Unbalanced Optimal Transport-based regularization
Application to inverse problems in epidemiology

Master internship in optimal transport for inverse problems applied to epidemiology

Supervision and contact: Barbara Pascal, barbara.pascal@cnrs.fr, CNRS junior researcher,
Jérôme Idier, jerome.idier@ls2n.fr, CNRS senior research.

Application: Send a CV, master grades, references and motivations to B. Pascal and J. Idier.

Location: Laboratoire des Sciences du Numérique de Nantes (LS2N), École Centrale Nantes.

Duration and dates: 4 to 6 months in 2025.

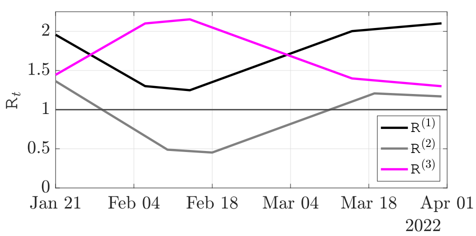
Context: The basic reproduction number of an epidemic, R_0, is defined as the average number of secondary infections caused by one standard contagious individual. Relaxed into a daily indicator, R_t at day t, called the *effective reproduction number*, it provides one of the most widely used tools to monitor the intensity of virus propagation in a population: when $\mathsf{R}_t > 1$ the number of cases is growing exponentially, while it is decreasing exponentially when $\mathsf{R}_t < 1$. In practice, health authorities collect daily new infection counts $\mathsf{Z}_t^{(d)}$, at days $t = 1, \ldots, T$ and for a collection of D territories (e.g., the 96 metropolitan French departments), and the $\mathsf{R}_t^{(d)}$s need to be extracted from these, possibly low quality, data. Leveraging the state-of-the-art epidemiological model proposed in [1], [2] performed the estimation of the reproduction number $\mathsf{R}_t^{(d)}$ at day t in territory d through the minimization of a penalized negative log-likelihood:

$$\widehat{\mathbf{R}} = \underset{\mathbf{R} \in \mathbb{R}_+^{D \times T}}{\operatorname{argmin}} \; - \log \mathcal{L}(\mathbf{Z}, \mathbf{R}) + \mu \mathcal{T}(\mathbf{R}) + \omega \mathcal{G}(\mathbf{R}) \tag{1}$$

where $\mathcal{L}(\mathbf{Z}, \mathbf{R})$ is the likelihood of $\mathbf{R}$ given reported infection counts $\mathbf{Z}$, encapsulating the epidemiological model, $\mathcal{T}(\mathbf{R})$ is a term promoting temporal regularity of $t \mapsto \mathsf{R}_t^{(d)}$ independently for each county d, $\mathcal{G}(\mathbf{R})$ favoring spatial consistency across reproduction numbers in *connected* counties (e.g., counties sharing terrestrial borders) and $\mu, \omega > 0$ are regularization parameters, balancing the fidelity to the model and the regularity constraints.

Challenge: This internship project focuses on the design of smart spartial regularizers $\mathcal{G}$. The most straightforward choice writes [2]

$$\mathcal{G}(\mathbf{R}) = \sum_{d \sim d'} \left\| \mathbf{R}^{(d)} - \mathbf{R}^{(d')} \right\|_1 = \sum_{d \sim d'} \sum_{t=1}^{T} \left| \mathsf{R}_t^{(d)} - \mathsf{R}_t^{(d')} \right| \tag{2}$$



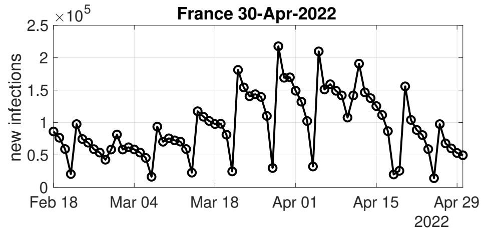
where the sum runs over all *connected* counties. One major disadvantage of this kind of penalization is that, since it is separable in t, it is highly sensitive to small temporal shifts between $\mathbf{R}^{(d)}$ and $\mathbf{R}^{(d')}$. As an illustration, the figure above represents the reproduction number $\mathbf{R}^{(d)}$ in three different counties, $d \in \{1, 2, 3\}$. $\mathbf{R}^{(1)}$ and $\mathbf{R}^{(2)}$ share a similar temporal pattern, simply slightly shifted in time and amplitude, while the behaviors of $\mathbf{R}^{(1)}$ and $\mathbf{R}^{(3)}$ are very different. Though, $\|\mathbf{R}^{(1)} - \mathbf{R}^{(2)}\|_1 \approx 56$ and $\|\mathbf{R}^{(1)} - \mathbf{R}^{(3)}\|_1 \approx 36$ demonstrating the poor ability of the ℓ_1 based $\mathcal{G}$ penalization (2) to handle global temporal patterns.

Objectives: The main objective of this internship is to tackle this limitation by recoursing to recent procedures leveraging *unbalanced optimal transport* to design regularizations better suited to the comparison of non-local patterns [3]. These methods amount to replacing the ℓ_1 norm in (2) by the *generalized Wasserstein distance*, defined as

$$W_1\left(\mathbf{R}^{(d)}, \mathbf{R}^{(d')}\right) = \min_{\mathbf{\Gamma} \geq 0} \langle \mathbf{\Gamma}, \boldsymbol{C} \rangle + \lambda \left(\left\| \mathbf{\Gamma} \mathbb{1}_T - \mathbf{R}^{(d)} \right\|_2^2 + \left\| \mathbf{\Gamma}^\top \mathbb{1}_T - \mathbf{R}^{(d')} \right\|_2^2 \right), \tag{3}$$

where $\mathbf{\Gamma}, \boldsymbol{C} \in \mathbb{R}^{T\times T}$, having nonnegative entries, are respectively a transport plan and an euclidean cost matrix; $^\top$ is the matrix transposition; $\mathbb{1}_T$ is a column vector with all entries equal to one; $\lambda > 0$ is a regularization parameter enabling to handle unbalanced transport.

Research program:
i) Plug the Wassertein distance W_1 into the variational formulation (1) in replacement of the standard ℓ_1 norm- penalization $\mathcal{G}$ and derive the associated optimization problem.
ii) Study the minimization problem and apply the formalism developed in [3] to design a fast algorithm to solve the unbalanced optimal transport regularized inverse problem (1).
iii) Investigate the influence of the relaxation parameter λ involved in the definition of unbalanced optimal transport.
iv) Compare the reproduction number estimation performance when using ℓ_1-norm based vs. unbalanced optimal transport penalizations, on synthetic and then real data made available by the Johns Hopkins University https://coronavirus.jhu.edu/ a example of which is provided on the right.



Prerequisite: The recruited intern is expected to be at ease with the basic concepts of statistics and optimization, as well as with Python programming. Mathematical background in convex nonsmooth optimization and/or about optimal transport would be appreciated.

References

[1] 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.

[2] P. Abry, N. Pustelnik, S. Roux, P. Jensen, P. Flandrin, R. Gribonval, C.-G. Lucas, É. Guichard, P. Borgnat, and N. Garnier. Spatial and temporal regularization to estimate COVID-19 reproduction number R (t): Promoting piecewise smoothness via convex optimization. *PLOS One*, 15(8):e0237901, 2020.

[3] J. Lee, N. P. Bertrand, and C. J. Rozell. Unbalanced optimal transport regularization for imaging problems. *IEEE Transactions on Computational Imaging*, 6:1219–1232, 2020.