Variational Diffusion Models for Blind MRI Inverse Problems

Cagan Alkan[1*] **Julio Oscanoa**[1*] **Daniel Abraham**[1] **Mengze Gao**[1] **Aizada Nurdinova**[1]
Kawin Setsompop[1] **John Pauly**[1] **Morteza Mardani**[2] **Shreyas Vasanawala**[1]
[1]Stanford University [2]NVIDIA Inc.
{calkan,joscanoa,abrahamd,gaomz,nurdaiza,
kawins,pauly,vasanawala}@stanford.edu
mmardani@nvidia.com
[*]denotes equal contribution

Abstract

Diffusion models have demonstrated state-of-the-art results in solving inverse problems in various domains including medical imaging. However, existing works generally consider the cases where the forward operator is fully known. Therefore, blind inverse problems with unknown forward operator parameters require modifications on existing methods. In this work, we present an extension of the recently developed regularization by denoising diffusion process (RED-diff) algorithm to blind inverse problems. Similarly to RED-diff, our method can reconstruct images without model re-training or fine-tuning for arbitrary acquisition settings. Tested in fieldmap-corrected MR image reconstruction, our blind RED-diff framework can successfully approximate the unknown forward model parameters and produce fieldmap-corrected reconstructions accurately.

1 Introduction

Diffusion models have shown superior performance for solving inverse problems as they are excellent generative priors (1; 2). Pre-trained diffusion models are used as strong data priors in a plug and play fashion at inference time. They are typically used in conjunction with data-consistency projection in the reverse diffusion process (3; 4; 2) or via an approximation of the posterior score function (5; 6). Recently, (7) have proposed a regularization by denoising (RED-diff) framework for solving generic inverse problems and (8) extended it for MRI reconstruction problem. RED-diff uses variational inference to approximate the posterior distribution which consequently corresponds to minimizing a data-consistency loss and score matching regularization via denoisers at different diffusion steps. The key strength of the diffusion based samplers for inverse problems is that they do not require training or fine-tuning for each specific task. However, the applicability of these diffusion based samplers are restricted to the cases where forward model is fully known. These algorithms require full characterization of the forward model, therefore they cannot be directly applied to the blind inverse problems which require estimation of the unknown forward model parameters.

In this paper, we extend the RED-diff framework to blind inverse problems. Using variational inference, we represent the sampling as alternating stochastic optimization that estimates both the image and forward model parameters. We evaluate the performance of blind RED-diff on MR image reconstruction with unknown magnetic fieldmap inhomogeneity.

NeurIPS 2023 Workshop on Deep Learning and Inverse Problems, New Orleans, USA.

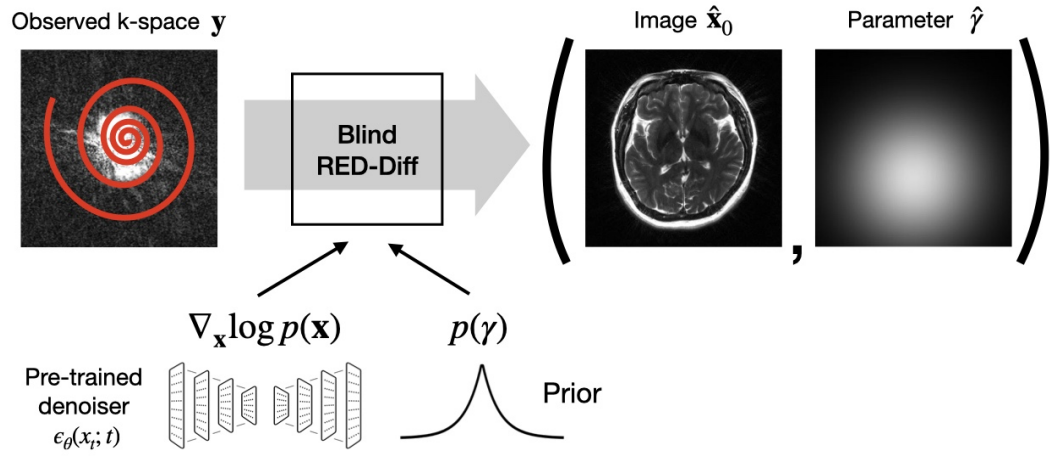

Figure 1: Blind RED-diff diagram. Our proposed algorithm extends RED-diff framework to blind inverse problems. Blind RED-diff combines data-consistency loss with score-matching regularization from denoisers at different time-steps and forward model parameter prior.

2 Methods

We consider the blind inverse problem:

$$\mathbf{y} = f_{\boldsymbol{\gamma}}(\mathbf{x}_0) + \boldsymbol{\eta} \tag{1}$$

where the forward model is parameterized by the unknown parameter $\boldsymbol{\gamma} \in \mathbb{R}^D$ that needs to be estimated, $\mathbf{x}_0$ is the ground truth image, $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \sigma_{\boldsymbol{\eta}}^2 \mathbf{I})$ is the measurement noise. Under this setting, the conditional distribution can be characterized as $p(\mathbf{y}|\mathbf{x}_0, \boldsymbol{\gamma}) \sim \mathcal{N}(f_{\boldsymbol{\gamma}}(\mathbf{x}_0), \sigma_{\boldsymbol{\eta}}^2 \mathbf{I})$. We try to minimize the KL-divergence using a variational approach

$$\min_q KL\left(q(\mathbf{x}_0, \boldsymbol{\gamma}|\mathbf{y}) || p(\mathbf{x}_0, \boldsymbol{\gamma}|\mathbf{y})\right) \tag{2}$$

where q is the joint variational distribution. When the image and forward model parameters are independent, the KL-divergence in Eq. 2 can be expressed as

$$KL\left(q(\mathbf{x}_0, \boldsymbol{\gamma}|\mathbf{y}) || p(\mathbf{x}_0, \boldsymbol{\gamma}|\mathbf{y})\right) \tag{3}$$

$$= \mathbb{E}_{q(\mathbf{x}_0|\mathbf{y})}\left[\log \frac{q(\mathbf{x}_0|\mathbf{y})}{p(\mathbf{x}_0)}\right] + \mathbb{E}_{q(\boldsymbol{\gamma}|\mathbf{y})}\left[\log \frac{q(\boldsymbol{\gamma}|\mathbf{y})}{p(\boldsymbol{\gamma})}\right] - \mathbb{E}_{q(\mathbf{x}_0, \boldsymbol{\gamma}|\mathbf{y})}\left[\log p(\mathbf{y}|\mathbf{x}_0, \boldsymbol{\gamma})\right] + \log p(\mathbf{y}) \tag{4}$$

$$= \underbrace{KL\left(q(\mathbf{x}_0|\mathbf{y}) || p(\mathbf{x}_0)\right)}_{\text{term (i)}} + \underbrace{KL\left(q(\boldsymbol{\gamma}|\mathbf{y}) || p(\boldsymbol{\gamma})\right)}_{\text{term (ii)}} \underbrace{- \mathbb{E}_{q(\mathbf{x}_0, \boldsymbol{\gamma}|\mathbf{y})}\left[\log p(\mathbf{y}|\mathbf{x}_0, \boldsymbol{\gamma})\right]}_{\text{term (iii)}} + \log p(\mathbf{y}) \tag{5}$$

where in line 4 we used the modeling assumption $p(\mathbf{x}_0, \boldsymbol{\gamma}) = p(\mathbf{x}_0)p(\boldsymbol{\gamma})$ and assumed that $q(\mathbf{x}_0, \boldsymbol{\gamma}|\mathbf{y}) = q(\mathbf{x}_0|\mathbf{y})q(\boldsymbol{\gamma}|\mathbf{y})$ based on the mean-field theory. Note that $q(\mathbf{x}_0|\mathbf{y}) := \mathcal{N}(\boldsymbol{\mu}_{\mathbf{x}}, \sigma_{\mathbf{x}}^2 \mathbf{I})$ and $q(\boldsymbol{\gamma}|\mathbf{y}) := \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}}^2 \mathbf{I})$ are the variational distributions. Consequently, the minimization in Eq. 2 consists of two KL-divergence terms, one for the image and another for the unknown forward model parameter, and a data-consistency term.

The KL-divergence in term (i) has the same form in (7). Therefore we can follow the same derivations as in Propositions 1 and 2 in (7) to represent it as a score matching loss which can be further simplified as:

$$KL\left(q(\mathbf{x}_0|\mathbf{y}) || p(\mathbf{x}_0)\right) \approx \mathbb{E}_{t,\boldsymbol{\epsilon}}\left[\omega(t)\frac{1}{\sigma_t^2}||\epsilon_{\boldsymbol{\theta}}(\mathbf{x}_t; t) - \boldsymbol{\epsilon}||_2^2\right] \tag{6}$$

$$= \mathbb{E}_{t,\boldsymbol{\epsilon}}\left[\lambda_t \left(\mathtt{sg}\left(\epsilon_{\boldsymbol{\theta}}(\mathbf{x}_t; t) - \boldsymbol{\epsilon}\right)\right)\right]^T \boldsymbol{\mu}_{\mathbf{x}} \tag{7}$$

where $\mathbf{x}_t = \alpha_t \boldsymbol{\mu}_{\mathbf{x}} + \sigma_t \boldsymbol{\epsilon}$ with α_t, σ_t, $\boldsymbol{\epsilon}$ defined for the variance preserving stochastic differential equation (VP-SDE) as in (9), $w(t)$ is a time-dependent weighting mechanism, and $\lambda_t := \frac{2T\sigma_{\boldsymbol{\eta}}^2 \alpha_t}{\sigma_t} \frac{d\omega(t)}{dt}$.

$\epsilon_{\boldsymbol{\theta}}(\mathbf{x}_t; t)$ is the diffusion model that estimates the score function $\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t)$ using denoising score matching (10). Similarly to (7), the notation (sg) indicates that the score is not differentiated with respect to $\boldsymbol{\mu}_{\mathbf{x}}$ during optimization.

The KL-divergence in term (ii) acts as a regularizer on $\boldsymbol{\gamma}$. When $p(\boldsymbol{\gamma})$ has a specific distribution, e.g. Gaussian or Laplace, we can obtain a closed-form expression for this term. For the i.i.d. Laplace prior assumption $p_i(\boldsymbol{\gamma}_i) \sim \mathcal{L}(\tilde{\mu}_{\boldsymbol{\gamma}_i}, \tilde{\sigma}_{\boldsymbol{\gamma}})$, $q(\boldsymbol{\gamma}_i|\mathbf{y}) := \mathcal{L}(\mu_{\boldsymbol{\gamma}_i}, \sigma_{\boldsymbol{\gamma}})$, and following (11) the regularizer can be expressed as

$$\mathcal{R}(\boldsymbol{\mu}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}}) = KL\left(q(\boldsymbol{\gamma}|\mathbf{y})||p(\boldsymbol{\gamma})\right) = \frac{\sigma_{\boldsymbol{\gamma}}}{\tilde{\sigma}_{\boldsymbol{\gamma}}} \log \frac{\tilde{\sigma}_{\boldsymbol{\gamma}}}{\sigma_{\boldsymbol{\gamma}}} - D + \frac{||\boldsymbol{\mu}_{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\mu}}_{\boldsymbol{\gamma}}||_1}{\tilde{\sigma}_{\boldsymbol{\gamma}}} + \frac{\sigma_{\boldsymbol{\gamma}}}{\tilde{\sigma}_{\boldsymbol{\gamma}}} \exp\left(-\frac{||\boldsymbol{\mu}_{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\mu}}_{\boldsymbol{\gamma}}||_1}{\sigma_{\boldsymbol{\gamma}}}\right). \tag{8}$$

The final term is simply the data consistency term, i.e., the reconstruction loss.

$$\mathbb{E}_{q(\mathbf{x}_0, \boldsymbol{\gamma}|\mathbf{y})}\left[\log p(\mathbf{y}|\mathbf{x}_0, \boldsymbol{\gamma})\right] = -\frac{1}{2\sigma_\eta^2} \mathbb{E}_{q(\mathbf{x}_0, \boldsymbol{\gamma}|\mathbf{y})}\left[||\mathbf{y} - f_{\boldsymbol{\gamma}}(\mathbf{x}_0)||_2^2\right] \tag{9}$$

Following (7), for simplicity we assume $\boldsymbol{\sigma}_{\boldsymbol{\gamma}} = 0$, then the optimization problem becomes

$$\min_{\boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\mu}_{\boldsymbol{\gamma}}} \frac{1}{2\sigma_\eta^2} ||\mathbf{y} - f_{\boldsymbol{\gamma}}(\boldsymbol{\mu}_{\mathbf{x}})||^2 + \mathbb{E}_{t, \boldsymbol{\epsilon}}\left[\omega(t) \frac{1}{\sigma_t^2} ||\epsilon_{\boldsymbol{\theta}}(\mathbf{x}_t; t) - \boldsymbol{\epsilon}||_2^2\right] + \mathcal{R}(\boldsymbol{\mu}_{\boldsymbol{\gamma}}, 0) \tag{10}$$

In essence, the solution to the blind inverse problem will find an image $\boldsymbol{\mu}_{\mathbf{x}}$ and the forward model parameter $\boldsymbol{\mu}_{\boldsymbol{\gamma}}$, while enforcing a high likelihood under the diffusion model prior and regularization imposed by the prior for the forward model parameter. Similar to (7), we search for $\hat{\mathbf{x}} = \boldsymbol{\mu}_{\mathbf{x}}$ and $\hat{\boldsymbol{\gamma}} = \boldsymbol{\mu}_{\boldsymbol{\gamma}}$ using first-order stochastic optimization in alternating fashion. Our method is described in Algorithm 1.

Algorithm 1: RED-diff for Blind Inverse Problems

Input: $\mathbf{y}$, $f_{\boldsymbol{\gamma}}(\cdot)$, $\{\tilde{\boldsymbol{\mu}}_{\boldsymbol{\gamma}}, \tilde{\sigma}_{\boldsymbol{\gamma}}, \lambda_{\boldsymbol{\gamma}}\}$, $\{\alpha_t, \sigma_t, \lambda_t\}$
Initialize : $\boldsymbol{\mu}_{\mathbf{x}_0}, \boldsymbol{\mu}_{\boldsymbol{\gamma}}$
for $t = T, \ldots, 1$ **do**
  $\boldsymbol{\epsilon}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
  $\mathbf{x}_t = \alpha_t \boldsymbol{\mu}_{\mathbf{x}} + \sigma_t \boldsymbol{\epsilon}_1$
  $loss_x = ||\mathbf{y} - f_{\boldsymbol{\mu}_{\boldsymbol{\gamma}}}(\boldsymbol{\mu}_{\mathbf{x}})||_2^2 + \lambda_t (\text{sg}\left[\epsilon_{\boldsymbol{\theta}}(\mathbf{x}_t; t) - \boldsymbol{\epsilon}\right])^T \boldsymbol{\mu}_{\mathbf{x}}$
  $\boldsymbol{\mu}_{\mathbf{x}} \leftarrow$ OptimizerStep$(loss_x)$
  $loss_{\boldsymbol{\gamma}} = ||\mathbf{y} - f_{\boldsymbol{\mu}_{\boldsymbol{\gamma}}}(\boldsymbol{\mu}_{\mathbf{x}})||_2^2 + \lambda_{\boldsymbol{\gamma}} \mathcal{R}(\boldsymbol{\mu}_{\boldsymbol{\gamma}}, 0)$
  $\boldsymbol{\mu}_{\boldsymbol{\gamma}} \leftarrow$ OptimizerStep$(loss_{\boldsymbol{\gamma}})$
end
Return: $\boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\mu}_{\boldsymbol{\gamma}}$

3 Experiments

We test the blind RED-diff algorithm in the MRI inverse problem of fieldmap-corrected image reconstruction, where we jointly reconstruct the image and estimate the off-resonance fieldmap. MRI data was retrospectively simulated using ground truth images from the fastMRI dataset (12). Ground truth sensitivity maps were calculated using ESPIRiT (13) on the fastMRI data. Ground truth fieldmaps were obtained from a separate brain dataset acquired with the PhysiCal sequence (14). To simulate the fieldmap corrupted multi-channel k-space data, a 16 shot variable density spiral was used with an undersampling factor of $R = 2$. The temporal sampling rate was set to $4\mu s$, and the total readout time was 15.6 ms, similar to what is used on GE scanners.

The forward model relating the ground truth brain images $\mathbf{x}_0$ to multi-channel spiral k-space data $\mathbf{y}$ is shown below:

$$\mathbf{y} = f_{\boldsymbol{\gamma}}(\mathbf{x}) + \eta = \mathbf{A}_{\boldsymbol{\gamma}} \mathbf{x}_{\mathbf{0}} + \eta = \mathbf{F} \mathbf{S} \mathbf{T}_{\boldsymbol{\gamma}} \mathbf{x}_{\mathbf{0}} + \boldsymbol{\eta}, \tag{11}$$

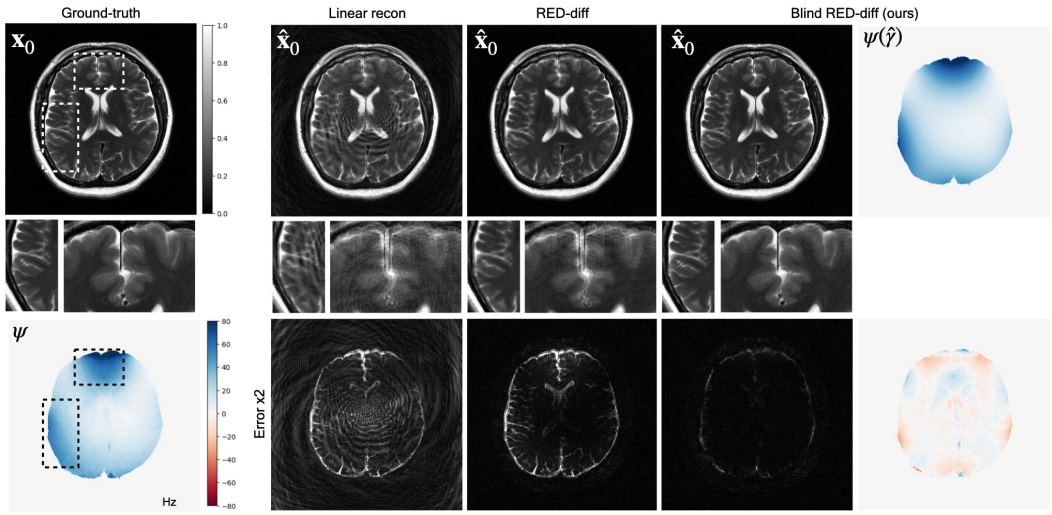

Figure 2: Results. Blind RED-diff is able to simultaneously resolve the off-resonance blurring and remove undersampling artifacts by concurrently estimating image $\mathbf{x}_0$ and the forward model parameter $\boldsymbol{\gamma}$. Conversely, RED-diff does not remove blurring artifacts. Blurring is mainly located in the regions of the ground truth field map $\boldsymbol{\psi}$ where off-resonance is stronger.

where the forward model $\mathbf{A}_{\boldsymbol{\gamma}}$ is composed by three operators: the transform operator $\mathbf{T}_{\boldsymbol{\gamma}}$, the sensitivity map operator $\mathbf{C}$ and the Non-Uniform Fourier Transform (15) operator $\mathbf{F}$. The transform operator $\mathbf{T}_{\boldsymbol{\gamma}}$ implements time-segmented off-resonance effects (16) caused by the field inhomogeneities map $\boldsymbol{\psi}(\boldsymbol{\gamma})$. We parameterize the field map with $\boldsymbol{\gamma}$. A more detailed model description can be found in Appendix A.2.

Algorithm 1 was implemented by modifying the csgm-mri-langevin (https://github.com/utcsilab/csgm-mri-langevin) and SMRD (https://github.com/batuozt/SMRD) libraries from Jalal et al. (2) and Ozturkler et al. (17) respectively. Reconstructions were run on a 24 GB NVIDIA Titan RTX. For the image $\mathbf{x}$ score function, we used the score function model from Jalal et al. (2). For the field map $\boldsymbol{\psi}$, we assumed a spatial polynomial model, where $\boldsymbol{\gamma}$ contains the polynomial coefficients. Empirically, we observed that a Laplace distribution approximates $p(\gamma)$. Further details can be found in Appendix A.2.

Results are shown in Figure 2. The linear reconstruction shows both off-resonance and undersampling artifacts. RED-diff is able to remove undersampling artifacts because of the diffusion prior. However, RED-diff cannot resolve the off-resonance blurring due to the limitations of the model. Conversely, our proposed blind RED-diff is able to remove both undersampling and off-resonance artifacts.

4 Discussion and Conclusion

In this work, we extended the RED-diff framework to blind inverse problems. Blind RED-diff requires a pretarined diffusion model and the functional description of the forward model. We evaluated the performance of our proposed method on fieldmap corrected MRI reconstruction where the B0-inhomogeneity map is not known at inference time. Our preliminary results demonstrated that the blind RED-diff framework can successfully approximate the unknown forward model parameters and produce fieldmap corrected reconstructions.

In our experiments, we used a simple Laplacian prior for the forward model parameters $\boldsymbol{\gamma}$ which leads to an ℓ_1 regularized solution. A much stronger diffusion based prior can also be used for representing the prior distribution. We are planning to extend our framework to include expressive diffusion priors as in Chung et. al. (18) for forward model parameters to improve the results.

One limitation of our formulation is the conditional independence assumption in the variational distributions for $\mathbf{x}$ and $\boldsymbol{\gamma}$ given the measurement $\mathbf{y}$. Although this assumption can be sufficient for cases like motion corrupted MRI and fieldmap corrected MRI reconstruction, it can be inaccurate

for scenarios where there is a strong dependence between $\mathbf{x}$ and $\boldsymbol{\gamma}$ such as the water-fat separation problem. These dependencies could be modeled with conditional diffusion models.

An important benefit of diffusion based samplers such as RED-diff is that they can be used to solve different inverse problems without retraining or fine-tuning as the score model is agnostic to the MRI forward model. We are planning to extend our framework to other inverse problems in MRI such as water-fat separation and quantitative parameter mapping.

Acknowledgments and Disclosure of Funding

This work was supported by NIH U01 EB029427.

References

[1] A. Jalal, S. Karmalkar, A. G. Dimakis, and E. Price, "Instance-optimal compressed sensing via posterior sampling," *International Conference on Machine Learning*, 2021.

[2] A. Jalal, M. Arvinte, G. Daras, E. Price, A. G. Dimakis, and J. I. Tamir, "Robust compressed sensing mri with deep generative priors," *Advances in Neural Information Processing Systems*, 2021.

[3] H. Chung and J. C. Ye, "Score-based diffusion models for accelerated mri," *Medical image analysis*, vol. 80, p. 102479, 2022.

[4] Y. Song, L. Shen, L. Xing, and S. Ermon, "Solving inverse problems in medical imaging with score-based generative models," in *International Conference on Learning Representations*, 2021.

[5] H. Chung, J. Kim, M. T. Mccann, M. L. Klasky, and J. C. Ye, "Diffusion posterior sampling for general noisy inverse problems," in *International Conference on Learning Representations*, 2022.

[6] J. Song, A. Vahdat, M. Mardani, and J. Kautz, "Pseudoinverse-guided diffusion models for inverse problems," in *International Conference on Learning Representations*, 2022.

[7] M. Mardani, J. Song, J. Kautz, and A. Vahdat, "A variational perspective on solving inverse problems with diffusion models," *arXiv preprint arXiv:2305.04391*, 2023.

[8] B. Ozturkler, M. Mardani, A. Vahdat, J. Kautz, and J. M. Pauly, "Regularization by denoising diffusion process for mri reconstruction," in *Medical Imaging with Deep Learning, short paper track*, 2023.

[9] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-based generative modeling through stochastic differential equations," in *International Conference on Learning Representations*, 2020.

[10] P. Vincent, "A connection between score matching and denoising autoencoders," *Neural computation*, vol. 23, no. 7, pp. 1661–1674, 2011.

[11] G. P. Meyer, "An alternative probabilistic interpretation of the huber loss," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5261–5269, June 2021.

[12] F. Knoll, J. Zbontar, A. Sriram, M. J. Muckley, M. Bruno, A. Defazio, M. Parente, K. J. Geras, J. Katsnelson, H. Chandarana, *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*, vol. 2, no. 1, p. e190007, 2020.

[13] M. Uecker, P. Lai, M. J. Murphy, P. Virtue, M. Elad, J. M. Pauly, S. S. Vasanawala, and M. Lustig, "Espirit—an eigenvalue approach to autocalibrating parallel mri: where sense meets grappa," *Magnetic resonance in medicine*, vol. 71, no. 3, pp. 990–1001, 2014.

[14] S. S. Iyer, C. Liao, Q. Li, M. K. Manhard, A. Berman, B. Bilgic, and K. Setsompop, "Physical: A rapid calibration scan for b0, b1+, coil sensitivity and eddy current mapping," in *Proceedings of the 28th Annual Meeting of ISMRM, virtual*, p. 0661, 2020.

[15] J. A. Fessler and B. P. Sutton, "Nonuniform fast fourier transforms using min-max interpolation," *IEEE transactions on signal processing*, vol. 51, no. 2, pp. 560–574, 2003.

[16] B. P. Sutton, D. C. Noll, and J. A. Fessler, "Fast, iterative image reconstruction for mri in the presence of field inhomogeneities," *IEEE transactions on medical imaging*, vol. 22, no. 2, pp. 178–188, 2003.

[17] B. Ozturkler, C. Liu, B. Eckart, M. Mardani, J. Song, and J. Kautz, "Smrd: Sure-based robust mri reconstruction with diffusion models," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 199–209, Springer, 2023.

[18] H. Chung, J. Kim, S. Kim, and J. C. Ye, "Parallel diffusion models of operator and image for blind inverse problems," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6059–6069, 2023.

A Appendix

A.1 MRI forward model

The multi-channel k-space $\mathbf{y}$ and the forward model $\mathbf{A}_{\boldsymbol{\gamma}}$ can be expressed as C stacked components, where C is the number of coil channels.

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_{C-1} \end{bmatrix}, \quad \mathbf{A}_{\boldsymbol{\gamma}} = \begin{bmatrix} \mathbf{A}_0 \\ \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_{C-1}, \end{bmatrix} \tag{12}$$

where $\mathbf{y}_c$ is the k-space data from coil channel c and, similarly, $\mathbf{A}_c$ is the forward model of coil channel c. We obviated the dependency of $\mathbf{A}_c$ to $\boldsymbol{\gamma}$ for notation simplicity. Then, the time-segmented forward model that relates k-space $\mathbf{y}_c$ to the underlying brain image $\mathbf{x}$ is the following:

$$\begin{aligned} \mathbf{y}_c &= \mathbf{A}_c\mathbf{x}_0 + \boldsymbol{\eta} \\ &= \hat{\mathbf{F}}\mathbf{S}_c\hat{\mathbf{T}}_{\boldsymbol{\gamma}}\mathbf{x}_0 + \boldsymbol{\eta} \\ &= \begin{bmatrix} \mathbf{F}_0 & & & \\ & \mathbf{F}_1 & & \\ & & \ddots & \\ & & & \mathbf{F}_{T-1} \end{bmatrix} \begin{bmatrix} \mathbf{S}_c & & & \\ & \mathbf{S}_c & & \\ & & \ddots & \\ & & & \mathbf{S}_c \end{bmatrix} \begin{bmatrix} \mathbf{T}_0(\gamma) & & & \\ & \mathbf{T}_1(\gamma) & & \\ & & \ddots & \\ & & & \mathbf{T}_{T-1}(\gamma) \end{bmatrix} \mathbf{x}_0 + \boldsymbol{\eta} \end{aligned} \tag{13}$$

$$\mathbf{S}_c = \text{diag}(\mathbf{s}_c), \tag{14}$$

$$\mathbf{T}_t(\gamma) = \text{diag}\left(\exp\left\{-i2\pi\boldsymbol{\tau}_t \text{vec}(\boldsymbol{\psi}(\boldsymbol{\gamma}))\right\}\right) \tag{15}$$

,

where $\mathbf{F}_t$ estimates the k-space points acquired during time segment t, T is the total number of time segments, $\boldsymbol{\tau}$ is the vector with the mean time of each time segment, $\mathbf{s}_c$ is the sensitivity map of coil channel c, and vec$(\cdot)$ is the vectorized function.

A.2 Field map model

We parameterize $\boldsymbol{\psi}$ as a masked spatial polynomial function, where $\boldsymbol{\gamma}$ contains the polynomial coefficients.

$$\boldsymbol{\psi}_{ij}(\gamma) = \sum_{p+q \le P} \boldsymbol{\gamma}_{pq} i^p j^q \times \mathbf{m}_{ij}, \tag{16}$$

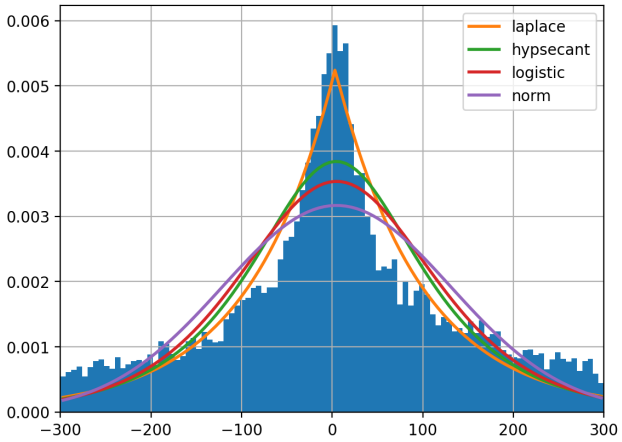

Figure 3: Distribution of polynomial coefficients fit to B0 field maps on the in-house brain dataset. We observe that a Laplace distribution approximates the prior for polynomial coefficients $p(\boldsymbol{\gamma})$ effectively.

where $\mathbf{m}$ represents a mask that selects the image support for numerical stability. In order to represent the prior $p(\boldsymbol{\gamma})$ on the polynomial coefficients, we fit $P = 5$th order polynomials on the B0 field maps. The B0 field map dataset was obtained from a separate brain acquisition using the PhysiCal sequence (14). The dataset contains a total of 2,420 2D slices from 11 subjects. Fig. 3 shows the distribution of the coefficients and candidate distributions fit on this data. Distribution fitting results indicate that a Laplace distribution with location parameter $\mu = 3.3$ and scale parameter $\sigma = 95.1$ can approximate this prior. Therefore, we assumed a Laplace distribution for the $\boldsymbol{\gamma}$ prior with these parameters which resulted in the ℓ_1 regularized update as shown in Eq. 8.