How to use `PGdual`, `GibbsPGdual` and `PGdec`

Gersende Fort *

Contents

*CNRS, Institut de Mathématiques de Toulouse,
118 route de Narbonne 31400 Toulouse, France
gersende.fort@math.univ-toulouse.fr

1 Associated publications

The algorithms `PGdual`, `GibbsPGdual` and `PGdec` were introduced in the papers

1. *Temporal Evolution of the Covid19 pandemic reproduction number: Estimations from Proximal optimization to Monte Carlo sampling*, by P. Abry, G. Fort, B. Pascal and N. Pustelnik. Accepted for publication in EMBC 2022 proceedings.

2. *Credibility Interval Design for Covid19 Reproduction Number from nonsmooth Langevin-type Monte Carlo sampling*, by H. Artigas, B. Pascal, G. Fort, P. Abry and N. Pustelnik. Accepted for publication in EUSIPCO 2022 proceedings.

3. *Estimation et Intervalles de crédibilité pour le taux de reproduction de la Covid19 par échantillonnage Monte Carlo Langevin Proximal*, by P. Abry, G. Fort, B. Pascal and N. Pustelnik. Accepted for publication in GRETSI 2022 proceedings.

4. *Credibility intervals for the reproduction number of the Covid-19 pandemic using Proximal Lanvevin samplers*, by P. Abry, G. Fort, B. Pascal and N. Pustelnik. Submitted.

5. *Covid19 Reproduction Number: Credibility Intervals by Blockwise Proximal Monte Carlo samplers*, by G. Fort, B. Pascal, P. Abry and N. Pustelnik. Submitted

2 The target distribution π and its image $\tilde{\pi}$ by $\overline{\mathsf{A}}$

Notations. Set

$$f(\theta) \stackrel{\text{def}}{=} -\sum_{t=1}^{T} \left(\mathsf{Z}_t \ln(\mathsf{R}_t + \mathsf{O}_t) - \Phi_t(\mathsf{R}_t + \mathsf{O}_t)\right),$$

and

$$\mathbf{R} \stackrel{\text{def}}{=} \begin{bmatrix} \mathsf{R}_1 \\ \mathsf{R}_2 \\ \cdots \\ \mathsf{R}_T \end{bmatrix}, \qquad \mathbf{O} \stackrel{\text{def}}{=} \begin{bmatrix} \mathsf{O}_1 \\ \mathsf{O}_2 \\ \cdots \\ \mathsf{O}_T \end{bmatrix}, \qquad \theta = \begin{bmatrix} \mathbf{R} \\ \mathbf{O} \end{bmatrix}, \qquad \Phi_t \stackrel{\text{def}}{=} \sum_{s=1}^{\tau_\phi} \phi(s) \mathsf{Z}_{t-s}.$$

Define the subset of $(\mathbb{R}_+)^T \times \mathbb{R}^T$

$$\mathcal{D} \stackrel{\text{def}}{=} \bigcap_t \{(\mathsf{R}_t, \mathsf{O}_t) \in \mathbb{R}_+ \times \mathbb{R} : \mathsf{R}_t + \mathsf{O}_t > 0 \text{ if } \mathsf{Z}_t > 0, \text{ or } \ \mathsf{R}_t + \mathsf{O}_t \geq 0 \text{ if } \mathsf{Z}_t \geq 0\}.$$

2.1 The target distribution π.

The target distribution is

$$\pi(\theta) \propto \exp\left(-f(\theta) - \lambda_{\mathsf{R}} \|\mathsf{D}_2 \mathbf{R}\|_1 - \lambda_{\mathsf{O}} \|\mathrm{Diag}(\Phi.) \mathbf{O}\|_1\right) \mathbb{1}_{\theta \in \mathcal{D}} = \exp\left(-f(\theta) - \lambda_{\mathsf{R}} \|\mathsf{A}\theta\|_1\right) \mathbb{1}_{\theta \in \mathcal{D}}$$

where

$$\mathsf{A} \stackrel{\text{def}}{=} \begin{bmatrix} \mathsf{D}_2 & 0_{(T-2)\times T} \\ 0_{T\times T} & \frac{\lambda_{\mathsf{O}}}{\lambda_{\mathsf{R}}} \mathrm{Diag}(\Phi.) \end{bmatrix} \in \mathbb{R}^{(2T-2)\times(2T)}.$$

2.2 Image $\tilde{\pi}$ of π by $\overline{\mathsf{A}}$.

Let $\overline{\mathsf{D}}$ be a $T \times T$ invertible matrix such that the rows 3 to T are equal to D_2. Then $(\overline{\mathsf{D}}\mathbf{R})_{3:T} = \mathsf{D}_2 \mathbf{R}$. Set

$$\overline{\mathsf{A}} \stackrel{\text{def}}{=} \begin{bmatrix} \overline{\mathsf{D}} & 0_{T\times T} \\ 0_{T\times T} & \frac{\lambda_{\mathsf{O}}}{\lambda_{\mathsf{R}}} \mathrm{Diag}(\Phi.) \end{bmatrix} \in \mathbb{R}^{(2T)\times(2T)}.$$

The image of π by $\overline{\mathsf{A}}$ is

$$\tilde{\pi}(\tilde{\theta}) \propto \exp\left(-f(\overline{\mathsf{A}}^{-1}\tilde{\theta}) - \lambda_{\mathsf{R}} \|\tilde{\theta}_{3:(2T)}\|_1\right) \mathbb{1}_{\mathcal{D}}(\overline{\mathsf{A}}^{-1}\tilde{\theta}). \tag{1}$$

3 `PGdual` algorithm

3.1 Composite log-density, proximal, gradient

For any $\tilde{\theta} \in \mathbb{R}^{2T}$ such that $\overline{\mathsf{A}}^{-1}\tilde{\theta} \in \mathcal{D}$,

$$\ln \tilde{\pi}(\tilde{\theta}) = -f(\overline{\mathsf{A}}^{-1}\tilde{\theta}) - \lambda_{\mathsf{R}} \|\tilde{\theta}_{3:(2T)}\|_1.$$

The gradient of $\tilde{\theta} \mapsto f(\overline{\mathsf{A}}^{-1}\tilde{\theta})$ is $\overline{\mathsf{A}}^{-\top} \, (\nabla f)\,(\overline{\mathsf{A}}^{-1}\tilde{\theta})$ where

$$(\nabla f)\,(\tau) = -\begin{bmatrix} \frac{\mathsf{Z}_1}{\tau_1+\tau_{T+1}} - \Phi_1 \\ \cdots \\ \frac{\mathsf{Z}_T}{\tau_T+\tau_{2T}} - \Phi_T \\ \frac{\mathsf{Z}_1}{\tau_1+\tau_{T+1}} - \Phi_1 \\ \cdots \\ \frac{\mathsf{Z}_T}{\tau_T+\tau_{2T}} - \Phi_T \end{bmatrix}, \qquad \tau = (\tau_1, \cdots, \tau_{2T}) \in \mathcal{D};$$

by convention, $0/0 = 0$.

Let $\gamma > 0$. The component $\#t$ of the proximal operator associated to the function $\tilde{\theta} \mapsto \gamma\lambda_{\mathsf{R}}\|\tilde{\theta}_{3:(2T)}\|_1$ and evaluated at $\tau = (\tau_1, \cdots, \tau_{2T}) \in \mathbb{R}^{2T}$, is τ_t if $t = 1, 2$; and

$$\mathrm{sign}(\tau_t)\ (|\tau_t| - \gamma\lambda_{\mathsf{R}})_+\,,$$

if $t \geq 3$.

3.2 The proposal mechanism in the dual space

Starting from $\tilde{\theta}$, the candidate for the next step is

$$\mu(\tilde{\theta}) + \sqrt{2\gamma}\,\mathcal{N}_{2T}(0, \mathrm{Id})$$

where

$$\mu(\tilde{\theta}) \overset{\text{def}}{=} \mathrm{Prox}_{\gamma\lambda_{\mathsf{R}}\|(\cdot)_{3:(2T)}\|_1}\Big(\tilde{\theta} - \gamma\,\overline{\mathsf{A}}^{-\top}\,\nabla f(\overline{\mathsf{A}}^{-1}\tilde{\theta})\Big).$$

Interpretations

• For each block $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{O}}$ of $\tilde{\theta}$, the candidate for the next step is

$$\mathrm{Prox}_{\gamma\lambda_{\mathsf{R}}\|(\cdot)_{3:T}\|_1}\Big(\tilde{\mathbf{R}} - \gamma\,\overline{\mathsf{D}}^{-\top}\,\nabla_1 f(\overline{\mathsf{A}}^{-1}\tilde{\theta})\Big),$$
$$\mathrm{Prox}_{\gamma\lambda_{\mathsf{R}}\|\cdot\|_1}\Big(\tilde{\mathbf{O}} - \gamma\,\frac{\lambda_{\mathsf{R}}}{\lambda_{\mathsf{O}}}\,\mathrm{Diag}(1/\Phi.)\,\nabla_2 f(\overline{\mathsf{A}}^{-1}\tilde{\theta})\Big),$$

where ∇_1 and ∇_2 denote the derivative w.r.t. $\mathbf{R}$ and $\mathbf{O}$ respectively.

• In the original space, by setting $\theta \overset{\text{def}}{=} \overline{\mathsf{A}}^{-1}\tilde{\theta}$, the candidate is

$$\overline{\mathsf{A}}^{-1}\mathrm{Prox}_{\gamma\lambda_{\mathsf{R}}\|(\cdot)_{3:(2T)}\|_1}\Big(\overline{\mathsf{A}}\theta - \gamma\,\overline{\mathsf{A}}^{-\top}\,\nabla f(\theta)\Big) + \sqrt{2\gamma}\,\mathcal{N}_{2T}(0, \overline{\mathsf{A}}^{-1}\overline{\mathsf{A}}^{-\top}).$$

Equivalently, for the **R**-block we have

$$\overline{\mathsf{D}}^{-1}\mathrm{Prox}_{\gamma\lambda_{\mathsf{R}}\|(\cdot)_{3:T}\|_1}\Big(\overline{\mathsf{D}}\mathbf{R}-\gamma\,\overline{\mathsf{D}}^{-\top}\,\nabla_1 f(\theta)\Big)+\sqrt{2\gamma}\,\mathcal{N}_T(0,\overline{\mathsf{D}}^{-1}\overline{\mathsf{D}}^{-\top}).$$

For the **O**-block we have

$$\frac{\lambda_{\mathsf{R}}}{\lambda_{\mathsf{O}}}\,\mathrm{Diag}(1/\Phi_{\cdot})\,\mathrm{Prox}_{\gamma\lambda_{\mathsf{R}}\|\cdot\|_1}\Big(\frac{\lambda_{\mathsf{O}}}{\lambda_{\mathsf{R}}}\,\mathrm{Diag}(\Phi_{\cdot})\mathbf{O}-\gamma\,\frac{\lambda_{\mathsf{R}}}{\lambda_{\mathsf{O}}}\,\mathrm{Diag}(1/\Phi_{\cdot})\,\nabla_2 f(\theta)\Big)+\sqrt{2\gamma}\;\frac{\lambda_{\mathsf{R}}}{\lambda_{\mathsf{O}}}\,\mathcal{N}_T(0,\,\mathrm{Diag}(1/\Phi_{\cdot}^2)),$$

which is equivalent to

$$\Big[\mathrm{Prox}_{\bar{\gamma}_t\lambda_{\mathsf{O}}\Phi_t|\cdot|_1}\Big(\mathsf{O}_t-\bar{\gamma}_t\,\nabla_{\mathsf{O}_t}f(\theta)\Big)+\sqrt{2\bar{\gamma}_t}\mathcal{N}(0,1)\Big]_{1\leq t\leq T};\qquad \bar{\gamma}_t\stackrel{\mathrm{def}}{=}\gamma\frac{\lambda_{\mathsf{R}}^2}{\lambda_{\mathsf{O}}^2}\Phi_t^{-2}.$$

3.3 Pseudo-code

Algorithm 1: `PGdual`

Data: λ_{R}, λ_{O}, T, `NbrMC`, $\theta^0 \in \mathcal{D}$, γ

Result: A $\mathcal{D}$-valued sequence $\{\theta^n, n \in \{0, \ldots, \mathtt{NbrMC}\}\}$

Define D_2, $\overline{\mathsf{A}}$ and $\overline{\mathsf{A}}^{-1}$;

Sample *NbrMC* $\times$ $(2T)$ standard Gaussian random variables, stored in *GaussRnd*; and *NbrMC* uniform random variables on $[0, 1]$, stored in *UnifRnd* ;

Initialize $\tilde{\theta}^0 = \overline{\mathsf{A}}\theta^0$, the chain in the image space. ;

Initialize *logpiCurrent* $= \ln \tilde{\pi}(\tilde{\theta}^0)$;

for $n = 0, \ldots, NbrMC - 1$ **do**

- Propose a candidate *Proposal*, equal to $\mu(\tilde{\theta}^n) + \sqrt{2\gamma}\, GaussRnd(:, n)$;
- **if** *Proposal* $\notin \mathcal{D}$ **then**
  - $\tilde{\theta}^{n+1} = \tilde{\theta}^n$
- **else**
  - Compute the log acceptance-rejection ratio *logalpha*: **begin**
    - compute $\log \tilde{\pi}(Proposal)$;
    - compute *logGaussNum*, the density of a Gaussian distribution with expectation $\mu(Proposal)$, covariance $2\gamma\mathrm{Id}$, at the point $\tilde{\theta}^n$;
    - compute *logGaussDenom* the density of a Gaussian distribution with expectation $\mu(\tilde{\theta}^n)$, covariance $2\gamma\mathrm{Id}$, at the point *Proposal* ;
    - set $logalpha = \log \tilde{\pi}(Proposal) - logpiCurrent + logGaussNum - logGaussDenom$
  - **if** *UnifRnd(n)* $< \exp(logalpha)$ **then**
    - set $\tilde{\theta}^{n+1} = Proposal$;
    - set $logpiCurrent = \log \tilde{\pi}(Proposal)$
  - **else**
    - $\tilde{\theta}^{n+1} = \tilde{\theta}^n$;
- Store $\theta^{n+1} = \overline{\mathsf{A}}^{-1}\tilde{\theta}^{n+1}$;
- Store *logpiCurrent* ;
- Update γ: **begin**
  - If $\mathrm{mod}(nn, \mathtt{param.frequency}) == 1$ and while $n <$ `MCMC.chain_burnin`: increase or decrease γ in order to target a mean acceptance rate `MCMC.target_ratio`.

4 `PGdec` algorithm

We write for all $\theta \in \mathcal{D}$,

$$f(\theta) + \lambda_{\mathsf{R}} \|\mathsf{A}\theta\|_1 = f(\theta) + \sum_{\ell=1}^{3} \lambda_{\mathsf{R}} \|\mathsf{D}_2^{(\ell)} \mathbf{R}\|_1 + \lambda_{\mathsf{O}} \|\mathrm{Diag}(\Phi_{\cdot}) \mathbf{O}\|_1$$

where $\mathsf{D}_2^{(\ell)}$ collects the rows $\ell, \ell+3, \ell+6, \cdots$ of the matrix D_2; let us say that $\mathsf{D}_2^{(\ell)}$ is a $T_\ell \times T$ matrix. Set

$$\mathsf{A}^{(\ell)} \stackrel{\text{def}}{=} \begin{bmatrix} \mathsf{D}_2^{(\ell)} & 0_{T_\ell \times T} \\ 0_{T \times T} & \frac{\lambda_{\mathsf{O}}}{\lambda_{\mathsf{R}}} \mathrm{Diag}(\Phi_{\cdot}) \end{bmatrix}$$

and observe that $f(\theta) + \lambda_{\mathsf{R}} \|\mathsf{A}^{(\ell)}\theta\|_1 = f(\theta) + \lambda_{\mathsf{R}} \|\mathsf{D}_2^{(\ell)} \mathbf{R}\|_1 + \lambda_{\mathsf{O}} \|\mathrm{Diag}(\Phi_{\cdot}) \mathbf{O}\|_1$ is part of $\log \pi$.

Since $\mathsf{D}_2^{(\ell)} \left(\mathsf{D}_2^{(\ell)}\right)^\top = \mathrm{Id}_{T_\ell}$ then

$$\mathsf{A}^{(\ell)} \left(\mathsf{A}^{(\ell)}\right)^\top = \begin{bmatrix} \mathrm{Id}_{T_\ell} & 0_{T_\ell \times T} \\ 0_{T \times T} & \left(\frac{\lambda_{\mathsf{O}}}{\lambda_{\mathsf{R}}}\right)^2 \mathrm{Diag}(\Phi_{\cdot}^2) \end{bmatrix}.$$

4.1 Composite log-density, gradient and proximal

For any $\theta \in \mathcal{D}$, consider the approximation of $\log \pi(\theta)$ given by

$$-f(\theta) - \lambda_{\mathsf{R}} \|\mathsf{A}^{(\ell)}\theta\|_1.$$

The gradient step of the smooth term $-f(\theta)$ is

$$-\nabla f(\theta) = \begin{bmatrix} \frac{\mathsf{Z}_1}{\mathsf{R}_1 + \mathsf{O}_1} - \Phi_1 \\ \cdots \\ \frac{\mathsf{Z}_T}{\mathsf{R}_T + \mathsf{O}_T} - \Phi_T \\ \frac{\mathsf{Z}_1}{\mathsf{R}_1 + \mathsf{O}_1} - \Phi_1 \\ \cdots \\ \frac{\mathsf{Z}_T}{\mathsf{R}_T + \mathsf{O}_T} - \Phi_T \end{bmatrix}, \qquad \theta \in \mathcal{D};$$

by convention, $0/0 = 0$.
Let $\gamma > 0$. The proximal operator of the function $\theta \mapsto \gamma \lambda_{\mathsf{R}} \|\mathsf{A}^{(\ell)}\theta\|_1$ is explicit: for any $\tau = (\tau_1, \cdots, \tau_{2T}) \in \mathbb{R}^{2T}$,

$$\mathrm{Prox}_{\gamma \lambda_{\mathsf{R}} \|\mathsf{A}^{(\ell)} \cdot\|_1}(\tau) = \begin{bmatrix} \left(\mathrm{Id}_T - \left(\mathsf{D}_2^{(\ell)}\right)^\top \mathsf{D}_2^{(\ell)}\right) \tau_{1:T} + \left(\mathsf{D}_2^{(\ell)}\right)^\top \mathrm{Prox}_{\gamma \lambda_{\mathsf{R}} \|\cdot\|_1}(\mathsf{D}_2^{(\ell)} \tau_{1:T}) \\ \mathrm{sign}(\tau_{T+1}) \left(|\tau_{T+1}| - \gamma \lambda_{\mathsf{O}} \Phi_1\right)_+ \\ \cdots \\ \mathrm{sign}(\tau_{2T}) \left(|\tau_{2T}| - \gamma \lambda_{\mathsf{O}} \Phi_T\right)_+ \end{bmatrix}.$$

4.2 Proposal mechanism

Given the current point θ, an index ℓ sampled at random in $\{1, 2, 3\}$ and a $(2T) \times (2T)$ invertible matrix C, the candidate is $\mu_\ell(\theta) + \sqrt{2\gamma}\,\mathsf{C}\,\mathcal{N}_{2T}(0, \mathrm{Id})$ where

$$\mu_\ell(\theta) \stackrel{\text{def}}{=} \mathrm{Prox}_{\gamma \lambda_{\mathsf{R}} \|\mathsf{A}^{(\ell)} \cdot\|_1} \left(\theta - \gamma \nabla f(\theta)\right).$$

4.3 Pseudo-code

Algorithm 2: `PGdec`

Data: λ_{R}, λ_{O}, T, $NbrMC$, $\theta^0 \in \mathcal{D}$, γ, C

Result: A $\mathcal{D}$-valued sequence $\{\theta^n, n \in \{0, \ldots, NbrMC\}\}$

Define D_2 and $\mathsf{D}_2^{(\ell)}$ for $\ell = 1, 2, 3$;

Sample $(2T) \times NbrMC$ centered Gaussian random variables with covariance matrix $\mathsf{CC}^\top$, stored in *GaussRnd*; $1 \times$ `NbrMC` uniform random variables on $[0, 1]$, stored in *UnifRnd*; and $1 \times NbrMC$ uniform random variables on $\{1, 2, 3\}$, stored in *BlockRnd* ;

for $n = 0, \ldots,$ `NbrMC` $- 1$ **do**

- Fix the index: $\ell = BlockRnd(n)$;
- Propose a candidate `Proposal`, equal to $\mu_\ell(\theta^n) + \sqrt{2\gamma}\, GaussRnd(:, n)$;
- **if** *Proposal* $\notin \mathcal{D}$ **then**
  - $\theta^{n+1} = \theta^n$
- **else**
  - Compute the log acceptance-rejection ratio *logalpha*: **begin**
    - compute $\log \pi(Proposal)$;
    - compute *logGaussNum*, the density of a Gaussian distribution with expectation $\mu(Proposal)$, covariance $2\gamma\mathsf{CC}^\top$, at the point θ^n ;
    - compute *logGaussDenom* the density of a Gaussian distribution with expectation $\mu(\theta^n)$, covariance $2\gamma\mathsf{CC}^\top$, at the point *Proposal* ;
    - set $logalpha = \log \pi(Proposal) - logpiCurrent + logGaussNum - logGaussDenom$
  - **if** *UnifRnd(n)* $< \exp(logalpha)$ **then**
    - set $\theta^{n+1} = Proposal$;
    - set $logpiCurrent = \log \pi(Proposal)$
  - **else**
    - $\theta^{n+1} = \theta^n$;
- Store θ^{n+1} ;
- Store *logpiCurrent* ;
- Update γ: **begin**
  - If mod(nn, `param.frequency`) == 1 and while $n <$ `MCMC.chain_burnin`: increase or decrease γ in order to target a mean acceptance rate `MCMC.target_ratio`.