Fast and Robust Diffusion Posterior Sampling for MR Image Reconstruction Using the Preconditioned Unadjusted Langevin Algorithm

Moritz Blumenthal*[1], Tina Holliber*[1], Jonathan I. Tamir[2,3], and Martin Uecker[1,4]

[1]Institute of Biomedical Imaging, Graz University of Technology, Graz, Austria
[2]Chandra Family Department of Electrical Engineering, University of Texas at Austin, USA
[3]Department of Diagnostic Medicine, Dell Medical School, University of Texas at Austin, USA
[4]BioTechMed-Graz, Graz, Austria

December 8, 2025

Abstract

Purpose: The Unadjusted Langevin Algorithm (ULA) in combination with diffusion models can generate high quality MRI reconstructions with uncertainty estimation from highly undersampled k-space data. However, sampling methods such as diffusion posterior sampling or likelihood annealing suffer from long reconstruction times and the need for parameter tuning. The purpose of this work is to develop a robust sampling algorithm with fast convergence.

*These authors contributed equally to this work.

Correspondence to: Martin Uecker, Graz University of Technology, Institute of Biomedical Imaging, Stremayrgasse 16/3, 8010 Graz, AUSTRIA, Email: uecker@tugraz.at

Theory and Methods: In the reverse diffusion process used for sampling the posterior, the exact likelihood is multiplied with the diffused prior at all noise scales. To overcome the issue of slow convergence, preconditioning is used. The method is trained on fastMRI data and tested on retrospectively undersampled brain data of a healthy volunteer.

Results: For posterior sampling in Cartesian and non-Cartesian accelerated MRI the new approach outperforms annealed sampling in terms of reconstruction speed and sample quality.

Conclusion: The proposed exact likelihood with preconditioning enables rapid and reliable posterior sampling across various MRI reconstruction tasks without the need for parameter tuning.

Keywords: diffusion posterior sampling, MRI, image reconstruction, parallel imaging, Bayesian reconstruction

1 Introduction

In recent years, diffusion models have shown impressive results in image generation by providing an approach for sampling from a high-dimensional probability distribution. In a Bayesian setting, they can be used as a learned prior to perform MRI reconstruction even from highly undersampled k-space data by sampling the posterior probability distribution [1–3]. This Bayesian approach to image reconstruction has several key advantages over other deep learning methods: (i) by decoupling the measurement model from the prior, no retraining is necessary when the measurement operator changes (e.g. due to different sampling trajectories, coil arrays, motion, field inhomogeneities); (ii) the measurement noise level uniquely determines the relative weighting of prior and likelihood, i.e. no tuning of a regularization parameter is required; and (iii) uncertainty maps computed based on the full posterior can show when the reconstruction becomes unreliable due to an insufficient amount of data [3].

While sampling the unconditional prior can be done efficiently using a reverse diffusion process based on a learned series of smoothed prior distributions [4–6], efficient sampling of the posterior distribution based on a realistic acquisition model such as SENSE [7] is still a challenging problem. Various modifications of the likelihood term were proposed [8, 9] to improve the sampling of the posterior distribution in a reverse diffusion process, including annealed [1] or noisy [10, 11] likelihoods. These methods are relatively slow and require a careful choice of step size and reverse-diffusion noise schedule to achieve good results. Crucially, these parameters must be tuned for different measurement models, thus losing some of the purported flexibility offered by the Bayesian approach. In particular, this affects non-Cartesian sampling, where the highly non-uniform sampling of k-space leads to slow convergence.

In this work, we observe that the practical difficulties encountered when sampling the posterior distribution are caused by the ill-conditioning of the problem. For convergence, small step sizes and many noise scales must be used which then prevents effective sampling. We tackle this problem by preconditioning: Degrees of freedom that correspond to large singular values of the SENSE model, i.e. which are strongly restricted by the measurements, are refined with small updates, while degrees of freedom corresponding to small singular values are refined with large updates allowing them to traverse the probability mass given by the prior distribution more freely. Our approach can then be used with a fixed pre-chosen step size and thus eliminate the need to tune it for different acquisitions.

We tested the proposed method on both Cartesian and non-Cartesian radial brain data, showing consistently better reconstruction quality and lower computation time compared to annealing without any need for parameter tuning.

2 Theory

2.1 Bayesian Reconstruction

MRI reconstruction can be formulated as the inverse problem

$$\mathbf{y} = A\mathbf{x} + \mathbf{n}, \qquad \mathbf{n} \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \mathbb{I}) \tag{1}$$

with k-space data $\mathbf{y}$, the SENSE model A, image $\mathbf{x}$ and white complex Gaussian noise $\mathbf{n}$ of unit variance (after prewhitening, normalization, and adapting the coil sensitivities accordingly). From the Bayesian perspective, the posterior probability density $p(\mathbf{x}|\mathbf{y})$ is given by

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})} \propto \exp\left(-\|\mathbf{y} - A\mathbf{x}\|_2^2 + \log p(\mathbf{x})\right) , \tag{2}$$

with the prior $p(\mathbf{x})$ and the likelihood $p(\mathbf{y}|\mathbf{x})$. From the posterior distribution, point estimators such as the maximum a posteriori (MAP) estimate, which corresponds to conventional image reconstruction with regularization, or the minimum mean square error (MMSE) estimate can be computed. To estimate the MMSE, multiple samples can be drawn from the posterior using the unadjusted Langevin algorithm (ULA) and averaged. For complex valued random vectors, the ULA update reads

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \gamma \nabla_{\bar{\mathbf{x}}} \log p(\mathbf{x}^k|\mathbf{y}) + \sqrt{2\gamma}\mathbf{z}^k \qquad \mathbf{z}^k \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \mathbb{I}) , \tag{3}$$

where $\nabla_{\bar{\mathbf{x}}} \log p(\mathbf{x}|\mathbf{y})$ is the score function of the posterior distribution and $\nabla_{\bar{\mathbf{x}}}$ is the complex conjugate gradient operator of Wirtinger calculus [12]. The corresponding complex likelihood score is given by

$$\nabla_{\bar{\mathbf{x}}} \log p(\mathbf{y}|\mathbf{x}) = A^H(\mathbf{y} - A\mathbf{x}) . \tag{4}$$

For convergence and to reduce the discretization bias, the step size γ must be sufficiently small compared to the inverse Lipschitz L^{-1} constant of the posterior score [13, 14]. The score $\nabla_{\bar{\mathbf{x}}} \log p(\mathbf{x}_0)$ of a smoothed version of the prior can be obtained from a training data set of images using denoising score matching (DSM) [4].

2.2 Posterior Sampling

The goal of posterior sampling is to draw samples from the posterior distribution in Eq. 2, where the prior distribution $p(\mathbf{x}_0)$ is learned. Because direct sampling of a high-dimensional multi-modal distribution is not practically possible, a reverse diffusion process is used. For sampling of the prior [4, 5], this can be done by successive sampling of a series of priors $p(\mathbf{x}_t)$ smoothed by convolutions with complex Gaussians of variance σ_t^2, starting from $t = 1$ with a simple complex Gaussian distribution of variance $\sigma_1^2 = \sigma_{\max}^2$ that can be sampled directly until the lowest noise scale $\sigma_0 = \sigma_{\min}$ is reached, where the learned score is assumed to approximate the true prior score well.

A straightforward idea is to formulate a diffusion process for the posterior in the same way as for the prior by adding Gaussian noise [15]. To be able to use the learned prior, at each time t the diffused posterior $p(\mathbf{x}_t|\mathbf{y})$ should factorize into the diffused prior $p(\mathbf{x}_t)$ and the conditional probability for measuring $\mathbf{y}$ when observing a perturbed sample $\mathbf{x}_t$. By marginalization over the unknown noiseless images $\mathbf{x}$ one arrives at

$$p^{\text{diffused}}(\mathbf{y}|\mathbf{x}_t) = \int p(\mathbf{y}|\mathbf{x})p(\mathbf{x}|\mathbf{x}_t)d\mathbf{x} \ . \tag{5}$$

We call this approach *Diffused Posterior*. However, this likelihood term needs to be approximated in practice, as it makes use of the posterior for the denoising problem $p(\mathbf{x}|\mathbf{x}_t)$, which is simpler than the full posterior but still intractable. This problem has led to various approximations for the term $p^{\text{diffused}}(\mathbf{y}|\mathbf{x}_t)$. Jalal et al. [1] introduced a weighting term κ_t added to the data noise during the diffusion process:

$$p^{\text{annealed}}(\mathbf{y}|\mathbf{x}_t) \propto \exp\left(-\frac{1}{1+\kappa_t}\|\mathbf{y} - A\mathbf{x}_t\|_2^2)\right), \tag{6}$$

where $\kappa_t \to 0$ for decreasing noise scale σ_t which we call *Annealed Likelihood*. Chung et al. [10] used the score output by the network to calculate the expectation $\mathbb{E}[\mathbf{x}_0|\mathbf{x}_t]$ in order to approximate $p(\mathbf{x}|\mathbf{x}_t) \approx \delta(\mathbf{x} - \mathbb{E}[\mathbf{x}_0|\mathbf{x}_t])$ at every noise scale.

While diffusion the posterior distribution is conceptually straightforward but practically challenging, an alternative approach is to use a different reverse process so long as the correct posterior distribution is obtained at $t = 0$. This insight was first described in Sohl-Dickstein et al. [16], in which the diffusion process was modified by multiplication with a function (e.g., the likelihood term). Hence, the learned prior distribution $p(\mathbf{x}_t)$ can be multiplied by a modified likelihood term $p(\mathbf{y}|\mathbf{x}_t)$, which smoothly varies with

t and is equal to the true likelihood $p(\mathbf{y}|\mathbf{x})$ at $t = 0$. Here, we investigate the use of the *Exact Likelihood* term,

$$p^{\text{exact}}(\mathbf{y}|\mathbf{x}_t) \propto \exp\left(-\|\mathbf{y} - A\mathbf{x}_t\|_2^2\right), \tag{7}$$

for the whole diffusion process, at every noise scale. To motivate this choice, we show the effect of the choice of the likelihood for a 2D toy model in Figure 1A that can be analytically computed.

2.3 Preconditioning of ULA

A naive use of the *Exact Likelihood* method is impractical because some directions are then constrained by the data already at high noise scales, implying a large Lipschitz constant corresponding to the maximum eigenvalue of $A^H A$ in the posterior score, exacerbated in non-Cartesian sampling. This then prevents the use of larger step sizes, leading to slow convergence. To solve this issue, we propose a preconditioned ULA (pULA) with the update rule [17–20]:

$$\mathbf{x}_t^{k+1} = \mathbf{x}_t^k + \gamma M_t \left[A^H(\mathbf{y} - A\mathbf{x}_t^k) + \nabla_{\bar{\mathbf{x}}} \log p_t(\mathbf{x}_t^k)\right] + \sqrt{2\gamma}\mathbf{z}^k \quad \mathbf{z}^k \sim \mathbb{C}\mathcal{N}(\mathbf{0}, M_t^{-1}), \tag{8}$$

with a preconditioning matrix M_t specifically adapted to linear inverse problems, i.e. $M_t = (A^H A + \sigma_t^{-2}\mathbb{I})^{-1}$. Here, the superscript denotes the kth Langevin update per diffusion time t. At $t = 1$, the reverse-diffusion process is started with a sample from the posterior distribution corresponding to a flat Gaussian prior with variance $\sigma_1^2 = \sigma_{\max}^2$, i.e.

$$\mathbf{x}_{t=1}^{k=0} \sim \mathbb{C}\mathcal{N}(M_1 A^H \mathbf{y}, M_1^{-1}). \tag{9}$$

At each noise level σ_t, we perform K Langevin steps of pULA before the noise level is reduced, and pULA is initialized with the last sample from the higher noise level.

Our choice of the precondition matrix M_t is based on the following observation: in combination with the likelihood, the preconditioner allows larger updates in directions of small singular values of A, i.e., those which are only weakly determined by the measurements. In these directions, the preconditioner effectively yields a step size proportional to the variance of the diffusion noise σ_t^2 as empirically found to work well for prior sampling [4].

The initialization and update rule of pULA require drawing samples $\mathbf{z}$ from the distribution $\mathbb{C}\mathcal{N}(\mathbf{0}, M_t^{-1})$. Those can be drawn efficiently using

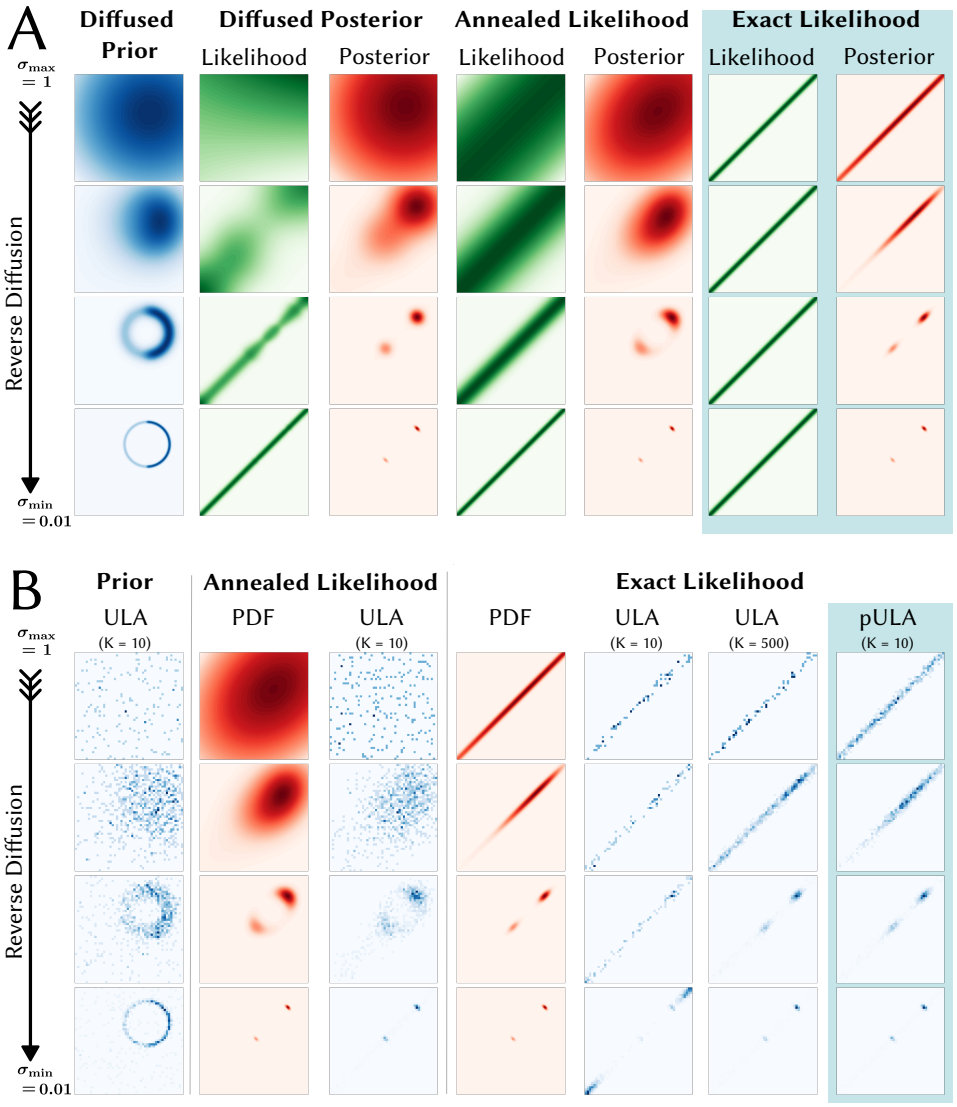

Figure 1: A: Analytical diffusion process for a 2D toy model with a 1D linear measurement $A = (1 \quad -1)^T$. The prior distribution is a mixture of 2D Gaussians forming a circle, which models the data manifold. Likelihood modifications corresponding to the diffused posterior and annealing are compared to the exact likelihood. All methods have the same posterior distribution at the minimum noise scale. B: Samples of the same 2D toy model with annealed (left) and exact (right) likelihood sampled with ULA for 10 and 500 iterations and pULA for 10 iterations. Samples are shown for noise levels $\sigma = [1, 0.2, 0.05, 0.01]$

[21]

$$\mathbf{z} = M_t \left(A^H \mathbf{n}_1 + \sigma_t^{-1} \mathbf{n}_2 \right) \quad \begin{pmatrix} \mathbf{n}_1 \\ \mathbf{n}_2 \end{pmatrix} \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \mathbb{I}) \,, \tag{10}$$

similar to the pseudo replica method [22]. For the initialization, the mean can be added to the zero-mean sample. Inserting Equation 10 into Equation 8, the update of pULA for the posterior score can be written as

$$\begin{aligned} \mathbf{x}_t^{k+1} = \mathbf{x}_t^k + \gamma M_t \Bigg[& A^H \left(\mathbf{y} + \sqrt{\frac{2}{\gamma}} \mathbf{n}_1^k - A \mathbf{x}_t^k \right) \\ & + \nabla_{\bar{\mathbf{x}}} \log p(\mathbf{x}_t^k) + \sqrt{\frac{2}{\gamma \sigma_t^2}} \mathbf{n}_2^k \Bigg] \,, \quad \begin{pmatrix} \mathbf{n}_1^k \\ \mathbf{n}_2^k \end{pmatrix} \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \mathbb{I}) \,. \end{aligned} \tag{11}$$

In this form, it is apparent that the pULA update requires only one application of M_t, which can be performed using the conjugate gradient (CG) method without explicitly forming the preconditioning matrix.

3 Methods

3.1 Data Processing and Network Training

For training, we used images reconstructed from the fully sampled k-space data of the multi-coil train folder of the fastMRI brain dataset [23] that have a 320×320 matrix size. The k-space data were pre-whitened using a noise covariance matrix estimate from the background patches of the fully sampled coil images. The data were then compressed to 12 virtual coils, and coil sensitivities were estimated with a low-resolution version of NLINV [24]. Using the corresponding reconstruction, a scaling for the k-space was estimated such that the final reconstruction is normalized to have a 99 percentile pixel magnitude of one. A foreground mask for the images was estimated using ESPIRiT [25]. Data for which this mask extends to the image boundary in the phase encoding direction were assumed to be corrupted, for example by motion, and removed. The final images for training were then reconstructed using FISTA with a small ℓ_1-Wavelet regularization. Using these high quality images, we trained a conditional denoising U-Net with conditional residual blocks [26] implemented in BART [27]. From the denoising U-Net, the score network is obtained using Tweedie's formula [22, 28].

3.2 MRI Data Acquisition

MRI data of a healthy volunteer was acquired on a 3T scanner (Magnetom Vida, Siemens Healthineers, Erlangen, Germany) after obtaining written informed consent and with approval of the local ethics committee. We acquired fully sampled multi-slice Cartesian T2-weighted brain data with a 20-channel head coil using a Turbo Spin Echo sequence (FOV = 250 mm, matrix size: 320×320, TR = 6000 ms, TE = 98 ms, ΔTE = 9.82 ms, FA = 150°, voxel size $0.8 \times 0.8 \times 3\,\mathrm{mm}^3$, ETL = 16, BW = $223\,\mathrm{Hz\,px}^{-1}$). In addition, a radial T1-weighted scan was acquired using a stack-of-stars FLASH sequence with a RAGA [29] sampling scheme (FOV = 250 mm, matrix size: 320×320, TR = 8.0 ms, TE = 3.32 ms, FA = 10°, voxel size: $0.8 \times 0.8 \times 3\,\mathrm{mm}^3$, BW = $780\,\mathrm{Hz\,px}^{-1}$). Two repetitions with 987 spokes each were acquired, *i.e.*, two RAGA full frames.

3.3 Numerical Experiments

We first tested the exact likelihood approach for a real-valued 2D toy model with a linear measurement operator $A = \begin{pmatrix} 1 & -1 \end{pmatrix}^T$ and an analytical prior defined by a Gaussian mixture model. pULA with the exact likelihood was compared to ULA with exact and annealed likelihood, where in total $N = 101$ noise scales were used with $\sigma_{\max} = 1$ and $\sigma_{\min} = 0.01$.

Then, the proposed exact likelihood method was compared to an ℓ_1-Wavelet regularized parallel imaging reconstruction and the annealed likelihood approach on a slice of the T2-weighted dataset. For this, the fully sampled k-space data were pre-whitened based on noise extracted from a background patch of the fully sampled coil images. The data were then retrospectively undersampled with both equispaced and randomized undersampling masks (acceleration 4 and 8) with a 16-line auto-calibration region. Similar to the processing of the training data, the undersampled data were coil compressed to 12 virtual channels, coils were estimated with NLINV, a foreground mask was computed with ESPIRiT, and a normalization scale was computed from the low-resolution NLINV reconstruction. Instead of scaling the k-space data, we absorb this normalization constant in the coil sensitivities, such that the k-space data stays white with unit variance noise and the scaling of the corresponding image remains consistent with normalization used for the training images.

The sampling techniques were implemented in BART. For the annealed

likelihood, we choose the annealing parameter κ_t such that

$$\frac{1}{1+\kappa_t} = \left(\frac{\sigma_{\max}^{-2}}{\lambda_{\max}(A^H A)}\right)^t, \tag{12}$$

where $\lambda_{\max}(A^H A)$ is the maximum eigenvalue of $A^H A$, which we estimate using power iterations. At $t = 1$, this choice balances the Lipschitz constants of the likelihood and the prior scores. For ULA of the annealed likelihood, we chose the step size $\gamma = \gamma_{\text{base}} \left[(1+\kappa_t)^{-1}\lambda_{\max}(A^H A) + \sigma_t^{-2}\right]^{-1}$ with a base scaling $\gamma_{\text{base}} = 0.5$ in all experiments. For pULA, we use the same step size for all noise levels, namely, $\gamma = 0.5$.

For the diffusion reconstruction of the T2-weighted dataset, we exponentially reduced the diffusion noise level σ_t from $\sigma_{\max} = 0.1$, $\sigma_{\max} = 1$, or $\sigma_{\max} = 10$ down to $\sigma_{\min} = 0.01$. Ten samples were drawn using ULA with the annealed likelihood and using pULA with exact likelihood, and the MMSE estimate was computed for both by averaging the samples. ULA was performed with $K = 8$ Langevin iterations, whereas pULA was performed with $K = 4$ Langevin iterations and 10 CG iterations for preconditioning. The regularization parameter for the undersampled ℓ_1-Wavelet reconstruction was chosen by a grid search to obtain the best PSNR with respect to the fully sampled reconstruction. For all reconstructions, the time per sample was measured. All images were normalized by multiplication with the RSS of the coils to obtain an RSS scaling of the final reconstructions. Afterward, PSNR and SSIM values and error maps were computed with magnitude images relative to the fully sampled ℓ_1-Wavelet reconstruction after multiplying all images with the foreground mask.

To investigate the robustness of the method under substantially different experimental conditions, we performed two additional experiments. First, the number of virtual coils used for sampling/reconstruction of the T1-weighted dataset was decreased to 4 or 1 after the pre-processing described above, i.e., the full pre-processing was performed using 12 virtual coils. Second, we performed reconstructions of the radial data, where we reduced the number of radial spokes from 987 spokes to 98, 49, and 24 spokes. To discard data during the transition towards the FLASH steady state, we used the data of the second frame only. The radial data was processed similar to the Cartesian data, but first an inverse FFT in partition direction was performed to allow for slice-wise processing. The gradient delays were corrected with RING [30]. To compute the ESPIRiT foreground mask, k-space data was gridded first, whereas coil sensitivities were directly estimated from the radial data with NLINV. In both cases, an ℓ_1-Wavelet reconstruction was

also performed, and 10 samples were drawn in the diffusion reconstructions to compute the MMSE estimator. For both experiments, $\sigma_{\max} = 1$ was used.

All numerical experiments have been performed on a system with an AMD EPYC 9334 CPU and an Nvidia H100 GPU (80 GB HBM3, SXM). We record and compare total runtime for all algorithms.

4 Results

Figure 1B shows the 2D toy example of the sampling results for the reverse diffusion process comparing the annealed and the exact likelihood method for (p)ULA. In total 1000 samples were drawn. The annealed posterior adapts during the diffusion process, whereas the exact posterior does not change. For time zero, all methods converge to the correct posterior distribution. However, there are different convergence speeds. For the exact likelihood, ULA needs a small step size and therefore more iterations ($K = 500$) to converge to the correct distribution. pULA with exact likelihood shows faster convergence, similar to the annealed approach ($K = 10$).

Sampling results for the Cartesian data are shown in Figure 2. While annealed and exact likelihood both yield good reconstructions, for identical undersampling factor and diffusion time, the exact likelihood with pULA is consistently outperforming the annealed likelihood method in terms of PSNR and SSIM as well as showing reduced visible differences in the error maps. This is despite using only $K = 4$ pULA iterations instead of $K = 8$ iterations per noise level in the annealed case, and overall requiring less computation time. The exact likelihood approach with pULA achieves good results even when starting with small noise scales.

The diffusion processes for 8× randomly undersampled k-space data are shown in Figure 3. When using the exact likelihood, even at high noise levels, noise is only visible in k-space locations that are not acquired. In contrast, for the annealed likelihood, noise appears in all k-space lines at the beginning of the reverse diffusion process. At a high noise scale, the annealed method first generates what seems to resemble a T1-weighted image with dark CSF, before the data term becomes dominant enough to guide the reconstruction towards the correct image. This behavior does not occur when using the exact likelihood.

Results for a different number of virtual coils when using ℓ_1-Wavelet regularization, the exact likelihood, and the annealed likelihood approach are shown in Figure 4. Both machine-learning methods show good results even

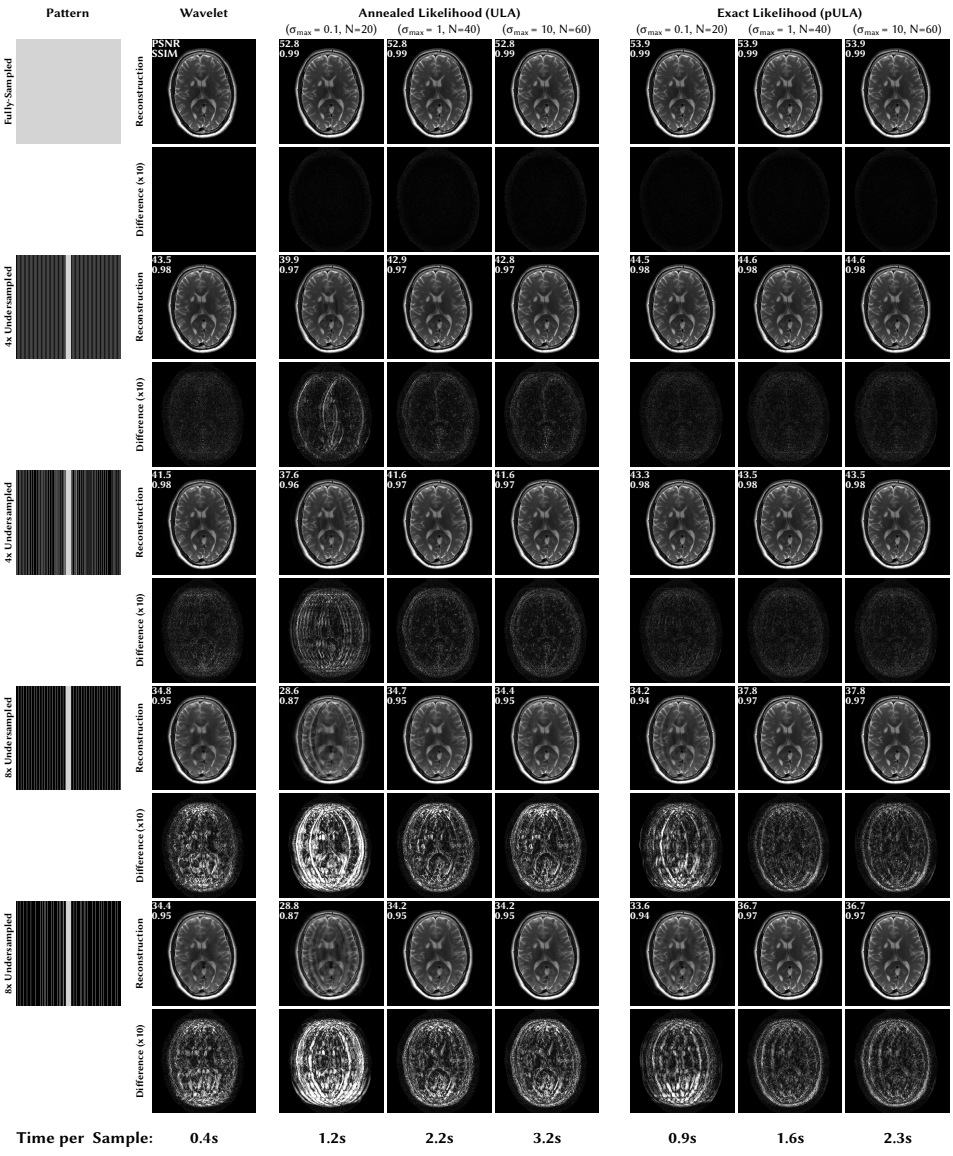

Figure 2: Reconstructions of a T2-weighted brain image for different undersampling patterns using ℓ_1-Wavelet regularization and diffusion posterior sampling with annealed and exact likelihood. Error maps and PSNR/SSIM values are computed relative to the fully-sampled ℓ_1-Wavelet reconstruction. Per noise level, $K = 8$ ULA or $K = 4$ pULA iterations were performed.

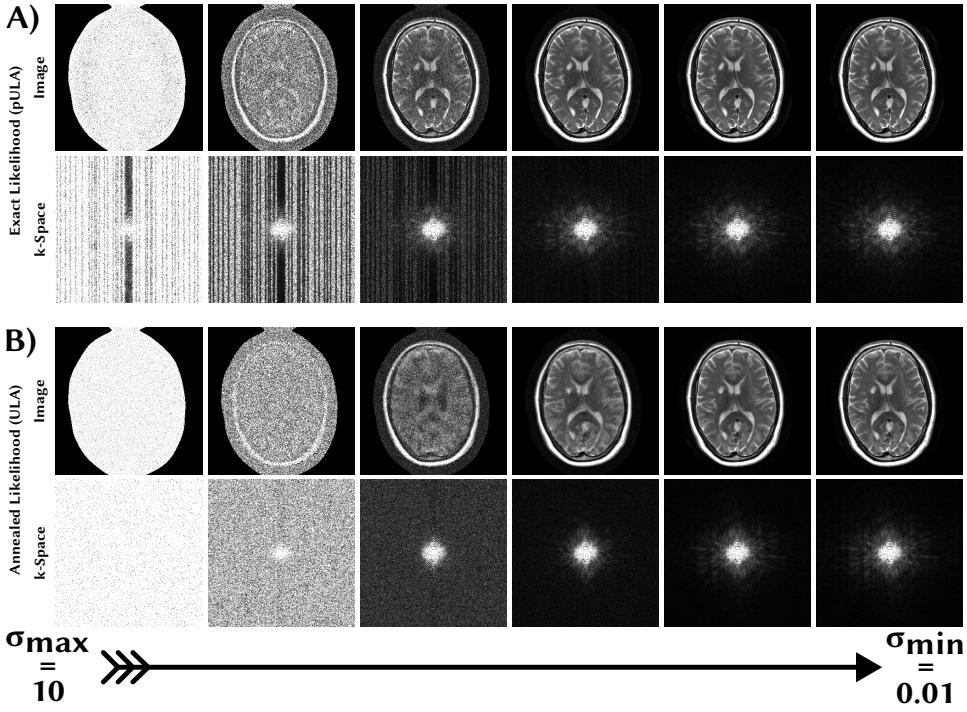

Figure 3: Reverse diffusion process with exact (A) and annealed (B) likelihood for a T2-weighted brain image sampled from 8x random undersampled data.

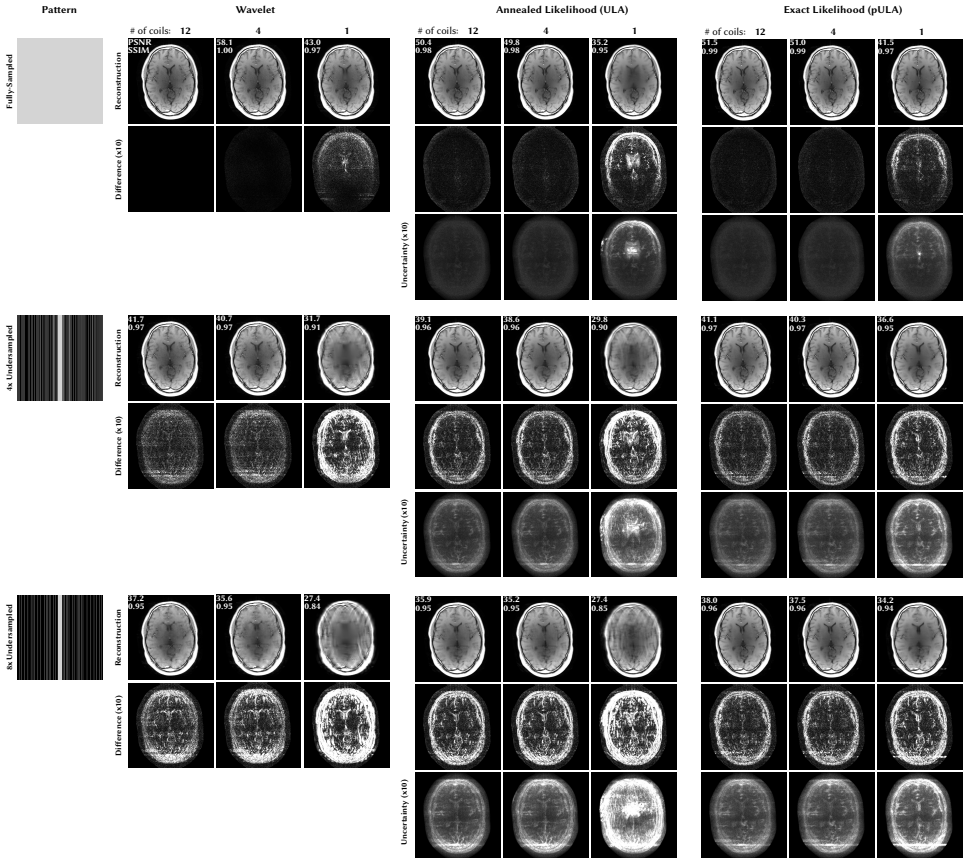

Figure 4: Reconstructions of a T1-weighted brain image for selected undersampling patterns and different numbers of virtual coils after coil compression. Uncertainty maps show the standard deviation over drawn samples.

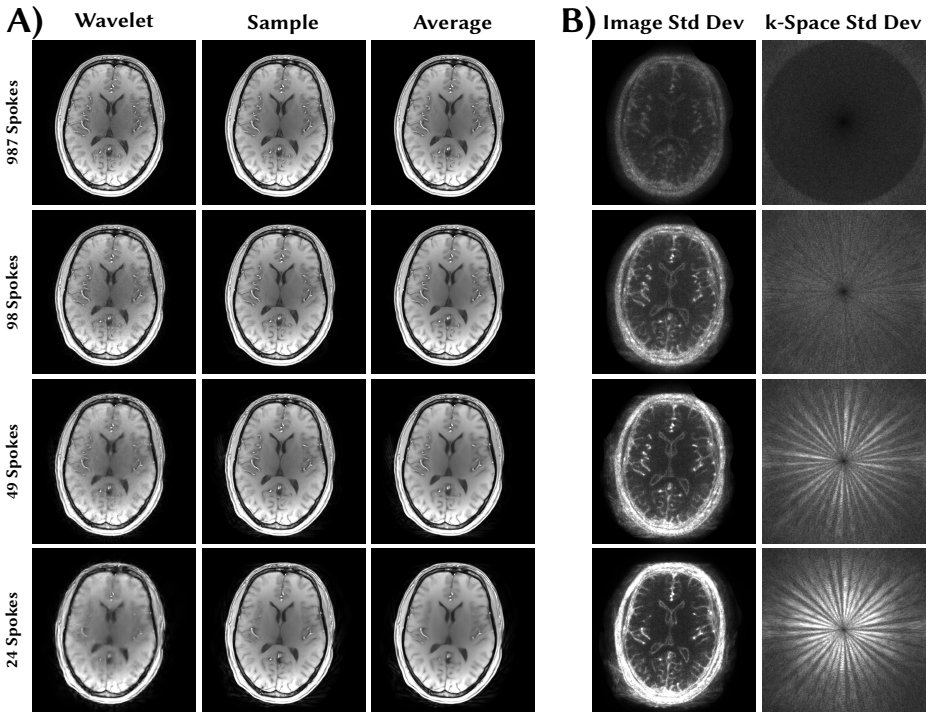

Figure 5: Brain image reconstructed from radially acquired FLASH data using different undersampling factors. A: Reconstruction with ℓ_1-Wavelet regularization, a single sample from the posterior with exact likelihood and the average of ten samples. B: Pixel-wise standard deviation map in image space and k-space.

with only four virtual coils. In the limit of only single coil reconstruction, the exact likelihood outperforms the annealed approach due to faster convergence. In the fully sampled single-coil experiment, the standard deviation maps reveal uncertainty in regions where information is lost because of coil compression, indicating areas where the network may be hallucinating.

The results of the non-Cartesian reconstruction are shown in Figure 5. While the individual samples are rich in details even in the 40-fold undersampled case, the average is slightly blurred. The pixel-wise standard deviation shows high uncertainty in regions containing small vessels.

5 Discussion

In this work, we addressed the problem that existing methods for posterior sampling for MRI reconstruction are rather slow and necessitate cumbersome parameter tuning to balance computation time and errors depending on the ill-posedness of the inverse problem. We investigated the use of an exact likelihood term rather than adjusting the likelihood in the diffusion process as done in previous publications.

In a 2D toy example, one can see that a naive use of the exact likelihood leads to slow convergence because the data restricts the problem space along one dimension, preventing large step sizes. While the annealed likelihood avoids this problem as it reduces the maximum eigenvalue, it comes at the cost of not utilizing the data properly at later diffusion times and requires careful parameter tuning. As we showed, slow convergence can be addressed effectively by preconditioning. The exact likelihood with pULA then shows similar performance in the 2D toy example as the annealed likelihood method. This observation can be extended to real MRI data where sampling with exact likelihood and pULA then consistently outperforms the annealing and shows robust behavior for different undersampling masks, numbers of virtual coils, and non-Cartesian radial MRI while using the same step size, number of iterations, and the same maximum noise level in all scenarios.

Due to the natural restriction to the problem space determined by the data, sampling only on small noise scales is sufficient for good reconstruction. Still, the annealed approach has to first explore a wider space, which includes irrelevant areas of the problem ruled out by the data. Due to the additional CG steps for preconditioning, pULA is computationally more expensive than the annealed approach for the same number of iterations and noise scales. Nevertheless, our results indicate that the exact likelihood approach with

pULA still converges faster, as it needs fewer network evaluations.

6 Conclusion

The proposed exact likelihood with preconditioned ULA enables fast and robust posterior sampling for different MRI reconstruction problems without parameter tuning. Especially ill-conditioned problems such as radial MRI benefit from the increased convergence speed.

Conflict of Interest

The authors declare no competing interests.

Data Availability Statement

In the spirit of reproducible research, the code to reproduce the results of this paper is available at `https://gitlab.tugraz.at/ibi/mrirecon/papers/dps-pula` (Version v0.1). All reconstructions have been performed with BART, available at `https://github.com/mrirecon/bart`. The data used in this study is available at Zenodo (DOI: 10.5281/zenodo.17739731).

Acknowledgements

The research was funded in whole or in part by the Austrian Science Fund (FWF) 10.55776/F100800 and National Science Foundation (NSF) CCF-2239687 (CAREER).

References

[1] Jalal A, Arvinte M, Daras G, Price E, Dimakis AG, Tamir J. Robust compressed sensing MRI with deep generative priors. In: *Advances in neural information processing systems*. Vol. 34. 2021:14938–14954.

[2] Chung H, Ye JC. Score-based diffusion models for accelerated MRI. *Medical Image Analysis* 2022; 80:102479.

[3] Luo G, Blumenthal M, Heide M, Uecker M. Bayesian MRI reconstruction with joint uncertainty estimation using diffusion models. *Magnetic Resonance in Medicine* 2023; 90:295–311.

[4] Song Y, Ermon S. Generative Modeling by Estimating Gradients of the Data Distribution. In: *Advances in Neural Information Processing Systems*. Vol. 32. 2019.

[5] Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. In: *Advances in neural information processing systems*. Vol. 33. 2020:6840–6851.

[6] Karras T, Aittala M, Aila T, Laine S. Elucidating the Design Space of Diffusion-Based Generative Models. In: *Advances in Neural Information Processing Systems*. Vol. 35. 2022:26565–26577.

[7] Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. SENSE: sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine* 1999; 42:952–962.

[8] Daras G, Chung H, Lai CH, et al. *A survey on diffusion models for inverse problems*. 2024. doi: 10.48550/arXiv.2410.00083.

[9] Chung H, Kim J, Ye JC. *Diffusion models for inverse problems*. 2025. doi: 10.48550/arXiv.2508.01975.

[10] Chung H, Kim J, Mccann MT, Klasky ML, Ye JC. Diffusion Posterior Sampling for General Noisy Inverse Problems. In: *ICLR 2023: The Eleventh International Conference on Learning Representations*. Vol. 11. 2023.

[11] Janati Y, Moulines E, Olsson J, Oliviero-Durmus A. Bridging diffusion posterior sampling and Monte Carlo methods: a survey. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 2025; 383:20240331.

[12] Kreutz-Delgado K. *The Complex Gradient Operator and the CR-Calculus*. 2009. doi: 10.48550/ARXIV.0906.4835.

[13] Holliber T, Blumenthal M, Uecker M. Unadjusted Langevin Sampling for Uncertainty Estimation in MRI Reconstruction - Theory and Numerical Validation. In: *Proceedings of the Annual Meeting of ISMRM*. 2025:2603.

[14] Dalalyan AS. Theoretical Guarantees for Approximate Sampling from Smooth and Log-Concave Densities. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 2017; 79:651–676.

[15] Kawar B, Elad M, Ermon S, Song J. Denoising Diffusion Restoration Models. In: *Advances in Neural Information Processing Systems*. Vol. 35. 2022:23593–23606.

[16] Sohl-Dickstein J, Weiss EA, Maheswaranathan N, Ganguli S. Deep Unsupervised Learning using Nonequilibrium Thermodynamics. In: *Proceedings of the 32nd International Conference on Machine Learning*. Vol. 37. 2015:2256–2265.

[17] Roberts GO, Stramer O. Langevin Diffusions and Metropolis-Hastings Algorithms. *Methodology And Computing In Applied Probability* 2002; 4:337–357.

[18] Corbineau MC, Kouamé D, Chouzenoux E, Tourneret JY, Pesquet JC. Preconditioned P-ULA for Joint Deconvolution-Segmentation of Ultrasound Images. *IEEE Signal Processing Letters* 2019; 26:1456–1460.

[19] Marnissi Y, Chouzenoux E, Benazza-Benyahia A, Pesquet JC. Majorize–Minimize Adapted Metropolis–Hastings Algorithm. *IEEE Transactions on Signal Processing* 2020; 68:2356–2369.

[20] Bhattacharya R, Jiang T. *Fast Sampling and Inference via Preconditioned Langevin Dynamics*. 2024. doi: `10.48550/arXiv.2310.07542`.

[21] Papandreou G, Yuille AL. Gaussian sampling by local perturbations. In: *Advances in neural information processing systems*. Vol. 23. 2010.

[22] Robbins H. An Empirical Bayes Approach to Statistics. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* 1954:157–163.

[23] Knoll F, Zbontar J, Sriram A, et al. fastMRI: A Publicly Available Raw k-Space and DICOM Dataset of Knee Images for Accelerated MR Image Reconstruction Using Machine Learning. *Radiology: Artificial Intelligence* 2020; 2:e190007.

[24] Uecker M, Hohage T, Block KT, Frahm J. Image reconstruction by regularized nonlinear inversion-joint estimation of coil sensitivities and image content. *Magnetic Resonance in Medicine* 2008; 60:674–682.

[25] Uecker M, Lai P, Murphy MJ, et al. ESPIRiT—an eigenvalue approach to autocalibrating parallel MRI: Where SENSE meets GRAPPA. *Magnetic Resonance in Medicine* 2014; 71:990–1001.

[26] Song Y, Sohl-Dickstein J, Kingma DP, Kumar A, Ermon S, Poole B. Score-Based Generative Modeling through Stochastic Differential Equations. In: *ICLR 2021: The Ninth International Conference on Learning Representations*. Vol. 9. 2021.

[27] Blumenthal M, Luo G, Schilling M, Holme HCM, Uecker M. Deep, deep learning with BART. *Magnetic Resonance in Medicine* 2023; 89:678–693.

[28] Efron B. Tweedie's Formula and Selection Bias. *Journal of the American Statistical Association* 2011; 106:1602–1614.

[29] Scholand N, Schaten P, Graf C, et al. Rational approximation of golden angles: Accelerated reconstructions for radial MRI. *Magnetic Resonance in Medicine* 2025; 93:51–66.

[30] Rosenzweig S, Holme HCM, Uecker M. Simple auto-calibrated gradient delay estimation from few spokes using Radial Intersections (RING). *Magnetic Resonance in Medicine* 2019; 81:1898–1906.