Proximal-Langevin samplers for nonsmooth composite posteriors: Application to the estimation of Covid19 reproduction number

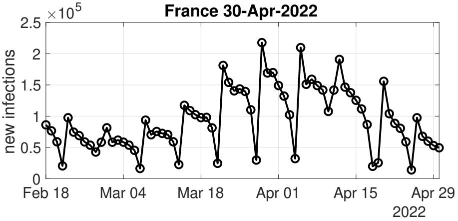
P. Abry[1,†], G. Fort[2,‡], B. Pascal[3,†], N. Pustelnik[1]

EUSIPCO 2023, Helsinki, Finland

---

1. CNRS, ENS de Lyon, Laboratoire de Physique, France,
2. CNRS, Institut de Mathématiques de Toulouse, France,
3. CNRS, LS2N, Nantes, France
† Supported by ANR-23-CE48-0009 "*OptiMoCSI*"
‡ Partly funded by *Fondation Simone et Cino Del Duca, Institut de France*

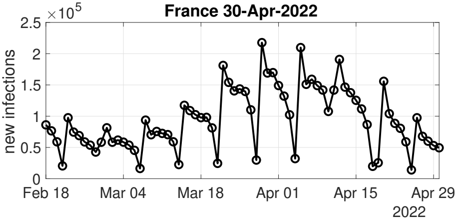
Daily counts of new infection cases



data collected
by Johns Hopkins University
from Public Health Agencies

Main challenges of epidemic surveillance

Daily counts of new infection cases



data collected
by Johns Hopkins University
from Public Health Agencies

Design of adapted sanitary measures and impact evaluation requires:

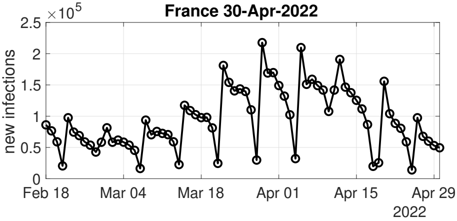
$\longrightarrow$ efficient monitoring tools — *epidemiological model,*

$\longrightarrow$ robustness to low quality of the data — *handle outlier values,*

$\longrightarrow$ reliable confidence levels — *credibility intervals.*

Main challenges of epidemic surveillance

Daily counts of new infection cases



data collected
by Johns Hopkins University
from Public Health Agencies

Design of adapted sanitary measures and impact evaluation requires:

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

$\longrightarrow$ robustness to low quality of the data *handle outlier values,*

$\longrightarrow$ reliable confidence levels *credibility intervals.*

Key indicator: reproduction number R_0 (Liu et al., 2018, *PNAS*)

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

$\Longrightarrow$ relaxed into an **effective time-varying reproduction number** R_t at day t

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

Propagation model of Covid19

Z_t: number of new infections at day t,

$$\mathbb{P}(\mathsf{Z}_t|\mathsf{Z}_{t-1},\mathsf{Z}_{t-2},\ldots) = \text{Poisson}\left(\mathsf{p}_t(\boldsymbol{\theta})\right), \quad \mathsf{p}_t(\boldsymbol{\theta}) = \mathsf{R}_t \sum_{u=1}^{\tau_\phi} \Phi_u \mathsf{Z}_{t-u} + \mathsf{O}_t$$

- Φ: serial interval function, i.e., infection delay distribution
- $\mathbf{R} = (\mathsf{R}_1, \cdots, \mathsf{R}_T)$ reproduction numbers at day $t = 1, \ldots, T$
- $\mathbf{O} = (\mathsf{O}_1, \cdots, \mathsf{O}_T)$ errors at day $t = 1, \ldots, T$

(Cori et al., 2013, *Am Journal of Epidemiology*; Pascal et al., 2022, IEEE Trans Sig Process)

Z_t: number of new infections at day t,

$$\mathbb{P}(Z_t|Z_{t-1}, Z_{t-2}, \ldots) = \text{Poisson}\left(p_t(\boldsymbol{\theta})\right), \quad p_t(\boldsymbol{\theta}) = R_t \sum_{u=1}^{\tau_\phi} \Phi_u Z_{t-u} + O_t$$

- Φ: serial interval function, i.e., infection delay distribution
- $\mathbf{R} = (R_1, \cdots, R_T)$ reproduction numbers at day $t = 1, \ldots, T$
- $\mathbf{O} = (O_1, \cdots, O_T)$ errors at day $t = 1, \ldots, T$

▶ $\boldsymbol{\theta} = (\mathbf{R}, \mathbf{O})$ to be estimated from $\mathbf{Z} = (Z_1, \ldots, Z_T)$ conditionally to $Z_0, Z_{-1}, \ldots$

(Cori et al., 2013, *Am Journal of Epidemiology*; Pascal et al., 2022, IEEE Trans Sig Process)

Z_t: number of new infections at day t,

$$\mathbb{P}(\mathsf{Z}_t|\mathsf{Z}_{t-1},\mathsf{Z}_{t-2},\ldots) = \text{Poisson}\left(\mathsf{p}_t(\boldsymbol{\theta})\right), \quad \mathsf{p}_t(\boldsymbol{\theta}) = \mathsf{R}_t \sum_{u=1}^{\tau_\phi} \Phi_u \mathsf{Z}_{t-u} + \mathsf{O}_t$$

- Φ: serial interval function, i.e., infection delay distribution
- $\mathbf{R} = (\mathsf{R}_1, \cdots, \mathsf{R}_T)$ reproduction numbers at day $t = 1, \ldots, T$
- $\mathbf{O} = (\mathsf{O}_1, \cdots, \mathsf{O}_T)$ errors at day $t = 1, \ldots, T$

▶ $\boldsymbol{\theta} = (\mathbf{R}, \mathbf{O})$ to be estimated from $\mathbf{Z} = (\mathsf{Z}_1, \ldots, \mathsf{Z}_T)$ conditionally to $\mathsf{Z}_0, \mathsf{Z}_{-1}, \ldots$

(Cori et al., 2013, *Am Journal of Epidemiology*; Pascal et al., 2022, IEEE Trans Sig Process)

Probability distribution of unknown parameters

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

Z_t: number of new infections at day t,

$$\mathbb{P}(\mathsf{Z}_t|\mathsf{Z}_{t-1},\mathsf{Z}_{t-2},\ldots) = \text{Poisson}\left(\mathsf{p}_t(\boldsymbol{\theta})\right), \quad \mathsf{p}_t(\boldsymbol{\theta}) = \mathsf{R}_t \sum_{u=1}^{\tau_\phi} \Phi_u \mathsf{Z}_{t-u} + \mathsf{O}_t$$

- Φ: serial interval function, i.e., infection delay distribution
- $\mathbf{R} = (\mathsf{R}_1, \cdots, \mathsf{R}_T)$ reproduction numbers at day $t = 1, \ldots, T$
- $\mathbf{O} = (\mathsf{O}_1, \cdots, \mathsf{O}_T)$ errors at day $t = 1, \ldots, T$

▶ $\boldsymbol{\theta} = (\mathbf{R}, \mathbf{O})$ to be estimated from $\mathbf{Z} = (\mathsf{Z}_1, \ldots, \mathsf{Z}_T)$ conditionally to $\mathsf{Z}_0, \mathsf{Z}_{-1}, \ldots$

(Cori et al., 2013, *Am Journal of Epidemiology*; Pascal et al., 2022, IEEE Trans Sig Process)

Probability distribution of unknown parameters

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

- $\mathcal{D} = \{\boldsymbol{\theta} \mid \forall t\colon\ \mathsf{R}_t \geq 0 \text{ and, if } \mathsf{Z}_t > 0, \mathsf{p}_t(\boldsymbol{\theta}) > 0;\ \text{else if } \mathsf{Z}_t = 0, \mathsf{p}_t(\boldsymbol{\theta}) \geq 0\}$

Z_t: number of new infections at day t,

$$\mathbb{P}(Z_t | Z_{t-1}, Z_{t-2}, \ldots) = \text{Poisson}\left(p_t(\boldsymbol{\theta})\right), \quad p_t(\boldsymbol{\theta}) = R_t \sum_{u=1}^{\tau_\phi} \Phi_u Z_{t-u} + O_t$$

- Φ: serial interval function, i.e., infection delay distribution
- $\mathbf{R} = (R_1, \cdots, R_T)$ reproduction numbers at day $t = 1, \ldots, T$
- $\mathbf{O} = (O_1, \cdots, O_T)$ errors at day $t = 1, \ldots, T$

▶ $\boldsymbol{\theta} = (\mathbf{R}, \mathbf{O})$ to be estimated from $\mathbf{Z} = (Z_1, \ldots, Z_T)$ conditionally to $Z_0, Z_{-1}, \ldots$

(Cori et al., 2013, *Am Journal of Epidemiology*; Pascal et al., 2022, IEEE Trans Sig Process)

Probability distribution of unknown parameters

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

- $\mathcal{D} = \{\boldsymbol{\theta} \mid \forall t\colon\ R_t \geq 0 \text{ and, if } Z_t > 0, p_t(\boldsymbol{\theta}) > 0;\ \text{else if } Z_t = 0, p_t(\boldsymbol{\theta}) \geq 0\}$
- $f(\boldsymbol{\theta}) := \begin{cases} \sum_{t=1}^{T} \left(-Z_t \ln p_t(\boldsymbol{\theta}) + p_t(\boldsymbol{\theta})\right) & \text{if } \boldsymbol{\theta} \in \mathcal{D}, \text{ with } 0 \cdot \ln(0) \stackrel{!}{=} 0 \\ \infty & \text{else} \end{cases}$

Z_t: number of new infections at day t,

$$\mathbb{P}(Z_t | Z_{t-1}, Z_{t-2}, \ldots) = \text{Poisson}\left(p_t(\boldsymbol{\theta})\right), \quad p_t(\boldsymbol{\theta}) = R_t \sum_{u=1}^{\tau_\phi} \Phi_u Z_{t-u} + O_t$$

- Φ: serial interval function, i.e., infection delay distribution
- $\mathbf{R} = (R_1, \cdots, R_T)$ reproduction numbers at day $t = 1, \ldots, T$
- $\mathbf{O} = (O_1, \cdots, O_T)$ errors at day $t = 1, \ldots, T$

▶ $\boldsymbol{\theta} = (\mathbf{R}, \mathbf{O})$ to be estimated from $\mathbf{Z} = (Z_1, \ldots, Z_T)$ conditionally to $Z_0, Z_{-1}, \ldots$

(Cori et al., 2013, *Am Journal of Epidemiology*; Pascal et al., 2022, IEEE Trans Sig Process)

Probability distribution of unknown parameters

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

- $\mathcal{D} = \{\boldsymbol{\theta} \mid \forall t\colon\ R_t \geq 0 \text{ and, if } Z_t > 0, p_t(\boldsymbol{\theta}) > 0;\ \text{else if } Z_t = 0, p_t(\boldsymbol{\theta}) \geq 0\}$
- $f(\boldsymbol{\theta}) := \begin{cases} \sum_{t=1}^{T} \left(-Z_t \ln p_t(\boldsymbol{\theta}) + p_t(\boldsymbol{\theta})\right) & \text{if } \boldsymbol{\theta} \in \mathcal{D}, \text{ with } 0 \cdot \ln(0) \stackrel{!}{=} 0 \\ \infty & \text{else} \end{cases}$
- $g(A\boldsymbol{\theta}) = \lambda_R \|D_2 \mathbf{R}\|_1 + \lambda_O \|\mathbf{O}\|_1, \qquad D_2 \in \mathbb{R}^{(T-2)\times T}$: discrete Laplacian matrix

(Artigas et al., 2022, EUSIPCO; Fort et al., 2023, IEEE Trans Sig Process)

Pandemic monitoring

Two quantities of interest:

- reproduction number $\mathbf{R} = (\mathrm{R}_1, \ldots, \mathrm{R}_T)$
- corrected number of new infections $\mathbf{Z}^{(D)} = \mathbf{Z} - \mathbf{O} = (\mathrm{Z}_1 - \mathrm{O}_1, \ldots, \mathrm{Z}_T - \mathrm{O}_T)$

Two quantities of interest:

- reproduction number $\mathbf{R} = (\mathrm{R}_1, \ldots, \mathrm{R}_T)$
- corrected number of new infections $\mathbf{Z}^{(D)} = \mathbf{Z} - \mathbf{O} = (\mathrm{Z}_1 - \mathrm{O}_1, \ldots, \mathrm{Z}_T - \mathrm{O}_T)$

Level of confidence required to support high impact sanitary decisions:

$\Longrightarrow$ estimate **credibility intervals at level 95%** under the statistical model

$$\boldsymbol{\theta} = (\mathbf{R}, \mathbf{O}) \sim \pi, \quad \text{with} \quad \pi(\boldsymbol{\theta}) \propto \exp\left(-f(\boldsymbol{\theta}) - g(\mathrm{A}\boldsymbol{\theta})\right) \mathbb{1}_{\mathcal{D}}(\boldsymbol{\theta})$$

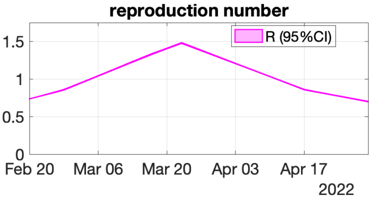
Two quantities of interest:

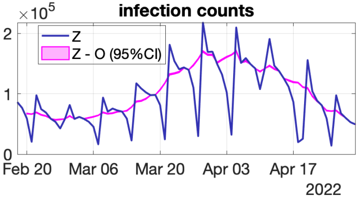
- reproduction number $\mathbf{R} = (R_1, \ldots, R_T)$
- corrected number of new infections $\mathbf{Z}^{(D)} = \mathbf{Z} - \mathbf{O} = (Z_1 - O_1, \ldots, Z_T - O_T)$

Level of confidence required to support high impact sanitary decisions:

$\Longrightarrow$ estimate **credibility intervals at level 95%** under the statistical model

$$\boldsymbol{\theta} = (\mathbf{R}, \mathbf{O}) \sim \pi, \quad \text{with} \quad \pi(\boldsymbol{\theta}) \propto \exp\left(-f(\boldsymbol{\theta}) - g(\mathrm{A}\boldsymbol{\theta})\right) \mathbb{1}_{\mathcal{D}}(\boldsymbol{\theta})$$





$R_T \in [0.70, 0.71] \quad \Longrightarrow \quad R_T < 1$ with probability at least 0.95

Credibility interval estimation of

$\boldsymbol{\theta} \equiv$ sample from a distribution[†] of the form:

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

(Artigas et al., 2022, EUSIPCO; Fort et al., 2023, IEEE Trans Sig Process)

[†] π is defined up to a normalizing constant.

Credibility interval estimation of

$$\boldsymbol{\theta} \equiv \text{sample from a distribution}^{\dagger} \text{ of the form:}$$

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

- $\boldsymbol{\theta} \in \mathbb{R}^d$ vector of parameters,
- f differentiable,
- $\mathcal{D} \subset \mathbb{R}^d$ admissible domain,
- g convex, non-smooth,
- $\mathrm{A} \in \mathbb{R}^{d \times d}$ invertible linear operator.

(Artigas et al., 2022, EUSIPCO; Fort et al., 2023, IEEE Trans Sig Process)

† π is defined up to a normalizing constant.

Credibility interval estimation of

$\boldsymbol{\theta} \equiv$ sample from a distribution[†] of the form:

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

- $\boldsymbol{\theta} \in \mathbb{R}^d$ vector of parameters,
- f differentiable,
- $\mathcal{D} \subset \mathbb{R}^d$ admissible domain,
- g convex, non-smooth,
- $\mathrm{A} \in \mathbb{R}^{d \times d}$ invertible linear operator.

(Artigas et al., 2022, EUSIPCO; Fort et al., 2023, IEEE Trans Sig Process)

Markov Chain Monte Carlo method

[†] π is defined up to a normalizing constant.

Credibility interval estimation of

$$\boldsymbol{\theta} \equiv \text{sample from a distribution}^{\dagger} \text{ of the form:}$$

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

- $\boldsymbol{\theta} \in \mathbb{R}^d$ vector of parameters,
- f differentiable,
- $\mathcal{D} \subset \mathbb{R}^d$ admissible domain,
- g convex, non-smooth,
- $\mathrm{A} \in \mathbb{R}^{d \times d}$ invertible linear operator.

(Artigas et al., 2022, EUSIPCO; Fort et al., 2023, IEEE Trans Sig Process)

Markov Chain Monte Carlo method

1) generate a Markov chain $\{\boldsymbol{\theta}^n, n \in \mathbb{N}\}$ such that

- $\boldsymbol{\theta}^{n+1}$ depends only on $\boldsymbol{\theta}^n$,
- at convergence, i.e., as $n \to \infty$, $\boldsymbol{\theta}^n \xrightarrow{\text{(in law)}} \pi$,

† π is defined up to a normalizing constant.

Credibility interval estimation of

$\boldsymbol{\theta} \equiv$ sample from a distribution[†] of the form:

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

- $\boldsymbol{\theta} \in \mathbb{R}^d$ vector of parameters,
- f differentiable,
- $\mathcal{D} \subset \mathbb{R}^d$ admissible domain,
- g convex, non-smooth,
- $\mathrm{A} \in \mathbb{R}^{d \times d}$ invertible linear operator.

(Artigas et al., 2022, EUSIPCO; Fort et al., 2023, IEEE Trans Sig Process)

Markov Chain Monte Carlo method

1) generate a Markov chain $\{\boldsymbol{\theta}^n, n \in \mathbb{N}\}$ such that
   - $\boldsymbol{\theta}^{n+1}$ depends only on $\boldsymbol{\theta}^n$,
   - at convergence, i.e., as $n \to \infty$, $\boldsymbol{\theta}^n \overset{\text{(in law)}}{\longrightarrow} \pi$,

2) compute credibility interval estimates from samples $\{\boldsymbol{\theta}^n, n \geq N\}$ for $N \gg 1$.

† π is defined up to a normalizing constant.

Hastings-Metropolis type algorithm $\mathsf{C} \in \mathbb{R}^{d \times d}$ symmetric positive definite; $\gamma > 0$

1) Gaussian proposal: $\boldsymbol{\theta}^{n+\frac{1}{2}} = \mu(\boldsymbol{\theta}^n) + \sqrt{2\gamma}\xi^{n+1}, \quad \xi^{n+1} \sim \mathcal{N}_d(0, \mathsf{C});$

freed from the constraint $\boldsymbol{\theta} \in \mathcal{D}$;

Hastings-Metropolis type algorithm $\quad \mathrm{C} \in \mathbb{R}^{d\times d}$ symmetric positive definite; $\gamma > 0$

1) Gaussian proposal: $\boldsymbol{\theta}^{n+\frac{1}{2}} = \mu(\boldsymbol{\theta}^n) + \sqrt{2\gamma}\xi^{n+1}, \quad \xi^{n+1} \sim \mathcal{N}_d(0, \mathrm{C});$

freed from the constraint $\boldsymbol{\theta} \in \mathcal{D}$;

2) accept-reject: if $\boldsymbol{\theta}^{n+\frac{1}{2}} \notin \mathcal{D}, \quad \boldsymbol{\theta}^{n+1} = \boldsymbol{\theta}^n$ (systematic reject);

else $\boldsymbol{\theta}^{n+1} = \boldsymbol{\theta}^{n+\frac{1}{2}}$ with probability depending on $\boldsymbol{\theta}^{n+\frac{1}{2}}$ and $\boldsymbol{\theta}^n$,

$\boldsymbol{\theta}^{n+1} = \boldsymbol{\theta}^n$ otherwise. (Kent, 1978, *Adv Appl Probab*)

Hastings-Metropolis type algorithm $\quad \mathsf{C} \in \mathbb{R}^{d\times d}$ symmetric positive definite; $\gamma > 0$

1) Gaussian proposal: $\boldsymbol{\theta}^{n+\frac{1}{2}} = \mu(\boldsymbol{\theta}^n) + \sqrt{2\gamma}\xi^{n+1}, \quad \xi^{n+1} \sim \mathcal{N}_d(0, \mathsf{C});$

freed from the constraint $\boldsymbol{\theta} \in \mathcal{D}$;

2) accept-reject: if $\boldsymbol{\theta}^{n+\frac{1}{2}} \notin \mathcal{D}, \quad \boldsymbol{\theta}^{n+1} = \boldsymbol{\theta}^n$ (systematic reject);

else $\boldsymbol{\theta}^{n+1} = \boldsymbol{\theta}^{n+\frac{1}{2}}$ with probability depending on $\boldsymbol{\theta}^{n+\frac{1}{2}}$ and $\boldsymbol{\theta}^n$,

$\boldsymbol{\theta}^{n+1} = \boldsymbol{\theta}^n$ otherwise. (Kent, 1978, *Adv Appl Probab*)

Case of smooth π: Tempered Langevin dynamics (Roberts & Tweedie, 1996, *Bernoulli*)

$$\mu(\boldsymbol{\theta}) = \boldsymbol{\theta} + \gamma\sqrt{\mathsf{C}}\,\nabla \ln \pi(\boldsymbol{\theta})$$

$\Longrightarrow$ move toward regions of high probability

Hastings-Metropolis type algorithm $\quad \mathrm{C} \in \mathbb{R}^{d\times d}$ symmetric positive definite; $\gamma > 0$

1) Gaussian proposal: $\boldsymbol{\theta}^{n+\frac{1}{2}} = \mu(\boldsymbol{\theta}^n) + \sqrt{2\gamma}\xi^{n+1}, \quad \xi^{n+1} \sim \mathcal{N}_d(0, \mathrm{C});$

freed from the constraint $\boldsymbol{\theta} \in \mathcal{D}$;

2) accept-reject: if $\boldsymbol{\theta}^{n+\frac{1}{2}} \notin \mathcal{D}, \quad \boldsymbol{\theta}^{n+1} = \boldsymbol{\theta}^n$ (systematic reject);

else $\boldsymbol{\theta}^{n+1} = \boldsymbol{\theta}^{n+\frac{1}{2}}$ with probability depending on $\boldsymbol{\theta}^{n+\frac{1}{2}}$ and $\boldsymbol{\theta}^n$,

$\boldsymbol{\theta}^{n+1} = \boldsymbol{\theta}^n$ otherwise. (Kent, 1978, *Adv Appl Probab*)

Case of smooth π: Tempered Langevin dynamics (Roberts & Tweedie, 1996, *Bernoulli*)

$$\mu(\boldsymbol{\theta}) = \boldsymbol{\theta} + \gamma\sqrt{\mathrm{C}}\,\nabla \ln \pi(\boldsymbol{\theta})$$

$\Longrightarrow$ move toward regions of high probability

Case of non-smooth π: proximal Langevin $\quad \pi \propto \exp\left(-f - g(\mathrm{A}\cdot)\right)\mathbb{1}_{\mathcal{D}}$

- f differentiable with gradient ∇f,
- g non-smooth, convex,

  with closed-form proximal operator $\mathrm{prox}_{\rho g} = (\mathrm{I} + \rho\partial g)^{-1}$, $\rho > 0$.

Purpose: compare different proximal design of the drift μ.

Purpose: drift term $\mu(\boldsymbol{\theta})$ adapted to $\pi \propto \exp\left(-f - g(\mathsf{A}\cdot)\right)\mathbb{1}_{\mathcal{D}}$, g non-smooth.

$$\operatorname{prox}_{\gamma g(\mathsf{A}\cdot)}(\boldsymbol{\theta}) = \underset{\boldsymbol{\varphi}\in\mathbb{R}^d}{\operatorname{argmin}} \left(\frac{1}{2}\|\boldsymbol{\theta} - \boldsymbol{\varphi}\|_2^2 + \gamma g(\mathsf{A}\boldsymbol{\varphi})\right)$$

Purpose: drift term $\mu(\boldsymbol{\theta})$ adapted to $\pi \propto \exp\left(-f - g(\mathrm{A}\cdot)\right)\mathbb{1}_{\mathcal{D}}$, g non-smooth.

$$\operatorname{prox}_{\gamma g(\mathrm{A}\cdot)}(\boldsymbol{\theta}) = \underset{\boldsymbol{\varphi}\in\mathbb{R}^d}{\operatorname{argmin}} \left(\frac{1}{2}\|\boldsymbol{\theta} - \boldsymbol{\varphi}\|_2^2 + \gamma g(\mathrm{A}\boldsymbol{\varphi})\right)$$

Primal methods: $\boldsymbol{\theta}^{n+\frac{1}{2}} = \mu(\boldsymbol{\theta}^n) + \sqrt{2\gamma}\xi^{n+1}, \quad \xi^{n+1} \sim \mathcal{N}_d(0, \mathrm{I})$

Purpose: drift term $\mu(\boldsymbol{\theta})$ adapted to $\pi \propto \exp\left(-f - g(\mathsf{A}\cdot)\right) \mathbb{1}_{\mathcal{D}}$, g non-smooth.

$$\operatorname{prox}_{\gamma g(\mathsf{A}\cdot)}(\boldsymbol{\theta}) = \underset{\boldsymbol{\varphi} \in \mathbb{R}^d}{\operatorname{argmin}} \left(\frac{1}{2} \|\boldsymbol{\theta} - \boldsymbol{\varphi}\|_2^2 + \gamma g(\mathsf{A}\boldsymbol{\varphi}) \right)$$

Primal methods: $\boldsymbol{\theta}^{n+\frac{1}{2}} = \mu(\boldsymbol{\theta}^n) + \sqrt{2\gamma}\xi^{n+1}, \quad \xi^{n+1} \sim \mathcal{N}_d(0, \mathsf{I})$

- Moreau drift: *smooth approximation of g by its Moreau envelop*

$$\mu^{\mathtt{M}}(\boldsymbol{\theta}) = \boldsymbol{\theta} - \gamma \nabla f(\boldsymbol{\theta}) - \frac{\gamma}{\rho} \mathsf{A}^\top (\mathsf{I} - \operatorname{prox}_{\rho g}) \mathsf{A}\boldsymbol{\theta}, \quad \rho = \gamma$$

(Durmus et al., 2018, *SIAM J Imaging Sci*; Luu et al., 2020, *Methodol Comput Appl Probab*)

Purpose: drift term $\mu(\boldsymbol{\theta})$ adapted to $\pi \propto \exp\left(-f - g(\mathsf{A}\cdot)\right) \mathbb{1}_{\mathcal{D}}$, g non-smooth.

$$\operatorname{prox}_{\gamma g(\mathsf{A}\cdot)}(\boldsymbol{\theta}) = \underset{\boldsymbol{\varphi} \in \mathbb{R}^d}{\operatorname{argmin}} \left(\frac{1}{2} \|\boldsymbol{\theta} - \boldsymbol{\varphi}\|_2^2 + \gamma g(\mathsf{A}\boldsymbol{\varphi}) \right)$$

Primal methods: $\boldsymbol{\theta}^{n+\frac{1}{2}} = \mu(\boldsymbol{\theta}^n) + \sqrt{2\gamma}\xi^{n+1}, \quad \xi^{n+1} \sim \mathcal{N}_d(0, \mathsf{I})$

- Moreau drift: *smooth approximation of g by its Moreau envelop*

$$\mu^{\mathtt{M}}(\boldsymbol{\theta}) = \boldsymbol{\theta} - \gamma \nabla f(\boldsymbol{\theta}) - \frac{\gamma}{\rho} \mathsf{A}^\top (\mathsf{I} - \operatorname{prox}_{\rho g}) \mathsf{A}\boldsymbol{\theta}, \quad \rho = \gamma$$

(Durmus et al., 2018, *SIAM J Imaging Sci*; Luu et al., 2020, *Methodol Comput Appl Probab*)

- PGdec drift: if $\mathsf{A}\mathsf{A}^\top = \nu \mathsf{I}$, with $\nu > 0 \Longrightarrow$ closed-form expression of $\operatorname{prox}_{\gamma g(\mathsf{A}\cdot)}$

$$\mu^{\mathtt{PGdec}}(\boldsymbol{\theta}) = \operatorname{prox}_{\gamma g(\mathsf{A}\cdot)} \left(\boldsymbol{\theta} - \gamma \nabla f(\boldsymbol{\theta}) \right)$$

extended to $g(\mathsf{A}\cdot) = \sum_{i=1}^{I} g_i(\mathsf{A}_i\cdot)$, with $\mathsf{A}_i\mathsf{A}_i^\top = \nu_i \mathsf{I}$, $\nu_i > 0$

(Fort et al., 2023, IEEE Trans Sig Process)

Purpose: drift term $\mu(\boldsymbol{\theta})$ adapted to $\pi \propto \exp\left(-f - g(\mathsf{A}\cdot)\right) \mathbb{1}_{\mathcal{D}}$, g non-smooth.

$$\operatorname{prox}_{\gamma g(\mathsf{A}\cdot)}(\boldsymbol{\theta}) = \underset{\boldsymbol{\varphi} \in \mathbb{R}^d}{\operatorname{argmin}} \left(\frac{1}{2} \|\boldsymbol{\theta} - \boldsymbol{\varphi}\|_2^2 + \gamma g(\mathsf{A}\boldsymbol{\varphi}) \right)$$

Primal methods: $$\boldsymbol{\theta}^{n+\frac{1}{2}} = \mu(\boldsymbol{\theta}^n) + \sqrt{2\gamma}\xi^{n+1}, \quad \xi^{n+1} \sim \mathcal{N}_d(0, \mathsf{I})$$

- Moreau drift: *smooth approximation of g by its Moreau envelop*

$$\mu^{\mathtt{M}}(\boldsymbol{\theta}) = \boldsymbol{\theta} - \gamma \nabla f(\boldsymbol{\theta}) - \frac{\gamma}{\rho} \mathsf{A}^\top (\mathsf{I} - \operatorname{prox}_{\rho g}) \mathsf{A}\boldsymbol{\theta}, \quad \rho = \gamma$$

(Durmus et al., 2018, *SIAM J Imaging Sci*; Luu et al., 2020, *Methodol Comput Appl Probab*)

- PGdec drift: if $\mathsf{A}\mathsf{A}^\top = \nu \mathsf{I}$, with $\nu > 0 \Longrightarrow$ closed-form expression of $\operatorname{prox}_{\gamma g(\mathsf{A}\cdot)}$

$$\mu^{\mathtt{PGdec}}(\boldsymbol{\theta}) = \operatorname{prox}_{\gamma g(\mathsf{A}\cdot)} \left(\boldsymbol{\theta} - \gamma \nabla f(\boldsymbol{\theta})\right)$$

extended to $g(\mathsf{A}\cdot) = \sum_{i=1}^{I} g_i(\mathsf{A}_i \cdot)$, with $\mathsf{A}_i \mathsf{A}_i^\top = \nu_i \mathsf{I}$, $\nu_i > 0$

(Fort et al., 2023, IEEE Trans Sig Process)

- Random Walk drift: $\mu^{\mathtt{RM}}(\boldsymbol{\theta}) = \boldsymbol{\theta}$

Purpose: drift term $\mu(\boldsymbol{\theta})$ adapted to $\pi \propto \exp\left(-f - g(\mathsf{A}\cdot)\right) \mathbb{1}_{\mathcal{D}}$, g non-smooth.

Purpose: drift term $\mu(\boldsymbol{\theta})$ adapted to $\pi \propto \exp(-f - g(\mathrm{A}\cdot))\,\mathbb{1}_{\mathcal{D}}$, g non-smooth.

$\mathrm{A} \in \mathbb{R}^{d \times d}$ **invertible**

dual drift term $\tilde{\mu}(\tilde{\boldsymbol{\theta}})$, $\tilde{\boldsymbol{\theta}} = \mathrm{A}\boldsymbol{\theta}$, adapted to $\tilde{\pi} \propto \exp\left(-f(\mathrm{A}^{-1}\cdot) - g\right)\mathbb{1}_{\mathcal{D}}(\mathrm{A}^{-1}\cdot)$

If $\mathrm{A} \in \mathbb{R}^{c \times d}$ with $c \leq d$ is **full rank**, consider invertible extension $\overline{\mathrm{A}} \in \mathbb{R}^{d \times d}$.

(Artigas, 2022, *EUSIPCO*; Fort et al., 2023, IEEE Trans Sig Process)

Purpose: drift term $\mu(\boldsymbol{\theta})$ adapted to $\pi \propto \exp\left(-f - g(\mathsf{A}\cdot)\right) \mathbb{1}_{\mathcal{D}}$, g non-smooth.

$\mathsf{A} \in \mathbb{R}^{d \times d}$ **invertible**

dual drift term $\tilde{\mu}(\tilde{\boldsymbol{\theta}})$, $\tilde{\boldsymbol{\theta}} = \mathsf{A}\boldsymbol{\theta}$, adapted to $\tilde{\pi} \propto \exp\left(-f(\mathsf{A}^{-1}\cdot) - g\right) \mathbb{1}_{\mathcal{D}}(\mathsf{A}^{-1}\cdot)$

If $\mathsf{A} \in \mathbb{R}^{c \times d}$ with $c \leq d$ is **full rank**, consider invertible extension $\overline{\mathsf{A}} \in \mathbb{R}^{d \times d}$.

Dual methods: $\boldsymbol{\theta}^{n+\frac{1}{2}} = \mathsf{A}^{-1}\tilde{\mu}(\tilde{\boldsymbol{\theta}}^n) + \sqrt{2\gamma}\xi^{n+1}, \quad \xi^{n+1} \sim \mathcal{N}_d(0, \mathsf{A}^{-1}\mathsf{A}^{-\top})$

Purpose: drift term $\mu(\boldsymbol{\theta})$ adapted to $\pi \propto \exp\left(-f - g(\mathsf{A}\cdot)\right) \mathbb{1}_{\mathcal{D}}$, g non-smooth.

$\mathsf{A} \in \mathbb{R}^{d \times d}$ **invertible**

dual drift term $\tilde{\mu}(\tilde{\boldsymbol{\theta}})$, $\tilde{\boldsymbol{\theta}} = \mathsf{A}\boldsymbol{\theta}$, adapted to $\tilde{\pi} \propto \exp\left(-f(\mathsf{A}^{-1}\cdot) - g\right) \mathbb{1}_{\mathcal{D}}(\mathsf{A}^{-1}\cdot)$

If $\mathsf{A} \in \mathbb{R}^{c \times d}$ with $c \leq d$ is **full rank**, consider invertible extension $\overline{\mathsf{A}} \in \mathbb{R}^{d \times d}$.

Dual methods: $\boldsymbol{\theta}^{n+\frac{1}{2}} = \mathsf{A}^{-1}\tilde{\mu}(\tilde{\boldsymbol{\theta}}^n) + \sqrt{2\gamma}\xi^{n+1}, \quad \xi^{n+1} \sim \mathcal{N}_d(0, \mathsf{A}^{-1}\mathsf{A}^{-\top})$

- Dual Moreau drift:

$$\tilde{\mu}^{\mathtt{M}}(\tilde{\boldsymbol{\theta}}) = \tilde{\boldsymbol{\theta}} - \gamma \mathsf{A}^{-\top}\nabla f(\mathsf{A}^{-1}\tilde{\boldsymbol{\theta}}) - \frac{\gamma}{\rho}(\mathsf{I} - \mathrm{prox}_{\rho g})\tilde{\boldsymbol{\theta}}, \quad \rho = \gamma$$

Purpose: drift term $\mu(\boldsymbol{\theta})$ adapted to $\pi \propto \exp\left(-f - g(\mathsf{A}\cdot)\right) \mathbb{1}_{\mathcal{D}}$, g non-smooth.

$\mathsf{A} \in \mathbb{R}^{d \times d}$ **invertible**

dual drift term $\tilde{\mu}(\tilde{\boldsymbol{\theta}})$, $\tilde{\boldsymbol{\theta}} = \mathsf{A}\boldsymbol{\theta}$, adapted to $\tilde{\pi} \propto \exp\left(-f(\mathsf{A}^{-1}\cdot) - g\right) \mathbb{1}_{\mathcal{D}}(\mathsf{A}^{-1}\cdot)$

If $\mathsf{A} \in \mathbb{R}^{c \times d}$ with $c \leq d$ is **full rank**, consider invertible extension $\overline{\mathsf{A}} \in \mathbb{R}^{d \times d}$.

Dual methods: $\boldsymbol{\theta}^{n+\frac{1}{2}} = \mathsf{A}^{-1}\tilde{\mu}(\tilde{\boldsymbol{\theta}}^n) + \sqrt{2\gamma}\xi^{n+1}, \quad \xi^{n+1} \sim \mathcal{N}_d(0, \mathsf{A}^{-1}\mathsf{A}^{-\top})$

- Dual Moreau drift:

$$\tilde{\mu}^{\mathtt{M}}(\tilde{\boldsymbol{\theta}}) = \tilde{\boldsymbol{\theta}} - \gamma\mathsf{A}^{-\top}\nabla f(\mathsf{A}^{-1}\tilde{\boldsymbol{\theta}}) - \frac{\gamma}{\rho}(\mathsf{I} - \mathrm{prox}_{\rho g})\tilde{\boldsymbol{\theta}}, \quad \rho = \gamma$$

- PGdual drift:

$$\tilde{\mu}^{\mathtt{PG}}(\tilde{\boldsymbol{\theta}}) = \mathrm{prox}_{\gamma g}\left(\tilde{\boldsymbol{\theta}} - \gamma\,\mathsf{A}^{-\top}\nabla f(\mathsf{A}^{-1}\tilde{\boldsymbol{\theta}})\right)$$

(Artigas, 2022, *EUSIPCO*; Fort et al., 2023, IEEE Trans Sig Process)

Purpose: drift term $\mu(\boldsymbol{\theta})$ adapted to $\pi \propto \exp\left(-f - g(\mathsf{A}\cdot)\right) \mathbb{1}_{\mathcal{D}}$, g non-smooth.

$\mathsf{A} \in \mathbb{R}^{d \times d}$ **invertible**

dual drift term $\tilde{\mu}(\tilde{\boldsymbol{\theta}})$, $\tilde{\boldsymbol{\theta}} = \mathsf{A}\boldsymbol{\theta}$, adapted to $\tilde{\pi} \propto \exp\left(-f(\mathsf{A}^{-1}\cdot) - g\right) \mathbb{1}_{\mathcal{D}}(\mathsf{A}^{-1}\cdot)$

If $\mathsf{A} \in \mathbb{R}^{c \times d}$ with $c \leq d$ is **full rank**, consider invertible extension $\overline{\mathsf{A}} \in \mathbb{R}^{d \times d}$.

Dual methods: $\boldsymbol{\theta}^{n+\frac{1}{2}} = \mathsf{A}^{-1}\tilde{\mu}(\tilde{\boldsymbol{\theta}}^n) + \sqrt{2\gamma}\xi^{n+1}, \quad \xi^{n+1} \sim \mathcal{N}_d(0, \mathsf{A}^{-1}\mathsf{A}^{-\top})$

- Dual Moreau drift:

$$\tilde{\mu}^{\mathrm{M}}(\tilde{\boldsymbol{\theta}}) = \tilde{\boldsymbol{\theta}} - \gamma\mathsf{A}^{-\top}\nabla f(\mathsf{A}^{-1}\tilde{\boldsymbol{\theta}}) - \frac{\gamma}{\rho}(\mathsf{I} - \mathrm{prox}_{\rho g})\tilde{\boldsymbol{\theta}}, \quad \rho = \gamma$$

- PGdual drift:

$$\tilde{\mu}^{\mathrm{PG}}(\tilde{\boldsymbol{\theta}}) = \mathrm{prox}_{\gamma g}\left(\tilde{\boldsymbol{\theta}} - \gamma\,\mathsf{A}^{-\top}\nabla f(\mathsf{A}^{-1}\tilde{\boldsymbol{\theta}})\right)$$

(Artigas, 2022, *EUSIPCO*; Fort et al., 2023, IEEE Trans Sig Process)

- Dual Random Walk drift: $\tilde{\mu}^{\mathrm{RM}}(\tilde{\boldsymbol{\theta}}) = \tilde{\boldsymbol{\theta}}$

Toy example: MCMC for penalized logistic regression

Model
- $\mathsf{X} \in \mathbb{R}^{N \times d}$: covariates matrix,
- $\boldsymbol{\theta}^* \in \mathbb{R}^d$: piecewise constant regression vector,
- $\mathsf{Y} \in \{0, 1\}^N$: binary response vector

$$\mathsf{Y}_j \sim \text{Bernoulli}\left((1 + \exp(-(\mathsf{X}\boldsymbol{\theta}^*)_j))^{-1} \right), \quad \text{independent.}$$

Toy example: MCMC for penalized logistic regression

Model

- $\mathsf{X} \in \mathbb{R}^{N \times d}$: covariates matrix,
- $\boldsymbol{\theta}^* \in \mathbb{R}^d$: piecewise constant regression vector,
- $\mathsf{Y} \in \{0,1\}^N$: binary response vector

$$\mathsf{Y}_j \sim \text{Bernoulli}\left((1+\exp(-(\mathsf{X}\boldsymbol{\theta}^*)_j))^{-1}\right), \quad \text{independent.}$$

A posteriori log-distribution $\mathsf{D}_1 \in \mathbb{R}^{d-1 \times d}$: discrete gradient

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

Toy example: MCMC for penalized logistic regression

Model

- $\mathsf{X} \in \mathbb{R}^{N\times d}$: covariates matrix,
- $\boldsymbol{\theta}^* \in \mathbb{R}^d$: piecewise constant regression vector,
- $\mathsf{Y} \in \{0,1\}^N$: binary response vector

$$\mathsf{Y}_j \sim \text{Bernoulli}\left((1+\exp(-(\mathsf{X}\boldsymbol{\theta}^*)_j))^{-1}\right), \quad \text{independent.}$$

A posteriori log-distribution $\mathsf{D}_1 \in \mathbb{R}^{d-1\times d}$: discrete gradient

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

`PGdec` : $\|\mathsf{D}_1\boldsymbol{\theta}\|_1 = \underbrace{\|\mathsf{D}_{1,e}\boldsymbol{\theta}\|_1}_{\text{even rows}} + \underbrace{\|\mathsf{D}_{1,o}\boldsymbol{\theta}\|_1}_{\text{odd rows}},$ $\mathsf{D}_{1,e}\mathsf{D}_{1,e}^\top = \mathsf{D}_{1,o}\mathsf{D}_{1,o}^\top = \mathsf{I};$

Toy example: MCMC for penalized logistic regression

Model

- $\mathsf{X} \in \mathbb{R}^{N\times d}$: covariates matrix,
- $\boldsymbol{\theta}^* \in \mathbb{R}^{d}$: piecewise constant regression vector,
- $\mathsf{Y} \in \{0,1\}^{N}$: binary response vector

$$\mathsf{Y}_j \sim \text{Bernoulli}\left((1+\exp(-(\mathsf{X}\boldsymbol{\theta}^*)_j))^{-1}\right), \quad \text{independent.}$$

A posteriori log-distribution $\mathsf{D}_1 \in \mathbb{R}^{d-1\times d}$: discrete gradient

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

`PGdec` : $\|\mathsf{D}_1\boldsymbol{\theta}\|_1 = \underbrace{\|\mathsf{D}_{1,e}\boldsymbol{\theta}\|_1}_{\text{even rows}} + \underbrace{\|\mathsf{D}_{1,o}\boldsymbol{\theta}\|_1}_{\text{odd rows}},$ $\quad \mathsf{D}_{1,e}\mathsf{D}_{1,e}^\top = \mathsf{D}_{1,o}\mathsf{D}_{1,o}^\top = \mathsf{I};$

`*dual` : $\overline{\mathsf{D}}_1 = \begin{pmatrix} -1 & 0\dots0 \\ 0 & \\ \vdots & \mathsf{D}_1 \\ 0 & \end{pmatrix} \Longrightarrow$ invertible extension of D_1.

Toy example: MCMC for penalized logistic regression

Model
- $\mathsf{X} \in \mathbb{R}^{N\times d}$: covariates matrix,
- $\boldsymbol{\theta}^* \in \mathbb{R}^d$: piecewise constant regression vector,
- $\mathsf{Y} \in \{0,1\}^N$: binary response vector

$$\mathsf{Y}_j \sim \text{Bernoulli}\left((1+\exp(-(\mathsf{X}\boldsymbol{\theta}^*)_j))^{-1}\right), \quad \text{independent.}$$

A posteriori log-distribution $\qquad \mathsf{D}_1 \in \mathbb{R}^{d-1\times d}$: discrete gradient

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

`PGdec` : $\|\mathsf{D}_1\boldsymbol{\theta}\|_1 = \underbrace{\|\mathsf{D}_{1,e}\boldsymbol{\theta}\|_1}_{\text{even rows}} + \underbrace{\|\mathsf{D}_{1,o}\boldsymbol{\theta}\|_1}_{\text{odd rows}},$ $\qquad \mathsf{D}_{1,e}\mathsf{D}_{1,e}^\top = \mathsf{D}_{1,o}\mathsf{D}_{1,o}^\top = \mathsf{I};$

`*dual` : $\overline{\mathsf{D}}_1 = \begin{pmatrix} -1 & 0\dots0 \\ 0 & \\ \vdots & \mathsf{D}_1 \\ 0 & \end{pmatrix} \Longrightarrow$ invertible extension of D_1.

Data $\quad N = 2.10^3$, $d = 20$

X: independent Rademacher r.v., rows normalized to 1.

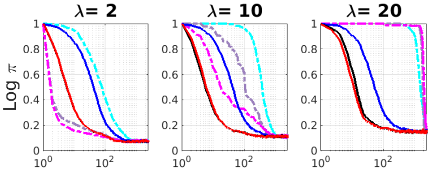
Convergence indicator:

$$\text{Log}\,\pi = \frac{\ln \pi_t(\boldsymbol{\theta}^n) - \ln \pi_t^*}{\ln \pi_t(\boldsymbol{\theta}^1) - \ln \pi_t^*}, \qquad \ln \pi_t^* = \max_{\boldsymbol{\theta} \in \mathbb{R}} \ln \pi_t(\boldsymbol{\theta})$$

high probability regions

Comparaison of the MCMC samplers

primal *dashed lines*	dual *solid lines*
`RW`	`RWdual`
`M`	`Mdual`
`PGdec`	`PGdual`



- gain to use 1^{st} order information vs. `RW`;
- primal samplers: the fastest at small λ;
- dual samplers: the fastest for medium to large λ, good for small λ.

$\Longrightarrow$ `Mdual` and `PGdual` fast convergence; robust to the choice of λ

Covid19 propagation model: $\boldsymbol{\theta} = (\mathbf{R}, \mathbf{O})$ of probability distribution

$$\pi(\boldsymbol{\theta}) \propto \exp\left(-\sum_{t=1}^{T}\left(-\mathsf{Z}_t \ln \mathsf{p}_t(\boldsymbol{\theta}) + \mathsf{p}_t(\boldsymbol{\theta})\right) - \lambda_{\mathsf{R}}\|\mathsf{D}_2\mathbf{R}\|_1 - \lambda_{\mathsf{O}}\|\mathbf{O}\|_1\right)\mathbb{1}_{\mathcal{D}}(\boldsymbol{\theta})$$

$$\mathsf{D}_2 \in \mathbb{R}^{(T-2)\times T} \ \textbf{full rank} \quad \Longrightarrow \quad \overline{\mathsf{D}}_2 \in \mathbb{R}^{T\times T} \ \textbf{invertible} \text{ extension}$$

Covid19 propagation model: $\boldsymbol{\theta} = (\mathbf{R}, \mathbf{O})$ of probability distribution

$$\pi(\boldsymbol{\theta}) \propto \exp\left(-\sum_{t=1}^{T}\left(-\mathsf{Z}_t \ln \mathsf{p}_t(\boldsymbol{\theta}) + \mathsf{p}_t(\boldsymbol{\theta})\right) - \lambda_{\mathsf{R}}\|\mathsf{D}_2\mathbf{R}\|_1 - \lambda_{\mathsf{O}}\|\mathbf{O}\|_1\right)\mathbb{1}_{\mathcal{D}}(\boldsymbol{\theta})$$

$\mathsf{D}_2 \in \mathbb{R}^{(T-2)\times T}$ **full rank** $\Longrightarrow$ $\overline{\mathsf{D}}_2 \in \mathbb{R}^{T\times T}$ **invertible** extension

MCMC dual samplers: [`RWdual`] random walk in the dual space
[`Mdual`] Moreau drift in the dual space
[`PGdual`] proximal-gradient type drift in the dual space

$$\mathbf{R}^{n+\frac{1}{2}} = \begin{cases} \mathbf{R}^n \\ \overline{\mathsf{D}}_2^{-1}\tilde{\mu}_{\mathsf{R}}^{\mathtt{M}}(\tilde{\boldsymbol{\theta}}^n) \\ \overline{\mathsf{D}}_2^{-1}\tilde{\mu}_{\mathsf{R}}^{\mathtt{PG}}(\tilde{\boldsymbol{\theta}}^n) \end{cases} + \sqrt{2\gamma_{\mathsf{R}}}\xi_{\mathsf{R}}^{n+1}; \quad \mathbf{O}^{n+\frac{1}{2}} = \begin{cases} \mathbf{O}^n \\ \tilde{\mu}_{\mathsf{O}}^{\mathtt{M}}(\tilde{\boldsymbol{\theta}}^n) \\ \tilde{\mu}_{\mathsf{O}}^{\mathtt{PG}}(\tilde{\boldsymbol{\theta}}^n) \end{cases} + \sqrt{2\gamma_{\mathsf{O}}}\xi_{\mathsf{O}}^{n+1}.$$

Covid19 propagation model: $\boldsymbol{\theta} = (\mathbf{R}, \mathbf{O})$ of probability distribution

$$\pi(\boldsymbol{\theta}) \propto \exp\left(-\sum_{t=1}^{T}\left(-\mathsf{Z}_t \ln \mathsf{p}_t(\boldsymbol{\theta}) + \mathsf{p}_t(\boldsymbol{\theta})\right) - \lambda_{\mathsf{R}}\|\mathsf{D}_2\mathbf{R}\|_1 - \lambda_{\mathsf{O}}\|\mathbf{O}\|_1\right)\mathbb{1}_{\mathcal{D}}(\boldsymbol{\theta})$$

$\mathsf{D}_2 \in \mathbb{R}^{(T-2)\times T}$ **full rank** $\implies$ $\overline{\mathsf{D}}_2 \in \mathbb{R}^{T\times T}$ **invertible** extension

MCMC dual samplers:

[RWdual] random walk in the dual space
[Mdual] Moreau drift in the dual space
[PGdual] proximal-gradient type drift in the dual space

$$\mathbf{R}^{n+\frac{1}{2}} = \begin{cases} \mathbf{R}^n \\ \overline{\mathsf{D}}_2^{-1}\tilde{\mu}_{\mathsf{R}}^{\mathtt{M}}(\tilde{\boldsymbol{\theta}}^n) \\ \overline{\mathsf{D}}_2^{-1}\tilde{\mu}_{\mathsf{R}}^{\mathtt{PG}}(\tilde{\boldsymbol{\theta}}^n) \end{cases} + \sqrt{2\gamma_{\mathsf{R}}}\xi_{\mathsf{R}}^{n+1}; \quad \mathbf{O}^{n+\frac{1}{2}} = \begin{cases} \mathbf{O}^n \\ \tilde{\mu}_{\mathsf{O}}^{\mathtt{M}}(\tilde{\boldsymbol{\theta}}^n) \\ \tilde{\mu}_{\mathsf{O}}^{\mathtt{PG}}(\tilde{\boldsymbol{\theta}}^n) \end{cases} + \sqrt{2\gamma_{\mathsf{O}}}\xi_{\mathsf{O}}^{n+1}.$$

Gaussian perturbation:

- $\xi_{\mathsf{R}}^{n+1} \sim \mathcal{N}(0, \overline{\mathsf{D}}_2^{-1}\overline{\mathsf{D}}_2^{-\top})$,
- $\xi_{\mathsf{O}}^{n+1} \sim \mathcal{N}(0, \mathsf{I})$;

Hyperparameters:

- $(\lambda_{\mathsf{R}}, \lambda_{\mathsf{O}}) = (3.5\,\sigma_{\mathbf{Z}}\sqrt{6}/4, 0.05)$,
- $\gamma_{\mathsf{O}} = \gamma(\lambda_{\mathsf{R}}/\lambda_{\mathsf{O}})^2$,
- γ adjusted to reach 25% acceptance rate.

Covid19 propagation model: $\boldsymbol{\theta} = (\mathbf{R}, \mathbf{O})$ of probability distribution

$$\pi(\boldsymbol{\theta}) \propto \exp\left(-\sum_{t=1}^{T}\left(-\mathsf{Z}_t \ln \mathsf{p}_t(\boldsymbol{\theta}) + \mathsf{p}_t(\boldsymbol{\theta})\right) - \lambda_{\mathsf{R}}\|\mathsf{D}_2\mathbf{R}\|_1 - \lambda_{\mathsf{O}}\|\mathbf{O}\|_1\right)\mathbb{1}_{\mathcal{D}}(\boldsymbol{\theta})$$

$$\mathsf{D}_2 \in \mathbb{R}^{(T-2)\times T} \text{ **full rank** } \quad \Longrightarrow \quad \overline{\mathsf{D}}_2 \in \mathbb{R}^{T\times T} \text{ **invertible** extension}$$

MCMC dual samplers:
- [RWdual] random walk in the dual space
- [Mdual] Moreau drift in the dual space
- [PGdual] proximal-gradient type drift in the dual space

$$\mathbf{R}^{n+\frac{1}{2}} = \begin{cases} \mathbf{R}^n \\ \overline{\mathsf{D}}_2^{-1}\tilde{\mu}_{\mathsf{R}}^{\mathtt{M}}(\tilde{\boldsymbol{\theta}}^n) \\ \overline{\mathsf{D}}_2^{-1}\tilde{\mu}_{\mathsf{R}}^{\mathtt{PG}}(\tilde{\boldsymbol{\theta}}^n) \end{cases} + \sqrt{2\gamma_{\mathsf{R}}}\xi_{\mathsf{R}}^{n+1}; \quad \mathbf{O}^{n+\frac{1}{2}} = \begin{cases} \mathbf{O}^n \\ \tilde{\mu}_{\mathsf{O}}^{\mathtt{M}}(\tilde{\boldsymbol{\theta}}^n) \\ \tilde{\mu}_{\mathsf{O}}^{\mathtt{PG}}(\tilde{\boldsymbol{\theta}}^n) \end{cases} + \sqrt{2\gamma_{\mathsf{O}}}\xi_{\mathsf{O}}^{n+1}.$$

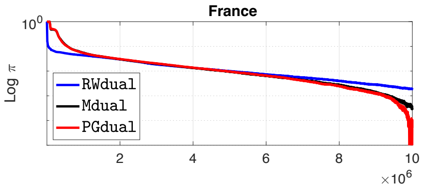
Gaussian perturbation:

- $\xi_{\mathsf{R}}^{n+1} \sim \mathcal{N}(0, \overline{\mathsf{D}}_2^{-1}\overline{\mathsf{D}}_2^{-\top})$,
- $\xi_{\mathsf{O}}^{n+1} \sim \mathcal{N}(0, \mathsf{I})$;

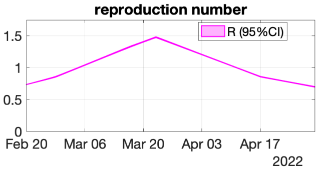
Hyperparameters:

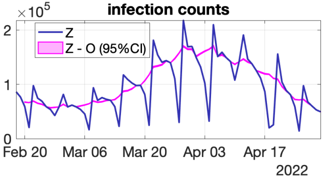
- $(\lambda_{\mathsf{R}}, \lambda_{\mathsf{O}}) = (3.5\,\sigma_{\mathbf{Z}}\sqrt{6}/4, 0.05)$,
- $\gamma_{\mathsf{O}} = \gamma(\lambda_{\mathsf{R}}/\lambda_{\mathsf{O}})^2$,
- γ adjusted to reach 25% acceptance rate.

Convergence of the Markov chains



Credibility intervals of regularized reproduction number and corrected counts



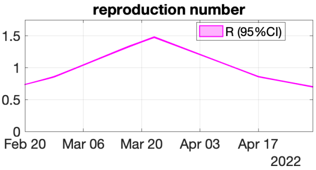


CI estimates of R_t and $Z_t^{(D)}$ published daily for 200+ countries during $\sim$2 years

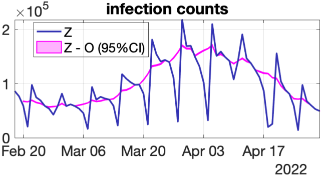
`https://perso.ens-lyon.fr/patrice.abry/`

`https://perso.math.univ-toulouse.fr/gfort/project/opsimore-2/`

Credibility intervals of regularized reproduction number and corrected counts



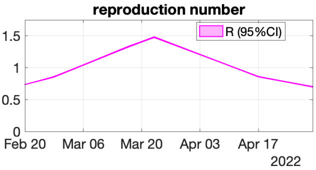
$\lambda_{\mathsf{R}} = 3.5 \times \mathrm{std}(\mathbf{Z})$



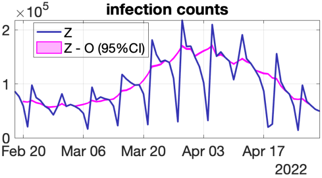
$\lambda_{\mathsf{O}} = 0.05$

$$\pi(\boldsymbol{\theta}) \propto \exp\left(-\sum_{t=1}^{T}\left(-\mathsf{Z}_t \ln \mathsf{p}_t(\boldsymbol{\theta}) + \mathsf{p}_t(\boldsymbol{\theta})\right) - \lambda_{\mathsf{R}}\|\mathsf{D}_2\mathbf{R}\|_1 - \lambda_{\mathsf{O}}\|\mathbf{O}\|_1\right)\mathbb{1}_{\mathcal{D}}(\boldsymbol{\theta});$$

Credibility intervals of regularized reproduction number and corrected counts

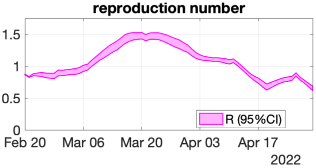


$\lambda_{\mathsf{R}} = 3.5 \times \mathrm{std}(\mathbf{Z})$

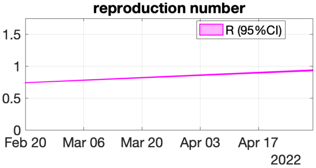


$\lambda_{\mathsf{O}} = 0.05$

$$\pi(\boldsymbol{\theta}) \propto \exp\left(-\sum_{t=1}^{T} \left(-\mathsf{Z}_t \ln \mathsf{p}_t(\boldsymbol{\theta}) + \mathsf{p}_t(\boldsymbol{\theta})\right) - \lambda_{\mathsf{R}} \|\mathsf{D}_2 \mathbf{R}\|_1 - \lambda_{\mathsf{O}} \|\mathbf{O}\|_1 \right) \mathbb{1}_{\mathcal{D}}(\boldsymbol{\theta});$$

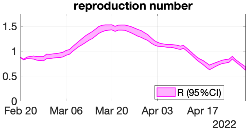


$\lambda_{\mathsf{R}} = 0.5 \times \mathrm{std}(\mathbf{Z})$

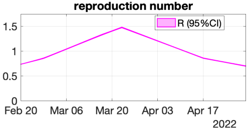


$\lambda_{\mathsf{R}} = 500 \times \mathrm{std}(\mathbf{Z})$

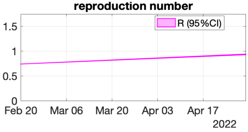
Expert choice: somehow arbitrary, subjective and costly



$\lambda_R = 0.5 \times \mathrm{std}(\mathbf{Z})$



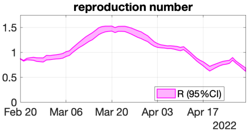
$\lambda_R = 3.5 \times \mathrm{std}(\mathbf{Z})$



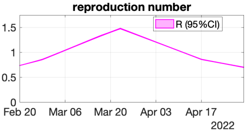
$\lambda_R = 500 \times \mathrm{std}(\mathbf{Z})$

Extension: a Hierarchical Bayesian epidemiological model

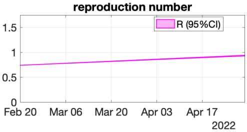
Expert choice: somehow arbitrary, subjective and costly



$\lambda_{\mathrm{R}} = 0.5 \times \mathrm{std}(\mathbf{Z})$



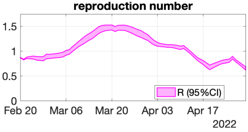
$\lambda_{\mathrm{R}} = 3.5 \times \mathrm{std}(\mathbf{Z})$



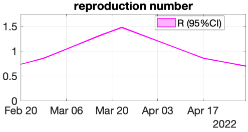
$\lambda_{\mathrm{R}} = 500 \times \mathrm{std}(\mathbf{Z})$

$$\ln \pi(\boldsymbol{\theta}|\boldsymbol{\lambda}) = \sum_{t=1}^{T} \left(\mathrm{Z}_t \ln \mathrm{p}_t(\boldsymbol{\theta}) - \mathrm{p}_t(\boldsymbol{\theta})\right) - \lambda_{\mathrm{R}} \|\mathrm{D}_2 \mathbf{R}\|_1 - \lambda_{\mathrm{O}} \|\mathbf{O}\|_1 + T \ln \lambda_{\mathrm{R}} + T \ln \lambda_{\mathrm{O}} + \mathrm{C}$$

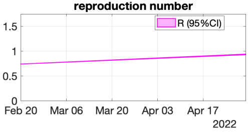
Expert choice: somehow arbitrary, subjective and costly



$\lambda_R = 0.5 \times \mathrm{std}(\mathbf{Z})$



$\lambda_R = 3.5 \times \mathrm{std}(\mathbf{Z})$

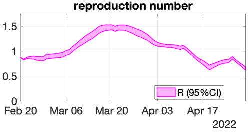


$\lambda_R = 500 \times \mathrm{std}(\mathbf{Z})$

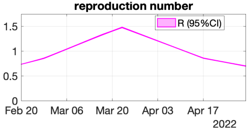
$$\ln \pi(\boldsymbol{\theta}|\boldsymbol{\lambda}) = \sum_{t=1}^{T} (Z_t \ln p_t(\boldsymbol{\theta}) - p_t(\boldsymbol{\theta})) - \lambda_R \|D_2 \mathbf{R}\|_1 - \lambda_O \|\mathbf{O}\|_1 + T \ln \lambda_R + T \ln \lambda_O + C$$

Purpose: account for uncertainty in the choice of the priors $\boldsymbol{\lambda} := (\lambda_R, \lambda_O)$

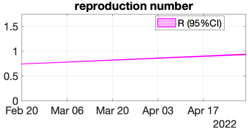
Expert choice: somehow arbitrary, subjective and costly



$\lambda_R = 0.5 \times \mathrm{std}(\mathbf{Z})$



$\lambda_R = 3.5 \times \mathrm{std}(\mathbf{Z})$



$\lambda_R = 500 \times \mathrm{std}(\mathbf{Z})$

$$\ln \pi(\boldsymbol{\theta}|\boldsymbol{\lambda}) = \sum_{t=1}^{T} (Z_t \ln p_t(\boldsymbol{\theta}) - p_t(\boldsymbol{\theta})) - \lambda_R \|D_2 \mathbf{R}\|_1 - \lambda_O \|\mathbf{O}\|_1 + T \ln \lambda_R + T \ln \lambda_O + C$$

Purpose: account for uncertainty in the choice of the priors $\boldsymbol{\lambda} := (\lambda_R, \lambda_O)$

$\Longrightarrow$ Conjugate Gamma hyperpriors: $\lambda_R \sim \Gamma(\alpha_R, \beta_R)$, $\lambda_O \sim \Gamma(\alpha_O, \beta_O)$

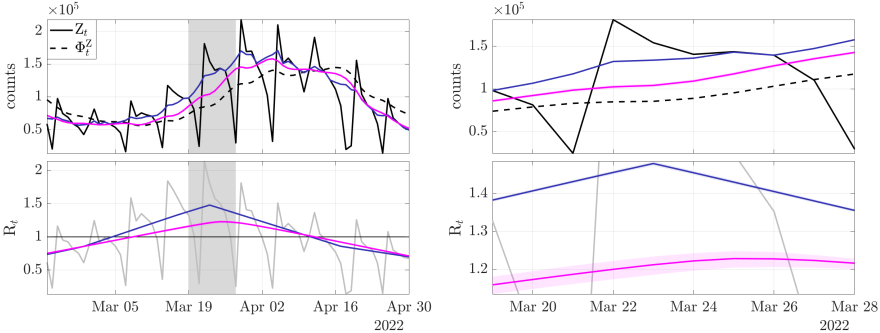
$$\lambda_R \mid \mathbf{R}, \mathbf{O} \sim \Gamma(T + \alpha_R, \|D_2 \mathbf{R}\|_1 + \beta_R), \quad \lambda_O \mid \mathbf{R}, \mathbf{O} \sim \Gamma(T + \alpha_O, \|\mathbf{O}\|_1 + \beta_O)$$

closed-form expression with direct sampling: Gibbs sampler alternating `PGdual` and Γ

Extension: a Hierarchical Bayesian epidemiological model

Principle: explore jointly the distribution of $(\mathbf{R}, \mathbf{O}, \lambda_\mathsf{R}, \lambda_\mathsf{O})$

$\Longrightarrow$ let $\boldsymbol{\lambda}$ adapt to data by varying around $(3.5 \times \mathrm{std}(\mathbf{Z}), 0.05)$



Area covered by the credibility intervals of **R**:

- `PGdual` targeting $\pi(\boldsymbol{\theta}|\boldsymbol{\lambda})$ by sampling $(\mathbf{R}, \mathbf{O})$: 0.46 ± 0.01
- `Gibbs` targeting $\pi(\boldsymbol{\theta}, \boldsymbol{\lambda})$ by sampling $(\mathbf{R}, \mathbf{O}, \lambda_\mathsf{R}, \lambda_\mathsf{O})$: 1.35 ± 0.07

Hierarchical distribution $\pi(\boldsymbol{\theta}, \boldsymbol{\lambda})$: more flexible, captures better intrinsic data variability

Conclusion and perspectives

Take home messages:

- Proximal Langevin based MCMC sampler for

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

  π composite distributions with constrained support

Take home messages:

- Proximal Langevin based MCMC sampler for

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

  π composite distributions with constrained support

- Speed of convergence on a toy example: faster when
  - taking into account first order information on $\ln \pi$
  - using adequate covariance in the Gaussian proposal

Take home messages:

- Proximal Langevin based MCMC sampler for

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

  π composite distributions with constrained support

- Speed of convergence on a toy example: faster when
  - taking into account first order information on $\ln \pi$
  - using adequate covariance in the Gaussian proposal

- Robust hierarchical Bayesian estimate: account better for uncertainty

Abry et al., EUSIPCO (2023); Fort et al., IEEE Trans. Signal Process. (2023);
Abry et al., CAMSAP (2023); Abry et al., Preprint (2024) with codes

bpascal-fr.github.io & github.com/gfort-lab

Conclusion and perspectives

Take home messages:

- Proximal Langevin based MCMC sampler for

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

  π composite distributions with constrained support

- Speed of convergence on a toy example: faster when
  - taking into account first order information on $\ln \pi$
  - using adequate covariance in the Gaussian proposal

- Robust hierarchical Bayesian estimate: account better for uncertainty

Abry et al., EUSIPCO (2023); Fort et al., IEEE Trans. Signal Process. (2023);
Abry et al., CAMSAP (2023); Abry et al., Preprint (2024) with codes

bpascal-fr.github.io & github.com/gfort-lab

Ongoing work:

- Online estimation scheme to process new counts on the fly
- New epidemiological models for low quality data: π not log-concave
- Monitoring different territories: spatiotemporal dynamics through graph inference