundary pair)** [\[Liu et al., 2023\]](#) The posterior of eq. [\(1\)](#)
835 given some boundary pair $(\mathbf{x}_0, \mathbf{x}_1)$ admits an analytic form:

$$p(\mathbf{x}_t \mid \mathbf{x}_0, \mathbf{x}_1) = \mathcal{N}(\boldsymbol{\mu}_t(\mathbf{x}_0, \mathbf{x}_1), \boldsymbol{\Sigma}_t), \quad (14)$$

836 where $\boldsymbol{\mu}_t = \frac{\bar{\sigma}_t^2}{\bar{\sigma}_t^2 + \sigma_t^2} \mathbf{x}_0 + \frac{\sigma_t^2}{\bar{\sigma}_t^2 + \sigma_t^2} \mathbf{x}_1$, $\boldsymbol{\Sigma}_t = \frac{\sigma_t^2 \bar{\sigma}_t^2}{\bar{\sigma}_t^2 + \sigma_t^2} \mathbf{I}$ and $\sigma_t^2 = \int_0^t g^2(\tau) d\tau$, $\bar{\sigma}_t^2 = \int_t^1 g^2(\tau) d\tau$ with
837 $\mathbf{G}_t = g(t) \mathbf{I}$.

838 **Proposition 2 (Optimal Transport ODE; OT-ODE)** [\[Liu et al., 2023\]](#) When $g^2(t) \rightarrow 0$, the SDE
839 between $(\mathbf{x}_0, \mathbf{x}_1)$ reduces to an ODE:

$$d\mathbf{x}_t = \mathbf{v}_t(\mathbf{x}_t \mid \mathbf{x}_0) dt, \quad (15)$$

840 where $\mathbf{v}_t(\mathbf{x}_t \mid \mathbf{x}_0) = \frac{g^2(t)}{\sigma_t^2} (\mathbf{x}_t - \mathbf{x}_0)$ whose solution is the posterior mean of eq. [\(14\)](#).

841 Consider the null space part of \mathbf{x}_t given by mean and covariance matrices from eqs. [\(8a\)](#) and [\(8b\)](#)
842 with $\alpha_t = \frac{\bar{\sigma}_t^2}{\bar{\sigma}_t^2 + \sigma_t^2}$, $\beta_t = \frac{\sigma_t^2}{\bar{\sigma}_t^2 + \sigma_t^2}$ for $\sigma_t^2 = \int_0^t g^2(\tau) d\tau$, $\bar{\sigma}_t^2 = \int_t^1 g^2(\tau) d\tau$, where $g(t)$ is the null space
843 diffusion coefficient given by eq. [\(10b\)](#), i.e., $g(t) = \left(\frac{d\beta_t}{dt} - 2\beta_t \frac{d}{dt} \log \alpha_t \right)^{\frac{1}{2}}$:

$$(\mathbf{I} - \mathbf{A}^+ \mathbf{A}) \mathbf{x}_t = \frac{\bar{\sigma}_t^2}{\bar{\sigma}_t^2 + \sigma_t^2} (\mathbf{I} - \mathbf{A}^+ \mathbf{A}) \mathbf{x}_0 + \left(\frac{\sigma_t^2 \bar{\sigma}_t^2}{\bar{\sigma}_t^2 + \sigma_t^2} \right)^{\frac{1}{2}} (\mathbf{I} - \mathbf{A}^+ \mathbf{A}) \boldsymbol{\epsilon} \quad (16)$$

844 for $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}_d, \mathbf{I}_{d \times d})$. Note that the null space part of \mathbf{x}_1 , being the PR of \mathbf{x}_0 , is zeroed out, i.e.,
845 $(\mathbf{I} - \mathbf{A}^+ \mathbf{A}) \mathbf{x}_1 = (\mathbf{I} - \mathbf{A}^+ \mathbf{A}) \mathbf{0}_{d \times d}$. This allows us to artificially rewrite \mathbf{x}_t as

$$(\mathbf{I} - \mathbf{A}^+ \mathbf{A}) \mathbf{x}_t = (\mathbf{I} - \mathbf{A}^+ \mathbf{A}) \left(\frac{\bar{\sigma}_t^2}{\bar{\sigma}_t^2 + \sigma_t^2} \mathbf{x}_0 + \frac{\sigma_t^2}{\bar{\sigma}_t^2 + \sigma_t^2} \mathbf{0}_{d \times d} \right) + \left(\frac{\sigma_t^2 \bar{\sigma}_t^2}{\bar{\sigma}_t^2 + \sigma_t^2} \right)^{\frac{1}{2}} (\mathbf{I} - \mathbf{A}^+ \mathbf{A}) \boldsymbol{\epsilon}. \quad (17)$$

846 Equivalently, $(\mathbf{I} - \mathbf{A}^+ \mathbf{A}) \mathbf{x}_t \sim \mathcal{N}((\mathbf{I} - \mathbf{A}^+ \mathbf{A}) \left(\frac{\bar{\sigma}_t^2}{\bar{\sigma}_t^2 + \sigma_t^2} \mathbf{x}_0 + \frac{\sigma_t^2}{\bar{\sigma}_t^2 + \sigma_t^2} \mathbf{x}_1 \right), \frac{\sigma_t^2 \bar{\sigma}_t^2}{\bar{\sigma}_t^2 + \sigma_t^2} (\mathbf{I} - \mathbf{A}^+ \mathbf{A}))$. Hence,
847 the posterior mean of SDB (SB) takes the form stated in proposition [1](#) and proposition [2](#) can be
848 applied directly to the null space part.

849 As a final remark, we note that, up to this point, the definition of $g(t)$ is interdependent with that
850 of α_t and β_t . It is not immediately clear that defining α_t, β_t as in theorem [2](#) fulfills $g^2(t) =$
851 $\frac{d\beta_t}{dt} - 2\beta_t \frac{d}{dt} \log \alpha_t$. We now show that this property is indeed satisfied.

852 Denote by $C = \sigma_t^2 + \bar{\sigma}_t^2$ and observe that $\beta_t = \sigma_t^2 \alpha_t$, $\bar{\sigma}_t^2 = C - \sigma_t^2$, $\frac{d\sigma_t^2}{dt} = g^2(t)$. Then,

$$\frac{d}{dt} \log \alpha_t = \frac{d}{dt} \log \frac{\bar{\sigma}_t^2}{\bar{\sigma}_t^2 + \sigma_t^2} \quad (18a)$$

$$= \frac{d}{dt} (\log \bar{\sigma}_t^2 - \log C) \quad (18b)$$

$$= \frac{g^2(t)}{\bar{\sigma}_t^2}, \quad (18c)$$

$$\frac{d\beta_t}{dt} = \frac{d}{dt} \left(\frac{\sigma_t^2 \bar{\sigma}_t^2}{\bar{\sigma}_t^2 + \sigma_t^2} \right) \quad (18d)$$

$$= \frac{d}{dt} \left(\frac{\sigma_t^2 (C - \sigma_t^2)}{C} \right) \quad (18e)$$

$$= \frac{1}{C} \frac{d}{dt} (C\sigma_t^2 - \sigma_t^4) \quad (18f)$$

$$= g^2(t) - \frac{2}{C} \sigma_t^2 g^2(t) \quad (18g)$$

$$= g^2(t) \left(1 - \frac{2}{C} \sigma_t^2 \right). \quad (18h)$$

853 By substituting these into the definition of $g^2(t)$, we obtain

$$\frac{d\beta_t}{dt} - 2\beta_t \frac{d}{dt} \log \alpha_t = \frac{d}{dt} (\sigma_t^2 \alpha_t) - 2\sigma_t^2 \alpha_t \frac{d}{dt} \log \alpha_t \quad (19a)$$

$$= \alpha_t \frac{d\sigma_t^2}{dt} + \sigma_t^2 \frac{d\alpha_t}{dt} - 2\sigma_t^2 \alpha_t \left(\frac{1}{\alpha_t} \frac{d\alpha_t}{dt} \right) \quad (19b)$$

$$= \alpha_t g^2(t) - \sigma_t^2 \frac{d\alpha_t}{dt} \quad (19c)$$

$$= \left(1 - \frac{\sigma_t^2}{C} \right) g^2(t) + \frac{\sigma_t^2}{C} g^2(t) \quad (19d)$$

$$= g^2(t), \quad (19e)$$

854 which completes the proof.

855

□

856 B Extended related works

857 B.1 Diffusion models for inverse problems

858 In recent years, deep neural networks have gained significant attention for solving various inverse
859 problems from different perspectives [Ongie et al., 2020, Chen et al., 2022]. Solving eq. (4) can be
860 interpreted as sampling from the posterior distribution $p(\mathbf{x} | \mathbf{y})$, and generative models that support
861 conditional generation are naturally suited to this task. Diffusion models, in particular, have emerged
862 as SOTA tools for this purpose, thanks to their flexible mathematical formulation and expressive
863 data priors [Daras et al., 2024]. The standard approach involves extending the diffusion process
864 to sample from $p(\mathbf{x}_0 | \mathbf{y})$ at $t = 0$ by applying Bayes' Theorem. The conditional score function
865 can be decomposed as $\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t | \mathbf{y}) = \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t) + \nabla_{\mathbf{x}_t} \log p(\mathbf{y} | \mathbf{x}_t)$, where the first term,
866 $\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t)$, can be approximated using a pretrained score network $s_\theta(\mathbf{x}_t, t)$, and the second term
867 models the relationship between \mathbf{x}_t and the measurement \mathbf{y} [Song et al., 2021b, Dhariwal and Nichol,
868 2021]. This framework has spurred considerable progress in the field, with numerous successful
869 methods emerging [Kawar et al., 2022, Chung et al., 2022, 2023a, Song et al., 2023a, Mardani et al.,
870 2024, Chung et al., 2024]. Since this approach does not require retraining the score network for each
871 specific problem, we refer to these methods as *unsupervised*.

872 Several works leverage the measurement system structure when applying pretrained diffusion models
873 to inverse problems. [Wang et al., 2023] (Denoising Diffusion Null-Space Models, DDNM) restrict
874 updates during generation to the null space component of \mathbf{x}_t , keeping the range part fixed. [Song

et al. [2023a] (Pseudoinverse-Guided Diffusion Models, IIGDM) approximate the likelihood score via a vector-Jacobian product, where the score network’s Jacobian is computed with automatic differentiation, and the vector reflects the range-nullspace decomposition. Garber and Tirer [2024] (Denoising Diffusion Models with Iteratively Preconditioned Guidance, DDPG) propose a method for interpolating between pseudoinverse-based and least-squares-based conditioning.

B.2 Diffusion and Schrödinger bridges.

Diffusion models, while effective for high-quality image synthesis, are limited by the simplicity of Gaussian priors for $p(\mathbf{x}_1)$. Recent developments in *bridge models* [Särkkä and Solin, 2019] generalize the diffusion process by allowing $p(\mathbf{x}_1)$ to be an arbitrary distribution. This is especially important in image restoration tasks, where paired samples $(\mathbf{x}_0, \mathbf{x}_1)$ —clean and distorted images—are available. Bridge models aim to generate samples from the posterior $p(\mathbf{x}_0 | \mathbf{x}_1)$ by initializing with a sample from $p(\mathbf{x}_1)$ rather than Gaussian noise. While conditioning the standard score network on \mathbf{x}_1 is a possible approach to achieve posterior sampling, it is often suboptimal [Batzolis et al., 2021].

Several methods have formulated the diffusion process as a stochastic bridge. Heng et al. [2021] propose a simulation-based algorithm using a fixed starting and ending point with an approximation of the true score. Liu et al. [2022] extend this with Doob’s h-transform [Doob, 1984] to bridge distributions. Simulation-free algorithms utilizing the h-transform are presented by Somnath et al. [2023] and Peluchetti [2022]. Delbracio and Milanfar [2023] construct a Brownian Bridge for direct restoration from \mathbf{x}_1 . More recently, Zhou et al. [2024] introduce the DDBM framework, extending the VE and VP processes, while He et al. [2024] study DDBM within the consistency framework [Song et al., 2023b]. Zheng et al. [2025] link DDBM to Denoising Diffusion Implicit Models [Song et al., 2021a].

The Schrödinger Bridge (SB) problem [Schrödinger, 1932], which aligns distributions via constrained forward and reverse SDEs, has also been explored. De Bortoli et al. [2021] apply Iterative Proportional Fitting (IPF) to solve the SB problem. Liu et al. [2023] propose a tractable class of SBs, leading to a simulation-free algorithm (I2SB). Chung et al. [2023b] extend it with an additional guidance term. Shi et al. [2024] build on IPF and introduce Iterative Markovian Fitting for SB solutions.

In a related approach, Luo et al. [2023a] derive a scalar case of eq. (I) that incorporates the start and end points, termed the *mean-reverting* SDE (IR-SDE). Through specific parameterization, they show that its score function is analytically tractable, connecting to the Ornstein-Uhlenbeck (OU) process [Gillespie, 1996]. Yue et al. [2024] extend this work with a generalized OU process and Doob’s h-transform (GOUB). Further extensions include Luo et al. [2023b], who apply IR-SDE in latent spaces, and Welker et al. [2022] and Richter et al. [2023], who adapt similar processes for speech tasks. Recently, Zhu et al. [2025] unify GOUB and DDBM within the framework of stochastic optimal control.

B.3 Baselines

B.3.1 Unsupervised

We begin with describing the unsupervised baselines, which rely on solving the following scalar reverse equation:

$$d\mathbf{x}_t = [\mathbf{f}(\mathbf{x}_t, t) - g^2(t)\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t | \mathbf{y})]dt + g(t)d\bar{\mathbf{w}}_t, \quad (20)$$

where \mathbf{f} and g represent the drift and diffusion coefficients respectively, while \mathbf{y} is the conditioning variable representing measurements in the inverse problem context. Unsupervised diffusion-based methods rely on decomposing the score function with Bayes’ Theorem through $\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t | \mathbf{y}) = \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t) + \nabla_{\mathbf{x}_t} \log p(\mathbf{y} | \mathbf{x}_t)$ and approximating $\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t)$ with a pretrained $\mathbf{s}_\theta(\mathbf{x}_t, t)$, while proposing different approaches for $\nabla_{\mathbf{x}_t} \log p(\mathbf{y} | \mathbf{x}_t)$.

Diffusion Posterior Sampling (DPS, Chung et al. [2023a]) approximates $p(\mathbf{y} | \mathbf{x}_t)$ with $p(\mathbf{y} | \hat{\mathbf{x}}_0(\mathbf{x}_t))$, where $\hat{\mathbf{x}}_0(\mathbf{x}_t) = \mathbb{E}[\mathbf{x}_0 | \mathbf{x}_t] = \mathbf{x}_t + g^2(t)\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t)$ is the Tweedie’s formula [Robbins, 1992], giving an approximation of the denoised image at timestep t . Similarly, $\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t)$ is also approximated with $\mathbf{s}_\theta(\mathbf{x}_t, t)$ in this case.

Pseudoinverse-Guided Diffusion Model (IIGDM, Song et al. [2023a]) proposes to approximate the loglikelihood score with $(\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}_0(\mathbf{x}_t))^\top (r_t^2 \mathbf{A}\mathbf{A}^\top + \sigma_y^2 \mathbf{I})^{-1} \mathbf{A} \frac{\partial \hat{\mathbf{x}}_0(\mathbf{x}_t)}{\partial \mathbf{x}_t}$, where $\sigma_y^2 \mathbf{I}$ is the measurement

system covariance and r_t^2 is a time-dependent term that should depend on the data (Appendix A.3., Song et al. [2023a]). Using automatic differentiation, one can compute $\frac{\partial \hat{\mathbf{x}}_0(\mathbf{x}_t)}{\partial \mathbf{x}_t}$. By combining it with the other terms, the entire approximation can be efficiently computed as a vector-Jacobian product.

In their basic formulation, Denoising Diffusion Null-space Models (DDNM, Wang et al. [2023]) rely on approximating $p(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{y})$ using the following update rule:

$$\mathbf{x}_{t-1} = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1 - \bar{\alpha}_t} \hat{\mathbf{x}}_0(\mathbf{x}_t, \mathbf{y}) + \frac{\sqrt{\alpha_t}(1 - \bar{\alpha}_{t-1})}{1 - \bar{\alpha}_t} \mathbf{x}_t + \sigma_t \epsilon, \quad (21)$$

where $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and α_t, β_t follow the notation from the original work. This rule utilizes the range-nullspace decomposition via $\hat{\mathbf{x}}_0(\mathbf{x}_t, \mathbf{y}) = \mathbf{A}^+ \mathbf{y} + (\mathbf{I} - \mathbf{A}^+ \mathbf{A}) \hat{\mathbf{x}}_0(\mathbf{x}_t)$, which preserves the true range space component, while updating the null space part with Tweedie’s estimate.

B.3.2 Supervised

We proceed with a description of baseline supervised bridge methods. All of these approaches assume a stochastic process conditioned on both the starting and ending point with the following forward and reverse equations:

$$d\mathbf{x}_t = \mathbf{f}(\mathbf{x}_t, \mathbf{x}_T, t, T)dt + g(t)d\mathbf{w}_t, \quad (22a)$$

$$d\mathbf{x}_t = \mathbf{f}'(\mathbf{x}_t, \mathbf{x}_T, t, T)dt + g(t)d\bar{\mathbf{w}}_t, \quad (22b)$$

where \mathbf{f}, \mathbf{f}' represent general drift coefficients respectively for the forward and reverse equation, g is the diffusion coefficient and \mathbf{x}_T represents the endpoint for the initial \mathbf{x}_0 .

Image-to-Image Schrodinger Bridges (I2SB, Liu et al. [2023]) formulate the distribution of \mathbf{x}_t , given starting and ending points $\mathbf{x}_0, \mathbf{x}_T$, as $p(\mathbf{x}_t | \mathbf{x}_0, \mathbf{x}_T) = \mathcal{N}(\frac{\bar{\sigma}_t^2}{\bar{\sigma}_t^2 + \sigma_t^2} \mathbf{x}_0 + \frac{\sigma_t^2}{\bar{\sigma}_t^2 + \sigma_t^2} \mathbf{x}_T, \frac{\sigma_t^2 \bar{\sigma}_t^2}{\bar{\sigma}_t^2 + \sigma_t^2} \mathbf{I})$, where $\sigma_t^2 = \int_0^t g^2(\tau) d\tau$, $\bar{\sigma}_t^2 = \int_t^T g^2(\tau) d\tau$ and show its equivalence to DDPM posterior sampling [Ho et al. 2020].

Image Restoration SDE (IR-SDE, Luo et al. [2023a]) is based on formulating the forward and reverse equations as

$$d\mathbf{x}_t = \theta_t(\mathbf{x}_T - \mathbf{x}_t)dt + \sigma(t)d\mathbf{w}_t, \quad (23a)$$

$$d\mathbf{x}_t = [\theta_t(\mathbf{x}_T - \mathbf{x}_t) - \sigma^2(t)\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t)]dt + \sigma(t)d\bar{\mathbf{w}}_t, \quad (23b)$$

where $\theta_t = \frac{\sigma_t^2}{\lambda^2}$, $\bar{\theta}_t = \int_0^t \theta_\tau d\tau$ and λ is a predefined constant. At $t = 1$, IR-SDE arrives at a Gaussian distribution (with non-zero covariance) centered at \mathbf{x}_T .

Generalized Ornstein-Uhlenbeck Bridges (GOUB, Yue et al. [2024]) show IR-SDE as a special case of their framework by incorporating the Doob’s h-transform [Doob, 1984], which pulls the process towards the desired endpoint. Formally, it is defined as $\mathbf{h}(\mathbf{x}_t, t, \mathbf{x}_T, T) = \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_T | \mathbf{x}_t)$ and incorporated into the forward and reverse equations:

$$d\mathbf{x}_t = \left((\theta_t + g^2(t) \frac{e^{-2\bar{\theta}_{t:T}}}{\bar{\sigma}_{t:T}^2})(\mathbf{x}_T - \mathbf{x}_t) \right) dt + g(t)d\mathbf{w}_t, \quad (24a)$$

$$d\mathbf{x}_t = \left((\theta_t + g^2(t) \frac{e^{-2\bar{\theta}_{t:T}}}{\bar{\sigma}_{t:T}^2})(\mathbf{x}_T - \mathbf{x}_t) - g^2(t)\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t | \mathbf{x}_T) \right) dt + g(t)d\bar{\mathbf{w}}_t, \quad (24b)$$

where $\theta_t = \frac{g^2(t)}{2\lambda^2}$, $\bar{\theta}_{s:t} = \int_s^t \theta_\tau d\tau$, $\bar{\sigma}_{s:t}^2 = \frac{g^2(t)}{2\theta_t}(1 - e^{-2\bar{\theta}_{s:t}})$ for a predefined constant λ .

Denoising Diffusion Bridge Model (DDBM, Zhou et al. [2024]) also utilize the h-transform, but instead show how to adapt prior unconditional processes, which map images to Gaussian noise, to construct a bridge between arbitrary distributions given paired data:

$$d\mathbf{x}_t = [\mathbf{f}(\mathbf{x}_t, t) + g^2(t)\mathbf{h}(\mathbf{x}_t, t, \mathbf{x}_T, T)]dt + g(t)d\mathbf{w}_t, \quad (25a)$$

$$d\mathbf{x}_t = [\mathbf{f}(\mathbf{x}_t, t) - g^2(t)(\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t | \mathbf{x}_T) - \mathbf{h}(\mathbf{x}_t, t, \mathbf{x}_T, T))]dt + g(t)d\bar{\mathbf{w}}_t, \quad (25b)$$

where $\mathbf{f}(\mathbf{x}_t, t), g(t)$ follow the original image-to-noise process.

C Extended methodology

C.1 Network parameterization

In a scalar setting, various reparameterizations of the score function were shown to provide different trade-offs in the final performance [Ho et al., 2020]. For example, the network could instead predict the added noise directly, which is bijectively obtained from the score. As these reparameterizations are related through simple scalar functions, one can choose them freely without additional considerations. In the matrix-valued setting, however, the choice of a parameterization is more subtle. In this case, observe that $\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t) = -\Sigma_t^{-1}(\mathbf{x}_t - \mathbf{H}_t \mathbb{E}[\mathbf{x}_0 | \mathbf{x}_t])$. Training a score-prediction model hence requires access to the inverse of Σ_t , which may in general be costly to obtain. By properly choosing the form of γ_t and β_t , one can show that $\mathbf{G}_t \mathbf{G}_t^\top \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t) = (f_t \mathbf{I} - 2\mathbf{F}_t)(\mathbf{H}_t \mathbb{E}[\mathbf{x}_0 | \mathbf{x}_t] - \mathbf{x}_t)$ for some function f_t , *i.e.*, an \mathbf{x}_0 -prediction model alleviates the need for computing Σ_t^{-1} . Therefore, we treat this parameterization as the default one for SDB.

C.2 Nonlinear measurement systems

While not the focus of this work, inverse problems with nonlinear system response \mathbf{A} may also be considered, *e.g.*, in the problem of GAN inversion [Creswell and Bharath, 2018]. Assuming that \mathbf{A} is differentiable, it may be locally linearized using Taylor expansion as $\mathbf{A}(\mathbf{x}) \approx \mathbf{A}(\mathbf{x}_0) + \mathbf{J}_{\mathbf{A}} \mathbf{x}_0 (\mathbf{x} - \mathbf{x}_0)$ around some point \mathbf{x}_0 , where $\mathbf{J}_{\mathbf{A}}$ is the Jacobian of \mathbf{A} . To adapt SDB to such cases, one can use $\mathbf{J}_{\mathbf{A}}$ as a local linear approximation of \mathbf{A} and treat it instead as the system response matrix. We consider further study of such models as future work.

C.3 Pseudocode

We provide pseudocode for SDB for both training and Euler-Maruyama sampling in algorithm 1 and algorithm 2 respectively.

Algorithm 1 SDB Training

Require: $p(\mathbf{x}_0), p(t), \mathbf{A}, \mathbf{A}^+, \Sigma^{1/2}, \alpha_t, \beta_t, \gamma_t, \mathbf{D}_\theta$

- 1: **for** each iteration **do**
- 2: $\mathbf{x} \sim p(\mathbf{x})$
- 3: $t \sim p(t)$
- 4: $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \in \mathbb{R}^m$
- 5: $\epsilon' \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \in \mathbb{R}^d$
- 6: $\mathbf{x}_t \leftarrow [\mathbf{A}^+ \mathbf{A} + \alpha_t (\mathbf{I} - \mathbf{A}^+ \mathbf{A})] \mathbf{x} + \gamma_t^{\frac{1}{2}} \mathbf{A}^+ \Sigma^{\frac{1}{2}} \epsilon + \beta_t^{\frac{1}{2}} (\mathbf{I} - \mathbf{A}^+ \mathbf{A}) \epsilon'$
- 7: $L_\theta \leftarrow \|\mathbf{D}_\theta(\mathbf{x}_t, t) - \mathbf{x}\|_1$
- 8: $\theta \leftarrow \text{optimizer}(\nabla_\theta L_\theta)$
- 9: **end for**
- 10: **return** θ

D Experimental setup

D.1 Training hyperparameters

We follow the training procedure proposed by [Luo et al., 2023a], using the ADAM optimizer [Kingma and Ba, 2015] with an initial learning rate of 1×10^{-4} , no weight decay, and $(\beta_1, \beta_2) = (0.9, 0.99)$. A multi-step learning rate scheduler is applied, halving the learning rate at the 36th, 60th, 72nd, and 90th epochs, as in the original work. All methods are trained using the ℓ_1 loss function with a batch size of 8. To ensure fairness across methods, we evaluate each model every 16 epochs and report the performance of the best checkpoint, rather than relying solely on the final one.

D.2 Computational requirements

All experiments were conducted on a cluster of NVIDIA A100 GPUs, with each method trained using a single GPU. The approximate training times for each task are as follows: 4 hours for super-resolution, 24 hours for MRI reconstruction, 48 hours for CT reconstruction, and 72 hours for inpainting.

Algorithm 2 SDB Sampling (Euler-Maruyama)

Require: $N, \mathbf{A}, \mathbf{A}^+, \Sigma^{1/2}, \alpha_t, \beta_t, \gamma_t, \mathbf{H}_t, \mathbf{F}_t, D_{\theta}, \mathbf{y}$

```

1:  $t \leftarrow 1$ 
2:  $\Delta t \leftarrow 1/N$ 
3:  $\epsilon' \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \in \mathbb{R}^d$ 
4:  $\hat{\mathbf{x}} \leftarrow \mathbf{A}^+ \mathbf{y}$ 
5:  $\mathbf{x}_t \leftarrow \hat{\mathbf{x}} + \beta_t^{\frac{1}{2}} (\mathbf{I} - \mathbf{A}^+ \mathbf{A}) \epsilon'$ 
6: for  $i \in \{1, \dots, N\}$  do
7:    $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \in \mathbb{R}^m$ 
8:    $\epsilon' \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \in \mathbb{R}^d$ 
9:    $\mathbf{x}_{t-\Delta t} \leftarrow \mathbf{x}_t + \Delta t [(f_t \mathbf{I} - 2\mathbf{F}_t)(\mathbf{H}_t D_{\theta}(\mathbf{x}_t, t) - \mathbf{x}_t) - \mathbf{F}_t \mathbf{x}_t] +$ 
        $\Delta t^{1/2} \left[ \left( \frac{d\gamma_t}{dt} \right)^{\frac{1}{2}} \mathbf{A}^+ \Sigma^{\frac{1}{2}} \epsilon + \left( \frac{d\beta_t}{dt} - 2\beta_t \frac{d}{dt} \log \alpha_t \right)^{\frac{1}{2}} (\mathbf{I} - \mathbf{A}^+ \mathbf{A}) \epsilon' \right]$ 
10:   $t \leftarrow t - \Delta t$ 
11: end for
12: return  $\mathbf{x}_0$ 

```

991 E Additional experimental results

992 E.1 Evaluation under a misspecified model (CT reconstruction)

993 Figure 4 presents the results of an analogous analysis of performance of bridge methods under a
994 misspecified model (here in CT reconstruction task), where the first part of the experiment considers
995 the default setting of σ_1^2 with increasing τ , while the second part shows the results for $\tau = 3.6$ and
996 increasing σ_1^2 . In a similar manner to section 5.2, all SDB variants achieve a clear advantage over
997 other bridge methods when the system’s parameters are perturbed.

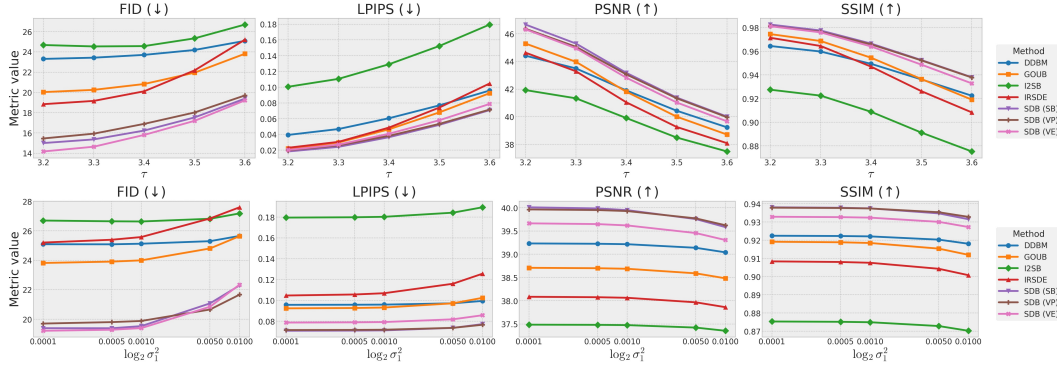


Figure 4: Quantitative comparison of SDB (SB) with other bridge methods in a misspecified CT reconstruction setting. The **top** row evaluates bridges trained with $\tau = 3.2$, $\sigma_1^2 = 0.0001$ on data generated from systems with increasing τ . The **bottom** row evaluates performance on data with $\tau = 3.6$ and increasing σ_1^2 . Perturbing the original system makes the problem harder in both cases.

998 E.2 Hyperparameter analysis

999 We conduct an ablation study to analyze the sensitivity of SDB to its hyperparameters. Following
1000 prior works [Zhou et al., 2024], our stochastic process is undefined at $t = 1$, requiring sampling to
1001 begin at $t = 1 - \epsilon_1$ for some $\epsilon_1 > 0$. Similarly, we terminate sampling at $t = \epsilon_2$, where $\epsilon_2 > 0$. For
1002 SDB (VP) and SDB (VE), ϵ_1 and ϵ_2 are the only hyperparameters. In contrast, SDB (SB) includes an
1003 additional design choice: the shape of the null-space diffusion coefficient $g(t)$. We parameterize it as

$$g^2(t) = \mathbb{1}_{t \leq 0.5} \hat{g}(t) + \mathbb{1}_{t > 0.5} \hat{g}(1 - t)$$

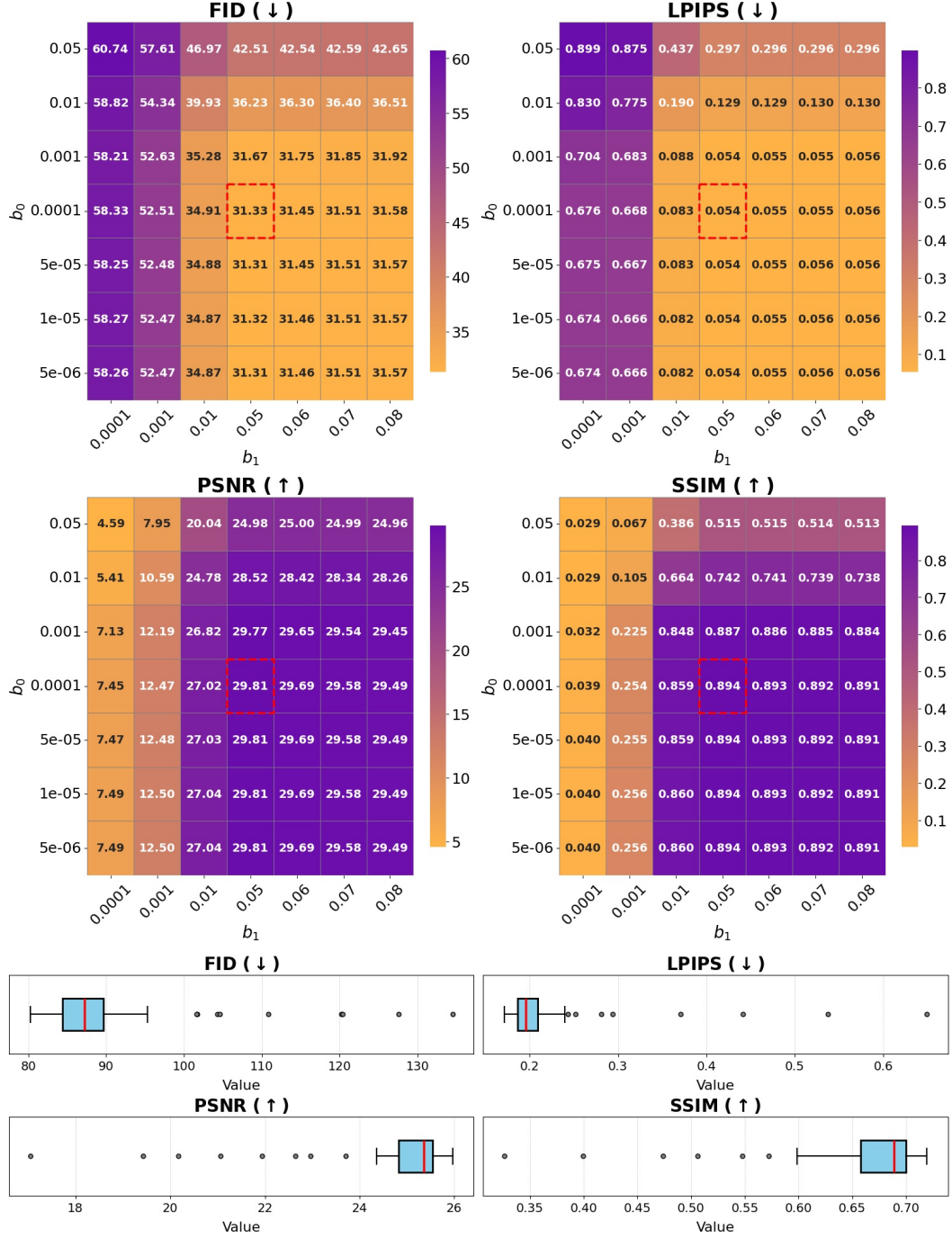


Figure 5: **Top:** Performance heatmaps for MRI reconstruction under varying (ϵ_1, ϵ_2) , using the optimal (b_0, b_1) configuration. The red box highlights the LPIPS-optimal setting. **Bottom:** Boxplots showing the best achievable performance across 64 (b_0, b_1) pairs on the superresolution task, where (ϵ_1, ϵ_2) are tuned separately for each (b_0, b_1) pair.

1004 where $\hat{g}(t) = (\sqrt{b_0} + t(\sqrt{b_1} - \sqrt{b_0}))^2$, which is a continuous version of the coefficient proposed by
 1005 Liu et al. [2023]. This introduces two additional hyperparameters, $b_0, b_1 > 0$. In the remainder of
 1006 this section, we focus on SDB (SB), noting that we observed qualitatively similar trends for the VP
 1007 and VE variants.

1008 We first examine how the choice of ϵ_1 and ϵ_2 influences the optimal configuration of (b_0, b_1) for
 1009 SDB (SB) in the MRI reconstruction task. Figure 5 (top) shows the performance trend when varying
 1010 (ϵ_1, ϵ_2) . We observe a simple and stable relationship with performance, where ϵ_1 has a stronger
 1011 influence than ϵ_2 . Moreover, a broad plateau emerges, allowing for easy tuning. We also highlight the
 1012 best configuration in terms of LPIPS, which lies close to the optimal Pareto frontier in this setting.

1013 Next, we analyze the stability of SDB (SB) across different (b_0, b_1) configurations. Since these
 1014 values determine the shape of the variance function, the optimal ϵ_1, ϵ_2 may differ across (b_0, b_1) pairs.
 1015 Figure 5 (bottom) presents boxplots of the best performance for 64 configurations of (b_0, b_1) on the
 1016 superresolution task. For each pair, we use grid search to identify the best ϵ_1, ϵ_2 values, leveraging the
 1017 small size of the DIV2K dataset. The results reveal strong stability, with performance concentrated
 1018 around the best observed value and only a few outliers.

1019 F Broader impact

1020 By incorporating measurement system parameters into its SDE, SDB offers positive societal impacts
 1021 by enabling more accurate and efficient reconstructions in critical applications such as medical
 1022 imaging, remote sensing, and scientific inverse problems, thereby supporting improved diagnostics,
 1023 sustainability, and broader accessibility. However, it also poses potential negative societal risks, in-
 1024 cluding misuse for surveillance or de-anonymization, amplification of bias from flawed measurement
 1025 models, and overreliance on plausible but incorrect outputs in high-stakes domains. Additionally,
 1026 the method could be repurposed for generating deceptive content from limited data. However, SDB
 1027 and prior bridge methods have not been demonstrated at industry-scale deployment levels, and in
 1028 practice, they are typically suited for problem-specific applications rather than large-scale settings
 1029 with vast amounts of data. This limits their potential for misuse in broader societal contexts.

1030 G Additional visual results

1031 We provide additional qualitative samples for inpainting (fig. 6), superresolution (fig. 7), MRI
 1032 reconstruction (fig. 8) and CT reconstruction (fig. 9).

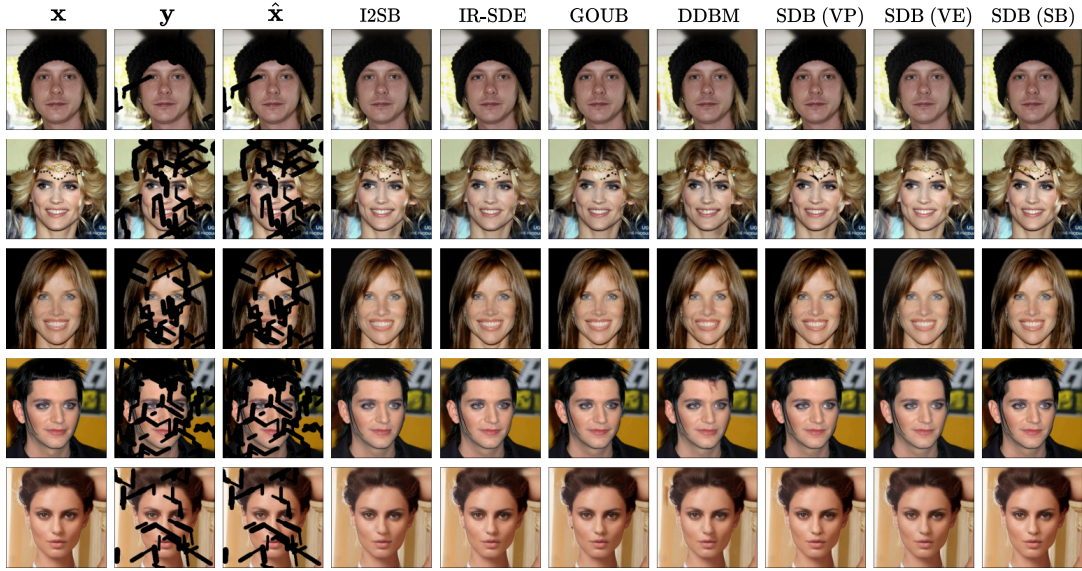


Figure 6: Qualitative comparison of SDB variants with the best-performing baselines (bridge methods). Rows depict the results for inpainting.

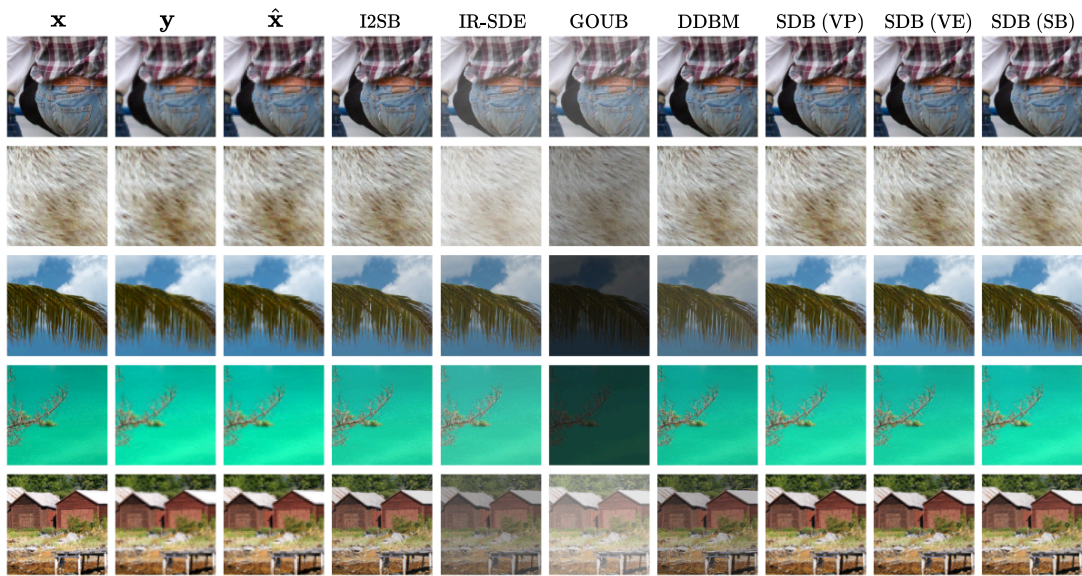


Figure 7: Qualitative comparison of SDB variants with the best-performing baselines (bridge methods). Rows depict the results for superresolution.

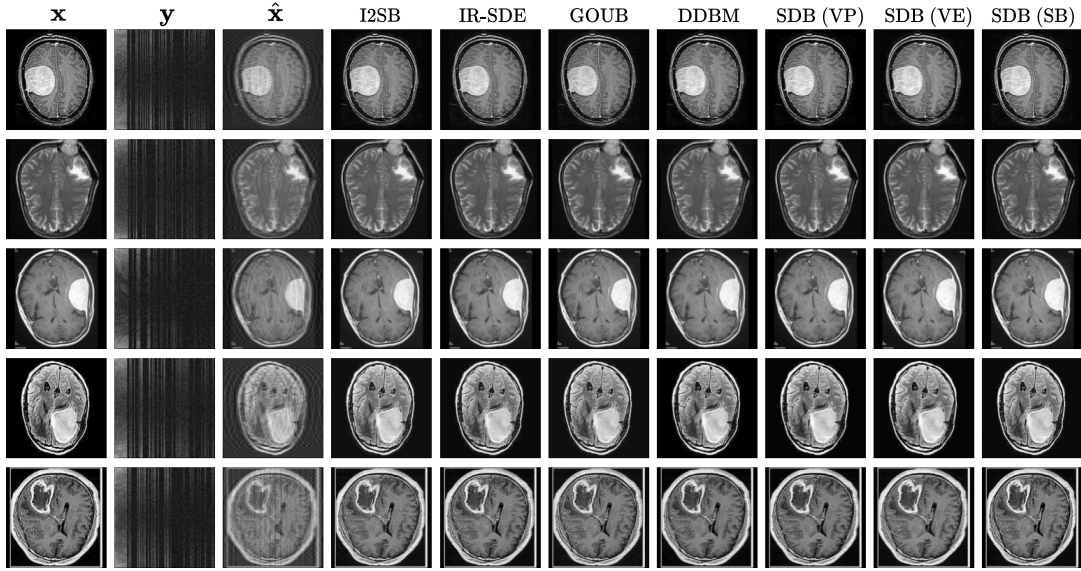


Figure 8: Qualitative comparison of SDB variants with the best-performing baselines (bridge methods). Rows depict the results for MRI reconstruction.

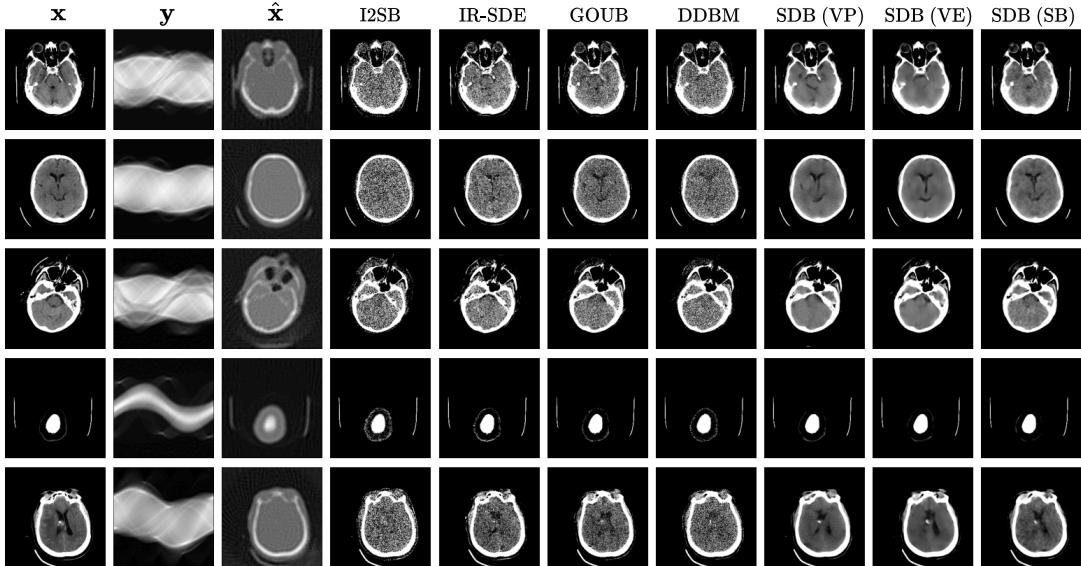


Figure 9: Qualitative comparison of SDB variants with the best-performing baselines (bridge methods). Rows depict the results for CT reconstruction.