

Optimal Fixed Lockdown for Pandemic Control

Qianqian Ma, Yang-Yu Liu, and Alex Olshevsky

Abstract—As a common strategy of contagious disease containment, lockdowns inevitably have economic cost. The ongoing COVID-19 pandemic underscores the trade-off arising from public health and economic cost. An optimal lockdown policy to resolve this trade-off is desired. Here we propose a mathematical framework of pandemic control through an optimal fixed stabilizing non-uniform lockdown, where our goal is to reduce the economic activity as little as possible while decreasing the number of infected individuals at a prescribed rate. This framework allows us to efficiently compute the optimal stabilizing lockdown policy for general epidemic spread models, including both the classical SIS/SIR/SEIR models and a model of COVID-19 transmissions. We demonstrate the power of this framework by analyzing publicly available data of inter-county travel frequencies to analyze a model of COVID-19 spread in the 62 counties of New York State. We find that an optimal stabilizing lockdown based on epidemic status in April 2020 would have reduced economic activity more stringently outside of New York City compared to within it, even though the epidemic was much more prevalent in New York City at that point. This finding holds for a variety of epidemic models and parameters from the literature, and is robust to errors in travel rates, different cost functions, and potential urban-rural transmission spread differentials.

INTRODUCTION

The COVID-19 pandemic has resulted in more than 395M confirmed cases and 5.7M deaths (up to Feb. 7th, 2022) [1] and has impacted the lives of more than 90% global population [2], [3]. Curbing the spread of a pandemic like COVID-19 continues to depend on the successful implementation of non-pharmaceutical interventions such as lockdowns, social distancing, shelter in place orders, contact tracing, isolation, and quarantine [4]–[7]. However, these interventions can also lead to substantial economic damage, motivating us to investigate the problem of curbing pandemic spread while minimizing the induced economic losses.

We consider the problem of designing an optimal fixed stabilizing lockdown that minimizes the economic damage while reducing the number of new infections to zero at a prescribed rate. Such a lockdown should be *non-uniform*, because shutting down different locations has different implications both for the economic cost and for pandemic spread. The difficulty is whereas a uniform lockdown can be found through

a search over a single parameter, a non-uniform lockdown is parametrized by many parameters associated with different locations. Despite of its significance and implications, a computationally fast framework to design stabilizing lockdown strategies is still lacking.

Here we propose such a framework by mapping the design of optimal lockdown policy to a classical problem in control theory — design an intervention that affects the eigenvalues of a matrix governing the dynamics of a dynamical system. It turns out that, even though general epidemic spreading dynamics are nonlinear, an eigenvalue bound for a linear approximation of the spreading dynamics at the initial condition nevertheless forces the number of infections at each location to go to zero asymptotically at a prescribed rate for all time. As we show, attaining such an eigenvalue bound at the initial condition of the epidemic is equivalent to enforcing an upper bound on the basic reproduction number over the entire lifetime of the epidemic. We provide two highly efficient algorithms that design the lockdown to achieve such an eigenvalue bound.

We apply these algorithms to design a stabilizing lockdown on both synthetic and real data (using data from SafeGraph [8] to fit a county-level model of New York State) for epidemic spread models of COVID-19 using disease parameters from the literature [9]–[11]. The lockdown we obtain using our methods is optimal in the sense that it has minimal cost among all possible lockdowns which achieve the same upper on the reproduction number over the lifetime of the epidemic. Unsurprisingly, we find that the heterogeneous lockdown is far more economical than a homogeneous lockdown. However, we find additional features of the optimal stabilizing lockdown that are counter-intuitive. For example, we find that in models of random graphs, degree centrality and population do not affect the strength of the lockdown of a location unless its population (or degree centrality) takes extremely smaller (or larger) values than others. Somewhat surprisingly, we show that an optimal stabilizing lockdown based on the epidemic status in April 2020 would have reduced activity more strongly outside of New York City (NYC) compared to within it, even though the epidemic was much more prevalent in NYC at that point. This finding passes a variety of robustness checks and holds for different combinations of epidemic transmission models (SIS/SIR/COVID-19), parameters governing epidemic spread, cost functions associated with lockdowns, as well as potential differences between how fast the disease spreads in urban vs rural locations.

This research was supported by NSF award ECCS-1740451.

Qianqian Ma is with the Department of Electrical and Computer Engineering, Boston University, Boston, MA, 02215, USA (e-mail: maqq@bu.edu). Yang-Yu Liu is with the Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, 02115, USA (e-mail: yyl@channing.harvard.edu). Alex Olshevsky is with the Department of Electrical and Computer Engineering and Division of Systems Engineering, Boston University, Boston, MA, 02215, USA (e-mail: alexols@bu.edu).

A. Related work

Our work is related to a number of recent papers motivated by the spread COVID-19, as well as some older work. Indeed, spatial spread of epidemic admits a natural network representation, where nodes represent different locations and edges encode traveling of residents between the locations. Such spatial epidemic network model has received attention in the studies of COVID-19 recently [4], [12]–[15]. We begin by discussing several papers most closely related to our work.

Our paper builds on the results of [11], which proposed a spatial epidemic transmission model and consider the effect of lockdowns; we use the same model of lockdown of [11] in this work. The major difference between this work in [11] is two-fold. First, we do not consider asymptotic stability in a model with births and deaths as our focus is on a shorter scale. Second, we propose new algorithms with improved running times; in particular, our main contribution is a linear time method that is applicable to the vast majority of cases we have considered. Similarly, the main difference of this work relative to [16] and the references therein are new algorithms (though the lockdown models differ somewhat), as well as the new observations on counter-intuitive phenomena satisfied by the optimal lockdown. Our work has some similarities with literature [17], which divided the population into 18 compartments and found that population heterogeneity could significantly impact disease-induced herd immunity.

Our Approach vs Traditional Optimal Control: An alternative approach would be to approach the lockdown problem using the techniques of optimal control. This approach is explored by a number of papers [18]–[24].

Our work has several differences with the traditional optimal control approach. The first is that we are looking for a fixed lockdown, whereas an optimal control based approach would offer a lockdown which varies for all time t . Fixed vs variable lockdowns offer a number of advantages and disadvantages, but one advantage of the former is that policymakers might be overly eager to relax lockdowns, resulting in poor decision-making; one could argue this is what happened in the United States in the summer of 2020 [25]. Thus we focus our investigation on a single-fixed lockdown maintained until the extinction of the epidemic.

Next, scalability is central to our results: our main result is an algorithm which, under certain conditions, runs in nearly linear time. This is not the case for the optimal control approach. For example, Khanafer and Basar [18] wrote down the optimal control formulation for the SIS case and remarked that solving the resulting equations “is intractable.”

As Khanafer and Basar pointed out, there are no methods that are guaranteed to find the optimal control efficiently. Nevertheless, in a number of recent works promising numerical results are obtained. For a direct formulation of the problem, one can turn to [20], which is also in the same spirit as our work, in that it studies control of COVID-19 using, among other things, movement restrictions. This results in a mixed-integer non-linear programming problem. Solving such problems is generally intractable, so [20] used a genetic algorithm as a heuristic.

Another possibility might be to use indirect methods (i.e., relying on the maximum principle) to solve the optimal control problem. The same complexity considerations, however, come up in this context again. Most of the literature on the optimal control of epidemics over networks seems to use variants of the forward-backward sweep method [22], [26]–[28] to solve the equations arising from the maximum principle, but it is known that this method can diverge even for simple examples [29]. Nevertheless, on many examples the method converges in reasonable time: for example, [26] reports excellent results on random networks of large size. Recent papers studying control of COVID-19 using numerical optimal control methods are [21]–[24]. These papers assumed a cost of a human life lost (sometimes set based on the average lifetime earnings) and considered the discounted total cost over the entire epidemic (alternatively, [21] considered the frontier of possible strategies over all possible ways to value life). However, the scalability of these approaches is unclear, due to their reliance on numerical methods without a clear convergence theory.

We remark that a fixed lockdown also has some disadvantages. For example, it does not allow us to modify the lockdown strategy in place as new information comes in about the underlying pandemic. Given that in this paper we will develop methods which can work in linear time, it may be appealing to handle time-varying lockdowns by using our method within an MPC framework as a way to bypass any computational inefficiency that may arise with an optimal control approach.

I. MODELS OF SPATIAL DISEASE SPREAD

All the epidemic spread models considered in this work are compartmental or network models [11], [30] with “locations” corresponding to neighborhoods, counties, or other geographic subdivisions. We consider n locations, with the variable x_i denoting the proportion of infected population at location i . Our framework can be applied to general epidemic spread models. For demonstration purpose, here we consider a simple model of COVID-19 which contains the classical Susceptible-Infectious-Recovered (SIR) model and the Susceptible-Exposed-Infectious-Recovered (SEIR) model as special cases. Our results also hold for the SIS model (see Appendix B for details).

We consider a simple model (similar to models in literature [9], [11], [18], [20], [31], [32]) of COVID-19 spreading that breaks infected individuals into two types: asymptomatic and symptomatic. This model allows individuals transmit the infection at different rates:

$$\begin{aligned} \dot{s}_i &= -s_i \sum_{j=1}^n a_{ij}(\beta^a x_j^a + \beta^s x_j^s) \\ \dot{x}_i^a &= s_i \sum_{j=1}^n a_{ij}(\beta^a x_j^a + \beta^s x_j^s) - (\epsilon + r^a) x_i^a \\ \dot{x}_i^s &= \epsilon x_i^a - r^s x_i^s \end{aligned} \quad (1)$$

Here s_i (x_i^a or x_i^s) stands for the proportion of susceptible (asymptomatic or symptomatic infected, respectively) population at location i , a_{ij} captures the rate at which infection flows

from location j to location i , β^a (or β^s) is the transmission rate of asymptomatic (or symptomatic) infected individuals, r^a (or r^s) is the recovery rate of asymptomatic (or symptomatic) infected individuals. We assume infected individuals are asymptomatic at first and ϵ is the rate at which they develop symptoms.

Note that our model of COVID-19 spreading can be considered as a generalization of the classical SIR model and the SEIR model of epidemic spread. Indeed, by setting $\beta^s = \epsilon = r^s = 0$, we recover the SIR model; and by setting $\beta^a = r^a = 0$, we recover the SEIR model. However, neither the SIR nor the SEIR model captures the existence of two classes of individuals who transmit infections at different rates as above. Our model can also be considered as a simplification of existing models in studying COVID-19 spreading [9], [11]. For example, in [11], asymptotic stability was considered in a slightly more general model including both births and deaths. Here for simplicity in our model we consider a fixed population size. In [9] eight classes of patients (instead of two) were introduced, depending on whether the infection is diagnosed, whether the patient is hospitalized, as well as other factors.

In matrix form, we can write our model as

$$\begin{pmatrix} \dot{s} \\ \dot{x}^a \\ \dot{x}^s \end{pmatrix} = \begin{pmatrix} 0 & -\beta^a \text{diag}(s)A & -\beta^s \text{diag}(s)A \\ 0 & \boxed{\beta^a \text{diag}(s)A - (\epsilon + r^a)} & \beta^s \text{diag}(s)A \\ 0 & \epsilon & -r^s \end{pmatrix} \begin{pmatrix} s \\ x^a \\ x^s \end{pmatrix}, \quad (2)$$

where scalars in the matrix should be understood as multiplying the identity matrix, and the matrix A stacks up the coefficients $[a_{ij}]$. Here s, x^a, x^s are now vectors. Let us write $M(t)$ for the bottom right $2n \times 2n$ submatrix (outlined by a box) in Eq. (2).

If we want the number of infections to decay to zero starting at time t_0 , a natural objective is to achieve a bound on the basic reproduction number, i.e., to achieve $R_0(t) \leq r$ for some $r < 1$ and all $t \geq t_0$, where $R_0(t)$ is the basic reproduction number¹. It turns out this is equivalent to the linear eigenvalue condition $\lambda(M(t_0)) \leq -\alpha$ for an appropriately chosen positive α (see Appendix A for a proof). Moreover, this eigenvalue condition ensures that the number of infections at each location goes to zero at an asymptotic rate of α , i.e., bounded asymptotically by a multiple of $e^{-\alpha t}$. These facts are standard, but for completeness we include a proof of them in the arxiv in [34].

In the remainder of this paper, we will attempt to design strategies that enforce decay of infections with a prescribed rate by modifying the matrix A through lockdowns to satisfy such an eigenvalue bound.

A. Construction of The Spread Matrix A

Methods of constructing A capturing spatial heterogeneity have been well studied [35]–[42]. Here we follow a recent

¹For a non-compartmental model, the basic reproduction number is defined as the average number of secondary infections produced by a typical case of an infection in a population where everyone is susceptible. For a compartmental model, the definition of basic reproduction number was given in [33].

work [11] that is particularly well-suited to model the lockdowns in curbing COVID-19. Denote the fixed population size at location i as N_i , and assume people travel from location i to location j at rate τ_{ij} .

We next describe how the matrix A is constructed. Observe that the susceptible individuals at location i can be infected not only at location i but also in other locations. Indeed, the rate of travel of susceptible population from location i to location l is $(1 - x_i)\tau_{il}$. Further, the rate of infection at location i is proportional to the fraction of infected people in the total population of location i . In terms of these variables, we have

$$\dot{s}_i = - \sum_{l=1}^n s_l \tau_{il} \left(\frac{\sum_{j=1}^n N_j \tau_{jl} x_j^a}{\sum_{k=1}^n N_k \tau_{kl}} \beta^a + \frac{\sum_{j=1}^n N_j \tau_{jl} x_j^s}{\sum_{k=1}^n N_k \tau_{kl}} \beta^s \right), \quad (3)$$

where the constants β^a, β^s capture the number of interactions a typical person will have upon traveling to a location, as well as the probability that an interaction will lead to an infection; since the latter may be different from asymptomatic and symptomatic individuals, the two numbers β^a, β^s may not be the same.

It is now straightforward to relate the coefficients a_{ij} 's to the flow rates τ_{ij} and the populations: this can be done by simply ‘‘pattern matching’’ Eq. (3) to Eq. (23) to obtain

$$a_{ij} = \sum_{l=1}^n \tau_{il} \tau_{jl} \frac{N_j}{\sum_{k=1}^n N_k \tau_{kl}}. \quad (4)$$

Observe that Eq. (4) can also be written in matrix form as $A = CB^T$ with

$$C = \tau, \quad B^T = D_1 \tau^T D_2, \quad (5)$$

where $\tau = (\tau_{ij})$, $D_1 = \text{diag}(\sum_k N_k \tau_{kl})^{-1}$ while $D_2 = \text{diag}(N_1, \dots, N_n)$.

B. Lockdown model

When a lockdown is ordered heterogeneously across different locations, this has two consequences. First, the transmission rates will be altered. For instance, ensuring that all buildings have a maximum enforced density limits the rate at which people can interact, as do mandatory face-covering, and other measures, resulting in a number of transmissions that is a fraction of what they otherwise would have been. We may account for this as follows. Taking Eq. (4), let us multiply both sides by β^a to obtain:

$$\beta^a a_{ij} = \sum_{l=1}^n \beta^a \tau_{il} \tau_{jl} \frac{N_j}{\sum_{k=1}^n N_k \tau_{kl}}.$$

This is pre-lockdown. Now the effect of the lockdown is to replace β^a in each term within the sum by $\beta^a f_l$, for some location-dependent $f_l \in [0, 1]$. The effect on β^s is similar. Secondly, travel rates to location l are also a fraction of what they were before since there is now reduced inducement to travel, i.e., τ_{il} should be replaced by $\tau_{il} g_l$ with $g_l \in [0, 1]$ for each location l .

To avoid overloading the notation, we will not change the definitions of β^a, β^s or the travel rates τ_{il} but instead achieve

the same effect by changing the definition of a_{ij} as:

$$a_{ij} = \sum_{l=1}^n z_l \tau_{il} \tau_{jl} \frac{N_j}{\sum_{k=1}^n N_k \tau_{kl}}, \quad (6)$$

where $z_l = f_l g_l \in [0, 1]$. In matrix notation, the post-lockdown A matrix is

$$A_z = C \text{diag}(z) B^T. \quad (7)$$

The quantities $z_1, \dots, z_n$ can be thought of as measuring the intensity of the lockdown at each location.

It goes without saying that designing the detailed rules that will bring locations to desired z_i is beyond the scope of the current work. Thus, if the methods we propose below recommend, say, $z_i = 0.5$, it will be up to policymakers to come up with detailed rules reducing by 50% the quantity $f_i g_i$.

C. Lockdown cost

Clearly, setting $z_l = 1$ corresponds to doing nothing and should have a zero cost in terms of lost employment. On the other hand, choosing $z_l = 0$ corresponds to a complete lockdown and should be avoided. We will later apply our framework to real data collected from counties in New York State; shutting down a county entirely would result in people being unable to obtain basic necessities, and thus the economic cost should approach $+\infty$ as $z_l \rightarrow 0$. With these considerations in mind, a natural choice of lockdown cost is

$$c(z_1, \dots, z_n) = \sum_{l=1}^n c_l \left(\frac{1}{z_l} - 1 \right). \quad (8)$$

Here c_l captures the relative economic cost of closing down location l . Throughout this paper, we will choose c_l to be the employment at local l , but other choices are also possible (e.g., c_l could be the GDP generated at location l).

Besides the cost function in Eq. (8), we will also consider cost functions that blow up with different exponents as $\sum_{l=1}^n c_l (z_l^{-k} - 1)$, as well as costs which threshold as $\sum_{l=1}^n \min(c_l (z_l^{-1} - 1), C_l)$, i.e., which alter our cost function by saturating at some location-dependent cost C_l rather than blowing up as $z_l \rightarrow 0$.

One feature of some of these cost functions is that they inherently discourage extreme disparities among nodes. Indeed, a lockdown that places all the burden on a small collection of nodes by setting their z_i close to zero will have cost that blows up. By contrast, some previous works such as [11], [20] used cost functions $\sum_i c_i (1 - z_i)$ which do not have this feature.

D. Final problem formulation

The optimal fixed stabilizing lockdown problem simply puts together all the features we have outlined above: we are looking for a fixed lockdown of minimum cost enforcing a guaranteed decay rate on the epidemic through eigenvalue bounds on the matrix A which is obtained from the travel rates τ_{ij} as described in Section I-A. More formally, we are looking for a vector z such that the matrix $A_z = C \text{diag}(z) B^T$ from Eq. (5) satisfies an eigenvalue constraint $\lambda_{\max}(A_z) \leq \alpha$

(where α is negative) as well as the condition $z \in [0, 1]^n$ while minimizing a cost function from Section I-C.

We note that there may be alternate objectives that are plausible. For example, a guaranteed bound on the reproduction number is not the same as minimizing the total number of cases over the lifetime of the epidemic: indeed, it was argued by many in the context of COVID-19 that simply letting a disease spread unimpeded through the population could result in a smaller number of deaths. Our objective rules out such strategies by insisting on attaining a small reproduction number for the entire lifetime of the epidemic beginning at the initial condition.

II. ALGORITHMS FOR THE OPTIMAL FIXED STABILIZING LOCKDOWN

We next state our main theoretical contribution. This requires some background on the so-called matrix balancing problem, briefly described in the next subsection.

A. The matrix balancing problem

Given a nonnegative matrix $P \in \mathbb{R}^{n \times n}$, we say it is balanced if it has identical row and column sums (i.e., the sum of the i 'th row is the same as the sum of the i 'th column for all $i = 1, \dots, n$). The matrix balancing problem is, given a nonnegative P , to find a nonnegative diagonal matrix D such that DPD^{-1} is balanced.

The problem of matrix balancing is quite old; for example, an asymmetric version of this problem was introduced in the classic work of Sinkhorn and Knopp in the 1960s [43]. It is impossible to survey all the literature on matrix balancing and related problems, though we refer the reader to the survey [44]. Recently, a powerful algorithm for matrix balancing was given in [45]. It was shown in that work that this problem can be solved in linear time, understood as follows: solving the problem to accuracy ϵ requires only $\tilde{O}(\|P\|_0 \log \kappa \log \epsilon^{-1})$ where $\|P\|_0$ is the number of nonzero entries in the matrix A , $\kappa = D_{\max}^*/D_{\min}^*$ is the imbalance of the optimal solution, and the $\tilde{O}$ hides logarithmic factors. Thus matrix balancing problems can be solved in nearly the same time as it takes to simply read the data, provided κ is bounded away from zero. In the event that we do not have an a-prior bound on κ , [45] gives complexity bounds of $\tilde{O}(\|P\|_0^{3/2})$ and $\tilde{O}(\|P\|_0 \text{diam}(A))$ where $\text{diam}(A)$ is the diameter of the graph corresponding to the matrix A .

We will summarize this complexity by saying that the running time “explicitly scales nearly linearly in the number of nonzero entries.” The “nearly” comes from the logarithmic terms; the word “explicit” comes because the scaling also depends on the imbalance of the optimal lockdown κ , and one can construct families of examples where the κ will have some kind of scaling with network size. Finally, we will say that a problem is reducible to matrix balancing if we can write down a matrix balancing problem whose solution can be translated in linear time to the solution of the original problem.

B. Main theoretical contribution

With the above preliminaries in place, we can now state our main theoretical contribution. Our main theorem provides efficient algorithms for solving the lockdown problem. Our key contribution is to give an algorithm for optimal stabilizing lockdown whose complexity has an explicit scaling which, under some assumptions, is nearly linear in the number of nonzero entries of the matrix A . That is to say, not only can the optimal heterogeneous lockdown be computed exactly, but doing so can take not much more operations as just reading through the data.

Our first step is to discuss an assumption required by one of our algorithms. The formal statement of the assumption is as follows.

Assumption 1 (High spread assumption): We have

$$\text{diag}(B^T \text{diag}(s(t_0))C) \geq \frac{\epsilon + r^a}{\beta^a + \beta^s \epsilon / r^s}.$$

To see why this condition is satisfied in a “high spread” regime, note that $A = CB^T$ and, given our choices of C and B in Eq. (5), the entry $C_{ii}B_{ii}$ corresponds to the spread of the epidemic in location i purely from the same-location trips of residents of location i . The left-hand side of the above equation is at least $\text{diag}(s(t_0))C_{ii}B_{ii}$, so if A_{ii} and the initial proportion of infected are large enough relative to the other parameters, this inequality will be satisfied.

We next turn to our main result.

Theorem 2: Suppose the graph corresponding to positive entries of the matrix A is strongly connected and $s(t_0) > 0$. Then:

- 1) Suppose Assumption 1 holds. Then the lockdown problem is reducible to matrix balancing.
- 2) If Assumption 1 does not hold, then the lockdown problem can be solved by computing an eigendecomposition and then performing $\tilde{O}(n^\omega)$ operations, where ω is the exponent of matrix multiplication and the $\tilde{O}(\cdot)$ notation hides factors which are logarithmic in the remaining problem parameters.

Our main result gives favorable complexities for lockdown problem depending on whether the high-spread assumption holds, though the complexity is a little worse if it does not. Analogous results hold for the SIS model, as explained in Appendix B of this paper. In practice, we find the optimal lockdown problem is solvable using the matrix balancing algorithm in the vast majority of the cases. Specifically, when we fit the models to New York State data, in 23 experiments out of 27 (where we varied the epidemic model from SIS/SIR/COVID-19 as well as different parameter values from the literature), the matrix balancing algorithm gave the correct answer.

The next section provides a proof of our main result.

III. PROOF OF THEOREM 2

We will present a series of lemmas and observations which will culminate in the proof of the main theorem. It is here that we will perform the reduction from the problem of computing the optimal lockdown to matrix balancing.

Our starting point is the following lemma on a “splitting” of a positive matrix.

Lemma 3:

- 1) A strongly connected matrix P with non-negative off-diagonal elements is continuous-time stable if and only if there exists $d > 0$ such that $Pd \leq 0$.
- 2) The nonnegative strongly connected matrix B is discrete-time stable if and only if there exists $d > 0$ such that $Bd \leq d$.
- 3) Suppose $P = L - D$ where L is nonnegative while D is a matrix with nonpositive off-diagonal elements whose inverse is elementwise nonnegative. Suppose further that both P and $D^{-1}L$ are both strongly connected. Then P is continuous time stable if and only if $B = D^{-1}L$ is discrete-time stable.

This lemma is a small variation on a well-known fact: usually, parts 1 and 2 are stated for strictly stable matrices, in which case all the inequalities need to be strict (see [46], Theorem 15.17 for the strict version of part (i) and Proposition 1 of [47] for the strict version of part (ii)). The non-strict version additionally requires that the matrices be strongly connected, which is not needed for the nonstrict version of this problem. Note that we do not claim that any part of this lemma is novel. For completeness, we nevertheless give a proof in [34] (See proof of Lemma 4.7.).

We will later need to interchange the order of products while still preserving the condition of being strongly connected. To that end, the following lemma will be useful.

Lemma 4: Suppose U, V are two nonnegative $n \times n$ matrices with no zero rows or columns such that VU is strongly connected. Then UV is strongly connected.

Proof: Consider a directed bipartite graph G on $2n$ vertices, with vertices $l_1, \dots, l_n$ and $r_1, \dots, r_n$ denoting the two sides of the bipartition, defined as follows. If $U_{ij} > 0$, then we put an edge from l_j to r_i ; and if $V_{ij} > 0$ we put an edge from r_j to l_i . Let G_1 be the graph on $l_1, \dots, l_n$ where we put the directed edge from l_i to l_j if there is a path of length two from l_i to l_j in G . Likewise, let G_2 be the directed graph on $r_1, \dots, r_n$ such that we put an edge from r_i to r_j whenever there is a path of length two in G .

Then the strong connectivity of VU is equivalent to having G_1 be strongly connected: indeed, $(VU)_{ab} > 0$ if and only if there exists a link from b to a in G_1 . Similarly, the strong connectivity of UV is equivalent to having G_2 be strongly connected. We will show that if G_1 is not strongly connected, neither is G_2 . The converse is established via a similar argument.

Indeed, suppose G_1 is not strongly connected. That means there exists a proper subset of the vertices, say $L_1 = \{l_1, \dots, l_k\}$, with no edges outgoing to $L_1^c = \{l_{k+1}, \dots, l_n\}$. Let r_1 be the set of out-neighbors of L_1 in G . Then we must have that r_1 is a proper subset of $\{r_1, \dots, r_n\}$ (for otherwise, the assumption that V has no zero rows/columns would contradict no edges going from L_1 to L_1^c in G_1) and the set of out-neighbors of r_1 in G is contained in L_1 .

But since (i) r_1 is a proper subset of the right-hand side (ii) the out-neighbors of r_1 in G are contained in L_1 (iii) the

out-neighbors of L_1 in G are contained in r_1 , we obtain that there are no edges leading from r_1 to r_1^c in G_2 . This proves G_2 is not strongly connected. ■

With these preliminary lemmas in place, we now turn to core of our reduction of lockdown to matrix balancing. The reduction will go through several “intermediate” problems, the first of which as follows.

Definition 1: We will refer to the following as the *stability scaling problem*: given a nonnegative strongly connected matrix P and positive diagonal matrix D , find positive scalars $q_1, \dots, q_n$ minimizing $\sum_{i=1}^n q_i^{-1}$ such that $\text{diag}(q_1, \dots, q_n)P - D$ is continuous-time stable.

We will find it useful below to refer to the *unconstrained lockdown problem*: this is the same as the lockdown problem described above but we do not constrain the entries of z to lie in $[0, 1]$ but rather only require them to be nonnegative. In particular, an unconstrained lockdown is allowed to increase intensity of activity in some locations by setting $z_i^* > 1$ (presumably to offset decreased activity in other locations while meeting a cost constraint). We will sometimes differentiate between these by talking about the *constrained case* (meaning the original formulation) and the *unconstrained case* (meaning that entries of z do not have to be upper bounded by one).

While this problem formulation is not of practical interest, our proofs will heavily use the fact that, under certain conditions, the constrained and unconstrained problems coincide. This is quite intuitive: if the epidemic spreads sufficiently fast everywhere, the unconstrained shutdown will never choose to increase the activity of any location.

The utility of these definitions should become clear after the following lemma.

Lemma 5: In the unconstrained case, the minimum lockdown model for COVID-19 dynamics can be reduced to stability scaling provided A is strongly connected and $s(t_0) > 0$. In the constrained case, the same holds under Assumption 1.

Proof: Let us begin by assuming that

$$\alpha < \min(r^s, \epsilon + r^a). \quad (9)$$

We will revisit this assumption later. We need to make the matrix

$$A_0 := \begin{pmatrix} \beta^a \text{diag}(s(t_0))A_z - (\epsilon + r^a) + \alpha & \beta^s \text{diag}(s(t_0))A_z \\ \epsilon & -r^s + \alpha \end{pmatrix} \quad (10)$$

stable, where we have introduced the notation that $A_z = C \text{diag}(z) B^T$. Let us write

$$A_0 = L - D,$$

where

$$L = \begin{pmatrix} \beta^a \text{diag}(s(t_0))A_z & \beta^s \text{diag}(s(t_0))A_z \\ 0 & 0 \end{pmatrix},$$

and

$$D = \begin{pmatrix} \epsilon + r^a - \alpha & 0 \\ -\epsilon & r^s - \alpha \end{pmatrix}.$$

We next apply part 3 of Lemma 3 to get that A_0 is continuous-time stable if and only if $D^{-1}L$ is discrete time stable. To do this, however, we need to verify that D^{-1} is elementwise nonnegative, and that both P and $D^{-1}L$ are strongly connected. This easily follows from observing that

$$D^{-1} = \begin{pmatrix} \frac{1}{\epsilon + r^a - \alpha} & 0 \\ \frac{\epsilon}{(\epsilon + r^a - \alpha)(r^s - \alpha)} & \frac{1}{r^s - \alpha} \end{pmatrix},$$

and recalling that $s(t_0) > 0$ as well as Eq. (9).

To summarize, we have shown that equivalently we need to make sure that $B = D^{-1}L$ is discrete-time stable. Since the eigenvalues of a matrix which is the product of two matrices do not change after changing the order of the product, this is the same as having LD^{-1} be discrete-time stable. But we have the simple expression

$$LD^{-1} = \begin{pmatrix} \text{diag}(s(t_0))A_z b_1 & \frac{\beta^s}{r^s - \alpha} \text{diag}(s(t_0))A_z \\ 0 & 0 \end{pmatrix}, \quad (11)$$

where $b_1 = \frac{\beta^s \epsilon + \beta^s (r^s - \alpha)}{(\epsilon + r^a - \alpha)(r^s - \alpha)}$.

But the eigenvalues of LD^{-1} are just the eigenvalues of its top $n \times n$ block. Thus we obtain that LD^{-1} is stable if and only if the matrix $\text{diag}(s(t_0))A_z b_1$ is discrete-time stable. Plugging in $A = C \text{diag}(z) B^T$, we now need that

$$\text{diag}(s(t_0))C \text{diag}(z) B^T b_1 \text{ is discrete-time stable.} \quad (12)$$

We next appear to part (3) of Lemma 3 again, using $D = \text{diag}(s(t_0))^{-1} b^{-1} I$ to obtain that we need

$$C \text{diag}(z) B^T - \text{diag}(s(t_0))^{-1} b^{-1} I \text{ is continuous-time stable,} \quad (13)$$

We can apply part (3) of Lemma 3 as strong connectivity follows from z being elementwise positive and strong connectivity of $A = CB^T$. But the last condition is equivalent to Equation (24), which we've already shown how to reduce to stability scaling.

We conclude the reduction by observing that, once again, this results in an instance of stability scaling with the matrix $P = B^T C$, which is strongly connected because $A = CB^T$ is strongly connected, allowing us to apply Lemma 4. Further, the matrix $\text{diag}(s(t_0))^{-1} b^{-1} I$ is a positive diagonal matrix due to the assumptions that $s(t_0) > 0$ and the assumption of Eq. (9).

For the constrained case, we argue that under Assumption 1, the optimal solution will have all $z_i^* \leq 1$, so we can equivalently consider the unconstrained case. Indeed, suppose e.g., that $z_j^* > 1$. In that case, the matrix $\text{diag}(z^*) B^T \text{diag}(s(t_0)) C$ has its (j, j) 'th entry at least b_1^{-1} . Since the j 'th row of $\text{diag}(z^*) B^T \text{diag}(s(t_0)) C$ is nonnegative, and since $B^T \text{diag}(s(t_0)) C$ is strongly connected by Lemma 4, we have that the j 'th row of $\text{diag}(z^*) B^T \text{diag}(s(t_0)) C$ is not zero. By the same argument as we made in the SIS case, the matrix $\text{diag}(z^*) B^T \text{diag}(s(t_0)) C - b_1^{-1} I$ cannot be continuous-time stable. This implies that Eq. (13) cannot be satisfied.

Finally, we conclude the proof by revisiting the assumption we made in the very beginning, namely the assumption of Eq. (9). We now argue the problem of optimal lockdown has no solution if that equation fails. Indeed, in that case, the matrix

A_0 has either its first $n \times n$ block nonnegative, or its last $n \times n$ block nonnegative. We show that the problem has no solution in the first case; the other case has a similar proof.

We want to argue that having $d > 0$ such that $A_0 d \leq 0$ cannot occur if the top $n \times n$ block of A_0 is nonnegative; by Lemma 3 part (1) this is enough to show A cannot be stable regardless of z . We can partition $d = [d_1, d_2]$, and we immediately see that we must have that $d_1 = 0$, since, by the strong connectivity of $\text{diag}(s(t_0))A_z$, every entry of d_1 is multiplied by some entry in the top left $n \times n$ submatrix of A_0 . Applying the same argument to the “top right” $n \times n$ block of A_0 , get that $d_2 = 0$, and this is a contradiction. This concludes the proof. ■

Thus we have just reduced the unconstrained lockdown problem to stability scaling; further, the constrained lockdown problem was reduced to the same under the high-spread condition.

Our next step is to reduce the stability scaling problem to a new problem, defined next, which has a somewhat simpler form.

Definition 2: Given a nonnegative matrix A , the problem of finding scalars $l_1, \dots, l_n$ such that $A - \text{diag}(l_1, \dots, l_n)$ is continuous-time stable while minimizing $\sum_{i=1}^n c_i l_i$ will be called the *diagonal subtraction problem*.

Lemma 6: Stability scaling can be reduced to diagonal subtraction.

Proof: Indeed, starting from

$$\min_{q_1 > 0, \dots, q_n > 0} \sum_{i=1}^n c_i q_i^{-1}$$

such that $\text{diag}(q_1, \dots, q_n)P - D$ is continuous time stable,

we apply Lemma 3, part (c) to obtain that the constraint is the same as

$$D^{-1} \text{diag}(q)P \text{ is discrete-time stable.}$$

Here we used crucially that the matrix P is nonnegative and strongly connected and that the diagonal matrix D is positive, ensuring that the assumptions of Lemma 3, part 3 are satisfied. We next use the same Lemma 3 part (c) again to obtain that this the last constraint is identical to

$$P - \text{diag}(q_1, \dots, q_n)^{-1}D \text{ is continuous time stable}$$

For simplicity, it is convenient to introduce the notation $\delta_i = q_i^{-1}$, $i = 1, \dots, n$. In terms of these new variables δ_i , we have the problem

$$\min_{\delta_1 > 0, \dots, \delta_n > 0} \sum_{i=1}^n c_i \delta_i$$

subject to $P - \text{diag}(d_1 \delta_1, \dots, d_n \delta_n)$ is continuous-time stable,

where $D = \text{diag}(d_1, \dots, d_n)$. Finally, introducing variables $u_i = d_i \delta_i$, $i = 1, \dots, n$, we have

$$\min_{u_1 > 0, \dots, u_n > 0} \sum_{i=1}^n c_i \frac{u_i}{d_i}$$

subject to $P - \text{diag}(u_1, \dots, u_n)$ is continuous-time stable ,

which is an instance of diagonal subtraction with cost coefficients c_i/d_i . ■

With all these preliminaries in place, we are now ready to turn to the proof of the first part of our main theoretical result. In Lemma 10, we have shown how to reduce the optimal lockdown problem to stability scaling. In the subsequent Lemma 6, we showed how to reduce stability scaling to diagonal subtraction. We now prove Theorem 2 by reducing diagonal subtraction to matrix balancing. After submission of this paper, we have found that this last assertion can be proved by putting together the results of [48] and [49]. We nevertheless include our original proof below for completeness, and so that the argument is available in a single place, while acknowledging that the reduction from diagonal subtraction to matrix balancing was obtained earlier.

Proof: [Proof of Theorem 2, parts 1]

Indeed, we need to solve

$$\min_{l_1 > 0, \dots, l_n > 0} \sum_{i=1}^n c_i l_i$$

such that $A - \text{diag}(l_1, \dots, l_n)$ is continuous time stable,

where A is a nonnegative matrix. We use part (a) of Lemma 3 to write this as

$$\min_{l_1, \dots, l_n, d_1, \dots, d_n} \sum_{i=1}^n c_i l_i \quad (14)$$

such that $Ad - \text{diag}(l_1 d_1, \dots, l_n d_n) \leq 0$

$$d > 0, l > 0$$

The difficulty here is that we are optimizing over l and d , and the constraint includes a product of the variables. We try to make this into a simpler problem by introducing the variables $\mu_i = l_i d_i$, $i = 1, \dots, n$. We change variables from (l, d) to (μ, d) to obtain the equivalent problem:

$$\min_{\mu_1, \dots, \mu_n, d_1, \dots, d_n} \sum_{i=1}^n c_i \mu_i / d_i \quad (15)$$

such that $Ad \leq \mu$

$$d > 0, \mu > 0$$

In this reformulation, the constraint is linear, and the nonlinearity has been moved to the objective.

Next, consider what happens when we fix some $d > 0$ and consider the best μ for that particular d . Because $c_i > 0$ for all i , and $d > 0$, we want to choose each element μ_i as small as possible. The best choice is clearly $\mu = Ad$; no component of μ can be lower than that by our constraints. Because A is a nonnegative matrix without zero rows by strong connectivity, this results in a feasible $\mu > 0$.

Thus we can transform this into a problem which optimizes over just the variables $d_1, \dots, d_n$:

$$\min_{d_1 > 0, \dots, d_n > 0} \sum_{i=1}^n c_i [Ad]_i / d_i.$$

Our next step is to change variables one more time. Since d is a positive vector, it is natural to write $d_i = e^{g_i}$. In terms

of the new variables $g_1, \dots, g_n$, we just need to minimize the function $f(g_1, \dots, g_n)$ defined as

$$\min f(g_1, \dots, g_n) := \min_{g_1, \dots, g_n} \sum_{i=1}^n \sum_{j=1}^n c_i a_{ij} e^{g_j - g_i}.$$

We now have an unconstrained minimization of a convex function. In particular, if we can find a point where $\nabla f(g_1, \dots, g_n) = 0$, we will have found the global optimum.

Let us consider the k 'th component of the gradient of $f(g_1, \dots, g_n)$. We have the equation

$$\begin{aligned} 0 = \frac{\partial f}{\partial g_k}(g_1, \dots, g_k) &= \sum_{j \neq k} -c_k a_{kj} e^{g_j - g_k} + \sum_{i \neq k} c_i a_{ik} e^{g_k - g_i} \\ &= \sum_{j=1}^n -c_k a_{kj} e^{g_j - g_k} + \sum_{i=1}^n c_i a_{ik} e^{g_k - g_i} \end{aligned} \quad (16)$$

Let $D_g = \text{diag}(e^{g_1}, \dots, e^{g_n})$ and $D_c = \text{diag}(c_1, \dots, c_n)$. Observing that

$$[D_g^{-1} D_c A D_g]_{uv} = e^{-g_u} c_u a_{uv} e^{g_v} = c_u a_{uv} e^{g_v - g_u},$$

we have that Eq. (16) can be written as

$$-[D_g^{-1} D_c A D_g \mathbf{1}]_k + [\mathbf{1}^\top D_g^{-1} D_c A D_g]_k = 0$$

In other words, the matrix $D_g^{-1} D_c A D_g$ needs to have its k 'th column sum equals to its k 'th row sum. Thus provided we can find a balancing of the matrix $D_c A$, we will have found the minimum of $f(g_1, \dots, g_n)$ as needed.

However, a matrix X can be balanced if and only if the underlying graph $G(X)$ is strongly connected [50]. Since we have assumed that c_i are all positive and A is strongly connected, we have that $D_c A$ can be balanced. We conclude that the unique minimum of $f(g_1, \dots, g_n)$ can be recovered from the balancing of this matrix. This concludes the reduction of diagonal subtraction to matrix balancing. ■

Having proved the first part of Theorem 2, we now turn to the second part. We will not be using any of the reductions used to prove the first part of the theorem. The first steps of our proof are very similar to the proof of the main result of [11], which gave a semidefinite formulation of a lockdown problem ensuring asymptotic stability in a model involving births and deaths. We diverge from [11] when we write the problem as a so-called covering semi-definite problem and analyze its complexity.

Proof: [Proof of Theorem 2, part 3] We can simply repeat the proof of Lemma 10 to reduce the constrained lockdown to stability scaling; except that, without Assumption 1, we now have to keep the constraint that $z \in [0, 1]^n$. Indeed, note that the only place where Assumption 1 was used in those two proofs was to argue that we can omit that constraint.

In this way, we can reduce the constrained lockdown problem to the following:

$$\min_{z_1, \dots, z_n} \sum_{i=1}^n c_i z_i^{-1}$$

$\text{diag}(a)C\text{diag}(z_1, \dots, z_n)B^\top - qI$ is continuous time stable

$$z \in [0, 1]^n,$$

where, from Eq. (13) we have that $a = s(t_0)$ while $q = b_1^{-1}$. Note that this optimization problem is only over the variables $z_1, \dots, z_n$; all other quantities appearing in it are parameters.

Suppose next that matrices C and B are chosen according to Eq. (5). In other words, we have the problem

$$\min_{z_1, \dots, z_n} \sum_{i=1}^n c_i z_i^{-1}$$

$\text{diag}(a)\tau\text{diag}(z_1, \dots, z_n)D_1\tau^\top D_2 - qI$ is continuous time stable

$$z \in [0, 1]^n,$$

where, recall, the matrices D_1, D_2 are defined immediately after Eq. (5).

Since the nonzero eigenvalues of a product of two matrices do not change when we flip the order in which they are multiplied, the second constraint is equivalent to

$\text{diag}(z_1, \dots, z_n)D_1\tau^\top D_2\text{diag}(a)\tau - qI$ is continuous time stable.

Applying part (3) of Lemma 3, we get that this is the same as

$$q^{-1}\text{diag}(z_1, \dots, z_n)D_1\tau^\top D_2\text{diag}(a)\tau \text{ is discrete time stable.}$$

Here we used the strong connectivity and positive diagonal of the matrix τ . Next, applying the same lemma again, we further get the equivalence to

$$\tau^\top D_2\text{diag}(a)\tau - D_1^{-1}\text{diag}(z_1, \dots, z_n)^{-1}qI \text{ is continuous time stable.}$$

This can be simplified further by observing that a symmetric matrix is stable if and only if it is non-positive-definite.

Indeed, let us introduce the notation $u_i = (D_1)_{ii}^{-1} q z_i^{-1}$, and $Q = \tau^\top D_2\text{diag}(a)\tau$. We can therefore write our problem as

$$\begin{aligned} \min \quad & \sum_{i=1}^n \frac{c_i (D_1)_{ii}}{q} u_i \\ \text{s.t.} \quad & 0 \geq Q - \text{diag}(u_1, \dots, u_n) \\ & u_i \in [q(D_1)_{ii}^{-1}, +\infty) \text{ for all } i = 1, \dots, n. \end{aligned}$$

Further defining $c'_i = c_i (D_1)_{ii}$ and the $2n \times 2n$ positive-semidefinite matrices $B_i = e_i e_i^\top + e_{n+i} e_{n+i}^\top$, we can write this as

$$\begin{aligned} \min \quad & \sum_{i=1}^n c'_i u_i \\ \text{s.t.} \quad & \sum_{i=1}^n u_i B_i \geq \begin{pmatrix} Q & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & q \text{diag}((D_1)_{11}^{-1}, \dots, (D_1)_{nn}^{-1}) \end{pmatrix} \end{aligned} \quad (17)$$

This is known as a ‘‘covering SDP.’’ In the recent paper, [51], it was shown that it can be solved, up to various logarithmic factors, in matrix multiplication time. We next discuss in detail the results in [51] and how they are applicable to Eq. (17).

Specifically, in [51], an algorithm for checking whether the system of equations

$$\sum_i x_i C_i \geq I, \sum_i x_i P_i \leq (1 - \delta)I, x \geq 0,$$

is feasible was given; here C_i, P_i are arbitrary nonnegative definite matrices. Formally, this is what is defined to be a covering SDP.

To begin applying this to our problem, let us choose $\delta = 1/2$ and $P_i = \frac{1}{2n} u^{-1} c'_i \mathbf{1}\mathbf{1}^\top$ for all i . In that case, we are checking the feasibility of

$$\sum_i x_i C_i \geq I, \sum_i c'_i x_i \leq u, x \geq 0. \quad (18)$$

The running time of the algorithm from [51] for checking whether this is feasible is $\tilde{O}(\sum_i \|C_i\|_0 + n^\omega)$, where the $\tilde{O}$ hides a multiplicative factor which is a power of logarithm is $O(n^2 \max_i \lambda_{\max}(C_i) / \lambda_{\min}(P_i))$ (see discussion under ‘‘Main Claim’’ in [51], and note that, in our case, the number of matrices and the dimension are proportional). In our case, we will choose $C_i = e_i e_i^\top + e_{n+i} e_{n+i}^\top$ so that

$$\lambda_{\max}(C_i) \leq 1, \lambda_{\min}(P_i) = \frac{1}{2} \max_i c'_i u^{-1},$$

and plugging the definition of c'_i , we have that the factor being hidden in the tilde is a power of logarithm of

$$\begin{aligned} n^2 \frac{1}{\max_i \lambda_{\max}(P_i)} &= 2n^2 \frac{1}{\max_i c'_i u^{-1}} \\ &= 2n^2 u \frac{1}{\max_i c'_i} \\ &= 2n^2 u \frac{1}{\max_i c_i (D_1)_{ii}} \\ &\leq 2n^2 u^{-1} \frac{1}{\max_i c_i \left(\sum_j N_j \tau_{ji} \right)^{-1}}. \end{aligned}$$

To summarize, we have discussed how long it takes to check feasibility of Eq. (18). The conclusion is that the running time $\tilde{O}(\sum_i \|C_i\|_0 + n^\omega)$, where the tilde notation hides a factor that is logarithmic in the problem parameters $n, u^{-1}, \min_i \sum_j c_i^{-1} N_j \tau_{ji}$.

From feasibility to optimization: We have just shown that checking feasibility of Eq. (18) can be done in polynomial time. We want to argue that this implies that minimizing $\sum_{i=1}^n c'_i x_i$ subject to the constraint $\sum_i x_i C_i \geq I, x \geq 0$ is polynomial-time. Of course, the way to do this is at obvious: $n \epsilon'$ -additive approximation to the latter can be found with lots of feasibility checks.

Let us suppose that we actually know that $x_i^* \geq l_i$ for some l_i (we do, in fact, know this because Eq. (17) forces

$$x_i^* \geq q(D_1)_{ii}^{-1} = q \sum_k N_k \tau_{ki} \quad (19)$$

$$\begin{aligned} &:= q N_i^{\text{visit}} \\ &:= l_i \end{aligned} \quad (20)$$

If c^* is the optimal cost, then we need to do $O(\log(c^* / \sum_i c'_i l_i) + \log \epsilon'^{-1})$ feasibility checks to get an ϵ additive approximation to the optimal cost. Upper bounding $c^* \leq n c_{\max} \max_i x_i^*$, we see that we need to do

$$O\left(\log \frac{n \max_i c_i \max_i x_i^*}{q \sum_i c'_i N_i^{\text{visit}}} + \log \epsilon'^{-1}\right),$$

feasibility checks.

Last step: getting to Eq. (17) We have just bounded the complexity of minimizing a linear objective subject to the constraint $\sum_i x_i C_i \geq I, x \geq 0$. But we cannot quite apply this directly to Eq. (17), since in Eq. (17), the right-hand side is not the identity matrix.

However, the right-hand side of Eq. (17) is symmetric, so it can be decomposed as $U \Lambda U^\top$ for orthogonal U and diagonal Λ . Since $A \geq B$ implies $Z A Z^\top \geq Z B Z^\top$, by choosing $Z = \Lambda^{-1/2} U^\top$ we can equivalently rewrite Eq. (17) as

$$\begin{aligned} \min \quad & \sum_{i=1}^n c'_i u_i \\ \text{s.t.} \quad & \sum_{i=1}^n u_i \Lambda^{-1/2} U^\top B_i U \Lambda^{-1/2} \geq I. \end{aligned}$$

Letting $C'_i = \Lambda^{-1/2} U^\top B_i U \Lambda^{-1/2}$, we thus obtain a problem to which the algorithm of Eq. ([51]) is applicable. Unfortunately, in the transformation the number of nonzero entries in the matrix C_i changes. Nevertheless, as explained in [51] (in footnote 6 in the arxiv version), despite the change of variables, the scaling in the running time remains with the number of nonzero entries in the matrices B_i . However, we now need to add the complexity of computing the eigendecomposition of the right-hand side of Eq. (17) to our complexity bounds.

Putting everything together: We can now put together everything we have observed above. The total complexity scales as the complexity of matrix multiplication and computing the eigendecomposition of a symmetric matrix, which takes $O(n^3)$ exact arithmetic operations [52]. This is multiplied by a power of a logarithm in the quantities

$$\epsilon'^{-1}, n, \max_i c_i, \min_i \sum_j c_i^{-1} N_j \tau_{ji}, \max_i x_i^*, u^{-1}, \frac{b_1}{\sum_i c'_i N_i^{\text{visit}}}, \lambda_{\max}(C')$$

where N_i^{visit} was defined in Eq. (20).

Now all of these are naturally viewed as problem parameters – except the final four, i.e., $\max_i x_i^*, u^{-1}, \frac{1}{\sum_i c'_i N_i^{\text{visit}}}, \lambda_{\max}(C')$ which are quantities we encountered in the course of the proof. Let us upper bound these in terms of other quantities.

First, we have that

$$\begin{aligned} \max_i x_i^* &= \max_i u_i^* \leq q \max_i q(D_1)_{ii}^{-1} (\min_i z_i^*)^{-1} \\ &= b_1^{-1} \left(\max_a \sum_k N_k \tau_{ka} \right) (\min_i z_i^*)^{-1}, \end{aligned}$$

which gives us the first bound we need.

Next, the smallest u will be is

$$\sum_i c'_i l_i = q \sum_i c'_i N_i^{\text{visit}}.$$

Thus we can ‘‘kill two birds with one stone’’ by first upper bounding u^{-1} as

$$\left(\sum_i c'_i l_i \right)^{-1} = \left(\sum_i c_i (D_1)_{ii}^{-1} q \sum_k N_k \tau_{ki} \right)^{-1}$$

which is a polynomial in

$$\max_i c_i^{-1}, \max_i \left(\sum_k N_k \tau_{ki} \right)^{-1}, b_1,$$

and this also upper bounds

$$\left(\sum_i c'_i N_i^{\text{visit}} \right)^{-1}$$

by a polynomial in the same quantities and q^{-1} . This gives the second and third bounds we need.

Finally,

$$\lambda_{\max}(C'_i) \leq \frac{1}{\min_i \lambda_{ii}} \leq \frac{1}{\min(\lambda_{\min}(Q), q \min_a N_a^{\text{visit}})},$$

which is in turn a polynomial in

$$\frac{1}{\lambda_{\min}(Q)}, q^{-1}, \left(\min_a N_k \tau_{ka} \right)^{-1}$$

To upper bound this, it only remains to lower bound $\lambda_{\min}$

However, since $Q = \tau^\top D_2 \text{diag}(a) \tau$, we have that

$$\begin{aligned} \lambda_{\min}(Q) &\geq \left(\min_i (D_2)_{ii} a_i \right) \lambda_{\min}(\tau^\top \tau) \\ &= \lambda_{\min}(\tau^\top \tau) \left(\min_i \beta^8 N_i s_i(t_0) \right). \end{aligned}$$

Adding up all the complexity: To summarize, to compute an ϵ' additive approximation to the optimal solution we need to compute an eigendecomposition and then perform a number of operations which is $O(n^\omega)$ times a polylog in the variables

$$\begin{aligned} &\epsilon'^{-1}, n, \max_i c_i, \max_i c_i^{-1}, (\min_i z_i^*)^{-1}, b_1, b_1^{-1}, (\lambda_{\min}(\tau^\top \tau))^{-1}, \\ &\max_a \sum_k N_k \tau_{ka}, \left(\min_a \sum_k N_k \tau_{ka} \right)^{-1}, \left(\min_i \beta^8 N_i s_i(t_0) \right)^{-1}. \end{aligned}$$

IV. SIMULATIONS

We now apply the algorithms we've developed to design an optimal stabilizing lockdown policy for the 62 counties the State of New York (NY). Our goal is to reduce active in each county in a non-uniform way to curb the spread COVID-19 while simultaneously minimizing economic cost. The data sources we employed are presented in the full version of our paper [34] (SI Sec. 6). When the "high-spreading" assumption is satisfied, we will apply the matrix-balancing algorithm, otherwise we will apply the covering semi-definite program. To provide valid estimation results, we employ three different sets of disease parameters provided in literature [10], [11], [9] (See SI Table 2 of [34]). Because the amount of simulations performed was too large to fit within the TAC page limits, we report here on the key results and include pointers to additional simulations in our arxiv version [34].

Comparison with other lockdown policies. We used the data of the 62 counties in NY on Apr. 1st, 2020 as initialization and estimated the number of active cases over 300 ~ 800 days (04/01/2020 - 01/26/2021 or 04/01/2020 - 06/01/2021) and the number of cumulative cases over 500 ~ 1,500 days (04/01/2020 - 08/14/2021 or 04/01/2020 - 05/10/2024) with different lockdown policies. Fig. 1a-c show the estimated active cases over times, and Fig. 1d-e show the estimated cumulative cases over time. Here results from different columns

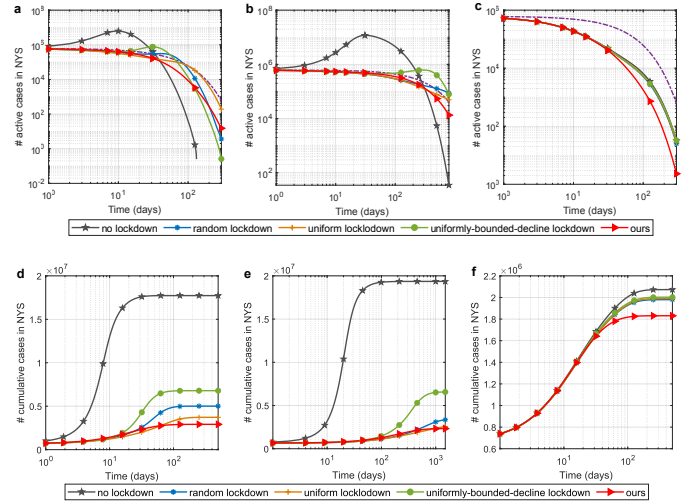


Fig. 1 Estimated number of active cases and cumulative cases for the COVID-19 model by applying different lockdown policies based on available data about COVID-19 outbreak in NY on April 1st, 2020. **a-c**, the estimated number of active cases in NY from Apr. 1st, 2020 to Jan. 26th, 2021 (or June 10th, 2022). **d-f**, the estimated cumulative cases in NY from Apr. 1st, 2020 to Aug. 14th, 2021 (or May 10th, 2024). In **a, d**, the disease parameters are set as in [10], the decay rate α is chosen as 0.0231 which corresponds to halving every 30 days. In **b, e**, the disease parameters are set as in [9], the decay rate is chosen as $\alpha = 0.2r^s = 0.0034$ so that $\alpha < \min(r^a, r^s)$. In **c, f**, the disease parameters are set as in [11], the decay rate α is chosen 0.0231 that corresponds to halving every 30 days.

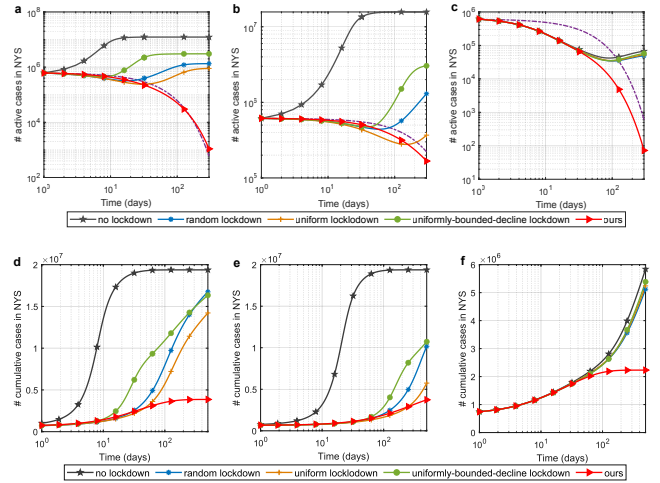


Fig. 2 Estimated number of active cases and cumulative cases for the SIS model by applying different lockdown policies based on available data about COVID-19 outbreak in NY on April 1st, 2020. **a-c**, the estimated number of active cases in NY from Apr. 1st, 2020 to Jan. 26th, 2021 (or June 10th, 2022). **d-f**, the estimated cumulative cases in NY from Apr. 1st, 2020 to Aug. 14th, 2021 (or May 10th, 2024). In **a, d**, the disease parameters are set as in [10], the decay rate α is chosen as 0.0231 which corresponds to halving every 30 days. In **b, e**, the disease parameters are set as in [9], the decay rate is chosen as $\alpha = 0.2r^s = 0.0034$ so that $\alpha < \min(r^a, r^s)$. In **c, f**, the disease parameters are set as in [11], the decay rate α is chosen 0.0231 that corresponds to halving every 30 days.

of Fig. 2 were calculated by using different sets of disease parameters adopted from literature [9]–[11].

We compared the optimal stabilizing lockdown policy calcu-

lated by our methods with several other benchmark policies: (1) no lockdown, i.e., $z_l = 1$ for all locations; (2) random lockdown, z_l is randomly chosen from a uniform distribution $\mathcal{U}[a, b]$ with the lower- and upper-bounds a and b chosen such that the overall cost of this policy is the same as that of our policy; (3) uniform lockdown, where $z_l = z$ is the same for all the locations and z is chosen such that the overall economic cost is the same as that of our policy; (4) uniformly-bounded-decline lockdown, where the “decline” is uniformly bounded across locations, i.e., the decay rate of the infections in each location is bounded by a constant α , where α is chosen such that the economic cost of this policy is the same as that of our policy. Note that among the four benchmark policies both the uniformly-bounded-decline lockdown and the random lockdown are heterogeneous locationwise.

From Fig. 1, Fig. 2, we can see that our optimal stabilizing lockdown policy outperforms all other lockdown policies in terms of the total final number of cumulative cases based on COVID-19 model and SIS model, respectively. Similar findings for SIR model is also reported in SI Fig. 2 in [34].

Optimal stabilizing lockdown rate z_l^* for each county.

Fig. 3 shows lockdown-rate profiles z_l calculated by various policies. First, we found that the optimal stabilizing lockdown profile is quite sensitive to the disease parameters. Second, surprisingly, the values of z_l^* for counties in NYC are relatively higher (corresponding to a less stringent lockdown) than that of counties outside NYC, regardless of the disease parameters. *Even though the epidemic was largely localized around NYC on the date we used to initialize the infection rates, the calculated optimal stabilizing lockdown profile indicates that it is cheaper to reduce the spread of COVID-19 by being harsher on neighboring regions with smaller populations.* It can be observed from Fig. 3 that this pattern only appears in our heterogenous optimal stabilizing lockdown policy. The result is counter-intuitive. Indeed, it is easy to see that in non-network models the lockdown strength is invariant to population². It is therefore strange to see that a higher-population higher-infection node (NYC) get assigned a lower intensity lockdown under our optimal fixed lockdown.

We have performed a number of additional simulations replicating this phenomenon and verifying its robustness: for lack of space, these can be found in [34]. In SI Sec. 7 of [34], we further replicated the same finding in a much simpler city-suburb model: we consider a city with large population and a neighboring suburb with small population and observe that the optimal stabilizing lockdown will choose to shut down the suburb harder. In SI Sec. 10 of [34], we further confirmed the same counterintuitive phenomenon using other cost functions. In SI Sec. 11 of [34], we checked the robustness of this counterintuitive phenomenon with respect to the uncertainty of the travel rate matrix τ by adding noises or removing part of the travelling data. It turns out that this phenomenon is quite robust against the uncertainty of the matrix τ . In particular, perturbing each τ_{ij} by noise with variance up to $10\tau_{ij}^2$ preserves

²Observe, for example, that scaling all the populations in Eq. (6) by the same number leaves the matrix A , and consequently the optimal stabilizing lockdown, unchanged.

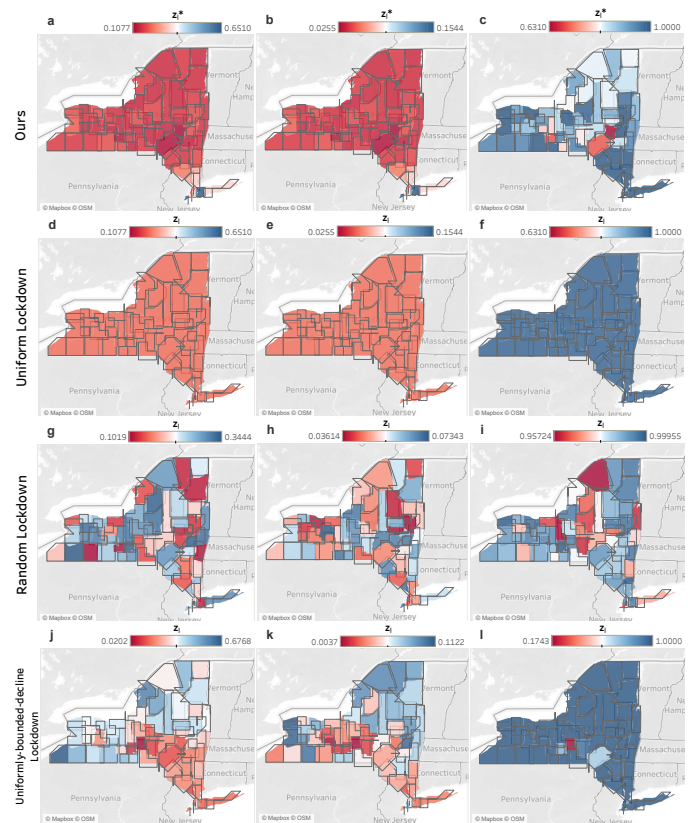


Fig. 3 Lockdown rate of each county given by different policies for the COVID-19 model based on available data about COVID-19 outbreak in NY on April 1st, 2020. **a-c**, optimal lockdown rate z_l^* given by our method. **d-f**, uniform lockdown rate z_l . **g-i**, random lockdown rate z_l . **j-l**, uniformly-bounded-decline lockdown rate z_l . In **a, d, g, j**, the disease parameters are set as in [10], the decay rate α is chosen as 0.0231 which corresponds to halving every 30 days. In **b, e, h, k** the disease parameters are set as in [9], the decay rate is chosen as $\alpha = 0.2r^S = 0.0034$ so that $\alpha < \min(r^A, r^S)$. In **c, f, i, l**, the disease parameters are set as in [11], the decay rate α is chosen 0.0231 that corresponds to halving every 30 days. It can be seen from **a-c** that the value of z_l^* for counties in NYC are relatively higher than other counties in New York State, which implies we should shutdown the outside of NYC harder than itself. Besides, it can be seen that such counter-intuitive phenomenon does not appear in any other lockdown policies.

the result, as does randomly setting half of the τ_{ij} to zero (see SI Fig. 18 of [34]).

We investigated this finding further in SI Sec. 8 of [34], where we let the metric to be optimized be the total number of infections. As a counterpoint, we do a greedy search over all two-parameter lockdowns that shut down NYC harder than the rest of New York State. Our results show that our lockdown has a smaller number of infections than the best two-parameter lockdown of the same cost.

Furthermore, in SI Sec. 12 of [34] we performed a robustness check of this finding in models where the disease spreads faster urban vs rural environments as a power of the population density. We presented these results by focusing on the most and least dense counties in New York, namely New York County and Hamilton county; there is a multiplicative factor of 30,000 difference in the densities between them. We considered models where the quantities β^A, β^S scale with a

power h of density. We found that in models where spread scales with density, our finding still holds if the epidemic spreads 2.8 times faster in New York County as compared with Hamilton County (corresponding to $h = 0.1$). By contrast, as we substantiate in the same SI section of [34], there does not appear to be a correlation between the density of a county in New York State and the growth rate of COVID-19. In particular, in real-world data the growth rate of COVID-19 has not been systematically higher in New York County as compared with Hamilton County over the lifetime of the epidemic.

Moreover, in SI Sec. 14 of [34], we considered whether this finding still holds if symptomatic individuals reduce their travel relative to asymptomatic individuals. We considered scenarios where the travel reduction is 10%, 50%, 90%, and we found that this finding remains in all of these cases.

The phenomenon in question likely occurs because shutting down a suburb with small population yields benefits proportional to the much larger population of the city, since a shutdown in the suburb affects the rate at which infection spreads in the city as city residents can infect each other through the suburb. Similarly, a possible explanation for the phenomenon we observe on New York State data is that shutdowns outside of NYC may be a cheaper way to curb the spread of infection within NYC.

We stress that this effect is due to the network interactions. In particular, this counterintuitive phenomenon does not occur in a hypothetical model of New York State where residents always stay within their own county: in that case, the lockdown problem reduces to a collection of single-node models which do not interact, and optimal stabilizing shutdown will be increasing in the proportion of infected (and insensitive to population), thus hitting NYC harder than the rest of New York State.

Additional observations. To fully understand the effect of the disease-related parameters to the optimal stabilizing lockdown-rate profile $\{z_l^*\}$ and the economic cost, we implemented additional numerical experiments to analyze the sensitivity. We plotted the obtained z_l^* with respect to degree, home-stay rate, population, and employment in SI Fig. 7 of [34]. We also implemented random permutation experiments (where we randomly permute one parameter while fixing everything else) in terms of degree, home-stay rate, population, employment and initial susceptible rate, and the results are shown in SI Fig. 8-9 of [34]. From these experiments, we found the distribution of z_l^* can be strongly affected by permutations of centrality, population, and the home stay rate. However, the distribution of z_l^* is not altered much by permuting employment and the initial susceptible rate. More details about these experiments are presented in SI Sec. 6 of [34].

From the empirical analysis of data in New York State, we hypothesize that the home-stay rate, degree centrality, and population are three major parameters that impact the optimal lockdown rate z_l^* of county- l . However, no inferences can be made about the effect of these parameters from empirical data because all of them vary together. To study how the value of z_l^* is related to these parameters, we implement experiments

on synthetically generated data. We describe the experiments and results next.

Impact of degree centrality. To study the impact of degree centrality, we considered geometric random graphs [53] (see SI Sec. 6.5 of [34] for other graphs). The population, the home-stay rate, and the initial susceptible rate of different nodes are set as the same values across all the nodes. The simulation results are presented in Fig. 4a-b. We found that degree centrality only matters for the value of z_l^* when there exist hotspots (i.e., hubs node with very high degrees). Beyond such hotspots, the effect of degree centrality is essentially ignorable.

Impact of population. To study the impact of population, we fix all other model parameters and vary the population of the nodes. We again considered geometric random graphs, where node degrees are similar. The simulation results are presented in Fig. 4c-d. We found that nodes with small populations are assigned smaller values of z_l^* , but once the population is large enough, z_l^* is almost independent of the population size.

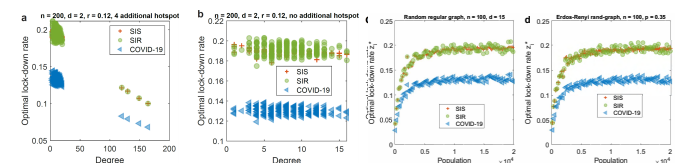


Fig. 4 Numerical results of optimal lockdown rates on synthetic networks. **a-b, the impact of the degree centrality** on the geometric random graph. It can be observed that centrality only matters for the value of z_l^* when there exist hotspots in all the three models. Without such hotspots, the effect of degree centrality is essentially ignorable. **c-d, the impact of population.** Nodes with small populations are assigned smaller values of z_l^* . Surprisingly, once the population is big enough, it doesn't affect z_l^* too much. In these figures, each point represents a node in the network.

V. CONCLUSION AND DISCUSSION

The main contribution of this paper is two-fold. Our first contribution is methodological: we give a modeling framework that gives rise to efficient methods for pandemic control through fixed lockdowns, with our main algorithm reduces the problem to matrix balancing, which, under certain assumption, can be solved in nearly linear time (and otherwise in matrix multiplication time). The favorable scaling of these methods allows us to scale up in a way that is not known for any other method. For example, at present data is not available to design lockdowns at the city level, but the method presented here can be scaled up to design a lockdown even at the neighborhood level for the entire United States. This could potentially be of use in future pandemics.

Our second contribution is to use our algorithms to observe several counter-intuitive properties of lockdowns. In particular, we observe that a model of epidemic spread in New York State will tend to shut down outside of NYC more stringently than NYC itself, even if the epidemic is largely localized to NYC. We compared the lockdown found by our model against exhaustive search of all two-parameter lockdowns which shut

down NYC harder than the rest of New York State to verify that indeed it outperforms. We further found that this result is robust against significant perturbations to the travel matrix, the epidemic parameters, as well as the epidemic model.

We caution that our results should not be viewed as an immediate policy recommendation. Indeed, in studying properties of the lockdowns that drive infections to zero at the fastest rate, we have not modeled a number of important factors. For example, hospital and ICU capacity in each county is important, as well as the ability of ambulances to shuttle patients between hospitals in neighboring counties, should some be overwhelmed. Additionally, if counties differ significantly in testing capabilities, this may affect transmission rates if a substantial and geographically varying number of patients know they are infected and follow guidelines to self-isolate. We have also not attempted to model substitution effects, where stringent lockdowns in some counties may cause people to increase their travel to counties with laxer restrictions.

Finally, we remark that since this paper was put online as a preprint, its results have been generalized to be applicable to general nonlinear networked systems in [54].

REFERENCES

- [1] Johns Hopkins University (JHU). Covid-19 dashboard. <https://coronavirus.jhu.edu/map.html>, 2020.
- [2] Pew Research Center. More than nine-in-ten people worldwide live in countries with travel restrictions amid COVID-19. <https://www.pewresearch.org/fact-tank/2020/04/01/more-than-nine-in-ten-people-worldwide-live-in-countries-with-travel-restrictions-amid-protect/unhbox/voidb@x\hbox{COVID}-19/>, 2020.
- [3] The New York Times. COVID-19 live updates: U.S. hospitalizations top 61,000, a record. <https://www.nytimes.com/live/2020/11/10/world/covid-19-coronavirus-live-updates?type=styl%20live-updates&label=virus&index=0&action=click&module=Spotlight&pgtype=Homepage#research-using-spring-cellphone-data-in-10-us-cities-could-help-influence-officials-facing-rising-cases-and-possible-restriction/>, 2020.
- [4] Matteo Chinazzi, Jessica T Davis, Marco Ajelli, Corrado Gioannini, Maria Litvinova, Stefano Merler, Ana Pastore y Piontti, Kunpeng Mu, Luca Rossi, Kaiyuan Sun, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*, 368(6489):395–400, 2020.
- [5] An Pan, Li Liu, Chaolong Wang, Huan Guo, Xingjie Hao, Qi Wang, Jiao Huang, Na He, Hongjie Yu, Xihong Lin, et al. Association of public health interventions with the epidemiology of the COVID-19 outbreak in Wuhan, China. *Jama*, 323(19):1915–1923, 2020.
- [6] Seth Flaxman, Swapnil Mishra, Axel Gandy, H Juliette T Unwin, Thomas A Mellan, Helen Coupland, Charles Whittaker, Harrison Zhu, Tresnia Berah, Jeffrey W Eaton, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, 584(7820):257–261, 2020.
- [7] Timothy C Germann, Kai Kadau, Ira M Longini, and Catherine A Macken. Mitigation strategies for pandemic influenza in the United States. *Proceedings of the National Academy of Sciences*, 103(15):5935–5940, 2006.
- [8] SafeGraph. Social Distancing Metrics. <https://docs.safegraph.com/docs/social-distancing-metrics>, 2020.
- [9] Giulia Giordano, Franco Blanchini, Raffaele Bruno, Patrizio Colaneri, Alessandro Di Filippo, Angela Di Matteo, and Marta Colaneri. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nature Medicine*, 26(6):855–860, 2020.
- [10] Andrea L. Bertozzi, Elisa Franco, George Mohler, Martin B. Short, and Daniel Sledge. The challenges of modeling and forecasting the spread of COVID-19. *Proceedings of the National Academy of Sciences*, 117(29):16732–16738, 2020.
- [11] John R Birge, Ozan Candogan, and Yiding Feng. Controlling epidemic spread: Reducing economic losses with targeted closures. *University of Chicago, Becker Friedman Institute for Economics Working Paper*, (2020-57), 2020.
- [12] Jayson S Jia, Xin Lu, Yun Yuan, Ge Xu, Jianmin Jia, and Nicholas A Christakis. Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature*, 582:389–394, 2020.
- [13] Christian Bongiorno and Lorenzo Zino. A multi-layer network model to assess school opening policies during the covid-19 vaccination campaign. *arXiv preprint arXiv:2103.12519*, 2021.
- [14] Lorenzo Zino and Ming Cao. Analysis, prediction, and control of epidemics: A survey from scalar to dynamic network models. *arXiv preprint arXiv:2103.00181*, 2021.
- [15] Francesco Parino, Lorenzo Zino, Maurizio Porfiri, and Alessandro Rizzo. Modelling and predicting the effect of social distancing and travel restrictions on covid-19 spreading. *arXiv preprint arXiv:2010.05968*, 2020.
- [16] Cameron Nowzari, Victor M Preciado, and George J Pappas. Analysis and control of epidemics: A survey of spreading processes on complex networks. *IEEE Control Systems Magazine*, 36(1):26–46, 2016.
- [17] Tom Britton, Frank Ball, and Pieter Trapman. A mathematical model reveals the influence of population heterogeneity on herd immunity to SARS-CoV-2. *Science*, 369(6505):846–849, 2020.
- [18] Ali Khanafer and Tamer Basar. On the optimal control of virus spread in networks. In *2014 7th International Conference on Network Games, Control and Optimization (NetGCoop)*, pages 166–172. IEEE, 2014.
- [19] Wolfgang Bock and Yashika Jayathunga. Optimal control and basic reproduction numbers for a compartmental spatial multipatch dengue model. *Mathematical Methods in the Applied Sciences*, 41(9):3231–3245, 2018.
- [20] Raffaele Carli, Graziana Cavone, Nicola Epicoco, Paolo Scarabaggio, and Mariagrazia Dotoli. Model predictive control to mitigate the covid-19 outbreak in a multi-region scenario. *Annual Reviews in Control*, 2020.
- [21] Daron Acemoglu, Victor Chernozhukov, Iván Werning, and Michael D Whinston. Optimal targeted lockdowns in a multi-group SIR model. *NBER Working paper*, (27102), 2020.
- [22] Pablo Fajgelbaum, Amit Khandelwal, Woookun Kim, Cristiano Mantovani, and Edouard Schaal. Optimal lockdown in a commuting network. *CEPR Discussion Papers*, (14923), 2020.
- [23] Fernando E Alvarez, David Argente, and Francesco Lippi. A simple planning problem for COVID-19 lockdown. *CEPR Discussion Paper*, (DP14658), 2020.
- [24] Alessandro Borri, Pasquale Palumbo, Federico Papa, and Corrado Possieri. Optimal design of lock-down and reopening policies for early-stage epidemics through sir-d models. *Annual Reviews in Control*, 2020.
- [25] Gery P Guy Jr, Florence C Lee, Gregory Sunshine, Russell McCord, Mara Howard-Williams, Lyudmyla Kompaniyets, Christopher Dunphy, Maxim Gakh, Regen Weber, Erin Sauber-Schatz, et al. Association of state-issued mask mandates and allowing on-premises restaurant dining with county-level covid-19 case and death growth rates? united states, march 1–december 31, 2020. *Morbidity and Mortality Weekly Report*, 70(10):350, 2021.
- [26] Fangzhou Liu and Martin Buss. Optimal control for information diffusion over heterogeneous networks. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 141–146. IEEE, 2016.
- [27] Adil El-Alami Laaroussi, Mostafa Rachik, and Mohamed Elhia. An optimal control problem for a spatiotemporal sir model. *International Journal of Dynamics and Control*, 6(1):384–397, 2018.
- [28] Luca Bolzoni, Elena Bonacini, Cinzia Soresina, and Maria Groppi. Time-optimal control strategies in sir epidemic models. *Mathematical biosciences*, 292:86–96, 2017.
- [29] Michael McAsey, Libin Mou, and Weimin Han. Convergence of the forward-backward sweep method in optimal control. *Computational Optimization and Applications*, 53(1):207–226, 2012.
- [30] C. Nowzari, V. M. Preciado, and G. J. Pappas. Optimal resource allocation for control of networked epidemic models. *IEEE Transactions on Control of Network Systems*, 4(2):159–169, 2017.
- [31] Renato Pagliara and Naomi Ehrlich Leonard. Adaptive susceptibility and heterogeneity in contagion models on networks. *IEEE Transactions on Automatic Control*, 2020.
- [32] Lorenzo Zino, Alessandro Rizzo, and Maurizio Porfiri. On assessing control actions for epidemic models on temporal networks. *IEEE Control Systems Letters*, 4(4):797–802, 2020.
- [33] Odo Diekmann, Johan Andre Peter Heesterbeek, and Johan AJ Metz. On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology*, 28(4):365–382, 1990.
- [34] Qianqian Ma, Yang-Yu Liu, and Alex Olshevsky. Optimal lockdown for pandemic control. *arXiv preprint arXiv:2010.12923*, 2020.

- [35] Bnaya Gross and Shlomo Havlin. Epidemic spreading and control strategies in spatial modular network. *Applied Network Science*, 5(1):1–14, 2020.
- [36] Ira M Longini Jr. A mathematical model for predicting the geographic spread of new infectious agents. *Mathematical Biosciences*, 90(1-2):367–383, 1988.
- [37] Lisa Sattenspiel, Klaus Dietz, et al. A structured epidemic model incorporating geographic mobility among regions. *Mathematical biosciences*, 128(1):71–92, 1995.
- [38] Julien Arino and P. van den Driessche. A multi-city epidemic model. *Mathematical Population Studies*, 10(3):175–193, 2003.
- [39] Jude Bayham, Nicolai V Kuminoff, Quentin Gunn, and Eli P Fenichel. Measured voluntary avoidance behaviour during the 2009 A/H1N1 epidemic. *Proceedings of the Royal Society B: Biological Sciences*, 282(1818):20150814, 2015.
- [40] Philip E Pare. *Virus spread over networks: Modeling, analysis, and control*. PhD thesis, University of Illinois at Urbana-Champaign, 2018.
- [41] Elliott H Bussell, Ciara E Dangerfield, Christopher A Gilligan, and Nik J Cuniffe. Applying optimal control theory to complex epidemiological models to inform real-world disease management. *Philosophical Transactions of the Royal Society B*, 374(1776):20180284, 2019.
- [42] Fabio Della Rossa, Davide Salzano, Anna Di Meglio, Francesco De Lellis, Marco Coraggio, Carmela Calabrese, Agostino Guarino, Ricardo Cardona-Rivera, Pietro De Lellis, Davide Liuzza, et al. A network model of Italy shows that intermittent regional strategies can alleviate the covid-19 epidemic. *Nature communications*, 11(1):1–9, 2020.
- [43] Richard Sinkhorn and Paul Knopp. Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics*, 21(2):343–348, 1967.
- [44] Martin Idel. A review of matrix scaling and sinkhorn's normal form for matrices and positive maps. *arXiv preprint arXiv:1609.06349*, 2016.
- [45] Michael B Cohen, Aleksander Madry, Dimitris Tsipras, and Adrian Vladu. Matrix scaling and balancing via box constrained Newton's method and interior point methods. 1:902–913, 2017.
- [46] Francesco Bullo. *Lectures on network systems*. Kindle Direct Publishing, 2019.
- [47] Anders Rantzer. Distributed control of positive systems. *Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference*, pages 6608–6611, 2011.
- [48] Van Sy Mai and Abdella Battou. Asynchronous distributed matrix balancing and application to suppressing epidemic. In *2019 American Control Conference (ACC)*, pages 2177–2182. IEEE, 2019.
- [49] Van Sy Mai, Abdella Battou, and Kevin Mills. Distributed algorithm for suppressing epidemic spread in networks. *IEEE Control Systems Letters*, 2(3):555–560, 2018.
- [50] Bahman Kalantari, Leonid Khachiyan, and Ali Shokoufandeh. On the complexity of matrix balancing. *SIAM Journal on Matrix Analysis and Applications*, 18(2):450–463, 1997.
- [51] Arun Jambulapati, Yin Tat Lee, Jerry Li, Swati Padmanabhan, and Kevin Tian. Positive semidefinite programming: mixed, parallel, and width-independent. *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pages 789–802, 2020.
- [52] Victor Y Pan and Zhao Q Chen. The complexity of the matrix eigenproblem. *Proceedings of the thirty-first annual ACM symposium on Theory of computing*, pages 507–516, 1999.
- [53] Mathew Penrose et al. *Random geometric graphs*, volume 5. Oxford university press, 2003.
- [54] Ron Ofir, Francesco Bullo, and Michael Margaliot. Minimum effort decentralized control design for contracting network systems. *arXiv preprint arXiv:2203.10392*, 2022.
- [55] Pauline Van den Driessche and James Watmough. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Mathematical biosciences*, 180(1-2):29–48, 2002.
- [56] Kevin D Smith and Francesco Bullo. Convex optimization of the basic reproduction number. *arXiv preprint arXiv:2109.07643*, 2021.
- [57] Ana Lajmanovich and James A Yorke. A deterministic model for gonorrhea in a nonhomogeneous population. *Mathematical Biosciences*, 28(3-4):221–236, 1976.
- [58] Gunnar Aronsson and Ingvar Mellander. A deterministic model in biomathematics. asymptotic behavior and threshold conditions. *Mathematical Biosciences*, 49(3):207–222, 1980.
- [59] Hongbin Guo, Michael Y Li, and Zhisheng Shuai. Global stability of the endemic equilibrium of multigroup SIR epidemic models. *Canadian Applied Mathematics Quarterly*, 14(3):259–284, 2006.
- [60] Ali Khanafer, Tamer Basar, and Bahman Ghahesifard. Stability of epidemic models over directed graphs: A positive systems approach. *Automatica*, 74:126–134, 2016.
- [61] Wenjun Mei, Shadi Mohagheghi, Sandro Zampieri, and Francesco Bullo. On the dynamics of deterministic epidemic propagation over networks. *Annual Reviews in Control*, 44:116–128, 2017.
- [62] Feliks Ruvimovič Gantmacher. *Applications of the Theory of Matrices*. Courier Corporation, 2005.

APPENDIX

A. Connection with the reproduction number

We now provide a proof of a claim that we made in the body of the paper, namely that the constraint $\lambda_{\max}(M(t_0)) \leq -\alpha$ is equivalent to a constraint on the basic reproduction number $R(t_0)$.

Proposition 7: Given any $r \in [0, 1]$, one can find an α in the domain $\alpha \in [0, \min(r^s, \epsilon + r^a)]$ such that the constraints

$$R(t_0) \leq r$$

and

$$\lambda_{\max}(M(t_0)) \leq -\alpha$$

are equivalent. The converse is also true: given α in the above domain, one can find an $r \in [0, 1]$ so that the above two constraints are equivalent.

Proof: The argument essentially reprises the proof of Lemma 5. Indeed, in that lemma we considered the matrix

$$M(t_0) = \begin{pmatrix} \beta^a \text{diag}(s(t_0))A_z - \epsilon - r^a & \beta^s \text{diag}(s(t_0))A_z \\ \epsilon & -r^s \end{pmatrix}.$$

In Eq. (12), it was shown that the constraint

$$\lambda_{\max}(M(t_0)) \leq \alpha$$

was equivalent to the constraint

$$\lambda_{\max}(\text{diag}(s(t_0))A_z b_1(\alpha)) \leq 1.$$

Here we write the expression $b_1(\alpha)$, unlike in Eq. (12), this way to highlight the dependence on α . Using the definition of $b_1(\alpha)$ from that lemma, the last condition can be rewritten as

$$\begin{aligned} \rho(\text{diag}(s(t_0))A_z) &\leq b_1(\alpha)^{-1} \\ &= \frac{(\epsilon + r^a - \alpha)(r^s - \alpha)}{\beta^s \epsilon + \beta^a(r^s - \alpha)} \\ &= \frac{\epsilon + r^a - \alpha}{\beta^a + \beta^s \epsilon / (r^s - \alpha)} \end{aligned} \quad (21)$$

We thus have that an upper bound the condition $\lambda_{\max}(M(t_0)) \leq -\alpha$ is equivalent to the upper bound $\rho(\text{diag}(s(t_0))A_z) \leq f(\alpha)$, where $f(\alpha)$ is a monotonically decreasing function of α . We next argue that a similar finding holds for the constraint $R(t_0) \leq r$.

To do this, we need to express the reproduction number in terms of the matrix $M(t_0)$. We use an expression from [55] in the form discussed in the recent paper [56]. Indeed, Definition 3 of [56] shows³, referring to [55], that if we split

$$M(t_0) = F + V$$

³Specifically, to apply those results here, the variable y of [56] represents the susceptible agents, while the variable x of that paper collects all the asymptomatic and symptomatic locations. The function $f(x, y)$ of [56], representing the inflow of infections, is taken to be zero.

where

$$F = \begin{pmatrix} \beta^a \text{diag}(s(t_0))A_z & \beta^s \text{diag}(s(t_0))A_z \\ 0 & 0 \end{pmatrix}$$

and

$$V = \begin{pmatrix} -\epsilon - r^a & 0 \\ \epsilon & -r^s \end{pmatrix}$$

then

$$R(t_0) = -\rho(FV^{-1}).$$

However, we have already computed this spectral radius in Eq. (11):

$$\begin{aligned} R(t_0) &= \rho(\text{diag}(s(t_0))A_z)b_1(0) \\ &= \rho(\text{diag}(s(t_0))A_z) \frac{\beta^s \epsilon + \beta^a r^s}{(\epsilon + r^a)r^s} \end{aligned}$$

Thus the constraint

$$R(t_0) \leq r$$

is equivalent to

$$\rho(\text{diag}(s(t_0))A_z) = r \frac{(\epsilon + r^a)r^s}{\beta^s \epsilon + \beta^a r^s} \quad (22)$$

We have thus shown that the constraint $R(t_0) \leq r$ is equivalent to the condition $\rho(\text{diag}(s(t_0))A_z) \leq g(r)$, where $g(r)$ is a monotonic function of r . Since both constraint on $R(t_0)$ and $\lambda_{\max}(M(t_0))$ are equivalent to constraints on $\rho(\text{diag}(s(t_0))A_z)$, they are equivalent to each other.

Finally, inspecting Eq. (21) and Eq. (22), we see that $\alpha = 0$ corresponds to $r = 1$, which is as one might expect. On the other hand, we see that as α approaches the endpoint of its domain, $\min(r^s, \epsilon + r^a)$, we have that the corresponding r approaches zero. This concludes the proof. ■

B. Results for The Network SIS Model

The SIS model differs from the COVID-19 model described in the main text or its special cases because it allows recovered individuals to get infected. An analogous form of our main result holds for the SIS model, as we describe next.

1) Problem formulation: The network SIS Model is described by the following set of ordinary differential equations

$$\dot{x}_i = (1 - x_i) \sum_{j=1}^n \beta a_{ij} x_j - \gamma x_i, \quad i = 1, \dots, n. \quad (23)$$

Here β denotes the transmission rate, which captures the rate at which an infected individual infects others, γ denotes the recovery rate, and a_{ij} captures the rate at which infection flows from the population at location j to location i .

For simplicity of notation, we can stack up the coefficients a_{ij} into a matrix as $A = [a_{ij}]$. Then we can write the network SIS model as

$$\dot{x} = \text{diag}(\mathbf{1} - x)\beta Ax - \gamma x,$$

where $\mathbf{1}$ denotes the vector of all-ones while $\text{diag}(u)$ makes a diagonal matrix out of the vector u .

It is naturally desirable to have $x_i(t) \rightarrow 0$ for all $i = 1, \dots, n$, i.e., to have the infection die out. This happens if and only if the matrix $\beta A - \gamma I$ is continuous-time stable [57]–[61]. To achieve an exponential decay rate α of each $x_i(t)$ (i.e., to be

bounded asymptotically by a multiple of $e^{-\alpha t}$), it is **necessary and sufficient** to have $\lambda_{\max}(\beta A - \gamma I) \leq \alpha$ (see formal proof in [34]). Note that even though the network SIS dynamics is nonlinear, the asymptotic stability is nevertheless equivalent to a linear eigenvalue condition.

Thus the optimal fixed lockdown condition is to find a vector $z \in [0, 1]^n$ such that $\lambda_{\max}(A_z) \leq \alpha$, where A_z is defined identically to the main body of the paper in Eq. (7).

2) Our results: We first state the form of the high-spread assumption for the SIS case.

Assumption 8 (High spread assumption): In the network SIS model, we have $\text{diag}(B^T C) \geq \gamma$

Our main result for the SIS case is then as follows.

Theorem 9: Suppose the graph corresponding to positive entries of the matrix A is strongly connected and $s(t_0) > 0$. Then:

- 1) Suppose Assumption 8 holds. Then the lockdown problem for the compartmental SIS dynamics is reducible to matrix balancing.
- 2) If Assumption 1 does not hold, then the lockdown problem for the compartmental SIS dynamics can be solved by computing an eigendecomposition and then performing $\tilde{O}(n^\omega)$ operations, where ω is the exponent of matrix multiplication and the $\tilde{O}(\cdot)$ notation hides factors which are logarithmic in the remaining problem parameters.

3) Proof Sketch of Theorem 9: The proof of Theorem 9 overlaps quite a bit with the proof of Theorem 2 in the main body of our paper. Indeed, almost all of the proof can be followed verbatim for the network SIS case. The only part that cannot be is Lemma 5, which explicitly references to the COVID-19 dynamics. We next give a proof of a version of this lemma for the SIS case.

Lemma 10: Suppose $A = CB^T$ is strongly connected. The minimum cost lockdown problem for the SIS network can be written as stability scaling. Under Assumption 8, the constrained lockdown problem for the network SIS problem can also be written as stability scaling.

Proof: We consider the unconstrained lockdown problem first. Recall that, in the SIS model, we are looking for a minimum cost positive vector z such that

$$C \text{diag}(z) B^T - (\gamma - \alpha) I \quad (24)$$

is continuous-time stable. Since the nonzero eigenvalues of a product of two matrices do not change after we change the order in which we multiply them, this is the same as requiring that

$$\text{diag}(z) B^T C - (\gamma - \alpha) I$$

is continuous-time stable. This is exactly the stability scaling problem provided we have two additional conditions. The first condition is that $\alpha < \gamma$ (because the diagonal matrix subtracted needs to be positive). The second condition is that $B^T C$ should be strongly connected. But by Lemma 4, the second condition is true because we assumed that $A = CB^T$ is strongly connected.

Observe that the reduction resulted in an instance of stability scaling with matrix $P = B^T C$. Since we have assumed $A = CB^T$ is strongly connected, we can apply Lemma 4 to obtain that P is strongly connected as needed. Under the assumption $\alpha < \gamma$, the matrix $(\gamma - \alpha)I$ is a positive diagonal matrix as required.

We thus have the reduction we want, from optimal lockdown to stability scaling, assuming $\alpha < \gamma$. But what if $\alpha \geq \lambda$? In that case, we claim that the minimum cost lockdown problem does not have a solution. Indeed, since at the optimal solution we must have all $z_i^* > 0$, we have that $C\text{diag}(z^*)B^T$ is an irreducible nonnegative matrix and its Perron-Frobenius eigenvalue is strictly positive by the Perron-Frobenius theorem [62]. Consequently, $C\text{diag}(z^*)B^T - (\gamma - \alpha)I$ has a positive eigenvalue and cannot be continuous-time stable.

Finally, we consider the constrained version. We argue that, under Assumption 8, the optimal solution to the constrained lockdown problem will have $z_i^* \leq 1$ for all i , so we can simply drop the constraint. Indeed, first observe that we can assume $\alpha \leq \gamma$, else the problem does not have a solution as explained above. Now suppose that $z_j^* > 1$; then $\text{diag}(z^*)B^T C - \text{diag}(\gamma - \alpha)I$ has a nonnegative j 'th row with at least one positive entry in that row. Indeed, the off-diagonal entries in the j 'th row are clearly nonnegative, while the diagonal entry is nonnegative by Assumption 1.

By Lemma 3, $\text{diag}(z^*)B^T C - \text{diag}(\gamma - \alpha)I$ cannot be continuous-time stable. As matrix $C\text{diag}(z)B^T - (\gamma - \alpha)I$ has the same nonzero eigenvalues as the matrix $\text{diag}(z^*)B^T C - \text{diag}(\gamma - \alpha)I$, it can not be stable either. ■

With this lemma established, all other parts of Theorem 2 can be followed verbatim, and the proof of Theorem 9 is complete.



Yang-Yu Liu is currently an Associate Professor at Harvard Medical School (HMS) and an Associate Scientist at Brigham and Women's Hospital (BWH). He received his PhD in Physics from University of Illinois at Urbana-Champaign in 2009, with thesis research focusing on phase transitions in disordered magnets. After that, he held positions as Postdoctoral Research Associate and then Research Assistant Professor in the Center for Complex Network Research at Northeastern University, before he joined HMS and BWH in 2013.

The primary goal of his postdoctoral research has been to combine tools from control theory, network science and statistical physics to address fundamental questions pertaining to the control of complex networks. His current research efforts focus on the study of human microbiome from the community ecology, dynamic systems, control theory, and machine learning perspectives.



Alex Olshevsky received the B.S. degree in applied mathematics and the B.S. degree in electrical engineering from the Georgia Institute of Technology, Atlanta, GA, USA, both in 2004, and the M.S. and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2006 and 2010, respectively. He was a postdoctoral scholar at Princeton University from 2010 to 2012, and an Assistant Professor at the University of Illinois at Urbana-Champaign from 2012 to 2016. He is currently an Associate Professor with the ECE department at Boston University.

Dr. Olshevsky is a recipient of the NSF CAREER Award, the Air Force Young Investigator Award, the INFORMS Computing Society Prize for the best paper on the interface of operations research and computer science, a SIAM Award for annual paper from the SIAM Journal on Control and Optimization chosen to be reprinted in SIAM Review, and an IMIA award for best paper on clinical medical informatics in 2019.



Qianqian Ma received the B.Eng. degree and the M.Eng. degree in control science and engineering from Harbin Institute of Technology, Harbin, China in 2014 and 2017, respectively. She is currently pursuing the Ph.D. degree at the Department of Electrical and Computer Engineering, Boston University, Boston, USA. Her research interests include distributed optimization, machine learning, and reinforcement learning.