Adaptive Identification of SIS Models

Chi Ho Leung, William E. Retnaraj, Ashish R. Hota, and Philip E. Paré*

Abstract—Effective containment of spreading processes such as epidemics requires accurate knowledge of several key parameters that govern their dynamics. In this work, we first show that the problem of identifying the underlying parameters of epidemiological spreading processes is often ill-conditioned and lacks the persistence of excitation required for the convergence of adaptive learning schemes. To tackle this challenge, we leverage a relaxed property called initial excitation combined with a recursive least squares algorithm to design an online adaptive identifier to learn the parameters of the susceptible-infected-susceptible (SIS) epidemic model from the knowledge of its states. We prove that the iterates generated by the proposed algorithm minimize an auxiliary weighted least squares cost function. We illustrate the convergence of the error of the estimated epidemic parameters via several numerical case studies and compare it with results obtained using conventional approaches.

I. INTRODUCTION

Accurately forecasting the progression of infectious diseases has become of vital importance following the catastrophic spread of several pandemics over the past century. Various dynamic models, beginning with the well-known SIS (susceptible-infected-susceptible) model [1] have been developed, often with tailored extensions [2] in response to different epidemic characteristics. These models are governed by parameters that dictate different metrics of interest, including the infection trajectory, the peak of infection, and the endemic equilibrium [3]. Therefore, predicting or forecasting the spread of the disease requires estimating these parameters within a limited time in an online manner.

Adaptive identification is an online system identification strategy that updates the parameter estimates based on incoming input and output data [4]. Traditional techniques for addressing adaptive identification problems often involve gradient descent and recursive least square filtering [5]. A longstanding obstacle in successful adaptive estimation and output error convergence is the requirement of a persistently excited regressor signal, which is often overly restrictive [4], [6]. Contemporary solutions to overcome these restrictions involve indirect adaptive control strategies [7] and robust adaptive control methods [8]. Other efforts to address the lack of uniform persistent excitation have explored the relaxation of this restrictive requirement, including the concept of initial excitation [9], [10].

Another challenge in online system identification is the identifiability of all unknown parameters. While the structural identifiability of a model can be analytically checked in a multitude of ways for both linear and nonlinear systems [11], a more pressing requirement is practical identifiability [11], [12], the lack of which makes parameter estimation a very difficult task despite it being theoretically possible. This challenge has been well-recognized in the context of epidemic models in recent papers [11], [13].

In this work we examine the performance of classical adaptive identification algorithms in learning the parameters governing the susceptible-infected-susceptible (SIS) epidemic model. We prove that for this class of models, the regressor matrix is not persistently exciting, which establishes that any algorithm that relies on this condition would not be able to succeed in the parameter identification task. We further show that the regressor matrix is ill-conditioned which leads to practical identifiability issues, and algorithms based on relaxed notions of persistent excitation also do not perform well.

After highlighting the above characteristics, we propose a novel algorithm that builds upon the recursive least squares (RLS) technique, combining the well-known RLS algorithm with the novel concept of an excitation set that is used to construct the main regressor. We explain the motivation behind this idea, introduce our algorithm, and show that the estimates obtained by the proposed algorithm minimize an auxiliary cost function that weighs exciting data points with unity weight and disregards regressors that are not sufficiently exciting. We compare our results with existing approaches and demonstrate that our algorithm is effective for adaptive parameter identification of SIS models with noise.

Notations: We denote a matrix X to be positive/negative semi-definite by $X \succeq 0$ and $X \preceq 0$, respectively. We denote $\kappa(X)$ as the condition number of matrix X, where the condition number is the ratio of the largest singular value of X to the smallest singular value of X. The weighted norm is denoted as $\|x\|_A = \sqrt{x^\top A x}$ for some positive semi-definite matrix A, and $\|\cdot\|$ is the 2-norm.

II. PROBLEM FORMULATION

In this section, we formally present the adaptive identification problem and introduce the necessary notions and tools required to address it. Subsequently, we justify the need for developing a novel adaptive identification algorithm by providing a motivating example that demonstrates the failure of the classic gradient descent adaptive identification law

*Chi Ho Leung and Philip E. Paré are with the Elmore Family School of Electrical and Computer Engineering, Purdue University, USA. William E. Retnaraj and Ashish R. Hota are with the Department of Aerospace Engineering and Electrical Engineering, respectively, IIT Kharagpur, India. E-mail: leung61@purdue.edu, retnaraj@ieee.org, ahota@ee.iitkgp.ac.in, philpare@purdue.edu.

in estimating the parameters of the SIS epidemic model. Consider the class of non-linear discrete-time systems where the parameters are linearly separable from the states:

$$x_{k+1} = x_k + \phi(x_k)\theta + \xi_k, \tag{1}$$

where $x_k \in \mathcal{X}$ is the state vector, $\theta \in \Theta \subseteq \mathbb{R}^p_{\geq 0}$ is the parameter space, $\phi : \mathcal{X} \mapsto \mathbb{R}^{n \times p}$ is the function that maps states to the regressor matrix, and ξ_k is some bounded unknown perturbation such that $\|\xi_k\| < \nu$ for all $k \geq 0$. We assume that our observation $y_k := x_{k+1} - x_k$ consists of the change in the state variables. We can now write the residuals of each recursive estimation step as:

$$r_k(\hat{\theta}) = y_k - \phi(x_k)\hat{\theta} = \xi_k. \tag{2}$$

Let $\alpha \in (0,1]$ be the exponential forgetting factor. Our goal is to design an adaptive identification law $\hat{\theta}_k = \hat{\theta}_{k-1} + f(x_{k-1}, x_k)$ that minimizes the empirical cost:

$$C_k^{(emp)}(\hat{\theta}_k) = \frac{1}{2} \sum_{i=0}^{k} \alpha^{k-i} \|r_i(\hat{\theta}_k)\|^2 \tag{3}$$

for all $k \geq 0$ given the measurements of y_k and x_k.

1) Preliminaries: In the following definitions, we denote $\Psi(\cdot, \cdot, \cdot)$ as the state transition function of the discrete-time dynamics $x_{k+1} = f(x_k)$, such that $x_k = \Psi(k, k_0, x_0)$ when the initial state is x_0 at time k_0. Adaptive algorithms are guaranteed to track target values and minimize estimation errors when operating under suitable assumptions, with persistence of excitation being one of the key requirements. There are different variants of definitions of persistent excitation depending on the application [9], [10], [14]. We define persistent excitation as follows.

Definition 1 (Persistent Excitation)**.** *Let $x_{k+1} = f(x_k)$. A function $\phi(x)$ is said to be persistently exciting w.r.t. f if:*

$$\sum_{k=l}^{l+L} \phi(\Psi(k, k_0, x_0))^\top \phi(\Psi(k, k_0, x_0)) \succeq \alpha I \quad \forall l \in \mathbb{Z}_{\geq 0} \tag{4}$$

for some positive constants L, α and initial condition x_0.

Another important aspect of the parameter identification problem is its practical identifiability, which depends on the geometric properties of the Fisher information matrix (FIM) of the cost function defined below.

Definition 2 (Fisher Information Matrix)**.** *The Fisher Information Matrix H of the kth empirical cost function (3) w.r.t. the parameters θ is defined as $H_{ij} = \partial_{\theta_i} \partial_{\theta_j} C_k^{(emp)}(\theta)$.*

Since the parameters are linearly separable from the observed states, the Fisher Information matrix can be written in the following manner.

Proposition 1. *The Fisher Information Matrix of the empirical cost $C_k^{(emp)}(\theta)$ is:*

$$H = \sum_{i=0}^{k} \alpha^{k-i} \phi(x_i)^\top \phi(x_i). \tag{5}$$

Further, H is positive semi-definite.

(a) Contour plot of identification error with estimation trajectory

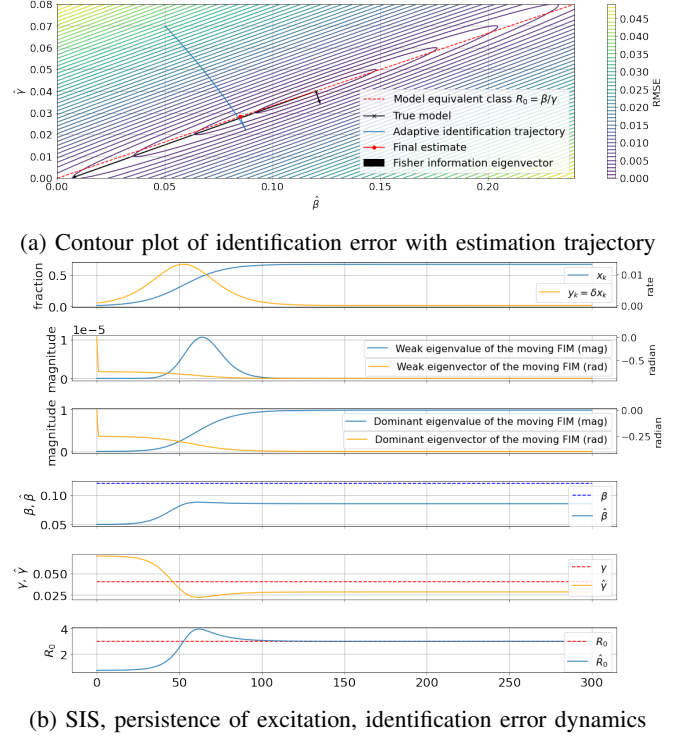(b) SIS, persistence of excitation, identification error dynamics

Fig. 1: The trajectory of the estimates on the RMSE contour plot is shown in Fig. 1a for the SIS model ($\beta = 0.12$, $\gamma = 0.04$, $x(k_0) = 0.01$, $\hat{\beta}(k_0) = 0.05$, $\hat{\gamma}(k_0) = 0.07$) without noise added. Plots displaying from the top in Fig. 1b: The infection proportion x_k and net infection rate $\Delta x_k = x_{k+1} - x_k$ over time; the next two, the moving FIM's eigenvalues and eigenvector, where the moving FIM is computed as $\sum_{i=l}^{l+3} \phi(x_i)^\top \phi(x_i)$; the next two, system parameters estimates $(\hat{\beta}, \hat{\gamma})$ which fail to converge to the true solution; and lastly, the estimated basic reproduction number computed as $\hat{\mathcal{R}}_0 = \frac{\hat{\beta}}{\hat{\gamma}}$.

The proof is straightforward and is omitted in the interest of space. Another useful notion related to practical identifiability is the condition number κ. Notice that $\kappa(M) \in [1, \infty)$ for any real-valued matrix M. The condition number of the FIM determines the ill-posedness of the identification problem. If the condition number is infinite, the identification problem is considered to be ill-posed. On the other hand, the problem is considered to be well-posed if the condition number is finite. However, if it is large, the problem may be ill-conditioned, meaning the FIM is close to singular.

2) Motivating Example: In this subsection, we examine the convergence behavior of the classic gradient descent adaptive law applied to the SIS spreading process.

Example 1. *Consider the SIS epidemic dynamics given by*

$$x_{k+1} = x_k + (1 - x_k)\beta x_k - \gamma x_k, \tag{6}$$

where the state x_k represents the proportion of infected individuals in the population, β is infection rate, and γ is the recovery rate. Consider the adaptive identification law borrowed from the formalism in [15]:

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \phi(x_k)^\top \left(y_k - \phi(x_k)\hat{\theta}_k \right), \tag{7}$$

where $\phi(x_k) := [(1 - x_k)x_k \quad - x_k]$, $y_k := x_{k+1} - x_k$, *and* $\theta := [\beta \quad \gamma]^\top$. *Note that* $\phi(x_k)^\top(y_k - \phi(x_k)\hat{\theta}_k)$ *is known as the negative gradient of the stage cost in* (3). *From Fig. 1, we can see that the classic gradient descent adaptive identification law fails to converge to the true parameters, and the estimates converge to a value in the* $\mathcal{R}_0$ *equivalent class, which is the set of parameters characterized by having the same basic reproduction number* $\mathcal{R}_0 = \frac{\beta}{\gamma}$. *This phenomenon is due to lack of persistent excitation and an ill-conditioned identification problem.*

In the second and third subplots of Fig. 1b, note that the FIM's eigenvalues span many orders of magnitude, that is from less than 10^{-5} to 1, over time. It is worth noting that this phenomenon is not unique to the SIS compartmental model, but exists in nonlinear multi-parameter models from various fields of study [16]. These issues motivate the development of a novel adaptive identification process.

III. MAIN RESULTS

One of the reasons that makes the traditional gradient descent method fail to converge in Example 1 is the lack of persistent excitation. The following proposition states that the lack of persistent excitation is a property of the SIS model regardless of the initial condition.

Proposition 2. *The regressor matrix* $\phi(x_k) := [(1-x_k)x_k \ -x_k]$, $y_k := x_{k+1} - x_k$ *is not persistently exciting with respect to the SIS dynamics in* (6).

Proof. We first notice that the SIS model in (6) has two equilibria: $x^{(1)} = 0$ and $x^{(2)} = 1 - \frac{\gamma}{\beta}$, with the epidemiological threshold $\mathcal{R}_0 = \frac{\beta}{\gamma}$. When $\mathcal{R}_0 \leq 1$, (6) is asymptotically stable around $x^{(1)}$; when $\mathcal{R}_0 > 1$, (6) is asymptotically stable around $x^{(2)}$. Fix any $L > 0$ and $\alpha > 0$; we want to show that there exists an l such that the smallest eigenvalue:

$$\lambda_{\min}\left(\sum_{k=l}^{l+L} \phi(x_k)^\top \phi(x_k)\right) < \alpha,$$

where

$$\phi(x_k)^\top \phi(x_k) = x_k^2 \begin{bmatrix} (1-x_k)^2 & -(1-x_k) \\ -(1-x_k) & 1 \end{bmatrix} =: A(x_k).$$

When $\mathcal{R}_0 > 1$, (6) is asymptotically stable around $x^{(2)}$. Therefore, for all $\epsilon > 0$, there exists a K such that $|x_k - x^{(2)}| < \epsilon$ for every $k \geq K$. Without loss of generality, assume x_k is approaching $x^{(2)}$ from above, then we can write $x_k = x^{(2)} + \epsilon_k$ where $0 < \epsilon_k < \epsilon \ \forall \ k \geq K$. Thus, since $\mathcal{R}_0^{-1} = (1 - x^{(2)})$,

$$A(x_k) = (x^{(2)} + \epsilon_k)^2 \begin{bmatrix} (1 - x^{(2)} - \epsilon_k)^2 & -(1 - x^{(2)} - \epsilon_k) \\ -(1 - x^{(2)} - \epsilon_k) & 1 \end{bmatrix}$$

$$= (x^{(2)})^2 \begin{bmatrix} \mathcal{R}_0^{-2} & -\mathcal{R}_0^{-1} \\ -\mathcal{R}_0^{-1} & 1 \end{bmatrix} + \epsilon_k \left((x^{(2)})^2 \begin{bmatrix} -2\mathcal{R}_0^{-1} + \epsilon_k & 1 \\ 1 & 0 \end{bmatrix} \right.$$

$$\left. + (2x^{(2)} + \epsilon_k) \begin{bmatrix} \left(\mathcal{R}_0^{-1} - \epsilon_k\right)^2 & -(\mathcal{R}_0^{-1} - \epsilon_k) \\ -(\mathcal{R}_0^{-1} - \epsilon_k) & 1 \end{bmatrix} \right)$$

$$= (x^{(2)})^2 aa^\top + o(\epsilon)I,$$

where $a = \begin{bmatrix} \mathcal{R}_0^{-1} & -1 \end{bmatrix}^\top$. We then rewrite $\sum_{k=l}^{l+L} A(x_k)$ as $(L + 1)(x^{(2)})^2 aa^\top + o(\epsilon)I$. Since $aa^\top$ is rank deficient, we can pick an ϵ such that $\lambda_{\min}((L + 1)(x^{(2)})^2 aa^\top + o(\epsilon)I) < \alpha$. When $\mathcal{R}_0 > 1$, we replace $x^{(2)}$ with 0, which leads to $A(x_k) = o(\epsilon)I$. Therefore, the same conclusion is achieved by picking a sufficiently small ϵ such that $\lambda_{\min}(\sum_{k=l}^{l+L} A(x_k)) = \lambda_{\min}(o(\epsilon)I) < \alpha$. $\square$

The lack of persistent excitation of SIS models is primarily due to the rank deficiency of the FIM induced from $\phi(x_k)$ as x_k approaches an equilibrium. Therefore, we can expect other compartmental models with the number of parameters p greater than the number of informative observed states n to exhibit a similar lack of persistent excitation. In addition, the period of excitation in spreading models usually associates with the short initial transient states of the infection. For this reason, the relaxed notion of persistent excitation introduced in [9], [10], [14] could potentially be useful for developing an effective adaptive identification algorithm.

Definition 3 (Initial Excitation). *Let* $x_{k+1} = f(x_k)$. *A function* ϕ *is said to be initially exciting with respect to* f *if:*

$$\sum_{k=k_0}^{L} \phi(\Psi(k, k_0, x_0))^\top \phi(\Psi(k, k_0, x_0)) \succeq \alpha I, \qquad (8)$$

for some positive constants L, α, *and initial condition* x_0.

Remark. *Note that* $\sum_{i=0}^{L} \phi(\Psi(k, k_0, x_0))^\top \phi(\Psi(k, k_0, x_0))$ *is the Fisher Information matrix of the empirical cost* (3) *at step* L *when* $\alpha = 1$.

However, initial excitation alone is not sufficient for addressing the challenges of adaptive identification for nonlinear spreading processes, illustrated by the following example.

Example 2. *In this example, we apply the initial excitation-based multi-model adaptive identification (IE-MMAI) algorithm introduced in [10] to the SIS epidemic with process and observation noise (Figure 2). IE-MMAI fails to converge due to a high FIM condition number, which indicates a highly-skewed error contour induced by the model structure as demonstrated in Fig. 1a.*

Models with the highly-skewed error contour are known to be "sloppy" [17] or practically unidentifiable [11] in the literature. Note that, consistent with our discussion in Example 1, the estimated reproduction number converges to the actual value even though the underlying parameters do not.

One natural solution to such a problem is to incorporate second-order information when computing the descent direction upon the arrival of every new data point. We introduce the following modification to the classic recursive least square algorithm to handle the lack of persistent excitation and practical non-identifiability when applying adaptive identification techniques to the SIS parameter estimation problem. The proposed algorithm is detailed in Algorithm 1, and we refer to it as the Greedily-weighted Recursive Least Squares (GRLS) Algorithm. We now introduce the notion
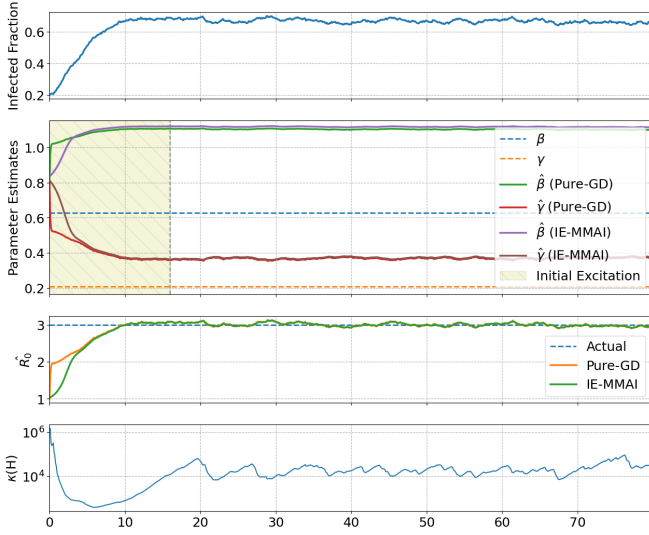
Fig. 2: Plots displaying (from top) the evolution of infected proportion for the SIS model ($\beta = 0.62929$, $\gamma = 0.20976$, $R_0 = 3.0$) with Gaussian process noise; parameter estimates over time from two chosen baseline algorithms: gradient descent (Pure-GD) (Example 1) and initial excitation-based multi-model adaptive identification (IE-MMAI) [10]; and the estimated reproduction number over time using each method; condition number of the Fisher Information Matrix of the empirical cost function (Definition 2) in the log scale over time for $\alpha = 0.94$.

of the *optimal excitation set* and the *greedy excitation set*, inspired by the notion of initial excitation, to illustrate the workings of the GRLS Algorithm.

Definition 4 (Optimal Excitation Set). *A subset of data points $\mathcal{E}$ in $\mathcal{D}_K = \{k \in \mathbb{Z} : k_0 \leq k \leq K\}$ is optimally exciting if:*

$$\mathcal{E}_{opt}(K) = \underset{E \subseteq \mathcal{D}_K}{\arg\min} \, \kappa\Big(\sum_{k \in E} \phi(\Psi(k, k_0, x_0))^\top \phi(\Psi(k, k_0, x_0)) \Big).$$

The main modification of Algorithm 1 to classic recursive least square filtering is that it attempts to store the data points that belong to the optimally exciting set so they will not be diluted by less informative new incoming data. However, solving for the optimally exciting set at each update iteration is expensive. Therefore, we propose a feasible approach to obtain a sub-optimal excitation set via a greedy algorithm.

Definition 5 (Greedy Excitation Set). *The Kth data point belongs to the greedy excitation set $\mathcal{E}_g(K)$ if it does not deteriorate the FIM's condition number induced by $\mathcal{E}_g(K-1)$:*

$$\kappa\Big(\sum_{\substack{k \in \mathcal{E}_g(K-1) \\ \cup \{K\}}} \phi(x_k)^\top \phi(x_k) \Big) \leq \kappa\Big(\sum_{k \in \mathcal{E}_g(K-1)} \phi(x_k)^\top \phi(x_k) \Big).$$

After the introduction of the greedy excitation set, we can briefly summarize the GRLS Algorithm as follows.

- Lines 3-14 in Algorithm 1 update the greedy exciting set and the corresponding hyper-parameters for computing $\hat{\theta}_{k+1}$ upon every incoming datum.
  - Line 4 compares the new condition number with the current condition number.
  - If $\kappa(H^{(e)}) \leq \kappa(H_k^{(e)})$, then we update the regressor matrices $\Phi_{k+1}^{(e)}$, the FIM $H_{k+1}^{(e)}$ and the corrector term $\upsilon_{k+1}^{(e)}$ of the excitation set on Lines 5-6;

Algorithm 1 Greedily-weighted Recursive Least Squares

Require: Input datum x_{k+1}
1: **if** $k < 0$ **then** Initialize $H_0^{(e)}$, P_0, $\hat{\theta}_0$, $\Phi_0^{(e)}$, $\upsilon_0^{(e)}$.
2: **else**
3: $\quad H^{(e)} \leftarrow H_k^{(e)} + \phi(x_k)^\top \phi(x_k)$
4: $\quad$ **if** $\kappa(H^{(e)}) \leq \kappa(H_k^{(e)})$ **then**
5: $\quad\quad \Phi_{k+1}^{(e)}, H_{k+1}^{(e)} \leftarrow [\Phi_k^{(e)^\top}, \phi(x_k)^\top]^\top, H^{(e)}$
6: $\quad\quad \upsilon_{k+1}^{(e)} \leftarrow \upsilon_k^{(e)} + \phi(x_k)^\top (x_{k+1} - x_k)$
7: $\quad\quad \Phi \leftarrow \sqrt{1-\alpha} \Phi_{k+1}^{(e)}$
8: $\quad\quad H, \upsilon \leftarrow (1-\alpha) H_{k+1}^{(e)}, (1-\alpha) \upsilon_{k+1}^{(e)}$
9: $\quad$ **else**
10: $\quad\quad \Phi_{k+1}^{(e)}, H_{k+1}^{(e)}, \upsilon_{k+1}^{(e)} \leftarrow \Phi_k^{(e)}, H_k^{(e)}, \upsilon_k^{(e)}$
11: $\quad\quad \Phi \leftarrow [\sqrt{1-\alpha}\Phi_{k+1}^{(e)^\top}, \phi(x_k)^\top]^\top$
12: $\quad\quad H \leftarrow (1-\alpha) H_{k+1}^{(e)} + \phi(x_k)^\top \phi(x_k)$
13: $\quad\quad \upsilon \leftarrow (1-\alpha) \upsilon_{k+1}^{(e)} + \phi(x_k)^\top (x_{k+1} - x_k)$
14: $\quad$ **end if**
15: $\quad P_{k+1} \leftarrow \frac{1}{\alpha} P_k - \frac{1}{\alpha} P_k \Phi^\top (\alpha I + \Phi P_k \Phi^\top)^{-1} \Phi P_k$
16: $\quad \hat{\theta}_{k+1} \leftarrow \hat{\theta}_k + P_{k+1}(\upsilon - H\hat{\theta}_k)$
17: **end if**

- else, $\Phi_{k+1}^{(e)}, H_{k+1}^{(e)}, \upsilon_{k+1}^{(e)}$ remain unchanged as specified in Line 10.
- Lines 7-8, 11-13 use $\Phi_{k+1}^{(e)}, H_{k+1}^{(e)}, \upsilon_{k+1}^{(e)}$ to compute Φ, H, υ which are needed for computing the inverse Hessian matrix P_{k+1} and the new estimate $\hat{\theta}_{k+1}$.
- Line 15 updates the inverse Hessian matrix with an exponential forgetting factor α.
- Line 16 updates the parameter estimates.

We are ready to characterize the optimality of Algorithm 1.

Theorem 1. *If P_0 is positive definite, then for all $k \in [0, \infty)$, $\hat{\theta}_{k+1}$ obtained by Algorithm 1 is the unique minimizer of the cost function:*

$$C_k(\hat{\theta}_k) = \sum_{i=0}^{k} w_{i,k} \|r_i(\hat{\theta}_k)\|^2 + \alpha^{k+1} \|\hat{\theta}_k - \theta_0\|_{P_0^{-1}}^2 \quad (9)$$

with the weighting function defined as:

$$w_{i,k} = \begin{cases} (1-\alpha) \sum_{l=i}^{k} \alpha^{k-l} & \text{if } i \in \mathcal{E}_g(k), \\ \alpha^{k-i} & \text{otherwise.} \end{cases}$$

Proof. The proof leverages mathematical induction to proceed, and it suffices to show the inductive step. We first note that $C_k(\hat{\theta})$ can be written in terms of: $C_k(\hat{\theta}) = \hat{\theta}^\top A_k \hat{\theta} + 2b_k^\top \hat{\theta} + c_k$, where A_k, b_k, c_k are:

$$A_k := \sum_{i=0}^{k} w_{i,k} \phi(x_i)^\top \phi(x_i) + \alpha^{k+1} P_0^{-1}$$

$$b_k := -\sum_{i=0}^{k} w_{i,k} \phi(x_i)^\top y(x_i) - \alpha^{k+1} P_0^{-1} \theta_0$$

$$c_k := \sum_{i=0}^{k} w_{i,k} y(x_i)^\top y(x_i) + \alpha^{k+1} \theta_0^\top P_0^{-1} \theta_0.$$

171

Then, A_k, b_k can be computed recursively as

$$A_k = \alpha A_{k-1} + \sum_{i \in U} \omega_i \phi(x_i)^\top \phi(x_i)$$

$$b_k = \alpha b_{k-1} - \sum_{i \in U} \omega_i \phi(x_i)^\top y_i,$$

where $U = \mathcal{E}_g(k) \bigcup \{k\}$, and $\omega_i = \begin{cases} (1-\alpha) & \text{if } i \in \mathcal{E}_g(k), \\ 1 & \text{otherwise.} \end{cases}$
By way of induction, assume $\exists\ k \in \mathbb{N}$ such that A_{k-1} is positive definite and the unique optimizer of $C_{k-1}(\hat\theta_{k-1})$ is $\hat\theta_k = -A_{k-1}^{-1} b_{k-1}$. We define $P_{k+1} := A_k^{-1}$. Since A_{k-1} is positive definite, we can apply the matrix inversion lemma [18, p, 304] and obtain a positive definite P_{k+1}:

$$P_{k+1} = A_k^{-1}$$

$$= \frac{1}{\alpha} \left(A_{k-1} + \frac{1}{\alpha} \left(\sum_{i \in U} \omega_i \phi(x_i)^\top \phi(x_i) \right) \right)^{-1}$$

$$= \frac{1}{\alpha} P_k - \frac{1}{\alpha} P_k \Phi^\top (\alpha I + \Phi P_k \Phi^\top)^{-1} \Phi P_k,$$

where $\Phi = \begin{cases} \sqrt{1-\alpha}\Phi_{k+1}^{(e)} & \text{if } k \in \mathcal{E}_g(k), \\ [\sqrt{1-\alpha}(\Phi_{k+1}^{(e)})^\top, \phi(x_k)^\top]^\top & \text{otherwise,} \end{cases}$
and $\Phi_{k+1}^{(e)} = [\phi(x_{k_1})^\top, \phi(x_{k_2})^\top, \ldots, \phi(x_{k_n})^\top]_{k_i \in \mathcal{E}_g(k)}^\top$,
which satisfies the computation of $\Phi_{k+1}^{(e)}$ and Φ on Lines 5, 7, 10, and 11 of Algorithm 1. By the quadratic minimization lemma [18], the unique minimizer of $C_k(\hat\theta)$ is:

$$\hat\theta_{k+1} = -A_k^{-1} b_k$$

$$= A_k^{-1} \left(-\alpha b_{k-1} + \sum_{i \in U} \omega_i \phi(x_i)^\top y_i \right)$$

$$= A_k^{-1} \left(\alpha A_{k-1} \hat\theta_k + \sum_{i \in U} \omega_i \phi(x_i)^\top y_i \right)$$

$$= A_k^{-1} \left(\left(A_k - \sum_{i \in U} \omega_i \phi(x_i)^\top \phi(x_i) \right) \hat\theta_k + \sum_{i \in U} \omega_i \phi(x_i)^\top y_i \right)$$

$$= \hat\theta_k + A_k^{-1} \left(\sum_{i \in U} \omega_i \phi(x_i)^\top y_i - \left(\sum_{i \in U} \omega_i \phi(x_i)^\top \phi(x_i) \right) \hat\theta_k \right)$$

$$= \hat\theta_k + P_{k+1}(v - H\hat\theta_k),$$

where $v := \sum_{i \in U} \omega_i \phi(x_i)^\top y_i$ and $H := \sum_{i \in U} \omega_i \phi(x_i)^\top \phi(x_i)$ matches the computation of v and H on Lines 8, 12, and 13 of Algorithm 1. By the principle of mathematical induction, P_n is positive definite and $\hat\theta_{n+1}$ is the unique minimizer of $C_n(\hat\theta_n)$ for all $n \in \mathbb{N}$. $\square$

While Theorem 1 characterizes the cost function which Algorithm 1 optimizes, the structure of the weighting function $w_{i,k}$ might not be immediately obvious. The following corollary clarifies the intuition behind $w_{i,k}$.

Corollary 2. *If $\alpha < 1$, P_0 is positive definite, and the cardinality of the excitation set in the limit, $\lim_{k \to \infty} |\mathcal{E}_g(k)|$, is finite, then, as $k \to \infty$, $\hat\theta_{k+1}$ obtained by Algorithm 1 is the unique minimizer of the cost function:*

$$C_k(\hat\theta) = \sum_{i=0}^{k} W_{i,k} \|r_i(\hat\theta)\|^2, \tag{10}$$

with

$$W_{i,k} = \begin{cases} 1 & \text{if } i \in \mathcal{E}_g(K) \\ \alpha^{k-i} & \text{otherwise.} \end{cases} \tag{11}$$

Proof. Note that, since $\alpha < 1$, the second term in the cost function in (9), $\alpha^{k+1} \|\hat\theta - \theta_0\|_{P_0^{-1}}^2$, goes to zero as $k \to \infty$. Furthermore, $\sum_{l=i}^{k} \alpha^{k-l}$ can be rewritten as $\sum_{l=0}^{k-i} \alpha^l$, and, since $\sum_{l=0}^{k-i} \alpha^l$ is a geometric sum, $\alpha < 1$, and $i \in \mathcal{E}_g(k)$ is finite by assumption, $\lim_{k \to \infty}(1-\alpha)\sum_{l=i}^{k} \alpha^{k-l} = 1$. Thus, the weighting function $w_{i,k}$ can be written as $W_{i,k}$ in (11). Therefore, by Theorem 1, Algorithm 1 obtains the optimal $\theta^* = \arg\min_{\hat\theta} C_k(\hat\theta)$ for the cost function in (10)-(11). $\square$

Corollary 2 characterizes the asymptotic behavior of the cost function which Algorithm 1 minimizes. The algorithm stores the data points which belong to the greedy excitation set by giving them a weight of 1 while assigning the rest of the unexciting points exponentially decaying weights.

IV. SIMULATIONS

We present parameter estimation results for both the noise-free case and including Gaussian process and observation noise for the SIS dynamics in (6). In Fig. 3, we compare the performance of the initial excitation approach, IE-MMAI [10], and a basic RLS with an exponential forgetting (EF-RLS), to the performance of the GRLS Algorithm we propose in this work. The same initial estimates of parameters, $\theta_0 = [\beta_0, \gamma_0]^\top = [1, 1]^\top$, are used for all algorithms with the exception of IE-MMAI, for which the $m = 3$ models were initialized randomly around θ_0. IE-MMAI, designed for only
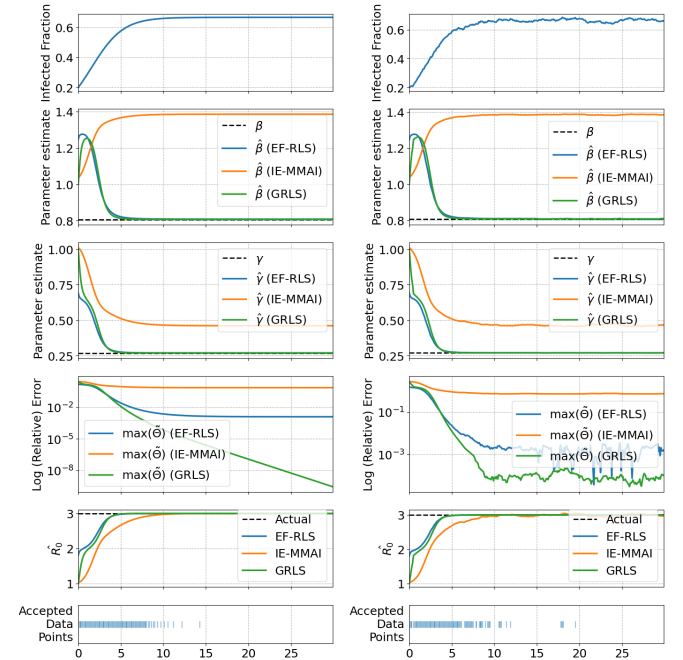

Fig. 3: Comparing the performance of EF-RLS, IE-MMAI, and GRLS: SIS Simulation ($\beta = 0.8076$, $\gamma = 0.2692$); parameter estimates over time for both a noise-free and a noisy simulation; maximum relative error in log scale for parameter estimates over time; reproduction number estimates over time; and data indices accepted into $\mathcal{E}_g$ by GRLS. Left panel shows the results without noise while the right panel shows it in presence of noise.

LTI systems and as a gradient descent-like first-order method, is justifiably sensitive to the choice of the initial parameter estimates, and often fails to converge due to the poor practical identifiability of SIS models, as discussed in Examples 1 and 2. Note that the estimated reproduction numbers still converge to the actual value despite the lack of convergence of the parameter estimates themselves, which is consistent with our discussion of SIS practical identifiability in Section II-2. On the other hand, even in the presence of noise, the parameter estimates converge for both EF-RLS and GRLS.

Data accepted into the GRLS greedy excitation set (5) used to construct the main regressor are diagrammatically depicted in the lowermost block of Fig. 3. A majority of points accepted by the algorithm is in the transient rise of states before the equilibrium is reached. The beginning of the epidemic garners a critical amount of information about the epidemic parameters.

Note that while EF-RLS is comparable in performance to GRLS for the noise-free case (left panel of Fig. 3), it becomes increasingly oscillatory upon losing excitation in the noisy case (right panel of Fig. 3). A closer look at the covariance matrix P_k in both algorithms reveals a steady increase in the condition number and maximum eigenvalue of P_k in EF-RLS, while those of GRLS saturate due to its tendency to avoid picking up non-exciting data points (see Fig. 4). This phenomenon in EF-RLS is observed in both the noise-free and noisy cases and is known in literature as covariance windup [19], which occurs when a non-unity forgetting factor in EF-RLS causes P_k to get closer to a singular matrix upon losing persistence of excitation. The linear increase in the maximum eigenvalue of P_k is consistent with past analysis [20]. Upon running the parameter estimation task for longer times, the EF-RLS estimates diverge.
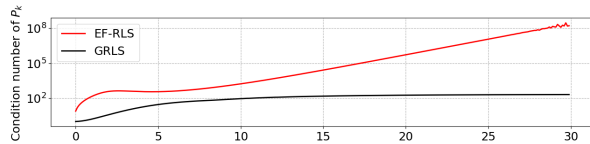

Fig. 4: Plot of the condition number of the covariance matrix P_k against time for both EF-RLS and GRLS running on data from the noise-free SIS simulation (Fig. 3).

While there are other ways to combat covariance windup in RLS and Kalman filtering applications [20], [21], our approach seeks to mitigate the issues uncovered by Example 2 as a first step to discover even more effective methods for online parameter estimation.

V. Conclusion

We have highlighted two problems that plague the application of adaptive identification tools to SIS models: the lack of persistence of excitation and the practical non-identifiability of epidemic models. We propose a novel algorithm (GRLS) based on recursive least squares and use the concept of initial excitation to construct an exciting set for the regressor. The GRLS Algorithm has superior performance compared to conventional algorithms and is able to identify epidemic parameters in the SIS model with process and observation noise. In particular, while estimates from EF-RLS become oscillatory in the presence of noise, eventually diverging, GRLS is able to maintain a stable estimate. Future work includes extending GRLS to estimate epidemic parameters on networks, time-varying systems, and other epidemic compartmental models, with the eventual goal of performing adaptive identification using real testing data.

References

[1] W. O. Kermack, A. G. McKendrick, and G. T. Walker, "A contribution to the mathematical theory of epidemics," *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, vol. 115, no. 772, pp. 700–721, 1927.

[2] G. Giordano, F. Blanchini, R. Bruno, P. Colaneri, A. Di Filippo, A. Di Matteo, and M. Colaneri, "Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy," *Nature Medicine*, vol. 26, no. 6, pp. 855–860, 2020.

[3] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, "Epidemic processes in complex networks," *Reviews of Modern Physics*, vol. 87, no. 3, p. 925, 2015.

[4] P. Ioannou and B. Fidan, *Adaptive Control Tutorial*. SIAM, 2006.

[5] S. A. U. Islam and D. S. Bernstein, "Recursive least squares for real-time implementation [lecture notes]," *IEEE Control Systems Magazine*, vol. 39, no. 3, pp. 82–85, 2019.

[6] K. S. Narendra and A. M. Annaswamy, "Persistent excitation in adaptive systems," *Int. Journal of Control*, vol. 45, no. 1, pp. 127–160, 1987.

[7] P. Ioannou and J. Sun, "Theory and design of robust direct and indirect adaptive-control schemes," *Int. Journal of Control*, vol. 47, no. 3, pp. 775–813, 1988.

[8] P. A. Ioannou and J. Sun, *Robust Adaptive Control*, 1996, vol. 1.

[9] S. K. Jha, S. B. Roy, and S. Bhasin, "Initial excitation-based iterative algorithm for approximate optimal control of completely unknown LTI systems," *IEEE Transactions on Automatic Control*, vol. 64, no. 12, pp. 5230–5237, 2019.

[10] A. Dhar, S. B. Roy, and S. Bhasin, "Initial excitation based discrete-time multi-model adaptive online identification," *European Journal of Control*, vol. 68, p. 100672, 2022.

[11] N. Cunniffe, F. Hamelin, A. Iggidr, A. Rapaport, and G. Sallet, "Observability, identifiability and epidemiology: A survey," *arXiv preprint arXiv:2011.12202*, 2023.

[12] F.-G. Wieland, A. L. Hauber, M. Rosenblatt, C. Tönsing, and J. Timmer, "On structural and practical identifiability," *Current Opinion in Systems Biology*, vol. 25, pp. 60–69, 2021.

[13] B. Prasse and P. Van Mieghem, "Predicting network dynamics without requiring the knowledge of the interaction graph," *Proceedings of the National Academy of Sciences*, vol. 119, no. 44, p. e2205517119, 2022.

[14] E. Panteley, A. Loria, and A. Teel, "Relaxed persistency of excitation for uniform asymptotic stability," *IEEE Transactions on Automatic Control*, vol. 46, no. 12, pp. 1874–1886, 2001.

[15] K. S. Narendra and A. M. Annaswamy, *Stable Adaptive Systems*. Courier Corporation, 2012.

[16] M. K. Transtrum, B. B. Machta, K. S. Brown, B. C. Daniels, C. R. Myers, and J. P. Sethna, "Perspective: Sloppiness and emergent theories in physics, biology, and beyond," *The Journal of Chemical Physics*, vol. 143, no. 1, p. 07B201_1, 2015.

[17] M. K. Transtrum, B. B. Machta, and J. P. Sethna, "Geometry of nonlinear least squares with applications to sloppy models and optimization," *Physical Review E*, vol. 83, no. 3, p. 036701, 2011.

[18] D. S. Bernstein, *Matrix Mathematics*. Princeton University Press, 2009.

[19] T. Fortescue, L. S. Kershenbaum, and B. E. Ydstie, "Implementation of self-tuning regulators with variable forgetting factors," *Automatica*, vol. 17, no. 6, pp. 831–835, 1981.

[20] L. Cao and H. Schwartz, "The Kalman filter based recursive algorithm: Windup and its avoidance," in *Proceedings of the American Control Conference*, 2001, pp. 3606–3611.

[21] B. Lai and D. S. Bernstein, "Exponential resetting and cyclic resetting recursive least squares," *IEEE Control Systems Letters*, vol. 7, pp. 985–990, 2023.