MATSIP TP – Estimation of the Reproduction Number of COVID-19

Jayotsana Sharma

September 2025

Exercise 2

(a) Plot of Z_t and Φ_t

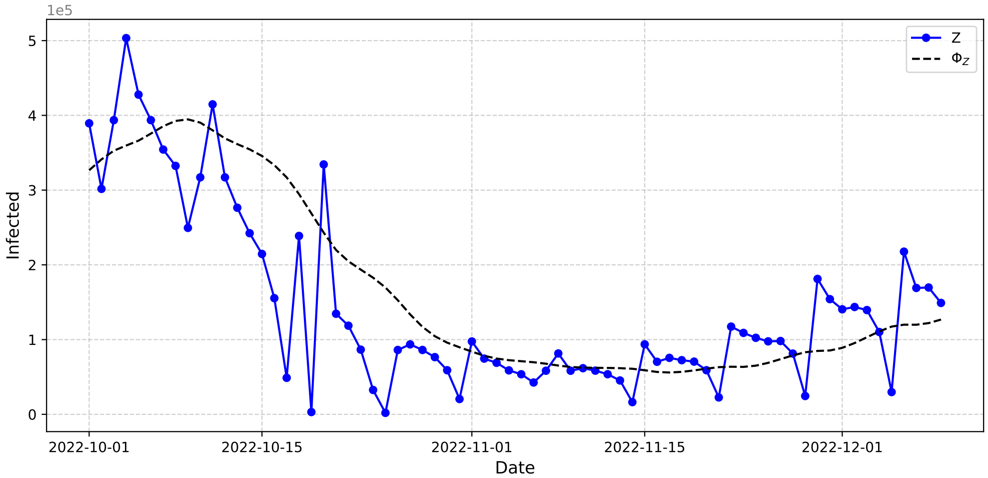

Figure 1: The daily new infections Z_t and Φ_t with respect to days

(b) Plot of $d_{KL}(Z_t \mid p_t)$

For a fixed Z_t, the divergence is

$$d_{KL}(Z_t \mid p_t) = \begin{cases} Z_t \ln \dfrac{Z_t}{p_t} + p_t - Z_t, & Z_t > 0,\ p_t > 0, \\ p_t, & Z_t = 0,\ p_t \geq 0, \\ +\infty, & \text{otherwise.} \end{cases}$$

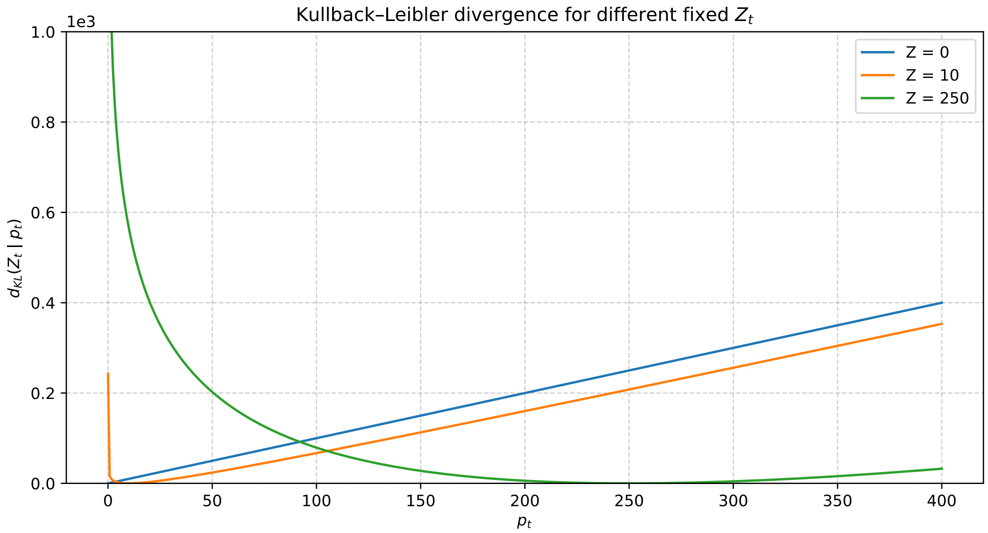

Figure 2: For $Z_t \in \{0, 10, 250\}$, the plot of $p_t \mapsto d_{KL}(Z_t \mid p_t)$

(c) Convexity and Differentiability

For $Z_t > 0$:

$$f(p) = Z_t \ln \frac{Z_t}{p} + p - Z_t, \quad f'(p) = -\frac{Z_t}{p} + 1, \quad f''(p) = \frac{Z_t}{p^2} > 0.$$

Hence f is strictly convex and infinitely differentiable (C^∞) on $(0, \infty)$. For $Z_t = 0$, $f(p) = p$ is linear and convex.

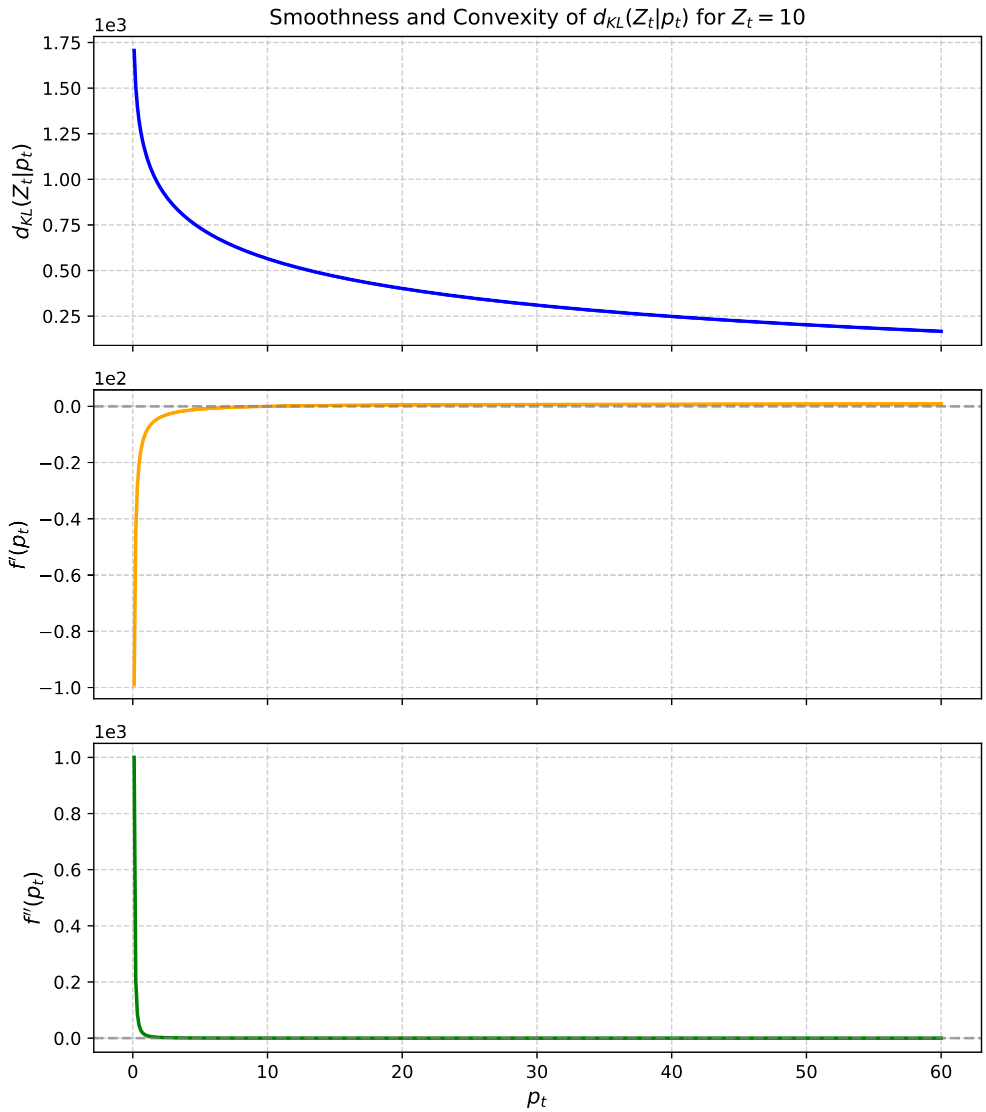

Figure 3: At $Z_t = 10$, plots of $d_{KL}(Z_t \mid p_t)$, $d'_{KL}(Z_t \mid p_t)$, and $d''_{KL}(Z_t \mid p_t)$ for convexity.

Summary.

- $f \in C^\infty((0,\infty))$ for $Z_t > 0$;
- convex on its whole effective domain;
- not continuous at $p_t = 0$ if $Z_t > 0$.

(d) Lipschitz Continuity of the Gradient

$$f''(p) = \frac{Z_t}{p^2} \;\Rightarrow\; \sup_{p>0} f''(p) = +\infty.$$

Therefore the gradient is *not globally Lipschitz*. It is only Lipschitz on any set $[\varepsilon, \infty)$ with constant $L = Z_t/\varepsilon^2$. For $Z_t = 0$, $f'(p) = 1$ is constant (0-Lipschitz).

Exercise 3.1 – Maximum Likelihood Estimator

(a) Single-Day Minimization

For $\Phi_t^Z > 0$,

$$f(R_t) = Z_t \ln \frac{Z_t}{R_t \Phi_t^Z} + R_t \Phi_t^Z - Z_t, \quad f'(R_t) = -\frac{Z_t}{R_t} + \Phi_t^Z.$$

Setting $f'(R_t) = 0$ gives

$$\boxed{R_t^\star = \frac{Z_t}{\Phi_t^Z}}, \qquad f''(R_t) = \frac{Z_t}{R_t^2} > 0.$$

Minimum value $f(R_t^\star) = 0$.

(b) Global MLE

Because $\mathrm{DKL}(R) = \sum_t d_{KL}(Z_t \mid R_t \Phi_t^Z)$ is separable,

$$\boxed{\hat{R}_{\text{MLE}} = \frac{Z}{\Phi^Z}}$$

(c) Numerical Result and Plot

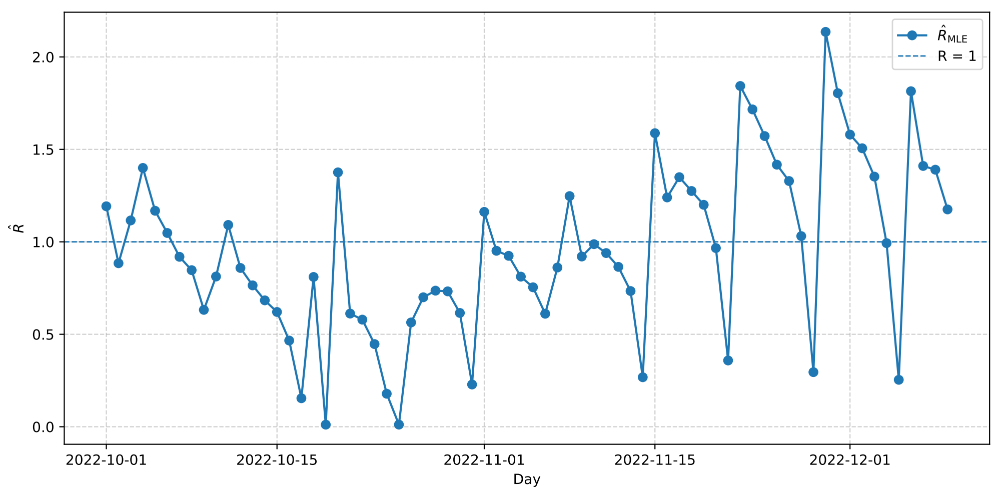

Figure 4: MLE $\hat{R}_t$ for daily. Values above 1 indicate epidemic growth; below 1, decay.

(d)

The MLE $R_t = Z_t/\Phi_t^Z$ fluctuates heavily due to reporting noise and small denominators. It is unbiased but *unstable*.

Exercise 3.2

We minimize

$$\min_R \ \mathrm{DKL}(Z \mid R \odot \Phi^Z) + \lambda \|D_2 R\|_1, \quad (D_2 R)_t = R_{t+2} - 2R_{t+1} + R_t.$$

(a) Forward–Backward

Because ∇DKL is not Lipschitz, forward–backward cannot be used.

(b) Prox of $p \mapsto d_{KL}(Z \mid p)$

(b) Proximity operator of $p \mapsto d_{\mathrm{KL}}(Z_t \mid p)$ For $\gamma > 0$, the proximity operator is defined as

$$\mathrm{prox}_{\gamma f}(v) = \arg\min_{p \geq 0} \left(f(p) + \frac{1}{2\gamma}(p - v)^2 \right).$$

Case 1: If $Z_t = 0$, then $f(p) = p$ on $p \geq 0$, hence

$$\mathrm{prox}_{\gamma f}(v) = \max(v - \gamma, 0).$$

Case 2: If $Z_t > 0$, we solve the first-order condition on $p > 0$:

$$-\frac{Z_t}{p} + 1 + \frac{1}{\gamma}(p - v) = 0 \quad \Longleftrightarrow \quad \frac{1}{\gamma}p^2 + \left(1 - \frac{v}{\gamma}\right)p - Z_t = 0.$$

This is a quadratic equation with coefficients

$$a = \frac{1}{\gamma}, \qquad b = 1 - \frac{v}{\gamma}, \qquad c = -Z_t.$$

The positive root gives

$$\mathrm{prox}_{\gamma d_{\mathrm{KL}}(Z_t|\cdot)}(v) = \frac{v - \gamma + \sqrt{(v-\gamma)^2 + 4\gamma Z_t}}{2}.$$

(c) Proximity operator of $R_t \mapsto d_{\mathrm{KL}}(Z_t \mid R_t \Phi_t Z)$

Let $a_t = \Phi_t Z$ (known, nonnegative). Define

$$h_t(R_t) = d_{\mathrm{KL}}(Z_t \mid a_t R_t).$$

Change variable $p = a_t R_t$. Then minimizing

$$d_{\mathrm{KL}}(Z_t \mid p) + \frac{1}{2\gamma}(p/a_t - v)^2$$

is equivalent to computing the proximity operator of d_{KL} at a scaled argument.

Hence,

$$p^\star = \mathrm{prox}_{\frac{\gamma}{a_t^2} d_{\mathrm{KL}}(Z_t|\cdot)}(a_t v), \qquad \mathrm{prox}_{\gamma h_t}(v) = \frac{p^\star}{a_t}.$$

(d) Proximity operator of the full data term $D_{\mathrm{KL}}(Z \mid R \odot \Phi Z)$

The functional separates over the temporal index t, hence

$$\mathrm{prox}_{\gamma\, D_{\mathrm{KL}}(Z|\cdot\odot\Phi Z)}(v) = \big(\mathrm{prox}_{\gamma\, h_t}(v_t)\big)_{t=1}^{T}, \qquad h_t(R_t) = d_{\mathrm{KL}}(Z_t \mid \Phi_t Z\, R_t).$$

Therefore, the proximity operator of the full Kullback–Leibler data term is obtained by applying, componentwise in t, the scalar formula derived in **(b)** (for $p \mapsto d_{\mathrm{KL}}(Z_t \mid p)$) together with the scaling rule of **(c)** (for h_t):

$$\mathrm{prox}_{\gamma\, D_{\mathrm{KL}}(Z|\cdot\odot\Phi Z)}(v) = \left(\frac{1}{a_t}\ \mathrm{prox}_{\frac{\gamma}{a_t^2} d_{\mathrm{KL}}(Z_t|\cdot)}(a_t v_t) \right)_{t=1}^{T}, \qquad a_t = \Phi_t Z.$$

(e) Prox of $\lambda\sigma^{-1}\|\cdot\|_1$

Soft-thresholding:

$$\operatorname{prox}_{\lambda\sigma^{-1}\|\cdot\|_1}(q) = \operatorname{sign}(q)\max(|q| - \lambda/\sigma, 0).$$

(f) Discrete Laplacian D_2

$$D_2 = \begin{bmatrix} 1 & -2 & 1 & 0 & \cdots \\ 0 & 1 & -2 & 1 & \cdots \\ & & \ddots & & \end{bmatrix}.$$

(g) Normalization trick

Let

$$\tilde{Z} = \frac{Z}{\alpha}, \qquad \tilde{\Phi} Z = \frac{\Phi Z}{\alpha}, \qquad \alpha > 0.$$

Then, by homogeneity of the Kullback–Leibler divergence, we have

$$D_{\mathrm{KL}}(\tilde{Z} \mid R \odot \tilde{\Phi} Z) = \sum_t d_{\mathrm{KL}}(\tilde{Z}_t \mid R_t \tilde{\Phi}_t Z) = \frac{1}{\alpha}\sum_t d_{\mathrm{KL}}(Z_t \mid R_t \Phi_t Z) = \frac{1}{\alpha} D_{\mathrm{KL}}(Z \mid R \odot \Phi Z).$$

Thus, running **Algorithm 1** with the normalized pair $(\tilde{Z}, \tilde{\Phi} Z)$ and a rescaled regularization parameter

$$\tilde{\lambda} = \frac{\lambda}{\alpha}$$

yields the same minimizer R as with the unscaled quantities.
In practice, it is recommended to set

$$\alpha = \operatorname{std}(Z) \quad \text{so that} \quad \tilde{\lambda} = \frac{\lambda}{\operatorname{std}(Z)},$$

in order to improve the numerical conditioning of the problem.

(h) and (i)

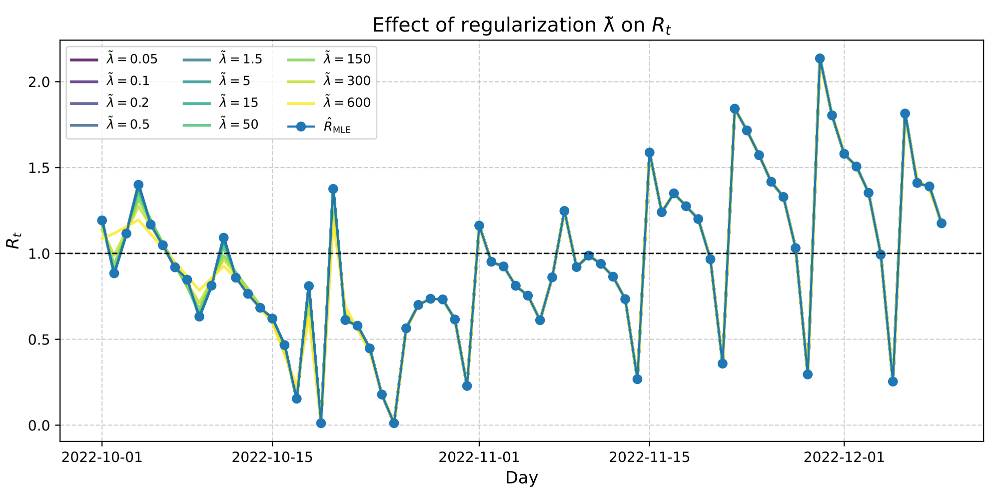

Figure 5: The variation of regularization parameter $\tilde{\lambda} = \tilde{\lambda} \in \{0.5, 3.5, 15, 50, 150, 250\}$ after normalization on R_t.

3.3 Tikhonov penalization

We now solve

$$\min_{R\in\mathbb{R}^T} \underbrace{DKL\big(Z \,\big|\, R\odot\Phi^Z\big)}_{=:g(R)} + \lambda \underbrace{\|D_2R\|_2^2}_{=:f(D_2R)},$$

where D_2 is the discrete Laplacian defined in the sheet and Φ^Z the (fixed) global infectiousness. We use the **Chambolle–Pock (CP) primal–dual** scheme exactly as in Algorithm 1, only changing the dual prox to account for the squared ℓ_2 penalty; the stepsize condition $\sigma\tau\|D_2\|_{\text{op}}^2 < 1$ is unchanged.

(a) Proximity of $x \mapsto \|x\|_2^2$. For $\gamma > 0$,

$$\text{prox}_{\gamma\|\cdot\|_2^2}(x) = \arg\min_u \tfrac{1}{2}\|u-x\|_2^2 + \gamma\|u\|_2^2 = \frac{x}{1+2\gamma}.$$

(b) CP with Tikhonov (ℓ_2^2) instead of ℓ_1. We write the problem as $\min_R g(R) + f(KR)$ with $K = D_2$ and $f(y) = \lambda\|y\|_2^2$. In CP we need

- the **dual proximal** $\text{prox}_{\sigma f^*}$,
- and the **primal proximal** $\text{prox}_{\tau g}$.

Dual prox (closed form "shrink by factor"). The convex conjugate of $f(y) = \lambda\|y\|_2^2$ is $f^*(q) = \frac{1}{4\lambda}\|q\|_2^2$. Hence

$$\text{prox}_{\sigma f^*}(v) = \arg\min_u \tfrac{1}{2}\|u-v\|_2^2 + \sigma\tfrac{1}{4\lambda}\|u\|_2^2 = \frac{v}{1+\sigma/(2\lambda)}.$$

Thus, the **dual update** becomes

$$Q^{k+1} = \frac{Q^k + \sigma\, D_2\overline{R}^k}{1+\sigma/(2\lambda)}.$$

Primal prox for the KL term (same as §3.2). The function $g(R) = \sum_t d_{KL}\big(Z_t \,\big|\, R_t\,\Phi_t^Z\big)$ is separable in t. For each t, the proximal map has the closed form (the same formula used in 3.2):

$$\boxed{\big[\text{prox}_{\tau g}(r)\big]_t = \frac{r_t - \tau\,\Phi_t^Z + \sqrt{(r_t-\tau\,\Phi_t^Z)^2 + 4\tau\,Z_t}}{2}}$$

(This is the proximal of $d_{KL}(Z_t\,|\,\cdot)$ composed with the affine map $p_t = R_t\Phi_t^Z$, simplified to a closed form in R_t; it is exactly the same formula given in the lab for the Poisson KL proximal with $Z\phi$.)

(c) Plot both estimates and comment. Both looks same in the Figure below:

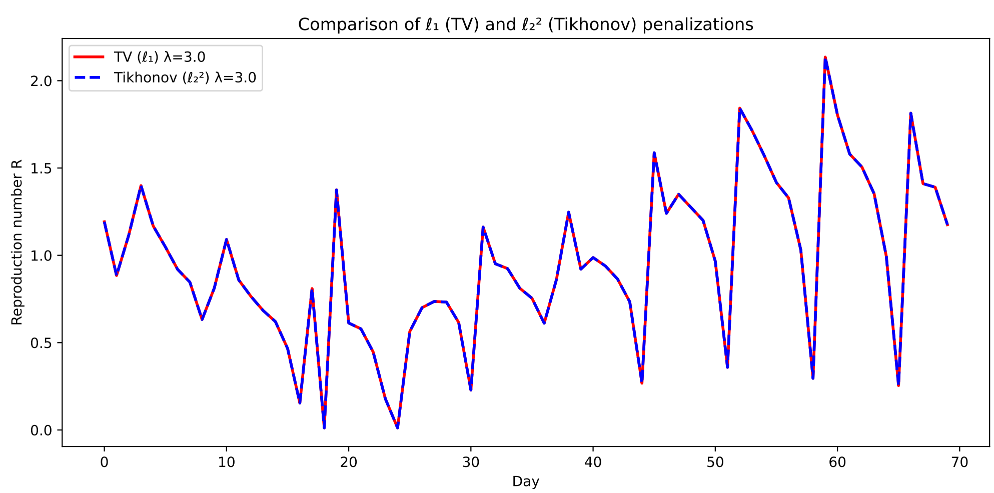

Figure 6: The regularization parameter $\tilde{\lambda} = 3$.