

Lab session – Estimation of the reproduction number of COVID-19

1 Preliminaries: subdifferentiable and proximal operators

Exercise 1. Let $f, g : \mathbb{R}^T \rightarrow \mathbb{R} \cup \{\infty\}$, two proper, continuous functions. Show that, if $\text{dom } f \cap \text{dom } g \neq \emptyset$, then

$$\forall \mathbf{x} \in \mathbb{R}^T, \quad \partial f(\mathbf{x}) + \partial g(\mathbf{x}) \subset \partial(f + g)(\mathbf{x}).$$

In particular, show that if f is continuously differentiable and convex on its domain, then

$$\forall \mathbf{x} \in \mathbb{R}^T, \quad \nabla f(\mathbf{x}) + \partial g(\mathbf{x}) \subset \partial(f + g)(\mathbf{x}). \quad (1)$$

Remark 1. This property does not require functions f and g to be convex. Its reciprocal does (see [1, Proposition 16.42] for the exact hypotheses required and a demonstration using convex duality, out of the scope of the present course!). Yet, the above inclusion is enough to interpret mixed proximal-gradient scheme: since, due to Equation (1), $\nabla f(\mathbf{x}^{[k]}) + \mathbf{u}^{[k]}, \mathbf{u}^{[k]} \in \partial g(\mathbf{x}^{[k]})$ is indeed one subgradient of $f + g$ at $\mathbf{x}^{[k]}$.

Exercise 2. Given some $\mathbf{y} \in \mathbb{R}^T$, let

$$f : \begin{cases} \mathbb{R}^T & \rightarrow \mathbb{R} \\ \mathbf{x} & \mapsto \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|_2^2. \end{cases}$$

compute the proximity operator of γf , for $\gamma > 0$. Plot it when $T = 1$ and comment briefly its behavior.

Exercise 3. Let $f : \mathbb{R}^T \rightarrow \mathbb{R}, \mathbf{x} \mapsto \|\mathbf{x}\|_1$, compute the proximity operator of γf , for $\gamma > 0$. Comment briefly its behavior.

2 Context, model and goals

From the very beginning of the COVID-19 pandemic National Health Authorities of all countries worldwide are monitoring the number of new infections each day, denoted by Z_t at day t . An example of such daily counts for five weeks in October and November 2022 in France is provided in Figure 1, as collected and made available by Johns Hopkins University¹.

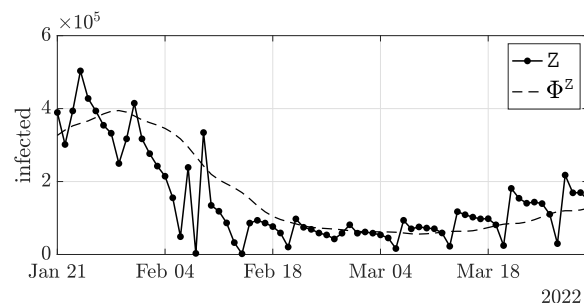


Figure 1: Daily new infection counts Z in France for 10 weeks.

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

En MATLAB

```
load('data_covid.mat')
plot(days, Z)
```

¹<https://coronavirus.jhu.edu/map.html>

```

En PYTHON
from scipy.io import loadmat
import matplotlib.pyplot as plt
data = loadmat('data_covid.mat', squeeze_me = True)
plt.plot(Z)

```

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 *standard reproduction number*, R_0 , defined as the average number of secondary cases generated by a typical infected individual throughout its period of contagiousness [3, 4]. 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;
- when $R_0 = 1$ the epidemic is stable.

The standard definition of the reproduction number 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 [3]. 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

$$Z_t | Z_1, \dots, Z_{t-1} \sim \text{Poisson}(R_t \Phi_t^Z), \quad \text{with } \Phi_t^Z = \sum_{u=1}^{\tau_\Phi} \phi(u) Z_{t-u} \quad (2)$$

where ϕ is the *serial interval function* associated to the pandemic, modeling the random delay between primary and secondary cases².

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

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

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

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

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

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

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

If $\mathbf{Z} \in \mathbb{N}^T$ is a vector of nonnegative integer observations, it quantifies the discrepancy between \mathbf{Z} and a vector of Poisson random variables of parameters \mathbf{p} .

Thus, the opposite log-likelihood associated to Cori's model (2) expresses in terms of the Kullback-Leibler divergence as

$$-\ln \mathcal{L}(\mathbf{Z} | \mathbf{R}) = d_{\text{KL}}(\mathbf{Z} | \mathbf{R} \odot \Phi^Z) \quad (3)$$

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

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

²For COVID-19 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.

3 Estimation of the reproduction number

3.1 Maximum Likelihood Estimator

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

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

This minimizer is denoted \hat{R}^{MLE} because it is the Maximum Likelihood Estimator of \mathbf{R} .

Indeed, as stated in Equation (3), the Kullback-Leibler divergence is the opposite log-likelihood of the Poisson model, thus minimizing $d_{\text{KL}}(\mathbf{Z} | \mathbf{R} \odot \Phi^Z)$ amounts to maximize the likelihood $\mathcal{L}(\mathbf{Z} | \mathbf{R})$.

- c. Compute numerically the Maximum Likelihood Estimate from the data \mathbf{Z} , named \mathbf{Z} in the file `data_covid.mat`, and the global infectiousness Φ^Z , named `PhiZ` in `data_covid.mat`. Plot it and comment about its temporal behavior.
- d. Would you advocate the use of the Maximum Likelihood Estimator to monitor the COVID-19 pandemic? Explain why.

3.2 Penalized likelihood

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

$$\hat{\mathbf{R}} \in \underset{\mathbf{R}}{\text{Argmin}} D_{\text{KL}}(\mathbf{Z} | \mathbf{R} \odot \Phi^Z) + \lambda \|\mathbf{D}_2 \mathbf{R}\|_1, \quad (4)$$

where $\mathbf{D}_2 : \mathbb{R}^T \rightarrow \mathbb{R}^{T-2}$ 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, (4) is a nonsmooth optimization problem. Thus one has to resort to proximal operators to solve it.

- a. Based on Question 2 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 $\Phi_t^Z = \sum_{u=1}^{\tau_{\Phi^Z}} \Phi^Z(u) Z_{t-u}$, deduce from 2 b. the expression of the proximity operator of $R_t \mapsto d_{\text{KL}}(Z_t | R_t \Phi_t^Z)$.
- d. Explain how to compute the proximity operator of

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

Because of the linear operator \mathbf{D}_2 inside the ℓ_1 norm in (4), 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 necessary to use a *splitting* scheme. We will thus turn to the primal-dual algorithm proposed in [2]. To ensure convergence of **Algorithm 1**, the descent steps $\tau > 0$ and $\sigma > 0$ must be chosen so that $\sigma\tau\|\mathbf{D}_2\|_{\text{op}}^2 < 1$, where $\|\mathbf{D}_2\|_{\text{op}}^2$ is the *operator norm* of \mathbf{D}_2 defined as

$$\|\mathbf{D}_2\|_{\text{op}} = \sup_{\mathbf{R} \in \mathbb{R}^T, \mathbf{R} \neq \mathbf{0}} \frac{\|\mathbf{D}_2 \mathbf{R}\|}{\|\mathbf{R}\|}$$

To compute the operator norm of a matrix $\mathbf{D}_2 \in \mathbb{R}^{T-2 \times T}$, you can use the MATLAB function `norm(D2)` or the PYTHON function `numpy.linalg.norm(D2, ord = 2)`. A standard choice for the descent steps is $\sigma = \tau = 0.99/\|\mathbf{D}_2\|_{\text{op}}$. Feel free to explore other choices satisfying the convergence condition.

Algorithm 1 Primal-dual minimization of the penalized Kullback-Leibler (4) 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: $\tau, \sigma > 0$ such that $\sigma\tau\|\mathbf{D}_2\|_{\text{op}}^2 < 1$

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

$$\begin{aligned}\mathbf{Q}^{[k+1]} &= \mathbf{Q}^{[k]} + \sigma\mathbf{D}_2\bar{\mathbf{R}}^{[k]} - \sigma\text{prox}_{\lambda\sigma^{-1}\|\cdot\|_1}(\sigma^{-1}\mathbf{Q}^{[k]} + \mathbf{D}_2\bar{\mathbf{R}}^{[k]}) \\ \mathbf{R}^{[k+1]} &= \text{prox}_{\tau\mathbf{D}_{\text{KL}}(\mathbf{Z}|\cdot\odot\Phi^{\mathbf{Z}})}(\mathbf{R}^{[k]} - \tau\mathbf{D}_2^*\mathbf{Q}^{[k+1]}) \\ \bar{\mathbf{R}}^{[k+1]} &= 2\mathbf{R}^{[k+1]} - \mathbf{R}^{[k]}\end{aligned}$$

$k \leftarrow k + 1$

end while

- e. Compute the proximity operator of the ℓ_1 norm multiplied by a scalar $\lambda\sigma^{-1}$, that is of $\lambda\sigma^{-1}\|\cdot\|_1 : \mathbf{Q} \in \mathbb{R}^N \mapsto \sum_{n=1}^N \lambda\sigma^{-1}|\mathbf{Q}_n| \in \mathbb{R}$ and explain how to compute it in practice.
- f. Construct the $T - 2 \times T$ matrix of the discrete Laplacian \mathbf{D}_2 .
- g. Show that by setting $\tilde{\mathbf{Z}} := \mathbf{Z}/\alpha$ for some $\alpha > 0$,

$$\mathbf{D}_{\text{KL}}(\tilde{\mathbf{Z}}|\mathbf{R}^{[k]} \odot \Phi^{\tilde{\mathbf{Z}}}) = \frac{1}{\alpha}\mathbf{D}_{\text{KL}}(\mathbf{Z}|\mathbf{R}^{[k]} \odot \Phi^{\mathbf{Z}}). \quad (5)$$

Justify that running **Algorithm 1** with \mathbf{Z} and $\Phi^{\mathbf{Z}}$ as input, with a given λ vs. with $\tilde{\mathbf{Z}}$ and $\Phi^{\tilde{\mathbf{Z}}}$ with $\tilde{\lambda} := \lambda/\text{std}(\mathbf{Z})$ give the same reproduction number estimate $\hat{\mathbf{R}}$.

From now on, when running **Algorithm 1**, replace \mathbf{Z} by $\mathbf{Z}/\text{std}(\mathbf{Z})$ and $\Phi^{\mathbf{Z}}$ by $\Phi^{\mathbf{Z}}/\text{std}(\mathbf{Z})$. Take into account that this amounts to replace λ by $\tilde{\lambda} := \lambda/\text{std}(\mathbf{Z})$. This normalization of input data improve the numerical robustness of the algorithmic scheme: it is to be seen as a purely numerical trick.

- h. Implement **Algorithm 1** and run it on the data \mathbf{Z} . Plot the estimated reproduction number \mathbf{R} and the evolution of the objective function

$$k \mapsto \mathbf{D}_{\text{KL}}(\mathbf{Z}|\mathbf{R}^{[k]} \odot \Phi^{\mathbf{Z}}) + \lambda\|\mathbf{D}_2\mathbf{R}^{[k]}\|_1 \quad (6)$$

along iterations to illustrate convergence of the scheme.

- i. Run **Algorithm 1** for different values of the regularization parameter $\tilde{\lambda}$ and comment on its influence on the estimated reproduction number \mathbf{R} . A possibility is to consider $\tilde{\lambda} \in \{0.5, 3.5, 15, 50, 150, 250\}$ ³.

3.3 Tikhonov penalization

- a. From **Exercise 2**, deduce the proximity operator of $\mathbf{x} \mapsto \|\mathbf{x}\|_2^2$.
- b. Replace the ℓ_1 penalization, $\|\mathbf{D}_2\mathbf{R}\|_1$, by an squared ℓ_2 penalization $\|\mathbf{D}_2\mathbf{R}\|_2^2$, also called Tikhonov regularization. Implement the Chambolle-Pock algorithm replacing $\|\cdot\|_1$ by $\|\cdot\|_2^2$.
Indication: the convergence condition does not change.
- c. Plot the two estimates on the same graph and comment.
Indication: the *optimal* parameter λ , i.e., the one for which the estimate is regular while still reflecting the pandemic dynamics, might change depending on whether the ℓ_1 or ℓ_2^2 penalization is used.

³This correspond to non normalized regularization parameters $\lambda \in \{0.5 \times \text{std}(\mathbf{Z}), 3.5 \times \text{std}(\mathbf{Z}), 15 \times \text{std}(\mathbf{Z}), 50 \times \text{std}(\mathbf{Z}), 150 \times \text{std}(\mathbf{Z}), 250 \times \text{std}(\mathbf{Z})\}$.

References

- [1] H. H. Bauschke and P. L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*, volume 408. Springer, 2011.
- [2] 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.
- [3] 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.
- [4] Q.-H. Liu, M. Ajelli, A. Aleta, S. Merler, Y. Moreno, and A. Vespignani. Measurability of the epidemic reproduction number in data-driven contact networks. *Proceedings of the National Academy of Sciences*, 115(50):12680–12685, 2018.