Total Variation Regularization for Compartmental Epidemic Models with Time-varying Dynamics

Wenjie Zheng
contact@zhengwenjie.net

Abstract

Traditional methods to infer compartmental epidemic models with time-varying dynamics can only capture *continuous* changes in the dynamic. However, many changes are *discontinuous* due to sudden interventions, such as city lockdown and opening of field hospitals. To model the discontinuities, this study introduces the tool of total variation regularization, which regulates the temporal changes of the dynamic parameters, such as the transmission rate. To recover the ground truth dynamic, this study designs a novel yet straightforward optimization algorithm, dubbed iterative Nelder-Mead, which repeatedly applies the Nelder-Mead algorithm. Experiments on the simulated data show that the proposed approach can qualitatively reproduce the discontinuities of the underlying dynamics. To extend this research to real data as well as to help researchers worldwide to fight against COVID-19, the author releases his research platform as an open-source package.

1 Introduction

Compartmental models (Kermack and McKendrick, 1927), such as SIR, are an important research subject in epidemiology to quantify infectious disease dynamics (Siettos and Russo, 2013). This family of models divides the population into several **compartments**, among which each individual transfers according to some **dynamic**. For instance, the SIR model prescribes three compartments, with the `Susceptible` compartment for those vulnerable, the `Infectious` compartment for those contagious, and the `Removed` compartment for those either obtained immunity after the recovery or died from the infection (Figure 1). The dynamic, on the other hand, governs the mechanic of how each individual transfers among these three compartments.

Within this fruitful research branch, a moderate quantity of effort is spent in capturing the *time-varing* aspect of the dynamic. That is, the dynamic is not a constant one through the entirety of the epidemic; rather, it varies due to changing population behaviors, public interventions, seasonal effects, viral evolution, etc. This type of models is not only, obviously, more realistic but also more coherent to empirical studies. For instance, the 1918 influenza pandemic displayed "three distinct waves" of infection within a 12-month period (He et al., 2011, p. 283). This kind of phenomenon can be explained only by a time-varying dynamic.

To capture the time-varying aspect, nearly all efforts are around the hypothesis that the **parameters** characterizing the dynamic are continuous deterministic functions or stochastic processes of time. One parameter particularly honored by this privilege is the **transmission rate**, which quantifies how often an infectious infects a susceptible. In the case of continuous deterministic function, it has been modeled as exponential functions (Chowell et al., 2004; Althaus, 2014), sigmoid functions (Camacho et al., 2014), sinusoid functions (Stocks et al., 2018), cubic B-splines (He et al., 2011), and Legendre polynomials (Smirnova et al., 2017). In the case of continuous stochastic process, it has been modeled as Wiener process (Dureau, Kalogeropoulos and Baguelin, 2013; Funk et al., 2016; Cazelles et al., 2018; Kucharski et al., 2020) and, more generally, Gaussian processes with periodic kernel or squared exponential kernel (Rasmussen et al., 2011; Xu et al., 2016).

Whilst continuous machinery is useful for capturing the time-varying aspect of the dynamic, it is not suitable to capture the sudden shocks on the dynamic, which entails *discontinuity*. For instance, during the recent Coronavirus Disease 2019 (COVID-19) pandemic, multiple regions (*e.g.*, Wuhan, Italy, France) were suddenly closed off. These measures, especially the draconian one in Wuhan,

were aimed at instantly reducing the transmission rate. Modeling them via a continuous function or process has the danger of smoothing out the transmission rate shift and thus underestimates the efficiency of these measures.

For the accurate detection of such sudden changes in the dynamic, this article does not impose the property of continuity let alone smoothness on the dynamic. Instead, the model estimation error is controlled by **total variation regularization**. Total variation regularization has the effect of detecting discontinuities within the investigated object (*e.g.*, function, process). It is used, among others, in image denoising, wherein it successfully restores the sharpness of images. The hence restored object has a well-known *staircase* visual effect. By applying total variation regularization on the calibration of epidemic dynamics, we can hope to reconstruct a dynamic mostly constant while still allowing some *phase shifts*.

It is worth noting that the approach proposed here is fundamentally different from the *piecewise* approach (such as in Funk et al., 2017). The latter artificially breaks the epidemic into several periods and then models each time period individually, whilst the former does not presume any locations to implant such breakages and, instead, lets the data speak for itself. The latter is reasonable for the modeling of public interventions, which generate *foreseeable* sudden impact on the epidemic dynamic. The former is more general and can additionally capture *invisible* phase shifts induced by, say, viral evolution. Furthermore, total variation regularization can still be preferable for the modeling of foreseeable phase shifts, for these shifts may not happen *immediately* after, say, the public intervention. The piecewise approach may neglect the delays and thus underestimate the efficiency of the interventions.

To apply total variation regularization in a *principled* way, this research adopted the well-known **state-space framework**[1] in epidemiology. This framework supposes that the underlying compartment status is a latent object and hence not directly observable. To infer the latent status, we can only collect secondary information which is derived from it. Using the machine learning terminology, it can be regarded as a Hidden Markov Model (HMM). State-space framework enables the **evidence synthesis** approach, which leverages data from multiple sources: Surveillance data (*i.e.*, prevalence and incidence) can be supplemented by additional serological, demographic, administrative, environmental, or phylogenetic data[2]. Interested readers are referred to Birrell et al. (2018, Section 3)'s review for some examples. A recent study using phylogenetic data to infer the epidemic of COVID-19 can be found in Kucharski et al. (2020)'s work.

To infer the latent dynamic in the state-space framework, researchers almost exclusively adopt some Monte Carlo methods such as Sequential Monte Carlo (*a.k.a.* particle filter) or particle Markov Chain Monte Carlo (pMCMC). The procedure starts with the combination of the prior provided by the aforementioned hypothesis on the latent dynamic and the likelihood provided by the observation mechanism, followed by the simulation of the posterior via some Monte Carlo method. This procedure is so streamlined that an entire software package[3] has been developed for it (Dureau, Ballesteros and Bogich, 2013).

Here, this article instead designs an **iterative Nelder-Mead** algorithm towards the **maximum a posteriori (MAP)** estimate, where the objective function of interest is the likelihood regularized by total variation[4]. In contrast to Monte Carlo methods estimating the posterior mean, MAP focuses on the posterior mode. Former experiences in image denoising suggest that MAP restores the discontinuities in the investigated object much better than the posterior mean does. Therefore, the key here is to design a suitable optimization algorithm to find the *global* optimum. Experiences conducted in this study show that the hereby proposed iterative Nelder-Mead algorithm is a qualified candidate for this purpose, and that it qualitatively reconstructs the underlying dynamic.

It is worth noting that, due to the nonparametric nature of the model adopted, the objective function under investigation here is neither convex nor unimodal, which differentiates this work from many others also applying the Nelder-Meal algorithm but on a unimodal objective thanks to a parametric model. When the objective is unimodel, all reasonable descent algorithms all converge to the

[1] Originally named State-Space Model (SSM). Here I mimicked several other researchers (such as Birrell et al., 2018) and named it as a framework to prevent any confusion with the modeling of the dynamics.

[2] The last one is related to coalescent theory.

[3] https://github.com/StateSpaceModels/ssm.

[4] The regularization can be regarded as an equivalent of prior. More on this later.

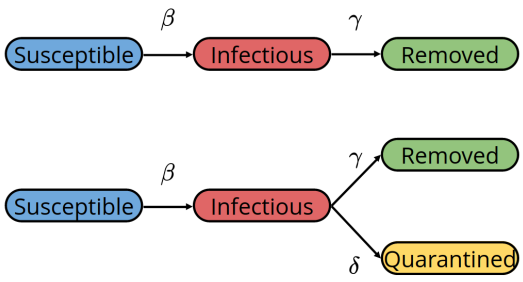

Figure 1: Upper: SIR. Lower: SIRQ.

same optimum (ignoring the convergence rate). Nevertheless, when the objective is not unimodel, algorithms can be easily trapped in some local optimum. It takes therefore much skill to reach the global optimum. In particular, this study discovered that the regularization hyperparameter does not only controls the bias of the model but also governs the topology of the objective function. That is, a too small value or a too large value can both render the objective difficult to optimize and hence trap the optimization algorithm.

The contribution of this study is 3-fold.

- It is the first one to propose total variation regularization to capture the discontinuity of time-varying epidemic dynamics. Moreover, it is the first one to apply the nonparametric approach simultaneously on more than one parameter. Thanks to the state-space framework it adopted, the methodology of this study is generalizable.
- It is the first one to (successfully) use *non* Monte Carlo methods for the inference of nonparametric compartmental models. The iterative Nelder-Mead algorithm designed for this purpose reveals the role of hyperparameters in tuning the topology of the objective function. Such knowledge might help solve the mysteries in the training of deep neural networks.
- Although the proposed approach is tested only on simulated data, I release my research platform as an open-source package to help researchers and practitioners worldwide to fight against COVID-19.

This paper is organized as follows. Section 2 introduces the most classic compartmental model **SIR** as well as designs an extension **SIRQ**. The former describes an environment without interventions, whilst the latter adds an additional compartment `Quarantined` to model the intervention. Section 3 describes the state-space framework, whose two components are the state equation and the observation equation. Section 4 presents the idea of total variation regularization and its potential to detect discontinuity. To solve the associated multimodal objective function, the iterative Nelder-Mead algorithm is proposed. Section 5 tests the proposed approach on simulated data, and the results show that the dynamic can be qualitatively recovered.

2 Compartmental models: SIR and SIRQ

SIR, standing for `Susceptible-Infectious-Removed`, is the basic compartmental model. It is important albeit simple. By using SIR as a stepping stone, we can understand more complex models such as SIRQ, which is promoted as a novel model here.

2.1 SIR model

The SIR model separates the population into three compartments: `susceptible`, `infectious`, and `removed`. Each individual (logically) transfers among these compartments according to his health status (Figure 1).

Susceptible: The population in this compartment are healthy people who are vulnerable to the disease.

Infectious: The population in this compartment are infected people who are free to infect those susceptible.

Removed: This compartment consists of two groups of people – those died from the disease or recovered from it and hence obtained immunity.

To describe the dynamic governing the mechanism, two types of setups are possible – stochastic or deterministic. The stochastic one assumes that each individual transfers according to some probability. The usual setup is that the probability of a healthy people transferring from `susceptible` to `infectious` follows an exponential distribution of parameter β, and the probability of an infected people transferring from `infectious` to `removed` follows an (independent) exponential distribution of parameter γ.

On the other hand, the deterministic one simplifies the above process by taking advantage of the law of large number. Instead of studying the dynamic on an individual basis, the deterministic setup considers it at the aggregate level. That is, the number of individual in each compartment varies according to some ordinary differential equation.

Let S_t, I_t, and R_t be the number of susceptible, infectious, and removed[5] at time t, respectively. Let $N_t = S_t + I_t + R_t$ be the number of total population, which is supposed to be constant (*i.e.*, there is no newborn or death because of reasons other than the infectious disease in question). Then the stochastic dynamic can be expressed by the following equations.

$$\begin{aligned}
&\Pr(S_{t+h} - S_t = -1, I_{t+h} - I_t = 1 | S_t, I_t, R_t) = \beta S_t I_t h / N_t + o(h), \\
&\Pr(I_{t+h} - I_t = -1, R_{t+h} - R_t = 1 | S_t, I_t, R_t) = \gamma I_t h + o(h).
\end{aligned}$$

The deterministic dynamic can be expressed by the following ordinary differential equations (ODE).

$$\begin{aligned}
\frac{\mathrm{d}S_t}{\mathrm{d}t} &= -\frac{\beta S_t I_t}{N_t}, \\
\frac{\mathrm{d}I_t}{\mathrm{d}t} &= \frac{\beta S_t I_t}{N_t} - \gamma I_t, \\
\frac{\mathrm{d}R_t}{\mathrm{d}t} &= \gamma I_t.
\end{aligned}$$

For large-scale epidemic, the deterministic dynamic is a good enough approximation of the stochastic dynamic (Kurtz, 1987) (see Figure 2). Interested readers are also referred to Siettos and Russo (2013, pp. 301) for a quick review and to Birrell et al. (2018, Section 3.2) for a detailed discussion.

The parameter β is called the transmission rate, and γ is called the removal rate. The ratio β/γ is associated with the most important quantity of infectious diseases – the **basic reproduction number** $\mathcal{R}_0 = \frac{\beta}{\gamma}$[7], which stands for the average number of victims an infectious is expected to infect at the very beginning of the outbreak. If $\mathcal{R}_0 > 1$, the disease becomes an epidemic; if $\mathcal{R}_0 < 1$, the disease dies out; if $\mathcal{R}_0 = 1$, it is an endemic (*i.e.*, the number of infectious neither grows nor deceases[8]).

2.2 SIRQ model

There are many extensions to SIR. Here, I designs another one, dubbed `susceptible-infectious-removed-quarantined` (SIRQ), to include the influence of public interventions. SIRQ creates a fourth compartment `quarantined`, which hosts the part of infectious getting quarantined or hospitalized (Figure 1). Therefore, the infectious can have two futures: either they stay wild and get nature selected (*i.e.*, removed) as in the SIR model, or they get quarantined, which also prevents them from infecting others.

[5] Here, I slightly abused the terminology by using the name of the compartment to denote people within that compartment.

[6] `www.zhengwenjie.net/sir`.

[7] Be careful not to confuse it with the number of population in the `removed` compartment at time 0, denoted by the plain R_0.

[8] Although the number of infectious remains constant, the total number of victims still increases. There is just a dynamic balance between those newly infected and those newly removed.

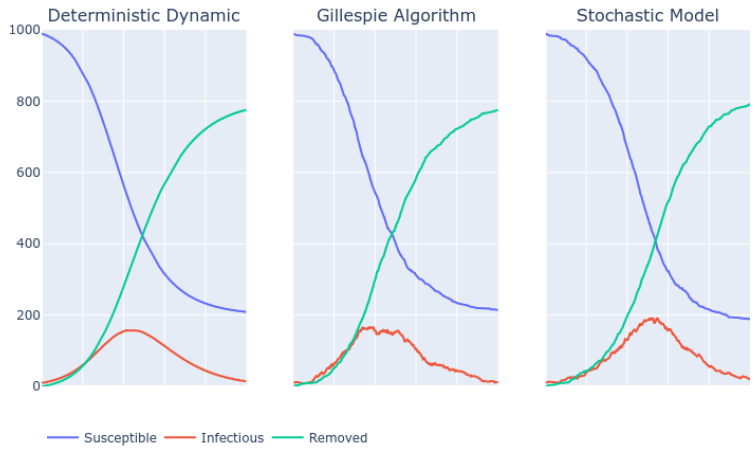

Figure 2: SIR model simulated with various algorithms. Figure from Wenjie Zheng's blog[6]. Left: deterministic dynamic, ODE with Euler scheme. Middle: stochastic dynamic, Gillespie algorithm (one instance). Right: stochastic dynamic, vanilla simulation (one instance).

The dynamic is very similar to the above. Here I write down only the deterministic one.

$$\begin{aligned}
\frac{\mathrm{d}S_t}{\mathrm{d}t} &= -\frac{\beta S_t I_t}{N_t}, \\
\frac{\mathrm{d}I_t}{\mathrm{d}t} &= \frac{\beta S_t I_t}{N_t} - \gamma I_t - \delta I_t, \\
\frac{\mathrm{d}R_t}{\mathrm{d}t} &= \gamma I_t, \\
\frac{\mathrm{d}Q_t}{\mathrm{d}t} &= \delta I_t,
\end{aligned}$$

where Q_t is the number of quarantined at time t, $N_t = S_t + I_t + R_t + Q_t$ is the total number of the population, and δ is the rate the infectious getting quarantined. The ratio γ/δ reflects the ratio of the infectious staying under radar. Here, the quarantined are assumed not infectious (though they may infect the healthcare personnel, we consider this possibility to to be low). Also, we do not further specify the outflow of the `quarantined` compartment, so it contains three types of people – those being quarantined, those died during the quarantine, and those recovered during the quarantine.

This model is particularly relevant to the situation of COVID-19. On the one hand, many infectious of COVID-19 are asymptomatic. They will hence not go to hospital, and they get recovered all by themselves. On the other hand, given the high transmissibility of COVID-19, there is not enough hospital resource for each patient. Many infectious have to stay wild and get nature selected. Besides the above two reasons, there is another one specific to China, mainland: the recall of the test kits is unsatisfying.

One usage of this model is to speculate the ratio of asymptomatic patients or the ratio of hospitalization for the evaluation of the government efficiency. The `removed` compartment is supposed to be undetectable, whilst the `quarantined` compartment can be accurately detected by the confirmed cases. Experiments show that, in the parametric setting, the ratio of asymptomatic patients or the like can be accurately inferred given only sparse information on the infectious and the quarantined.

Concerning the basic reproduction number, it has two choices. The controlled version uses $\frac{\beta}{\gamma+\theta}$, whose value determines whether the disease will become an epidemic or die out. The uncontrolled version uses $\frac{\beta}{\gamma}$, which stands for the outcome if the quarantine measure is ever called off.

3 State-space framework

The state-space framework has become the *de facto* state of the art for the usage of compartment models. Many studies, such as the one by Wu et al. (2020), use this framework implicitly. This

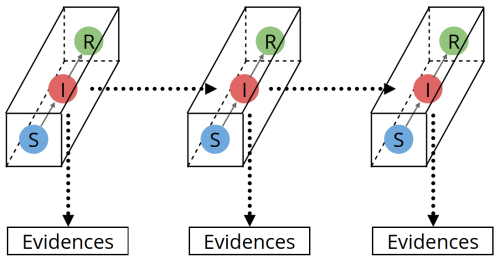

Figure 3: A state-space framework example featuring SIR.

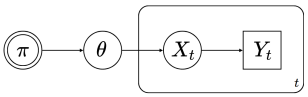

Figure 4: Graphical model for the state-space framework.

section will first lay down a solid mathematical foundation of the state-space framework—which will later facilitate the introduction of total variation regularizatio—and then complement it with some concrete examples.

3.1 Mathematical foundation

The state-space framework is very similar to the Hidden Markov Model in machine learning. It assumes that the compartment status is in the invisible state space, and that we can only observe secondary information derived from the states. Let us illustrate this concept with SIR. The vector (S_t, I_t, R_t) forms the (invisible) state at time t. The dynamic characterized by the parameter (β, γ) connects the temporally successive states (*e.g.*, (S_t, I_t, R_t) and $(S_{t+1}, I_{t+1}, R_{t+1})$). The state at time t determines the evidence data that we could collect at time t (Figure 3).

In the state-space framework, there are usually two tasks. The first one is to infer the underlying state status, which tells us the severity of the current epidemic. The second one is to characterize the properties (*e.g.*, $\mathcal{R}_0$) of the epidemic itself, which helps us evaluate the ferocity of our enemy. These two tasks are interdependent: the fulfillment of one entails the other.

It is a principled way to infer these two tasks together through the graphical model (Figure 4). The epidemic dynamic parameter θ (in SIR, $\theta = (\beta, \gamma)$) governs the temporal transition of the states X_t (in SIR, $X_t = (S_t, I_t, R_t)$):

$$X_{t+1}|X_t \sim p_\theta(\cdot|X_t) \quad \text{(state equation)}$$

This probability distribution can be degenerate, in which case it is reduced to a *deterministic* epidemic dynamic. The previous section is entirely dedicated for the description of p_θ. Then, the invisible states X_t generates some empirical evidence, denoted as Y_t:

$$Y_t|X_t \sim p_{\bar{\eta}}(\cdot|X_t) \quad \text{(observation equation)},$$

where $\bar{\eta}$ is often manually selected by the researchers to prevent any unintended complexity. Bayesian statisticians may go further by supposing the parameter θ lies in some probability space. Thus, they will include in the model a prior:

$$\theta \sim \bar{\pi}(\cdot) \quad \text{(prior)},$$

where the prior $\bar{\pi}$ is preset. To infer the state X_t and the parameter θ, we can build their posterior distribution, which is proportional to the joint distribution. Let x_i be the value observed of X_i and let $x_{1:T}$ denote the tuple $x_1, \ldots, x_T$, then the posterior distribution

$$p(\theta, x_{1:T}|y_{1:T}) \propto p_{\bar{\eta}}(y_{1:T}|x_{1:T}) p_\theta(x_{1:T}) \bar{\pi}(\theta) \quad \text{(posterior)}.$$

The above assumes the Markov property. In practice, the state-space framework can be more general by abandoning the Markov property or by introducing dependence between Y_t. Since the basic

is already enough for understanding this article, we will stop here and turn to some examples. In particular, the candidates for the state equation are already described in the previous section in the form of SIR and SIRQ. The following will only present some candidates for the observation equation and the prior.

3.2 Candidates for the observation equation

The observation equation is used to introduce evidences for the inference of the dynamic and the latent states. Traditionally we use only surveillance data, but now an increasing number of studies start to use exotic data such as serological, demographic, administrative, environmental, and phylogenetic data (Birrell et al., 2018).

The first example is how researchers used traveling data to infer the total number of COVID-19 patients in Wuhan, China. In January, 2020, there are a large number of people infected by COVID-19 in Wuhan city. Unaware of the infection, they traveled abroad and later got diagnosed. Imai et al. (2020) modeled this natural experiment as sampling with replacement. That is, the number of patients diagnosed abroad follows a binomial distribution $\mathrm{Bin}(m, I/N)$, where m stands for the total number of outbound travelers, I the number of infectious, and N the total population. Rigorously speaking, the natural experiment is more like sampling *without* replacement, but the difference is neglectable when $m \ll N$. Wu et al. (2020) modeled it, alternatively, as a Poisson distribution $\mathrm{Poi}(mI/N)$. Incidentally, there is no fundamental difference between the above two options, for $\mathrm{Bin}(m, I/N) \approx \mathrm{Poi}(mI/N)$ when $m > 20$ and $I \ll N$ thanks to the law of rare events.

The second example is based more on my personal opinion than on the literature: I advocate to use the number of confirmed cases to infer the `quarantined` compartment. It is attempting to use the confirmed cases for the `infectious` compartment. Nonetheless, a smarter and more sensible arrangement is to use them for the `removed` compartment in SIR or the `quarantined` compartment in SIRQ. When a person is confirmed infected, he most likely will be admitted to the hospital and thus lose the ability to infect others (ignoring the minor risk of infecting the healthcare personnel). This is essentially the `removed` compartment in SIR used for. If the hospitalization is imperfect (*i.e.*, a part of patients are not confirmed and have to be nature selected), this is a perfect scenario for the SIRQ model, where the confirmed cases can be associated with the `quarantined` compartment. Xu et al. (2016, Section 5.2) thought alike and used the confirmed cases to infer the `removed` compartment. In this article, I will use the confirmed cases for the `quarantined` compartment. The observation equation can either feature the Gaussian distribution or the Poisson distribution. In fact, there is no fundamental difference, for $\mathtt{Poi}(\lambda) \approx \mathcal{N}(\lambda, \lambda)$ for sufficiently large λ.

The third example is to use the serological data to infer the `removed` compartment by detecting the antibody in the blood. In an imperfect quarantine scenario modeled by SIRQ, some patients survived the disease without formal medical intervention. This part of patients are never administratively confirmed, but their existence and the death of the unconfirmed together comment on the severity of the epidemic. To fairly evaluate the epidemic, it is essential to estimate the portion of unconfirmed cases. Since recovered people will have antibody in the blood, the serological data can help us screen out this group of people. To get an unbiased estimate, we can sample the whole population, then it is reduced to an elementary statistics problem.

3.3 Candidates for the prior

The prior concerns the preset distribution on the parameter characterizing the dynamic (the parameter characterizing the observation equation is preset). This parameter can be finite-dimensional (vector) or infinite-dimensional (function). The finite-dimensional cases usually use an uninformative prior (*i.e.*, constant), and the posterior degenerates to the plain likelihood. It is only in the infinite-dimensional cases that the selection of prior becomes nontrivial.

In the infinite-dimensional case, the parameter θ_t is time-varying. We are to sample functions for θ_t in some function space. In other words, θ_t is a stochastic process. There are mainly two candidate spaces for this purpose. The first defines the stochastic process by a stochastic differential equation:

$$\mathrm{d}h(\theta_t) = \mu_{t,\theta}\,\mathrm{d}t + \sigma_{t,\theta}\,\mathrm{d}B_t,$$

where $\mu_{t,\theta}$ is the drift, $\sigma_{t,\theta}$ is the volatility, B_t is a standard Wiener process, and $h(\cdot)$ is a preset deterministic function. The most common choice for θ_t is Brownian motion where $h(\cdot) = \cdot$ and

geometric Brownian motion where $h(\cdot) = \log(\cdot)$. If the process is expected to converge, Dureau, Kalogeropoulos and Baguelin (2013, p. 4) also proposed the Ornstein Uhlenbeck process.

The second function space is the Gaussian process. A Gaussian process has the property that its arbitrary segments follow a multivariate Gaussian distribution, hence the name. It is wildly used in nonparametric Bayesian statistics. The distribution of the Gaussian process is uniquely defined by its expectation (function) and its kernel (covariance function) $K(\cdot, \cdot)$. Xu et al. (2016) investigated two types of kernels: squared exponential

$$K(x, y) = \alpha^2 \exp[-(x - y)^2/(2\ell^2)]$$

and periodic

$$K(x, y) = \alpha^2 \exp(-\ell^{-1}(1 - \cos(2\pi\omega^{-1}|x - y|))).$$

In the above description, the parameter θ_t is quite abstract. In practice, this θ_t has concrete meaning. For example, in SIRQ, $\theta_t = (\beta_t, \gamma_t, \delta_t)$, in which case, we can apply independent priors individually on each component.

4 Total variation regularization and iterative Nelder-Mead

Prior can limit the model complexity and hence control the model estimation error in the context of bias-variance tradeoff. An alternative approach is to apply a regularization on the log-likelihood. This section introduces the concept of **total variation regularization**, which is wildly used in image denoising. It has the advantage of detecting the discontinuities in the investigated object, and thus it is expected here to capture sudden shocks on the dynamics. In contrast to MCMC, which calculates the posterior mean, this section designs a novel algorithm, dubbed **iterative Nelder-Mead**, to calculate the (regularization) posterior mode.

4.1 Total variation regularization

Section 3.1 formulates the posterior as

$$p_{\bar{\eta}}(y_{1:T}|x_{1:T})p_\theta(x_{1:T})\bar{\pi}(\theta).$$

By applying logarithm, we get

$$\underbrace{\log p_{\bar{\eta}}(y_{1:T}|x_{1:T}) + \log p_\theta(x_{1:T})}_{\text{log-likelihood}} + \underbrace{\log \bar{\pi}(\theta)}_{\text{regularization}},$$

where the last term can be regarded as a regularization. In other words, prior is *one* type of regularization.

An alternative regularization would be substituting the prior for a norm applied on θ:

$$\underbrace{\log p_{\bar{\eta}}(y_{1:T}|x_{1:T}) + \log p_\theta(x_{1:T})}_{\text{log-likelihood}} + \underbrace{\|\theta\|}_{\text{regularization}}.$$

This norm can be, among others, total variation

$$\|\theta\|_{\text{TV}} := \sup \sum_i |\theta_{t_{i+1}} - \theta_{t_i}|,$$

where the supreme runs over the set of all partitions, or **quadratic variation**

$$[\theta] := \lim_{\|P\|\to 0} \sum_i (\theta_{t_{i+1}} - \theta_{t_i})^2,$$

where P ranges over all partitions and the norm is the mesh.

The above frames the prior as a type of regularization; the converse is also true. Total variation regularization or quadratic variation regularization can be regarded as a prior on the space of functions with finite total variation or finite quadratic variation, respectively. In particular, quadratic variation regularization is equivalent to specifying θ as a Brownian motion.

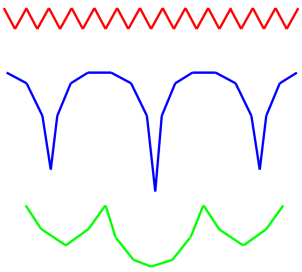

Figure 5: Topology of the objective function with various regularization weights. Top: under-regularized. Middle: over-regularized. Bottom: well-regularized.

4.2 Iterative Nelder-Mead

It is wildly perceived, in the computer vision community, that the posterior mean does not recover the discontinuity as well as the posterior mode does. This subsection describes the challenges to calculate the posterior mode as well as provides solutions to it.

The first challenge is the unavailability of the gradient. Although the part of the likelihood associated with the observation equation can be easily differentiated, the part associated with the state equation does not have closed-forms. Therefore, we have to rely on some zero-order algorithm. One algorithm I see fit is the **Nelder-Mead** method (*a.k.a.* **downhill simplex method**), which is prevalent in civil engineering. For example, to build a suspension bridge, an engineer has to choose the thickness of each strut, cable, and pier. These elements are interdependent, and it is not easy to determine the impact of changing any specific element. Analogously, in our context, changing the value of the dynamic at any single point is unlikely to have a significant impact on the whole epidemic (the integral does not depend on the function values on a null set).

The second challenge is **multimodality** of the objective function, which refers to functions with multiple modes. Two factors contribute to this multimodality. On the one hand, when the observation is sparse, there are many candidate dynamics able to reproduce the observation accurately; each candidate then forms a valley. On the other hand, constant dynamics do not suffer from the regularization penalty, and the modification on a single point will not affect the overall dynamic (null set) but does increase total variation, so each constant dynamic also forms a valley. The optimization algorithm can therefore be easily trapped in some local optimum. To mitigate the influence of multimodality, I designed the **iterative Nelder-Mead** algorithm, which repeatedly rerun Nelder-Mead on the new local optimum (Algorithm 1).

Algorithm 1 Iterative Nelder-Mead

Require: initiate point $\mathbf{x}$, objective function $\mathbf{f}$
  repeat
    $\mathbf{x} \leftarrow$ Nelder-Mead$(\mathbf{f}, \mathbf{x})$
  until stop condition fulfilled

Iterative Nelder-Mead mitigates the problem of multimodality to some extent, but it does not affect the hardship of the problem itself. During the experiments, I discovered that the regularization hyperparameter (weight) plays a critical role in defining the hardship of the problem by altering the topology of the objective function. Indeed, when the regularization weight is small, the objective function has many equally good local minima: the desired local minimum hides among its peers. When the regularization weight is large, the objective function contains fewer local minima, but each minimum is like an abyss: once the solver loses its way into one abyss, it has no chance to ever escape. Therefore, the regularization weight should be something in-between – highlight the desired local minimum while preserving the smoothness of the objective function (Figure 5).

5 Experiments

Three types of dynamics and three types of (simulated) data are investigated in this study. The three types of dynamics are constant SIRQ with no regularization, time-varying SIR with regularization, and time-varying SIRQ with regularization. All dynamics use the ODE versions specified in Section 2 and are implemented with the Euler-Maruyama scheme.

The three types of data are virulence data, surveillance data, and serological data:

Virulence. Sample the population and test for the pathogen (*e.g.*, virus). This data is used for the estimation of the current number of `infectious` people

Surveillance. The confirmed and thus quarantined cases of the disease. This data is used for the estimation of the number of `removed` (in SIR) or `quarantined` (in SIRQ) people.

Serological. Sample the population and test for the antibody. This data is used for the estimation of the `removed` people in SIRQ provided that the infectious disease in question is a novel one. Otherwise, this data should be used not only for the `removed` but also for the `immune` compartment in an MSIR model. To reduce the sampling cost, this data can be collected along with the virulence data.

Experiment results are promising and serve well as a proof of concept. Limited by the budget, I conducted the experiments on simulated data only. To further test the model and the algorithm on real datasets as well as to help the whole world fight against the COVID-19 pandemic, I release my research platform as an open-source package[9].

5.1 Constant SIRQ

The SIRQ model has three parameters $\theta = (\beta, \gamma, \delta)$. The simulation scenario is set in a small town with 1000 inhabitants. It starts with 10 `infectious` people and 0 `removed` or `quarantined` people. Then the epidemic develops with a dynamic $(\beta, \gamma, \delta) = (0.3, 0.03, 0.07)$. Two types of data evidences are available for the inference of the epidemic. The virulence data is 9 samples during the life of the epidemic, with a sample size of 10 people each. The surveillance data is the number of confirmed cases at 8 different moments. The maximum likelihood estimate yields the value $(\hat{\beta}, \hat{\gamma}, \hat{\delta}) = (0.307, 0.030, 0.073)$, very close to the ground truth.

This experiment does not include any time-varying factor. Rather, it demonstrates the power of this simple model. With virulence data and surveillance data only, we are able to precisely estimate the basic reproduction rate and the quarantine ratio δ/γ. This model is particularly useful in the COVID-19 pandemic, where many are asymptomatic patients.

5.2 Time-varying SIR with regularization

The model tested in this subsection has one time-varying parameter and one constant parameter $\theta_t = (\beta_t, \gamma)$. The transmission rate β_t is supposed to be time-varying to reflect the gatherings in holidays, the adoption of social distancing, and so forth. The removal rate γ is set to be constant because the aggressiveness of the virus and the resistance of the population are believed to be stable (unless the virus mutates). In the experiment, the ground truth β_t firstly increases because of holiday gatherings and then drops because of rising public awareness, whilst the ground truth γ is held stable with little variation (Figure 6).

The epidemic takes place in a city of 100k inhabitants. It starts with 100 `infectious` people and 0 `removed` people. The only data available to infer the epidemic is the 9-sample virulence data during the life of the epidemic, with each sample containing 1k people. The lack of the number of confirmed cases suggests that this is a foreign country trying to evaluate an epidemic struck country with information censorship; the virulence data corresponds to the exported cases.

The data is considered *sparse* in contrast to the 100-dimensional parameter, which justifies the necessity of regularization. I applied total variation regularization and solved it with iterative Nelder-

[9] Temporally hosted on https://github.com/WenjieZ/2019-nCoV. Documentations are under development.

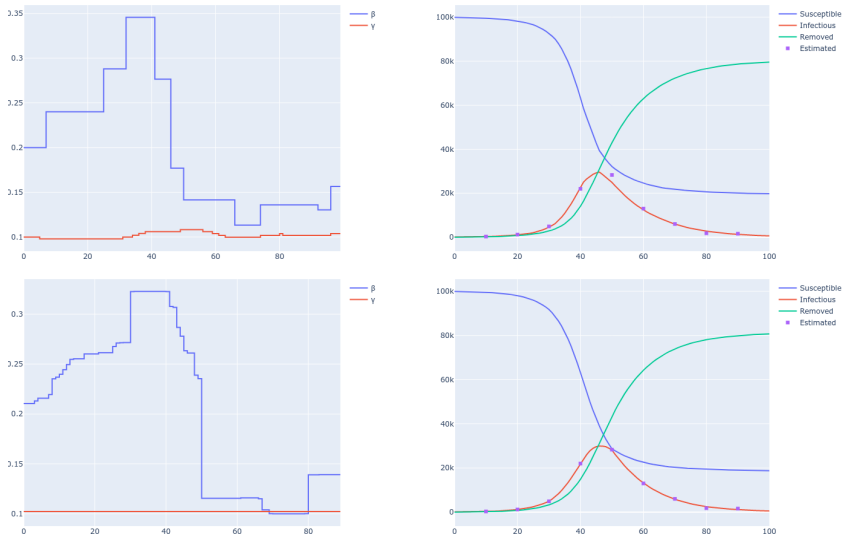

Figure 6: Time-varying SIR. Left: the dynamic. Right: the resulted epidemic and the evidence (dots). Top: ground truth. Bottom: estimation. β denotes the transmission rate; γ denotes the removal rate.

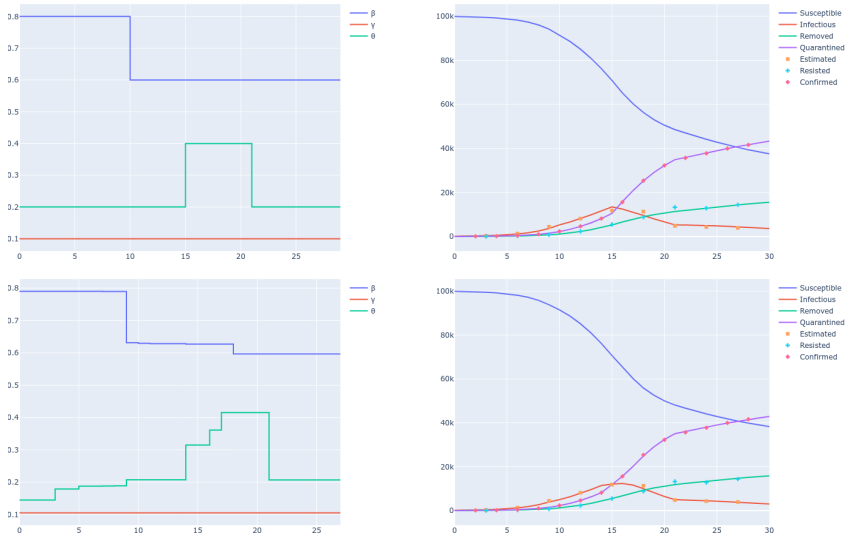

Figure 7: Time-varying SIRQ. Left: the dynamic. Right: the resulted epidemic and the evidences (dots). Top: ground truth. Bottom: estimation. β denotes the transmission rate; γ denotes the removal rate; θ should be replaced with δ, which denotes the quarantine rate.

Mead. The estimated dynamic qualitatively recovers the ground truth. In addition, the estimated dynamic can perfectly generate evidence identical to the sample collected (Figure 6).

5.3 Time-varying SIRQ with regularization

The model tested in this subsection has *two* time-varying parameters and one constant parameter $\theta_t = (\beta_t, \gamma, \delta_t)$. The transmission rate β_t is supposed to be time-varying to reflect the gatherings in holidays, the adoption of social distancing, and so forth. The removal rate γ is set to be constant because the aggressiveness of the virus and the resistance of the population are believed to be stable (unless the virus mutates). The quarantine rate δ_t is supposed to be time-varying to reflect the opening of field hospitals. In the experiment, the ground truth β_t is mostly constant with one sudden drop because of the city lockdown, whilst the ground truth δ_t rises suddenly thanks to the opening of a field hospital and then drops to the previous level because the hospital is full (Figure 7).

The epidemic takes place in a city of 100k inhabitants. It starts with 100 `infectious` people and 0 `removed` or `quarantined` people. The data available here is much richer than previous. It includes a 9-sample virulence data with each containing 1k people, the confirmed cases, and a 9-sample serological data with each also containing 1k people. These information are not too difficult to obtain for countries with excellent governance.

One challenge to infer this model is the coexistence of *two* time-varying parameters. Still, total variation regularization and iterative Nelder-Mead succeeded in qualitatively reproducing the ground truth, and the estimated dynamic generates evidences perfectly fit to the reality.

6 Conclusions and outlook

The combination of total variation regularization and iterative Nelder-Mead successfully detects the discontinuities of the underlying time-varying dynamics. When the epidemic follows an SIR dynamic with time-varying transmission rate β_t, the proposed combination qualitatively recovers the dynamic with the help of sparse virulence data. When the epidemic follows an SIRQ dynamic with time-varing transmission rate β_t and time-varying quarantine rate δ_t, the proposed combination qualitatively recovers the dynamic with the help of virulence, surveillance, and serological data.

There are three directions to improve this study. Firstly, it is unclear whether the local optimum achieved by iterative Nelder-Mead is the global one. Since the optimization algorithm plays a critical role in the solution finding process, one part of research effort should be directed to improve the searching ability of the optimization algorithm. Secondly, it is currently unclear whether these dynamics could be *quantitatively* perfectly recovered with more data or under better conditions. One research direction thus consists of precisely inferring the underlying dynamics so that public policies (*e.g.*, city lockdown) could be evaluated. It also helps compare the efficiency of different policies in the same country and the efficiency of the same policy in different countries. Thirdly and most importantly, it is urgent to leverage the machinery provided by this study against COVID-19.

References

Althaus, C. L. (2014), 'Estimating the reproduction number of ebola virus (ebov) during the 2014 outbreak in west africa', *PLoS Currents* (September), 1–9.

Birrell, P. J., De Angelis, D. and Presanis, A. M. (2018), 'Evidence synthesis for stochastic epidemic models', *Statistical Science* **33**(1), 34–43.

Camacho, A., Kucharski, A., Funk, S., Breman, J., Piot, P. and Edmunds, W. (2014), 'Potential for large outbreaks of ebola virus disease', *Epidemics* **9**, 70–78.

Cazelles, B., Champagne, C., Dureau, J. and Koelle, K. (2018), 'Accounting for non-stationarity in epidemiology by embedding time-varying parameters in stochastic models', *PLoS Computational Biology* **14**(8).

Chowell, G., Hengartner, N., Castillo-Chavez, C., Fenimore, P. and Hyman, J. (2004), 'The basic reproductive number of ebola and the effects of public health measures: the cases of congo and uganda', *Journal of Theoretical Biology* **229**(1), 119–126.

Dureau, J., Ballesteros, S. and Bogich, T. (2013), Ssm: Inference for time series analysis with state space models.

Dureau, J., Kalogeropoulos, K. and Baguelin, M. (2013), 'Capturing the time-varying drivers of an epidemic using stochastic dynamical systems', *Biostatistics* **14**(3), 541–555.

Funk, S., Camacho, A., Kucharski, A. J., Eggo, R. M. and Edmunds, W. J. (2016), 'Real-time forecasting of infectious disease dynamics with a stochastic semi-mechanistic model', *Epidemics* .

Funk, S., Ciglenecki, I., Tiffany, A., Gignoux, E., Camacho, A., Eggo, R. M., Kucharski, A. J., Edmunds, W. J., Bolongei, J., Azuma, P., Clement, P., Alpha, T. S., Sterk, E., Telfer, B., Engel, G., Parker, L. A., Suzuki, M., Heijenberg, N. and Reeder, B. (2017), 'The impact of control strategies and behavioural changes on the elimination of ebola from lofa county, liberia', *Philosophical Transactions Biological Sciences* **372**(1721).

He, D., Dushoff, J., Day, T., Ma, J. and Earn, D. J. D. (2011), 'Mechanistic modelling of the three waves of the 1918 influenza pandemic', *Theoretical Ecology* **4**(2), 283–288.

Imai, N., Dorigatti, I., Cori, A., Donnelly, C., Riley, S. and Ferguson, N. M. (2020), 'Estimating the potential total number of novel coronavirus cases in wuhan city, china'. [Online; accessed 30. Mar. 2020].
URL: *http://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/news–wuhan-coronavirus*

Kermack, W. O. and McKendrick, A. G. (1927), 'A contribution to the mathematical theory of epidemics', *Proceedings Mathematical Physical and Engineering Sciences* **115**(772), 700–721.

Kucharski, A. J., Russell, T. W., Diamond, C., Liu, Y., Edmunds, J., Funk, S., Eggo, R. M., Sun, F., Jit, M., Munday, J. D., Davies, N., Gimma, A., van Zandvoort, K., Gibbs, H., Hellewell, J., Jarvis, C. I., Clifford, S., Quilty, B. J., Bosse, N. I., Abbott, S., Klepac, P. and Flasche, S. (2020), 'Early dynamics of transmission and control of covid-19: a mathematical modelling study', *The Lancet Infectious Diseases* .

Kurtz, T. G. (1987), Approximation of population processes.

Rasmussen, D. A., Ratmann, O., Koelle, K. and Lemey, P. (2011), 'Inference for nonlinear epidemiological models using genealogies and time series', *PLoS Computational Biology* **7**(8).

Siettos, C. I. and Russo, L. (2013), 'Mathematical modeling of infectious disease dynamics', *Virulence* **4**(4), 295–306.

Smirnova, A., deCamp, L. and Chowell, G. (2017), 'Forecasting epidemics through nonparametric estimation of time-dependent transmission rates using the seir model', *Bulletin of Mathematical Biology* .

Stocks, T., Britton, T. and Höhle, M. (2018), 'Model selection and parameter estimation for dynamic epidemic models via iterated filtering: application to rotavirus in germany', *Biostatistics* .

Wu, J. T., Leung, K. and Leung, G. M. (2020), 'Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, china: a modelling study', *Lancet* .

Xu, X., Kypraios, T. and O'Neill, P. D. (2016), 'Bayesian non-parametric inference for stochastic epidemic models using gaussian processes', *Biostatistics* **17**(4), 619–633.