Optimal incentives to mitigate epidemics: A Stackelberg mean field game approach

Alexander Aurell René Carmona Gökçe Dayanıklı Mathieu Laurière
{aaurell, rcarmona, gokced, lauriere}@princeton.edu
Dept. of Operations Research and Financial Engineering, Princeton University, Princeton, NJ 08544

Keywords: SIR epidemics, Mean field game, Stackelberg equilibrium, Machine learning

This extended abstract summarizes the work [1].

Introduction & Impact Statement. Throughout the year 2020 governments implemented non-pharmaceutical disease mitigation strategies to decrease the spread of Covid-19 in societies while waiting for vaccination to begin. We have seen local measures implemented that vary over time with the threat posed by the disease, are essentially feedback strategies on the level of alertness. The alertness level is assessed by a set of indicators whose exact form vary from society to society. These indicators reflect the current overall health of the society and the trends of the disease spread. Examples include the multi-tier systems implemented in Alaska [2], Ohio [3], New York [4], and other states.

Implementing policies backed by science should reduce the disease spread. However, this conclusion relies on the public's willingness to comply. Inspired by the challenging situation faced by local governments we ask whether regulations can be implemented in an optimized way while simultaneously taking the reaction, possible non-compliance, and selfishness of the population into account. To this end, we study in [1] a *Principal-Agent model* for the interplay between a policy maker and a society. In the model, the society reacts to policies in a non-optimal way; each individual seeks her best response while anticipating the action of the others. Assuming all individuals act in this way, the society will settle in a *Nash equilibrium*. Moreover, the individual might reject the policy altogether (*non-compliance*) if it is too strict for her (e.g., very harsh lockdown). The policy maker anticipates how the society reacts and with this knowledge designs a policy that is optimal with respect to the policy maker's risk metric. For the full picture, with the compartmental models of epidemics as a basis and on top of that the regulator's optimization procedure, we develop a numerical approach with machine learning to the computation of optimal policies.

The population's response to policies and regulations. During the ongoing pandemic, it has become clear that one of the main challenges for policy makers is to anticipate the population's reaction to various policies. The population does not behave like a simple dynamical system because it is composed of many individuals taking decisions (e.g., to go or not to go out). Our first contribution is to propose a framework to study how the population reacts to any given policy using ideas from game theory. More precisely, we use the notion of Nash equilibrium [5], in which each individual's behavior is optimal in the sense that no one can gain from a unilateral change of action. This represents a situation in which each agent selfishly optimizes their own utility function while assuming that all other agents also do so. Since it is very challenging to find a Nash equilibrium where there is a large finite number of agents, we consider an infinite population and hence phrase the Nash equilibrium using the paradigm of mean field games (MFG for short) [6, 7, 8]. For a given policy of the regulator, the goal is to find an optimal behavior $\hat{\alpha}$ and a flow of population distribution $\hat{\rho}$ such that: (1) the optimal behavior of an infinitesimal agent facing the population evolution $\hat{\rho}$ is given by $\hat{\alpha}$, and (2) if all the agents use $\hat{\alpha}$ as a control, the induced evolution for the population is $\hat{\rho}$. We show that this fixed point problem has a solution in our model, which is the equilibrium behavior of the population. Computing $\hat{\alpha}$ and $\hat{\rho}$ is a challenging task due to the coupling between the individual optimization and the macroscopic evolution. We develop a numerical approach based on a characterization of the Nash equilibrium through a forward-backward system of stochastic differential equations. In this system, the dimension of the variable is equal to the number of states in the epidemic model, and if one uses traditional numerical methods, the computational cost increases dramatically with this number. For this reason, the numerical method we propose is based on neural network approximation, which works well even in high dimension, hence allowing us to handle more realistic and complex models with a large number of states. We train the neural networks using Monte Carlo simulations. Then we test the numerical approach by successfully replicating the ODE solution of a benchmark linear quadratic test case based on the SIR model, where the agents can be in the three states **S**usceptible, **I**nfected and **R**ecovered. In Fig. 1, the results of the late lockdown policy with the ODE solver and our numerical approach can be seen. Other examples that include early lockdown and no lockdown policies are presented in [1].

Regulator's policy optimization. Building on our understanding of the population's reaction to any given policy, we propose a framework to study the *optimal* policy to minimize the regulator's objective function. From a government's viewpoint, the goal is to find a policy which will minimize a kind of social cost, while taking into account the fact that the population may not follow perfectly the recommendations. In other words, the government tries to steer the equilibrium adopted by the population towards a socially optimal situation. In our model, the regulator aims to minimize the proportion of infected people, tries to set a policy closer to the recommended by healthcare professionals and takes the risk of the policy being rejected by the society into account. When we incorporate the regulator's optimization on top of the Nash equilibrium, the problem becomes even more challenging, both theoretically and numerically. To solve this problem, we borrow ideas from contract theory [9, 10, 11] and we view the problem as a contract between the regulator and the population. Under suitable assumptions, we prove existence and uniqueness of an optimal policy. We then extend the numerical methods developed for the population equilibrium to tackle the regulator's policy optimization with a more complex setting for the population. Considering the setting of Fig. 2, we show in Fig. 3 the advantage obtained by optimizing the regulator's policy. Other examples can be found in [1].

Conclusion and related work. To understand the interplay between action and response, which is in general highly non-linear, is a difficult task. The new tools has the potential to inform policy makers in the control of the spread of epidemics, and the localized re-opening of an economy after shut-down. Our work [1] is contributing to the growing body of literature on MFGs in epidemics modeling. The field is growing and we mention here only a few references: vaccination [12], testing [13], and social distancing protocols [14]. A realistic compartmental model for epidemic spread needs to be very high dimensional, stratified over age groups, locations, risk groups, etc. While a differential equation-based solution to the Principal-Agent problem can be available, it is likely too expensive to compute. The machine learning algorithm we present in [1] offers scalability. Furthermore, it is not bound to the full Principal-Agent problem but can also be used to study scenarios, i.e., evaluating the effect of various levels of lockdown measures.

Figures

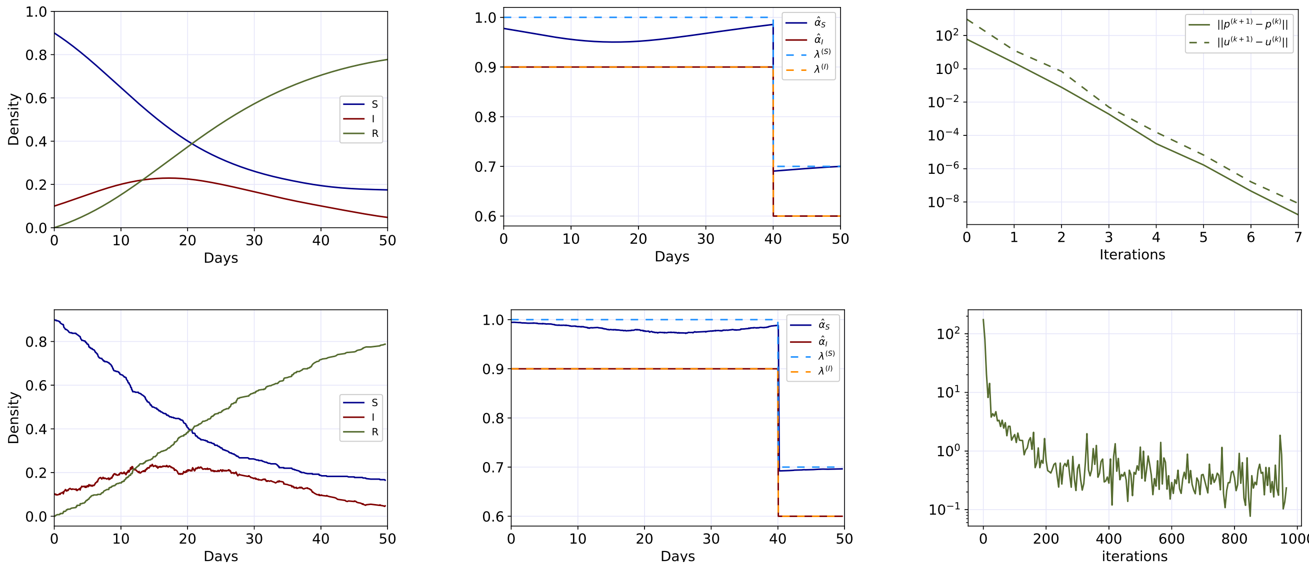

Figure 1: **SIR scenario.** The agents in the MFG aim to minimize their socialization and treatment costs through selection of their best contact factor, a variable determining their level of socialization. A susceptible agent is infected at a rate depending on her own contact factor, the mean of the contact factor of infected agents, and the proportion of population infected (the form of this rate is given by the same formula as the rate from S to E in Fig. 2). Infected agents recover at constant rate and the recovered state is absorbing. *Top row:* Semi-explicit solution output by an ODE solver. *Bottom row:* Output by our machine learning algorithm. In each row we see the evolution of the population state distribution (*left*), the evolution of the controls/actions (where λ is the contact factor recommended by the government and $\hat{\alpha}$ is the contact factor implemented by the population) (*middle*), and the convergence of the algorithm (*right*). The machine learning algorithm is replicating the true solution.

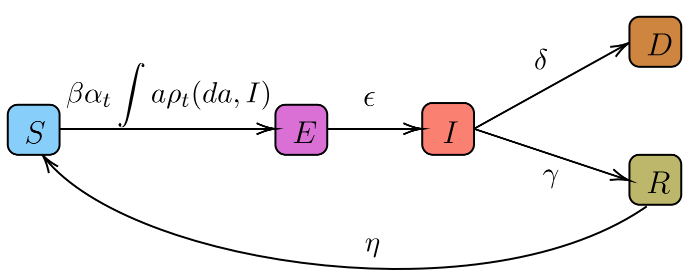

Figure 2: **SEIRD scenario model diagram.** The 5 states in the population are **S**, **E**xposed, **I**, **R**, **D**ead. Exposed agents are assumed not to be infectious yet. The agents in the MFG aim to minimize their socialization and possible treatment costs by choosing a contact factor $(\alpha_t)_{t\in[0,T]}$. The rate at which a susceptible agent becomes exposed at time t depends on her own contact factor, α_t, and a measure of the infected agents' mean contact factor and the proportion of the population currently infected, $\int a\rho_t(da, I)$. Exposed agents become infected, infected agents either die or recover and recovered agents can become susceptible again, with constant transition rates $\varepsilon, \delta, \gamma, \eta > 0$.

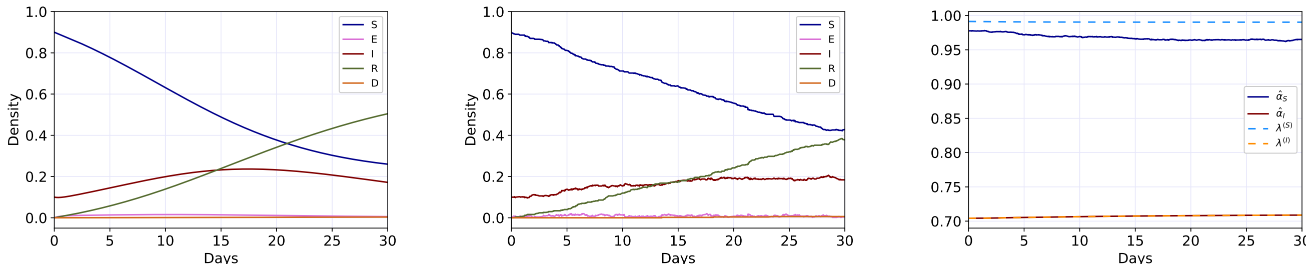

Figure 3: **SEIRD scenario numerical results.** Under policy optimization, the regulator recommends the infected part of the population to restrict their socialization. The infected agents follow these regulations. Furthermore, the susceptible part of the population decides to decrease their contact factor in order to protect themselves, even though the regulator decides not to put any restrictions on them. When we compare the proportion of the population with the *free spread* scenario where the regulator does not impose any restrictions and the population does not optimize their individual behavior, we can see that less people get sick during the time frame. *Left:* Evolution of the state distribution during free spread. *Middle:* Same, but during controlled spread. *Right:* The contact factors during cotrolled spread.

References

[1] Alexander Aurell, René Carmona, Gökçe Dayanıklı, and Mathieu Laurière. Optimal incentives to mitigate epidemics: a Stackelberg mean field game approach. *arXiv preprint arXiv:2011.03105*, 2020.

[2] Alaska's COVID-19 Alert Levels. http://dhss.alaska.gov/dph/Epi/id/Pages/COVID-19/alertlevels.aspx. (Accessed on 01/25/2021).

[3] Ohio Alert Indicators. https://coronavirus.ohio.gov/static/OPHASM/Summary-Alert-Indicators.pdf. (Accessed on 01/25/2021).

[4] Cluster Action Initiative. https://forward.ny.gov/cluster-action-initiative. (Accessed on 01/25/2021).

[5] John Nash. *Non-cooperative games*. PhD thesis, Princeton University, 1950.

[6] Jean-Michel Lasry and Pierre-Louis Lions. Jeux à champ moyen. I–le cas stationnaire. *Comptes Rendus Mathématique*, 343(9):619–625, 2006.

[7] Jean-Michel Lasry and Pierre-Louis Lions. Jeux à champ moyen. II–Horizon fini et contrôle optimal. *Comptes Rendus Mathématique*, 343(10):679–684, 2006.

[8] Minyi Huang, Roland P Malhamé, Peter E Caines, et al. Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle. *Communications in Information & Systems*, 6(3):221–252, 2006.

[9] Bengt Holmstrom and Paul Milgrom. Aggregation and linearity in the provision of intertemporal incentives. *Econometrica: Journal of the Econometric Society*, pages 303–328, 1987.

[10] Yuliy Sannikov. *Contracts: The Theory of Dynamic Principal–Agent Relationships and the Continuous-Time Approach*, volume 1 of *Econometric Society Monographs*, page 89–124. Cambridge University Press, 2013.

[11] Jakša Cvitanić, Dylan Possamaï, and Nizar Touzi. Dynamic programming approach to principal–agent problems. *Finance and Stochastics*, 22(1):1–37, 2018.

[12] Laetitia Laguzet and Gabriel Turinici. Individual vaccination as Nash equilibrium in a SIR model with application to the 2009–2010 influenza A (H1N1) epidemic in France. *Bulletin of Mathematical Biology*, 77(10):1955–1984, 2015.

[13] Emma Hubert, Thibaut Mastrolia, Dylan Possamaï, and Xavier Warin. Incentives, lockdown, and testing: from Thucydides's analysis to the COVID-19 pandemic. preprint, 2020.

[14] Romuald Elie, Emma Hubert, and Gabriel Turinici. Contact rate epidemic control of COVID-19: an equilibrium view. *Mathematical Modelling of Natural Phenomena*, 15:35, 2020.