



**HAL**  
open science

# CREDIBILITY INTERVAL DESIGN FOR COVID19 REPRODUCTION NUMBER FROM NONSMOOTH LANGEVIN-TYPE MONTE CARLO SAMPLING

Hugo Artigas, Barbara Pascal, Gersende Fort, Patrice Abry, Nelly Pustelnik

► **To cite this version:**

Hugo Artigas, Barbara Pascal, Gersende Fort, Patrice Abry, Nelly Pustelnik. CREDIBILITY INTERVAL DESIGN FOR COVID19 REPRODUCTION NUMBER FROM NONSMOOTH LANGEVIN-TYPE MONTE CARLO SAMPLING. 2021. hal-03371837v2

**HAL Id: hal-03371837**

**<https://hal.science/hal-03371837v2>**

Preprint submitted on 20 Feb 2022 (v2), last revised 13 Jun 2022 (v3)

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

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

# CREDIBILITY INTERVAL DESIGN FOR COVID19 REPRODUCTION NUMBER FROM NONSMOOTH LANGEVIN-TYPE MONTE CARLO SAMPLING

Hugo Artigas<sup>(1)</sup>, Barbara Pascal<sup>(2)</sup>, Gersende Fort<sup>(3)</sup>, Patrice Abry<sup>(4)</sup>, Nelly Pustelnik<sup>(4,5)</sup>

<sup>(1)</sup> Ecole Polytechnique, Paris, France

<sup>(2)</sup> Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, France (barbara.pascal@univ-lille.fr)

<sup>(3)</sup> CNRS, Institut de Mathématiques de Toulouse, France (gersende.fort@math.univ-toulouse.fr)

<sup>(4)</sup> Univ Lyon, Ens de Lyon, Univ Claude Bernard, CNRS, Laboratoire de Physique, Lyon, France. (firstname.lastname@ens-lyon.fr)

<sup>(5)</sup> ISPGROUP/ICTEAM, UCLouvain, Belgium

## ABSTRACT

Monitoring the Covid19 pandemic is critical to design sanitary policies. Recently, reliable estimates of the pandemic reproduction number were obtained from a nonsmooth convex optimization procedure designed to fit epidemiology requirements and to be robust to the low quality of the data (outliers, pseudo-seasonalities, ...). Applied to daily new infection counts made public by National Health Agencies and centralized by Johns Hopkins University, robust estimates of the reproduction number for 200+ countries are updated and published every day. To further improve estimation procedures and also, and mostly, increase their usability by epidemiologists, the present work exploits the Bayesian paradigm and derives a new Monte Carlo method to sample from a nonsmooth convex a posteriori distribution. This new sampler stems from an original combination of the Langevin Monte Carlo algorithm with Proximal operators. Its relevance and practical efficiency to produce meaningful credibility intervals for the Covid19 reproduction number are assessed from several indices quantifying the statistics of the Monte Carlo chains, and making use of real daily new infection counts.

**Index Terms**— Nonsmooth Langevin Monte Carlo sampler, Bayesian credibility interval, reproduction number, Covid19.

## 1. INTRODUCTION

**Context.** Monitoring the time evolution of the Covid19 pandemic constitutes a critical stake to design counter measures. Pandemic intensity is often assessed by the reproduction number,  $R$ , that quantifies the number of second infections stemming from one same primary infection (cf., e.g., [1, 2, 3, 4, 5]). The online and daily estimate of  $R$  turned however extremely difficult during the Covid19 pandemic, mostly because of issues most countries faced in collecting reliable daily new infection counts yielding low-quality data (missing counts, outliers, seasonalities, ...). Therefore, assessing the confidence that can be granted to point estimates is a critical and difficult challenge, motivating the present work.

**Related works.** While refined pandemic assessment can efficiently be achieved when the pandemic has passed from elaborated compartmental models and Bayesian estimates (cf., e.g., [6, 7]), recently it has been shown that within pandemic, reliable epidemic intensity estimates can be obtained from nonsmooth convex optimization procedures [8, 9]. The functional to minimize is built from the epidemic

propagation model proposed in [5], whose quality is to focus on a unique parameter, the reproduction number  $R$ , while preserving the key epidemic propagation feature, the so-called *serial interval function*  $\Phi(t)$ , that quantifies the probability that symptoms today are caused by infection in the past few days. With additional temporal regularity and positivity constrained in the functional, the procedure in [8, 9] delivers epidemiologically realistic estimates that are robust to the low quality of the data. It does not however provide assessment of credibility intervals, an issue what we intend to address here. **Goals, contributions and outline.** The goal of the present work is to explore the potential of Langevin-type stochastic sampling schemes to produce credibility intervals for the evolution of the Covid19 reproduction number  $R$ . To that end, the pandemic model in [5] and the one in [8, 9] are first recast into a Bayesian framework, with careful analysis and writing of the corresponding a posteriori distribution (cf. Section 2). The core methodological contribution is detailed in Section 3: The *Proximal-Gradient dual* stochastic sampling scheme, refining the classical Metropolis Adjusted Langevin Algorithm, is devised to sample a generic class of a posteriori distributions, including the one introduced in Section 2. The originality of the proposed sampler is to account for the difficulties stemming from the specificities of the a posteriori distributions considered here.

To assess the relevance of the proposed samplers in producing credibility intervals for  $R$ , and to compare it against simpler *Random Walk*-type samplers, indices quantifying the statistics of the sampled *chains* are defined in Section 4. They are measured on real Covid19 data, consisting of daily infection counts reported by the National Health Authorities of countries around the world and collected and made available by the Johns Hopkins University (cf. Section 4 for a description). The practical usability of the sampling strategies and of the achieved credibility intervals is discussed both from a technical and epidemiology surveillance perspectives, using Covid19 data from several countries and different time periods.

## 2. REPRODUCTION NUMBER BAYESIAN MODEL

The pandemic model used in this paper elaborates on the one proposed in [5] in order to account for the low quality of the data as fully detailed and motivated in [9]. In this work, this model is recast into a Bayesian framework. The statistical model associated with the vector of  $T$  observations  $\mathbf{Z} = (\mathbf{Z}_1, \dots, \mathbf{Z}_T)^\top \in \mathbb{N}^T$ , consisting of new daily infection counts, is indexed by the unknown

$$\theta := (\mathbf{R}, \mathbf{O}) = (R_1, \dots, R_T, O_1, \dots, O_T) \in (\mathbb{R}_+)^T \times \mathbb{R}^T$$

H. Artigas and G. Fort are partly funded by the *Fondation Simone et Cino Del Duca, Institut de France*.

P. Abry partially supported by Grant 80PRIME-2021 CNRS.

with  $R_t$  the reproduction number at time  $t$  and  $O_t$  the *outliers* modeling the low quality of the data at time  $t$  (irrelevant/missing counts, mis-reported counts, pseudo-seasonal effects, ...). For any  $\theta$ , the conditional distribution of  $Z_t$  given the past  $Z_1, \dots, Z_{t-1}$  and initial values  $(Z_{-\tau_\phi+1}, \dots, Z_0)$  is a Poisson distribution with intensity

$$p_t(\theta) := R_t \sum_{u=1}^{\tau_\phi} \Phi_u Z_{t-u} + O_t, \quad (1)$$

where the *serial interval function*  $(\Phi_u)_{1 \leq u \leq \tau_\phi}$  accounts for epidemiological mechanisms: It quantifies the random delays between the onset of symptoms in a primary case and in secondary cases [5, 10, 6].  $(\Phi_u)_{1 \leq u \leq \tau_\phi}$  is assumed known and following [11, 12], it is classically modeled by a Gamma distribution truncated over  $\tau_\phi = 26$  days with mean and standard deviation of 6.6 and 3.5 days. By convention, a Poisson distribution with null intensity is the Dirac mass at zero. This implies that the negative log-likelihood of the observations  $\mathbf{Z}$  is given by  $(0 \ln 0 = 0)$  by convention)

$$f(\theta) := \begin{cases} -\sum_{t=1}^T (Z_t \ln p_t(\theta) - p_t(\theta)) & \text{if } \theta \in \mathcal{D}, \\ +\infty & \text{otherwise,} \end{cases} \quad (2)$$

defined on the measurable set

$$\mathcal{D} := \{\theta \in (\mathbb{R}_+)^T \times \mathbb{R}^T : p_t(\theta) > 0 \text{ for } t \text{ s.t. } Z_t > 0\} \cup \{\theta \in (\mathbb{R}_+)^T \times \mathbb{R}^T : p_t(\theta) \geq 0 \text{ for } t \text{ s.t. } Z_t = 0\}. \quad (3)$$

Interpreting the regularity constraints in [9], the negative logarithm of the a priori distribution of  $\theta$  is defined on  $(\mathbb{R}_+)^T \times \mathbb{R}^T$  by<sup>1</sup>

$$g(\theta) := \lambda_R \|\mathbf{D}_2 \mathbf{R}\|_1 + \lambda_O \|\mathbf{O}\|_1; \quad (4)$$

$\|\cdot\|_1$  denotes the  $L^1$ -norm.  $\mathbf{D}_2$  is the discrete-time second order derivative  $(T-2) \times T$  matrix:

$$\mathbf{D}_2 := \frac{1}{\sqrt{6}} \begin{bmatrix} 1 & -2 & 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & -2 & 1 & 0 & \dots & 0 \\ \dots & & & & & \dots & \\ 0 & \dots & & & 1 & -2 & 1 \end{bmatrix}, \quad (5)$$

with normalized rows, and  $\lambda_R, \lambda_O$  are (fixed) positive *regularization hyperparameters*, balancing the strengths of the different constraints one against the others and against the likelihood. Under the a priori distribution,  $\mathbf{R}$  and  $\mathbf{O}$  are independent and distributed resp. as a (non stationary) Laplace AR(2) process and a Laplace distribution. Even if the optima of the a priori distribution satisfy  $\mathbf{D}_2 \mathbf{R} = 0$  and  $\mathbf{O} = 0$ , for samples  $(\mathbf{R}, \mathbf{O})$  from the a priori distribution, neither  $\mathbf{D}_2 \mathbf{R}$  nor  $\mathbf{O}$  are sparse.

Combining the likelihood (2) and the prior (4) leads to the a posteriori density with respect to the Lebesgue measure:

$$\pi(\theta) := \exp(-f(\theta) - g(\theta)) \mathbb{1}_{\mathcal{D}}(\theta), \quad (6)$$

where  $\mathbb{1}_{\mathcal{D}}$  is the  $\{0, 1\}$ -valued indicator function of  $\mathcal{D}$ . Up to an additive constant, the negative log-likelihood of a Poisson variable  $Z_t$  at  $\theta$  is the Kullback-Leibler distance  $d_{\text{KL}}(Z_t | p_t(\theta))$  between  $Z_t$  and  $p_t(\theta)$ . The negative log-density  $-\ln \pi$  is the criterion minimized in [9] for the reconstruction of  $\theta$ . It is proved in [9, 13] that when  $\sum_{u=1}^{\tau_\phi} \Phi_u Z_{t-u}$  is positive for at least two values  $t_*, t_{**}$ , a minimizer of  $-\ln \pi$  exists and the set of the minimizers is included in a level set  $\{\theta : f(\theta) = f_*\}$  of  $f$  (and thus, of  $g$ ). The a priori distribution implies that the R part of the optima of  $\ln \pi$  cannot vary too much across successive days – so that epidemiologists can extract local trends indicating whether the pandemic is increasing or decreasing.

<sup>1</sup>In the Bayesian setting, the distributions are often defined up to a multiplicative constant. We adopt this convention here.

### 3. NONSMOOTH LANGEVIN MONTE CARLO SAMPLERS

When the posterior distribution is known up to a normalizing constant (as in Section 2), estimation through the Bayesian paradigm entails the use of numerical tools such as Markov Chain Monte Carlo (MCMC) samplers designed to define an empirical distribution approximating the posterior distribution. The computation of  $\alpha$ -credibility regions (see, e.g., [14, chapter 5]) or of Bayesian estimators such as the posterior mean, the median or the Maximum A Posteriori (MAP), follows from a Monte Carlo (MC) approximation.

In this work, we propose original MCMC samplers targeting a generic density  $\pi$  of the form (6) and satisfying A1-A2.

**Assumptions on the target density  $\pi$ .**

**A1.**  $f$  and  $g$  are finite on a measurable set  $\mathcal{D}$  and  $f$  is continuously differentiable on the interior of  $\mathcal{D} \subseteq \mathbb{R}^d$ .

**A2.**  $g$  is blockwise: for  $j \in \{1, \dots, J\}$ , there exist matrices  $A_j \in \mathbb{R}^{c_j \times d_j}$ , with  $c_j \leq d_j$ , and proper, convex, lower semi-continuous functions  $g_j : \mathbb{R}^{c_j} \rightarrow ]-\infty, +\infty]$  such that  $\sum_{j=1}^J d_j = d$  and

$$\forall \theta := (\theta_1, \dots, \theta_J) \in \mathbb{R}^{d_1} \times \dots \times \mathbb{R}^{d_J}, \quad g(\theta) = \sum_{j=1}^J g_j(A_j \theta_j),$$

where, for each  $j \in \{1, \dots, J\}$ , (i) the proximal operator<sup>2</sup> of  $g_j$  possesses a closed-form expression<sup>3</sup>, (ii)  $A_j$  is full rank, so that it can be augmented into an invertible matrix  $\bar{A}_j \in \mathbb{R}^{d_j \times d_j}$ , (iii)  $g_j$  extends into  $\bar{g}_j : \mathbb{R}^{d_j} \rightarrow ]-\infty, +\infty]$  such that  $\forall \theta_j, g_j(A_j \theta_j) = \bar{g}_j(\bar{A}_j \theta_j)$ .

A1 requires the differentiability of  $f$  but not the Lipschitz property for its gradient; A2 does not assume that the proximal operator of  $g$  has a closed-form expression.

**Example.** For the target density  $\pi$  defined in Section 2, both A1 and A2 are valid. We have  $d = 2T$ ,  $J = 2$ ,  $\theta_1 = \mathbf{R}$ ,  $\theta_2 = \mathbf{O}$ ,  $g_1 = \lambda_R \|\cdot\|_1$ ,  $g_2 = \lambda_O \|\cdot\|_1$ ,  $A_1 = \mathbf{D}_2$  (see (5)), and  $A_2 = \mathbf{I}$ , the identity matrix.  $A_1$  can be augmented in many different ways, among which:

$$\bar{\mathbf{D}}_2 := \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ -2/\sqrt{5} & 1/\sqrt{5} & 0 & \dots & 0 \\ & & \mathbf{D}_2 & & \end{bmatrix}. \quad (7)$$

Another possibility is  $\bar{\mathbf{D}}_o$ , obtained from  $\bar{\mathbf{D}}_2$  by making its first two rows orthogonal, and orthogonal to the rows of  $\mathbf{D}_2$ . Then,  $\bar{g}_1 = \lambda_R \|(\cdot)_{3:T}\|_1$ , computed only on components 3 to  $T$ , and  $\bar{g}_2 = g_2$ .

**Metropolis Adjusted Langevin Algorithm (MALA) sampler and its limitations.** If  $-\ln \pi = f$  (i.e.,  $g = 0$ ) is a smooth function on  $\mathcal{D}$ , a popular MCMC sampler which takes benefit of a first order knowledge of  $\pi$ , relies on Langevin dynamics, stemming from *tempered Langevin diffusion* [16]. Given a  $d \times d$  matrix  $\Gamma$  and a positive step size  $\gamma$  chosen by the user, at each iteration  $\#(n+1)$  starting from the current point  $\theta^n$ , a jump is proposed to the point

$$\theta^{n+\frac{1}{2}} := \theta^n - \gamma \Gamma^\top \nabla f(\theta^n) + \sqrt{2\gamma} \Gamma \xi^{n+1}, \quad (8)$$

where  $\xi^{n+1} \sim \mathcal{N}_d(0, 1)$  is a standard  $\mathbb{R}^d$ -valued Gaussian distribution.  $A^\top$  denotes the transpose of the matrix  $A$ . This proposal is accepted ( $\theta^{n+1} = \theta^{n+\frac{1}{2}}$ ) or rejected ( $\theta^{n+1} = \theta^n$ ) through a Metropolis mechanism, and this yields the *Metropolis Adjusted Langevin Algorithm* (MALA, [17]). The idea of MALA methods is to drift the

<sup>2</sup> $\text{prox}_g(y) := \arg\min_{x \in \mathbb{R}^\ell} (g(x) + \|x - y\|^2/2)$  is well-defined for any proper, convex, lower semi-continuous function  $g : \mathbb{R}^\ell \rightarrow ]-\infty, +\infty]$ .  $\|\cdot\|$  denotes the Euclidean norm in  $\mathbb{R}^\ell$ .

<sup>3</sup>See [15] to have an exhaustive list of known proximal operators.

proposed moves towards areas of high probability under  $\pi$ , by using the information provided by  $\nabla \ln \pi$ .

**Proximal-Gradient dual sampler** (PGdual). In the present work,  $-\ln \pi$  is not smooth (see A2): this calls for the use of Proximal Langevin-based MC methods. The literature provides examples of MC samplers combining a Langevin approach and a Proximal approach (see, e.g., [18, 19, 20, 21, 22, 23, 24, 25]) but none of these algorithms directly apply to the framework A1-A2. Indeed, the density  $\pi$  may be positive for non-sparse  $(D_2\mathbf{R}, \mathbf{O})$ . Further, A2 assumes that  $g_j$  has an explicit proximal operator for any  $j \in \{1, \dots, J\}$  which does not imply that this holds true for  $g$ , notably when  $g$  involves a linear operator, e.g., in (4). Finally, we want to promote a method which takes benefit of both the blockwise separable expression of  $g$  and the full rank property of matrices  $\mathbf{A}_j$  by computing the proximal operator separately on each range space of  $\bar{\mathbf{A}}_j$ . These limitations led us to propose the following original and general Proximal-Gradient dual (PGdual) Algorithm (sketched in Algorithm 1), which is among the Metropolis-Hastings samplers with Gaussian proposal; the main originality is the definition of the drift term. At iteration  $\#(n+1)$ , a candidate is proposed, whose block  $\#j$  is

$$\theta_j^{n+\frac{1}{2}} := \mu_j(\theta^n) + \sqrt{2\gamma_j} \Gamma_j \xi_j^{n+1}, \quad \forall j \in \{1, \dots, J\}, \quad (9)$$

where  $\xi_j^{n+1} \sim \mathcal{N}_{d_j}(0, \mathbf{I})$  and

$$\mu_j(\theta) := \bar{\mathbf{A}}_j^{-1} \text{prox}_{\gamma_j \bar{g}_j} \left( \bar{\mathbf{A}}_j \theta_j - \gamma_j \bar{\mathbf{A}}_j^{-\top} \nabla_j f(\theta) \right); \quad (10)$$

$\nabla_j$  denotes the gradient with respect to the block  $\#j$  and  $\mathbf{A}^{-\top}$  is the inverse of  $\mathbf{A}^\top$ .  $\mu_j(\theta^n)$  performs a proximal-gradient iteration associated to a part of the original composite function in the dual space, consisting in the range of the matrix  $\bar{\mathbf{A}}_j$ ; then sends each block  $\#j$  back in the direct space through the mapping  $\bar{\mathbf{A}}_j^{-1}$ . Formally,  $-\ln \pi$  transferred in the dual space reads

$$(\tilde{\theta}_1, \dots, \tilde{\theta}_J) \mapsto f(\bar{\mathbf{A}}_1^{-1} \tilde{\theta}_1, \dots, \bar{\mathbf{A}}_J^{-1} \tilde{\theta}_J) + \sum_{j=1}^J \bar{g}_j(\tilde{\theta}_j).$$

Introducing  $q_j(\theta, \theta'_j)$  the Gaussian kernel on  $\mathbb{R}^{d_j}$  centered at  $\mu_j(\theta)$ , with covariance matrix  $2\gamma_j \Gamma_j \Gamma_j^\top$  and evaluated at  $\theta'_j$ , the acceptance-rejection (AR) step defined Line 5 makes Algorithm 1 to have  $\pi$  as unique invariant distribution.

There is a parallel between the preconditioned gradient appearing in the drift term (10) of the proposed point (9), and the scaled Langevin dynamics (8). For this reason, we advocate  $\Gamma_j = \bar{\mathbf{A}}_j^{-1}$  for any  $j$ , whatever the proximal operator is.

Our recent work [13] shows how to apply PGdual in a more general setting than A 2, when there are many penalty terms for each block  $\#j$ . It also discusses how PGdual is related to other proximal-gradient based extensions of the MALA sampler.

**Example (to follow).** For  $\pi$  defined in Section 2, the  $\mu_j$ 's are

$$\mu_1(\theta) := \bar{\mathbf{D}}^{-1} \text{prox}_{\lambda_R \|(\cdot)_{3:T}\|_1} \left( \bar{\mathbf{D}} \mathbf{R} - \gamma_1 \bar{\mathbf{D}}^{-\top} \nabla_1 f(\theta) \right), \quad (11)$$

$$\mu_2(\theta) := \text{prox}_{\lambda_O \|\cdot\|_1} (\mathbf{O} - \gamma_2 \nabla_2 f(\theta)), \quad (12)$$

where  $\bar{\mathbf{D}}$  is a  $(T \times T)$  invertible augmentation of  $D_2$ , see, e.g., Eq (7).

#### 4. CREDIBILITY INTERVALS FROM COVID19 DATA

**Covid19 data.** The data used here are part of a large data set available from the *Johns Hopkins University*<sup>4</sup>, consisting of daily new

<sup>4</sup> <https://coronavirus.jhu.edu/>

---

#### Algorithm 1: Proximal-Gradient dual (PGdual)

---

**Data:**  $d_j \times d_j$  matrices  $\Gamma_j, \gamma_j > 0, N_{\max} \in \mathbb{N}_*, \theta^0 \in \mathcal{D}$

**Result:** A  $\mathcal{D}$ -valued sequence  $\{\theta^n, n \in 0, \dots, N_{\max}\}$

```

1 for  $n = 0, \dots, N_{\max} - 1$  do
2   for  $j = 1, \dots, J$  do
3     Sample  $\xi_j^{n+1} \sim \mathcal{N}_{d_j}(0, \mathbf{I})$ ;
4     Set  $\theta_j^{n+\frac{1}{2}} = \mu_j(\theta^n) + \sqrt{2\gamma_j} \Gamma_j \xi_j^{n+1}$ ;
5   Set  $\theta^{n+1} = \theta^{n+\frac{1}{2}}$  with probability
      
$$1 \wedge \frac{\pi(\theta^{n+\frac{1}{2}})}{\pi(\theta^n)} \prod_{j=1}^J \frac{q_j(\theta^{n+\frac{1}{2}}, \theta_j^n)}{q_j(\theta^n, \theta_j^{n+\frac{1}{2}})}$$

      and  $\theta^{n+1} = \theta^n$  otherwise.

```

---

infection counts for around 200 countries, by National Health Authorities. The focus is here on the evolution of  $\mathbf{R}_t$  on a realistic time period of  $T = 35$  days (5 weeks) of pandemic. To account for the time support of the serial function  $\Phi$ , we use  $\tau_\phi = 26$  additional observations for the initialization  $\mathbf{Z}_{-\tau_\phi+1}, \dots, \mathbf{Z}_0$ . Counts from January, 1st 2022 to February, 4th 2022 for Italy are used for the performance assessments (see Figure 1, Figure 2). However, tools and methods proposed here have been applied to numerous countries for various time periods of interest (see, e.g., Figure 3).

**MCMC samplers.** The proposal mechanisms of PGdual are

$$\mathbf{R}^{n+\frac{1}{2}} = \mu_1(\theta^n) + \sqrt{2\gamma_1} \bar{\mathbf{D}}^{-1} \xi_1^{n+1}, \quad \mathbf{O}^{n+\frac{1}{2}} = \mu_2(\theta^n) + \sqrt{2\gamma_2} \xi_2^{n+1}$$

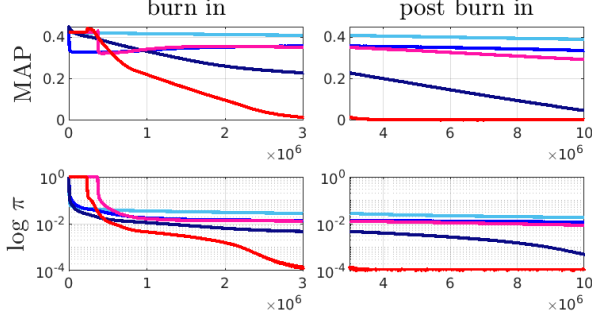
where  $\mu_1, \mu_2$  are defined by (11), (12). Two versions of PGdual are compared: PGdual Invert when  $\bar{\mathbf{D}} = \bar{\mathbf{D}}_2$ , and PGdual Ortho when  $\bar{\mathbf{D}} = \bar{\mathbf{D}}_o$  (see Section 3 for the definition of  $\bar{\mathbf{D}}_2$  and  $\bar{\mathbf{D}}_o$ ). Further, PGdual is compared to the Random Walk (RW) sampler, whose proposal mechanism is

$$\mathbf{R}^{n+\frac{1}{2}} = \mathbf{R}^n + \sqrt{2\gamma_1} \xi_1^{n+1}, \quad \mathbf{O}^{n+\frac{1}{2}} = \mathbf{O}^n + \sqrt{2\gamma_2} \xi_2^{n+1}.$$

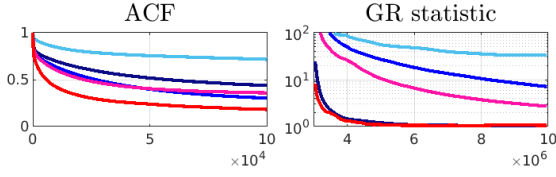
We also explore the benefit of covariance matrices for the  $\mathbf{R}$ -part of the Random-Walk chains: RW Invert and RW Ortho correspond to  $\mathbf{R}^{n+\frac{1}{2}} = \mathbf{R}^n + \sqrt{2\gamma_1} \bar{\mathbf{D}}^{-1} \xi_1^{n+1}$  resp. in the case  $\bar{\mathbf{D}} = \bar{\mathbf{D}}_2$  and  $\bar{\mathbf{D}} = \bar{\mathbf{D}}_o$ ; the proposal mechanism for the  $\mathbf{O}$ -part is as in RW. These comparisons will outline the role of the drift terms  $\mu_j$ , and of the covariance matrix of the Gaussian proposition.

**Settings.** All chains are run for  $N_{\max} = 10^7$  iterations. Except in Figure 1[left], displayed quantities are computed from the Markov path after discarding its first 30% points corresponding to a burn in period. All the curves displayed on Figure 1 and Figure 2 are mean values over 15 independent runs: they start from different points  $\theta^0$  obtained by Gaussian perturbations of the intuitive but poor (in terms of value of  $\pi$ ) value  $\mathbf{R}^{\text{init}} := (1, \dots, 1)^\top$  and  $\mathbf{O}^{\text{init}} := (0, \dots, 0)^\top$ . The step size  $\gamma_1$  is adapted during the burn in period in order to target a mean AR ratio equal to 0.25, and then it is frozen. For fair comparisons among methods, we use here the same target AR ratio which is known to be optimal for Random Walk algorithms (see, e.g., [26]; similar results for nonsmooth Langevin samplers do not exist). We set  $\gamma_2 = \sigma_{\mathbf{Z}}^3 \gamma_1 / T$  where  $\sigma_{\mathbf{Z}}$  is the standard deviation of the data set  $\mathbf{Z}$ ; and  $(\lambda_R, \lambda_O) = (3.5 \sigma_{\mathbf{Z}} \sqrt{6}/4, 0.05)$  as in [9].

**MCMC performance.** Figure 1[top] displays the evolution of the normalized distance to the MAP  $\hat{\mathbf{R}}$  along iterations:  $n \mapsto \|\mathbf{R}^n - \hat{\mathbf{R}}\| / \|\hat{\mathbf{R}}\|$  for the five MCMC samplers. Figure 1[bottom] displays



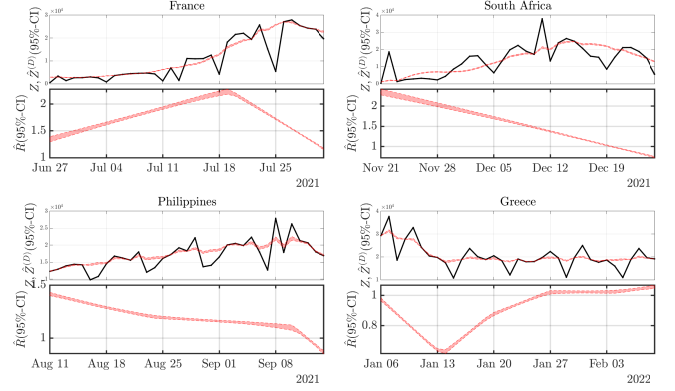
**Fig. 1. MCMC sampler performance.** RW in light blue, RW Invert in blue and RW Ortho in dark blue; PGdual Invert in pink and PGdual Ortho in red. During the burn in period [left] and after [right], evolution of the distance to the MAP along iterations [top] and to  $\max \ln \pi$  along iterations [bottom].



**Fig. 2. MCMC sampler performance.** RW in light blue, RW Invert in blue and RW Ortho in dark blue; PGdual Invert in pink and PGdual Ortho in red. [left] Mean absolute value of the ACF vs the first  $10^5$  lags. [right] The GR statistic vs iterations.

the evolution of the normalized log-density along iterations  $n \mapsto (\ln \pi(\theta^n) - \max \ln \pi) / (\ln \pi(\theta^0) - \max \ln \pi)$  for the five MCMC samplers.  $\hat{R}$  is computed by the method in [9]. Figure 2 compares the samplers through two criteria related to the convergence: the lag- $k$  auto-correlation function (ACF) of a Markov chain with initial distribution  $\pi$  - it is related to the effective sample size of a Markov chain (see [27]); and the Gelman-Rubin statistic (GR, [28]). The ACF plots are obtained as mean absolute values of the  $2T$  real valued ACF of  $(R_t, O_t)_{1 \leq t \leq T}$ . GR can be read as a ANOVA-type criterion, reaching an optimal value of one when the different paths are homogeneous.

These plots show that: (i) RW has a poor behavior compared to the other samplers, thus illustrating the role of the drift and the covariance matrix  $\bar{D}^{-1} \bar{D}^{-T}$  in the efficiency of the samplers: RW fails to converge before  $N_{\max}$  iterations contrary to PGdual Ortho. Indeed, from the results on Figure 1 and Figure 2: the distance to  $\hat{R}$  and the evolution of  $\ln \pi$  along the path show that RW moves slowly to high density regions thus being not efficient at all in the exploration of the target distribution  $\pi$ ; in addition, its ACF decreases to 0 very slowly and there is a strong heterogeneity between the 15 runs of this method (see the GR statistic). (ii) The behavior of the Random-Walk based methods is improved thanks to a genuine choice of the covariance matrix: RW Invert and - with a stronger evidence - RW Ortho take benefit of correlated Gaussian noises for the  $R$ -part of the chain. (iii) For the drift part of the proposal mechanism, there is a gain in using first order information on  $\ln \pi$ : if RW Invert and PGdual Invert have a similar behavior on Figure 1 and Figure 2[left] with a slight advantage for PGdual Invert, PGdual Ortho is definitely better than RW Ortho and than all the other methods. When the adapted step size is small enough, it has the highest speed of convergence to the optimum of  $\ln \pi$ ; at the end of the burn in period, it reached high density



**Fig. 3. Credibility intervals** for different countries and different time periods. For each country, observed counts  $Z$  (black solid lines) and 95% credibility interval denoised counts  $Z^{(D)}$  (red pipe) [top]; 95% credibility interval estimates for  $R$  [Bottom].

regions; it provides the faster decaying rate of the ACF; its GR statistic converges rapidly to the optimal value.

**Credibility intervals.** Let us further illustrate the relevance of credibility interval-based estimations for  $R$  and  $O$ . For practical purposes, the credibility intervals for  $O$  are translated into credibility intervals for the denoised counts  $Z^{(D)}$  by subtraction to the original counts  $Z$ : intuitively,  $Z^{(D)} = Z - O$ . Credibility intervals are reported for PGdual Ortho only - since shown above to yield the best performance - for several countries and various time periods of interest. The 95% credibility intervals are obtained from the empirical quantiles 0.025 and 0.975 of the Markov paths. In Figure 3, for each country, the top plot shows the daily counts  $Z$  (black) with the credibility interval estimates of the denoised counts  $Z^{(D)}$  (red), while the bottom plot represents the credibility interval estimates for  $R$ .

For France, the selected time period was chosen when the 4th pandemic wave started to severely strike the country (mid-june 2021) until reaching  $R \simeq 2$  early July. Figure 3 shows that credibility intervals excluded “ $R < 1$ ” before the end of June. Early July, the French Health Authorities made the announcement of a mandatory “sanitary pass” for any social activities after Aug. 9th. An immediate and massive vaccination *phase* started in the French population, which yields the *plateau* of the 4th wave, clearly observed in Figure 3 to be reached in the 3rd week of July. Credibility intervals further validate a clear decay of  $R$  after July 20th, that is, two weeks before the counts of infections actually started to decrease, thus showing the relevance and interest of the credibility interval estimates for  $R$ . For the Philippines and South Africa, the periods of late April 2021 and late August 2021 correspond to massive strikes of the pandemic. For both countries, Figure 3 confirms that the changes of slopes in  $R$  announce a few weeks in advance the maximum of the wave and the start of the decay in new infections. For Greece, the credibility intervals enable to detect with high confidence that “ $R > 1$ ” since January 27th 2022, which was far from being obvious by directly observing the infection counts.

Estimates and credibility intervals are updated on a regular basis and made available for the current period and for France, at [perso.math.univ-toulouse.fr/gfort/project/opsimore/](https://perso.math.univ-toulouse.fr/gfort/project/opsimore/).

## 5. CONCLUSIONS AND FUTURE WORKS

The present work constitutes a significant step toward the actual use of the tools envisaged in [8, 9] to monitor the reproduction number. Introducing a Bayesian formalism permits to produce reliable credi-

bility intervals for the reproduction number, as well as Bayesian estimators. It is based on an original Proximal-based Langevin MCMC algorithm designed to handle a convex but nonsmooth a posteriori potential  $-\ln \pi$ , whose nonsmooth part consists of the sum of several terms and which is able to account for the epidemiological model. Satisfactory agreements between credibility intervals and point estimates (MAP, posterior mean, . . .), obtained using real Covid19 data, emphasize the reliability of this novel MCMC sampler. Future investigations include the automated data-driven tuning of the design parameters (e.g., step sizes  $\gamma_i$ ), and the theoretical derivation of  $\text{PG}_{\text{dual}}$  convergence rates which are, to the best of our knowledge, an open question for many MALA-based samplers designed for nonsmooth target distributions.

## 6. REFERENCES

- [1] O. Diekmann, J. A. P. Heesterbeek, and J. A. J. Metz, “On the definition and the computation of the basic reproduction ratio  $R_0$  in models for infectious diseases in heterogeneous populations,” *J. Math. Biol.*, vol. 28, pp. 365–382, 1990.
- [2] J. Wallinga and P. Teunis, “Different Epidemic Curves for Severe Acute Respiratory Syndrome Reveal Similar Impacts of Control Measures,” *Am. J. Epidemiol.*, vol. 160, pp. 509–516, 2004.
- [3] P. van den Driessche and J. Watmough, “Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission,” *Math Biosci.*, vol. 180, pp. 29–48, 2002.
- [4] T. Obadia, R. Haneef, and P.-Y. Boëlle, “The  $R_0$  package: A toolbox to estimate reproduction numbers for epidemic outbreaks,” *BMC Medical Inform Decis. Mak.*, vol. 12, pp. 147, 2012.
- [5] A. Cori, N. M. Ferguson, C. Fraser, and S. Cauchemez, “A new framework and software to estimate time-varying reproduction numbers during epidemics,” *Am. J. Epidemiol.*, vol. 178, pp. 1505–1512, 2013.
- [6] Q.-H. Liu, M. Ajelli, A. Aleta, S. Merler, Y. Moreno, and A. Vespignani, “Measurability of the epidemic reproduction number in data-driven contact networks,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 115, pp. 12680–12685, 2018.
- [7] F. Brauer, C. Castillo-Chavez, and Z. Feng, *Mathematical models in epidemiology*, Springer, New York, 2019.
- [8] P. Abry et al., “Spatial and temporal regularization to estimate COVID-19 reproduction number  $R(t)$ : Promoting piecewise smoothness via convex optimization,” *PLOS One*, vol. 15, 2020, e0237901.
- [9] B. Pascal, P. Abry, N. Pustelnik, S. Roux, R. Gribonval, and P. Flandrin, “Nonsmooth convex optimization to estimate the Covid-19 reproduction number space-time evolution with robustness against low quality data,” Tech. Rep., arXiv 2109.09595, 2022.
- [10] R.N. Thompson et al., “Improved inference of time-varying reproduction numbers during infectious disease outbreaks,” *Epidemics*, vol. 29, pp. 100356, 2019.
- [11] F. Riccardo et al., “Epidemiological characteristics of COVID-19 cases in Italy and estimates of the reproductive numbers one month into the epidemic,” medRxiv:2020.04.08.20056861, 2020.
- [12] G. Guzzetta et al., “The impact of a nation-wide lockdown on COVID-19 transmissibility in Italy,” arXiv:2004.12338 [q-bio.PE], 2020.
- [13] G. Fort, B. Pascal, P. Abry, and N. Pustelnik, “Covid19 Reproduction Number: Credibility Intervals by Blockwise Proximal Monte Carlo Samplers,” Tech. Rep., HAL, 2022.
- [14] C. P. Robert, *The Bayesian choice: a decision-theoretic motivation*, Springer-Verlag, 1994.
- [15] G. Chierchia, E. Chouzenoux, P.L. Combettes, and J.-C. Pesquet, “The proximity operator repository,” <http://proximity-operator.net/>.
- [16] J. Kent, “Time-Reversible Diffusions,” *Adv Appl Probab*, vol. 10, no. 4, pp. 819–835, 1978.
- [17] G. O. Roberts and R. L. Tweedie, “Exponential convergence of Langevin distributions and their discrete approximations,” *Bernoulli*, vol. 2, pp. 341 – 363, 1996.
- [18] M. Pereyra, P. Schniter, E. Chouzenoux, J.-C. Pesquet, J.-Y. Tourneret, A. O. Hero, and S. McLaughlin, “A survey of stochastic simulation and optimization methods in signal processing,” *IEEE J. Selected Topics Signal Process.*, vol. 10, pp. 224–241, 2016.
- [19] Y.F. Atchadé, “A Moreau-Yosida approximation scheme for a class of high-dimensional posterior distributions,” Tech. Rep., arXiv: Statistics Theory, 2015.
- [20] A. Schreck, G. Fort, S. Le Corff, and É. Moulines, “A Shrinkage-Thresholding Metropolis Adjusted Langevin Algorithm for Bayesian Variable Selection,” *IEEE J. Selected Topics Signal Process.*, vol. 10, pp. 366–375, 2016.
- [21] A. Salim and P. Richtarik, “Primal Dual Interpretation of the Proximal Stochastic Gradient Langevin Algorithm,” in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, Eds., 2020, vol. 33, pp. 3786–3796.
- [22] A. Durmus, É. Moulines, and M. Pereyra, “Efficient Bayesian Computation by Proximal Markov Chain Monte Carlo: When Langevin Meets Moreau,” *SIAM J Imaging Sci*, vol. 11, pp. 473–506, 2018.
- [23] A. Durmus, S. Majewski, and B. Miasojedow, “Analysis of Langevin Monte Carlo via Convex Optimization,” *J. Mach. Learn. Res.*, vol. 20, pp. 73:1–73:46, 2019.
- [24] N. Chatterji, J. Diakonikolas, M. I. Jordan, and P. Bartlett, “Langevin Monte Carlo without smoothness,” in *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, S. Chiappa and R. Calandra, Eds., 2020, vol. 108 of *Proceedings of Machine Learning Research*, pp. 1716–1726.
- [25] T.D. Luu, J. Fadili, and C. Chesneau, “Sampling from Nonsmooth Distributions Through Langevin Diffusion,” *Methodol Comput Appl Probab*, 2020.
- [26] G.O. Roberts and J.S. Rosenthal, “Optimal scaling for various Metropolis-Hastings algorithms,” *Statistical Science*, vol. 16, pp. 351 – 367, 2001.
- [27] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods (Springer Texts in Statistics)*, Springer-Verlag, Berlin, Heidelberg, 2005.
- [28] S.P. Brooks and A. Gelman, “General methods for monitoring convergence of iterative simulations,” *J Comput Graph Stat*, vol. 7, pp. 434–455, 1998.