# Regularization by denoising: Bayesian model and Langevin-within-split Gibbs sampling

Elhadji C. Faye, Mame Diarra Fall, Nicolas Dobigeon

▶ **To cite this version:**

Elhadji C. Faye, Mame Diarra Fall, Nicolas Dobigeon. Regularization by denoising: Bayesian model and Langevin-within-split Gibbs sampling. IEEE Transactions on Image Processing, 2025, 34, pp.221-234. 10.1109/TIP.2024.3520012 . hal-04480090v2

**HAL Id: hal-04480090**

**https://hal.science/hal-04480090v2**

Submitted on 16 Jan 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Regularization by denoising: Bayesian model and Langevin-within-split Gibbs sampling

Elhadji C. Faye, Mame Diarra Fall and Nicolas Dobigeon, *Senior Member, IEEE*

***Abstract*—This paper introduces a Bayesian framework for image inversion by deriving a probabilistic counterpart to the regularization-by-denoising (RED) paradigm. It additionally implements a Monte Carlo algorithm specifically tailored for sampling from the resulting posterior distribution, based on an asymptotically exact data augmentation (AXDA). The proposed algorithm is an approximate instance of split Gibbs sampling (SGS) which embeds one Langevin Monte Carlo step. The proposed method is applied to common imaging tasks such as deblurring, inpainting and super-resolution, demonstrating its efficacy through extensive numerical experiments. These contributions advance Bayesian inference in imaging by leveraging data-driven regularization strategies within a probabilistic framework.**

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

## I. Introduction

This paper is interested in conducting Bayesian inference about an image $\mathbf{x} \in \mathbb{R}^n$ given the measurements $\mathbf{y} \in \mathbb{R}^m$ related to $\mathbf{x}$ through a statistical model specified by the likelihood function

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

In (1), the potential function $f(\mathbf{x}, \mathbf{y})$ is a fidelity term, i.e., accounting for the consistency of $\mathbf{x}$ with respect to (w.r.t.) the measured data $\mathbf{y}$. In what follows, this potential function will be assumed to be convex and $L_f$-smooth, i.e., continuously differentiable and its gradient is Lipschitz continuous with Lipschitz constant $L_f$. This problem is in line with the most frequently encountered imaging inverse problems such as denoising, deblurring, and inpainting relying on a linear forward model and a Gaussian perturbation. For such tasks, the potential function writes $f(\mathbf{x}, \mathbf{y}) = \frac{1}{2\sigma^2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2$ where $\mathbf{A} \in \mathbb{R}^{m \times n}$ is the degradation matrix. Estimating $\mathbf{x}$ from $\mathbf{y}$ is generally an ill-posed or, at least, ill-conditioned problem. The Bayesian paradigm consists in assigning a prior distribution to $\mathbf{x}$, which summarizes the prior knowledge about $\mathbf{x}$ and acts as a regularization. This prior distribution writes

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

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

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

This posterior distribution provides a comprehensive description of the solutions and allows various Bayesian estimators and uncertainty measures to be derived. In particular, computing the maximum a posteriori (MAP) estimator boils down to solving the minimization problem

$$\min_{\mathbf{x}} f(\mathbf{x}, \mathbf{y}) + \beta g(\mathbf{x}). \tag{4}$$

Numerous works from the literature have focused on the difficult task of designing a relevant prior distribution $p(\mathbf{x})$ or, equivalently, a relevant potential function $g(\mathbf{x})$. These regularizations usually promote specific expected or desired properties about $\mathbf{x}$. More specifically, conventional optimization methods solving (4) are generally based on explicit model-based regularizations, such as total variation (TV) promoting piecewise constant behavior [1], Sobolev promoting smooth content [2] or sparsity-promoting regularizations based on the use of $\ell_p$-norm with $p \leq 1$ [3], [4]. However, designing an appropriate model-based regularization remains an empirical and subjective choice. Moreover, their ability of characterizing complex image structures is generally limited or comes at the price of a significant increase of the resulting algorithmic burden. More recently, a different route has been taken by devising smart strategies avoiding the empirical design of handcrafted model-based regularizations. The seminal work by Venkatakrishnan *et al.* has introduced the concept of *plug-and-play* (PnP) as an implicit prior [5]. This framework naturally emerges whenever the algorithmic scheme designed to solve (4) embeds the proximal operator associated with $g(\cdot)$. Possible schemes include the alternating direction method of multipliers (ADMM) [4], half quadratic splitting (HQS) [6] or Douglas-Rachford splitting [7]. Interestingly, this proximal mapping can be interpreted as a denoising task under the assumption of an additive white Gaussian noise. PnP approaches replace this proximal step by a more general denoiser $\mathsf{D} : \mathbb{R}^n \to \mathbb{R}^n$, including non-local means (NLM) [8], block-matching and 3D filtering (BM3D) [9] or any more recently proposed learning-based denoisers such as DnCNN [10] or DRUNet [11]. Thanks

E. C. Faye and M. D. Fall are with Institut Denis Poisson, University of Orléans, Orléans, France (e-mail: {elhadji-cisse.faye, diarra.fall}@univ-orleans.fr).

N. Dobigeon is with University of Toulouse, IRIT/INP-ENSEEIHT, CNRS, 2 rue Charles Camichel, BP 7122, 31071 Toulouse Cedex 7, France (e-mail: Nicolas.Dobigeon@enseeiht.fr).

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org., provided by the author. The material includes a PDF document containing the detailed proofs of the propositions presented in this manuscript.

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

to its effectiveness and its simplicity, this framework has gained in popularity for a wide range of applications in the context of imaging problems [11]–[14].

In the same vein as PnP, the regularization-by-denoising (RED) framework relies on the sole ability of performing a denoising task to define an explicit image-adaptive Laplacian-based prior [15]. The RED framework requires specific assumptions on the denoiser $\mathsf{D}(\cdot)$, such as differentiability, local homogeneity, strong passivity and Jacobian symmetry [15], [16]. Moreover, the theoretical analyses of most of the iterative algorithms designed to solve the resulting RED-based optimization problems require the denoiser $\mathsf{D}(\cdot)$ to satisfy nonexpansiveness properties [17], [18]. Empirical observations show that such algorithms can even exhibit numerical instabilities when using a denoiser which violates this assumption [17]. One strategy to ensure convergence when using learning-based denoisers consists in imposing spectral normalization during the training phase [13], [19]. More generally, even if it may be difficult nay impossible to ensure that denoisers (e.g. learning-based) satisfy all these quite restrictive assumptions, RED has proven to be empirically effective when performing many restoration tasks and has motivated several subsequent research works. In particular, in the seminal works [5] and [15], the canonical definitions of the so-called PnP and RED frameworks differ by the fact they rely on implicit and explicit functionals, respectively. However, more recent works have introduced a denoiser-based regularization that relaxes the symmetric Jacobian constraint imposed by RED while inherently preserving a similar key property of its gradient [14], [20], [21]. Using this so-called "gradient step denoiser" within a PnP scheme is shown to ensure convergence of the iterates towards a stationary point of an explicit functional. Beyond this convergence guarantee, a noticeable contribution of these works is the following: despite their distinct roots, the PnP and RED frameworks are shown to be much more related than a crude distinction between implicit and explicit functionals.

All the variational approaches discussed above treat $\mathbf{x}$ in a deterministic way and generally produce only point estimates approximating the solution of the minimization problem (4). As an alternative, the Bayesian framework models the image $\mathbf{x}$ as a random variable and generally seeks a comprehensive description of the posterior distribution $p(\mathbf{x}|\mathbf{y})$. As such, Bayesian methods are able to go beyond a sole point estimation, by enhancing it with a quantification of uncertainty in a probabilistic manner in terms of variance and credibility intervals. This ability to quantify uncertainty is particularly useful for decision-making and reliability assessment [22]. As an archetypal illustration of this usefulness, one can mention the field of astronomy where uncertainty quantification contributes to the identification of reliable structures in cosmic images provided by instruments such as radio interferometers [23], [24]. Another field particularly concerned by this need of uncertainty quantification is medical imaging [25], [26]. Indeed, it allows to assess risks and variabilities in diagnostic decisions by computing uncertainties from probabilistic models. This is critical when handling high-stakes decisions, such as in the cases of cancer detection or post-stroke brain analysis. Accurate uncertainty estimations offered by Bayesian techniques improve decision-making and model reliability, ensuring meaningful interpretations. Exploring the posterior distribution is generally carried out by generating samples asymptotically distributed according to this target distribution using Markov chain Monte Carlo (MCMC) methods. Most of the works dedicated to the development of MCMC algorithms for inverse problems in imaging relies on conventional model-based prior distributions. As their deterministic counterparts, they encode expected characteristics of the image prescribed beforehand and chosen based on quite empirical arguments. Very few recent works have attempted to depart from this paradigm by incorporating data- or task-driven regularizations as prior distributions. For instance, available training samples can be used to learn a mapping from an instrumental latent distribution towards the image prior. Benefiting from advances in the machine learning literature, this mapping can be chosen as a deep generative model, such as a variational autoencoder [27] or a normalizing flow [28]. Devising a PnP prior in the context of Monte Carlo sampling has been investigated in [29], resulting in the so-called PnP unadjusted Langevin algorithm (PnP-ULA). Its rationale follows the same motivation as its deterministic counterpart, namely avoiding the explicit definition of the prior distribution by the ability of performing a denoising task.

Surprisingly, despite its interesting performance and abundance of works devising on the RED paradigm within a variational framework, up to authors' knowledge, RED has never been considered from a Bayesian perspective and embedded into a Monte Carlo algorithm. The main objective of this paper is to fill this gap. More precisely, the main contributions reported hereafter can be summarized as follows. First, Section II introduces a probabilistic counterpart of RED by defining a new distribution that can be subsequently chosen as a prior distribution in a Bayesian inversion task. Then, Section III introduces a new Monte Carlo algorithm that is shown to be particularly well suited to sample from the resulting posterior distribution. It follows an asymptotically exact data augmentation (AXDA) scheme [30], resulting in a nonstandard instance of the split Gibbs sampler (SGS) [31]. It is worth noting that similar splitting-based samplers have been implemented in very recent works to allow denoising diffusion models to be easily embedded as generative priors into Bayesian models [32]–[34]. All of them, as well as the proposed algorithm, depart from the the canonical instance of the split Gibbs sampler whose theoretical properties have been deeply investigated in [35]. More precisely, they replace the exact sampling of the auxiliary (i.e., splitting) variable by an *ad hoc* sampling step conveniently afforded to the specific nature of the prior model. Thus Section III also reports a thorough theoretical analysis to ensure and quantify convergence of the proposed algorithm. This analysis relies on standard assumptions frequently encountered in the optimization and Monte Carlo sampling literature. Some of these assumptions may appear to be technical as they may be difficult to assess in practical scenarios, in particular when leveraging recent deep networks whose comprehensive theoretical characterization is out of reach. However, this analysis ensures that the proposed algorithmic scheme is sound in a more formal

setting. Besides, the rationale of the proposed approach is also put into perspective w.r.t. recently proposed Monte Carlo algorithms, in particular PnP-ULA, drawing some connections between AXDA and RED leveraging the Tweedie's formula [36]. Then, the proposed algorithm is instantiated to solve three ubiquitous inversion tasks, namely deblurring, inpainting and super-resolution. Extensive numerical experiments are conducted in Section IV to compare the performance of the proposed algorithm to state-of-the-art variational and Monte Carlo methods. One would like to emphasize that the primary ambition of the work reported in this paper is not to naively promote the RED engine as the consistently state-of-the-art regularization to be universally considered for solving all imaging inverse problems. Instead, it is rather to show that there is a way to reframe this framework into a fully Bayesian setting, benefiting from the advantages of RED while being aware of its limitations discussed above.

**Notations and conventions.** The Euclidean norm on $\mathbb{R}^n$ is denoted by $\|\cdot\|$. We denote by $\mathcal{N}(\boldsymbol{\mu}, \mathbf{Q}^{-1})$ the Gaussian distribution with mean vector $\boldsymbol{\mu}$ and precision matrix $\mathbf{Q}$. The $(n \times n)$-identity matrix is denoted $\mathbf{I}_n$. For any matrix $\mathbf{S} \in \mathcal{M}_n(\mathbb{R})$, if we denote $\mathbf{0}$ the zero matrix, the notation $\mathbf{0} \preccurlyeq \mathbf{S}$ means that $\mathbf{S}$ is semi-definitive positive. The Wasserstein distance of order 2 between two probability measures $\tau$ and $\tau'$ on $\mathbb{R}^n$ with finite 2-moments is defined by $W_2(\tau, \tau') = (\inf_{\zeta \in \mathcal{T}(\tau,\tau')} \int_{\mathbb{R}^n \times \mathbb{R}^n} \| V - V' \|^2 \, \mathrm{d}\zeta(V, V'))^{1/2}$, where $\mathcal{T}(\tau, \tau')$ is the set of transport plans of $\tau$ and $\tau'$.

## II. Bayesian formulation of RED inversion

This section starts by recalling some background about RED. Then it proposes a probabilistic counterpart of the regularization, that can be subsequently used as a prior within a Bayesian framework.

### A. Regularization by denoising (RED)

The RED engine defines $g(\cdot)$ as the explicit image-adaptive Laplacian-based potential [15]

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

where $\mathsf{D}_\nu : \mathbb{R}^n \to \mathbb{R}^n$ is a denoiser with $\nu$ controlling the denoising strength, designed for the removal of additive white Gaussian noise. Although it offers a significant flexibility in the choice of the denoisers that can be used, RED requires $\mathsf{D}_\nu(\cdot)$ to obey the following assumptions, referred to as the RED conditions.

(C1) *Local homogeneity*: $\forall \mathbf{x} \in \mathbb{R}^n$,

$$\mathsf{D}_\nu\left((1+\epsilon)\mathbf{x}\right) = (1+\epsilon)\mathsf{D}_\nu(\mathbf{x}) \tag{6}$$

for any sufficiently small $\epsilon > 0$.

(C2) *Differentiability*: the denoiser $\mathsf{D}_\nu(\cdot)$ is differentiable with Jacobian denoted $\nabla\mathsf{D}_\nu(\cdot)$.

(C3) *Jacobian symmetry* [16]: $\forall \mathbf{x} \in \mathbb{R}^n$, $\nabla\mathsf{D}_\nu(\mathbf{x})^\top = \nabla\mathsf{D}_\nu(\mathbf{x})$.

(C4) *Strong passivity*: the Jacobian spectral radius satisfies $\eta\left(\nabla\mathsf{D}_\nu(\mathbf{x})\right) \leq 1$.

The major implication of local homogeneity (C1) is that the directional derivative of $\mathsf{D}_\nu(\cdot)$ along $\mathbf{x}$ can be computed by applying the denoiser itself, i.e.,

$$\nabla\mathsf{D}_\nu(\mathbf{x})\mathbf{x} = \mathsf{D}_\nu(\mathbf{x}). \tag{7}$$

The Jacobian symmetry (C3) and the strong passivity (C4) ensure that applying the denoiser does not increase the norm of the input:

$$\|\mathsf{D}_\nu(\mathbf{x})\| = \|\nabla\mathsf{D}_\nu(\mathbf{x})\mathbf{x}\| \leq \eta\left(\nabla\mathsf{D}_\nu(\mathbf{x})\right) \cdot \|\mathbf{x}\| \leq \|\mathbf{x}\|. \tag{8}$$

Interestingly, two additional keys and highly beneficial properties follow: *i)* the RED potential $g_{\text{red}}(\cdot)$ is a convex functional and *ii)* the gradient of $g_{\text{red}}(\cdot)$ is expressed as the denoising residual

$$\nabla g_{\text{red}}(\mathbf{x}) = R(\mathbf{x}) = \mathbf{x} - \mathsf{D}_\nu(\mathbf{x}) \tag{9}$$

which avoids differentiating the denoising operation itself. Thus, one of the most appealing opportunity offered by RED is its ability to embed powerful denoisers, such as those based on deep neural networks, without requiring to differentiate them. It is worth noting that if the denoising function $\mathsf{D}_\nu(\cdot)$ does not meet the condition (C3), then there is no regularizer $g(\cdot)$ whose gradient can be written as the residual $R(\cdot)$ [16]. Unfortunately many popular denoisers, such as trainable nonlinear reaction-diffusion (TNRD), NLM, BM3D, and DnCNN, are characterized by non-symmetric Jacobian. Yet, RED-based restoration algorithms are shown to empirically converge and to reach excellent performance when solving various inverse problems even when those conditions are partially satisfied [15], [20].

### B. Probabilistic counterpart of RED

To formulate the RED-based inversion within a statistical framework, one requirement consists in introducing a prior distribution defined from the RED potential $g_{\text{red}}(\cdot)$ given by (5). More precisely, one defines

$$p_{\text{red}}(\mathbf{x}) \propto \exp\left[-\frac{\beta}{2}\mathbf{x}^\top (\mathbf{x} - \mathsf{D}_\nu(\mathbf{x}))\right]. \tag{10}$$

The functional $p_{\text{red}}(\cdot)$ does not necessarily define a probability density function (pdf). For $p_{\text{red}}(\cdot)$ to be a valid pdf, i.e. $\int_{\mathbb{R}^n} p_{\text{red}}(\mathbf{x})\mathrm{d}\mathbf{x} < \infty$, certain conditions must be satisfied.

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

This technical assumption is not restrictive and it is easy to show that it would be violated only in trivial cases. Indeed, let $\bar{\mathbf{x}} \in \mathbb{R}^n$ denote an image such that all the eigenvalues of $\boldsymbol{\Lambda}(\bar{\mathbf{x}})$ are zero. Then the matrix $\nabla\mathsf{D}_\nu(\bar{\mathbf{x}})$ is symmetric (see RED condition (C3)), with real coefficients and all eigenvalues equal to 1. The spectral theorem yields $\nabla\mathsf{D}_\nu(\bar{\mathbf{x}}) = \mathbf{I}_n$. From the local homogeneity (C1) and its corollary (7), one has $\mathsf{D}_\nu(\bar{\mathbf{x}}) = \bar{\mathbf{x}}$, i.e., $\bar{\mathbf{x}}$ is already a noise-free image which does not need to be further denoised.

The next result states that mild assumptions are sufficient to guarantee that the function (10) defines a proper distribution.

**Proposition 1.** *If Assumption 1 and Conditions (C3)–(C4) hold, then*

$$\int_{\mathbb{R}^n} p_{red}(\mathbf{x})\mathrm{d}\mathbf{x} < +\infty \tag{11}$$

*and $p_{red}(\cdot)$ in (10) defines a proper pdf.*

*Proof.* See Appendix A. □

### C. RED posterior distribution

Combining the RED prior $p_{\text{red}}(\mathbf{x})$ defined by (10) and the likelihood function $p(\mathbf{y}|\mathbf{x})$ defined by (1), the RED posterior distribution $\pi$ of interest writes

$$\begin{aligned}\pi(\mathbf{x}) &\triangleq p(\mathbf{x}|\mathbf{y}) \\ &\propto \exp\left[-f(\mathbf{x},\mathbf{y}) - \frac{\beta}{2}\mathbf{x}^\top\left(\mathbf{x}-\mathsf{D}_\nu(\mathbf{x})\right)\right].\end{aligned} \tag{12}$$

As stated earlier, deriving the MAP estimator associated with the RED posterior (12) consists in solving the optimization problem (4). In the seminal paper [15], this problem is tackled thanks to first-order optimization methods such as steepest decent (SD), fixed-point (FP) iteration and ADMM. More recently, it has been reformulated as a convex optimization problem using a projection onto the fixed point set of demi-contractive denoisers [20]. Instead, the work conducted in this manuscript proposes to follow a different route by proposing to sample from this posterior distribution. While these samples offer a comprehensive description of the RED posterior, they can be subsequently used to derive Bayesian estimators or credibility intervals. Because of the non-standard form of the RED posterior, sampling according to (12) requires to develop a dedicated algorithm introduced in the following section.

## III. Proposed algorithm

### A. Langevin-within-split Gibbs sampler

Generating samples efficiently from the posterior distribution with pdf $\pi(\mathbf{x})$ defined by (12) is not straightforward, in particular due to the use of the denoiser $\mathsf{D}_\nu(\cdot)$. When $\pi(\cdot)$ is proper and smooth with $\mathbf{x} \mapsto \nabla \log \pi(\mathbf{x})$ Lipschitz continuous, one solution would consist in resorting to the ULA [37]. This strategy will be shown to be intimately related to PnP-ULA in Section III-D. However, it may suffer from several shortcomings, such as poor mixing properties and higher resulting computational times (see experimental results in Section IV-D). Conversely, the work in this manuscript derives a dedicated Monte Carlo algorithm to sample from a posterior distribution written as (3). This algorithm will be shown to be particularly well suited to sample from the RED posterior (12), i.e., when $g(\cdot) = g_{\text{red}}(\cdot)$.

The proposed sampling scheme first leverages an asymptotically exact data augmentation (AXDA) as introduced in [30]. Inspired by optimization-flavored counterparts, AXDA employs a variable splitting technique to simplify and speed up the sampling according to possibly complex distributions. More precisely, it introduces an auxiliary variable $\mathbf{z} \in \mathbb{R}^n$ and considers the augmented distribution

$$\begin{aligned}\pi_\rho(\mathbf{x},\mathbf{z}) &= p(\mathbf{x},\mathbf{z}|\mathbf{y};\rho^2) \\ &\propto \exp\left[-f(\mathbf{x},\mathbf{y}) - \beta g(\mathbf{z}) - \frac{1}{2\rho^2}||\mathbf{x}-\mathbf{z}||^2\right]\end{aligned} \tag{13}$$

where $\rho$ is a positive parameter that controls the dissimilarity between $\mathbf{x}$ and $\mathbf{z}$. This data augmentation (13) is approximate in the sense that the marginal distribution

$$\begin{aligned}\pi_\rho(\mathbf{x}) &= \int_{\mathbb{R}^n} \pi_\rho(\mathbf{x},\mathbf{z})d\mathbf{z} \\ &\propto \int_{\mathbb{R}^n} \exp\left[-f(\mathbf{x},\mathbf{y}) - \beta g(\mathbf{z}) - \frac{1}{2\rho^2}||\mathbf{x}-\mathbf{z}||^2\right]\mathrm{d}\mathbf{z}\end{aligned} \tag{14}$$

coincides with the target distribution $\pi(\mathbf{x})$ only in the limiting case $\rho \to 0$. The conditional distributions[1] associated to the augmented posterior $\pi_\rho(\mathbf{x},\mathbf{z})$ are given by

$$p(\mathbf{x}|\mathbf{y},\mathbf{z}) \propto \exp\left[-f(\mathbf{x},\mathbf{y}) - \frac{1}{2\rho^2}||\mathbf{x}-\mathbf{z}||^2\right] \tag{15}$$

$$p(\mathbf{z}|\mathbf{x}) \propto \exp\left[-\beta g(\mathbf{z}) - \frac{1}{2\rho^2}||\mathbf{x}-\mathbf{z}||^2\right]. \tag{16}$$

The so-called split Gibbs sampler (SGS) alternatively samples according to these two conditional distributions to generate samples asymptotically distributed according to (13) [31], [38]. Interestingly, this splitting allows the two terms $f(\cdot,\mathbf{y})$ and $g(\cdot)$ defining the full potential to be dissociated and involved into two distinct conditional distributions. SGS shares strong similarities with ADMM and HQS methods and is expected to lead to simpler, scalable and more efficient sampling schemes.

Specifically, sampling according to the conditional posterior (15) can be interpreted as solving the initial estimation problem defined by the likelihood function (1) with now a Gaussian distribution with mean $\mathbf{z}$ and diagonal covariance matrix $\rho^2\mathbf{I}_n$ assigned as a prior. As stated earlier, a large family of imaging inverse problems, such as deblurring, inpainting and super-resolution is characterized by the quadratic potential function $f(\mathbf{x},\mathbf{y}) = \frac{1}{2\sigma^2}\|\mathbf{A}\mathbf{x}-\mathbf{y}\|_2^2$ also considered in this work. This leads to the Gaussian conditional distribution

$$p(\mathbf{x}|\mathbf{y},\mathbf{z}) = \mathcal{N}(\mathbf{x};\boldsymbol{\mu}(\mathbf{z}),\mathbf{Q}^{-1}) \tag{17}$$

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

$$\left\{\begin{array}{rcl}\mathbf{Q} &=& \frac{1}{\sigma^2}\mathbf{A}^\top\mathbf{A} + \frac{1}{\rho^2}\mathbf{I} \\ \boldsymbol{\mu}(\mathbf{z}) &=& \mathbf{Q}^{-1}\left(\frac{1}{\sigma^2}\mathbf{A}^\top\mathbf{y} + \frac{1}{\rho^2}\mathbf{z}\right).\end{array}\right. \tag{18}$$

In this case, sampling according to this conditional distribution can be efficiently achieved using dedicated algorithms that depend on the structure of the precision matrix $\mathbf{Q}$. Interested readers are invited to consult [39] for a recent overview of these methods. It is worth noting that when the potential function $f(\cdot,\mathbf{y})$ is not quadratic, the proposed framework can embed proximal Monte Carlo algorithms to sample from (15), as in [40], [41].

In the specific case considered in this work where $g(\cdot) = g_{\text{red}}(\cdot)$, the conditional distribution (16) can be interpreted as the posterior distribution associated to a Bayesian denoising problem equipped with a RED prior. Its objective boils down to inferring an object $\mathbf{z}$ from the observations $\mathbf{x}$ contaminated by

[1]The conditional distributions associated to $\pi_\rho(\mathbf{x},\mathbf{z})$ are $p(\mathbf{x}|\mathbf{y},\mathbf{z};\rho^2)$ and $p(\mathbf{z}|\mathbf{x};\rho^2)$. To lighten the notations, the coupling parameter $\rho^2$ will be omitted in what follows, i.e., one employs $p(\mathbf{x}|\mathbf{y},\mathbf{z})$ and $p(\mathbf{z}|\mathbf{x})$.

an additive white Gaussian noise with a covariance matrix $\rho^2\mathbf{I}_n$. Sampling according to this conditional is not straightforward, in particular due to the regularization potential $g_{\text{red}}(\cdot)$ whose definition involves the denoiser $\mathsf{D}_\nu(\cdot)$. This work proposes to take advantage of the property (9) by sampling from (16) following a Langevin Monte Carlo (LMC) step, i.e.,

$$\mathbf{z}^{(t+1)} = \mathbf{z}^{(t)} + \gamma\nabla\log p\left(\mathbf{z}^{(t)} \mid \mathbf{x}\right) + \sqrt{2\gamma}\boldsymbol{\varepsilon}^{(t)} \tag{19}$$

where $\gamma > 0$ is a fixed step-size and $\left\{\boldsymbol{\varepsilon}^{(t)}\right\}_{t\in\mathbb{N}}$ is a sequence of independent and identically distributed $n$-dimensional standard Gaussian random variables. Given the particular form of the conditional distribution (16) and the property (9), this recursion writes explicitly as

$$\begin{aligned}\mathbf{z}^{(t+1)} = &\left(1 - \gamma\beta - \frac{\gamma}{\rho^2}\right)\mathbf{z}^{(t)} \\ &+ \frac{\gamma}{\rho^2}\mathbf{x}^{(t)} + \gamma\beta\mathsf{D}_\nu\left(\mathbf{z}^{(t)}\right) + \sqrt{2\gamma}\boldsymbol{\varepsilon}^{(t)}.\end{aligned} \tag{20}$$

The proposed so-called Langevin-within-SGS (LwSGS) instantiated to sample according to the RED posterior (12) is summarized in Algo 1.

**Algorithm 1** LwSGS to sample from the RED posterior

**Input:** denoiser $\mathsf{D}_\nu(\cdot)$, regularization parameter $\beta$, coupling parameter $\rho^2$, step-size $\gamma$, number of burn-in iterations $N_{\text{bi}}$, total number of iterations $N_{\text{MC}}$

**Initialization:** $\mathbf{x}^{(0)}$, $\mathbf{z}^{(0)}$

1: **for** $t = 0$ to $N_{\text{MC}} - 1$ **do**

*% Sampling the splitting variable according to (20)*

2: $\boldsymbol{\varepsilon}^{(t)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$

3: $\mathbf{z}^{(t+1)} = \left(1 - \gamma\beta - \frac{\gamma}{\rho^2}\right)\mathbf{z}^{(t)} + \frac{\gamma}{\rho^2}\mathbf{x}^{(t)} + \gamma\beta\mathsf{D}_\nu\left(\mathbf{z}^{(t)}\right) + \sqrt{2\gamma}\boldsymbol{\varepsilon}^{(t)}$

*% Sampling the variable of interest according to (15)*

4: $\mathbf{x}^{(t+1)} \sim \mathcal{N}\big(\boldsymbol{\mu}(\mathbf{z}^{(t+1)}), \mathbf{Q}^{-1}\big)$

5: **end for**

**Output:** collection of samples $\{\mathbf{x}^{(t)}, \mathbf{z}^{(t)}\}_{t=N_{\text{bi}}+1}^{N_{\text{MC}}}$

Because of the discretization followed by the LMC step (19), the samples produced by (20) are biased and are not exactly distributed according to (16). To mitigate this bias and ensure that LMC exactly targets (16), one well-admitted strategy consists in including a Metropolis-Hasting (MH) step, resulting in the Metropolis adjusted Langevin algorithm [42]. Then, combined with the sampling according to (15), the overall resulting sampling algorithm would become a canonical instance of Metropolis-within-Gibbs algorithm whose samples would be ensured to be distributed according to the augmented posterior distribution (13). However, performing this MH step within each iteration of the SGS requires to compute multiple corresponding MH ratios and to accept or reject the proposed samples, which may significantly increase the computational burden of the SGS. In this work, one proposes to bypass this MH correction, yet at the price of an approximation which is controlled. Indeed, the bias induced by the use of a LMC step within a SGS iteration will be investigated in the theoretical analysis conducted in Section III-C.

### *B. Related Monte Carlo algorithms*

The proposed LwSGS algorithm shares some similarities with some recently developed Monte Carlo algorithms, mainly motivated by the will of *i)* avoiding the explicit definition of prior models or *ii)* fastening inference possibly by distributed computations over several nodes. Interestingly these two motivations are the same as the ones explaining the success of splitting-based optimization algorithms (e.g., HQS and ADMM). On one hand, adopting the AXDA strategy underlying (13), the conditional distribution (16) of the auxiliary variable can be interpreted as the posterior distribution associated with a Bayesian denoising problem. Akin to the PnP paradigm, this simple observation suggests replacing the exact sampling from (16) by the use of an off-the-shelf stochastic denoiser, e.g., chosen as a diffusion-based or score-based model. This is the strategy followed in several recent works [32]–[34], with the noticeable advantage of avoiding an explicit prior model as in the proposed RED-LwSGS approach. Whereas the PnP-SGS introduced in [32] is not accompanied by a theoretical analysis, e.g., to ensure the existence of a stationary distribution, the so-called PnP diffusion model and diffusion PnP method designed in [33] and [34], respectively, are granted with such theoretical insights.

On another hand, mainly motivated by numerical efficiency considerations, the authors in [43] have adopted the splitting strategy offered by AXDA to derive a distributed SGS (DSGS) when the posterior distribution comprises multiple composite terms. Again, as with LwSGS, the core idea of DSGS can be sketched as replacing the exact sampling of one conditional distribution of the augmented posterior distribution by a more efficient surrogate sampling technique. Given the particular form of the posterior distributions considered in [43], a suitable choice of this surrogate is shown to be one step of a proximal stochastic gradient Langevin algorithm (PSGLA) [44]. When this sampling is not corrected by a MH step, this leads to an inexact instance of SGS coined as PSGLA-within-SGS. The synchronous distributed version of PSGLA-within-SGS accounts for the hypergraph structure of the involved composite terms to efficiently distribute the variables over multiple workers under controlled communication costs. In the same vein, another synchronous distributed MCMC algorithm referred to as DG-LMC has been introduced in [45] to conduct Bayesian inference when the target log-posterior also writes as a sum of multiple composite terms. Also leveraging AXDA, it adopts a splitting scheme different from LwSGS and PSGLA-within-SGS. Yet, it can be interpreted as an inexact SGS for which multiple conditional distributions are approximately sampled thanks to LMC steps. While the existence of a stationary distribution targeted by PSGLA-within-SGS and its convergence have not been demonstrated in [43], such a thorough theoretical analysis has been conducted for DG-LMC in [45]. Even if these results are a precious asset to conduct a similar analysis of LwSGS, they should be carefully adapted to fit the splitting scheme adopted by LwSGS.

### C. Theoretical analysis

This section provides theoretical insights regarding the proposed LwSGS algorithm. For sake of generality and unless otherwise stated, this analysis is conducted for any regularization potential $g(\cdot)$ satisfying standard assumptions detailed below and frequently encountered in the optimization and Monte Carlo sampling literature. Although it may be difficult to assess the validity of these assumptions in practice (in particular when using learning-based regularizations), their statement enables a rigorous and comprehensive theoretical study of the proposed sampling algorithm. Moreover, along this section, these assumptions will be also discussed and examined under the prism of the RED paradigm, i.e., with $g(\cdot) = g_{\text{red}}(\cdot)$ and when the proposed LwSGS aims at targeting the RED posterior distribution (12).

As stated above, because of the absence of MH correction after the LMC step, Algo. 1 does not fall into the class of Metropolis-within-Gibbs samplers. Thus, the primer objective of this analysis is to demonstrate that the samples produced by Algo. 1 are asymptotically distributed according to a unique invariant distribution[2] $\pi_{\rho,\gamma}$ following an ergodic transition kernel denoted $P_{\rho,\gamma}$. Thanks to an appropriate synchronous coupling, the convergence analysis of LwSGS reduces to that of the Markov chain produced by the sampling (20) according to the conditional distribution $p(\mathbf{z}|\mathbf{x})$. The proofs of the main results stated in what follows, namely Propositions 2 and 3, are reported in the supplementary materials [46]. One first introduces and discusses two assumptions regarding the regularization potential $g(\cdot)$.

**Assumption 2** (Twice differentiability)**.** *The potential function $g(\cdot)$ is twice continuously differentiable and there exists $M_g > 0$ such that $\forall \mathbf{z} \in \mathbb{R}^n$, $\|\nabla^2 g(\mathbf{z})\| \leq M_g$.*

As stated in Section II-A, under the RED conditions (C1) and (C3), the gradient of the RED potential is given by (9). This implies that the regularization potential $g_{\text{red}}(\cdot)$ is twice continuously differentiable with Hessian matrix $\nabla^2 g_{\text{red}}(\mathbf{z}) = \mathbf{I}_n - \nabla \mathsf{D}_\nu(\mathbf{z})$. Moreover, thanks to the Jacobian symmetry (C3) and strong passivity (C4) conditions, one has for all $\mathbf{z} \in \mathbb{R}^n$, $\|\nabla^2 g_{\text{red}}(\mathbf{z})\| \leq 2$. In other words, Assumption 2 always holds for RED.

**Assumption 3** (Strong convexity)**.** *The potential function $g(\cdot)$ is $m_g$-strongly convex, i.e., there exists $m_g > 0$ such that $m_g \mathbf{I}_n \preceq \nabla^2 g$.*

In the RED framework, a sufficient condition for the strong convexity of the potential $g_{\text{red}}(\cdot)$ is to ensure that the denoiser $\mathsf{D}_\nu(\cdot)$ is contractive, i.e., $\forall (\mathbf{z}_1, \mathbf{z}_2) \in \mathbb{R}^n \times \mathbb{R}^n$, $\|\mathsf{D}_\nu(\mathbf{z}_1) - \mathsf{D}_\nu(\mathbf{z}_2)\|_2 \leq \epsilon \|\mathbf{z}_1 - \mathbf{z}_2\|_2$ for some Lipschitz constant $\epsilon < 1$. Under this condition, $g_{\text{red}}(\cdot)$ can be shown to be $m_g$-strongly convex with $m_g = 1 - \epsilon > 0$. Unfortunately, most existing denoisers do not follow this contraction property [14]. To ensure the strong convexity of the potential $g_{\text{red}}(\cdot)$ when using a deep network-based denoiser, one solution would consist in explicitly including a regularization term into the training loss which constrains the Lipschitz constant [13], [18], [47]. Finally, it is worth noting that when Assumption 3 is satisfied, Assumption 1 is also satisfied, which implies that $p_{\text{red}}(\cdot)$ is well-defined.

Under these assumptions, the convergence of the proposed LwSGS algorithm is stated in the following proposition.

**Proposition 2.** *Let $\gamma \in \mathbb{R}_+^*$ such that $\gamma \leq (\beta M_g + 1/\rho^2)^{-1}$. Then, under Assumptions 2 and 3, the kernel $P_{\rho,\gamma}$ admits a unique stationary distribution $\pi_{\rho,\gamma}$. Moreover, for any $\mathbf{v} = (\mathbf{x}, \mathbf{z})^\top \in \mathbb{R}^n \times \mathbb{R}^n$ and any $t \in \mathbb{N}^*$, we have*

$$W_2^2(\delta_{\mathbf{v}} P_{\rho,\gamma}^t, \pi_{\rho,\gamma}) \leq C_1 \big(1 - \gamma\beta m_g\big)^{2(t-1)} W_2^2(\delta_{\mathbf{v}}, \pi_{\rho,\gamma}),$$

*where $C_1 = 1 + \frac{1}{\rho^2}\|\mathbf{Q}^{-1}\|^2$.*

*Proof.* See [46, Section 1]. □

From this proposition, it appears that the rate of geometric ergodicity in Wasserstein distance is given by $r = 1 - \gamma\beta m_g$. This rate quantifies the speed of convergence of the Markov chain produced by the kernel $P_{\rho,\gamma}$ towards its stationary distribution $\pi_{\rho,\gamma}$. It depends on two parameters of the model, namely the regularization parameter $\beta$ and the strong convexity parameter $m_g$. Noticeably, it also depends on an algorithmic parameter whose tuning can be left to the end-user, namely the LMC discretization step size $\gamma$: the larger $\gamma$, the faster the convergence. However, the asymptotic convergence of the samples produced by the kernel $P_{\rho,\gamma}$ towards the distribution $\pi_{\rho,\gamma}$ is only possible if this step size $\gamma$ is sufficiently small. Indeed, this proposition establishes convergence for any step size $\gamma \leq (\beta M_g + 1/\rho^2)^{-1}$. From Assumptions 2 and 3, $-\log p(\mathbf{z}|\mathbf{x})$ is $\beta m_g$-strongly convex and $(\beta M_g + 1/\rho^2)$-smooth, i.e., $\beta m_g \mathbf{I}_n \preceq -\nabla^2 \log p(\mathbf{z}|\mathbf{x}) \preceq (\beta M_g + 1/\rho^2)\mathbf{I}_n$. From [45] and [37], a sufficient condition on the step size $\gamma$ to ensure contraction w.r.t. the Wasserstein distance is $\gamma \leq 2/(\beta m_g + \beta M_g + 1/\rho^2)$. Thus the step size $\gamma$ plays a critical role in determining the stability and the convergence speed of the algorithm. A larger step size may lead to a faster convergence but with possible instability if it exceeds the threshold $(\beta M_g + 1/\rho^2)^{-1}$. On the other hand, a smaller $\gamma$ ensures stability but it slows down the convergence. Besides, the upper bound given in this proposition depends on the coupling parameter $\rho$ and the smoothing parameter $M_g$. Thus these two model parameters implicitly impact the convergence rate. Specifically, a smaller $\rho$ and a larger $M_g$ tighten the upper bound on the step size, leading to slower convergence. Conversely, a larger $\rho$ and a smaller $M_g$ allow for larger step sizes $\gamma$, which can accelerate convergence.

Once the asymptotic convergence of the samples produced by Algo. 1 has been ensured, the second stage of the theoretical analysis consists in analyzing the bias between the stationary distribution $\pi_{\rho,\gamma}$ and the targeted augmented distribution $\pi_\rho$. This bias, which results from the use of the LMC step in Algo. 1 to target the conditional distribution $p(\mathbf{z}|\mathbf{x})$ defined by (16), is quantified in the following proposition.

[2]With a slight abuse of notation, one uses the same notations for a probability distribution and its associated pdf.

**Proposition 3.** *Let $\gamma \in \mathbb{R}_+^*$ such that $\gamma \leq 2(\beta m_g + \beta M_g + 1/\rho^2)^{-1}$. Then, under Assumptions 2 and 3, we have*

$$W_2^2(\pi_\rho, \pi_{\rho,\gamma}) \leq n\gamma C_2 \tilde{M}^2 \Big(1 + \frac{\gamma^2 \tilde{M}^2}{12} + \frac{\gamma \tilde{M}^2}{2\tilde{m}}\Big),$$

*where $\tilde{m} = \beta m_g + 1/\rho^2$, $\tilde{M} = \beta M_g + 1/\rho^2$ and $C_2 = \frac{2}{\beta m_g}(1 + \frac{1}{\rho^2}\|\mathbf{Q}^{-1}\|^2)$.*

*Proof.* See [46, Section 2]. □

This bias is upper-bounded by a term driven by the step size. More precisely, as $\gamma$ decreases towards zero, the squared Wasserstein distance $W_2^2(\pi_\rho, \pi_{\rho,\gamma})$ is upper-bounded by $\mathcal{O}(n\gamma)$. This result is in agreement with those stated in [45, Proposition 4] and [37, Corollary 7]. Besides, it is worth recalling that adopting AXDA leads to sampling from an approximate posterior distribution $\pi_\rho$ defined by (14) while the "ideal" target distribution is $\pi$ in (12). Thus the price to pay for benefiting from the advantages brought by AXDA (see Section III-B) lies in another bias that could be measured as $W_2^2(\pi_\rho, \pi)$. The quality of this approximation is obviously governed by the coupling parameter $\rho$. Previous studies have deeply investigated its impact and have shown that this bias is of order $\mathcal{O}(\rho)$ for sufficiently small values of $\rho$ [30], [45]. Since it also impacts the convergence speed of LwSGS (see Proposition 2 and the related discussion), this coupling parameter follows a certain trade-off between efficiency and accuracy.

### *D. Revisiting PnP-ULA and AXDA from the RED paradigm*

This section draws connections between AXDA, the proposed algorithm and PnP-ULA [29]. As a reminder, in a nutshell, PnP-ULA targets a posterior distribution of the form

$$p_\epsilon(\mathbf{x}|\mathbf{y}) \propto \exp\left[-f(\mathbf{x}, \mathbf{y})\right] p_\epsilon(\mathbf{x}) \tag{21}$$

with $\epsilon > 0$ where

$$p_\epsilon(\mathbf{x}) \propto \int_{\mathbb{R}^n} p(\mathbf{z}) \exp\left[-\frac{1}{2\epsilon}\|\mathbf{x} - \mathbf{z}\|^2\right] \mathrm{d}\mathbf{z} \tag{22}$$

defines the regularized counterpart of the prior distribution $p(\mathbf{x})$. Interestingly, when the prior writes $p(\mathbf{x}) \propto \exp\left[-\beta g(\mathbf{x})\right]$ and $\rho^2 = \epsilon$, the posterior distribution $p_\epsilon(\mathbf{x}|\mathbf{y})$ targeted by PnP-ULA (21) perfectly matches the marginal distribution $\pi_\rho(\mathbf{x})$ in (14) resulting from an AXDA strategy and targeted by a SGS.

Besides, from an algorithmic point of view, ULA recursions applied to (21) conventionally write

$$\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} - \gamma \nabla f(\mathbf{x}^{(t)}, \mathbf{y}) + \gamma \nabla \log p_\epsilon(\mathbf{x}^{(t)}) + \sqrt{2\gamma}\boldsymbol{\varepsilon}^{(t)}.$$

Thanks to the Tweedie's identity [36], the score function can be replaced by the denoising residual, i.e., $\epsilon \nabla \log p_\epsilon(\mathbf{x}) = \mathsf{D}_\epsilon^*(\mathbf{x}) - \mathbf{x}$ where $\mathsf{D}_\epsilon^*(\cdot)$ is a minimum mean square error (MMSE) denoiser. This leads to the (simplified) PnP-ULA scheme

$$\begin{aligned}\text{(PnP-ULA)}: \ \mathbf{x}^{(t+1)} &= \mathbf{x}^{(t)} - \gamma \nabla f(\mathbf{x}^{(t)}, \mathbf{y}) \\ &\quad + \frac{\gamma}{\epsilon}\left(\mathsf{D}_\epsilon^*(\mathbf{x}^{(t)}) - \mathbf{x}^{(t)}\right) + \sqrt{2\gamma}\boldsymbol{\varepsilon}^{(t)}.\end{aligned} \tag{23}$$

This simplified scheme departs from the canonical PnP-ULA scheme studied in [29] by only omitting an additional term $\frac{\gamma}{\lambda}\left[\Pi_{\mathbb{S}}(\mathbf{x}^{(t)}) - \mathbf{x}^{(t)}\right]$ where $\Pi_{\mathbb{S}}(\cdot)$ denotes the projection onto the convex and compact set $\mathbb{S}$. This term has been included into the PnP-ULA scheme for technical reasons to derive convergence results. Its impact will be empirically shown to be marginal in practice (see Section IV-F). Conversely, using (9), ULA recursions applied to the RED posterior (12) writes

$$\begin{aligned}\text{(RED-ULA)}: \ \mathbf{x}^{(t+1)} &= \mathbf{x}^{(t)} - \gamma \nabla f(\mathbf{x}^{(t)}, \mathbf{y}) \\ &\quad + \gamma\beta\left(\mathsf{D}_\nu(\mathbf{x}^{(t)}) - \mathbf{x}^{(t)}\right) + \sqrt{2\gamma}\boldsymbol{\varepsilon}^{(t)}.\end{aligned} \tag{24}$$

It clearly appears that the RED-ULA scheme defined by the previous recursion (24) coincides with PnP-ULA (23) when $\beta = \frac{1}{\epsilon}$ and the denoiser embedded into RED is chosen as the MMSE denoiser, i.e., $\mathsf{D}_\nu(\cdot) = \mathsf{D}_\epsilon^*(\cdot)$. Since in practice the use of an MMSE denoiser is infeasible, PnP-ULA is implemented with an off-the-shelf denoiser. Thus, RED-ULA is no more than the practical implementation of PnP-ULA.

Moreover, an important corollary deals with the respective prior distributions defining the posteriors targeted by the three considered algorithms, namely PnP-ULA, RED-ULA and RED-SGS[3]. Indeed, on one hand, PnP-ULA and RED-ULA target the same posterior distribution (21) (provided the use of an MMSE denoiser). On another hand, PnP-ULA and RED-SGS also target the same posterior distribution (12). It yields that the regularized prior implicitly induced by AXDA coincides with the RED prior based on an MMSE denoiser, i.e., $p_\epsilon(\mathbf{x}) = p_{\text{red}}^*(\mathbf{x})$ and, in particular,

$$-\log p_\epsilon(\mathbf{x}) = \frac{1}{2\epsilon}\mathbf{x}^\top[\mathbf{x} - \mathsf{D}_\epsilon^*(\mathbf{x})]. \tag{25}$$

Note that differentiating this identity obviously leads again to the celebrated Tweedie's identity.

## IV. Experiments

### *A. Experimental setup*

Experiments have been conducted based on two popular image data sets. namely the Flickr Faces High Quality (FFHQ) $256 \times 256$ data set [48] and the ImageNet $512 \times 512$ data set [49]. Note that using the ImageNet $512 \times 512$ data set aims at assessing the scability of the proposed method when facing large-scale imaging problems. All images have been normalized to the range $[0, 1]$. The performance of the proposed RED-LwSGS algorithm is assessed w.r.t. to three inversion tasks

- *deblurring*: the operator $\mathbf{A}$ is assumed to be a $n \times n$ circulant convolution matrix associated with a spatially invariant blurring kernel. It is chosen as a Gaussian kernel of size $25 \times 25$ with standard deviation $1.6$.
- *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 accross the three color channels.
- *super-resolution*: the operator $\mathbf{A}$ is decomposed as $\mathbf{A} = \mathbf{SB}$ where the $n \times n$ matrix $\mathbf{B}$ stands for a spatially invariant Gaussian blur of size $7 \times 7$ 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

[3]The algorithmic scheme denoted RED-SGS is defined as a canonical Gibbs sampler with targets the augmented distribution (13). It can be interpreted as the RED-LwSGS for which the sampling according to (16) would be exact.

each dimension (i.e., $m = nd^2$). It is worth noting that for this task, the AXDA model needs to be adapted to account for the specific structure of the matrix $\mathbf{A}$ (more details are given in Appendix B).

For all tasks, the degraded images have been corrupted by an additive Gaussian noise to reach a signal-to-noise ratio of $\mathrm{SNR}$ =30dB for the FFHQ $256 \times 256$ data set and $\mathrm{SNR}$ =5dB for the ImageNet $512 \times 512$ data set, consisting in a severely less favorable experimental scenario.

### B. Compared methods

The proposed method has been compared to several state-of-the-art inversion methods:

- RED-ADMM [15]: ADMM with RED,
- RED-HQS: HQS algorithm with RED,
- PnP-ADMM [5]: ADMM with a PnP regularization,
- PnP-ULA [29]: ULA with a PnP regularization,
- TV-SP [31]: SGS with a TV regularization,
- TV-MYULA [50]: Moreau-Yosida ULA with a TV regularization,
- DiffPIR [51]: HQS with a diffusion-based PnP regularization,
- DPS [52]: diffusion posterior sampling,
- DPIR [11]: HQS algorithm with a PnP regularization.

All the RED- and PnP-based algorithms as well as DPIR have been implemented using the pre-trained deep network DRUNet [11] as the denoiser $\mathsf{D}_\nu(\cdot)$. It has been taken directly from the corresponding repository and has been used without further fine-tuning for the considered inversion tasks. It is worth noting that this model was initially trained on different data sets from the FFHQ and ImageNet data sets used in the experiments. Conversely, DPS and DiffPIR have been implemented using the pre-trained UNet diffusion-based models that were specifically trained on the FFHQ and ImageNet data sets in [53] and [54], respectively. Thus marginal performance gains may be expected for the last two methods since the training and data sets coincide. Yet, in all experiments, the test images have never been seen by the model while training, which prevents any bias due to potentially overfitted pre-trained models. See Appendix C-A for complementary information.

Besides, optimization-based methods such as the ADMM and HQS-based algorithms provide only point estimates of the restored images. Conversely, sampling-based methods such as PnP-ULA, TV-SP, TV-MYULA and the proposed Lw-SGS algorithm are Monte Carlo methods that are able to enrich point estimates with credibility intervals. For these sampling methods, the results reported below correspond to the MMSE estimates approximated by averaging the generated samples after the burn-in period. Complementary information regarding the algorithm implementations are reported in Appendices C-B and C-C.

### C. Figures-of-merit

Beyond performing visual inspection, the methods are compared with respect to several quantitative figures-of-merit. Peak signal noise ratio (PSNR) (dB) and structural similarity index (SSIM) are considered as image quality metrics (the higher the score, the better the reconstruction). They are complemented with the learned perceptual image patch similarity (LPIPS) and the Fréchet inception distance (FID) for which the lower the score, the better the reconstructed image. Moreover, to assess the effectiveness of the studied sampling algorithms, they are also compared in term of integrated autocorrelation time (IAT), which is an indicator for good or bad mixing (the lower, the better) [55]. Finally, all methods are compared in terms of computational times when the algorithms are implemented on a server equipped with 48 CPU cores Intel 2.8Ghz, 384Go RAM, Nvidia A100 GPU.

### D. Experimental results

The results obtained by the compared algorithms when performing the three considered tasks are reported in Tables I and II for the two datasets FFHQ and ImageNet, respectively. These results show that the proposed RED-LwSGS method achieves very competitive performance for all three tasks. In the case of inpainting, the forward operator masking $80\%$ pixels is non-invertible and the problem is expected to require further prior regularization than the two other tasks. For this task, 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 on the same model-based regularization. For super-resolution, which is a more challenging problem than deblurring and inpainting, RED-LwSGS demonstrates similar performance to DiffPIR, RED-ADMM, RED-HQS and DPIR on the FFHQ $256 \times 256$ data set, and comparable performance to PnP-ULA on the ImageNet $512 \times 512$ data set.

Figures 1 and 2 visually assess the performance by depicting the results obtained on test images drawn from the FFHQ and ImageNet datasets. For all the considered tasks, the proposed RED-LwSGS method produces high-quality, realistic images that closely match ground-truth details. As mentioned above, RED-LwSGS as any other sampling method generates samples asymptotically distributed according to the posterior distribution. These samples can be used to quantify estimation uncertainty. The two rightmost panels in Fig. 1 and 2 illustrate this advantage by depicting the estimated pixelwise standard deviations obtained by the proposed algorithm and PnP-ULA. It is worth noting that this added value cannot be provided by optimization-based methods such as RED-ADMM, RED-HQS, PnP-ADMM, DiffPIR, DPS and DPIR which only offer point estimates. Noticeably, pixels located in homogeneous regions are characterized by lower uncertainty, while pixels in textured regions, edges, or complex structures appear to be estimated with more difficulty.

Tables I and II also reported the computational times required by the compared algorithms when tackling each restoration task. Among the class of sampling methods, RED-LwSGS stands out for its smallest computational times. Moreover, as for the other sampling-based methods, RED-LwSGS remains within a factor of 50 compared to the optimization-based methods, namely RED-ADMM, RED-HQS, PnP-ADMM, DiffPIR and DPIR. Sampling methods are known to be generally more

TABLE I
FFHQ $256 \times 256$ DATA SET (SNR $= 30$dB): AVERAGE PERFORMANCE OVER A TEST SET OF 100 IMAGES AND CORRESPONDING STANDARD DEVIATIONS. **BOLD**: BEST SCORE, UNDERLINE: SECOND SCORE.

| | | Obs. | RED-LwSGS | PnP-ULA | TV-MYULA | TV-SP | RED-ADMM | RED-HQS | PnP-ADMM | DiffPIR | DPS | DPIR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Deblurring** | **PSNR**(dB) ↑ | 29.86 | 40.13±1.931 | 38.06±1.986 | 38.03±2.028 | 39.89±1.320 | **41.28**±2.459 | 40.80±2.560 | 34.03±7.601 | 37.15±2.066 | 35.35±2.688 | 38.42±2.536 |
| | **SSIM** ↑ | 0.899 | 0.982±0.003 | 0.975±0.006 | 0.981±0.006 | 0.970±0.004 | **0.985**±0.004 | 0.984±0.005 | 0.869±0.154 | 0.957±0.010 | 0.949±0.018 | 0.972±0.012 |
| | **LPIPS** ($\times 10^{-2}$) ↓ | 4.87 | 0.260±0.174 | 0.490±0.362 | 0.530±0.382 | 0.210±0.111 | **0.180**±0.150 | 0.220±0.195 | 3.320±8.037 | 0.741±2.491 | 6.405±1.825 | 0.811±1.402 |
| | **FID** ↓ | 95.89 | 8.181±7.316 | 13.78±10.03 | 17.39±10.67 | 5.900±4.422 | **4.348**±6.106 | 7.148±7.975 | 18.54±10.81 | 6.055±3.127 | 11.76±5.162 | 3.392±5.014 |
| | **IAT** ↓ | - | 75.17±12.42 | 75.21±12.43 | 75.22±12.42 | **75.14**±12.42 | - | - | - | - | - | - |
| | **Time**(s) ↓ | - | 44±5 | 52±7 | 64±5 | 40±5 | 2±0 | 2±0 | 2±0 | 1±0 | 30±2 | 4±1 |
| **Inpainting** | **PSNR**(dB) ↑ | 7.21 | 30.73±2.932 | 31.46±2.650 | 27.71±1.881 | 27.27±1.781 | **31.63**±2.672 | 31.36±2.293 | 31.32±3.142 | 31.26±2.25 | 31.26±2.26 | 31.50±2.59 |
| | **SSIM** ↑ | 0.068 | 0.908±0.028 | 0.906±0.026 | 0.830±0.040 | 0.815±0.041 | 0.911±0.024 | 0.901±0.025 | **0.915**±0.042 | 0.890±0.025 | 0.890±0.025 | 0.904±0.027 |
| | **LPIPS** ($\times 10^{-2}$) ↓ | 58.31 | 2.358±1.722 | 2.049±1.317 | 5.673±2.628 | 6.128±2.745 | **1.954**±1.329 | 2.119±1.428 | **1.954**±1.538 | 2.108±0.539 | 2.312±2.258 | 2.145±2.737 |
| | **FID** ↓ | 554.33 | 78.09±30.36 | 90.33±39.73 | 181.33±57.57 | 183.15±59.26 | 79.65±31.54 | 85.40±32.3 | 90.33±39.73 | 52.79±23.03 | 51.38±20.46 | 119.55±47.74 |
| | **IAT** ↓ | - | 75.24±12.48 | 75.51±12.55 | 76.97±13.51 | 113.86±12.17 | - | - | - | - | - | - |
| | **Time**(s) ↓ | - | 74±1 | 79±1 | 150±7 | 71±1 | 3±0 | 2±0 | 2±0 | 2±0 | 43±1 | 7±2 |
| **Super-res.** | **PSNR**(dB) ↑ | - | 30.43±2.161 | 29.01±2.013 | 28.99±2.017 | 28.94±2.019 | 30.49±2.222 | 30.54±2.206 | 30.13±2.184 | **30.99**±2.212 | 28.94±1.798 | 30.21±2.279 |
| | **SSIM** ↑ | - | 0.872±0.036 | 0.847±0.037 | 0.847±0.037 | 0.846±0.037 | 0.875±0.036 | **0.876**±0.036 | 0.867±0.037 | 0.868±0.034 | 0.833±0.054 | 0.869±0.038 |
| | **LPIPS** ($\times 10^{-2}$) ↓ | - | 3.519±2.139 | 5.025±2.359 | 4.925±2.435 | 5.125±2.435 | 3.418±2.038 | 3.418±2.037 | 3.510±2.140 | **1.112**±0.821 | 5.842±4.591 | 2.720±3.910 |
| | **FID** ↓ | - | 100.09±34.62 | 202.15±48.02 | 202.29±48.30 | 199.16±47.63 | 95.48±33.51 | 95.10±33.33 | 110.04±36.31 | 65.16±30.15 | 18.32±42.57 | 105.47±36.21 |
| | **IAT** ↓ | - | **75.75**±12.62 | 75.88±12.64 | 75.85±12.64 | 75.86±12.63 | - | - | - | - | - | - |
| | **Time**(s) ↓ | - | 115±25 | 128±40 | 133±26 | 112±23 | 3±1 | 3±1 | 3±1 | 2±0 | 50±4 | 3±1 |

TABLE II
IMAGENET $512 \times 512$ DATA SET (SNR $= 5$dB): AVERAGE PERFORMANCE OVER A TEST SET OF 100 IMAGES AND CORRESPONDING STANDARD DEVIATIONS. **BOLD**: BEST SCORE, UNDERLINE: SECOND SCORE.

| | | Obs. | RED-LwSGS | PnP-ULA | TV-MYULA | TV-SP | RED-ADMM | RED-HQS | PnP-ADMM | DiffPIR | DPIR | DPS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Deblurring** | **PSNR**(dB) ↑ | 27.32 | 30.69±2.752 | 29.04±3.714 | 30.14±3.602 | 29.61±3.947 | 30.71±3.753 | 28.90±3.593 | 30.86±4.917 | 29.99±9.100 | **31.35**±3.831 | 30.16±3.627 |
| | **SSIM** ↑ | 0.709 | 0.834±0.085 | 0.754±0.168 | 0.842±0.059 | 0.753±0.109 | 0.804±0.099 | 0.714±0.143 | **0.853**±0.089 | 0.771±0.266 | 0.839±0.076 | 0.826±0.062 |
| | **LPIPS** ($\times 10^{-2}$) ↓ | 27.96 | **19.23**±6.658 | 23.11±11.60 | 19.63±5.324 | 23.46±8.509 | 20.81±8.103 | 26.60±10.59 | 20.00±6.670 | 21.16±22.914 | 20.08±7.147 | 20.65±7.292 |
| | **FID** ↓ | 66.51 | 30.10±3.318 | 40.51±6.490 | 27.98±7.161 | 29.06±6.036 | **21.34**±3.611 | 31.20±5.088 | 30.76±6.273 | 46.55±6.774 | 21.77±2.959 | 31.82±4.823 |
| | **IAT** ↓ | - | 272.59±13.43 | 275.93±32.85 | 275.74±12.04 | **271**.75±12.32 | - | - | - | - | - | - |
| | **Time**(s) ↓ | - | 254±69 | 258±74 | 256±132 | 127±57 | 4±0 | 4±0 | 4±0 | 26±0 | 5±1 | 72±19 |
| **Inpainting** | **PSNR**(dB) ↑ | 6.81 | 27.63±5.389 | 28.05±4.772 | 23.15±3.606 | 25.84±4.009 | **28.63**±5.085 | 27.85±4.719 | 26.93±6.304 | 28.26±4.466 | 28.44±5.020 | 27.83±4.019 |
| | **SSIM** ↑ | 0.063 | 0.805±0.110 | 0.812±0.098 | 0.630+0.143 | 0.726±0.105 | **0.838**±0.103 | 0.815±0.100 | 0.834±0.116 | 0.794±0.110 | 0.825±0.108 | 0.802±0.202 |
| | **LPIPS** ($\times 10^{-2}$) ↓ | 77.37 | 24.38±9.151 | 26.17±8.323 | 45.19±7.309 | 35.58±6.022 | 22.86±8.908 | 25.41±7.824 | **19.66**±9.759 | 23.44±7.290 | 27.26±9.654 | 25.01±9.891 |
| | **FID** ↓ | 440.79 | 42.73±38.74 | 67.36±53.50 | 168.38±107.26 | 91.52±64.72 | 55.42±45.64 | 65.27±50.40 | **34.17**±46.09 | 41.05±31.98 | 70.95±58.91 | 45.98±54.93 |
| | **IAT** ↓ | - | **254.54**±314.46 | 265.04±313.45 | 311.61±329.12 | 290.86±317.32 | - | - | - | - | - | - |
| | **Time**(s) ↓ | - | 200±5 | 199±1 | 585±248 | 82±3 | 20±0 | 20±0 | 20±0 | 27±0 | 20±0 | 75±11 |
| **Super-res.** | **PSNR**(dB) ↑ | - | 26.10±4.089 | **26.43**±4.575 | 25.82±4.071 | 26.05±4.262 | 25.63±4.117 | 25.62±4.112 | 23.52±5.218 | 25.14±4.323 | 25.24±4.419 | 25.35±4.028 |
| | **SSIM** ↑ | - | 0.706±0.123 | **0.715**±0.137 | 0.667±0.122 | 0.694±0.132 | 0.645±0.126 | 0.644±0.126 | 0.658±0.158 | 0.608±0.164 | 0.633±0.135 | 0.667±0.193 |
| | **LPIPS** ($\times 10^{-2}$) ↓ | - | 35.71±6.865 | **32.93**±8.840 | 39.80±7.153 | 35.79±7.258 | 37.93±6.959 | 38.08±6.942 | 40.81±10.45 | 39.34±9.944 | 40.16±7.702 | 40.02±7.983 |
| | **FID** ↓ | - | 103.10±70.80 | 120.97±102.60 | **92.02**±63.86 | 93.54±67.67 | 109.80±73.90 | 110.21±74.16 | 196.86±132.19 | 119.67±93.82 | 127.04±86.78 | 109.23±80.37 |
| | **IAT** ↓ | - | 277.89±312.83 | **276.07**±315.65 | 277.29±313.02 | 280.70±315.22 | - | - | - | - | - | - |
| | **Time**(s) ↓ | - | 301±64 | 352±54 | 334±135 | 187±60 | 7±0 | 7±0 | 7±0 | 8±2 | 7±0 | 60±2 |

computationally intensive than optimization-based methods. This is due not only to the cost per iteration but also to the larger number of iterations required to obtain reliable estimates. However, contrary to optimization methods, which typically only retain the final iterate, every sample generated by a sampling algorithm beyond its burn-in phase contributes valuable information. Specifically, the set of samples provides a full characterization of the posterior distribution, enabling both point estimates and confidence measures. This additional computational burden is the cost of obtaining uncertainty quantifications, which are crucial for certain applications. In particular, one computationally demanding step in the proposed LwSGS algorithm is the sampling of the splitting variable according to (20), which requires invoking the denoiser $\mathsf{D}(\cdot)$. In this work, the denoiser used is DRUNet, a deep neural network. To reduce this computational burden, a simpler, less computationally intensive denoiser could be employed, which would make the algorithm more efficient without sacrificing too much in terms of performance. Yet, from these experimental results, the price to pay for offering an uncertainty quantification on top of point estimation seems reasonable.

Finally, the convergence properties of the compared sampling-based algorithms have been assessed by monitoring the autocorrelation function (ACF) of the median components of the chains generated by those algorithms. By denoting $\mathbf{x}^{(t)} = \left[x_1^{(t)}, \ldots, x_n^{(t)}\right]^\top$, the median component has been defined as the produced pixelwise Monte Carlo chain $\left\{x_i^{(t)}\right\}_{t=N_{\mathrm{bi}}+1}^{N_{\mathrm{MC}}}$ with the median variance. Faster decreasing ACF means that the samples are less correlated and generally implies faster convergence of the Markov chain. Fig. 3 depicts these ACFs for the three restoration tasks conducted on one image from the FFHQ data set. For the deblurring task, it is not clear which of the compared methods is the more efficient, i.e., with the fastest ACF decay. Conversely, for the inpainting and super-resolution tasks, the ACF of RED-LwSGS decreases faster than the ACFs obtained with the two other Monte Carlo algorithms. This finding is confirmed by the IAT measures reported in Tables I and II.

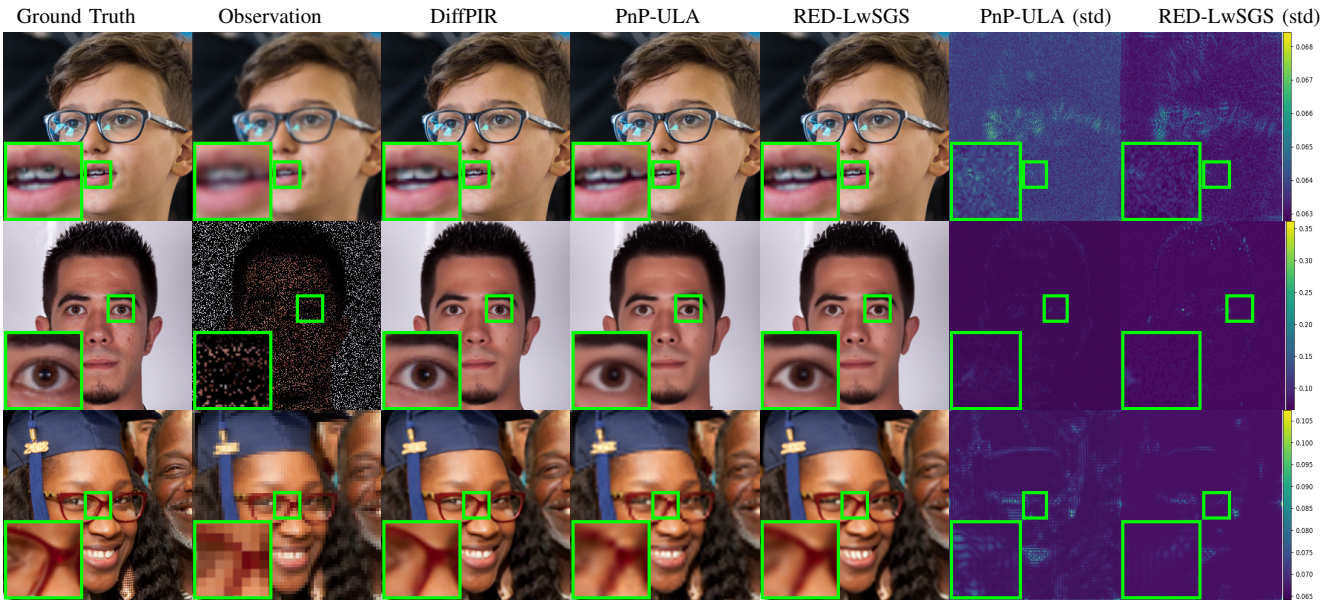


Fig. 1. FFHQ $256 \times 256$ data set ($\mathrm{SNR} = 30\mathrm{dB}$): images recovered by the compared methods for deblurring (top), inpainting (middle) and super-resolution (bottom).

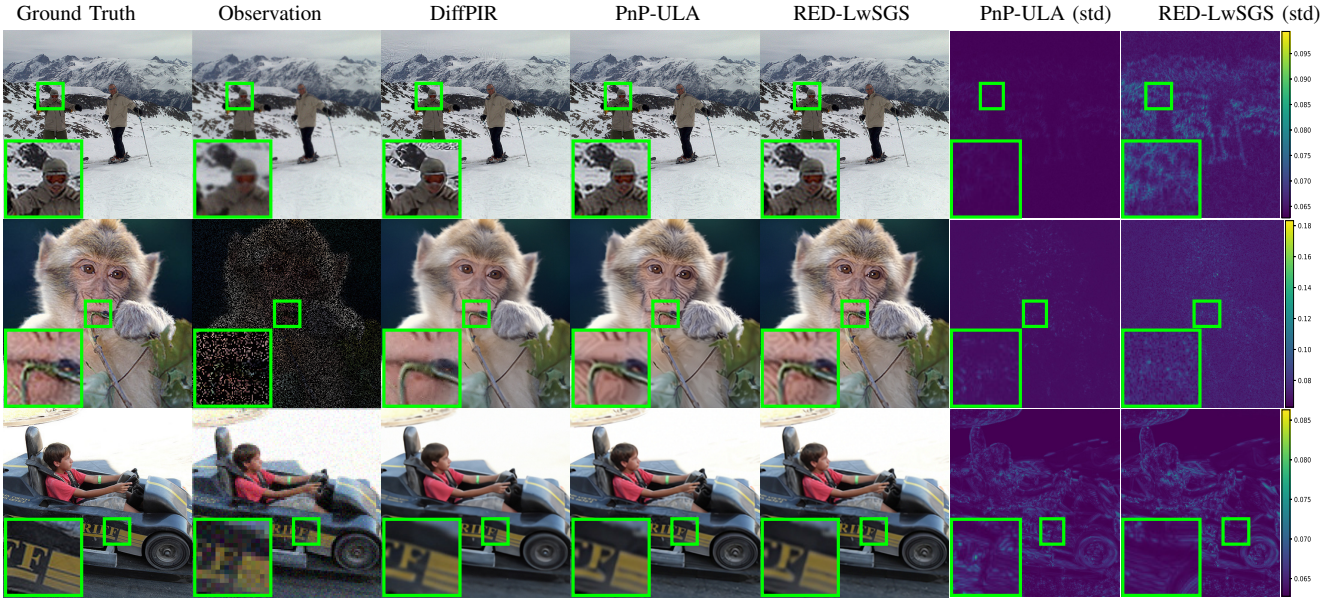


Fig. 2. ImageNet $512 \times 512$ data set ($\mathrm{SNR} = 5\mathrm{dB}$): images recovered by the compared methods for deblurring (top), inpainting (middle) and super-resolution (bottom).

### *E. Does DRUNet meet the RED conditions?*

The explicit formula (9) of the RED gradient requires the RED conditions (C1)–(C4) to be satisfied, i.e., the denoiser $\mathsf{D}_\nu(\cdot)$ should be differentiable, locally homogeneous, with symmetric Jacobian and strongly passive. It is legitimate to assess whether these conditions are verified for the deep denoiser considered in the experiments, namely DRUNet. First, the canonical implementation of DRUNet makes $\mathsf{D}_\nu(\cdot)$ not continuously differentiable w.r.t. the input due to the use of ReLU activation functions. However, to ensure (C2), the overall architecture can be made continuously differentiable by replacing them with SoftPlus activation functions, which are $\mathcal{C}^\infty$, as suggested in [14]. Besides, the three other RED conditions are empirically assessed through numerical experiments. The local homogeneity condition (C1) of the denoiser is evaluated by computing the two following normalized mean square errors [16]

$$\mathrm{NMSE}_{\mathrm{LH},1} = \mathbb{E}\left[\frac{||\mathsf{D}_\nu\big((1+\epsilon)\mathbf{x}\big) - (1+\epsilon)\mathsf{D}_\nu(\mathbf{x})||^2}{||(1+\epsilon)\mathsf{D}_\nu(\mathbf{x})||^2}\right]$$

and

$$\mathrm{NMSE}_{\mathrm{LH},2} = \mathbb{E}\left[\frac{||\nabla\mathsf{D}_\nu(\mathbf{x})\mathbf{x} - \mathsf{D}_\nu(\mathbf{x})||^2}{||\mathsf{D}_\nu(\mathbf{x})||^2}\right]$$

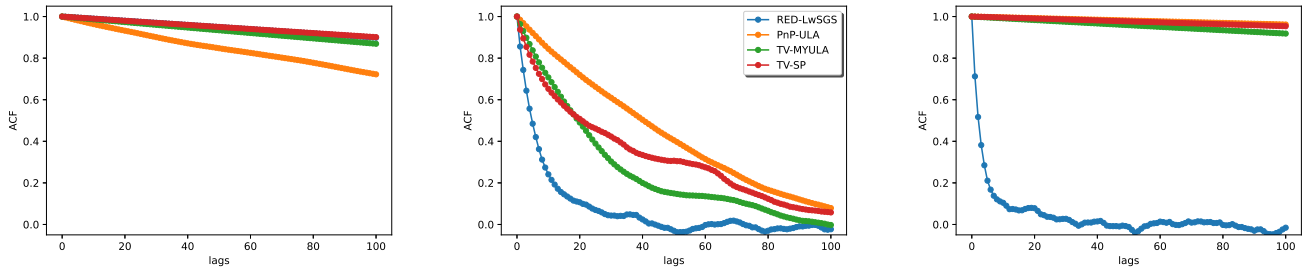


Fig. 3. FFHQ $256 \times 256$ data set: (absolute) autocorrelation function (ACF) of the samples generated by the compared algorithms for deblurring (left), inpainting (middle) and super-resolution (right).

TABLE III
NUMERICAL EXPERIMENTS TO ASSESS THE RED CONDITIONS.

| | | FFHQ | ImageNet |
|---|---|---|---|
| **Homogeneity** | $\text{NMSE}_{\text{LH},1}$ | $1.78 \times 10^{-8}$ | $5.34 \times 10^{-8}$ |
| | $\text{NMSE}_{\text{LH},2}$ | $1.75 \times 10^{-2}$ | $5.28 \times 10^{-2}$ |
| **Jacobian symmetry** | $\text{NMSE}_{\text{JS}}$ | $5.81 \times 10^{-2}$ | $8.39 \times 10^{-2}$ |
| **Strong passivity** | MSR | 1.44 | 1.63 |

TABLE IV
NUMERICAL COMPARAISONS OF RED-ULA AND PNP-ULA.

| | | PSNR | SSIM | LPIPS | IAT | % Proj. |
|---|---|---|---|---|---|---|
| **RED-ULA** | | 38.03 | 0.9809 | 0.0053 | 75.22 | - |
| **PnP-ULA** | $\mathbb{S} = [0, 1]$ | 38.46 | 0.9813 | 0.0048 | 75.21 | 96.71 |
| | $\mathbb{S} = [-1, 2]$ | 38.03 | 0.9809 | 0.0053 | 75.22 | 0 |

which are motivated by the definition (6) and the property (7), respectively. These metrics should be close to zero to ensure local homogeneity. Further, following (C3), the denoiser should have a symmetric Jacobian. This characteristic is empirically evaluated by computing

$$\text{NMSE}_{\text{JS}} = \mathbb{E}\left[\frac{||\nabla \mathsf{D}_\nu(\mathbf{x}) - \nabla \mathsf{D}_\nu^\top(\mathbf{x})||^2}{||\nabla \mathsf{D}_\nu(\mathbf{x})||^2}\right]$$

which should be close to zero for a symmetric Jacobian [16]. Finally, to assess the strong passivity condition (C4), one considers the mean spectral radius

$$\text{MSR} = \mathbb{E}\left[\eta\left(\nabla \mathsf{D}_\nu(\mathbf{x})\right)\right]$$

which can be computed by the power iteration and should be smaller than 1. When computing these metrics, the $(i, j)$-element of the gradient of $\mathsf{D}_\nu(\cdot)$ has been approximated as

$$[\nabla \mathsf{D}_\nu(\mathbf{x})]_{i,j} \approx \frac{[\mathsf{D}_\nu(\mathbf{x} + \epsilon \mathbf{e}_j)]_i - [\mathsf{D}_\nu(\mathbf{x} - \epsilon \mathbf{e}_j)]_i}{2\epsilon}$$

where $\mathbf{e}_j$ denotes the $j$th canonical basis vector, i.e., the $j$th column of $\mathbf{I}_n$, and $\epsilon > 0$ is small enough. These metrics are reported in Table III when the four considered scores have been computed over 100 patches of size $32 \times 32$ extracted from images of the two data sets, namely FFHQ $256 \times 256$ and ImageNet $512 \times 512$. They show that the DRUNet denoiser seems to satisfy the local homogeneity and symmetric Jacobian conditions. However, DRUNet has a spectral radius greater than 1. Yet, to ensure a strongly passive deep neural network-based denoiser, various strategies could be been envisioned, such as spectral normalization [13].

### *F. RED-ULA vs. PnP-ULA*

Section III-D has drawn some connections between PnP-ULA and RED-ULA. In particular, it has shown that in practice the two algorithms basically reduce to the same algorithmic scheme, except that PnP-ULA embeds an additional projection step onto an arbitrary pre-defined set $\mathbb{S}$. This projection aims at ensuring that the drift satisfies an asymptotic growth condition. To experimentally validate this equivalence, Table IV reports the performance of PnP-ULA and RED-ULA when tackling the deblurring task. These performances have been computed over 100 images of the FFHQ dataset. Two configurations for the set $\mathbb{S}$ are considered, $\mathbb{S} = [0, 1]$ and $\mathbb{S} = [-1, 2]$. The rates of activation of the constraint, i.e., the proportion of samples generated by PnP-ULA that do not satisfy the drift condition and should be projected onto the set $\mathbb{S}$, are also reported in terms of percentage. These results show that, when $\mathbb{S} = [-1, 2]$, the projection embedded in PnP-ULA is never activated and the performance is the same as the one obtained by RED-ULA, which confirms that the two algorithms are identical. When $\mathbb{S} = [0, 1]$, this projection is applied to almost all the samples generated by PnP-ULA, without significantly affecting the performance.

## V. CONCLUSION

This work built a Bayesian counterpart of the regularization-by-denoising (RED) engine, offering a data-driven framework to define prior distributions in Bayesian inversion tasks. It defined a new probability distribution from the RED potential, which was subsequently embedded into a Bayesian model as a prior distribution. Since the resulting RED posterior distribution was not standard, a dedicated Monte Carlo algorithm was designed. By leveraging an asymptotically exact data augmentation (AXDA), this algorithm was a particular instance of the split Gibbs sampler which had the great advantage of decoupling the data-fitting term and the RED potential. One stage of SGS was performed following a Langevin Monte Carlo step, which leads to the so-called Langevin-within-split Gibbs sampling. A thorough theoretical analysis was conducted

to assess the convergence guarantees of the algorithm. Some tight connections were drawn between AXDA and RED to show that the implicit prior resulting from an AXDA scheme coincides with the RED prior defined by a MMSE denoiser. Extensive numerical experiments showed that the proposed approach competes favorably with state-of-the-art variational and Monte Carlo methods when tackling conventional inversion tasks, namely deblurring, inpainting and super-resolution. The proposed approach was shown to provide a comprehensive characterization of the solutions which could be accompanied by an uncertainty quantification. By bridging the gap between the RED paradigm and Bayesian inference, this work opened new avenues for incorporating data-driven regularizations into Monte Carlo algorithms.

Future research could explore the possibility of relaxing some of the RED conditions while still ensuring the existence of the resulting posterior distribution. A promising avenue could be to redefine the prior distribution from the perspective of "gradient step denoisers" as discussed in [14], [20], [21]. This approach might lead to less restrictive and more flexible prior models. Furthermore, another promising direction for future research is the extension of the proposed framework and its theoretical analysis to nonlinear inverse problems such as those recently considered in [33] and [34]. This would significantly broaden its applicability to more challenging reconstruction tasks encountered in various application fields such as medical imaging and astronomy.

## APPENDIX A
## PROOF OF PROPOSITION 1

Under the RED conditions, the RED prior (10) 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

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

## APPENDIX B
## AUGMENTED DISTRIBUTION AND RED-LWSGS FOR SUPER-RESOLUTION

This appendix details the AXDA model and the corresponding sampling algorithm when tackling the super-resolution task. In this case, the operator $\mathbf{A}$ can be written as $\mathbf{A} = \mathbf{SB}$ where $\mathbf{B} \in \mathbb{R}^{n\times n}$ is a circulant matrix standing for a spatially invariant blur and $\mathbf{S} \in \mathbb{R}^{m\times n}$ stands for a regular downsampling operator. When directly adopting the splitting trick proposed in Section III-A, sampling according to the conditional distribution (15) remains difficult because the precision matrix is neither diagonal (as for the inpainting task) nor diagonalizable in the Fourier domain (as for the deblurring task). To overcome this difficulty, one suitable AXDA consists in introducing two splitting variables, which allows the operators $\mathbf{B}$ and $\mathbf{S}$ to be decoupled. This leads to the augmented posterior distribution

$$\pi_{\rho_1,\rho_2}(\mathbf{x},\mathbf{z}_1,\mathbf{z}_2) \propto \exp\left[-\frac{1}{2\sigma^2}||\mathbf{S}\mathbf{z}_1 - \mathbf{y}||_2^2 - \beta g_{\text{red}}(\mathbf{z}_2) - \frac{1}{2\rho_1^2}||\mathbf{B}\mathbf{x} - \mathbf{z}_1||_2^2 - \frac{1}{2\rho_2^2}||\mathbf{x} - \mathbf{z}_2||^2\right].$$

The associated SGS alternatively samples according to the three conditional distributions

$$p(\mathbf{z}_1|\mathbf{x},\mathbf{y}) \propto \exp\left[-\tfrac{1}{2\sigma^2}\|\mathbf{S}\mathbf{z}_1 - \mathbf{y}\|_2^2 - \tfrac{1}{2\rho_1^2}||\mathbf{B}\mathbf{x} - \mathbf{z}_1||_2^2\right] \tag{26}$$

$$p(\mathbf{x}|\mathbf{z}_1,\mathbf{z}_2) \propto \exp\left[-\tfrac{1}{2\rho_1^2}\|\mathbf{B}\mathbf{x} - \mathbf{z}_1\|_2^2 - \tfrac{1}{2\rho_2^2}\|\mathbf{x} - \mathbf{z}_2\|^2\right] \tag{27}$$

$$p(\mathbf{z}_2|\mathbf{x}) \propto \exp\left[-\tfrac{1}{2}\mathbf{z}_2^\top(\mathbf{z}_2 - \mathsf{D}_\nu(\mathbf{z}_2)) - \tfrac{1}{2\rho_2^2}\|\mathbf{x} - \mathbf{z}_2\|^2\right]$$

It appears that (26) and (27) define the conditional posteriors associated with the inpainting and deblurring tasks, respectively.

## APPENDIX C
## EXPERIMENTAL DETAILS

### A. Pre-trained models

This section provides some detailes regarding the implementation of the pre-trained models employed in the experiments. These models play a crucial role in the performance of both the sampling-based and optimization-based methods used for solving inverse problems.

Regarding the PnP- and RED-based methods, all experiments have been performed with DRUNet as the pre-trained denoiser [11]. This denoiser $\mathsf{D}_\nu(\cdot)$ has the ability to handle different noise levels with a single model thanks to the parameter $\nu$ which controls the strength of the denoising. During the experiments, this parameter has been adjusted following the strategy recommended in [11]. More precisely, a decreasing sequence $\nu^{(1)} > \nu^{(2)} > \cdots > \nu^{(N_{\text{bi}})}$ has been uniformly sampled according to a logarithmic scale between a large value fixed as $\nu^{(1)} = 49$ and a value $\nu^{(K)}$ adjusted w.r.t. to the image noise level. For the optimization-based methods, this sequence is used as a parameter of the denoiser along the $K$ iterations or the algorithms. For the sampling-based methods, to ensure the stationary of the kernel, this parameter is set to a fixed value beyond the burn-in period, i.e., $\forall t > K$, $\nu^{(t)} = \nu^{(K)}$ where $K = N_{\text{bi}}$ stands for number of the burn-in iterations. In other words, the ULA and LwSGS sampling methods are ensured to sample from a well-defined stationary distribution. It is worth noting that the posterior distribution ideally targeted by the ULA and LwSGS sampling methods should have been written as $\pi(\mathbf{x}) = p(\mathbf{x} \mid \mathbf{y}; \beta, \nu^{(K)})$ since it depends on the wo model parameters $\beta$ and $\nu^{(K)}$. They have been omitted throughout this article to lighten the notations.

Regarding DiffPIR and DPS, they embed the UNet diffusion-based pre-trained models of [54] and [53] to conduct experiments on the ImageNet and FFHQ datasets, respectively.

The pre-trained models and their training sets are summarized in Table V for each experiment conducted in this work.

TABLE V
TRAINING AND TESTING DATA SETS OF THE PRE-TRAINED MODELS.

| Test sets | Methods | Model | Training sets |
|---|---|---|---|
| FFHQ | RED-{LwSGS, ADMM, HQS}, PnP-{ULA, ADMM}, DPIR | DRUNet [11] | DIV2K Flick2K BSD |
| | DiffPIR, DPS | UNet [53] | FFHQ |
| ImageNet | RED-{LwSGS, ADMM, HQS}, PnP-{ULA, ADMM}, DPIR | DRUNet [11] | DIV2K Flick2K BSD |
| | DiffPIR, DPS | UNet [54] | ImageNet |

### *B. Implementation details regarding RED-LwSGS*

This appendix provides additional details regarding the implementation of the proposed LwSGS algorithm. The regularization parameter and the coupling parameters have been adjusted to reach the best performance. For experiments on the FFHQ data set, the regularization parameter $\beta$ is set to $8.0 \times 10^{-2}, 1.25 \times 10^{-1}$ and $1.0$ for the deblurring, inpainting and super-resolution tasks, respectively, while it has been fixed to $4.89 \times 10^{-3}$, $1.167 \times 10^{-1}$ and $4.966 \times 10^{-2}$ for the experiments conducted on the ImageNet data set. The other parameters are fixed as $(N_{\mathrm{MC}}, N_{\mathrm{bi}}, \rho^2, \gamma) = (5000, 2000, 6 \times 10^{-8}, \frac{0.99}{2\beta+1/\rho^2})$ for deblurring, $(N_{\mathrm{MC}}, N_{\mathrm{bi}}, \rho^2, \gamma) = (10000, 4500, 1.5, \frac{0.99}{2\beta+1/\rho^2})$ for inpainting and $(N_{\mathrm{MC}}, N_{\mathrm{bi}}, \rho_1^2, \rho_2^2, \gamma) = (12500, 3500, 2 \times 10^{-1}, 1, \frac{0.8}{2\beta+1/\rho_2^2})$ for super-resolution which requires a double splitting.

### *C. Implementation details regarding the compared methods*

This appendix provides additional details regarding the implementation of the compared methods. First, RED-ADMM, RED-HQS, PnP-ADMM, PnP-ULA and Diff-PIR are implemented using the same denoiser as the proposed method (see Appendix C-A). For the sampling-based methods, i.e., PnP-ULA, TV-MYULA, TV-SP, the total number of iterations have been set as for the proposed RED-LwSGS (see Appendix C-B). For optimization-based algorithms, i.e, RED-ADMM, RED-HQS, DPIR and PnP-ADMM, the total number of iterations is set as follows: $150$ for deblurring and $350$ for inpainting and super-resolution. Finally all model and algorithmic parameters have been adjusted to reach the best performance. More precisely, for the experiments conducted on the FFHQ data:

- PnP-ULA: the parameters $(N_{\mathrm{bi}}, \beta)$ have been set to $(2500, 7.3 \times 10^{-4})$ for deblurring, $(5000, 10^{-4})$ for inpainting and $(6000, 2.75 \times 10^{-4})$ for super-resolution.
- TV-MYULA: the parameters $(N_{\mathrm{bi}}, \beta)$ have been set to $(2500, 10^{-1})$ for deblurring, $(5000, 8 \times 10^{-8})$ for inpainting and $(8500, 1)$ for super-resolution.
- TV-SP: the parameters $(N_{\mathrm{bi}}, \rho, \beta)$ have been set to $(2000, 9 \times 10^{-4}, 1)$ for deblurring and $(4500, 1, 3.5 \times 10^{-3})$ for inpainting while for super-resolution, which requires a double splitting, the parameters $(N_{\mathrm{bi}}, \rho_1^2, \rho_2^2, \beta)$ have been set to $(3500, 10^{-5}, 10^{-5}, 3.76)$.
- RED-ADMM: the hyperparameters $(\alpha, \lambda, \beta)$ have been set to $(2, 2 \times 10^{-3}, 9 \times 10^{-4})$ for deblurring, $(2, 10^{-2}, 4 \times 10^{-2})$ for inpainting and $(2, 8 \times 10^{-3}, 10^{-6})$ for super-resolution.
- RED-HQS: similarly to RED-ADMM, the hyperparameters are $(\alpha, \lambda, \beta) = (2, 10^{-2}, 4 \times 10^{-3})$ for deblurring, $(\alpha, \lambda, \beta) = (2, 2 \times 10^{-2}, 1.8 \times 10^{-2})$ for inpainting and $(\alpha, \lambda, \beta) = (2, 8 \times 10^{-3}, 10^{-6})$ for super-resolution.
- PnP-ADMM: the parameter $\rho$ is set to $10^{-4}$ for deblurring, $10^{-3}$ for inpainting and $6.5 \times 10^{-2}$ for super-resolution.
- DiffPIR: the parameters $(\lambda, \zeta)$ have been set to $(2, 3 \times 10^{-1})$ for deblurring and $(1, 1)$ for inpainting and super-resolution.
- DPIR: the parameter $\lambda$ had been set to $1 \times 10^{-3}$ for deblurring and inpainting and $0.9$ for super-resolution.
- DPS: it has been implemented following the original paper by using a 1000-step DDPM sampler backbone and the parameter $\zeta$ is set to 1.

## 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 $\ell_p$ $(0 < p < 1)$ 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] J. Douglas and H. H. Rachford, "On the numerical solution of heat conduction problems in two and three space variables," *Trans. Am. Math. Soc.*, vol. 82, no. 2, pp. 421–439, 1956.

[8] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. Int. Conf. Computer Vision Pattern Recognition (CVPR)*, vol. 2, 2005, pp. 60–65.

[9] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, 2007.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] Y. Sun, J. Liu, and U. Kamilov, "Block coordinate regularization by denoising," in *Adv. in Neural Information Process. Systems (NeurIPS)*, vol. 32, 2019.

[18] J. Liu, S. Asif, B. Wohlberg, and U. Kamilov, "Recovery analysis for plug-and-play priors using the restricted eigenvalue condition," in *Adv. in Neural Information Process. Systems (NeurIPS)*, vol. 34, 2021, pp. 5921–5933.

[19] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2018.

[20] R. Cohen, M. Elad, and P. Milanfar, "Regularization by denoising via fixed-point projection (RED-PRO)," *SIAM J. Imag. Sci.*, vol. 14, no. 3, pp. 1374–1406, 2021.

[21] S. Hurault, A. Leclaire, and N. Papadakis, "Gradient step denoiser for convergent plug-and-play," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2022.

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

[23] 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.

[24] T. I. Liaudat, M. Mars, M. A. Price, M. Pereyra, M. M. Betcke, and J. D. McEwen, "Scalable Bayesian uncertainty quantification with data-driven priors for radio interferometric imaging," *RAS Techniques and Instruments*, vol. 3, no. 1, pp. 505–534, 2024.

[25] M. D. Fall, É. Barat, and T. Dautremer, "Hierarchical nonparametric Bayesian modeling for clinical 3D PET data reconstruction," 2024, preprint.

[26] L. Huang, S. Ruan, Y. Xing, and M. Feng, "A review of uncertainty quantification in medical image analysis: probabilistic and non-probabilistic methods," *Medical Image Analysis*, p. 103223, 2024.

[27] 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.

[28] 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.*, vol. 17, no. 2, pp. 820–860, 2024.

[29] 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.

[30] 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.

[31] 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.

[32] F. Coeurdoux, N. Dobigeon, and P. Chainais, "Plug-and-play split Gibbs sampler: embedding deep generative priors in Bayesian inference," *IEEE Trans. Image Processing*, vol. 33, pp. 3496–3507, May 2024.

[33] X. Xu and Y. Chi, "Provably robust score-based diffusion posterior sampling for plug-and-play image reconstruction," *arXiv preprint arXiv:2403.17042*, 2024.

[34] Z. Wu, Y. Sun, Y. Chen, B. Zhang, Y. Yue, and K. L. Bouman, "Principled probabilistic imaging using diffusion models as plug-and-play priors," *arXiv preprint arXiv:2405.18782*, 2024.

[35] M. Vono, D. Paulin, and A. Doucet, "Efficient MCMC sampling with dimension-free convergence rate using ADMM-type splitting," *J. Mach. Learning Research*, vol. 23, no. 1, pp. 1100–1168, 2022.

[36] B. Efron, "Tweedie's formula and selection bias," *J. Amer. Stat. Soc.*, vol. 106, no. 496, pp. 1602–1614, 2011.

[37] A. Durmus and E. Moulines, "High-dimensional Bayesian inference via the unadjusted Langevin algorithm," *Bernoulli*, vol. 25, no. 4A, pp. 2854–2882, 2019.

[38] 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.

[39] M. Vono, N. Dobigeon, and P. Chainais, "High-dimensional Gaussian sampling: a review and a unifying approach based on a stochastic proximal point algorithm," *SIAM Rev.*, vol. 64, no. 1, pp. 3–56, 2022.

[40] 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.

[41] 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.

[42] G. O. Roberts and O. Stramer, "Langevin diffusions and Metropolis-Hastings algorithms," *Methodology and computing in applied probability*, vol. 4, pp. 337–357, 2002.

[43] P.-A. Thouvenin, A. Repetti, and P. Chainais, "A distributed Gibbs sampler with hypergraph structure for high-dimensional inverse problems," *J. Comput. Graph. Stat.*, vol. 33, no. 3, pp. 814–832, 2023.

[44] A. Durmus, S. Majewski, and B. Miasojedow, "Analysis of Langevin Monte Carlo via convex optimization," *J. Mach. Learning Research*, vol. 20, no. 1, pp. 2666–2711, 2019.

[45] V. Plassier, M. Vono, A. Durmus, and E. Moulines, "DG-LMC: a turn-key and scalable synchronous distributed MCMC algorithm via Langevin Monte Carlo within Gibbs," in *Proc. Int. Conf. Machine Learning (ICML)*. PMLR, 2021, pp. 8577–8587.

[46] E. C. Faye, M. D. Fall, and N. Dobigeon, "Regularization by denoising: Bayesian model and Langevin-within-split Gibbs sampling – Supplementary materials," 2024.

[47] M. Terris, A. Repetti, J.-C. Pesquet, and Y. Wiaux, "Building firmly nonexpansive convolutional neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*, 2020, pp. 8658–8662.

[48] 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.

[49] 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.

[50] 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.

[51] 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.

[52] H. Chung, J. Kim, M. T. Mccann, M. L. Klasky, and J. C. Ye, "Diffusion posterior sampling for general noisy inverse problems," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2023.

[53] J. Choi, S. Kim, Y. Jeong, Y. Gwon, and S. Yoon, "ILVR: Conditioning method for denoising diffusion probabilistic models," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2021.

[54] P. Dhariwal and A. Nichol, "Diffusion models beat GANs on image synthesis," in *Adv. in Neural Information Process. Systems (NeurIPS)*, vol. 34, 2021, pp. 8780–8794.

[55] M. D. Fall and E. Barat, "A simple and efficient method for sampling mixture models based on Dirichlet and Pitman-Yor processes," 2024, preprint. [Online]. Available: https://hal.science/hal-04696216