Credibility Intervals for the Reproduction Number of the Covid-19 Pandemic Using Proximal Langevin Samplers

Patrice Abry, Gersende Fort, Barbara Pascal, Nelly Pustelnik

CREDIBILITY INTERVALS FOR THE REPRODUCTION NUMBER OF THE COVID-19 PANDEMIC USING PROXIMAL LANGEVIN SAMPLERS

Patrice Abry[1]*, Gersende Fort*[2]*, Barbara Pascal*[3]*, Nelly Pustelnik*[1]

[1] ENSL, CNRS, Laboratoire de Physique, F-69342 Lyon, France. firstname.lastname@ens-lyon.fr

[2] Institut de Mathématiques de Toulouse, Université de Toulouse, CNRS, F-31062 Toulouse, France gersende.fort@math.univ-toulouse.fr

[3] Nantes Université, École Centrale Nantes, CNRS, LS2N, UMR 6004, F-44000 Nantes, France barbara.pascal@cnrs.fr

ABSTRACT

Monitoring the time evolution of the intensity of the Covid-19 pandemic within the pandemic and despite the limited quality of the data is both crucial and challenging. In the context of a pandemic, providing a level of confidence in the estimation of epidemiological indicators is essential to inform decision makers. The present work proposes a Bayesian estimation of the pandemic Covid-19 reproduction number and of the denoised reported counts through credibility intervals. The challenge stems from an epidemiological Bayesian model robust to errors in reported counts, which yields a non differentiable a posteriori log-density. Four different Hastings-Metropolis algorithms combining Langevin approaches and proximal operators are compared on a toy example; the most efficient one is plugged into a Metropolis-within-Gibbs algorithm performing a credibility intervals-based estimation of Covid-19 pandemic indicators, exemplified for several countries worldwide.

***Index Terms*—** Markov chain Monte Carlo, log-concave composite density, Langevin Monte Carlo, Proximal operators, Bayesian credibility intervals, Covid-19, Reproduction number.

1. INTRODUCTION

Context. Real-time monitoring of the evolution of the Covid-19 pandemic is both crucial and difficult: Crucial as it is a prerequisite to the design of efficient health policies; Difficult because it must be carried out while the pandemic is still developing (and not retrospectively, after the pandemic is over) and from data - made available by National Health Authorities - that were and still are of limited quality (missing values, outlier values, pseudo-seasonalities, $\cdots$). Within pandemic phases, epidemiologists often have recourse to the notion of time-varying reproduction number, R_t, to quantify the intensity of the pandemics [1]. Research works, conducted during the earliest phase of the pandemics [2, 3], have lead to propose a robust and efficient estimation of R_t based on a *variational formulation* solved by *nonsmooth convex minimization*: the criterion results from augmenting the epidemiological model in [1] with an error management strategy specifically built to handle low quality data [3]. Further and more recently, a *Bayesian* reading of this criterion allowed us to obtain *credibility intervals* (CI) estimates of R_t computed via Markov chain Monte Carlo (MCMC) sampling [4, 5]. However, the design of such samplers for the Bayesian model at hand turns out to be complex: the resulting *a posteriori* density π with respect to the Lebesgue measure on $\mathbb{R}^d$ is concave and takes the general form:

$$\pi \propto \exp(-(f + g(\mathsf{C}\cdot)))\mathbb{1}_{\mathcal{D}}, \tag{1}$$

where $\mathbb{1}_{\mathcal{D}}$ is the $\{0,1\}$-valued indicator function of the support $\mathcal{D}$ of π, $-f$ is a differentiable log-likelihood, C is a linear operator and $-g(\mathsf{C}\cdot)$ is a nonsmooth log-prior. The composite structure of $\ln \pi$ and the composition of the nonsmooth component with a linear operator makes the MCMC sampling challenging.

Related work. Numerous applications require to design MCMC methods targeting log-concave composite distributions [6, 7, 4]. From the seminal paper by [8] on Langevin Monte Carlo (LMC) methods, many strategies taking benefit of first order information on π to design Markov chains approximating π were proposed. When π is smooth, the most popular is the Metropolis-Adjusted Langevin Algorithm sampler (MALA) [9] which uses a LMC iteration as a proposal mechanism in a Hastings-Metropolis (HM) sampler ([10], see also [11]): $\boldsymbol{\tau}_{n+\frac{1}{2}} \sim \mu(\boldsymbol{\tau}_n) + \mathcal{N}(0, 2\gamma\,\mathrm{I})$ with $\mu(\boldsymbol{\tau}) := \boldsymbol{\tau} + \gamma \nabla \ln \pi(\boldsymbol{\tau})$. This proposition step is then followed with an accept-reject (AR) step: $\boldsymbol{\tau}_{n+1} = \boldsymbol{\tau}_{n+1/2}$ if the candidate is accepted and $\boldsymbol{\tau}_{n+1} = \boldsymbol{\tau}_n$ otherwise. The AR step guarantees that the MCMC sampler admits π as unique invariant density (see e.g., [11]) - and that the chain $\{\boldsymbol{\tau}_n, n \geq 0\}$ is $\mathcal{D}$-valued. To handle a nonsmooth density π, [12] proposes the use of the *proximal* operator prox ([13], see also [14, Chapter 1]) leading to P−ULA (or P−MALA when an AR step is included). Because the composition of a nonsmooth function g with a linear operator C impairs an explicit analytical computation of $\mathrm{prox}_{g(\mathsf{C}\cdot)}$, even when prox_g is known in closed-form, a direct application of P−ULA to a density π of the form (1) is not possible and one has to resort to more advanced schemes. In [15] and [16], the drifts $\boldsymbol{\mu}$ rely on a differentiable approximation of $g(\mathsf{C}\cdot)$ through its Moreau envelope (see e.g. [14, Chapter 1]), which leads to a regular approximation of $f + g(\mathsf{C}\cdot)$, whose gradient involves the proximal operator of g; $\boldsymbol{\mu}$ is then *the sum* of a gradient term related to f and of a proximal term associated with g. The drifts proposed by [4] and [5] define $\boldsymbol{\mu}$ as *the composition* of a gradient term related to f and of a proximal term related to g. There are pro and cons for the introduction of the AR step (see e.g. [17]) - when not introduced, one set $\boldsymbol{\tau}_{n+1} = \boldsymbol{\tau}_{n+1/2}$. Note however that when $\mathcal{D}$ is not $\mathbb{R}^d$, AR forces the chain $\{\boldsymbol{\tau}_n, n \geq 0\}$ to remain in $\mathcal{D}$ which is not guaranteed by Gaussian proposal distributions $\mathcal{N}(\boldsymbol{\mu}(\boldsymbol{\tau}), 2\gamma\,\mathrm{I})$. More elaborated LMC-based proposal mechanisms yield a $\mathcal{D}$-valued Markov chain by using projections or reflections on the boundaries of $\mathcal{D}$, see e.g. [18, 19] nevertheless, they require assumptions on π and on the topology of $\mathcal{D}$.

Another class of methods for nonsmooth target densities is proposed in [20]: it relies on variable splitting and data augmentation schemes inspired by the Alternating Direction Method of Multipliers algorithm. In this paper, for a fair comparison of the methods, we restrict our attention to MCMC samplers without data augmentation – a technique which may get simpler distributions, easier to sample

P. Abry is partially supported by Grant 80PRIME-2021 CNRS. G. Fort is partly funded by the *Fondation Simone et Cino Del Duca, Institut de France*.

from but at the price of a subtle balance between augmentation of the dimension of the sampling space and faster convergence rate inherited from a reduced correlation between samples.

Outline, Goals and contributions. The first goal of the present work is to frame the relations between the proximal-Langevin based HM algorithms proposed in [4, 5, 15] and [16], and to compare their performance on a toy example. The second goal is to use the most promising sampler for solving the estimation of credibility intervals for the daily reproduction numbers R_t of the Covid-19 pandemic. Section 2 provides, as a first contribution, a detailed and organized review of existing and proposed proximal-Langevin based HM algorithms. Then, as a second contribution of this paper, performance of these different HM samplers are compared in Section 3, on a toy density. One of the samplers we proposed in a previous work (see [5]) is put forward from performance assessment in Section 3. A last novelty of this paper, is to plug this sampler into a Metropolis-within-Gibbs strategy for the estimation of the reproduction number R_t from real Covid-19 new infection counts, made available from the John Hopkins repository for 200+ territories in the world (cf. Section 4). Its relevance and potential interests are discussed in details.

2. PROXIMAL-LANGEVIN HASTINGS-METROPOLIS

This section is devoted to Langevin-based MCMC samplers, when the target density π with respect to the Lebesgue measure on $\mathbb{R}^d$ satisfies: **(H1)** π is of the form (1); **(H2)** $f : \mathcal{D} \to \mathbb{R}$ is continuously differentiable on $\mathcal{D}$; **(H3)** $g : \mathcal{D} \to \mathbb{R}$ is a convex lower semi-continuous function, and possesses a proximal operator having a closed-form expression, which is assumed not to be the case for the proximal operator of $g(\mathsf{C}\cdot)$. We restrict our attention to HM samplers, and discuss how to design more efficient proposal distributions than in the Random Walk HM, from first order informations on the target density π. Since $-\ln\pi$ is a composite function (see **H1, H2** and **H3**), we consider combinations of gradient steps with respect to f (see **H2**) and implicit gradient steps based on the proximal operator of g (see **H3**). Let us describe four possible proposal mechanisms, some of them requiring the additional assumption: **(H4)** C is a $d \times d$ invertible matrix. Set $\Gamma := \mathsf{C}^{-1}\mathsf{C}^{-\top}$ and let $\gamma, \rho > 0$.

Moreau (`M`). First, let us use the `semi-FBLMC` iteration proposed in [16]. It relies on the Moreau envelope of g with parameter ρ, which is a continuously differentiable function g_ρ satisfying $\nabla[g_\rho](\boldsymbol{\tau}) = \rho^{-1}\left(\boldsymbol{\tau} - \mathrm{prox}_{\rho\, g}(\boldsymbol{\tau})\right)$. Given the current value $\boldsymbol{\tau}_n$ of the chain, a jump to $\boldsymbol{\tau}_{n+\frac{1}{2}} \sim \boldsymbol{\mu}_{\mathtt{M}}(\boldsymbol{\tau}_n) + \mathcal{N}_d(0, 2\gamma\,\mathrm{I})$ is proposed, where $\boldsymbol{\mu}_{\mathtt{M}}(\boldsymbol{\tau}_n) := \boldsymbol{\tau}_n - \gamma\nabla f(\boldsymbol{\tau}_n) - \gamma\rho^{-1}\left(\mathsf{C}\boldsymbol{\tau}_n - \mathrm{prox}_{\rho\, g}(\mathsf{C}\boldsymbol{\tau}_n)\right)$.

Moreau in an image space (`Mdual`). The gradient of the Moreau envelope of $g(\mathsf{C}\cdot)$ has no closed-form expression (see **H3**) and we propose a change of variable to remove C from the nonsmooth component of $-\ln\pi$ by using the following observation (see e.g. [21, Section 3] and [22, Section 4.2.] for similar ideas): a Markov chain $\{\boldsymbol{\tau}_n, n \geq 0\}$ with stationary distribution π can be obtained from a Markov chain $\{\tilde{\boldsymbol{\tau}}_n, n \geq 0\}$ targeting $\tilde{\pi}(\cdot) \propto \pi(\mathsf{C}^{-1}\cdot)$ by setting $\boldsymbol{\tau}_n := \mathsf{C}^{-1}\tilde{\boldsymbol{\tau}}_n$. We have $-\ln\tilde{\pi}(\tilde{\boldsymbol{\tau}}) = f(\mathsf{C}^{-1}\tilde{\boldsymbol{\tau}}) + g(\tilde{\boldsymbol{\tau}})$ with support $\{\tilde{\boldsymbol{\tau}} : \mathsf{C}^{-1}\tilde{\boldsymbol{\tau}} \in \mathcal{D}\}$. A Moreau envelope approach for the smoothing of $-\ln\tilde{\pi}$ is known as a `ULA` iteration (see [12, 15]). It proposes a jump to $\tilde{\boldsymbol{\tau}}_{n+\frac{1}{2}} \sim \boldsymbol{\mu}_{\mathtt{Mdual}}(\tilde{\boldsymbol{\tau}}_n) + \mathcal{N}_d(0, 2\gamma\,\mathrm{I})$ where $\boldsymbol{\mu}_{\mathtt{Mdual}}(\tilde{\boldsymbol{\tau}}_n) := \tilde{\boldsymbol{\tau}}_n - \gamma\mathsf{C}^{-\top}\nabla f(\mathsf{C}^{-1}\tilde{\boldsymbol{\tau}}_n) - \frac{\gamma}{\rho}\left(\tilde{\boldsymbol{\tau}}_n - \mathrm{prox}_{\rho\, g}(\tilde{\boldsymbol{\tau}}_n)\right)$. In the original sampling space, this is equivalent to a jump from $\boldsymbol{\tau}_n$ to $\boldsymbol{\tau}_{n+\frac{1}{2}} \sim \boldsymbol{\tau}_n - \gamma\Gamma\nabla f(\boldsymbol{\tau}_n) - \gamma\rho^{-1}\left(\boldsymbol{\tau}_n - \mathrm{prox}^{\Gamma^{-1}}_{\rho\, g(\mathsf{C}\cdot)}(\boldsymbol{\tau}_n)\right) + \mathcal{N}_d(0, 2\gamma\Gamma)$; prox^A_h is the proximal operator of h in the metric induced by the positive definite matrix A (see e.g. [23, Section XV.4]).

Proximal-Gradient decomposition (`PGdec`). This method applies when $g(\mathsf{C}\cdot) = \sum_{j=1}^J g_j(\mathsf{C}_j\cdot)$ and $g_j(\mathsf{C}_j\cdot)$ has a proximal operator with a closed form expression; see Section 3 for an example. In [5], we introduced the following proposal distribution : sample J_n at random in $\{1, \cdots, J\}$ and set $\boldsymbol{\tau}_{n+\frac{1}{2}} \sim \boldsymbol{\mu}_{\mathtt{PGdec}}(\boldsymbol{\tau}_n, J_n) + \mathcal{N}_d(0, 2\gamma\,\mathrm{I})$ where $\boldsymbol{\mu}_{\mathtt{PGdec}}(\boldsymbol{\tau}_n, J_n) := \mathrm{prox}_{\gamma g_{J_n}(\mathsf{C}_{J_n}\cdot)}\left(\boldsymbol{\tau}_n - \gamma\nabla f(\boldsymbol{\tau}_n)\right)$.

Proximal-Gradient in an image space (`PGdual`). As above for `Mdual`, let us apply a Proximal-Gradient step in the image of $\mathcal{D}$ by C: given $\tilde{\boldsymbol{\tau}}_n$, sample $\tilde{\boldsymbol{\tau}}_{n+\frac{1}{2}} \sim \boldsymbol{\mu}_{\mathtt{PGdual}}(\tilde{\boldsymbol{\tau}}_n) + \mathcal{N}_d(0, 2\gamma\,\mathrm{I})$ where $\boldsymbol{\mu}_{\mathtt{PGdual}}(\tilde{\boldsymbol{\tau}}_n) := \mathrm{prox}_{\gamma g}\left(\tilde{\boldsymbol{\tau}}_n - \gamma\mathsf{C}^{-\top}\nabla f(\mathsf{C}^{-1}\tilde{\boldsymbol{\tau}}_n)\right)$. In the original sampling space, this is equivalent to a jump from $\boldsymbol{\tau}_n$ to $\boldsymbol{\tau}_{n+\frac{1}{2}} \sim \mathrm{prox}^{\Gamma^{-1}}_{\gamma g(\mathsf{C}\cdot)}\left(\boldsymbol{\tau}_n - \gamma\Gamma\nabla f(\boldsymbol{\tau}_n)\right) + \mathcal{N}_d(0, 2\gamma\Gamma)$, which is a Variable Metric Forward-Backward iteration ([24], see also [25]).

Accept-reject step. None of these proposal strategies takes into account the support $\mathcal{D}$ of π; in order to obtain a $\mathcal{D}$-valued sequence $\{\boldsymbol{\tau}_n, n \geq 0\}$, an accept-reject step is included at each iteration.

Remark. When $g = 0$ (i.e. when $-\ln\pi$ is continuously differentiable), `M` and `PGdec` are the same algorithms and are equal to `MALA` ([8, 26]); and `Mdual` and `PGdual` are the same algorithms and are equal to a *tempered* `MALA` ([27, 28]).

3. TOY EXAMPLE

Definition. Let π_{t} be the penalized likelihood in a logistic regression model, defined on $\mathcal{D} := \mathbb{R}^d$:

$$\ln\pi_{\mathsf{t}}(\boldsymbol{\tau}) := \mathsf{Y}^\top\mathsf{X}\boldsymbol{\tau} - \sum_{j=1}^N \ln\left(1 + \exp((\mathsf{X}\boldsymbol{\tau})_j)\right) - \lambda\|\mathsf{D}_1\boldsymbol{\tau}\|_1;$$

$\mathsf{Y} \in \{0,1\}^N$ being the vector collecting the binary responses, X the $N \times d$ matrix of covariates, D_1 the $(d-1) \times d$ discrete differentiation matrix and $\lambda > 0$ the regularization parameter.

Samplers. By augmenting D_1 with a raw into an invertible matrix denoted by $\overline{\mathsf{D}}_1$, π_{t} satisfies (**H1**-**H4**) and the four methods proposed in Section 2 apply with $\mathsf{C} := \overline{\mathsf{D}}_1$ and $g(\boldsymbol{\tau}) := \lambda\|\boldsymbol{\tau}_{2:d}\|_1$ for any $\boldsymbol{\tau} \in \mathbb{R}^d$. For `PGdec`, the penalization decomposes into $\|(\overline{\mathsf{D}}_1\boldsymbol{\tau})_{2:d}\|_1 = \|\mathsf{D}_{1,p}\boldsymbol{\tau}\|_1 + \|\mathsf{D}_{1,i}\boldsymbol{\tau}\|_1$, where $\mathsf{D}_{1,p}$ (resp. $\mathsf{D}_{1,i}$) concatenates the lines of D_1 of even (resp. odd) indices; these two matrices satisfy $\mathsf{D}_{1,x}\mathsf{D}_{1,x}^\top = \nu\,\mathrm{I}$ for $x \in \{i,p\}$ and some $\nu > 0$ (here $\nu = 1$) which makes the computation of the proximity operator of $\|\mathsf{D}_{1,p}\cdot\|_1$ and $\|\mathsf{D}_{1,i}\cdot\|_1$ explicit [29]. To implement the samplers working in the image space, the first raw added to D_1 to obtain the invertible $\overline{\mathsf{D}}_1$ is equal to the projection of the vector $(-1, 0, \cdots, 0) \in \mathbb{R}^d$ on the space orthogonal to the range of D_1. For `M` and `Mdual`, we choose $\rho := \gamma$. For all these methods, the value of γ is adapted during the $5\,10^3$ first iterations in order to reach an accept-reject rate of 0.25. We also run `RW` and `RWdual` which correspond resp. to $\boldsymbol{\tau}_{n+\frac{1}{2}} \sim \boldsymbol{\tau}_n + \mathcal{N}(0, \mathrm{I})$ and $\boldsymbol{\tau}_{n+\frac{1}{2}} \sim \boldsymbol{\tau}_n + \mathcal{N}(0, \Gamma)$, where $\Gamma := \overline{\mathsf{D}}_1^{-1}\overline{\mathsf{D}}_1^{-\top}$. The comparison to the four proposed approaches enable to asses the relevance of methods exploiting first order information about $\mathcal{L} := \ln\pi_{\mathsf{t}}$.

Compared performances. To compare the performances of the different samplers, Log π measures the relative distance to the maximum $\mathcal{L}_\star$ ($-\mathcal{L}$ being strictly convex) of $\mathcal{L}$; the evolution of $\left(\mathcal{L}(\boldsymbol{\tau}^n) - \mathcal{L}_\star\right) / \left(\mathcal{L}(\boldsymbol{\tau}^1) - \mathcal{L}_\star\right)$ is plotted over the first $2\,500$ iterations of the chain. ACF represents the absolute value of the autocorrelation function in terms of the lag (from 0 to 600), computed from $17\,500$ points obtained after a burn-in of $2\,500$ points. The ACF reflects the ergodicity of the chain: the smaller the better [11]. The data set is simulated: $N = 2.10^3$, $d = 20$; the matrix of covariates X is built from independent Rademacher random variables (r.v.), the rows of X being afterward normalized to 1; the

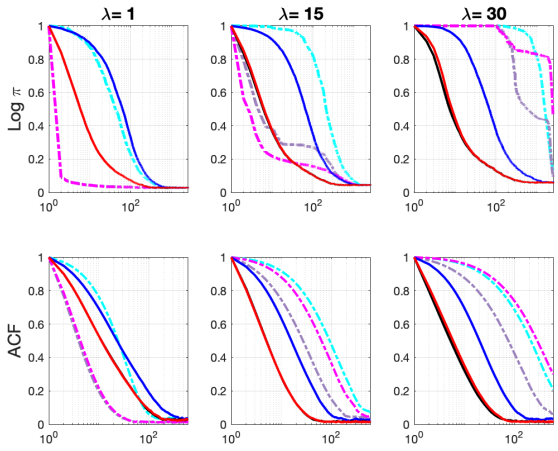

Fig. 1: **Evolution of *Log* π with respect to the number of iterations, and of the *ACF* with respect to the lag.** In dashed lines: `RW` in cyan, `M` in gray et `PGdec` in magenta. Methods based on the change of space are plotted in solid lines: `RWdual` in blue, `Mdual` in black and `PGdual` in red. The curves for `PGdual` and `Mdual` are often superimposed.

components of Y are independent Bernoulli r.v. with probability of success $(1+\exp(-(\mathsf{X}\boldsymbol{\tau}^\star)_j))^{-1}$ where $\boldsymbol{\tau}^\star$ is blockwise constant, with six components equal to -2, then seven at 1.5, and then seven at -1. Finally, three different values of the regularization parameter λ are considered: $\lambda \in \{1, 15, 30\}$. Performances obtained as averages over 50 independent realizations are reported in Figure 1. It illustrates the gain of exploiting first order information on π to accelerate the move of the sampler toward high probability regions. For `RW`, `PGdec` and `M`, this gain decreases when λ is large. Exploiting partial information, as done by `PGdec`, remains possible when λ is small; for larger values of λ, `PGdec` is less efficient than `M` in the stationary regime of the chain (see the *ACF* criterion). The schemes proceeding in the image space, `PGdual` and `Mdual`, appears to be extremely robust to very different values of λ, with a slight advantage for `PGdual` when λ is medium to large. At last, altogether the plots of Figure 1 advocates to prefer the use of *dual* samplers, in particular `PGdual` and `Mdual` whose performance are systematically good, and better for medium to large values of λ.

4. COVID-19 PANDEMIC INTENSITY ASSESSMENT

This section provides CI estimates of the reproduction numbers $\mathbf{R} := (\mathsf{R}_1, \cdots, \mathsf{R}_T)$ from day 1 to day T, defined by estimated quantiles of an *a posteriori* distribution on $\mathbf{R}$. We first introduce the Bayesian model, and then describe the MCMC sampler used to explore the *a posteriori* distribution and compute empirical quantiles.

Model. The model we consider is built on the ones by [1] and [3]. [1] proposed an epidemiology model designed to estimate the reproduction numbers $\mathbf{R}$ from the observed counts of new daily infections, denoted by $\mathbf{Z} = (\mathsf{Z}_1, \cdots, \mathsf{Z}_T)$. Given the past $\mathsf{Z}_1, \cdots \mathsf{Z}_{t-1}$, Z_t is a Poisson distribution with intensity $\mathbf{R}_t\Phi_t$ where $\Phi_t := \sum_{u=1}^{\tau_\phi} \phi(u)\mathsf{Z}_{t-u}$ and $\phi(\cdot)$ is the probability density function of a Gamma distribution with mean 6.68 days, standard deviation 3.53 days and a mode at 4.8 days. At time t, Φ_t is a weighted mean of the number of new infections over the past τ_ϕ days, with weights $\phi(\cdot)$ which model the delay between a primary infection and a second one. In [3], an extension of this model is designed to estimate simultaneously the reproduction numbers $\mathbf{R}$ and the errors $\mathbf{O} := (\mathsf{O}_1, \cdots, \mathsf{O}_T)$ when counting and reporting the true number of daily new infections. By using the model of [3], we defined in [5] an *a posteriori* distribution π of the parameter $\boldsymbol{\theta} := (\mathbf{R}, \mathbf{O})$ as follows. The support of π is

$$\mathcal{D} := \{\boldsymbol{\theta} \in (\mathbb{R}_+)^T \times \mathbb{R}^T : \mathsf{R}_t\Phi_t + \mathsf{O}_t \geq 0 \text{ if } \mathsf{Z}_t = 0\}$$
$$\cup \{\boldsymbol{\theta} \in (\mathbb{R}_+)^T \times \mathbb{R}^T : \mathsf{R}_t\Phi_t + \mathsf{O}_t > 0 \text{ if } \mathsf{Z}_t > 0\}.$$

On $\mathcal{D}$, $-\ln\pi(\boldsymbol{\theta}) = F(\boldsymbol{\theta}) + G(\boldsymbol{\theta})$, up to an additive constant. The function $G(\boldsymbol{\theta}) := \lambda_{\mathsf{R}}\|\mathsf{D}_2\mathbf{R}\|_1 + \lambda_{\mathsf{O}}\|\mathbf{O}\|_1$ where $\lambda_{\mathsf{R}}, \lambda_{\mathsf{O}} > 0$ and D_2 is proportional to the second order derivative operator $((\mathsf{D}_2\mathbf{R})_t := (\mathsf{R}_{t+2} - 2\mathsf{R}_{t+1} + \mathsf{R}_t)/\sqrt{6}$ for $t = 1, \cdots, T-2)$, plays the role of a negative log-prior on $\boldsymbol{\theta}$. The function $F(\boldsymbol{\theta})$ is the negative log-likelihood of Poisson distributions with intensities $\mathsf{R}_t\Phi_t + \mathsf{O}_t$ for $t = 1, \cdots, T$:

$$F(\boldsymbol{\theta}) := \sum_{t=1}^{T} \{(\mathsf{R}_t\Phi_t + \mathsf{O}_t) - \mathsf{Z}_t \ln(\mathsf{R}_t\Phi_t + \mathsf{O}_t)\}.$$

MCMC sampler. We consider a Metropolis-within-Gibbs sampler: since both conditional distributions of $\mathbf{R}$ given $\mathbf{O}$ (denoted by $\pi_1(\cdot|o)$) and $\mathbf{O}$ given $\mathbf{R}$ (denoted by $\pi_2(\cdot|r)$) can not be exactly sampled from, each iteration of the Gibbs sampler calls two HM kernels. Given the current value of the chain $(\mathbf{R}_n, \mathbf{O}_n)$, *(i)* update first the $\mathbf{R}$-component by sampling from a HM kernel $P_1(\mathbf{O}_n, \cdot)$ targeting $\pi_1(\cdot|\mathbf{O}_n)$ and obtain $\mathbf{R}_{n+1}$ and then *(ii)* update the $\mathbf{O}$-component by sampling from a HM kernel $P_2(\mathbf{R}_{n+1}, \cdot)$ targeting $\pi_2(\cdot|\mathbf{R}_{n+1})$ and obtain $\mathbf{O}_{n+1}$. From the expressions of F, G and $\mathcal{D}$, it may be seen that $-\ln\pi_1(\cdot|\mathbf{O}_n) = F(\cdot, \mathbf{O}_n) + \lambda_{\mathsf{R}}\|(\overline{\mathsf{D}}_2\cdot)_{3:T}\|_1$ on $\{r \in (\mathbb{R}_+)^T : (r, \mathbf{O}_n) \in \mathcal{D}\}$, where $\overline{\mathsf{D}}_2$ is an invertible augmentation of D_2 obtained by adding two rows on the top of D_2; and $-\ln\pi_2(\cdot|\mathbf{R}_{n+1}) = F(\mathbf{R}_{n+1}, \cdot) + \lambda_{\mathsf{O}}\|\cdot\|_1$ on $\{o \in \mathbb{R}^T : (\mathbf{R}_{n+1}, o) \in \mathcal{D}\}$. We can therefore use any of the HM kernels described in Section 2. Based on the results of Section 3, we will use the `PGdual` HM kernel.

MCMC simulations. In the present work, $\tau_\phi := 26$ and $(\lambda_{\mathsf{R}}, \lambda_{\mathsf{O}}) := (3.5\,\sigma_{\mathbf{Z}}\sqrt{6}/4, 0.05)$ with $\sigma_{\mathbf{Z}}$ denoting the standard deviation of $\mathbf{Z}$. This setting avoids the dependence of the order of magnitude of population or pandemic intensity on the choice of λ_{R} and thus permits to use the same hyperparameters for most countries (see [3]). The first two rows of $\overline{\mathsf{D}}_2$ are orthogonal and orthogonal to the rows of D_2; they are obtained from the vectors $(1, 0, \cdots, 0)$ and $(-2, 1, 0, \cdots, 0)/\sqrt{5}$. As discussed in [4], the augmentation $\overline{\mathsf{D}}_2$ is not unique and this definition is numerically efficient.

Gibbs chains run for 10^7 iterations with a burn-in phase of 3.10^6 iterations. They are initialized at the Maximum a Posteriori, computed with a proximal convex non-smooth minimization algorithm (see [3]). For each `PGdual` HM kernel, the step size γ is adapted during the burn-in phase in order to reach a mean accept-reject rate of 25% [30]. The empirical quantiles are estimated from the ouput of the Gibbs sampler, after burn-in; the 95% CIs are computed for each component R_t of $\mathbf{R}$ and O_t of $\mathbf{O}$, from the 2.5% and 97.5% empirical quantiles.

Covid-19 data. The Covid-19 data used here are those made available at the *Johns Hopkins University*[1]. This repository collects and organizes, remarkably every day since the earliest phase of the pandemic till today, Covid19 pandemic data made available by the National health Authorities of 200+ countries or territories across the world. Use is made here of only daily new infection counts Z_t for the most recent 10 weeks ($T = 70$ days, prior to the present day, Oct. 11th, 2022) and a few countries. Yet, tools devised here can be applied to any period of times and any country of interest. Counts Z_t are not preprocessed prior to estimation, except for low-a priori informed 7-day sliding median filtering, that removes totally irrelevant values, i.e., counts beyond $\pm$ 7 local standard deviations.

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

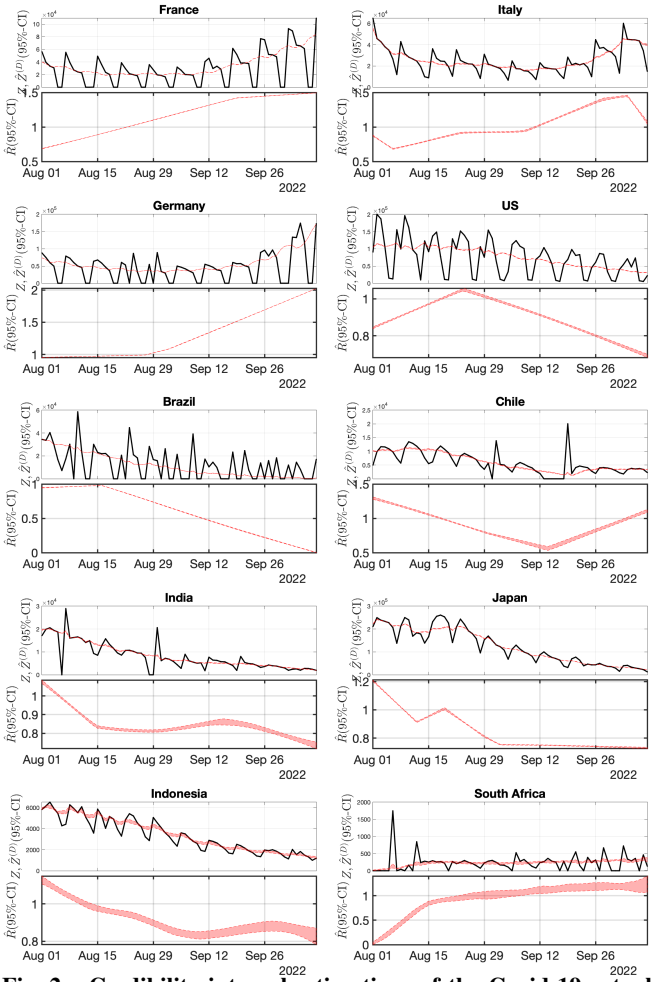

Fig. 2: **Credibility interval estimations of the Covid-19 actual infection counts and of the reproduction numbers. For several countries** Top: Observed counts (black) and 95% CIs estimations of actual counts of new infections (red); Bottom: 95% CIs estimations of the R_t's.

Covid-19 pandemic intensity assessment. Fig. 2 displays for different countries across the world, the reported daily new infection counts Z_t (top plots, black solid line). 95%-CIs of pseudo denoised counts $\mathbf{Z}^{(D)} := (\mathsf{Z}_1^{(D)}, \cdots, \mathsf{Z}_T^{(D)})$ are superimposed (top plots, red filled bands), where $\mathsf{Z}_t^{(D)} := \mathsf{Z}_t - \mathsf{O}_t$ and CIs for $\mathsf{Z}_t^{(D)}$ stem from those of O_t. Finally, the 95%-CIs for R_t are displayed (bottom plots, red filled bands). Fig. 2 permits to draw a number of conclusions.

First, and at the methodological level, 95%-CIs yield realistic estimates of both $\mathsf{Z}_t^{(D)}$ and R_t for many different countries, and this, with a common choice of parameters $(\lambda_{\mathsf{R}}, \lambda_{\mathsf{O}})$ for all countries. Observe also that the CIs are very narrow (of the order of a few times 10^{-2} for R_t), which indicates that given the actual Covid-19 data, the modeling proposed here, and the choice of $(\lambda_{\mathsf{R}}, \lambda_{\mathsf{O}})$, there remains little uncertainty on estimates. Further, Fig. 2 shows that CIs tend to be larger both at error in reported counts locations, and at changes in the pandemic intensity behavior, as measured by R_t (e.g. when it switches from increasing to decreasing).

Second and at the application level, Fig. 2 shows that $\mathbf{Z}^{(D)}$ provides a smooth temporal evolution, that satisfactorily *cancels* the 7-day pseudo-seasonality when it exists in data, with filling up missing counts and decreasing mis-reported subsequent counts. This is obtained with no a priori information related to seasonalities or more generally, to the calendar. Other erroneous values not associated with the pseudo seasonality are also satisfactorily removed. Further, the CI estimates of $\mathbf{R}$ also provide a very regular evolution of $t \mapsto \mathsf{R}_t$ (almost piecewise linear). In agreement with a potential use of these analyses by epidemiologists, it indicates not only whether R_t is above or below 1, thus betraying an increase/decrease of the intensity of the pandemics, but also whether the trend is locally increasing or decreasing, quantifying an acceleration/deceleration in the growth/decrease of the pandemics. For example, currently, the pandemics for most European countries is within an accelerating growth phase (R_t above one and with linearly increasing trend), while for most other countries of the world, R_t is below 1 and with stable or decreasing trends, indicating a slowly regressing pandemics at the time of writing.

It is also worth mentioning that the model, notably with a L^1-norm constraint on $\mathsf{D}_2\mathbf{R}$, the second derivative of the $\mathsf{R}_t's$, promotes smooth quasi-piecewise linear estimates of R_t. Strictly speaking, while it only performs estimations at current date, it implicitly achieves *nowcasting* (as opposed to *forecasting*), i.e., a prediction of the short-term evolution of the pandemics. Given the intrinsic time scales of the pandemics, controlled by the serial distribution function $\phi(\cdot)$, an estimated change in the intensity of R_t predicts a change in the evolution of the daily new infection counts about two weeks after. Hence, a change from linear increase to linear decrease in R_t today announces that a current *Covid wave* should reach its maximum within the next two/three weeks. At the time of writing, the maximum of the 8th wave for Europe can not yet be detected. The decreasing R_t recently observed for Italy, for instance, predicts that the maximum of infections should be reached in the coming days. For France, the lesser increase in R_t, observed after Sept. 18th, indicates a slower acceleration of the pandemics, and thus might be read as preliminaries to a pandemic deceleration. These observations, at the time of writing, may constitute the first signs of the soon to come maximum of the 8th wave for Europe. Conversely, the start of the current 8th wave in France was visible as early as the first days of August (as for most West-European countries) by a change from decrease to increase in R_t.

Finally, the estimate of R_t across a chosen period permits a retrospective analysis of the impacts of sanitary policies in given countries, by quantifying the subsequent evolution of R_t. As an interesting outcome, theses analyses show that the pandemics developed across West-European countries, essentially with the same dynamics and despite sometimes very different sanitary policies.

5. CONCLUSIONS AND PERSPECTIVES

This paper presents MCMC samplers designed to explore log-concave composite densities possibly with a constrained support. Their performances are compared on a toy problem. Their relevance and interest have been illustrated on the assessment of the intensity in the Covid-19 pandemics, from daily new infection counts: credibility intervals for the reproduction number of the pandemic are updated automatically on a daily-basis for many countries and made available on the authors web pages[2]. Future works will be about hyperparameters: either automated and data-driven selections, or a Bayesian approach to analyze how the credibility intervals for the R_t's are affected by uncertainty on $(\lambda_{\mathsf{R}}, \lambda_{\mathsf{O}})$. They will also include a sequential statistical analysis to take into account the daily new observations.

[2]perso.math.univ-toulouse.fr/gfort/project/opsimore-2/, https://perso.ens-lyon.fr/patrice.abry.

6. REFERENCES

[1] 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.

[2] 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.

[3] 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," *IEEE Trans. Signal Process.*, vol. 70, pp. 2859–2868, 2022.

[4] H. Artigas, B. Pascal, G. Fort, P. Abry, and N. Pustelnik, "Credibility interval design for Covid19 reproduction number from nonsmooth Langevin-type Monte Carlo sampling," in *Proceedings of EUSIPCO*, 2022.

[5] G. Fort, B. Pascal, P. Abry, and N. Pustelnik, "Covid19 Reproduction Number: Credibility Intervals by Blockwise Proximal Monte Carlo Samplers," Tech. Rep. hal-03611079v1, HAL, 2022.

[6] S. Makni, J. Idier, T. Vincent, B. Thirion, G. Dehaene-Lambertz, and P. Ciuciu, "A fully Bayesian approach to the parcel-based detection-estimation of brain activity in fMRI," *Neuroimage*, vol. 41, no. 3, pp. 941–969, 2008.

[7] C. Vacar and J.-F. Giovannelli, "Unsupervised joint deconvolution and segmentation method for textured images: a Bayesian approach and an advanced sampling algorithm," *EURASIP Journal on Advances in Signal Processing*, vol. 2019, no. 1, pp. 1–17, 2019.

[8] G. Parisi, "Correlation functions and computer simulations," *Nucl. Phys. B*, vol. 180, no. 3, pp. 378–384, 1981.

[9] K. L. Mengersen and R. L. Tweedie, "Rates of convergence of the Hastings and Metropolis algorithms," *Ann. Stat.*, vol. 24, no. 1, pp. 101–121, 1996.

[10] W.K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.

[11] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer-Verlag, Berlin, Heidelberg, 2005.

[12] M. Pereyra, "Proximal Markov chain Monte Carlo algorithms," *Statistics and Computing*, vol. 26, pp. 745–760, 2016.

[13] J.J. Moreau, "Proximité et dualité dans un espace hilbertien," *Bulletin de la Société Mathématique de France*, vol. 93, pp. 273–299, 1965.

[14] R.T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer Verlag, Heidelberg, Berlin, New York, 1998.

[15] 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.

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

[17] R. Dwivedi, Y. Chen, M.J. Wainwright, and B. Yu, "Log-concave sampling: Metropolis-Hastings algorithms are fast," *Journal of Machine Learning Research*, vol. 20, no. 183, pp. 1–42, 2019.

[18] S. Bubeck, R. Eldan, and J. Lehec, "Sampling from a Log-Concave Distribution with Projected Langevin Monte Carlo," *Discrete Comput Geom*, vol. 59, pp. 757–783, 2018.

[19] S. Melidonis, P. Dobson, Y. Altmann, M. Pereyra, and K.C. Zygalakis, "Efficient bayesian computation for low-photon imaging problems," Tech. Rep. arXiv:2206.05350, arXiv, 2022.

[20] M. Vono, N. Dobigeon, and P. Chainais, "Split-and-augmented Gibbs sampler—Application to large-scale inference problems," *IEEE Trans. Signal Process.*, vol. 67, no. 6, pp. 1648–1661, 2019.

[21] R.J. Tibshirani and J. Taylor, "The solution path of the generalized lasso," *Ann. Stat.*, vol. 39, no. 3, pp. 1335 – 1371, 2011.

[22] A.S. Dalalyan, "Theoretical guarantees for approximate sampling from smooth and log-concave densities," *J. Roy. Statist. Soc. B*, vol. 79, no. 3, pp. 651–676, 2017.

[23] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms*, Springer Verlag, Heidelberg, 1996, Two volumes - 2nd printing.

[24] H.H.-G. Chen and R.T. Rockafellar, "Convergence rates in forward-backward splitting," *SIAM J. Optim.*, vol. 7, pp. 421–444, 1997.

[25] P.L. Combettes and B.C. Vũ, "Variable metric forward–backward splitting with applications to monotone inclusions in duality," *Optimization*, vol. 63, no. 9, pp. 1289–1318, 2014.

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

[27] J. Kent, "Time-Reversible Diffusions," *Adv Appl Probab*, vol. 10, no. 4, pp. 819–835, 1978.

[28] G.O. Roberts and O. Stramer, "Langevin Diffusions and Metropolis-Hastings Algorithms," *Methodol. Comput. Appl. Probab.*, vol. 4, pp. 337–357, 2002.

[29] 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.

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