Recursive approximation of complex behaviours with IoT-data imperfections

Bekiroglu, Korkut; Srinivasan, Seshadhri; Png, Ethan; Su, Rong; Lagoa, Constantino

2020

Bekiroglu, K., Srinivasan, S., Png, E., Su, R., & Lagoa, C. (2020). Recursive approximation of complex behaviours with IoT-data imperfections. *IEEE/CAA Journal of Automatica Sinica*, 7(3), 656-667. doi:10.1109/jas.2020.1003126

<https://hdl.handle.net/10356/145763>

<https://doi.org/10.1109/JAS.2020.1003126>

© 2020 The Chinese Association of Automation. All rights reserved. This paper was published in *IEEE/CAA Journal of Automatica Sinica* and is made available with permission of the Chinese Association of Automation.

Downloaded on 26 Apr 2023 20:11:41 SGT

Recursive Approximation of Complex Behaviours With IoT-Data Imperfections

Korkut Bekiroglu, Seshadhri Srinivasan, Ethan Png, Rong Su, *Senior Member, IEEE*, and Constantino Lagoa

Abstract—This paper presents an approach to recursively estimate the simplest linear model that approximates the time-varying local behaviors from imperfect (noisy and incomplete) measurements in the internet of things (IoT) based distributed decision-making problems. We first show that the problem of finding the lowest order model for a multi-input single-output system is a cardinality (ℓ_0) optimization problem, known to be NP-hard. To solve the problem a simpler approach is proposed which uses the recently developed atomic norm concept and the modified Frank-Wolfe (mFW) algorithm is introduced. Further, the paper computes the minimum data-rate required for computing the models with imperfect measurements. The proposed approach is illustrated on a building heating, ventilation, and air-conditioning (HVAC) control system that aims at optimizing energy consumption in commercial buildings using IoT devices in a distributed manner. The HVAC control application requires recursive thermal dynamical model updates due to frequently changing conditions and non-linear dynamics. We show that the method proposed in this paper can approximate such complex dynamics on single-board computers interfaced to sensors using unreliable communication channels. Real-time experiments on HVAC systems and simulation studies are used to illustrate the proposed method.

Index Terms—Adaptability, distributed decision systems, imperfect measurements, internet of things (IoT), low order model identification.

I. INTRODUCTION

A. Motivation

SCIENTISTS have recently put significant effort in decentralizing most of these centralized problems to improve the general performance of the systems [1]. Furthermore, in many distributed decision-making problems, the cost of computa-

tion may escalate because of the structure of the local agents' models. Moreover, complex models require exhaustive computations for control systems applications. Therefore, implementable distributed control of complex processes requires the use of "simple models" that can i) provide good local approximations of the behavior; and ii) lead to control algorithms that are computationally efficient and easily implementable. Such problems are widespread in the internet of things (IoT), due to limited computing resources, imprecise sensors, and data imperfections resulting from unreliable networks [2]. Also, adaptability is an essential feature in many emerging applications, including IoT, and therefore, recursive identification becomes pivotal for such systems. The problem becomes interesting in the building heating, ventilation, and air-conditioning systems due to the multiple thermal zones in a building that have complex non-linear behaviors and subjected to frequent disturbances [3], [4].

B. Literature Review

Data imperfection can arise in many applications. For instance, incomplete measurements arise in many applications such as fault detection in building HVAC systems [5]. In the system identification literature, subspace identification methods are widely used in both linear [6] and non-linear system identification [7]. However, these techniques cannot handle sparsely sampled data, and furthermore, they are not amenable to recursive implementation. Therefore, considerable effort has been recently devoted to handling missing measurements in system identification problems while simultaneously enforcing system order [8], [9]. Most of these methods use a rank minimization (RM) method, a non-convex NP-hard problem that is mostly relaxed by using the concept of nuclear norm [9], [10]. In this scenario, the missing measurement is a decision variable of the optimization problem which compounds the computation complexity further. Recently, a method to identify multi-input and single-output (MISO) models with missing output data was proposed in [11]. However, the recursive identification was not considered in the presence of missing measurements. In addition to linear system identification, non-linear system identification literature also has many mature methods [7] to approximate the non-linear behavior of systems. Even though these methods have recursive model identification, according to our best knowledge literature still needs local low order linear model approximation of non-linear systems from the incomplete dataset. As it is mentioned in the motivation Section I-A, having a local low order linear model approximation is an essential step for distributed control applications such as model predictive control (MPC) to diminish the complexity.

Manuscript received January 4, 2020; revised February 21, 2020; accepted March 5, 2020. This work was supported by the Building and Construction Authority through the NRF GBIC Program (NRF2015ENC-GBICRD001-057). Recommended by Associate Editor Xin Luo. (*Corresponding author: Korkut Bekiroglu.*)

Citation: K. Bekiroglu, S. Srinivasan, E. Png, R. Su, and C. Lagoa, "Recursive approximation of complex behaviours with IoT-data imperfections," *IEEE/CAA J. Autom. Sinica*, vol. 7, no. 3, pp. 656–667, May 2020.

K. Bekiroglu is with the College of Engineering, SUNY Polytechnic Institute, Utica 13503, NY, USA (e-mail: korkut.bekiroglu@sunypoly.edu).

S. Srinivasan is with the Berkeley Education Alliance for Research in Singapore (BEARS), Singapore 138602, Singapore (e-mail: seshadhri.srinivasan@bears-berkeley.sg).

E. Png and R. Su are with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore (e-mail: ethanpng@ntu.edu.sg; rsu@ntu.edu.sg).

C. Lagoa is with the Electrical Engineering Department of Pennsylvania State University, University Park, PA 16802 USA (e-mail: lagoa@engr.psu.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JAS.2020.1003126

C. Contributions

In this paper, we present an algorithm which recursively computes simple linear models that approximate complex non-linear and time-varying local behaviors in the presence of imperfect (noisy and incomplete) measurements. A review of the literature reveals the lack of techniques for computing lower-order system models recursively from input-output data with missing measurements, a common occurrence in many distributed systems. In contrast to existing results, this paper presents an approach that directly identifies the lowest order MISO model recursively in the presence of imperfect measurements. We should note that the MIMO version of the proposed work requires further research. The proposed method only assumes that the system is stable and persistently excited, unlike existing approaches that require additional information on model order, accuracy, and continuous data. The main contributions of the investigation can be summarized as:

- 1) A method to recursively identify complex non-linear and time-varying behaviors using simple (low order) MISO linear models for distributed decision-making problems.
- 2) Compute the minimum data-rate required for identifying such lower order models from imperfect measurements.
- 3) Illustration of the proposed method in a pilot IoT based building management system which shows the efficiency and performance of the proposed approach.

D. Organization

The problem statement is presented in Section II. The relaxation for the cardinality optimization problem and a modified Frank-Wolfe algorithm to solve the relaxed problem are presented in Section III. Section IV presents the theory for the minimum data requirement for system identification. Illustrative examples are provided in Section V. Conclusions and further directions of this investigation are presented in Section VI.

Notations:

- x Lower case light letters denotes scalars
- $\mathbf{x}$ Lower case bold letters denotes vector
- $\mathbf{x}(i)$ Denotes i th element of the vector $\mathbf{x}$
- $\mathbb{R}, \mathbb{C}, \mathbb{Z}^+$ The set of real, complex, and positive integer numbers
- $\mathcal{R}(p)$ Denotes real part of $p \in \mathbb{C}$
- $\mathcal{I}(p)$ Denotes imaginary part of $p \in \mathbb{C}$
- $\text{diag}(\sigma_{t_0} \dots \sigma_{t_n})$ Diagonal matrix with given $\sigma_{t_0} \dots \sigma_{t_n}$
- $\mathcal{Z}^{-1}$ Denotes the inverse z -transform
- $\|\mathbf{c}\|_{\ell_0}$ The cardinality of $\{\mathbf{c} : \mathbf{c}_p \neq 0\}$
- $\|\mathbf{g}\|_{\ell_2}$ The 2 norm: $\|\mathbf{c}\|_{\ell_2} = \sqrt{\sum_{i=1}^N c_i^2} = \sqrt{\mathbf{c}^T \mathbf{c}}$
- $\|\mathbf{c}\|_{\ell_\infty}$ The infinity norm: $\|\mathbf{c}\|_{\ell_\infty} = \max(|\mathbf{c}|)$
- $\mathbf{g} * \mathbf{u}$ Denotes convolution of the signals $\mathbf{g}$ and $\mathbf{u}$
- $\mathbf{T}_u$ A lower triangular Toeplitz matrix associated with any finite sequence $\{\mathbf{u}(k), k = 0, 1, \dots, n-1\}$, or any column vector $\mathbf{u} = [\mathbf{u}(0), \mathbf{u}(1), \dots, \mathbf{u}(n-1)]^T$

$$\mathbf{T}_u \approx \begin{bmatrix} \mathbf{u}(0) & 0 & \dots & 0 \\ \mathbf{u}(1) & \mathbf{u}(0) & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ \mathbf{u}(n-1) & \mathbf{u}(n-2) & \dots & \mathbf{u}(0) \end{bmatrix}$$

II. PROBLEM STATEMENT

Consider the distributed decision-making problem with each controlled entities having the MISO dynamical model

$$\begin{aligned} \mathbf{y}(t_k) &= \mathcal{F}(\mathbf{y}(t_1), \mathbf{y}(t_2), \dots, \mathbf{y}(t_{n_y})) \\ &\quad \mathbf{u}_1(t_1), \mathbf{u}_1(t_2), \dots, \mathbf{u}_1(t_{n_u}) \\ &\quad \mathbf{u}_2(t_1), \mathbf{u}_2(t_2), \dots, \mathbf{u}_2(t_{n_u}) \\ &\quad \vdots \\ &\quad \mathbf{u}_N(t_1), \mathbf{u}_N(t_2), \dots, \mathbf{u}_N(t_{n_u}), t_k) + \varepsilon(t_k) \end{aligned} \quad (1)$$

where $\mathcal{F}$ represents a non-linear time-varying function, $\mathbf{y}(t_k)$, $\mathbf{u}_1(t_k), \mathbf{u}_2(t_k), \dots, \mathbf{u}_N(t_k)$, and $\varepsilon(t)$ denotes the sensor measurements (output), inputs, and the measurement noise, respectively, and n_y, n_u are the time lags for the system output and inputs. Our idea is to approximate the non-linear and time-varying model in (1) of each controlled entities with the simplest linear model which captures the essential dynamics with imperfect sensor measurements but known inputs¹. The following definitions are given to formulating the proposed problem.

Definition 1 (Linear Local Measurement Model): Considering $\mathbf{y}(t_k)$ to be the measurements obtained with commensurate sampling instants $t_k : t_0 < t_1 < t_2 < \dots < t_m$. Then the local behaviors of MISO dynamics in (1) can be approximated with a strictly proper linear time-invariant (LTI) as

$$\mathbf{y}(t_k) = \sum_{i=1}^N (\mathbf{g}_i * \mathbf{u}_i)(t_k) + \varepsilon(t_k), \quad k = 0, \dots, m \quad (2)$$

where $*$ denotes convolution, m is the number of measurements, $\mathbf{g}_i$ represents the impulse response from input i to output, $\mathbf{u}_i$ represents the input signal of i th input, and $\mathbf{y}(t_k)$ is output signal at time t_k corrupted by some bounded noise ε .

Definition 2 (Stable Dynamic Region of the Model): The dynamics of the simple linear model are characterized by a set of poles $\forall i \in \mathbb{Z}^+, p_i \in D_\rho$, inside a closed centered disc around the origin with $\rho \leq 1$. The set of poles are deemed as well behaved, i.e., $p_i \in D_\rho = \{p_i \in \mathbb{C} \text{ or } p_i \in \mathbb{R} : |p_i| \leq \rho\}$ ².

Definition 3 (Data Loss Operator): Data-loss is modelled by a binary variable $\sigma_{t_k} \in \{0, 1\}$, with $\sigma_{t_k} = 1$, if the packet is delivered and 0, otherwise. The network communication for m packets is modelled as a $m \times m$ matrix $\mathcal{L}_m = \text{diag}[\sigma_{t_0} \dots \sigma_{t_m}]$, and the channel is assumed to be perfect, if $\text{trace}(\mathcal{L}_m) = m+1$, and imperfect otherwise. The data loss operator models the constraints due to data imperfections in the model identification problem. It is a $\mathcal{L}_m B$ matrix multiplication operator which models the data loss. The operator takes as input the tuple $\{\mathcal{L}_m, B\}$ and provides an output that replaces the rows of B with zeros for the rows with all zeros in the matrix $\mathcal{L}_m$ and leaves the other rows unaltered.

Without loss of generality, the measurement model in (2)

¹ The imperfect measurements can result from a variety of reasons which include hardware aspects such as sensor failures, noise, faults, etc., or network induced imperfections such as latencies, packet dropout, quantization, and others.

² The set of poles can be defined based on priori such as rise time, overshoot, etc. Only the stable poles are chosen for this paper, reasonable one for HVAC systems, but can include unstable poles.

can be extended to model the IoT data constraint by introducing an IoT data loss operator $\mathcal{L}_m$, i.e.,

$$\tilde{\mathbf{y}}(t_k) = \mathcal{L}_m \left(\sum_{i=1}^N (\mathbf{g}_i * \mathbf{u}_i)(t_k) + \varepsilon(t_k) \right) \quad (3)$$

where $\tilde{\mathbf{y}}(t_k) = \mathcal{L}_m \mathbf{y}(t_k)$ is the imperfect output measurements obtained by using the network.

Remark 1: The linear approximation of the non-linear and time-varying systems is ill-defined. Given a slow time-varying non-linear system, one can locally (in-state and time) approximate its behavior by an LTI system. Since there are many ways to approximate this behavior, we choose an approximation that provides a trade-off between complexity and error.

Finally, we tried to address the following system identification problem in Problem 1 under the following Assumptions 1–3.

Problem 1: Suppose there exists a lowest order model $\mathbf{g}_i$ that approximates the non-linear and time-varying local behaviour of a system. The problem is to compute $\hat{\mathbf{g}}_i$, an estimate of the $\mathbf{g}_i$ having poles in the stable dynamic region D_ρ (see, Definition 2) from imperfect measurements $\tilde{\mathbf{y}}(t_k)$ under the given assumptions.

Assumption 1: The lowest order transfer function $\hat{\mathbf{g}}_i$ has no repeated poles. This assumption does not limit the capability of the proposed method since any repeated poles can be approximated with non-repeated poles, but with an arbitrary small perturbation [12].

Assumption 2: Initial conditions are assumed to be zero to simplify the analysis of the proposed method. Non-zero initial conditions can also be addressed with a small modification of the optimization problem³.

Assumption 3: The measurement noise in (2) is bounded, i.e., $\|\varepsilon\|_2 \leq \Xi$.

In what follows, we show that the problem of finding the lowest order linear model recursively with imperfect measurements that captures complex non-linear behavior of each of the controlled entities is a ℓ_0 optimization model. We propose an elegant way and an efficient algorithm to solve the complex problem. Furthermore, we also investigate the data that needs to be transmitted for approximating the non-linear behaviours.

III. RECURSIVE SYSTEM IDENTIFICATION METHODOLOGY WITH IMPERFECT DATA

In this section, the optimization problem of the minimum order system identification of a MISO system is presented.

A. Minimum Order Model Identification Problem

The impulse response of a linear system $\mathbf{g}_i$ in (2) can be expressed as a linear combination of a set of first-order

systems' impulse responses, denoted by *Atoms*, by using partial fraction expansion [13]. To obtain the minimum order transfer function, the cardinality of *Atoms* (number of these first-order systems) has to be minimized. However, one should note that even though each transfer function of input i to output shares the same poles, they could have different zeros. This requirement can be addressed using the following approach

$$\min_{c_p^i, c_p \in \mathbb{C}} \|\mathbf{c}\|_{\ell_0} \quad (4a)$$

$$\text{s.t. } \left\| \sum_{i=1}^N \left(\sum_{p \in D_\rho} c_p^i (\mathbf{g}_p^i * u_i) - \mathbf{y} \right) \right\|_{\ell_2}^2 \leq \Xi \quad (4b)$$

$$\max_{i=1,2,\dots,N} (|c_p^i|) \leq c_p \quad (4c)$$

where ℓ_0 norm minimization of a vector $\mathbf{c}$ in (4a) is analogous to minimizing cardinality ($\text{cardinality}\{\mathbf{c} : c_p \neq 0\}$) where c_p is the p th element of vector $\mathbf{c}$. The constraint in (4b) enforces the accuracy of the model for a bounded noise for real output data $\mathbf{y}$ (note that we assume complete data in this step and incomplete data will be address in (13)) and (4c) enforces a block sparsity of c_p^i which means the sparsity pattern of each input vector are same. In other words, it is ensured that each transfer function from output to input i use same poles. Note that c_p^i represents a constant for each first order system for input i . There are four challenges in the optimization problem in (4a), they are: i) the cardinality minimization problem is a NP-hard problem [14]; ii) the computation complexity arising from c_p^i being complex; iii) the search-space (D_ρ) contains infinite number of poles; and iv) the need to perform recursive identification compounds the complexity further.

B. Convex Relaxation

To relax the optimization problem in (4a), this paper utilizes an atomic norm-based approach. The complete definition and properties of the atomic norm can be found in [15]. From the definition of *Atoms*, if one chooses simple objects as *Atoms* in an atomic set and tries to force using few of them in expressing a certain object, this naturally promotes a sparse representation of the object at hand. The proposed method is developed by exploiting this property. In our analysis, we choose the impulse response as the *Atoms* as given in Appendix-A.

The noisy measurement model $\mathbf{y}(t_k)$ of the system (2) can now be represented using the following matrix and vector notations as

$$\mathbf{y} = \sum_{i=1}^N \mathbf{T}_{\mathbf{u}_i} \mathbf{g}_i + \varepsilon \quad (5)$$

where $\mathbf{T}_{\mathbf{u}_i} \in \mathbb{R}^{m \times m}$ is the Toeplitz matrix of i th input and $\mathbf{g}_i$ is the impulse response for the input u_i to output $\mathbf{y} \in \mathbb{R}^m$. The impulse response of the transfer function $\mathbf{g}_i$ in (5) is approximated as

$$\mathbf{g}_i \approx \sum_{a \in A(D_\rho)} c_a^i \mathcal{Z}^{-1}(a), \quad \forall c_a^i \in \mathbb{R}, \quad i = 1, 2, \dots, N \quad (6)$$

³ Discrete linear systems with distinct poles have two facts: i) the same poles appear in the impulse response and the initial condition response; ii) the number of distinct poles used in these two is equal to the order of the system. Hence, sparsifying the impulse response of the system (i.e., identifying the system of lowest order) is equivalent to minimizing the number of unique poles that are used in the total response, i.e., the response that includes both the response to initial conditions and the zero state response to the input.

where $\mathcal{Z}^{-1}$ denotes the inverse z -transform. The sparsity pattern of each c_a^i in (7) with c_a^i being the i th element of vector $\mathbf{c}_a$ is identical to enforce the same pole in each of the transfer function, $\mathbf{g}_i$

$$\mathbf{c}_a = [c_a^1 c_a^2 \dots c_a^{N-1} c_a^N]. \quad (7)$$

Rewriting (5), we have

$$\mathbf{y} = \hat{\mathbf{T}}\hat{\mathbf{g}} + \varepsilon \quad (8)$$

with

$$\begin{aligned} \hat{\mathbf{g}} &= [\mathbf{g}_1 \mathbf{g}_2 \dots \mathbf{g}_N]^T \\ &= \sum_{a \in A(D_p)} [c_a^1 \mathcal{Z}^{-1}(a) c_a^2 \mathcal{Z}^{-1}(a) \dots c_a^N \mathcal{Z}^{-1}(a)]^T \end{aligned} \quad (9)$$

and

$$\hat{\mathbf{T}} = [\mathbf{T}_{u_1} \mathbf{T}_{u_2} \dots \mathbf{T}_{u_N}]. \quad (10)$$

One can realize that each $\mathbf{g}_i$ utilizes the same atom. Using the reformulation in (8), the non-convex optimization problem in (4a) can be relaxed using the following atomic norm

$$\|\hat{\mathbf{g}}\|_A = \{\inf \sum_{a \in A(D_p)} \|\mathbf{c}_a\|_{\ell_\infty} : \text{s.t. (9)}\}. \quad (11)$$

Having defined the atomic norm that encourages block sparsity for the MISO system transfer function, the following convex constrained optimization problem (12), a convex relaxation problem of (4a), can be solved. This relaxes the requirement of solving an NP-hard problem.

$$\begin{aligned} \min_{\hat{\mathbf{g}}} & \frac{1}{2} \|\hat{\mathbf{T}}\hat{\mathbf{g}} - \mathbf{y}\|_{\ell_2}^2 \\ \text{s.t. } & \|\hat{\mathbf{g}}\|_A \leq \tau \end{aligned} \quad (12)$$

where τ is a bound on the convex surrogate of block sparsity of vectors $\mathbf{c}$ in (4a). The IoT induced constraints are brought in by introducing the data loss operator $\mathcal{L}_m$ as

$$\begin{aligned} \min_{\hat{\mathbf{g}}} & \frac{1}{2} \|\mathcal{L}_m(\hat{\mathbf{T}}\hat{\mathbf{g}} - \mathbf{y})\|_{\ell_2}^2 \\ \text{s.t. } & \|\hat{\mathbf{g}}\|_A \leq \tau. \end{aligned} \quad (13)$$

With this trivial modification, the optimization model for obtaining the minimum order transfer function for the distributed decision-making system with imperfect data is obtained. Though the non-convex ℓ_0 optimization problem is now relaxed to a convex problem, the presence of infinitely many poles in the set D_p increases the search space significantly. The presence of infinitely many poles is solved in two ways: gridding the region of the well-behaved poles [16] and a randomized Frank-Wolfe algorithm (rFWA) [11]. While gridding is a computationally inefficient method to search, the rFWA is not suitable for recursive computations. Therefore, this investigation introduces a heuristic step that enhances the computational speed and makes the algorithm suitable for recursive system identification. We call this modified version of rFWA as a modified Frank Wolfe algorithm (mFWA).

C. Solving Convex Optimization by Using mFWA

In this section, we introduce a method to solve the convex problem in (13) that is a special case of the atomic norm constraint problem presented in [17];

$$\begin{aligned} \min_{\mathbf{g}} & f(\mathbf{g}) \\ \text{s.t. } & \|\mathbf{g}\|_A \leq \tau \end{aligned} \quad (14)$$

where $f(\mathbf{g})$ is a convex and smooth function. In our case, the function $f(\cdot)$ is the $\|\hat{\mathbf{T}}\hat{\mathbf{g}} - \mathbf{y}\|_{\ell_2}^2/2$, i.e., the 2-norm of the error and the definition of $\|\mathbf{g}\|_A$ are given in (11).

The various steps involved in the mFWA are shown in Algorithm 1. The inputs to the algorithm are the initial model $\hat{\mathbf{g}}^t$, the compact pole location set D^t , bounds on the iteration error ϵ , the convergence criteria γ , τ bound on the atomic norm of the impulse response and data, and the pole locations of initial model $\mathcal{D}^t, \mathcal{P}^t$. The convergence criteria, iteration bounds, and iteration counter k are first initialized. The following computations are performed until the convergence criteria and iteration bounds are met.

i) The poles (P_k^t) are selected by uniformly sampling D^t , a region having stable poles (Step 4).

ii) Given the set of poles (P_k^t), the corresponding atom set denoted by $A(P_k^t)$ is calculated from the definition of the atomic set in (21) (Step 5).

iii) The gradient vector of the objective function in (13) is calculated as

$$\nabla f(\hat{\mathbf{g}}_k) = \mathcal{L}_m(\hat{\mathbf{T}}\hat{\mathbf{g}}_k - \mathbf{y})^T \hat{\mathbf{T}}. \quad (15)$$

This gradient vector is partitioned for each impulse response i as $\nabla f(\hat{\mathbf{g}}_k^i) \in \mathbb{R}^m$ (Step 6).

iv) Different from rFWA, we modified this step in our algorithm for MISO systems. The minimum over the set of the inner product of the gradient and the $+$ and $-$ signs of corresponding *Atoms* is computed for each input i . This gives the descent direction for the optimization problem (Step 6).

Remark 2 (Promoting Block Sparsity): The gradient vector for each impulse response $\nabla f(\hat{\mathbf{g}}_k^i) \in \mathbb{R}^m$ is treated with an atom $\pm a$ in atomic set $A(P_k^t)$. This enforces each impulse response to use the same pole. However, because of the structure of the atomic set A in (22), the sign combination of this chosen impulse response is tested to find the descent direction for each impulse response.

v) The optimal step-size α_k for each iteration k with given τ and μ_k^i for the best atom a_k with $\alpha \in [0, 1]$ is computed as (Step 7) where

$$\alpha^* = \max(0, \min(\alpha_k, 1))$$

where

$$\alpha_k = \frac{S^T Q}{Q^T Q}$$

where

$$S = \mathcal{L}_m(\hat{\mathbf{T}}\hat{\mathbf{g}}_k - \mathbf{y})$$

and

$$Q = \mathcal{L}_m(\hat{\mathbf{T}}(\tau[\mu_k^1 a_k \dots \mu_k^N a_k]^T - \hat{\mathbf{g}}_k)). \quad (16)$$

vi) Then this step is different than the classical rFWA for SISO systems since this step is added to promote the recursive system identification. In here, if the $\alpha_k \neq 0$ then the selected atom is used to update the impulse response. Therefore, the pole corresponding to the atom, i.e., (p_k) is appended to the existing pole set $\mathcal{P}^t$ for the recursive computation $t + 1$ (Step 8).

vii) The following heuristic is also added to rFWA to improve the quality of the algorithm for the MISO system identification. Although the bound τ on the atomic norm of $\hat{\mathbf{g}}$ enforces the block sparsity of vectors $\mathbf{c}_a^i$ in (7), magnitude of each impulse response $\mathbf{g}^i$ should be adapted in each iteration with an additional constant. Therefore, the scaling factor $\mu_k^i = \|\mathbf{g}_k^i\|_{\ell_\infty}$ is calculated in each step to scale the magnitude of impulse response from input i to output (Step 10). After calculating each impulse response, the impulse response $\mathbf{g}_k^i$ is scaled with μ_k^i (Step 9)

viii) Update the impulse response of each input i as $\mathbf{g}_{k+1}^i$, and compute the matrix $\hat{\mathbf{g}}_{k+1}$ (Step 11).

ix) The loop is repeated until convergence. The output of the algorithm is the $\hat{\mathbf{g}}^{t+1}$.

Algorithm 1: mFWA ($\hat{\mathbf{g}}^t, D^t, \epsilon, \gamma, \mathcal{D}^t, \mathcal{P}^t, \tau$)

```

1: Inputs:  $\hat{\mathbf{g}}^t, D^t, \epsilon, \gamma, \mathcal{D}^t, \mathcal{P}^t, \tau$ 
2: Set  $\mathcal{E} = 0, \nabla J = 0, k = 0, \mu_0 = 1$ 
3: while  $\mathcal{E} \leq \epsilon$  &&  $\nabla J \leq \gamma$ 
4:    $P_k^t \leftarrow \mathcal{U}(D^t)$ : Uniformly sampled poles
5:    $A(P_k^t) \leftarrow$  Calculate Atoms for poles ( $P_k^t$ ) using (22)
6:    $a_k^i \leftarrow \operatorname{argmin}_{a \in A(P_k^t)} \langle \nabla f(\hat{\mathbf{g}}_k^i), \pm a \rangle \triangleright \forall \text{ input } i \text{ and } a \in A(P_k^t)$ 
7:    $\alpha_k \leftarrow \operatorname{argmin}_{\alpha \in [0,1]} f(\hat{\mathbf{g}}_k + \alpha[\tau[\mu_k^1 a_k^1 \dots \mu_k^N a_k^N]^T - \hat{\mathbf{g}}_k]) \triangleright \hat{\mathbf{g}}_k \text{ in (9)}$ 
8:    $\mathcal{P}^t \leftarrow \text{Append}(\{p_k\} \mid \alpha_k \neq 0 : a_k \in A(p_k), p_k \in P_k^t) \triangleright a_k^i \text{ shares same}$ 
   pole
9:    $\mathbf{g}_{k+1}^i \leftarrow \mathbf{g}_k^i + \alpha_k[\tau\mu_k^i a_k^i - \mathbf{g}_k^i]$ 
10:   $\mu_{k+1}^i = \|\mathbf{g}_{k+1}^i\|_{\ell_\infty}$ 
11:   $\hat{\mathbf{g}}_{k+1} \leftarrow [\mathbf{g}_{k+1}^1 \mathbf{g}_{k+1}^2 \dots \mathbf{g}_{k+1}^N]^T$  using (9)
12:  Compute:
13:     $\mathcal{E} = \|\mathcal{L}_m(\hat{\mathbf{T}}\hat{\mathbf{g}} - \mathbf{y})\|_{\ell_2}^2$ 
14:     $\nabla J = \|\hat{\mathbf{g}}_{k+1} - \hat{\mathbf{g}}_k\|_{\ell_2}^2$ 
15:     $k := k + 1$ 
16: end while
17:  $\mathcal{P}^{t+1} = \mathcal{P}^t$ 
18: Output:  $\hat{\mathbf{g}}^{t+1}, \mathcal{P}^{t+1}$ 

```

For the rFWA, it has been already proven that rFWA converges in expectation [13]. In addition to convergence in expectation, the rFWA for SISO case converges with a fixed rate. Since the randomization step in rFWA and mFWA is the same (Steps 4 and 6), we have the same rate of convergence for our mFWA. For the completeness, we highlighted Lemma 1 and the Theorem 2 given in Appendix-B for our recursive identification algorithm.

In what follows, we will illustrate how the mFWA proposed in this investigation can be used for recursive system identification.

D. Recursive Minimum Order System Identification Methodology

In this section, we introduce one of our main contributions. The proposed recursive system identification methodology is shown in Fig. 1. The method works in two modes: off-line and on-line. In the off-line mode, the data is used to learn a minimum order nominal model $\hat{\mathbf{g}}^0$ from a set of data $\mathcal{D}^0$ by solving (13). In the off-line mode, the bound on the approximation error, an input of the Algorithm 2, can be also learned by analyzing the fitting error. This error is the

approximation errors of linear estimation of non-linear and time-varying behavior. In this way, the user can have an initial guess of τ in (14). Then this model (impulse response) obtained in the off-line mode and its poles are inputs to the on-line mode. The heart of the model adaptation algorithm is the mFWA which takes the nominal model from the previous step, the current set of data $\mathcal{D}^1$ which also has missing measurements, and the pole locations of this nominal model to obtain the lowest order model $\hat{\mathbf{g}}^1$. The procedure repeats when a new set of data is available. In this recursive system identification algorithm hot-starts the next iteration using the poles of the current model.

Fig. 1. Recursive identification methodology.

While in the off-line mode, the poles are randomly chosen in D_ρ (see, Definition 2), in the on-line mode, the poles are randomly chosen in the norm ball around each pole of the nominal model computed in the previous step. The norm-ball is defined as follows.

Definition 4 (Norm Ball Around a Pole p): Given the pole p , a closed centered norm ball around the pole p with radius β is the set of poles $\forall i \in \mathbb{Z}^+, p_i \in D_{\beta p}$ such that $\{p_i \in \mathbb{C} \text{ or } p_i \in \mathbb{R} \text{ with } \|p_i - p\| \leq \beta\}$.

The concept of norm-ball enables searching the poles in the on-line mode around the $D_{\beta p}$ region instead of the whole region D_ρ , and this increases the speed of computation significantly as the most likely locations are checked by the mFWA first. Our main algorithmic contribution is given in Algorithm 2.

Algorithm 2: Recursive Method

```

1: Inputs:  $D^t, \epsilon, \gamma, \mathcal{D}^t, \tau, \mathcal{P}^0 = []$ 
2: if  $t = 0$  then  $\triangleright$  (off-line mode)
3:    $\mathcal{D}^0 = \{\tilde{\mathbf{y}}^0, \mathbf{u}_i^0\}$ 
4:    $[\hat{\mathbf{g}}, \mathcal{P}^1] = \text{mFWA}(\hat{\mathbf{g}}^0, D_\rho, \epsilon, \gamma, \mathcal{D}^0, \mathcal{P}^0, \tau)$ 
5:    $\hat{\mathbf{g}}^1 \leftarrow \text{Model order reduction}(\hat{\mathbf{g}})$ 
6: else  $\triangleright$  (on-line mode)
7:    $\mathcal{D}^t = \{\tilde{\mathbf{y}}^t, \mathbf{u}_i^t\}$ 
8:   Given  $\mathcal{P}^t \rightarrow$  generate  $D_{\beta p}$  (see, Definition 4)
9:    $D^t \leftarrow \bigcup D_{\beta p} : \{p \in \mathcal{P}^t\}$ 
10:   $[\hat{\mathbf{g}}, \mathcal{P}^{t+1}] = \text{mFWA}(\hat{\mathbf{g}}^t, D^t, \epsilon, \gamma, \mathcal{D}^t, \mathcal{P}^t, \tau)$ 
11:   $\hat{\mathbf{g}}^{t+1} \leftarrow \text{Model order reduction}(\hat{\mathbf{g}})$ 
12: end if
13: Output:  $\hat{\mathbf{g}}^{t+1}, \mathcal{P}^{t+1}$ 

```

Furthermore, in between each mode, we include a model reduction step because of the structure of the Frank-Wolfe algorithm. One can see in Step 9 in Algorithm 1, the poles in each step are appended to the current impulse response. This process directly affects the parsimony of the system in the recursive iteration. Therefore, during the recursive system identification, some of the poles in the previous step are eliminated using the following method.

Remark 3 (Model Order Reduction): Before initiating the new identification step by utilizing the previous nominal model with a new set of data (see, Algorithm 1), this previously identified impulse response is used to construct a Hankel matrix. Since the rank of the Hankel matrix is equal to the system order, the singular value decomposition (SVD) is calculated and the tails of these singular values are eliminated. Since this method is a well-known method, we refer readers to the following studies [18]–[20].

IV. MINIMUM DATA REQUIRED FOR LINEAR SYSTEM IDENTIFICATION

After presenting the recursive system identification algorithm in the previous section, in this section, we analyze how much data we need to identify a n th order system. In other words, we investigate how much missing data can we handle to identify the true approximated system. Prediction error of an n th order system from (8) is $\varepsilon(\mathbf{c}) = \mathbf{y} - \hat{\mathbf{T}}\mathbf{g}\mathbf{c}$, where $\mathbf{c} \in \mathbb{R}^n$ and $\mathbf{y}$, $\hat{\mathbf{g}}$ are with suitable dimensions. The multiplication of $\hat{\mathbf{g}}$ and the unknown vector $\mathbf{c}$ is the linear combination of n impulse responses. Then the following optimization computes the corresponding impulse responses in $\hat{\mathbf{g}}$.

$$\mathbf{c}^* = \min_{\mathbf{c} \in \mathbb{R}^n} E(\varepsilon(\mathbf{c})\varepsilon(\mathbf{c})^T) = \min_{\mathbf{c} \in \mathbb{R}^n} E((\mathbf{y} - \hat{\mathbf{T}}\mathbf{g}\mathbf{c})(\mathbf{y} - \hat{\mathbf{T}}\mathbf{g}\mathbf{c})^T)$$

Assuming the input $\mathbf{u}$ is a quasi-stationary signal, the uniqueness of $\mathbf{c}^*$ is guaranteed as long as the matrix $\hat{\mathbf{T}}\mathbf{g}$ has full column rank and can be determined as the solution of $\mathbf{y} = \hat{\mathbf{T}}\mathbf{g}\mathbf{c}$. However the data loss operator $\mathcal{L}_m$ in (13) reduces the rank of the matrix $\hat{\mathbf{T}}\mathbf{g}$ ($\mathcal{L}_m \hat{\mathbf{T}}\mathbf{g}$). From the definition of the persistent excitation condition [21], one can state that the rank of the correlation matrix ($R_u = E(u_i(t_k)u_i^H(t_k))$) is equal to the dimension of the $\mathbf{c}$.

$$R_u = E(u_i(t_k)u_i^H(t_k)) = \begin{bmatrix} Ru_i(0) & Ru_i(1) & \dots & Ru_i(m-1) \\ Ru_i(1) & Ru_i(0) & \dots & Ru_i(m-2) \\ \vdots & \ddots & \dots & \vdots \\ Ru_i(m-1) & \dots & R(1) & Ru_i(0) \end{bmatrix}. \quad (17)$$

Using this fact, one can propose the following theorem.

Theorem 1 (Necessary Condition for Identifiability): Suppose there are m imperfect measurement packets and our aim is to identify a system whose highest order is n , where $m \gg n$. The necessary conditions for identifiability of the system with the imperfect measurement is given by

$$\text{rank}(\mathcal{L}_m) \geq n.$$

The proof of the theorem is given in Appendix-C.

V. EXAMPLES

This section presents two case-studies, a real-time

experiment, and simulations that demonstrate how complex behaviors can be approximated as simple models by using the approaches presented in this paper for distributed decision making. Both the examples concern the building with variable air volume (VAV)-based HVAC control, a complex non-linear system. Even though imperfect measurement can be seen in most of the IoT based implementations, HVAC control is one of the good cases to test the proposed work [22]. Note that estimating the energy model for building automation is an important problem.

A. Case-Study: IoT-Based HVAC Control

1) *Pilot Description:* The pilot considered is the VAV controlled HVAC system located at S1B1 building, NTU, Singapore. It has 85 zones and 6 AHUs sharing a centralized chiller [23]. The test-bed has AHU that are manufactured by York and have flow sensors from ABB. The fan power is 11 kW induction motor and the centralized chillers' flow is controlled using a valve. The fan supplies air to individual zones through a common duct. The fan speed is controlled using a speed control drive and the power varies as a square of the fan speed. The test-bed has control systems supplied by Siemens A/s to handle the AHU control and Quantum systems for the chiller control. Currently, the test-bed has a centralized control architecture with each zone's thermal behavior having a nonlinear dynamics (bi-linear) due to coupling between state and control inputs. We place IoT sensors to collect local information as explained later and there are issues of data-loss due to communication imperfections. In the experiment, we describe data collection, sensors, communication and other aspects with which data is collected. Currently, a centralized controller is used for allocating the mass-flow rate to each of the zones by controlling the opening of VAV. The air-flow in the zones are coupled through mass-flow rate constraints and fan capacity. Following [24] the zone thermal dynamics is given by

$$c_i \dot{\mathbf{X}}_i = \dot{m}_i c_a (\mathbf{X}_c - \mathbf{X}_i) + R_i (\mathbf{X}_{oa} - \mathbf{X}_i) + \dot{Q}, \quad \forall i \in \{1, \dots, N_z\} \quad (18)$$

where c_i , c_a , and N_z denote the thermal capacitance of the zone i , specific heat capacity of air [$\text{kJkg}^{-1}\text{C}^{-1}$] and the number of zones, respectively. Additionally, $\dot{m}_i$, $\mathbf{X}_c$, $\mathbf{X}_i$, R_i , and $\mathbf{X}_{oa}$ denote the mass-flow rate in zone i [kgs^{-1}], temperature of the cool air from AHU [$^{\circ}\text{C}$], temperature of the zone i [$^{\circ}\text{C}$], thermal resistance of zone i [$\text{kW}^{\circ}\text{C}^{-1}$], and ambient temperature [$^{\circ}\text{C}$], respectively. The term $\dot{Q}$ models the heat load rate and is influenced by factors such as occupancy which are time-varying. Consequently, model adaptation is required for capturing influences of such time-varying factors.

2) *Distributed Control:* Current building automation systems are centralized structures and use conventional control techniques such as proportional-integral controller or thermostats which are not optimal. Recent research has proven that using the information on weather, occupancy, etc., to control mass-flow rate in individual zones can enhance energy efficiency. In this backdrop, the model predictive controller which uses the zone thermal models, coupling in the AHU and chiller circuits to compute the optimal mass-flow rate has emerged as a promising solution. Implementing MPC with

centralized architecture is complex due to the non-linear and time-varying zone thermal dynamics. Distributed control wherein the optimal mass-flow rates are computed using single board computers interfaced with sensors that offer scalability and simplicity. Further, by using the IoT such boards can be networked to give a distributed implementation. However, the non-linear and time-varying zone thermal dynamics makes the realization of MPC on a single-board computer complex. Therefore, simple linear models that approximate complex thermal dynamics recursively are required for realizing distributed HVAC control.

3) *IoT-Based Implementation:* To recursively approximate the non-linear zone thermal dynamics with simple linear models, the linear model presented in [24] and the method presented in this paper were used. To model the zone thermal dynamics, we conducted experiments to record the VAV mass-flow rate in $l\ s^{-1}$ was measured and it was converted to $kg\ s^{-1}$ using a suitable conversion factor. In addition, influences of weather were measured using a web-service and the occupancy was measured using CO_2 sensors. The temperature and humidity were measured with DHT11 sensors interfaced with the Raspberry Pi board and the sensor measurements were transmitted to the controller board that implements the MPC and our model approximation approach proposed in the paper. We use the NRF2401L wireless module to transmit the measurement by constructing the sensor module as shown in Fig. 2. The input measurements from these sensors are given in Fig. 3. The sensor module in

Fig. 2. Sensor module with wireless interface.

Fig. 3. Mass-flow rate and ambient temperature recorded during the experiment.

Fig. 2 is used to measure the mass-flow rate from which control input could be computed as

$$\dot{Q}_c = \dot{m}_z C_p (T_s - T_z) \quad (19)$$

where $\dot{Q}_c$ is the cooling energy rate, $\dot{m}_z$ is the zone mass-flow rate, C_p is the specific heat capacity of air, T_s and T_z are the supply air temperature, and zone temperature, respectively. Since the data is transmitted to the controller module using a wireless network, there are packet-losses which make the identification challenge. Our experiment showed that about 20–25% of the data-loss occurred due to wireless communication and sensors. The need to have the lowest order model to reduce computations adds another dimension to the complexity. The parameters used in our experiments are $\tau = 3$, $\rho = 0.999$, and $\beta = 0.15$. The measurements are $\dot{m}_i$, X_{oa} , and X_i .

4) *Experiment:* The experiments were performed on 28th Feb. 2018 for a period of 200 minutes. The mass-flow rate and ambient temperature measured are shown in Fig. 3. To collect the data, we used Apache24 (a HTTP server), PHP7+HTML5 (web), and MySQL (database). We chose these platforms as they are open-source that provides free to use software without licensing constraints. During the experiments, the temperature set-point of the thermostat controlling the VAV is varied from 23 to 27 °C. The model adaptation when the temperature settings changed is reported here. During the experiments the occupancy, heating load and other factors were allowed to be changed and only the set-point changes were made during specific instants. The first data-set (see, Table I(a)) shows the nominal impulse response models and initial condition response estimation computed by using the off-line imperfect data. The identified initial condition response, impulse response with cooling energy and ambient temperature as inputs and zone temperature as the output are also shown in the second and third column of Table I. In addition, the fourth column shows the actual response versus the one computed by the identified model (initial condition plus input responses that are the convolution of impulse responses and inputs). The performance of the proposed method for the first set of data is then evaluated with mean squared error ($MSE = (\sum_{i=1}^m (X_i - \hat{X}_i)^2)/m$) which is 1.59 and presented in the last column of Table I. The off-line model is the nominal model of the proposed approach and the models are adapted during the on-line mode.

During the on-line mode, imperfect measurements are used to update the model recursively. Four iterations of the model updates are shown in the Table I(b)–(e) with different percentages of packet-loss. The nominal model is updated by flipping the mass-flow rate from *high* to *low*, and the 1st batch of the online data is recorded. The impulse responses with the cooling energy and ambient temperature as inputs and the actual measurements as well as the predicted measurements from the identified model are also shown. One can see that with the model adaptation step the modeling error, i.e., MSE reduces to 0.0047. The model obtained in the 2nd step is updated in the 3rd step and approximately 20% packet drop-out is considered. One can see that the model accuracy increases due to recursive identification by about 20%. The

TABLE I

(a) NOMINAL MODEL FROM OFF-LINE DATA; (b) UPDATED MODEL USING NOMINAL MODEL IN FIRST ONLINE MODE; (c) ONLINE ADAPTED MODEL FROM THE PREVIOUS STEP; (d) ONLINE ADAPTED MODEL FROM THE PREVIOUS STEP; AND (e) UPDATED MODEL WITH THE FIFTH DATA-SET ((a) TO (e) ARE NUMBERED TOP ROW TO BOTTOM ROW)

	Initial condition response	Impulse response from cooling energy (KW)	Impulse response from ambient temperature in (X_{oa})	Output room temperature	MSE
1st data set (a)					1.59
2nd data set (b)					0.0047
3rd data set (c)					0.0026
4th data set (d)					0.0002
5th data set (e)					0.0025

4th set of experiments are conducted with 30% data-loss (the highest anticipated loss from the network) and the error increases to 0.0002. Still, the modeling error is low and indicates the efficiency of the approach to recursively approximate complex non-linear behaviors. The 5th batch of data is collected from the experiments with packet-loss of 20%. But due to the recursive model updates, the MSE is 0.0025 over 20 samples. This shows that the accuracy is sensitive to packet losses. Since the model is simple (lowest order and linear), the MPC can be designed to reduce the energy consumption in a distributed manner. Note that in this study, we only identify the nominal model from noisy and incomplete measurements. The identified models were not implemented yet in the controller design stage since we only

focused on the identification part of this study.

B. Simulation Case-Study

In the literature, prediction error method (PEM) and subspace identification methods have been widely studied for system identification. Similarly, two approaches or routines are used for handling imperfect data: *Missdata* and *Merge*. The *Missdata* routine linearly interpolates the missing measurements, and a full data-set is used for identification. On the other hand, the routine *Merge* removes the NAN values and combines data-sets into a set of “multi-experiment data object” which is used for identification. While the proposed algorithm does not require any filling or data-processing approaches such as *Missdata* and *Merge* for handling

imperfect data, other approaches in the literature are limited by this requirement.

To illustrate this aspect, the discrete-time version of the non-linear zone thermal dynamics in (18) sampling period δ is given by [24]

$$\mathbf{X}_i(t_k + 1) + \xi_1 \mathbf{X}_i(t_k) + \xi_2 \dot{m}_i(t_k)(\mathbf{X}_i(t_k) - \mathbf{X}_c) = \xi_3 \mathbf{w}_i(t_k) \quad \forall t_k \in \{0, 1, \dots\} \quad (20)$$

with $\xi_1 = \delta R_i / c_i$, $\xi_2 = \delta c_a / c_i$, and $\mathbf{w}_i(t_k) = \frac{\delta}{c_i} \dot{Q}_i(t_k) + R_i \mathbf{X}_{oa}(t_k)$, respectively. Our idea is to approximate the non-linear behavior using a simple linear model given by

$$\mathbf{X}_i(t_k + 1) + \xi_1 \mathbf{X}_i(t_k) + \xi_2 \mathbf{u}_i(t_k) = \xi_3 \mathbf{w}_i(t_k) \quad (21)$$

where $\mathbf{u}_i(t_k) = \dot{m}_i(\mathbf{X}_i(t_k) - \mathbf{X}_c)$ is used to enforce linear assumption on the system. Consequently, the non-linear dynamics are approximated by a first-order model which can be used to implement sophisticated control actions. Without loss of generality, the approach can work for a system of any order and not restrained to first-order ones, the current example is used to simplify our analysis. The system model considered is given by [25] with $\xi_1 = 0.64$, $\xi_2 = -2.64$, and $\xi_3 = 0.1$, where $\mathbf{X}$ is the room temperature, $\mathbf{u}$ is the cooling energy input, $\mathbf{w}$ is the outside temperature, and ε is the noise. Given the zone thermal model and the sampling rate is 15 minutes, the model is excited with a PRBS input with a length of 3000 minutes. Furthermore, the temperature measure based on this input is contaminated with following white noise $\mathcal{N} \sim (0, 0.25 \times \max(\mathbf{X}_{\text{True Measurement}}))$. Then 30% of the temperature measurements are randomly eliminated to generate data-loss on the output. The proposed approach is used to identify the model with parameter $\tau = 1.9$, to provide a good trade-off between error and model order. The impulse response is computed as $\tilde{\mathbf{X}}$ with inputs $\mathbf{u}$ and $\mathbf{w}$. The model identified by the proposed method with imperfect measurement is $\tilde{\xi}_1 = 0.6782$, $\tilde{\xi}_2 = -2.34$, and $\tilde{\xi}_3 = 0.06$.

C. Benchmark Comparison

The model identified is compared with the PEM and sub-space approach (n4sid) by using *Merge* technique, and a comparison of the impulse responses obtained is shown in Fig. 4. It is evident from the impulse response that the proposed method performs better than existing approaches even with imperfect measurements. A comparison of the model fitting is provided in Table II which also illustrates the utility of the proposed approach.

Similarly, the robustness of the approach to data-loss is tested by randomly exiting the system 500 times with the following conditions:

- 1) the sampling rate is 15 minutes;
- 2) the model is excited with a PRBS input with the length of 2250 minutes;
- 3) white noise is $\mathcal{N}(0, 0.15 \times \max(\mathbf{X}_{\text{True Measurement}}))$
- 4) when 50%, 60%, 70%, 80%, 90%, 100% of temperature measurements are available;
- 5) $\tau_1 = 4.5$ and $\tau_2 = 0.2$.

The robustness is studied by defining the Goodness of Fit measure, given by

Fig. 4. Impulse response estimates.

TABLE II
PROPOSED APPROACH VS PEM AND SUB-SPACE IDENTIFICATION

		$\ \mathbf{X} - \tilde{\mathbf{X}}\ _{\ell_2}$	$\ g - \tilde{g}\ _{\ell_2}^{u_1}$	$\ g - \tilde{g}\ _{\ell_2}^{u_2}$
Merge	Proposed	4.71	0.38	0.04
	n4sid	11.73	0.80	0.18
	PEM	9.39	0.63	0.15
Interpolation	Misdata	24.42	-	-

TABLE III
AVERAGE GOODNESS OF FIT FOR DIFFERENT LEVEL OF DATA-LOSS

	Data-loss rates					
	50%	40%	30%	20%	10%	0%
Min	0.0057	0.0034	0.0041	0.0012	0.0021	0.0019
Average	0.1883	0.1497	0.1003	0.0951	0.0732	0.0653
Max	2.7972	2.6829	2.7481	2.6884	2.5611	0.4497

$$\text{Goodness of Fit} = \frac{\|\mathbf{X} - \tilde{\mathbf{X}}\|_{\ell_2}^2}{m}$$

and the performance of the algorithm is tabulated in Table III. One can observe that the quality of fitting is non-linear with respect to data-loss and the average performance is still within reasonable limits though the maximum is sensitive to data loss. These results illustrate the robustness of the proposed approach vis-à-vis, data-loss.

VI. CONCLUSION

This investigation proposed an approach to recursively estimate complex non-linear behaviors using simple linear models from imperfect measurements for distributed decision-making problems. The method focused on the Internet of Things based control of building heating, ventilation, and air-conditioning system that performs distributed decision making with unreliable networks. We first showed that the problem of computing the lowest order model is a ℓ_0 optimization problem. The influence of multiple inputs and imperfect data made the problem even more challenging. An approach to solve this challenging problem was proposed using the atomic norm concept and a modified Frank-Wolfe algorithm. Further, the minimum amount of data required to identify the lowest

order model was also studied to handle imperfect measurements. The proposed approach was illustrated in a building HVAC system both with experimental and simulation examples. Our results showed the ability of the method to compute the lower order linear model from imperfect measurements. Further, the method showed better performance than existing methods such as prediction error method and sub-space identification that require restrictive assumptions. Extensions to multi-input and multi-output systems are the future course of this investigation. Also, the proposed work can be integrated with building operation, management, or predictive control.

APPENDIX

A. Atomic Set

In the literature, to address the issue of complex terms ($c_p^i, c_p \in \mathbb{C}$ in (4a)), instead of taking linear combination of first order system responses, second order ones have been used [26], leading to real-valued coefficients. To include this modification, we define an atomic set which is denoted by $A(\cdot)$, which takes set of poles, named $P \in D_p$ or $P \in D_{\beta_p}$ (see, Definition 4 in Section III-D) as an input and generates the *Atoms*. The set consists of the first, second order responses, and the one to represent the poles in the unit circle as

$$A(P) = A_1(P) \cup A_2(P) \cup A_3(P) \cup A_4(P)$$

with the scaling factors, $\alpha_p s$, are used as a weighting for each atom to normalize the Hankel nuclear norm of it in the set A [16], [26]⁴. This leads to

$$\begin{aligned} \varphi_p &= \frac{1-p^{2m}}{1-p^2} \quad \text{and} \quad \varphi_a = \frac{1-|p|^{2m}}{1-|p|^2} \\ \Gamma &= \frac{\mathcal{R}(\varphi_p) - \varphi_a - \mathcal{R}(p^2(\bar{p})^{2m}\varphi_p) + |p|^{2N+2}\varphi_a}{1-|p|^2} \\ \alpha_p^1 &= \sqrt{2(\mathcal{R}(\varphi_p^2) + \varphi_a^2) + 2\sqrt{2\Gamma(|\varphi_p|^2 - \varphi_a^2)}}^{-1} \\ \alpha_p^2 &= \sqrt{2(\varphi_a^2 - \mathcal{R}(\varphi_p^2)) + 2\sqrt{2\Gamma(|\varphi_p|^2 - \varphi_a^2)}}^{-1} \\ \alpha_p^3 &= (1-p^2)/(1-p^{2m+2}). \end{aligned}$$

The atomic set is

$$\begin{aligned} A_1(P) &= \left\{ \pm \alpha_p^1 \left(\frac{1}{z-p} + \frac{1}{z-p^*} \right) : p \in P \right\} \\ A_2(P) &= \left\{ \pm \alpha_p^2 \left(\frac{-j}{z-p} + \frac{j}{z-p^*} \right) : p \in P \right\} \\ A_3(P) &= \left\{ \pm \alpha_p^3 \frac{1}{z-p} : p \in P, p \text{ real} \right\} \\ A_4(P) &= \{+1, -1\} \end{aligned} \quad (22)$$

where m is the length of measurement vector $\mathbf{y} \in \mathbb{R}^m$.

B. Convergence Lemma and Theorem

The converges in expectation lemma is

Lemma 1: Algorithm 1 converges in expectation and almost

surely (a.s.) for a fixed number of *Atoms* say N_k satisfying $N_k \geq 1$ for all k , i.e., let f^* be the optimum of problem (12) or (13), then

$$\lim_{k \rightarrow \infty} f(\mathbf{g}_k) - f^* = 0, \text{ a.s.}$$

and

$$\lim_{k \rightarrow \infty} E[f(\mathbf{g}_k)] - f^* = 0.$$

Also, the convergence rate is

Theorem 2: Let $C_2 > 0$, $L > 0$, and $0 < s < 1$ be given constants. Let the number of *Atoms* N_k be large enough so that for any $\mathbf{g}$ with $\|\mathbf{g}\| \leq \tau$. Then, there exists a constant C_1 so that

$$E[f(\mathbf{g}_{k+1}) - f(\mathbf{g}^*)] \leq \frac{C_1}{k+L} + \frac{C_2}{k+L+1}$$

where $\mathbf{g}^*$ is an optimal solution of Problem (13).

The proof of Lemma 1 and Theorem 2 are omitted since it is given in [13].

C. Proof of Theorem 1

Proof: The correlation matrix of the output from (13) is

$$R_y = E(\mathbf{y}\mathbf{y}^T) = E(\mathcal{L}_m \hat{\mathbf{T}} \hat{\mathbf{g}} \hat{\mathbf{g}}^T \hat{\mathbf{T}}^T \mathcal{L}_m^T).$$

From the definition of the convolution sum, we have

$$\begin{aligned} R_y &= E(\mathcal{L}_m \hat{\mathbf{T}} \hat{\mathbf{g}} \mathbf{u} \mathbf{u}^T \hat{\mathbf{T}}^T \mathcal{L}_m^T) = \mathcal{L}_m \hat{\mathbf{T}} \hat{\mathbf{g}} E(\mathbf{u} \mathbf{u}^T) \hat{\mathbf{T}}^T \mathcal{L}_m^T \\ &= \mathcal{L}_m \hat{\mathbf{T}} \hat{\mathbf{g}} R_u \hat{\mathbf{T}}^T \mathcal{L}_m^T. \end{aligned} \quad (23)$$

Considering that the system is persistently excited by the quasi-stationary input $\mathbf{u}$, the following necessary condition needs to hold $\text{rank}(R_u) \geq n$ [21]. Clearly, for the rank deficient matrix $\mathcal{L}_m$, for the system to be persistently excited, the rank conditions can be expanded

$$\text{rank}(R_u) \geq \text{rank}(\mathcal{L}_m) \geq n. \quad (24)$$

Moreover, it is known that if $\mathcal{L}_m \in \mathbb{R}^{m \times m}$, $R_u \in \mathbb{R}^{m \times m}$, then $\text{rank}(\mathcal{L}_m \hat{\mathbf{T}} \hat{\mathbf{g}} R_u \hat{\mathbf{T}}^T \mathcal{L}_m^T) \leq \min(\text{rank}(\mathcal{L}_m), \text{rank}(R_u))$, and Assumption 2, we have

$$\text{rank}(\mathcal{L}_m \hat{\mathbf{T}} \hat{\mathbf{g}} R_u \hat{\mathbf{T}}^T \mathcal{L}_m^T) \leq \text{rank}(\mathcal{L}_m).$$

Therefore, the persistent excitation condition leads to

$$\text{rank}(R_u) \geq \text{rank}(\mathcal{L}_m) \geq \text{rank}(\mathcal{L}_m \hat{\mathbf{T}} \hat{\mathbf{g}} R_u \hat{\mathbf{T}}^T \mathcal{L}_m^T) \geq n.$$

Necessary condition for identifiability of n th order system is given by $\text{rank}(\mathcal{L}_m) \geq n$. ■

Remark 4: This necessary condition is sufficient for noise-free measurements or one can find a sufficient condition for minimum data requirement for noisy measurement if the noise frequency is known.

REFERENCES

- [1] J. Liu and H. Xiao, "Distributed output feedback model predictive control for a team of coupled linear subsystems," *IET Control Theory & Applications*, vol. 11, no. 11, pp. 1807–1812, 2017. [Online]. Available: <https://digital-library.theiet.org/content/journals/10.1049/iet-cta.2016.0751>
- [2] L. D. Xu, W. He, and S. Li, "Internet of things in industries: a survey," *IEEE Trans. Industrial Informatics*, vol. 10, no. 4, pp. 2233–2243, 2014.

⁴ Justification of Hankel norm normalization can be found in [16], [26].

- [3] H. Dibowski, J. Ploennigs, and K. Kabitzsch, "Automated design of building automation systems," *IEEE Trans. Industrial Electronics*, vol. 57, no. 11, pp. 3606–3613, 2010.
- [4] Y.-L. Chang and C.-C. Tsai, "Adaptive generalised predictive temperature control for air conditioning systems," *IET Control Theory & Applications*, vol. 5, no. 6, pp. 813–822, 2011. [Online]. Available: <https://digital-library.theiet.org/content/journals/10.1049/iet-cta.2010.0085>
- [5] D. Li, Y. Zhou, G. Hu, and C. J. Spanos, "Handling incomplete sensor measurements in fault detection and diagnosis for building hvac systems," *IEEE Trans. Automation Science and Engineering*, pp. 1–14, 2019.
- [6] P. Van Overschee and B. De Moor, "N4SID: subspace algorithms for the identification of combined deterministic-stochastic systems," *Automatica*, vol. 30, no. 1, pp. 75–93, 1994.
- [7] G. Mercère, M. Lovera, L. Bako, and S. Lecoeuche, "Recursive subspace identification of Hammerstein models based on least squares support vector machines," *IET Control Theory & Applications*, vol. 3, no. 9, pp. 1209–1216, 2009. [Online]. Available: <https://digital-library.theiet.org/content/journals/10.1049/iet-cta.2008.0339>
- [8] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.
- [9] M. Sznaier, M. Ayazoglu, and T. Inanc, "Fast structured nuclear norm minimization with applications to set membership systems identification," *IEEE Trans. Automatic Control*, vol. 59, no. 10, pp. 2837–2842, 2014. [Online]. Available: <http://ieeexplore.ieee.org/articleDetails.jsp?arnumber=6778049>
- [10] Z. Liu and L. Vandenberghe, "Interior-point method for nuclear norm approximation with application to system identification," *SIAM J. Matrix Analysis and Applications*, vol. 31, no. 3, pp. 1235–1256, 2009.
- [11] K. Bekiroglu, S. Srinivasan, E. Png, R. Su, K. Poolla, and C. Lagoa, "An internet of things compliant model identification methodology for smart buildings," in *Proc. IEEE 56th Annual Conf. Decision and Control*, 2017, pp. 4440–4445.
- [12] P. Van Dooren, K. Gallivan, and P.-A. Absil, " H_2 -optimal model reduction with higher-order poles," *SIAM J. Matrix Analysis and Applications*, vol. 31, pp. 2738, 2010.
- [13] B. Yilmaz, K. Bekiroglu, C. Lagoa, and M. Sznaier, "A randomized algorithm for parsimonious model identification," *IEEE Trans. Automatic Control*, vol. 63, no. 2, pp. 532–539, 2018.
- [14] D. L. Donoho, "Compressed sensing," *IEEE Trans. Information Theory*, vol. 52, pp. 1289–1306, 2006.
- [15] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The convex geometry of linear inverse problems," *Foundations of Computational Mathematics*, vol. 12, no. 6, pp. 805–849, 2012.
- [16] P. Shah, B. N. Bhaskar, G. Tang, and B. Recht, "Linear system identification via atomic norm regularization," in *Proc. IEEE 51st Conf. Decision and Control*, 2012, pp. 6265–6270.
- [17] A. Tewari, P. K. Ravikumar, and I. S. Dhillon, "Greedy algorithms for structurally constrained high dimensional problems," *Advances in Neural Information Processing Systems*, pp. 882–890, 2011.
- [18] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [19] R. K. Lim and R. W. Longman, "State-space system identification with identified Hankel matrix," *Department of Mechanical and Aerospace Engineering Technical Report No. 3045*, no. 3045, pp. 1–36, 1998.
- [20] I. Boz, K. Bekiroglu, M. Solaimanian, P. Tavassoti-Kheiry, C. Lagoa, T.-k. Pexbouhan, and C. Lagoa, "Validation of model order assumption and noise reduction method for the impact resonance testing of asphalt concrete," *J. Nondestructive Evaluation*, vol. 36, no. 3, pp. 58, 2017.
- [21] L. Ljung, "System identification," in *Signal Analysis and Prediction*. Springer, 1998, pp. 163–173.
- [22] S. Baldi, S. Yuan, P. Endel, and O. Holub, "Dual estimation: constructing building energy models from data sampled at low rate," *Applied Energy*, vol. 169, pp. 81–92, 2016.
- [23] E. Png, S. Srinivasan, K. Bekiroglu, J. Chaoyang, R. Su, and K. Poolla, "An internet of things upgrade for smart and scalable heating, ventilation and air-conditioning control in commercial buildings," *Applied Energy*, vol. 239, pp. 408–424, 2019.
- [24] N. Radhakrishnan, S. Srinivasan, R. Su, and K. Poolla, "Learning based hierarchical distributed HVAC scheduling with operational constraints," *IEEE Trans. Control Systems Technology*, vol. 26, no. 5, pp. 1892–1900, 2017.
- [25] A. Aswani, N. Master, J. Taneja, V. Smith, A. Krioukov, D. Culler, and C. Tomlin, "Identifying models of HVAC systems using semiparametric regression," in *Proc. American Control Conf.*, 2012, pp. 3675–3680.
- [26] K. Bekiroglu, B. Yilmaz, C. Lagoa, and M. Sznaier, "Parsimonious model identification via atomic norm minimization," in *Proc. European Control Conf.*, pp. 2392–2397, 2014.

Korkut Bekiroglu received the B.S. degree from the Department of Electrical and Electronics Engineering of Black Sea Technical University, Turkey, in 2005, the M.S. degree from the Northeastern University, USA, in 2010, and the Ph.D. degree from Penn State University, USA and University Park, USA, in 2015, both in the Department of Electrical Engineering. After graduation, in August 2015, he was appointed as a Postdoctoral Researcher in the Methodology Center at Pennsylvania State University, USA until 2016. He is currently an Assistant Professor of electrical engineering technology at the College of Engineering, SUNY Polytechnic Institute, USA. Before joining SUNY Polytechnic Institute, Dr. Bekiroglu was a Research Fellow working on the Green Building Innovation Cluster (GBIC) project at Nanyang Technological University, Singapore. His research interests include system identification, model predictive control, optimization, robust optimization, discrete event dynamical systems, mechatronics, building automation, signal processing, and robust control.

Seshadhri Srinivasan is currently a Postdoctoral at Berkeley Education Alliance for Research in Singapore (BEARS), Singapore, working on building signal processing, smart grid data imputation, fault-diagnosis of buildings, occupancy-based energy control of HVAC systems, multi-level optimization of Building Energy Management, Singapore. He is part of the SinBerBEST1, SinBerBEST2, and GBIC project. Dr. Srinivasan serves the IEEE CSS as a Member of Standardization Committee. Currently, he is working with Prof. Kameshwar Poolla, Dept. of Mechanical Engineering, UC Berkeley on multi-level optimal control and deep learning. His research in-

terests include mathematical optimization, cyber-physical systems, and industrial automation. He received the Ph.D. degree with National Institute of Technology-Tiruchirappalli, India on networked control systems and has been part of various projects during his different stints as a Researcher and as an Academician.

Ethan Png received the M.Sc. computation degree from The University of Manchester Institute of Science and Technology (UMIST), UK, in 2004. His research interests include visible light communication, IoT, intelligent self-driving bicycles, artificial neural networks, medical robots, IC design, and condensed matter physics (semiconductors).

Rong Su received the bachelor of engineering degree from University of Science and Technology of China, and master of applied science degree, and Ph.D. degree from University of Toronto, Canada, respectively. He was affiliated with University of Waterloo, Canada, and Technical University of Eindhoven, Netherlands before he joined Nanyang Technological University, Singapore in 2010. Dr. Su's research interests include multi-agent systems, discrete-event system theory, model-based fault diagnosis, operation planning and scheduling with applications in flexible manufacturing, intelligent transportation, human-robot interface, power manage-

ment, and green building. In the aforementioned areas he has more than 110 journal and conference publications and 2 patents, and has been involved in several projects sponsored by Singapore National Research Foundation (NRF), Singapore Agency of Science, Technology and Research (A*STAR), Singapore Ministry of Education (MoE), Singapore Civil Aviation Authority (CAAS), and Singapore Economic Development Board (EDB). Prof. Su is a Senior Member of IEEE, and an Associate Editor for *Transactions of the Institute of Measurement and Control*, *Journal of Control and Decision*, and *Journal of Discrete Event Dynamic Systems: Theory and Applications*. He is also the Chair of the Technical Committee on Smart Cities in the *IEEE Control Systems Society*, and a Board Member of the School of EEE at NTU.

Constantino Lagoa received the B.S. and M.S. degrees from the Instituto Superior Tecnico, Technical University of Lisbon, Portugal, in 1991 and 1994, respectively, and the Ph.D. degree from the University of Wisconsin at Madison, USA, in 1998. He joined the Electrical Engineering Department of Pennsylvania State University, University Park, USA, in August 1998, where he currently holds the position of Professor. He has a wide range of research interests including robust optimization and control, chance constrained optimization, controller design under risk specifications, system identification, control of computer networks, and discrete event dynamical systems. Dr. Lagoa is currently an Associate Editor of *Automatica* and was an Associate Editor for *IEEE Transactions on Automatic Control* and *IEEE Transactions on Control Systems Technology*. He is a Member of the IFAC Technical Committee on Robust Control and has served as a Member of the International Program Committee of several conferences.