Covid19 Reproduction Number: Credibility Intervals by Blockwise Proximal Monte Carlo Samplers

Gersende Fort ©, Barbara Pascal ©, Member, IEEE, Patrice Abry ©, Fellow, IEEE, and Nelly Pustelnik ©, Member, IEEE

Abstract—Monitoring the COVID19 pandemic constitutes a critical societal stake that received considerable research efforts. The intensity of the pandemic on a given territory is efficiently measured by the reproduction number, quantifying the rate of growth of daily new infections. Recently, estimates for the time evolution of the reproduction number were produced using an inverse problem formulation with a nonsmooth functional minimization. While it was designed to be robust to the limited quality of the COVID19 data (outliers, missing counts), the procedure lacks the ability to output credibility interval based estimates. This remains a severe limitation for practical use in actual pandemic monitoring by epidemiologists that the present work aims to overcome by use of Monte Carlo sampling. After interpretation of the nonsmooth functional into a Bayesian framework, several sampling schemes are tailored to adjust the nonsmooth nature of the resulting posterior distribution. The originality of the devised algorithms stems from combining a Langevin Monte Carlo sampling scheme with Proximal operators. Performance of the new algorithms in producing relevant credibility intervals for the reproduction number estimates and denoised counts are compared. Assessment is conducted on real daily new infection counts made available by the Johns Hopkins University. The interest of the devised monitoring tools are illustrated on Covid19 data from several different countries.

Index Terms—Markov Chain Monte Carlo sampling, nonsmooth convex optimization, Bayesian inverse problems, credibility intervals, Covid19, reproduction number.

I. INTRODUCTION

CONTEXT. The Covid19 pandemic is causing unprecedented health, social, and economic crises. This triggered massive research efforts to design efficient procedures aiming to assess the intensity of the pandemic, a prerequisite to develop efficient sanitary policies [1]. Several indices are commonly used to measure the strength of a pandemic, such as, e.g., the reproduction number of interest here. However, often, the value of the index alone is not sufficient and credibility intervals of these indices constitute valuable information for the decision makers, notably in periods of rapid pandemic evolution or of changes in trends, an issue not always addressed in pandemic monitoring and at the heart of the present work.

Related works. Pandemic monitoring can be conducted with numerous tools from different scientific fields, (cf. [2] for a review), amongst which compartmental models, such as the founding Susceptible-Infectious-Recovered scheme. Within pandemic period, when data are scarce and of limited quality, the reproduction number, $R_t$, is often used by epidemiologists as an efficient practical proxy for the pandemic intensity: it measures the number of second infections caused by one primary infection (cf. e.g., [3], [4], [5], [6], [7]). It thus plays a key role in the pandemic evolution assessment: the number of new infections today, $Z_t$, depends on $R_t$ and on a weighted average of the new infection counts on previous days $\{\ldots, Z_{t-3}, Z_{t-2}, Z_{t-1}\}$; the weight function being the so-called serial interval function $\Phi$, that quantifies the distribution of the random delays between the onsets of symptoms in a primary and secondary cases [6], [7], [8], [9]. It has recently been shown that, within pandemic, reliable estimates for the temporal estimation of $R_t$ can be obtained from an inverse problem formulation resulting in a nonsmooth convex optimization problem [10], [11]. The functional to minimize is built from combining the pandemic model in [7], with time regularity constraints. While the procedure was engineered to produce realistic estimations of the temporal evolution of the reproduction number that are robust to the limited quality of the Covid19 pandemic data (severely corrupted with outliers, missing or negative counts and pseudo-seasonalities), it does not however provide credibility intervals, a critical issue towards its practical and actual use by epidemiologists, the main issue that we aim to address in the present work.

Goals, contributions and outline. The overall goal of the present work is to devise Monte Carlo sampling strategies to perform the estimation by means of credibility intervals of the pandemic reproduction number and of denoised infection counts. To that end, Section II details the proposed statistical model used to embed into a stochastic framework the epidemiological model in [7] and its robust extension to data corruption [11]. Its originality stems from using non-differentiable priors to ensure robustness to data corruption. The uniqueness of the maximum a posteriori is thoroughly studied. Further, Section III devises original sampling schemes tailored to
handle the non-differentiability of the target distribution; We propose two blockwise Proximal-Gradient based extensions of the Langevin Metropolis algorithms: PGdec and PGdual. We establish their ergodicity, and carry out a comparative study. Using real Covid19 data, made available at the Johns Hopkins University repository and described in Section IV, the performance of up to twelve variations of the sampling strategies are assessed and compared, using well-thought indices quantifying their efficiency (cf. Section V). Finally, in Section VI, the relevance of the proposed blockwise Proximal-Gradient samplers is illustrated for several different countries representative of the evolution of the pandemic across the world, for a 5-week recent period. Daily updates of these credibility interval estimates as well as MATLAB routines for their calculations are available at https://github.com/gfort-lab/OpSiMorE.

Notations. Vectors are column-vectors, and are denoted in bold font, e.g., \( \mathbf{r} \in \mathbb{R}^T \). For \( p \leq q \), the vector \( \mathbf{x}_{p,q} \) concatenates the scalars \( x_i \) or vectors \( \mathbf{x}_i \) for \( i = p, \ldots, q \). The matrices are denoted in bold font. For a matrix \( \mathbf{A}, \mathbf{A}^\top \) (resp. det(\( \mathbf{A} \)) and the inverse of \( \mathbf{A} \)). We set \( \mathbf{A}^{-1} \) denotes the transpose of \( \mathbf{A} \) (resp. the determinant and the inverse of \( \mathbf{A} \)).

The model depends on initial values \( \theta \in \Gamma \times \theta_o \). Vectors are column-vectors, and are denoted in \( \mathbb{P}Gdual \) for \( \mathbb{O} \) discrete-time second order \( 0 \) \( \propto \) is the (1)

\[ (\mathbf{f} \mathbf{Z} (\theta) := \sum_{t=1}^{T} \{ T \ln Z_t \ln I_t (\theta, \mathbf{Z}) + I_t (\theta, \mathbf{Z}) \}, \]

where \( I_t (\theta, \mathbf{Z}) \) is a positive intensity. Up to an additive constant, \( f \mathbf{Z} (\theta) = \sum_{t=1}^{T} d_{KL} (Z_t I_t (\theta, \mathbf{Z})) \), where \( d_{KL} \) denotes the Kullback-Leibler divergence related to the log-likelihood of a Poisson process.\(^1\) We choose

\[ I_t (\theta, \mathbf{Z}) := \mathbb{R}^T \sum_{u=1}^{T} \Phi u Z_{t-u} + O_t. \]

\( \Phi := (\Phi_u)_{1 \leq u \leq \tau_o} \) is the serial interval function, describing the average infectiousness profile after infection [7], [8], [9]. \( \Phi \) is assumed known and, following [12], [13], is classically modeled as a Gamma distribution truncated over \( \tau_o = 26 \) days with mean and standard deviation of 6.6 and 3.5 days. The reproduction number \( R_t \) at time \( t \) acts on the rate at which a person infected at time \( t - u \) generates new infections at time \( t \): this rate is equal to \( R_t \Phi u \).

The support of the density \( \pi \mathbf{Z} (\cdot) \) is the measurable subset \( D_{\mathbf{Z}} \) of \( \mathbb{P}Gdual \) ensuring that the intensities \( I_t (\theta, \mathbf{Z}) \) are positive, or non-negative when \( Z_t = 0 \) (with the convention that \( 0 \ln 0 = 0 \) in the expression of \( f \mathbf{Z} \)). This yields

\[ D_{\mathbf{Z}} := \{ \theta : I_t (\theta, \mathbf{Z}) > 0 \text{ for } s.t. Z_t > 0 \} \cup \{ \theta : I_t (\theta, \mathbf{Z}) \geq 0 \text{ for } s.t. Z_t = 0 \}. \]

\( \Phi := (\Phi_u)_{1 \leq u \leq \tau_o} \) is the serial interval function, describing the average infectiousness profile after infection [7], [8], [9]. \( \Phi \) is assumed known and, following [12], [13], is classically modeled as a Gamma distribution truncated over \( \tau_o = 26 \) days with mean and standard deviation of 6.6 and 3.5 days. The reproduction number \( R_t \) at time \( t \) acts on the rate at which a person infected at time \( t - u \) generates new infections at time \( t \): this rate is equal to \( R_t \Phi u \).

The support of the density \( \pi \mathbf{Z} (\cdot) \) is the measurable subset \( D_{\mathbf{Z}} \) of \( \mathbb{P}Gdual \) ensuring that the intensities \( I_t (\theta, \mathbf{Z}) \) are positive, or non-negative when \( Z_t = 0 \) (with the convention that \( 0 \ln 0 = 0 \) in the expression of \( f \mathbf{Z} \)). This yields

\[ D_{\mathbf{Z}} := \{ \theta : I_t (\theta, \mathbf{Z}) > 0 \text{ for } s.t. Z_t > 0 \} \cup \{ \theta : I_t (\theta, \mathbf{Z}) \geq 0 \text{ for } s.t. Z_t = 0 \}. \]

C. A Bayesian Model

Let us describe a Bayesian framework such that \( \pi \mathbf{Z} \) is a posterior distribution. The model depends on initial values \( Z_{1-\tau_o}, R_1, R_2 \), which are omitted in the notations.

\(^1\)Its definition is, for some \( z \in \mathbb{N} \),

\( (\forall \mathbf{z} \in \mathbb{R}) \quad d_{KL} (\mathbf{z} | \mathbf{I}) := \begin{cases} z \ln \frac{z}{I} & + z - I, \\ I & \text{if } z > 0, I > 0, \\ + \infty & \text{if } z = 0, I \geq 0, \\ \end{cases} \) otherwise.
Prior distribution. Conditionally to \((R_1, R_2)\), the prior distribution of \((R_{3:T}, O)\) has the density
\[
\left( \prod_{t=3}^{T} p_1(R_t|R_{t-1}, R_{t-2}) \right) \left( \prod_{t=1}^{T} p_2(O_t) \right)
\]
w.r.t. the Lebesgue measure on \(\mathbb{R}^{T-2} \times \mathbb{R}^T\); we set
\[
p_1(r_t|r_{t-1}, r_{t-2}) := \frac{\lambda_R}{2\sqrt{6}} \exp\left( -\frac{\lambda_R}{\sqrt{6}} |r_t - 2r_{t-1} + r_{t-2}| \right)
\]
\[
p_2(o_t) := \frac{\lambda_O}{2} \exp(-\lambda_O |o_t|).
\]

It is easily seen that this prior density is equal to \(\exp(-g(\theta))\) up to a multiplicative constant (depending on \(\lambda_R\) and \(\lambda_O\)). Such a prior states that \(R_{3:T}\) and \(O\) are mutually independent conditionally to \((R_1, R_2)\). Further, the outliers \(O\) are independent and distributed as a Laplace distribution (with parameter \(\lambda_O > 0\)) as commonly encountered in the literature (see, e.g., [14, 15, 16]). Finally \(R_{3:T}\) is distributed as an AR(2) process with Laplace distribution (with parameter \(\lambda_R > 0\)). \(\lambda_R\) and \(\lambda_O\) are (fixed) positive regularization hyperparameters. \(g\) favors some large values among many small ones of the components of both vectors \(D\) and \(O\); it models smooth piecewise linear time evolutions for \(R\), or equivalently a sparse set of components where the discrete second order derivative in time of \(R\) is non zero.

Likelihood. Conditionally to \((R, O)\), the observations are not independent and the joint distribution of \(Z\) is
\[
\prod_{t=1}^{T} p_3(Z_t|Z_{t-\tau_6:t-1}, R_t, O_t).
\]

When \(R_t \geq 0\) and \(I_t(\theta, Z) > 0\) (see (5) for the definition of \(I_t\), following [7, 11]), the pandemic diffusion is modeled as a Poisson distribution with time-varying intensity \(I_t(\theta, Z)\):
\[
p_3(Z_t|Z_{1:t-1}, R_t, O_t) := \frac{(I_t(\theta, Z))^{Z_t}}{Z_t!} \exp(-I_t(\theta, Z)).
\]

Such a model claims that the mean value of the counts \(Z_t\) at time \(t\) is \(\sum_{u=1}^{\tau_6} R_u Z_{t-u} + O_t\). When \(R_t \geq 0\) and \(I_t(\theta, Z) = 0\), \(p_3\) is the Dirac mass at zero; it is equal to (7) when \(Z_t = 0\) and equal to 0 otherwise. Finally, when \(R_t < 0\) or \(I_t(\theta, Z) < 0\), \(p_3\) is a distribution on the negative integers; this implies that when \(Z \in \mathbb{N}^T\), \(p_3(Z_t|Z_{t-\tau_6:t-1}, R_t, O_t) = 0\).

This description implies that the likelihood of a vector of observations \(Z\) taking values in \(\mathbb{N}^T\) is \(\exp(-F(\theta))\).

Posterior distribution. This model implies that \(\pi_Z(\theta)\) is the a posteriori distribution of \((R_{3:T}, O)\) given \(Z\) when \(Z \in \mathbb{N}^T\).

D. Bayesian Estimators

In the Bayesian approach to Decision Theory, the maximum, the median, and the expectation of the a posteriori distribution, are Bayes estimators \(\hat{\theta}\) associated to a loss function \(\ell\)
\[
\hat{\theta}(Z) \in \text{Argmin}_{\theta \in \mathcal{D}} \int_{\mathcal{D}} \ell(\tau, \theta) \pi_Z(\theta) d\theta;
\]
\(\ell\) is, respectively, the 0 – 1 loss, the \(L^1\)-norm and the squared \(L^2\)-norm (see e.g. [17, Sec. 2.3 and 2.5]). Computing the Maximum a Posteriori (MAP) fits the minimization problem proposed in [11] for the reconstruction of \(\theta = (R, O)\):
\[
\min_{\theta} \sum_{t=1}^{T} d_{dl}(Z_t|I_t(\theta, Z)) + \lambda_R \|DR\|_1 + \lambda_O \|O\|_1. \tag{8}
\]

The optimization problem is a nonsmooth convex minimization problem both encapsulating the transmission process, favoring piecewise linear behavior of the reproduction number along time and sparsity of the outliers. The minimization is performed with the Chambolle-Pock primal-dual algorithm allowing to handle both the non-differentiability and linear operators [18, 19].

Properties of the MAP are established in Proposition 1.

Proposition 1: If there are at least two positive averaged counts \(\sum_{u=1}^{\tau_6} \Phi_u Z_{t-u} + \sum_{u=1}^{\tau_6} \Phi_u Z_{t-u}^{-},\) and one positive count \(Z_t,\) a MAP exists. If \(\theta^* = (R^*, O^*)\) are two MAP, then \(I_t(\theta^*, Z) = I_t(\theta^*, Z),\) \(O_t^* \geq 0\) and \(|D^R|\) for all \(t \in \{1, \ldots, T\}\).

Proof: The first statement is adapted from [11]; The second statement is established in [11]. The sign conditions result from a first order expansion of the \(L^1\)-norm. For a detailed proof, see Section VIII in the Supplementary material.

Proposition 1 implies that the MAP is either unique, or that there are uncountably many MAP. In addition, it shows that \(f_Z\) and \(g\) are constant over the set of the minimizers. Thus, following the same lines as in [20], a sufficient condition for the uniqueness of the MAP is derived (see Section VIII in the Supplementary material).

The expression of the distribution \(\pi_Z(\theta)\) is so complex that it is known only up to a normalizing constant. Consequently, the computation of most statistics of \(\pi_Z(\cdot)\) relies on Monte Carlo samplers, in order to produce samples \(\{\theta^n, n \geq 0\}\) in \(\mathcal{D}_Z\) approximating \(\pi_Z\) (see e.g. [21, Sec. 2.3]): for example, the estimation of the median and more generally the quantiles of \(\pi_Z\) can rely on the order statistics of the samples, and the mean a posteriori can be approximated by the Monte Carlo sum \(N^{-1} \sum_{n=1}^{N} \theta^n\).

III. Blockwise Proximal-Gradient Monte Carlo Samplers

The aim is now to devise Monte Carlo sampling strategies for the distribution defined in (1). However, this section will address a broader class of densities, defined on \(\mathbb{R}^d\) with respect to the Lebesgue measure, and expressed as \(\pi(\theta) \propto \exp(-F(\theta))\) where \(F := f + g\) and \(f, g, D\) satisfy the smoothness and blockwise structure assumptions A1 and A2 defined below.

A. Smoothness and Blockwise Structure

A1: \(f\) and \(g\) are finite on \(D \subseteq \mathbb{R}^d\) and \(f\) is continuously differentiable on the interior of \(D\).

Additionally, \(g\) has a blockwise structure that we aim to use in the design of the proposed samplers. This blockwise structure stems both from the decomposition of \(\theta\) into \(J\) blocks
\((\theta_1, \ldots, \theta_J) \in \mathbb{R}^{d_1} \times \ldots \times \mathbb{R}^{d_J}\) and from the sum of several functions of \(\theta_j\) possibly combined with a linear operator. A2: For \(j \in \{1, \ldots, J\}\), there exist matrices \(A_{i,j} \in \mathbb{R}^{c_{i,j} \times d_i}\), and proper, convex, lower semi-continuous functions \(g_{i,j} : \mathbb{R}^{c_{i,j}} \to ]-\infty, +\infty[\) such that \(\sum_{j=1}^J d_j = d\) and 

\[
\forall \theta := (\theta_1^T, \ldots, \theta_J^T)^T, \quad g(\theta) := \sum_{j=1}^J \sum_{i=1}^{I_j} g_{i,j}(A_{i,j} \theta_j).
\]

In addition, the proximity operator of \(g_{i,j}\) has a closed form expression.

In Bayesian inverse problems, \(f\) may stand for the data fidelity term and the nonsmooth part \(g\) stands for penalty terms acting on blocks of the parameter \(\theta\). Different splittings of the prior defined by (2) fits Assumption A2.

Example 2: The prior \(g\) given by (2) satisfies A2:

\[
\lambda_R \|[\text{DR}]_1 + \lambda_0 \|[O]_1 = g_{1,1}(A_{1,1} R) + g_{1,2}(A_{1,2} O),
\]

where \(\theta_1 := R, \theta_2 := O, A_{1,1} := D, A_{1,2} := I_T, g_{1,1} := \lambda_R \|[ \cdot \|_1\) and \(g_{1,2} := \lambda_0 \|[ \cdot \|_1\).

Example 3: The prior \(g\) given by (2) satisfies A2:

\[
\lambda_R \|[\text{DR}]_1 + \lambda_0 \|[O]_1 = \sum_{i=1}^3 g_{i,1}(A_{i,1} R) + g_{i,2}(A_{i,2} O),
\]

where \(\theta_1 := R\) and \(\theta_2 := O, A_{1,1}\) collects the rows \(i, i + 3, i + 6, \ldots\) of the matrix \(D, A_{1,2} := I_T, g_{i,1} := \lambda_R \|[ \cdot \|_1\) and \(g_{i,2} := \lambda_0 \|[ \cdot \|_1\).

The second example follows block splitting strategies described in [22, 23].

B. Proximal Algorithms and Metropolis-Hastings Algorithms

The design of an optimization strategy to minimize \(F\) on \(D\) and the design of a sampler to approximate the target distribution \(\pi\) both rely on the activation of an operator \(\mu : \mathbb{R}^d \to \mathbb{R}^d\). To be more specific, when minimizing \(F\), we aim to design a sequence of the form:

\[
\theta^{n+1} = \mu(\theta^n),
\]

where \(\mu\) is an operator built from \(F\) and \(D\) in such a way that the sequence \((\theta^n)_{n \in \mathbb{N}}\) converges to a minimizer of \(F\) (cf. [24], [25], [26] for an exhaustive list of algorithmic schemes). When building a Metropolis-Hastings algorithm (say with Gaussian proposal), a new point is proposed as:

\[
\theta^{n+1/2} = \mu(\theta^n) + \xi^{n+1} \quad \xi^{n+1} \sim N_{{\mathbb{R}^d}}(0_d, \mathbf{C}),
\]

\(\mathbf{C} \in \mathbb{R}^{d \times d}\) is a positive definite matrix. The Langevin dynamics is recovered in the specific case where \(F\) is smooth with \(\mu(\theta) := \theta - \gamma \nabla F(\theta)\) being a gradient ascent over \(\ln \pi\) with step size \(\gamma > 0\) and \(\mathbf{C} := 2\gamma I_d\). Scaled Langevin samplers are also popular: given a \(d \times d\) matrix \(\Gamma\), set

\[
\mu(\theta) := \theta - \gamma \Gamma \nabla F(\theta), \quad \mathbf{C} := 2\gamma \Gamma \Gamma^T;
\]

they are inherited from the so-called tempered Langevin diffusions [27] (see also [28] for a pioneering work on its use in the Markov Chain Monte Carlo literature). Either this proposed point is the new point \(\theta^{n+1} = \theta^{n+1/2}\) (thus yielding the Langevin Monte Carlo algorithm [29]; see also [30, 31]) or there is an acceptance-rejection Metropolis mechanism (thus yielding the Metropolis Adjusted Langevin Algorithm (MALA) [32]). The general Metropolis-Hastings procedure with Gaussian proposal, is summarized in Algorithm 1 where we denote by \(q(\theta, \gamma)\) the density of the distribution \(N_{{\mathbb{R}^d}}(\mu(\theta), \mathbf{C})\) evaluated at \(\tau \in \mathbb{R}^d\):

\[
q(\theta, \tau) := \frac{\exp \left\{-0.5(\tau - \mu(\theta))^T \mathbf{C}^{-1}(\tau - \mu(\theta))\right\}}{\sqrt{2\pi} \sqrt{\det(C)}}.
\]

The constraint \(\theta \in D\) is managed by the acceptance-rejection step (since \(\pi(\theta^{n+1/2}) = 0\) when \(\theta^{n+1/2} \notin D\) but not necessarily in the proposal mechanism.

The MALA algorithms drift the proposed moves towards areas of high probability for the distribution \(\pi\), using first order information on \(\pi\). Building on this idea, many strategies were proposed in the literature in the setting defined by A1: \(\mu\) can either be a gradient step when \(F\) is smooth, or a proximal step (i.e., \(\mu = \text{prox}_{\gamma F}\) also referred as an implicit subgradient descent step), or a Moreau-Yosida envelope gradient step (i.e., \(\mu = I_d - \gamma \nabla^2 F\)) where the Moreau envelope of a function \(F\) with parameter \(\gamma > 0\) is defined as \(\tilde{F} := \inf_{y} \gamma F(y) + \frac{1}{2\gamma} \|\cdot - y\|^2\).

In [33], explicit subgradient steps possibly combined with a proximal step are used. In [34], \(\mu\) relies on a Gaussian smoothing of convex functions with Hölder-continuous sub-gradients; this method applies under regularity conditions and convexity assumptions on \(g\) which are not implied by A1–A2. In [35], the authors add a Moreau-Yosida envelope term and a gradient term. In [36], \(\mu\) composes a Moreau-Yosida envelope of \(g(A_1)\) and a gradient step. Let us cite [37], [38], [39] who also use proximal operators in order to define trans-dimensional Monte Carlo samplers – an objective which is out of the scope defined by A1. See also [40] and [41, Sec. 3] for a survey on MCMC samplers using optimization techniques, and [42, Sec. 4] for optimization techniques combined with Multiple Try Metropolis strategies.
However, in the optimization context (9) with a non-smooth objective function $F$, explicit sub-gradient method is suboptimal with rate $O(1/\sqrt{n})$ compared to proximal-based strategy such as forward-backward (relying on an implicit sub-gradient method) whose rate is $O(1/n)$ [43]. Proximal-based algorithms are thus especially adapted to handle non-smooth functions but also smooth functions that do not have a Lipschitz gradient. The only limitation of using proximal-based algorithms relies on the possible difficulty to derive a closed-form expression of the proximity operator. However, for most common functions encountered in signal processing, closed-forms exist and have been summarized on the website ProxRepository http://proximity-operator.net/scalarfunctions.html - see also Lemma 4 below, for standard functions composed with linear operators. When $F = f + g$ where $f$ and $g$ satisfy Assumption A1, deriving a proximal activation $\mu$ is often a tedious task as no closed form expression for $\text{prox}_F$ exists in a general framework [44], [45].

The standard solution consists in a proximal-gradient activation: $\mu(\theta) := \text{prox}_g(\theta - \gamma \nabla f(\theta))$. When dealing with a blockwise structure for $g$ as in A2, the choice of $\mu$ has to manage both the additive structure of $g$ and the combination of the $g_{i,j}$'s with a linear operator. Unfortunately, the proximity operator has a closed form expression in very limited cases recalled below in Lemma 4.

**Lemma 4.** Let $A \in \mathbb{R}^{c \times d}$.
1. Let $g := \frac{1}{2} \|z\|^2$ with $z \in \mathbb{R}^c$. For every $\gamma > 0$, $\text{prox}_{\gamma g}(A) = \gamma(A^TA + I_d)^{-1}(\gamma(A^TA z) + I_d)$.
2. Let $g$ be a proper lower semi-continuous convex function. Let $A^A$ be invertible. For every $\gamma > 0$, $\text{prox}_{\gamma g}(A) = I_d - A^TA^{-1}(I_d - \text{prox}_{\gamma(A^A)^{-1}})A$ where, for every $x \in \mathbb{R}^d$, $\text{prox}_{\gamma(A^A)^{-1}}(x) = \text{argmin}_{y \in \mathbb{R}^d} g(y) + \frac{1}{2\gamma}(x - y)^T(A^A)^{-1}(x - y)$.
3. Let $g(A) := \sum_{k=1}^c g_k(A_{\zeta_k})$ where $g_k$ is convex, lower semi-continuous, and proper fom $\mathbb{R}^d \to ]-\infty, +\infty[$ and $A_{\zeta_k} \in \mathbb{R}^d$ denotes the row $\# k$ of $A$. Suppose that $A^A = A$, where $A := \text{diag}(\chi_1, ..., \chi_c)$ and $\chi_k > 0$. Then, for every $\gamma > 0$: for all $\eta \in \mathbb{R}^d$

$$\text{prox}_{\gamma g}(A)(\eta) = \eta - A^T(A^T)^{-1}(A^T \eta - \text{prox}_{\gamma(A^A)^{-1}}(A^T \eta))$$

where for all $\zeta = \zeta_1 : c$, we set $A g(\zeta) := \sum_{k=1}^n \chi_k g_k(\zeta_k)$.

**Proof:** 1) See [25]. 2) This result is a direct consequence of [18, Proposition 23.25 (ii)] and [18, Proposition 23.343-(ii)-(iii)]. 3) Result extracted from [22] and a direct consequence of 2) for specific choices of $A$ and $g$.

**C. Novel Blockwise Proximal Monte Carlo Samplers: PGdec and PGdual**

None of the algorithms recalled in the previous section directly apply to the context of Assumptions A1 to A2. We propose two novel Metropolis-Hastings algorithms: the PGdec sampler and the PGdual sampler, which use the proximal operator to handle the nonsmooth part of $g$ of $-\ln \pi$, its additive structure and its combination with linear operators. Such strategies can be used as building blocks for more general MCMC samplers such as Multiple Try based samplers (see e.g. [46] and [47]) and more generally, adaptive Monte Carlo samplers (see e.g. [48]). An extension to Gibbs sampler is proposed and discussed in Section III-F.

- The PGdec sampler. Additionally to Assumptions A1 and A2, we further assume:

**A3:** Each function $g_{i,j}(A_{i,j})$ possesses a proximal operator having a closed form expression.

**A3** assumes that each component $g_{i,j}(A_{i,j})$ has a tractable proximal operator which does not imply that the sum $g = \sum_{i=1}^n g_{i,j}$ admits a tractable proximal operator.

The Proximal-Gradient Decomposition sampler (PGdec) is described by Algorithm 2. It is a Metropolis-Hastings sampler with Gaussian proposal: conditionally to the current point $\theta^n$, PGdec proposes a move to $\theta^{n+1/2} = (\theta^{n+1/2}_1, ..., \theta^{n+1/2}_J)$ sampling independently the $J$ blocks from Gaussian distributions (see $\theta^{n+1/2}$ in line 5 of Algorithm 2); Then an acceptance-rejection step is applied (see (15)). The originality of our method is the definition of $\mu$: for every $j \in \{1, ..., J\}$ and $i \in \{1, ..., I_j\}$,

$$\mu_{i,j}^{\text{PGdec}}(\theta) := \text{prox}_{\gamma_{i,j} g_{i,j}(A_{i,j})}(\theta - \gamma_{i,j} \nabla_j f(\theta)), \quad (13)$$

where $\gamma_{i,j}$ is a positive step size and $\nabla_j$ denotes the differential operator w.r.t. the block $\# j$ of $\theta$. The proposed drift takes benefit of the blockwise separable expression of $g$ and computes at each iteration the proximal operator associated to part of the sum in order to perform the proximal activation with a closed form expression. Conditionally to $\theta^n$, for each block $\# j$, one of the component $g_{i,j}$ is selected at random in $\{1, ..., I_j\}$ (see line 3); then, $\theta^{n+1/2}$ is sampled from a $\mathbb{R}^{d_j}$-valued Gaussian distribution with expectation $\mu_{i,j}^{\text{PGdec}}(\theta^n)$ and covariance matrix $C_{i,j}$. We denote by $q_{i,j}(\theta, \tau)$ the density of the distribution $N_{\mu_{i,j}^{\text{PGdec}}(\theta), C_{i,j}}$ evaluated at $\tau \in \mathbb{R}^{d_j}$:

$$q_{i,j}(\theta, \tau) := \exp \left( \frac{-0.5 (\tau - \mu_{i,j}^{\text{PGdec}}(\theta))^T C_{i,j}^{-1} (\tau - \mu_{i,j}^{\text{PGdec}}(\theta))}{\sqrt{2\pi}^\frac{d_j}{2} \sqrt{\det(C_{i,j})}} \right).$$

**Remark:** Let $j \in \{1, ..., J\}$ and $i \in \{1, ..., I_j\}$. Given $\theta^n = (\theta^{n-1}_j, \theta^n_{J+1:j}) \in \mathbb{R}^{d_j}$, $\mu_{i,j}^{\text{PGdec}}(\theta^n)$ successively computes a gradient step w.r.t. the smooth function $f$ and the variable $\theta_j$, and a proximal step with respect to the function $g_{i,j}(A_{i,j})$; the step size is $\gamma_{i,j}$ for both steps. Hence, $\mu_{i,j}^{\text{PGdec}}(\theta^n) = f(\theta^n_{J+1:j}) + g_{i,j}(A_{i,j})$.

**Example 6 (Example 2 to follow):** The proximal operator of $R \mapsto \lambda_R \|DR\|_1$ is not explicit, so that, when decomposing $g$ as proposed in Example 2, PGdec can not be applied to approximate $\pi_Z$.

**Example 7 (Example 3 to follow):** For all $i \in \{1, 2, 3\}$, $A_{i,i} A_{i,i}$ is the identity matrix so that, by Lemma 4, $\text{prox}_{\gamma_{i,i} A_{i,i}} \|A_{i,i}\|_1$ is explicit and given by $(I_T - A_{i,i}^TA_{i,i}) + A_{i,i}^T \text{prox}_{\gamma_{i,i} A_{i,i}} 1_{\|A_{i,i}\|_1}$, and in addition, $\text{prox}_{\gamma_{i,i} A_{i,i}} \|A_{i,i}\|_1$ has a closed
Algorithm 2: Blockwise Metropolis-Hastings samplers.

**Data:** $d_i \times d_i$ positive definite matrices $\mathbf{C}_{i,j}$; $\gamma_j > 0$; a positive integer $N_{max}$; $\theta^0 \in \mathcal{D}$

**Result:** A $\mathcal{D}$-valued sequence

$$\{\theta^n, n \in \{0, \ldots, N_{max}\}\}$$

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 $\xi_j^{n+1} \sim N(0, \theta_j^{n+1})$;

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

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

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

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

form expression. Hence, when decomposing $g$ as proposed in Example 3, PGdual can be applied to approximate $\pi_Z$.

- **The PGdual sampler**. The Proximal-Gradient dual sampler (PGdual) is defined along the same lines as Algorithm 2. It is designed for situations when for any $i, j$, the dimensions of the matrices $\mathbf{A}_{i,j}$ satisfy $c_{i,j} \leq d_i$ and $\mathbf{A}_{i,j}$ can be augmented in an invertible $d_j \times d_j$ matrix $\bar{\mathbf{A}}_{i,j}$.

For every $j \in \{1, \ldots, J\}$ and $i \in \{1, \ldots, I_j\}$, let $\bar{\mathbf{A}}_{i,j}$ be a $d_j \times d_j$ invertible matrix such that for any $\theta_j \in \mathbb{R}^{d_j}$, $(\mathbf{A}_{i,j}\theta_j)_{d_j-c_{i,j}+d_j} = \mathbf{A}_{i,j}\theta_j$. For $x \in \{x_1, \ldots, x_{d_j}\} \in \mathbb{R}^{d_j}$, define $\bar{g}_{i,j}(x) := \bar{g}_{i,j}(x_{d_j-c_{i,j}+1:d_j})$. PGdual uses the drift functions

$$\mu_{i,j}^{PGdual}(\theta) := \bar{\mathbf{A}}_{i,j}^{-1} \text{prox}_{\gamma_j\theta_j} \left(\mathbf{A}_{i,j}\theta_j - \gamma_j\bar{\mathbf{A}}_{i,j}^{-1}\nabla f(\theta)\right).$$

**Remark 8:** For every $j \in \{1, \ldots, J\}$, select $i_j \in \{1, \ldots, I_j\}$, and consider the partial objective function:

$$\theta \mapsto f(\theta_1, \ldots, \theta_J) + \sum_j \bar{g}_{i_j,j}(\bar{\mathbf{A}}_{i_j,j}\theta_j).$$

(17)

Consider the one-to-one maps $\tilde{\theta}_j := \bar{\mathbf{A}}_{i_j,j}\theta_j$ for any $j$, and the application

$$\tilde{\theta} := \tilde{\theta}_{1:j} \mapsto f(\bar{\mathbf{A}}_{i_1:1,j}, \ldots, \bar{\mathbf{A}}_{i_J,j,1,j}\tilde{\theta}_j) + \sum_j \bar{g}_{i_j,j}(\tilde{\theta}_j).$$

(18)

The PG step w.r.t. $\tilde{\theta}_j$ reads:

$$\text{prox}_{\gamma_j\tilde{g}_{i,j}} \left(\tilde{\theta}_j - \gamma_j\bar{\mathbf{A}}_{i_j,j}^{-1}\nabla f(\bar{\mathbf{A}}_{i_1:1,j}, \ldots, \bar{\mathbf{A}}_{i_{j-1},j,1:j}\tilde{\theta}_j)\right)$$

$$= \text{prox}_{\gamma_j\tilde{g}_{i,j}} \left(\bar{\mathbf{A}}_{i_j,j}\tilde{\theta}_j - \gamma_j\bar{\mathbf{A}}_{i_j,j}^{-1}\nabla f(\theta)\right).$$

(19)

Therefore, applying a PG step w.r.t. the variable $\tilde{\theta}_j$ in this “dual” space, and going back to the original space by applying $\bar{\mathbf{A}}_{i,j}^{-1}$ leads to the drift $\mu_{i,j}^{PGdual}$ in (16).

From (16) and Lemma 4 Item 2, we have

$$\mu_{i,j}^{PGdual}(\theta) := \text{prox}_{\gamma_j\bar{g}_{i,j}} \left(\bar{\mathbf{A}}_{i,j}\theta_j - \gamma_j\bar{\mathbf{A}}_{i,j}^{-1}\nabla f(\theta)\right)$$

(20)

thus showing that it is a Variable Metric Proximal-Gradient step in the metric induced by $\bar{\mathbf{A}}_{i,j}$, for the minimization of $\theta_j \mapsto f(\theta) + \bar{g}_{i,j}(\bar{\mathbf{A}}_{i,j}\theta_j)$. See [49], [50]; see also [51, Sec. 2] in the specific case $J = 1, J = 1$.

**Example 9 (Example 2 to follow)**: Denote by $\bar{\mathbf{D}}_T$ the $T \times T$ invertible matrix obtained by augmenting $\mathbf{D}$ with two vectors in $\mathbb{R}^T$ as follows:

$$\bar{\mathbf{D}}_T := \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ 0 & \ldots & 0 & 1 & 0 \\ 0 & \ldots & 0 & 0 & 1 \end{bmatrix}.$$
Theorem 10 makes the PGdual idea explicit by proposing a matrix $\hat{A}_{i,j}$ augmenting $A_{i,j}$, computing the associated drift $\mu_{i,j}^{\text{PGdual}}$ and comparing it to $\mu_{i,j}^{\text{PGdec}}$.

**Theorem 10:** Assume A1 and A2. Let $j \in \{1, \ldots, J\}$ and $i \in \{1, \ldots, I_j\}$. Assume that $c_{i,j} < d_j$ and $A_{i,j} A_{i,j}^\top$ is invertible.

Let $U_{i,j}$ be a $(d_j - c_{i,j}) \times d_j$ matrix such that $U_{i,j} A_{i,j} = 0_{(d_j - c_{i,j}) \times c_{i,j}}$ and $U_{i,j} U_{i,j}^\top$ is invertible.

Then, the matrix $\hat{A}_{i,j} := [U_{i,j}; A_{i,j}] \in \mathbb{R}^{d_j \times d_j}$ is invertible. For any $\theta \in \mathcal{D}$, $\mu_{i,j}^{\text{PGdual}}(\theta)$ given by (16) is equal to

$$
\begin{align*}
\hat{A}_{i,j}^\top (A_{i,j} \hat{A}_{i,j}^{-1})^{-1} \text{prox}_{\gamma_i \omega_{i,j}} \left( A_{i,j} \left( \theta_j - \gamma_i \Omega_{i,j} \nabla_j f(\theta) \right) \right) \\
+ (I_{d_j} - \Pi_{i,j}) \left( \theta_j - \gamma_i \Omega_{i,j} \nabla_j f(\theta) \right),
\end{align*}
$$

where

$$
\Omega_{i,j} := U_{i,j}^\top (U_{i,j} U_{i,j}^\top)^{-2} U_{i,j}, \quad \hat{\Omega}_{i,j} := A_{i,j}^\top (A_{i,j} A_{i,j}^\top)^{-2} A_{i,j}.
$$

Additionally, when $U_{i,j} U_{i,j}^\top = I_{d_j - c_{i,j}}$, $A_{i,j} A_{i,j}^\top = I_{c_{i,j}}$, and under A3 then $\mu_{i,j}^{\text{PGdual}} = \mu_{i,j}^{\text{PGdec}}$ where $\mu_{i,j}^{\text{PGdec}}$ is given by (22).

**Proof:** The main ingredients are the equalities

$$
\hat{A}_{i,j}^{-1} = \left[ U_{i,j}^\top (U_{i,j} U_{i,j}^\top)^{-1} A_{i,j} A_{i,j}^\top \right]^{-1}.
$$

and $U_{i,j}^\top (U_{i,j} U_{i,j}^\top)^{-1} U_{i,j} + \Pi_{i,j} = I_{d_j}$. The proof follows from standard matrix algebra. A detailed proof is given in Section IX of the Supplementary material.

The result remains true when $c_{i,j} = d_j$; in that case, $\hat{A}_{i,j} = A_{i,j}$, $I_{d_j} - \Pi_{i,j} = 0_{d_j \times d_j}$ and $\hat{\Omega}_{i,j} = A_{i,j}^{-1} A_{i,j}^\top$. Theorem 10 provides sufficient conditions for the drift function $\mu_{i,j}^{\text{PGdec}}$ and the drift function $\mu_{i,j}^{\text{PGdual}}$ to be equal. Observe that the conditions on $U_{i,j}$ are satisfied as soon as the rows of $U_{i,j}$ are orthonormal and orthogonal to the rows of $A_{i,j}$.

Let us derive two strategies for the application of PGdual to sample the target density defined by (1).

**Example 11 (Examples 2 and 9, to follow):** A first strategy is to decompose $g$ as in Example 2; Example 9 provides a possible augmentation of $D$. Another one is proposed by Theorem 10:

$$
\mathbf{D}_0 := \begin{bmatrix}
U_{1,1} & D
\end{bmatrix} \in \mathbb{R}^{T \times T}
$$

where $U_{1,1}$ is obtained by making orthogonal the first two rows of $\mathbf{D}_i$ (see (21)) and making them orthogonal to the rows of $D$. With this choice, $\mathbf{D}_0$ has a condition number one order of magnitude smaller compared to $\mathbf{D}_1$, which is likely to impact the convergence of Markov chains (see Section VI).

This strategy acts globally on $\ln \pi_Z$; note, at each iteration, a PG approach for the function $f_Z + \lambda_R [D \cdot ||1 + \lambda_0|| \cdot 1]$. It will be numerically explored in Section V.

**Example 12 (Example 3 to follow):** A second strategy is to decompose $g$ as in Example 3, and define the matrices $\hat{A}_{i,j}$ as described in Theorem 10. This strategy defines $\mu_{i,j}^{\text{PGdual}}$ by considering part of $\ln \pi_Z$: at each iteration, having selected $i_1 \in \{1,2,3\}$, it proposes a PG approach for the function $f_Z + \lambda_R [A_{i_1,1} \cdot ||1 + \lambda_0|| \cdot 1]$ (see Remark 8).

**E. Convergence Analysis of PGdec and PGdual**

We prove that both PGdec and PGdual produce a sequence of points $\{\theta^n, n \geq 0\}$ which is an ergodic Markov chain having $\pi$ as its unique invariant distribution.

**Proposition 13:** Assume A1, A2 and A3. Assume also that $\pi$ is continuous on $D$. Then the sequence $\{\theta^n, n \geq 0\}$ given by Algorithm 2 applied with $\mu = \mu_{i,j}^{\text{PGdec}}$ is a Markov chain, taking values in $D$ and with unique invariant distribution $\pi$. In addition, for any initial point $\theta^0 \in D$ and any measurable function $h$ such that $\int h(\theta) \pi(\theta) d\theta < \infty$,

$$
\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} h(\theta^n) = \int h(\theta) \pi(d\theta),
$$

with probability one.

The same result holds when $\mu = \mu_{i,j}^{\text{PGdual}}$.

**Proof:** We have $\theta^{n+1} \in D$ since $\pi(\theta^{n+1/2}) = 0$ when $\theta^{n+1/2} \notin D$. The other properties result from [52, Lemmas 1.1 and 1.2] and [53, Propositions 10.1.1 and 10.4.4 and Th. 17.0.1]. The Harris recurrence property (required for the law of large numbers to hold for any $\theta^0$) can be proved along the same lines as [54, Th. 6(v)]. See the detailed proof in Section IX of the Supplementary material.

**Remark 14:** A slight adaption of the proof shows that when selecting the indices $i_j$, the uniform distribution on $\{1, \ldots, I_j\}$ can be replaced with a probability distribution $\{p_{i,j}(\theta^n), i = 1, \ldots, I_j\}$ depending on the current value $\theta^n$, up to a modification of the acceptance-rejection ratio (15) (see Section IX of the Supplementary material).

**F. Gibbs PGdec and Gibbs PGdual Samplers**

The additive structure of $g$ in A2, naturally suggests the extension of PGdec and PGdual to the Gibbs sampler algorithm [55], [56]. Since whatever $\theta = \theta_{1,J} \in \mathbb{R}^d$, exact sampling from the conditional distributions on $\mathbb{R}^{d_j}$

$$
\tau \mapsto \pi_j(\tau | \theta) \propto \exp(-f(\theta_{1:j-1}, \tau, \theta_{j+1:J})
$$

for $j = 1, \ldots, J$, is not always explicit, we propose a Metropolis-within-Gibbs strategy; see [57]. The pseudo-code of the so-called Gibbs Blockwise Proximal sampler is given by Algorithm 3 in the case of a systematic scan order of the $J$ components (our method easily extends to other scan orders; details are left to the reader): at each iteration $\#(n+1)$, and for each block $\#(i, i)$ sample at random $i_j \in \{1, \ldots, I_j\}$, (ii) sample a candidate from $\mathcal{N}_{d_j}(\mu_{i,j}(\theta), C_{i,j})$, where $\theta := (\theta_{1:j-1}, \theta_{j+1:j}) \in \mathbb{R}^{d}$ is the current value of the chain, and (iii) accept/reject this candidate via a Metropolis step targeting the distribution $\pi_j(\cdot | \theta)$.

The Gibbs PGdec and the Gibbs PGdual samplers correspond to Algorithm 3 applied resp. with $\mu_{i,j} = \mu_{i,j}^{\text{PGdec}}$ and $\mu_{i,j} = \mu_{i,j}^{\text{PGdual}}$. 


Algorithm 3: Gibbs Blockwise Proximal sampler.

**Data:** \( d_j \times d_j \) positive definite matrices \( C_{i,j} \); \( \gamma_j > 0 \);

**Result:** A \( \mathcal{D} \)-valued sequence \( \{\theta^n, n \in \{0, \ldots, N_{\text{max}}\}\} \)

1. Set \( \vartheta = \theta^0 \);
2. for \( n = 0, \ldots, N_{\text{max}} - 1 \) do
3. for \( j = 1, \ldots, J \) do
4. Sample \( i_j \in \{1, \ldots, I_j\} \) with probability \( 1/I_j \);
5. Sample \( \xi_{j}^{n+\frac{1}{2}} \sim N(d_j, C_{i,j}) \);
6. Set \( \theta_j^{n+\frac{1}{2}} = \mu_{i_j,j}(\vartheta) + \xi_{j}^{n+\frac{1}{2}} \);
7. Set \( \theta_j^{n+1} = \theta_j^{n+\frac{1}{2}} \) with probability
8. \( \frac{\pi_j(\theta_j^{n+\frac{1}{2}} | \vartheta) q_{i_j,j}((\theta_{1,j-1}^{n+\frac{1}{2}}, \theta_{j+1,j}^{n}), \theta_j^{n})}{\pi_j(\theta_j^{n+\frac{1}{2}}, q_{i_j,j}(\vartheta, \theta_j^{n+\frac{1}{2}}))} \)
9. and \( \theta_j^{n+1} = \theta_j^{n} \) otherwise ;
10. Update \( \vartheta = (\theta_1^{n+1}, \theta_J^{n+1}) \)

IV. COVID-19 Data

To illustrate, assess, and compare the relevance and performance of the Monte Carlo procedures for credibility interval estimation proposed here, use is made of real Covid-19 data made available by the Johns Hopkins University. The repository collects, impressively since the outbreak of the pandemic, daily new infections and new death counts from the National Public Health Authorities of 200+ countries and territories of the world. Counts are updated on a daily basis since the early stage of the pandemic (Jan. 1st, 2020) until today. This repository thus provides researchers with a remarkable dataset to analyze the pandemic.

As mentioned in the Introduction section, because of the sanitary crisis context, data made available by most National Public Health Authorities in the world are of very limited quality as they are corrupted by outliers and missing or negative counts. Data quality also varies a lot across countries or even within a given country depending on the phases and periods of the pandemic.

The present work uses the new infection counts only, as the estimation of the space-time evolution of the pandemic reproduction number \( R \) is targeted. A mild and non-informative preprocessing is applied to data by replacing negative counts by a null value.

V. MONTE CARLO SAMPLER ASSESSMENTS

This section aims to assess and compare the performance for several variations of the samplers \( \text{PGdec} \) and \( \text{PGdual} \) introduced in Section III, using the real Covid19 data described in Section IV. The target distribution is \( \pi_Z \), given by (1).

Fig. 1. Displayed: the normalized distance to \( \hat{R}_{\text{MAP}} \) versus the iteration index, during the burn-in phase (row 1) and after the burn-in phase (row 2); the ACF criterion versus the lag (row 3) and the GR criterion versus the iteration index (row 4); for the \( \text{RW} \)-based samplers (column 1), \( \text{PGdec} \)-based samplers (column 2) and \( \text{PGdual} \)-based samplers (column 3). For each Algo in \{ \text{RW}, \text{PGdec}, \text{PGdual} \}: Algo I is in cyan, Algo O is in red, Gibbs Algo I is in blue and Gibbs Algo O is in black.

A. Assessment Set-Up

Performance assessments were conducted on data from several countries and for different periods of interests. For space reasons, they are reported only for the United Kingdom and for a recent time period (Dec. 6th, 2021 to Jan. 9th, 2022), with corresponding \( Z \) in Fig. 2[top right plot]. Following [11], we set \( \lambda_0 = 0.05 \) and \( \lambda_R = 3.5 \times \sqrt{5} \sigma_Z / 4 \) with \( \sigma_Z \) the standard deviation of \( Z \). Samplers are run with \( N_{\text{max}} = 1e7 \) iterations, including a burn-in phase of \( 3e6 \) iterations. Except for the plots in Fig. 1[first row], reported performances are computed by discarding the points produced during the burn-in phase. The initial values \( Z_{-\tau_0+1}, \ldots, Z_{-1}, Z_0 \) are set to the observed counts. For each algorithm, performances are computed from averages over 15 independent runs. Initial points consist of random perturbations around the non-informative point \( R^0 := (1, \ldots, 1)^\top \in \mathbb{R}^T \) and \( O^0 := (0, \ldots, 0)^\top \in \mathbb{R}^T \). For all samplers, \( \theta \) is seen as a two-block vectors \( \theta_1 = R, \theta_2 = O \); hence \( J = 2 \). For PGdec, \( g \) is decomposed as in Example 3; for PGdual, it is decomposed as in Example 2.

B. Samplers

PGdec and Gibbs PGdec. These two samplers are run by decomposing \( g \) as described in Example 3. This yields two different drift functions for the blocks \( R \) and \( O \); see Example 7. For the \( R \)-part, the drift functions are defined, for \( i = 1, 2, 3, \)
as:

\[
\mu_{1,1}^{\text{PGdec}}(\theta) := \left( I - A_{i,1}^T A_{i,1} \right) (R - \gamma_1 \nabla_1 f_Z(\theta)) \\
+ A_{i,1} \text{prox}_{\gamma_1 \lambda_0 \| \cdot \|_1} \left( A_{i,1} R - \gamma_1 A_{i,1} \nabla_1 f_Z(\theta) \right),
\]

obtained from (22) with \( A_{i,1} A_{i,1}^T \), the identity matrix. For the
\( O \)-part,

\[
\mu_{1,2}^{\text{PGdec}}(\theta) := \text{prox}_{\gamma_2 \lambda_0 \| \cdot \|_1} \left( O - \gamma_2 \nabla_2 f_Z(\theta) \right).
\]

\( \nabla_1 \) (resp. \( \nabla_2 \)) denotes the gradient operator w.r.t \( R \) (resp. \( O \)).

\( \text{PGdual and Gibbs PGdual} \). These two samplers are run by
decomposing \( g \) as described in Example 2. Consequently, for
the \( R \)-part, the drift function is given by

\[
\mu_{1,1}^{\text{PGdual}}(\theta) := D^{-1} \text{prox}_{\gamma_1 \lambda_0 \| \cdot \|_1} \left( D R - \gamma_1 D^T \nabla_1 f_Z(\theta) \right)
\]

where \( D := D_1 \) (see Example 9, the sampler is referred to as
\( \text{PGdual Inv} \) (II)) or \( D := D_3 \) (see Example 11,
the sampler is referred to as \( \text{PGdual Ortho} \) (O)); for the \( O \)-part,
for both \( \text{PGdual I} \) and \( \text{PGdual O} \),

\[
\mu_{1,2}^{\text{PGdual}}(\theta) := \text{prox}_{\gamma_2 \lambda_0 \| \cdot \|_1} \left( O - \gamma_2 \nabla_2 f_Z(\theta) \right).
\]

**RW and Gibbs RW.** For comparisons, we also run Random
Walk-based samplers (RW) with Gaussian proposals: \( \text{RW} \) and
Gibbs \( \text{RW} \) are defined respectively from Algorithm 2
and Algorithm 3, with \( I_1 = I_2 = 1 \) and \( \mu_{1,2}(\theta) = \theta \) for \( j = 1, 2 \).

**Covariance matrices** \( C_{i,j} \). For the \( R \)-part (block \( j = 1 \)),
based on the comment in Section III-C, we choose \( C_{1,1} := 2\gamma_1 D_1^{-1} D_1^T \)
for \( \text{PGdual I} \) and Gibbs \( \text{PGdual I} \); and
\( C_{1,1} := 2\gamma_1 D_0^{-1} D_0^T \) for \( \text{PGdual O} \) and Gibbs \( \text{PGdual O} \). The \( \text{PGdual} \) closely
based samplers and the RW-based samplers are run with the same
covariance matrices. For \( \text{PGdec} \), this yields
\( C_{1,1} := 2\gamma_1 D_1^{-1} D_1^T \) for \( \text{PGdual I}, \text{Gibbs PGdual I} \) and
\( C_{1,1} := 2\gamma_1 D_0^{-1} D_0^T \) for \( \text{PGdual O}, \text{Gibbs PGdual O} \) for
the three cases \( i = 1, 2, 3 \). For \( \text{RW} \), \( C_{1,1} := 2\gamma_1 D_1^{-1} D_1^T \) for
\( \text{RW I} \) and Gibbs \( \text{RW I} \); and \( C_{1,1} := 2\gamma_1 D_0^{-1} D_0^T \) for \( \text{RW O} \) and
Gibbs \( \text{RW O} \). For \( \text{PGdec} \) and \( \text{PGdual}, \gamma_1 \) is the same
step size as in (25) and (27). For the \( O \)-part (block \( j = 2 \)), we
choose \( C_{1,2} := 2\gamma_2 I_1^T \) for \( \text{PGdual O} \) and \( \text{PGdual}, \gamma_2 \) is the same
step size as in (26) and (28).

**Twelve sampling strategies.** The present section will
compare twelve different samplers, constructed from the three
drift strategies (\( \text{PGdec, PGdual, RW} \)), times two families
(\( \text{Metropolis-Hastings, Gibbs} \)), times two choices for
the covariance matrix \( C_{1,1} (I, O) \).

**Step sizes.** Different strategies are compared for the definition of
the step sizes \( (\gamma_1, \gamma_2) \). All of them consist in adapting
the step sizes during the burn-in phase in such a way that the mean
acceptance-rejection rate reaches approximately 0.25 (which is
known to be optimal for some \( \text{RW} \) Metropolis-Hastings
samplers, [58]). We observed that convergence occurs before 5e5
iterations, see Fig. 1 (row 1, columns 2 and 3); the proposed
samples are all rejected and the chain does not move from its
initial point during the first iterations, when the step size is too
large. At the end of the burn-in phase, the step sizes are frozen
and no longer adapted. For the Metropolis-Hastings samplers
\( \text{PGdec, PGdual, and RW} \), \( \gamma_1 \) is adapted and \( \gamma_2 := \gamma_1 \) (case
PGdec) of \( \gamma_2 := (\lambda_R/\lambda_0)^2 \gamma_1 \) (cases PGdual, RW). For
the Gibbs samplers, the acceptance-rejection steps are specific to
each block \( R \) and \( O \). A first consequence is that a move on
one block can be accepted while the other one is not; this
may yield larger step sizes \((\gamma_1, \gamma_2)\) which in turn favor larger
moves of the chains. A second consequence is that we use the
acceptance-rejection rate for the \( R \)-part (resp. for the \( O \)-part) to
adapt \( \gamma_1 \) (resp. \( \gamma_2 \)): the two step sizes have their own efficiency
criterion.

**C. Performance Assessment Criteria**

The sampler performances are assessed and compared using
three different criteria, see Fig. 1.

**Distance to the MAP.** We compute the normalized distance
\( \| R^n - R_{\text{MAP}} \| / \| R_{\text{MAP}} \| \) where \( R_{\text{MAP}} \) denotes the MAP estimator,
(computed as in [11]): it is displayed vs. the iteration index
\( n \), in the burn-in phase (row 1) and after the burn-in phase (row 2).
This criterion quantifies the ability of the chains to perform a
relevant exploration of the distribution: The chains have to visit
the support of \( \pi_Z \), they show better ergodicity rates when they
are able to rapidly escape from low density regions to move to
higher density regions. Paths start from a non-informative
initial point, considered as a point in a low density region. This
criterion permits to quantify a relevant behavior of the Markov
chain when (i) it drifts rapidly towards zero during the burn-in
phase (and (ii) it fluctuates in a large neighborhood of zero after
the burn-in phase.

**Autocorrelation function (ACF).** We then compute the ACF
for each of the 2T components of the vector \( \theta \) along the Markov
path, for a lag from 1 to 1e5. On row 3, we report the mean
value, over these 2T components, of the absolute value of the
ACF versus the lag. This criterion quantifies a relevant behavior
of the Markov chain when it converges rapidly to zero; it is
indeed related to the effective sample size of a Markov chain
(see [59]) and to the mixing properties of the chain (see e.g. [53,
Th. 17.4.4 and Sec. 17.4.3.]). For example, a weaker ACF means
that less iterations of the sampler are required to reach a given
estimation accuracy.

**Gelman-Rubin (GR) statistic.** Finally, we compute the GR
statistic (see [60], [61]) which quantifies a relevant behavior of
the Markov chain when it converges rapidly to one. It measures
how the sampler forgets its initial value and provides homogeneous
approximations of the target distribution \( \pi_Z \) after a given
number of iterations. On row 4, we report this statistic versus
the iteration index.

**D. Performance Comparisons**

**Covariance matrices.** Normalized Distance to MAP indices
(Fig. 1 [rows 1 & 2]) and GR indices (Fig. 1 [row 4]) show that, for
all algorithms \( \text{RW, PGdec, and PGdual} \), the strategy \( O \) is more
efficient than the strategy \( I \), as corresponding indices decay
more rapidly (to 0 and 1 respectively) for the former than for
the latters. A plausible explanation is that the smaller condition number of $\mathbf{D}_0$ compared to $\mathbf{D}_1$ is crucial to improve the mixing behavior of the Markov chains.

**Metropolis-Hastings vs. Gibbs samplers.** The evolution of the ACF criteria (Fig. 1[row 3]) shows that the Gibbs strategies are globally more efficient than the Metropolis-Hastings ones. Further, GR indices confirm better efficiency of the Markov chains obtained from Gibbs strategies.

**Drift functions.** The benefit of using functions $\mu$ miming optimization algorithms to drift the proposed points towards the higher probability regions of $\pi_Z$ is clearly quantified in Fig. 1 across all performance indices.

During the burn-in phase, the normalized distance to MAP indices decay toward 0 significantly faster for all PGdual-based algorithms, compared to RW-based ones. Also, the PGdual algorithms reach an expected plateau early in the burn-in phase, while this is barely the case at the end of the burn-in phase for RW algorithms.

After the burn-in phase (second row), the samplers using optimization-based drift functions $\mu$ show better behavior after reaching the high density regions as they permit a broader and more rapid exploration of the support of the distribution around its maximum, with large amplitude moves from $\mathbf{R}_{\text{MAP}}$, and faster returns to $\mathbf{R}_{\text{MAP}}$. This is notably clearly visible for the Gibbs PGdual strategy.

ACF indices decay more rapidly for the PGdual strategies than for the RW ones, irrespective of the choice of the covariance matrix or of the Metropolis-Hastings or Gibbs versions. Also, ACF indices for PGdual algorithms are less sensitive to the choice of the covariance matrices than the RW ones.

Finally, the GR statistic indices (row 4) clearly illustrates that Markov chains produced by PGdual samplers have better mixing properties, compared to others, showing hence sensitivity to the choice of the initial point.

**Optimal sampler.** Combined together, these observations, globally consistent with those stemming from other countries or time periods, lead to the following generic comments.

While yielding essentially equivalent performance, across time periods and countries, the choice of the covariance matrices (algorithms 0 and 1) significantly outperforms algorithms relying on Identity covariance matrices (not reported here as showing poor performances). This non-trivial construction actually stemmed from the thorough and detailed mathematical analysis conducted in Section III-C. Theorem 10 advocates to perform the augmentation of the $(T - 2) \times T$ matrix $\mathbf{D}$ into a $T \times T$ invertible matrix $\mathbf{D}$, by adding two rows which are orthogonal to the rows of $\mathbf{D}$.

The observation that Gibbs samplers show better performances may stem from the fact that they benefit from larger values of the step sizes learned on the fly by the algorithms which favor larger jumps when proposing $\theta^{n+1/2}$ from $\theta^n$ and imply better mixing properties.

Finally, the PGdual algorithms show homogeneous performances for example when varying the covariance matrix, and are thus less sensitive to parameter tuning, an important practical feature. Also, the PGdual algorithms show systematically better performances than the PGdec ones. This may result from the definition of $\mu_{\text{PGdec}}$, which, because of the block-splitting approach (see (25)), uses only partial information on $\pi_Z$ at each iteration.

As an overall conclusion, systematically observed across all studied time periods and countries, the Gibbs PGdual 0 algorithm, devised from the careful analysis of the theoretical properties of the distributions defined by A1–A2, is consistently found to be the most efficient strategy for a relevant assessment of the Covid19 pandemic time evolution.

**VI. Credibility Intervals for $R$**

**Goal and set-up.** The present section aims to illustrate the relevance of credibility interval-based estimations for the reproduction number $R$ and for the outliers $O$ from real Covid19 data. The statistical model for $(R, O)$ is described in Section II-B. Because it is of greater interest for epidemiologists, the credibility intervals are translated in credibility intervals for the denoised counts $Z^{(D)}$ by their simple subtraction to the original counts $Z$, i.e., intuitively $Z^{(D)} = Z - O$.

Credibility interval-based estimations are reported for Gibbs PGdual 0 sampling scheme only, as Section V established that it achieved the best performance amongst the twelve sampling schemes tested. Estimations are computed for a time period of five weeks ($T = 35$ days), which corresponds to a few typical pandemic time scales, of the order of 7 days, induced by the serial interval function $\Phi$, cf. Section II. This period is set to a recent phase of the pandemic (Dec. 13th, 2021 to Jan. 17th, 2022). Estimations are reported for several countries, arbitrarily chosen as representatives of the pandemic across the world, but conclusions are valid for most countries.

**Computation of the credibility intervals.** Credibility intervals are computed as follows: For each day $t \in [T_1, T_2]$ in the period of interest $[T_1, T_2]$, the chosen sampling scheme (Gibbs PGdual 0) outputs $7.10^6$ points of a Markov chain approximating the distribution of $R_t$. Quantiles of that distribution are estimated using the empirical cumulative distribution function $\hat{F}$.

For a chosen credibility level of $1 - \alpha$, the upper and lower bounds of the credibility intervals are defined by the empirical $1 - \alpha/2$ and $\alpha/2$ quantiles. For illustration purposes, $1 - \alpha$ is set here to 95%. The same procedure is applied to produce credibility intervals for the outliers $O_t$. Finally, credibility intervals for the estimated denoised new infection counts $Z^{(D)}$ are obtained by subtracting the credibility intervals for $O_t$ to the count $Z_t$. Fig. 2 reports, top plots, the daily counts of new infections (black lines) to which are superimposed the 95% credibility interval-based estimations for the denoised counts, $Z^{(D)}$, (red pipes). Fig. 2 further reports, bottom plots, the 95% credibility interval-based estimations for $R$.

**Relevance of credibility interval-based estimations.** Fig. 2, together with the examination of equivalent plots for other countries, yields the following generic conclusions.

The estimated denoised counts show far smoother evolution along time compared to the raw counts, hence providing far...
more realistic assessments of the intensity of the pandemics. Notably, for most countries, the zero or low counts, associated with week-ends or non-working days, followed by days with over-evaluated counts by compensation, are smoothed out by the outlier estimation procedure, while the values of the counts for the regular (or non corrupted) days are left unchanged. This is the direct benefit of the nonlinear filtering procedure underlying the estimation formulation in (1), as opposed to traditional denoising procedures performed by classical linear filtering (such as moving average).

For the credibility intervals for $R$, their sizes range, depending on countries and time periods, from below 1% to above 10%. A careful examination shows that the credibility interval size is mostly driven by data quality: the credibility interval size increases when outliers are detected. Further, for a given country, the size of the credibility intervals varies only mildly along time over a five-week period. Changes in size are often associated with changes in the trends of the estimates of $R$, or with the occurrence of outliers. These credibility intervals provide hence a relevant assessment not only of the intensity of the pandemics, but also of the confidence that can be granted to this assessment, by providing epidemiologists with a range of likely values of $R$, rather than a single value. This permits to compare the evolution of the pandemics across several countries on a better scientifically grounded basis.

These credibility interval-based estimations permit a double analysis of the pandemic: They permit retrospectively to evaluate the impacts of sanitary measures on the pandemic evolution. Additionally, the smooth nature of the estimation of $R$ (close to piecewise linear) performs an implicit short term forecast (or a nowcast) of the evolution of the pandemic intensity: For instance, for several countries (e.g., France, Mali, Brazil, Singapore,...) the estimate of $R$ is decreasing for the last 5 to 10 days of the studied period, predicting that daily new infections will reach a maximum of the current wave within the coming days and then will start to decrease.

These credibility interval estimates can be complemented with other estimates such as the Maximum, median or Mean a Posteriori (cf. e.g., [62], [63]).

Finally, let us emphasize that these estimates, denoised counts and credibility intervals, are obtained using a single and same set of hyperparameters $\lambda_R$ and $\lambda_O$ common to all countries.

VII. Conclusion

The proposed tools perform a relevant credibility interval-based estimation of the Covid19 reproduction number and denoised new infection counts, by combining a statistical modeling of the time evolution of the pandemic with Monte Carlo Metropolis sampling strategies. Robustness against the low quality of the Covid19 is achieved by engineering the Bayesian model to impose sparsity in the changes of a smooth time evolution for the reproduction number and in the outlier occurrences, modeling data corruption. This is obtained at the price of the non-differentiability of their a posteriori distribution, thus precluding the use of the classical Metropolis Adjusted Langevin Algorithm to produce credibility intervals. This lead us to propose several Proximal-Gradient Monte Carlo algorithms tailored to the sampling of non-differentiable distributions, that also constitute valid sampling schemes for a much broader range of applications than that of the strict Covid19 pandemic monitoring, e.g., in image processing or more generally, in any Bayesian inverse problems with several nonsmooth priors.

Estimation performances were assessed and compared on real Covid19 data, using a set of well-selected indices quantifying the efficiency of these different sampling schemes.

Finally, it was shown for several countries and for a recent five-week period that the achieved credibility interval-based estimations of both denoised new infection counts and reproduction number provide practitioners with an efficient and robust tool for the actual and practical monitoring of the Covid19 pandemic. Such estimates are updated on a daily basis on the author’s web-pages. Automated and data-driven estimations of the hyperparameters $\lambda_R$ and $\lambda_O$ are under current investigations.

In an effort toward reproducible research and open science, MATLAB codes implementing PGdec, PGdual for both Metropolis-Hastings and Gibbs versions are made publicly available at https://github.com/gfort-lab/OpSiMorE.
REFERENCES

[1] A. Flahault, “COVID-19 cacophony: Is there any orchestra conductor?,” Lancet, vol. 395, no. 10229, 2020, Art. no. 1037.

[2] J. Arino, Descending, Modelling and Forecasting the Spatial and Temporal Spread of COVID-19: A Short Review. Cham, Switzerland: Springer, 2022, pp. 25–51.

[3] O. Diekmann, J. A. P. Heesterbeeck, 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.

[4] J. Wallinga and P. Teunis, “Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures,” Amer. J. Epidemiol., vol. 160, pp. 509–516, 2004.

[5] 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.

[6] T. Obadia, R. Haneef, and P.-Y. Boelle, “The R0 package: A toolbox to estimate reproduction numbers for epidemic outbreaks,” BMC Med. Inform. Decis. Mak., vol. 12, 2012, Art. no. 147.

[7] A. Cori, N. M. Ferguson, C. Fraser, and S. Cauchemez, “A new framework and software to estimate time-varying reproduction numbers during epidemics,” Amer. J. Epidemiol., vol. 178, pp. 1505–1512, 2013.

[8] R. Thompson et al., “Improved inference of time-varying reproduction numbers during infectious disease outbreaks,” Epidemics, vol. 29, 2019, Art. no. 100356, [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1755436519300350

[9] Q.-H. Liu, M. Ajell, A. Alera, S. Merler, Y. Moreno, and A. Vespignani, “Measurability of the epidemic reproduction number in data-driven contact networks,” Proc. Nat. Acad. Sci. USA, vol. 115, pp. 12680–12685, 2018.

[10] P. Abry et al., “Spatial and temporal regularization to estimate COVID-19 reproduction number $R(1)$: Promoting piecewise smoothness via convex optimization,” PLoS One, vol. 15, 2020, Art. no. e0237901.

[11] B. Pascal, P. Abry, N. Pustelnik, S. Roux, R. Gribonval, and P. Flan- drin, “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. 2589–2568, 2022.

[12] F. Riccardo et al., “Epidemiological characteristics of COVID-19 cases in Italy and estimates of the reproductive numbers one month into the epidemic,” Euro Surveill., vol. 20, 2020, Art. no. 2000790, [Online]. Available: https://www.medrxiv.org/content/early/2020/04/11/2020.04.08.20058661

[13] G. Guzzetta et al., “The impact of a nation-wide lockdown on COVID-19 transmissibility in Italy,” 2020, arXiv:2004.12188.

[14] P. Moulin and J. Liu, “Analysis of multiresolution image denoising many: Springer, 2018.

[15] J. Kent, “Time-reversible diffusions,” Adv. Appl. Probability, vol. 10, no. 4, pp. 819–835, 1978.

[16] G. Roberts and O. Stramer, “Langevin diffusions and Metropolis-Hastings algorithms,” Methodol. Comput. Appl. Probability, vol. 4, pp. 337–357, 2002.

[17] G. Parisi, “Correlation functions and computer simulations,” Nucl. Phys. B, vol. 180, no. 3, pp. 378–384, 1981.

[18] A. Durmus and E. Moulines, “Nonasymptotic convergence analysis for the unadjusted Langevin algorithm,” Ann. Appl. Probability, vol. 27, pp. 1551–1587, 2017.

[19] R. Dwivedi, Y. Chen, M. J. Wainwright, and B. Yu, “Log-concave sam- pling: Metropolis-Hastings algorithms are fast!,” in Proc. 31st Conf. Learn. Theory, vol. 75, 2018, pp. 793–797.

[20] G. O. Roberts and R. L. Tweedie, “Exponential convergence of Langevin distributions and their discrete approximations,” Bernoulli, vol. 2, pp. 349–363, 1996.

[21] 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.

[22] N. Chatterji, J. Diakonikolas, M. I. Jordan, and P. Bartlett, “Langevin Monte Carlo without smoothness,” in Proc. 33rd Int. Conf. Artif. Intell. Statist., vol. 108, 2020, pp. 1716–1726.

[23] A. Durmus, E. Moulines, and M. Pereyra, “Efficient Bayesian computation by proximal Markov Chain Monte Carlo: When Langevin meets Moreau,” SIAM J. Imag. Sci., vol. 11, pp. 473–506, 2018.

[24] T. Luu, J. Fadili, and C. Chesneau, “Sampling from non-smooth distribu- tions through Langevin diffusion,” Methodol. Comput. Appl. Probability, vol. 23, pp. 1173–1201, 2021.

[25] Y. Achâdée, “A Moreau-Yosida approximation scheme for a class of high-dimensional posterior distributions,” 2015, arXiv:1505.07072v2.

[26] A. Schreck, G. Fort, S. L. Corff, and É. Moulines, “A shrinking-thresholding Metropolis adjusted Langevin algorithm for Bayesian vari- able selection,” IEEE J. Sel. Topics Signal Process., vol. 10, no. 2, pp. 366–375, Mar. 2016.

[27] A. Salim and P. Richtarik, “Primal dual interpretation of the proximal stochastic gradient Langevin algorithm,” in Advances in Neural Informa- tion Processing Systems, vol. 33, H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, Eds. Red Hook, NY, USA: Curran Associates, 2020, pp. 3786–3796.

[28] M. Pereyra et al., “A survey of stochastic simulation and optimization methods in signal processing,” IEEE J. Sel. Topics Signal Process., vol. 10, no. 2, pp. 224–241, Mar. 2016.

[29] D. Luengo, L. Martino, M. Bugallo, V. Elvira, and S. Sarkka, “A survey of Monte Carlo methods for parameter estimation,” EURASIP J. Adv. Signal Process., vol. 2020, 2020, Art. no. 25.

[30] J. Liu, F. Liang, and W. Wong, “The multiple-try method and local optimi- zation in Metropolis sampling,” J. Amer. Stat. Assoc., vol. 95, no. 449, pp. 121–134, 2000.

[31] Y. Nesterov, Introductory Lectures on Convex Optimization: A Basic Course, vol. 87. Berlin, Germany: Springer, 2003.

[32] Y.-L. Yu, “On decomposing the proximal map,” in Proc. Annua. Conf. Neural Inf. Process. Syst., Oaxaca, Mexico, 2013, pp. 91–99. [Online]. Available: http://papers.nips.cc/paper/4863-on-decomposing-the-proximal-map

[33] N. Pustelnik and L. Condat, “Proximity operator of a sum of functions: Application to depth map estimation,” IEEE Signal Process. Lett., vol. 24, no. 12, pp. 1827–1831, Dec. 2017.

[34] L. Martino, “A review of multiple try MCMC algorithms for signal processing,” Digit. Signal Process., vol. 75, pp. 154–152, Dec. 2018.

[35] M. Bédard, R. Douc, and E. Moulines, “Scaling analysis of multiple-try MCMC methods,” Stochastic Process Their Appl., vol. 122, no. 3, pp. 758–786, 2012.

[36] C. Andrieu and J. Thoms, “A tutorial on adaptive MCMC,” Statist. Com- put., vol. 18, pp. 343–373, 2008.

[37] H.-G. Chen and R. Rockafellar, “Convergence rates in forward-backward splitting,” SIAM J. Optim., vol. 7, pp. 421–444, 1997.
[50] P. Combettes and B. Vũ, “Variable metric forward–backward splitting with applications to monotone inclusions in duality,” *Optimization*, vol. 63, no. 9, pp. 1289–1318, 2014.

[51] P. Abry, G. Fort, B. Pascal, and N. Pustelnik, “Credibility intervals for the reproduction number of the Covid-19 pandemic using proximal Langevin samplers,” 2022. [Online]. Available: https://hal.archives-ouvertes.fr/hal-03902144

[52] 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.

[53] S. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2009.

[54] G. O. Roberts and J. Rosenthal, “Harris recurrence of Metropolis-within-Gibbs and trans-dimensional Markov chains,” *Ann. Appl. Probability*, vol. 16, no. 4, pp. 2123–2139, 2006.

[55] S. Geman and D. Geman, “Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 6, pp. 721–741, Nov. 1984.

[56] A. Gelfand and A. Smith, “Sampling-based approaches to calculating marginal densities,” *J. Amer. Stat. Assoc.*, vol. 85, no. 410, pp. 398–409, 1990.

[57] K. Chan and C. J. Geyer, “Discussion: Markov chains for exploring posterior distributions,” *Ann. Stat.*, vol. 22, no. 4, pp. 1747–1758, 1994.

[58] A. Gelman, W. Gilks, and G. Roberts, “Weak convergence and optimal scaling of random walk Metropolis algorithms,” *Ann. Appl. Probability*, vol. 7, no. 1, pp. 110–120, 1997.

[59] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods* (Springer Texts in Statistics Series). Berlin, Germany: Springer, 2005.

[60] A. Gelman and D. Rubin, “Inference from iterative simulation using multiple sequences,” *Stat. Sci.*, vol. 7, no. 4, pp. 457–472, 1992.

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

[62] 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 *Proc. 30th Eur. Signal Process. Conf.*, 2022, pp. 2196–2200.

[63] P. Abry, G. Fort, B. Pascal, and N. Pustelnik, “Temporal evolution of the COVID19 pandemic reproduction number: Estimations from proximal optimization to Monte Carlo sampling,” in *Proc. IEEE 44th Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2022, pp. 167–170.