

Efficient Bayesian computation for low-photon imaging problems

Savvas Melidonis¹ Paul Dobson² Yoann Altmann³ Marcelo Pereyra^{1,4}
Konstantinos C. Zygalakis^{2,4}

June 14, 2022

Abstract

This paper studies a new and highly efficient Markov chain Monte Carlo (MCMC) methodology to perform Bayesian inference in low-photon imaging problems, with particular attention to situations involving observation noise processes that deviate significantly from Gaussian noise, such as binomial, geometric and low-intensity Poisson noise. These problems are challenging for many reasons. From an inferential viewpoint, low-photon numbers lead to severe identifiability issues, poor stability and high uncertainty about the solution. Moreover, low-photon models often exhibit poor regularity properties that make efficient Bayesian computation difficult; e.g., hard non-negativity constraints, non-smooth priors, and log-likelihood terms with exploding gradients. More precisely, the lack of suitable regularity properties hinders the use of state-of-the-art Monte Carlo methods based on numerical approximations of the Langevin stochastic differential equation (SDE), as both the SDE and its numerical approximations behave poorly. We address this difficulty by proposing an MCMC methodology based on a reflected and regularised Langevin SDE, which is shown to be well-posed and exponentially ergodic under mild and easily verifiable conditions. This then allows us to derive four reflected proximal Langevin MCMC algorithms to perform Bayesian computation in low-photon imaging problems. The proposed approach is demonstrated with a range of experiments related to image deblurring, denoising, and inpainting under binomial, geometric and Poisson noise.

Key words: Low-photon imaging, computational imaging, inverse problems, Bayesian inference, Markov chain Monte Carlo methods, uncertainty quantification, proximal algorithms.

AMS subject classifications: 65C40, 68U10, 62F15, 65C60, 65J22, 62E17, 62F30, 62H10, 68W25

1 Introduction

Photon-limited imaging problems arise in modalities that measure the number of photons emitted or reflected by an object or scene of interest. Canonical examples include emission tomographic imaging [37], fluorescence microscopy [9, 37], astronomical imaging [9, 69], and single-photon light detection and ranging (LIDAR) [6, 33, 58, 68]. While these modalities have traditionally operated in moderately mild regimes, modern photon-limited imaging applications increasingly operate in low-photon to photon-starved regimes, particularly in the context of quantum-enhanced imaging technologies that exploit the particle nature of light and information in individual photons in order to go beyond the limitations of classical imaging paradigms [29].

An important characteristic of low-photon imaging problems is that their (discrete-valued) data exhibit statistical properties that deviate significantly from the Gaussian statistics encountered in many other imaging modalities. Mildly low-photon problems often exhibit Poisson or compound-Poisson statistics, whereas more challenging photon-starved

Funding: This work was supported by the UK Research and Innovation (UKRI) Engineering and Physical Sciences Research Council (EP-SRC) grants EP/V006134/1, EP/V006177/1 and EP/T007346/1, the UK Royal Academy of Engineering under the Research Fellowship Scheme (RF201617/16/31) and the Leverhulme Trust (RF/2020-310).

¹School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh, EH14 4AS, Scotland, UK (sm2041@hw.ac.uk, m.pereyra@hw.ac.uk).

²School of Mathematics, University of Edinburgh, Edinburgh, EH9 3FD, Scotland, UK (pdobson@ed.ac.uk, kzygalak@exseed.ed.ac.uk).

³School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh, EH14 4AS, Scotland, UK (y.altmann@hw.ac.uk).

⁴Maxwell Institute for Mathematical Sciences, Bayes Centre, 47 Potterrow, Edinburgh, Scotland, UK.

problems exhibit approximately Bernoulli/binomial or geometric data [4, 5]. Poisson-type statistics arise from using sensors capable of discriminating photon-detection events within a given time period, assuming detector dead times can be neglected [5, 58, 68]. The more challenging Bernoulli, binomial and geometric statistics arise from sensors that cannot accurately quantify photons beyond the first detection, such as Single-Photon Detectors [29] and in particular Single-Photon Avalanche Diodes (SPADs) [5, 25, 39, 58].

Accurate recovery of images from low-photon data is difficult for many reasons. First, from an inferential perspective, the data have very limited information about the solution. This leads to image estimation problems that are severely ill-conditioned or ill-posed, and which exhibit high levels of intrinsic uncertainty. Second, from a computational perspective, the data fidelity terms associated with Poisson, binomial and geometric statistics have poor regularity properties (e.g., exploding gradients) and involve non-negativity constraints on the solution space. This makes it difficult to use gradient-based computation algorithms that scale efficiently to large problems [24, 32]. These difficulties can be mitigated to some extent by using tools from proximal optimisation as proposed in [27]. Some of these ideas will be revisited in this paper in the context of Bayesian computation, and combined with new strategies for addressing these difficult problems.

The literature considers three main frameworks to solve low-photon imaging problems: Bayesian statistics [4, 5, 46, 47, 72, 76], machine learning [3, 17, 45, 64, 65], and the variational framework [19, 27, 34]. This paper focuses on the Bayesian statistical framework because it enables the use of Bayesian decision theory to analyse these difficult imaging problems [60]. This theory allows deriving estimators, quantifying uncertainty, automatically adjusting unknown model parameters, and performing model selection in the absence of ground truth, all of which are relevant in problems with high intrinsic uncertainty [12, 24, 53, 71].

Bayesian computation for low-photon imaging problems is usually addressed in the following three ways: proximal optimisation, variational inference (VI), e.g., variational Bayes (VB) [55] or Expectation-Propagation (EP) [40, 74, 75], and Markov chain Monte Carlo (MCMC) simulation [61]. Proximal optimisation strategies are mainly used for maximum-a-posteriori (MAP) estimation in Bayesian models that are log-concave, where MAP estimation is a convex problem that can be solved efficiently by using provably convergent and scalable algorithms [19, 27, 34]. Some basic forms of uncertainty visualisation and quantification can also be formulated as convex optimisation problems [53, 59]. Despite their theoretical rigour, these methods can only support very limited inferences and log-concave priors. Some recent works consider generalisations to data-driven plug-and-play (PnP) priors that can deliver more accurate results, usually at the expense of weaker theoretical guarantees [3, 17, 45, 64, 65]. Moreover, to support more advanced inferences without significantly increasing computing effort, VI low-photon imaging methods approximate the Bayesian posterior distribution by a tractable surrogate model (e.g., a Gaussian model) useful for approximate inference. VI are often computationally efficient and deliver accurate point estimators, but they are by nature highly problem-specific [46, 47] and may be unreliable for uncertainty quantification because of local convergence issues and approximation errors [55]. EP methods can suffer from similar limitations, and their computation cost is higher, but they can provide more accurate posterior estimates. Conversely, modern MCMC methods provide a general Bayesian computation strategy to perform inference in imaging problems that can support a wide range of inferences and models, with detailed convergence guarantees, albeit at computational cost that is potentially significantly higher than VI and optimisation strategies. To the best of our knowledge, very few works consider MCMC methods for low-photon imaging problems. For Poisson imaging problems, the state of the art MCMC methods are the Poisson hierarchical Gibbs sampler (PHGS) given in [8], the primal-dual preconditioned Crank-Nicolson Langevin MCMC algorithm [76], and the split-and-augmented (SPA) Gibbs sampler [72]. For Bernoulli, binomial and geometric data, the state of the art are the Gibbs samplers described in [4, 5]. These MCMC methods are either highly model-specific, they do not account for the non-negativity constraint in practice or they do not benefit from important recent developments such as acceleration [54].

The aim of this paper is to propose a general computationally efficient MCMC methodology to perform Bayesian inference in low-photon imaging problems, with special attention to models that are log-concave. Similarly to modern MCMC methods for Gaussian imaging problems, the proposed strategy is based on approximations of the Langevin stochastic differential equation (SDE) [18, 24, 54], which we modify in key ways to enable computation for low-photon imaging problems with non-negativity constraints. More precisely, in this paper we develop a new approach based on reflected Langevin SDEs, for which we develop convergence theory and a series of MCMC algorithms.

The remainder of this paper is organised as follows: Section 2 defines notation and introduces the class of problems

and Bayesian models considered. Section 3 recalls the Langevin SDE, underlines the several issues preventing their direct application to Bayesian low-photon imaging problems, and presents the proposed Reflected Langevin SDE as well as its theoretical properties. Following on from this, Section 4 then presents the proposed MCMC samplers and discusses connections with state of the art MCMC algorithms for models with Gaussian likelihoods [24, 31, 52, 54]. Section 5 demonstrates the proposed approach on a range of challenging experiments related to image deblurring with Poisson data, denoising with binomial data, and inpainting with geometric data, including comparisons with MAP estimation by convex optimisation [27] and the SPA sampler [72].

2 Problem Statement

2.1 Observation Models

We consider imaging inverse problems where we aim to estimate an unknown image $x \in \mathbb{R}^n$ from an observation $y \in \mathbb{R}^m$ related to x through a statistical model with likelihood function $p(y|x)$. In the problems we are interested in, the recovery of x from y is ill-posed or severely ill-conditioned¹ and involves non-negativity constraints on x . The typical scenario is image recovery from Poisson distributed observations in imaging applications when the mean photon arrival rate is low. In this case, we acquire a set of discrete photon measurements $y = [y_1, \dots, y_m] \in \mathbb{N}_0^m = \{0, 1, \dots\}^m$ related to the ground truth image x through the statistical model

$$y|x \sim \mathcal{P}(Ax) , \quad (2.1)$$

where $A \in \mathbb{R}_+^{m \times n}$, $\mathbb{R}_+^{m \times n} = \{x \in \mathbb{R}^{m \times n} : \min_i x_i \geq 0\}$, is a linear operator which models the physical properties of the observation process. A common difficulty in applications is that AA^T is often poorly conditioned or rank deficient. The respective negative log-likelihood is given by

$$f_{\mathcal{P}}(x) = \sum_{i=1}^m \left[(Ax)_i - y_i \log((Ax)_i) + \log(y_i!) + \iota_{\mathbb{R}_+^n}(x) \right] , \quad (2.2)$$

where $\iota_{\mathbb{R}_+^n}(\cdot)$ is the indicator function on $\mathbb{R}_+^n = \{x \in \mathbb{R}^n : \min_i x_i > 0\}$ that requires x to be positive. Because of the presence of A and the logarithm, the Poisson log-likelihood is not quadratic and often non-separable. Also, $x \mapsto \nabla f_{\mathcal{P}}(x)$ is not necessarily globally Lipschitz continuous. As a result, the estimation of x from y is highly challenging.

In some single-photon imaging applications, the Poisson likelihood assumption does not hold and the relation between the unknown image and the observations is better modeled by binomial or geometric models [5, 39, 58, 68]. For instance, if the detector dead time² can't be neglected but can be reset at the end of a given period (referred to as repetition period here), then the measurements $y = [y_1, \dots, y_m] \in \mathbb{N}_0^m$ can be related to the ground image $x \in \mathbb{R}^n$ by the statistical model [5, 58, 68]

$$y|x, t \sim \mathcal{Bin}(t, 1 - e^{Ax}) , \quad (2.3)$$

where $\mathcal{Bin}(\cdot, \cdot)$ stands for the product of independent binomial distributions, $t = [t_1, \dots, t_m] \in \mathbb{N}^m$ gathers the numbers of repetition periods for each detector, and A is again a linear operator. Then negative log-likelihood is thus given by

$$f_{\mathcal{B}}(x) = \sum_{i=1}^m \left[-y_i \log(1 - e^{-(Ax)_i}) + (t_i - y_i)(Ax)_i + \iota_{\mathbb{R}_+^n}(x) \right] . \quad (2.4)$$

In the first-photon imaging context [39] where the recordings are interrupted for each pixel after the first detection, the number of repetition periods required to record the first detection event in each pixel follows a geometric distribution specified (in vector form) by

$$t|x, y = 1 \sim \mathcal{Geo}(1 - e^{-Ax}) , \quad (2.5)$$

¹The estimation problem is said to be ill-posed or ill-conditioned when it does not admit a unique solution or it admits a solution that is not stable w.r.t. small perturbations in y .

²A detector's dead time period corresponds to the period following a photon detection during which additional photons cannot be detected.

and the negative log-likelihood is given by

$$f_{\mathcal{G}}(x) = \sum_{i=1}^m \left[(t_i - 1)(Ax)_i - \log \left(1 - e^{-(Ax)_i} \right) + \iota_{\mathbb{R}_{++}^n}(x) \right] . \quad (2.6)$$

Note that in (2.5), the variables treated as observations are in t , while in (2.3), the observations are in y and t is only a fixed experimental parameter. In extreme photon-starved regimes, the binomial and geometric random noise processes are highly non-linear, have very poor signal-to-noise properties and involve log-likelihoods that are not Lipschitz continuous over the non-negative domain. These models differ from the more classical Poisson model but share similar challenges, which is why we consider them as exemplar observation models in this work (to illustrate that our general samplers are not likelihood-specific).

2.2 Bayesian inference for imaging applications

The identifiability issues of the aforementioned likelihoods imply that additional information is needed to reduce the uncertainty about x and deliver accurate well-posed solutions. Within the Bayesian framework, regularization is introduced via the prior distribution over x , denoted as $p(x)$. This distribution promotes expected properties about x (e.g. sparsity, piece-wise regularity or smoothness). Inferences are then based on the posterior distribution, given via Bayes' theorem [60] by

$$\pi(x) \triangleq p(x|y) = \frac{p(y|x)p(x)}{\int_{\mathbb{R}^N} p(y|x)p(x)dx} .$$

In this work, we focus on log-concave models of the form

$$\pi(x) = \frac{e^{-f_y(x) - g(x) - \iota_{\mathbb{R}_{++}^n}(x)}}{\int_{\mathbb{R}^d} e^{-f_y(x) - g(x) - \iota_{\mathbb{R}_{++}^n}(x)} dx} , \quad (2.7)$$

where $f_y : \mathbb{R}^n \rightarrow \mathbb{R}$ (corresponding to the data log-likelihood) and $g : \mathbb{R}^n \rightarrow (-\infty, +\infty]$ are two lower bounded functions and $\iota_{\mathbb{R}_{++}^n}(\cdot)$ is the indicator function on \mathbb{R}_{++}^n which is convex and guarantees that the positivity constraint is satisfied. The two functions f_y and g satisfy the following conditions:

1. f_y is convex and can be expressed as $f_y(x) = F(Ax)$ for some $F : \mathbb{R}_{++}^n \rightarrow \mathbb{R}$ which is Lipschitz continuous on the set $\{x \in \mathbb{R}_{++}^n : x_i > b \text{ for all } i\}$ for any $b > 0$.
2. g is proper, convex and lower semi-continuous, but potentially non-smooth.

Calculating π exactly is often not feasible in imaging because of dimensionality of x . Instead, summaries such as Bayesian estimators posterior probabilities or expectations are used to deliver some important information about π . In particular, for log-concave models as (2.7), recent works often rely on the MAP estimator [53, 55]

$$\hat{x}_{MAP} = \operatorname{argmax}_x \pi(x) = \operatorname{argmin}_x \{f_y(x) + g(x) + \iota_{\mathbb{R}_{++}^n}(x)\} . \quad (2.8)$$

which can efficiently be computed, even in high dimensions, by using proximal convex optimization techniques [16, 20]. For instance, to solve the minimization problem of the form (2.8) under a Poisson likelihood, an ADMM variant was considered in [27] which does not assume Lipschitz continuity in the gradient while a quadratic approximation of the Poisson likelihood that turns the original problem into a constrained l_2 denoising problem was proposed in [34].

Log-concavity also plays a central role in the calculation of moment-based estimators such as the minimum mean square error (MMSE) estimator since it guarantees the existence of all posterior moments. It also enables the efficient calculation of other advanced quantities related to uncertainty quantification, calibration of model parameters, and model selection in the absence of ground truth [12, 24, 53, 71]. More precisely, to calculate such estimators or perform more advanced analyses, it is necessary to use high-dimensional MCMC methods to simulate samples from

the posterior π followed by Monte Carlo integration [32]. In recent years, highly efficient MCMC methods have been developed to perform Bayesian inference for imaging sciences based on techniques from proximal optimization. These methods are known as proximal MCMC methods [18, 24, 31, 52, 54, 72] and are useful for Bayesian imaging since they are easy to implement, have significantly reduced computational time (compared to traditional Gibbs and Metropolis-Hastings samplers) and offer theoretical convergence guarantees. However, they have been implemented for Gaussian imaging problems and cannot directly be applied to low-photon imaging problems that involve poor regularity properties and non-negativity constraints. In this work, we will present how proximal MCMC methods can be adjusted to tackle the challenging class of models (2.7).

3 Langevin SDEs for sampling

Before presenting the proposed reflected SDE we first recall how to use the Langevin SDE for sampling from distributions of the form

$$\pi(x) = \frac{e^{-U(x)}}{\int_{\mathbb{R}^N} e^{-U(x)} dx}, \quad (3.1)$$

for some function $U : \mathbb{R}^n \rightarrow \mathbb{R}$ with e^{-U} integrable. If π is given by (2.7) then $U(x) = f_y(x) + g(x) + \iota_{\mathbb{R}_{++}^n}(x)$. Before we discuss the difficulties in such a model, we first consider the case where U is Lipschitz continuously differentiable on \mathbb{R}^n . As is shown in [44] the class of SDEs which target π is given by

$$\begin{aligned} dX_t &= [-(S(X_t) + J(X_t))\nabla_x U(X_t) + \Gamma(X_t)]dt + \sqrt{2S(X_t)}dW_t, \\ \Gamma_i(x) &= \sum_{j=1}^n \partial_{x_j} (S_{ij}(x) + J_{ij}(x)), \end{aligned} \quad (3.2)$$

for some positive semidefinite diffusion matrix $S(x)$ and skew-symmetric matrix $J(x)$. Here, W_t is a n -dimensional Brownian motion. We concentrate on the overdamped Langevin SDE which corresponds to taking $J = 0$ and $S(x) = I_n$, that is

$$dX_t = -\nabla_x U(X_t)dt + \sqrt{2}dW_t. \quad (3.3)$$

Another choice which is popular for sampling algorithms is the underdamped Langevin SDE which corresponds to taking $X_t = (q_t, p_t)$, $U(q, p) = U(q) + |p|^2/(2u)$,

$$S = \begin{pmatrix} 0 & 0 \\ 0 & \gamma u \end{pmatrix}, \quad J = \begin{pmatrix} 0 & -u \\ u & 0 \end{pmatrix},$$

for some constants $u, \gamma > 0$, that is

$$\begin{aligned} dq_t &= p_t dt \\ dp_t &= -u\nabla_x U(q_t)dt - \gamma p_t dt + \sqrt{2\gamma u}dW_t. \end{aligned} \quad (3.4)$$

3.1 Sampling in a domain

In the case where $U(x) = f_y(x) + g(x) + \iota_{\mathbb{R}_{++}^n}(x)$ we no longer have that the SDE (3.3) or even the more general form of SDE (3.2) admits π as an invariant measure. Indeed there are several issues: (i) the function g does not need to be differentiable, so $\nabla_x g$ is not well-defined; (ii) π is only positive on a subset of the domain but the SDE (3.3) is irreducible so any invariant measure of (3.3) admits a positive density on \mathbb{R}^n ; (iii) the function $\nabla_x f$ is not Lipschitz continuous and therefore the SDE may not be well-posed. Let us discuss each of these issues in turn.

∇g is not well defined. Since g is not differentiable, we approximate g by a function g^λ which is obtained by applying the Moreau-Yosida (MY) envelope on g , that is,

$$g^\lambda(x) = \min_{y \in \mathbb{R}^n} \{g(y) + \frac{1}{2\lambda} \|y - x\|^2\} \quad (3.5)$$

where $g^\lambda \rightarrow g$ as $\lambda \rightarrow 0$. It is shown in [63, Proposition 12.19] that g^λ is Lipschitz continuously differentiable with Lipschitz constant λ^{-1} and gradient

$$\nabla g^\lambda(x) = \frac{1}{\lambda} (x - \text{prox}_g^\lambda(x)) , \quad (3.6)$$

where

$$\text{prox}_g^\lambda(x) = \text{argmin}_{y \in \mathbb{R}^n} \{g(y) + \frac{1}{2\lambda} \|y - x\|^2\}.$$

The use of (3.6) to approximate terms such as the function g is investigated in [24].

Positivity constraint. Since the SDE (3.3) is non-degenerate, if it admits an invariant measure, then this measure must have a strictly positive density. Therefore we can not use the SDE (3.3) to sample from a measure which is supported on \mathbb{R}_+^n . In principle, by taking the matrix $S(x)$ to be degenerate one can construct an SDE which has an invariant measure supported on the manifold \mathbb{R}_+^n , however this introduces further complications such as not having a unique invariant measure (see [13] for a discussion of the invariant measures of degenerate SDEs). Therefore, we propose to introduce reflections on the boundary of \mathbb{R}_+^n and construct a reflected SDE (RSDE) which targets an invariant measure on \mathbb{R}_+^n .

∇f not well-defined. The function $\nabla_x f_y$ is not necessarily Lipschitz-continuous and is not necessarily finite on the boundary of \mathbb{R}_+^n . Thus, it is not clear that the SDE (3.3) is well-defined on \mathbb{R}_+^n (even assuming at this stage that reflections are included in a suitable manner and that g is replaced by g^λ). Even if the SDE was well-defined, these remain challenging for its numerical approximation since commonly used discrete time approximations need not be stable in the absence of a Lipschitz condition (see [48]). To address this difficulty, we propose to introduce the approximation f_y^b defined by $f_y^b(x) = F(Ax + b)$ of f_y . Note that by the assumptions of F , the function $\nabla_x f_y^b$ is globally Lipschitz continuous. For these reasons, we propose to approximate π by the measure $\pi^{\lambda,b}$ which is supported on \mathbb{R}_+^n and defined for $x \in \mathbb{R}_+^n$ by

$$\pi^{\lambda,b}(x) = \frac{e^{-U^{\lambda,b}(x)}}{\int_{\mathbb{R}_+^n} e^{-U^{\lambda,b}(\tilde{x})} d\tilde{x}}, \quad U^{\lambda,b}(x) = f_y^b(x) + g^\lambda(x). \quad (3.7)$$

In the following lemma we show that $\pi^{\lambda,b}$ is an approximation of π that has interesting asymptotic properties as λ and b vanish.

Lemma 3.1. *Suppose that the conditions of Section 2.2 hold. Let f_y^b and g^λ be as above then with $\pi^{\lambda,b}$ given by (3.7) and π defined by (2.7), we have*

$$\lim_{\lambda \rightarrow 0, b \rightarrow 0} \|\pi - \pi^{\lambda,b}\|_{TV} = 0.$$

Proof. The proof is deferred to Appendix A. □

3.2 Reflected SDE

In Section 3.1 we discussed the issue of sampling from the distribution π given by (2.7) and introduced an approximation $\pi^{\lambda,b}$ defined on the domain \mathbb{R}_+^n . In this section we introduce a RSDE which has $\pi^{\lambda,b}$ as an invariant distribution and discuss its properties. In Section 3.3 we verify that this RSDE has the desired invariant distribution and converges exponentially.

We consider the following RSDE defined on the convex domain $\mathbb{R}_+^n \subseteq \mathbb{R}^n$

$$dX_t = -\nabla_x U^{\lambda,b}(X_t) dt + \sqrt{2} dW_t + d\kappa_t, \quad (3.8)$$

here W_t is a n -dimensional Brownian motion, κ_t is local time which only increases on $\partial\mathbb{R}_+^n$, $U^{\lambda,b} : \mathbb{R}_+^n \rightarrow \mathbb{R}$ and is C^2 with $\nabla_x U^{\lambda,b}$ being globally Lipschitz continuous. Observe that this RSDE is non-negative as it takes values in \mathbb{R}_+^n . We define the semigroup, \mathcal{P}_t , corresponding to this RSDE by

$$\mathcal{P}_t \varphi(x) = \mathbb{E}[\varphi(X_t^x)] \quad \text{for all } \varphi \in B_b(\mathbb{R}_+^n), x \in \mathbb{R}_+^n, t \geq 0.$$

For background on semigroups and operators we refer the reader to [26]. We (formally) define the generator corresponding to this semigroup to be

$$\mathcal{L}\varphi(x) = -\langle \nabla_x U^{\lambda,b}(x), \nabla_x \varphi(x) \rangle + \Delta_x \varphi(x)$$

for $\varphi : \mathbb{R}_+^n \rightarrow \mathbb{R}$ sufficiently smooth. Under the above assumptions it is shown in [70, Theorem 4.1] there is a (pathwise) unique strong solution to (3.8). We emphasise that the domain does not need to be smooth provided that it is convex, in particular the domain \mathbb{R}_+^n is permitted. Let us expand on what we mean by local time in this context, $\kappa : [0, \infty) \rightarrow \mathbb{R}^n$ is a càdlàg function with bounded variation, $\kappa_0 = 0$, the set $\{t \geq 0 : X_t^i > 0 \text{ for all } i\}$ has measure zero with respect to the measure $d|\kappa|$ and we can write

$$\kappa_t = \int_0^t \hat{n}(s) d|\kappa_s|$$

where the function $\hat{n}(s)$ is a unit normal vector at X_s for almost all s with respect to the measure $d|\kappa|$. Since the domain needs not to be smooth we shall explain what we mean by a normal vector. A normal vector \hat{n} at x is any unit vector which is orthogonal to some supporting hyperplane H of \mathbb{R}_+^n . A supporting hyperplane of \mathbb{R}_+^n is a hyperplane such that \mathbb{R}_+^n is contained in one of the two closed half-spaces bounded by H and \mathbb{R}_+^n has at least one boundary point on the hyperplane. In this setting, the set of (inward) normal vectors at 0 is the set

$$\{\hat{n} = (\hat{n}_1, \dots, \hat{n}_n) : |\hat{n}| = 1, n_i \geq 0 \text{ for all } i\}.$$

If the set \mathbb{R}_+^n had a smooth boundary and the coefficients are suitably smooth (see [28] for example), it is well known that the semigroup $\mathcal{P}_t\varphi$ corresponding to X_t solves a Neumann boundary value problem. As the case we are interested in has a non-smooth boundary for $n > 1$ more care is required. This domain is investigated in [22] under the assumption that the drift coefficient $\nabla_x U^{\lambda,b} \in C^1(\mathbb{R}_+^n)$ is globally Lipschitz, and it is shown that the semigroup $\mathcal{P}_t\varphi \in C^1(\mathbb{R}_+^n)$ for $\varphi \in C_b(\mathbb{R}_+^n)$ and satisfies the Neumann boundary conditions $\partial_{x_i} \mathcal{P}_t\varphi(x) = 0$ whenever $x_i = 0$. Therefore the semigroup $u_t(x) = \mathcal{P}_t\varphi$ solves the Neumann boundary value problem:

$$\begin{cases} \partial_t u_t(x) = \mathcal{L}u_t(x), & x \in \mathbb{R}_+^n, \\ \frac{\partial u_t}{\partial x_i}(x) = 0, & \text{for any } x \in \mathbb{R}_+^n \text{ with } x_i = 0, \\ u_0(x) = \varphi(x), & x \in \mathbb{R}_+^n. \end{cases} \quad (3.9)$$

In Proposition A.1 (see Appendix A) we apply the results of [22] to describe the smoothness of the semigroup, indeed for any $\varphi \in C_b(\mathbb{R}_+^n)$ we have that $\mathcal{P}_t\varphi \in C^1(\mathbb{R}_+^n)$ and by ellipticity we have $\mathcal{P}_t\varphi \in C^\infty(\mathbb{R}_+^n)$.

3.3 Exponential Ergodicity of RSDEs

In order to show exponential ergodicity we shall apply [50, Theorem 6.1]. Before giving the details of this theorem we provide an informal summary of the main assumption, the existence of a Lyapunov function. For our purposes, a Lyapunov function is a function $V : \mathbb{R}_+^n \rightarrow \mathbb{R}_+$ which is positive, C^2 , $V(x) \rightarrow \infty$ as $x \rightarrow \infty$ and satisfies for some $C_1 > 0, C_2 \in \mathbb{R}$

$$\mathcal{L}V \leq -C_1V + C_2 \quad (3.10)$$

$$\frac{\partial V}{\partial x_i}(x) = 0 \quad \text{whenever } x_i = 0, \text{ for some } i \in \{1, \dots, n\}. \quad (3.11)$$

The first condition (3.10) is a sufficient condition to have that \mathcal{P}_tV is bounded in t hence that the process is non-explosive and can not spend too much time in the tails. The second condition (3.11) ensures that the boundary condition arising from the non-negativity constraint is satisfied. This condition can be viewed as a necessary condition to have that V is in the domain $D(L)$ of the generator. However as V is unbounded it is not clear that V belongs to the domain of the generator, indeed we do not even have that \mathcal{P}_tV is well-defined at this stage. For this reason

we introduce a sequence of operators \mathcal{L}_ℓ defined on functions from some bounded set O_ℓ to \mathbb{R} . We construct such functions by considering the process $X_{t \wedge T^\ell}$ where T^ℓ is the first exit time of O_ℓ , i.e. the process is equal to X_t up until the first exit time of O_ℓ and then remains constant. Set \mathcal{L}_ℓ to the generator of this process. Since O_ℓ is bounded we have that $V \in C_b(O_\ell)$ and that V belongs to the domain of \mathcal{L}_ℓ . We now give the more precise statement of this assumption, which is phrased in terms of the extended generator. We assume there is a sequence of open (as a subset of \mathbb{R}_+^n) and bounded sets O_ℓ such that $O_\ell \subseteq O_{\ell+1}$ for all $\ell \geq 1$ and $\bigcup_{\ell=1}^\infty O_\ell = \mathbb{R}_+^n$. Let T^ℓ be the first exit time of O_ℓ and set \mathcal{L}_ℓ to be the extended generator of the process $\{X_{t \wedge T^\ell}^x\}_{t \geq 0}$, that is a function $\varphi : \mathbb{R}_+^n \times \mathbb{R}_+ \rightarrow \mathbb{R}$ is in the domain of \mathcal{L}_ℓ if there exists a function $\psi : \mathbb{R}_+^n \times \mathbb{R}_+ \rightarrow \mathbb{R}$ such that for each $x \in \mathbb{R}_+^n, t > 0$

$$\mathbb{E}[\varphi(X_{t \wedge T^\ell}^x, t)] = \varphi(x, 0) + \mathbb{E} \left[\int_0^t \psi(X_{s \wedge T^\ell}^x, s) ds \right], \quad \mathbb{E} \left[\int_0^t |\psi(X_{s \wedge T^\ell}^x, s)| ds \right] < \infty,$$

and we write $\mathcal{L}_\ell \varphi = \psi$. In our setting we can set $O_\ell = [0, \ell]^n$ and the domain of \mathcal{L}_ℓ contains all C^2 functions with $\partial_{x_i} f(x) = 0$ whenever $x_i = 0$ by Itô's formula. Now we shall state the conditions we require to apply [50, Theorem 6.1] which is [50, (CD3)] restated in our notation.

Hypothesis 3.2. *There exists a function $V : \mathbb{R}_+^n \rightarrow \mathbb{R}_+$ which is positive, measurable, $V(x) \rightarrow \infty$ as $x \rightarrow \infty$ and for some $C_1 > 0, C_2 < \infty$*

$$\mathcal{L}_\ell V(x) \leq -C_1 V(x) + C_2, \quad x \in O_\ell. \quad (3.12)$$

Theorem 3.3. *([50, Theorem 6.1]) Suppose that X_t is a right process, and that all compact sets are petite (see [50] for definitions). If Hypothesis 3.2 holds, then X_t^x admits a unique invariant measure π and there exists $\beta < 1$ and $B < \infty$ such that*

$$\sup_{|\varphi| \leq 1+V} |\mathcal{P}_t \varphi(x) - \pi(\varphi)| \leq B(1+V(x))\beta^t, \quad t \geq 0, x \in \mathbb{R}_+^n.$$

We can now apply this theorem for the RSDE (3.8).

Theorem 3.4. *Let \mathcal{P}_t be the semigroup corresponding to the RSDE (3.8). Assume that $\nabla_x U^{\lambda, b}$ is continuously differentiable and globally Lipschitz. Suppose that $U^{\lambda, b}$ is convex and $e^{-U^{\lambda, b}(x)}$ is integrable over \mathbb{R}_+^n then X_t^x admits a unique invariant measure π and there exist $V : \mathbb{R}_+^n \rightarrow [0, \infty), \beta < 1$ and $B < \infty$ such that*

$$\sup_{|\varphi| \leq 1+V(x)} |\mathcal{P}_t \varphi(x) - \pi(\varphi)| \leq B(1+V(x))\beta^t, \quad t \geq 0, x \in \mathbb{R}_+^n.$$

Moreover, the invariant measure π is given by (3.7) for any $x \in \mathbb{R}_+^n$.

Proof. The proof is deferred to Appendix A. □

Remark 3.5. *Since all the models considered in Section 2.1 correspond to convex negative log-likelihoods, by imposing a log concave prior which is integrable we have that the assumptions of Theorem 3.4 are satisfied and hence the corresponding RSDE is exponentially ergodic in each of these cases. Note that the proof of the above theorem does not explicitly require that $U^{\lambda, b}$ is convex but only that there exist $\alpha > 0$ and $R > 0$ such that for $|x| \geq R$,*

$$\langle x, \nabla U^{\lambda, b}(x) \rangle \geq \alpha |x|. \quad (3.13)$$

It is shown in [7, Lemma 2.2] that any convex C^1 -function $U^{\lambda, b}$ which is integrable must satisfy (3.13), however as (3.13) only needs to hold outside a compact set it is clear that there are also non-convex functions $U^{\lambda, b}$ for which Theorem 3.4 holds.

3.4 Discrete time approximations of RSDEs

In Section 3.2 we introduced an RSDE which is exponentially ergodic and has the measure π as an invariant distribution however this leads to the question of how do we approximate this process. In this section we discuss different approaches to approximate the local time term in the RSDE. There are three approximations we will discuss: (i) Penalty scheme; (ii) Projection scheme; (iii) Reflection scheme. Note there are more complicated schemes such as

the half-plane approximation (see [30] for details) however, as this is more expensive and leads to the same order of accuracy as the reflection scheme, we do not consider it.

The penalty scheme replaces the local time term in (3.8) by $\beta_\varepsilon(X_t)dt$ where $\beta_\varepsilon(x) = (x^+ - x)/\varepsilon$ and x^+ is the vector whose i -component is $\max(x_i, 0)$. Then for each fixed ε we have an SDE defined on the whole space \mathbb{R}^n . An alternative way to view this scheme is to consider the measure π given by (2.7) and apply the proximal operator to the indicator function $\iota_{\mathbb{R}_+^n}$ in the same way as we approximate g . This results in a measure $\pi^{\lambda, b, \varepsilon}$ supported on \mathbb{R}^n which can then be viewed as the invariant measure of the Langevin SDE

$$dX_t = -\nabla_x f^b(X_t)dt - \nabla_x g^\lambda(X_t)dt + \beta_\varepsilon(X_t)dt + \sqrt{2}dW_t.$$

This SDE can then be approximated by standard numerical schemes for SDEs such as the Euler-Maruyama scheme. In order to have a scheme which converges to the RSDE as the time step tends to zero we need to let ε be a function of the time step h . A potential issue with this approach is that it allows for X_t to be negative which can lead to numerical instabilities. For a further discussion about the convergence of this scheme in terms of mean square error, the reader is invited to consult [56].

In contrast to the penalty scheme, the projection scheme ensures that the samples are always in \mathbb{R}_+^n by applying a projection at the end of each step. This scheme is defined by

$$\begin{aligned} Y_{t_{k+1}} &= \bar{X}_{t_k} - \nabla_x U(X_{t_k})h + \sqrt{2h}\xi_k, \\ \bar{X}_{t_{k+1}}^i &= (Y_{t_{k+1}}^i)^+ \quad \text{for all } i \in \{1, \dots, n\}. \end{aligned}$$

Here $\pi_D^\gamma(x) = \pi_{\partial D}^\gamma(x) + (F(x))^+ \gamma(\pi_{\partial D}^\gamma(x))$, $t_k = kh$, and ξ_k are i.i.d random variables with mean zero, covariance given by the identity matrix and finite third moments. It has been shown in [21] that the weak error converges with order $h^{\frac{1}{2}-\epsilon}$ for any $\epsilon > 0$ and moreover that the rate $h^{\frac{1}{2}}$ is a lower bound for reflecting Brownian motion in an interval. In the case where ξ_k is bounded, the weak error converges with order $h^{\frac{1}{2}}$ [21, Theorem 3.4].

The reflected Euler scheme is similar to the projected Euler scheme but the process is in the interior of the domain with probability one, i.e. $\bar{X}_{t_k}^i \neq 0$ for any i, k with probability 1. This scheme is defined by

$$\begin{aligned} Y_{t_{k+1}} &= \bar{X}_{t_k} - \nabla_x U^{\lambda, b}(X_{t_k})h + \sqrt{2h}\xi_k, \\ \bar{X}_{t_{k+1}} &= |Y_{t_{k+1}}| \quad \text{for all } i \in \{1, \dots, n\}. \end{aligned}$$

Here $|\cdot|$ is understood to be applied componentwise, i.e. for $x \in \mathbb{R}^n$ we define the vector $|x|$ by $(|x|)_i = |x_i|$. It is shown in [10, Theorem 1] that this scheme has weak error with order h , comparing favourably with the projection scheme.

From the options that we have discussed above, we choose to adopt the reflected scheme as a way of incorporating the non-negativity constraint without increasing the computational complexity of the methods or introducing significant additional bias. The resulting proximal Markov chain Monte Carlo algorithms are presented in Section 4. From an algorithmic viewpoint, these MCMC algorithms are closely related to their non-reflected counterparts, with the only minor change being the careful introduction of a component-wise absolute-value operation. However, from a Bayesian computation methodology perspective, this minor change has the profound effect of allowing us to approximate the reflected SDE, which we have shown to be well-posed and to converge exponentially fast to the desired target density.

3.5 Related work

Let us mention some related works on the use of RSDEs for sampling. In the case of a smooth bounded domain, convergence of RSDEs to their invariant distribution and numerical approximations as studied in [14, 43]. In both of these papers the invariant distribution is unknown (as they work with general RSDEs) but can be shown to exist and be unique since the domain is bounded and the noise is non-degenerate. Convergence of the numerical scheme is then investigated by using suitable expansions and PDE estimates. In the case of convex bounded domains the papers [11, 41] establishes exponential convergence of the RSDE to the invariant distribution and convergence of the projected Euler scheme. Reflected Langevin dynamics have also been used in the context of optimisation, see [67], in the case of a bounded smooth domain. Finally we mention that reflections can also be used with Hamiltonian

dynamics [15, 23], although we note here that a Metropolis-Hastings step has been included to ensure the algorithm has the desired invariant distribution. Since all of these algorithms are for bounded domains they are not applicable to our setting and extra care is required as the domain is not smooth.

4 Reflected MCMC methods

In general, it is not possible to solve (3.8), and discrete approximations of the overdamped and underdamped Langevin dynamics (Section 3) need to be considered instead. In this Section, numerical schemes will be described that aim to solve (3.8) in a high dimensional setting. These schemes are inspired by the Reflected Euler scheme provided in Section 3.4 and adjust current proximal MCMC methodology [18, 24, 42, 54] that has proven to be scalable and efficient under Gaussian noise but cannot guarantee that the non-negative constraint is satisfied under Poisson, binomial or geometric noise processes.

4.1 Reflected MYULA (PMALA)

In the absence of the non-negativity constraint requirement, a simple way to discretize the SDE in (3.3) is to follow the Euler-Maruyama (EM) scheme leading to the known Unadjusted Langevin Algorithm (ULA) [62]. Under a non-differentiable $U = -\log \pi$, the authors in [52] approximated the non-smooth term g of U by its MY envelope g^λ given in (3.5). By using (3.6), ULA could be applied on (3.3) leading to the known Moreau-Yosida Unadjusted Langevin Algorithm (MYULA). Taking into consideration the non-negativity constraint requirement, we propose a reflected version of MYULA, namely Reflected MYULA (R-MYULA), where each sample is simply reflected by taking its absolute value (see Section 3.4). The algorithm is presented in (1). If necessary, the asymptotic bias introduced by the discretization of the RSDE in (3.8) and the approximation $\pi^{\lambda, \beta}$ of π can be corrected by complementing the R-MYULA with a Metropolis-Hastings (MH) step [24, 52, 62] leading to the Reflected Proximal Metropolis Adjusted Langevin Algorithm (R-PMALA).

A main computational drawback of R-MYULA and R-PMALA is the fact that in order to converge one needs to choose $\delta \leq 1/L$ where $L = L_{f_y^b} + 1/\lambda$ is the Lipschitz constant of $\nabla \log \pi^{\lambda, b}$ and $L_{f_y^b}$ is the Lipschitz constant of f_y^b constituting a function of b . The parameters λ and b define a trade-off between the computational efficiency and accuracy since letting $\lambda \rightarrow 0$ and $b \rightarrow 0$ brings $\pi^{\lambda, b}$ close to π , which reduces the asymptotic bias, but leads the discretization time-step to be diminished and consequently reduces the chain efficiency [24]. Thus, in cases where the likelihood is very informative ($L_{f_y^b}$ is large) or a high level of accuracy in the approximated g^λ is required, the stepsize will be small leading to a highly correlated Markov chain and an MCMC method that explores very slowly the target distribution $\pi^{\lambda, b}$.

Algorithm 1 Reflected MYULA

```

Set  $X_0 \in \mathbb{R}_{++}^n$ ,  $b > 0$ ,  $\lambda > 0$ ,  $\delta \in (0, \lambda/(\lambda L_{f_y^b} + 1)]$ ,  $N \in \mathbb{N}$ 
for  $k = 0 : (N - 1)$  do
     $Z_{k+1} \sim \mathcal{N}(0, \mathbb{I}_n)$ 
     $Y_{k+1} = (1 - \frac{\delta}{\lambda}) X_k - \delta \nabla f_y^b(X_k) + \frac{\delta}{\lambda} \text{prox}_y^\lambda(X_k) + \sqrt{2\delta} Z_{k+1}$  (MYULA)
     $X_{k+1} = |Y_{k+1}|$ 
end for
return  $\{X_k : k \in \{1, \dots, N\}\}$ 

```

4.2 Reflected SKROCK

The authors in [54] dealt with the step-size limitation of MYULA by applying a more sophisticated discretization scheme to simulate the Langevin SDE (3.3) based on explicit stabilized Runge-Kutta integrators [1, 2]. This scheme is known as the stochastic second kind orthogonal Runge-Kutta Chebyshev (SKROCK) method, and allows for accelerated and efficient sampling from the posterior distribution [2, 54]. Under the non-negativity requirement, we propose

Algorithm 2 Reflected SKROCK

Set $X_0 \in \mathbb{R}_{++}^n$, $b > 0$, $\lambda > 0$, $N \in \mathbb{N}$, $s \in \{3, \dots, 15\}$, $\eta > 0$

Compute $l_s = (s - 0.5)^2(2 - 4/3\eta) - 1.5$

Compute $\omega_0 = 1 + \frac{\eta}{s^2}$, $\omega_1 = \frac{T_s(\omega_0)}{T'_s(\omega_0)}$, $\mu_1 = \frac{\omega_1}{\omega_0}$, $\nu_1 = s\omega_1/2$, $k_1 = s\omega_1/\omega_0$

Choose $\delta \in (0, \delta_s^{max}]$, where $\delta_s^{max} = l_s/(L_{f_y^b} + 1/\lambda)$

for $k = 0 : (N - 1)$ **do**

$Z_{k+1} \sim \mathcal{N}(0, \mathbb{I}_n)$

$K_0 = X_k$

$W_1 = X_k + \nu_1 \sqrt{2\delta} Z_{k+1}$

$Y_1 = X_k - \mu_1 \delta \nabla f_y^b(W_1) - \frac{\mu_1 \delta}{\lambda} (W_1 - \text{prox}_g^\lambda(W_1)) + k_1 \sqrt{2\delta} Z_{k+1}$

$K_1 = |Y_1|$

for $j = 2 : s$ **do**

Compute $\mu_j = \frac{2\omega_1 T_{j-1}(\omega_0)}{T_j(\omega_0)}$, $\nu_j = \frac{2\omega_0 T_{j-1}(\omega_0)}{T_j(\omega_0)}$, $k_j = -\frac{T_{j-2}(\omega_0)}{T_j(\omega_0)} = 1 - \nu_j$

$Y_j = -\mu_j \delta \nabla f_y^b(K_{j-1}) - \frac{\mu_j \delta}{\lambda} (K_{j-1} - \text{prox}_g^\lambda(K_{j-1})) + \nu_j K_{j-1} + k_j K_{j-2}$

$K_j = |Y_j|$

end for

$X_{k+1} = K_s$

end for

return $\{X_k : k \in \{1, \dots, N\}\}$

Algorithm 3 Reflected MYUULA

Set $X_0 \in \mathbb{R}_{++}^n$, $V_0 \in \mathbb{R}^n$, $b > 0$, $\lambda > 0$, $\gamma > 0$, $u = \lambda/(\lambda L_{f_y^b} + 1)$, $\delta = O(1)$, $N \in \mathbb{N}$

Set $\Sigma = \begin{pmatrix} u \left(\delta - \frac{1}{\gamma^2} e^{-2\gamma\delta} - \frac{3}{\gamma^2} + e^{-\gamma\delta} \right) \mathbb{I}_n & \frac{u}{\gamma} (1 + e^{-2\gamma\delta} - 2e^{-\gamma\delta}) \mathbb{I}_n \\ \frac{u}{\gamma} (1 + e^{-2\gamma\delta} - 2e^{-\gamma\delta}) \mathbb{I}_n & u(1 - e^{-2\gamma\delta}) \mathbb{I}_n \end{pmatrix}$

for $k = 0 : (N - 1)$ **do**

$(Z_{k+1}^1, Z_{k+1}^2) \sim \mathcal{N}(0, \Sigma)$

$V_{k+1} = V_k e^{-\gamma\delta} - \frac{u}{\gamma} (1 - e^{-\gamma\delta}) (-\nabla f_y^b(X_k) - \frac{1}{\lambda} (X_k - \text{prox}_g^\lambda(X_k)))$

$Y_{k+1} = X_k + \frac{1}{\gamma} (1 - e^{-\gamma\delta}) V_k - \frac{u}{\gamma} \left(\delta - \frac{1}{\gamma} (1 - e^{-\gamma\delta}) \right) (-\nabla f_y^b(X_k) - \frac{1}{\lambda} (X_k - \text{prox}_g^\lambda(X_k)))$

$X_{k+1} = |Y_{k+1}|$

end for

return $\{X_k : k \in \{1, \dots, N\}\}$

the Reflected SKROCK (R-SKROCK) algorithm by applying reflection after every gradient evaluation. The algorithm is presented in (2).

4.3 Reflected MYUULA

In a similar manner with MYULA, the authors in [31] used a MY smoothed approximation U^λ of a non-differentiable $U = -\log \pi$ to solve the underdamped dynamics in (3.4). Literature regarding formulations of Langevin-based algorithms to integrate (3.4) can be found, for instance, in [18, 49, 66]. In this work, we present a reflected version of the the well-known Euler exponential integrator [18, 36] to solve (3.4). The algorithm is presented in (3).

4.4 Implementation guidelines

We now discuss suitable ranges and recommended values for the parameters of Algorithm 1, Algorithm 2 and Algorithm 3. Our recommendations seek to provide general rules that are simple and robust, rather than optimal values for specific models.

Setting λ and b

Similarly to [24], we recommend to set $\lambda \in [L_{f_y^b}^{-1}, 10L_{f_y^b}^{-1}]$ and use $\lambda = L_{f_y^b}^{-1}$ in our numerical experiments (larger values of λ lead to faster convergence at the cost of additional bias). Our experiments suggest that the value of b is more critical in terms of controlling bias than λ . We recommend setting b to be 1% of the expected Mean Intensity Value (MIV) of x (i.e., $\|x\|_1/n$), which provides a good trade-off of efficiency and accuracy.

Setting δ , γ , u , s , and η

The convergence theory for MYULA [24] requires setting $\delta \in (0, \delta_{max})$ with $\delta_{max} = 1/(L_{f_y^b} + 1/\lambda)$ (the value δ_{max} is derived from bounding the Lipschitz constant of $\nabla f_y^b + (x - \text{prox}_g^\lambda(x))/\lambda$). In our experiments, given the dimensionality involved, and because we did not observe significant bias, we choose $\delta = \delta_{max}$ to improve convergence speed. Regarding R-MYUULA, following the spirit of [18, 66], we set $\delta = 2$, $u = 1/(L_{f_y^b} + 1/\lambda)$, and $\gamma = 2$. For R-SKROCK, similarly to [54], we recommend setting $\delta \in (0, \delta_s^{max}]$, where $\delta_s^{max} = l_s/(L_{f_y^b} + 1/\lambda)$, $l_s = (s - 0.5)^2(2 - 4/3\eta) - 1.5$ and $s \in \{3, \dots, 15\}$; we use $\delta = \delta_s^{max}$ with $s = 10$ and $\eta = 0.05$ in our experiments.

Lastly, to maximise the asymptotic computational efficiency of R-PMALA we set δ to achieve an acceptance rate close to 57% within the range of admissible values for geometric ergodicity $(0, \delta_{max})$ with $\delta_{max} = 1/(L_{f_y^b} + 1/\lambda)$ [57].

5 Numerical Experiments

In this section, the MCMC methods proposed in Section 4 are demonstrated through a range of experiments related to Poisson image deconvolution, binomial denoising and geometric image inpainting. These experiments were selected to represent a variety of challenging configurations in terms of ill-posedness, ill-conditioning and dimensionality of y and x . We report results for R-MYULA, R-SKROCK, R-MYUULA and R-PMALA and comparisons with MAP estimation by using the ADMM algorithm PIDAL [27]. Because PIDAL was originally designed to work with Poisson likelihoods, in the binomial and geometric experiments we make minor changes to PIDAL so that it performs MAP inference w.r.t. to the correct model. Moreover, in the Poisson deconvolution experiment, we also report a comparison with the state-of-the-art Gibbs sampler SPA [72], which was successfully applied to total-variation Poisson image deblurring in [72] (SPA combines MYULA with an augmentation-relaxation strategy to improve convergence speed, at the expense of more bias).

To make the comparisons fair, and without loss of generality, we conduct all our experiments by using the same total-variation prior that is considered in PIDAL and SPA. This prior is given by $p(x|\theta) \propto \exp\{-\theta TV(x)\}$, where $TV(\cdot)$ denotes the isotropic total-variation pseudo-norm, and $\theta > 0$ is a regularisation parameter. We automatically estimate θ from y for each experiment by marginal maximum likelihood estimation by using [71, Algorithm 1]. Furthermore, for fairness, all methods are run with the same computing time budget. In particular, we generate 10^6 samples for the R-MYULA and R-MYUULA, $10^6/s$ samples with $s = 10$ for the R-SKROCK, 6.5×10^5 samples for the R-PMALA (there is an additional computational overhead associated with the Metropolis-Hastings (MH) step [52]), and 6.8×10^5 for SPA. In all cases, we use a 5% burn-in period. Also, to check the asymptotic bias introduced by the approximation $\pi^{\lambda, b}$ of π , the discretization and the reflection, for each experiment we also run a long R-PMALA chain targeting π . In all cases, we observed a very good agreement between the unadjusted methods and the R-PMALA reference, indicating that the asymptotic bias is indeed very small.

The Monte Carlo samples produced by the MCMC methods are then used to compute the following quantities: 1) the posterior mean $E(x|y)$, which is the minimum mean square error (MMSE) Bayesian estimator of $x|y$; 2) uncertainty visualisation plots presenting the marginal standard deviation of pixels at different resolutions [12]; and 3)

the sample autocorrelation function (ACF) of the fastest and slowest mixing components of the Markov chain (these correspond to the one-dimensional subspaces where the Markov chain achieves the highest and slowest convergence rates, corresponding to the subspaces with lowest and highest variance respectively³). For completeness, the ACF for a pixel with typical variance is also included in the plots (we use the median). Lastly, as a way of illustrating convergence speed, for each experiment we also plot the evolution of the estimated normalized root MSE (NRMSE) for the posterior mean estimate, as a function of the number of gradient evaluations.

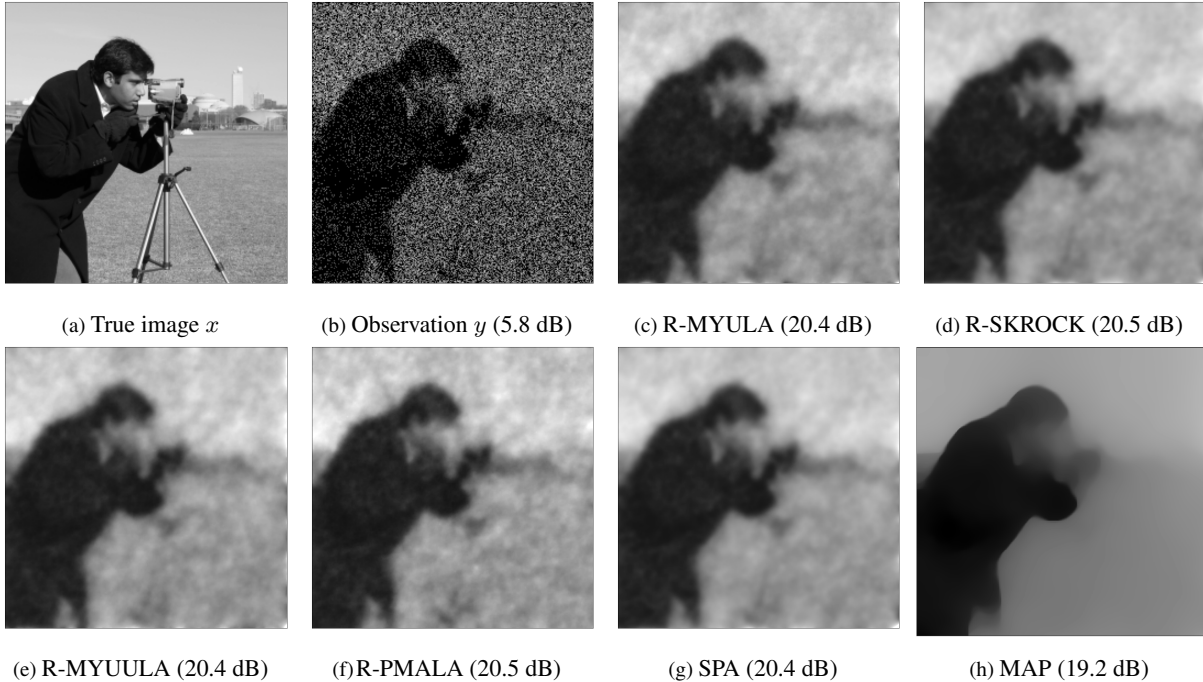


Figure 1: Poisson experiment: (a) True image x of size 256×256 ; (b) Blurred observation y ; (c)-(f) MMSE estimators; (h) MAP estimator.

5.1 Non-blind Poisson image deconvolution

In this experiment, we consider the estimation of a high-resolution image $x \in \mathbb{R}^n$ from a blurred observation $y \sim \mathcal{P}(Ax)$, where the blur operator is known. We chose A to be nearly singular leading to an ill-conditioned problem with highly noise-sensitive solutions. The posterior distribution is given by

$$\pi \triangleq p(x|y) \propto \exp(-f_{\mathcal{P}}(x) - \theta TV(x)) . \quad (5.1)$$

where we recall that $f_{\mathcal{P}}(\cdot)$ is the Poisson negative log-likelihood defined in (2.2).

Figure 1a presents the ground truth cameraman image x of size $n = 256 \times 256$ pixels, whose intensities we scaled for this experiment so that its MIV is 1 (recall that in the Poisson model the noise power is directly determined by the intensities of the image x , so in order to test different signal-to-noise ratios it is necessary to scale the ground truth image). Figure 1b shows a realization y with peak signal-to-noise ratio (PSNR) ~ 5.8 dB generated by the observation model (2.1) with A being a 5×5 uniform blur. For this experiment, the maximum marginal likelihood estimate of θ is $\hat{\theta} = 5.65$. All the methods are implemented by using this value, and by following the recommendations of Section 4.4. For R-PMALA, achieving the optimal acceptance probability of 57% is not possible in this experiment

³To identify the directions with smallest and largest uncertainty, someone would need to compute the posterior covariance matrix, an infeasible task in high dimensions. Instead, approximations of the posterior covariance are used by assuming that the latter is diagonalizable on the same basis as the forward operator A .

as the maximum admissible value $\delta_{max} = 1/(L_{f_y^b} + 1/\lambda)$ leads to an acceptance probability of 85% (we use this value of δ for our experiments). Regarding SPA [72], we test a large range of values in a supervised manner and find that $(\rho, \alpha) = (0.035, 0.035)$ provide a competitive trade-off between asymptotic accuracy and convergence speed.

Figure 1 presents also the MMSE estimates given by the proposed methods and the MAP estimate calculated by PIDAL. The MMSE estimates are visually similar with an apparent staircase effect while the MAP estimate returns an oversmoothed reconstruction. It should be noted that this oversmoothed result is related to the particular choice of θ . A value for θ that improves the MAP estimate can be found (e.g. through cross-validation) but this would require knowledge of the true image (the MMSE results can also be marginally improved by tailoring θ in this way). The PSNR values of each method are also reported in Figure 1. The MMSE estimates have almost identical PSNR values and outperform the MAP estimate in this case.

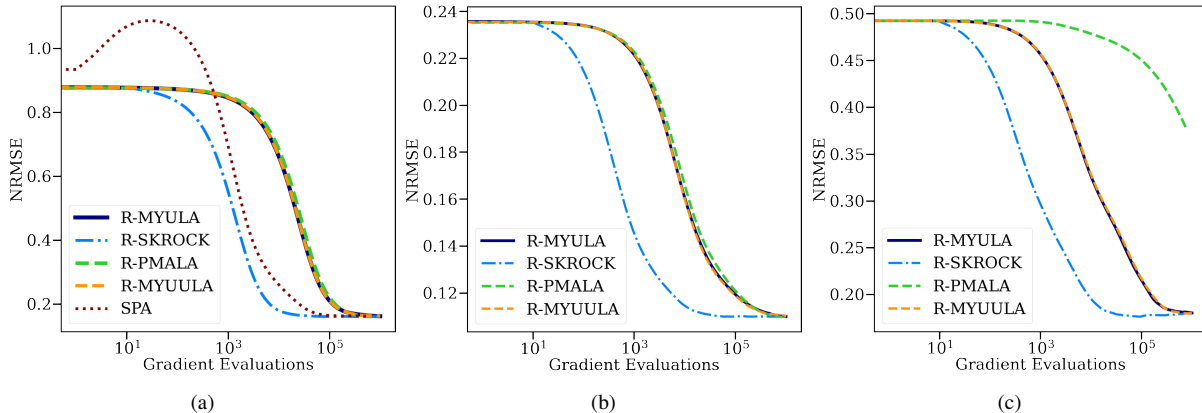


Figure 2: Evaluation of NRMSE for (a) the Poisson deblurring experiment (b) the binomial denoising experiment and (c) the geometric inpainting experiment. The number of gradient evaluations are presented in log-scale (base 10).

Figure 2a presents the evolution of the NRMSE estimation for the MMSE solutions as a function of the number of gradient evaluations (in log-scale). We observe that R-SKROCK has the highest convergence speed followed by SPA whose NRMSE increases for a small number of iterations. This is due to a transient regime of the auxiliary variables introduced in SPA. It can also be observed that R-MYULA and R-MYUULA have similar convergence speed, and R-PMALA has slightly slower convergence speed due to its Metropolis nature.

In Figure 3, we present uncertainty visualization plots that are useful to quantify the uncertainty related to image structures at different spatial scales. Specifically, for different scales j , we downsample (by averaging) the stored samples by a factor of 2^j before computing the standard deviation. The estimates of the pixel-wise standard deviations ($j = 0$) obtained by each algorithm are presented in the first column of Figure 3. It is observed that the estimates obtained by R-MYULA, R-PMALA and R-MYUULA are less accurate of the respective one obtained by R-SKROCK in agreement with the experiments under Gaussian noise in [54]. High uncertainty is concentrated around the edges of the cameraman which is expected from the particular choice of prior. Some uncertainty also exists within the background of the scene, since the likelihood is highly uninformative for these pixel regions and the prior does not carry any information to recover important details (e.g. the buildings). Note that because of the Poisson likelihood and the positivity constraints, the uncertainties appear larger in the background (higher average intensity) than on the coat of the cameraman (lower average intensity). However, the relative uncertainties (i.e., when dividing by the pixelwise true intensities) would be higher in dark than bright pixels. For larger structures of pixels ($j \neq 0$), the highest uncertainty is observed at the background of the cameraman's image, since there both the likelihood and the prior are the most weak.

To conclude this experiment, the sample ACFs are computed for the slowest and the fastest directions in the Fourier domain⁴. In Figure 4, we see that independence is reached in approximately 20 iterations (after thinning) in the median and fastest directions, and it is much slower for the few very uncertain coefficients. In addition, the superiority of R-SKROCK and SPA in sense of convergence properties can be observed along all directions. The ACF

⁴The slowest (fastest) direction corresponds to the Fourier coefficient with the highest (lowest) variance.

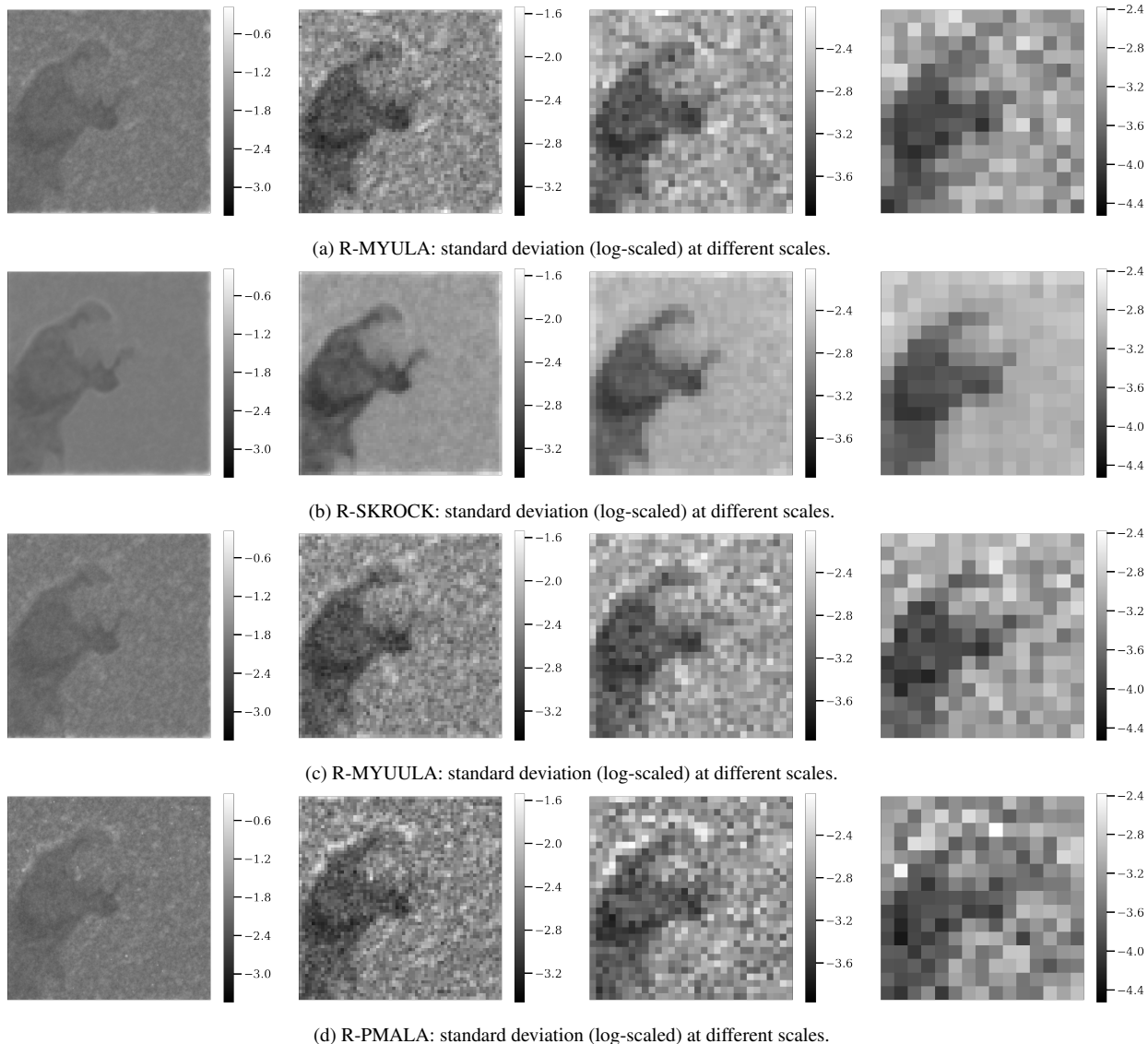


Figure 3: Marginal posterior standard deviation for the Poisson deblurring problem at different scales (0,2,4,8 from left to right). The scale i corresponds to a downsampling by a factor 2^i of the original sample size. Most of the uncertainty is located around the edges and the cameraman's background.

Poisson deblurring cases	MMSE		MAP	
	NRMSE	PSNR	NRMSE	PSNR
MIV = 1, g.e. = 1×10^6	0.1612	20.53 dB	0.1873	19.25 dB
MIV = 10, g.e. = 2×10^6	0.1143	23.54 dB	0.1186	23.22 dB
MIV = 100, g.e. = 5×10^6	0.088	25.74 dB	0.089	25.72 dB

Table 1: Comparison of MMSE and MAP estimator for different Poisson deblurring cases. After a large number of gradient evaluations (g.e.), the algorithms exhibit identical performance.

of the R-SKROCK samples along the slowest direction decay the fastest among all the methods.

Finally, Table 1 reports comparisons between the MMSE estimation and the MAP estimation under less severe Poisson noise (i.e. higher MIV for the target image). We run the MCMC algorithms enough time so the calculated

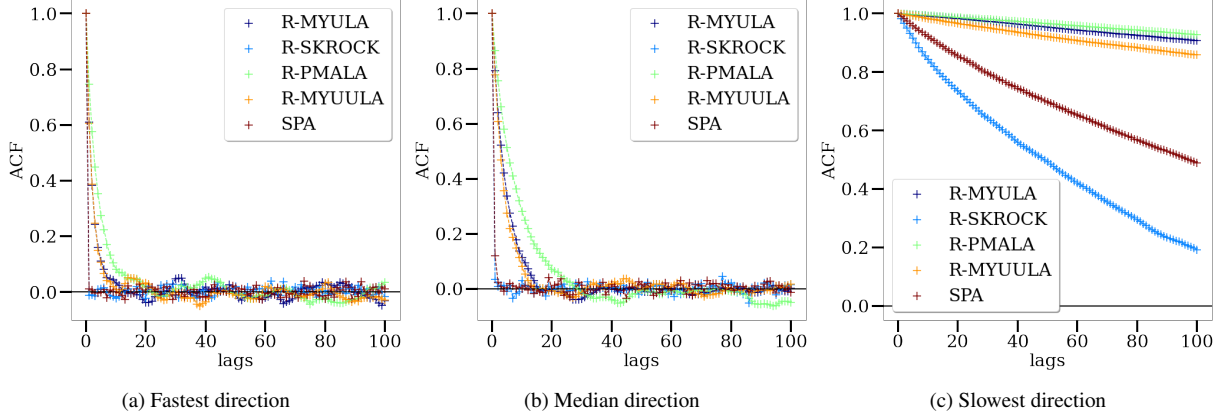


Figure 4: ACF for the fastest, median, and slowest direction for the Poisson deblurring problem. The ACF is shown for lags up to 100 for all images in the Fourier domain.

MMSE estimates of each method have identical performance. It is observed that the MMSE estimate outperforms the MAP estimate, particularly in cases where the likelihood is uninformative (i.e. low MIV).

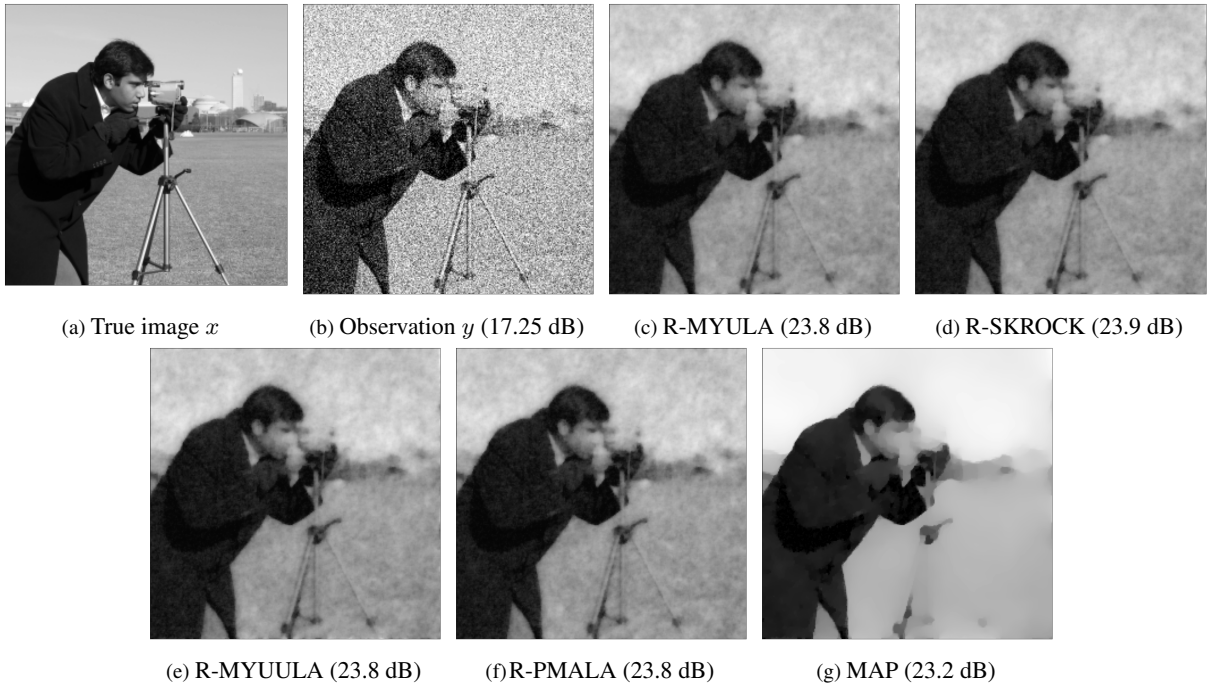


Figure 5: Binomial experiment: (a) True image x ; (b) Observation y ; (c)-(f) MMSE estimators; (g) MAP estimator.

5.2 Binomial denoising

We consider now a binomial denoising problem, where we seek to estimate a high-resolution image $x \in \mathbb{R}^n$ from a realization $y \sim \text{Bin}(t, 1 - e^{-x})$. The posterior distribution is given by

$$\pi \triangleq p(x|y) \propto \exp(-f_{\mathcal{B}}(x) - \theta TV(x)) , \quad (5.2)$$

where $f_B(\cdot)$ is the binomial negative log-likelihood in (2.4).

Figure 5b presents a binomial realization y of the cameraman image generated by using the observation model (2.3) with $A = I$, $t_i = 10 \forall i = 1, \dots, n$ and the MIV of x equal to 1 (recall that in the binomial model the noise power is directly determined by the intensities of the image x and the repetition periods t). For this experiment, the maximum marginal likelihood estimate of θ is $\hat{\theta} = 6.4$. All the methods are implemented by using this value, and by following the recommendations of Section 4.4. For R-PMALA, we achieved an acceptance rate of 83% by setting $\delta = \delta_{max} = 1/(L_y^b + 1/\lambda)$.

Figure 5 also presents the MMSE estimates given by the proposed methods and the MAP estimate calculated by adjusting PIDAL for binomial likelihoods. The MMSE estimates are visually similar with an apparent staircase effect while the MAP estimate returns an oversmoothed reconstruction. The PSNR values of each method are reported in the respective captions in Figure 5. The MMSE estimates have almost identical PSNR values which are higher than

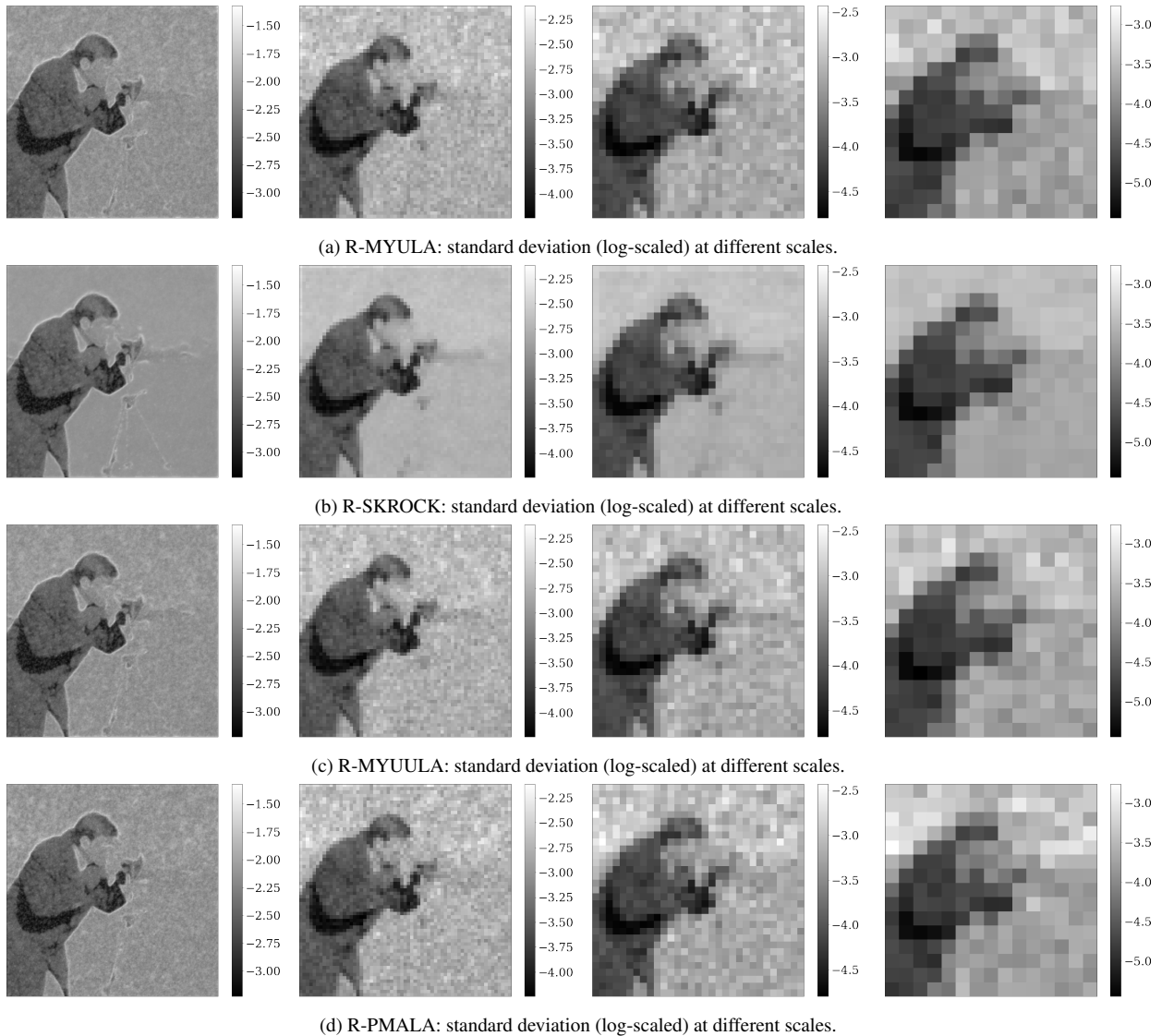


Figure 6: Marginal posterior standard deviation for the binomial denoising problem at different scales (0,2,4,8 from left to right). The scale i corresponds to a downsampling by a factor 2^i of the original sample size. Most of the uncertainty is located around the edges and the cameraman's background.

the respective one of the MAP estimate.

Figure 2b presents the evolution of the NRMSE estimation for the MMSE solutions as a function of the number of gradient evaluations (in log-scale). We observe that R-SKROCK has the highest convergence speed, R-MYULA and R-MYUULA have similar convergence speed, and R-PMALA has the slowest convergence among the proposed methods.

The estimates of the pixel-wise standard deviations ($j = 0$) are presented in the first column of Figure 6. It is observed that the estimates obtained by R-MYULA, R-PMALA and R-MYUULA are slightly more noisy than respective one obtained by R-SKROCK. High uncertainty is concentrated around the edges (e.g. cameraman figure and buildings) and some uncertainty also exists in homogeneous regions, since the likelihood is quite uninformative for these pixel regions. For larger structures of pixels ($j \neq 0$), the highest uncertainty is spotted at the background part of the cameraman’s image. see Figure 6. R-SKROCK returns a smooth estimate of the posterior variance while R-PMALA returns a slightly noisier estimate since its Metropolised nature leads to higher estimation variance [12, 24].

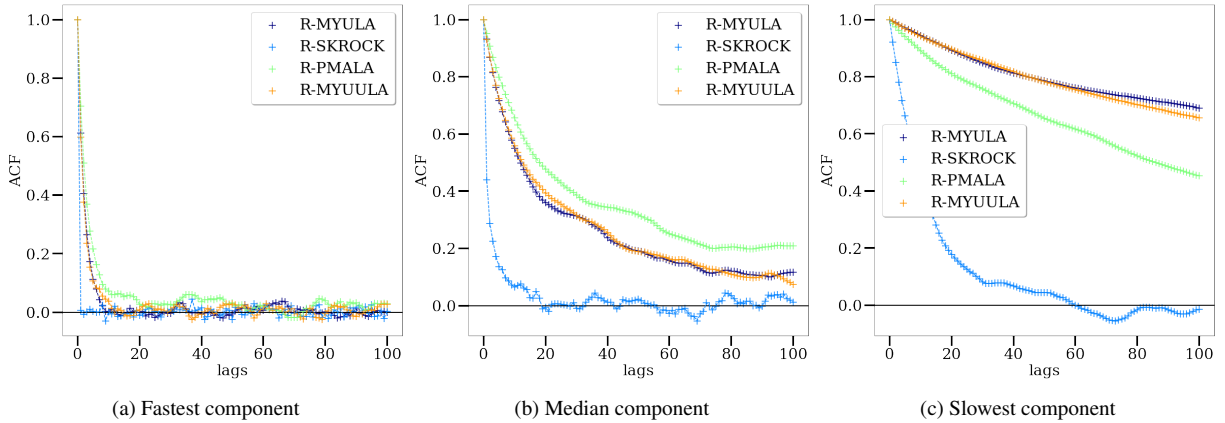


Figure 7: ACF for the fastest, median, and slowest components for the binomial denoising problem. The ACF is shown for lags up to 100 for all images in the pixel domain.

Finally, in Figure 7, we see that independence is reached fast for the components of low or median uncertainty, and is much slower for the few very uncertain pixels. The superiority of R-SKROCK in sense of convergence properties compared with the other methods can also be observed. Particularly, the ACF of the R-SKROCK samples along the slowest component decay the fastest among the ones given by R-MYULA, R-MYUULA, and R-PMALA.

Table 2 reports NRMSE and PSNR comparisons between the MMSE estimation and the MAP estimation for different cases of binomial noise. We run the MCMC algorithms enough time so the calculated MMSE estimates of each method have identical performance. It is observed that the MMSE estimator outperforms the MAP estimator in cases where the likelihood is quite uninformative (low MIV or low number of repetition periods t) while the MAP estimator is more competitive under more informative likelihoods.

Binomial denoising cases	MMSE		MAP	
	NRMSE	PSNR	NRMSE	PSNR
MIV = 1, t=10, g.e.= 5×10^6	0.1092	23.94 dB	0.1189	23.20 dB
MIV = 1, t=100, g.e.= 5×10^6	0.0592	29.25 dB	0.0573	29.54 dB
MIV = 0.1, t=100, g.e.= 3×10^6	0.0946	25.19 dB	0.099	24.76 dB
MIV = 0.1, t=1000, g.e.= 5×10^6	0.0496	30.79 dB	0.045	31.66 dB

Table 2: Comparison of MMSE and MAP estimator for different binomial denoising cases. After a large number of gradient evaluations (g.e.), the algorithms exhibit identical performance.

5.3 Geometric Inpainting

We consider now a very challenging inpainting problem, where we seek to estimate a high-resolution image $x \in \mathbb{R}^n$ from a set of geometric measurements $t \sim \text{Geo}(1 - e^{-Ax})$, where A is a $m \times n$ matrix containing $m \leq n$ randomly

selected rows of the $n \times n$ identity matrix. The posterior distribution is given by

$$\pi \triangleq p(x|t) \propto \exp(-f_G(x) - \theta TV(x)) , \quad (5.3)$$

where $f_G(\cdot)$ is the geometric log-likelihood in (2.5).

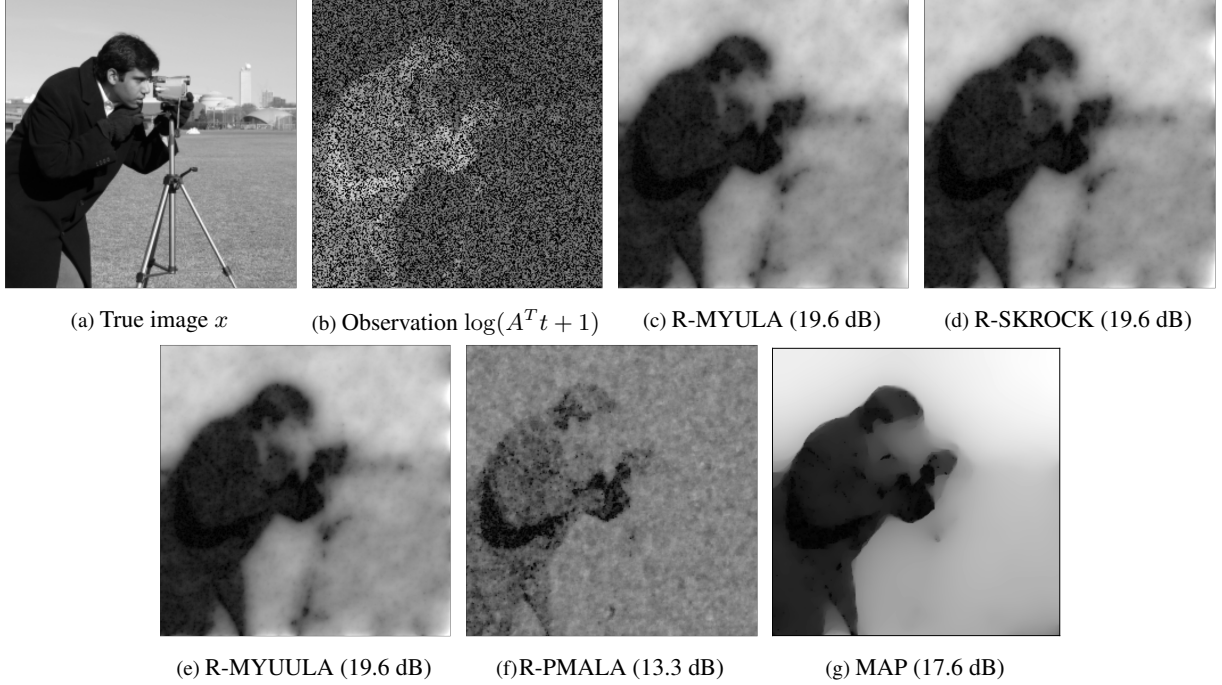


Figure 8: Geometric experiment: (a) True image x ; (b) Observation $A^T t$ is displayed as $\log(A^T t + 1)$ for visualization purposes; (c)-(f) MMSE estimators as computed by the different MCMC algorithms; (g) MAP estimator.

Figure 8 presents a (backprojected) realization of x generated by using the model (2.5) with A being a masking operator with 30% mixing pixels and the MIV of x being 10^{-2} . For this experiment, the maximum marginal likelihood estimate of θ is $\hat{\theta} = 632$. All the methods are implemented by using this value, and by following the recommendations of Section 4.4. Regarding R-PMALA, it struggled to reach stationarity up to the available computational time because of the extremely uninformative likelihood. It required to take a very small step size ($\delta \ll \delta_{max}$) to reach an acceptable acceptance rate of 57%.

Figure 8 presents the MMSE estimates given by the proposed methods and the MAP estimate calculated by the adjusted PIDAL for geometric likelihoods. The MMSE estimates given by R-MYULA, R-SKROCK, R-MYUULA are visually similar with an apparent staircase effect while the MAP estimate returns an oversmoothed reconstruction. Since R-PMALA failed to properly converge in the available computing budget, it returned a noisy reconstruction. The PSNR values of each method are reported in the respective captions in Figure 8. The comparison between MMSE and MAP estimation is more significant in this challenging experiment. Specifically, the PSNR values given by the MMSE estimates of R-MYULA, R-SKROCK, R-MYUULA are significantly higher than the PSNR value of the MAP estimator.

Figure 2c presents the evolution of the NRMSE estimation for the MMSE solutions as a function of the number of gradient evaluations (in log-scale). It can be observed that R-SKROCK has again the highest convergence speed, R-MYULA and R-MYUULA have similar convergence speed, and R-PMALA is by far the slowest algorithm.

The estimates of the pixel-wise standard deviations ($j = 0$) are presented in the first column of Figure 9. It is observed that the estimates obtained by R-MYULA and R-MYUULA are slightly less accurate of the SKROCK estimate. High uncertainty is concentrated around the edges (e.g. cameraman figure) and the cameraman's background,

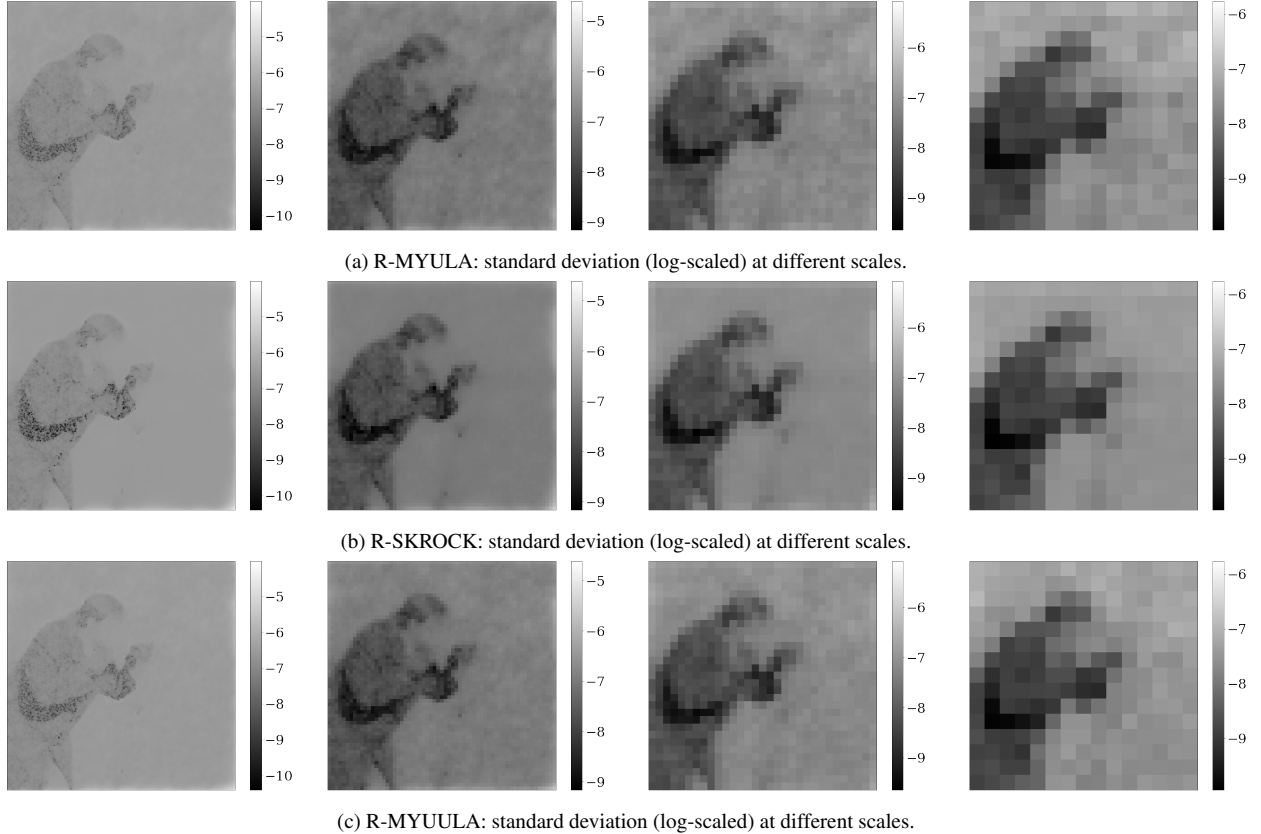


Figure 9: Marginal posterior standard deviation for the geometric inpainting problem at different scales (0,2,4,8 from left to right). The scale i corresponds to a downsampling by a factor 2^i of the original sample size. Most of the uncertainty is located at places where the likelihood is highly uninformative e.g. at the cameraman’s background.

since the likelihood is highly uninformative for these pixel regions. For higher scales ($j \neq 0$), the highest uncertainty is spotted at the background part of the cameraman’s image as expected.

To conclude this experiment, in Figure 10 we see that independence is reached fast for the components of low or median uncertainty, and is much slower for the few very uncertain pixels. Similarly to the previous experiments, R-SKROCK has better convergence properties compared with the other methods, see the decay of the ACF samples along the median and slowest components.

6 Conclusion

This work presented a new MCMC methodology based on a reflected and regularised Langevin SDE, specialised for performing Bayesian computation in low-photon imaging inverse problems involving a non-negativity constraint on the solution space. This reflected SDE is shown to be well-posed, have the desired invariant distribution, and to be exponentially ergodic under mild and easily verifiable conditions. This allowed deriving three unadjusted Langevin MCMC algorithms to perform Bayesian computation, as well as a Metropolised Langevin algorithm. The proposed approach was demonstrated with a range of experiments involving binomial, geometric, and low-intensity Poisson noise processes. In these experiments, we compared the computational accuracy and efficiency of the proposed MCMC algorithms, and illustrated their use for point estimation as well as uncertainty visualisation analyses. Moreover, we observed that in all the experiments with weakly informative data, the MMSE estimator computed by MCMC sampling outperformed MAP estimation (by convex optimization [27]) in terms of accuracy. The methods also compared

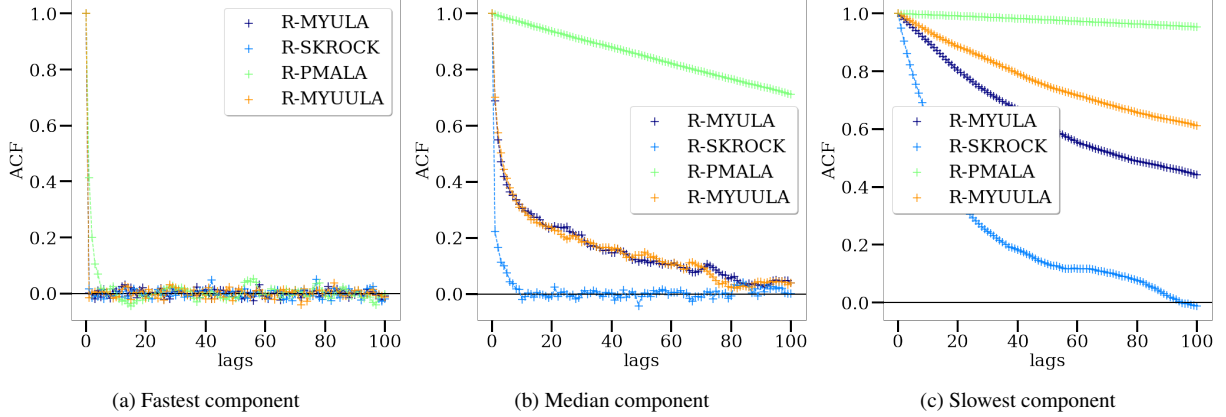


Figure 10: ACF for the fastest, median, and slowest components for the geometric inpainting problem. The ACF is shown for lags up to 100 for all images in the pixel domain.

favourably against the state-of-the-art SPA proximal MCMC algorithm in the context of Poisson image deblurring [72]. Future works will include a more detailed theoretical analysis of the proposed methodology and algorithms (e.g. non-asymptotic convergence rates [24]), as well as new variants suitable for data-driven priors such as plug-and-play priors encoded by image denoising neural networks [42] and priors encoded by generative neural network models [38].

A Proof of exponential ergodicity

Proof of Lemma 3.1. Observe that $f_y^b(x), g^\lambda(x)$ both converge pointwise to $f_y(x), g(x)$ respectively as $(\lambda, b) \rightarrow (0, 0)$. For f_y^b this follows since F is continuous and for g^λ this follows from [24, Equation (10)]. Moreover from [24] we have $g^\lambda \leq g$ and f_y^b is lower bounded by assumption. Therefore by the dominated convergence theorem we have $\int e^{-f_y^b - g^\lambda} dx \rightarrow \int e^{-f_y - g} dx$ as $(\lambda, b) \rightarrow (0, 0)$. Then by Scheffé's Lemma, see [73, (5.10)], we have

$$\lim_{\lambda \rightarrow 0, b \rightarrow 0} \int_{\mathbb{R}_+^d} |\pi(x) - \pi^{\lambda, b}(x)| dx = 0.$$

That is, $\pi^{\lambda, b}$ converges to π in total variation. \square

Proposition A.1. *Let \mathcal{P}_t be the semigroup corresponding to the SDE (3.8). Assume that $\nabla_x U^{\lambda, b}$ is continuously differentiable and globally Lipschitz with constant $L \geq 0$. Then for all $\varphi \in C_b^1(\mathbb{R}_+^n)$ we have $\mathcal{P}_t \varphi \in C_b^1(\mathbb{R}_+^n)$ for all $t > 0$ and moreover there exists $C > 0$ such that for all $t > 0$ and $\varphi \in C_b(\mathbb{R}_+^n)$ we have*

$$\|\nabla_x \mathcal{P}_t \varphi\|_\infty \leq \frac{C e^{Ltn}}{\sqrt{t}} \|\varphi\|_\infty.$$

Proof of Proposition A.1. Fix $\varphi \in C_b(\mathbb{R}_+^n)$ then by [22] we have that $\mathcal{P}_t \varphi \in C^1(\mathbb{R}_+^n)$ and we have the following representation by [22, Theorem 2]

$$\frac{\partial}{\partial x_j} \mathcal{P}_t \varphi(x) = \frac{1}{\sqrt{2t}} \sum_{i=1}^d \mathbb{E} \left[\mathbf{E}_{0,j} \left[\varphi(X_t^x) \int_0^t \mathbb{1}_{\xi_s=i} \mathbb{1}_{\tau>s} \rho_{0,s} dW_s^i \right] \right]. \quad (\text{A.1})$$

Here $\mathbf{E}_{0,j}$ is the expectation of the process ξ_s for with $\xi_0 = j$, and ξ_s is a continuous time Markov chain on $\{1, \dots, d\}$ with generator

$$\mathbf{L}\varphi(i) = \sum_{k=1}^d |c_t(i, k)| [\varphi(k) - \varphi(i)], \quad c_t(i, k) = \frac{\partial^2 U^{\lambda, b}}{\partial x_i \partial x_j}(X_t^x).$$

Let η_ℓ be the sequence of jump moments of ξ , i.e. $\eta_{\ell+1} = \inf\{s > \eta_\ell : \xi_s \neq \xi_{\eta_\ell}\}$. Let τ be the stopping time

$$\tau = \inf\{s > 0 : X_s^{x, \xi_s} = 0\}.$$

With this notation we can define $\rho_{0,s}$ as

$$\rho_{0,s} = \exp\left(\int_0^s \sum_{k \neq \xi_r} |c_r(\xi_r, k)| dr + \int_0^s c_r(\xi_r, \xi_r) dr\right) \prod_{0 < \eta_k \leq s} \text{sign}(c_{\eta_k}(\xi_{\eta_{k-1}}, \xi_{\eta_k})).$$

Since $\nabla_x U^{\lambda, b}(x)$ is globally Lipschitz we have that $c_t(i, k)$ is bounded in (t, i, k) , i.e. there exists $L > 0$ such that $c_t(i, k) \leq L$ which gives the following bound on $\rho_{0,s}$:

$$|\rho_{0,s}| \leq e^{sLn}.$$

Using this bound in (A.1) we have

$$\left| \frac{\partial}{\partial x_j} \mathcal{P}_t \varphi(x) \right| \leq \frac{e^{tLn}}{t} \|\varphi\|_\infty \sum_{i=1}^d \mathbb{E}[|W_s^i|] = \sqrt{\frac{2}{\pi}} \frac{e^{tLn} n}{\sqrt{t}} \|f\|_\infty.$$

□

As a consequence of the above smoothing result we show that every compact set is small.

Definition A.2. A set $C \subseteq \mathbb{R}_+^n$ is called a small set if there exist $t > 0$ a probability measure ν with $\nu(C) = 1$ and positive constant $\eta > 0$ such that

$$\mathcal{P}_t \mathbb{1}_E(x) \geq \eta \nu(E) \quad \text{for all } x \in C, E \subseteq \mathbb{R}_+^n \text{ measurable.}$$

Proposition A.3. Let \mathcal{P}_t be the semigroup corresponding to the SDE (3.8). Assume that $\nabla_x U^{\lambda, b}$ is continuously differentiable and globally Lipschitz. Then every compact set $K \subseteq \mathbb{R}_+^n$ is small.

Proof of Proposition A.3. First we show using a Girsanov transformation that the law of X_t^x is equivalent to the law of a reflected Brownian motion and hence admits a positive density on \mathbb{R}_+^n with respect to Lebesgue measure. Define the exponential martingale

$$Z_t = \exp\left(\frac{1}{\sqrt{2}} \int_0^t \nabla U^{\lambda, b}(X_s) dW_s - \frac{1}{4} \int_0^t \nabla_x U^{\lambda, b}(X_s)^2 ds\right)$$

we observe that this is a true martingale by Novikov's condition which is satisfied since $\nabla_x U^{\lambda, b}(X_s^x)$ is bounded. Then by Girsanov's theorem

$$\tilde{W}_t = W_t - \frac{1}{\sqrt{2}} \int_0^t \nabla U^{\lambda, b}(X_s) ds$$

is a Brownian motion under \mathbb{Q} where $\frac{d\mathbb{Q}}{d\mathbb{P}} = Z_t$. Therefore under \mathbb{Q} we have

$$X_t = x - \int_0^t \nabla U^{\lambda, b}(X_s) ds + \sqrt{2} W_t + L_t = x + \sqrt{2} \tilde{W}_t + L_t$$

is a reflected Brownian motion. By [35, Lemma 7.9] we have for any measurable $E \subseteq \mathbb{R}_+^n$ and $x \in \mathbb{R}_+^n$ that $\mathbb{Q}(X_t^x \in E) = 0$ if and only if E has zero Lebesgue measure. Hence $\mathbb{P}(X_t^x \in E) = 0$ if and only if E has zero Lebesgue measure since \mathbb{P} is equivalent to \mathbb{Q} . In particular we have that the law of X_t^x admits a density $p_t(\cdot; x)$ with respect to Lebesgue measure on \mathbb{R}_+^n and the density is strictly positive.

Fix a compact set $K \subseteq \mathbb{R}_+^n$ and $t > 0$. Define the measure ν by

$$\nu(E) = \frac{\inf_{x \in K} \mathbb{P}(X_t^x \in E \cap K)}{\inf_{x \in K} \mathbb{P}(X_t^x \in K)}, \text{ for any measurable } E \subseteq \mathbb{R}_+^n,$$

it remains to show that ν is a probability measure, i.e. that $\inf_{x \in K} \mathbb{P}(X_t^x \in K) > 0$. Since the law of X_t^x admits a strictly positive density we know that $\mathbb{P}(X_t^x \in K) > 0$ for each $x \in K$. If we have that $x \mapsto \mathbb{P}(X_t^x \in K)$ is continuous then since K is compact, $\mathbb{P}(X_t^x \in K)$ attains its minimum value and that value must be positive, i.e. $\inf_{x \in K} \mathbb{P}(X_t^x \in K) > 0$. It remains to show $x \mapsto \mathbb{P}(X_t^x \in K)$ is continuous. Let $\varphi(x) = \mathbb{1}_K(x)$ and approximate this function by the sequence $\{\varphi_k\}_{k \geq 0}$ where

$$\varphi_k(x) = \max\left(1 - \frac{1}{k}d(x, K), 0\right)$$

and $d(x, K)$ denotes the distance between x and the set K . Note that $\varphi_k \in C_b(\mathbb{R}_+^n)$, $\|\varphi_k\|_\infty \leq 1 = \|\varphi\|_\infty$ and $\varphi_k(x) \rightarrow \varphi(x)$ as $k \rightarrow \infty$ for each $x \in \mathbb{R}_+^n$. By the dominated convergence theorem $\mathcal{P}_t \varphi_k(x) \rightarrow \mathcal{P}_t \varphi(x)$ for each $x \in \mathbb{R}_+^n, t > 0$. On the other hand by Proposition A.1 the derivative of $\mathcal{P}_t \varphi_k$ is bounded uniformly in k and x and in particular, for fixed $t > 0$ the sequence $\{\mathcal{P}_t \varphi_k\}$ is uniformly bounded and equicontinuous therefore by Arzelà–Ascoli Theorem there exists some $\psi \in C_b(\mathbb{R}_+^n)$ such that $\mathcal{P}_t \varphi_k$ converges locally uniformly to ψ . Since $\mathcal{P}_t \varphi_k$ converges pointwise to $\mathcal{P}_t \varphi$ we have that $\psi = \mathcal{P}_t \varphi$ and hence $\mathcal{P}_t \varphi \in C_b(\mathbb{R}_+^n)$. \square

Before we prove Theorem 3.4 we need the following elementary lemma from [7, Lemma 2.2]. Note that although this Lemma is proved for functions on the whole space \mathbb{R}^n the proof still holds when restricted to the space \mathbb{R}_+^n .

Lemma A.4. *If $U^{\lambda, b}$ is differentiable, convex and $\int_{\mathbb{R}_+^n} e^{-U^{\lambda, b}(x)} dx < \infty$ then there exist $\alpha > 0$ and $R > 0$ such that for $|x| \geq R$, (3.13) holds.*

Proof of Theorem 3.4. By Proposition A.3 we have that every compact set is small and hence is petite (by [51, Proposition 5.5.3]) therefore to verify the conditions of Theorem 3.3 it remains to find a function $V(x)$ which satisfies Hypothesis 3.2. Set

$$V(x) = \exp\left(\gamma(1 + \|x\|^2)^{\frac{1}{2}}\right).$$

Note that

$$\partial_{x_i} V(x) = \frac{\gamma_i x_i}{(1 + \|x\|^2)^{\frac{1}{2}}} V(x)$$

In particular, we see that $\partial_{x_i} V(x) = 0$ if $x_i = 0$ and hence V is in the domain of the extended generator \mathcal{L}_ℓ for each ℓ . It remains to show that $V(x)$ satisfies (3.12). For any $x \in O_\ell$ we have

$$\frac{\mathcal{L}_\ell V(x)}{V(x)} = -\frac{\gamma}{(1 + \|x\|^2)^{\frac{1}{2}}} \langle \nabla_x U(x), x \rangle + \frac{\gamma n}{(1 + \|x\|^2)^{\frac{1}{2}}} + \frac{\gamma^2 \|x\|^2}{1 + \|x\|^2}. \quad (\text{A.2})$$

For $|x| \geq R$, using Lemma A.4 we have

$$\frac{\mathcal{L}_\ell V(x)}{V(x)} \leq -\frac{\gamma}{(1 + \|x\|^2)^{\frac{1}{2}}} \alpha \|x\| + \frac{\gamma n}{(1 + \|x\|^2)^{\frac{1}{2}}} + \frac{\gamma^2 \|x\|^2}{1 + \|x\|^2}.$$

Since $z/\sqrt{1+z^2}$ is increasing for $z > 0$ we can use that $\|x\| \geq R$ to obtain the bound

$$\frac{\mathcal{L}_\ell V(x)}{V(x)} \leq \gamma \left(\gamma - \alpha \frac{R}{(1 + R^2)^{\frac{1}{2}}} \right) + \frac{\gamma n}{(1 + R^2)^{\frac{1}{2}}}. \quad (\text{A.3})$$

Set $\gamma = \alpha/4$ and let R be sufficiently large that $\frac{\gamma n}{(1+R^2)^{\frac{1}{2}}} \leq \alpha/4$, $R/(1+R^2)^{\frac{1}{2}} \geq 3/4$ and that $\frac{\gamma n}{(1+R^2)^{\frac{1}{2}}} \leq \alpha^2/16$ then we have

$$\frac{\mathcal{L}_\ell V(x)}{V(x)} \leq -\frac{1}{16} \alpha^2.$$

Therefore Hypothesis 3.2 is satisfied and by Theorem 3.3 we have that X_t^x is exponentially ergodic.

In order to show that π is given by (3.7) it is sufficient to show that for some core \mathcal{C} of the domain of \mathcal{L} we have

$$\int_{\mathbb{R}_+^n} \mathcal{L}\varphi(x)e^{-U^{\lambda,b}(x)}dx = 0 \quad \text{for all } \varphi \in \mathcal{C}.$$

Here we are viewing \mathcal{L} as the infinitesimal generator of the semigroup \mathcal{P}_t in the space $C_0(\mathbb{R}_+^n)$ (i.e. the set of continuous functions vanishing at infinity endowed with the supremum norm). Let

$$\mathcal{C} = \{\varphi \in C^\infty((0, \infty)^n) \cap C_b^1(\mathbb{R}_+^n) : \mathcal{L}\varphi \in C_0(\mathbb{R}_+^n), \partial_{x_i}\varphi(x) = 0 \text{ if } x_i = 0\}.$$

That is \mathcal{C} is the set of functions which are infinitely differentiable in the interior of \mathbb{R}_+^n , are C^1 up to the boundary, is bounded with bounded first order derivative, the generator applied to φ belongs to $C_0(\mathbb{R}_+^n)$ and satisfies the boundary conditions of the PDE (3.9). Since \mathcal{P}_t preserves the set \mathcal{C} , and \mathcal{C} is contained within the domain of \mathcal{L} and is dense in $C_0(\mathbb{R}_+^n)$ we have that \mathcal{C} is a core of \mathcal{L} by [26, Proposition 3.3]. Fix $f \in \mathcal{C}$ then by integration by parts

$$\begin{aligned} \int_{\mathbb{R}_+^n} \mathcal{L}\varphi(x)e^{-U^{\lambda,b}(x)}dx &= \sum_{i=1}^n \int_{\mathbb{R}_+^n} (\partial_{x_i} U^{\lambda,b}(x)\partial_{x_i}\varphi(x) + \partial_{x_i}^2 \varphi(x)) e^{-U^{\lambda,b}(x)}dx \\ &= \sum_{i=1}^n \int_{\mathbb{R}_+^n} [-\partial_{x_i} U^{\lambda,b}(x)\partial_{x_i}\varphi(x) + \partial_{x_i} U^{\lambda,b}(x)\partial_{x_i}\varphi(x)] e^{-U^{\lambda,b}(x)}dx \\ &\quad + \left[\partial_{x_i}\varphi(x)e^{-U^{\lambda,b}(x)} \right] \Big|_{x_i=0} \\ &= 0. \end{aligned}$$

□

References

- [1] Assyr Abdulle. *Explicit Stabilized Runge–Kutta Methods*, pages 460–468. Encyclopedia of Applied and Computational Mathematics, 2015.
- [2] Assyr Abdulle, Ibrahim Almuslimani, and Gilles Vilmart. Optimal explicit stabilized integrator of weak order 1 for stiff and ergodic Stochastic Differential Equations. *SIAM/ASA Journal on Uncertainty Quantification*, 6(2):937–964, 1 2018.
- [3] Abdullah H. Al-Shabli, Xiaojian Xu, Ivan Selesnick, and Ulugbek S. Kamilov. Bregman plug-and-play priors. *arXiv preprint arXiv:2202.02388*, 2022.
- [4] Yoann Altmann, Reuben Aspden, Miles Padgett, and Steve McLaughlin. A Bayesian approach to denoising of single-photon binary images. *IEEE Transactions on Computational Imaging*, 3(3):460–471, 2017.
- [5] Yoann Altmann, Steve McLaughlin, and Miles Padgett. Unsupervised restoration of subsampled images constructed from Geometric and Binomial data. In *2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 1–5, 2017.
- [6] Yoann Altmann, Ximing Ren, Aongus McCarthy, Gerald S. Buller, and Steve McLaughlin. Lidar waveform-based analysis of depth images constructed using sparse single-photon data. *IEEE Transactions on Image Processing*, 25(5):1935–1946, may 2016.
- [7] Dominique Bakry, FRANCK BARTHE, and Patrick Cattiaux. A simple proof of the Poincaré inequality for a large class of probability measures including the log-concave case. *Electronic Communications in Probability*, 13, 02 2008.

- [8] Johnathan M. Bardsley and Per Christian Hansen. MCMC algorithms for computational UQ of nonnegativity constrained linear inverse problems. *SIAM Journal on Scientific Computing*, 42(2):A1269–A1288, 2020.
- [9] Mario Bertero, Patrizia Boccacci, and Valeria Ruggiero. Inverse imaging with Poisson data. *IOP Publish*, 2018.
- [10] Mireille Bossy, Emmanuel Gobet, and Denis Talay. A symmetrized Euler scheme for an efficient approximation of reflected diffusions. *Journal of Applied Probability*, 41(3):877–889, 2004.
- [11] Sébastien Bubeck, Ronen Eldan, and Joseph Lehec. Sampling from a log-concave distribution with projected Langevin Monte Carlo. *Discrete & Computational Geometry*, 59(4):757–783, 2018.
- [12] Xiaohao Cai, Marcelo Pereyra, and Jason D McEwen. Uncertainty quantification for radio interferometric imaging – I. proximal MCMC methods. *Monthly Notices of the Royal Astronomical Society*, 480(3):4154–4169, 7 2018.
- [13] Thomas Cass, Dan Crisan, Paul Dobson, and Michela Ottobre. Long-time behaviour of degenerate diffusions: UFG-type SDEs and time-inhomogeneous hypoelliptic processes. *Electronic Journal of Probability*, 26:1–72, 2021.
- [14] Patrick Cattiaux, José R León, and Clémentine Prieur. Invariant density estimation for a reflected diffusion using an Euler scheme. *Monte Carlo Methods and Applications*, 23(2):71–88, 2017.
- [15] Apostolos Chalkis, Vissarion Fisikopoulos, Marios Papachristou, and Elias Tsigaridas. Truncated log-concave sampling with reflective Hamiltonian Monte Carlo. *arXiv preprint arXiv:2102.13068*, 2021.
- [16] Antonin Chambolle and Thomas Pock. An introduction to continuous optimization for imaging. *Acta Numerica*, 25:161–319, 2016.
- [17] Stanley H. Chan, Xiran Wang, and Omar A. Elgendy. Plug-and-play ADMM for image restoration: Fixed-point convergence and applications. *IEEE Transactions on Computational Imaging*, 3(1):84–98, 2017.
- [18] Xiang Cheng, Niladri S. Chatterji, Peter L. Bartlett, and Michael I. Jordan. Underdamped Langevin MCMC: A non-asymptotic analysis. In Sébastien Bubeck, Vianney Perchet, and Philippe Rigollet, editors, *Proceedings of the 31st Conference On Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pages 300–323. PMLR, 06–09 Jul 2018.
- [19] Emilie Chouzenoux, Anna Jezierska, Jean-Christophe Pesquet, and Hugues Talbot. A convex approach for image restoration with exact Poisson-Gaussian likelihood. *SIAM Journal on Imaging Sciences*, 8, 03 2015.
- [20] Patrick Louis Combettes and Jean-Christophe Pesquet. Proximal Splitting Methods in Signal Processing. In Bauschke, H.H.; Burachik, R.S.; Combettes, P.L.; Elser, V.; Luke, D.R.; Wolkowicz, and H. (Eds.), editors, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 185–212. Springer, 2011.
- [21] C. Costantini, B. Pacchiarotti, and F. Sartoretto. Numerical approximation for functionals of reflecting diffusion processes. *SIAM Journal on Applied Mathematics*, 58(1):73–102, 1998.
- [22] Jean-Dominique Deuschel and Lorenzo Zambotti. Bismut–Elworthy’s formula and random walk representation for SDEs with reflection. *Stochastic Processes and their Applications*, 115(6):907–925, 2005.
- [23] P Dobson, I Fursov, G Lord, and M Ottobre. Reversible and non-reversible Markov chain Monte Carlo algorithms for reservoir simulation problems. *Computational Geosciences*, 24(3):1301–1313, 2020.
- [24] Alain Durmus, Eric Moulines, and Marcelo Pereyra. Efficient Bayesian computation by proximal Markov chain Monte Carlo: when Langevin meets Moreau. *SIAM Journal on Imaging Sciences*, 11(1):473–506, 2018.
- [25] M. Eisaman, Jingwen Fan, Alan Migdall, and Sergey Polyakov. Invited review article: Single-photon sources and detectors. *The Review of scientific instruments*, 82:071101, 07 2011.

- [26] Stewart N Ethier and Thomas G Kurtz. *Markov processes: characterization and convergence*, volume 282. John Wiley & Sons, 2009.
- [27] Mário A. T. Figueiredo and José M. Bioucas-Dias. Restoration of Poissonian Images using Alternating Direction Optimization. *IEEE Transactions on Image Processing*, 19(12):3133–3145, 2010.
- [28] Mark Iosifovich Freidlin. *Functional Integration and Partial Differential Equations.(AM-109), Volume 109*. Princeton university press, 2016.
- [29] Graham Gibson, Steven Johnson, and Miles Padgett. Single-pixel imaging 12 years on. *Optics Express*, 28, 08 2020.
- [30] Emmanuel Gobet. Euler schemes and half-space approximation for the simulation of diffusion in a domain. *ESAIM: Probability and Statistics*, 5:261–297, 2001.
- [31] Jacob Vorstrup Goldman, Torben Sell, and Sumeetpal Sidhu Singh. Gradient-based Markov chain Monte Carlo for Bayesian inference with non-differentiable priors. *Journal of the American Statistical Association*, 0(0):1–12, 2021.
- [32] Peter J. Green, Krzysztof Łatuszyński, Marcelo Pereyra, and Christian P. Robert. Bayesian computation: a summary of the current state, and samples backwards and forwards. *Statistics and Computing*, 25:835–862, 7 2015.
- [33] Abderrahim Halimi, Yoann Altmann, Aongus McCarthy, Ximing Ren, Rachael Tobin, Gerald S. Buller, and Steve McLaughlin. Restoration of intensity and depth images constructed using sparse single-photon data. In *2016 24th European Signal Processing Conference (EUSIPCO)*, pages 86–90, 2016.
- [34] Zachary T. Harmany, Roummel F. Marcia, and Rebecca M. Willett. This is SPIRAL-TAP: Sparse Poisson intensity reconstruction algorithms—theory and practice. *IEEE Transactions on Image Processing*, 21(3):1084–1096, 2012.
- [35] J. M. Harrison and R. J. Williams. Brownian models of open queueing networks with homogeneous customer populations. *Stochastics*, 22(2):77–115, 1987.
- [36] Marlis Hochbruck and Alexander Ostermann. Exponential integrators. *Acta Numerica*, 19:209–286, 2010.
- [37] Thorsten Hohage and Frank Werner. Inverse problems with Poisson data: Statistical regularization theory, applications and algorithms. *Inverse Problems*, 32:093001, 09 2016.
- [38] Matthew Holden, Marcelo Pereyra, and Konstantinos C. Zygalakis. Bayesian imaging with data-driven priors encoded by neural networks: Theory, methods, and algorithms. *arXiv preprint arXiv:2103.10182*, 2021.
- [39] Ahmed Kirmani, Dheera Venkatraman, Dongeek Shin, Andrea Colaço, Franco N. C. Wong, Jeffrey H. Shapiro, and Vivek K Goyal. First-photon imaging. *Science*, 343(6166):58–61, 2014.
- [40] Young-Jun Ko and Matthias W. Seeger. Expectation propagation for rectified linear Poisson regression. In Geoffrey Holmes and Tie-Yan Liu, editors, *Asian Conference on Machine Learning*, volume 45 of *Proceedings of Machine Learning Research*, pages 253–268, Hong Kong, 20–22 Nov 2016. PMLR.
- [41] Andrew Lamperski. Projected Stochastic Gradient Langevin algorithms for constrained sampling and non-convex learning. In *Conference on Learning Theory*, pages 2891–2937. PMLR, 2021.
- [42] Rémi Laumont, Valentin De Bortoli, Andrés Almansa, Julie Delon, Alain Durmus, and Marcelo Pereyra. Bayesian imaging using Plug & Play priors: when Langevin meets Tweedie. *SIAM Journal on Imaging Sciences*, 15(2):701–737, 2022.
- [43] B Leimkuhler, A Sharma, and Michael V Tretyakov. Simplest random walk for approximating Robin boundary value problems and ergodic limits of reflected diffusions. *arXiv preprint arXiv:2006.15670*, 2020.

- [44] Yi-An Ma, Emily B Fox, Tianqi Chen, and Lei Wu. Irreversible samplers from jump and continuous Markov processes. *Statistics and Computing*, 29(1):177–202, 2019.
- [45] Willem Marais and Rebecca Willett. Proximal-gradient methods for Poisson image reconstruction with BM3D-based regularization. In *2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 1–5, 2017.
- [46] Yosra Marnissi, Yuling Zheng, Emilie Chouzenoux, and Jean-Christophe Pesquet. A variational Bayesian approach for image restoration—application to image deblurring with Poisson–Gaussian noise. *IEEE Transactions on Computational Imaging*, 3(4):722–737, 2017.
- [47] Yosra Marnissi, Yuling Zheng, and Jean-Christophe Pesquet. Fast variational Bayesian signal recovery in the presence of Poisson–Gaussian noise. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3964–3968, 2016.
- [48] Jonathan C Mattingly, Andrew M Stuart, and Desmond J Higham. Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stochastic processes and their applications*, 101(2):185–232, 2002.
- [49] Robert I. McLachlan and G. Reinout W. Quispel. Splitting methods. *Acta Numerica*, 11:341–434, 2002.
- [50] Sean P. Meyn and R. L. Tweedie. Stability of Markovian processes III: Foster-Lyapunov criteria for continuous-time processes. *Advances in Applied Probability*, 25(3):518–548, 1993.
- [51] Sean P Meyn and Richard L Tweedie. *Markov chains and stochastic stability*. Springer Science & Business Media, 2012.
- [52] Marcelo Pereyra. Proximal Markov chain Monte Carlo algorithms. *Statistics and Computing*, 26:745–760, 2016.
- [53] Marcelo Pereyra. Maximum-a-posteriori estimation with bayesian confidence regions. *SIAM Journal on Imaging Sciences*, 10(1):285–302, 2017.
- [54] Marcelo Pereyra, Luis Vargas Miele, and Konstantinos C. Zygalakis. Accelerating proximal Markov chain Monte Carlo by using an explicit stabilised method. *SIAM Journal on Imaging Sciences*, 13(2):905–935, 2020.
- [55] Marcelo Pereyra, Philip Schniter, Emilie Chouzenoux, Jean-Christophe Pesquet, Jean-Yves Tournet, Alfred O. Hero, and Steve McLaughlin. A survey of stochastic simulation and optimization methods in signal processing. *IEEE Journal of Selected Topics in Signal Processing*, 10(2):224–241, 3 2016.
- [56] Roger Pettersson. Penalization schemes for reflecting Stochastic Differential Equations. *Bernoulli*, 3(4):403–414, 1997.
- [57] Natesh S. Pillai. Optimal scaling for the Proximal Langevin Algorithm in high dimensions. *arXiv preprint arXiv:2204.10793*, 2022.
- [58] Joshua Rapp and Vivek K Goyal. A few photons among many: Unmixing signal and noise for photon-efficient active imaging. *IEEE Transactions on Computational Imaging*, 3(3):445–459, 2017.
- [59] Audrey Repetti, Marcelo Pereyra, and Yves Wiaux. Scalable Bayesian uncertainty quantification in imaging inverse problems via convex optimization. *SIAM Journal on Imaging Sciences*, 12(1):87–118, 2019.
- [60] Christian Robert. *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer Science & Business Media, 2007.
- [61] Christian P. Robert and George Casella. *Monte Carlo Statistical Methods (Springer Texts in Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2005.

- [62] Gareth O Roberts and Richard L Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, pages 341–363, 1996.
- [63] R.T. Rockafellar, M. Wets, and R.J.B. Wets. *Variational Analysis*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2009.
- [64] Ernest Ryu, Jialin Liu, Sicheng Wang, Xiaohan Chen, Zhangyang Wang, and Wotao Yin. Plug-and-play methods provably converge with properly trained denoisers. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 5546–5557. PMLR, 09–15 Jun 2019.
- [65] Yash Sanghvi, Abhiram Gnanasambandam, and Stanley Chan. Photon-limited deblurring using algorithm unrolling. In *NeurIPS 2021 Workshop on Deep Learning and Inverse Problems*, 2021.
- [66] Jesus Maria Sanz-Serna and Konstantinos C Zygalakis. Wasserstein distance estimates for the distributions of numerical approximations to ergodic Stochastic Differential Equations. *Journal of Machine Learning Research*, 22(242):1–37, 2021.
- [67] Kanji Sato, Akiko Takeda, Reiichiro Kawai, and Taiji Suzuki. Convergence error analysis of Reflected Gradient Langevin Dynamics for globally optimizing non-convex constrained problems. *arXiv preprint arXiv:2203.10215*, 2022.
- [68] Dongeek Shin, Ahmed Kirmani, Vivek K Goyal, and Jeffrey H. Shapiro. Photon-efficient computational 3-D and reflectivity imaging with single-photon detectors. *IEEE Transactions on Computational Imaging*, 1(2):112–125, 2015.
- [69] J-L Starck and Fionn Murtagh. *Astronomical image and data analysis*. Springer Science & Business Media, 2007.
- [70] Hiroshi Tanaka. Stochastic Differential Equations with reflecting boundary condition in convex regions. *Hiroshima Mathematical Journal*, 9(1):163 – 177, 1979.
- [71] Ana Fernandez Vidal, Valentin De Bortoli, Marcelo Pereyra, and Alain Durmus. Maximum likelihood estimation of regularisation parameters in high-dimensional inverse problems: an empirical Bayesian approach. part I: Methodology and experiments. *SIAM Journal on Imaging Sciences*, 13(4):1945–1989, 2020.
- [72] M. Vono, N. Dobigeon, and P. Chainais. Bayesian image restoration under Poisson noise and log-concave prior. In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1712–1716, 2019.
- [73] David Williams. *Probability with martingales*. Cambridge university press, 1991.
- [74] Dan Yao, Stephen McLaughlin, and Yoann Altmann. Patch-based image restoration using expectation propagation. *SIAM Journal on Imaging Sciences*, 15(1):192–227, 2022.
- [75] Chen Zhang, Simon Arridge, and Bangti Jin. Expectation propagation for Poisson data. *Inverse Problems*, 35(8):085006, jul 2019.
- [76] Qingping Zhou, Tengchao Yu, Xiaoqun Zhang, and Jinglai Li. Bayesian inference and uncertainty quantification for image reconstruction with Poisson data. *SIAM Journal on Imaging Sciences*, 13(1):29–52, 2020.