Credibility intervals for Covid19 reproduction number from Nonsmooth Langevin-type Monte Carlo sampling

Gersende Fort
CNRS
Institut de Mathématiques de Toulouse, France

Maxwell Institute for Mathematical Sciences, Heriot-Watt University, Edinburgh, October 2022.

In collaboration with

- Patrice Abry, CNRS, Lab. de Physique de l'ENS Lyon, France
- Barbara Pascal, CNRS, LS2N, Nantes, France
- Nelly Pustelnik, CNRS, Lab. de Physique de l'ENS Lyon, France

Talk based on the papers:

- *Covid19 Reproduction Number: Credibility Intervals by Blockwise Proximal Monte Carlo samplers*
  by G. Fort, B. Pascal, P. Abry and N. Pustelnik
  HAL 03611079, submitted
- *Temporal evolution of the Covid19 pandemic reproduction number: Estimations from Proximal optimization to Monte Carlo sampling*
  by P. Abry, G. Fort, B. Pascal and N. Pustelnik
  EMBC 2022 (HAL 03565440)
- *Credibility intervals design for Covid19 reproduction number from nonsmooth Langevin-type Monte Carlo sampling*
  by H. Artigas, B. Pascal, G. Fort, P. Abry and N. Pustelnik
  EUSIPCO 2022 (HAL 03371837)
- *Estimation et intervalles de crédibilité pour le taux de reproduction de la Covid19 par échantillonnage Monte Carlo Langevin proximal*
  by P. Abry, G. Fort, B. Pascal and N. Pustelnik
  GRETSI 2022 (HAL 03611891)

Partly funded by
Fondation Simone et Cino Del Duca, Project OpSiMorE

I. Reproduction number of the Covid19
The model

Credibility intervals for the Reproduction number R, why ?

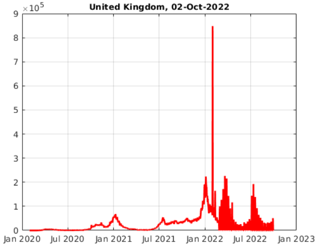
- Monitoring the Covid19 pandemic constitutes a critical societal stake: Covid19 pandemic caused/is causing unprecedented health, social, and economic crises.
- Need to assess the intensity of the / a pandemic, prerequisite for efficient sanitary policies.

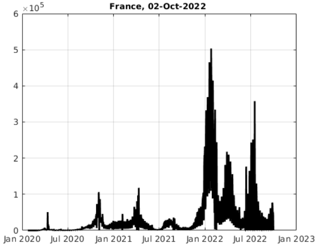
- The reproduction number measures
- the strength of the pandemic by quantifying rate of growth of daily new infections
- the number of second infections caused by one primary infection.

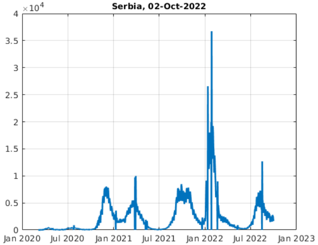
- Estimation of the daily R_t
- by a value of the index
- by credibility intervals: valuable information for the decision makers, notably in periods of rapid evolution or of changes in trends.

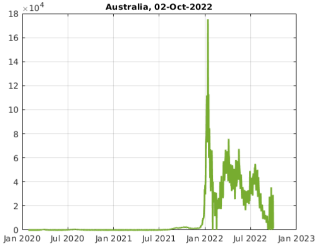
The data: daily new infections

- Real data, from Johns-Hopkins University repository
- Examples for UK, France, Serbia and Australia



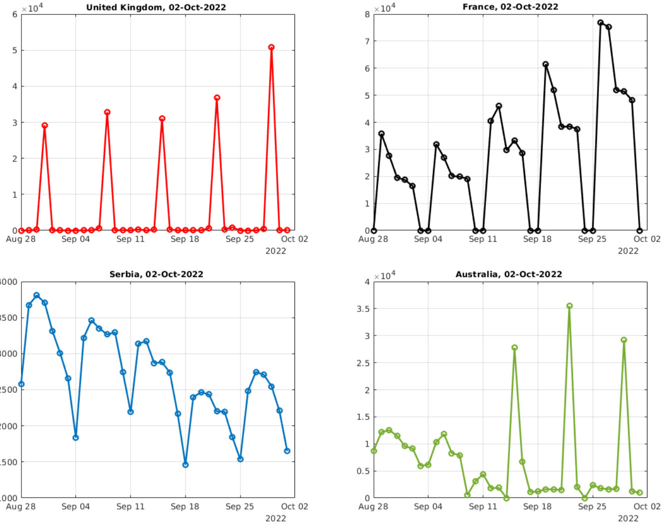






The data: daily new infections - zoom on the last 35 days

- Examples for UK, France, Serbia and Australia



The statistical model (1/4)

From Cori et al Cori et al (2013)

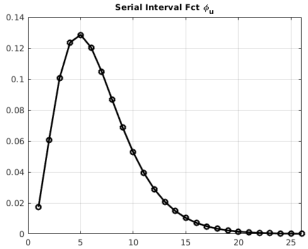
- The data $Z_1, \cdots, Z_T$: non negative integers
- Parameter: $(\mathsf{R}_1, \cdots, \mathsf{R}_T) \in (\mathbb{R}_+)^T$
- Conditionally to the past

$$Z_t \Big| Z_{1:(t-1)} \sim \mathcal{P}\left(\mathsf{R}_t \ \Phi_t^{\mathsf{Z}}\right) \qquad \text{where } \Phi_t^{\mathsf{Z}} := \sum_{u=1}^{\tau_\phi} \phi_u Z_{t-u}$$

- $\tau_\phi = 26$ days
- $\phi_u := \mathrm{PDF}_{\mathrm{Gamma}}(u)$

shape $= 1/0.28$, scale $= 1.87$

mean 6.68 days
std 3.53 days
mode 4.8 days



The statistical model (2/4)

From Abry et al Abry et al (2020)

- A priori distribution on the R_t's

- regularization needed as many parameters as observations
- piecewise linear time evolutions of $t \mapsto \mathsf{R}_t$

- L^1 penalization of the discrete time second derivative of $t \mapsto \mathsf{R}_t$

$$\ln \text{prior}: = -\lambda_{\mathsf{R}} \; \|\mathsf{D}_2 \, \mathsf{R}_{1:T}\|_1 \qquad \text{up to an additive constant}$$

where

$$\mathsf{D}_2 := \frac{1}{\sqrt{6}} \begin{bmatrix} 1 & -2 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & 0 & \cdots & 0 \\ \cdots & & & & & & \\ 0 & \cdots & & 0 & 1 & -2 & 1 \end{bmatrix} \in \mathbb{R}^{(T-2)\times T} \qquad \mathsf{R}_{1:T} := \begin{bmatrix} \mathsf{R}_1 \\ \mathsf{R}_2 \\ \cdots \\ \mathsf{R}_T \end{bmatrix}$$

- $\lambda_R := 3.5 \, \frac{\sqrt{6}}{4} \, \mathrm{std}(Z_1, \cdots, Z_T)$

The statistical model (3/4)

From Pascal et al Pascal et al (2021)

- The model from Cori et al + the regularized R_t's

- Model the errors on the counts via $\mathsf{O}_1, \cdots, \mathsf{O}_T$ in $\mathbb{R}^T$

- corrupted data, with pseudo-seasonalities, under-evaluations / over-evaluations ($\rightarrow$ negative counts)
- a priori distribution **and** modification of the likelihood

$$Z_t \Big| Z_{1:(t-1)} \sim \mathcal{P}\left(\mathsf{R}_t \ \Phi_t^{\mathsf{Z}} + \mathsf{O}_t\right) \qquad \text{where } \Phi_t^{\mathsf{Z}} := \sum_{u=1}^{\tau_\phi} \phi_u Z_{t-u}$$

$$\ln \text{prior:} = -\lambda_{\mathsf{R}} \ \|\mathsf{D}_2 \, \mathsf{R}_{1:T}\|_1 - \lambda_{\mathsf{O}} \ \|\mathsf{O}_{1:T}\|_1 \qquad \text{up to an additive constant}$$

- A constraint set

$$\mathcal{D} := \bigcap_t \{\mathsf{R}_{1:T}, \mathsf{O}_{1:T} \text{ s.t. positive intensity if } Z_t > 0 \text{ and non negative intensity if } Z_t \geq 0\}$$

- $\lambda_{\mathsf{O}} := 0.05$

The statistical model (4/4): hereafter

- A posteriori distribution of

$$\theta := (\mathsf{R}_{1:T}, \mathsf{O}_{1:T}) \in (\mathbb{R}_+)^T \times \mathbb{R}^T$$

given $Z_1, \cdots, Z_T$

- with log-density

$$\theta \mapsto \begin{cases} -f(\theta) - g(\mathsf{A}\,\theta) & \text{on } \mathcal{D} \\ -\infty & \text{otherwise} \end{cases} \qquad \mathsf{A} := \begin{bmatrix} \mathsf{D}_2 & 0_{(T-2)\times T} \\ 0_{T\times T} & \frac{\lambda_{\mathsf{O}}}{\lambda_{\mathsf{R}}} \mathrm{I}_T \end{bmatrix} \in \mathbb{R}^{(2T-2)\times(2T)}$$

where

$$f(\theta) := \sum_{t=1}^{T} \{(\mathsf{R}_t \Phi_t^{\mathsf{Z}} + \mathsf{O}_t) - Z_t \ln(\mathsf{R}_t \Phi_t^{\mathsf{Z}} + \mathsf{O}_t)\} \qquad g(\theta) := \lambda_{\mathsf{R}} \, \|\theta\|_1$$

II. Estimation of the R_t's via MCMC

Bayesian estimation

- Quantiles and other statistics Maximum a Posteriori, mean a posteriori, etc

- for each component R_t and O_t
- based on the marginal distributions of the distribution of $\theta := (R_{1:T}, O_{1:T})$

- Joint distribution of the form

$$-\ln \text{posterior:} \quad \theta \mapsto \begin{cases} f(\theta) + g(A\,\theta) & \text{on } \mathcal{D} \\ +\infty & \text{otherwise} \end{cases} \qquad A := \begin{bmatrix} D_2 & 0_{(T-2)\times T} \\ 0_{T\times T} & \frac{\lambda_O}{\lambda_R} I_T \end{bmatrix}$$

$$f(\theta) := \sum_{t=1}^{T} \{(R_t \Phi_t^Z + O_t) - Z_t \ln(R_t \Phi_t^Z + O_t)\} \qquad g(\theta) := \lambda_R \, \|\theta\|_1$$

Bayesian estimation via MCMC

- Quantiles and other statistics Maximum a Posteriori, mean a posteriori, etc

- for each component R_t and O_t
- based on the marginal distributions of the distribution of $\theta := (R_{1:T}, O_{1:T})$

- Joint distribution of the form

$$-\ln \text{posterior:} \quad \theta \mapsto \begin{cases} f(\theta) + g(A\,\theta) & \text{on } \mathcal{D} \\ +\infty & \text{otherwise} \end{cases} \qquad A := \begin{bmatrix} D_2 & 0_{(T-2)\times T} \\ 0_{T\times T} & \frac{\lambda_O}{\lambda_R} I_T \end{bmatrix}$$

Markov Chain Monte Carlo samplers

Bayesian estimation via MCMC

- Quantiles and other statistics Maximum a Posteriori, mean a posteriori, etc

- for each component R_t and O_t
- based on the marginal distributions of the distribution of $\theta := (\mathsf{R}_{1:T}, \mathsf{O}_{1:T})$

- Joint distribution of the form

$$-\ln \text{posterior:} \quad \theta \mapsto \begin{cases} f(\theta) + g(\mathsf{A}\,\theta) & \text{on } \mathcal{D} \\ +\infty & \text{otherwise} \end{cases} \qquad \mathsf{A} := \begin{bmatrix} \mathsf{D}_2 & 0_{(T-2)\times T} \\ 0_{T\times T} & \frac{\lambda_{\mathsf{O}}}{\lambda_{\mathsf{R}}} \mathrm{I}_T \end{bmatrix}$$

- f is C^1, convex
- $g := \lambda_R \|\cdot\|_1$ is lower semi-continuous, proper, convex
  - has a unique proximity operator, which is explicit
  - the proximity operator of $g(\mathsf{A}\cdot)$ is not explicit
- A is a $(2T-2) \times 2T$ matrix
  - full row rank
  - $g(\mathsf{A}\theta) = \sum_{i=1}^{3} g_i(\mathsf{A}_i\theta)$ and $g_i(\mathsf{A}_i\cdot)$ has an explicit proximity operator.

The Maximum a Posteriori estimator of the (R_t, O_t)'s (1/2)

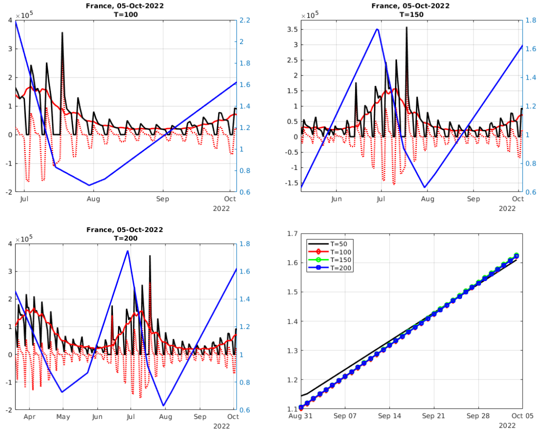
- Does it exist ? Unique ? Pascal et al (2022), Fort et al (2022)
  - If $\Phi_t^Z > 0$ and $\Phi_{t'}^Z$ for $t < t' \leq T$, and $Z_{t''} > 0$ then a MAP exists.
  - It two MAP, then: same Poisson intensity and same sign $O_t O'_t \geq 0; (D_2 R_t)\,(D_2 R'_t) \geq 0$.

- Computation: a Chambolle-Pock iterative algorithm proposed by Pascal et al (2022) see also Abry et al (2020)

- MAP for France, Serbia and Australia over the last 100 days:
(left y-axis) the data Z_t, (dots) $\hat{O}_t$ by MAP and (line) $Z_t - \hat{O}_t$
(right y-axis) $\hat{R}_t$ by MAP.

The Maximum a Posteriori estimator of the (R_t, O_t)'s (2/2)

- Role of T for the MAP estimate of the last R_t's:
  - [top,bottom left] The MAP estimate is computed from $T = 100$, $T = 150$ and $T = 200$ observations.
  - [bottom right] The three estimates for the last 35 days are displayed



Which MCMC sampler ?

$$-\ln \pi(\mathsf{R}_{1:T}, \mathsf{O}_{1:T}) = \sum_{t=1}^{T} \{(\mathsf{R}_t \Phi_t^{\mathsf{Z}} + \mathsf{O}_t) - Z_t \ln(\mathsf{R}_t \Phi_t^{\mathsf{Z}} + \mathsf{O}_t)\} + \lambda_{\mathsf{R}} \|\mathsf{A}\theta\|_1 \qquad \text{on } \mathcal{D}$$

$$\mathsf{A} := \begin{bmatrix} \mathrm{D}_2 & 0_{(T-2)\times T} \\ 0_{T\times T} & \frac{\lambda_{\mathsf{O}}}{\lambda_{\mathsf{R}}} \mathrm{I}_T \end{bmatrix}$$

- Hastings-Metropolis family
  - a Markov kernel for the update of $\theta := (\mathsf{R}_{1:T}, \mathsf{O}_{1:T})$
  - with a proposal mechanism using *first order* information on $\ln \pi$
- Gibbs family
  - update in turn $\mathsf{R}_{1:T}$ given $\mathsf{O}_{1:T}$, and then $\mathsf{O}_{1:T}$ given $\mathsf{R}_{1:T}$
  - the conditional distributions are not explicit → Metropolis-within-Gibbs methods

In both cases, faced with the design of Hastings-Metropolis sampler when the target distribution is of the form

$$-\ln \mathsf{target}(\tau) = f(\tau) + g(\mathsf{C}\tau) \qquad \tau \in \mathcal{D}$$

with C full row rank $\qquad f$ is $C^1 \qquad \mathrm{prox}_{\gamma\, g}$ exists and is unique

Which sampling space ?

$\pi(\tau) \propto \exp(-f(\tau) - g(\mathsf{C}\tau))\, 1_{\mathcal{D}}(\tau)$ where $\mathsf{C} \propto \mathrm{I}$ or C is a full row rank $(d-2) \times d$ matrix

- First strategy: Sample in the original space

- Obtain a Markov chain $\{\tau_n, n \geq 0\}$ with target π.

- Second strategy: Move to an *image space* and go back

$$\tilde{\tau} := \bar{\mathsf{C}}\,\tau \qquad \bar{\mathsf{C}} := \begin{bmatrix} \cdots \\ \cdots \\ \mathsf{C} \end{bmatrix} \text{invertible matrix,}$$

Image of π by $\bar{C}$:

$$\tilde{\pi}(\tilde{\tau}) \propto \exp(-f(\bar{\mathsf{C}}^{-1}\tilde{\tau}) - g(\tilde{\tau}_{3:d}))\, 1_{\mathcal{D}}(\bar{\mathsf{C}}^{-1}\tilde{\tau})$$

- Sample a Markov chain $\{\tilde{\tau}_n, n \geq 0\}$ with target distribution $\tilde{\pi}$

- Go back: $\tau_n := \bar{\mathsf{C}}^{-1}\,\tilde{\tau}_n$ is a Markov chain with target distribution π.

III. Proximal-Langevin based proposal distributions

$$-\ln \mathsf{target}(\tau) = f(\tau) + g(\mathsf{C}\tau) \qquad \tau \in \mathcal{D} \subseteq \mathbb{R}^d$$

with

- f is C^1
- $\mathrm{prox}_{\gamma g}$ exists, is unique and is explicit
- C is an **invertible** $d \times d$ matrix

Case $\ln \pi$ smooth: Langevin dynamic

- **Langevin proposal** Roberts and Tweedie (1996); scaling in Roberts and Rosenthal (2002)

$$\tau_{n+\frac{1}{2}} = \tau_n + \gamma_{n+1} \nabla \ln \pi(\tau_n) + \sqrt{2\gamma_{n+1}}\, \mathcal{N}(0, \mathrm{I})$$

- Tempered Langevin proposal Kent (1978), Roberts and Stramer (2002)

$$\tau_{n+\frac{1}{2}} = \tau_n + \gamma_{n+1}\, \Gamma\, \nabla \ln \pi(\tau_n) + \sqrt{2\gamma_{n+1}}\, \sqrt{\Gamma}\, \mathcal{N}(0, \mathrm{I})$$

Case $\ln \pi$ composite: a proximal-gradient approach $\rightarrow$ `PGdual`

Via the "move and go back" approach, we propose a proximal-gradient step

- From the result in convex optimization:

$$\tau_\star = \operatorname{argmin}(f(\mathsf{C}^{-1}\cdot) + g) \iff 0 \in \mathsf{C}^{-\top}\nabla f(\mathsf{C}^{-1}\tau_\star) + \partial g(\tau_\star)$$
$$\iff \operatorname{prox}_{\gamma g}(\tau_\star - \gamma\,\mathsf{C}^{-\top}\nabla f(\mathsf{C}^{-1}\tau_\star)) = \tau_\star \qquad \forall \gamma > 0.$$

- The proposal for sampling $\tilde{\pi}$

$$\tilde{\tau}_{n+\frac{1}{2}} = \operatorname{prox}_{\gamma g}(\tilde{\tau}_n - \gamma\,\mathsf{C}^{-\top}\nabla f(\mathsf{C}^{-1}\tilde{\tau}_n)) + \sqrt{2\gamma}\,\mathcal{N}(0, \mathrm{I})$$

- The proposal for sampling π: $\tau_{n+\frac{1}{2}} := \mathsf{C}^{-1}\tilde{\tau}_{n+\frac{1}{2}}$

Interpretation:

Tempered Langevin + variable metric Proximal = Variable Metric Proximal-Gradient

$$\tau_{n+\frac{1}{2}} = \operatorname{prox}^{\mathsf{C}^\top\mathsf{C}}_{\gamma g(\mathsf{C}\cdot)}(\tau_n - \gamma\,\mathsf{C}^{-1}\mathsf{C}^{-\top}\nabla f(\tau_n)) + \sqrt{2\gamma}\,\mathsf{C}^{-1}\mathcal{N}(0, \mathrm{I})$$

Case $\ln \pi$ composite: a Proximal-Gradient approach → PGdec

We propose a Proximal-Gradient (PG) step, when

- block-splitting $g(\mathsf{C}\tau) = \sum_{i=1}^{I} g_i(\mathsf{C}_i\tau)$
- explicit proximity operator $\mathrm{prox}_{\gamma\, g_i(\mathsf{C}_i \cdot)}$

- The proposal for sampling π:

(a) Sample uniformly $i \in \{1, \cdots, I\}$

(b) Update via a PG step which uses the component $\#i$ of the non-smooth fct

Langevin + Proximal = Proximal-Gradient

$$\tau_{n+\frac{1}{2}} = \mathrm{prox}_{\gamma\, g_i(\mathsf{C}_i \cdot)} \left(\tau_n - \gamma \nabla f(\tau_n)\right) + \sqrt{2\gamma}\, \mathcal{N}(0, \mathrm{I})$$

Case $\ln \pi$ composite: a Moreau envelope approach $\rightarrow$ MYULA Durmus et al (2018)

Via the "move and go back" approach, combined with MYULA

- From non-smooth convex optimization, the Moreau envelope

$$g_\rho(\tilde{\tau}) := \min_{\mathbb{R}^d} \left(\rho g(\cdot) + \frac{1}{2} \| \cdot - \tilde{\tau} \|^2 \right) \qquad \nabla g_\rho(\tilde{\tau}) = \frac{1}{\rho} \left(\tilde{\tau} - \mathrm{prox}_{\rho\, g}(\tilde{\tau}) \right)$$

- The proposal for sampling $\tilde{\pi}$

$$\tilde{\tau}_{n+\frac{1}{2}} = \tilde{\tau}_n - \gamma\, \mathsf{C}^{-\top} \nabla f(\mathsf{C}^{-1}\, \tilde{\tau}_n) - \frac{\gamma}{\rho} \left(\tilde{\tau}_n - \mathrm{prox}_{\rho\, g}(\tilde{\tau}_n) \right) + \sqrt{2\gamma}\, \mathcal{N}(0, \mathrm{I})$$

- The proposal for sampling π: $\quad \tau_{n+\frac{1}{2}} := \mathsf{C}^{-1} \tilde{\tau}_{n+\frac{1}{2}}$

Interpretation:

Tempered Langevin + variable metric Moreau envelope see e.g. Hiriart-Urruty and Lemaréchal (1996, Chapter XV)

$$\tau_{n+\frac{1}{2}} = \left(1 - \frac{\gamma}{\rho}\right) \tau_n - \gamma\, \mathsf{C}^{-1} \mathsf{C}^{-\top} \nabla f(\tau_n)) + \frac{\gamma}{\rho} \mathrm{prox}^{\mathsf{C}^\top \mathsf{C}}_{\rho\, g(\mathsf{C} \cdot)}(\tau_n) + \sqrt{2\gamma}\, \mathsf{C}^{-1}\, \mathcal{N}(0, \mathrm{I})$$

Case $\ln \pi$ composite: a Moreau envelope $\longrightarrow$ `semi-FBLMC` Luu et al (2021)

- Since $\mathrm{prox}_{\gamma g(\mathsf{C}\cdot)}$ not explicit:

- Do not use the Moreau envelope of $\tau \mapsto g(\mathsf{C}\tau)$

- Use the Moreau envelope of g, evaluated at $\mathsf{C}\tau$

$$\nabla[g_\rho(\mathsf{C}\cdot)](\tau) = \mathsf{C}^\top \nabla[g_\rho](\mathsf{C}\tau) = \frac{1}{\rho}\mathsf{C}^\top \left(\mathsf{C}\tau - \mathrm{prox}_{\rho g}(\mathsf{C}\tau)\right)$$

- The proposal for sampling π:

$$\tau_{n+\frac{1}{2}} = \left(\mathrm{I} - \frac{\gamma}{\rho}\mathsf{C}^\top\mathsf{C}\right)\tau_n - \gamma\,\nabla f(\tau_n) + \frac{\gamma}{\rho}\mathsf{C}^\top \mathrm{prox}_{\rho g}(\mathsf{C}\tau) + \sqrt{2\gamma}\,\mathcal{N}(0,\mathrm{I})$$

To summarize

- $-\ln \pi = f$ i.e. $\ln \pi$ is smooth

Algorithm	–	–	Drift	Cov	hyperparam
Langevin			$\tau - \gamma \nabla f(\tau)$	$2\gamma \mathrm{I}$	γ
Tempered Langevin			$\tau - \gamma\Gamma \, \nabla f(\tau)$	$2\gamma\Gamma$	γ, Γ

- $-\ln \pi = f + g(\mathsf{C}\cdot)$ i.e. $\ln \pi$ is non-smooth $\qquad g(\mathsf{C}\cdot) = \sum_{i=1}^{I} g_i(\mathsf{C}_i\cdot)$

Set $\quad \Gamma := \mathsf{C}^{-1}\mathsf{C}^{-\top}$

Algorithm	original space	move and go back	Drift	Cov	hyperparam
`PGdec`	✓		$\mathrm{prox}_{\gamma\, g_i(\mathsf{C}_i\cdot)}\left(\tau - \gamma \nabla f(\tau)\right)$	$2\gamma \mathrm{I}$	γ
`PGdual`		✓	$\mathrm{prox}^{\Gamma^{-1}}_{\gamma\, g(\mathsf{C}\cdot)}\left(\tau - \gamma\Gamma \, \nabla f(\tau)\right)$	$2\gamma\Gamma$	γ
`semi-FBLMC`	✓		$\tau - \gamma \nabla f(\tau) - \frac{\gamma}{\rho}\mathsf{C}^\top\left(\mathsf{C}\tau - \mathrm{prox}_{\rho\, g}(\mathsf{C}\tau)\right)$	$2\gamma \mathrm{I}$	γ, ρ
`MYULA`		✓	$\tau - \gamma\Gamma \, \nabla f(\tau) - \frac{\gamma}{\rho}\left(\tau - \mathrm{prox}^{\Gamma^{-1}}_{\rho\, g(\mathsf{C}\cdot)}(\tau)\right)$	$2\gamma\Gamma$	γ, ρ

To summarize

- $-\ln \pi = f$ i.e. $\ln \pi$ is smooth

Algorithm	–	–	Drift	Cov	hyperparam
Langevin			$\tau - \gamma \nabla f(\tau)$	$2\gamma \mathrm{I}$	γ
Tempered Langevin			$\tau - \gamma \Gamma \nabla f(\tau)$	$2\gamma \Gamma$	γ, Γ

- $-\ln \pi = f + g(\mathsf{C}\cdot)$ i.e. $\ln \pi$ is non-smooth $\qquad g(\mathsf{C}\cdot) = \sum_{i=1}^{I} g_i(\mathsf{C}_i\cdot)$

Set $\qquad \Gamma := \mathsf{C}^{-1}\mathsf{C}^{-\top}$

Algorithm	original space	move and go back	Drift	Cov	hyperparam
`PGdec`	✓		$\mathrm{prox}_{\gamma\, g_i(\mathsf{C}_i\cdot)}\,(\tau - \gamma \nabla f(\tau))$	$2\gamma \mathrm{I}$	γ
`PGdual`		✓	$\mathrm{prox}^{\Gamma^{-1}}_{\gamma\, g(\mathsf{C}\cdot)}\,(\tau - \gamma \Gamma \nabla f(\tau))$	$2\gamma \Gamma$	γ
`semi-FBLMC`	✓		$\tau - \gamma \nabla f(\tau)$ $- \frac{\gamma}{\rho} \mathsf{C}^\top \left(\mathsf{C}\tau - \mathrm{prox}_{\rho\, g}(\mathsf{C}\tau)\right)$	$2\gamma \mathrm{I}$	γ, ρ
`MYULA`		✓	$\tau - \gamma \Gamma \nabla f(\tau)$ $- \frac{\gamma}{\rho} \left(\tau - \mathrm{prox}^{\Gamma^{-1}}_{\rho\, g(\mathsf{C}\cdot)}(\tau)\right)$	$2\gamma \Gamma$	γ, ρ

To summarize

- $-\ln \pi = f$ i.e. $\ln \pi$ is smooth

Algorithm	–	–	Drift	Cov	hyperparam
Langevin			$\tau - \gamma \nabla f(\tau)$	$2\gamma \mathrm{I}$	γ
Tempered Langevin			$\tau - \gamma \Gamma \nabla f(\tau)$	$2\gamma\Gamma$	γ, Γ

- $-\ln \pi = f + g(\mathsf{C}\cdot)$ i.e. $\ln \pi$ is non-smooth $\qquad g(\mathsf{C}\cdot) = \sum_{i=1}^{I} g_i(\mathsf{C}_i \cdot)$

Set $\quad \Gamma := \mathsf{C}^{-1}\mathsf{C}^{-\top}$

Algorithm	original space	move and go back	Drift	Cov	hyperparam
`PGdec`	✓		$\operatorname{prox}_{\gamma\, g_i(\mathsf{C}_i \cdot)}(\tau - \gamma \nabla f(\tau))$	$2\gamma \mathrm{I}$	γ
`PGdual`		✓	$\operatorname{prox}^{\Gamma^{-1}}_{\gamma\, g(\mathsf{C}\cdot)}(\tau - \gamma \Gamma \nabla f(\tau))$	$2\gamma\Gamma$	γ
`semi-FBLMC`	✓		$\tau - \gamma \nabla f(\tau) - \frac{\gamma}{\rho} \mathsf{C}^\top \left(\mathsf{C}\tau - \operatorname{prox}_{\rho\, g}(\mathsf{C}\tau)\right)$	$2\gamma \mathrm{I}$	γ, ρ
`MYULA`		✓	$\tau - \gamma \Gamma \nabla f(\tau) - \frac{\gamma}{\rho}\left(\tau - \operatorname{prox}^{\Gamma^{-1}}_{\rho\, g(\mathsf{C}\cdot)}(\tau)\right)$	$2\gamma\Gamma$	γ, ρ

Propose and then: with or without an accept-reject step ?

- Without an AR step → `Langevin MC` Parisi (1981); `ULA` Roberts and Tweedie (1996); `MYULA` Durmus et al (2018)

$$\tau_{n+1} = \tau_{n+\frac{1}{2}}$$

- With an AR step → `MALA` Roberts and Tweedie (1996); Bou-Rabee and Hairer (2012); Eberle (2014); Dwivedi et al (2019); `MYMALA` Durmus et al (2018)

$$\tau_{n+1} = \tau_{n+\frac{1}{2}} \quad \text{or} \quad \tau_{n+1} = \tau_n$$

- In our case: **with an AR step**
  - the support of π is $\mathcal{D}$. This "constraint" is managed by the AR step.
  - Better mixing time in TV, as a function of d and the tolerance ϵ for `MALA` vs `ULA` Dwivedi et al (2019)

Comparison on a toy example (1/2)

- Logistic regression:
  - $\mathsf{Y} \in \mathbb{R}^T$ collects the T $\{0,1\}$-valued response variables.
  - X is the $T \times d$ covariate matrix.
  - $\theta \in \mathbb{R}^d$: unknown regression vector.
  - a priori distribution: piecewise constant

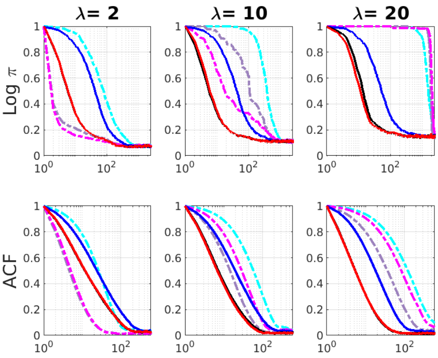
- A posteriori distribution, on $\mathbb{R}^d$

$$-\ln \pi(\theta) := -\mathsf{Y}^\top \mathsf{X}\theta + \sum_{t=1}^{T} \ln\left(1 + \exp((\mathsf{X}\theta)_t)\right) + \lambda \|\mathsf{D}_1\theta\|_1$$ up to an additive constant

- $T = 2\,000$; $d = 20$;

X: independent Rademacher and norm of the rows equal to one;
$\theta_\star$: piecewise constant
Y: independent Bernoulli with success probability $(1 + \exp(-(\mathsf{X}\theta_\star)_t))^{-1}$

Comparison on a toy example (2/2)



RW semi-FBLMC PGdec
RW "move/go back" MYULA PGdual

[top] Evolution $n \mapsto \frac{\ln \pi(\theta_n) - \max \ln \pi}{\ln \pi(\theta_1) - \max \ln \pi}$ along the first $2\,500$ samples.
[bottom] the ACF function, computed from $17\,500$ samples.

Show

- the benefit of first order informations on $\ln \pi$, mainly for small and medium λ

- PGdec < semi-FBLMC in the stationary phase

- PGdec: partial informations OK, when λ small

- PGdual and MYULA: robust to λ; and almost equivalent

- PGdual and MYULA: the best when λ medium to large.

Convergence analysis of PGdec and PGdual

Law of large numbers Proposition 13, Fort et al (2022)

Assume that π *is continuous on* $\mathcal{D}$; f *is continuously differentiable on* $\mathcal{D}$ *and* g *is convex, proper.*

PGdec. *Assume that* $\mathrm{prox}_{\gamma\, g_i(\mathsf{C}_i\cdot)}$ *exists and has a closed form expression.*
The sequence $\{\tau_n, n \geq 0\}$ *given by* `PGDec` *is a Markov chain, taking values on* $\mathcal{D}$ *and with unique invariant distribution* π. *In addition, for all initial point* $\tau_0 \in \mathcal{D}$ *and any measurable function* h *such that* $\int |h(\tau)|\, \pi(\tau) \mathrm{d}\tau < \infty$

$$\lim_{N \to +\infty} \frac{1}{N} \sum_{n=1}^{N} h(\tau_n) = \int h(\tau)\, \pi(\tau) \mathrm{d}\tau \qquad \textit{a.s.}$$

PGdual. *Assume that* $\mathrm{prox}_{\gamma\, g}$ *exists and has a closed form expression.*
The same ergodicty result holds for the sequence $\{\tau_n, n \geq 0\}$ *given by* `PGdual`.

Proof: under the stated assumptions, the chain is irreducible, aperiodic and π is its unique invariant distribution. In addition, the chain is positive Harris-recurrent.

Case log-π composite: other MCMC samplers in the literature

- `Perturbed Langevin MC`: bases on Gaussian smoothing of g (Chatterji et al (2020))

$$g_\mu(\cdot) := \mathbb{E}\left[g(\cdot + \mu \mathcal{N}(0, \mathrm{I}))\right]$$

When g is Hölder-continuous.

- AXDA (Vono et al (2019))
  - From the result

$$\min_{\tau} \left(f(\tau) + g(\mathsf{C}\tau)\right) \Longleftrightarrow \min_{\tau=\tau'} \left(f(\tau) + g(\mathsf{C}\tau')\right)$$

  - Data Augmentation scheme + Metropolis-within-Gibbs

$$-\ln \pi_{\mathrm{da}}(\tau, \tau') \propto f(\tau) + g(\mathsf{C}\tau') + \phi(\tau, \tau')$$

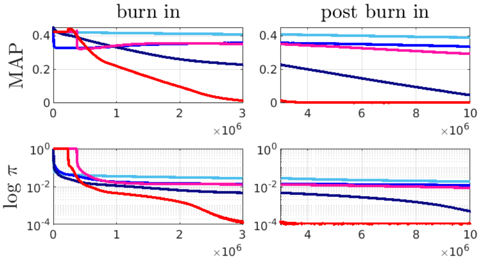
IV. Credibility intervals for the Covid19 reproduction number

Design parameters

$$-\ln \pi(\theta) = f(\theta) + \lambda_{\mathsf{R}} \|\mathsf{A}\theta\|_1 \qquad \mathsf{A} := \begin{bmatrix} \mathsf{D}_2 & 0_{(T-2)\times T} \\ 0_{T\times T} & \frac{\lambda_{\mathsf{O}}}{\lambda_{\mathsf{R}}} \mathrm{I}_T \end{bmatrix} \in \mathbb{R}^{(2T-2)\times(2T)}$$

- Augmentation of D_2 into a $T \times T$ invertible matrix: we compare two strategies "invert" and "ortho"
- Step size γ_{n+1} adapted during the first iterations (burn-in) and then fixed.
- Estimation of the quantiles: empirical quantiles, from the output of the Markov chain after a burn-in period.
- For the Moreau envelope: $\rho = \gamma$.

From D_2 to $\bar{D}_2$ (1/2)



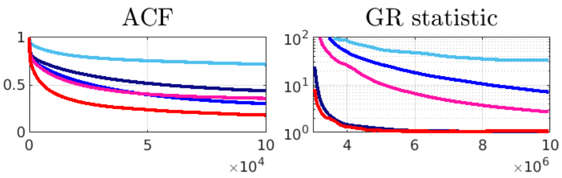
RW RW Invert RW Ortho
PGdual Invert PGdual Ortho.

During the burn in period [left] and after [right],
(top) evolution of the distance from θ_n to the MAP along iterations
(bottom) evolution of the distance from $\ln \pi(\theta_n)$ to $\max \ln \pi$ along iterations

Show

- Among the RW's, more efficient when the Gaussian noise is not i.i.d.
- $\bar{D}_2^{(o)}$ more efficient for RW's and PGdual
- PGdual more efficient than the RW's

From D_2 to $\bar{D}_2$ (2/2)



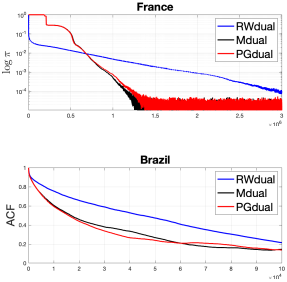
RW RW Invert RW Ortho
PGdual Invert PGdual Ortho.

(left) Mean absolute value of the ACF vs the first 10^5 lags
(right) The Gelman-Rubin statistic vs iterations.

Show

- same conclusions as before
it promotes PGdual "ortho"

PGdual and MYMALA (1/2)



Show

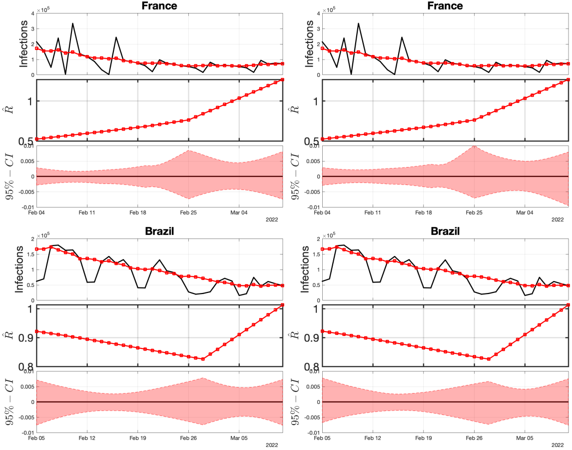
- the benefit of first-order methods

- equivalent results for PGdual and MYMALA

RW "ortho" when the Gaussian noise is no iid
MYMALA "ortho"
PGdual "ortho"

(top) $n \mapsto (\log \pi(\theta_n) - \max \log \pi)/(\log \pi(\theta_1) - \max \log \pi)$
(bottom) ACF

PGdual and MYMALA (2/2)

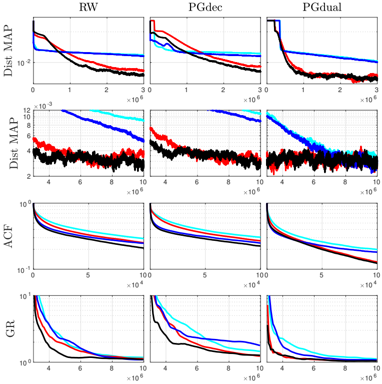


Show

- Same conclusions as before:

equivalent results for PGdual and MYMALA.

Gibbs or Hastings-Metropolis ?



HM "Invert" HM "Ortho"
Gibbs "Invert" Gibbs "Ortho"

Show

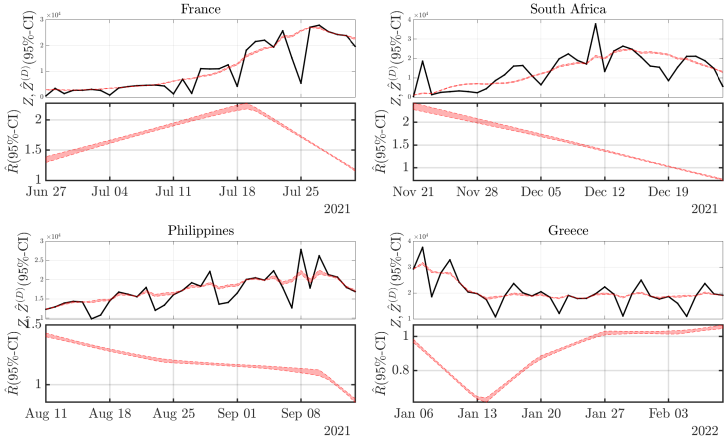
- the benefit of "ortho" w.r.t. "invert"

- Gibbs "ortho" among the best, whatever the criterion and the algorithm

- `PGdual` looks the best.

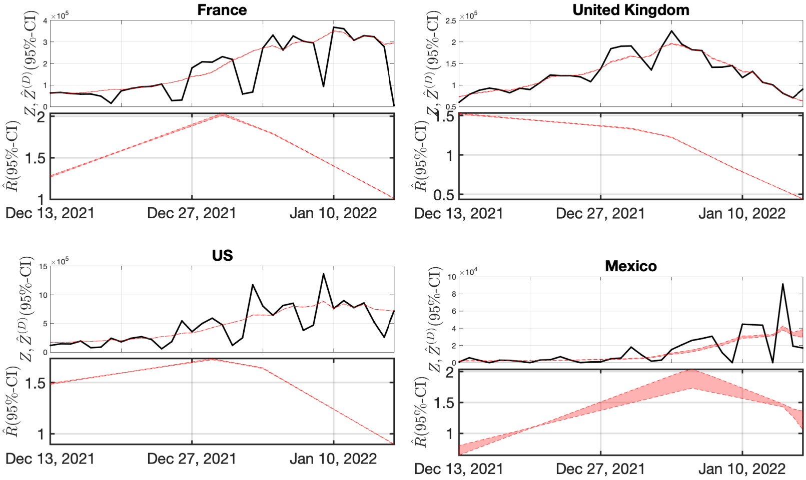
(rows 1 & 2) evolution of the distance from θ_n to the MAP along iterations
(row 3) ACF
(row 4) Gelman Rubin statistics

Credibility intervals for different countries and different time periods



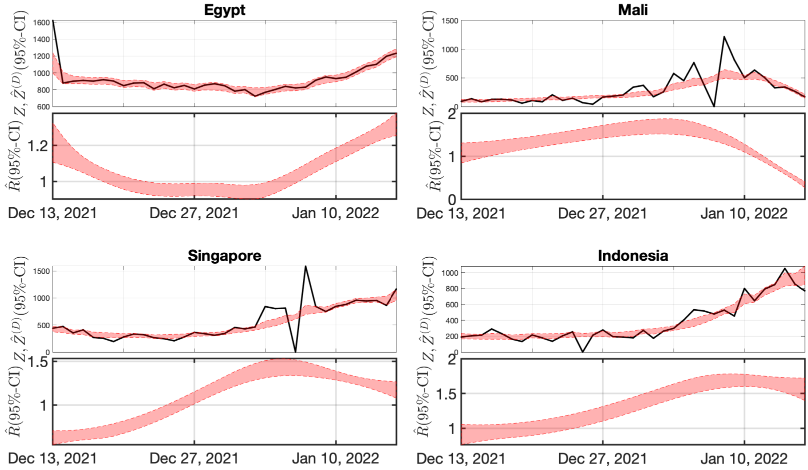
For each country,
(top) observed counts $Z_1, \cdots, Z_T$; and 95% credibility interval for the $Z_t - \hat{O}_t$'s
(bottom) 95% credibility interval for $\mathsf{R}_1, \cdots, \mathsf{R}_T$
With `PGdual` "ortho"

Credibility Intervals for different countries



With Gibbs + `PGdual` "ortho"

Credibility Intervals for different countries



With Gibbs + `PGdual` "ortho"

V. Conclusions and future works

- In Fort et al. (2022), extension of `PGdec` and `PGdual` to blockwise structure of the non-smooth component:

$$g(\mathsf{C}\theta) = \sum_{j=1}^{J} \sum_{i=1}^{I_j} g_{ij}(\mathsf{C}_{ij}\theta^{(j)}) \qquad \theta = \mathsf{blocks}(\theta^{(1)}, \cdots, \theta^{(J)})$$

- How to choose λ_R and λ_O ?
- What about sequential analysis $\rightarrow$ SMC
- Noisy data $\rightarrow$ noisy data and missing data.

VI. Bibliography

Bibliography

Abry et al., *Spatial and temporal regularization to estimate COVID-19 reproduction number $R(t)$: Promoting piecewise smoothness via convex optimization*, PLOSOne, 2020.

N. Bou-Rabee and M. Hairer. *Nonasymptotic mixing of the MALA algorithm*, IMA J. of Numerical Analysis, 2012.

Chatterji, N.S. and Diakonikolas, J. and Jordan, M.I. and Bartlett, P.L. *Langevin Monte Carlo without smoothness*, AISTATS, 2020.

A. Cori, N. M. Ferguson, C. Fraser and S. Cauchemez *A new framework and software to estimate time-varying reproduction numbers during epidemics* Am. J. Epide-miol., 2013.

Durmus, A. and Moulines, E. and Pereyra, M. *Efficient Bayesian computation by proximal Markovchain Monte Carlo: When Langevin meets Moreau* .SIAM J. Imaging Sci., 2018.

R. Dwivedi, Y. Chen, M.J. Wainwright and B. Yu *Log-concave sampling: Metropolis-Hastings algorithms are fast*, JMLR, 2019.

A. Eberle. *Error bounds for Metropolis-Hastings algorithms applied to perturbations of Gaussian measures in high dimensions*, Ann. of Appl. Probab., 2014.

G. Fort, B. Pascal, P. Abry, and N. Pustelnik, *Covid19 Reproduction Number : Credibility Intervals by Block-wise Proximal Monte Carlo Samplers*, hal-03611079v1, 2022.

T.D. Luu, J. Fadili, and C. Chesneau, *Sampling fromNon-smooth Distributions Through Langevin Diffusion*, Methodol Comput Appl Probab, 2020

Parisi, G. *Correlation functions and computer simulations* .Nucl. Phys. B, 1981.

B. Pascal, P. Abry, N. Pustelnik, S. Roux, R. Gribonval,and P. Flandrin, *Nonsmooth convex optimization to esti-mate the Covid-19 reproduction number space-time evo-lution with robustness against low quality data*, IEEE Trans Signal Processing, 2022.

G.O. Roberts and O. Stramer. *Langevin diffusions and Metropolis-Hastings algorithms*, Methodology and computing in applied probability, 2002.

G.O. Roberts and J.S. Rosenthal , 2002.

G.O. Roberts and R.L. Tweedie. *Exponential convergence of Langevin distributions and their discrete approximations.* Bernoulli, 1996.

Vono, M. and Dobigeon, N. and Chainais, P.. *Split-and-augmented Gibbs sampler - Application to large-scale inference problems.* IEEE Trans Signal Processing, 2019.