

Optimisation convexe non-lisse pour l'estimation régularisée du R_0 de l'épidémie de Covid-19

Barbara PASCAL, CRIStal - Lille **Patrice ABRY**, LP ENSL - Lyon
Nelly PUSTELNIK, LP ENSL - Lyon **Stéphane ROUX**, LP ENSL - Lyon
Rémi GRIBONVAL, LIP - Lyon **Patrick FLANDRIN**, LP ENSL - Lyon

Le **coefficient de reproduction** R_0 , quantifiant le nombre moyen de personnes contaminées par un individu infecté, mesure l'intensité de la pandémie de Covid-19. Son estimation, à partir des nombres de nouveaux cas $Z_t^{(d)}$, au jour t dans le département d , est donc un enjeu majeur. La principale difficulté est de gérer des données $Z_t^{(d)}$, cf. Fig. 1, qui s'avèrent non seulement très bruitées, mais présentent en outre des **variations ponctuelles aberrantes**.

Nous proposons, en nous appuyant sur la littérature en épidémiologie [Cori, 2013], de modéliser la propagation stochastique du virus par un processus **poissonnien discret** dont le paramètre **instantané** dépend des données passées et d'une variable de perturbation ponctuelle dite d'*outliers*, de sorte que $Z_t^{(d)} \sim \mathcal{P}(P_t^{(d)})$, où $P_t^{(d)} = R_t^{(d)}(\Phi Z)_t^{(d)} + O_t^{(d)}$, avec $(\Phi Z)_t^{(d)} = \sum_{s \geq 1} \phi(s)Z_{t-s}^{(d)}$, ϕ **fonction d'intervalle de série**, étant connue, $R_t^{(d)}$ et $O_t^{(d)}$ désignant respectivement le coefficient de reproduction et la variable d'*outlier* au jour t dans le département d . Dans le travail présenté, l'éviction des données aberrantes et l'estimation du coefficient de reproduction sont alors réalisées **simultanément** par la résolution du problème d'optimisation convexe non-lisse :

$$\begin{aligned}
 \hat{R}, \hat{O} = \operatorname{argmin}_{R, O \in \mathbb{R}^{D \times T}} & \sum_{d=1}^D \sum_{t=1}^T \left[P_t^{(d)} - Z_t^{(d)} \ln \left(P_t^{(d)} \right) + \iota_{\geq 0}(R_t^{(d)}) + \iota_{> 0}(P_t^{(d)}) \right] \\
 & + \lambda_{\text{time}} \sum_{d=1}^D \sum_{t=2}^{T-2} \left| (\mathbf{D}_2 R)_t^{(d)} \right| + \lambda_O \sum_{d=1}^D \sum_{t=1}^T \left| O_t^{(d)} \right| + \lambda_{\text{space}} \sum_{d \sim d'} \sum_{t=1}^T \left| (\mathbf{G} R)_t^{(d, d')} \right|,
 \end{aligned} \tag{1}$$

avec $(\mathbf{D}_2 R)_t^{(d)} = R_{t+1}^{(d)} + R_{t-1}^{(d)} - 2R_t^{(d)}$ le laplacien discret et $(\mathbf{G} R)_t^{(d, d')} = R_t^{(d)} - R_t^{(d')}$ un opérateur de gradient sur graphe. Le terme d'attache aux données, s'appuyant sur la divergence de Kullback-Leibler, encode le modèle de Poisson et assure la positivité du coefficient de reproduction et du paramètre de Poisson instantané estimés *via* les indicatrices. Les termes de pénalisation utilisent la norme ℓ_1 pour favoriser des propriétés de régularité et de parcimonie, privilégiant des variations temporelles **linéaire par morceaux** du coefficient de reproduction, afin de **détecter précocement la tendance** d'évolution de la pandémie, et des variations spatiales constantes par morceaux, traduisant la similarité de la dynamique épidémique entre territoires voisins, et l'occurrence **ponctuelle** de valeurs aberrantes pour les *outliers*. En Fig. 2, des exemples d'estimées $\hat{R}_t^{(d)}$ obtenues par une procédure sans *outlier* [Abry, 20'] (cyan et magenta) ou en minimisant (1) (bleu et rouge) démontrent la meilleure gestion des valeurs aberrantes par la procédure développée. Du point de vue théorique, nous démontrons l'existence d'une solution au problème (1) et l'unicité du paramètre de Poisson instantané $\hat{P}_t^{(d)} = \hat{R}_t^{(d)}(\Phi Z)_t^{(d)} + \hat{O}_t^{(d)}$ grâce à des arguments de **coercivité et de convexité**. Cette observation conforte la **cohérence** de notre approche et témoigne de sa **fidélité au modèle épidémiologique** de Cori, assurant l'**interprétabilité** des résultats obtenus.

Les travaux présentés ont fait l'objet d'une prépublication : B. Pascal, P. Abry, N. Pustelnik, S. Roux, R. Gribonval, et P. Flandrin, "Nonsmooth convex optimization to estimate the Covid-19 reproduction number space-time evolution with robustness against low quality data.", 2021, arxiv:2109.09595.

Contact : barbara.pascal@univ-lille.fr

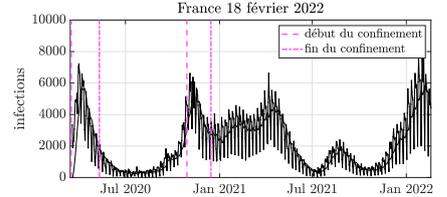


FIGURE 1 – Nouvelles infections

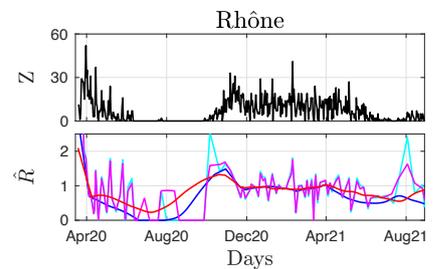


FIGURE 2 – Estimations de $R(t)$ par résolution de (1).