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

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

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

Hugo Artigas[(1)], *Barbara Pascal*[(2)], *Gersende Fort*[(3)], *Patrice Abry*[(4)], *Nelly Pustelnik*[(4)]

[(1)] Ecole Polytechnique, Paris, France

[(2)] Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, France (barbara.pascal@univ-lille.fr)

[(3)] CNRS, Institut de Mathématiques de Toulouse, France (gersende.fort@math.univ-toulouse.fr)

[(4)] Univ Lyon, Ens de Lyon, Univ Claude Bernard, CNRS, Laboratoire de Physique, Lyon, France.

(firstname.lastname@ens-lyon.fr)

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, pseudoseasonalities, ...). Applied to daily new infection counts made public by National Health Agencies and centralized at 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 new Monte Carlo methods to sample from a nonsmooth convex *a posteriori* distribution. These new samplers stem from an original combination of the Langevin Monte Carlo algorithm with Proximal operators. Their 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 nonsmooth convex optimization 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 nonsmooth convex optimization procedures 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: *Proximal-Gradient Decomposition* stochastic sampling schemes, refining the classical Metropolis Adjusted Langevin Algorithms, are devised to sample a generic class of *a posteriori* distributions, including the one introduced in Section 2. The originality of the proposed samplers is to account for the difficulties stemming from the specificity of the *a posteriori* distributions considered here.

To assess the relevance of the proposed stochastic sampling schemes in producing credibility intervals for R, and to compare them 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. The period of June-July 2021 in France is used as a pandemic-interesting case study, but the proposed tools are ready for applications to any country or period of time.

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, the model in [9] is recast into a Bayesian framework. The statistical model associated with the vector of T observations $\mathbf{Z} = (\mathsf{Z}_1, \ldots, \mathsf{Z}_T)^\top \in \mathbb{N}^T$,

H. Artigas and G. Fort are partly funded by the *Fondation Simone et Cino Del Duca, French Academy of Sciences.*

consisting of new daily infection counts, is indexed by the unknown

$$\theta := (\mathbf{R}^\top, \mathbf{O}^\top)^\top = (\mathsf{R}_1, \dots, \mathsf{R}_T, \mathsf{O}_1, \dots, \mathsf{O}_T)^\top \in (\mathbb{R}_+)^T \times \mathbb{R}^T$$

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 θ, the conditional distribution of Z_t given the past $\mathsf{Z}_{t-1}, \dots, \mathsf{Z}_1$ and initial values $(\mathsf{Z}_0, \cdots, \mathsf{Z}_{1-\tau_\phi})$ is a Poisson distribution with intensity

$$p_t(\theta) := \mathsf{R}_t \sum_{u=1}^{\tau_\phi} \Phi_u \mathsf{Z}_{t-u} + \mathsf{O}_t \ , \tag{1}$$

where the *serial function* $(\Phi_u)_{1\le u\le \tau_\phi}$ accounts for epidemiology mechanisms: It quantifies the random delays between the onset of symptoms in a primary case and in secondary cases [5, 10, 6]. $(\Phi_u)_{1\le u\le \tau_\phi}$ 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 normalized negative log-likelihood of the observations $\mathbf{Z}$ is given by [1]

$$f(\theta) := \begin{cases} -\frac{1}{T}\sum_{t=1}^T \left(\mathsf{Z}_t \ln p_t(\theta) - p_t(\theta)\right) & \text{if } \theta \in \mathcal{D}, \\ +\infty & \text{otherwise,} \end{cases} \tag{2}$$

defined on the measurable set

$$\begin{aligned} \mathcal{D} := \{\theta \in (\mathbb{R}_+)^T \times \mathbb{R}^T : p_t(\theta) > 0 \text{ for } t \text{ s.t. } \mathsf{Z}_t > 0\} \\ \cup \{\theta \in (\mathbb{R}_+)^T \times \mathbb{R}^T : p_t(\theta) \ge 0 \text{ for } t \text{ s.t. } \mathsf{Z}_t = 0\} \ . \end{aligned} \tag{3}$$

Interpreting the regularity constraints in [9], the logarithm of the *a priori* distribution of θ is defined on $(\mathbb{R}_+)^T \times \mathbb{R}^T$ by

$$g(\theta) := -\frac{\lambda_{\mathrm{T}}}{4T}\|\mathsf{D}_2\mathbf{R}\|_1 - \frac{\lambda_{\mathrm{O}}}{T}\|\mathbf{O}\|_1 \ ; \tag{4}$$

D_2 is the discrete-time second order derivative $(T-2)\times T$ matrix:

$$\mathsf{D}_2 := \begin{bmatrix} 1 & -2 & 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & -2 & 1 & 0 & \dots & 0 \\ \dots & & & & & & \dots \\ 0 & \dots & & & 1 & -2 & 1 \end{bmatrix} \ , \tag{5}$$

and $\lambda_{\mathrm{T}}, \lambda_{\mathrm{O}}$ 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 $(\mathbf{R}_\star, \mathbf{O}_\star)$ satisfy $\mathsf{D}_2\mathbf{R} = 0$ and $\mathbf{O} = 0$, samples from the *a priori* distribution are not necessarily sparse.

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

$$\pi(\theta) := \exp\left(-f(\theta) - g(\theta)\right) \mathbb{1}_{\mathcal{D}}(\theta) \ . \tag{6}$$

Upon noting that, up to an additive constant, the negative log-likelihood of a Poisson variable Z_t at θ, is the Kullback-Leibler distance $d_{\mathrm{KL}}(\mathsf{Z}_t | p_t(\theta))$ between Z_t and $p_t(\theta)$, the negative log-density $-\ln\pi$ is the criterion minimized in [9] for the reconstruction of θ. It is proved in [9] that when $\sum_{u=1}^{\tau_\phi} \Phi_u \mathsf{Z}_{t-u} > 0$ for any $1 \le t \le T$, a minimum of $-\ln\pi$ exists and the set of the minima is included in a level set $\{\theta : f(\theta) = f_*\}$ of f. The *a priori* distribution implies that 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.

[1] 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 does not have a closed form (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 α-credible regions (see, e.g. [13, 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 π of the form (6) and satisfying A1-A2.

Assumptions on the target π.

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

A2. *g is block-wise: for $j \in \{1, \dots, J\}$, $i \in \{1, \dots, I_j\}$, there exist matrices $\mathsf{A}_{i,j} \in \mathbb{R}^{c_i \times d_j}$, and proper, convex, lower semi-continuous functions $g_{i,j} : \mathbb{R}^{c_i} \to]-\infty, +\infty]$ such that $\sum_{j=1}^J d_j = d$ and*

$$\forall \theta := (\theta_1^\top, \dots, \theta_J^\top)^\top, \qquad g(\theta) := \sum_{j=1}^J \sum_{i=1}^{I_j} g_{i,j}(\mathsf{A}_{i,j}\theta_j) \ ,$$

and where each function $g_{i,j}(\mathsf{A}_{i,j}\cdot)$ possesses a proximal operator[2] having a closed form.

A 1 requires the differentiability of f but not the Lipschitz-gradient property, and A2 does not assume that the proximal operator of g exists or has a closed-form expression. A2 is verified when, e.g., *(i)* $g_{i,j}$ is limited to a certain class of functions such as the square ℓ_2-norm and if $(\mathsf{A}_{i,j}^\top \mathsf{A}_{i,j} + \mathsf{Id})$ is efficiently invertible [14], or *(ii)* for a general function $g_{i,j}$ having an explicit proximal operator (see [15] to have an exhaustive list) when $\mathsf{A}_{i,j}\mathsf{A}_{i,j}^\top = \nu_{i,j}\mathsf{Id}$ for some $\nu_{i,j} > 0$ whose expression reads (see [16]) for every $\theta_j \in \mathbb{R}^{d_j}$,

$$\mathrm{prox}_{\gamma g_{i,j}(\mathsf{A}_{i,j}\cdot)}(\theta_j) = \theta_j + \nu_{i,j}^{-1}\mathsf{A}_{i,j}^\top\left(\mathrm{prox}_{\gamma\nu_{i,j} g_{i,j}}(\mathsf{A}_{i,j}\theta_j) - \mathsf{A}_{i,j}\theta_j\right).$$

A generalization when $\mathsf{A}_{i,j}\mathsf{A}_{i,j}^\top$ is a diagonal matrix is provided in [17] at the price of separability of $g_{i,j}$. The reader could refer to [17, 18] for typical examples in signal and image processing where condition A2 is satisfied, e.g., Total Variation penalization.

For the target density π defined in Section 2, both A1 and A2 are valid. For ease of notations, proof is given here for $(T-2)/3$ a positive integer. We have $d = 2T$, $J = 2$, $I_1 = 3$, $I_2 = 1$, $\theta_1 = \mathbf{R}$ and $\theta_2 = \mathbf{O}$. For $i \in \{1, 2, 3\}$, $\mathsf{A}_{i,1} \in \mathbb{R}^{(T-2)/3 \times T}$ collects the rows $i, i+3, i+6, \cdots$ of the matrix D_2 and satisfies $\mathsf{A}_{i,1}\mathsf{A}_{i,1}^\top = 6\mathsf{Id}$; $\mathsf{A}_{1,2} = \mathsf{Id}$; $g_{i,1} := \lambda_{\mathrm{T}}\|\cdot\|_1/(4T)$; and $g_{1,2}(\cdot) := \lambda_{\mathrm{O}}\|\cdot\|_1/T$.

Metropolis Adjusted Langevin Algorithm (MALA) sampler and its limitations. If $-\ln\pi = f$ (i.e., $g = 0$) is a smooth convex function on $\mathcal{D}$, a popular MCMC sampler which takes benefit of a first order knowledge of π, relies on Langevin dynamics. Given a positive step size $\gamma > 0$ chosen by the user, at each iteration $\#n$ starting from the current point θ^n, a jump is proposed to the point

$$\theta^{n+\frac{1}{2}} := \theta^n - \gamma\,\nabla f(\theta^n) + \sqrt{2\gamma}\,\xi^{n+1} \ , \tag{7}$$

where $\xi^{n+1} \sim \mathcal{N}_d(0, \mathsf{Id})$ is a standard $\mathbb{R}^d$-valued Gaussian distribution. This proposal is accepted ($\theta^{n+1} = \theta^{n+\frac{1}{2}}$) or rejected ($\theta^{n+1} =$

[2] $\mathrm{prox}_g(y) := \mathrm{argmin}_{x\in\mathbb{R}^\ell}(g(x) + \|x-y\|^2/2)$ is well-defined for any proper, convex, lower semi-continuous function $g : \mathbb{R}^\ell \to]-\infty, +\infty]$.

θ^n) through a Metropolis mechanism, and this yields the *Metropolis Adjusted Langevin Algorithm* (MALA, [19]). The Metropolis step makes the chain $\{\theta^n, n \geq 0\}$ to be Markovian with unique invariant distribution π. The idea of MALA methods is to drift the proposed moves towards areas of high probability under π, by using the information provided by $\nabla \ln \pi$.

Proximal-Gradient Decomposition (PGD) sampler. In the present work, $-\ln \pi$ is not smooth (see A1-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., [20, 21, 22, 23, 24, 25, 26, 27]) but none of these algorithms directly apply to the framework A1-A2. Indeed, the density π may be positive for non-sparse vectors. Further, A2 assumes that $g_{i,j}(\mathsf{A}_{i,j}\cdot)$ has a proximal operator with a closed form for any $j \in \{1, \ldots, J\}$ and $i \in \{1, \ldots, I_j\}$ which does not imply that this holds true for g. Finally, we want to promote a method which takes benefit of the block-wise separable expression of g by using at each iteration the proximal operator associated to a partial sum.

These limitations led us to propose the following original and general Proximal-Gradient Decomposition (PGD) Algorithm (sketched in Algorithm 1). At iteration $\#n$, part of the terms in g are chosen at random by selecting an index i_j for each block j (see Line 3). Then, a jump is proposed to $\theta^{n+\frac{1}{2}}$ whose block $\#j$ is

$$\theta_j^{n+\frac{1}{2}} := \mu_{i_j,j}(\theta^n) + \sqrt{2\gamma_j}\Gamma_j \xi_j^{n+1} \ , \quad \forall\, j \in \{1, \ldots, J\} \ , \quad (8)$$

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

$$\mu_{i,j}(\theta) := \mathrm{prox}_{\gamma_j g_{i,j}(\mathsf{A}_{i,j}\cdot)}\left(\theta_j - \gamma_j \nabla_j f(\theta)\right) \ . \quad (9)$$

The drift term $\mu_{i_j,j}(\theta^n)$ performs a proximal-gradient iteration associated to a part of the original composite function, combining a gradient step w.r.t. the variable θ_j, and a proximal step relative to the function $g_{i_j,j}(\mathsf{A}_{i_j,j}\cdot)$, started from θ_j^n and with step size γ_j. The candidate $\theta_j^{n+1/2}$ is sampled from a Gaussian distribution with drift $\mu_{i_j,j}(\theta^n)$ (see (8)). Introducing $q_{i,j}(\theta, \theta_j')$ the Gaussian kernel on $\mathbb{R}^{d_j}$ centered at $\mu_{i,j}(\theta)$, with covariance matrix $2\gamma_j\Gamma_j\Gamma_j^\top$ and evaluated at θ_j', the acceptance-rejection (AR) step defined Line 6 makes the algorithm to have π as unique invariant distribution.

Algorithm 1: Proximal-Gradient Decomposition (PGD)

Data: $d_j \times d_j$ positive definite matrices Γ_j, $\gamma_j > 0$, $N_{\max} \in \mathbb{N}_\star$, $\theta^0 \in \mathcal{D}$

Result: A $\mathcal{D}$-valued sequence $\{\theta^n, n \in [N_{\max}]\}$

1 **for** $n = 0, \ldots, N_{\max} - 1$ **do**

2 **for** $j = 1, \ldots, J$ **do**

3 Sample $i_j \in \{1, \ldots, I_j\}$ with probability $1/I_j$;

4 Sample $\xi_j^{n+1} \sim \mathcal{N}_{d_j}(0, \mathsf{Id})$;

5 Set $\theta_j^{n+\frac{1}{2}} = \mu_{i_j,j}(\theta^n) + \sqrt{2\gamma_j}\,\Gamma_j\,\xi_j^{n+1}$;

6 Set $\theta^{n+1} = \theta^{n+\frac{1}{2}}$ with probability

$$1 \wedge \frac{\pi(\theta^{n+\frac{1}{2}})}{\pi(\theta^n)} \prod_{j=1}^{J} \frac{q_{i_j,j}(\theta^{n+\frac{1}{2}}, \theta_j^n)}{q_{i_j,j}(\theta^n, \theta_j^{n+\frac{1}{2}})}$$

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

Let us detail $\mu_{i,j}$ when f, g, and $\mathcal{D}$ are respectively defined by (2), (4), and (3). The proximal operator in (10) being tractable since $\mathsf{A}_{i,1}\mathsf{A}_{i,1}^\top = 6\mathsf{Id}$, we choose

$$\mu_{i,1} = \mathrm{prox}_{\frac{\gamma_1\lambda_{\mathrm{T}}}{4T}\|\mathsf{A}_{i,1}\cdot\|_1}\left(\mathbf{R}^n - \gamma_1\,\nabla_1 f(\theta^n)\right) \ , \quad (10)$$

$$\mu_{1,2} = \mathrm{prox}_{\gamma_2\frac{\lambda_{\mathrm{O}}}{T}\|\cdot\|_1}\left(\mathbf{O}^n - \gamma_2\,\nabla_2 f(\theta^n)\right) . \quad (11)$$

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*[3], consisting of daily new infection counts measured for the entire population of more than 200 countries, by National Health Authorities. The focus is here on the evolution of R_t on a realistic time period of $T = 56$ days (8 weeks) of pandemic. To account for the time support of the serial function Φ, we use $\tau_\phi = 26$ additional observations for the initialization $\mathsf{Z}_0, \cdots, \mathsf{Z}_{-\tau_\phi+1}$. Counts from June, 8 2021 to August, 2nd 2021 for France (see Figure 2[top]) are used here as the period is of pandemic evolution interest. However, tools and methods proposed here can be applied to any country and any time period of interest.

MCMC samplers. Two versions of `PGD` are compared, corresponding to $(\Gamma_1, \Gamma_2) := (\mathsf{Id}, \mathsf{Id})$ and $(\Gamma_1, \Gamma_2) := (\overline{\mathsf{D}}_2^{-1}, \mathsf{Id})$; they are named resp. `naive-PGD` and `PGD`. $\overline{\mathsf{D}}_2$ is a genuine invertible extension of D_2 (see (5)) chosen as:

$$\overline{\mathsf{D}}_2 := \begin{bmatrix} 1/\sqrt{6} & 0 & 0 & \cdots & 0 \\ -2/\sqrt{2} & 1/\sqrt{2} & 0 & \cdots & 0 \\ & & \mathsf{D}_2 & & \end{bmatrix} . \quad (12)$$

The ability of `PGD` (see Section 3) to construct relevant credibility intervals for the reproduction number R_t of the Covid19 pandemic is compared to the Random Walk (`RW`) sampler and to another strategy referred to as `PGdual` - available for π defined in Section 2 but not for *any* distribution described by A1-A2. These comparisons will outline the role of the covariance structure Γ_1 and of the drift term in the Gaussian proposal mechanisms: all the samplers repeat *(a)* the proposition of a move of the form (8) with a specific drift term, and *(b)* a Metropolis AR step.

In `RW`, the proposal mechanism is

$$\mathbf{R}^{n+\frac{1}{2}} = \mathbf{R}^n + \sqrt{2\gamma_1}\,\Gamma_1\xi_1^{n+1}, \quad \mathbf{O}^{n+\frac{1}{2}} = \mathbf{O}^n + \sqrt{2\gamma_2}\,\xi_2^{n+1} \ ;$$

in `naive-RW`, $\Gamma_1 := \mathsf{Id}$ while in `RW`, $\Gamma_1 := \overline{\mathsf{D}}_2^{-1}$.

Finally, `PGdual` corresponds to the proposal mechanism

$$\mathbf{R}^{n+\frac{1}{2}} = \tilde{\mu}_1(\theta^n) + \sqrt{2\gamma_1}\,\overline{\mathsf{D}}_2^{-1}\xi_1^{n+1}, \mathbf{O}^{n+\frac{1}{2}} = \tilde{\mu}_2(\theta^n) + \sqrt{2\gamma_2}\,\xi_2^{n+1}$$

where $\tilde{\mu}_2 = \mu_{1,2}$ given by (11) and

$$\tilde{\mu}_1(\theta) := \overline{\mathsf{D}}_2^{-1}\,\mathrm{prox}_{\frac{\gamma_1\lambda_{\mathrm{T}}}{4T}\|(\cdot)_{3:T}\|_1}\left(\overline{\mathsf{D}}_2\,\mathbf{R} - \gamma_1\overline{\mathsf{D}}_2^{-\top}\,\nabla_1 f(\theta)\right) \ ;$$

for $\tau \in \mathbb{R}^T$, $\tau_{3:T}$ collects the components 3 to T of τ. We have $(\mathbf{R}, \mathbf{O}) \sim \pi$ if and only if $(\overline{\mathsf{D}}_2\mathbf{R}, \mathbf{O}) \sim \bar{\pi}$ where $\bar{\pi}(\tau, o) \propto \exp(-f(\overline{\mathsf{D}}_2^{-1}\tau, o) - \lambda_{\mathrm{T}}\|\tau_{3:T}\|_1/(4T) - \lambda_{\mathrm{O}}\|o\|_1/T)$ on the set $\{(\tau, o) \in \mathbb{R}^{2T} : (\overline{\mathsf{D}}_2^{-1}\tau, o) \in \mathcal{D}\}$. The proposal mechanism of `PGdual` combines: *(a)* in the (τ, o)-space, sample a Gaussian distribution centered at a Proximal-Gradient step associated to the composite function $-\ln \bar{\pi}$ and started at $(\tau^n, o^n) := (\overline{\mathsf{D}}_2\mathbf{R}^n, \mathbf{O}^n)$, then *(b)* move this point back to the θ-space by applying $\overline{\mathsf{D}}_2^{-1}$.

Settings. All chains start from an intuitive (but poor, in terms of value of π) initialization $\mathbf{R}^0 := (1, \ldots, 1)^\top$ and $\mathbf{O}^0 :=$

[3] https://coronavirus.jhu.edu/

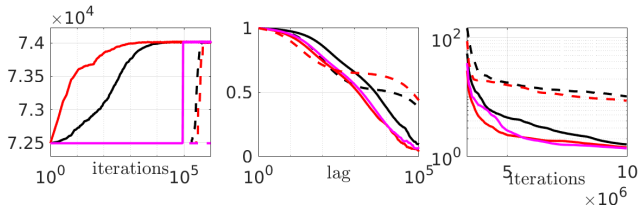

Fig. 1. **MCMC sampler performance.** `PGD` (red solid line), `naive-PGD` (red dashed line), `PGdual` (magenta solid line), `RW` (black solid line) and `naive-RW` (black dashed line). [left] Evolution of the log-density $\ln \pi$ along the first $3e5$ iterations. [middle] Mean absolute value of the ACf vs the first $1e5$ lags. [right] Gelman-Rubin statistics vs iterations, after the burn-in period.

$(0, \dots, 0)^\top$ and are run for $N_{\max} = 1e7$ iterations. Except in Figure 1[left], displayed quantities are computed from the Markov path after discarding its first $N_{\max}/3$ points corresponding to a burn-in phase. The step size γ_1 is adapted until iteration $\#N_{\max}/3$ 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 optimal AR ratio for Random Walk algorithms (see, e.g., [28]), since similar results for nonsmooth Langevin samplers do not exist. We set $\gamma_2 = \mathrm{std}(\mathbf{Z})\,\gamma_1$ where $\mathrm{std}(\mathbf{Z})$ denotes the standard deviation of the data set $\mathbf{Z}$; and $(\lambda_{\mathrm{T}}, \lambda_{\mathrm{O}}) = (3.5 \times \mathrm{std}(\mathbf{Z}), 0.05)$ as in [9].

MCMC performance. Figure 1[left] displays the evolution (mean value over ten independent runs) of the log-density $n \mapsto \ln \pi(\theta^n)$ along the first $3e5$ iterations, and for five MCMC samplers. Figure 1[middle, right] compare the samplers through two criteria related to the convergence: the lag-k auto-correlation function (ACf) of a Markov chain with initial distribution π - it is related to the effective sample size of a Markov chain (see [29]); and the Gelman-Rubin statistics (GRs, [30]). The ACf plots are obtained as mean absolute values of the $2T$ real valued ACf of $(\mathsf{R}_t, \mathsf{O}_t)_{1\le t\le T}$. GRs, computed for the full vector $(\mathbf{R}, \mathbf{O}) \in \mathbb{R}^{2T}$ and from 15 independent runs, can be read as anova-type criteria that quantify inter vs. intra path homogeneities. These plots show that: *(i)* Both `naive-RW` and `naive-PGD` have poor behaviors compared to the other samplers, thus illustrating the role of Γ_1 in the efficiency of the samplers: Compared to the other samplers, they move slowly to high density regions thus being less efficient in the exploration of the target distribution π, their ACf decrease to 0 more slowly, and their GRs converge far more slowly to the optimal value (equal to one). We also observed that `naive-RW` and `naive-PGD` fail to converge before $N_{\max}$ iterations contrary to the other methods, a poor asymptotic behavior which explains the results on Figure 1[right]: when each algorithm is run 15 times, there is a strong heterogeneity between the 15 outputs of `naive-RW` and similarly for the ones of `naive-PGD`. *(ii)* For the drift part of the proposal mechanism, there is a gain in using first order information on $\ln \pi$: `PGD` reaches high density regions more rapidly than `RW` when it starts from low density regions (see Figure 1[left]), and both `PGD` and `PGdual` improve on `RW` in the asymptotic regime of the Markov chain (see Figure 1[middle,right]). *(iii)* Nevertheless, `PGdual` is sensitive to the definition of γ_1 and it is not able to escape from a low density region as rapidly as `PGD` (see Figure 1[left]). To emphasize the role of γ_1, `PGdual` is also run with a constant step size γ_1 equal to the limiting value obtained after adaption: the sampler is not able to accept a move (see the magenta dotted line on Figure 1[left]). The same experiment - not shown here - is run for `PGD`: It does not suffer from this drawback.

Credibility intervals. The performance analyses reported above, and repeated for several countries and different time periods, led to conclude that `PGD` is so far the most efficient sampling strategy to explore the *a posteriori* distribution (6). Therefore, we report only credibility intervals and posterior means obtained from `PGD`. Figure 2[top] shows the observations $(\mathsf{Z}_t)_{1\le t\le T}$ as well as the denoised counts $\mathsf{Z}_t - \widehat{\mathsf{O}}_t$; $\widehat{\mathsf{O}}_t$ is the posterior mean point estimate of O_t. The sequence $(\mathsf{Z}_t - \widehat{\mathsf{O}}_t)_{1\le t\le T}$ exhibits clearer trends than the counts $(\mathsf{Z}_t)_{1\le t\le T}$ themselves. Figure 2[bottom] displays the 95% credibility region for $\mathbf{R} \in \mathbb{R}^T$, obtained from the empirical quantiles 0.025 and 0.975 of a Markov path. It also shows the MAP computed as in [9] and the posterior mean estimators for R_t. The selected time period was chosen as the 4th pandemic wave started to severely strike France (after mid-june) until reaching $\mathsf{R} \simeq 2$ early July. Figure 2 shows that credibility intervals excluded "$\mathsf{R} < 1$" 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 2 to be reached in the 3rd week of July. Credibility intervals further validates 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 on-the-fly estimation credibility intervals for R.

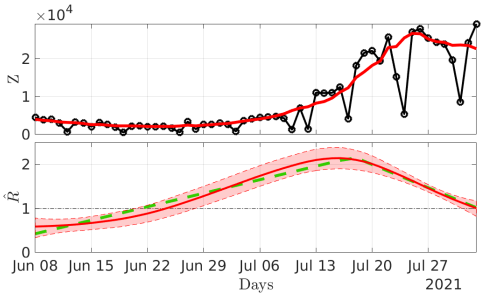

Fig. 2. **Credibility intervals.** [Top] Observed counts $\mathbf{Z}$ (black solid line with circles) and denoised counts $(\mathsf{Z}_t - \widehat{\mathsf{O}}_t)$ (red solid line) with $\widehat{\mathsf{O}}_t$ estimated by the posterior mean. [Bottom] 95% credibility interval (shadowed area) and two different point estimations: MAP (green dashed line), posterior mean (red solid line).

Estimates and credibility intervals are updated on a regular basis and made available for the current period at 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. Recasting the model in [8, 9] into a Bayesian formalism permits to produce reliable credibility intervals for the reproduction number, as well as Bayesian estimators. It is based on an original Proximal-based Langevin MCMC algorithm designed to handle both a drift-term that accounts for the epidemiological model and a convex but nonsmooth *a posteriori* potential $-\ln \pi$, whose nonsmooth part additionally consists of the sum of several terms. Satisfactory agreements between credibility intervals and point estimates (MAP, posterior mean, ...), obtained using real Covid19 data, emphasizes the reliability of this novel MCMC sampler. Future investigations include the automated data-driven tuning of the design parameters (e.g., step sizes γ_i), and the theoretical derivation of `PGD` 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 R0 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 R0 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," *Proceedings of the National Academy of Sciences*, 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, N. Pustelnik, S. Roux, P. Jensen, P. Flandrin, R. Gribonval, C.-G. Lucas, É. Guichard, P. Borgnat, and N. Garnier, "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, 2021.

[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, M. Ajelli, X. D. Andrianou, A. Bella, M. Del Manso, M. Fabiani, S. Bellino, S. Boros, A. M. Urdiales, and V. Marziano 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] C. P. Robert, *The Bayesian choice: a decision-theoretic motivation*, Springer-Verlag, 1994.

[14] P.-L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke et al., Ed., pp. 185–212. Springer-Verlag, New York, 2011.

[15] G. Chierchia, E. Chouzenoux, P.L. Combettes, and J.-C. Pesquet, "The proximity operator repository," *http://proximity-operator.net/*.

[16] P. L. Combettes and J.-C. Pesquet, "A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery," *IEEE J. Selected Topics Signal Process.*, vol. 1, pp. 564–574, 2007.

[17] N. Pustelnik, C. Chaux, and J.-C. Pesquet, "Parallel proXimal algorithm for image restoration using hybrid regularization," *IEEE Trans. Image Process.*, vol. 20, pp. 2450–2462, 2011.

[18] B. Pascal, N. Pustelnik, P. Abry, and J.-C. Pesquet, "Block-coordinate proximal algorithms for scale-free texture segmentation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2018.

[19] G. O. Roberts and R. L. Tweedie, "Exponential convergence of Langevin distributions and their discrete approximations," *Bernoulli*, vol. 2, pp. 341 – 363, 1996.

[20] M. Pereyra, P. Schniter, É. 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.

[21] Y.F. Atchadé, "A Moreau-Yosida approximation scheme for a class of high-dimensional posterior distributions," Tech. Rep., arXiv: Statistics Theory, 2015.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] T.D. Luu, J. Fadili, and C. Chesneau, "Sampling from Non-smooth Distributions Through Langevin Diffusion," *Methodol Comput Appl Probab*, 2020.

[28] G.O. Roberts and J.S. Rosenthal, "Optimal scaling for various Metropolis-Hastings algorithms," *Statistical Science*, vol. 16, pp. 351 – 367, 2001.

[29] C.P. Robert and G. Casella, *Monte Carlo statistical methods*, Springer Verlag, 2004.

[30] S.P. Brooks and A. Gelman, "General methods for monitoring convergence of iterative simulations," *J Comput Graph Stat*, vol. 7, pp. 434–455, 1998.