The Basic Reproduction Number as a Loop Gain Matrix

A. Colombo, *Senior Member, IEEE*

Abstract—**The COVID-19 pandemic and the disordered reactions of most governments made the importance of mathematical modelling and model-based predictions evident, even outside the scientific community. The basic reproduction number $\mathcal{R}_0$ quickly entered the common jargon, as a concise but effective tool to communicate the spreading power of a disease and estimate, at least roughly, the possible outcomes of the epidemic. However, while $\mathcal{R}_0$ is easily defined for simple models, its proper definition is more subtle for larger, state-of-the-art models. Here we show that it is nothing else than the spectral radius of the gain matrix of a linear system, and that this matrix generalizes $\mathcal{R}_0$ in the computation of the vector-valued final epidemic size and epidemic threshold, in a large class of finite-dimensional SIR-like models.**

Index Terms—**Biological systems, compartmental and positive systems, network analysis and control.**

I. INTRODUCTION

IN DECEMBER 2019 a wave of pneumonia cases spread through the city of Wuhan (China) [1], [2]. These were soon attributed to a new strain of coronavirus, SARS CoV-2, a virus related to the strains that caused the MERS and SARS epidemics in recent years [3], but with different pathology and epidemiology that make it a relatively unpredictable pathogen. In a year the infection spread to affecting over 100 million people worldwide [4]. Chinese authorities reacted by setting up a massive campaign of testing and social restrictions. As the pathogen spread through other countries, governments were for the most part initially reticent to implement restrictive measures along the line of what was done in China, only to follow suit under the pressure of exponentially increasing infections and consequent fatalities.

An unexpected by-product of the disease's disruptive dynamics was a surge in public consciousness about the importance of reliable modelling of the disease's dynamics, to provide predictions and guide the response to the pandemic. Common mathematical models, such as the SIR model by Kermack and McKendrick [5], became common parlance, and

the basic reproduction number $\mathcal{R}_0$ entered the pages of daily newspapers.

A finite-dimensional SIR model is a system of three differential equations describing the total number of susceptible (S), infected (I) and removed (R) individuals in a population. In virtually all studies on COVID-19 a simple variant of the model is used, which assumes no endemic equilibrium (all infected eventually heal or die), and no vital dynamics (birth and death for reasons other than the disease is neglected). Despite the simplicity of the setting, these SIR-like models can have hundreds of differential equations and thousands of parameters to model the geographic distribution or social stratification of the population [6]–[9], and have been successfully used to model the diffusion of past epidemics [10], [11]. In these cases, the time evolution of the S, I, and R compartments can be written as:

$$\dot{S} = -D(S)TI,$$
$$\dot{I} = AI + BD(S)TI,$$
$$\dot{R} = C\tilde{A}I, \tag{1}$$

where $S \in \mathbb{R}_+^m$, $I \in \mathbb{R}_+^n$, $R \in \mathbb{R}_+^p$. Here and in the following we use the notation $D(x)$ for a diagonal matrix with elements of vector x in the diagonal. Symbols A, $\tilde{A}$, B, and T instead denote matrices of suitable dimensions. We assume that

 (i) A is stable (all eigenvalues have strictly negative real part),
 (ii) A is Metzler (all elements except diagonal ones are nonnegative), and all elements of $\tilde{A}$, B, C, and T are nonnegative,
(iii) columns of B and C sum to 1, and columns of $A + \tilde{A}$ sum to 0,
(iv) all rows of T have at least one strictly positive element.

Assumption (i) guarantees the absence of endemic equilibria. Indeed, if such an equilibrium existed, by definition it would satisfy $\dot{S} = 0$ and $\dot{I} = 0$ for $I \neq 0$. The first equation requires $D(S)TI = 0$, which would imply that, in the second equation, $AI = 0$. This contradicts (i). Assumption (ii) follows from the nature of the right-hand side of (1), which describes flows between and within compartments, while assumption (iii) ensures mass conservation, i.e., $\|S(t)\|_1 + \|I(t)\|_1 + \|R(t)\|_1 = 1$, for all t. Assumptions (ii) and the structure of (1) further imply that the system is positive. Finally assumption (iv) ensures that all subcompartments of S can be infected.

The matrix product $BD(S)T$ in (1) is known in the epidemiological literature (e.g., [12], [13]) as the *transmission* matrix: it models all epidemiological events that lead to a new infection, through the flow $D(S)TI$ from compartment S to I. Matrix A is instead known as the *transition* matrix: its elements encode the flows of individuals within compartment I, as well as from compartment I to R. The flow from I to R, in particular, is given by $\tilde{A}I$, and is distributed among the subcompartments of R through matrix C.

This family of compartmental models covers most of the finite dimensional, deterministic epidemic models that have been proposed and validated for the modelling of COVID-19, including socially stratified, spatially explicit (spatially structured) and networked models, though it excludes vital dynamics. As an illustrative example, consider the case of a toy metapopulation consisting of two SEIR subpopulations, where the exposed individuals (E) are not infective, while the infective individuals ($\mathcal{I}$) can come into contact with susceptibles of either subpopulation. The model equations for one subpopulation are

$$
\begin{aligned}
\dot{S}_i &= -S_i(a_{i,i}\mathcal{I}_i + a_{i,j}\mathcal{I}_j) \\
\dot{E}_i &= S_i(a_{i,i}\mathcal{I}_i + a_{i,j}\mathcal{I}_j) - b_i E_i \\
\dot{\mathcal{I}}_i &= b_i E_i - c_i \mathcal{I}_i \\
\dot{R}_i &= c_i \mathcal{I}_i,
\end{aligned}
\tag{2}
$$

with $i = 1, j = 2$, the other subpopulation model is obtained by swapping indices. The model has 8 state variables. To cast it into form (1) we collect them into three vectors: $S := (S_1, S_2)^\top$, $I := (E_1, \mathcal{I}_1, E_2, \mathcal{I}_2)^\top$, $R := (R_1, R_2)^\top$. The model matrices then become

$$
A := \begin{pmatrix} -b_1 & 0 & 0 & 0 \\ b_1 & -c_1 & 0 & 0 \\ 0 & 0 & -b_2 & 0 \\ 0 & 0 & b_2 & -c_2 \end{pmatrix}, B := \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix},
$$

$$
T := \begin{pmatrix} 0 & a_{1,1} & 0 & a_{1,2} \\ 0 & a_{2,1} & 0 & a_{2,2} \end{pmatrix}, C := \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix},
$$

$$
\tilde{A} := \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & c_1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & c_2 \end{pmatrix}.
$$

The sign of the elements of vectors and matrices plays an important role in the analysis of this system, so we need to introduce some notation to avoid confusion. Given a vector x or a matrix M, we write $x \geq 0$ or $M \geq 0$ if all elements of x or M are nonnegative. Such a vector or matrix is called *nonnegative*. In the case that all entries of x or M are strictly positive, we say that x or M is *positive*, writing $x > 0$ or $M > 0$.

The contributions of this letter are as follows. In Section II, we rewrite system (1) to highlight the feedback structure of its I compartment's dynamics, and we use this structure to compute a loop gain matrix, called $GD(S)$, which we show to share the same spectral properties as the *Next Generation Matrix* (NGM), the tool typically used to define the basic reproduction number in ODE-based epidemic models (see, e.g., [13]).

Our main contributions are in Theorems 2 and 3, which establish a relation between $GD(S)$ and the asymptotic distribution of non-infected individuals among the compartments of

S, and prove this distribution to be the globally asymptotically stable equilibrium of a discrete-time map over a compact set, and in Theorem 5, which proves that, for the class of systems we consider, a decrease of any term of matrix T (the contact rates) implies a decrease of the final epidemic size. These results suggest how $GD(S)$ (more naturally than the NGM) plays the role of a matrix-valued basic reproduction number in the general class of systems described by (1).

We also show through Theorem 4 and its corollary how the epidemic threshold (the threshold of susceptible individuals, distributed among compartments of S, above which a new infection can trigger and epidemic) is related to the spectral radius of $GD(S)$. In the light of the link between $GD(S)$ and the Next Generation Matrix, which we established, this is a known result. The relevance stands in the link itself: if the epidemic threshold depends only on the dynamics of the linear feedback system (4), the vast body of results about the robust stability and control of linear feedback systems can be used to assess, for instance, the robustness of a sub-threshold population in uncertain models with nominal equations of the form (1).

II. THE BASIC REPRODUCTION NUMBER AS A LOOP GAIN MATRIX

The basic reproduction number, $\mathcal{R}_0$, is defined as the expected number of new infections that a typical infected individual in (1) will cause, in a fully susceptible population. It was defined algebraically in [13], [14] as the spectral radius of the NGM (or, more precisely, the NGM with large domain), which in the notation of (1) is written as

$$
- BD(S_0)TA^{-1}, \tag{3}
$$

where S_0 is the value of S at time 0. Note that the NGM is a nonnegative matrix, a fact that follows from B, $D(S_0)$ and T being nonnegative, from our assumption of A being stable, and from the following result.
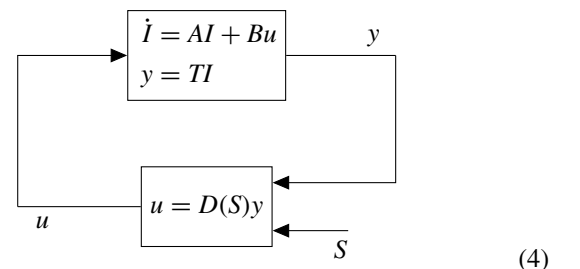
Lemma 1 (See [15]): Let A be a Metzler matrix whose eigenvalues have negative real part. Then $A^{-1} \leq 0$.

The Perron Frobenius theorem (see, e.g., [16]) thus ensures that its spectral radius equals its largest positive eigenvalue.

Starting from (3), we can give a further, more intuitive representation of $\mathcal{R}_0$ as the spectral radius of the gain matrix of the linear system that models the instantaneous diffusion of the disease. We begin from the following algebraic result.

Lemma 2: Let M and N be matrices such that MN and NM are square. Then the set of nonzero eigenvalues of MN is equal to the set of nonzero eigenvalues of NM.

Now, we consider once again system (1), and specifically the equation of $\dot{I}$. We can rewrite the dynamics of the infected as the following feedback loop

$$
\tag{4}
$$

The upper block is a linear time-invariant system. The input vector u, which has the same size as S, represents the (normalized) instantaneous flow of newly infected individuals, which are distributed through the compartment I as dictated by matrix B. The output vector y, of the same size as S and u, is the fraction of susceptible individuals in each subcompartment of S that come in contact with an infected, per unit time. The gain matrix of this linear system is

$$G := -TA^{-1}B,$$

and describes the ratio of y over u at steady state.

The lower block is a time-dependent gain, simply multiplying the vector y by the square, time-variant matrix $D(S)$, that is, translating the number y of infectious contacts into the flow u of newly infected individuals according to the current distribution of susceptible individuals. It follows that the loop gain matrix of the feedback loop for a fixed value of S is equal to

$$GD(S).$$

If we evaluate it at $S = S_0$, where S_0 encodes the structure of the metapopulation in the absence of the disease, then $GD(S_0)$ indeed describes how a given distribution y of new infections in a fully susceptible population initially propagates under the dynamics of (1). It easily follows from Lemma 2 and the definition of the NGM that

Theorem 1: $\mathcal{R}_0$ is the spectral radius of the loop gain matrix $GD(S_0)$.

Proof: From Lemma 2, setting $M := BD(S_0)$ and $N := TA^{-1}$. ∎

Note that, in the common case of a vector I of dimension larger than S (i.e., in metapopulations where each infected compartment is divided in more than one subcompartment), matrix $GD(S_0)$ has smaller dimension than $-BD(S_0)TA^{-1}$. It may resemble, in this sense, the NGM with small domain from [13]. The two matrices are however not identical, in general, even though they share the same set of nonzero eigenvalues.

We see in the next section how the gain matrix $GD(S_0)$ generalizes usage of $\mathcal{R}_0$ in the computation of typical epidemic quantities for all models of the form (1).

III. RELATIONS BETWEEN $GD(S_0)$ AND THE EPIDEMIC QUANTITIES

A. The Final Epidemic Size

A typical use of $\mathcal{R}_0$ is in estimating, based on initial conditions and the structure of the population, the final epidemic size, i.e., the total number of individuals that will be infected during the course of the epidemic. As long as I vanishes asymptotically, by mass conservation the asymptotic values of S or R, which we call S_∞ and R_∞, can be used interchangeably for this purpose.

This is a simple exercise in the case where S, I and R are scalar. First notice that, using (3), assuming $S_0 = 1$, and taking $B = 1$ according to our hypotheses, the basic reproduction number for the scalar system is $\mathcal{R}_0 = -\frac{BS_0T}{A} = -\frac{T}{A}$. Now,

taking the first two equations in (1) we can write

$$\frac{dI}{dS} = -\frac{A}{ST} - 1,$$

which gives $I(t) = -\frac{A}{T}\ln S(t) - S(t) + C$. Assuming that S_0 is asymptotic to 1 (therefore I_0 is asymptotic to 0) we obtain $C = 1$. Finally, assuming $\lim_{t\to\infty} I(t) = 0$ we obtain

$$-\ln S_\infty = \mathcal{R}_0(1 - S_\infty), \qquad (5)$$

or equivalently $-\ln(1 - R_\infty) = \mathcal{R}_0 R_\infty$. We can generalize the above calculation as follows, using vector Δ with $\Delta_i := S_{i,\infty}/S_{i,0}$ to represent the final epidemic size weighted on the initial population distribution, that is, the fraction of individuals in each subcompartment of S that remain uninfected throughout the epidemic.

Theorem 2: The asymptotic value Δ satisfies

$$-\ln(\Delta) + TA^{-1}I_0 - GD(S_0)(\mathbf{1} - \Delta) = 0, \qquad (6)$$

where $\mathbf{1}$ is a vector with elements equal to 1 and $\ln(\Delta)$ is the elementwise natural logarithm of Δ.

Proof: We have

$$\int_0^\infty \frac{dI(\tau)}{d\tau} d\tau = I_\infty - I_0$$
$$= A\int_0^\infty I(\tau)d\tau + B\int_0^\infty D(S(\tau))TI(\tau)d\tau.$$

We know that $\dot{S} = -D(s)TI$ and, by the assumption of no endemic equilibrium, that $I_\infty = 0$, therefore

$$-I_0 = A\int_0^\infty I(\tau)d\tau - B\int_0^\infty \frac{dS(\tau)}{d\tau}d\tau$$
$$= A\int_0^\infty I(\tau)d\tau - B(S_\infty - S_0).$$

Multiplying both sides by TA^{-1} we obtain

$$-TA^{-1}I_0 = \int_0^\infty TI(\tau)d\tau + TA^{-1}B(S_0 - S_\infty). \qquad (7)$$

Finally, from the law of $\dot{S}$ we have $D^{-1}(S)\dot{S} = -TI$, that is

$$\int_0^\infty TI(\tau)d\tau = \ln\left(\frac{S_0}{S_\infty}\right) = -\ln(\Delta).$$

Substituting this in (7) and simplifying we obtain the desired relation. ∎

We can further prove existence and uniqueness of the solution of (6), and a means to compute it, as follows. Take $x \in \mathbb{R}^m$ and define the map

$$x \mapsto F(x) := -TA^{-1}I_0 + GD(S_0)(\mathbf{1} - e^{-x}),$$

where e^{-x} is the element-wise exponential of $-x$. Call X the box $\{x : 0 \le x \le \bar{x}\}$ for some $\bar{x} > -TA^{-1}I_0 + GD(S_0)\mathbf{1}$. The following theorem generalizes the result of [17, Th. 5.4] to the class of systems in (1), using a similar proof structure.

Theorem 3: Defining $\Delta := e^{-x}$ and provided that $I_0 > 0$, the value of Δ solution of (6) exists and is the unique fixed point of $F(x)$ in X, and is the limit of the sequence $x^{(k)} := F(x^{(k-1)})$, starting from any $x^{(0)} \in X$.

Proof: X is positively-invariant under F. Consider the sequences $x^{(k)} = F(x^{(k-1)})$ with $x^{(0)} = 0$ and $x^{(0)} = \bar{x}$.

The map F is order-preserving, $F(0) \geq 0$, and $F(\bar{x}) < \bar{x}$, therefore the sequences are totally ordered and monotonic. Since they are also bounded within the compact set X they converge to fixed points p and q, respectively. Furthermore, since $0 \leq x \leq \bar{x}$ for all $x \in X$, points p and q are ordered ($p \leq q$) and are lower and upper bounds to the limiting value of $x^{(k)} = F(x^{(k-1)})$ for all $x^{(0)} \in X$.

We now show by contradiction that p and q must coincide. Consider a straight half-line starting from q and going through p. The line intersects the boundary of X at some point w such that $w_i = 0$, for some $i \in \{1, \ldots, m\}$. Call $F_i(x)$ the i-th component of F. Using Lemma 1 and assumptions (i), (ii) and (iv), we have that $-TA^{-1} \geq 0$ and all rows have at least one strictly positive element. This implies that $-TA^{-1}I_0 > 0$ (since $I_0 > 0$) and $GD(S_0)(\mathbf{1} - e^{-w}) \geq 0$ (since $B(\mathbf{1} - e^{-w}) \geq 0$). This means that $F_i(w) > w_i = 0$. Moreover, $F_i(x)$ is a concave function of x, hence if we take a weight α such that $p = \alpha w + (1-\alpha)q$, we obtain $p_i = F_i(p) \geq \alpha F_i(w) + (1-\alpha)F_i(q) = \alpha F_i(w) + (1-\alpha)q_i$. Then either $\alpha = 0$ (hence $p = q$), or $p_i > \alpha w_i + (1-\alpha)q_i = p_i$, which is a contradiction. Therefore F has a unique fixed point in X. ∎

In the statement of Theorem 2, I_0 appears as an additive term in the implicit definition of Δ. This means that the distribution of individuals in S_∞ is minimally affected by the distribution of the initial seed I_0 of infected individuals: as $\|I_0\|_1 \to 0$, S_∞ tends to a distribution independent of I_0 and implicitly given by the relation

$$-\ln\Delta = GD(S_0)(\mathbf{1} - \Delta), \quad 0 < \Delta < 1.$$

This is indeed the generalization of (5) to arbitrary models of the form (1), and $GD(S_0)$ now plays the role of $\mathcal{R}_0$ in the implicit expression of the final epidemic size. The above formula (or its more general version (6)) can be used, given an estimate of the initial population distribution S_0, to evaluate the most effective among a set of policies affecting matrix G (e.g., distancing or isolation of given population classes), by solving an optimization problem

$$\max_{G\in\mathcal{G}, 0<\Delta<1} \text{Cost}(\Delta) \quad s.t. \quad GD(S_0)(\mathbf{1}-\Delta) + \ln\Delta = 0, \quad (8)$$

where $\text{Cost}(\Delta)$ is a suitable cost function. As a toy example, consider again the two-population model (2), with $a_{1,1} = 2.4$, $a_{1,2} = 1.2$, $a_{2,1} = 1$, $a_{2,2} = 2$, $b_1 = b_2 = 2$, $c_1 = 1$, $c_2 = 1.2$, $S_1(0) = 0.5$, $S_2(0) = 0.5$ (two populations of identical size, population 1 has slightly higher probability of contracting the disease from contacts with members of either population, and slightly lower recovery rate). Assume that a fixed amount of protective devices (e.g., face masks) are to be distributed among the two populations, and their effect is to reduce the transmission rates, proportionally to the amount of devices that are distributed. Assume that we aim to minimize the final epidemic size, so that $\text{Cost}(\Delta) := S_0^\top\Delta = \|S_\infty\|_1$. Calling α_1 and α_2 the effect of the distributed devices in the two populations, respectively, elements of T change as $a_{i,j} \mapsto a_{i,j}(1 - \alpha_i)$. Now, assuming that the total amount of available devices binds $\alpha_1 + \alpha_2 = 0.4$, we can construct the set $\mathcal{G}$ in (8) as $\mathcal{G} := \{-T_{\alpha_1,\alpha_2}A^{-1}B : \alpha_1 + \alpha_2 = 0.4\}$, where $-T_{\alpha_1,\alpha_2}$ denotes matrix T after rescaling of its elements. A nonlinear solver (in this case, MATLAB fmincon

with default options) finds $\alpha_1 = 0.4, \alpha_2 = 0$ as the optimal resource allocation, corresponding to $\|S_\infty\|_1 \simeq 0.73$. Note that an even resource allocation ($\alpha_1 = \alpha_2 = 0.2$) would have resulted in $\|S_\infty\|_1 \simeq 0.66$: the optimal resource allocation, as opposed to an even allocation, can in this case spare about 7% of the population from contracting the disease.

B. The Epidemic Threshold

An epidemic threshold is a parameter value, in an epidemic model, below which the introduction of an infinitesimal number of infected individuals in an otherwise disease-free population cannot trigger an epidemic. In the scalar SIR model without vital dynamics (e.g., [18]), this is given by

$$S_0 < \frac{1}{\mathcal{R}_0}. \quad (9)$$

When S_0 is below threshold, herd immunity prevents an epidemic outbreak. Mathematically speaking, this means that I_0 is an asymptotically stable equilibrium of (4) with $S = S_0$. An alternative interpretation of the epidemic threshold in the same model is as the upper bound to the value of S_∞. Once an epidemic has started, the set of individuals not affected by the disease at the end of the epidemic wave tends to a value below the epidemic threshold:

$$S_\infty < \frac{1}{\mathcal{R}_0}. \quad (10)$$

In vector models such as (1) the fact that a given S be above or below the epidemic threshold depends not only on $\|S\|_1$, but also on the relative value of its elements. Characterizing the set is therefore a bit more involved. We can generalize the relations (9) and (10) as follows.

Call Σ_1 the set of S vectors such that (4) with input S is asymptotically stable, and call Σ_2 the set of S vectors such that there exists $R \geq 0$ for which (S, I, R), with $I = 0$ is a limit point of at least one orbit of (1) with initial condition $(S(0), I(0), R(0)) : I(0) > 0$. The two sets coincide and are characterized as follows.

Theorem 4:

$$\Sigma_1 = \Sigma_2 = \Sigma := \{S : \det(I - cGD(S)) > 0, \forall c \in [0, 1]\}.$$

The proof of this theorem uses the following result

Lemma 3 (See, e.g., Corollary 1.5, Chapter 2, in [19]): Given a Metzler matrix M, its dominant eigenvalue λ_{pf} is a weakly nondecreasing function of any positive perturbation of the elements of M

Proof of Theorem 4: We consider the first two equations in (1). The value of S is nonincreasing, and since S is bounded above 0 it must tend to some limit value $S_\infty \leq S(t)$, for all $t \geq 0$. We can thus rewrite the equation of I as

$$\dot{I} = (A + BD(S_\infty)T)I + BD(S - S_\infty)TI, \quad (11)$$

which we can see as the linear system

$$\dot{I} = (A + BD(S_\infty)T)I + d \quad (12)$$

with a time-dependent perturbation $d \geq 0$. The value of I instead must go to 0 as $t \to \infty$ by the assumption of no endemic equilibrium. We now show by contradiction that (12)

does not admit any solution $I(t) \to 0$ with $I(0) > 0$, unless matrix $(A + BD(S_\infty)T)$ is stable.

First, note that $(A + BD(S_\infty)T)$ is Metzler. Assume that its dominant eigenvalue is nonnegative, call $v(0)$ a vector lying in its stable eigenspace, and assume that $v(0) > 0$. Then we can pick two finite perturbations $\pm p(0)$ such that $v(0) \pm p(0) \geq 0$ and such that $p(0)$ is a linear combination of the eigenvectors of $(A + BD(S_\infty)T)$ relative to null or positive eigenvalues. Consider how the flow of an initial condition $v(0) \pm p(0)$ evolves under the vector field (12) with $d = 0$. We have $\lim_{t \to \infty} v(t) \pm p(t) = \lim_{t \to \infty} \pm p(t)$, with $\lim_{t \to \infty} |p(t)| > 0$. Since $\pm p(t)$ cannot belong to the positive orthant at the same time, but the positive orthant of (12) is positively invariant, we conclude that the stable eigenspace of $(A + BD(S_\infty)T)$ cannot contain any vector $v > 0$, unless the dominant eigenvalue of $(A + BD(S_\infty)T)$ is negative.

Now we take, as by assumption, an arbitrary $I(0) > 0$. By the above reasoning this vector must have nonzero component in the center or unstable eigenspaces of $(A + BD(S_\infty)T)$, if they exist. This implies that if $(A + BD(S_\infty)T)$ is not stable then the solution of the linear equation (12) cannot go to 0 for any $d \geq 0$, and the same must hold for the solution of (11). We thus conclude that $(A + BD(S_\infty)T)$ is stable at any limit point S_∞. This means that $\Sigma_1 = \Sigma_2$.

We can now characterize the set Σ. By the matrix determinant lemma, we have

$$\det(A + BD(S)T) = \det(A)\det(I + TA^{-1}BD(S))$$
$$= \det(A)\det(I - GD(S)),$$

and given that A is stable by assumption, $A + BD(S)T$ is stable only provided that $\det(I + TA^{-1}BD(S)) > 0$. Also, by Lemma 3, if $A + BD(S)T$ is stable then so is $A + BD(cS)T$, for any $c \in [0, 1]$. These facts together give that stability of $A + BD(S)T$ implies that $\det(I - GD(cS)) > 0$, for all $c \in [0, 1]$.

To prove that the converse also holds, note that by the above mentioned matrix determinant lemma $\det(I - GD(cS)) > 0$, for all $c \in [0, 1]$, implies that $\det(A + BD(cS)T) \neq 0$, for all $c \in [0, 1]$. Since $A + BD(cS)T$ is Metzler, the Perron Frobenius theorem ensures its dominant eigenvalue is real. We know that $A + BD(cS)T$ is stable for $c = 0$, so it can only lose stability through a zero crossing of its real dominant eigenvalue, which would imply $\det(A + BD(cS)T) = 0$. Therefore we have

$$A + BD(S)T \text{ is stable} \Leftrightarrow \det(I - GD(cS)) > 0, \; \forall\, c \in [0, 1].$$

We have seen before that S_∞ must be such that $A + BD(S)T$ is stable, so the above statement is equivalent to

$$S \in \Sigma \Leftrightarrow \det(I - cGD(S)) > 0, \; \forall\, c \in [0, 1].$$

∎

The result in Theorem 4, which generalizes the epidemic threshold, can also be restated as follows:

Corollary 1: $\Sigma := \{S : \rho(GD(S)) < 1\}$, where ρ is the spectral radius.

Proof: The eigenvalues of $cGD(S)$ are a continuous function of c. Given that, for $c = 0$, all eigenvalues of $I - cGD(S)$ are strictly positive, and given that the dominant eigenvalue of $cGD(S)$ is real (the matrix is nonnegative), $\det(I - cGD(S)) > 0$,

for all $c \in [0, 1]$, implies that all eigenvalues of $I - cGD(S)$ remain strictly positive for all $c \in [0, 1]$, including $c = 1$. This means that the dominant eigenvalue of $GD(S)$ must be smaller than 1. As this is also the spectral radius by the Perron Frobenius theorem, we have $\rho(GD(S)) < 1$. ∎

The results above prove the almost global asymptotic stability of the states with $S \in \Sigma$ and $I = 0$, and characterize the set Σ by the inequality $\rho(GD(\bar{S})) = \mathcal{R}_0 < 1$. Equivalent characterizations have been given before albeit with different proofs see, e.g., [20] (Theorem 2). What makes the above formulation novel, besides the different proof, is that it elucidates the relation between these results and the classical theory of linear feedback systems: calling F the loop transfer function of (4) for a fixed S, $\rho(GD(S)) \leq \|F\|_\infty$, where the latter denotes the $\mathcal{H}_\infty$ norm of the transfer function. Therefore, the stability of (4) by the Small Gain Theorem implies $S \in \Sigma$. This means that questions related to $S \in \Sigma$, such as the ability of a new disease to trigger an epidemic or the minimum number of infected at the end of an epidemic wave, when the nonlinear model (1) is subject to parametric or modelling uncertainties, can be attacked using linear robust stability techniques [21].

C. Social Distancing Always Reduces the Total Infected Population

To conclude, we remark that in the first few months of the COVID-19 pandemic, when data was insufficient to reliably identify any predictive model, numerous publications stated that a reduction of social contacts (in (1), a reduction of the value of the elements of T) was nonetheless a fundamental tool to reduce the impact of the pandemic, as numerical simulations showed that it contributed to reducing the final epidemic size. While this is certainly a true and important message to convey to decision makers and to the public, its truth did not need the use of numerical simulations to be supported, as it holds for any, arbitrary SIR-like model of the form (1):

Theorem 5: Let $I_0 > 0$, and let T_{jk} be element (j, k) of matrix T. Then, $\frac{dS_\infty}{dT_{jk}} \leq 0$ and $\frac{dS_\infty}{dT_{jk}} \neq 0$.

Proof: We have seen that S_∞ and T are related through the vector-valued equation

$$H(s_\infty, T) := \ln\left(\frac{S_0}{S_\infty}\right) + TA^{-1}I_0 + TA^{-1}B(S_0 - S_\infty) = 0.$$

We have that $\frac{dH(S_\infty, T)}{dT_{jk}} = \frac{\partial H(S_\infty, T)}{\partial T_{jk}} + \frac{\partial H(S_\infty, T)}{\partial S_\infty}\frac{dS_\infty}{dT_{jk}} = 0$, and $\frac{\partial H(S_\infty, T)}{\partial S_\infty} = -D^{-1}(S_\infty) - TA^{-1}B = -(I - GD(S_\infty))D^{-1}(S_\infty)$. We know that $S_\infty > 0$, otherwise it would not satisfy $H(S_\infty, T) = 0$ due to the log term, so $D^{-1}(S_\infty)$ exists. Also, from Theorem 4 we know that $(I - GD(S_\infty))$ is nonsingular. This implies that $\frac{\partial H(S_\infty, T)}{\partial S_\infty}$ is invertible, therefore

$$\frac{dS_\infty}{dT_{jk}} = -\left(\frac{\partial H(S_\infty, T)}{\partial S_\infty}\right)^{-1}\frac{\partial H(S_\infty, T)}{\partial T_{jk}}.$$

Putting the expression for $H(s_\infty, T)$ in the above formula we obtain

$$\frac{dS_\infty}{dT_{jk}} = D(S_\infty)(I - GD(S_\infty))^{-1}\Big(\delta(j,k)A^{-1}I_0$$
$$+ \; \delta(j,k)A^{-1}B(S_0 - S_\infty)\Big),$$

where $\delta(j,k)$ is a matrix with all elements equal to zero, except element (j,k), which is equal to 1. Now, by assumption we have $I_0 > 0$, while by Lemma 1 $A^{-1} \leq 0$. Since all other terms in the expression $(\delta(j,k)A^{-1}I_0 + \delta(j,k)A^{-1}B(S_0 - S_\infty))$ are nonnegative and A^{-1} is nonsingular, we have

$$\Big(\delta(j,k)A^{-1}I_0 + \delta(j,k)A^{-1}B(S_0 - S_\infty)\Big) \leq 0.$$

Furthermore, having $A^{-1} \leq 0$ and nonsingular, and having $I_0 > 0$, imply that $A^{-1}I_0 < 0$, therefore

$$\Big(\delta(j,k)A^{-1}I_0 + \delta(j,k)A^{-1}B(S_0 - S_\infty)\Big) \neq 0.$$

We then know, from Corollary 1, that $I - GD(S_\infty)$ is stable and invertible, and we can easily verify that $-(I - GD(S_\infty))$ is Metzler, so by Lemma 1 we know that $(I - GD(S_\infty))^{-1} \geq 0$. Therefore,

$$(I - GD(S_\infty))^{-1}\Big(\delta(j,k)A^{-1}I_0$$
$$+ \; \delta(j,k)A^{-1}B(S_0 - S_\infty)\Big) \leq 0.$$

Nonsingularity of $(I - GD(S_\infty))^{-1}$ implies

$$(I - GD(S_\infty))^{-1}\Big(\delta(j,k)A^{-1}I_0$$
$$+ \; \delta(j,k)A^{-1}B(S_0 - S_\infty)\Big) \neq 0.$$

Finally, we have already seen that $D(S_\infty) \geq 0$ and is nonsingular, therefore

$$\frac{dS_\infty}{dT_{jk}} \leq 0, \quad \frac{dS_\infty}{dT_{jk}} \neq 0.$$

∎

IV. Conclusion

Taking a large class of SIR-like models, which covers many modern state-of-the-art networked, spatially explicit, and socially stratified models, we have shown how the basic reproduction number, commonly termed $\mathcal{R}_0$, is linked to the gain matrix of a linear system. This highlights a stronger connection than commonly acknowledged between a large class of epidemic models and the linear-feedback systems. We have shown how this matrix (and not just its spectral radius), plays the role of $\mathcal{R}_0$ in the computation of the final epidemic size and the epidemic threshold, and how the epidemic threshold is essentially a simple by-product of the small gain theorem. Finally, we have formally proved that for all systems in the above-mentioned class, a reduction in the rate of contacts (at any time during the epidemic) reduces the final epidemic size.

References

[1] T. P. Velavan and C. G. Meyer, "The COVID-19 epidemic," *Tropical Med. Int. Health*, vol. 25, no. 3, pp. 278–280, Feb. 2020.

[2] H. Wang et al., "Phase-adjusted estimation of the number of coronavirus disease 2019 cases in Wuhan, China," *Cell Discov.*, vol. 6, no. 1, p. 10, Feb. 2020.

[3] N. Petrosillo, G. Viceconte, O. Ergonul, G. Ippolito, and E. Petersen, "COVID-19, SARS and MERS: Are they closely related?" *Clin. Microbiol. Infection*, vol. 26, no. 6, pp. 729–734, Jun. 2020.

[4] (Jan. 2021). *WHO Coronavirus Disease (COVID-19) Dashboard*. [Online]. Available: https://covid19.who.int/

[5] W. O. Kermack and A. G. McKendrick, "A contribution to the mathematical theory of epidemics," *Proc. Roy. Soc. A, Math. Phys. Eng. Sci.*, vol. 115, no. 772, pp. 700–721, Aug. 1927.

[6] G. F. de Arruda, G. Petri, and Y. Moreno, "Social contagion models on hypergraphs," *Phys. Rev. Res.*, vol. 2, no. 2, Apr. 2020. Art. no. 023032.

[7] R. Li et al., "Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2)," *Science*, vol. 368, no. 6490, pp. 489–493, Mar. 2020.

[8] M. Chinazzi et al., "The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak," *Science*, vol. 368, pp. 395–400, Apr. 2020.

[9] M. Gatto et al., "Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures," *Proc. Nat. Acad. Sci.*, vol. 117, no. 19, pp. 10484–10491, Apr. 2020.

[10] M. Li, J. Ma, and P. van den Driessche, "Model for disease dynamics of a waterborne pathogen on a random network," *J. Math. Biol.*, vol. 71, no. 4, pp. 961–977, Oct. 2014.

[11] Z. S. Y. Wong, C. M. Bui, A. A. Chughtai, and C. R. Macintyre, "A systematic review of early modelling studies of Ebola virus disease in West Africa," *Epidemiol. Infection*, vol. 145, no. 6, pp. 1069–1094, Feb. 2017.

[12] O. Diekmann and J. A. P. Heesterbeek, *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation* (Wiley Series in Mathematical and Computational Biology). New York, NY, USA: Wiley, 2000.

[13] O. Diekmann, J. A. P. Heesterbeek, and M. G. Roberts, "The construction of next-generation matrices for compartmental epidemic models," *J. Roy. Soc. Interface*, vol. 7, no. 47, pp. 873–885, Nov. 2009.

[14] O. Diekmann, J. Heesterbeek, and 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, Jun. 1990.

[15] C. Briat, "Sign properties of Metzler matrices with applications," *Linear Algebra Appl.*, vol. 515, pp. 53–86, Feb. 2017.

[16] L. Farina and S. Rinaldi, *Positive Linear Systems: Theory and Applications*. New York, NY, USA: Wiley, 2000.

[17] W. Mei, S. Mohagheghi, S. Zampieri, and F. Bullo, "On the dynamics of deterministic epidemic propagation over networks," *Annu. Rev. Control*, vol. 44, pp. 116–128, Sep. 2017.

[18] B. Ridenhour, J. M. Kowalik, and D. K. Shay, "Unraveling R_0: Considerations for public health applications," *Amer. J. Public Health*, vol. 104, no. 2, pp. e32–e41, Feb. 2014.

[19] A. Berman and R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*. Philadelphia, PA, USA: Soc. In. Appl. Math., Jan. 1994.

[20] P. van den Driessche and J. Watmough, "Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission," *Math. Biosci.*, vol. 180, pp. 29–48, Nov./Dec. 2002.

[21] K. Zhou and J. C. Doyle, *Essentials of Robust Control*. London, U.K.: Prentice-Hall, 1998.