

On Recursive Proper Orthogonal Decomposition via Perturbation Theory with Applications to Distributed Sensing in Cyber-Physical Systems

Chao Xu, Lixiang Luo and Eugenio Schuster

Abstract—Distributed sensing of cyber-physical systems has become feasible with recent developments in sensor technology, wireless communication and distributed computing. Distributed sensing generates huge amounts of data from the events occurring in the physical side, which should be promptly reflected in the cyber side so that actions can be made timely by the computing systems. Due to the dense temporal-spatial distribution of the measured data, great challenges have been posed in terms of data storage, information processing and communications. The proper orthogonal decomposition (POD) method is a powerful tool to extract dominant information from distributed observational data, which has been widely used in signal processing and pattern analysis of fluid turbulence. The classical POD method implements dominant information extraction when the entire data set is known. However, in real-time measurements, new data is collected and incorporated into the historic data set at each sampling time. We propose a recursive proper orthogonal decomposition (rPOD) method based on the operator perturbation theory, where the accumulative truncation error can be controlled by a gradient search algorithm. This method is illustrated with two state-of-the-art problems governed by the heat conduction equation (1D) and the Navier-Stokes equations (2D) respectively.

I. INTRODUCTION

Cyber-Physical Systems (CPS) are the integration of information/cyber systems (including measurement, communication, computation and control) with physical processes. Embedded computers and networks monitor and control the physical processes, usually with feedback loops where physical processes affect computations and vice versa [1].

Distributed sensing is widely used in monitoring physical processes where huge amounts of data are then generated from temporally-spatially distributed measurements, greatly challenging data storage, information processing and communications (e.g., [2], [3], [4]). The proper orthogonal decomposition (POD) [5] method is a powerful tool to extract dominant information from distributed observational data, which has been widely used in signal processing and pattern analysis of fluid turbulence. The classical POD method enables dominant information extraction when the entire data set is known. However, in real-time measurements, new data is collected and incorporated into the historic data set at each sampling time. In order to update the dominant modes, solving the POD problem at each sampling time with newly collected observational data may be infeasible when the data vector dimension is quite large.

Recursive POD methods can be very valuable for systems controlled in closed loop where the dominant POD modes vary over time as the system is excited by new inputs. For instance, receding horizon control of complex physical processes may benefit from these methods. Traditional approaches towards the update of the POD modes at each sampling time include the increase of the data set by the newly measured data, which implies an increasing computational burden, or the forward shift of the time window defining the fixed-length data set, which implies the neglect of older data. A recursive approach could take into account all the historic data while keeping the computational burden low.

For signals arising in distributed sensing, the spatial dimension is usually quite large and the method of snapshots [6], [7] is usually needed to reduce the order of the POD problem. By combining the method of snapshots with the operator perturbation theory [8], we propose in this work a recursive POD method. The operator perturbation theory was previously exploited in [9] to recursively obtain the eigen-decomposition of time-varying covariance matrices of signal arrays with restricted dimensions. In order to control the accumulative truncation error imposed by the operator perturbation theory, we propose a method based on gradient search techniques to track the estimate error of the recursive POD method and if necessary to implement an estimate correction. In [10], a recursive POD method is proposed based exclusively on gradient search techniques. This is a general method that also updates the POD modes by inheriting information from historic data, but the POD mode update usually takes longer than in the operator perturbation theory approach. In addition, the search gradient approach does not provide an update of the corresponding POD eigenvalues. As an alternative to reshaping the POD modes recursively, a dimensionality update method has been recently proposed in [11] to increase or reduce the number of POD modes in order to achieve certain approximation accuracy. We finally apply the proposed method to carry out dynamic pattern extraction of data arising from the simulations of both a 1D heat conduction model and a 2D fluid past a cylinder model.

We organize this paper as follows. In Section II, we summarize the POD method including both integral and matrix versions. In Section III, we motivate the perturbation formulation in the the POD problem. We summarize the rPOD algorithms in Section IV and validate the solutions in Section V. We close this paper in Section VI by stating conclusions and potential research topics.

This work was supported by the NSF CAREER award program (ECCS-0645086).

C. Xu (chx205@lehigh.edu), L. Luo and E. Schuster are with the Department of Mechanical Engineering and Mechanics, Lehigh University, 19 Memorial Drive West, Bethlehem, PA 18015, USA.

II. PROPER ORTHOGONAL DECOMPOSITION

A. Empirical eigenvalue problem

We give a basic introduction of the POD method over a finite spatial interval $\Omega \in \mathbb{R}$. Assuming that the spatial-temporal evolution of a variable over Ω is denoted by $x(\xi, t)$, where $\xi \in \Omega$ is the spatial coordinate and $t \in [t_1, t_F]$ is the time coordinate. We assume that $x(\xi, t)$ is square integrable for any given $t \in [t_1, t_F]$, i.e., $\int_{\Omega} x^2(\xi, t) d\xi < \infty$ or $x(\xi, \cdot) \in L^2(\Omega)$. We take measurements of the evolution $x(\xi, t)$ over certain discrete time instants $t_n \in [t_1, t_F]$, where $n = 1, 2, \dots, N$, and define $x(\xi, t_n) = x_n(\xi)$ as a *snapshot*, for any integer $1 \leq n \leq N$. All the snapshots $\{x_n(\xi)\}_{n=1}^N$ form a ‘‘curve’’ ensemble and the POD problem is to extract dominant features from all the measurements. Before mathematically stating the POD method, we introduce two necessary definitions below:

- 1) L^2 -inner product $(\cdot, \cdot)_{L^2}$: Given any two snapshots $x_i(\xi)$ and $x_j(\xi)$, the inner product is defined as $(x_i, x_j)_{L^2} = \int_{\Omega} x_i(\xi)x_j(\xi)d\xi$, and the induced L^2 -norm is denoted as $\|\cdot\|_{L^2}$;
- 2) Ensemble average $\langle \cdot \rangle$: Given snapshots $x_n(\xi)$, $n = 1, 2, \dots, N$, the ensemble average is defined by $\langle x_n \rangle_{n=1}^N = \frac{1}{N} \sum_{n=1}^N x_n(\xi)$.

Thus, the POD method can be formulated as the following normalized optimization problem

$$\max_{\varphi_i \in L^2(\Omega)} \frac{\langle (x_n, \varphi_i)_{L^2} \rangle_{n=1}^N}{\|\varphi_i\|^2}, \quad (1)$$

i.e., as choosing a basis function $\varphi_i(\xi)$ ($i = 1, \dots, N$) to maximize the averaged projection of the data ensemble $\{x_n\}_{n=1}^N$ onto φ_i . By introducing the orthonormal constraint for the to-be-obtained basis function φ_i , i.e., $\|\varphi_i\| = 1$, it is possible to obtain the following augmented cost function

$$J[\varphi_i] = \langle (x_n, \varphi_i)_{L^2}^2 \rangle_{n=1}^N - \lambda_i (\|\varphi_i\| - 1), \quad (2)$$

where λ_i is the Lagrangian multiplier. The optimality condition can be stated as

$$\left. \frac{dJ[\varphi_i + \delta_i \phi_i]}{d\delta_i} \right|_{\delta_i=0} = 0, \quad \forall \varphi_i + \delta_i \phi_i \in L^2(\Omega), \delta_i \in \mathbb{R}. \quad (3)$$

By defining $\mathfrak{R}_N \varphi_i = \int_{\Omega} \langle x_n(\xi)x_n(\xi') \rangle_{n=1}^N \varphi_i(\xi') d\xi'$, we can rewrite

$$(\mathfrak{R}_N \varphi_i, \varphi_i)_{L^2} = \int_{\Omega} \mathfrak{R}_N \varphi_i(\xi) \varphi_i(\xi) d\xi = \langle (x_n, \varphi_i)_{L^2}^2 \rangle_{n=1}^N.$$

Then, we can rewrite $J[\varphi_i] = (\mathfrak{R}_N \varphi_i, \varphi_i)_{L^2} - \lambda_i [(\varphi_i, \varphi_i)_{L^2} - 1]$ and we can expand the cost variation as

$$\begin{aligned} & J[\varphi_i + \delta_i \phi_i] \\ &= (\mathfrak{R}_N \varphi_i, \varphi_i)_{L^2} + 2\delta_i (\mathfrak{R}_N \varphi_i, \phi_i)_{L^2} + \delta_i^2 (\mathfrak{R}_N \phi_i, \phi_i)_{L^2} \\ & \quad - \lambda_i [(\varphi_i, \varphi_i)_{L^2} + 2\delta_i (\varphi_i, \phi_i)_{L^2} + \delta_i^2 (\phi_i, \phi_i)_{L^2}] + \lambda_i. \end{aligned}$$

Therefore, the POD optimality condition (3) becomes $(\mathfrak{R}_N \varphi_i - \lambda_i \varphi_i, \varphi_i)_{L^2} = 0$ which is equivalent to the following integral equation

$$\int_{\Omega} \langle x_n(\xi)x_n(\xi') \rangle_{n=1}^N \varphi_i(\xi') d\xi' = \lambda_i \varphi_i(\xi). \quad (4)$$

This is called the *empirical eigenvalue problem* and the optimal basis is given by the eigenfunctions $\{\varphi_i(\xi)\}_{i=1}^N$ of the integral equation. The integral equation (4) can be denoted simply by

$$\mathfrak{R}_N \varphi_i = \lambda_i \varphi_i. \quad (5)$$

B. Numerical solution

Given the spatial grid division $\xi_1 < \xi_2 < \dots < \xi_M$, the snapshots $\{x_n(\xi)\}$ are replaced by M -dimensional vectors, $\{\mathbf{x}_n\}$, where $\mathbf{x}_n = [x_n(\xi_1), \dots, x_n(\xi_M)]^T$. The eigenvector φ_i is replaced by an M -dimensional vector \mathbf{v}_i . Then, the empirical eigenvalue problem (4) can be discretized as

$$\mathcal{R}_N \begin{bmatrix} v_{i_1} \\ \vdots \\ v_{i_M} \end{bmatrix} = \lambda_i \begin{bmatrix} v_{i_1} \\ \vdots \\ v_{i_M} \end{bmatrix} \quad (6)$$

where

$$\mathcal{R}_N = \begin{bmatrix} \langle x_n(\xi_1)x_n(\xi_1) \rangle_{n=1}^N & \dots & \langle x_n(\xi_1)x_n(\xi_M) \rangle_{n=1}^N \\ \vdots & \ddots & \vdots \\ \langle x_n(\xi_M)x_n(\xi_1) \rangle_{n=1}^N & \dots & \langle x_n(\xi_M)x_n(\xi_M) \rangle_{n=1}^N \end{bmatrix}.$$

C. Method of snapshots

Usually, the scale of the empirical eigenvalue problem (4) is huge for high dimensional systems. Sirovich proposed a numerical procedure that can save time in solving the eigenvalue problem governed by the integral equation (4). By assuming that the eigenfunctions $\{\varphi_i(\xi)\}_{i=1}^N$ can be expressed by linear combinations of the snapshots,

$$\varphi_i(\xi) = \sum_{n=1}^N \sigma_{n,i} x_n(\xi), \quad (7)$$

where the snapshots $x_n(\xi)$ are assumed to be linear independent. By substituting (7) into the integral equation (4) we can obtain

$$\int_{\Omega} \frac{1}{N} \sum_{n=1}^N x_n(\xi)x_n(\xi') \sum_{k=1}^N \sigma_{k,i} x_k(\xi') d\xi' = \lambda_i \sum_{n=1}^N \sigma_{n,i} x_n(\xi),$$

which can be rewritten as

$$\frac{1}{N} \sum_{k,n=1}^N \int_{\Omega} x_k(\xi') x_n(\xi') d\xi' x_n(\xi) \sigma_{n,i} = \lambda_i \sum_{n=1}^N \sigma_{n,i} x_n(\xi). \quad (8)$$

Therefore, based on (8), we can formulate a matrix eigenvalue problem $\sum_{k=1}^N \frac{1}{N} \int_{\Omega} x_k(\xi') x_n(\xi') d\xi' \sigma_{n,i} = \lambda_i \sigma_{n,i}$.

III. PERTURBATION OF EIGENVALUE PROBLEMS

A. Motivation of perturbations

Assuming that N_0 measurements are available to compute the autocorrelation operator $\overline{\mathfrak{R}}_{N_0}$ and its discrete representation $\overline{\mathcal{R}}_{N_0}$, and another N_t new measurements are collected, then the empirical eigenvalue problem for the updated data ensemble can be written as

$$\int_{\Omega} \langle x_n(\xi)x_n(\xi') \rangle_{n=1}^{N_0+N_t} \varphi_i(\xi') d\xi' = \lambda_i \varphi_i(\xi). \quad (9)$$

We note that the autocorrelation kernel can be decomposed as

$$\begin{aligned} \langle x_n(\xi)x_n(\xi') \rangle_{n=1}^{N_0+N_t} &= \frac{1}{N_0+N_t} \sum_{n=1}^{N_0+N_t} x_n(\xi)x_n(\xi') \quad (10) \\ &= \frac{N_0}{N_0+N_t} \langle x_n(\xi)x_n(\xi') \rangle_{n=1}^{N_0} + \frac{N_t}{N_0+N_t} \frac{\sum_{n=N_0+1}^{N_0+N_t} x_n(\xi)x_n(\xi')}{N_t}. \end{aligned}$$

When $N_0 \gg N_t$, it is possible to consider the second term of the third line in (10) as a small perturbation of the first term. We define the first term of the second line in (10) as $\bar{\mathcal{R}}_{N_0}(i, j) \approx \langle x_n(\xi_i)x_n(\xi_j) \rangle_{n=1}^{N_0}$ ($N_0 \gg N_t$). Thus, the empirical eigenvalue problem (9) can be considered as a perturbed problem of (4), which can be solved by the method of snapshots. In the rest of this section, we discuss the perturbation problems for both integral and matrix equations.

B. Perturbed matrix equation

We assume that the eigen-pair $(\mathbf{v}_{0i}, \bar{\lambda}_{0i})$ solves the following eigenvalue problem:

$$\bar{\mathcal{R}}_{N_0} \mathbf{v}_{0i} = \bar{\lambda}_{0i} \mathbf{v}_{0i}, \quad i = 1, 2, \dots, M, \quad (11)$$

where the eigenvectors satisfy $\mathbf{v}_{0i}^T \mathbf{v}_{0j} = \delta_i^j$ (δ_i^j is one if $i = j$ and zero otherwise). The perturbed eigenvalue problem is

$$(\bar{\mathcal{R}}_{N_0} + \tilde{\mathcal{R}})(\mathbf{v}_{0i} + \tilde{\mathbf{v}}_{0i}) = (\bar{\lambda}_{0i} + \tilde{\lambda}_{0i})(\mathbf{v}_{0i} + \tilde{\mathbf{v}}_{0i}). \quad (12)$$

By taking into account (11) and neglecting higher order terms, we can rewrite (12) as

$$\bar{\mathcal{R}}_{N_0} \tilde{\mathbf{v}}_{0i} + \tilde{\mathcal{R}} \mathbf{v}_{0i} = \bar{\lambda}_{0i} \tilde{\mathbf{v}}_{0i} + \tilde{\lambda}_{0i} \mathbf{v}_{0i}. \quad (13)$$

We assume that the perturbed vector $\tilde{\mathbf{v}}_{0i}$ can be expressed as a linear combination of unperturbed eigenvectors,

$$\tilde{\mathbf{v}}_{0i} = \sum_{j=1, j \neq i}^M c_{ij} \mathbf{v}_{0j}, \quad (14)$$

where the constants c_{ij} are to be determined. We substitute (14) into the perturbed equation (13) and note (11) to obtain

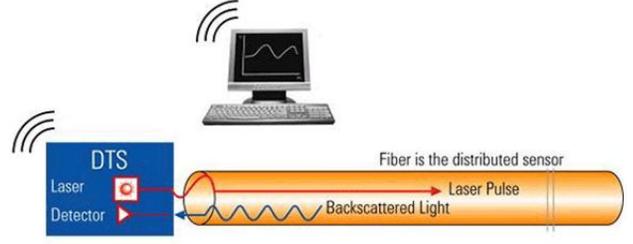
$$\sum_{j=1}^M c_{ij} \bar{\lambda}_{0j} \mathbf{v}_{0j} + \tilde{\mathcal{R}} \mathbf{v}_{0i} = \sum_{j=1}^M c_{ij} \bar{\lambda}_{0i} \mathbf{v}_{0j} + \tilde{\lambda}_{0i} \mathbf{v}_{0i}. \quad (15)$$

Left multiplying (15) with $(\mathbf{v}_{0i})^T$ and $(\mathbf{v}_{0j})^T$, we obtain

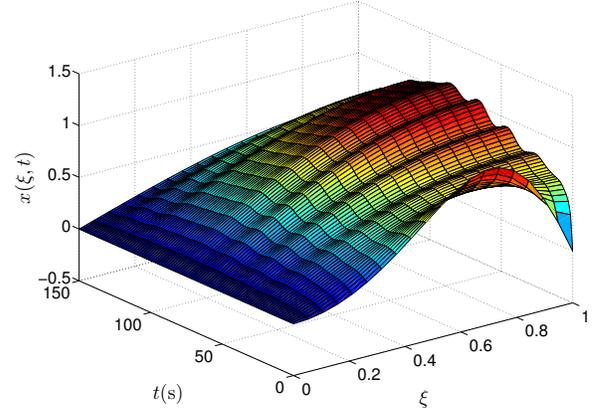
$$\tilde{\lambda}_{0i} = (\mathbf{v}_{0i})^T \tilde{\mathcal{R}} \mathbf{v}_{0i}, \quad c_{ij} = \frac{(\mathbf{v}_{0j})^T \tilde{\mathcal{R}} \mathbf{v}_{0i}}{\bar{\lambda}_{0i} - \bar{\lambda}_{0j}}, \quad i \neq j. \quad (16)$$

Remark 1: When $\bar{\lambda}_{0i} - \bar{\lambda}_{0j}$ is very small, we either set an upper bound for $|c_{ij}|$ or use a higher order truncation for the perturbation problem (12). We follow the same rule in the rest of this paper to compute the coefficient c_{ij} .

Remark 2: We can appreciate that it requires huge amounts memory to compute the correlation matrix $\tilde{\mathcal{R}}$ in (16) for data over very dense spatial grids. By noting the definition of $\tilde{\mathcal{R}}$ below equation (6), we are able to rewrite it as $\tilde{\mathcal{R}} = \frac{1}{N_0+N_t} \tilde{\mathcal{X}} \tilde{\mathcal{X}}^T$, where $\tilde{\mathcal{X}} = [\mathbf{x}_{N_0+1}, \dots, \mathbf{x}_{N_0+N_t}]$. By rewriting $\tilde{\mathcal{R}}$, we compute the updates (16) based on $\tilde{\lambda}_{0i} = \frac{\mathbf{v}_{0i}^T \tilde{\mathcal{X}} \tilde{\mathcal{X}}^T \mathbf{v}_{0i}}{N_0+N_t}$ and $c_{ij} = \frac{(\mathbf{v}_{0j})^T \tilde{\mathcal{X}} \tilde{\mathcal{X}}^T \mathbf{v}_{0i}}{(N_0+N_t)(\bar{\lambda}_{0i} - \bar{\lambda}_{0j})}$, $i \neq j$. Then, we can avoid the memory issue in storing the matrix $\tilde{\mathcal{R}}$.



(a)



(b)

Fig. 1. (a) One dimensional distributed temperature sensing. (b) Simulated data of 1D heat conduction over $0 \leq t \leq 150$ (s).

IV. RECURSIVE POD METHODS

A. Posterior error control

The recursion based on the perturbation theory does not have an error control criterion. We are motivated by [10] to consider a posterior error estimation associated with the recursive POD method. After each recursion, we obtain the modified POD modes which can be used to check the approximation error, i.e.,

$$e(\mathbf{v}_k) = \sum_{n=N_0+1}^{N_0+N_t} \left\| \mathbf{x}_n - \sum_{k=1}^l (\mathbf{x}_n^T \mathbf{v}_k) \mathbf{v}_k \right\|^2. \quad (17)$$

If the error index $e(\mathbf{v}_k)$ is higher than expected, it is possible to modify the modes \mathbf{v}_k , $k = 1, 2, \dots, l$, based on the following gradient search scheme

$$\mathbf{v}_k^{(s+1)} = \mathbf{v}_k^{(s)} - \varepsilon \left[\frac{\partial e(\mathbf{v}_k)}{\partial \mathbf{v}_k} \right]_{\mathbf{v}_k^{(s)}}^T, \quad (18)$$

where ε is the search parameter and

$$\frac{\partial e(\mathbf{v}_k)}{\partial \mathbf{v}_k} = -4 \sum_{n=N_0+1}^{N_0+N_t} \mathbf{x}_n^T \mathbf{v}_k \left(\mathbf{x}_n - \sum_{k=1}^l (\mathbf{x}_n^T \mathbf{v}_k) \mathbf{v}_k \right)^T. \quad (19)$$

In the computations of gradient (19), we have used the following definitions and properties: i- Given a scalar function $f: \mathbb{R}^l \rightarrow \mathbb{R}$, then $\frac{\partial f}{\partial \mathbf{x}} = \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right)$; ii- Given any vector $\mathbf{a} \in \mathbb{R}^l$ and any matrix $\mathbf{A} \in \mathbb{R}^{l \times l}$, then $\frac{\partial \mathbf{x}^T \mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{x}^T$, $\frac{\partial \mathbf{a}^T \mathbf{x}}{\partial \mathbf{x}} = \frac{\partial \mathbf{x}^T \mathbf{a}}{\partial \mathbf{x}} = \mathbf{a}^T$ and $\frac{\partial \mathbf{A}^T \mathbf{x}}{\partial \mathbf{x}} = \mathbf{A}$.

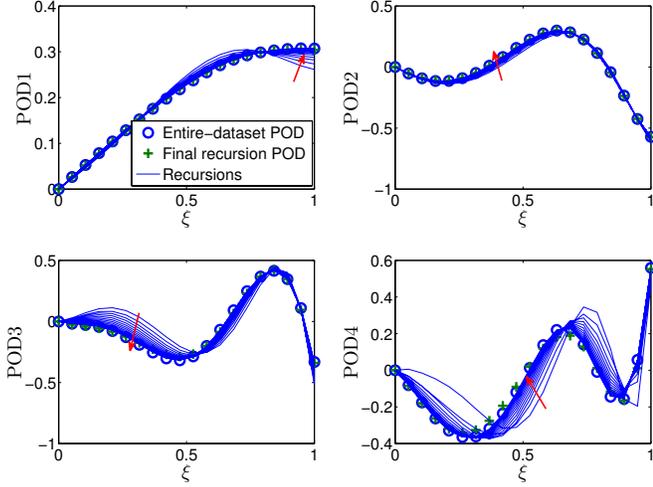


Fig. 2. Comparison of POD modes based on different methods.

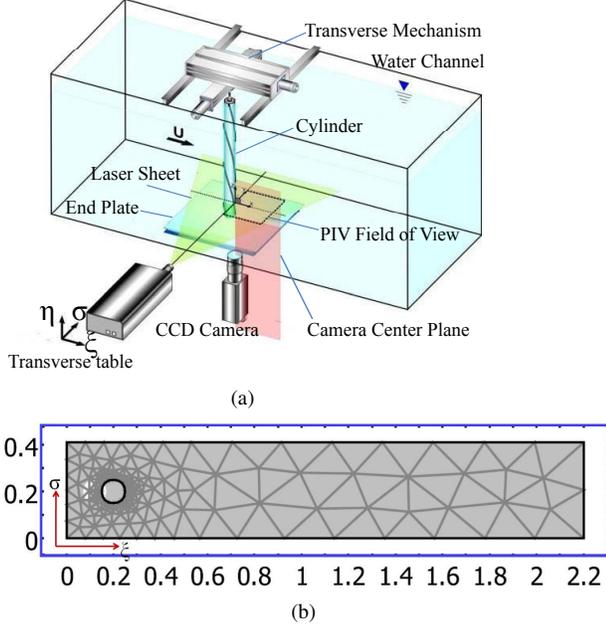


Fig. 3. (a) PIV scheme of free-surface wave interaction with a horizontal cylinder, a courtesy of the Fluids Research Laboratory at Lehigh University. (b) Two dimensional approximation: fluid flow past a cylinder. Channel dimensions: $\Omega = [0, 2.2] \times [0, 0.41]$. Cylinder dimension: radius 0.05m. Coordinates of the cylinder center: (0.2, 0.205).

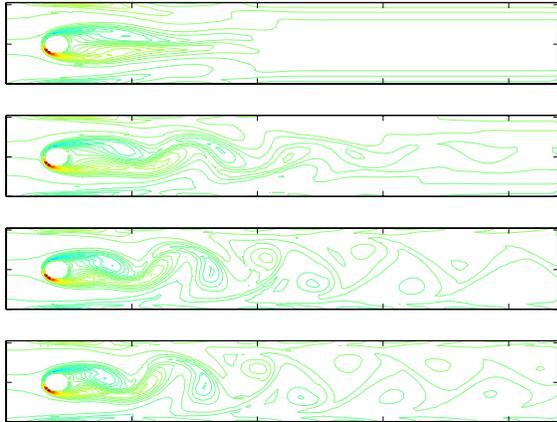


Fig. 4. Snapshots of the vorticity at time $t = 1s, 2s, 3s$ and $4s$, respectively.

B. Recursive algorithm

We summarize below the algorithm for recursive POD computation under receding horizon measurements.

- 1) Compute the autocorrelation matrix $\bar{\mathcal{R}}_{N_0}$,

$$\bar{\mathcal{R}}_{N_0}(i, j) = \langle x_n(\xi_i)x_n(\xi_j) \rangle_{n=1}^{N_0}, \quad (20)$$

and the eigenvalues and eigenvectors of the matrix $\bar{\mathcal{R}}_{N_0}$ denoted by $(\bar{\lambda}_{0i}, \mathbf{v}_{0i})$, $i = 1, 2, \dots, M$;

- 2) Compute the perturbed autocorrelation matrix $\tilde{\mathcal{R}}$,

$$\tilde{\mathcal{R}}(i, j) = \frac{1}{N_0 + N_t} \sum_{n=N_0+1}^{N_0+N_t} x_n(\xi_i)x_n(\xi_j); \quad (21)$$

- 3) Compute the perturbations of both the eigenvalues and eigenvectors

$$\tilde{\lambda}_{0i} = (\mathbf{v}_{0i})^T \tilde{\mathcal{R}} \mathbf{v}_{0i}, \quad \tilde{\mathbf{v}}_{0i} = \sum_{j=1}^M c_{ij} \mathbf{v}_{0j}, \quad (22)$$

where $c_{ij} = \frac{(\mathbf{v}_{0j})^T \tilde{\mathcal{R}} \mathbf{v}_{0i}}{\lambda_{0i} - \lambda_{0j}}$, $i \neq j$. Then, we obtain the

following recursion $\mathbf{v}_{0i} = \mathbf{v}_{0i} + \sum_{j=1}^M c_{ij} \tilde{\mathbf{v}}_{0j}$, $\bar{\lambda}_{0i} = \bar{\lambda}_{0i} + (\mathbf{v}_{0i})^T \tilde{\mathcal{R}} \mathbf{v}_{0i}$, and $N_0 = N_0 + N_t$;

- 4) If $e(\mathbf{v}_k)$ (defined in (17)) is smaller than error tolerance, we accept the recursion result. Otherwise, we use the gradient searching law (18) to adjust the POD modes;
- 5) Go back to Step 2).

V. STATE-OF-THE-ART EXAMPLES

A. Distributed temperature sensing

We consider the distributed temperature sensing of a one dimensional heat conduction problem shown in Fig. 1(a), where the mathematical model of the heat conduction over the domain $0 \leq \xi \leq 1$ (normalized representation) is governed by the following equations:

$$\begin{cases} \frac{\partial x(\xi, t)}{\partial t} = \frac{\partial}{\partial \xi} \left[D(\xi, t) \frac{\partial x(\xi, t)}{\partial \xi} \right] + \varepsilon(t)V(\xi, t)x(\xi, t), \\ x(0, t) = \frac{\partial x(1, t)}{\partial \xi} = 0, \quad x(\xi, 0) = x_0(\xi), \end{cases}$$

where $x(\xi, t)$ represents the spatial (ξ) temporal (t) temperature distribution, $\varepsilon(t)$ is a time-varying coefficient, $D(\xi)$ and $V(\xi)$ are appropriate spatial functions. Instead of using real experimental data, we employ the finite element method to simulate the mathematical model and generate data for the validation of the rPOD algorithm. We use the following parameters: $D(\xi) = \frac{1}{50\pi^2} (1 + \frac{1}{5} \sin t)$, $\varepsilon(t) = \frac{1}{5\pi^2} \sin(\frac{\pi}{100}t)$, $V(\xi, t) = \frac{3}{5} + \cos(\frac{\pi}{10}t) \cos(\frac{\pi}{10}\xi)$, $x_0(\xi) = \sin(\pi\xi^2)$.

The temporal-spatial simulation data is shown in Fig. 1(b). We start by using the first $N_0 = 25$ data measurements to generate dominant POD modes. In each iteration, $N_t = 5$ new data measurements are incorporated into the data set and the POD modes are updated based on the perturbation theory. We compare in Fig. 2 the first four dominant POD modes obtained either directly from the entire data set of 100 measurements (denoted by blue circles) or from the

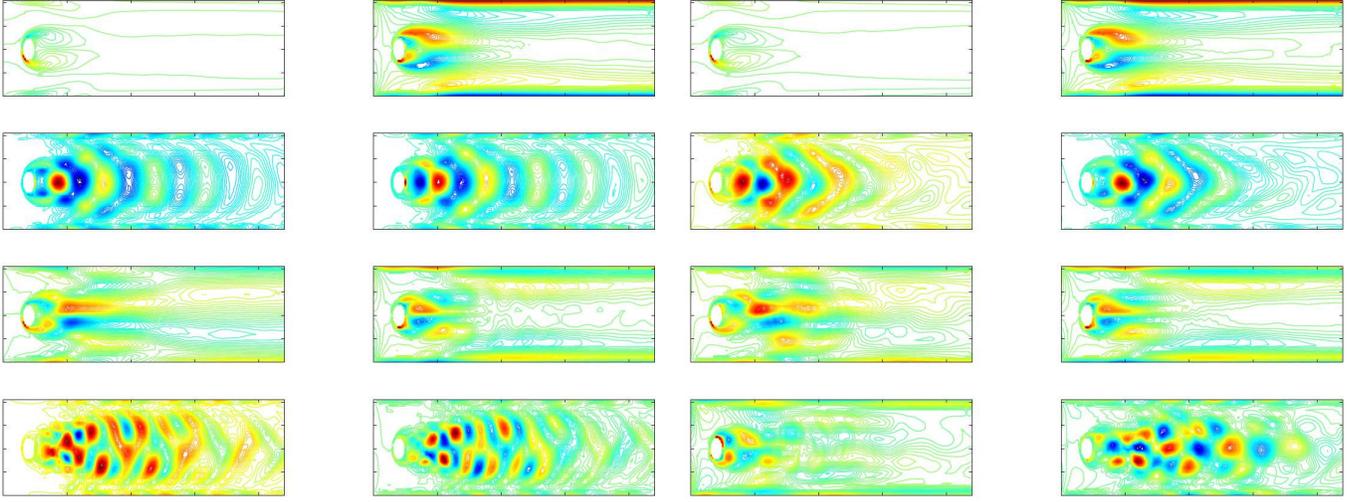


Fig. 5. The first eight POD modes of the snapshots of the vorticity field based on the entire data (from left to right, from top to bottom).

Fig. 7. Truncating-perturbation-based POD modes of the snapshots of the vorticity field (from left to right, from top to bottom).

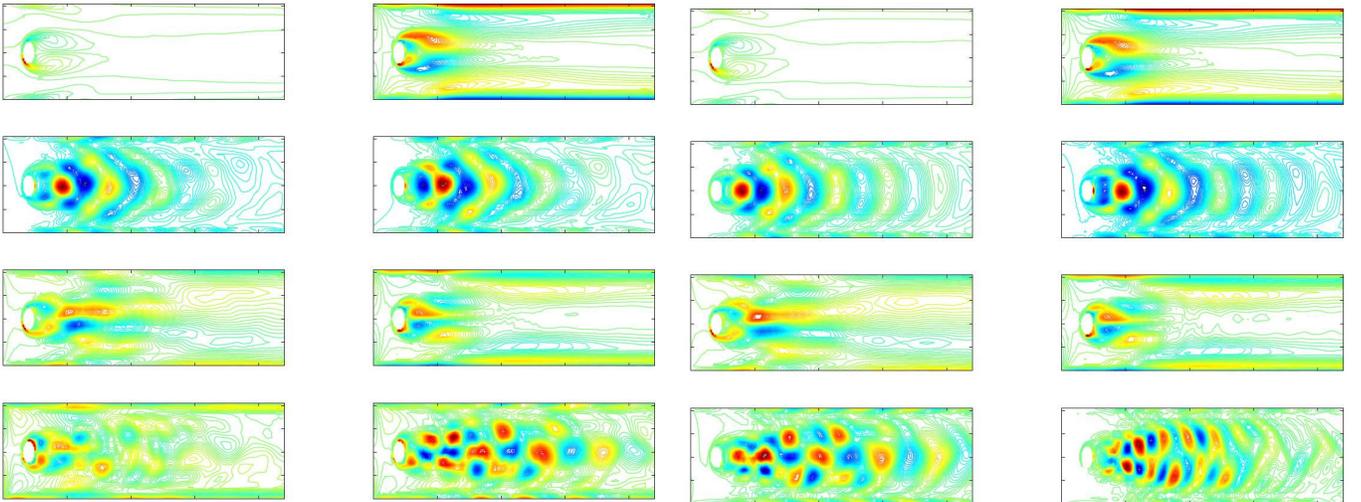


Fig. 6. The first eight POD modes of the snapshots of the vorticity field based on the historic data $N_0 = 60$ (from left to right, from top to bottom).

Fig. 8. POD modes with gradient adjust of the snapshots of the vorticity field (from left to right, from top to bottom).

recursions (15 recursions in total). It is shown in the figure that every POD mode evolves over time except for the second one. Thus, it is clearly necessary to update the POD modes with newly measured data. The red arrows in the figure show the iteration directions. The outmost curves represent the POD modes obtained recursively based on 100 measurements (denoted by green crosses).

B. Two dimensional fluid flow past a cylinder

The schematic of a particle image velocimetry (PIV) setup is shown in Fig. 3 (a). The setup is used to study flow patterns behind a vertical cylinder with helical windings. The flow is created in a transparent water channel, and the water is seeded with light-reflecting particles. The images of the particles, which are illuminated by a laser, are captured by a high-resolution camera. These images are then processed on a computer to yield the global instantaneous flow velocity

measurements. Instead of dealing with experimental data, in this paper we use a two dimensional model approximation of the flow to generate the simulation data employed to validate the proposed rPOD algorithm. A two dimensional fluid flow past a cylinder is shown in Fig. 3 (b), which is extracted from the PIV view plane shown in Fig. 3 (a). An adaptive grid for finite element method (FEM) computation is included in Fig. 3 (b). The mathematical model is governed by the viscous incompressible Navier-Stokes equation:

$$\frac{\partial x^u}{\partial t} + x^u \frac{\partial x^u}{\partial \xi} + x^v \frac{\partial x^u}{\partial \sigma} + \frac{\partial p}{\partial \xi} = \frac{1}{\text{Re}} \Delta x^u, \quad \text{in } Q, \quad (23)$$

$$\frac{\partial x^v}{\partial t} + x^u \frac{\partial x^v}{\partial \xi} + x^v \frac{\partial x^v}{\partial \sigma} + \frac{\partial p}{\partial \sigma} = \frac{1}{\text{Re}} \Delta x^v, \quad \text{in } Q, \quad (24)$$

$$\frac{\partial x^u}{\partial \xi} + \frac{\partial x^v}{\partial \sigma} = 0, \quad \text{in } Q, \quad (25)$$

where the temporal-spatial domain is defined by $Q = (0, T) \times \Omega$ and $\Omega = \{(\xi, \sigma) | \xi \in [0, 2.2], \sigma \in [0, 0.41]\}$, $x(\xi, \sigma, t) = \{x^u(\xi, \sigma, t), x^v(\xi, \sigma, t)\}$ is the two component fluid velocity vector, $p = p(\xi, \sigma, t)$ is the pressure field and t is the time. The Reynolds number Re is defined by $Re = \frac{UD}{\nu}$, where U is the upstream velocity, D is the diameter of the cylinder and ν is the kinematic viscosity of the Newtonian fluid flow. The boundary conditions at the upper and lower walls, and around the cylinder are assumed to be nonslip. The inlet flow has a velocity profile $(x^u, x^v) = (U, 0)$. With the normalized parameters given as $D = 0.1$, $U = 1.5$ and $\nu = 5 \times 10^{-3}$, we simulate the fluid flow dynamics and the snapshots of the vorticity field $\omega = \frac{\partial x^v}{\partial \xi} - \frac{\partial x^u}{\partial \sigma}$ in Ω at different times are shown in Fig. 4.

We have $N_s = 80$ snapshots in total. To represent each snapshot of the vorticity field, we use a regular Euclidean coordinate with a size of $N_\xi = 441, N_\sigma = 83$. Then, each snapshot is represented by an $N_\xi \times N_\sigma$ matrix. In order to use the POD theory for 1D data, we rearrange each matrix into a data array of length $N_\xi N_\sigma$ following the rule of one line after another. The first eight POD modes are shown in Fig. 5. We use $N_0 = 60$ snapshots as the historic data set to extract initial dominant POD modes (shown in Fig. 6). We note that the POD modes of order higher than six (seventh and eighth are shown in the figures) are much different from those in Fig. 5. This indicates that the POD modes evolve over time. Then, for each POD update, we collect five ($N_t = 5$) snapshots to enrich the historic data set. After four recursions, the POD modes are shown in Fig. 7. By checking the posterior error defined by (17), we note that the error accumulation is not negligible. Then, we follow the update law (19) to adjust the POD modes obtained from the perturbation theory approach, where the search parameter is set to be $\varepsilon = 0.01$. After four iterations, modifications are incorporated into the POD modes shown in Fig. 8. In order to facilitate the comparison of the POD modes obtained from the entire-data set ($N_s = 80$), from the historic-data ensemble ($N_0 = 60$) and from the recursive algorithm based on 80 measurements, we compare the first four POD modes in Fig. 9. We can see that the iterations do not perfectly converge to the POD modes based on the entire-data set, but show a good compromise between historic-data and entire-data sets.

VI. CONCLUSIONS

We study in this paper distributed sensing and real time pattern extraction arising in cyber-physical systems. Classical POD methods have the capability of extracting dominant features of collected data. However, in real time measurements, huge amounts of data are collected and updates of the POD modes are required. With the assumption of small perturbations, we are able to reformulate the POD problem at each update time into an operator/matrix perturbation problem. The first order truncation is used to write the update law for the eigenvalues and eigenvectors. The update law provides modifications for the POD modes by only requiring newly collected data and previously saved POD solutions.

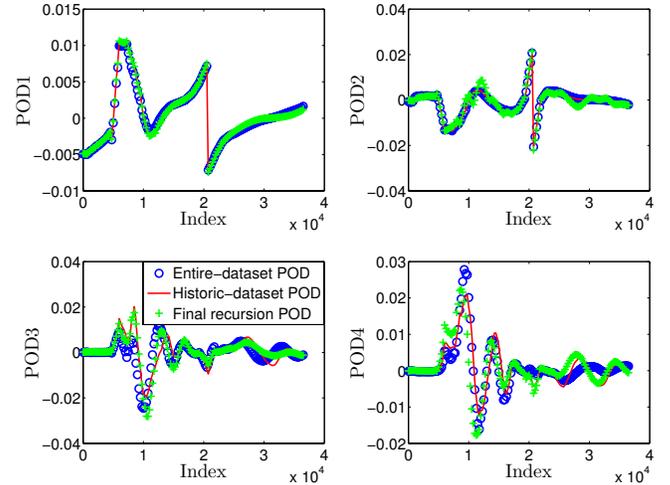


Fig. 9. Comparison of different POD modes (vector form).

The posterior error estimation is used to control consistent error accumulation due to the perturbation approximation. A modification law based on the gradient search technique is used to adjust the perturbation truncation results. This method gives accurate POD mode extraction in real time with relatively low computational cost. It has potential to be applied in several fields, such as receding horizon control of distributed parameter systems (e.g., flow past a cylinder), and dynamic/moving sensing in monitoring and information acquisition (e.g., water quality monitoring with moving sensors).

REFERENCES

- [1] E. Lee, "Cyber physical systems: Design challenges," *International Symposium on Object/Component/Service-Oriented Real-Time Distributed Computing (ISORC)*, 2008.
- [2] T. Kurashima, T. Horiguchi, and M. Tateda, "Distributed-temperature sensing using stimulated Brillouin scattering in optical silica fibers," *Optics Letters*, vol. 15, pp. 1038–1040, 1990.
- [3] M. Raffel, C. Willert, and J. Kompenhans, *Particle Image Velocimetry: A Practical Guide (Experimental Fluid Mechanics)*. New York: Springer, 1998.
- [4] I. Hutchinson, *Principles of Plasma Diagnostics (Second Edition)*. New York: Springer, 2002.
- [5] P. Holmes, J. Lumley, and G. Berkooz, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*. New York: Cambridge University Press, 1996.
- [6] L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human faces," *Journal of the Optical Society of America A: Optics, Image Science, and Vision*, vol. 4, pp. 519–524, 1987.
- [7] L. Sirovich, "Turbulence and the dynamics of coherent structures: part I, II, III," *Quarterly of Applied Mathematics*, vol. 45, pp. 573–590, 1987.
- [8] F. Rellich, *Perturbation Theory of Eigenvalue Problems*. New York: Gordon and Breach, 1969.
- [9] B. Champagne, "Adaptive eigendecomposition of data covariance matrices based on first order perturbations," *IEEE Transactions on Signal Processing*, vol. 42, pp. 2758–2770, 1994.
- [10] D. Sahoo, S. Park, D. Wee, A. Annaswamy, and A. Ghoniem, "A recursive proper orthogonal decomposition algorithm for flow control problems," *Adaptive Control Laboratory, MIT. Technical Report 0208*, 2002.
- [11] A. Varshney, S. Pitschiah, and A. Armaou, "Feedback control of dissipative PDE systems using adaptive model reduction," *AICHE Journal*, vol. 55, pp. 906–918, 2009.