System Identification of Spatiotemporally Invariant Systems

Azeem Sarwar, Petros G. Voulgaris, and Srinivasa M. Salapaka

Abstract— We present a distributed projection algorithm for system identification of spatiotemporally invariant systems. Each subsystem communicates only with its immediate neighbor to share its current estimate along with a cumulative improvement index. Based on the cumulative improvement index, the best estimate available is picked in order to carry out the next iterate. For small estimation error, the scheme switches over to a "smart" averaging routine. The proposed algorithm guarantees to bring the local estimates arbitrarily close to one another. We demonstrate that the proposed scheme has a clear advantage over the standard projection algorithm and is amenable to indirect distributed adaptive control of spatiotemporally invariant systems. Our proposed algorithm is also suitable to address the estimation problem in distributed networks that arise in a variety of applications, such as environment monitoring, target localization and potential sensor network problems.

I. Introduction

In this article we look at the problem of system identification of spatiotemporal systems from a control theoretic perspective. In particular, we restrict our focus to a certain class of spatiotemporal systems that are spatially invariant. Spatial invariance is a strong property of a given system, which means that the dynamics of the system do not vary as we translate along some spatial axis. Spatially invariant systems have received considerable attention recently, and control design of such systems has been worked out in some detail (see e.g. [1], and [2]). Many distributed systems, for example array of identical microcantilevers to be employed in Atomic Force Microscope (AFM) application [3] or large segmented telescope [4], can be approximated as spatially invariant.

A lot of recent work has been done in the area of distributed parameter estimation. Incremental adaptive strategies have been proposed in [5] that require a cyclic pattern of collaboration among the subsystems (nodes). Each subsystem updates the estimate that it receives from its predecessor based on the information available to it and passes it to its successor to do the same. The same authors have proposed diffusion techniques in [6] that requires more network connectivity. Each subsystem combines its current estimate with the estimate of its neighbor, based on some performance criterion, to come up with an aggregate. This aggregate is then used for carrying out the next estimation update. A similar space-time diffusion approach can be found in [7]. From a control theoretic perspective, we are interested in ascertaining if the following is achieved for a distributed system:

1) estimation error (difference between actual and predicted information) goes to zero, regardless of the convergence of estimates to the true value
2) estimates get close to each other arbitrarily as time grows (at least locally) allowing one to possibly employ results of [8] for adaptation

Most of the work cited above (and the references therein) concerns with the convergence of the estimates to the true parameter and seldom address the above mentioned objectives. Our motivation for this paper is exactly summarized in the above mentioned objectives, with the underlying system being spatiotemporally invariant. What we propose in this paper is a distributed projection algorithm for system identification of spatiotemporally invariant systems that builds on the standard projection algorithm. The main idea, however, can be applied to other gradient based schemes such as least squares. Our goal is to achieve the above mentioned objectives together with providing better performance than the standard projection algorithm if run independently on every subsystem.

A. Setup

The basic setup is captured in Figure 1, where an infinite string of interconnected subsystems is shown. P_i refers to the i_{th} subsystem, and u_i, y_i refer to the respective input and output. The subsystem P_i is taken as a single input single output linear time invariant. We note here that all subsystems have identical dynamics, however, they may be operating independently. Equivalently, $P_i = P_j \ \forall \ i, j$.
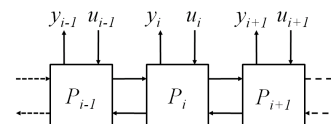
Azeem Sarwar is with the Department of Mechanical Science and Engineering, University of Illinois at Urbana Champaign, USA asarwar@illinois.edu

Petros G. Voulgaris is with the Faculty of Aerospace Engineering and with the Coordinated Sciences Laboratory, University of Illinois at Urbana Champaign, USA, and a Visiting ADGAS Chaired Professor, Petroleum Institute, Mechanical Engineering, Abu Dhabi, UAE. voulgari@illinois.edu

Srinivasa M. Salapaka is with the Faculty of Mechanical Science and Engineering, University of Illinois at Urbana Champaign, USA salapaka@illinois.edu

Fig. 1.   Basic Setup

B. Notations

The set of reals is denoted by $\mathbb{R}$ and the set of integers is denoted by $\mathbb{Z}$. The set of non-negative integers is denoted by $\mathbb{Z}^+$. The space l_2 denotes the set of sequences $f = \{f(t)\}_{t=0}^{t=\infty}$ with $\|f\|_2 := (\sum_t |f(t)|^2)^{1/2} < \infty$. The norm $(\|\cdot\|)$ throughout this paper is taken to be Euclidian. $\mathscr{B}(x,r)$ denotes a ball in l_2 space with center x and radius r.

The paper is organized as follows: Section II presents the main idea of the distributed projection algorithm along with some simulation results. We conclude our discussion in Section III along with some remarks on the future work.

II. DISTRIBUTED PROJECTION ALGORITHM

We assume that the deterministic dynamical system can be described by a model that may be expressed succinctly in the following simple form:

$$y_i(t) = \phi_i(t-1)^T \theta_0 \tag{1}$$

where $y_i(t)$ denotes the (scalar) system output of subsystem i at time 't', $\phi_i(t-1)$ denotes a vector that is a linear function of

$$
\begin{aligned}
Y(t) &= [\{y_i(t-1), y_i(t-2), \cdots\}, \{y_{i-1}(t-1), y_{i-1}(t-2), \cdots\}, \\
&\quad \{y_{i+1}(t-1), y_{i+1}(t-2), \cdots\}, \cdots] \\
U(t) &= [\{u_i(t-1), u_i(t-2), \cdots\}, \{u_{i-1}(t-1), u_{i-1}(t-2), \cdots\}, \\
&\quad \{u_{i+1}(t-1), u_{i+1}(t-2), \cdots\}, \cdots]
\end{aligned}
$$

θ_0 denotes a parameter vector that is unknown. We introduce the following notations that shall be used frequently in the sequel. $\hat{\theta}_i(t-1)$ is the estimate of the true parameter vector θ_0 at time 't' available at the i_{th} subsystem. $\hat{\theta}(0)$ is the initial estimate/guess which is given.

A. Main Idea

The main idea of the algorithm is depicted in Figure 2. We assume that all the subsystems start with the same initial guess $\hat{\theta}_i(0) = \hat{\theta}(0)$. Each subsystem essentially runs the same algorithm which is being described in this section. At time '$t+1$', subsystem i has access to its own estimate $\hat{\theta}_i(t)$, and the estimate of its immediate neighbors $\hat{\theta}_{i-1}(t)$, and $\hat{\theta}_{i+1}(t)$. Before updating its current estimate based on the information available at time '$t+1$', subsystem i resets its current estimate to the best available estimate from the set $\{\hat{\theta}_i(t), \hat{\theta}_{i-1}(t), \hat{\theta}_{i+1}(t)\}$. This estimate, call it $\hat{\theta}^i(t)$, is then taken as the current estimate by the subsystem i in order to establish the next estimate. If there is a tie i.e. there are more than one estimates that qualify as the best, an average is taken. In the sequel, the best estimate available to the i_{th} subsystem for next iterate shall be referred to as $\hat{\theta}^i(\cdot)$. The update equation of the algorithm is given in Equation (2). The cumulative improvement index helps in identifying the best estimate available.

$$
\begin{aligned}
\hat{\theta}_i(t+1) &= \\
&\hat{\theta}^i(t) + \frac{a(t)\phi_i(t)}{c + \phi_i(t)^T \phi_i(t)} \left[y_i(t+1) - \phi_i(t)^T \hat{\theta}^i(t) \right] \tag{2}
\end{aligned}
$$
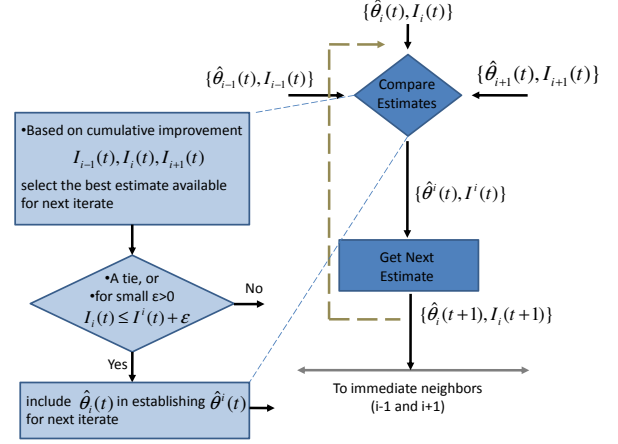

Fig. 2. Flowchart depicting the main idea of distributed projection algorithm

where

$$
\hat{\theta}^i(t) \in \left\{ \hat{\theta}_i(t), \hat{\theta}_{i-1}(t), \hat{\theta}_{i+1}(t), \frac{\hat{\theta}_i(t) + \hat{\theta}_{i-1}(t)}{2}, \frac{\hat{\theta}_i(t) + \hat{\theta}_{i+1}(t)}{2}, \right.
$$
$$
\left. \frac{\hat{\theta}_{i+1}(t) + \hat{\theta}_{i-1}(t)}{2}, \frac{\hat{\theta}_i(t) + \hat{\theta}_{i-1}(t) + \hat{\theta}_{i+1}(t)}{3} \right\}
$$

where $c > 0$, and $0 < a(t) < 2$.

B. Cumulative Improvement Index $I_i(\cdot)$

Each subsystem shares its current estimate along with the associated cumulative improvement index with its immediate neighbors. Cumulative improvement index $I_i(\cdot)$ is a measure of how much an estimate has improved from the original guess $\hat{\theta}(0)$. In order to explain the cumulative improvement index $I_i(\cdot)$ mathematically, let us first introduce some notations.

$$
\begin{aligned}
\tilde{\theta}^i(t) &:= \hat{\theta}^i(t) - \theta_0 \tag{3} \\
\tilde{\theta}_i(t) &:= \hat{\theta}_i(t) - \theta_0 \tag{4} \\
e_i(t) &:= y_i(t) - \phi_i(t-1)^T \hat{\theta}^i(t-1) \\
&= -\phi_i(t-1)^T \tilde{\theta}^i(t-1) \tag{5}
\end{aligned}
$$

We shall refer to $e_i(t)$ as the estimation error. Consider Equation (2). Subtracting θ_0 from both sides of the (2), and using (1) along with (3), we obtain

$$
\tilde{\theta}_i(t+1) = \tilde{\theta}^i(t) - \frac{a(t)\phi_i(t)}{c + \phi_i(t)^T \phi_i(t)} \phi_i(t)^T \tilde{\theta}^i(t) \tag{6}
$$

Using (5), we have

$$
\|\tilde{\theta}_i(t+1)\|^2 - \|\tilde{\theta}^i(t)\|^2 =
$$
$$
a(t) \left[-2 + \frac{\phi_i(t)^T \phi_i(t)}{c + \phi_i(t)^T \phi_i(t)} \right] \frac{e_i(t+1)^2}{c + \phi_i(t)^T \phi_i(t)} \tag{7}
$$

Note that

$$a(t)\left[-2+\frac{\phi_i(t)^T\phi_i(t)}{1+\phi_i(t)^T\phi_i(t)}\right]<0 \qquad (8)$$

Expanding $\left\|\tilde{\theta}^i(t)\right\|$, and its successive terms, we can write

$$\begin{aligned}
\left\|\tilde{\theta}_i(t+1)\right\|^2 &= \left\|\tilde{\theta}(0)\right\|^2 + \\
&\sum_{j=1}^{t} a(t)\left[-2+\frac{\phi_k(j-1)^T\phi_k(j-1)}{c+\phi_k(j-1)^T\phi_k(j-1)}\right]\frac{e_k(j)^2}{c+\phi_k(j-1)^T\phi_k(j-1)} \\
&+a(t)\left[-2+\frac{\phi_i(t)^T\phi_i(t)}{c+\phi_i(t)^T\phi_i(t)}\right]\frac{e_i(t+1)^2}{c+\phi_i(t)^T\phi_i(t)}
\end{aligned} \qquad (9)$$

where the subscript k captures the evolution path of $\hat{\theta}^i(t)$, i.e. the sequence of the subsystems that were involved in establishing $\hat{\theta}^i(t)$. We now define the cumulative improvement index in the following way:

$$\begin{aligned}
I^i(t) &:= \\
&\sum_{j=1}^{t} a(t)\left[-2+\frac{\phi_k(j-1)^T\phi_k(j-1)}{c+\phi_k(j-1)^T\phi_k(j-1)}\right]\frac{e_k(j)^2}{c+\phi_k(j-1)^T\phi_k(j-1)}
\end{aligned} \qquad (10)$$

From (10) we have

$$I_i(t+1) = I^i(t) + a(t)\left[-2+\frac{\phi_i(t)^T\phi_i(t)}{c+\phi_i(t)^T\phi_i(t)}\right]\frac{e_i(t+1)^2}{c+\phi_i(t)^T\phi_i(t)} \qquad (11)$$

It is clear from Equation (10), that $I_i(t+1) \leq 0$. Its magnitude is exactly the square of the distance $\hat{\theta}_i(t+1)$ has traveled from $\hat{\theta}(0)$ towards θ_0. Another thing to note is that $I_i(t+1) \leq I_i(t)$. As evident from Equation (9), the cumulative improvement index is just a scalar that can be updated iteratively at each time step. The bigger the magnitude of the cumulative improvement index, the better an estimate is.

C. The ε-Rule:

The key idea of the algorithm is that we can write any estimate as the sum of the initial guess and some improvement terms. The cumulative improvement index is a sum of nonpositive real numbers. At the same time, $\left\|\tilde{\theta}_i(t)\right\|$ is a bounded function (bounded below by 0, and above by $\left\|\tilde{\theta}(0)\right\|$). As $t \to \infty$, the cumulative improvement index $I_i(t)$ will contain infinitely many nonpositive terms. Since, however, $\left\|\tilde{\theta}_i(t)\right\|$ is bounded, $\lim_{t\to\infty} I_i(t)$ should exist. This implies that the improvement in estimates will get slower and slower as time grows. We note here that it is possible, for example, to have $\left\|\tilde{\theta}_i(t)\right\| \approx \left\|\tilde{\theta}_{i-1}(t)\right\|$, but $\hat{\theta}_i(t)$ and $\hat{\theta}_{i-1}(t)$ could still be far apart. This is depicted in Figure 3. For such a situation, when the neighboring estimates (in space as well as time) are at almost 'equal distance' from the true value but could still be far apart, we introduce the ε-rule, which is as follows:
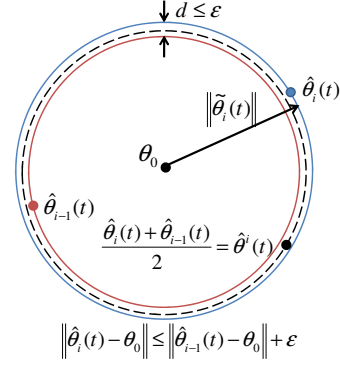

Fig. 3. Figure Showing ε-Rule

- There is a small number $\varepsilon > 0$ specified priori and known to every one.
- At a given time t, if $\left\|\tilde{\theta}_i(t)\right\| \leq \left\|\tilde{\theta}^i(t)\right\| + \varepsilon$, redefine $\hat{\theta}^i(t)$ to include $\hat{\theta}_i(t)$ (if it was not already present). For example, if $\hat{\theta}^i(t) = \hat{\theta}_{i-1}(t)$, redefine it as $\hat{\theta}^i(t) = \frac{\hat{\theta}_{i-1}(t)+\hat{\theta}_i(t)}{2}$ etc.

The purpose of the ε-rule is to bring the estimates close to each other locally when their respective rate of improvement slows down. At the same time, this rule makes sure that the next estimate will be as good as the current if not better in terms of its distance from the true value. The algorithm, therefore, guarantees improvement throughout its execution making sure that the identification error at each subsystem will eventually become small. The proposed estimation scheme also serves to bring the estimates 'close' to one another asymptotically as shown below.

D. Properties of Distributed Projection Algorithm

The properties of distributed projection algorithm are summarized below.

Lemma 1: For the algorithm (2) and subject to (1) it follows that

1) $\left\|\hat{\theta}_i(t)-\theta_0\right\| \leq \left\|\hat{\theta}^i(t-1)-\theta_0\right\| \leq \left\|\hat{\theta}(0)-\theta_0\right\| \quad t \geq 1$
2) $\lim_{t\to\infty} \frac{e_i(t)}{[c+\phi_i(t-1)^T\phi_i(t-1)]^{1/2}} = 0$
3) $\lim_{t\to\infty} \left\|\hat{\theta}_i(t)-\hat{\theta}_k(t)\right\| = 0$ for $k \in \{i-1,\ i+1\}$
4) $\lim_{t\to\infty} \left\|\hat{\theta}_i(t)-\hat{\theta}_i(t-1)\right\| = 0 \ \forall\ i$. This together with 3) implies

$$\lim_{t\to\infty} \left\|\hat{\theta}_i(t)-\hat{\theta}_k(t+l)\right\| = 0 \text{ for } k \in \{i,\ i-1,\ i+1\}$$

and for finite l

Proof: The proof of 1) is evident from the discussion provided above.

For 2), we observe that $\left\|\tilde{\theta}_i(t)\right\|^2$ is a bounded nonincreasing function , and by summing (7) we get

$$\|\tilde{\theta}_i(t)\|^2 = \|\tilde{\theta}(0)\|^2$$
$$+\sum_{j=1}^{t-1} a(t)\left[-2+\frac{\phi_k(j-1)^T\phi_k(j-1)}{c+\phi_k(j-1)^T\phi_k(j-1)}\right]\frac{e_k(j)^2}{c+\phi_k(j-1)^T\phi_k(j-1)}$$
$$+a(t)\left[-2+\frac{\phi_i(t-1)^T\phi_i(t-1)}{c+\phi_i(t-1)^T\phi_i(t-1)}\right]\frac{e_i(t)^2}{c+\phi_i(t-1)^T\phi_i(t-1)} \quad (12)$$

Since $\|\tilde{\theta}_i(t)\|^2$ is nonnegative and (8) holds, we conclude that 2) holds.

In order to prove 3), first note the following:

1) Given $\varepsilon_0 > 0 \; \exists \; T_{\varepsilon_0} \mid \forall \, t \geq T_{\varepsilon_0}$, and $\forall \, i$ we have

$$\left|\frac{e_i(t)}{[c+\phi_i(t-1)^T\phi_i(t-1)]^{1/2}}\right| \leq \varepsilon_0 \quad (13)$$

i.e. the (normalized) estimation error will be small (and shall remain small) for all subsystems after time T_{ε_0}.

2) Given $\varepsilon_1 > 0 \; \exists \; T_{\varepsilon_1} \mid \forall \, t \geq T_{\varepsilon_1} \; \forall i$, we have

$$\left|\|\tilde{\theta}_i(t-1)\| - \|\tilde{\theta}_i(t)\|\right| \leq \varepsilon_1 \quad (14)$$

i.e. the improvement in the parameter estimates will be small (and shall remain small) for all subsystems after time T_{ε_1}. Also note that if $\varepsilon_1 \leq \varepsilon$, where ε is the constant used for ε-rule, then ε-rule will stay in place $\forall \, t \geq T_{\varepsilon_1}$.

3) if $\|\tilde{\tilde{\theta}}_i(t)\| = \|\tilde{\theta}^i(t)\|$, and $\hat{\theta}_i(t) \neq \hat{\theta}^i(t)$, then:

$$\left\|\theta_0 - \frac{\hat{\theta}_i(t)+\hat{\theta}^i(t)}{2}\right\| < \frac{1}{2}\|\theta_0 - \theta_i(t)\|$$
$$+\frac{1}{2}\|\theta_0 - \theta^i(t)\|$$

(triangle inequality is strict since the estimates are not not aligned)

$$= \|\theta_0 - \theta_i(t)\| = \|\theta_0 - \theta^i(t)\| \quad (15)$$

To see the main idea, let $\hat{\theta}^i(t) = \frac{\hat{\theta}_i(t)+\hat{\theta}_{i-1}(t)}{2}$, and $\hat{\theta}_i(t) \neq \hat{\theta}_{i-1}(t)$ then

$$\|\tilde{\theta}_i(t+1)\|^2 = \left\|\frac{\tilde{\theta}_i(t)+\tilde{\theta}_{i-1}(t)}{2}\right\|^2$$
$$+a(t+1)\left[-2+\frac{\phi_i(t)^T\phi_i(t)}{c+\phi_i(t)^T\phi_i(t)}\right]\underbrace{\frac{e_i(t+1)^2}{c+\phi_i(t)^T\phi_i(t)}}_{\leq \varepsilon_0^2} \quad (16)$$

$$\|\tilde{\theta}_i(t+1)\|^2 \leq \left\|\frac{\tilde{\theta}_i(t)+\tilde{\theta}_{i-1}(t)}{2}\right\|^2 < \|\tilde{\theta}_i(t)\|^2 \quad (17)$$

Since $\left|\|\tilde{\theta}_i(t)\| - \|\tilde{\theta}_i(t+1)\|\right| \leq \varepsilon_1$, the improvement caused by averaging should be less than ε_1. We will show below that if the estimates are far enough then the improvement caused by averaging will exceed ε_1.

For $t \geq \max\{T_{\varepsilon_1}, T_{\varepsilon_0}\}$ and $\varepsilon_1 \leq \varepsilon$, we demonstrate below that there will be a share of the estimates from the immediate neighbors at the i_{th} subsystem, or the immediate neighbors will have a share of the estimate from i_{th} subsystem. We have one of the following possibilities for the subsystem i:

1) $\hat{\theta}^i(t) = \hat{\theta}_i(t)$ in which case we will have $\hat{\theta}^{i+1}(t) \in \{\frac{\hat{\theta}_{i+1}(t)+\hat{\theta}_i(t)}{2}, \frac{\hat{\theta}_{i+1}(t)+\hat{\theta}_i(t)+\hat{\theta}_{i+2}(t)}{3}\}$, and $\hat{\theta}^{i-1}(t) \in \{\frac{\hat{\theta}_{i-1}(t)+\hat{\theta}_i(t)}{2}, \frac{\hat{\theta}_{i-1}(t)+\hat{\theta}_i(t)+\hat{\theta}_{i-2}(t)}{3}\}$

2) $\hat{\theta}^i(t) = \frac{\hat{\theta}_i(t)+\hat{\theta}_{i-1}(t)}{2}$ in which case we will have $\hat{\theta}^{i+1}(t) \in \{\frac{\hat{\theta}_{i+1}(t)+\hat{\theta}_i(t)}{2}, \frac{\hat{\theta}_{i+1}(t)+\hat{\theta}_i(t)+\hat{\theta}_{i+2}(t)}{3}\}$

3) $\hat{\theta}^i(t) = \frac{\hat{\theta}_i(t)+\hat{\theta}_{i+1}(t)}{2}$ in which case we will have $\hat{\theta}^{i-1}(t) \in \{\frac{\hat{\theta}_{i-1}(t)+\hat{\theta}_{i-2}(t)}{2}, \frac{\hat{\theta}_{i-1}(t)+\hat{\theta}_i(t)+\hat{\theta}_{i-2}(t)}{3}\}$,

4) $\hat{\theta}^i(t) = \frac{\hat{\theta}_{i+1}(t)+\hat{\theta}_i(t)+\hat{\theta}_{i-1}(t)}{3}$

We will provide an upper bound on $\|\hat{\theta}_i(t) - \hat{\theta}_k(t)\|$, where $k \in \{i-1, i+1\}$. Note that the farther apart the estimates are, the closer the resulting averaged estimate shall be to the true parameter. We shall assume that we have situation 4) as this covers the rest (1-3). This becomes obvious if we further assume that $\hat{\theta}_i(t) = \hat{\theta}_{i+1}(t)$ and let $e_i(t) = 0$. This ensures that the improvement can be attributed to the process of averaging alone. We also assume that $\|\tilde{\theta}_i(t)\| = \|\tilde{\theta}_{i-1}(t)\|$. Our goal is to see how far $\hat{\theta}_{i-1}(t)$ can be from $\hat{\theta}_i(t)$ so that the resulting improvement does not exceed ε_1. The calculation is presented in Figure 4. As shown in Figure 4(a), we are interested in establishing the distance c. It is, however, easier to calculate the distance $\bar{c}$ where $\bar{c} > c$ (see Figure 4(b)). Solving the right triangle $AB\theta_0$ we see that $\bar{c} = 2\sqrt{3a\varepsilon_1 - 9/4\varepsilon_1^2}$. Clearly, this distance is controlled by ε_1. We, therefore, conclude that $\|\hat{\theta}_i(t) - \hat{\theta}_k(t)\| \leq 2\sqrt{3a\varepsilon_1 - 9/4\varepsilon_1^2}$, where $k \in \{i-1, i+1\}$. To prove 4) we can bound the successive iterates at subsystem i as follows:

$$\|\hat{\theta}_i(t+1) - \hat{\theta}_i(t)\|$$
$$\leq \left\|\frac{\hat{\theta}_i(t)+\hat{\theta}_{i-1}(t)+\hat{\theta}_{i+1}(t)}{3} + \frac{a(t+1)\phi_i(t)}{c+\phi_i(t)^T\phi_i(t)}e_i(t+1) - \hat{\theta}_i(t)\right\|$$
$$(18)$$
$$\leq \frac{1}{3}\|\hat{\theta}_{i-1}(t) - \hat{\theta}_i(t)\| + \frac{1}{3}\|\hat{\theta}_{i+1}(t) - \hat{\theta}_i(t)\| + \varepsilon_0 \quad (19)$$
$$\leq \frac{4}{3}\sqrt{3a\varepsilon_1 - 9/4\varepsilon_1^2} + \varepsilon_0 \quad (20)$$

Similarly we can calculate an upper bound on $\|\hat{\theta}_i(t+1) - \hat{\theta}_k(t)\|$ for $k \in \{i-1, i+1\}$ controlled by ε_1, and ε_0. Clearly as ε_0 and ε_1 approach zero with $t \to \infty$, the difference between the estimates also approach zero. This completes the proof.

∎

E. Information Exchange

In this section we provide details about the information exchange necessary to take place between the subsystems to execute the proposed algorithm.

1) $\{\hat{\theta}_i(t), I_i(t)\}$:

After each iterate or update, each subsystem must provide its estimate and the associated cumulative improvement index to its immediate neighbors. The cumulative improvement index $I_i(t)$ can be computed iteratively, e.g. at each step, all one needs to calculate is the term
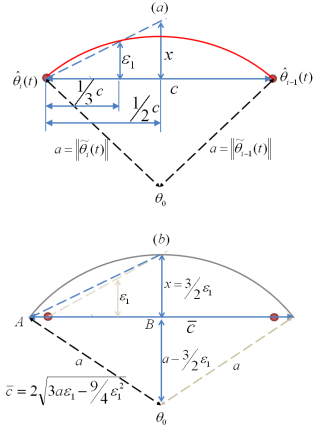
Fig. 4. Calculation of Upper Bound on the Distance of Local Estimates

Fig. 5. Circulant System

$$a(t)\left[-2+\frac{\phi_i(t-1)^T\phi_i(t-1)}{c+\phi_i(t-1)^T\phi_i(t-1)}\right]\frac{e_i(t)^2}{c+\phi_i(t-1)^T\phi_i(t-1)}$$

and add it to $I^i(t-1)$ to obtain, $I_i(t)$

2) $[\{y_i(t-1),y_i(t-2),\cdots\},\{y_{i-1}(t-1),y_{i-1}(t-2),\cdots\},$
$\{y_{i+1}(t-1),y_{i+1}(t-2),\cdots\},\cdots],$
$[\{u_i(t-1),u_i(t-2),\cdots\},\{u_{i-1}(t-1),u_{i-1}(t-2),\cdots\},$
$\{u_{i+1}(t-1),u_{i+1}(t-2),\cdots\},\cdots]:$

Depending on the structure of the model of the system, each system must receive information about the outputs and inputs from other subsystems that affect it in order to construct the regression vector $\phi_i(t-1)$.

3) More information exchange is required whenever ε-rule kicks in or averaging takes place in order to keep a track of the cumulative improvement index. Define:

$$E_i(t) := \hat{\theta}_i(t) - \hat{\theta}(0) \qquad (21)$$

Note that we can equivalently write,

$$E_i(t) = \sum_{\tau=1}^{t}\left[\frac{\phi_k(\tau-1)}{c+\phi_k(\tau-1)^T\phi_k(\tau-1)}\right]e_k(\tau) \qquad (22)$$

To demonstrate the requirement of this additional information exchange, assume that at time t, we need to take $\hat{\theta}^i(t) = \frac{\hat{\theta}_i(t)+\hat{\theta}_{i-1}(t)}{2}$ for the subsystem i. The calculation of $I^i(t)$ is given as follows:

$$\hat{\theta}^i(t) = \frac{\hat{\theta}_i(t)+\hat{\theta}_{i-1}(t)}{2} = \frac{\hat{\theta}(0)+E_i(t)+\hat{\theta}(0)+E_{i-1}(t)}{2}$$
$$= \hat{\theta}(0)+\frac{E_i(t)+E_{i-1}(t)}{2} \qquad (23)$$

subtracting θ_0 from both sides and taking inner product with itself, we get

$$\left\|\tilde{\theta}^i(t)\right\|^2 = \left\|\tilde{\theta}(0)\right\|^2 + \tilde{\theta}(0)^T[E_i(t)+E_{i-1}(t)]$$
$$+\frac{1}{4}[E_i(t)+E_{i-1}(t)]^T[E_i(t)+E_{i-1}(t)] \quad (24)$$
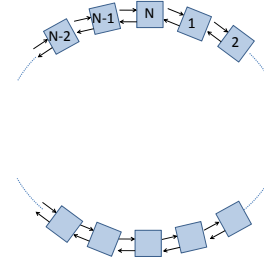
From the above expression, we can write $I^i(t)$ as follows

$$I^i(t) = \tilde{\theta}(0)^T[E_i(t)+E_{i-1}(t)]$$
$$+\frac{1}{4}[E_i(t)+E_{i-1}(t)]^T[E_i(t)+E_{i-1}(t)] \qquad (25)$$

The expression $\tilde{\theta}(0)^T E_i(t)$ can be iteratively computed as follows:

$$\tilde{\theta}(0)^T E_i(t)$$
$$= \tilde{\theta}(0)^T\left(\sum_{\tau=1}^{t}\left[\frac{\phi_k(\tau-1)}{c+\phi_k(\tau-1)^T\phi_k(\tau-1)}\right]e_k(\tau)\right)$$
$$= \sum_{\tau=1}^{t}\left[\frac{\hat{\theta}(0)^T\phi_k(\tau-1)-y_k(\tau)}{c+\phi_k(\tau-1)^T\phi_k(\tau-1)}\right]e_k(\tau)$$
$$= \tilde{\theta}(0)^T E_j(t-1)+\left[\frac{\hat{\theta}(0)^T\phi_i(t-1)-y_i(t)}{c+\phi_i(t-1)^T\phi_i(t-1)}\right]e_i(t)$$

Below we summarize the required information exchange:

a) $E_i(t)$
The vector $E_i(t)$ should be shared amongst the immediate neighbors in order to evaluate the expression $E_i(t)^T E_{i-1}(t)$. The record of this vector can be kept iteratively.

b) $\tilde{\theta}(0)^T E_i(t)$
Each subsystem should also share the scalar $\tilde{\theta}(0)^T E_i(t)$ with its immediate neighbors. The record of this scalar can be kept iteratively.

F. Simulation

In this section we present implementation of our algorithm on a circulant system. Basic structure of a circulant system is presented in Figure 5. Each subsystem has the same dynamics. For $i=N$, $i+1=1$. Likewise, for $i=1$, $i-1=N$. The dynamics of the i_{th} subsystem are given as follows:

$$y_i(t+1) = -0.6y_i(t)+0.1y_{i-1}(t)+0.1y_{i+1}(t)+u_i(t) \quad (26)$$

The initial guess for all the subsystems was taken as $\hat{\theta}(0) = [1,1,1,1]^T$. The input for the subsystems was chosen as given below:
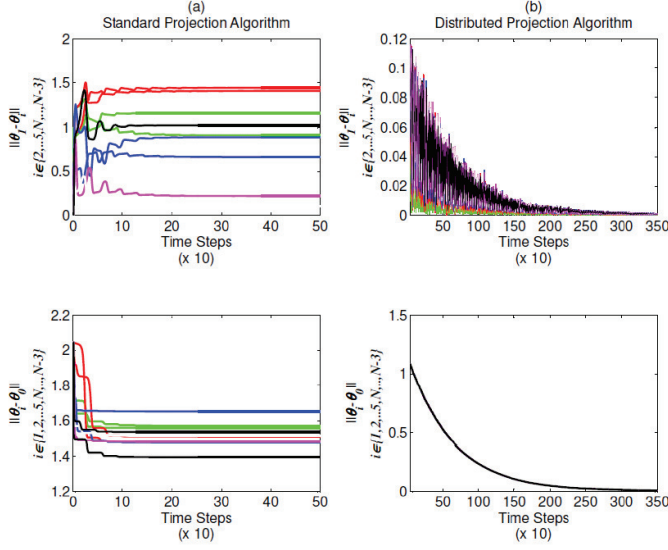
$$u_i(t) = cos(t+i)+sin(t-i/2) \qquad (27)$$

Fig. 6. Comparison of Distributed Projection Algorithm (a) with Standard Projection Algorithm (b)

A system comprising of $N = 100$ subsystems was simulated. Standard projection algorithm (where there is no sharing of estimates amongst the subsystems) was simulated along with the proposed distributed projection algorithm. For the implementation of distributed projection algorithm, a constant value of $\varepsilon = 10^{-5}$ was chosen. The results are captured in Figure 6. The top plots for Figure 6 present the distance of the estimates of the subsystem 1 from its neighbors, four to its right and four to its left. The bottom plots, on the other hand, present the distance of the subsystems $(N-3, \cdots, N, 1, \cdots 5)$ from the true parameter θ_0. Clear difference can be seen amongst the two plots. For the standard projection algorithm, the parameter estimates cease to change after about $20(\times 10)$ time steps. Also, the parameter estimates of the neighbors for subsystem 1 are quite far apart from it. The performance of distributed projection algorithm, on the other hand, is quite outstanding. It is quite interesting to note that the input signal was not 'exciting' for the standard projection algorithm, whereas convergence to true parameter is seen for the distributed projection algorithm.

III. CONCLUSIONS

We have presented a distributed projection algorithm for system identification of spatiotemporally invariant systems. Each subsystem receives information from all of its neighbors affecting it in order to construct the regression vector. Each subsystem, however, communicates only with its immediate neighbor to share its current estimate along with the related information (cumulative improvement index etc.). The best available estimate is picked in order to carry out the next iterate. For small estimation error, the scheme switches over to a "smart" averaging routine. The scheme ensures

continuous decay of estimation error and it serves to bring the local estimates arbitrarily close to one another. It was also seen that the parameter estimates converge to the true value even when the standard projection algorithm fails to do so. This calls for an investigation into the excitation conditions necessary for true parameter identification for the distributed projection algorithm. This work is part of our future research work.

We comment here that although our focus is on system identification of spatiotemporally invariant systems, the proposed algorithm is very well equipped to address the estimation problem in distributed networks. Distributed networks embedded with cooperative algorithms have been proposed to address estimation problems that arise in a variety of applications, such as environment monitoring, target localization and potential sensor network problems [9].

We remark at the end that the idea presented in this paper can be extended easily to other identification schemes such as least squares. This is obvious since the key point is the ability to write the given estimate as the sum of the original guess and some improvement terms.

IV. ACKNOWLEDGMENTS

REFERENCES

[1] Bamieh, B.F. Paganini and M. A. Dahleh, "Distributed Control of Spatially Invariant Systems", *IEEE Transactions on Automatic Control*, Vol. 47, No. 7, pp.1091-1107. 2002.

[2] Raffaello D'Andrea and Geir E. Dullerud, "Distributed Control Design for Spatially Interconnected Systems" *IEEE Transaction on Automatic Control*, Vol. 48, No. 9, 2003.

[3] A. Sarwar, P.G. Voulgaris and S.M. Salapaka. *Modeling and Distributed Control of Electrostatically Actuated Microcantilever Arrays* In Proc. American Control Conference 2007. (New York)

[4] S. Jiang , P. G. Voulgaris, L.E. Holloway, L.A. Thompson "Distributed control of large segmented telescopes", In Proc. ACC, 2006

[5] C. G. Lopes, A. H. Syed "Incremental Adaptive Strategies Over Distributed Networks", IEEE Transactions on Signal Processing, Vol 55, No. 8, Aug 2007

[6] C. G. Lopes, A. H. Syed "Diffusion Least-Mean Squares Over Adaptive Networks",In Proc. IEEE ICASSP, Hawaii, 2007,

[7] L. Xiao, S. Boyd, S.Lall "A Space-Time Diffusion Scheme for Peer-to-Peer Least Squares Estimation",In Proc. The Fifth International Conference on Information Processing in Sensor Networks, 2006

[8] Azeem Sarwar, Petros G. Voulgaris, and Srinivasa M. Salapaka. "Stability of Slowly Varying Spatiotemporal Systems", *In Proc. IEEE Conference on Decision and Control*, Cancun, Mexico, Dec, 2008

[9] D. Estrin, G. Pottie and M. Srivastava. "Intrumenting the world with wireless sensor networks," *In Proc. IEEE ICASSP*, pp. 2033-2036, Salt Lake City, UT, May 2001.