Bregman geometry-aware split Gibbs sampling for Bayesian Poisson inverse problems*

Elhadji C. Faye† Mame Diarra Fall‡ Nicolas Dobigeon§ Éric Barat¶

November 18, 2025

Abstract

This paper proposes a novel Bayesian framework for solving Poisson inverse problems by devising a Monte Carlo sampling algorithm which accounts for the underlying non-Euclidean geometry. To address the challenges posed by the Poisson likelihood – such as non-Lipschitz gradients and positivity constraints – we derive a Bayesian model which leverages exact and asymptotically exact data augmentations. In particular, the augmented model incorporates two sets of splitting variables both derived through a Bregman divergence based on the Burg entropy. Interestingly the resulting augmented posterior distribution is characterized by conditional distributions which benefit from natural conjugacy properties and preserve the intrinsic geometry of the latent and splitting variables. This allows for efficient sampling via Gibbs steps, which can be performed explicitly for all conditionals, except the one incorporating the regularization potential. For this latter, we resort to a Hessian Riemannian Langevin Monte Carlo (HRLMC) algorithm which is well suited to handle priors with explicit or easily computable score functions. By operating on a mirror manifold, this Langevin step ensures that the sampling satisfies the positivity constraints and more accurately reflects the underlying problem structure. Performance results obtained on denoising, deblurring, and positron emission tomography (PET) experiments demonstrate that the method achieves competitive performance in terms of reconstruction quality compared to optimization- and sampling-based approaches.

1 Introduction

Reconstructing images from measurements corrupted by Poisson noise is a fundamental problem in computational imaging, with critical applications in low-light photography, astronomy [1], emission tomography [2], and fluorescence microscopy [3, 4]. In such settings, the data acquisition process involves photon-limited imaging, where the discrete and stochastic nature of photon arrivals leads to signal-dependent noise accurately modeled by a Poisson distribution [5]. Mathematically, the measurements $\mathbf{y} = (y_i)_{1\leq i\leq m} \in \mathbb{N}^m$ are modeled as

$$y_i \sim \mathcal{P}(\alpha \mathbf{h}_i^\top \mathbf{x}), \quad 1 \leq i \leq m, \tag{1}$$

where $\mathbf{x} \in \mathbb{R}_+^n$ denotes the unknown intensity image, $\alpha > 0$ controls the intensity level and reflects the severity of shot noise affecting the measurements, and each $\mathbf{h}_i \in \mathbb{R}^n$ defines the forward

*This work was funded in part by the ANR AI.iO Project (ANR-20-THIA-0017), the BACKUP project (ANR-23-CE40-0018-01) and the Artificial Natural Intelligence Toulouse Institute (ANITI, ANR-23-IACL-0002).

†Institut Denis Poisson, UMR CNRS University of Orléans, University of Tours, Orléans, France (elhadji-cisse.faye@univ-orleans.fr).

‡Univ Rouen Normandie, INSA Rouen Normandie, Université Le Havre Normandie, Normandie Univ, LITIS UR 4108, F-76000 Rouen, France (diarra.fall@univ-rouen.fr).

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

¶CEA, University of Paris-Saclay, France (eric.barat@cea.fr).

model, collected row-wise in a matrix $\mathbf{H} \in \mathbb{R}_+^{m \times n}$. From equation (1), the negative log-likelihood of the model is given by

$$\begin{aligned} -\log p(\mathbf{y}|\mathbf{x}) &= f(\mathbf{x};\mathbf{y}) \\ &= \sum_{i=1}^{m} \left[\alpha \mathbf{h}_i^\top \mathbf{x} - y_i \log(\alpha \mathbf{h}_i^\top \mathbf{x}) + \log(y_i!)\right]. \end{aligned} \tag{2}$$

Recovering the signal or image of interest $\mathbf{x}$ from the measurements $\mathbf{y}$ is typically an ill-posed inverse problem due to noise, incomplete data, and ill-conditioning of $\mathbf{H}$, necessitating suitable regularization strategies.

To address the ill-posedness of Poisson inverse problems, numerous optimization-based methods have been developed. These approaches typically formulate the reconstruction task as a variational problem, combining the data fidelity term $f(\cdot;\mathbf{y})$ derived from the Poisson likelihood (2) with a regularization term $g(\cdot)$ that encodes prior knowledge about the image. It is important to note that the Poisson log-likelihood does not exhibit a Lipschitz-continuous gradient. This is a sufficient condition for classical algorithms to be applicable with guaranteed convergence. To mitigate these issues, researchers have explored various approaches. One classical method is the Poisson Image Deconvolution by Augmented Lagrangian (PIDAL) [6] algorithm, which employs the Alternating Direction Method of Multipliers (ADMM) [7] framework with total variation (TV) regularization [8]. Alternatively, Bauschke *et al.* addressed this problem by proposing a proximal gradient descent (PGD) algorithm within the framework of Bregman divergence, called Bregman proximal gradient (BPG) [9]. The advantage of BPG lies in relaxing the smoothness requirement on the negative log-likelihood needed for PGD convergence and instead introduces the *NoLip* condition. Whatever their algorithmic structures, conventional instances of these optimization approaches often employ explicit handcrafted priors, such as total variation (TV) [8] or sparsity-inducing priors [10–12], which may not capture the complex structures present in natural images. To overcome these limitations, alternative approaches involving implicit data-driven priors have been introduced. Among these, Plug-and-Play (PnP) [13] and Regularization-by-Denoising (RED) [14] methods have gained significant success. These approaches modify traditional proximal splitting schemes by replacing proximal or gradient steps with the application of powerful denoisers, such as BM3D [15], demonstrating improved performance in Poisson reconstruction tasks [16]. Such strategies have gained in popularity after the recent advances in the field of deep learning, leveraging the expressive power of neural network-based denoisers [17]. In this direction, Hurault *et al.* propose a PnP algorithm specifically tailored for Poisson inverse problems, which ensures convergence by embedding a denoising operator within a Bregman geometry [18]. The method addresses the incompatibility between standard PnP schemes, which generally assume Gaussian noise and rely on Euclidean proximity operators, and the structure of Poisson noise, whose negative log-likelihood is neither Lipschitzian nor smooth. While these optimization-based methods have shown success in various scenarios, they primarily provide point estimates of the reconstructed image and do not inherently quantify the uncertainty associated with the reconstruction. To overcome this limitation, Bayesian methods offer a probabilistic alternative, where the solution $\mathbf{x}$ is treated as a random variable equipped with a prior distribution $p(\mathbf{x}) \propto \exp\{-g(\mathbf{x})\}$, and inference is conducted from the posterior distribution

$$\begin{aligned} \pi(\mathbf{x}) \triangleq p(\mathbf{x}|\mathbf{y}) &\propto p(\mathbf{y}|\mathbf{x})\, p(\mathbf{x}) \\ &\propto \exp\{-f(\mathbf{x};\mathbf{y}) - g(\mathbf{x})\}. \end{aligned} \tag{3}$$

As with the aforementioned optimization-based methods, Bayesian inference for Poisson inverse problems faces challenges due to the non-smoothness, non-convexity, and constraints such as non-negativity associated with the data fitting term $f(\cdot;\mathbf{y})$. To address these challenges, the authors of [19] propose a split Gibbs sampler (SGS) that samples from an asymptotically exact approximation of the target distribution [20]. By introducing auxiliary variables through a splitting strategy, SGS decomposes the original sampling problem into simpler subproblems that can be more easily managed, even in high-dimensional settings. More recently, Melidonis *et*

al. propose a reflected and regularized Langevin stochastic differential equation (SDE) that is well-posed and exponentially ergodic under mild conditions [21]. This framework leads to the development of four reflected proximal Langevin Markov chain Monte Carlo (MCMC) algorithms, demonstrating effectiveness in image deblurring, denoising, and inpainting tasks under various noise models, including Poisson noise. Building upon this, Klatzer *e*t al. develop a novel PnP Langevin sampling methodology tailored for low-photon Poisson imaging problems [22]. This approach incorporates two strategies: *i)* a PnP Langevin method with reflections and a Poisson likelihood approximation, and *ii)* a mirror sampling algorithm utilizing Riemannian geometry to handle constraints and the irregularity of the likelihood without approximations.

This work introduces a novel Bayesian model for addressing Poisson inverse problems that can be instantiated for a wide range of regularization potentials $g(\cdot)$, in particular those associated with implicit, data-driven priors. This model builds on several key ingredients. First, it leverages an exact data augmentation strategy, a modeling trick that has proven effective in previous works on Poisson inversion, to reformulate the likelihood in a tractable form while preserving the physical interpretability of the model. Second, it follows a geometry-aware double augmentation scheme that not only decouples the inference into simpler subproblems but also maintains favorable conjugacy properties for exact sampling. Third, a split Gibbs sampler (SGS) is implemented to handle the resulting augmented posterior efficiently. Three of the four steps of this SGS boil down to sampling according to standard conditional distributions, which can be achieved straightforwardly. The remaining step consists in sampling according to the conditional distribution which embeds the regularization potential $g(\cdot)$ whose specification remains at the discretion of the end-user. For the sake of generality, we therefore propose to resort to a versatile yet powerful sampling step that can handle, in practice, a broad class of prior distributions – provided their score function $\nabla \log g(\mathbf{x})$ is either explicit or readily computable. More precisely, this sampling step is based on the Hessian Riemannian Langevin Monte Carlo (HRLMC) algorithm [23], a mirror Langevin algorithm which explicitly accounts for the underlying geometry. Once combined, these components form a coherent Bayesian framework for geometry-aware Monte Carlo sampling in Poisson imaging. The proposed so-called HRLMC-within-SGS (HRLwSGS) algorithm allows for sampling from posterior distributions granted with data-driven prior models such as those based on denoising operators. The remainder of this paper is structured as follows. Section 2 reviews technical preliminaries necessary to the derivation of the proposed Bayesian framework subsequently exposed in Section 3. Section 4 describes the Monte Carlo algorithm proposed to sample from the target posterior distribution. Section 5 reports some numerical results illustrating the efficiency of the method when solving various Poisson inversion tasks. Section 6 summarizes the main findings and concludes the paper.

2 Background

This section reviews the two key concepts underlying the methodology derived in the sequel of this paper: Bregman divergences and the asymptotically exact data augmentation (AXDA) framework. These elements form the foundation of the Bayesian model derived to solve Poisson inverse problems efficiently.

2.1 Bregman divergence

Given a strictly convex and differentiable function $h : \mathbb{R}^n \to \mathbb{R}$, the Bregman divergence between two points $\mathbf{x}$ and $\mathbf{z}$ is defined as [24]

$$d_h(\mathbf{x}, \mathbf{z}) = h(\mathbf{x}) - h(\mathbf{z}) - \langle \nabla h(\mathbf{z}), \mathbf{x} - \mathbf{z} \rangle. \tag{4}$$

In the context of inverse problems, Bregman divergences were introduced in [25, 26] and are often used to encode constraints or non-Euclidean geometries. Unlike conventional metrics, Bregman divergences are generally asymmetric and do not necessarily satisfy the triangle inequality. Nevertheless, they capture important geometric properties and enable projections that are adapted to the underlying structure of the data. Several choices of the convex generator $h(\cdot)$ yield

Bregman divergences adapted to different problem geometries. For instance, when $h(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|^2$, the resulting Bregman divergence reduces to the squared Euclidean distance

$$d_{\mathrm{E}}(\mathbf{x}, \mathbf{z}) = \frac{1}{2}\|\mathbf{x} - \mathbf{z}\|^2, \tag{5}$$

which corresponds to the standard geometry underlying numerous problems underlying Gaussian noise models. Alternatively, choosing $h(\mathbf{x}) = \sum_{i=1}^{n} \mathbf{x}_i \log \mathbf{x}_i - \mathbf{x}_i$ yields the generalized Kullback-Leibler (KL) divergence

$$d_{\mathrm{KL}}(\mathbf{x}, \mathbf{z}) = \sum_{i=1}^{n} \left(\mathbf{x}_i \log \frac{\mathbf{x}_i}{\mathbf{z}_i} - \mathbf{x}_i + \mathbf{z}_i \right), \tag{6}$$

which is particularly suitable for modeling non-negative variables such as intensities and for tackling inverse problems under Poisson noise. Another relevant choice is the Burg entropy [27] defined on $\mathbb{R}^n_{++}$ by

$$h(\mathbf{x}) = -\sum_{i=1}^{n} \log \mathbf{x}_i. \tag{7}$$

The associated Bregman divergence, also known as the Itakura-Saito divergence, is given by

$$d_{\mathrm{IS}}(\mathbf{x}, \mathbf{z}) = \sum_{i=1}^{n} \left(\frac{\mathbf{x}_i}{\mathbf{z}_i} - \log \frac{\mathbf{x}_i}{\mathbf{z}_i} - 1 \right), \tag{8}$$

which is known to be particularly appropriate for modeling strictly positive variables and arises naturally in information geometry and Poisson-like models. This divergence respects the multiplicative structure of the positive orthant and enforces positivity implicitly and has been shown to be particularly well suited for modeling multiplicative and Poisson noise processes [28–30].

2.2 Asymptotically exact data augmentation (AXDA)

The AXDA framework [20] introduces auxiliary variables to reformulate complex or intractable distributions into more manageable forms. In Bayesian inference, to facilitate efficient sampling from the posterior distribution (3), AXDA introduces an auxiliary variable $\mathbf{z} \in \mathbb{R}^n$ such that the joint (augmented) distribution writes

$$\pi_\rho(\mathbf{x}, \mathbf{z}) \propto \exp\left\{-f(\mathbf{x}; \mathbf{y}) - g(\mathbf{z})\right\} \kappa_\rho(\mathbf{z}, \mathbf{x}), \tag{9}$$

where $\kappa_\rho(\cdot, \cdot)$ is such that π_ρ defines a proper joint distribution. Vono *et al.* [20] discuss various design choices for the kernel function $\kappa_\rho(\cdot, \cdot)$, which plays a central role in the formulation of the augmented posterior distribution and the subsequent sampling scheme. Among the possible constructions, a general class of admissible geometry-aware kernels can be defined as

$$\kappa_\rho(\mathbf{z}, \mathbf{x}) = \exp\left\{ -\frac{1}{\rho} d_h(\mathbf{z}, \mathbf{x}) + \varphi(\mathbf{z}) \right\}, \tag{10}$$

where $\rho > 0$ is a coupling parameter, d_h denotes the Bregman divergence (4) associated with the strictly convex and $\mathcal{C}^2$ function $h \colon \mathbb{R}^n \to \mathbb{R}$ defining the Bregman geometry and $\varphi : \mathbb{R}^n \to \mathbb{R}$ is a normalization potential ensuring integrability of the kernel. A natural and widely used choice is the quadratic generator $h(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|^2$ which leads to the standard Euclidean distance defined in (5) and yields the following Gaussian kernel

$$\kappa_\rho(\mathbf{z}, \mathbf{x}) \propto_{\mathbf{z}} \exp\left\{ -\frac{1}{2\rho} \|\mathbf{z} - \mathbf{x}\|^2 \right\}. \tag{11}$$

This choice has been adopted in many works from the literature dedicated to inverse or regression problems underlying Gaussian noises [31–35] or even Poisson noises [19].

As established in [20], the total variation distance between the marginal density $p_\rho(\mathbf{x}|\mathbf{y}) \triangleq \int \pi_\rho(\mathbf{x}, \mathbf{z})\mathrm{d}\mathbf{z}$ and the target distribution $\pi(\mathbf{x})$ vanishes as $\rho \to 0$, i.e

$$\|\pi(\mathbf{x}) - p_\rho(\mathbf{x})\|_{\mathrm{TV}} \xrightarrow[\rho \to 0]{} 0.$$

This result guarantees that the original distribution π is asymptotically recovered from the marginal p_ρ in the limit of vanishing coupling parameter ρ. The split Gibbs sampler (SGS) alternatively samples according to the two conditional distributions associated with the augmented distribution π_ρ to generate samples asymptotically distributed according to (9) [36]. 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. Thus, SGS shares strong similarities with ADMM [7] and HQS [37] methods since this divide-and-conquer strategy leads to simpler, scalable and more efficient sampling schemes.

3 Bayesian model

This section builds the Bayesian model step-by-step. As already emphasized in the introduction, this model can embed a large variety of prior distributions $p(\mathbf{x}) \propto \exp\{-g(\mathbf{x})\}$. Thus for the sake of generality and simplicity, this model is derived without specifying the regularization potential $g(\cdot)$, even if particular choices of this potential will be considered later. We first describe the exact data augmentation strategy, which reformulates the Poisson likelihood in a way that enables efficient sampling while preserving physical interpretability. We then introduce a second, geometry-aware augmentation that allows for a favorable decoupling of the variables and leads to conditional distributions with tractable forms. Finally, we present the full augmented posterior distribution that serves as the basis for the proposed sampling scheme.

3.1 Exact data augmentation

To mitigate the challenges associated with the non-Lipschitz property of the Poisson likelihood, we propose leveraging a data augmentation strategy initially adopted in the seminal and popular works introducing the maximum likelihood - expectation maximization (ML-EM) algorithms for emission and transmission tomography [38, 39]. Such a smart augmentation scheme in the data space has received some attention in later statistical research related to tomography [40, 41], while authors in [42–45] consider an alternative augmentation in the image space.

The key idea is to reformulate the original Poisson likelihood using unobserved latent variables, yielding a tractable and physically interpretable reformulation of the inverse problem. This strategy consists in introducing latent variables $\mathbf{n} = \{n_{ij}\}_{i,j} \in \mathbb{N}^{m \times n}$ such that

$$n_{ij}|x_j \sim \mathcal{P}(\alpha h_{ij} x_j) \tag{12}$$

where n_{ij} are mutually independent for all (i, j). The conditional distribution of the latent variables given the unknown image becomes factorized and tractable as

$$\begin{aligned} p(\mathbf{n}|\mathbf{x}) &= \prod_{i=1}^{m} \prod_{j=1}^{n} \frac{(\alpha h_{ij} x_j)^{n_{ij}} e^{-\alpha h_{ij} x_j}}{n_{ij}!} \\ &= \exp\left\{ \sum_{i=1}^{m} \sum_{j=1}^{n} \left[n_{ij} \log(\alpha h_{ij} x_j) - \alpha h_{ij} x_j - \log(n_{ij}!) \right] \right\}. \end{aligned}$$

In this formulation, the observed measurements $\mathbf{y} \in \mathbb{N}^m$ are deterministically related to the latent variables $\mathbf{n}$ via the coherence condition

$$\sum_{j=1}^{n} n_{ij} = y_i, \quad \text{for all } i = 1, \ldots, m. \tag{13}$$

This constraint ensures consistency between the latent process and the measured data. Accordingly, the conditional distribution of $\mathbf{y}$ given $\mathbf{n}$ can be expressed as

$$-\log p(\mathbf{y} \mid \mathbf{n}) = \iota_{C_{\mathbf{y}}}(\mathbf{n}). \tag{14}$$

where $\iota_{C_{\mathbf{y}}}$ denotes the indicator function of the feasible set $C_{\mathbf{y}}$

$$\iota_{C_{\mathbf{y}}}(\mathbf{n}) = \begin{cases} 0, & \text{if } \mathbf{n} \in C_{\mathbf{y}}, \\ +\infty, & \text{else.} \end{cases}$$

and $C_{\mathbf{y}} = \left\{ \mathbf{n} \in \mathbb{N}^{m \times n} : \sum_{j=1}^{n} n_{ij} = y_i, \ \forall i \right\}$. Importantly, this augmentation is exact in the sense that marginalizing out the latent variables recovers the original likelihood, as stated in the following theorem.

Theorem 1. *The joint likelihood defined as*

$$p(\mathbf{y}, \mathbf{n}|\mathbf{x}) = p(\mathbf{y}|\mathbf{n})\, p(\mathbf{n}|\mathbf{x}) \propto \exp\{-f(\mathbf{x}, \mathbf{n}; \mathbf{y})\} \tag{15}$$

with

$$f(\mathbf{x}, \mathbf{n}; \mathbf{y}) = \sum_{i=1}^{m} \sum_{j=1}^{n} [-n_{ij} \log(\alpha h_{ij} x_j) + \alpha h_{ij} x_j + \log(n_{ij}!)] + \iota_{C_{\mathbf{y}}}(\mathbf{n})$$

satisfies

$$\sum_{\mathbf{n} \in \mathbb{N}^{m \times n}} p(\mathbf{y}|\mathbf{n})\, p(\mathbf{n}|\mathbf{x}) = p(\mathbf{y}|\mathbf{x}) \tag{16}$$

where $p(\mathbf{y}|\mathbf{x})$ is the Poisson likelihood defined in (2).

Proof. See Appendix A. □

The introduction of these latent variables transforms the Poisson likelihood (2) into an augmented likelihood (15) that will be shown to be suitable to splitting and Monte Carlo sampling strategies. Precisely, assuming a prior distribution defined in (3), the joint posterior over $(\mathbf{x}, \mathbf{n})$ is given by

$$\begin{aligned} p(\mathbf{x}, \mathbf{n}|\mathbf{y}) &\propto p(\mathbf{y}|\mathbf{n})\, p(\mathbf{n}|\mathbf{x})\, p(\mathbf{x}) \\ &\propto \exp\{-f(\mathbf{x}, \mathbf{n}; \mathbf{y}) - g(\mathbf{x})\}. \end{aligned} \tag{17}$$

Theorem 1 ensures that the marginal distribution of $\mathbf{x}$ under the augmented model coincides exactly with the true posterior. This guarantees that inference performed in the augmented space using $(\mathbf{x}, \mathbf{n})$ targets the correct Bayesian posterior over $\mathbf{x}$, while enabling tractable sampling steps.

3.2 Double variable splitting for AXDA

To further separate the contributions of the prior and the likelihood in the augmented posterior distribution (17), we adopt the AXDA framework recalled in Section 2.2. This methodology involves introducing an auxiliary variable $\mathbf{z}_1 \in \mathbb{R}^n_{++}$ in the target posterior distribution defined in Equation (17). This leads to the following augmented posterior distribution denoted by π_ρ:

$$\pi_\rho(\mathbf{x}, \mathbf{n}, \mathbf{z}_1) \propto \exp\{-f(\mathbf{x}, \mathbf{n}; \mathbf{y}) - g(\mathbf{z}_1)\}\ \kappa_\rho(\mathbf{z}_1, \mathbf{x}). \tag{18}$$

A central modeling component in the augmented formulation (18) is the choice of the kernel function κ_ρ, which governs the coupling between the auxiliary variable $\mathbf{z}_1$ and the variable of interest $\mathbf{x}$. As already pointed out in Section 2.2, a popular and widely-adopted choice consists in using a Gaussian kernel (11). While this kernel leads to simple updates when coupled with a Gaussian likelihood, it remains ill-suited to inverse problems involving Poisson noise which are considered in this work. First, in the absence of conjugacy with respect to the Poisson likelihood,

it does not result in particularly simple Gibbs steps. Moreover, a Gaussian choice would disregard the positivity constraints inherent in intensity variables and would fail to reflect the geometry of Poisson counts. To overcome these limitations, this work proposes to define the kernel κ_ρ using a Bregman divergence that is specifically adapted to the statistical and geometric structure of Poisson inverse problems. It is worth noting that, according to the authors' knowledge, this is the first time that a non-Gaussian kernel κ_ρ is adopted within an AXDA framework to solve real-world estimation problems beyond the toy examples considered in [20]. More precisely, a particularly relevant choice for the convex generator h in the presence of positivity constraints is the Burg entropy, defined in (7) [27]. The corresponding coupling kernel has the following structure

$$\kappa_\rho(\mathbf{z}_1, \mathbf{x}) = \exp\left\{-\frac{1}{\rho} d_{\mathrm{IS}}(\mathbf{z}_1, \mathbf{x}) + \varphi(\mathbf{z}_1)\right\} \tag{19}$$

where $d_{\mathrm{IS}}(\cdot, \cdot)$ is the Itakura-Saito divergence defined in (8) and $\varphi(\mathbf{z}_1)$ is the normalizing potential

$$\varphi(\mathbf{z}_1) = -\sum_{j=1}^{n} \log z_{1j}. \tag{20}$$

This construction introduces a non-Euclidean geometry that preserves strict positivity and reflects the information structure of photon-limited measurements. Examining the Itakura-Saito divergence, as defined in (8), highlights that the kernel naturally embeds the latent space within the positive orthant, promoting multiplicative consistency between the auxiliary variable $\mathbf{z}_1$ and the variable of interest $\mathbf{x}$.

The resulting doubly augmented posterior distribution enriched with the latent variables $\mathbf{n}$ and the splitting variable $\mathbf{z}_1$ writes

$$\pi_\rho(\mathbf{x}, \mathbf{n}, \mathbf{z}_1) \propto \exp\left\{ - f(\mathbf{x}, \mathbf{n}; \mathbf{y}) - g(\mathbf{z}_1) - \frac{1}{\rho} d_{\mathrm{IS}}(\mathbf{z}_1, \mathbf{x}) + \varphi(\mathbf{z}_1)\right\}. \tag{21}$$

Despite the structural advantages brought by the introduction of the first auxiliary variable $\mathbf{z}_1$, which decouples the data-fitting potential $f(\cdot; \mathbf{y})$ from the regularization potential $g(\cdot)$, this formulation still presents notable limitations. In particular, the conditional distribution of the image variable $\mathbf{x}$ given $\mathbf{z}_1$ and $\mathbf{n}$ does not yield a conjugate structure amenable to efficient sampling by Gibbs steps in high-dimensional settings. The absence of conjugacy implies that this conditional distribution cannot be sampled directly using standard methods, and instead necessitates more elaborate procedures such as rejection sampling, which are often inefficient because computationally demanding.

Fortunately, restoring tractable conditional updates by Gibbs steps can be achieved by judiciously leveraging the AXDA framework once more. Introducing a second auxiliary variable $\mathbf{z}_2 \in \mathbb{R}^n_{++}$ leads to a double splitting strategy, whereby this newly introduced splitting variable $\mathbf{z}_2$ serves as a stochastic mediator between $\mathbf{x}$ and $\mathbf{z}_1$. More precisely, we build on the doubly augmented distribution (21) as

$$\pi_\rho(\mathbf{x}, \mathbf{n}, \mathbf{z}_{1:2}) \propto \exp\left\{ - f(\mathbf{x}, \mathbf{n}; \mathbf{y}) - g(\mathbf{z}_1) - \frac{1}{\rho} d_{\mathrm{IS}}(\mathbf{z}_1, \mathbf{z}_2) + \varphi(\mathbf{z}_1)\right\} \check{\kappa}_\rho(\mathbf{z}_2, \mathbf{x}) \tag{22}$$

where the second coupling kernel $\check{\kappa}_\rho$ also derives from an Itakura-Saito divergence, yet evaluated with reflected arguments

$$\check{\kappa}_\rho(\mathbf{z}_2, \mathbf{x}) = \exp\left\{-\frac{1}{\rho} d_{\mathrm{IS}}(\mathbf{x}, \mathbf{z}_2) + \varphi(\mathbf{z}_2)\right\}. \tag{23}$$

This finally leads to the following triply augmented posterior distribution

$$\pi_\rho(\mathbf{x}, \mathbf{n}, \mathbf{z}_{1:2}) \propto \exp\left\{ - f(\mathbf{x}, \mathbf{n}; \mathbf{y}) - g(\mathbf{z}_1) - \frac{1}{\rho} d_{\mathrm{IS}}(\mathbf{z}_1, \mathbf{z}_2) + \varphi(\mathbf{z}_1) - \frac{1}{\rho} d_{\mathrm{IS}}(\mathbf{x}, \mathbf{z}_2) + \varphi(\mathbf{z}_2)\right\}. \tag{24}$$

Combining this double splitting strategy with the exact augmentation introduced in Section 3.1 leads to an augmented posterior distribution (24) which exhibits a tractable conditional structure.

As will be demonstrated in the next sections, it yields four conditional distributions, three of which belong to known and tractable families. In contrast, a single splitting would have lead to more complex dependence and conditional distributions. This property enables an efficient Gibbs-like sampling scheme to be used, which samples iteratively from the full conditionals of $(\mathbf{x}, \mathbf{n}, \mathbf{z}_1, \mathbf{z}_2)$. It is also worth noting that the non-symmetry of the non-Gaussian coupling kernels κ_ρ and $\check{\kappa}_\rho$ questions the place of the splitting variables $\mathbf{z}_1$ and $\mathbf{z}_2$. The alternative choices which would have consisted in exchanging the roles played by the splitting variables or the variable of interest would have lead to a more complex nay invalid sampling scheme, as further discussed in Section 4.3. Finally, by introducing the additional splitting variable $\mathbf{z}_2$, which interacts with both $\mathbf{z}_1$ and $\mathbf{x}$, the Markov chain transitions are expected to exhibit smoother behaviors. This second splitting variable $\mathbf{z}_2$ acts as a probabilistic buffer that facilitates information exchange between the prior and the likelihood terms, improving the convergence behavior and stability of the sampling process. The details of this sampling procedure are provided in Section 4.

4 Hessian Riemannian Langevin Monte Carlo within split-Gibbs sampler (HRLwSGS)

To perform posterior inference, we develop a Gibbs sampling algorithm targeting the joint distribution over latent variables $\mathbf{x}$, $\mathbf{n}$, $\mathbf{z}_1$, and $\mathbf{z}_2$, conditioned on the observed counts $\mathbf{y}$. Our model structure leads to four conditional distributions, three of which admit closed-form classical sampling schemes, while the fourth requires a specialized geometry-aware sampling method. We therefore organize the description of the Gibbs sampler into two parts:

- Classical sampling steps: multinomial, gamma, inverse-gamma,
- Non-standard sampling step performed via the Hessian Riemannian Langevin Monte Carlo (HRLMC) algorithm.

4.1 Standard sampling steps

The latent counts $\mathbf{n} = \{n_{ij}\}$ are introduced as latent variables linking the observed data $\mathbf{y}$ to the unknown image $\mathbf{x}$ via the equations (12) and (13). The corresponding conditional distribution is given by

$$\begin{aligned} p(\mathbf{n}|\mathbf{y}, \mathbf{x}) &\propto \exp\left\{-f(\mathbf{x}, \mathbf{n}; \mathbf{y})\right\} \\ &\propto \exp\left\{\sum_{i=1}^{m}\sum_{j=1}^{n}\left[n_{ij}\log(\alpha h_{ij}x_j) - \log(n_{ij})\right] + \iota_{C_{\mathbf{y}}}(\mathbf{n})\right\}. \end{aligned} \tag{25}$$

Conditioned on $\mathbf{y}$ and $\mathbf{x}$, the latent variables $\mathbf{n}$ follow a product of constrained Poisson distributions. By applying the well-known result that Poisson variables conditioned on their sum follow a multinomial distribution, we obtain

$$\mathbf{n}_i|\mathbf{x}, y_i \sim \text{Multinomial}\left\{y_i, \left\{\frac{h_{ij}x_j}{\sum_{k=1}^{n} h_{ik}x_k}\right\}_{j=1}^{n}\right\}, \quad i = 1, \ldots, m \tag{26}$$

where $\mathbf{n}_i = (n_{i1}, \ldots, n_{in}) \in \mathbb{N}^n$. Each row $\mathbf{n}_{i\cdot}$ is thus sampled independently, making this step computationally efficient.

Given the counts $\mathbf{n}$ and the auxiliary variable $\mathbf{z}_2$, the posterior for $\mathbf{x}$ results from a combination of an augmented Poisson likelihood and a prior defined by the Bregman divergence associated with the Burg entropy

$$\begin{aligned} p(\mathbf{x}|\mathbf{z}_2, \mathbf{n}) &\propto \exp\left\{-f(\mathbf{x}, \mathbf{n}; \mathbf{y}) - \frac{1}{\rho}d_{\text{IS}}(\mathbf{x}, \mathbf{z}_2)\right\} \\ &\propto \exp\left\{\sum_{i=1}^{m}\sum_{j=1}^{n}\left[n_{ij}\log(\alpha h_{ij}x_j) - \alpha h_{ij}x_j\right] - \frac{1}{\rho}\sum_{j=1}^{n}\left(\frac{x_j}{z_{2j}} - \log\frac{x_j}{z_{2j}}\right)\right\} \end{aligned}$$

The conditional distribution factorizes as

$$x_j|n_{\cdot j}, z_{2j} \sim \text{Gamma}\left(\sum_{i=1}^{m} n_{ij} + \frac{1}{\rho} + 1,\ \alpha \sum_{i=1}^{m} h_{ij} + \frac{1}{\rho z_{2j}}\right), \quad j = 1, \ldots, n. \tag{27}$$

The variable $\mathbf{z}_2$ acts as an intermediary in the prior coupling between $\mathbf{x}$ and $\mathbf{z}_1$. Its conditional distribution is given by

$$\begin{aligned} p(\mathbf{z}_2|\mathbf{z}_1, \mathbf{x}) &\propto \exp\left\{ -\frac{1}{\rho} d_{\text{IS}}(\mathbf{z}_1, \mathbf{z}_2) - \frac{1}{\rho} d_{\text{IS}}(\mathbf{x}, \mathbf{z}_2) + \varphi(\mathbf{z}_2) \right\} \\ &\propto \exp\left\{ -\frac{1}{\rho} \sum_{j=1}^{n} \left(\frac{x_j}{z_{2j}} - \log \frac{1}{z_{2j}} \right) - \frac{1}{\rho} \sum_{j=1}^{n} \left(\frac{z_{1j}}{z_{2j}} - \log \frac{1}{z_{2j}} \right) - \sum_{j=1}^{n} \log z_{2j} \right\} \end{aligned}$$

This conditional distribution is also fully explicit and separable

$$z_{2j}|z_{1j}, x_j \sim \text{InvGamma}\left(\frac{2}{\rho},\ \frac{x_j + z_{1j}}{\rho}\right), \quad j = 1, \ldots, n. \tag{28}$$

In the hierarchical model, the variable $\mathbf{z}_2$ acts as a latent intermediary between $\mathbf{z}_1$ and $\mathbf{x}$. This auxiliary role enables efficient block-wise sampling and facilitates posterior exploration by decoupling the variables while still allowing information to propagate through the hierarchy. Notably, when $\rho \to 0$, the inverse-gamma distribution becomes increasingly concentrated, and its mean converges to: $\mathbb{E}[z_{2j}] \longrightarrow \frac{x_j + z_{1j}}{2}$, as $\rho \to 0$. In this regime, $\mathbf{z}_2$ effectively interpolates between $\mathbf{x}$ and $\mathbf{z}_1$, behaving like a stochastic average. Consequently, the hierarchical model induces a form of local averaging that stabilizes the sampling process: each update of $\mathbf{z}_2$ carries combined information from the current states of $\mathbf{x}$ and $\mathbf{z}_1$. This behavior is particularly useful for reducing variance in posterior inference and promoting better mixing in Gibbs-type algorithms.

4.2 Non-standard sampling

The conditional distribution of $\mathbf{z}_1$ given $\mathbf{z}_2$ writes

$$p(\mathbf{z}_1|\mathbf{z}_2) \propto \exp\left\{ -g(\mathbf{z}_1) - \frac{1}{\rho} d_{\text{IS}}(\mathbf{z}_1, \mathbf{z}_2) + \varphi(\mathbf{z}_1). \right\}$$

It involves the regularization term $g(\cdot)$ associated with the prior distribution defined in (3). Its structure can be easily interpreted as the posterior distribution associated with a non-Gaussian denoising problem with the regularization function $g(\cdot)$. The objective is to reconstruct $\mathbf{z}_1$ from a noisy observation $\mathbf{z}_2$, which is assumed to be affected by a multiplicative noise consistent with an inverse-gamma distribution underlying the Itakura-Saito divergence of the data-fitting term. In other words, the discrepancy between the clean and observed variables is measured through the Bregman divergence $d_h(\mathbf{z}_1, \mathbf{z}_2)$, induced by the Burg entropy, which reflects the geometry of positive-valued data. Introducing the potential function

$$U(\mathbf{z}_1) = g(\mathbf{z}_1) + \sum_{j=1}^{n} \log z_{1j} + \frac{1}{\rho} \sum_{j=1}^{n} \left(\frac{z_{1j}}{z_{2j}} - \log z_{1j} \right), \tag{29}$$

this conditional distribution can be rewritten as $p(\mathbf{z}_1|\mathbf{z}_2) \propto \exp\{-U(\mathbf{z}_1)\}$. For most regularization potentials $g(\cdot)$, this distribution does not belong to a standard family. To enable the use of a wide class of regularizations, including implicit data-driven prior distributions, we propose sapmling from this distribution using the Hessian Riemannian Langevin Monte Carlo (HRLMC) algorithm [23, 46], which is particularly suited to constrained domains such as $\mathbb{R}^n_{++}$. The mirror Langevin algorithm is a discretization of Langevin dynamics on a Riemannian manifold, where the geometry is defined by the Hessian of a convex potential function $\phi(\cdot)$. This approach allows for

efficient sampling from distributions supported on manifolds or constrained subsets of Euclidean space. The HRLMC updating rule writes

$$\mathbf{z}_1^{(t+1)} = \nabla\phi^\star\left(\nabla\phi(\mathbf{z}_1^{(t)}) - \gamma\nabla U(\mathbf{z}_1^{(t)}) + \sqrt{2\gamma\nabla^2\phi(\mathbf{z}_1^{(t)})}\boldsymbol{\varepsilon}^{(t)}\right), \tag{30}$$

where $\boldsymbol{\varepsilon}^{(t)} \sim \mathcal{N}(0, \mathbf{I}_n)$ and $\phi^\star(\cdot)$ is the Legendre transform of $\phi(\cdot)$. In the considered Poisson inversion context, the mirror function $\phi(\cdot)$ is set as the Burg entropy (7) and its gradient and Hessian are given[1] by $\nabla\phi(\mathbf{z}) = -\frac{1}{\mathbf{z}}$ and $\nabla^2\phi(\mathbf{z}) = \frac{1}{\mathbf{z}^2}\mathbf{I}_n$, respectively. This sampling strategy guarantees positivity and adapts to the manifold geometry. When the potential $U(\cdot)$ is given by (29), its gradient writes

$$\nabla U(\mathbf{z}_1) = \beta\nabla g(\mathbf{z}_1) + \frac{1}{\rho\mathbf{z}_2} + \left(1 - \tfrac{1}{\rho}\right)\frac{1}{\mathbf{z}_1}. \tag{31}$$

Thus, the practical implementation of the HRLMC iterations only requires the gradient $\nabla g(\cdot)$ of the regularization potential to be explicit or easily computable – a quantity also referred to as the prior score. This encompasses several recent and powerful data-driven priors, ranging from denoiser-based regularizations (see Appendix B) to those derived from generative models. Moreover, the HRLMC algorithm has been shown to exhibit favorable convergence properties under certain conditions, including relative strong convexity and Lipschitz-smoothness of the potential function $U(\cdot)$. These conditions are further discussed in Appendix C, in particular when the potential $g(\cdot)$ derives from the RED paradigm.

Algorithm 1 HRLwSGS algorithm for Bayesian Poisson inversion.

Input: Observation $\mathbf{y}$, degradation matrix $\mathbf{H} = \{h_{ij}\}$, regularization parameter β, coupling parameter ρ, step-size γ, number of burn-in iterations N_{bi}, total number of iterations N_{MC}

Initialization: $\mathbf{n}^{(0)}$, $\mathbf{x}^{(0)}$, $\mathbf{z}_1^{(0)}$, $\mathbf{z}_2^{(0)}$

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

2: Sample count variable from (26):

$$\mathbf{n}^{(t+1)} \sim \prod_{i=1}^{m} \text{Multinomial}\left\{y_i, \left(\frac{h_{i1}x_1^{(t)}}{\sum_{j=1}^m h_{ij}x_j^{(t)}}, \ldots, \frac{h_{in}x_n^{(t)}}{\sum_{j=1}^m h_{ij}x_j^{(t)}}\right)\right\}$$

3: Sample variable of interest from (27):

$$\mathbf{x}^{(t+1)} \sim \prod_{j=1}^{n} \text{Gamma}\left(x_j^{(t)}; \sum_{i=1}^{m} n_{ij}^{(t+1)} + \frac{1}{\rho} + 1,\ \alpha\sum_{i=1}^{m} h_{ij} + \frac{1}{\rho z_{2j}^{(t)}}\right)$$

4: Sample splitting variable from (30) using HRLMC:

$$\mathbf{z}_1^{(t+1)} = \nabla\phi^*\left(\nabla\phi(\mathbf{z}_1^{(t)}) - \gamma\nabla U(\mathbf{z}_1^{(t)}) + \sqrt{2\gamma\nabla^2\phi(\mathbf{z}_1^{(t)})}\boldsymbol{\varepsilon}^{(t)}\right), \quad \boldsymbol{\varepsilon}^{(t)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

5: Sample splitting variable from (28):

$$\mathbf{z}_2^{(t+1)} \sim \prod_{j=1}^{n} \text{InvGamma}\left(z_{2j}^{(t)};\ \frac{2}{\rho},\ \frac{x_j^{(t+1)} + z_{1j}^{(t+1)}}{\rho}\right)$$

6: **end for**

Output: Collection of samples $\left\{\mathbf{x}^{(t)}, \mathbf{z}_1^{(t)}, \mathbf{z}_2^{(t)}\right\}_{t=N_{\text{bi}}+1}^{N_{\text{MC}}}$

[1]The inversion and exponentiation operations should be understood as component-wise.

4.3 Discussion on the splitting strategy, its theoretical and practical implications

In the proposed framework, the introduction of the first auxiliary variable makes two splitting strategies based on data augmentation theoretically possible. These differ in the direction of the divergence: $d_h(\mathbf{z}_1, \mathbf{x})$ or $d_h(\mathbf{x}, \mathbf{z}_1)$. While both formulations yield valid augmented distributions, they lead to significantly different conditional structures, which in turn drive the theoretical properties of the sampler, its implementation, and its practical robustness.

In the first strategy, based on the divergence $d_h(\mathbf{z}_1, \mathbf{x})$ adopted in Section 3, the conditional distribution $p(\mathbf{x}|\mathbf{z}_1, \mathbf{n})$ is a Generalized Inverse Gaussian (GIG) distribution. While sampling from this distribution is slightly more complex than from a standard gamma distribution, it remains tractable. Furthermore, the potential $U(\mathbf{z}_1)$ associated with the conditional $p(\mathbf{z}_1|\mathbf{x})$ exhibits favorable properties: its dependence on $\mathbf{z}_1$ is smoother and more regular. This structure simplifies the convergence analysis of the HRLMC sampler. In particular, the relative strong convexity constant remains positive under mild assumptions on the regularization potential $g(\cdot)$, and the commutator bound between $\nabla^2\phi(\cdot)$ and $\nabla^2 U(\cdot)$ stays moderate as long as $\nabla^2 g(\cdot)$ is bounded. Consequently, the required Lipschitz constant of $g(\cdot)$ can be relatively large without jeopardizing convergence.

In contrast, the alternative strategy, which relies on the divergence $d_h(\mathbf{x}, \mathbf{z}_1)$, yields a conditional distribution $p(\mathbf{x}|\mathbf{z}_1, \mathbf{n})$ that factorizes into gamma distributions. This allows faster and simpler sampling steps for $\mathbf{x}$, which may appear advantageous in practice. However, this gain in simplicity comes at a theoretical cost. The potential $U(\mathbf{z}_1)$ now involves terms such as x_j/z_{1j}, which cause gradients and Hessians to become ill-conditioned when z_{1j} approaches zero. As a result, the convergence analysis requires stricter assumptions on the boundedness of both $\mathbf{x}$ and $\mathbf{z}_1$, and above all, a tighter control of the Lipschitz constant L_{D} of the regularization $g(\cdot)$. For example, when $g(\cdot)$ derives from a prior associated with RED, ensuring the positivity of the relative strong convexity constant imposes that $L_{\mathrm{D}} < \epsilon_{\mathbf{z}_1}^4 / C_{\mathbf{z}_1}^4$, with $0 < \epsilon_{\mathbf{z}_1} \leq z_{1j} \leq C_{\mathbf{z}_1}$. This condition becomes particularly restrictive when the support of $\mathbf{z}_1$ is wide.

While both strategies are theoretically sound and elegant in structure, the first approach, based on $d_h(\mathbf{z}_1, \mathbf{x})$, strikes a better balance between theoretical guarantees and practical flexibility. It is compatible with a wider range of regularization functions $g(\cdot)$, including those with moderate Lipschitz constants. The second strategy, is simpler from an operational point of view but suffers from stricter theoretical constraints. For these reasons, the first strategy is generally more suitable for high-dimensional inverse problems involving data-driven potentials $g(\cdot)$, and is the one adopted in this work.

5 Numerical Experiments

In this section, we demonstrate the effectiveness of the proposed Bayesian sampling method by conducting a series of experiments related to Poisson inverse problems, namely image denoising, image deblurring, and positron emission tomography (PET) reconstruction. These tasks have been selected to demonstrate the ability of the proposed method to handle inverse problems of varying difficulty, including variations in the conditioning of the forward operator $\mathbf{H}$, the dimensionality of the images $\mathbf{x}$ and $\mathbf{y}$, and the severity of the noise α.

For denoising and deblurring, we use 30 RGB images of size 256×256 from the ImageNet dataset [47]. For the PET experiment, we use a standard 128×128 synthetic phantom. In all experiments, the measurements are corrupted with Poisson noise, which is typical for photon-limited imaging scenarios. For the denoising and deblurring tasks, the noise level is controlled by the scaling factor $\alpha > 0$, which adjusts the intensity of the clean image before applying the Poisson degradation. It is worth noting that a lower value of α corresponds to a higher level of noise, i.e., a lower signal-to-noise ratio. To evaluate the proposed method under varying noise conditions, two distinct noise levels have been considered: $\alpha = 10$ (severe noise) and $\alpha = 40$ (moderate noise).

The HRLwSGS sampling algorithm described in Section 4 is implemented when the potential $g(\cdot)$ is defined as the RED potential (see Appendix B.1 for more details) and the corresponding

denoiser $\mathsf{D}_\nu(\cdot)$ is the DRUNet. It is worth noting that this denoiser is used off-the-shelf, i.e., it has been pretrained on images corrupted by Gaussian noise and not to specifically handled Poisson noise [48]. The resulting inversion method is referred to as RED-HRLwSGS. To illustrate the versatility of the proposed Bayesian framework, which can accommodate various forms of regularization, the proposed algorithm has been also instantiated using a Bregman score denoiser [18] (see Appendix B.2). The corresponding variant of the algorithm will be referred to as BSD-HRLwSGS. We compare these methods with several state-of-the-art algorithms specifically designed for solving inverse problems under Poisson noise. These include the optimization-based ADMM algorithm, TV-PIDAL [6] and two Monte Carlo sampling methods: *i)* TV-SPA which leverages an AXDA strategy and a TV regularization [19] and *ii)* RPnP-ULA which relies on a reflected Langevin scheme and the use of a denoiser, chosen as DRUNet (RED-RPnP-ULA) or BSD (BSD-RPnP-ULA) [21]. The PIDAL optimization-based method provides point estimates, typically corresponding to the maximum a posteriori (MAP) solution. In contrast, sampling-based methods such as RPnP-ULA, TV-SPA and the proposed algorithm HRLwSGS draw samples from the posterior distribution. For these methods, the solution to the Poisson inversion problem has been chosen as the minimum mean square error (MMSE) estimator, which is approximated by averaging the generated samples as follows

$$\hat{\mathbf{x}}_{\mathrm{MMSE}} = \frac{1}{N_{\mathrm{MC}} - N_{\mathrm{bi}}} \sum_{t=N_{\mathrm{bi}}+1}^{N_{\mathrm{MC}}} \mathbf{x}^{(t)} \tag{32}$$

where N_{MC} is the total number of iterations including N_{bi} burn-in iterations. Besides, these sampling-based methods not only enable point estimation, but also the quantification of uncertainty. Therefore, the results provided by these methods will be granted with uncertainty quantification in the form of pixelwise posterior standard deviation and coverage maps. For all tasks, the total number of iterations and the burn-in iterations of the sampling-based methods have been set to $N_{\mathrm{MC}} = 25\,000$ and $N_{\mathrm{bi}} = 10\,000$ respectively.

In addition to visual inspection, the methods are compared with respect to several quantitative figures-of-merit. Peak signal-to-noise ratio (PSNR) in decibels (dB) and structural similarity index (SSIM) [49] are considered as image quality metrics (the higher the score, the better the reconstruction). To capture perceptual differences that are more aligned with human visual perception, we also include the Learned Perceptual Image Patch Similarity (LPIPS) [50] metric, with lower scores denoting closer resemblance to the reference image.

Further information regarding the experimental settings and parameters of the compared methods are reported in Appendix D.

5.1 Poisson image denoising

This experiment assesses the performance of the compared algorithms where restoring images only corrupted by Poisson noise. In this simplified experimental scheme, the degradation operator is the identity matrix, $\mathbf{H} = \mathbf{I}$, meaning that the noise is applied directly to the image, with no additional transformations, such as convolution or masking, being applied. This setup enables a dedicated evaluation of the denoising capabilities of each method.

Quantitative results are reported in Table 1. For a severe noise level ($\alpha = 10$), BSD-HRLwSGS achieves the best overall performance, with the highest PSNR (25.50 dB) and SSIM (0.741). This demonstrates its ability to preserve both fidelity and structural details under severe degradation. RED-RPnP-ULA produces competitive results (PSNR of 25.10dB and SSIM of 0.723), but remains slightly inferior in terms of structural similarity. TV-PIDAL achieves noticeably lower accuracy, highlighting the limitations of classical variational approaches in this challenging regime. Finally, TV-SPA, despite its Bayesian formulation, struggles to cope with strong noise, as evidenced by its reduced quantitative scores and weaker perceptual quality.

Under moderate noise conditions ($\alpha = 40$), all methods perform better than in the low-noise regime. Among them, BSD-HRLwSGS consistently delivers the best overall results, with the highest PSNR (26.88dB), SSIM (0.825), and lowest LPIPS (0.229), indicating superior perceptual quality. Although RED-RPnP-ULA attains a competitive PSNR (26.74dB) and

Table 1: Denoising experiment: average performance and corresponding standard deviations.

		PSNR(dB)↑	**SSIM**↑	**LPIPS** ↓
$\alpha = 10$	TV-PIDAL	24.12±2.21	0.715±0.076	0.377±0.025
	TV-SPA	22.49±2.32	0.690±0.059	0.445±0.067
	RED-RPnP-ULA	<u>25.10±2.00</u>	<u>0.723±0.064</u>	**0.325±0.050**
	BSD-RPnP-ULA	23.71±2.01	0.701±0.081	0.387±0.032
	RED-HRLwSGS	24.89±2.08	0.719±0.081	0.333±0.072
	BSD-HRLwSGS	**25.50±1.92**	**0.741±0.064**	<u>0.329±0.067</u>
$\alpha = 40$	TV-PIDAL	25.54±2.86	0.736±0.082	0.338±0.064
	TV-SPA	23.87±3.65	0.713±0.078	0.382±0.058
	RED-RPnP-ULA	<u>26.74±2.33</u>	<u>0.785±0.054</u>	<u>0.269±0.047</u>
	BSD-RPnP-ULA	25.37±2.71	0.756±0.036	0.332±0.053
	RED-HRLwSGS	26.01±2.29	0.779±0.067	0.274±0.061
	BSD-HRLwSGS	**26.88±2.62**	**0.825±0.032**	**0.229±0.041**

favorable perceptual scores, its results are slightly inferior to those of BSD-HRLwSGS. Thus, BSD-HRLwSGS clearly outperforms both classical variational methods and other sampling-based strategies.

Figure 1 presents representative qualitative results for both noise levels ($\alpha = 40$ for the first two columns and $\alpha = 10$ for the last two columns). These visual comparisons are consistent with the quantitative evaluation: the proposed HRLwSGS method yields reconstructions that are visually closer to the ground truth, with sharper details, reduced artifacts, and improved perceptual quality.

5.2 Poisson image deconvolution

We evaluate the proposed method on the Poisson image deconvolution task, where the degradation operator $\mathbf{H}$ is modeled as a circulant convolution matrix associated with a spatially invariant Gaussian kernel of size 25×25 and standard deviation 1.6. This setting simulates realistic optical blur and constitutes a more challenging inverse problem than the denoising scenario due to the ill-posedness induced by the convolution. Quantitative performance is reported in Table 2 for two noise levels $\alpha = 10$ and $\alpha = 40$, while qualitative results are provided in Figure 2, and uncertainty quantification is analyzed in Figure 3.

At the high noise level ($\alpha = 10$), the proposed RED-HRLwSGS method achieves the best overall reconstruction quality, with the highest PSNR (23.22dB) and SSIM (0.681), while BSD-HRLwSGS provides a comparable structural consistency (SSIM of 0.677). RED-RPnP-ULA yields similar fidelity (PSNR of 23.08dB, SSIM of 0.672) but slightly lower perceptual quality (LPIPS of 0.398). Conversely, BSD-RPnP-ULA exhibits a noticeable drop across all metrics, and TV-based methods remain significantly inferior. Visual results in Figure 2 corroborate these observations, showing that RED-HRLwSGS better preserves structures and textures, while avoiding the ringing artifacts visible in BSD-HRLwSGS and RED-RPnP-ULA reconstructions.

At the lower noise level ($\alpha = 40$), all methods demonstrate significant improvement. RED-HRLwSGS achieves the highest PSNR (24.52dB) and SSIM (0.772), whereas RED-RPnP-ULA provides the best perceptual quality (LPIPS of 0.340) with nearly equivalent fidelity (PSNR of 24.11dB). BSD-HRLwSGS remains competitive (PSNR of 23.93dB and SSIM of 0.751). TV-PIDAL still lags behind in perceptual fidelity.

Beyond point estimation, the proposed Bayesian framework enables posterior uncertainty quantification, as illustrated in Figure 3. For $\alpha = 10$, we report both the pixelwise posterior standard deviation and the coverage map, where each pixel value corresponds to the minimum probability ensuring that the highest posterior density (HPD) interval contains the true intensity. Uncertainty maps reveal localized uncertainty concentrated along textured and edge regions. Compared to BSD-RPnP-ULA and TV-SPA, the uncertainty maps obtained with RED-HRLwSGS, BSD-HRLwSGS and RED-RPnP-ULA appear more spatially coherent and interpretable, which

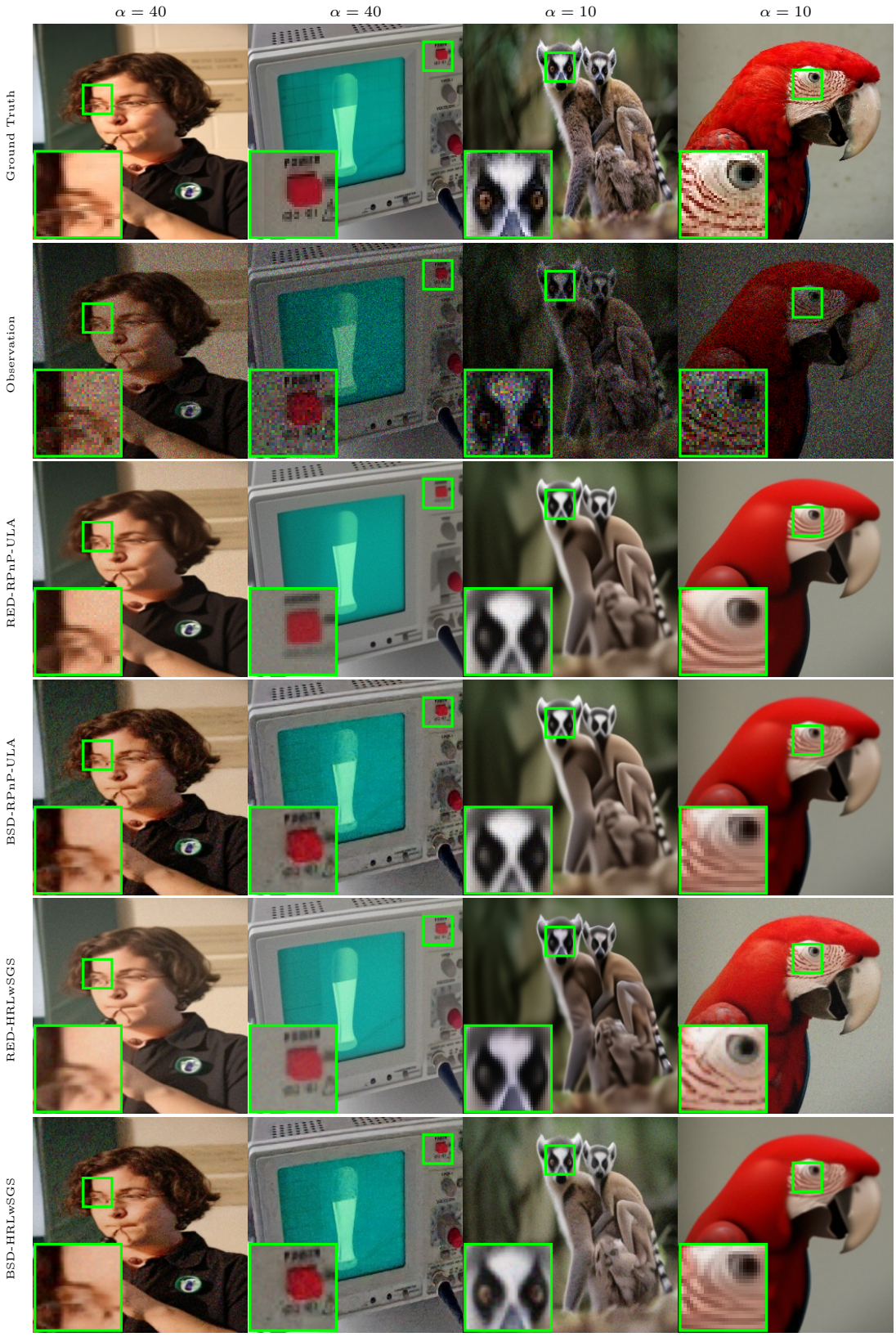

Figure 1: Denoising experiment: visual results. Due to space contraints, the results recovered by TV-SPA and TV-PIDAL are not reproduced, as their quality is significantly inferior.

is consistent with the expected behavior for deconvolution problems. The coverage maps support these observations: the probability that the HPD intervals contain the true intensity remains high (close to 1) across most of the image, with lower values mainly along edges.

Table 2: Deblurring experiment: average performance and corresponding standard deviations.

		PSNR(dB)↑	**SSIM**↑	**LPIPS** ↓
$\alpha = 10$	TV-PIDAL	22.21±2.02	0.621±0.081	0.455±0.048
	TV-SPA	21.32±2.12	0.629±0.065	0.460±0.043
	RED-RPnP-ULA	23.08±1.87	0.672±0.091	0.398±0.066
	BSD-RPnP-ULA	22.45±1.68	0.630±0.027	0.451±0.038
	RED-HRLwSGS	**23.22**±1.89	**0.681**±0.094	**0.396**±0.062
	BSD-HRLwSGS	22.98±1.88	0.677±0.065	0.435±0.044
$\alpha = 40$	TV-PIDAL	<u>24.13</u>±3.00	0.728±0.073	<u>0.356</u>±0.057
	TV-SPA	22.35±2.88	0.684±0.077	0.437±0.045
	RED-RPnP-ULA	24.11±2.78	<u>0.762</u>±0.088	**0.340**±0.053
	BSD-RPnP-ULA	23.31±2.13	0.739±0.068	0.424±0.035
	RED-HRLwSGS	**24.52**±1.41	**0.772**±0.088	0.382±0.054
	BSD-HRLwSGS	23.44±2.09	0.755±0.081	0.422±0.048

5.3 Positron Emission Tomographic Reconstruction

We evaluate the performance of the proposed method when performing a PET reconstruction task using a synthetic phantom of size 128×128. The forward operator $\mathbf{H}$ simulates a parallel-beam acquisition model with 512 projection angles uniformly distributed in $[0, \pi)$. The resulting system matrix $\mathbf{H} \in \mathbb{R}^{93184 \times 16384}$ is constructed using the ASTRA toolbox [51, 52] and defines a severely ill-posed inverse problem due to the sparsity of the measurements. This configuration emulates a realistic tomographic setting and poses a significant computational challenge due to the high dimensionality and ill-posedness of the problem.

Quantitative results are summarized in Table 3, and representative reconstructions are displayed in Figure 4. The proposed RED-HRLwSGS attains the highest PSNR (30.52dB), indicating superior reconstruction fidelity in terms of PSNR. Both TV-PIDAL and RED-RPnP-ULA reach PSNRs of 27.16dB; TV-PIDAL is associated with a high SSIM (0.856) while RED-RPnP-ULA achieves the best structural and perceptual metrics (SSIM of 0.866 and LPIPS of 0.112). For instance, TV-PIDAL exhibits a relatively high LPIPS (0.190) despite its competitive PSNR.

The Bayesian baseline TV-SPA performs poorly across all metrics (PSNR of 22.46dB, SSIM of 0.511, LPIPS of 0.206). BSD-RPnP-ULA shows intermediate performance (23.13dB of PSNR, SSIM of 0.786, LPIPS of 0.163) and the BSD-HRLwSGS variant attains a PSNR of 24.98dB with an SSIM of 0.819 and an LPIPS of 0.163, suggesting limited benefit from the BSD prior. RED-HRLwSGS provides the most favorable trade-off between fidelity and perceptual realism (30.52dB of PSNR, SSIM of 0.809, LPIPS of 0.143). Overall, HRLwSGS variants offer a combination of high reconstruction fidelity, acceptable perceptual scores and the advantage of enabling uncertainty quantification.

Figure 4 provides a qualitative comparison of the reconstructed images (first row), the pixelwise posterior standard deviations (second row) and the coverage probability maps (third row). These visualizations highlight both reconstruction fidelity and the reliability of the uncertainty quantification across the compared samplers. For the proposed RED-HRLwSGS, the reconstructed image is sharp and visually similar to the ground truth, with well-preserved edges and smooth homogeneous regions. The associated uncertainty map shows localized uncertainty mainly along structural contours, while homogeneous regions remain stable with very low variance. Its coverage map is spatially consistent and concentrated near one in most areas. The BSD-HRLwSGS reconstruction exhibits slightly higher noise levels and minor artefacts near the edges

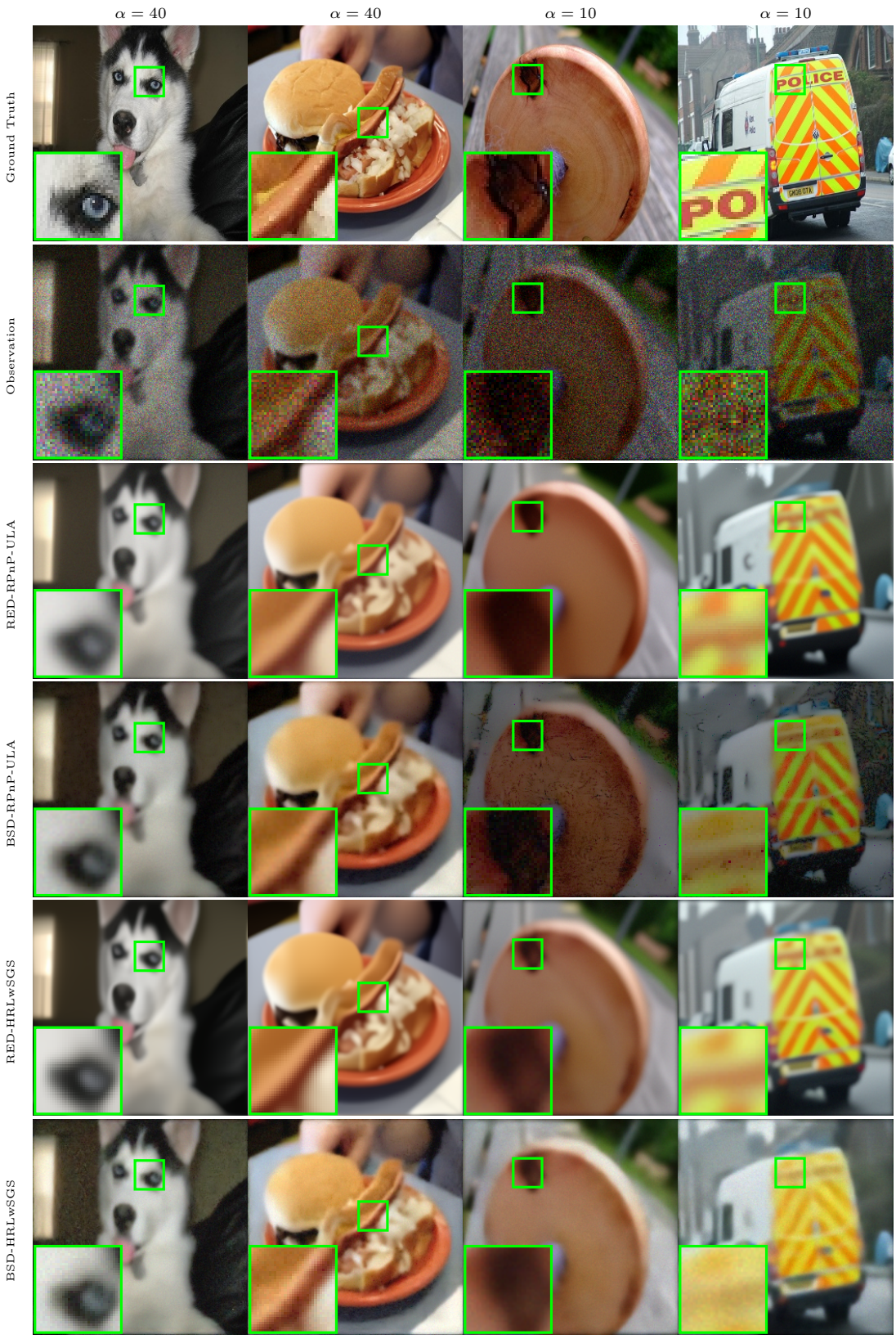

Figure 2: Deblurring experiment: visual results. Due to space contraints, the results recovered by TV-SPA and TV-PIDAL are not reproduced, as their quality is significantly inferior.

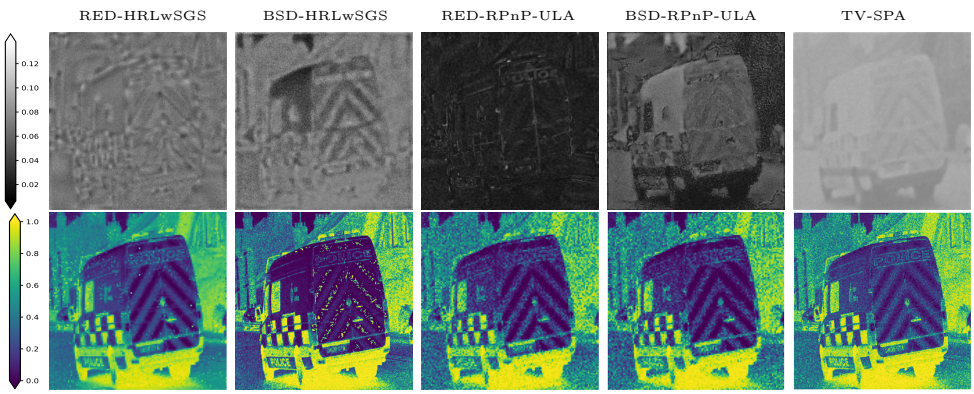

Figure 3: Deblurring experiment ($\alpha = 10$): standard deviations (top) and coverage probability maps (bottom).

compared to the RED-HRLwSGS reconstruction, as reflected by the larger posterior standard deviations. Nevertheless, the uncertainty remains structured and well localized. The coverage map corroborates these observations, showing moderately reduced confidence around the object boundaries but overall satisfactory reliability.

Table 3: PET reconstruction experiment: performance.

Method	PSNR (dB)	SSIM	LPIPS
TV-PIDAL	<u>27.16</u>	<u>0.856</u>	0.190
TV-SPA	22.46	0.511	0.206
RED-RPnP-ULA	<u>27.16</u>	**0.866**	**0.112**
BSD-RPnP-ULA	23.13	0.786	0.163
RED-HRLwSGS	**30.52**	0.809	<u>0.143</u>
BSD-HRLwSGS	24.98	0.819	0.163

Figure 5 shows the performance of the sampling-based algorithms through autocorrelation functions (ACF) and calibration results. The ACF is a key indicator of MCMC sampling efficiency: a rapid decay towards zero indicates fast-mixing chains, low correlation between successive samples, and effective exploration of the posterior distribution. In the fastest scenario, TV-SPA and BSD-RPnP-ULA methods exhibit an almost instantaneous drop in the ACF, indicating weak correlation and superior mixing. By contrast, HRLwSGS-based approaches (namely BSD-HRLwSGS and RED-HRLwSGS) show slower decay, suggesting stronger temporal dependencies. In the median case, this tendency partially reverses. BSD-RPnP-ULA still achieves the fastest ACF decay, closely followed by RED-HRLwSGS, whose ACF steadily decreases after a moderate number of iterations. The remaining methods struggle to reach decorrelation, revealing slower convergence and poorer mixing behavior. Finally, in the slowest case, the difference becomes particularly striking. RED-HRLwSGS succeeds in maintaining a fast and stable ACF decay, demonstrating efficient posterior distribution exploration, whereas all other algorithms exhibit persistent correlations, which are symptomatic of limited sampling efficiency.

Calibration measures how accurately predicted uncertainties reflect true variability. This is evaluated by comparing a target coverage c with the achieved coverage, i.e., the fraction of pixels in which the true value lies within the posterior interval at the c-level. A perfectly calibrated model will follow the diagonal line, curves below this line indicate overconfidence, while curves above it indicate underconfidence. In medical imaging, slight conservatism is favored over overconfidence. The results show that RED-HRLwSGS achieves near-ideal calibration across the full range, providing reliable uncertainty estimates. BSD-RPnP-ULA and RED-RPnP-ULA show alternating under- and over-coverage, and BSD-HRLwSGS slightly overestimates uncertainty. These results highlight RED-HRLwSGS as the most robust method for efficient sampling and

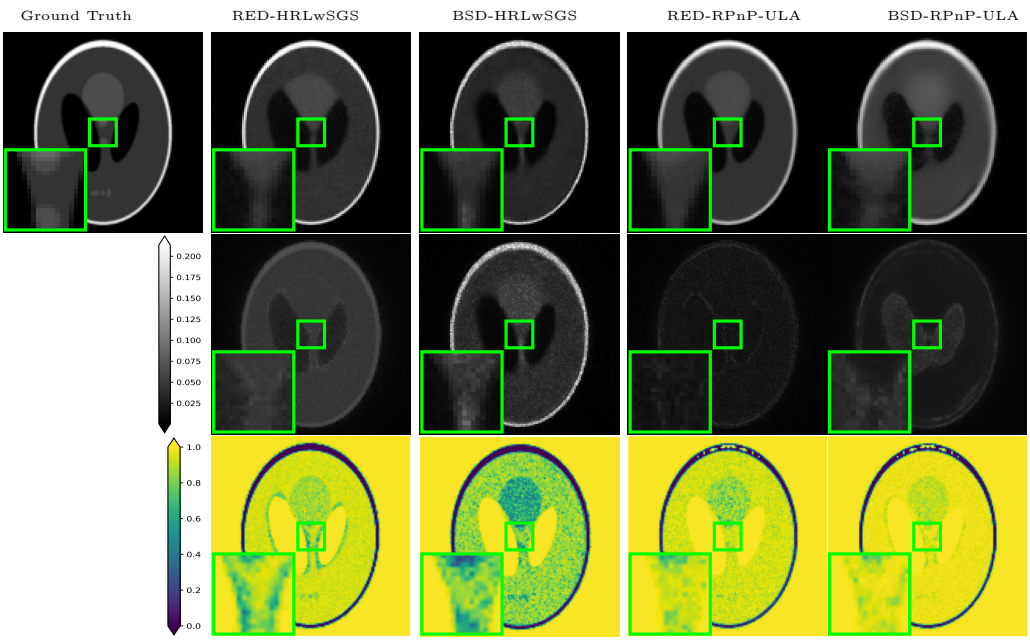

Figure 4: PET reconstruction experiment: reconstructed images (top), standard deviations (middle), and coverage probability maps (bottom).

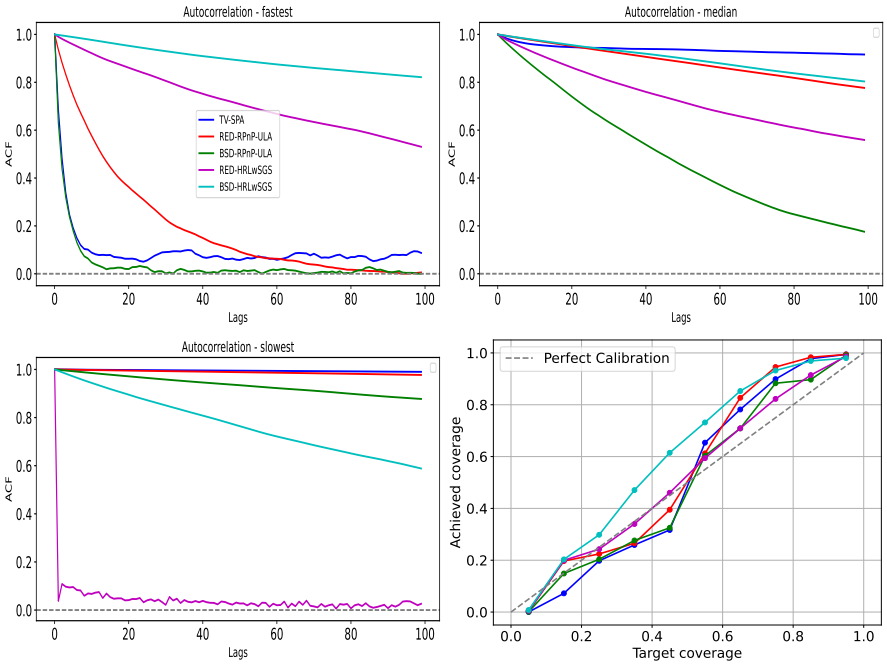

Figure 5: PET reconstruction experiment: autocorrelation and calibration curves.

trustworthy uncertainty quantification in PET imaging.

6 Conclusion

This paper introduced a novel Bayesian model for addressing Poisson inverse problems, combining exact and geometry-aware data augmentations. The model preserved the interpretability of

the Poisson likelihood while ensuring favorable conjugacy properties and strict positivity via a Bregman divergence. The resulting augmented posterior exhibited a tractable conditional structure, enabling a sampling algorithm with three fully explicit updates and one geometry-aware Langevin step. Specifically, efficient sampling was achieved using a split Gibbs sampler that incorporated a Hessian Riemannian Langevin Monte Carlo (HRLMC) step, which was versatile enough to handle a wide range of priors, particularly those defined by an explicit or easily computable score function.

Experimental results on denoising, deblurring, and positron emission tomography (PET) reconstruction tasks demonstrated that the method achieved reconstruction quality competitive with or superior to state-of-the-art optimization-based and Bayesian sampling approaches. Moreover, it provided valuable and reliable posterior uncertainty quantification, which was crucial in high-stakes applications like medical imaging.

A Proof of Theorem 1

From the joint formulation, we marginalize out $\mathbf{n}$:

$$\sum_{\mathbf{n}\in\mathbb{N}^{m\times n}} p(\mathbf{y}|\mathbf{n})\, p(\mathbf{n}|\mathbf{x}) = \sum_{\mathbf{n}\in\mathbb{N}^{m\times n}} \exp\left\{-f(\mathbf{x},\mathbf{n};\mathbf{y})\right\}.$$

Due to the indicator function, the sum restricts to $\mathbf{n}$ satisfying $\sum_j n_{ij} = y_i$:

$$\begin{aligned}\sum_{\mathbf{n}\in\mathbb{N}^{m\times n}} \exp\left\{-f(\mathbf{x},\mathbf{n};\mathbf{y})\right\} &= \exp\left\{-\sum_{i=1}^{m}\sum_{j=1}^{n}\alpha h_{ij}x_j\right\}\cdot \sum_{\mathbf{n}\in C_{\mathbf{y}}}\prod_{i=1}^{m}\prod_{j=1}^{n}\frac{(\alpha h_{ij}x_j)^{n_{ij}}}{n_{ij}!}\\ &= \exp\left\{-\sum_{i=1}^{m}\alpha\mathbf{h}_i^{\top}\mathbf{x}\right\}\cdot \prod_{i=1}^{m}\sum_{\mathbf{n}\in C_{\mathbf{y}}}\prod_{j=1}^{n}\frac{(\alpha h_{ij}x_j)^{n_{ij}}}{n_{ij}!}\end{aligned}$$

This inner sum is the multinomial coefficient corresponding to a partition of a Poisson variable of total intensity $\sum_j \alpha h_{ij}x_j$:

$$\sum_{\mathbf{n}\in C_{\mathbf{y}}}\prod_{j=1}^{n}\frac{(\alpha h_{ij}x_j)^{n_{ij}}}{n_{ij}!} = \frac{1}{y_i!}\left(\sum_{j=1}^{n}\alpha h_{ij}x_j\right)^{y_i} = \frac{1}{y_i!}\left(\alpha\mathbf{h}_i^{\top}\mathbf{x}\right)^{y_i}$$

Thus,

$$\sum_{\mathbf{n}\in\mathbb{N}^{m\times n}} \exp\left\{-f(\mathbf{x},\mathbf{n};\mathbf{y})\right\} = \exp\left\{-\sum_{i=1}^{m}\left[\alpha\mathbf{h}_i^{\top}\mathbf{x} - y_i\log(\alpha\mathbf{h}_i^{\top}\mathbf{x}) + \log(y_i!)\right]\right\},$$

which is the original Poisson likelihood $p(\mathbf{y}|\mathbf{x})$. This completes the proof.

B Denoiser-based priors

This section reviews two classes of regularization which leverage denoisers, namely regularization-by-denoising (RED) and Bregman score denoisers (BSD). These regularizations have been considered in this paper to derive particular instances of the proposed HRLwSGS algorithm, namely RED-HRLwSGS and BSD-HRLWsGS.

B.1 Regularization-by-denoising (RED)

The RED framework defines the regularization term $g(\cdot)$ using a denoising operator $\mathsf{D}_\nu : \mathbb{R}^n \to \mathbb{R}^n$ [14],

$$g(\mathbf{x}) = \frac{1}{2}\mathbf{x}^{\top}(\mathbf{x} - \mathsf{D}_\nu(\mathbf{x})) \tag{33}$$

where ν controls the denoising strength. This formulation explicitly integrates the denoising operator into the regularization term, allowing the use of advanced denoisers when solving inverse problems. When the denoiser D_ν is locally homogeous and has a symmetric Jacobian [53], the gradient of the RED regularizer (i.e. the prior score function in a Bayesian context) has a simple form

$$\nabla g(\mathbf{x}) = \mathbf{x} - \mathsf{D}_\nu(\mathbf{x}).$$

A probabilistic formulation of RED has been recently proposed [35], which enables the use of a RED potential to define a corresponding prior distribution within a Bayesian framework.

B.2 Bregman score denoiser (BSD)

The Bregman score denoiser is a denoising operator designed to operate under a Bregman geometry [18] . It generalizes the Gaussian gradient-step denoiser [54] to handle noise models associated with Bregman divergences. Formally, the denoiser $\mathcal{B}_\nu \colon \mathbb{R}^n \to \mathbb{R}^n$ is defined by

$$\mathcal{B}_\nu(\mathbf{y}) = \mathbf{y} - \left(\nabla^2 h(\mathbf{y})\right)^{-1} \cdot \nabla g_\nu(\mathbf{y}) \tag{34}$$

where $h \colon \mathbb{R}^n \to \mathbb{R}$ is a strictly convex and $\mathcal{C}^2$ function defining the Bregman geometry, and $g_\nu \colon \mathbb{R}^n \to \mathbb{R}$ is a (possibly nonconvex) potential function. Following the formulation of Hurault *et al.* [18], the noisy observations are assumed to be corrupted by a Bregman noise, resulting in the likelihood function

$$p(\mathbf{y}|\mathbf{x}) \propto \exp\left\{-\gamma d_h(\mathbf{x}, \mathbf{y})\right\} \tag{35}$$

where d_h denotes the Bregman divergence (4) generated by h. This framework encompasses classical Gaussian models as a special case, and naturally extends to non-Euclidean settings. As considered in [18], when h is the Burg's entropy and $\gamma > 1$, the distribution $p(\mathbf{y}|\mathbf{x})$ writes as a product of inverse Gamma distributions, i.e,

$$y_j|x_j \sim \mathcal{IG}(\gamma - 1, \gamma x_j), \quad j = 1, \ldots, n. \tag{36}$$

Importantly, under mild regularity assumptions, $\mathcal{B}_\nu$ also admits a variational interpretation as a Bregman proximal operator. Specifically, for any $\mathbf{y} \in \operatorname{int}\operatorname{dom}(h)$, we have

$$\mathcal{B}_\nu(\mathbf{y}) \in \arg\min_{\mathbf{x} \in \mathbb{R}^n} \left\{d_h(\mathbf{x}, \mathbf{y}) + \psi_\nu(\mathbf{x})\right\} \tag{37}$$

where ψ_ν is an explicit function derived from g_ν. The Bregman score denoiser can then be considered as a natural extension of score-based denoising to inverse problems governed by non-Euclidean geometries, such as those arising under Poisson noise. It provides a flexible and theoretically grounded framework for both variational and sampling-based image restoration methods.

In practice, Hurault *et al.* propose to define the regularization function g_ν as the following quadratic potential [18, 54]

$$g_\nu(\mathbf{x}) = \frac{1}{2}\|\mathbf{x} - \mathsf{N}_\nu(\mathbf{x})\|^2 \tag{38}$$

where $\mathsf{N}_\nu \colon \mathbb{R}^n \to \mathbb{R}^n$ denotes a deep convolutional neural network, instantiated in practice with the DRUNet architecture [48]. The gradient ∇g_ν of the regularization potential can be computed via automatic differentiation, enabling the seamless integration of complex neural architectures into the inversion pipeline.

C Properties of the HRLMC and related assumptions

This section discusses the assumptions required to ensure the convergence of the HRLMC algorithm. Both works [23] and [46] investigate the convergence properties of Langevin algorithms defined in non-Euclidean geometries, where updates are performed using mirror maps derived from strictly convex functions. These algorithms are particularly well-suited for sampling over constrained domains such as the positive orthant $\mathbb{R}^n_+$, as in the case of Poisson inverse problems.

Zhang *et al.* introduce this discretization scheme of the Riemannian Langevin diffusion and analyze its convergence in Wasserstein distance [23]. Under mild assumptions, a non-asymptotic upper-bound on the sampling error can be derived and HRLMC exhibits convergence guarantees analogous to those of LMC algorithms operating under Euclidean geometry. More recently, Li *et al.* [46] conducted an improved analysis showing that the bias of the discretized sampler vanishes as the step size approaches zero. This analysis requires only a subset of the assumptions initially stated in [23], i.e.,

(A1) Modified self concordance

$$\left\|\nabla^2\phi(\mathbf{z})^{1/2} - \nabla^2\phi(\mathbf{z}')^{1/2}\right\|_{HS} \leq \sqrt{\alpha}\,\|\nabla\phi(\mathbf{z}) - \nabla\phi(\mathbf{z}')\|_2$$

(A2) Relative strong convexity

$$m\|\nabla\phi(\mathbf{z}) - \nabla\phi(\mathbf{z}')\|_2^2 \leq \langle\nabla U(\mathbf{z}) - \nabla U(\mathbf{z}'), \nabla\phi(\mathbf{z}) - \nabla\phi(\mathbf{z}')\rangle$$

(A3) Relative Lipschitz-smoothness

$$\|\nabla U(\mathbf{z}) - \nabla U(\mathbf{z}')\|_2 \leq M\|\nabla\phi(\mathbf{z}) - \nabla\phi(\mathbf{z}')\|_2.$$

In Section 4, the HRLMC algorithm is implemented using the mirror map $\phi(\cdot)$ defined in (7) to target the distribution whose potential function $U(\cdot)$ is defined in (29). In what follows, we show that these quantities satisfy the conditions (A1)–(A3) when $g(\cdot)$ is specifically chosen as the RED potential (33). To do so, we further assume that the denoiser D_ν is L_{D}-Lipschitz and the variables z_j are bounded, i.e., it exists $\epsilon_z > 0$ and $C_z < \infty$ such that $\epsilon_z \leq z_j \leq C_z$.

(A1) Modified self-concordance The choice of the mirror map leads to

$$\nabla^2\phi(\mathbf{z})^{1/2} - \nabla^2\phi(\mathbf{z}')^{1/2} = \operatorname{diag}\left(\frac{1}{z_1} - \frac{1}{z_1'}, \ldots, \frac{1}{z_n} - \frac{1}{z_n'}\right),$$

and thus

$$\left\|\nabla^2\phi(\mathbf{z})^{1/2} - \nabla^2\phi(\mathbf{z}')^{1/2}\right\|_{HS} = \sqrt{\sum_{j=1}^{n}\left(\frac{1}{z_j} - \frac{1}{z_j'}\right)^2}.$$

Similarly

$$\|\nabla\phi(\mathbf{z}) - \nabla\phi(\mathbf{z}')\|_2 = \sqrt{\sum_{j=1}^{n}\left(\frac{1}{z_j'} - \frac{1}{z_j}\right)^2},$$

which shows that Assumption (A1) is verified with $\alpha = 1$.

(A2) Relative strong convexity The gradient of the potential writes

$$\nabla U(\mathbf{z}) = \nabla g(\mathbf{z}) + \left(\frac{1}{\rho} - 1\right)\nabla\phi(\mathbf{z}) + \frac{1}{\rho\mathbf{x}}.$$

When $g(\cdot)$ is chosen as the RED potential, its gradient is given by $\nabla g(\mathbf{z}) = \mathbf{z} - \mathsf{D}_\nu(\mathbf{z})$ and the inner product in (A2) expands as

$$\beta\langle\nabla g(\mathbf{z}) - \nabla g(\mathbf{z}'), \nabla\phi(\mathbf{z}) - \nabla\phi(\mathbf{z}')\rangle + \left(\frac{1}{\rho} - 1\right)\|\nabla\phi(\mathbf{z}) - \nabla\phi(\mathbf{z}')\|_2^2.$$

The first term decomposes as

$$\langle\mathbf{z} - \mathbf{z}', \nabla\phi(\mathbf{z}) - \nabla\phi(\mathbf{z}')\rangle - \langle\mathsf{D}_\nu(\mathbf{z}) - \mathsf{D}_\nu(\mathbf{z}'), \nabla\phi(\mathbf{z}) - \nabla\phi(\mathbf{z}')\rangle.$$

with

$$\langle \mathbf{z}-\mathbf{z}', \nabla\phi(\mathbf{z})-\nabla\phi(\mathbf{z}')\rangle = \sum_{j=1}^{n} \frac{(z_j-z_j')^2}{z_j z_j'} \geq \frac{1}{C_z^2}\|\mathbf{z}-\mathbf{z}'\|_2^2,$$

and

$$|\langle \mathsf{D}_\nu(\mathbf{z})-\mathsf{D}_\nu(\mathbf{z}'), \nabla\phi(\mathbf{z})-\nabla\phi(\mathbf{z}')\rangle| \leq L_{\mathsf{D}} C_z^2 \|\nabla\phi(\mathbf{z})-\nabla\phi(\mathbf{z}')\|_2^2.$$

Thus Assumption (A2) is ensured with

$$m = \frac{1}{\rho} - 1 + \beta\left(\frac{\epsilon_z^4}{C_z^2} - L_{\mathsf{D}} C_z^2\right).$$

The sign of m depends on the parameters ρ, β, L_{D}, ϵ_z, and C_z. Since $\epsilon_z < C_z$, we have $\frac{\epsilon_z^4}{C_z^2} < \epsilon_z^2 < C_z^2$, hence $\beta\left(\frac{\epsilon_z^4}{C_z^2} - L_{\mathsf{D}} C_z^2\right) \leq \beta C_z^2(1-L_{\mathsf{D}})$. This term is positive if $\frac{\epsilon_z^4}{C_z^4} \geq L_{\mathsf{D}}$, in which case $m \geq 0$ whenever $\rho \leq 1$. If $\frac{\epsilon_z^4}{C_z^4} \leq L_{\mathsf{D}}$, the term is negative and $m \geq 0$ requires

$$\rho \leq \frac{1}{1+\beta\left(L_{\mathsf{D}} C_z^2 - \frac{\epsilon_z^4}{C_z^2}\right)}.$$

(A3) Relative Lipschitz-smoothness Using $\|\nabla g(\mathbf{z}) - \nabla g(\mathbf{z}')\|_2 \leq (1+L_{\mathsf{D}})\|\mathbf{z}-\mathbf{z}'\|_2$ and $\|\mathbf{z}-\mathbf{z}'\|_2 \leq C_z^2\|\nabla\phi(\mathbf{z})-\nabla\phi(\mathbf{z}')\|_2$, we obtain Assumption (A3) with

$$M = \beta(1+L_{\mathsf{D}})C_z^2 + \left|\frac{1}{\rho} - 1\right|.$$

D Experimental settings and parameter values

This appendix gathers the numerical values of the parameters associated with the methods compared in the experiments. For each task and each noise level, we report the configurations adopted for the algorithms based on total variation regularization (TV-PIDAL, TV-SPA), as well as for the methods relying on deep denoisers (RPnP-ULA and HRLwSGS).

Proposed HRLwSGS algorithm – In the case of HRLwSGS, the update of the auxiliary variable $\mathbf{z}_1$ is driven by the potential $U(\mathbf{z}_1)$. The form of its gradient written in (31) highlights the balance between three contributions. The term $\beta\nabla g(\mathbf{z}_1)$ encodes the action of the regularization, with β weighting its relative importance. When β is too small, its contribution becomes negligible compared to the other terms, since $1/\rho$ dominates for small ρ, effectively suppressing regularization. Conversely, a sufficiently large β rebalances the dynamics, allowing the regularization to play its intended role, thereby improving both the mixing stability and the perceptual quality of the reconstructions. The parameter ρ controls the degree of coupling between $\mathbf{x}$, $\mathbf{z}_1$, and $\mathbf{z}_2$, that is, between the regularization and the likelihood subproblems. The chosen values reflect an empirical compromise, ensuring both adequate data fidelity and effective prior usage. Overall, the pair (β, ρ) acts as a balance between data fidelity and the strength of regularization.

For the denoising strength parameter ν, we adopt the strategy described in [35]. The parameters used in the HRLwSGS algorithm depend on the type of regularization, the considered restoration task, and the noise level α. The parameters $(\rho, \beta, \delta, \nu)$ were empirically tuned for each configuration. In the denoising task, we set $\rho = 10^{-4}$, $\beta = 4\times10^3$, $\delta = 2\times10^{-7}$, and $\nu = (20, 10)$ for $\alpha = 40$, while for $\alpha = 10$, we use $\rho = 10^{-4}$, $\beta = 4\times10^3$, $\delta = 3\times10^{-7}$, and $\nu = (25, 20)$. For deblurring, the chosen parameters are $\rho = 10^{-4}$, $\beta = 3\times10^3$, $\delta = 2.5\times10^{-7}$, and $\nu = (45, 9)$ for $\alpha = 40$, and $\rho = 10^{-4}$, $\beta = 2\times10^3$, $\delta = 4\times10^{-7}$, and $\nu = (25, 10)$ for $\alpha = 10$. In the tomography setting, we fix $\rho = 10^{-2}$, $\beta = 5\times10^2$, $\delta = 2\times10^{-3}$, and $\nu = (20, 10)$. For the BSD-LwSGS, the hyperparameters were selected similarly. In denoising, when $\alpha = 40$, we use $\rho = 1.6\times10^{-4}$, $\beta = 10^3$, $\delta = 2\times10^{-7}$, and $\nu = (0.01, 0.01)$, while for $\alpha = 10$, we set $\rho = 10^{-4}$, $\beta = 2\times10^3$, $\delta = 2\times10^{-7}$, and

$\nu = (0.05, 0.05)$. For deblurring, we choose $\rho = 10^{-4}$, $\beta = 10^3$, $\delta = 2 \times 10^{-7}$, and $\nu = (0.02, 0.01)$ for $\alpha = 40$, and $\rho = 10^{-4}$, $\beta = 10^3$, $\delta = 2 \times 10^{-7}$, and $\nu = (0.09, 0.045)$ for $\alpha = 10$. Finally, in tomography, the parameters are fixed to $\rho = 10^{-2}$, $\beta = 3 \times 10^1$, $\delta = 2 \times 10^{-4}$, and $\nu = (0.05, 0.05)$.

RPnP-ULA algorithm – For RPnP-ULA, we fixed $\varepsilon = (5/255)^2$. With RED-RPnP-ULA, for denoising, the configuration was $\beta = 1.0$ with a constant denoising level $\nu = (25, 25)$. In deblurring with $\alpha = 10$, the parameters were $\beta = 0.9$ and $\nu = (25, 10)$. For deblurring with $\alpha = 40$, we used $\beta = 10^{-2}$ and denoising levels $\nu = (13, 13)$. Finally, for PET, the configuration was $\beta = 1$ and $\nu = (8, 8)$. For BSD-RPnP-ULA, we set $\beta = 2.0$ and $\nu = (0.1, 0.05)$ for deblurring with $\alpha = 10$. For deblurring with $\alpha = 40$, we used $\beta = 2.0$ and $\nu = (0.1, 0.05)$.

TV-PIDAL algorithm – The algorithm was run with a maximum of 1000 iterations. The regularization and coupling parameters varied depending on the task: for denoising, we used $\beta = 1.0$ and $\rho = 2.0$; for deblurring, $\beta = 1.0$ and $\rho = 1.7$; and for positron emission tomography (PET), $\beta = 8$ and $\rho = 3$.

TV-SPA algorithm – The parameters for denoising were set to $\rho_1 = \rho_2 = 1.5$ and $\beta = 1.5$. In deblurring, we used $\rho_1 = \rho_2 = 1.5$, $\beta = 2.0$ for high noise, and $\rho_1 = \rho_2 = 1.5$, $\beta = 1.5$ for moderate noise. In the PET setting, the parameters were $\rho_1 = \rho_2 = 10^{-4}$ and $\beta = 4 \times 10^5$.

References

[1] R. J. Hanisch and R. L. White, "The restoration of HST images and spectra II," *Space Telescope Science Institute, Baltimore*, 1994.

[2] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. Med. Imag.*, vol. 1, no. 2, pp. 113–122, 2007.

[3] D. A. Agard and J. W. Sedat, "Three-dimensional architecture of a polytene nucleus," *Nature*, vol. 302, no. 5910, pp. 676–681, 1983.

[4] P. Sarder and A. Nehorai, "Deconvolution methods for 3-d fluorescence microscopy images," *IEEE signal processing magazine*, vol. 23, no. 3, pp. 32–45, 2006.

[5] J. R. Janesick, *Photon transfer*. SPIE press, 2007, no. PUBDB-2021-04195.

[6] M. A. Figueiredo and J. M. Bioucas-Dias, "Restoration of Poissonian images using alternating direction optimization," *IEEE Trans. Image Process.*, vol. 19, no. 12, pp. 3133–3145, 2010.

[7] 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.

[8] 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.

[9] H. H. Bauschke, M. N. Dao, and S. B. Lindstrom, "Regularizing with Bregman–Moreau envelopes," *SIAM Journal on Optimization*, vol. 28, no. 4, pp. 3208–3228, 2018.

[10] A. Lee, F. Caron, A. Doucet, and C. Holmes, "A hierarchical Bayesian framework for constructing sparsity-inducing priors," *arXiv preprint arXiv:1009.1914*, 2010.

[11] T. Park and G. Casella, "The Bayesian lasso," *J. Amer. Stat. Assoc.*, vol. 103, no. 482, pp. 681–686, 2008.

[12] N. Dobigeon, A. O. Hero, and J.-Y. Tourneret, "Hierarchical Bayesian sparse image reconstruction with application to MRFM," *IEEE Trans. Image Process.*, vol. 18, no. 9, pp. 2059–2070, 2009.

[13] 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.

[14] 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.

[15] 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.

[16] W. Marais and R. Willett, "Proximal-gradient methods for Poisson image reconstruction with BM3D-based regularization," in *Proc. IEEE Int. Workshop Comput. Adv. Multi-Sensor Adaptive Process. (CAMSAP)*. IEEE, 2017, pp. 1–5.

[17] P. Milanfar and M. Delbracio, "Denoising: A powerful building-block for imaging, inverse problems, and machine learning," *Philosophical Transactions A*, vol. 383, no. 2299, p. 20240326, 2025.

[18] S. Hurault, U. Kamilov, A. Leclaire, and N. Papadakis, "Convergent Bregman plug-and-play image restoration for Poisson inverse problems," *Adv. in Neural Information Process. Systems (NeurIPS)*, vol. 36, pp. 27 251–27 280, 2023.

[19] 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.

[20] ——, "Asymptotically exact data augmentation: Models, properties, and algorithms," *J. Comput. Graph. Stat.*, vol. 30, no. 2, pp. 335–348, 2020.

[21] S. Melidonis, P. Dobson, Y. Altmann, M. Pereyra, and K. Zygalakis, "Efficient Bayesian computation for low-photon imaging problems," *SIAM J. Imag. Sci.*, vol. 16, no. 3, pp. 1195–1234, 2023.

[22] T. Klatzer, S. Melidonis, M. Pereyra, and K. C. Zygalakis, "Efficient Bayesian computation using plug-and-play priors for Poisson inverse problems," *arXiv preprint arXiv:2503.16222*, 2025.

[23] K. S. Zhang, G. Peyré, J. Fadili, and M. Pereyra, "Wasserstein control of mirror Langevin Monte Carlo," in *Proc. Conf. Learning Theory (COLT)*. PMLR, 2020, pp. 3814–3841.

[24] L. M. Bregman, "The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming," *USSR computational mathematics and mathematical physics*, vol. 7, no. 3, pp. 200–217, 1967.

[25] P. P. Eggermont, "Maximum entropy regularization for fredholm integral equations of the first kind," *SIAM Journal on Mathematical Analysis*, vol. 24, no. 6, pp. 1557–1576, 1993.

[26] M. Burger and S. Osher, "Convergence rates of convex variational regularization," *Inverse problems*, vol. 20, no. 5, p. 1411, 2004.

[27] J. P. Burg, *Maximum entropy spectral analysis.* Stanford University, 1975.

[28] A. Banerjee, S. Merugu, I. S. Dhillon, and J. Ghosh, "Clustering with Bregman divergences," *J. Mach. Learning Research*, vol. 6, no. Oct, pp. 1705–1749, 2005.

[29] C. Févotte, N. Bertin, and J.-L. Durrieu, "Nonnegative matrix factorization with the Itakura-Saito divergence: With application to music analysis," *Neural computation*, vol. 21, no. 3, pp. 793–830, 2009.

[30] Y. C. Cavalcanti, T. Oberlin, N. Dobigeon, C. Févotte, S. Stute, M.-J. Ribeiro, and C. Tauber, "Factor analysis of dynamic PET images: beyond Gaussian noise," *IEEE Trans. Med. Imag.*, vol. 38, no. 9, pp. 2231–2241, 2019.

[31] 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)*, 2018, pp. 1–6.

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

[33] Z. Wu, Y. Sun, Y. Chen, B. Zhang, Y. Yue, and K. Bouman, "Principled probabilistic imaging using diffusion models as plug-and-play priors," in *Adv. in Neural Information Process. Systems (NeurIPS)*, vol. 37, 2024, pp. 118 389–118 427.

[34] Y. Sun, Z. Wu, Y. Chen, B. T. Feng, and K. L. Bouman, "Provable probabilistic imaging using score-based generative priors," *IEEE Trans. Comput. Imag.*, 2024.

[35] E. C. Faye, M. D. Fall, and N. Dobigeon, "Regularization by denoising: Bayesian model and Langevin-within-split Gibbs sampling," *IEEE Trans. Image Process.*, vol. 34, pp. 221–234, Jan. 2025.

[36] 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.

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

[38] Y. Vardi and L. Shepp, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. Med. Imag.*, vol. 1001, no. 2, 1982.

[39] K. Lange, R. Carson *et al.*, "EM reconstruction algorithms for emission and transmission tomography," *J. Comput. Assist. Tomogr.*, vol. 8, no. 2, pp. 306–16, 1984.

[40] M. Filipović, E. Barat, T. Dautremer, C. Comtat, and S. Stute, "PET reconstruction of the posterior image probability, including multimodal images," *IEEE Trans. Med. Imag.*, vol. 38, no. 7, pp. 1643–1654, 2018.

[41] F. Goncharov, É. Barat, and T. Dautremer, "Nonparametric posterior learning for emission tomography," *SIAM/ASA J. Uncertainty Quantification*, vol. 11, no. 2, pp. 452–479, 2023.

[42] M. D. Fall, "Bayesian Nonparametrics and Biostatistics: The case of PET imaging," *The International Journal of Biostatistics*, vol. 15, no. 2, 2019.

[43] M. D. Fall, É. Barat, C. Comtat, T. Dautremer, T. Montagu, and S. Stute, "Dynamic and clinical PET data reconstruction: A nonparametric Bayesian approach," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*. IEEE, 2013, pp. 345–349.

[44] M. D. Fall, É. Barat, C. Comtat, T. Dautremer, T. Montagu, and A. Mohammad-Djafari, "A discrete-continuous Bayesian model for emission tomography," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*. IEEE, 2011, pp. 1373–1376.

[45] A. Sitek, "Reconstruction of emission tomography data using origin ensembles," *IEEE Trans. Med. Imag.*, vol. 30, no. 4, pp. 946–956, 2010.

[46] R. Li, M. Tao, S. S. Vempala, and A. Wibisono, "The mirror Langevin algorithm converges with vanishing bias," in *Proc. Int. Conf. Algorithmic Learning Theory (ALT)*. PMLR, 2022, pp. 718–742.

[47] 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.

[48] 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.

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

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

[51] W. Van Aarle, W. J. Palenstijn, J. Cant, E. Janssens, F. Bleichrodt, A. Dabravolski, J. De Beenhouwer, K. Joost Batenburg, and J. Sijbers, "Fast and flexible x-ray tomography using the ASTRA toolbox," *Optics express*, vol. 24, no. 22, pp. 25 129–25 147, 2016.

[52] W. Van Aarle, W. J. Palenstijn, J. De Beenhouwer, T. Altantzis, S. Bals, K. J. Batenburg, and J. Sijbers, "The ASTRA toolbox: A platform for advanced algorithm development in electron tomography," *Ultramicroscopy*, vol. 157, pp. 35–47, 2015.

[53] 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.

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