Nonsmooth convex optimization to estimate the Covid-19 reproduction number space-time evolution with robustness against low quality data
Barbara Pascal, Patrice Abry, Nelly Pustelnik, Stéphane G. Roux, Rémi Gribonval, Patrick Flandrin

To cite this version:
Barbara Pascal, Patrice Abry, Nelly Pustelnik, Stéphane G. Roux, Rémi Gribonval, et al.. Nonsmooth convex optimization to estimate the Covid-19 reproduction number space-time evolution with robustness against low quality data. IEEE Transactions on Signal Processing, 2022, 70, pp.2859-2868. hal-03348154

HAL Id: hal-03348154
https://hal.science/hal-03348154
Submitted on 20 Sep 2021
Nonsmooth convex optimization to estimate the Covid-19 reproduction number space-time evolution with robustness against low quality data

Barbara Pascal Member, IEEE, Patrice Abry Fellow, IEEE, Nelly Pustelnik Member, IEEE, Stéphane Roux, Rémi Gribonval Fellow, IEEE, Patrick Flandrin Life Fellow, IEEE.

Abstract—Daily pandemic surveillance, often achieved through the estimation of the reproduction number, constitutes a critical challenge for national health authorities to design countermeasures. In an earlier work, we proposed to formulate the estimation of the reproduction number as an optimization problem, combining data-model fidelity and space-time regularity constraints, solved by nonsmooth convex proximal minimizations. Though promising, that first formulation significantly lacks robustness against the Covid-19 data low quality (irrelevant or missing counts, pseudo-seasonalities, . . . ) stemming from the emergency and crisis context, which significantly impairs accurate pandemic evolution assessments. The present work aims to overcome these limitations by carefully crafting a functional permitting to estimate jointly, in a single step, the reproduction number and outliers defined to model low quality data. This functional also enforces epidemiology-driven regularity properties for the reproduction number estimates, while preserving convexity, thus permitting the design of efficient minimization algorithms, based on proximity operators that are derived analytically. The explicit convergence of the proposed algorithm is proven theoretically. Its relevance is quantified on real Covid-19 data, with robustness against low quality data (missing samples, outliers, seasonalities, . . . ) stemming from the emergency and crisis context.

Index Terms—Covid-19, reproduction number, space-time evolution, nonsmooth convex optimization, outlier robustness.

I. INTRODUCTION

Context. The ongoing COVID-19 pandemic has produced an unprecedented health and economic crisis, urging for the construction of efficient monitoring procedures of its spreading across territories, a crucial step in designing sanitary, social and economic policies by national authorities [1]. It is however often not the value of the infection level per se that matters to design pandemic counter-measures (lockdown, . . . ), but its evolution along time and variations across territories. However, in the context of the Covid-19 pandemic outburst and of the sanitary crisis, collecting daily new infection counts, the basis material in any pandemic surveillance strategy, results in low-quality data (missing samples, outliers, seasonalities, . . . ). Surprisingly, after 18 months of pandemic, the quality of the data collected and made available by most national health authorities in the world remains limited, which severely impairs an accurate and timely pandemic evolution assessment, the issue at the heart of the present work.

Related works. Pandemic surveillance has been envisaged with several categories of tools from different fields of sciences (cf. [2] for a review). Yet, it is classically performed using compartmental models, elaborating on the classical Susceptible-Infectious-Recovered reference. For realistic use and to match social realities (social groups, contact, . . . ), these models need to be refined (cf. e.g., [3, 4]), essentially by increasing the number of compartments, which implies increasing (in a quadratic way) the numbers of parameters to be estimated. Parameter estimation is usually achieved within Bayesian frameworks, maximizing the likelihood attached to the models, at the expense of heavy computational burdens. Such models are thus often used a posteriori (i.e., after the epidemic) with consolidated and accurate datasets. The low quality of the Covid-19 data collected by most national public health authorities across the world significantly impairs the use of such Bayesian schemes and thus the reliable estimations of these parameters. This thus precludes the use of such models for daily basis update of the pandemic evolution assessment.

Instead of compartmental models, epidemiologists massively use the so-called reproduction number, \( R \), that measures how many new individuals are on average infected by an already contaminated person, a key marker of the pandemic strength (cf. e.g., [5–9]). Estimated to around 3 during the outburst of the Covid-19 pandemic [10, 11], it is also used as a function of time \( R_T \) to monitor the temporal evolution of the pandemic. The strength of such a pandemic monitoring is to be based on a single parameter to estimate \( R \), while accounting for the basic and universal pandemic propagation mechanisms: The number of new infections, \( Z_T \), at day \( T \), depends on \( R_T \) and on the numbers of new infections measured at previous days, \( \{Z_{T-1}, Z_{T-2}, Z_{T-3}, . . . \} \) weighted by the so-called serial interval function \( \Phi(t) \). The latter quantifies the probability that someone found Covid-positive today was actually contaminated several days ago [3, 8, 9, 12]. Following [9], \( R_T \) can be estimated using a Bayesian scheme:

\[ R_T = \frac{1}{\Phi(t)} \sum_{t=1}^{T} \frac{Z_t}{R_T} \]

where \( \Phi(t) \) is the serial interval distribution.

https://shiny.dide.imperial.ac.uk/epiestim/
Instead, in an earlier work [13], we proposed an estimation procedure of the spatio-temporal evolution of the reproduction number, written as a nonsmooth convex optimization problem, combining data-model fidelity with time and space regularity constraints for the estimation of \( R \) to be pandemic realistic, and solved by proximal algorithms [14–18]. Though promising, that proximal-optimization based estimation significantly lacks robustness against the low quality of the Covid-19 data, a generic and crucial issue addressed in the present contribution.

**Goals, contributions and outline.** The overall goal of this work is to devise an inverse problem-type convex-optimization procedure to assess the spatio-temporal evolution of the reproduction number \( R \), from daily new infection counts, measured simultaneously on several connected territories, such as the different counties of a same country. The proposed estimation procedure is designed to be: i) robust to the poor quality of Covid-19 data; ii) efficiently applicable online to update estimates, on a daily basis or whenever new data are available, at moderate computational costs even for large datasets.

To that end, Section II recalls the pandemic model used here [9], frames its classical estimation and related issues. Section III describes the core methodological contributions of the paper: i) It discusses the poor quality of the data and proposes an extension of the original pandemic modelling that handles poor quality data as outliers (cf. Section III-A); ii) It proposes an original and theoretical design of a functional whose minimization yields in a single step a joint estimation of both the reproduction number and the outliers, by combining the log-likelihood of the extended model as a data fidelity term with temporal and spatial regularity constraints for epidemiology-realistic estimates of \( R \) and sparsity constraints for the outliers (cf. Section III-B). This functional is carefully crafted to preserve convexity so that fast and efficient proximal algorithms can be devised for minimization; iii) The theoretical properties of the functional and of its solution (existence, unicity) are studied theoretically in Section III-C; iv) This permits to devise a proximal type iterative algorithm, with explicit analytical derivation of the proximal operators, whose convergence and stopping criterion are discussed theoretically and practically (cf. Section III-D).

The relevance of the proposed estimation procedure to assess the spatio-temporal evolution of the pandemic is illustrated on real Covid-19 data (daily counts of new infections) made available by national public health authorities (such as e.g., Santé-Publique-France, SPF) or collected by Johns Hopkins University, described in Section IV.

Estimation performance are reported in Section IV i) First, the role and choices of the hyperparameters balancing the regularization terms and the data-model fidelity term are discussed in detail (cf. Section IV-A); ii) Second, the benefits of the proposed strategy to estimate the temporal evolution of \( R_T \) at the level of a given country are illustrated by comparisons against different estimation strategies, such as a two-step procedure (outliers estimated and removed first, followed by estimation of \( R \)) (cf. Section IV-B); iii) Third, the relevance of the space-time evolution of \( R \) estimated across territories that are connected (shared borders, massive population commutations,…), here the 96 départements (counties) that organize administratively metropolitan France, is illustrated and discussed precisely (cf. Section V-C).

The proposed procedure is efficient enough to permit the automated daily updates of the estimates of \( R_T \) for 200+ countries, for the 96 continental French départements of France and for the 50 US states. Links to the resulting animated and interactive maps of estimates are provided and open-source estimation procedures will be made publicly available.

**II. Pandemic Model and Estimation**

**A. Model**

The pandemic model used here, focused on the reproduction number \( R \), was proposed in [9] and relies on two key ideas: i) Conditionally to past counts \( Z_{1:T-1} \equiv \{Z_1,\ldots,Z_{T-1}\} \), the count of daily new infections, \( Z_T \), is modeled as a random variable drawn from a Poisson distribution \( \text{Poiss}(p_T) \), ii) whose parameter \( p_T \) depends on the past counts \( Z_{1:T-1} \), on the causal serial interval function \( \Phi_T \) and on the current \( R_T \):

\[
P(Z_T|Z_{1:T-1};R_T) \sim \text{Poiss}
\left(p_T \triangleq R_T \times \sum_{t=1}^{\tau_0} \Phi_t Z_{t-1}\right)
\]  

The serial interval function \( \Phi_t \) is a key element of the model that accounts for the epidemiology evolution mechanisms, as it models the random delays between the onset of symptoms in a primary case and the onset of symptoms in secondary cases, [3, 8, 9, 12]. For the Covid-19 pandemic and for earlier pandemics of same types, it was shown that \( \Phi \) can be approximated as a Gamma function, with shape and rate parameters 1.87 and 0.28, respectively, corresponding to mean and standard deviations of 6.6 and 3.5 days, indicating a high risk of infecting other persons from 3 to 10 days after the symptoms have appeared [19–21]. It is assumed here that \( \Phi \) is known and follows this Gamma approximation.

**B. Estimation**

Given the analytic expression of the Poisson distribution in Model (1), the log-likelihood \( \ln P(Z_{1:T-1};R_{1:T}) = Z_T \ln p_T - p_T - \ln Z_T \), permits to compute the Maximum-Likelihood Estimate,

\[
\hat{R}_{T}^\text{MLE} = \text{argmin}(\ln P(Z_T|Z_{1:T-1};R_{1:T}))
\]  

closed-form expression as: \( \hat{R}_{T}^\text{MLE} = Z_T / \sum_{t=1}^{\tau_0} \Phi_t Z_{t-1} \). It can be interpreted as a ratio of moving averages, whose size and shape are consistent with pandemic modeling (thus with \( \Phi_T \). The \( \hat{R}_{T}^\text{MLE} \) computed from real Covid-19 data (cf. Section IV) are displayed in Figs. 2 and 3 (middle plots) and display a far too large variability along time to be of any practical use in actual pandemic monitoring. Such variability mostly stems from the low quality of the data. The goal of the present work is thus to improve the estimation of \( R \) despite the low quality of the available Covid-19 data.

**III. Robust estimation: Methodology**

**A. Covid-19 data low quality and multivariate infection counts**

Covid-19 daily new infection counts made available by public health authorities are for most countries severely corrupted, with missing samples, non meaningful negative counts
followed by retrospective cumulated counts spread across the following days, ... They also show pseudo-seasonality effects, with significantly less counts on non working days. Seasonality is however uneasy to model, as non working days are not only week-ends but also additional day-offs that depend on countries. Also, the way the pandemic has been monitored along week-ends and non-working days has significantly changed along time, even within a given country. Therefore, rather than trying to devise necessarily ad-hoc and parametric models for noise in Covid-19 data, we have instead chosen to refer to these different types of data corruption under the generic term of outlier, denoted $O_T$, and we propose to model observed daily new infection counts $Z_{iT}$ still from a Poisson distribution conditionally to past counts $Z_{i1:T-1}$, yet with a Poisson parameter $p_T$ that depends both on the current reproduction number $R_T$ and current outliers $O_T$.

Additionally, when new infection counts are monitored simultaneously for several connected territories, e.g., the countries or states of a given country, the observed daily counts consist of $D$ time series $\{Z_{i,t}^{(d)}, t \in [1, T], d \in [1, D]\}$, with $[1, T]$ the integers $\{1, 2, \ldots, T\}$. The pandemic spread on each territory $d$ is thus characterized by both a reproduction number time series and an outlier time series $\{(R_{i,t}^{(d)}, O_{i,t}^{(d)}), t \in [1, T]\}$, that must thus be estimated jointly for all territories.

These considerations lead to a first key contribution of the present work: the modeling of corrupted daily infection counts across territories by extending Model (1) to multivariate daily new infection Poisson counts, each with an instantaneous Poisson parameter $p_t^{(d)}$, with $(\Phi Z_{i,t}^{(d)}) = \sum_{u=1}^{\infty} P_{t,u} Z_{i,t}^{(d)}$.

\[ \mathbb{P}(Z_{i,t}^{(d)} | Z_{1:t-1}^{(d)}; R_t^{(d)}, O_t^{(d)}) \sim \text{Poiss}(p_t^{(d)} \approx R_t^{(d)} \times (\Phi Z_{i,t}^{(d)} + O_t^{(d)})). \tag{3} \]

**B. Crafting the convex nonsmooth estimation functional**

The goal is now to estimate jointly $R = \{R_t^{(d)}, t \in [1, T], d \in [1, D]\}$ and $O = \{O_t^{(d)}, t \in [1, T], d \in [1, D]\}$ from $\{Z_{i,t}^{(d)}, t \in [1, T], d \in [1, D]\}$. The next original contribution is thus to frame the estimation into an inverse problem strategy, with a careful crafting of the functional to minimize combining the negative log-likelihood of extended Model (3) with positivity and regularity constraints in time and space on $R$ and structure on $O$.

i) **Negative log-likelihood function:** The negative log-likelihood associated to the extended Poisson Model (3) is defined in a separable manner as

\[ F(R, O | Z) \triangleq \sum_{d=1}^{D} \sum_{t=1}^{T} f(R_{i,t}^{(d)}, O_{i,t}^{(d)} | Z_{i,t}^{(d)}, (\Phi Z)_{i,t}^{(d)}), \tag{4} \]

with

\[ f(r, o | z, (\Phi z)) \triangleq \begin{cases} \ell\{0,0\}(r, o) & \text{if } z = \Phi z = 0, \\ d_{KL}(z | r \times (\Phi z) + o) & \text{otherwise}, \end{cases} \]

The indicator function of the singleton $\{(0,0)\}$ permits to force the (MLE) estimates of $R$ and $O$ to vanish when the pandemic stops, i.e., when $\Phi Z_{i,t}^{(d)} = Z_{i,t}^{(d)} = 0$, $\ell\{0,0\}(r, o) = \begin{cases} 0 & \text{if } r = o = 0 \\ \infty & \text{otherwise}, \end{cases}$

while the standard Kullback-Leibler divergence (KLD) is defined as $D_{KL}(Z|P) = \sum_{t=1}^{T} \sum_{d=1}^{D} d_{KL}(Z_{i,t}^{(d)} | P_{i,t}^{(d)})$, with

\[ d_{KL}(z|p) \triangleq \begin{cases} z \ln \frac{z}{p} + p - z & \text{if } z > 0, p > 0, \\ p & \text{if } z = 0, p > 0, \\ \infty & \text{else}. \end{cases} \tag{6} \]

2) **Regularity and positivity constraints on $R$:** Unlike Model (1), from which a unique MLE estimate can be derived (Eq. (2)), for Model (3) the extended likelihood (Eq. (4)) is no longer strictly convex, thus does not have a unique maximum, and hence requires additional constraints on the estimates of $R$ and $O$.

i) **Time regularity:** By nature and to be of any potential use for actual pandemic monitoring, the estimated reproduction numbers for a given territory must vary only slowly and smoothly along time. Following [13], it has been chosen to enforce piecewise linear time evolution of the estimate of the reproduction numbers. Such a property ensures that, besides an estimate of $R$ at day $t$, one also gets an estimation of a local trend, assessing whether the pandemic is growing or decreasing. Following [22][23], to favor piecewise linear temporal evolution of the estimates, we penalize the $\ell_1$-norm of the multivariate time-domain Laplacian $D_2 : \mathbb{R}^{D \times T} \to \mathbb{R}^{D \times (T-2)}$ of the estimates of $R$: \[ ||D_2 R||_1 \triangleq \sum_{d=1}^{D} \sum_{t=2}^{T-1} |(D_2 R)_{i,t}^{(d)}|, \tag{7} \]

with $(D_2 R)_{i,t}^{(d)} \triangleq \frac{1}{7} (R_{i,t-1}^{(d)} - R_{i,t}^{(d)} + \frac{2}{7} R_{i,t+1}^{(d)}).$ ii) **Positivity:** $R$ is by nature positive. Yet, Eq. (4) indicates that, per se, the proposed extended negative log-likelihood ensures the positivity of $p_t^{(d)} = R_t^{(d)} \times (\Phi Z_{i,t}^{(d)} + O_t^{(d)})$ but not of $R_t^{(d)}$ itself. Thus, to enforce the positivity of $R_t^{(d)}$, an indicator function is added in the constraints:

\[ \ell_{\geq 0}(R) \triangleq \begin{cases} \infty & \text{if } \exists t \in [1, T], d \in [1, D] \text{, s.t. } R_t^{(d)} < 0, \\ 0 & \text{otherwise}. \end{cases} \tag{8} \]

iii) **Spatial regularity:** When different territories $d$ are connected, it is natural to expect that their pandemic dynamics are correlated and thus that the corresponding estimates of $R$ show some spatial regularity. Following [13], we use a graph, with $D$ vertices corresponding to the different territories and $E$ edges to connections between pairs of territories. We favor piecewise constant spatial estimates of $R_t^{(d)}$ across territories by penalizing the Graph Total Variation, defined as:

\[ \text{GTV}(R) = \sum_{t=1}^{T} \sum_{d_1 \sim d_2} \left| R_t^{(d_1)} - R_t^{(d_2)} \right|, \tag{9} \]

where $\sum_{d_1 \sim d_2}$ runs over vertices connected by an edge.

The Graph Total Variation can be rewritten as a composition of a discrete difference operator and a sparsity-promoting $\ell_1$-norm, $\text{GTV}(R) = ||G R||_1$, defining the linear operator $G$:

\[ G : \mathbb{R}^{D \times T} \to \mathbb{R}^{E \times T}, \quad R \mapsto \left\{ R_t^{(d_1)} - R_t^{(d_2)}, t \in [1, T], d_1 \sim d_2 \right\}. \tag{10} \]
3) Constraints on the outliers: The discussion in Section III-A suggests that outliers have a sparse structure. They occur with significant values only on specific days (sundays, day-offs,...) and are essentially negligible elsewhere. Sparsity is classically enforced in inverse problem settings by constraining the \( \ell_1 \)-norm. We thus further propose to impose the following constraint to the estimates of the outliers:

\[
\| O \|_1 \triangleq \sum_{d=1}^{D} \sum_{t=1}^{T} | O^{(d)} _t | .
\]  

(11)

4) Estimation as an optimization problem: Combining the data-fidelity term of Eq. (4) and the penalizations of Eqs. (7), (8), (9), and (11), we propose to obtain the estimates of \( R \) and \( O \) by minimizing:

\[
( \hat{R}, \hat{O} ) \in \operatorname{Argmin}_{R,O \in \mathbb{R}^{D \times T}} F( R, O | Z ) + \lambda_T \| D_2 R \| _1 + \iota_{\geq 0}(R) + \lambda_S \| G R \| _1 + \lambda_O \| O \| _1 ,
\]  

(12)

with \( \lambda_T > 0, \lambda_S > 0 \) and \( \lambda_O > 0 \), regularisation hyperparameters, balancing the strengths of the different constraints one against the others and against the likelihood.

5) Reformulation of the optimization problem: To study the theoretical properties of Problem (12), and to derive a minimization algorithm, it is useful to recast Eq. (12) into a generic formulation:

\[
( \hat{R}, \hat{O} ) = \operatorname{Argmin}_{R,O \in \mathbb{R}^{D \times T}} F( R, O | Z ) + H( L( R, O ) ),
\]  

(13)

with \( L \) a linear operator and \( H \) a function enforcing constraints promoting certain sparsity / non-negativity, defined as:

\[
L : \left\{ \begin{array}{c}
\mathbb{R}^{D \times T} \\
( R, O ) \end{array} \right. \rightarrow \mathbb{R}^{D \times (T-2) \times D \times T \times E \times T \times D \times T}
\]  

(14)

\[
H( Q_1, Q_2, Q_3, Q_4 ) \triangleq \| Q_1 \| _1 + \iota_{\geq 0}(Q_2) + \| Q_3 \| _1 + \| Q_4 \| _1 .
\]  

(15)

C. Convexity, existence of a minimizer and uniqueness

Showing that Eq. (12) provides well-defined estimates of \( R \) and \( O \) requires to prove the existence of a minimizer of \( F( \cdot ) + H( L \cdot ) \). To that aim, preliminary results are necessary.

Lemma 1. The extended Kullback-Leibler data-fidelity term \( F( \cdot | Z ) \) is jointly convex with respect to the pair \( ( R, O ) \).

Proof. \( F \) being fully separable in time and space, it is sufficient to demonstrate the joint convexity of \( f( r, o | z, ( \Phi z ) ) \) with respect to the pair \( ( r, o ) \). If \( z = ( \Phi z ) = 0 \), the convexity of \( f( r, o | z, ( \Phi z ) ) \) w.r.t. \( ( r, o ) \) reduces to the convexity of the indicator function of the convex set \{\((0,0)\)\}. Else, since the one-component KLD \( d_{\text{KL}}(z|p) \) is convex with respect to its second argument [24], and \( ( r, o ) \rightarrow r \times ( \Phi z ) + o \) being linear, \( ( r, o ) \rightarrow f( r, o | z, ( \Phi z ) ) \) is convex.

Corollary 1. The penalized Kullback-Leibler functional (Eq. (12)) is jointly convex with respect to the pair \( ( R, O ) \).

Proof. First, Lemma 1 ensures the convexity of the data-fidelity term \( F \). Then, the function \( H \), appearing in Eq. (13), involves \( \ell_1 \)-norms and an indicator function which are convex, thus it is convex and the composition of \( H \) with the linear operator \( L \) is convex as well. As the sum of two convex terms, the penalized Kullback-Leibler functional is convex.

Further, because the Kullback-Leibler data-fidelity term \( F \) (Eq. (4)), as well as the indicator function in Eq. (8), may take infinite values, the domain of the convex functional in Eq. (12) is not the whole space \( \mathbb{R}^{D \times T} \), but restricts to a domain \( \Omega \) of feasible variables, described in Lemma 2.

Lemma 2. The domain \( \Omega \) of feasible points for Problem (12) consists of pairs \( ( R, O ) \) such that, \( \forall u \in [1, t], \forall d \in [1, D] \)

(\( i \)) \( R^{(u)}_t ( \Phi z )^{(d)}_t + O^{(d)}_t \geq 0 \),

(\( ii \)) \( Z^{(d)}_t = ( \Phi z )^{(d)}_t = 0 \) then \( R^{(d)}_t = O^{(d)}_t = 0 \),

(\( iii \)) \( R^{(d)}_t \geq 0 \).

It is convex and closed.

Proof. Conditions (\( i \)) and (\( ii \)) reflect the fact that the extended Kullback-Leibler data-fidelity term \( F \) (Eq. (4)) must be finite for feasible time series, while Condition (\( iii \)) stems from the positivity constraint enforced by the indicator function in Eq. (8). Further, each Condition (\( i \)) to (\( iii \)) defines a closed convex set of feasible time series, the finite intersection of closed convex sets being closed and convex as well, \( \Omega \) is a closed convex set.

Then, the existence of a minimizer can be proven.

Theorem 1. The penalized Kullback-Leibler functional in Eq. (12) is lower bounded. Further, if \( ( \Phi z )^{(d)}_t > 0 \) for all \( t \in [1, T], d \in [1, D] \), then the minimization problem of Eq. (12) has a least, one solution.

Proof. Since, for feasible points, the standard Kullback-Leibler divergence, the indicator functions and the \( \ell_1 \)-norms are nonnegative, the functional is lower bounded by zero. Let \( \| ( R, O ) \| _2 \triangleq \sqrt{ \| R \| _2^2 + \| O \| _2^2 } \) be the canonical norm on the product space \( \mathbb{R}^{D \times T} \). Let 1 (resp. 0) the multivariate time series with all entries equal to 1 (resp. 0), \( (1, 0) \in \Omega \) is a feasible point. Let

\[
\mu \triangleq F( 1, 0 | Z ) + H( L( 1, 0 ) ) = F( 1, 0 | Z ) \geq 0
\]  

(16)

As shown below, the functional \( F( \cdot | Z ) + H( L \cdot ) \) is coercive, thus there exists \( \alpha > 0 \) such that

\[
\| ( R, O ) \| _2 > \alpha \Rightarrow F( R, O | Z ) + H( L( R, O ) ) > \mu .
\]  

(17)

Let \( \Delta_\alpha \) be the closed ball of radius \( \alpha \) in \( \mathbb{R}^{D \times T} \) for the canonical norm \( \| \cdot \| _2 \). Since the feasible set \( \Omega \) is closed, the intersection \( \Omega \cap \Delta_\alpha \) is bounded and closed, hence compact as a subset of a finite-dimensional Hilbert space. Further, the functional \( F( \cdot | Z ) + H( L \cdot ) \) being continuous over \( \Omega \cap \Delta_\alpha \) it reaches its minimum over this set at some point \(( \hat{R}, \hat{O} ) \in \Omega \cap \Delta_\alpha \). Then, since \( (1, 0) \in \Omega \cap \Delta_\alpha \), it follows that \( F( \hat{R}, \hat{O} | Z ) + H( L( \hat{R}, \hat{O} ) ) \leq \mu \). From Eq. (17), if \( ( R, O ) \in \Omega \setminus \Delta_\alpha \), then one has \( F( R, O | Z ) + H( L( R, O ) ) > \mu \), showing that the relative minimum \(( \hat{R}, \hat{O} ) \) on \( \Omega \cap \Delta_\alpha \) is to be a global minimum on the whole domain of feasible points \( \Omega \). To conclude the proof, we need to compute \( \alpha \) satisfying (17), which follows from:
In finite dimension, the equivalence of the norms yields con-
formly vanishing, i.e.,

Proof. Proposition 1. any minimizer

Lemma 3 (\[14\], Chapter 11). Let \(g : \mathbb{R}^n \to \mathbb{R}, h : \mathbb{R}^m \to \mathbb{R}\) two proper, lower-semicontinuous, convex functions and \(B : \mathbb{R}^m \to \mathbb{R}^n\) a linear operator. Consider the following convex minimization problem

\[
\hat{x} \in \arg\min_{x \in \mathbb{R}^m} g(Bx) + h(x).
\] (18)

If \(g\) is strictly convex, then there exists \(\hat{y} \in \mathbb{R}^n\), such that for any minimizer \(\hat{x}\) of (18), \(\hat{y} = B\hat{x}\).

Proposition 1. Let \((\hat{R}, \hat{O})\) denote a solution of Problem (12). Let \(\hat{P} \triangleq \{p_i^{(d)} \mid t \in \{1, \ldots, T\}, d \in \{1, \ldots, D\}\}\), then \(\hat{P}\) is unique.

Proof. In the trivial situation when the daily counts are uniformly vanishing, i.e., \(Z = \Phi Z \equiv 0\), the unique minimizer coincides with the \(\hat{y}\) feasible point and consists in estimated reproduction numbers and outliers equal to zero at any time and in any territory.

Otherwise, two situations are to be considered

(i) If \(Z\) and \(\Phi Z\) never vanish simultaneously: In this case \(F(R, O, Z) = D_{KL}(Z | B(R, O))\), with \(B : (R, O) \mapsto R \times (\Phi Z) + O\). Further, the standard Kullback-Leibler divergence \(g(P) = D_{KL}(Z | P)\) is strictly convex. Hence, setting \(h = H(L, \cdot)\), Proposition 1 yields the result.

(ii) If there exist some times \(u\) and territories \(d\) s.t. \(Z^{(d)}_{t_i} = (\Phi Z)^{i}_{t_i} = 0\): For these times and territories, the estimated instant Poisson parameter is forced to cancel by the indicator function involved in Eq. (4). Then, restricting to the indices for which \(Z\) and \(\Phi Z\) are not vanishing simultaneously, the data-fidelity term \(F\) is strictly convex and the uniqueness of \(\hat{P}\) follows by the same reasoning.

D. Minimization via a proximal algorithm

From Corollary 1 the penalized Kullback-Leibler functional is convex, and, from Theorem 1 it has at least one minimizer. The \(\ell_1\)-norm and the indicator functions in the objective function make the overall functional non-smooth, preventing from the use of standard gradient descent. Minimizing \(\ell_1\)-norm thus requires more advanced tools, handling non-differentiable functions, such as proximal algorithms.

1) Adjoint and proximal operators: Let \(L\) a bounded linear operator \(L\), its adjoint operator is denoted \(L^*\) and its operator norm is \(\|L\|_\text{op} = \sup \{\|LX\|_2 \mid \|X\|_2 \leq 1\}\). The convex conjugate of the proper, lower-semicontinuous function \(H : X \mapsto \mathbb{R}\), is defined as \(H^*(X) = \sup_{X \in \mathbb{R}} \langle X, X \rangle - H(X)\) \([25]\), with \(\langle \cdot, \cdot \rangle\) a Hilbert space. For \(\tau \in \mathbb{R}^+\) the proximal operator of \(\tau H\) at \(X \in \mathbb{R}\) reads by definition: \(\text{prox}_{\tau H}(X) \triangleq \arg \min_{X \in \mathbb{R}} \frac{1}{2}\|X - Y\|^2 + \tau H(X)\).

2) Primal-dual algorithmic scheme: In this work, we chose to implement the Chambolle-Pock primal-dual minimization scheme \([26]\). Making use of the compact reformulation in Eq. (13), we perform here an explicit particularization of the Chambolle-Pock algorithm detailed in Algorithm 1. From \([26]\), Algorithm 1 converges toward a minimizer of Problem (13), if the descent parameters \(\tau, \sigma\) satisfy

\[
\tau\sigma \|L\|^2_{\text{op}} < 1.
\] (19)

Therefore, the actual and practical use of Algorithm 1 and enforcement of Condition (19), requires to evaluate, at least, an upper bound of \(\|L\|_{\text{op}}\).

**Proposition 2.** Let \(L\) be defined in Eq. (14), then

\[
\|L\|^2_{\text{op}} \leq \max\{\lambda^2_1 \|D_2\|^2_{\text{op}} + \lambda_8 \|G\|_{\text{op}}^2 + 1, \lambda^2_3\}.\] (20)

**Proof.** By definition of the operator norm \(\|\|_{\text{op}}\), one has

\[
\|L(R, O)\|^2_2 = \lambda^2_1 \|D_2R\|^2_2 + \|R\|^2_2 \|G\|_{\text{op}}^2 + \lambda^2_3 \|O\|^2_2
\]
\[
\leq \lambda^2_1 \|D_2\|^2_{\text{op}} + 1 + \lambda_8 \|G\|_{\text{op}}^2 \|R\|^2_{\text{op}} + \lambda^2_3 \|O\|^2_2
\]
\[
\leq \max\{\lambda^2_2 \|D_2\|^2_{\text{op}} + 1 + \lambda_8 \|G\|_{\text{op}}^2, \lambda^2_3\}\|\|L(R, O)\|^2_2.
\]

In practice, equal descent step sizes are chosen, so that they saturate Condition (19), replacing the unknown \(\|L\|_{\text{op}}\) by its upper bound derived at Proposition 2 leading to

\[
\tau = \sigma = \frac{0.99}{\sqrt{\max\{\lambda^2_1 \|D_2\|^2_{\text{op}} + \lambda_8 \|G\|_{\text{op}}^2 + 1, \lambda^2_3\}}}.
\] (22)

3) Closed-form expression of the proximal and adjoint operators: For proximal algorithms, such as Algorithm 1 the key ingredient to achieve an actual and fast implementation is to derive a closed-form expression for each proximal operator involved.

**Proximal operator of the regularizing function:** The dual variable update in Algorithm 1 involves the proximal operator of the convex conjugate of the penalty \(H\) defined in (15). In practice, first, the proximal operator of \(H\) is computed, and then, the proximal operator of \(H\) is derived applying Moreau’s identity, which states that, for \(\sigma > 0\), \(\text{prox}_{\sigma H}(X) = X - \sigma \text{prox}_{H/\sigma}(X/\sigma)\). Then, function \(H/\sigma\) being fully separable, both over its four entries and over the components of each entry, its proximal operator can be computed component-wise. These computations involve two standard closed-form proximal operators: i) the proximal operator of the absolute value, known as the soft-thresholding, \(\text{prox}_{\sigma}(|q|) = \max\{|q| - \sigma, 0\} \frac{q}{|q|}\), and ii) the proximal operator of the indicator function, which amounts to a projection: \(\text{prox}_{\sigma} q = \max\{0, q\}\)
Algorithm 1: Primal-dual minimization of the penalized Kullback-Leibler \cite{12} for the estimation of reproduction numbers and outliers.

Require: Infection counts: \( Z \in \mathbb{R}^{D \times T} \) and \((\Phi Z) \in \mathbb{R}^{D \times T}\).

Choose descent parameters: \( \tau, \sigma > 0 \), precision: \( \epsilon \).

Max. iterations: \( k_{\text{max}} \), length of window: \( k_{\text{smooth}} \).

Initialization \( R^{[0]} = Z, O^{[0]} = 0 \), \( Q^{[0]} = L(R^{[0]}, O^{[0]}), \tilde{R}^{[0]} = R^{[0]}, \tilde{O}^{[0]} = O^{[0]} \).

\( \Phi_0 = F(R^{[0]}, O^{[0]} | Z) + H(L(R^{[0]}, O^{[0]})) \).

\( \Psi_0 = \epsilon + 1, \tilde{\Psi}_0 = \epsilon + 1, k = 0 \).

while \( \Psi_k \geq \epsilon \) and \( k < k_{\text{max}} \) do

Update the dual, primal and auxiliary variables:

\[
Q^{[k+1]} = \text{prox}_{\sigma H^*}(Q^{[k]} + \sigma L(\tilde{R}^{[k]}, \tilde{O}^{[k]}))
\]

\[
(R^{[k+1]}, O^{[k+1]}) = \text{prox}_{\tau F}((R^{[k+1]}, O^{[k+1]}) - \tau L^* Q^{[k+1]})
\]

\[
(\tilde{R}^{[k+1]}, \tilde{O}^{[k+1]}) = 2(R^{[k+1]}, O^{[k+1]}) - (R^{[k]}, O^{[k]})
\]

Computation of the convergence criteria:

\[
\Phi_{k+1} = F(R^{[k+1]}, O^{[k+1]} | Z) + H(L(R^{[k+1]}, O^{[k+1]}))
\]

\[
\Psi_{k+1} = |\Phi_{k+1} - \Phi_k| / \Phi_k
\]

\[
\tilde{\Psi}_{k+1} = \max \{ \Psi_{k} : |k - k_{\text{smooth}}| + \ell \leq k + 1 \} \quad (23)
\]

\( k \leftarrow k + 1 \)

end while

Adjoint of the linear operator: Updating the primal variable requires first to compute \( L^* : \mathbb{R}^{D \times (T-2)} \rightarrow (\mathbb{R}^{D \times T})^2 \), which can be expressed in terms of the adjoint operators of \( D_2 \) and \( G \), via

\[
L^* : (Q_1, Q_2, Q_3, Q_4) \mapsto (\lambda_2 D_2^* Q_1 + Q_2 + \lambda_3 G^* Q_3, \lambda_4 Q_4).
\]

(24)

Proximal operator of the data-fidelity term: The extended negative log-likelihood \( F \) of Eq. (3) being separable over the times \( t \in [1, T] \) and territories \( d \in [1, D] \), its proximal operator can be computed in a component-wise manner. Each term in \( (\Phi Z) \) corresponds to the composition of a linear map with the standard Kullback-Leibler divergence \( d_{KL}(z|\cdot) \), whose proximal operator writes explicitly \cite{24}:

\[
\text{prox}_{d_{KL}(z|\cdot)}(p) = (p - \tau + \sqrt{(p - \tau)^2 + 4\tau z}) / 2. \quad (25)
\]

Proposition 3. For \( \beta \triangleq (\Phi Z)^2 + 1 \), \( a : (r, o) \mapsto r \times (\Phi z) + o \),

\[
\text{prox}_{\tau f(\cdot, \cdot, (\Phi z))}(r, o) = \begin{cases} 
    (0, 0), & \text{if } z = (\Phi z) = 0 \\
    (r, o) - a^* \beta^{-1} (I - \text{prox}_{\tau d_{KL}(z|\cdot)}) a(r, o), & \text{otherwise}
\end{cases}
\]

from which derives straightforwardly the proximal operator of the global data-fidelity function \( F \).

Proof. Since \( a a^* = \beta > 0 \), the formula in Eq. (26) is obtained applying Proposition 3.4 of \cite{27}.

4) Stopping criterion: The final step in the actual of an iterative algorithm is to choose a stopping criterion. Classically, the convergence of Algorithm \cite{11} is quantified by the decrease of the normalized increments of the objective functional.

Fig. 1 shows that such increments (dashed blue curves) display large and erratic fluctuations, thus precluding their use to construct a reliable stopping criterion. Instead, and to provide a robust convergence criterion, we propose to make use of a non-linearly smoothed normalized increments (solid black curve in Fig. 1), obtained as a maximum over a sliding window over the past \( k_{\text{smooth}} \) iterations, as detailed in Eq. (25). Hence, in our concrete experiments, we set \( k_{\text{smooth}} = 500 \) and Algorithm 1 is stopped whenever the smoothed increment reaches a precision \( \epsilon \), set to \( 10^{-7} \), or when the number of iterations exceeds a maximum of \( k_{\text{max}} \) iterations, set to \( 10^7 \).

IV. COVID-19 DATA

To show the relevance and efficiency of the proposed procedure for the estimation of the evolution of the reproduction number of the pandemic, two real Covid-19 datasets made available by national health authorities are used, different in nature and from different international repositories.

1) Dataset1: Country population level daily new infection counts: Johns Hopkins University\cite{13} provides access to the cumulated counts of daily new infections (together with deceased and recovered persons), at the entire population level, on a per country basis, for 200+ countries worldwide, since January 1st, 2020, hence, essentially since the earliest stages of the Covid-19 pandemic. In the present work, use will be made of the daily new infection counts only, \( \{Z_{1:T}\} \), from Feb., 15th, 2020 until until August, 31st, 2021, hence \( T = 586 \) days of pandemic.

2) Dataset2: French Counties hospital Level daily new infection counts: Santé-Publique-France\cite{14} the French national health authorities, provides several different datasets related to the Covid pandemic in France. In the present work, use will be made of the daily hospital-recorded new infection counts, across France on a per département-basis, départements (or counties) consisting of the usual French administrative granularity. These new infections counts are stacked into a \( Z \in \mathbb{R}^{T \times D} \) data matrix, for the 101 French départements, and estimation of the reproduction number will be conducted jointly for all départements. Such data are however available only after March 19th, 2020 and analyzed until August, 31st, 2021, hence \( T = 531 \) days of pandemic.

https://coronavirus.jhu.edu/
https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_time_series/
https://www.santepubliquefrance.fr/
https://www.data.gouv.fr/fr/datasets/
V. Robust estimation from Covid-19 data

A. Hyperparameter tuning

One of the issue at hand for a practical use of Eq. 12 to estimate jointly $R$ and $O$ lies in the selection of the hyperparameters, $\lambda_T$, $\lambda_S$, and $\lambda_O$ that balance the contributions of the regularizations terms one against the others as well as against the data-model fidelity term. Such choices have significant impacts on the achieved estimates. Accurate parameter tuning is thus critical yet may prove time consuming notably when this selection needs to be made for each of the 200+ countries studied here, with possibly very different population sizes or pandemic intensities.

Automated data driven selection of hyperparameters in functional minimization has been addressed theoretically in several different settings (cf. e.g. [28] and references therein), often relying on a Stein Unbiased Risk Estimator (SURE). This is working well only for large size data, which is however not the case for the Covid-19 pandemic.

Instead, in the present work, hyperparameter selection was driven by the following dimensional analysis. Inspecting Eq. [12] shows that the functional to minimize is covariant under the change $(Z, O; \lambda_T, \lambda_S, \lambda_O) \to (\alpha Z, \alpha O; \alpha \lambda_T, \alpha \lambda_S, \alpha \lambda_O)$ for any $\alpha > 0$, thus leading to identical minima. This led us to propose to use one same set of hyperparameters $(\lambda_T, \lambda_S, \lambda_O)$ valid for all countries, irrespective of their population sizes or pandemic intensities, by applying minimizations in [12] to $\alpha Z$ instead of $Z$, where $\alpha$ are arbitrary multiplicative constants, that may depend on each country. We found empirically that setting systematically $\alpha$ to the empirical standard deviation of the observed $Z$ constitutes an efficient way to account both for the population size and for the severity of the pandemic. Precisely, $\lambda_T = 3.5$ is chosen as a universal (for all countries) time constant, after inspections of numerous different countries over the whole pandemic period, as a valid trade-off between too small — that would lead to too much variability in the estimation of $R$ and would prevent the use of $R$ to actually assess the strength of the pandemic and that would lead to too few variability hence preventing the possibility of detecting changes in the pandemic dynamics that may follow some sanitary policy decisions (lockdown, vaccination). Further, $\lambda_O$ is set to 0.025 to enable the outlier term in the functional to be versatile enough to account for pseudo-seasonal effects modeled as sparse irrelevant measures. Finally, for different $D = 96$ territories, the connectivity of the graph for metropolitan France contains 475 edges, indicating that the time and space regularizations will have comparable impacts when $96 \lambda_{\text{time}} \simeq 2 \times 475 \lambda_{\text{space}}$, which frames the selection of $\lambda_{\text{space}}$. Inspection of the results lead to choose $\lambda_{\text{space}} = 0.002$, thus corresponding to a mild spatial regularization.

B. Temporal evolution for each country independently

Because Public Health Policies were not or weakly coordinated across countries, even for closely intertwined spaces such as Western European countries, population level daily new infection counts (Dataset1, JHU) are first analyzed inde-

![Fig. 2: France (six first months in 2021). Top: $Z$ and $Z-\hat{O}_O$, (black and red) ; Middle: $\hat{R}_O$, $\hat{O}_O$, $\hat{R}_{noO}$ (black, red and blue) ; Bottom: $O_O$, $\hat{O}_{noO}$ (red and blue).](image)
observed across numerous other countries.

Fig. 3 further compares the estimates $\hat{R}_{MLE}$, $(\hat{R}_O, \hat{O}_O)$ and $(\hat{R}_{noO}, \hat{O}_{noO})$ for several different countries. Fig. 3 clearly shows that i) $\hat{R}_{MLE}$ constitute a very irregular estimate of $R$ along time that cannot be used at all in practice ; ii) compared to $(\hat{R}_{noO}, \hat{O}_{noO})$, $(\hat{R}_O, \hat{O}_O)$ better accounts for both outliers and pseudo-seasonal effects and yields smooth and regular along time and outlier-robust estimations of $R$. $\hat{R}_O$ is thus far more likely to represent the actual, less biased and realistic assessment of the pandemic time evolution in a given country. Estimates $(\hat{R}_O, \hat{O}_O)$ are updated on a daily basis for 200+ countries across the world, and made publicly available via interactive and animated maps available (cf. Section VII).

C. Joint space-time Evolution for metropolitan France

Let us now consider multivariate daily infections counts, $Z \in \mathbb{R}^{T \times D}$, for the $D = 96$ départements of metropolitan France, considered as the vertices of the graph $G$ in Eq. 12 with edges between départements sharing a terrestrial boundary. Estimates $(\hat{R}_{O,S}, \hat{O}_{O,S}) \in \mathbb{R}^{T \times D}$ are obtained by applying Eq. 12 to $Z \in \mathbb{R}^{T \times D}$, after standardization independently per département, with $(\lambda_T, \lambda_S, \lambda_O) = (3.5, 0.002, 0.025)$ (cf. Section V-A).

To illustrate the relevance of $(\hat{R}_{O,S}, \hat{O}_{O,S})$, they are compared against estimates obtained without spatial regularization and without accounting for outliers $(\hat{R}_{noO,noS}, \hat{O}_{noO,noS})$ ($\lambda_T, \lambda_S, \lambda_O = (3.5, 0, +\infty)$), without spatial regularization but accounting for outliers $(\hat{R}_{O,noS}, \hat{O}_{O,noS})$ ($\lambda_T, \lambda_S, \lambda_O = (3.5, 0, 0.025)$), with spatial regularization but without accounting for outliers $(\hat{R}_{noO,S}, \hat{O}_{noO,S})$ ($\lambda_T, \lambda_S, \lambda_O = (3.5, 0.002, +\infty)$).

These estimates are first compared as functions of time for two départements in Fig. 4, which shows that estimates that do not account for outliers are far too irregular for being useful in pandemic monitoring, even with spatial regularization. To the converse, $\hat{R}_{O,S}$ that results from both accounting for outliers and spatial regularization yields very regular estimates likely to reflect a relevant assessment of the pandemic. The proposed estimation procedure is of particular interest when applied to territories with granularity level as is the case for the French départements (with on average slightly less than one million inhabitants) or during low activity phases of the pandemic, such as early summer 2020 (days 95 to 155). Spatial regularization is of particular relevance for theses two chosen départements. Indeed, département Paris (Id=75) corresponds to the city of Paris, thus a small territory yet with massive connection to the surrounding départements : estimates of $R$ within Paris, can thus not notably differ from those of these so-called Paris-Crown départements. The same holds for département Rhône (Id=69), of very small surface (for historical reasons) yet acting, with the city of Lyon as a massive transit hub, for a number of surrounding départements.

Further, the estimates of $R$ are for continental France are reported in Fig. 3 for three different days. Estimates of $\hat{R}_{O,S}$ with spatial regularization and accounting for outliers (right most plots) permits a far clearer assessment of the status of the pandemic across metropolitan France. Estimates $\hat{R}_{noO,S}$ and $\hat{R}_{noO,noS}$ show far too much spatial variability to be realistic estimates. Estimates $\hat{R}_{noO,S}$ improve spatial regularity, but still show significant variability induced by data low quality. Estimates $\hat{R}_{O,S}$, with spatial regularization
Fig. 5: Compared estimates of $R^{(d)}_t$ across metropolitan France for three different days along the pandemic.

and accounting for outliers (right most plots), likely yield the most likely and relevant estimates each day and permit a far clearer assessment of the status of the pandemic across the connected metropolitan France départements. March, 31st, corresponds to the pandemic 1st-wave maximum and $R_{O,S}$ shows a clear North-East/South-West gradient in the strength of the pandemic; March, 31st, 2021 corresponds to the start of the third lockdown period, referred to as couvre-feu (corfew) in France and $R_{O,S}$ shows that the pandemic was significantly progressing uniformly across all France; August, 31st, 2021 corresponds to the time of submission of this work and shows that the pandemic is globally decreasing in France, yet non uniformly and still active in large areas. Estimates $R_{O,S}$ are updated automatically on a daily basis for the 96 continental France départements and made publicly available via interactive and animated maps (cf. Section VI).

VI. CONCLUSION

The present work developed and assessed an inverse problem-type procedure to estimate the spatio-temporal evolution of the Covid-19 reproduction number, $R$, robust to the low quality of the pandemic data. The devised procedure relies first on a detailed analysis of the data quality that suggested that most count misreports, be they mild and seasonal (sundays, week-end, days-off) or large and random (report failure), can be efficiently accounted for as sparse outliers. Second, it is based on devising carefully a functional that balances a data-model fidelity term (Poisson distribution and outliers), and regularity in time and space properties that $R$ needs to fulfill to be of actual relevance in practical pandemic monitoring. Fast and efficient algorithms were devised to minimize this functional and their convergence was studied. Applied to real Covid-19, the procedure was shown to yield meaningful estimates of $R$, that thus provides a relevant assessment of the spatio-temporal evolution of the pandemic that can be involved as part of a decision strategy for designing Covid-19 counter measures. Indeed, at a current time, the procedure outputs estimates of $R$ for the entire pandemic period, thus permitting to assess a posteriori if and how the implementation of a sanitary policy decision (lockdown, curfew,...) impacted the evolution of the pandemic. Also and importantly, while the proposed procedure is not forecasting $R$, it is actually implementing a nowcasting approach: Not only estimates of $R$ today are provided, but the imposed piecewise smoothness constraints provide epidemiologist with a local trend around at current time indicating whether the pandemic is progressing or regressing. Spatial regularization also permits to robustly assess homogeneity or heterogeneity of the pandemic across related territories, thus permitting to decide on local or global (nation-wide) measures.

Estimates are updated automatically on a daily basis for 200+ countries worldwide and for several related territories (Counties in France, States in the USA). They are made publicly available via downloadable text files or interactive and animated maps. Procedures will be made publicly available.

Future investigations to improve estimation and monitoring include the construction of confidence intervals for the estimates and automated data-driven selection of the hyperparameters, using adaptive strategy that adjust the variations and phases of the pandemic.

We believe that the present work constitutes a significant step forward toward the practical use of the proposed procedure permits an actual real-time and on-the-fly monitoring of the Covid-19 pandemics. These tools are ready to be put at work immediately at the outbreak of potential future pandemics. Finally, the automated production of these daily up-dates is also intended as a scientific tool to favor trans-disciplinary scientific work against Covid-19 impact on the society at large.

REFERENCES


