Proximal schemes for the estimation of the reproduction number of Covid19:

From convex optimization to Monte Carlo sampling

Variational approaches for signal and image processing

SiSyPhe day, ENS Lyon

November 18th 2022

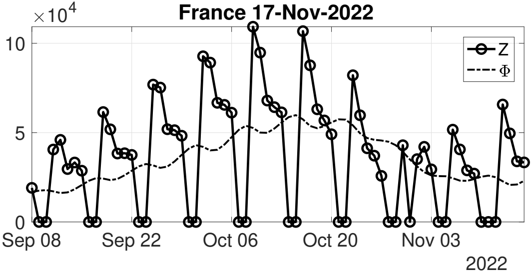
Barbara Pascal

Joint work with P. Abry, N. Pustelnik, S. Roux, R. Gribonval, P. Flandrin;
G. Fort, H. Artigas; Juliana Du

Outline

- Pandemic study: modeling at the service of monitoring
- Reproduction number estimation from minimization of penalized likelihood
- Bayesian framework for credibility interval estimation
- Conclusion & Perspectives

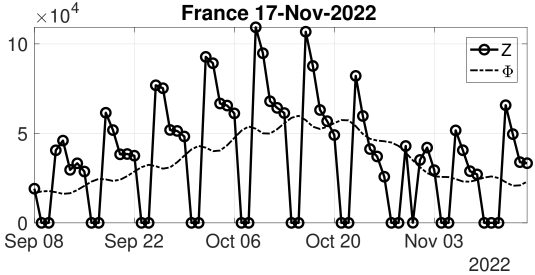
Counts of daily new infections



data from National Health Agencies collected by Johns Hopkins University

$\Longrightarrow$ number of cases not informative enough: need to capture the **dynamics**

Counts of daily new infections



data from National Health Agencies collected by Johns Hopkins University

$\Longrightarrow$ number of cases not informative enough: need to capture the **dynamics**

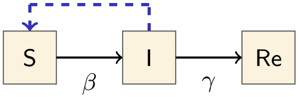
Design adapted counter measures and evaluate their effectiveness

$\longrightarrow$ efficient monitoring tools *epidemiological model,*

$\longrightarrow$ robust to low quality of the data *managing erroneous counts,*

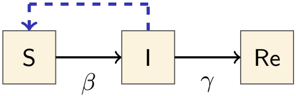
$\longrightarrow$ accompanied by reliable confidence level *credibility intervals.*

Susceptible-Infected-Recovered (SIR), among *compartmental models*



- ODE: $\frac{\mathrm{d}\mathsf{S}_t}{\mathrm{d}t} = -\beta \mathsf{S}_t \mathsf{I}_t$, $\frac{\mathrm{d}\mathsf{I}_t}{\mathrm{d}t} = \beta \mathsf{S}_t \mathsf{I}_t - \gamma \mathsf{I}_t$, $\frac{\mathrm{d}\mathsf{Re}_t}{\mathrm{d}t} = \gamma \mathsf{I}_t$
- Stochastic model: likelihood maximization to infer β, γ

Susceptible-Infected-Recovered (SIR), among *compartmental models*



– ODE: $\frac{\mathrm{d}\mathsf{S}_t}{\mathrm{d}t} = -\beta \mathsf{S}_t \mathsf{I}_t$, $\frac{\mathrm{d}\mathsf{I}_t}{\mathrm{d}t} = \beta \mathsf{S}_t \mathsf{I}_t - \gamma \mathsf{I}_t$, $\frac{\mathrm{d}\mathsf{Re}_t}{\mathrm{d}t} = \gamma \mathsf{I}_t$

– Stochastic model: likelihood maximization to infer β, γ

Limitations:

- refinement needed to get socially realistic model
- quadratic increase of the number of parameters
- Bayesian framework: heavy computational burden
- need consolidated and accurate datasets

✗ not adapted to real-time monitoring of Covid19 pandemic

Reproduction number in Cori model

"averaged number of secondary cases generated by a typical infectious individual"

(Cori et al., 2013, *Am. Journal of Epidemiology*; Liu et al., 2018, *PNAS*)

Reproduction number in Cori model

"averaged number of secondary cases generated by a typical infectious individual"

(Cori et al., 2013, *Am. Journal of Epidemiology*; Liu et al., 2018, *PNAS*)

Interpretation: at day t

$R_t > 1$ the virus propagates at exponential speed,

$R_t < 1$ the epidemic shrinks with an exponential decay,

$R_t = 1$ the epidemic is stable.

$\Longrightarrow$ one single indicator accounting for the overall pandemic mechanism

Reproduction number in Cori model

"averaged number of secondary cases generated by a typical infectious individual"

(Cori et al., 2013, *Am. Journal of Epidemiology*; Liu et al., 2018, *PNAS*)

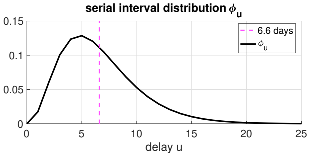
Interpretation: at day t

$R_t > 1$ the virus propagates at exponential speed,
$R_t < 1$ the epidemic shrinks with an exponential decay,
$R_t = 1$ the epidemic is stable.

$\Longrightarrow$ one single indicator accounting for the overall pandemic mechanism

Principle: Z_t new infections at day t

$$\mathbb{E}\left[Z_t\right] = R_t \Phi_t, \quad \Phi_t = \sum_{u=1}^{\tau_\Phi} \phi_u Z_{t-u}$$



with Φ_t global "infectiousness" in the population

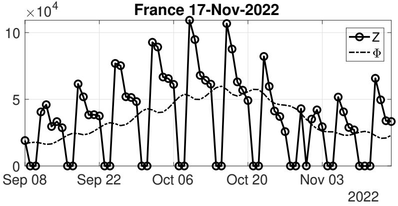
$\{\phi_u\}_{u=1}^{\tau_\Phi}$ distribution of delay between onset of symptoms in primary and secondary cases

Gamma distribution truncated at 25 days, of mean 6.6 days and standard deviation 3.5 days

Data: daily counts $\mathbf{Z} = (Z_1, \ldots, Z_T)$

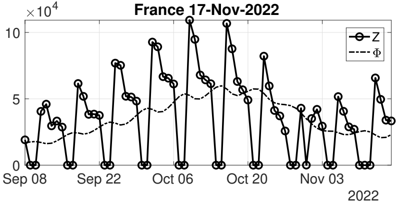
Model: Poisson distribution

$$\mathbb{P}(\mathsf{Z}_t | \mathbf{Z}_{t-\tau_\Phi:t-1}, \mathsf{R}_t) = \frac{(\mathsf{R}_t \Phi_t)^{\mathsf{Z}_t} \mathrm{e}^{-\mathsf{R}_t \Phi_t}}{\mathsf{Z}_t!}$$



Pandemic study: modeling at the service of monitoring

Data: daily counts $\mathbf{Z} = (Z_1, \ldots, Z_T)$



Model: Poisson distribution

$$\mathbb{P}(Z_t | \mathbf{Z}_{t-\tau_\Phi:t-1}, R_t) = \frac{(R_t \Phi_t)^{Z_t} e^{-R_t \Phi_t}}{Z_t!}$$

Maximum Likelihood Estimate (MLE)

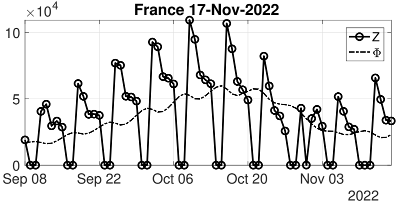
$$\begin{aligned}
&\ln\left(\mathbb{P}(Z_t | \mathbf{Z}_{t-\tau_\Phi:t-1}, R_t)\right) \\
&= Z_t \ln(R_t \Phi_t) - R_t \Phi_t - \ln(Z_t!) \\
&\underset{Z_t \gg 1}{\simeq} Z_t \ln(R_t \Phi_t) - R_t \Phi_t - Z_t \ln(Z_t) + Z_t \\
&\underset{\text{(def.)}}{=} -d_{KL}(Z_t | R_t \Phi_t) \quad \text{(Kullback-Leibler)}
\end{aligned}$$

Pandemic study: modeling at the service of monitoring

Data: daily counts $\mathbf{Z} = (Z_1, \ldots, Z_T)$

Model: Poisson distribution

$$\mathbb{P}(Z_t | \mathbf{Z}_{t-\tau_\Phi:t-1}, R_t) = \frac{(R_t \Phi_t)^{Z_t} e^{-R_t \Phi_t}}{Z_t!}$$



Maximum Likelihood Estimate (MLE)

$$\begin{aligned}
&\ln\left(\mathbb{P}(Z_t | \mathbf{Z}_{t-\tau_\Phi:t-1}, R_t)\right) \\
&= Z_t \ln(R_t \Phi_t) - R_t \Phi_t - \ln(Z_t!) \\
&\underset{Z_t \gg 1}{\simeq} Z_t \ln(R_t \Phi_t) - R_t \Phi_t - Z_t \ln(Z_t) + Z_t \\
&\underset{\text{(def.)}}{=} -d_{KL}(Z_t | R_t \Phi_t) \quad \text{(Kullback-Leibler)}
\end{aligned}$$

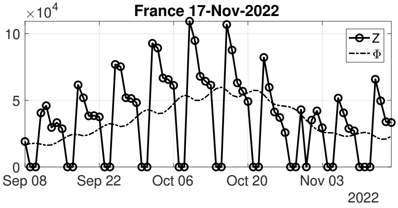
$$\Longrightarrow \widehat{R}_t^{\text{MLE}} = Z_t / \Phi_t = Z_t / \sum_{u=1}^{\tau_\Phi} \phi_u Z_{t-u}$$

ratio of moving averages

Data: daily counts $\mathbf{Z} = (Z_1, \ldots, Z_T)$

Model: Poisson distribution

$$\mathbb{P}(Z_t | \mathbf{Z}_{t-\tau_\Phi:t-1}, R_t) = \frac{(R_t \Phi_t)^{Z_t} e^{-R_t \Phi_t}}{Z_t!}$$



Maximum Likelihood Estimate (MLE)

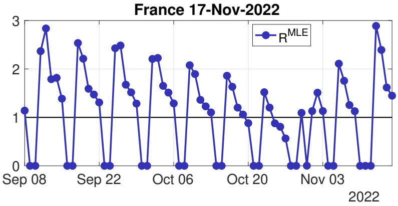
$$\ln\left(\mathbb{P}(Z_t | \mathbf{Z}_{t-\tau_\Phi:t-1}, R_t)\right) = Z_t \ln(R_t \Phi_t) - R_t \Phi_t - \ln(Z_t!)$$

$$\underset{Z_t \gg 1}{\simeq} Z_t \ln(R_t \Phi_t) - R_t \Phi_t - Z_t \ln(Z_t) + Z_t$$

$$\underset{\text{(def.)}}{=} -d_{KL}(Z_t | R_t \Phi_t) \quad \text{(Kullback-Leibler)}$$

$$\Longrightarrow \widehat{R}_t^{MLE} = Z_t / \Phi_t = Z_t / \sum_{u=1}^{\tau_\Phi} \phi_u Z_{t-u}$$

ratio of moving averages



- huge variability along time/ no local trend
- not robust to pseudo-periodicity/ misreported counts

Solution 0: (state-of-the-art) smoothing over a temporal window

$$\widehat{\mathrm{R}}_{t,s}^{\mathrm{MLE}}, \text{ with } s = 7 \text{ days}$$

(Cori et al., 2013, *Am. Journal of Epidemiology*)

$\Longrightarrow$ not able to detect rapid surge, nor fast decrease following sanitary restrictions

Solution 0: (state-of-the-art) smoothing over a temporal window

$$\widehat{\mathrm{R}}^{\mathrm{MLE}}_{t,s}, \text{ with } s = 7 \text{ days}$$

(Cori et al., 2013, *Am. Journal of Epidemiology*)

$\Longrightarrow$ not able to detect rapid surge, nor fast decrease following sanitary restrictions

---

Solution 1: regularization through nonlinear filtering

$$\widehat{\mathbf{R}}^{\mathrm{PKL}} = \underset{\mathbf{R}\in\mathbb{R}_+^T}{\operatorname{argmin}} \sum_{t=1}^{T} \mathrm{d}_{\mathrm{KL}}\left(\mathrm{Z}_t \,|\mathrm{R}_t\Phi_t\right) + \lambda_{\mathrm{R}}\mathcal{P}(\mathbf{R}) \quad \text{(penalized Kullback-Leibler)}$$

with $\mathcal{P}(\mathbf{R})$ favoring some temporal regularity (Abry et al., 2020, *PlosOne*)

Solution 0: (state-of-the-art) smoothing over a temporal window

$$\widehat{\mathrm{R}}^{\mathrm{MLE}}_{t,s}, \text{ with } s = 7 \text{ days}$$

(Cori et al., 2013, *Am. Journal of Epidemiology*)

$\Longrightarrow$ not able to detect rapid surge, nor fast decrease following sanitary restrictions

---

Solution 1: regularization through nonlinear filtering

$$\widehat{\mathbf{R}}^{\mathrm{PKL}} = \underset{\mathbf{R}\in\mathbb{R}_+^T}{\operatorname{argmin}} \sum_{t=1}^{T} \mathrm{d}_{\mathrm{KL}}\left(\mathrm{Z}_t \,|\mathrm{R}_t\Phi_t\right) + \lambda_{\mathrm{R}}\mathcal{P}(\mathbf{R}) \quad \text{(penalized Kullback-Leibler)}$$

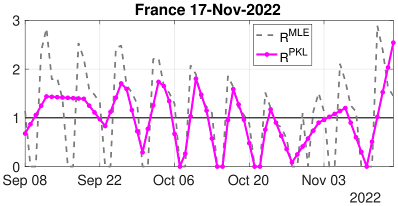
with $\mathcal{P}(\mathbf{R})$ favoring some temporal regularity (Abry et al., 2020, *PlosOne*)

$$\mathcal{P}(\mathbf{R}) = \|\mathbf{D}_2\mathbf{R}\|_1$$

$$(\mathbf{D}_2\mathbf{R})_t = R_{t+1} - 2R_t + R_{t-1}$$

2nd order derivative & ℓ_1-norm

$\Longrightarrow$ piecewise linearity

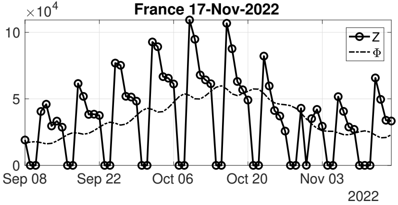


captures global **trend**, more **regular** than MLE, but **pseudo-oscillations**

New infection counts **Z** are corrupted by

- missing samples,
- non meaningful negative counts,
- retrospected cumulated counts,
- pseudo-seasonality effects.

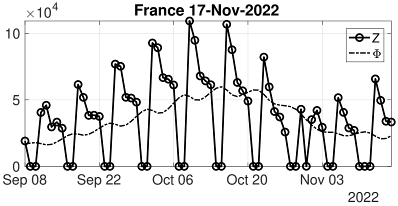
$\Longrightarrow$ full parametric modeling out of reach



New infection counts **Z** are corrupted by

- missing samples,
- non meaningful negative counts,
- retrospected cumulated counts,
- pseudo-seasonality effects.

$\Longrightarrow$ full parametric modeling out of reach



Solution 1': first correct **Z**, then apply penalized Kullback-Leibler on corrected $\mathbf{Z}^{(C)}$

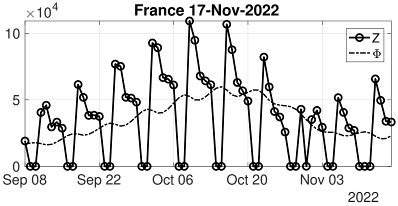
$\Longrightarrow$ two-step procedure not optimal: accumulates correction & regularization biases

(Abry et al., 2020, *Eng. Med. Biol. Conf.*)

New infection counts **Z** are corrupted by

- missing samples,
- non meaningful negative counts,
- retrospected cumulated counts,
- pseudo-seasonality effects.

$\Longrightarrow$ full parametric modeling out of reach



Solution 1': first correct **Z**, then apply penalized Kullback-Leibler on corrected $\mathbf{Z}^{(C)}$

$\Longrightarrow$ two-step procedure not optimal: accumulates correction & regularization biases

(Abry et al., 2020, *Eng. Med. Biol. Conf.*)

Solution 2: **one-step** procedure performing jointly

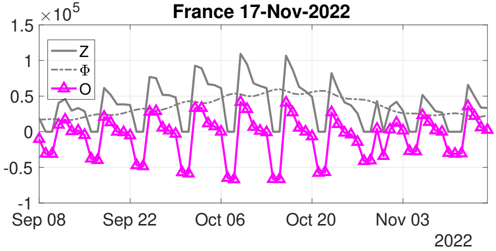
correction of corrupted Z_t & estimation of regularized R_t

(Pascal et al., 2022, *Trans. Sig. Process.*)

Extended Cori Model: additional latent variable O_t accounting for misreport

$$Z_t \sim \text{Poiss}\left(R_t \Phi_t + O_t\right), \quad R_t \Phi_t + O_t \geq 0$$

nonzero values of O_t concentrated on specific days (Sundays, day-offs, ...)



Interpretation:

$$\text{Poiss}\left(R_t \Phi_t + O_t\right) \sim \begin{cases} \text{Poiss}\left(R_t \Phi_t\right) + \text{Poiss}\left(O_t\right) & \text{if } O_t \geq 0, \\ \text{Poiss}\left(\alpha_t R_t \Phi_t\right), \; \alpha_t = 1 - \frac{-O_t}{R_t \Phi_t} \in [0, 1] & \text{if } O_t < 0. \end{cases}$$

Data: reported counts $\mathbf{Z} = (Z_1, \ldots, Z_T)$

Model: corrected Poisson $\mathbb{P}(Z_t | \mathbf{Z}_{t-\tau_\Phi:t-1}, R_t, O_t) = \dfrac{(R_t \Phi_t + O_t)^{Z_t} e^{-(R_t \Phi_t + O_t)}}{Z_t!}$

Data: reported counts $\mathbf{Z} = (Z_1, \ldots, Z_T)$

Model: corrected Poisson $\mathbb{P}(Z_t | \mathbf{Z}_{t-\tau_\Phi:t-1}, R_t, O_t) = \dfrac{(R_t\Phi_t + O_t)^{Z_t} e^{-(R_t\Phi_t + O_t)}}{Z_t!}$

Generalized Penalized Kullback-Leibler

$$(\widehat{\mathbf{R}}, \widehat{\mathbf{O}}) = \underset{(\mathbf{R},\mathbf{O}) \in \mathbb{R}_+^T \times \mathbb{R}^T}{\operatorname{argmin}} \sum_{t=1}^{T} d_{KL}\left(Z_t \mid R_t\Phi_t + O_t\right) + \lambda_R \|\mathbf{D}_2\mathbf{R}\|_1 + \iota_{\geq 0}(\mathbf{R}) + \lambda_O \|\mathbf{O}\|_1$$

$\Longrightarrow$ estimates piecewise linear, non-negative R_t and sparse O_t

Data: reported counts $\mathbf{Z} = (Z_1, \ldots, Z_T)$

Model: corrected Poisson $\quad \mathbb{P}(Z_t | \mathbf{Z}_{t-\tau_\Phi:t-1}, R_t, O_t) = \dfrac{(R_t\Phi_t + O_t)^{Z_t} e^{-(R_t\Phi_t + O_t)}}{Z_t!}$

Generalized Penalized Kullback-Leibler

$$(\widehat{\mathbf{R}}, \widehat{\mathbf{O}}) = \underset{(\mathbf{R},\mathbf{O}) \in \mathbb{R}_+^T \times \mathbb{R}^T}{\operatorname{argmin}} \sum_{t=1}^{T} d_{KL}\left(Z_t \,|\, R_t\Phi_t + O_t\right) + \lambda_R \|\mathbf{D}_2\mathbf{R}\|_1 + \iota_{\geq 0}(\mathbf{R}) + \lambda_O \|\mathbf{O}\|_1$$

$\Longrightarrow$ estimates piecewise linear, non-negative R_t and sparse O_t

properties of the objective function:

- sum of convex functions composed with linear operators $\Longrightarrow$ globally convex;
- feasible domain: $(\forall t, R_t \geq 0)$ & (if $Z_t > 0, R_t\Phi_t + O_t > 0$, else $R_t\Phi_t + O_t \geq 0$);
- $p_t \mapsto d_{KL}\left(Z_t \,|\, p_t\right)$ is strictly-convex.

Data: reported counts $\mathbf{Z} = (Z_1, \ldots, Z_T)$

Model: corrected Poisson $\quad \mathbb{P}(Z_t | \mathbf{Z}_{t-\tau_\Phi:t-1}, R_t, O_t) = \dfrac{(R_t \Phi_t + O_t)^{Z_t} e^{-(R_t \Phi_t + O_t)}}{Z_t!}$

Generalized Penalized Kullback-Leibler

$$(\widehat{\mathbf{R}}, \widehat{\mathbf{O}}) = \underset{(\mathbf{R},\mathbf{O}) \in \mathbb{R}_+^T \times \mathbb{R}^T}{\operatorname{argmin}} \sum_{t=1}^{T} d_{KL}\left(Z_t \,|\, R_t \Phi_t + O_t\right) + \lambda_R \|\mathbf{D}_2 \mathbf{R}\|_1 + \iota_{\geq 0}(\mathbf{R}) + \lambda_O \|\mathbf{O}\|_1$$

$\Longrightarrow$ estimates piecewise linear, non-negative R_t and sparse O_t

properties of the objective function:

- sum of convex functions composed with linear operators $\Longrightarrow$ globally convex;
- feasible domain: $(\forall t, R_t \geq 0)$ & (if $Z_t > 0, R_t \Phi_t + O_t > 0$, else $R_t \Phi_t + O_t \geq 0$);
- $p_t \mapsto d_{KL}\left(Z_t \,|\, p_t\right)$ is strictly-convex.

Theorem (Pascal et al., 2022, *Trans. Sig. Process.*)

+ The minimization problem has at least one solution $(\widehat{\mathbf{R}}, \widehat{\mathbf{O}})$.
+ The estimated time-varying Poisson intensity $\widehat{p}_t = \widehat{R}_t \Phi_t + \widehat{O}_t$ is unique.

$$\underset{(\mathbf{R},\mathbf{O})\in\mathbb{R}_+^T\times\mathbb{R}^T}{\text{minimize}} \sum_{t=1}^{T} \mathrm{d}_{\mathrm{KL}}\left(\mathrm{Z}_t \,|\, \mathrm{R}_t\Phi_t + \mathrm{O}_t\right) + \lambda_{\mathrm{R}}\|\mathbf{D}_2\mathbf{R}\|_1 + \iota_{\geq 0}(\mathbf{R}) + \lambda_{\mathrm{O}}\|\mathbf{O}\|_1$$

- each term of the functional is convex;
- ℓ_1-norm and indicative functions $\Longrightarrow$ nonsmooth; ✗ gradient descent
- gradient of $\mathrm{p}_t \mapsto \mathrm{d}_{\mathrm{KL}}\left(\mathrm{Z}_t \,|\, \mathrm{p}_t\right)$ is not Lipschitzian; ✗ forward-backward
- linear operator $\mathbf{D}_2 \Longrightarrow$ no explicit form for $\mathrm{prox}_{\|\mathbf{D}_2\cdot\|_1}$ ♣ need splitting

$$\underset{(\mathbf{R},\mathbf{O})\in\mathbb{R}_+^T\times\mathbb{R}^T}{\text{minimize}} \sum_{t=1}^{T} \mathrm{d}_{\mathrm{KL}}\left(\mathrm{Z}_t \,|\, \mathrm{R}_t\Phi_t + \mathrm{O}_t\right) + \lambda_{\mathrm{R}}\|\mathbf{D}_2\mathbf{R}\|_1 + \iota_{\geq 0}(\mathbf{R}) + \lambda_{\mathrm{O}}\|\mathbf{O}\|_1$$

- each term of the functional is convex;
- ℓ_1-norm and indicative functions $\Longrightarrow$ nonsmooth; ✗ gradient descent
- gradient of $\mathrm{p}_t \mapsto \mathrm{d}_{\mathrm{KL}}\left(\mathrm{Z}_t \,|\, \mathrm{p}_t\right)$ is not Lipschitzian; ✗ forward-backward
- linear operator $\mathbf{D}_2 \Longrightarrow$ no explicit form for $\mathrm{prox}_{\|\mathbf{D}_2\cdot\|_1}$ ♣ need splitting

$$\Longleftrightarrow \underset{(\mathbf{R},\mathbf{O})\in\mathbb{R}_+^T\times\mathbb{R}^T}{\text{minimize}} \quad f(\mathbf{R},\mathbf{O}|\mathbf{Z}) + h(\mathbf{A}(\mathbf{R},\mathbf{O})), \quad \mathbf{A} \text{ linear; } f, h \text{ proximable}$$

$$\mathbf{A}(\mathbf{R},\mathbf{O}) = (\lambda_{\mathrm{R}}\mathbf{D}_2\mathbf{R}, \mathbf{R}, \lambda_{\mathrm{O}}\mathbf{O}); \quad h(\mathbf{Q}_1,\mathbf{Q}_2,\mathbf{Q}_3) = \|\mathbf{Q}_1\|_1 + \iota_{\geq 0}(\mathbf{Q}_2) + \|\mathbf{Q}_3\|_1$$

$$\underset{(\mathbf{R},\mathbf{O})\in\mathbb{R}_+^T\times\mathbb{R}^T}{\text{minimize}} \sum_{t=1}^{T} \mathrm{d}_{\mathrm{KL}}\left(\mathrm{Z}_t \,|\, \mathrm{R}_t\Phi_t + \mathrm{O}_t\right) + \lambda_{\mathrm{R}}\|\mathbf{D}_2\mathbf{R}\|_1 + \iota_{\geq 0}(\mathbf{R}) + \lambda_{\mathrm{O}}\|\mathbf{O}\|_1$$

- each term of the functional is convex;
- ℓ_1-norm and indicative functions $\Longrightarrow$ nonsmooth; ✗ gradient descent
- gradient of $\mathrm{p}_t \mapsto \mathrm{d}_{\mathrm{KL}}\left(\mathrm{Z}_t \,|\, \mathrm{p}_t\right)$ is not Lipschitzian; ✗ forward-backward
- linear operator $\mathbf{D}_2 \Longrightarrow$ no explicit form for $\mathrm{prox}_{\|\mathbf{D}_2\cdot\|_1}$ ♣ need splitting

$$\Longleftrightarrow \underset{(\mathbf{R},\mathbf{O})\in\mathbb{R}_+^T\times\mathbb{R}^T}{\text{minimize}} \quad f(\mathbf{R},\mathbf{O}|\mathbf{Z}) + h(\mathbf{A}(\mathbf{R},\mathbf{O})), \quad \mathbf{A} \text{ linear; } f, h \text{ proximable}$$

$$\mathbf{A}(\mathbf{R},\mathbf{O}) = (\lambda_{\mathrm{R}}\mathbf{D}_2\mathbf{R}, \mathbf{R}, \lambda_{\mathrm{O}}\mathbf{O}); \quad h(\mathbf{Q}_1,\mathbf{Q}_2,\mathbf{Q}_3) = \|\mathbf{Q}_1\|_1 + \iota_{\geq 0}(\mathbf{Q}_2) + \|\mathbf{Q}_3\|_1$$

Primal-dual algorithm (Chambolle et al., 2011, *Int. Conf. Comput. Vis.*)

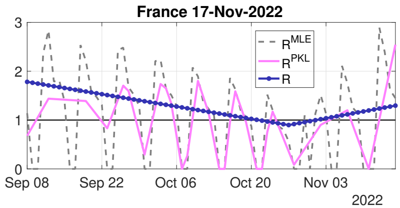
for $k = 1, 2 \ldots$ **do**

$$\mathbf{Q}^{[k+1]} = \mathrm{prox}_{\sigma h^*}(\mathbf{Q}^{[k]} + \sigma\mathbf{A}(\overline{\mathbf{R}}^{[k]}, \overline{\mathbf{O}}^{[k]})) \qquad \text{dual}$$

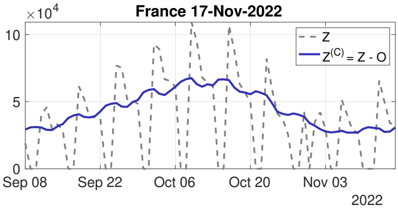
$$(\mathbf{R}^{[k+1]}, \mathbf{O}^{[k+1]}) = \mathrm{prox}_{\tau f(\cdot|\mathbf{Z})}((\mathbf{R}^{[k+1]}, \mathbf{O}^{[k+1]}) - \tau\mathbf{A}^*\mathbf{Q}^{[k+1]}) \qquad \text{primal}$$

$$(\overline{\mathbf{R}}^{[k+1]}, \overline{\mathbf{O}}^{[k+1]}) = 2(\mathbf{R}^{[k+1]}, \mathbf{O}^{[k+1]}) - (\mathbf{R}^{[k]}, \mathbf{O}^{[k]}) \qquad \text{auxiliary}$$

Reproduction number $\widehat{\mathbf{R}}$

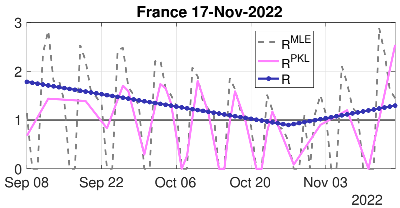


Corrected infection counts $\mathbf{Z}^{(C)}$

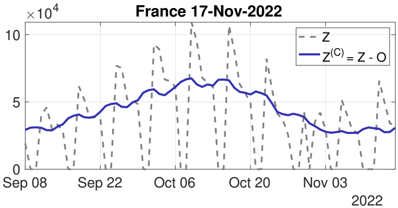


$\Longrightarrow$ no more pseudo-seasonality, local trends well captured, smooth behavior

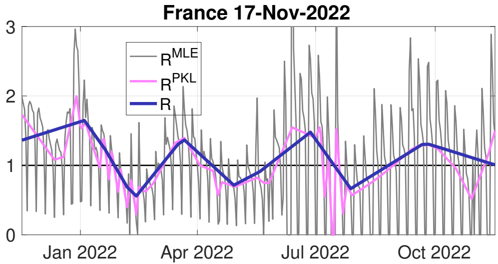
Reproduction number $\widehat{\mathbf{R}}$



Corrected infection counts $\mathbf{Z}^{(C)}$



$\Longrightarrow$ no more pseudo-seasonality, local trends well captured, smooth behavior



fast numerical scheme: 15 to 30 sec for 70 days to 1 year

New infection counts per county: $\mathbf{Z} = \left\{ \mathrm{Z}_t^{(d)},\, d \in [1, D],\, t \in [1,\, T] \right\}$

$\Rightarrow$ multivariate time-varying reproduction number $\mathrm{R}_t^{(d)}$

New infection counts per county: $\mathbf{Z} = \left\{ \mathsf{Z}_t^{(d)},\ d \in [1, D],\ t \in [1, T] \right\}$

$\Rightarrow$ multivariate time-varying reproduction number $\mathsf{R}_t^{(d)}$

Multivariate extended penalized Kullback-Leibler

$$\left(\widehat{\mathbf{R}}, \widehat{\mathbf{O}}\right) = \underset{(\mathbf{R},\mathbf{O}) \in \mathbb{R}_+^{D \times T} \times \mathbb{R}^{D \times T}}{\operatorname{argmin}} \sum_{d=1}^{D} \sum_{t=1}^{T} \mathsf{d}_{\mathsf{KL}} \left(\mathsf{Z}_t^{(d)} \,\middle|\, \mathsf{R}_t^{(d)} \Phi_t^{(d)} + \mathsf{O}_t^{(d)} \right) + \lambda_{\mathsf{R}} \|\mathbf{D}_2 \mathbf{R}\|_1 + \iota_{\geq 0}(\mathbf{R}) + \lambda_{\text{space}} \|\mathbf{G}\mathbf{R}\|_1 + \lambda_{\mathsf{O}} \|\mathbf{O}\|_1$$

$\Longrightarrow$ $\|\mathbf{G}\mathbf{R}\|_1$ favors **piecewise constancy** in space

New infection counts per county: $\mathbf{Z} = \left\{ \mathrm{Z}_t^{(d)},\ d \in [1, D],\ t \in [1, T] \right\}$

$\Rightarrow$ multivariate time-varying reproduction number $\mathrm{R}_t^{(d)}$

Multivariate extended penalized Kullback-Leibler

$$\left(\widehat{\mathbf{R}}, \widehat{\mathbf{O}}\right) = \underset{(\mathbf{R},\mathbf{O}) \in \mathbb{R}_+^{D\times T} \times \mathbb{R}^{D \times T}}{\operatorname{argmin}} \sum_{d=1}^{D} \sum_{t=1}^{T} \mathrm{d}_{\mathrm{KL}} \left(\mathrm{Z}_t^{(d)} \middle| \mathrm{R}_t^{(d)} \Phi_t^{(d)} + \mathrm{O}_t^{(d)} \right)$$
$$+ \lambda_{\mathrm{R}} \|\mathbf{D}_2 \mathbf{R}\|_1 + \iota_{\geq 0}(\mathbf{R}) + \lambda_{\text{space}} \|\mathbf{G}\mathbf{R}\|_1 + \lambda_{\mathrm{O}} \|\mathbf{O}\|_1$$

$\Longrightarrow \|\mathbf{G}\mathbf{R}\|_1$ favors **piecewise constancy** in space

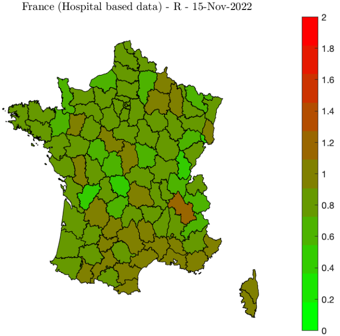
Graph Total Variation

$$\|\mathbf{G}\mathbf{R}\|_1 = \sum_{t=1}^{T} \sum_{d_1 \sim d_2} \left| \mathrm{R}_t^{(d_1)} - \mathrm{R}_t^{(d_2)} \right|$$

sum over neighboring counties

here: $d_1 \sim d_2 \Leftrightarrow$ share terrestrial border

$$\widetilde{\mathbf{A}}(\mathbf{R}, \mathbf{O}) = (\lambda_{\mathrm{R}} \mathbf{D}_2 \mathbf{R}, \mathbf{R}, \lambda_{\text{space}} \mathbf{G}\mathbf{R}, \lambda_{\mathrm{O}} \mathbf{O})$$



Bayesian framework for credibility interval estimation

<u>Pointwise estimate</u> of parameter $\boldsymbol{\theta} = (\mathbf{R}, \mathbf{O})$ from observations $\mathbf{Z}$

$$\underset{(\mathbf{R},\mathbf{O})\in\mathbb{R}_+^T\times\mathbb{R}^T}{\text{minimize}} \quad f(\boldsymbol{\theta}|\mathbf{Z}) + h(\mathbf{A}\boldsymbol{\theta})$$

(Pascal et al., 2022, *Trans. Sig. Process.*)

Pointwise estimate of parameter $\boldsymbol{\theta} = (\mathbf{R}, \mathbf{O})$ from observations $\mathbf{Z}$

$$\underset{(\mathbf{R},\mathbf{O})\in\mathbb{R}_+^T\times\mathbb{R}^T}{\text{minimize}} \quad f(\boldsymbol{\theta}|\mathbf{Z}) + h(\mathbf{A}\boldsymbol{\theta})$$ (Pascal et al., 2022, *Trans. Sig. Process.*)

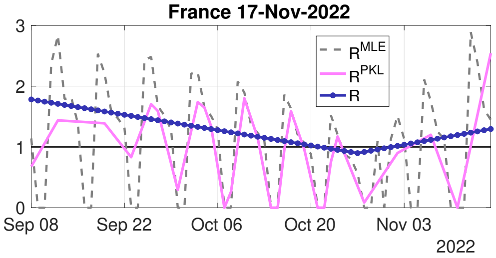
Q: what is the value of R today? **A:** solve the minimization problem and output $\widehat{\mathrm{R}}_T$.

Bayesian framework for credibility interval estimation

Pointwise estimate of parameter $\boldsymbol{\theta} = (\mathbf{R}, \mathbf{O})$ from observations $\mathbf{Z}$

$$\underset{(\mathbf{R},\mathbf{O})\in\mathbb{R}_+^T\times\mathbb{R}^T}{\text{minimize}} \quad f(\boldsymbol{\theta}|\mathbf{Z}) + h(\mathbf{A}\boldsymbol{\theta}) \quad \text{(Pascal et al., 2022, \textit{Trans. Sig. Process.})}$$

Q: what is the value of R today? **A:** solve the minimization problem and output $\widehat{\mathrm{R}}_T$.



$\widehat{\mathrm{R}}_T = 1.2955$

Bayesian framework for credibility interval estimation

Pointwise estimate of parameter $\boldsymbol{\theta} = (\mathbf{R}, \mathbf{O})$ from observations $\mathbf{Z}$

$$\underset{(\mathbf{R},\mathbf{O}) \in \mathbb{R}_+^T \times \mathbb{R}^T}{\text{minimize}} \quad f(\boldsymbol{\theta}|\mathbf{Z}) + h(\mathbf{A}\boldsymbol{\theta})$$

(Pascal et al., 2022, *Trans. Sig. Process.*)

Bayesian reformulation: interpret $(\widehat{\mathbf{R}}, \widehat{\mathbf{O}})$ as the MAP of

$$\pi(\boldsymbol{\theta}) \propto \exp(-f(\boldsymbol{\theta}|\mathbf{Z}) - h(\mathbf{A}\boldsymbol{\theta}))$$

- $\exp(-f(\boldsymbol{\theta}|\mathbf{Z})) \sim$ likelihood of the observation
- $\exp(-h(\mathbf{A}\boldsymbol{\theta})) \sim$ prior on the parameter of interest

Pointwise estimate of parameter $\boldsymbol{\theta} = (\mathbf{R}, \mathbf{O})$ from observations $\mathbf{Z}$

$$\underset{(\mathbf{R},\mathbf{O})\in\mathbb{R}_+^T\times\mathbb{R}^T}{\text{minimize}} \quad f(\boldsymbol{\theta}|\mathbf{Z}) + h(\mathbf{A}\boldsymbol{\theta}) \quad \text{(Pascal et al., 2022, \textit{Trans. Sig. Process.})}$$

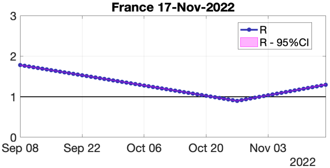
Bayesian reformulation: interpret $(\widehat{\mathbf{R}}, \widehat{\mathbf{O}})$ as the MAP of

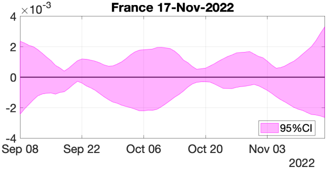
$$\pi(\boldsymbol{\theta}) \propto \exp(-f(\boldsymbol{\theta}|\mathbf{Z}) - h(\mathbf{A}\boldsymbol{\theta}))$$

- $\exp(-f(\boldsymbol{\theta}|\mathbf{Z})) \sim$ likelihood of the observation
- $\exp(-h(\mathbf{A}\boldsymbol{\theta})) \sim$ prior on the parameter of interest

$\Longrightarrow$ instead of focusing on $\widehat{\mathrm{R}}_t$, the **pointwise** MAP, probe π to get

$\mathrm{R}_t \in [\underline{\mathrm{R}}_t, \overline{\mathrm{R}}_t]$ with 95% probability, i.e., **credibility interval** estimates





$$\widehat{\mathrm{R}}_T \in [1.2987, 1.3047]$$

Log-likelihood from Poisson model $\mathcal{D} = \{\boldsymbol{\theta} \,|\, \forall t,\ \mathsf{R}_t\Phi_t + \mathsf{O}_t \geq 0\}$

$$f(\boldsymbol{\theta}\,|\mathbf{Z}) := \begin{cases} -\sum_{t=1}^{T}(\mathsf{Z}_t \ln(\mathsf{R}_t\Phi_t + \mathsf{O}_t) - (\mathsf{R}_t\Phi_t + \mathsf{O}_t) + \mathcal{C}(\mathsf{Z}_t)) & \text{if } \boldsymbol{\theta} \in \mathcal{D}, \\ +\infty & \text{otherwise,} \end{cases}$$

Log-likelihood from Poisson model $\mathcal{D} = \{\boldsymbol{\theta} \,|\, \forall t,\ \mathsf{R}_t\Phi_t + \mathsf{O}_t \geq 0\}$

$$f(\boldsymbol{\theta}) \quad := \begin{cases} -\sum_{t=1}^{T}(\mathsf{Z}_t \ln(\mathsf{R}_t\Phi_t + \mathsf{O}_t) - (\mathsf{R}_t\Phi_t + \mathsf{O}_t)) & \text{if } \boldsymbol{\theta} \in \mathcal{D}, \\ +\infty & \text{otherwise,} \end{cases}$$

Prior distribution of $\boldsymbol{\theta} = (\mathbf{R}, \mathbf{O}) = (\mathsf{R}_1, \ldots, \mathsf{R}_T, \mathsf{O}_1, \ldots, \mathsf{O}_T) \in (\mathbb{R}_+)^T \times \mathbb{R}^T$

- reproduction number: $\mathsf{R}_t - 2\mathsf{R}_{t-1} + \mathsf{R}_{t-2} \sim \mathsf{Laplace}(\lambda_\mathsf{R})$

Bayesian framework for credibility interval estimation

Log-likelihood from Poisson model $\mathcal{D} = \{\boldsymbol{\theta} \,|\, \forall t,\ \mathsf{R}_t\Phi_t + \mathsf{O}_t \geq 0\}$

$$f(\boldsymbol{\theta}) \quad := \begin{cases} -\sum_{t=1}^{T}(\mathsf{Z}_t \ln(\mathsf{R}_t\Phi_t + \mathsf{O}_t) - (\mathsf{R}_t\Phi_t + \mathsf{O}_t)) & \text{if } \boldsymbol{\theta} \in \mathcal{D}, \\ +\infty & \text{otherwise,} \end{cases}$$

Prior distribution of $\boldsymbol{\theta} = (\mathbf{R}, \mathbf{O}) = (\mathsf{R}_1, \ldots, \mathsf{R}_T, \mathsf{O}_1, \ldots, \mathsf{O}_T) \in (\mathbb{R}_+)^T \times \mathbb{R}^T$

- reproduction number: $\mathsf{R}_t - 2\mathsf{R}_{t-1} + \mathsf{R}_{t-2} \sim \mathrm{Laplace}(\lambda_{\mathsf{R}})$
- outliers $\mathsf{O}_t \sim \mathrm{Laplace}(\lambda_{\mathsf{O}})$

Log-likelihood from Poisson model $\mathcal{D} = \{\boldsymbol{\theta} \,|\, \forall t,\ \mathsf{R}_t\Phi_t + \mathsf{O}_t \geq 0\}$

$$f(\boldsymbol{\theta}) \quad := \begin{cases} -\sum_{t=1}^{T}(\mathsf{Z}_t \ln(\mathsf{R}_t\Phi_t + \mathsf{O}_t) - (\mathsf{R}_t\Phi_t + \mathsf{O}_t)) & \text{if } \boldsymbol{\theta} \in \mathcal{D}, \\ +\infty & \text{otherwise,} \end{cases}$$

Prior distribution of $\boldsymbol{\theta} = (\mathbf{R}, \mathbf{O}) = (\mathsf{R}_1, \ldots, \mathsf{R}_T, \mathsf{O}_1, \ldots, \mathsf{O}_T) \in (\mathbb{R}_+)^T \times \mathbb{R}^T$

- reproduction number: $\mathsf{R}_t - 2\mathsf{R}_{t-1} + \mathsf{R}_{t-2} \sim \text{Laplace}(\lambda_\mathsf{R})$
- outliers $\mathsf{O}_t \sim \text{Laplace}(\lambda_\mathsf{O})$

$$\Rightarrow g(\boldsymbol{\theta}) = \lambda_\mathsf{R}\|\mathbf{D}_2\mathbf{R}\|_1 + \lambda_\mathsf{O}\|\mathbf{O}\|_1, \quad \mathbf{D}_2 = \underbrace{\frac{1}{\sqrt{6}}\begin{bmatrix} 1 & -2 & 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & -2 & 1 & 0 & \ldots & 0 \\ \ldots & & & & & & \ldots \\ 0 & \ldots & & & 1 & -2 & 1 \end{bmatrix}}_{\text{Laplacian}}$$

Bayesian framework for credibility interval estimation

Log-likelihood from Poisson model $\mathcal{D} = \{\boldsymbol{\theta} \,|\, \forall t,\ \mathsf{R}_t\Phi_t + \mathsf{O}_t \geq 0\}$

$$f(\boldsymbol{\theta}) \quad := \begin{cases} -\sum_{t=1}^{T}(\mathsf{Z}_t \ln(\mathsf{R}_t\Phi_t + \mathsf{O}_t) - (\mathsf{R}_t\Phi_t + \mathsf{O}_t)) & \text{if } \boldsymbol{\theta} \in \mathcal{D}, \\ +\infty & \text{otherwise,} \end{cases}$$

Prior distribution of $\boldsymbol{\theta} = (\mathbf{R}, \mathbf{O}) = (\mathsf{R}_1, \ldots, \mathsf{R}_T, \mathsf{O}_1, \ldots, \mathsf{O}_T) \in (\mathbb{R}_+)^T \times \mathbb{R}^T$

- reproduction number: $\mathsf{R}_t - 2\mathsf{R}_{t-1} + \mathsf{R}_{t-2} \sim \text{Laplace}(\lambda_\mathsf{R})$
- outliers $\mathsf{O}_t \sim \text{Laplace}(\lambda_\mathsf{O})$

$$\Rightarrow g(\boldsymbol{\theta}) = \lambda_\mathsf{R}\|\mathbf{D}_2\mathbf{R}\|_1 + \lambda_\mathsf{O}\|\mathbf{O}\|_1, \quad \mathbf{D}_2 = \underbrace{\frac{1}{\sqrt{6}}\begin{bmatrix} 1 & -2 & 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & -2 & 1 & 0 & \ldots & 0 \\ \ldots & & & & & & \ldots \\ 0 & \ldots & & & 1 & -2 & 1 \end{bmatrix}}_{\text{Laplacian}}$$

Posterior distribution of unknown parameters $\boldsymbol{\theta} = (\mathbf{R}, \mathbf{O})$

$$\pi(\boldsymbol{\theta}) \propto \exp\left(-f(\boldsymbol{\theta}) - g(\boldsymbol{\theta})\right) \mathbb{1}_{\mathcal{D}}(\boldsymbol{\theta})$$

- f, g convex
- f smooth, g nonsmooth

Purpose: sampling the random variable $\boldsymbol{\theta} = (\mathbf{R}, \mathbf{O}) \in \mathbb{R}^{2T}$ according to the posterior[†]

$$\pi(\boldsymbol{\theta}) \propto \exp\left(-f(\boldsymbol{\theta}) - g(\boldsymbol{\theta})\right) \mathbb{1}_{\mathcal{D}}(\boldsymbol{\theta})$$

[†] π is defined up to a normalizing constant

Markov Chain Monte Carlo sampling

Purpose: sampling the random variable $\boldsymbol{\theta} = (\mathbf{R}, \mathbf{O}) \in \mathbb{R}^{2T}$ according to the posterior[†]

$$\pi(\boldsymbol{\theta}) \propto \exp\left(-f(\boldsymbol{\theta}) - g(\boldsymbol{\theta})\right) \mathbb{1}_{\mathcal{D}}(\boldsymbol{\theta})$$

Principle: 1) generate a random sequence $\{\boldsymbol{\theta}^n, \, n \in \mathbb{N}\}$ such that

- $\boldsymbol{\theta}^{n+1}$ only depends on $\boldsymbol{\theta}^n$,
- at convergence, i.e., as $n \to \infty$, $\boldsymbol{\theta}^n \sim \pi$,

2) compute Bayesian estimators, e.g., credibility intervals, on samples $\{\boldsymbol{\theta}^n, \, n \geq N\}$

[†] π is defined up to a normalizing constant

Purpose: sampling the random variable $\boldsymbol{\theta} = (\mathbf{R}, \mathbf{O}) \in \mathbb{R}^{2T}$ according to the posterior[†]

$$\pi(\boldsymbol{\theta}) \propto \exp\left(-f(\boldsymbol{\theta}) - g(\boldsymbol{\theta})\right) \mathbb{1}_{\mathcal{D}}(\boldsymbol{\theta})$$

Principle: 1) generate a random sequence $\{\boldsymbol{\theta}^n,\, n \in \mathbb{N}\}$ such that

- $\boldsymbol{\theta}^{n+1}$ only depends on $\boldsymbol{\theta}^n$,
- at convergence, i.e., as $n \to \infty$, $\boldsymbol{\theta}^n \sim \pi$,

2) compute Bayesian estimators, e.g., credibility intervals, on samples $\{\boldsymbol{\theta}^n,\, n \geq N\}$

State-of-the-art: *Hastings-Metropolis random walk*

(i) propose a random move according to

$$\boldsymbol{\theta}^{n+\frac{1}{2}} = \boldsymbol{\theta}^n + \sqrt{2\gamma}\Gamma\xi^{n+1}, \quad \xi^{n+1} \sim \mathcal{N}_{2T}(0, \mathrm{I})$$

with γ positive step size, $\Gamma \in \mathbb{R}^{2T\times 2T}$

[†] π is defined up to a normalizing constant

Purpose: sampling the random variable $\boldsymbol{\theta} = (\mathbf{R}, \mathbf{O}) \in \mathbb{R}^{2T}$ according to the posterior[†]

$$\pi(\boldsymbol{\theta}) \propto \exp\left(-f(\boldsymbol{\theta}) - g(\boldsymbol{\theta})\right) \mathbb{1}_{\mathcal{D}}(\boldsymbol{\theta})$$

Principle: 1) generate a random sequence $\{\boldsymbol{\theta}^n,\, n \in \mathbb{N}\}$ such that

- $\boldsymbol{\theta}^{n+1}$ only depends on $\boldsymbol{\theta}^n$,
- at convergence, i.e., as $n \to \infty$, $\boldsymbol{\theta}^n \sim \pi$,

2) compute Bayesian estimators, e.g., credibility intervals, on samples $\{\boldsymbol{\theta}^n,\, n \geq N\}$

State-of-the-art: *Hastings-Metropolis random walk*

(i) propose a random move according to

$$\boldsymbol{\theta}^{n+\frac{1}{2}} = \boldsymbol{\theta}^n + \sqrt{2\gamma}\Gamma\xi^{n+1}, \quad \xi^{n+1} \sim \mathcal{N}_{2T}(0, \mathrm{I})$$

with γ positive step size, $\Gamma \in \mathbb{R}^{2T \times 2T}$

(ii) accept: $\boldsymbol{\theta}^{n+1} = \boldsymbol{\theta}^{n+\frac{1}{2}}$, with probability $1 \wedge \dfrac{\pi(\boldsymbol{\theta}^{n+\frac{1}{2}})}{\pi(\boldsymbol{\theta}^n)}$, or reject: $\boldsymbol{\theta}^{n+1} = \boldsymbol{\theta}^n$

[†] π is defined up to a normalizing constant

Metropolis Adjusted Langevin Algorithm (MALA)

Langevin dynamics: $\boldsymbol{\theta}^{n+\frac{1}{2}} = \mu(\boldsymbol{\theta}^n) + \sqrt{2\gamma}\xi^{n+1}$, (Kent, 1978, *Adv Appl Probab*)

$\mu(\boldsymbol{\theta})$ adapted to $\pi(\boldsymbol{\theta}) = \exp(-f(\boldsymbol{\theta}) - g(\boldsymbol{\theta}))\mathbb{1}_{\mathcal{D}}(\boldsymbol{\theta})$

Langevin dynamics: $\boldsymbol{\theta}^{n+\frac{1}{2}} = \mu(\boldsymbol{\theta}^n) + \sqrt{2\gamma}\xi^{n+1}$, (Kent, 1978, *Adv Appl Probab*)

$\mu(\boldsymbol{\theta})$ adapted to $\pi(\boldsymbol{\theta}) = \exp(-f(\boldsymbol{\theta}) - g(\boldsymbol{\theta}))\mathbb{1}_{\mathcal{D}}(\boldsymbol{\theta})$

Case 1: $g = 0$ and $-\ln \pi = f$ is smooth (Roberts & Tweedie, 1996, *Bernoulli*)

$$\mu(\boldsymbol{\theta}) = \boldsymbol{\theta} - \gamma \Gamma\Gamma^\top \nabla f(\boldsymbol{\theta}) = \boldsymbol{\theta} + \gamma \Gamma\Gamma^\top \nabla \ln \pi(\boldsymbol{\theta})$$

$\Longrightarrow$ move towards areas of higher probability

Metropolis Adjusted Langevin Algorithm (MALA)

Langevin dynamics: $\boldsymbol{\theta}^{n+\frac{1}{2}} = \mu(\boldsymbol{\theta}^n) + \sqrt{2\gamma}\xi^{n+1}$, (Kent, 1978, *Adv Appl Probab*)

$$\mu(\boldsymbol{\theta}) \text{ adapted to } \pi(\boldsymbol{\theta}) = \exp(-f(\boldsymbol{\theta}) - g(\boldsymbol{\theta}))\mathbb{1}_{\mathcal{D}}(\boldsymbol{\theta})$$

Case 1: $g = 0$ and $-\ln\pi = f$ is smooth (Roberts & Tweedie, 1996, *Bernoulli*)

$$\mu(\boldsymbol{\theta}) = \boldsymbol{\theta} - \gamma\Gamma\Gamma^\top\nabla f(\boldsymbol{\theta}) = \boldsymbol{\theta} + \gamma\Gamma\Gamma^\top\nabla\ln\pi(\boldsymbol{\theta})$$

$\Longrightarrow$ move towards areas of higher probability

Case 2: $-\ln\pi = f + g$ is nonsmooth

$$\mu(\boldsymbol{\theta}) = \mathrm{prox}_{\gamma g}^{\Gamma\Gamma^\top}(\boldsymbol{\theta} - \gamma\Gamma\Gamma^\top\nabla f(\boldsymbol{\theta}))$$

combining *Langevin* and *proximal*[†] approaches

† $\mathrm{prox}_{\gamma g}^{\Gamma\Gamma^\top}(y) = \operatorname*{argmin}_{x\in\mathbb{R}^d}\left(\frac{1}{2}\|x-y\|^2_{\Gamma\Gamma^\top} + \gamma g(x)\right)$: preconditioned proximity operator of g

Posterior density of $\boldsymbol{\theta} = (\mathbf{R}, \mathbf{O})$: $\quad \pi(\boldsymbol{\theta}) \propto \exp\left(-f(\boldsymbol{\theta}) - g(\boldsymbol{\theta})\right) \mathbb{1}_{\mathcal{D}}(\boldsymbol{\theta})$

- smooth negative log-likelihood

$$\text{if } \boldsymbol{\theta} \in \mathcal{D}, \quad f(\boldsymbol{\theta}) = -\sum_{t=1}^{T} \left(\mathsf{Z}_t \ln \mathsf{p}_t(\boldsymbol{\theta}) - \mathsf{p}_t(\boldsymbol{\theta})\right), \quad \mathsf{p}_t(\boldsymbol{\theta}) = \mathsf{R}_t(\Phi \mathsf{Z})_t + \mathsf{O}_t$$

- nonsmooth convex lower-semicontinuous negative a priori log-distribution

$$g(\boldsymbol{\theta}) = \lambda_{\mathsf{R}} \|\mathbf{D}_2 \mathbf{R}\|_1 + \lambda_{\mathsf{O}} \|\mathbf{O}\|_1 = h(\mathbf{A}\boldsymbol{\theta})$$

$\mathbf{A} : \boldsymbol{\theta} \mapsto (\mathbf{D}_2\mathbf{R}, \mathbf{O})$ linear operator, $h(\cdot_1, \cdot_2) = \lambda_{\mathsf{R}} \|\cdot_1\|_1 + \lambda_{\mathsf{O}} \|\cdot_2\|_1$

Proximal-Gradient dual sampler `PGdual`

Posterior density of $\boldsymbol{\theta} = (\mathbf{R}, \mathbf{O})$: $\quad \pi(\boldsymbol{\theta}) \propto \exp\left(-f(\boldsymbol{\theta}) - g(\boldsymbol{\theta})\right) \mathbb{1}_{\mathcal{D}}(\boldsymbol{\theta})$

- smooth negative log-likelihood

$$\text{if } \boldsymbol{\theta} \in \mathcal{D}, \quad f(\boldsymbol{\theta}) = -\sum_{t=1}^{T} \left(\mathsf{Z}_t \ln \mathsf{p}_t(\boldsymbol{\theta}) - \mathsf{p}_t(\boldsymbol{\theta})\right), \quad \mathsf{p}_t(\boldsymbol{\theta}) = \mathsf{R}_t(\Phi \mathsf{Z})_t + \mathsf{O}_t$$

- nonsmooth convex lower-semicontinuous negative a priori log-distribution

$$g(\boldsymbol{\theta}) = \lambda_{\mathsf{R}} \|\mathbf{D}_2 \mathbf{R}\|_1 + \lambda_{\mathsf{O}} \|\mathbf{O}\|_1 = h(\mathbf{A}\boldsymbol{\theta})$$

$\mathbf{A} : \boldsymbol{\theta} \mapsto (\mathbf{D}_2\mathbf{R}, \mathbf{O})$ linear operator, $h(\cdot_1, \cdot_2) = \lambda_{\mathsf{R}} \|\cdot_1\|_1 + \lambda_{\mathsf{O}} \|\cdot_2\|_1$

Case 3: $-\ln \pi = f + h(\mathbf{A}\cdot)$ (Fort et al., 2022, *preprint*)

closed-form expression of $\text{prox}_{\gamma h}$ but not of $\text{prox}_{\gamma h(\mathbf{A}\cdot)}$

1) extend $\mathbf{A}$ into **invertible** $\overline{\mathbf{A}}$, and h in $\bar{h}$ such that $\bar{h}(\overline{\mathbf{A}}\boldsymbol{\theta}) = h(\mathbf{A}\boldsymbol{\theta})$

2) reason on the **dual** variable $\tilde{\boldsymbol{\theta}} = \overline{\mathbf{A}}\boldsymbol{\theta}$

Langevin: drift toward higher probability regions

$$\operatorname*{argmax}_{\boldsymbol{\theta}\in\mathbb{R}^{2T}} \ln \pi(\boldsymbol{\theta}) = \operatorname*{argmin}_{\boldsymbol{\theta}\in\mathbb{R}^{2T}} f(\boldsymbol{\theta}) + \bar{h}(\overline{\mathbf{A}}\boldsymbol{\theta}) = \mathbf{A}^{-1} \operatorname*{argmin}_{\tilde{\boldsymbol{\theta}}\in\mathbb{R}^{2T}} f(\overline{\mathbf{A}}^{-1}\tilde{\boldsymbol{\theta}}) + \bar{h}(\tilde{\boldsymbol{\theta}})$$

Langevin: drift toward higher probability regions

$$\underset{\boldsymbol{\theta}\in\mathbb{R}^{2T}}{\operatorname{argmax}} \ln \pi(\boldsymbol{\theta}) = \underset{\boldsymbol{\theta}\in\mathbb{R}^{2T}}{\operatorname{argmin}}\ f(\boldsymbol{\theta}) + \bar{h}(\overline{\mathbf{A}}\boldsymbol{\theta}) = \mathbf{A}^{-1}\underset{\tilde{\boldsymbol{\theta}}\in\mathbb{R}^{2T}}{\operatorname{argmin}}\ f(\overline{\mathbf{A}}^{-1}\tilde{\boldsymbol{\theta}}) + \bar{h}(\tilde{\boldsymbol{\theta}})$$

$$\Longrightarrow \quad \mu(\boldsymbol{\theta}) = \underbrace{\overline{\mathbf{A}}^{-1}}_{\text{back to } \boldsymbol{\theta}}\ \underbrace{\operatorname{prox}_{\gamma\bar{h}}\left(\overline{\mathbf{A}}\boldsymbol{\theta} - \gamma\overline{\mathbf{A}}^{-\top}\nabla f(\boldsymbol{\theta})\right)}_{\text{proximal-gradient on } \tilde{\boldsymbol{\theta}}}$$

Proximal-Gradient dual sampler `PGdual`

Langevin: drift toward higher probability regions

$$\underset{\boldsymbol{\theta}\in\mathbb{R}^{2T}}{\text{argmax}} \ln \pi(\boldsymbol{\theta}) = \underset{\boldsymbol{\theta}\in\mathbb{R}^{2T}}{\text{argmin}}\ f(\boldsymbol{\theta}) + \bar{h}(\overline{\mathbf{A}}\boldsymbol{\theta}) = \mathbf{A}^{-1} \underset{\tilde{\boldsymbol{\theta}}\in\mathbb{R}^{2T}}{\text{argmin}}\ f(\overline{\mathbf{A}}^{-1}\tilde{\boldsymbol{\theta}}) + \bar{h}(\tilde{\boldsymbol{\theta}})$$

$$\Longrightarrow \quad \mu(\boldsymbol{\theta}) = \underbrace{\overline{\mathbf{A}}^{-1}}_{\text{back to } \boldsymbol{\theta}}\ \underbrace{\text{prox}_{\gamma\bar{h}}\left(\overline{\mathbf{A}}\boldsymbol{\theta} - \gamma\overline{\mathbf{A}}^{-\top}\nabla f(\boldsymbol{\theta})\right)}_{\text{proximal-gradient on } \tilde{\boldsymbol{\theta}}}$$

Two strategies to extend $\mathbf{A} = \begin{pmatrix} \mathbf{D}_2 & 0 \\ 0 & \mathbf{I} \end{pmatrix} \in \mathbb{R}^{(2T-1)\times 2T}$ into $\overline{\mathbf{A}} = \begin{pmatrix} \overline{\mathbf{D}} & 0 \\ 0 & \mathbf{I} \end{pmatrix} \in \mathbb{R}^{2T\times 2T}$:

Proximal-Gradient dual sampler PGdual

Langevin: drift toward higher probability regions

$$\underset{\boldsymbol{\theta}\in\mathbb{R}^{2T}}{\operatorname{argmax}} \ln \pi(\boldsymbol{\theta}) = \underset{\boldsymbol{\theta}\in\mathbb{R}^{2T}}{\operatorname{argmin}}\ f(\boldsymbol{\theta}) + \bar{h}(\overline{\mathbf{A}}\boldsymbol{\theta}) = \mathbf{A}^{-1} \underset{\tilde{\boldsymbol{\theta}}\in\mathbb{R}^{2T}}{\operatorname{argmin}}\ f(\overline{\mathbf{A}}^{-1}\tilde{\boldsymbol{\theta}}) + \bar{h}(\tilde{\boldsymbol{\theta}})$$

$$\Longrightarrow \quad \mu(\boldsymbol{\theta}) = \underbrace{\overline{\mathbf{A}}^{-1}}_{\text{back to } \boldsymbol{\theta}} \underbrace{\operatorname{prox}_{\gamma\bar{h}}\left(\overline{\mathbf{A}}\boldsymbol{\theta} - \gamma\overline{\mathbf{A}}^{-\top}\nabla f(\boldsymbol{\theta})\right)}_{\text{proximal-gradient on } \tilde{\boldsymbol{\theta}}}$$

Two strategies to extend $\mathbf{A} = \begin{pmatrix} \mathbf{D}_2 & 0 \\ 0 & \mathrm{I} \end{pmatrix} \in \mathbb{R}^{(2T-1)\times 2T}$ into $\overline{\mathbf{A}} = \begin{pmatrix} \overline{\mathbf{D}} & 0 \\ 0 & \mathrm{I} \end{pmatrix} \in \mathbb{R}^{2T\times 2T}$:

Invert

$$\overline{\mathbf{D}}_2 := \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -2/\sqrt{5} & 1/\sqrt{5} & 0 & \cdots & 0 \\ & \mathbf{D}_2 & & & \end{bmatrix}$$

Langevin: drift toward higher probability regions

$$\underset{\boldsymbol{\theta}\in\mathbb{R}^{2T}}{\operatorname{argmax}} \ln \pi(\boldsymbol{\theta}) = \underset{\boldsymbol{\theta}\in\mathbb{R}^{2T}}{\operatorname{argmin}}\ f(\boldsymbol{\theta}) + \bar{h}(\overline{\mathbf{A}}\boldsymbol{\theta}) = \mathbf{A}^{-1}\underset{\tilde{\boldsymbol{\theta}}\in\mathbb{R}^{2T}}{\operatorname{argmin}}\ f(\overline{\mathbf{A}}^{-1}\tilde{\boldsymbol{\theta}}) + \bar{h}(\tilde{\boldsymbol{\theta}})$$

$$\Longrightarrow \quad \mu(\boldsymbol{\theta}) = \underbrace{\overline{\mathbf{A}}^{-1}}_{\text{back to } \boldsymbol{\theta}} \underbrace{\operatorname{prox}_{\gamma\bar{h}}\left(\overline{\mathbf{A}}\boldsymbol{\theta} - \gamma\overline{\mathbf{A}}^{-\top}\nabla f(\boldsymbol{\theta})\right)}_{\text{proximal-gradient on } \tilde{\boldsymbol{\theta}}}$$

Two strategies to extend $\mathbf{A} = \begin{pmatrix} \mathbf{D}_2 & 0 \\ 0 & \mathrm{I} \end{pmatrix} \in \mathbb{R}^{(2T-1)\times 2T}$ into $\overline{\mathbf{A}} = \begin{pmatrix} \overline{\mathbf{D}} & 0 \\ 0 & \mathrm{I} \end{pmatrix} \in \mathbb{R}^{2T\times 2T}$:

Invert

$$\overline{\mathbf{D}}_2 := \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -2/\sqrt{5} & 1/\sqrt{5} & 0 & \cdots & 0 \\ & \mathbf{D}_2 & & & \end{bmatrix}$$

Ortho

$$\overline{\mathbf{D}}_o := \begin{bmatrix} v_1^\top \\ v_2^\top \\ \mathbf{D}_2 \end{bmatrix} \qquad \begin{array}{c} v_1, v_2 \in \mathbb{R}^{2T} \\ v_1 \perp v_2,\ v_1, v_2 \in (\mathbf{D}_2^\top)^\perp \end{array}$$

Proximal-Gradient dual sampler `PGdual`

Langevin: drift toward higher probability regions

$$\underset{\boldsymbol{\theta}\in\mathbb{R}^{2T}}{\operatorname{argmax}} \ln \pi(\boldsymbol{\theta}) = \underset{\boldsymbol{\theta}\in\mathbb{R}^{2T}}{\operatorname{argmin}}\ f(\boldsymbol{\theta}) + \bar{h}(\overline{\mathbf{A}}\boldsymbol{\theta}) = \mathbf{A}^{-1} \underset{\tilde{\boldsymbol{\theta}}\in\mathbb{R}^{2T}}{\operatorname{argmin}}\ f(\overline{\mathbf{A}}^{-1}\tilde{\boldsymbol{\theta}}) + \bar{h}(\tilde{\boldsymbol{\theta}})$$

$$\Longrightarrow \quad \mu(\boldsymbol{\theta}) = \underbrace{\overline{\mathbf{A}}^{-1}}_{\text{back to } \boldsymbol{\theta}}\ \underbrace{\operatorname{prox}_{\gamma\bar{h}}\left(\overline{\mathbf{A}}\boldsymbol{\theta} - \gamma\overline{\mathbf{A}}^{-\top}\nabla f(\boldsymbol{\theta})\right)}_{\text{proximal-gradient on } \tilde{\boldsymbol{\theta}}}$$

Two strategies to extend $\mathbf{A} = \begin{pmatrix} \mathbf{D}_2 & 0 \\ 0 & \mathsf{I} \end{pmatrix} \in \mathbb{R}^{(2T-1)\times 2T}$ into $\overline{\mathbf{A}} = \begin{pmatrix} \overline{\mathbf{D}} & 0 \\ 0 & \mathsf{I} \end{pmatrix} \in \mathbb{R}^{2T\times 2T}$:

`Invert`

$$\overline{\mathbf{D}}_2 := \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -2/\sqrt{5} & 1/\sqrt{5} & 0 & \cdots & 0 \\ & \mathbf{D}_2 & & & \end{bmatrix}$$

`Ortho`

$$\overline{\mathbf{D}}_o := \begin{bmatrix} v_1^\top \\ v_2^\top \\ \mathbf{D}_2 \end{bmatrix} \qquad \begin{matrix} v_1, v_2 \in \mathbb{R}^{2T} \\ v_1 \perp v_2,\ v_1, v_2 \in (\mathbf{D}_2^\top)^\perp \end{matrix}$$

Proposed `PGdual` **drift terms** on $\boldsymbol{\theta} = (\mathbf{R}, \mathbf{O})$:

reproduction numbers $\mu_{\mathrm{R}}(\boldsymbol{\theta}) = \overline{\mathbf{D}}^{-1} \operatorname{prox}_{\gamma_{\mathrm{R}}\lambda_{\mathrm{R}}\|(\cdot)_{3:T}\|_1}\left(\overline{\mathbf{D}}\,\mathbf{R} - \gamma_{\mathrm{R}}\overline{\mathbf{D}}^{-\top}\nabla_{\mathrm{R}} f(\boldsymbol{\theta})\right)$

outliers $\mu_{\mathrm{O}}(\boldsymbol{\theta}) = \operatorname{prox}_{\gamma_{\mathrm{O}}\lambda_{\mathrm{O}}\|\cdot\|_1}\left(\mathbf{O} - \gamma_{\mathrm{O}}\nabla_{\mathrm{O}} f(\boldsymbol{\theta})\right)$

Data: $\overline{\mathbf{D}} = \overline{\mathbf{D}}_2$ (`Invert`) or $\overline{\mathbf{D}} = \overline{\mathbf{D}}_o$ (`Ortho`)

$\gamma_{\mathrm{R}}, \gamma_{\mathrm{O}} > 0$, $N_{\max} \in \mathbb{N}_\star$, $\boldsymbol{\theta}^0 = (\mathbf{R}^0, \mathbf{O}^0) \in \mathcal{D}$

Result: A $\mathcal{D}$-valued sequence $\{\boldsymbol{\theta}^n = (\mathbf{R}^n, \mathbf{O}^n), n \in 0, \ldots, N_{\max}\}$

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

Sample $\xi_{\mathrm{R}}^{n+1} \sim \mathcal{N}_T(0, \mathrm{I})$ and $\xi_{\mathrm{O}}^{n+1} \sim \mathcal{N}_T(0, \mathrm{I})$;

Set $\mathbf{R}^{n+\frac{1}{2}} = \mu_{\mathrm{R}}(\boldsymbol{\theta}^n) + \sqrt{2\gamma_{\mathrm{R}}}\,\overline{\mathbf{D}}^{-1}\overline{\mathbf{D}}^{-\top}\,\xi_{\mathrm{R}}^{n+1}$;

$\mathbf{O}^{n+\frac{1}{2}} = \mu_{\mathrm{O}}(\boldsymbol{\theta}^n) + \sqrt{2\gamma_{\mathrm{O}}}\,\xi_{\mathrm{O}}^{n+1}$;

$\boldsymbol{\theta}^{n+\frac{1}{2}} = (\mathbf{R}^{n+\frac{1}{2}}, \mathbf{O}^{n+\frac{1}{2}})$;

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

$$1 \wedge \frac{\pi(\boldsymbol{\theta}^{n+\frac{1}{2}})}{\pi(\boldsymbol{\theta}^n)} \frac{q_{\mathrm{R}}(\boldsymbol{\theta}^{n+\frac{1}{2}}, \boldsymbol{\theta}_{\mathrm{R}}^n)}{q_{\mathrm{R}}(\boldsymbol{\theta}^n, \boldsymbol{\theta}_{\mathrm{R}}^{n+\frac{1}{2}})} \frac{q_{\mathrm{O}}(\boldsymbol{\theta}^{n+\frac{1}{2}}, \boldsymbol{\theta}_{\mathrm{O}}^n)}{q_{\mathrm{O}}(\boldsymbol{\theta}^n, \boldsymbol{\theta}_{\mathrm{O}}^{n+\frac{1}{2}})},$$

$q_{\mathrm{R}/\mathrm{O}}$: Gaussian kernel stemming from nonsymmetric proposal

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

Algorithm 1: Proximal-Gradient dual: `PGdual Invert` and `PGdual Ortho`

Gaussian proposal: $\boldsymbol{\theta}^{n+\frac{1}{2}} = \mu(\boldsymbol{\theta}^n) + \sqrt{2\gamma}\Gamma\xi^{n+1}$

- random walks: $\mu(\boldsymbol{\theta}) = \boldsymbol{\theta}$

  `RW`: $\Gamma = \mathsf{I}$; `RW Invert`: $\Gamma = \overline{\mathbf{D}}_2^{-1}\overline{\mathbf{D}}_2^{-\top}$; `RW Ortho`: $\Gamma = \overline{\mathbf{D}}_o^{-1}\overline{\mathbf{D}}_o^{-\top}$

- Proximal-Gradient dual: $\mu_{\mathsf{R}}(\boldsymbol{\theta})$, $\mu_{\mathsf{O}}(\boldsymbol{\theta})$, $\Gamma = \overline{\mathbf{D}}^{-1}\overline{\mathbf{D}}^{-\top}$

  `PGdual Invert`: $\overline{\mathbf{D}} = \overline{\mathbf{D}}_2$; `PGdual Ortho`: $\overline{\mathbf{D}} = \overline{\mathbf{D}}_o$

Practical settings: $N_{\max} = 10^7$ iterations, 15 independent runs

Gaussian proposal: $\boldsymbol{\theta}^{n+\frac{1}{2}} = \mu(\boldsymbol{\theta}^n) + \sqrt{2\gamma}\Gamma\xi^{n+1}$

- random walks: $\mu(\boldsymbol{\theta}) = \boldsymbol{\theta}$

  `RW`: $\Gamma = \mathrm{I}$; `RW Invert`: $\Gamma = \overline{\mathbf{D}_2}^{-1}\overline{\mathbf{D}_2}^{-\top}$; `RW Ortho`: $\Gamma = \overline{\mathbf{D}_o}^{-1}\overline{\mathbf{D}_o}^{-\top}$

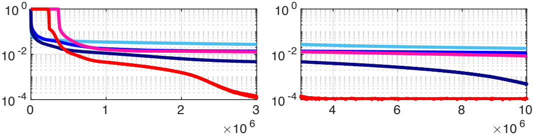
- Proximal-Gradient dual: $\mu_{\mathrm{R}}(\boldsymbol{\theta})$, $\mu_{\mathrm{O}}(\boldsymbol{\theta})$, $\Gamma = \overline{\mathbf{D}}^{-1}\overline{\mathbf{D}}^{-\top}$

  `PGdual Invert`: $\overline{\mathbf{D}} = \overline{\mathbf{D}_2}$; `PGdual Ortho`: $\overline{\mathbf{D}} = \overline{\mathbf{D}_o}$

Practical settings: $N_{\max} = 10^7$ iterations, 15 independent runs

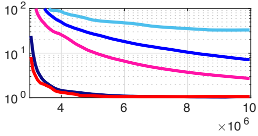
log-density

burn in | post burn in



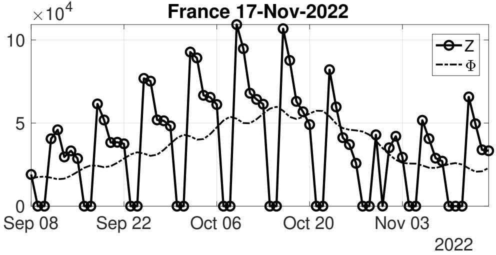
$\left(\ln \pi(\boldsymbol{\theta}^n) - \max \ln \pi\right) / \left(\ln \pi(\boldsymbol{\theta}^0) - \max \ln \pi\right)$

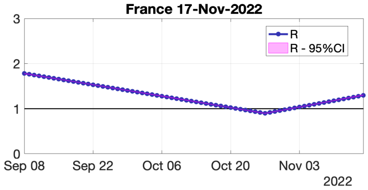
Gelman-Rubin

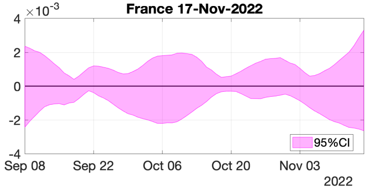


ANOVA-type criterion

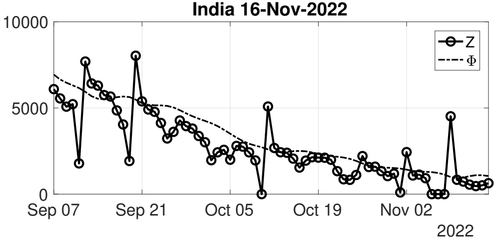
Sanitary situation in France

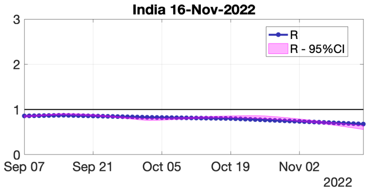


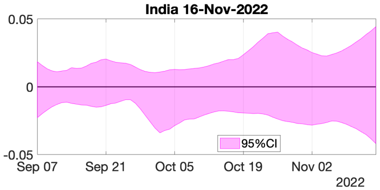




Worldwide Covid19 monitoring

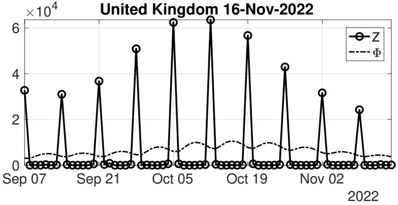






Why not United Kingdom?

Why not United Kingdom?

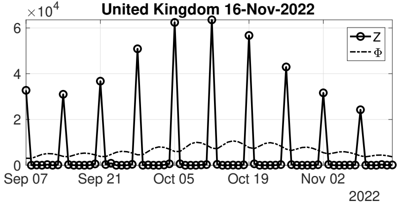


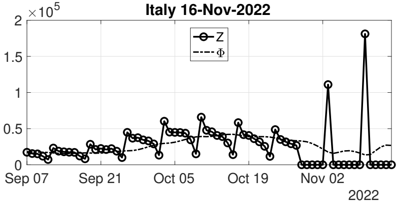
rate of erroneous counts: $6/7$!

PGdual credibility interval estimation of the reproduction number

Why not United Kingdom?

And Italy?





rate of erroneous counts: $6/7$!

seems to adopt the same reporting rate ...

$\Longrightarrow$ call for new tools, robust to very scarce data

Conclusion

✓ Extended Cori model handling erroneous reported counts via a latent variable

$$\mathrm{Z}_t | \mathbf{Z}_{t-\tau_\Phi:t-1}, \mathrm{R}_t, \mathrm{O}_t \sim \mathrm{Poiss}(\mathrm{R}_t \Phi_t + \mathrm{O}_t)$$

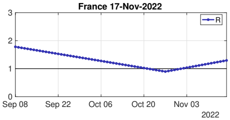
Conclusion

✓ Extended Cori model handling erroneous reported counts via a latent variable

$$\mathrm{Z}_t | \mathbf{Z}_{t-\tau_\Phi:t-1}, \mathrm{R}_t, \mathrm{O}_t \sim \mathrm{Poiss}(\mathrm{R}_t \Phi_t + \mathrm{O}_t)$$

✓ Estimation of piecewise linear R_t and corrected counts via convex optimization

$$\underset{(\mathbf{R},\mathbf{O}) \in \mathbb{R}_+^T \times \mathbb{R}^T}{\mathrm{minimize}} \sum_{t=1}^{T} \mathrm{d}_{\mathrm{KL}} \left(\mathrm{Z}_t \,|\, \mathrm{R}_t \Phi_t + \mathrm{O}_t \right) + \lambda_{\mathrm{R}} \|\mathbf{D}_2 \mathbf{R}\|_1 + \iota_{\geq 0}(\mathbf{R}) + \lambda_{\mathrm{O}} \|\mathbf{O}\|_1$$



$\widehat{\mathrm{R}}_T = 1.1959$

(Pascal et al., 2022, *Trans. Sig. Process.*;)

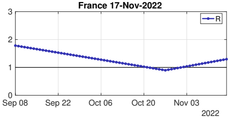
✓ Extended Cori model handling erroneous reported counts via a latent variable

$$\mathrm{Z}_t | \mathbf{Z}_{t-\tau_\Phi:t-1}, \mathrm{R}_t, \mathrm{O}_t \sim \mathrm{Poiss}(\mathrm{R}_t \Phi_t + \mathrm{O}_t)$$

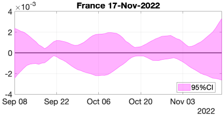
✓ Estimation of piecewise linear R_t and corrected counts via convex optimization

$$\underset{(\mathbf{R},\mathbf{O}) \in \mathbb{R}_+^T \times \mathbb{R}^T}{\text{minimize}} \sum_{t=1}^{T} \mathrm{d}_{\mathrm{KL}} \left(\mathrm{Z}_t \,|\, \mathrm{R}_t \Phi_t + \mathrm{O}_t \right) + \lambda_{\mathrm{R}} \|\mathbf{D}_2 \mathbf{R}\|_1 + \iota_{\geq 0}(\mathbf{R}) + \lambda_{\mathrm{O}} \|\mathbf{O}\|_1$$

✓ Bayesian credibility interval estimates via proximal Langevin MCMC samplers



$\widehat{\mathrm{R}}_T = 1.1959$



$\widehat{\mathrm{R}}_T \in [1.1978, 1.2016]$

(Pascal et al., 2022, *Trans. Sig. Process.*; Fort et al., 2022, *arXiv:2203.09142*)

$\longrightarrow$ Avoid mixing errors O_t with the pandemic mechanism $R_t\Phi_t$: anomaly models

$$Z_t|\mathbf{Z}_{t-\tau_\Phi:t-1}, R_t, O_t \sim \text{Poiss}((1-e_t)R_t\Phi_t + e_t O_t), \quad e_t \in \{0,1\}$$

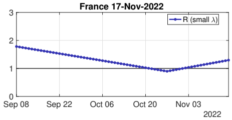
Perspectives

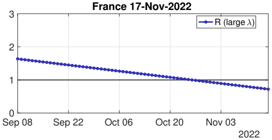
⟶ Avoid mixing errors O_t with the pandemic mechanism $\mathsf{R}_t\Phi_t$: anomaly models

$$\mathsf{Z}_t|\mathbf{Z}_{t-\tau_\Phi:t-1}, \mathsf{R}_t, \mathsf{O}_t \sim \mathrm{Poiss}((1-\mathsf{e}_t)\mathsf{R}_t\Phi_t + \mathsf{e}_t\mathsf{O}_t), \quad \mathsf{e}_t \in \{0,1\}$$

⟶ Selection of regularization parameters λ_R, λ_O

$$\underset{(\mathbf{R},\mathbf{O})\in\mathbb{R}_+^T\times\mathbb{R}^T}{\text{minimize}} \sum_{t=1}^{T} \mathsf{d}_{\mathsf{KL}}\left(\mathsf{Z}_t \,|\, \mathsf{R}_t\Phi_t + \mathsf{O}_t\right) + \lambda_\mathsf{R}\|\mathbf{D}_2\mathbf{R}\|_1 + \iota_{\geq 0}(\mathbf{R}) + \lambda_\mathsf{O}\|\mathbf{O}\|_1$$





Juliana Du PhD thesis

⟶ Synthetic data