# Efficient Bayesian Computational Imaging with a Surrogate Score-Based Prior: Supplementary Materials

**Anonymous Author(s)**
Affiliation
Address
email

## Contents



## 1 Forward models

In our experiments, we considered forward models of the form

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \epsilon, \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \sigma_{\mathbf{y}}^2, \mathbf{I}). \tag{1}$$

This corresponds to the log-likelihood function

$$\log p(\mathbf{y} \mid \mathbf{x}) \propto -\frac{1}{2\sigma_{\mathbf{y}}^2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 . \tag{2}$$

In this section, we describe the forward models of the inverse problems mentioned in the main text: accelerated MRI, denoising, and reconstruction from low spatial frequencies ("deblurring").

Submitted to 37th Conference on Neural Information Processing Systems (NeurIPS 2023). Do not distribute.

### 1.1 Accelerated MRI

Accelerated MRI collects sparse spatial-frequency measurements in $\kappa$-space of an underlying anatomical image. As the acceleration rate increases, the number of measurements decreases. In accelerated MRI, the forward model can be written as

$$\mathbf{y} = \mathbf{M} \odot \mathcal{F}(\mathbf{x}^*) + \epsilon, \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \sigma_{\mathbf{y}}^2 \mathbf{I}), \tag{3}$$

where $\mathbf{x} \in \mathbb{C}^D$ and $\mathbf{y} \in \mathbb{C}^M$. $\mathcal{F}$ denotes the 2D Fourier transform, and $\mathbf{M} \in \{0, 1\}^D$ is a binary sampling mask that reduces the number of non-zero measurements to $M << D$. Often $\sigma_{\mathbf{y}}$ is assumed to be small (e.g., corresponding to an SNR of at least 30 dB). We use Poisson-disc sampling [5] to obtain a sampling mask. $16\times$-acceleration, for example, corresponds to a sampling mask with only $1/16$ nonzero elements.

**Experimental setup.** In our experiments, we assumed that $|\sigma_{\mathbf{y}}|$ is $0.05\%$ of the DC (zero-frequency) amplitude. This corresponds to a maximum SNR of 40 dB. The only exception is for comparison to baselines (Fig. 5), since baseline methods do not account for measurement noise. In this case, we let $|\sigma_{\mathbf{y}}| = 0.1\%$ of the DC amplitude along the horizontal direction of the true image, which amounts to a very low level of noise.

### 1.2 Denoising

The denoising forward model is simply

$$\mathbf{y} = \mathbf{x} + \epsilon, \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \sigma_{\mathbf{y}}^2, \mathbf{I}), \tag{4}$$

where $\mathbf{x} \in \mathbb{R}^D$, and $\sigma_{\mathbf{y}}$ determines the level of i.i.d. Gaussian noise added to the clean image to get $\mathbf{y} \in \mathbb{R}^D$.

**Experimental setup.** In our presented experiments on denoising, $\sigma_{\mathbf{y}} = 0.2$, which is $20\%$ of the dynamic range of the image.

### 1.3 Deblurring

We refer to the task of reconstruction from low spatial-frequency measurements as deblurring. The forward model is given by

$$\mathbf{y} = \mathbf{D}\mathbf{x} + \epsilon, \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \sigma_{\mathbf{y}}^2, \mathbf{I}), \tag{5}$$

where $\mathbf{x} \in \mathbb{C}^D$ and $\mathbf{y} \in \mathbb{C}^M$. $\mathbf{D} \in \mathbb{C}^{M \times D}$ is the 2D discrete Fourier transform (DFT) matrix with only the first $M$ basis functions.

**Experimental setup.** In our presented experiments on deblurring, the measurements are the lowest $6.25\%$ of the DFT components, and $|\sigma_{\mathbf{y}}| = 1$.

## 2 Variational distributions

### 2.1 RealNVP

The architecture of the RealNVP is determined by the number of affine-coupling layers and the width of each layer. For images up to $64 \times 64$, we use 32 affine-coupling layers and set the number of hidden neurons in the first layer to $1/8$ of the image dimensionality (e.g., $32 \cdot 32 \cdot 3/8$ for $32 \times 32$ RGB images). We use batch normalization in the network. Please refer to the original DPI [4] PyTorch implementation[1] for details on the architecture. Our implementation is an adaptation of this codebase in JAX.

### 2.2 Gaussian

Other experiments use a multivariate Gaussian distribution with a diagonal covariance matrix as the variational distribution. In this case, the parameters are the mean image and the pixel-wise standard deviation. We initialize the mean at 0.5 and the standard deviation at 0.1 for all pixels. To sample, we take the absolute value of the standard deviation and construct the diagonal covariance matrix.

[1] `https://github.com/HeSunPU/DPI`

## 3 Experiment details

For the sake of reproducibility, we detail the experimental setup behind each figure. Our code will be made publicly available.

Some common implementation details are that the exact prior ($\log p_\theta^{\text{ODE}}$) is always estimated with $16$ trace estimators. The RealNVP always has 32 affine-coupling layers.

### 3.1 MRI efficiency experiment (Tab. 1, Fig. 2)

**Score model.** For each image size, the score model was an NCSN++ architecture with 64 filters in the first layer and trained with the VP SDE with $\beta_{\text{min}} = 0.1$, $\beta_{\text{max}} = 10$.

**Variational optimization.** For each task (i.e., each image size and prior), the variational distribution was a multivariate Gaussian with diagonal covariance. The batch size was 64, learning rate 0.0002, and gradient clip 1. A convergence criterion based on the loss value is difficult to define due to high variance of the loss (we used 1 time sample to estimate $b_\theta(\mathbf{x})$). We defined a convergence criterion based on the change in the mean of the variational distribution. Specifically, every 10000 steps, we evaluated a snapshot of the variational Gaussian and computed $\delta = \|\mu_{\text{curr}} - \mu_{\text{prev}}\| / \|\mu_{\text{prev}}\|$, where $\mu_{\text{curr}}$ and $\mu_{\text{prev}}$ are the current and previous snapshot means, respectively. If $\delta < \varepsilon$ for some threshold $\varepsilon$ two snapshots in a row, then the optimization was considered converged. Since convergence rate depends on the image size and the prior used, we set a different $\varepsilon$ for each task:

- $16 \times 16$ (surrogate): $\varepsilon = 0.002$
- $32 \times 32$ (surrogate): $\varepsilon = 0.003$
- $64 \times 64$ (surrogate): $\varepsilon = 0.005$
- $128 \times 128$ (surrogate): $\varepsilon = 0.007$
- $256 \times 256$ (surrogate): $\varepsilon = 0.009$
- $16 \times 16$ (exact): $\varepsilon = 0.0025$
- $32 \times 32$ (exact): $\varepsilon = 0.0027$
- $64 \times 64$ (exact): $\varepsilon = 0.005$

We were conservative in defining the convergence and checked that optimization under the surrogate actually achieved better sample quality than optimization under the exact prior (see Main Fig. 2).

**Data.** The test image is from the fastMRI [6] single-coil knee test dataset and was resized to $64 \times 64$ with antialiasing.

### 3.2 256x256 MRI examples (Fig. 1)

The $4\times$-acceleration result is from the efficiency experiment (Main Tab. 1 and Fig. 2) on the $256 \times 256$ test image. The $16\times$-acceleration result came from a similar setup, where the variational distribution was Gaussian with diagonal covariance. Optimization was done with a batch size of 64, learning rate of 0.00001, and gradient clip of 0.0002. We ran optimization for 270K steps (optimization for $4\times$-acceleration was done in 100K steps with the convergence criterion).

In the figure caption, we report that the true image is within three standard deviations of the inferred posterior mean for $96\%$ and $99\%$ of the pixels for $16\times$- and $4\times$-acceleration, respectively. This was computed based on the mean and standard deviation of 128 samples from the inferred posterior. We find the same result when using the exact mean and standard deviation of the inferred posterior: with respect to the inferred posterior, the true image is within three standard deviations of the mean for $96.7\%$ and $99.0\%$ of the pixels for $16\times$- and $4\times$-acceleration, respectively.

### 3.3 Ground-truth posteior (Fig. 3a)

**Data.** The mean and covariance of the ground-truth Gaussian prior were fit with PCA (with 256 principal components) to training data from the CelebA dataset [1]. The CelebA images were resized to $16 \times 16$ with antialiasing.

**Score model.** The score model was based on the DDPM++ deep continuous archictecture of Song et al. [2] with 128 filters in the first layer. It was trained with the VP SDE with $\beta_{\text{min}} = 0.1$ and $\beta_{\text{max}} = 20$ for 100K steps.

**Variational optimization.** The variational distribution was a RealNVP. Under the surrogate prior, optimization was done with a learning rate of 0.00005 and gradient clip of 1. Under the exact prior, the learning rate was 0.0002 and gradient clip 1. Both priors used a batch size of 64.

### 3.4 32x32 image denoising (Fig. 3b)

**Variational optimization.** For both CelebA denoising (i) and CIFAR-10 denoising (ii), the variational distribution was a RealNVP. Optimization under the surrogate prior was done with a learning rate of 0.00001 and gradient clip of 1. For CelebA, the batch size was 64 and training was done for 1.72M steps (convergence was probably achieved earlier, but we continued training to be conservative). For CIFAR-10, the batch size was 128 and training was done for 550K steps. For both (i) and (ii), optimization under the exact prior was done with a learning rate of 0.0002 and gradient clip of 1 for 20K steps.

**Score model.** For both (i) and (ii), the score model had an NCSN++ architecture with 64 filters in the first layer. For the CelebA prior, it was trained with the VP SDE with $\beta_{\text{min}} = 0.1$ and $\beta_{\text{max}} = 20$ and with images that were resized without antialiasing. For the CIFAR-10 prior, it was trained with the VP SDE with $\beta_{\text{min}} = 0.1$ and $\beta_{\text{max}} = 10$.

**Data.** Both the CelebA image and the CIFAR-10 image are $32 \times 32$. The CelebA image was resized without antialiasing.

### 3.5 Bound gap (Fig. 4)

Visualization of the bound gap is shown for optimization of the RealNVP from Fig. 3b(i) (i.e., $32 \times 32$ CelebA denoising). For the plots comparing the lower-bound to the ODE log-probability, we used 2048 time samples to estimate $b_\theta(\mathbf{x})$.

### 3.6 Image-restoration metrics (Fig. 5)

**Score model.** The score model is the same as the one used for the $64 \times 64$ image in the MRI efficiency experiment (Main Fig. 2).

**Variational optimization.** The variational distribution was a RealNVP. Optimization was done with a learning rate of 0.00001 and gradient clip of 0.0002. We used the same convergence criterion as the one used in the MRI efficiency experiment with $\varepsilon = 0.005$.

**Baseline hyperparameters.** For SDE+Poj, we used the `projection` CS solver provided by Song et al. [3] with the hyperparameters `snr=0.517, coeff=1`. For Score-ALD, we used the `langevin` CS solver with the hyperparameters `n_steps_each=3, snr=0.212, projection_sigma_rate=0.713`. For DPS, we used `scale=0.5`. This was the best scale out of $[10, 1, 0.9, 0.5, 0.3, 0.1, 0.001]$ for a test image in terms of PSNR with respect to the true image.

## References


[1] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of International Conference on Computer Vision (ICCV)*, December 2015.

[2] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *ICLR*, 2021. URL `https://openreview.net/forum?id=PxTIG12RRHS`.

[3] Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon. Solving inverse problems in medical imaging with score-based generative models. In *ICLR*, 2022. URL `https://openreview.net/forum?id=vaRCHVj0uGI`.

[4] He Sun and Katherine L Bouman. Deep probabilistic imaging: Uncertainty quantification and multi-modal solution characterization for computational imaging. In *AAAI*, pages 2628–2637, 2021.

[5] Muhammad Usman and Philipp G Batchelor. Optimized sampling patterns for practical compressed mri. In *SAMPTA'09*, pages Poster–session, 2009.

[6] Jure Zbontar, Florian Knoll, Anuroop Sriram, Tullie Murrell, Zhengnan Huang, Matthew J Muckley, Aaron Defazio, Ruben Stern, Patricia Johnson, Mary Bruno, et al. fastmri: An open dataset and benchmarks for accelerated mri. *arXiv preprint arXiv:1811.08839*, 2018.