

Lab session – Nonsmooth convex optimization

1 Context, model and goals

From the very beginning of the Covid19, National Health Authorities are monitoring the number of new infections each day, denoted by Z_t at day t . An example of such daily counts for the past five weeks in France is provided in Figure 1.

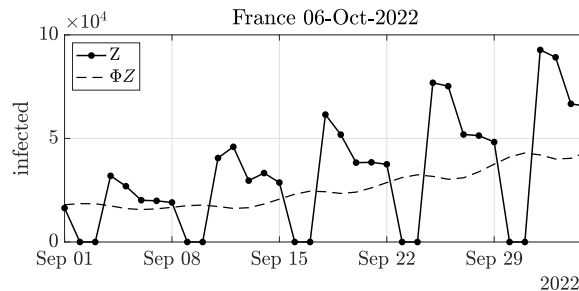


FIGURE 1 – Daily new infection counts Z in France for the past 5 weeks.

a. Load the MATLAB file `data_covid.mat` and plot the daily new infections Z with respect to days.

Yet, the number of new infections itself is not informative enough about the dynamics of the pandemic. Thus, looking only at Z_t does not make it possible to detect an epidemic resurgence early enough, and therefore to react quickly enough to avoid a dramatic explosion in the number of infected people.

This is why we rapidly turn to the monitoring of the *reproduction number* R_0 , defined as the average number of secondary cases generated by a typical infected individual throughout its period of contagiousness. The reproduction number thus quantifies the intensity of the pandemic :

- when $R_0 > 1$ the virus is spreading at exponential speed ;
- when $R_0 < 1$ the epidemic is vanishing.

The standard definition of R_0 is relaxed into a time-dependent reproduction number R_t at day t , which is linked to the number of new infections by Cori's model [2]. In this epidemiological model the number of new infections at day t depends on the past daily counts Z_1, \dots, Z_{t-1} and follow a Poisson distribution

$$Z_t | Z_1, \dots, Z_{t-1} \sim \text{Poisson}(R_t(\Phi Z)_t), \quad \text{with } (\Phi Z)_t = \sum_{u=1}^{\tau_\Phi} \Phi(u) Z_{t-u} \quad (1)$$

where Φ is the *serial interval function* associated to the pandemic, modeling the random delay between primary and secondary cases¹.

Reminder 1. The Poisson distribution of parameter $p \in \mathbb{R}_+$ is a probability distribution on nonnegative integers. A random variable Z follows a Poisson distribution if and only if

$$\forall k \in \mathbb{N}, \quad \mathbb{P}(Z = k | p) = e^{-p} \frac{p^k}{k!}.$$

Thus, given an observation $Z \in \mathbb{N}$ assumed to follow a Poisson model of parameter p , the log-likelihood $\ln \mathcal{L}(Z | p)$ writes

$$\ln \left(e^{-p} \frac{p^Z}{Z!} \right) = -p + Z \ln(p) - \ln(Z!) \underset{Z \gg 1}{\simeq} -p + Z \ln(p) - Z \ln(Z) + Z = - \left(Z \ln \left(\frac{Z}{p} \right) + p - Z \right),$$

where the approximation $\ln(Z!) \simeq Z \ln(Z) - Z$, valid for $Z \rightarrow \infty$, can be derived from Stirling formula.

1. For Covid19 pandemic the serial interval function is modeled by a Gamma distribution with mean 6.6 days and standard deviation 3.5 days cropped at $\tau_\Phi = 26$ days.

Definition 1. The Kullback-Leibler between a vector of observations $\mathbf{Z} \in \mathbb{N}^T$ and a vector of parameters $\mathbf{p} \in \mathbb{R}_+^T$ is defined in a fully separable manner as

$$D_{\text{KL}}(\mathbf{Z}|\mathbf{p}) = \sum_{t=1}^T d_{\text{KL}}(Z_t|p_t), \text{ where } d_{\text{KL}}(Z_t|p_t) = \begin{cases} Z_t \ln\left(\frac{Z_t}{p_t}\right) + p_t - Z_t & \text{if } Z_t > 0, p_t > 0 \\ p_t & \text{if } Z_t = 0, p_t \geq 0 \\ \infty & \text{otherwise.} \end{cases}$$

Thus, the opposite log-likelihood associated to Cori's model (1) expresses in terms of the Kullback-Leibler divergence as

$$-\ln \mathcal{L}(\mathbf{Z}|\mathbf{R}) = D_{\text{KL}}(\mathbf{Z}|\mathbf{R} \odot \Phi \mathbf{Z}) \quad (2)$$

where $\mathbf{R} \odot \Phi \mathbf{Z}$ denotes the component-wise vector product, $(\mathbf{R} \odot \Phi \mathbf{Z})_t = R_t(\Phi \mathbf{Z})_t$.

- b. For Z_t fixed at $Z_t \in \{0, 10, 250\}$ plot the function $p_t \mapsto d_{\text{KL}}(Z_t|p_t)$.
- c. Show that the function $p_t \mapsto d_{\text{KL}}(Z_t|p_t)$ is convex and differentiable.
- d. Is the gradient of $p_t \mapsto d_{\text{KL}}(Z_t|p_t)$ Lipschitzian? Justify your answer.

2 Estimation of the reproduction number

2.1 Maximum Likelihood Estimator

- a. Given fixed Z_t and $(\Phi \mathbf{Z})_t = \sum_{u=1}^{T_\Phi} \Phi(u) Z_{t-u}$, compute the minimum and the minimizer of the function $d_{\text{KL}} : R_t \mapsto d_{\text{KL}}(Z_t|R_t(\Phi \mathbf{Z})_t)$.
- b. Deduce the minimum and the minimizer $\hat{\mathbf{R}}^{\text{MLE}}$ of the Kullback-Leibler divergence

$$D_{\text{KL}} : \begin{cases} \mathbb{R}^T & \rightarrow \mathbb{R} \\ \mathbf{R} & \mapsto \sum_{t=1}^T d_{\text{KL}}(Z_t|R_t(\Phi \mathbf{Z})_t). \end{cases}$$

This minimizer is denoted $\hat{\mathbf{R}}^{\text{MLE}}$ and it is the *Maximum Likelihood Estimator* of \mathbf{R} .

Indeed, as stated in Equation (2), the Kullback-Leibler divergence is the opposite log-likelihood of the Poisson model, thus minimizing $D_{\text{KL}}(\mathbf{Z}|\mathbf{R} \odot \Phi \mathbf{Z})$ amounts to maximize the likelihood $\mathcal{L}(\mathbf{Z}|\mathbf{R})$.

- c. Compute numerically the Maximum Likelihood Estimate from the data \mathbf{Z} and $\Phi \mathbf{Z}$ provided in `data_covid.mat`. Plot it and comment about its temporal behavior.
- d. Would you advocate the use of the MLE to monitor the Covid19 pandemic? Explain why.

2.2 Penalized likelihood

In order to enforce some regularity on the temporal behavior of R_t we consider the penalized likelihood estimator

$$\hat{\mathbf{R}} \in \underset{\mathbf{R}}{\text{Argmin}} D_{\text{KL}}(\mathbf{Z}|\mathbf{R} \odot \Phi \mathbf{Z}) + \lambda \|\mathbf{D}_2 \mathbf{R}\|_1, \quad (3)$$

where $\mathbf{D}_2 : \mathbb{R}^T \rightarrow \mathbb{R}^{T-1}$ is the discrete Laplacian operator acting on \mathbb{R}^T as

$$\forall t \in \{1, 2, \dots, T-2\}, \quad (\mathbf{D}_2 \mathbf{R})_t = R_{t+2} - 2R_{t+1} + R_t.$$

The ℓ_1 penalization favors sparsity of the second order derivative of the estimate and thus $\hat{\mathbf{R}}$ is expected to be piecewise linear, with only a few days at which the slope of $t \mapsto R_t$ is changing.

Because of the presence of the ℓ_1 norm in the objective function, (3) is a nonsmooth optimization problem. Thus one has to resort to proximal operators to solve it.

- a. Based on Question 1 c. explain why it is not possible to use the *forward-backward* algorithm.

To circumvent this limitation, we will use only proximity operators.

- b. For fixed Z_t , compute the proximity operator of $p_t \mapsto d_{\text{KL}}(Z_t|p_t)$.
- c. Given Z_t and $(\Phi \mathbf{Z})_t = \sum_{u=1}^{T_\Phi} \Phi(u) Z_{t-u}$, deduce from 1. the expression of the proximity operator of $R_t \mapsto d_{\text{KL}}(Z_t|R_t(\Phi \mathbf{Z})_t)$.

- d. Explain how to compute the proximity operator of

$$D_{\text{KL}} : \begin{cases} \mathbb{R}^T & \rightarrow \mathbb{R} \\ \mathbf{R} & \mapsto \sum_{t=1}^T d_{\text{KL}}(Z_t | R_t(\Phi Z)_t). \end{cases}$$

Because of the linear operator \mathbf{D}_2 inside the ℓ_1 norm in (3), we do not have a closed-form expression of the proximity operator of the penalization $\|\mathbf{D}_2 \mathbf{R}\|_1$ and hence it is neither possible to use the Douglas-Rachford algorithm. We will thus turn to the primal-dual algorithm proposed in [1].

Algorithm 1 Primal-dual minimization of the penalized Kullback-Leibler (3) for the estimation of reproduction numbers.

Require: Infection counts : $\mathbf{Z} \in \mathbb{R}^T$ and $\Phi \mathbf{Z} \in \mathbb{R}^T$

Choose descent parameters : $\tau, \sigma > 0$

Max. iterations : k_{\max}

Initialization $\mathbf{R}^{[0]} = \mathbf{Z} / \Phi \mathbf{Z}$

$\mathbf{Q}^{[0]} = \mathbf{D}_2 \mathbf{R}^{[0]}, \bar{\mathbf{R}}^{[0]} = \mathbf{R}^{[0]}$

while $k < k_{\max}$ **do**

$$\mathbf{Q}^{[k+1]} = \mathbf{Q}^{[k]} + \sigma \mathbf{D}_2 \bar{\mathbf{R}}^{[k]} - \sigma \text{prox}_{\lambda \sigma^{-1} \|\cdot\|_1} \left(\sigma^{-1} \mathbf{Q}^{[k]} + \mathbf{D}_2 \bar{\mathbf{R}}^{[k]} \right)$$

$$\mathbf{R}^{[k+1]} = \text{prox}_{\tau D_{\text{KL}}(\mathbf{Z} | \cdot \odot \Phi \mathbf{Z})} (\mathbf{R}^{[k+1]} - \tau \mathbf{D}_2^* \mathbf{Q}^{[k+1]})$$

$$\bar{\mathbf{R}}^{[k+1]} = 2\mathbf{R}^{[k+1]} - \mathbf{R}^{[k]}$$

$k \leftarrow k + 1$

end while

- d. Compute the proximity operator of the ℓ_1 norm multiplied by a scalar $\lambda \sigma^{-1}$, that is of $\lambda \sigma^{-1} \|\cdot\|_1 : \mathbf{Q} \in \mathbb{R}^N \mapsto \sum_{n=1}^N \lambda \sigma^{-1} |Q_n| \in \mathbb{R}$ and explain how to compute it in practice.
- f. Construct the $T - 2 \times T$ matrix of the discrete Laplacian \mathbf{D}_2 .
- f. Implement Algorithm 1 and run it on the data \mathbf{Z} . Plot the evolution of the objective function $D_{\text{KL}}(\mathbf{Z} | \mathbf{R}^{[k]} \odot \Phi \mathbf{Z}) + \lambda \|\mathbf{D}_2 \mathbf{R}^{[k]}\|_1$ along iterations to illustrate convergence of the scheme.
- g. Run Algorithm 1 for different values of the regularization parameter λ and comment on the influence of λ . A possibility is to consider $\lambda \in \{0.5 \times \text{std}(\mathbf{Z}), 3.5 \times \text{std}(\mathbf{Z}), 50 \times \text{std}(\mathbf{Z})\}$.

Références

- [1] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1) :120–145, 2011.
- [2] A. Cori, N. M. Ferguson, C. Fraser, and S. Cauchemez. A new framework and software to estimate time-varying reproduction numbers during epidemics. *American Journal of Epidemiology*, 178(9) :1505–1512, 2013.