INNER SOURCE IDENTIFICATION FOR FIELD ESTIMATION IN WIRELESS SENSOR NETWORKS

Haotian Zhang, Bruce Krogh, and José M. F. Moura

Carnegie Mellon University
Dept. of Electrical and Computer Engineering
Pittsburgh, PA 15213, USA
{haotian@andrew, krogh@ece, moura@ece}.cmu.edu

ABSTRACT

In previous work, we presented a method for constructing dynamic input-output models for real-time state estimation in correlated dynamic fields using wireless sensor networks (WSNs). The input signals correspond to the sensors on the boundary of the physical space under study, reflecting the assumption there are no independent sources in the interior of the region. In this paper, we present a method to identify sensors not on the boundary that should be used as inputs in the dynamic models when there are inner sources, i.e., sources inside the physical space. We extend concepts from the behavioral model theory of Jan Willems to handle noisy time series. We construct a block-Hankel structured matrix based on the sensor data, and then calculate as an independence indicator the angle between each row of the matrix and the subspace determined by the span of the preceding rows. The inner sources are identified based on these angle values. Experimental results with real temperature data collected with a WSN, and including heat sources as inner sources, illustrate the proposed method.

1. INTRODUCTION

Wireless sensor networks (WSNs) are promising for large-scale environmental monitoring and control applications, such as temperature regulation in office buildings [1] or data centers [2]. In [3] and [4], we propose a method for constructing input-output models to estimate values of correlated dynamic fields using real-time measurements from a WSN. The input signals for these models are identified as the sensors on the boundary of the physical space being monitored, based on the assumption that there are no independent sources inside this region. In many applications, this is not a realistic assumption–there will typically be sources (e.g., heat sources) within the region. This paper is concerned with the problem of identifying which sensor signals should be used as inputs in the input-output model used for real-time state estimation.

In 1986 and 1987, Jan C. Willems published a sequence of papers [5, 6, 7] that proposed a new methodology for obtaining exact or approximate dynamical models for linear time-invariant systems from a set of observed *deterministic* time series. This work was further developed into a general approach, called *the behavioral approach*, for modeling and analysis of multivariable dynamical systems [8]. We recast our WSN problem of identifying the inner sources in Willems' behavioral approach, except that we are concerned with finding a method to distinguish the inputs from outputs in a set of observed *noisy* time series, e.g., a set of sensor measurements series.

One of the key features in Willems' methodology is that dynamical systems are considered without distinguishing between inputs and outputs a priori, i.e., without identifying which components of the set of observed time series are inputs and which are outputs [5]. Causality is a matter of representation and not an a priori axiom that needs to be imposed on the studied systems. The inputs and outputs are distinguished, as a byproduct, in the process of exact modelling, based on the linear independence of those time series [6]. However, the approach proposed in [6] is valid only for deterministic/noiseless time series, rather than noisy time series, as it is the case with sensor measurements in WSN. In this paper we develop a general method extending Willems' approach to distinguish inputs from a set of noisy time series.

This paper is organized as follows. Section 2 states the problem we consider, and describes briefly Willems' algorithm for deterministic/noiseless time series. In Section 3, we propose an algorithm to identify inputs within a set of *noisy* measurement series. In Section 4, we modify and apply the proposed algorithm to data from WSNs to identify inner sources in our field estimation approach developed in [4]. Section 5 presents experimental results applying this method to identify inner sources based on data from a WSN of 18 sensors in a large laboratory space with internal heat sources. Section 6 concludes the paper.

2. PROBLEM FORMULATION

Our overall goal is to estimate a random field from sensor measurements provided by a collection of sensors in a WSN. We assume that at the locations where it is desired to estimate

Research supported by NSF grants # ECS-0225449 and # CNS-0428404.

the field there are no sensors. An important step in our solution is to fit a dynamical system to model the random field. An intermediate stage in this process is to distinguish within the sensor measurements which ones correspond to field sources. We refer to this as the inner source identification problem.

Suppose that the data collected by a WSN are noisy measurements of the inputs and outputs of an LTI dynamic systems, but the separation into input and output signals is not given. Our target is to identify the input signals from these noisy measurements.

Assume that we have a set of q finite time series: $\mathbf{w_i} = [w_i(1) \ \cdots \ w_i(T)]$, $i = 1, \cdots, q$. The task is to identify which of these q finite time series should be regarded as inputs. Let $\tilde{\mathbf{w}}(t) = [w_1(t) \ \cdots \ w_q(t)]^T$ and $\tilde{\mathbf{w}} = [\tilde{\mathbf{w}}(1) \ \cdots \ \tilde{\mathbf{w}}(T)]$. Assume that the AR model of the q time series is: $R_0\tilde{\mathbf{w}}(t) + R_1\tilde{\mathbf{w}}(t+1) + \cdots + R_L\tilde{\mathbf{w}}(t+L) = 0$, where $R_0, R_1, \cdots, R_L \in \mathbb{R}^{g\times q}$, g is the number of system laws, and L is the *memory length* of the system. We consider a moving window, with window width $\Delta > L$, in order to identify the driving components of $\tilde{\mathbf{w}}$, i.e., the inputs. Our method examines the following $q(\Delta+1)\times(T-\Delta)$ *block-Hankel structured matrix* $H(\tilde{\mathbf{w}})$ from the data:

$$\begin{bmatrix} \tilde{\mathbf{w}}(1) & \tilde{\mathbf{w}}(2) & \cdots & \cdots & \tilde{\mathbf{w}}(T-\Delta) \\ \tilde{\mathbf{w}}(2) & \tilde{\mathbf{w}}(3) & \cdots & \cdots & \tilde{\mathbf{w}}(T-\Delta+1) \\ \tilde{\mathbf{w}}(3) & \tilde{\mathbf{w}}(4) & \cdots & \cdots & \tilde{\mathbf{w}}(T-\Delta+2) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \tilde{\mathbf{w}}(\Delta+1) & \tilde{\mathbf{w}}(\Delta+2) & \cdots & \cdots & \tilde{\mathbf{w}}(T) \end{bmatrix} \quad (1)$$

Willems developed an algorithm in [6] to identify a set of input signals in the AR model from a given set of *noiseless* time series by considering the linear dependence of each row of $H(\tilde{\mathbf{w}})$ with respect to the preceding rows, starting from the top row. This results in a dependency column vector $\mathbf{d}$, composed of $*$'s and $\bullet$'s. A $*$ or a $\bullet$ in the i^{th} component shows that the i^{th} row of $H(\tilde{\mathbf{w}})$ is linearly independent of, or dependent on, the preceding rows, respectively. In the case considered in [6] where $T = \infty$, the Hankel structure of $H(\tilde{\mathbf{w}})$ implies that if $\mathbf{d}$ has a $\bullet$ in its i^{th} row, it will have a $\bullet$ in its $(i+nq)$ row for all possible n's satisfying $(i+nq) < q(\Delta+1)$. Let j be the index of a given time series, i.e., $1 \leq j \leq q$. If there is a $*$ in the $(j+nq)^{th}$ row of $\mathbf{d}$ for all $1 \leq n \leq \Delta$, then the j^{th} time series in $\tilde{\mathbf{w}}$ should be regarded as an input. This approach proposed in [6] is valid only for *noiseless* time series, rather than noisy time series such as the signals from a WSN. When the data is corrupted by noise, even very weak noise, Willems' algorithm gives the meaningless result that all time series are inputs. In the next section, we propose a generalization of this method to distinguish inputs from a set of noisy time series.

3. INPUT IDENTIFICATION WITHIN A SET OF NOISY TIME SERIES

Willems' method in [6] makes "hard" decisions, represented by $*$ or $\bullet$, on the independence of each row of the Hankel matrix $H(\tilde{\mathbf{w}})$. We explore now replacing the hard decisions by soft decisions with the hope that soft decisions on the independence of each row of the Hankel matrix $H(\tilde{\mathbf{w}})$ will lead to an identification of the inner sources when the measurements are noisy. Such soft decisions can be realized by evaluating the *angle* θ between the vector h corresponding to the current row of the Hankel matrix $H(\tilde{\mathbf{w}})$ and the subspace $\mathcal{S}$ spanned by the preceding rows. Let

$$\theta(h, \mathcal{S}) = cos^{-1}\left(\left|\frac{\langle h, \Pi_{\mathcal{S}}h\rangle}{|h||\Pi_{\mathcal{S}}h|}\right|\right), \quad (2)$$

where $\Pi_{\mathcal{S}}(\bullet)$ is the projection operator on $\mathcal{S}$ and $\langle\cdot,\cdot\rangle$ is the inner product between the two vectors.

If a row h of $H(\tilde{\mathbf{w}})$ is independent from the preceding rows, the angle θ between this row h and the subspace $\mathcal{S}$ specified by the preceding rows should be nonzero; if this row is dependent on the preceding rows, the angle θ should be 0. Larger values of the angle θ represent stronger independence of the corresponding row of $H(\tilde{\mathbf{w}})$.

We calculate the angle θ for each row of the Hankel matrix $H(\tilde{\mathbf{w}})$. Let $\overline{\theta}_i$ be the average of these angles θ corresponding to delayed rows of the same time series. We collect these average angles into the *average angle vector* $\mathbf{v}_a = [\overline{\theta}_1 \ \cdots \overline{\theta}_q]$.

Let T_0 be the total length of these time series; we only need to work with the first T samples of each time series in $\tilde{\mathbf{w}}$, where $T < T_0$. Let Δ be the width of the moving window. Since the row length of the Hankel matrix $H(\tilde{\mathbf{w}})$ is $T-\Delta$, T must be larger than Δ.

Algorithm 1 (Input Identification)

1. **Initialization** *Given Δ, let $T = \Delta + c$, where c is a small integer constant whose value chosen empirically. Build the Hankel matrix $H(\tilde{\mathbf{w}})$ from the data set $\tilde{\mathbf{w}}$ for the T and Δ, as in (1).*

2. For $(q+1) \leq i \leq q(\Delta+1)$, compute $\theta_i = \theta(h_i, \mathcal{S}_{i-1})$ by Equation (2), where h_i is the i^{th} row of $H(\tilde{\mathbf{w}})$, and $\mathcal{S}_{i-1}$ is the subspace spanned by the first $(i-1)$ rows of $H(\tilde{\mathbf{w}})$.

3. For $1 \leq i \leq q$, let

$$v_i = \frac{1}{\Delta}\sum_{n=1}^{\Delta+1}\theta_{i+nq}. \quad (3)$$

4. For $1 \leq j \leq q$, if $v_j > 0$, then declare the time series $\mathbf{w_j}$ an inner source.

5. End.

In Step 2, we do not compute the angle θ for the first q rows of $H(\tilde{\mathbf{w}})$ because these rows are the first segments of the q time series and they do not provide reliable information on the dependence relationship among these time series.

4. INNER SOURCE IDENTIFICATION IN WIRELESS SENSOR NETWORKS

We proposed in [4] an approach to estimate the real-time values of a physical field at specific locations of interest $R = \{r^1, \ldots, r^n\}$ based on measurements at sensor locations $S = \{s^1, \ldots, s^m\}$, where R is not necessarily contained in S. To solve this real-time estimation problem in WSNs, we defined the concept of cut point sets, and derived reduced-order models based on the cut point sets to implement the real-time field estimation. The cut point set can be regarded as the "boundary" of the field region of interest where the field values are measured by wireless sensors. For the reduced-order model derived from a cut point set X_c, its input usually includes all the sensors in X_c, and its output usually includes all the sensors inside X_c, i.e., encompassed by the sensors in X_c. However, if one or more sensors inside X_c represent sources of the studied field, called inner sources, these sensors should be regarded inputs, rathern than outputs, of the reduced-order model. Therefore, it is an important problem to identify inner sources from the sensors inside the cut point set X_c. Readers can refer to [4] for details.

We modify Algorithm 1 to identify inner sources in WSNs. In our application for field estimation with a WSN, the q time series include measurements of the sensors that belong to the cut point set X_c and of the q' sensors inside X_c, i.e., the measurable vertices inside the extended state vertex set X'_s. We compute the average angle vector $\mathbf{v}_a$ based on all these time series, but we only track its components corresponding to the q' inner sensors, since our task is to identify the inner sources; the sensors on the boundary, i.e., in the cut point set X_c, are already identified as inputs to the dynamic model. Let $\mathbf{v}'_a$ be the average angle vector corresponding to the concerned q' components of $\mathbf{v}_a$.

Algorithm 2 (Inner Source Identification in WSNs)

1. *Same as Algorithm 1.*
2. *Same as Algorithm 1.*
3. *Same as Algorithm 1.*
4. *Let* $\mathbf{v}' = [v_{n_1} \ \cdots \ v_{n_{q'}}]$, *where* $v_{n_j}, j = 1, \cdots, q'$, *correspond to the* q' *time series in* $\tilde{\mathbf{w}}'$.
5. *For* $1 \leq j \leq q'$, *if* $v_{n_j} > 0$, *then declare the time series* $\mathbf{w_{n_j}}$ *an inner source.*
6. *End.*

5. EXPERIMENTAL RESULTS

We collected temperature data using 18 wireless sensor nodes in a laboratory at Carnegie Mellon University with two controllable electrical heaters. Figure 1 shows the layout of the experiment. The locations of the heaters are indicated by small rectangles labeled 1 and 2 and the sensor locations are indicated by circles labeled 1 to 18. Figure 2 shows the experimental set-up. Both heaters are controlled by remote power control devices [9]. We can set an on-off power sequence for each heater to generate a desired temperature variation pattern in the room.

The total duration of the experiment we carried out is 15 hours, and the power on-off sequences for the two controllable heaters are shown in Figure 3. The sampling rate for each sensor is 0.5 Hz, but the samples for different sensors are not synchronized. We use linear interpolation to obtain synchronized sensor measurements for all 18 sensors with a sampling rate of 1 Hz. We show the 18 time series in Figure 4,

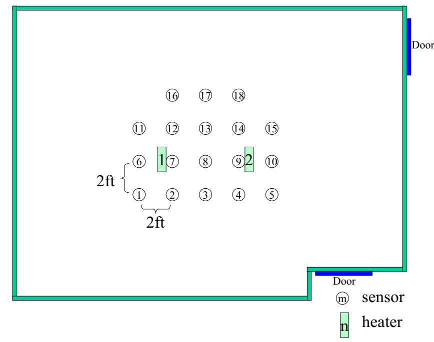

Fig. 1. Experimental layout of sensors.

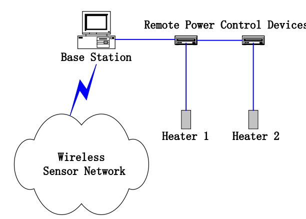

Fig. 2. Experimental set-up.

where the top curve shown in red (if color is available) is the first inner source S_7, the second top curve, in purple, is the second inner source S_9, and the remaining curves, all drawn in blue, are the measurements for the other 16 sensors.

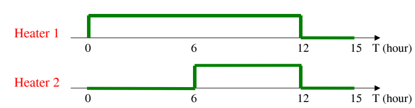

Fig. 3. Power on-off sequence for the heaters.

Each time series in our experimental data has about 54,000 samples. Because we only need a short segment of these time series to identify the inputs, we use the following two sets of time series:

$\tilde{\mathbf{w}}_\mathbf{a}$ — includes 18 time series of 1000 samples each, starting at the 10,000 s sample and ending at the 10,999 s sample. In this segment, there is only one inner source S_7.

$\tilde{\mathbf{w}}_\mathbf{b}$ — includes 18 time series of 1000 samples each, starting at the 30,000 s sample and ending at the 30,999 s sample. In this segment, there are two inner sources S_7 and S_9.

For both $\tilde{\mathbf{w}}_\mathbf{a}$ and $\tilde{\mathbf{w}}_\mathbf{b}$, we apply Algorithm 2 for different values of T, i.e., different values of $c = T - \Delta$. As explained in the last section, we compute the average angle vector $\mathbf{v}_a$ based on the measurement of all the 18 sensors, but we only track its components corresponding to the 6 inner sensors, i.e., $\{S_7, S_8, S_9, S_{12}, S_{13}, S_{14}\}$, to identify the inner sources. The results for $\tilde{\mathbf{w}}_\mathbf{a}$ and $\tilde{\mathbf{w}}_\mathbf{b}$ are shown in Figure 5 and Figure 6, respectively.

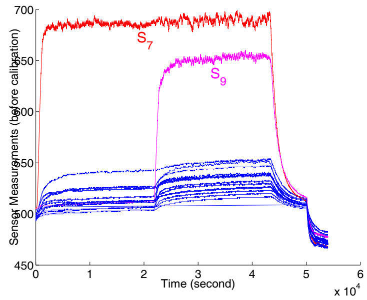

Fig. 4. Experimental data for 18 sensors.

Figure 5 shows that, when $0 < c < 46$, there is only one inner source, the inner source S_7 represented by the line plot with circle marks, in red, that is successfully and correctly identified as an inner source, since the value of its corresponding component in $\mathbf{v}_a$ is the only nonzero component of $\mathbf{v}_a$.

Figure 5 shows the results when $0 < c < 150$. In this case, both heaters are turned on, and so there are two inner sources collocated with sensors S_7 and S_9. The two line plots in Figure 5, the red one with circle marks and the blue one with star marks, successfully identify both sensors S_7 and S_9 as the correct inner sources.

Figure 5 and Figure 6 show that Algorithm 2 successfully identified the inner sources within the group of real data noisy time series; the constant c is chosen to have a small positive value.

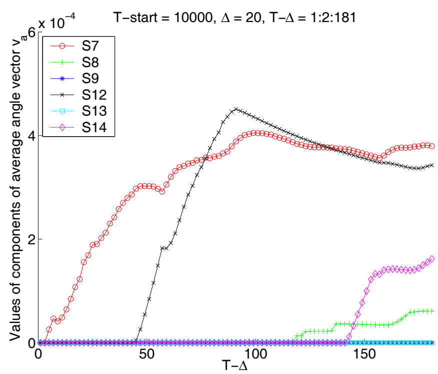

Fig. 5. Result for experimental data $\tilde{\mathbf{w}}_\mathbf{a}$: one inner source S_7.

6. CONCLUSION

This paper considers the problem in wireless sensor networks of identifying which sensor signals should be considered as input signals in models used to estimate a correlated field at locations where there are no available sensors. These signals are called inner sources. We present an algorithm to identify the inner sources when the time series are noisy, making it possible to relax the assumption in [6] that all inputs signals are on the boundary. Experimental results with real data collected by a sensor network with 18 sensors and two heaters verify the validity of the proposed algorithm.

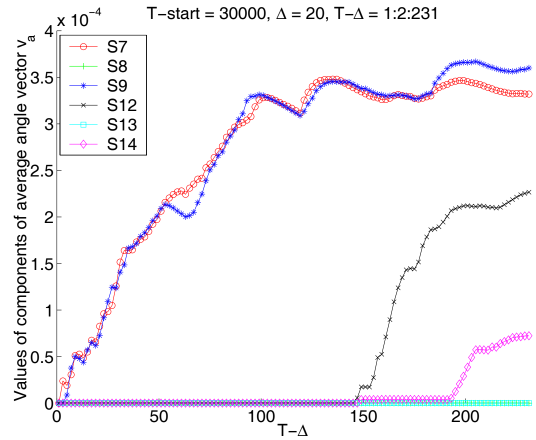

Fig. 6. Result for experimental data $\tilde{\mathbf{w}}_\mathbf{b}$: two inner sources S_7 and S_9.

7. REFERENCES

[1] A. T. P. So, *Perspectives in Control Engineering*, chapter Building control and automation systems, pp. 393–416, IEEE Press, 2001.

[2] C. D. Patel, "A vision of energy aware computing from chips to data centers," in *The International Symposium on Micro-Mechanical Enigneering*, Dec 2003.

[3] H. Zhang, B. H. Krogh, and J. M. F. Moura, "Estimation in virtual sensor-actuator arrays using reduced-order physical models," in *IEEE Conf. on Decision and Control*, Dec 2004.

[4] H. Zhang, J. M. F. Moura, and B. H. Krogh, "Estimation in sensor networks: a graph approach," in *Fourth International Symposium on Information Processing in Sensor Networks (IPSN 2005)*, Los Angeles, CA, USA, Apr. 25-27 2005, pp. 203–209.

[5] J. C. Willems, "From time series to linear system - part i. finite dimensional linear invariant systems," *Automatica*, vol. 22, no. 5, pp. 561–580, 1986.

[6] J. C. Willems, "From time series to linear system - Part II. Exact modelling," *Automatica*, vol. 22, no. 6, pp. 675–694, 1986.

[7] J. C. Willems, "From time series to linear system - part iii. approximate modelling," *Automatica*, vol. 23, no. 1, pp. 87–115, 1987.

[8] I. Markovsky, J. C. Willems, S. Van Huffel, and B. L. M. De Moor, *Exact and Approximate Modeling of Linear Systems: A Behavioral Approach*, SIAM Publications, 2005.

[9] *http://www.wti.com/rps-10.htm*.