Convex Optimization of the Basic Reproduction Number

Kevin D. Smith, Student Member, IEEE, and Francesco Bullo, Fellow, IEEE

Abstract—The basic reproduction number $R_0$ is a fundamental quantity in epidemiological modeling, reflecting the typical number of secondary infections that arise from a single infected individual. While $R_0$ is widely known to scientists, policymakers, and the general public, it has received comparatively little attention in the controls community. This note provides two novel characterizations of $R_0$: a stability characterization and a geometric program characterization. The geometric program characterization allows us to write $R_0$-constrained and budget-constrained optimal resource allocation problems as geometric programs, which are easily transformed into convex optimization problems. We apply these programs to allocating vaccines and antidotes in numerical examples, finding that targeting $R_0$ instead of the spectral abscissa of the Jacobian matrix (a common target in the controls literature) leads to qualitatively different solutions.

Index Terms—Compartmental models, convex optimization, epidemics, geometric programming, optimal resource allocation.

I. INTRODUCTION

Perhaps the most important parameter in an epidemic is the basic reproduction number. This number, denoted $R_0$, is the number of secondary infections that arise from a typical infected individual within an otherwise completely susceptible population. $R_0$ is a widely-known term, especially since 2020, when articles with “$R_0$” in the title ran in mainstream publications like The New York Times and The Wall Street Journal. Since $R_0$ is an intuitive and widely-known quantity, one might also expect it to appear frequently in the controls literature on epidemics, but this is not the case.

Instead, the literature tends to focus on two other major approaches to epidemic control. First, in the optimal control framework, parameters or control inputs are chosen to minimize some cost function integrated along the model trajectory [1], [2], [3], [4]. These trajectories seldom admit closed-form solutions, so this approach generally requires model-specific analysis and numerical solutions of Pontryagin’s conditions [1], [2], potentially large-scale optimization to embed discrete-time dynamics [3], or linearization and a discount factor to ensure convergence [4].

The second major approach is the spectral optimization framework, in which resources are allocated to minimize the spectral abscissa of the model’s Jacobian matrix about some disease-free equilibrium [5], [6], [7], [8], [9]. If the Jacobian is stable, then the abscissa represents the rate at which the trajectory converges towards this equilibrium; so minimizing the (negative) abscissa leads to a faster-decaying epidemic. Spectral optimization is based on a linear approximation of the model, but it is nonetheless an appealing framework for resource allocation, since the spectral abscissa can be directly evaluated from model parameters (without computing a trajectory).

The spectral abscissa is closely related to $R_0$. They are equivalent threshold parameters for whether the epidemic spreads or decays: in compartmental epidemic models (under reasonable assumptions), the epidemic enters an exponential growth phase if and only if the abscissa is positive, if and only if $R_0 > 1$ [10]. Furthermore, intuitively, both quantities reflect the rate at which the epidemic spreads or decays. But it is important to note that the abscissa and $R_0$ are different quantities. In fact, through proper choice of infection and recovery rates in the Kermack–McKendrick SIR model, one can achieve any pair of values for the abscissa $\alpha$ and reproduction number $R_0$ such that $R_0 > 0$ and $\text{sgn}(\alpha) = \text{sgn}(R_0 - 1)$. Thus, while the intuition for these two quantities is similar, minimizing the abscissa will generally lead to a different allocation of resources than minimizing $R_0$ directly.

To the best of authors knowledge, there is no work in the literature that focuses on directly minimizing or constraining $R_0$ in the resource allocation problem. Motivated by the ubiquity of $R_0$ in epidemiology and its popularity in the public discourse around COVID-19, this note provides theoretical foundations to fill in this gap.

Contributions: We propose a modification of the spectral optimization framework to operate on $R_0$ instead of on the spectral abscissa. We offer the following three primary contributions.

1) We provide two novel characterizations of $R_0$ in compartmental epidemic models. One characterization relates $R_0$ to the stability of perturbations to the Jacobian matrix, and the other expresses $R_0$ as a geometric program, which can be transformed into a convex optimization problem.

2) We define two $R_0$-based optimal resource allocation problems: the $R_0$-constrained allocation problem, which identifies the lowest-cost allocation to restrict $R_0$ below a given upper bound; and the budget-constrained allocation problem, which minimizes $R_0$ with a limited allowance for resource cost. We provide a geometric programming transcription for both of these problems, allowing them to be solved efficiently with off-the-shelf software.

3) We present numerical results based on a county-level multigroup susceptible-exposed-infected-removed (SEIR) model in California, parameterized using real-world cell phone mobility data. The experiments study the allocation of vaccines and antidotes, a classical problem in spectral optimization. We explain and emphasize the differences between the allocations based on $R_0$ and the corresponding allocations based on the abscissa.

Organization: Section II introduces the general family of compartmental epidemic models that we consider (Section II-A), formally defines $R_0$ (Section II-B), briefly reviews geometric programming (Section II-C), and states three key lemmas about Metzler and Hurwitz matrices (Section II-D). Section III presents our main theoretical results, including the two new characterizations of $R_0$ (Section III-A), and the two $R_0$-based optimal resource allocation problems and their geometric program transcriptions (Section III-B). Finally, Section IV presents the numerical experiments and Section V concludes this article.
**Notation:** The matrix $A \in \mathbb{R}^{n \times n}$ is *Metzler* if all its off-diagonal entries are nonnegative and is *Hurwitz* if all its eigenvalues have negative real part. Let $\rho(A)$ denote the spectral radius of $A$. Given $A \in \mathbb{R}^{n \times n}$, let $\text{diag}(A)$ denote the vector in $\mathbb{R}^n$ composed of the diagonal elements of $A$. Given $x \in \mathbb{R}^n$, let $\text{diag}(x)$ denote the diagonal matrix whose diagonal is $x$. Thus, $\text{diag}(\text{diag}(x)) = x$, and $\text{diag}(\text{diag}(A))$ is a copy of $A$ with all off-diagonal entries set to zero. Given a set $S$, we write $\text{cl}(S)$ to denote the closure of $S$.

## II. PRELIMINARIES

### A. Compartmental Epidemic Models

Compartmental models are a general and widely-used family of epidemic models that divide a population into compartments based on disease state and other demographic factors. This article focuses on deterministic epidemic models, in which the number of individuals in each compartment is governed by a system of differential equations. Perhaps the most well-known example is Kermack and McKendrick’s SIR model, which has three compartments (susceptible, infected, and recovered), but compartmental models can be arbitrarily complex to capture nuances in the spread of infection between different parts of the population in different disease states. Compartmental models are frequently based on an underlying stochastic model, such that the state variables approximate the expected number of individuals in each compartment.

We consider the general compartmental model in [10], with $n$ infected compartments and $m$ noninfected compartments. Let $x \in \mathbb{R}^n$ be the expected numbers of individuals in each infected compartment, and let $y \in \mathbb{R}^m$ be the expected numbers of noninfected individuals. The resulting dynamics are

$$
\dot{x} = f(x,y) + v(x,y) \quad (1a)
$$

$$
\dot{y} = g(x,y) \quad (1b)
$$

where $f$, $v$, and $g$ are continuously differentiable and defined on nonnegative domains. The dynamics of the infected subsystem are decomposed into two vector fields $f$ and $v$; where $f$ contains the rates at which new infections appear, and $v$ contains rates of transitions that do not correspond to new infections. For example, if infected individuals must pass through a latent disease state before entering an active infectious state (as in the SEIR model), then $f$ captures new infections as they appear in the latent state, while transitions from latent to active infections are contained in $v$, since the latter are not altogether new infections. This explicit separation of rates corresponding to new infections from all other transitions is crucial to the computation of $R_0$, and it reflects extra physical interpretation that cannot be inferred from the expression for $\dot{x}$ alone.

**Assumption 1 (Regularity of $f$, $v$, and $g$):** The vector fields $f$, $v$, and $g$ have the following properties:

1) $f(x,y) \geq 0$, for all $x$ and $y$;
2) $f(0_n, y) = 0_n$ and $v(0_n, y) = 0_n$ for all $y$;
3) for all $x$, $y$, and $i$, $x_i = 0$ implies that $v_i(x,y) \geq 0$;
4) for all $x$, $y$, and $j$, $y_j = 0$ implies that $g_j(x,y) \geq 0$.

Assumption 1 collects weak conditions that are obvious from the physical interpretations of $f$, $v$, and $g$. Condition 1 follows from the interpretation of $f$ as a rate at which new infections are created. Condition 2 ensures that no individuals can transfer into or out of an infected compartment (through new infections or otherwise) if the population is completely free of disease; thus every disease-free state is an equilibrium of (1a). Finally, conditions 3 and 4 reflect the fact that individuals cannot transition out from an empty compartment.

We also assume that (1) admit a disease-free equilibrium point $(0_n, y^*)$ that is locally asymptotically stable in the absence of new infections. That is, if new infections are “switched off” by dropping the vector field $f$ from the dynamics, then the population will return to $(0_n, y^*)$ even if a small number of infected individuals are introduced.

**Assumption 2 (Existence of a Stable Equilibrium):** There exists $y^* \geq 0_n$ such that $g(0_n, y^*) = 0_n$, and the following Jacobian matrix is Hurwitz:

$$
D \begin{bmatrix} v(0_n,y^*) \\ g(0_n,y^*) \end{bmatrix} = \begin{bmatrix} D_x v(0_n,y^*) & D_y v(0_n,y^*) \\ D_x g(0_n,y^*) & D_y g(0_n,y^*) \end{bmatrix}.
$$

The point $(0_n, y^*)$ satisfying Assumption 2 is not necessarily unique, and while it is also an equilibrium point of the full model, it may be unstable when $f$ is no longer ignored.

Under Assumptions 1 and 2, linearizing the dynamics of (1a) about $(0_n, y^*)$ decouples them from $y$, and we obtain

$$
\dot{x} = (F + V)x \quad (2)
$$

where $F = D_x f(0_n, y^*)$ is nonnegative and $V = D_y v(0_n, y^*)$ is Hurwitz and Metzler. We refer the reader to [10, Lemma 1] for details of this linearization.

### B. Basic Reproduction Numbers

The basic reproduction number is well-known in epidemiology as the typical number of secondary infections that arise from a single infected individual, within an otherwise completely susceptible population. Diekmann et al. [11] introduced the next generation operator to compute this quantity in general models with structured populations. This approach was later applied by Driessche and Watmough [10] specifically to the compartmental model (1).

**Definition 1:** For a compartmental epidemic model (1a) and (1b) satisfying Assumptions 1 and 2 and with linearization (2) about $(0_n, y^*)$, the basic reproduction number is

$$
R_0 = \rho(FV^{-1}). \quad (3)
$$

We refer the reader to [11, sec. 2] and [10, sec. 3] for derivations of (3) from the epidemiological definition of $R_0$.

### C. Geometric Programming

Geometric programs are a family of generally nonconvex optimization problems that can be transformed into convex optimization problems by a change of variables. Geometric programs enjoy a multitude of applications in engineering and control theory, including the design of optimal positive systems [12], a problem which is closely related to the resource allocation considered in this note. We refer the reader to [13] as a standard introduction to geometric programming and briefly introduce the key concepts in what follows.

A *monomial function* is a map $\mathbb{R}_+^n \rightarrow \mathbb{R}_+$ of the form $f(x) = x_1^{b_1} x_2^{b_2} \cdots x_n^{b_n}$, where $c > 0$ and $b_i \in \mathbb{R}$. A *posynomial function* is a sum of monomial functions. Note that posynomials are closed under addition and multiplication, and that a posynomial divided by a monomial is a posynomial. Given a posynomial function $f_0$, a set of posynomial functions $f_i, i \in \{1, \ldots, m\}$, and a set of monomial functions $g_i, i \in \{1, \ldots, p\}$, a geometric program in standard form is

$$
\begin{align*}
\text{minimize:} & \quad f_0(x) \\
\text{subject to:} & \quad f_i(x) \leq 1, \quad i \in \{1, \ldots, m\} \\
& \quad g_i(x) = 1, \quad i \in \{1, \ldots, p\}.
\end{align*}
$$

The problem becomes convex after the change of variables $x_i = e^{u_i}$. Off-the-shelf software is available for geometric programs, including the CVX package in MATLAB [14].
D. Properties of Hurwitz and Metzler Matrices

We now reproduce three lemmas regarding properties of Metzler and Hurwitz matrices that will be necessary for our main results. The first lemma is a standard result characterizing the stability of Metzler matrices (see [15, Th. 10.14]).

Lemma 1 (Metzler Hurwitz Lemma): Let $M \in \mathbb{R}^{n \times n}$ be a Metzler matrix. The following are equivalent:

1) $M$ is Hurwitz;
2) $M$ is invertible and $-M^{-1} \geq 0$;
3) there exists $w > 0_n$ such that $Mw < 0_n$.

We borrow the next two results from [10]; the first is a slight restatement of [10, Lemma 5], so we do not include a proof.

Lemma 2 (Properties of Hurwitz and Metzler Matrices): Let $H, M \in \mathbb{R}^{n \times n}$ be Metzler matrices, such that $H$ is Hurwitz and $-MH^{-1}$ is Metzler. The following are equivalent:

1) $M$ is Hurwitz;
2) $-MH^{-1}$ is Hurwitz.

The second result is abstracted from the proof of [10, Th. 2] and we include a self-contained proof.

Lemma 3 (Stability of Perturbed Metzler Matrices): Let $H \in \mathbb{R}^{n \times n}$ be Metzler and Hurwitz, and let $E \in \mathbb{R}^{n \times n}$ be a nonnegative perturbation matrix. The following are equivalent:

1) $H + E$ is Hurwitz;
2) $\rho(-EH^{-1}) < 1$.

Proof: Let $A = -(H + E)H^{-1} = -(I_n + EH^{-1})$. Note that $A$ is Metzler since $-H^{-1} \geq 0$ by Lemma 1, so $-EH^{-1} \geq 0$. Then, by Lemma 2, $H + E$ is Hurwitz if and only if $A$ is Hurwitz. If $\rho(-EH^{-1}) < 1$, then $A$ is clearly Hurwitz. But if $\rho(-EH^{-1}) \geq 1$, then $A$ is not Hurwitz: since $-EH^{-1} \geq 0$, the Perron-Frobenius theorem guarantees that its dominant eigenvalue is real and nonnegative, so $-(I_n + EH^{-1})$ has an eigenvalue with nonnegative real part. ■

III. OPTIMIZATION FRAMEWORK FOR $R_0$

A. Geometric Program for $R_0$

The main theoretical result of this article is the following theorem, which provides two novel characterizations of $R_0$.

Theorem 1 (Characterizations of $R_0$): Consider the linearized epidemiologic dynamics (2) with $F \in \mathbb{R}^{n \times n}$ and $V \in \mathbb{R}^{n \times n}$ Hurwitz and Metzler. Write $V = V_{dl} - V_d$, where $V_d \geq 0$ is diagonal and $V_{dl} \geq 0$ has zero diagonal. The following are characterizations of the basic reproduction number.

1) Stability characterization

$$R_0 = \inf_{r > 0} \{r : F + rV \text{ is Hurwitz}\}. \quad (4)$$

2) Geometric program characterization

$$R_0 = \inf_{w > 0_n} \{r : \text{diag}(rVw)^{-1}(F + rVw)w \leq \underline{1}_n\}. \quad (5)$$

Proof: To prove that (4) follows from (3), we compute

$$\inf_{r > 0} \{r : F + rV \text{ is Hurwitz}\} = \inf_{r > 0} \{r : \rho((F + rV)^{-1}) < 1\}$$

$$= \inf_{r > 0} \{r : \rho(FV^{-1}) < r\}$$

$$\geq \inf_{r > 0} \{r : R_0 < r\} = R_0$$

where the first step follows from Lemma 3. We now use (4) to prove (5).

Let $W = \{w > 0_n : Vw < 0_n\}$ and $\bar{W} = \{w > 0_n : Vw \leq 0_n\}$. By Lemma 5 (in Appendix A)

$$R_0 = \inf_{r > 0} \{r : F + rV \text{ is Hurwitz}\}$$

$$= \inf_{r > 0} \{r : \exists w \in W \text{ s.t. } (F + rV)w < 0_n\}$$

$$= \inf_{r > 0} \{r \geq 0 : 3w \in \bar{W} \text{ s.t. } (F + rV)w \leq 0_n\}$$

In the last step, we note that the $Vw \leq 0_n$ constraint is implied by $(F + rV)w \leq 0_n$, so we are free to remove it. Manipulating the $(F + rV)w \leq 0_n$ constraint into the standard form for geometric programming yields (5).

Remark 1 (Degenerate Cases, Pt. I): The infimum in (5) is not always attained. For example, if $F = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$ and $V = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, then $R_0 = \inf_{r > 0} \{r : r \geq \frac{-1}{w} \geq 1\} = 1$. But there is no feasible point $w > 0_2$ that satisfies the inequality constraint with $r = 1$. Thus, in general, we cannot replace the infimum in (5) with a minimum.

B. Optimal Resource Allocation

The geometric program characterization (5) sets us up to efficiently optimize model parameters to either minimize or constrain $R_0$. In a manner analogous to [5], we consider two forms of the resource allocation problem: $R_0$-constrained allocation, and budget-constrained allocation. In both forms of the resource allocation problem, we suppose that the model parameters $F, V_{dl}$, and $V_d$ depend on a vector of "resources" $\theta \geq 0_n$, and that the cost of a particular allocation of resources is given by a cost function $c(\theta)$. Furthermore, the resources must satisfy some collection of constraints $h(\theta) \leq \underline{1}_q$. The dependence on $\theta$ must obey the following conditions.

Assumption 3 (Resource Dependence): The resource dependence of the parameters $F(\theta), V_{dl}(\theta), V_d(\theta), c(\theta)$, and $h(\theta)$ have the following properties:

1) $F(\theta), V_{dl}(\theta), c(\theta)$, and $h(\theta)$ are element-wise posynomial functions.
2) $V_d(\theta)$ is an element-wise monomial function.
3) The set of feasible allocations $\{\theta \geq 0_n : h(\theta) \leq \underline{1}_q\}$ is bounded, and if $\theta$ is in this set, then $V_{dl}(\theta) - V_d(\theta)$ is Hurwitz.

Conditions 1 and 2 are necessary to transcribe the allocation problem as a geometric program, while condition 3 ensures that the matrix parameters $F, V_{dl}$, and $V_d$ satisfy the antecedent of Theorem 1 for any feasible allocation. Condition 3 also ensures the feasible $\theta$ are confined to a compact set. Under these assumptions, for all $\theta \in h^{-1}(\underline{1}_q)$, the resource dependence of $R_0$ can be written as

$$R_0(\theta) = \rho \left( F(\theta)(V_{dl}(\theta) - V_d(\theta))^{-1} \right). \quad (6)$$

Additional resources will typically reduce the rate of new infections or increase the rate at which existing infections are removed. This property is not included in Assumption 3, since it is not needed for any of the results in this section. However, if this property is true, then it is useful (albeit unsurprising) to note that $R_0(\theta)$ is weakly decreasing in $\theta$.

Lemma 4 (Monotonicity): Suppose that $F(\theta), V_{dl}(\theta)$, and $V_d(\theta)$ satisfy Assumption 3. If additionally $F(\theta)$ and $V_{dl}(\theta)$ are nonincreasing and $V_d(\theta)$ is nondecreasing in $\theta$, then for $\theta, \theta' \in h^{-1}(\underline{1}_q)$ with $\theta' \geq \theta$, we have $R_0(\theta') \leq R_0(\theta)$.

Proof: Let $\theta' \geq \theta$. Since $0 \leq F(\theta') \leq F(\theta), 0 \leq V_{dl}(\theta') \leq V_{dl}(\theta)$, and $V_d(\theta') \geq V_d(\theta)$, we can write $F(\theta') = F(\theta) + \Delta F$ and $V(\theta') = V(\theta) + \Delta V(\theta)$ for some matrices $\Delta F, \Delta V \geq 0$. Then

$$V^{-1}(\theta) - V^{-1}(\theta') = (V(\theta') + \Delta V)^{-1} - V^{-1}(\theta')$$

$$= -\left( (V(\theta') + \Delta V)^{-1} (\Delta V) V^{-1}(\theta') \right) \leq 0$$

where the last inequality follows from Lemma 1, since $V(\theta)$ and $V(\theta')$ are Hurwitz and Metzler, and thus $V^{-1}(\theta) \leq 0$ and $V^{-1}(\theta') \leq 0$. Then

$$-F(\theta)V^{-1}(\theta) = -F(\theta')V^{-1}(\theta') \geq 0$$

$$\geq -F(\theta')V^{-1}(\theta') \geq -F(\theta)V^{-1}(\theta').$$
Since $-F(\theta)V^{-1}(\theta) \geq 0$ and $-F'(\theta)V^{-1}(\theta') \geq 0$, we are guaranteed that

$$R_0(\theta) = \rho (-F(\theta)V^{-1}(\theta)) \geq \rho (-F'(\theta')V^{-1}(\theta')) = R_0(\theta')$$

since the spectral radius is weakly increasing in the elements of a nonnegative matrix [16, Th. 8.1.18].

We now define the two optimal allocation problems. In the $R_0$-constrained allocation problem, we identify the cheapest allocation of resources to ensure that $R_0 \leq r_{\text{max}}$, where $r_{\text{max}} > 0$ is some arbitrary threshold. In the budget-constrained allocation problem, some budget $c_{\text{max}} > 0$ is available to spend on resources, and we would like to deploy these limited resources to minimize $R_0$.

**Definition 2 (Optimal Allocation Problems):** Let $F(\theta), V_{\text{ad}}(\theta), V_0(\theta), c(\theta)$, and $h(\theta)$ satisfy Assumption 3. We define the following optimization problems.

1. Given $r_{\text{max}} > 0$, we say that $\theta^*$ is an optimal $R_0$-constrained allocation if $\theta^*$ is a minimizer of

$$\min_{\theta \geq \theta_k} \left\{ c(\theta) : h(\theta) \leq 1, \text{ and } R_0(\theta) \leq r_{\text{max}} \right\}. \quad (7)$$

2. Given $c_{\text{max}} > 0$, we say that $\theta^*$ is an optimal budget-constrained allocation if $\theta^*$ is a minimizer of

$$\min_{\theta \geq \theta_k} \left\{ R_0(\theta) : h(\theta) \leq 1, \text{ and } c(\theta) \leq c_{\text{max}} \right\}. \quad (8)$$

Assumption 3 ensures that $R_0(\theta)$ in (6) is well-defined over the feasible sets; furthermore, $R_0(\theta)$ is continuous, since the matrix inverse is a continuous function of its elements. Thus, the feasible sets are compact, so the minima of both problems exist.

Using Theorem 1, we can construct a pair of geometric programs to solve for optimal $R_0$-constrained and budget-constrained allocations. We define a map $p : \mathbb{R}_{\geq 0}^n \times \mathbb{R}_{\geq 0}^n \times \mathbb{R}_{\geq 0}^n \to \mathbb{R}_{\geq 0}$ by

$$p(r,w,\theta) = \text{diag} (r V_{\text{ad}}(\theta) w) - (F(\theta) + r V_{\text{ad}}(\theta)) w.$$ \quad (9)

Under Assumption 3, $p(r,w,\theta)$ is posynomial, so the following are geometric programs.

**Problem 1 ($R_0$-Constrained Allocation GP):** Given $r_{\text{max}} > 0$ and a tolerance parameter $\tau > 0$,

$$\text{minimize : } c(\theta)$$

$$\text{variables : } r > 0, \quad w > 0_n, \quad \theta > 0_k$$

$$\text{subject to : } p(r, w, \theta) \leq 1_n, \quad h(\theta) \leq 1_q, \quad r \leq r_{\text{max}} + \tau.$$ \quad (10)

**Problem 2 (Budget-Constrained Allocation GP):** Given $c_{\text{max}} > 0$,

$$\text{minimize : } r$$

$$\text{variables : } r > 0, \quad w > 0_n, \quad \theta > 0_k$$

$$\text{subject to : } p(r, w, \theta) \leq 1_n, \quad h(\theta) \leq 1_q, \quad c(\theta) \leq c_{\text{max}}.$$ \quad (11)

**Theorem 2 (Geometric Program Transcription):** Let $\theta^* \geq 0_k$, $r_{\text{max}} > 0$, and $c_{\text{max}} > 0$. Let $F_1(\tau)$ for $\tau > 0$ and $F_2$ be the sets of feasible points $(r, w, \theta)$ for Problems 1 and 2. The following are true.

1. $\theta^*$ is an optimal $R_0$-constrained allocation if and only if the infimum of Problem 1 converges to $c(\theta^*)$ as $\tau \to 0$, and there exists $r^*, w^*$ such that $(r^*, w^*, \theta^*) \in c(F_1(\tau))$ for all $\tau > 0$.

2. $\theta^*$ is an optimal budget-constrained allocation if and only if $R_0(\theta^*)$ is the infimum of Problem 2 and there exists $r^*, w^*$ such that $(r^*, w^*, \theta^*) \in c(F_2)$.

See Appendix B for the proof.

We note that Problem 1 is an arbitrarily accurate approximation of the $R_0$-constrained allocation problem, controlled by the parameter $\tau \geq 0$. This approximation is necessary due to the closed inequality constraint on $R_0$ and the representation of $R_0$ by the infimum in (5), which is not always attained.

**Remark 2 (Degenerate Cases, Pt. II):** In some cases, Problem 1 may be infeasible when $\tau = 0$, for example, if $F(\theta) = F$ and $V(\theta) = V$ are the matrices defined in Remark 1 and $r_{\text{max}} = 1$. Fortunately, the feasible set is nonempty for all $\tau > 0$, so we can still consider the limit of solutions to Problem 1 as $\tau \to 0_+$. This feasibility problem arises due to the constraint on $R_0$, so it is not an issue in Problem 2.

In practice, the issue of an empty feasible set is not of significant concern, since numerical optimization already has inherently limited precision. We suggest solving Problem 1 with $\tau = 0$ (and only using a small positive value if the solver reports primal infeasibility).

**IV. NUMERICAL EXAMPLES**

In the following experiments, we compare $R_0$-minimizing allocations with abscissa-minimizing allocations. The code used to generate these results is available online.\(^1\)

**A. Epidemic Model**

We adopt a standard multigroup SEIR model (with vital dynamics) for an epidemic in the state of California, where each group corresponds to one of the state’s $n = 58$ counties. The SEIR model has two infected states (exposed and infectious) and two noninfected states (susceptible and recovered). Letting $s, e, z, r \in \mathbb{R}_{\geq 0}^n$ denote the expected number of people in each group and disease state, the model dynamics for each group $i \in \{1, \ldots, n\}$ are

$$\dot{s}_i = -\beta_i s_i \sum_{j=1}^n a_{ij} z_j, \quad \dot{e}_i = \beta_i s_i \sum_{j=1}^n a_{ij} z_j - \gamma_i e_i, \quad \dot{z}_i = \gamma_i e_i - \delta_i z_i, \quad \dot{r}_i = \delta_i z_i.$$ \quad (12)

It is clear that the model has a disease-free equilibrium $(s_0, e_0, z_0, r_0)$. Linearizing about this point, we obtain

$$\begin{bmatrix} c \approx -\delta(\gamma) & \text{diag}(\beta) \text{diag}(s_0) A & \text{diag}(\beta) \text{diag}(s_0) \end{bmatrix} \begin{bmatrix} c \\ z \\ \tau \end{bmatrix}.$$ \quad (13)

Because the $\text{diag}(\beta) \text{diag}(s_0) A$ term is the only one corresponding to the creation of new infections, we decompose this Jacobian into the two matrices

$$F = \begin{bmatrix} 0 & \text{diag}(\beta) \text{diag}(s_0) A \\ 0 & 0 & \text{diag}(\gamma) & -\text{diag}(\beta) \end{bmatrix}, \quad V = \begin{bmatrix} -\text{diag}(\gamma) & 0 \\ 0 & -\text{diag}(\gamma) & -\text{diag}(\beta) \end{bmatrix}.$$ \quad (14)

where $F$ is nonnegative and $V$ is Hurwitz and Metzler.

The model requires a matrix of intergroup contact rates $A \in \mathbb{R}_{n \times n}$, which we estimated using data from SafeGraph.\(^2\) In particular, we used the Social Distancing Metrics dataset to estimate a matrix $P \in \mathbb{R}_{n \times n}$, where $p_{ij}$ is the daily fraction of people from county $i$ who visited county $j$, averaged over each day in 2020. Then, $(P P^T)_{ij}$ approximates the probability that two random individuals from counties $i$ and $j$ are...
colocated in the same county on a given day. We set \( A = \alpha P P^T \), where the scalar \( \alpha = 2.3667 \times 10^{-7} \) was chosen to ensure \( R_0 = 2.5 \) when \( \beta = 0.1, \gamma = 0.2, \) and \( \delta = 0.1 \). Note that \( \alpha \) is always multiplied by \( \beta \), so the only effect of this scalar is to allow us to work with round numbers for \( \beta \) and \( R_0 \).

The remaining model parameters are the transmission rates \( \beta > 0 \), incubation rates \( \gamma > 0 \), and recovery rates \( \delta > 0 \), for each group. We used uniform model parameters across each group for simplicity. We generated 2000 different models by choosing \( \beta, \gamma, \) and \( \delta \) for each of 10 \((\gamma)\) or 20 \((\beta)\) evenly spaced values in the range \([0.025, 0.5], [0.05, 0.5], \) and \([0.05, 0.5]\), respectively. The \( \gamma \) and \( \delta \) range was chosen to allow for a wide range of mean incubation and recovery times (between 2 and 20 days), while the \( \beta \) range was coarsely tuned so that the models have a wide but realistic range of pre-intervention \( R_0 \) (95% between 0.23 and 19.38).

B. Optimal Allocation of Pharmaceuticals

We consider the following optimal resource allocation scenario from [5], in which there are two types of pharmaceutical interventions: vaccines, which reduce the local transmission rates \( \beta_i \); and antidotes, which increase the local recovery rates \( \delta_i \). By allocating vaccines to patch \( i \), we can optimize the local transmission rate within a range \( \beta_i \in [\beta_L, \beta_U] \), where \( \beta_U > 0 \). The cost of this vaccine allocation is, for all \( i \)

\[
f_i(\beta_i) = \frac{\beta_i^{-1} - \beta_i^{-1}}{\beta_i^{-1} - \beta_i^{-1}}. \tag{10}
\]

Note that the most aggressive allocation has a cost of \( f_i(\beta) = 1 \), while allocation of no vaccines at all has a cost \( f_i(\beta) = 0 \). The form of (10) ensures diminishing returns in the investment of vaccines at each patch. Similarly, by allocating antidotes to patch \( i \), the local recovery rate can be optimized in the range \( \delta_i \in [\delta_L, \delta_U] \), with \( \delta_U > 0 \). The cost of the antidote allocation is, for all \( i \)

\[
g_i(\delta_i) = \frac{(\delta_i - \delta_i)^{-1} - (\delta_i - \delta_i)^{-1}}{(\delta_i^0 - \delta_i)^{-1} - (\delta_i^0 - \delta_i)^{-1}} \tag{11}
\]

where the parameters \( \delta_i^0 > \delta_i \) control the shape of the cost curve. The total cost, summing over the local costs of vaccines and antidotes over all patches, is constrained by a budget \( c_{\text{max}} \).

In order to perform budget-constrained resource allocation, we must encode the following budget constraint in the standard form for geometric programming:

\[
\sum_{i=1}^n f_i(\beta_i) + g_i(\delta_i) \leq c_{\text{max}}. \tag{12}
\]

Since \( g_i \) have nonposynomial dependence on \( \delta_i \), we replace \( 1 - \delta_i \) with auxiliary variables \( \eta_i \), constrained by \( \delta_i - \delta_i \leq \eta_i \leq \delta_i - \delta_i \). Then, the posynomial budget constraint is

\[
\sum_{i=1}^n \kappa^{-1} \beta_i^{-1} + \beta_i^{-1} \kappa^{-1} \eta_i^{-1} \leq 1 \tag{12}
\]

where we define a positive constant \( \kappa = c_{\text{max}} + \sum_{i=1}^n \beta_i^{-1} - \beta_i^{-1} \kappa^{-1} (\delta_i - \delta_i)^{-1} \kappa^{-1} \delta_i^{-1} - (\delta_i - \delta_i)^{-1} \delta_i^{-1} \kappa^{-1} \delta_i^{-1}. \]

Altogether, the resource vector is \( \theta^T = [\beta^T, \eta^T] \), and the constraints are \( \beta \leq \beta \leq \beta, \delta_i - \delta_i \leq \eta_i \leq \delta_i - \delta_i \), and (12).

For each experiment, we selected cost parameters based on the preintervention SEIR model parameters \( \beta \) and \( \delta \). Since pharmaceuticals and vaccines never increase the transmission rate or decrease the recovery rate, we set \( \beta_i = \beta_i \) and \( \delta_i = \delta_i \). We chose \( \beta_i = 0.1 \beta_i \) and \( \delta_i = 2 \delta_i \) to reflect a 90% reduction in transmissibility and 50% reduction in mean recovery time at maximum investment, and we selected \( \delta_i = 2 \) so that \( \delta_i > \delta_i \).

C. Results and Discussion

We first set a budget of \( c_{\text{max}} = 0.1 \) and performed budget-constrained resource allocation to minimize \( R_0 \) and the absicssa for each of the 2000 models. We then simulated the nonlinear postintervention dynamics for both the \( R_0 \)-minimized and absicssa-minimized models until convergence.

In 1270 models, both the \( R_0 \)-minimized and absicssa-minimized models had \( R_0 > 1 \), so the number of infected individuals experienced an initial exponential growth phase before peaking and decaying. Fig. 1 (left) compares the number of active infections at the peak between the \( R_0 \)-minimized and absicssa-minimized trajectories. In 1068 (84.1%) of these models, minimizing \( R_0 \) led to a smaller peak than minimizing the absicssa. Similarly, Fig. 1 (right) compares the number of cumulative infections at the end of the simulation. Minimizing \( R_0 \) resulted in fewer cumulative cases in 1056 (83.1%) in the example models. In the remaining models, one or both of the \( R_0 \)-minimizing or absicssa-minimizing allocations led to \( R_0 < 1 \), so the trajectory immediately decays toward a disease-free equilibrium. It is not meaningful to compare peaks in these models; however, in 96.4% of them, minimizing \( R_0 \) resulted in fewer cumulative infections.

Next, we selected three particular models to examine the allocations under various budgets. We chose a low-\( R_0 \) model (\( \beta = 0.05, \gamma = 0.2, \delta = 0.2; R_0 = 0.625 \)), a mid-\( R_0 \) model (\( \beta = 0.1, \gamma = 0.2, \delta = 0.1; R_0 = 2.5 \)), and a high-\( R_0 \) model (\( \beta = 0.15, \gamma = 0.2, \delta = 0.075; R_0 = 5.0 \)), and we repeated the budget-constrained allocations at various budgets. Fig. 2 (left) plots the cumulative infections for the postintervention models. Cumulative infections in the \( R_0 \)-minimized and absicssa-minimized models are very similar at low budgets, but past a budget of 2, minimizing the \( R_0 \) leads to a modest decrease in cumulative infections when compared to minimizing the absicssa. (It is not meaningful to plot the peak infections, since \( R_0 < 1 \) in all postintervention models with budgets above 2.) Fig. 2 (right) illustrates a difference in allocation strategies between the two targets, as minimizing \( R_0 \) results in greater spending on vaccines.
\[
\forall R \in \mathbb{R}^I \left( \max_{\hat{\imath} \in L_0} \kappa_0 = 0 \right) \implies \inf \{ r > 0 : (F + rV)w \leq \Omega_0 \} = \inf \{ r > 0 : (F + rV)\hat{\imath} \leq \Omega_0 \} = \inf \{ r > 0 : (F + rV)w \leq \Omega_0 \} = \inf \{ r > 0 : (F + rV)\hat{\imath} \leq \Omega_0 \} = \inf \hat{R}
\]

where \( \hat{R} = \{ r^*(w) : w \in \hat{W} \} \), and \( r^* : \hat{W} \rightarrow \mathbb{R}_{\geq 0} \) is the map defined by

\[
r^*(w) = \inf \{ r > 0 : (F + rV)w \leq \Omega_0 \} \quad \forall w \in \hat{W}.
\]

It is straightforward to solve for \( r^*(w) \)

\[
r^*(w) = \max_{\hat{\imath} \in I^F} \left( \frac{(Fw)_{\hat{\imath}}}{|Vw|_{\hat{\imath}}} \right) \quad \forall w \in \hat{W}.
\]

Similar to \( \hat{R}_0 \), we have the following expression for \( R_0^* \):

\[
R_0^* = \inf \{ r > 0 : (F + rV)w \leq \Omega_0 \}
\]

\[
= \inf_{w \in \hat{W}} \left( \inf \{ r > 0 : (F + rV)w \leq \Omega_0 \} \right)
\]

\[
= \inf_{w \in \hat{W}} \left( \inf \{ r > 0 : (F + rV)w \leq \Omega_0 \} \right)
\]

\[
= \inf_{w \in \hat{W}} \left( \inf \{ r > 0 : (F + rV)w \leq \Omega_0 \} \right) = \inf \hat{R}
\]

where \( \hat{R} = \{ r^*(w) : w \in \hat{W} \} \). The remainder of the proof is to show that \( R_0^* \) is a lower bound on \( \hat{R} \).

Let \( \hat{r} \in \hat{R} \), so that \( \hat{r} > 0 \) and \( (F + \hat{r}V)\hat{w} \leq \Omega_0 \) for some \( \hat{w} \in \hat{W} \). Let \( x > 0 \), such that \( Vx < \Omega_0 \) (which must exist because \( V \) is Hurwitz), and for all \( t \geq 0 \), let \( w(t) = \hat{w} + tx \). We can also show that

\[
\frac{(Fx)_{\hat{\imath}}}{|Vw(t)|} + \frac{(F\hat{w})_{\hat{\imath}}}{|V\hat{w}|} > 0 \quad \forall \hat{\imath} \in I^F.
\]

Then, for all \( t > 0 \)

\[
|\hat{r}^*(w(t))| - |\hat{r}^*(\hat{w})| = \max_{\hat{\imath} \in I^F} \left( \frac{(Fx)_{\hat{\imath}}}{|Vw(t)|} \right) - \max_{\hat{\imath} \in I^F} \left( \frac{(F\hat{w})_{\hat{\imath}}}{|V\hat{w}|} \right) 
\]

\[
\leq \max_{\hat{\imath} \in I^F} \kappa_{\hat{\imath}} t.
\]

Thus, given any \( \epsilon > 0 \), we can choose \( t < \epsilon (\max_{\hat{\imath} \in I^F} \kappa_{\hat{\imath}})^{-1} \) to ensure that \( |\hat{r}^*(w(t))| - |\hat{r}^*(\hat{w})| < \epsilon \). Because \( w(t) \in \hat{W} \) for all \( t > 0 \), it is the case that \( \hat{r}^*(w(t)) \in \hat{R} \) for all \( t > 0 \), so that every open ball around \( \hat{r}^*(\hat{w}) \) contains a point in \( \hat{R} \). Then, \( \hat{r}^*(\hat{w}) \in \epsilon(\hat{R}) \), which implies that \( \hat{r}^*(\hat{w}) \geq R_0 \). But \( \hat{r}^*(\hat{w}) \leq \hat{r} \), and \( \hat{r} \) was chosen arbitrarily from \( \hat{R} \), so \( R_0 \) is a lower bound on \( \hat{R} \). But \( \hat{R}_0 \) is the greatest such lower bound, so we conclude that \( R_0 \geq R_0^* \).
Since the determinant of $M(r, \theta)$ is a polynomial in $r$ of degree $n$, for some scalars $a_1, a_2, \ldots, a_n \in \mathbb{C}$, we can write $|M(r, \theta)| = (r - a_1)(r - a_2) \cdots (r - a_n)$. Due to (4) in Theorem 1, $M(R_0(\theta), \theta)$ must be singular, so $R_0(\theta)$ is a root; then we can assign $a_1, a_2, \ldots, a_n = R_0(\theta)$ up to some multiplicity $\ell$. Define a “pseudodeterminant” $\mu(r, \theta) = (r - a_{1,z})(r - a_{2,z}) \cdots (r - a_{n,z})$ as the product of the remaining factors, which is real and nonzero for all $r > R_0(\theta)$. Then

$$M^{-1}(r, \theta) = \frac{\text{ad}_{[\mu]}(M(r, \theta))}{(r - R_0(\theta))^{\ell}} \mu(r, \theta) \quad \forall r > R_0(\theta).$$

Now, pick $z > 0$ arbitrarily, and define

$$w(r, \theta) = -(r - R_0(\theta))^{\ell} M^{-1}(r, \theta)z \quad \forall r > R_0(\theta)$$

(14)

$$w'(r, \theta) = \lim_{r \to R_0(\theta)^+} w(r, \theta) = -(\text{ad}_{[\mu]}(M(R_0(\theta), \theta)) \mu(R_0(\theta), \theta)) z. \quad (15)$$

For any $r > R_0(\theta)$, (4) in Theorem 1 implies that $M(r, \theta)$ is Hurwitz, so $-M(r, \theta)w(r, \theta) > 0$, and thus, $w(r, \theta) > 0$. Furthermore, $M(r, \theta)w(r, \theta) < 0$, so expanding $M(r, \theta)$ with (13) and rearranging, we obtain $p(r, w(r, \theta), \theta) < 0$. Now we use $w'(r, \theta)$ to formally establish relationships between the feasible sets of both pairs of optimization problems.

**Lemma 6 (Relating the Feasible Sets):** For each $\tau > 0$, let $\Theta_1(\tau) \subset \mathbb{R}^k$ be the set of such that $(r, w, \theta) \in \text{cl}(F_1(\tau))$ for some $r, w, \theta$. The following hold:

1. $\theta \in \Theta_1(\tau)$ implies $w(\theta) \leq \| q \|\Delta r_0$ and $R_0(\theta) \leq r_{\text{max}}$. Fix any $\tau > 0$, and let $\tau > 0$. By (15), we can choose $\tau > R_0(\theta)$ such that $|w(r, \theta, \theta)| < \epsilon$ and $|r - R_0(\theta)| < \min\{r, \epsilon\}$. Since $p(r, w(r, \theta), \theta) \leq \| q \|\Delta r_0$ and $r < R_0(\theta + \tau) \leq r_{\text{max}} + \tau$, we have $(r, w(r, \theta), \theta) \in F_1(\tau)$, so every neighborhood of $(R_0(\theta), w(\theta), \theta)$ (by choice of $\epsilon$) contains a point in $F_1(\tau)$. We prove (2) by a similar argument (without $\tau$), noting that $\theta \in G_2$ implies $c(\theta) \leq c_{\text{max}}$.

To prove 3, we note that $\Theta_1(\tau)$ is the infimum of Problem 1 for all $\tau > 0$, and we define $c'$ as the minimum cost of the $R_0$-constrained allocation problem. Noting that $\Theta_1(\tau)$ are nested downward as $\tau \to 0$

$$c' = \min \Theta_1(\tau) = \lim_{\tau \to 0} \min \Theta_1(\tau) = \lim_{\tau \to 0} c'(\tau).$$

The second step is due to Lemma 6, and the third step is a general property of intersections of nested sets. Let $\theta^0$ be an optimal $R_0$-constrained allocation. Then, $\theta^0 \in G_1$, so by Lemma 6, $(R_0(\theta^0), w^{(\theta^0)}, \theta^0) \in \text{cl}(F_1(\tau))$ for all $\tau > 0$, and we have shown that $c'(\tau) \to c' = c(\theta^0)'$ as $\tau \to 0$. On the other hand, if there exist $r^*, w^*$ such that $(r^*, w^*, \theta^*) \in \text{cl}(F_1(\tau))$ for all $\tau > 0$, then Lemma 6 guarantees $\theta^0 \in G_2$, and $c'(\tau) \to c(\theta^0)$ implies that $c(\theta^0) = c'$. We now prove 2. Let $\theta^0$ be an optimal budget-constrained allocation. Then, $\theta^0 \in G_2$, so Lemma 6 implies that $(R_0(\theta^0), w^{(\theta^0)}, \theta^0) \in \text{cl}(F_2)$. Consider any other point $(r, w, \theta) \in \text{cl}(F_2)$, and note that Lemma 6 also implies $\theta \in G_2$, so that $R_0(\theta^0) \leq R_0(\theta)$. But $R_0(\theta) < r$ by (5), so $R_0(\theta^0) < r$. Thus, $R_0(\theta^0)$ is the minimum value of $r$ over $\text{cl}(F_2)$.

Finally, suppose that $(R_0(\theta^0), w^{(\theta^0)}, \theta^0) \in \text{cl}(F_2)$ and that $R_0(\theta^0)$ is the infimum of Problem 2. Lemma 6 guarantees that $\theta^0 \in G_2$. Consider any other point $\theta \in G_2$, and note that $(R_0(\theta), w^{(\theta)}, \theta) \in \text{cl}(F_2)$ as well, so that $R_0(\theta^0) \leq R_0(\theta)$. Therefore, $\theta^0$ is a minimizer for (8), so it is an optimal budget-constrained allocation.

**References**

[1] R. E. Rowthorn, R. Laxminarayan, and C. A. Gilligan, “Optimal control of epidemics in metapopulations,” J. Roy. Soc. Interface, vol. 6, no. 41, pp. 1135–1144, 2009.

[2] S. Lee, G. Chowell, and C. Castillo-Chávez, “Optimal control for pandemic influenza: The role of limited antiviral treatment and isolation,” J. Theor. Biol., vol. 265, pp. 136–150, 2010.

[3] M. Hayhoe, F. Barreras, and V. M. Preciado, “Multitask learning and nonlinear optimal control of the COVID-19 outbreak: A geometric programming approach,” Ann. Rev. Control, vol. 52, pp. 495–507, 2021.

[4] V. L. J. Somers and I. R. Manchester, “Sparse resource allocation for control of spreading processes via convex optimization,” IEEE Contr. Syst. Lett., vol. 5, no. 2, pp. 547–552, Apr. 2021.

[5] V. M. Preciado, M. Zargham, C. Entyohia, A. Jadbabaie, and G. J. Pappas, “Optimal resource allocation for network protection against spreading processes,” IEEE Trans. Control Netw. Syst., vol. 1, no. 1, pp. 99–108, Mar. 2014.

[6] J. A. Torres, S. Roy, and Y. Wan, “Sparse resource allocation for linear network spread dynamics,” IEEE Trans. Autom. Control, vol. 62, no. 4, pp. 1714–1728, Apr. 2017.

[7] C. Nowzari, V. M. Preciado, and G. J. Pappas, “Optimal resource allocation for control of networked epidemic models,” IEEE Trans. Control Netw. Syst., vol. 4, no. 2, pp. 159–169, Jun. 2017.

[8] V. S. Mai, A. Battou, and K. Mills, “Distributed algorithm for suppressing epidemic spread in networks,” IEEE Control Syst. Lett., vol. 2, no. 3, pp. 555–560, Jul. 2018.

[9] A. R. Hota, J. Godbole, and P. E. Paré, “A closed-loop framework for inference, prediction, and control of SIR epidemics on networks,” IEEE Trans. Netw. Sci. Eng., vol. 8, no. 3, pp. 2262–2278, Jul./Sep. 2021.

[10] P. V. den Driessche and J. Wathmough, “Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission,” Math. Biosc., vol. 180, no. 1, pp. 29–48, 2002.

[11] O. Diekmann, J. A. P. Heesterbeek, and J. A. J. Metz, “On the definition and the computation of the basic reproduction ratio $R_0$ in models for infectious diseases in heterogeneous populations,” J. Math. Biol., vol. 28, no. 4, pp. 365–382, 1990.

[12] M. Ogura, M. Kishida, and J. Lam, “Geometric programming for optimal positive linear systems,” IEEE Trans. Autom. Control, vol. 65, no. 11, pp. 4648–4663, Nov. 2020.

[13] S. Boyd, S.-J. Kim, L. Vandenberghe, and A. Hassibi, “A tutorial on geometric programming,” Optim. Eng., vol. 8, no. 1, pp. 67–127, 2007.

[14] M. Grant and S. Boyd, “CVX: Matlab software for disciplined convex programming, version 2.1,” Mar. 2014. [Online]. Available: http://cvxr.com/cvx/.

[15] F. Bullo, Lectures on Network Systems, 1st ed. Seattle, WA, USA: Kindle Direct Pub., Jan. 2022.

[16] R. A. Horn and C. R. Johnson, Matrix Analysis, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2012.