Bayesian formulation of regularization by denoising - Model and Monte Carlo sampling

Elhadji C. Faye, Mame Diarra Fall, Aladine Chetouani, Nicolas Dobigeon

Bayesian formulation of regularization by denoising – Model and Monte Carlo sampling

Elhadji C. Faye, Mame Diarra Fall
Institut Denis Poisson
Université d'Orléans
Orléans, France
{elhadji-cisse.faye, diarra.fall}@univ-orleans.fr

Aladine Chetouani
Laboratoire PRISME
Université d'Orléans
Orléans, France
aladine.chetouani@univ-orleans.fr

Nicolas Dobigeon
IRIT/INP-ENSEEIHT
University of Toulouse
Toulouse, France
nicolas.dobigeon@enseeiht.fr

***Abstract*—Image restoration aims at recovering a clean image from degraded observations. This paper presents a novel Bayesian framework for image restoration using a regularization-by-denoising (RED) prior. It introduces a probabilistic counterpart to the RED paradigm, and proposes a new Monte Carlo algorithm to efficiently sample from the resulting posterior distribution. The proposed method benefit from the recent developments of deep learning-based denoisers. Extensive numerical experiments illustrate the efficiency of the proposed method, showcasing its competitive performance against state-of-the-art methods.**

***Index Terms*—Inverse problems, Bayesian inference, deep learning, Markov chain Monte Carlo algorithms.**

I. Introduction

Image restoration (IR) is an inverse problem that aims at recovering an unknown image $\mathbf{x} \in \mathbb{R}^n$ from a degraded version $\mathbf{y} \in \mathbb{R}^m$. The degraded image $\mathbf{y}$ is generally observed according to the linear model $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$ where $\mathbf{A}$ is the degradation matrix and $\mathbf{n}$ is assumed to be additive white Gaussian noise (AWGN). By specifying this degradation matrix, various image restoration tasks can be considered. The relationship between $\mathbf{x}$ and $\mathbf{y}$ can be described by a statistical model prescribed by the likelihood function

$$p(\mathbf{y}|\mathbf{x}) \propto \exp\left[-f(\mathbf{x}, \mathbf{y})\right] \tag{1}$$

where $f(\mathbf{x}, \mathbf{y}) = \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2/(2\sigma^2)$ is a fidelity term, i.e., it accounts for the consistency of $\mathbf{x}$ with respect to the measured data $\mathbf{y}$. Since IR is generally an ill-posed or, at least, ill-conditioned problem, the Bayesian paradigm assigns a prior distribution to $\mathbf{x}$ that encapsulates the prior knowledge about $\mathbf{x}$. This distribution can be expressed as follows

$$p(\mathbf{x}) \propto \exp\left[-\beta g(\mathbf{x})\right] \tag{2}$$

where $g : \mathbb{R}^n \to \mathbb{R}$ stands for the regularization term with regularization parameter $\beta > 0$. From the likelihood $p(\mathbf{y}|\mathbf{x})$ and the prior $p(\mathbf{x})$, the posterior distribution $p(\mathbf{x}|\mathbf{y})$ is derived from the Bayes' rule

$$p(\mathbf{x}|\mathbf{y}) \propto \exp\left[-f(\mathbf{x}, \mathbf{y}) - \beta g(\mathbf{x})\right]. \tag{3}$$

Part of this work was supported by the Artificial Natural Intelligence Toulouse Institute (ANITI, ANR-19-PI3A-0004), the AI.iO Project (ANR-20-THIA-0017) and the BACKUP project (ANR-23-CE40-0018- 01).

Bayesian inference exploits this posterior distribution to derive various estimators. In particular, the maximum a posteriori (MAP) solution is obtained by solving a minimization problem

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}}\, f(\mathbf{x}, \mathbf{y}) + \beta g(\mathbf{x}). \tag{4}$$

Traditional methods often use explicit regularizations $g(\cdot)$ such as total variation (TV) promoting piecewise constant behavior [1], Sobolev favoring smooth content [2], or ℓ_p-norm with $p \leq 1$ promoting sparsity [3], [4]. However their design is subjective and may not fully capture complex image structures. To overcome these limitations, Venkatakrishnan *et al.* introduced the concept of *plug-and-play* (PnP) as an implicit prior [5]. Most PnP approaches leverage a variable splitting strategy such as half-quadratic splitting (HQS) [6] or the alternating direction method of multipliers (ADMM) [4], to explicitly exhibit a proximal mapping. This mapping is subsequently interpreted as a denoising step, which allows the underlying optimization subproblem to be replaced by an off-the-shelf denoiser. PnP has demonstrated its effectiveness in various imaging applications [7]–[10] by capitalizing on the inherent benefits of deep neural network-based denoisers, such as DnCNN [11] or DRUNet [7]. In the same spirit as PnP, the regularization-by-denoising (RED) framework defines an explicit image-adaptive Laplacian-based prior that relies solely on the ability to perform a denoising task [12].

However, all the optimization algorithms mentioned above treat $\mathbf{x}$ in a deterministic way and generally produce only point estimates. An alternative to these approaches is to solve inverse problems in a fully Bayesian framework. This latter treats $\mathbf{x}$ as a random variable, offering a richer description thanks to the posterior distribution $p(\mathbf{x}|\mathbf{y})$ which allows uncertainty quantification to be performed [13], [14]. Markov chain Monte Carlo (MCMC) methods are commonly used to explore this distribution. However most of the works devoted to the development of MCMC algorithms for inverse problems in imaging rely on conventional model-based prior distributions. Like their deterministic counterpart, they encode the expected image characteristics on the basis of empirical arguments. Recent works have explored data-driven regularization methods, such as deep generative models like variational autoencoders or normalizing flows [15], [16]. While a probabilistic counterpart

of PnP has been devised for Monte Carlo sampling [17], a corresponding formulation for the RED paradigm is still lacking. Although RED has demonstrated its superiority over PnP in a variational context, its derivation within a Bayesian framework to embed Monte Carlo algorithms remains unexplored.

This paper presents a framework for Bayesian inference using a RED prior. Section II proposes a probabilistic interpretation of RED, defining a new distribution that can serve as a prior in Bayesian inversion tasks. This section also introduces a new Monte Carlo algorithm specifically designed to efficiently sample from the resulting posterior distribution. Extensive numerical experiments described in Section III show that the proposed approach competes favorably with state-of-the-art variational and Monte Carlo methods on ubiquitous inversion tasks. Section IV concludes the paper.

II. Proposed Method

Motivated by PnP, Romano *et al.* introduced regularization by denoising (RED) [12], a PnP-like framework to exploit denoisers as an implicit regularization. Given a denoiser $\mathsf{D}_\nu : \mathbb{R}^n \to \mathbb{R}^n$ with ν controlling the denoising strength, the regularization term associated with RED is expressed as

$$g_{\text{red}}(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top \left(\mathbf{x} - \mathsf{D}_\nu(\mathbf{x})\right). \tag{5}$$

Although it offers great flexibility in the choice of denoisers that can be used, RED requires to satisfy the so-called RED conditions: $\mathsf{D}_\nu(\cdot)$ should be differentiable, locally homogeneous, with a symmetric Jacobian and ensuring the passivity condition [12], [18]. Under these conditions, the gradient of $g_{\text{red}}(\cdot)$ is expressed as the denoising residual

$$\nabla g_{\text{red}}(\mathbf{x}) = \mathbf{x} - \mathsf{D}_\nu(\mathbf{x}) \tag{6}$$

which bypasses the need to differentiate the denoising operator itself. As PnP, RED appears to be particularly appealing since it can take benefit from recently designed efficient denoisers, boosted by the advances achieved by deep learning.

A. RED prior and posterior distributions

The formulation of RED-based inversion into a fully statistical framework requires the definition of a new prior distribution. It is derived from the RED potential $g_{\text{red}}(\cdot)$ as

$$p_{\text{red}}(\mathbf{x}) \propto \exp\left[-\frac{\beta}{2}\mathbf{x}^\top \left(\mathbf{x} - \mathsf{D}_\nu(\mathbf{x})\right)\right]. \tag{7}$$

It is worth noting that the functional $p_{\text{red}}(\cdot)$ does not necessarily define a probability density function (pdf). For $p_{\text{red}}(\cdot)$ to define a proper pdf, some conditions must be met. This point is discussed in what follows.

Assumption 1. *The matrix $\mathbf{\Lambda}(\mathbf{x}) = \mathbf{I}_n - \nabla\mathsf{D}_\nu(\mathbf{x})$, $\forall \mathbf{x} \in \mathbb{R}^n$, has at least one non-zero eigenvalue.*

This technical assumption is generally not restrictive, typically applicable except in trivial cases where all eigenvalues of $\mathbf{\Lambda}(\bar{\mathbf{x}})$ are zero, meaning $\bar{\mathbf{x}}$ is already a noise-free image and requires no further denoising. The following result states that this assumption combined with the RED conditions are sufficient to guarantee that the function (7) defines a proper distribution.

Proposition 1. *If Assumption 1 and RED conditions hold, then $p_{red}(\cdot)$ in (7) defines a proper pdf.*

Proof. Under the RED conditions, the RED prior (7) can be rewritten according to the pseudo-quadratic form

$$p_{\text{red}}(\mathbf{x}) \propto \exp\left[-\tfrac{\beta}{2}\mathbf{x}^\top \mathbf{\Lambda}(\mathbf{x})\mathbf{x}\right]$$

with $\mathbf{\Lambda}(\mathbf{x}) = \mathbf{I}_n - \nabla\mathsf{D}_\nu(\mathbf{x})$. From Assumption 1, there exists $\lambda_{\min} > 0$ such that $\lambda_{\min}\mathbf{I}_n \preceq \mathbf{\Lambda}(\mathbf{x})$, $\forall \mathbf{x} \in \mathbb{R}^n$. This implies that $\lambda_{\min}\mathbf{x}^\top\mathbf{x} \leq \mathbf{x}^\top\mathbf{\Lambda}(\mathbf{x})\mathbf{x}$ and

$$\int_{\mathbb{R}^n} p_{\text{red}}(\mathbf{x})\mathrm{d}\mathbf{x} \leq \int_{\mathbb{R}^n} \exp\left[-\tfrac{\beta}{2}\lambda_{\min}\|\mathbf{x}\|^2\right]\mathrm{d}\mathbf{x} < \infty.$$

□

Combining the RED prior $p_{\text{red}}(\mathbf{x})$ defined by (7) and the likelihood function $p(\mathbf{y}|\mathbf{x})$ defined by (1), the RED posterior distribution of interest is written as

$$\pi(\mathbf{x}) \triangleq p(\mathbf{x}|\mathbf{y}) \propto \exp\left[-f(\mathbf{x},\mathbf{y}) - \beta g_{\text{red}}(\mathbf{x})\right]. \tag{8}$$

Sampling from this posterior is not easy, due to the use of the denoiser. In the next section, we propose a dedicated Monte Carlo algorithm that is particularly well suited to this task.

B. Monte Carlo sampling

The proposed sampling strategy consists first in subjecting the target posterior (3) to an asymptotically exact data augmentation (AXDA) [19]. In a fashion similar to HQS and ADMM, this procedure introduces an auxiliary variable $\mathbf{z} \in \mathbb{R}^n$ and considers the augmented distribution

$$\begin{aligned}\pi_\rho(\mathbf{x},\mathbf{z}) &= p(\mathbf{x},\mathbf{z}|\mathbf{y};\rho^2) \\ &\propto \exp\left[-f(\mathbf{x},\mathbf{y}) - \beta g_{\text{red}}(\mathbf{z}) - \frac{1}{2\rho^2}||\mathbf{x}-\mathbf{z}||^2\right]\end{aligned} \tag{9}$$

where ρ is a positive parameter controlling the dissimilarity between $\mathbf{x}$ and $\mathbf{z}$. This data augmentation (9) is approximate in the sense that the marginal distribution $\pi_\rho(\mathbf{x}) = \int_{\mathbb{R}^n} \pi_\rho(\mathbf{x},\mathbf{z})d\mathbf{z}$ coincides with the target posterior distribution $\pi(\mathbf{x})$ only in an asymptotic regime $\rho \to 0$. Then the split Gibbs sampler (SGS) [20], [21] alternatively samples according to the two conditional distributions associated to the augmented posterior $\pi_\rho(\mathbf{x},\mathbf{z})$, and defined as follows

$$p(\mathbf{x}|\mathbf{y},\mathbf{z};\rho^2) \propto \exp\left[-f(\mathbf{x},\mathbf{y}) - \frac{1}{2\rho^2}||\mathbf{x}-\mathbf{z}||^2\right] \tag{10}$$

$$p(\mathbf{z}|\mathbf{x};\rho^2) \propto \exp\left[-\beta g_{\text{red}}(\mathbf{z}) - \frac{1}{2\rho^2}||\mathbf{x}-\mathbf{z}||^2\right]. \tag{11}$$

This splitting allows the two terms $f(\cdot,\mathbf{y})$ and $g_{\text{red}}(\cdot)$ defining the full potential to be dissociated and involved into two distinct conditional distributions, leading to a simpler and more efficient sampling algorithm. Given the quadratic form of the data-fitting

term $f(\mathbf{x},\mathbf{y}) = \frac{1}{2\sigma^2}\|\mathbf{A}\mathbf{x}-\mathbf{y}\|_2^2$, the conditional distribution (10) writes

$$p(\mathbf{x}|\mathbf{y},\mathbf{z};\rho^2) = \mathcal{N}(\mathbf{x};\boldsymbol{\mu}(\mathbf{z}),\mathbf{Q}^{-1}) \tag{12}$$

where the precision matrix $\mathbf{Q}$ and the mean vector $\boldsymbol{\mu}(\cdot)$ are

$$\left\{ \begin{array}{rcl} \mathbf{Q} & = & \frac{1}{\sigma^2}\mathbf{A}^\top\mathbf{A} + \frac{1}{\rho^2}\mathbf{I} \\ \boldsymbol{\mu}(\mathbf{z}) & = & \mathbf{Q}^{-1}\left(\frac{1}{\sigma^2}\mathbf{A}^\top\mathbf{y} + \frac{1}{\rho^2}\mathbf{z}\right). \end{array} \right. \tag{13}$$

When the potential function is not quadratic, the proposed framework can embed proximal Monte Carlo algorithms to sample from (10), as in [22], [23]. Besides, one can notice that equation (11) corresponds to the posterior distribution associated with a Bayesian denoising task aimed at estimating the quantity $\mathbf{z}$ from a noisy observation $\mathbf{x}$ contaminated by an additive white Gaussian noise. Sampling from this conditional is challenging, mainly due to the regularization potential $g_{\text{red}}(\cdot)$ which incorporates the denoiser $\mathsf{D}_\nu(\cdot)$. We propose to take advantage of the property (6) by sampling from (11) using a Langevin Monte Carlo (LMC) step, i.e.,

$$\begin{aligned} \mathbf{z}^{(t+1)} &= \mathbf{z}^{(t)} + \gamma\nabla\log p\left(\mathbf{z}^{(t)} \mid \mathbf{x};\rho^2\right) + \sqrt{2\gamma}\boldsymbol{\varepsilon}^{(t)} \\ &= \mathbf{z}^{(t)} - \gamma\beta\left(\mathbf{z}^{(t)} - \mathsf{D}_\nu(\mathbf{z}^{(t)})\right) \\ &\qquad + \frac{\gamma}{\rho^2}\left(\mathbf{x}-\mathbf{z}^{(t)}\right) + \sqrt{2\gamma}\boldsymbol{\varepsilon}^{(t)} \end{aligned} \tag{14}$$

where $\left\{\boldsymbol{\varepsilon}^{(t)}\right\}_{t\in\mathbb{N}}$ is a sequence of independent and identically distributed n-dimensional standard Gaussian random variables and $\gamma > 0$ is a fixed step-size that controls a trade-off between asymptotic accuracy and convergence speed. Due to the discretization, the samples produced by (14) are biased and not exactly distributed according to (11). This approximation error could be mitigated at the extra cost of combining (14) with a Metropolis-Hastings correction step, resulting in a Metropolis adjusted Langevin algorithm [24]. This paper rather adopts an LMC step without any MH adjustment. The proposed algorithm, instantiated to sample according to the RED posterior (8), is called Langevin-within-SGS using RED priors (RED-LwSGS). The convergence of the proposed algorithm is deeply investigated in the extended report [25].

III. Experiments

A. Experimental setup

Experimental protocol – The proposed method is evaluated using the Flickr Faces High Quality (FFHQ) [26] and Imagenet [27] data sets, comprising 100 RGB images with dimensions of 256×256 pixels ($n = 256^2$). Two restoration tasks are considered, namely image inpainting and single image super-resolution. The corresponding tasks are specified as follows. For image inpainting, the operator $\mathbf{A}$ stands for a binary mask with $m \ll n$. It is designed such that 80% of the total pixels are randomly masked across the three color channels. Conversely, for single image super-resolution, the operator $\mathbf{A}$ is decomposed as $\mathbf{A} = \mathbf{S}\mathbf{B}$ where the $n \times n$ matrix $\mathbf{B}$ stands for a spatially invariant Gaussian blur of size 7×7 with standard deviation 1.6 and the operator $\mathbf{S}$ is a $m \times n$ binary matrix which performs a regular subsampling of factor $d = 4$ in each direction (i.e., $m = nd^2$). For all tasks, the degraded images have been corrupted by an additive Gaussian noise to reach a $\text{SNR} = 30\text{dB}$. All images are normalized to the range of $[0, 1]$.

TABLE I
Values of the parameters used during the experiments.

	N_{MC}	N_{bi}	β	γ	ρ^2	ρ_1^2	ρ_2^2
Inpainting	10000	4500	0.125	$\frac{0.99}{2\beta+1/\rho^2}$	1.5	–	–
Super-res.	12500	3500	1.0	$\frac{0.8}{2\beta+1/\rho_2^2}$	–	0.2	1

Compared methods – The proposed method has been compared to several state-of-the-art optimization-based methods, including ADMM with RED (RED-ADMM) [12], HQS algorithm with RED (RED-HQS) and ADMM with a PnP regularization (PnP-ADMM) [5]. It has also been compared to state-of-the-art Monte Carlo methods, including the unadjusted Langevin algorithm (ULA) with a PnP regularization (PnP-ULA) [17], SGS with a TV regularization (TV-SP) [20] and Moreau-Yosida ULA with TV regularization (TV-MYULA) [28]. We have also considered a denoising diffusion model for PnP image restoration (DiffPIR) [29]. To conduct fair comparisons, the same pre-trained deep denoiser DRUNet [7] has been employed for all PnP- and RED-based algorithms, namely RED-ADMM, RED-HQS, PnP-ADMM, PnP-ULA and RED-LwSGS. DRUNet has been taken directly from an open repository[1] and applied without further fine-tuning for the inversion tasks. The test images have never been seen by the model during training, to avoid any bias due to potential over-fitting. It is worth noting that the optimization-based methods provide only point estimates of the restored images, following the MAP estimation (4). Conversely, the results obtained by the Monte Carlo methods and reported below correspond to the minimum mean square error (MMSE) estimators approximated by averaging the generated samples, i.e.,

$$\hat{\mathbf{x}}_{\text{MMSE}} = \frac{1}{N_{\text{MC}} - N_{\text{bi}}}\sum_{t=N_{\text{bi}}+1}^{N_{\text{MC}}} \mathbf{x}^{(t)} \tag{15}$$

where N_{bi} is the number of burn-in iterations and N_{MC} is the total number of iterations.

When tackling the super-resolution task, the proposed RED-LwSGS algorithm should be slightly adapted to follow a double splitting [25]. The regularization parameter β and the coupling parameters have been adjusted to reach the best performance. Table I details the parameter values for all tasks.

Figures-of-merit – To quantitatively compare the algorithm performance, two standard distortion metrics are considered, namely Peak Signal-to-Noise Ratio (PSNR) (dB) and Structural Similarity Index (SSIM) [30], with higher scores indicating

[1] Available online at https://github.com/cszn/DPIR

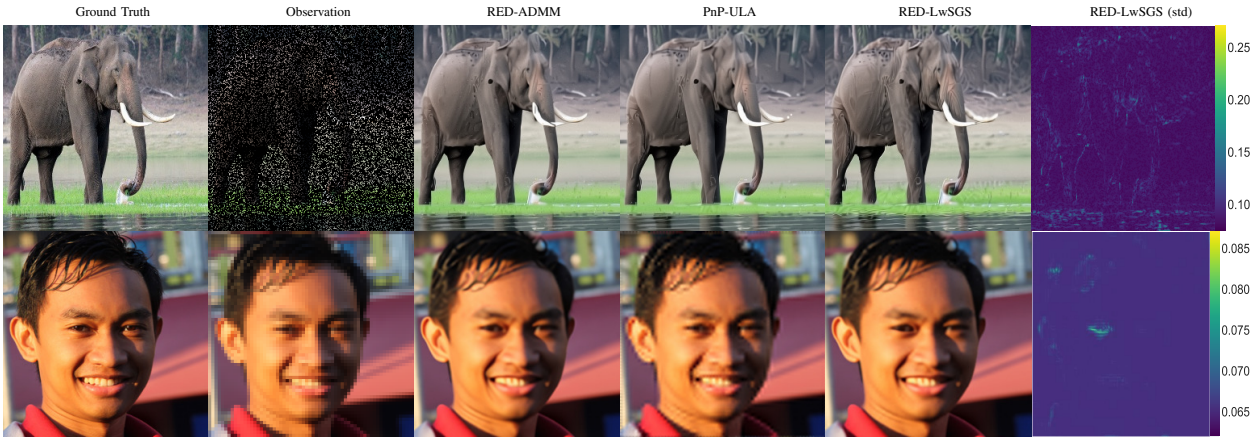

Fig. 1. Images recovered by the compared methods for inpainting (top) and super-resolution (bottom).

better reconstruction. In addition, the Learned Perceptual Image Patch Similarity (LPIPS) [31] is also considered as a proxy of the human-perceived similarity. Additionally, all methods are compared in terms of computational times when the algorithms are implemented on a server equipped with 48 Intel 2.8GHz CPU cores, 384GB RAM, and Nvidia A100 GPU.

B. Experimental results

Table II reports the average PSNR (dB), SSIM and LPIPS results of different methods when performing image inpainting and single image super-resolution. These results show that for the both tasks, the proposed RED-LwSGS method performs quite competitively. In the inpainting experiment, algorithms relying on data-driven regularizations, such as RED-LwSGS and PnP-ULA, appear to include more informative priors when compared to TV-MYULA and TV-SP which rely to the same model-based regularization. Concerning super-resolution, a more challenging problem than inpainting, RED-LwSGS exhibits similar performance to DiffPIR, RED-ADMM, and RED-HQS when other MCMC algorithms are shown to be ineffective. Table II also includes the computation times of each compared algorithm for different restoration tasks. Noticeably, the computational time of RED-LwSGD is similar to that of TV-SP. It remains within a factor of less than 50 compared with PnP-ADMM, RED-ADMM, RED-HQS and DiffPIR. The price to pay to get quantified uncertainties appears very reasonable.

Visual comparisons of the results reached by RED-ADMM, PnP-ADMM and RED-LwSGS can be conducted from Figure 1. The proposed method produces high-quality, sharp and realistic images for the two considered tasks. As already stated, the proposed RED-LwSGS generates samples asymptotically distributed according to the posterior distribution. These samples can be used to measure estimation uncertainty. Figure 1 illustrates this advantage by depicting the estimated pixel-wise standard deviations obtained by the proposed algorithm. As expected, pixels located within uniform regions exhibit lower uncertainty, whereas those within textured areas, edges, or intricate structures tend to be estimated with higher difficulty.

TABLE II
FFHQ DATASET: AVERAGE PERFORMANCE AND CORRESPONDING STANDARD DEVIATIONS. **BOLD**: BEST SCORE, UNDERLINE: SECOND SCORE.

		PSNR (dB)↑	**SSIM**↑	**LPIPS**↓	**Times** (s)↓
Inpainting	**Observation**	7.2069	0.0678	0.5831	-
	RED-LwSGS	30.73±2.932	0.908±0.028	0.023±0.017	74±1
	PnP-ULA	31.46±2.650	0.906±0.026	0.020±0.013	79±1
	TV-MYULA	27.71±1.881	0.830±0.040	0.056±0.026	150±7
	TV-SP	27.27±1.781	0.815±0.041	0.061±0.027	71±1
	RED-ADMM	**31.63**±2.672	0.911±0.024	**0.019**±0.013	3±0
	RED-HQS	31.36±2.293	0.901±0.025	0.021±0.014	2±0
	PnP-ADMM	31.32±3.142	**0.915**±0.042	**0.019**±0.015	2±0
	DiffPIR	31.26±2.25	0.890±0.025	0.021±0.005	2±0
Super-resolution	**RED-LwSGS**	30.43±2.161	0.872±0.036	0.035±0.021	115±25
	PnP-ULA	29.01±2.013	0.847±0.037	0.050±0.024	128±40
	TV-MYULA	28.99±2.017	0.847±0.037	0.049±0.024	133±26
	TV-SP	28.94±2.019	0.846±0.037	0.051±0.024	112±23
	RED-ADMM	30.49±2.222	0.875±0.036	0.034±0.020	3±1
	RED-HQS	30.54±2.206	**0.876**±0.036	0.034±0.020	3±1
	PnP-ADMM	30.13±2.184	0.867±0.037	0.035±0.021	3±1
	DiffPIR	**30.99**±2.212	0.868±0.034	**0.011**±0.008	2±0

IV. CONCLUSION

This paper presented a Bayesian formulation of the regularization by denoising paradigm, enabling a data-driven approach to define prior distributions to solve inversion tasks. To sample from the resulting posterior distribution, a new algorithm was developed. It leveraged an asymptotically exact data augmentation scheme and could be interpreted as a particular instance of the split Gibbs sampler (SGS). It provided a comprehensive description of the posterior distribution, enabling uncertainty to be quantified. Extensive numerical experiments demonstrated that the proposed approach competed favorably with state-of-the-art variational and Monte Carlo methods when performing inpainting and super-resolution tasks. This framework was shown to benefit from the most recent advances in deep learning at the price of a reasonable computational cost.

REFERENCES

[1] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: nonlinear phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

[2] W. C. Karl, "Regularization in image restoration and reconstruction," in *Handbook of image and video processing*. Elsevier, 2005, pp. 183–V.

[3] F. Cao, M. Cai, Y. Tan, and J. Zhao, "Image super-resolution via adaptive ℓ_p $(0 < p < 1)$ regularization and sparse representation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 7, pp. 1550–1561, 2016.

[4] M. V. Afonso, J. M. Bioucas-Dias, and M. A. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681–695, 2010.

[5] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *Proc. IEEE Global Conf. Signal Info. Process. (GlobalSIP)*. IEEE, 2013, pp. 945–948.

[6] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Trans. Image Process.*, vol. 4, no. 7, pp. 932–946, 1995.

[7] K. Zhang, Y. Li, W. Zuo, L. Zhang, L. Van Gool, and R. Timofte, "Plug-and-play image restoration with deep denoiser prior," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 44, no. 10, pp. 6360–6376, 2021.

[8] S. H. Chan, X. Wang, and O. A. Elgendy, "Plug-and-play ADMM for image restoration: Fixed-point convergence and applications," *IEEE Trans. Comput. Imag.*, vol. 3, no. 1, pp. 84–98, 2016.

[9] E. Ryu, J. Liu, S. Wang, X. Chen, Z. Wang, and W. Yin, "Plug-and-play methods provably converge with properly trained denoisers," in *Proc. Int. Conf. Machine Learning (ICML)*. PMLR, 2019, pp. 5546–5557.

[10] S. Hurault, A. Leclaire, and N. Papadakis, "Proximal denoiser for convergent plug-and-play optimization with nonconvex regularization," in *Proc. Int. Conf. Machine Learning (ICML)*. PMLR, 2022, pp. 9483–9505.

[11] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, 2017.

[12] Y. Romano, M. Elad, and P. Milanfar, "The little engine that could: Regularization by denoising (RED)," *SIAM J. Imag. Sci.*, vol. 10, no. 4, pp. 1804–1844, 2017.

[13] J. M. Bardsley, *Computational Uncertainty Quantification for Inverse Problems: An Introduction to Singular Integrals*. SIAM, 2018.

[14] X. Cai, M. Pereyra, and J. D. McEwen, "Uncertainty quantification for radio interferometric imaging –I. Proximal MCMC methods," *Monthly Notices of the Royal Astronomical Society*, vol. 480, no. 3, pp. 4154–4169, 2018.

[15] M. Holden, M. Pereyra, and K. C. Zygalakis, "Bayesian imaging with data-driven priors encoded by neural networks," *SIAM J. Imag. Sci.*, vol. 15, no. 2, pp. 892–924, 2022.

[16] Z. Cai, J. Tang, S. Mukherjee, J. Li, C. B. Schönlieb, and X. Zhang, "NF-ULA: Langevin Monte Carlo with normalizing flow prior for imaging inverse problems," *SIAM J. Imag. Sci.*, 2024.

[17] R. Laumont, V. D. Bortoli, A. Almansa, J. Delon, A. Durmus, and M. Pereyra, "Bayesian imaging using plug & play priors: when Langevin meets Tweedie," *SIAM J. Imag. Sci.*, vol. 15, no. 2, pp. 701–737, 2022.

[18] E. T. Reehorst and P. Schniter, "Regularization by denoising: Clarifications and new interpretations," *IEEE Trans. Comput. Imag.*, vol. 5, no. 1, pp. 52–67, 2018.

[19] M. Vono, N. Dobigeon, and P. Chainais, "Asymptotically exact data augmentation: Models, properties, and algorithms," *J. Comput. Graph. Stat.*, vol. 30, no. 2, pp. 335–348, 2020.

[20] M. Vono, N. Dobigeon, and P. Chainais, "Split-and-augmented Gibbs sampler – Application to large-scale inference problems," *IEEE Trans. Signal Process.*, vol. 67, no. 6, pp. 1648–1661, 2019.

[21] L. J. Rendell, A. M. Johansen, A. Lee, and N. Whiteley, "Global consensus monte carlo," *J. Comput. Graph. Stat.*, vol. 30, no. 2, pp. 249–259, 2020.

[22] M. Vono, N. Dobigeon, and P. Chainais, "Sparse Bayesian binary logistic regression using the split-and-augmented Gibbs sampler," in *Proc. IEEE Workshop Mach. Learning for Signal Process. (MLSP)*, Aalborg, Denmark, Sept. 2018.

[23] M. Vono, N. Dobigeon, and P. Chainais, "Bayesian image restoration under Poisson noise and log-concave prior," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*, Brighton, U.K., April 2019.

[24] R. Dwivedi, Y. Chen, M. J. Wainwright, and B. Yu, "Log-concave sampling: Metropolis-Hastings algorithms are fast!" in *Proc. Conf. Learning Theory (COLT)*. PMLR, 2018, pp. 793–797.

[25] E. C. Faye, M. D. Fall, and N. Dobigeon, "Regularization by denoising: Bayesian model and Langevin-within-split Gibbs sampling," *arXiv preprint*, 2024. [Online]. Available: https://arxiv.org/abs/2402.12292

[26] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proc. Int. Conf. Computer Vision Pattern Recognition (CVPR)*, 2019, pp. 4401–4410.

[27] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. Int. Conf. Computer Vision Pattern Recognition (CVPR)*, 2009, pp. 248–255.

[28] A. Durmus, E. Moulines, and M. Pereyra, "Efficient Bayesian computation by proximal Markov chain Monte Carlo: when Langevin meets Moreau," *SIAM J. Imag. Sci.*, vol. 11, no. 1, pp. 473–506, 2018.

[29] Y. Zhu, K. Zhang, J. Liang, J. Cao, B. Wen, R. Timofte, and L. V. Gool, "Denoising diffusion models for plug-and-play image restoration," in *Int. Conf. Computer Vision Pattern Recognition Workshops (NTIRE)*, 2023.

[30] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.

[31] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proc. Int. Conf. Computer Vision Pattern Recognition (CVPR)*, 2018, pp. 586–595.