2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Probabilistic Reduced-Dimensional Vector Autoregressive Modeling for Dynamics Prediction and Reconstruction with Oblique Projections

Yanfang Mo, *Member, IEEE*, Jiaxin Yu, *Student Member, IEEE*, and S. Joe Qin, *Fellow, IEEE*

***Abstract*— In this paper, we propose a probabilistic reduced-dimensional vector autoregressive (PredVAR) model with oblique projections. This model partitions the measurement space into a dynamic subspace and a static subspace that do not need to be orthogonal. The partition allows us to apply an oblique projection to extract dynamic latent variables (DLVs) from high-dimensional data with maximized predictability. We develop an alternating iterative PredVAR algorithm that exploits the interaction between updating the latent VAR dynamics and estimating the oblique projection, using expectation maximization (EM) and a statistical constraint. In addition, the noise covariance matrices are estimated as a natural outcome of the EM method. A simulation case study of the nonlinear Lorenz oscillation system illustrates the advantages of the proposed approach over two alternatives.**

I. Introduction

Extracting reduced-dimensional dynamics is crucial in many industries, including chemicals, power systems, finance, and transportation, where operational data are usually high-dimensional with dynamic features [1]–[5]. Classic data analytic tools like principal component analysis (PCA) and canonical correlation analysis (CCA) are oblivious to the system dynamics [3] and should be extended. Dynamic PCA (DPCA) [6] and subspace-based models [7] can consider the auto-correlations in time series. A new dynamic latent variable (DLV) model was developed in [8] to characterize the dynamic relations in DLVs and the static cross-correlations in residuals. Moreover, dynamic-inner PCA (DiPCA) [9] and dynamic-inner CCA (DiCCA) [10] algorithms produce rank-ordered DLVs by maximizing prediction power. Furthermore, Qin developed a latent vector autoregressive modeling algorithm with a CCA objective (LaVAR-CCA) [11] and a vector autoregressive model [12] for the DLVs, where a state-space generalization is subsequently developed in [13].

Uncertainty estimation requires a statistical viewpoint for dimension reduction and dynamics extraction. The LaVAR-CCA [12] has a statistical interpretation of profile likelihood. Many dynamic factor models (DFMs) use time-related statistics to estimate model parameters [2], [4], [14]–[16]. However, more attention should be paid to dynamics modeling in the statistical literature [12], [17], [18]. More importantly, the structured dynamics-noise relationship can facilitate signal reconstruction and prediction [3], [4].

Yanfang Mo is with the Hong Kong Institute for Data Science, City University of Hong Kong, Hong Kong Yanfang.MO@cityu.edu.hk

Jiaxin Yu is with the School of Data Science and Hong Kong Institute for Data Science, City University of Hong Kong, Hong Kong.

S. Joe Qin is with the Department of Computing and Decision Science and Institute of Data Science, Lingnan University, Hong Kong. Corresponding author. joeqin@LN.edu.hk. He acknowledges the financial support from a Natural Science Foundation of China Project (U20A20189), a General Research Fund by RGC of Hong Kong (No. 11303421), a Collaborative Research Fund by RGC of Hong Kong (Project No. C1143-20G), an ITF - Guangdong-Hong Kong Technology Cooperation Funding Scheme (Project Ref. No. GHP/145/20), an InnoHK initiative of The Government of the HKSAR for the Laboratory for AI-Powered Financial Technologies, and a Shenzhen-Hong Kong-Macau Science and Technology Project Category C (9240086).

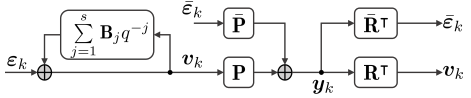

Fig. 1. A block diagram of a PredVAR model.

Our probabilistic model partitions the measurement space into a signal subspace admitting the low-dimensional DLV dynamics and a static noise subspace, which do not need to be orthogonal to each other. Thus, an oblique-projection perspective is adopted for the signal-noise structure. This perspective is well-known in the signal-processing literature [19], where model parameters are deemed known, which is not the case for this work. A challenge lies in identifying a proper oblique projection to extract DLVs and degrade the effect of noise on estimating the system dynamics.

Also, the estimated DLV dynamics can facilitate oblique projection identification. Thus, an alternating iterative scheme is developed for the latent dynamics and oblique projection estimations by exploring their interactions. Specifically, two expectation-maximization (EM) steps are used to estimate the DLV dynamics and the signal subspace. Meanwhile, a statistical constraint is uniquely imposed to characterize the relationship between the estimated signal and static noise and is thus used to identify the noise subspace. The contributions in this work are as follows.

1) A probabilistic reduced-dimensional vector regressive (PredVAR) model with oblique projections is proposed.

2) An iterative algorithm is developed to update the DLV dynamics and oblique projection estimations alternately. It uniquely uses a statistical constraint together with an EM procedure in the oblique projection identification.

3) A simulated case study is conducted to demonstrate the strength of our approach compared to two benchmarks, a one-shot algorithm that first identifies the oblique projection and then estimates the DLV dynamics and a counterpart that focuses on the orthogonal projection.

II. Model Formulation

Denote the measurement time series from a dynamic system by $\{\boldsymbol{y}_k \in \Re^p\}_{k=1}^{N+s}$. It is typical for operational data that the dynamics are excited in a subspace of dimension $\ell <$

p, referred to as reduced-dimensional dynamics [12]. Thus, a dynamic latent process $\{\boldsymbol{v}_k \in \Re^{\ell}\}$ exists, together with a loadings matrix $\mathbf{P} \in \Re^{p\times\ell}$ that is of full column rank.

The reduced-dimensional dynamics and the signal-noise structure motivate us to construct the model in [12] in the PredVAR model as

$$\boldsymbol{y}_k = \mathbf{P}\boldsymbol{v}_k + \bar{\mathbf{P}}\bar{\boldsymbol{\varepsilon}}_k, \qquad \bar{\boldsymbol{\varepsilon}}_k \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\bar{\boldsymbol{\varepsilon}}}), \quad (1)$$

$$\boldsymbol{v}_k = \sum_{j=1}^{s} \mathbf{B}_j \boldsymbol{v}_{k-j} + \boldsymbol{\varepsilon}_k, \qquad \boldsymbol{\varepsilon}_k \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}). \quad (2)$$

where $[\mathbf{P} \quad \bar{\mathbf{P}}]$ is required to be nonsingular, the innovations vector $\boldsymbol{\varepsilon}_k \in \Re^{\ell}$ and the static noise $\bar{\boldsymbol{\varepsilon}}_k \in \Re^{p-\ell}$ are assumed to be serially and mutually independent ($E\{\boldsymbol{\varepsilon}_k \bar{\boldsymbol{\varepsilon}}_k^{\intercal}\} = \mathbf{0}$). An s-order VAR model is used to describe the DLV dynamics. Equations (2) and (1) are called inner and outer models. By (2), the one-step-ahead prediction of $\boldsymbol{v}_k$ is defined as

$$\tilde{\boldsymbol{v}}_k = E\{\boldsymbol{v}_k \mid \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{k-1}\} = \sum_{j=1}^{s} \mathbf{B}_j \boldsymbol{v}_{k-j}. \quad (3)$$

Associated with $\mathbf{P}$ and $\bar{\mathbf{P}}$, the DLV and static weight matrices $\mathbf{R} \in \Re^{p\times\ell}$ and $\bar{\mathbf{R}} \in \Re^{p\times(p-\ell)}$ satisfy

$$\begin{bmatrix}\mathbf{R} & \bar{\mathbf{R}}\end{bmatrix}^{\intercal} \begin{bmatrix}\mathbf{P} & \bar{\mathbf{P}}\end{bmatrix} = \mathbf{I}. \quad (4)$$

Therefore, pre-multiplying $\mathbf{R}^{\intercal}$ to (1) and $\mathbf{P}$ to (5) lead to

$$\boldsymbol{v}_k = \mathbf{R}^{\intercal}\boldsymbol{y}_k \text{ and } \mathbf{P}\boldsymbol{v}_k = \mathbf{P}\mathbf{R}^{\intercal}\boldsymbol{y}_k, \quad (5)$$

giving an oblique projection of $\boldsymbol{y}_k$ since $(\mathbf{P}\mathbf{R}^{\intercal})^2 = \mathbf{P}\mathbf{R}^{\intercal}$.

The PredVAR model is illustrated in Fig. 1. The VAR model for DLVs describes the low-dimensional dynamics and initiates exploring more complicated dynamics descriptions (see [13], [20]–[22]). By (1), (2), and (5), a VAR model with reduced-rank coefficients $\{\mathbf{P}\mathbf{B}_j\mathbf{R}^{\intercal}, j \in [\![s]\!] \triangleq \{1, 2, \ldots, s\}\}$ is obtained

$$\boldsymbol{y}_k = \sum_{j=1}^{s} \mathbf{P}\mathbf{B}_j\mathbf{R}^{\intercal}\boldsymbol{y}_{k-j} + \boldsymbol{e}_k, \text{ where} \quad (6)$$

$$\boldsymbol{e}_k = \mathbf{P}\boldsymbol{\varepsilon}_k + \bar{\mathbf{P}}\bar{\boldsymbol{\varepsilon}}_k \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{e}} = \mathbf{P}\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}\mathbf{P}^{\intercal} + \bar{\mathbf{P}}\boldsymbol{\Sigma}_{\bar{\boldsymbol{\varepsilon}}}\bar{\mathbf{P}}^{\intercal}). \quad (7)$$

By (4), (7), and the assumption $E\{\boldsymbol{\varepsilon}_k \bar{\boldsymbol{\varepsilon}}_k^{\intercal}\} = \mathbf{0}$, the constraint below is derived and imposed in this paper

$$\mathbf{R}^{\intercal}\boldsymbol{\Sigma}_{\boldsymbol{e}}\bar{\mathbf{R}} = \mathbf{0}. \quad (8)$$

The imposed statistical constraint (8) plays a role in estimating the parameter tuple $(\mathbf{P}; \boldsymbol{v}; \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}; \mathbf{B}_j, j \in [\![s]\!]; \bar{\mathbf{P}}; \boldsymbol{\Sigma}_{\bar{\boldsymbol{\varepsilon}}})$, discussed in the next section. Two tuples of PredVAR model parameters are observationally equivalent if they can generate the same measurement data. In this regard, the following theorem deserves attention when estimating a PredVAR model with given measurements [23].

Theorem 1: For arbitrary nonsingular matrices $\mathbf{M} \in \Re^{\ell\times\ell}$ and $\bar{\mathbf{M}} \in \Re^{(p-\ell)\times(p-\ell)}$, the two PredVAR models are observationally equivalent with their respective parameter tuples as $(\mathbf{P}; \boldsymbol{v}; \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}; \mathbf{B}_j, j \in [\![s]\!]; \bar{\mathbf{P}}; \boldsymbol{\Sigma}_{\bar{\boldsymbol{\varepsilon}}})$ and $(\mathbf{P}\mathbf{M}^{-1}; \mathbf{M}\boldsymbol{v}; \mathbf{M}\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}\mathbf{M}^{\intercal}; \mathbf{M}\mathbf{B}_j\mathbf{M}^{-1}, j \in [\![s]\!]; \bar{\mathbf{P}}\bar{\mathbf{M}}^{-1}; \bar{\mathbf{M}}\Sigma_{\bar{\boldsymbol{\varepsilon}}}\bar{\mathbf{M}}^{\intercal})$.

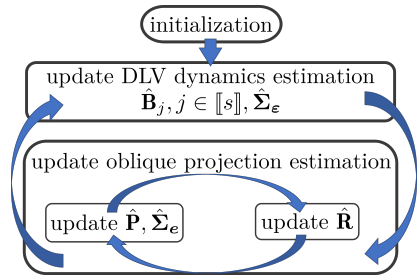

Fig. 2. An illustration of our alternating procedure.

Note that the oblique projections related to $\mathbf{P}$ and $\mathbf{R}$ are the same for observationally equivalent models in Theorem 1 [4]. So are the matrices $\mathbf{P}\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}\mathbf{P}^{\intercal}$, $\bar{\mathbf{P}}\boldsymbol{\Sigma}_{\bar{\boldsymbol{\varepsilon}}}\bar{\mathbf{P}}^{\intercal}$, $\boldsymbol{\Sigma}_{\boldsymbol{e}}$, and $\mathbf{P}\mathbf{B}_j\mathbf{R}^{\intercal}, j \in [\![s]\!]$. Moreover, diverse restrictions can be imposed to improve the identifiability and for numerical concerns, as shown in [23].

III. Model Estimation

To identify the PredVAR model, the DLV dynamics and the oblique projection estimations are alternately conducted via EM and a statistical constraint, as depicted in Fig. 2. For convenience, the symbol $\hat{\cdot}$ is used to distinguish between an estimated parameter and its true value.

A. Identifying DLV Dynamics with Oblique Projections

Given $(\mathbf{P}, \bar{\mathbf{P}})$, to use EM for estimating $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$ and $\mathbf{B}_j, j \in [\![s]\!]$, the following likelihood should be maximized:

$$\prod_{k=s+1}^{s+N} p(\boldsymbol{v}_k \mid \boldsymbol{v}_{k-1}, \boldsymbol{v}_{k-2}, \ldots, \boldsymbol{v}_{k-s}).$$

Since $\boldsymbol{\varepsilon}_k \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})$, the M-step requires minimizing the following function

$$L^{DLV}(\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}, \mathbb{B}) = N \ln|\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}| + \sum_{k=s+1}^{s+N} (\boldsymbol{v}_k - \tilde{\boldsymbol{v}}_k)^{\intercal}\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}(\boldsymbol{v}_k - \tilde{\boldsymbol{v}}_k),$$

where $\tilde{\boldsymbol{v}}_k$ is the one-step-ahead prediction defined in (3). Take the derivatives of $L^{DLV}(\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}, \mathbb{B})$ with respect to $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}$ and $\mathbf{B}_j, j \in [\![s]\!]$ and set them to zeros. It follows that

$$\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} = \frac{1}{N}\sum_{k=s+1}^{s+N} (\boldsymbol{v}_k - \tilde{\boldsymbol{v}}_k)(\boldsymbol{v}_k - \tilde{\boldsymbol{v}}_k)^{\intercal},$$

$$\mathbf{B}_j = \sum_{k=s+1}^{s+N} \left(\boldsymbol{v}_k\boldsymbol{v}_{k-j}^{\intercal} - \sum_{i\in[\![s]\!], i\neq j} \mathbf{B}_i\boldsymbol{v}_{k-i}\boldsymbol{v}_{k-j}^{\intercal}\right) \times \left(\sum_{k=s+1}^{s+N} \boldsymbol{v}_{k-j}\boldsymbol{v}_{k-j}^{\intercal}\right)^{-1}.$$

The last equation is rearranged as

$$\sum_{i\in[\![s]\!]} \mathbf{B}_i \sum_{k=s+1}^{s+N} \boldsymbol{v}_{k-i}\boldsymbol{v}_{k-j}^{\intercal} = \sum_{k=s+1}^{s+N} \boldsymbol{v}_k\boldsymbol{v}_{k-j}^{\intercal}, \quad j \in [\![s]\!].$$

In the E-step, by (5), the following updating formula is obtained with an estimate $\hat{\mathbf{R}}$:

$$\hat{\boldsymbol{v}}_k = \hat{\mathbf{R}}^{\intercal}\boldsymbol{y}_k.$$

Form the following augmented matrices as in [12]:

$$\mathbf{Y}_i = [\boldsymbol{y}_{i+1}\ \boldsymbol{y}_{i+2}\ \cdots\ \boldsymbol{y}_{i+N}]^\intercal, i \in \{0\} \cup [\![s]\!];$$
$$\hat{\mathbf{V}}_i = [\hat{\boldsymbol{v}}_{i+1}\ \hat{\boldsymbol{v}}_{i+2}\ \cdots\ \hat{\boldsymbol{v}}_{i+N}]^\intercal, i \in \{0\} \cup [\![s]\!];$$
$$\hat{\mathbb{V}} = [\hat{\mathbf{V}}_{s-1}\ \hat{\mathbf{V}}_{s-2}\ \cdots\ \hat{\mathbf{V}}_0];$$
$$\hat{\mathbb{B}} = [\hat{\mathbf{B}}_1\ \hat{\mathbf{B}}_2\ \cdots\ \hat{\mathbf{B}}_s]^\intercal.$$

Then, the previous formulas can be rewritten as

$$\hat{\mathbf{V}}_i = \mathbf{Y}_i \hat{\mathbf{R}}, i \in \{0\} \cup [\![s]\!] \tag{9}$$
$$\hat{\mathbb{B}} = (\hat{\mathbb{V}}^\intercal \hat{\mathbb{V}})^{-1} \hat{\mathbb{V}}^\intercal \hat{\mathbf{V}}_s; \tag{10}$$
$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}} = (\hat{\mathbf{V}}_s - \hat{\mathbb{V}}\hat{\mathbb{B}})^\intercal (\hat{\mathbf{V}}_s - \hat{\mathbb{V}}\hat{\mathbb{B}})/N. \tag{11}$$

The updating formula for $\mathbb{B}$ is the same as that in LaVAR-CCA. This is unsurprising because LaVAR-CCA has a likelihood interpretation as stated in [12].

B. Identifying Oblique Projection with DLV Dynamics

Given the DLV dynamics, to use EM for estimating $\hat{\mathbf{P}}$ and $\hat{\mathbf{R}}$, the following likelihood is maximized:

$$\prod_{k=s+1}^{s+N} p(\boldsymbol{y}_k \mid \boldsymbol{v}_{k-1}, \boldsymbol{v}_{k-2}, \ldots, \boldsymbol{v}_{k-s}),$$

while respecting (7) and (8) for $\boldsymbol{e}_k$. This is equivalent to minimizing the function below, subject to (8).

$$L^{proj}(\boldsymbol{\Sigma}_{\boldsymbol{e}}, \mathbf{P}) = N \ln |\boldsymbol{\Sigma}_{\boldsymbol{e}}| + \sum_{k=s+1}^{s+N} (\boldsymbol{y}_k - \mathbf{P}\tilde{\boldsymbol{v}}_k)^\intercal \boldsymbol{\Sigma}_{\boldsymbol{e}}^{-1} (\boldsymbol{y}_k - \mathbf{P}\tilde{\boldsymbol{v}}_k),$$

The imposed constraint (8) allows us to uniquely determine the subspace spanned by $\mathbf{R}$, given $\mathbf{P}$ and $\boldsymbol{\Sigma}_{\boldsymbol{e}}$. Precisely, with singular value decompositions (SVDs), obtain $\bar{\mathbf{R}}$ from an orthonormal basis of the null space of $\mathbf{P}$ and then an orthonormal basis of the null space of $\boldsymbol{\Sigma}_{\boldsymbol{e}}\bar{\mathbf{R}}$ gives $\mathbf{R}$.

Ignoring (8), an EM procedure can be used to estimate $\mathbf{P}$ and $\boldsymbol{\Sigma}_{\boldsymbol{e}}$. Similarly to before, take the derivatives of $L^{proj}(\boldsymbol{\Sigma}_{\boldsymbol{e}}, \mathbf{P})$ with respect to $\mathbf{P}$ and $\boldsymbol{\Sigma}_{\boldsymbol{e}}^{-1}$ and set them to zero. Then, the M-step requires the updating formulas

$$\hat{\mathbf{P}} = \mathbf{Y}_s^\intercal \hat{\mathbb{V}}\hat{\mathbb{B}}(\hat{\mathbb{B}}^\intercal \hat{\mathbb{V}}^\intercal \hat{\mathbb{V}}\hat{\mathbb{B}})^{-1}; \tag{12}$$
$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{e}} = (\mathbf{Y}_s - \hat{\mathbb{V}}\hat{\mathbb{B}}\hat{\mathbf{P}}^\intercal)^\intercal (\mathbf{Y}_s - \hat{\mathbb{V}}\hat{\mathbb{B}}\hat{\mathbf{P}}^\intercal)/N. \tag{13}$$

The E-step requires (9). The above analysis leads to Algorithm 1, where $\hat{\mathbf{P}}$ and $\hat{\mathbf{R}}$ are alternately updated by the EM method and the statistical constraint (8) until convergence. Note that (12) is different from that in LaVAR-CCA [12], which updates $\hat{\mathbf{P}}$ by the formula $\hat{\mathbf{P}} = \mathbf{Y}_s^\intercal \hat{\mathbf{V}}_s (\hat{\mathbf{V}}_s^\intercal \hat{\mathbf{V}}_s)^{-1}$ and does not directly involve the DLV dynamics.

C. A Dynamics-Projection Alternate Updating Scheme

Exploiting the interaction between the latent dynamics estimation in Section III-A and the oblique projection estimation in Section III-B, we develop an alternating iterative scheme to identify a PredVAR model with an illustration in Fig. 2 and pseudo-codes in Algorithm 2. The algorithm's convergence follows from the classic result on EM [24] if the $\mathbf{R}$-estimates are replaced with true values. Moreover, the estimate $\hat{\mathbf{R}}$ is imposed to be orthonormal and solely depends on the statistical constraint (8) given other parameters.

Algorithm 1: Update $\hat{\mathbf{P}}$ and $\hat{\mathbf{R}}$ with $\hat{\mathbf{B}}$

Input: $\{\boldsymbol{y}_k \in \Re^p\}_{k=1}^{N+s}$; $\hat{\mathbf{B}}_j, j \in [\![s]\!]$; $\hat{\mathbf{R}}$;
Output: $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{e}}$; $\hat{\mathbf{P}}$; $\hat{\mathbf{R}}$ with $\hat{\mathbf{R}}^\intercal \hat{\mathbf{R}} = \mathbf{I}$;
while *the termination condition is unsatisfied* **do**
  Update $\hat{\mathbf{P}}$ in (12) and $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{e}}$ in (13);
  Perform SVD on $\hat{\mathbf{P}} = \mathcal{U}_1 \mathcal{D}_1 \mathcal{V}_1^\intercal$ with singular values in ascending order; $\hat{\bar{\mathbf{R}}} = \mathcal{U}_1(:, 1:(p-\ell))$;
  Perform SVD on $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{e}} \hat{\bar{\mathbf{R}}} = \mathcal{U}_2 \mathcal{D}_2 \mathcal{V}_2^\intercal$ with singular values in ascending order; $\hat{\mathbf{R}} = \mathcal{U}_2(:, 1:\ell)$;
  Calculate the DLVs in (9);

Algorithm 2: A PredVAR Algorithm

Input: scaled measurements $\{\boldsymbol{y}_k \in \Re^p\}_{k=1}^{N+s}$; s; l;
Output: $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}}$; $\hat{\mathbf{B}}_j, j \in [\![s]\!]$; $\hat{\mathbf{R}}$; $\hat{\mathbf{P}}$; $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{e}}$;
Perform SVD on $\mathbf{Y}_s^\intercal \mathbf{Y}_s / N = \mathcal{U}_0 \mathcal{D}_0 \mathcal{V}_0^\intercal$; $\hat{\mathbf{R}} = \mathcal{U}_0(:, 1:(p-\ell))$;
while *the termination condition is unsatisfied* **do**
  Estimate the DLVs in (9);
  Update $\mathbb{B}$ in (10) and $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$ in (11);
  Update $\hat{\mathbf{R}}$, $\hat{\mathbf{P}}$, and $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{e}}$ in Algorithm 1;

Note that, although $\mathbf{R}^\intercal \mathbf{P} = \mathbf{I}$ is not enforced in the algorithm, it follows from (9), (10), and (12) that

$$\hat{\mathbf{R}}^\intercal \hat{\mathbf{P}} \overset{(12)}{=} \hat{\mathbf{R}}^\intercal \mathbf{Y}_s^\intercal \hat{\mathbb{V}}\hat{\mathbb{B}}(\hat{\mathbb{B}}^\intercal \hat{\mathbb{V}}^\intercal \hat{\mathbb{V}}\hat{\mathbb{B}})^{-1} \overset{(9)}{=} \mathbf{V}_s^\intercal \hat{\mathbb{V}}\hat{\mathbb{B}}(\hat{\mathbb{B}}^\intercal \hat{\mathbb{V}}^\intercal \hat{\mathbb{V}}\hat{\mathbb{B}})^{-1}$$
$$\overset{(10)}{=} \mathbf{V}_s^\intercal \hat{\mathbb{V}}(\hat{\mathbb{V}}^\intercal \hat{\mathbb{V}})^{-1} \mathbf{V}_s (\mathbf{V}_s^\intercal \hat{\mathbb{V}}(\hat{\mathbb{V}}^\intercal \hat{\mathbb{V}})^{-1} \mathbf{V}_s)^{-1} = \mathbf{I},$$

and follows from (9), (11), (13), and $\hat{\mathbf{P}}^\intercal \hat{\mathbf{R}} = \mathbf{I}$ that

$$\hat{\mathbf{R}}^\intercal \hat{\boldsymbol{\Sigma}}_{\boldsymbol{e}} \hat{\mathbf{R}} \overset{(13)}{=} \hat{\mathbf{R}}^\intercal (\mathbf{Y}_s - \hat{\mathbb{V}}\hat{\mathbb{B}}\hat{\mathbf{P}}^\intercal)^\intercal (\mathbf{Y}_s - \hat{\mathbb{V}}\hat{\mathbb{B}}\hat{\mathbf{P}}^\intercal)\hat{\mathbf{R}}/N$$
$$\overset{(9)}{=} (\hat{\mathbf{V}}_s - \hat{\mathbb{V}}\hat{\mathbb{B}})^\intercal (\hat{\mathbf{V}}_s - \hat{\mathbb{V}}\hat{\mathbb{B}})/N \overset{(11)}{=} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}}$$

after the convergence of Algorithm 2.

IV. Simulation Case Study

A. Data Description

The Lorenz oscillator described in [12] is used to generate the latent dynamics of synthesized data. Note that the Lorenz oscillator with three coordinates $\boldsymbol{v} \in \Re^3$ is a nonlinear chaotic system instead of following the regression dynamics as in (2). Assume that there are six sensors and each sensor measures a mix of the DLVs $\boldsymbol{v} \in \Re^3$ and static noise $\bar{\boldsymbol{\varepsilon}} \in \Re^3$ injected via respective channels.

Collect 10000 subsequent data points from the Lorenz oscillator and set the variance of the noise $\bar{\boldsymbol{\varepsilon}}$ as that of the

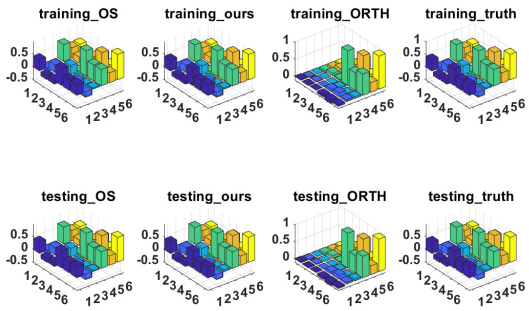

Fig. 3. Measurement reconstruction covariance.

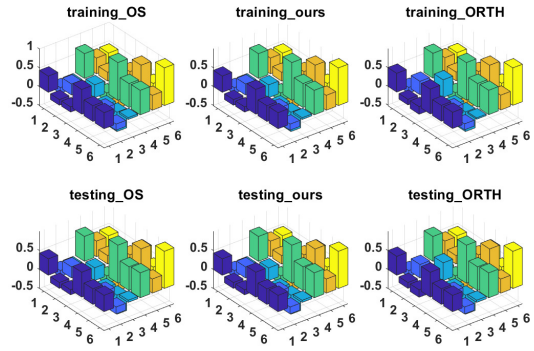

Fig. 4. Measurement prediction covariance.

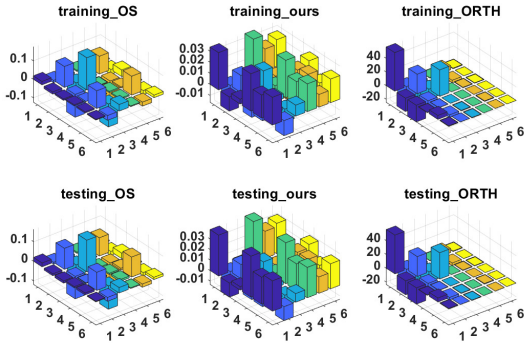

Fig. 5. Signal reconstruction covariance.

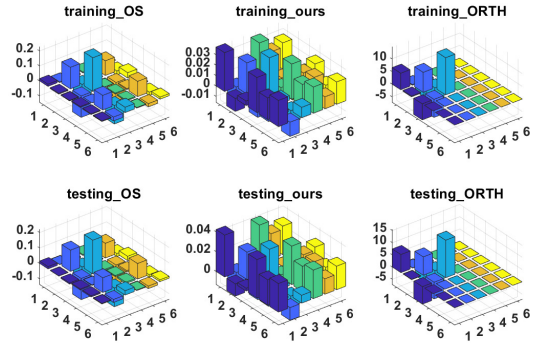

Fig. 6. Signal prediction covariance.

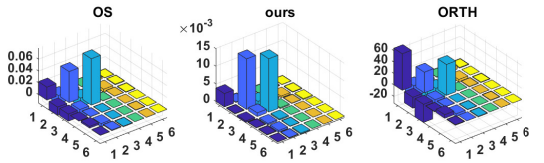

Fig. 7. Signal prediction covariance by EM (non-orthogonal generation).

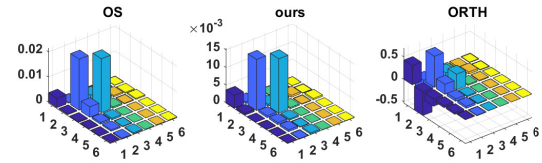

Fig. 8. Signal prediction covariance by EM (ORTH generation).

collected data. The loadings matrices are set as

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \text{ and } \bar{\mathbf{P}} = \begin{bmatrix} -0.2997 & -0.4611 & -0.2868 \\ -0.2403 & 0.2559 & 0.6444 \\ -0.1334 & 0.5749 & -0.5168 \\ -0.2997 & -0.4611 & -0.2868 \\ -0.5400 & -0.2052 & 0.3576 \\ -0.6733 & 0.3697 & -0.1592 \end{bmatrix}.$$

The signal and noise subspaces are highly oblique as the canonical angles [25] between the subspaces spanned by the columns of $\mathbf{R}$ and $\mathbf{P}$ are 23.99°, 51.27°, and 60.97°, which should be zeros for orthogonal cases. Obtain measurement samples $\{\boldsymbol{y}_k\}$ by (1). Unless otherwise specified, the first and last 3000 samples are for training and testing, respectively.

B. Benchmark Algorithms

The first benchmark (denoted as OS) is called a one-shot algorithm. First, the eigendecomposition method in [4] is used to identify $\mathbf{P}$ and $\mathbf{R}$. Then, with the estimated oblique projection, update the parameters on the DLV dynamics as in Lines 3 and 4 of Algorithm 2. As the OS algorithm does not have an alternate updating procedure, the comparison between the OS algorithm and ours will show the improving effect of the dynamics-projection interaction.

The second benchmark (denoted as ORTH) emphasizes the orthogonal projection on the signal subspace. That is, the DLVs $\boldsymbol{v}_k$ are estimated via the natural filter:

$$\boldsymbol{v}_k^{or} = (\mathbf{P}^\intercal \mathbf{P})^{-1} \mathbf{P}^\intercal \boldsymbol{y}_k = \mathbf{P}^\dagger \boldsymbol{y}_k,$$

where $\mathbf{P}^\dagger$ is the Moore-Penrose inverse. The alternate updating procedure is still applicable and used. The comparison between this algorithm and ours will verify the improving effect of the oblique projection perspective.

C. Comparison

The simulation results in the training and testing sets are presented separately. Similar performances regarding the two sets are observed for each algorithm. This fact suggests the efficacy of all algorithms. Nevertheless, the three algorithms have different efficiency levels.

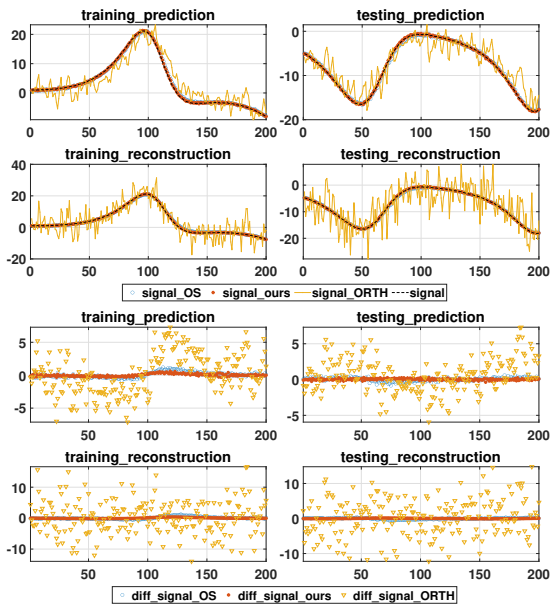

Fig. 9. Reconstruction and prediction for Sensor 2.

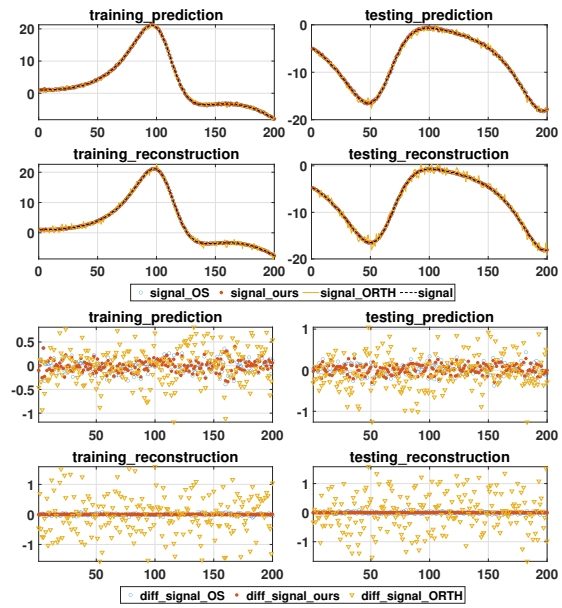

Fig. 10. Reconstruction and prediction for Sensor 2 (ORTH generation).

Fig. 3 concerns the measurement reconstruction and depicts the covariance matrix of the series $\{\boldsymbol{y}_k - \hat{\mathbf{P}}\hat{\mathbf{R}}^\intercal \boldsymbol{y}_k\}$. For the reference purpose, the covariance matrix (denoted as truth) of $\{\boldsymbol{y}_k - \mathbf{P}\boldsymbol{v}_k\}$ is visualized, referring to the difference between the measurement and the true signal. As can be seen, the measurement reconstruction covariance for the OS algorithm or ours is similar to the reference covariance. However, the ORTH algorithm leads to a measurement reconstruction covariance matrix with smaller values, especially for those involving the first three sensors. This observation indicates that the ORTH algorithm does not eliminate sufficient static noise from the measurement for DLV dynamics identification, compared with two others. Fig. 4 depicts the covariance matrix of $\{\boldsymbol{y}_k - \hat{\mathbf{P}}\hat{\tilde{\boldsymbol{v}}}_k\}$ and shows little difference among the three algorithms regarding measurement prediction. Combing this observation with Fig. 4, we argue that more noise is involved in the DLV dynamics estimation of the ORTH algorithm than the other two counterparts.

Fig. 5 depicts the covariance matrix of $\hat{\mathbf{P}}\hat{\boldsymbol{v}}_k - \mathbf{P}\boldsymbol{v}_k$, referring to the difference between the reconstructed and true signals. Fig. 6 depicts the covariance matrix of $\hat{\mathbf{P}}\hat{\tilde{\boldsymbol{v}}}_k - \mathbf{P}\boldsymbol{v}_k$, referring to the difference between the predicted and true signals. As can be seen, the proposed algorithm attains the best reconstruction and prediction performances regarding covariance, while the ORTH algorithm achieves the worst. Fig. 7 depicts $\hat{\mathbf{P}}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}}\hat{\mathbf{P}}^\intercal$, where $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}}$ is estimated from the EM procedure in each algorithm. Again, the proposed algorithm achieves the best performance in signal prediction, and the ORTH algorithm is the worst by likelihood analysis. These observations suggest that the proposed algorithm is the best to eliminate the static noise from the measurement and improve the DLV dynamics estimation.

The reconstructed ($\hat{\mathbf{P}}\hat{\mathbf{R}}^\intercal \boldsymbol{y}_k$), predicted ($\hat{\mathbf{P}}\hat{\tilde{\boldsymbol{v}}}_k$), and true ($\mathbf{P}\boldsymbol{v}_k$) signals in Sensor 2 are plotted in Fig. 9. Also, the difference between the reconstructed and true signals and the difference between the predicted and true signals are plotted. The OS algorithm and ours can reconstruct or predict the signal well, but ours attains better performance. This fact exhibits the strength of our alternating iterative procedure. Also, we see many spikes in the signal curves generated by the ORTH algorithm. This observation again verifies that too much noise is reserved in the DLV dynamics estimation of the ORTH algorithm. Nevertheless, the curves generated by the ORTH algorithm can track the signal curves. This fact is consistent with the observation in Fig. 5 and Fig. 6 that the signal prediction performance is better than the signal reconstruction performance for the ORTH algorithm.

D. Empirical Consistency Analysis

This subsection shows how the training sample set affects the oblique subspaces estimation. Table I records the Frobenius norm of the difference between the true projection matrix $\mathbf{P}\mathbf{R}^\intercal$ and its estimate from an algorithm by using the first x samples in the data set, with x varying from 1000 to 10000. Overall, the projection estimations of all three algorithms benefit from sample augmentation, though the effect diminishes and occasionally reverses as the sample number increases. Moreover, for the same training set, the proposed algorithm still attains a projection closest to the ground truth than the two benchmarks. We observe similar phenomena for signal subspace identification. Table II records the average canonical angle between the true signal subspace spanned by $\mathbf{P}$ and the estimated signal subspace spanned by $\hat{\mathbf{P}}$ for each algorithm. A smaller angle usually indicates a better signal subspace estimation. The two tables show the empirical consistency for all algorithms. Moreover, the proposed algorithm can extract more accurate information with fewer measurements, benefiting from the dynamics-projection interaction and the oblique projection perspective.

TABLE I

DISTANCE TO THE TRUE PROJECTION MATRIX

Num. \ Alg.	OS	PredVAR	ORTH
1000	1.2362	0.8774	2.0075
2000	0.9113	0.8367	1.9381
3000	0.9338	0.8488	1.7015
4000	0.9016	0.8352	1.7338
5000	0.9031	0.8369	1.7537
6000	0.8679	0.8270	1.8011
7000	0.8601	0.8262	1.7787
8000	0.8589	0.8313	1.7774
9000	0.8476	0.8271	1.7765
10000	0.8571	0.8330	1.7610

TABLE II

ANGLE-DISTANCE TO THE TRUE SIGNAL SUBSPACE

Num. \ Alg.	OS	PredVAR	ORTH
1000	10.42°	4.31°	22.68°
2000	5.30°	2.96°	16.32°
3000	4.34°	3.05°	4.28°
4000	3.51°	1.76°	2.90°
5000	3.44°	1.49°	4.15°
6000	2.74°	1.19°	3.47°
7000	2.93°	1.38°	2.86°
8000	2.43°	1.32°	2.10°
9000	1.44°	0.94°	1.87°
10000	1.74°	1.02°	1.94°

E. Data with Orthogonal Signal and Noise Subspaces

To further demonstrate the merit of the oblique-projection perspective, the measurements are generated using the same data points from the Lorenz oscillator and noise data $\{\bar{\boldsymbol{\varepsilon}}_k\}$ but changing the static loadings matrix $\bar{\mathbf{P}}$ as

$$\begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}^{\mathsf{T}}.$$

In this case, the signal and noise subspaces are orthogonal (ORTH generation). Fig. 8 depicts $\hat{\mathbf{P}}\hat{\bar{\boldsymbol{\Sigma}}}_{\boldsymbol{\varepsilon}}\hat{\mathbf{P}}^{\mathsf{T}}$, where $\hat{\bar{\boldsymbol{\Sigma}}}_{\boldsymbol{\varepsilon}}$ is estimated from the corresponding EM procedure in each algorithm. As seen, the proposed algorithm again pursues the smallest signal prediction covariance. Also, the reconstructed, predicted, true signals and their differences in Sensor 2 are plotted in Fig. 10. The performance of the ORTH algorithm improves in this case but is still worse than the other two. Notably, there are still spikes in the signal curves generated by the ORTH algorithm. The underlying reason may be that the Lorenz attractor is a nonlinear system, which leads to model mismatch with a linear VAR model to capture its dynamics. The ORTH algorithm is sensitive to the model mismatch. In contrast, the additional freedom can endow oblique projection with the ability to compensate for the DLV model mismatch against orthogonal projection that requires $\mathbf{R}$ and $\mathbf{P}$ to form the same subspace.

REFERENCES

[1] M. Sznaier, "Control oriented learning in the era of big data," *IEEE Control Syst. Lett.*, vol. 5, pp. 1855–1867, 2020.

[2] D. Pena and G. E. P. Box, "Identifying a simplifying structure in time series," *J. Am. Stat. Assoc.*, vol. 82, pp. 836–843, 1987.

[3] S. J. Qin, Y. Dong, Q. Zhu, J. Wang, and Q. Liu, "Bridging systems theory and data science: A unifying review of dynamic latent variable analytics and process monitoring," *Annu Rev Control*, vol. 50, pp. 29–48, 2020.

[4] Z. Gao and R. S. Tsay, "Modeling high-dimensional time series: A factor model with dynamically dependent factors and diverging eigenvalues," *J. Am. Stat. Assoc.*, pp. 1–17, 2021.

[5] G. C. Reinsel, R. P. Velu, and K. Chen, *Multivariate Reduced-Rank Regression: Theory, Methods and Applications*, vol. 225. Springer Nature, 2023.

[6] W. Ku, R. H. Storer, and C. Georgakis, "Disturbance detection and isolation by dynamic principal component analysis," *Chemometrics Intell. Lab. Syst.*, vol. 30, pp. 179–196, 1995.

[7] W. Li and S. J. Qin, "Consistent dynamic PCA based on errors-in-variables subspace identification," *J. Process Control*, vol. 11, pp. 661–678, 2001.

[8] G. Li, S. J. Qin, and D. Zhou, "A new method of dynamic latent-variable modeling for process monitoring," *IEEE Trans. Ind. Electron.*, vol. 61, pp. 6438–6445, 2014.

[9] Y. Dong and S. J. Qin, "A novel dynamic PCA algorithm for dynamic data modeling and process monitoring," *J. Process Control*, vol. 67, pp. 1–11, 2018.

[10] Y. Dong and S. J. Qin, "Dynamic latent variable analytics for process operations and control," *Comput. Chem. Eng.*, vol. 114, pp. 69–80, 2018.

[11] S. J. Qin, "Latent vector autoregressive modeling for reduced dimensional dynamic feature extraction and prediction," in *IEEE Conf. Decis. Control (CDC)*, pp. 3689–3694, IEEE, 2021.

[12] S. J. Qin, "Latent vector autoregressive modeling and feature analysis of high dimensional and noisy data from dynamic systems," *AIChE J.*, p. e17703, 2022.

[13] J. Yu and S. J. Qin, "Latent state space modeling of high-dimensional time series with a canonical correlation objective," *IEEE Control Syst. Lett.*, vol. 6, pp. 3469–3474, 2022.

[14] G. E. P. Box and G. C. Tiao, "A canonical analysis of multiple time series," *Biometrika*, vol. 64, pp. 355–365, 1977.

[15] C. Lam and Q. Yao, "Factor modeling for high-dimensional time series: Inference for the number of factors," *Ann. Stat.*, pp. 694–726, 2012.

[16] D. Peña, E. Smucler, and V. J. Yohai, "Forecasting multiple time series with one-sided dynamic principal components," *J. Am. Stat. Assoc.*, 2019.

[17] Q. Wen, Z. Ge, and Z. Song, "Data-based linear Gaussian state-space model for dynamic process monitoring," *AIChE J.*, vol. 58, no. 12, pp. 3763–3776, 2012.

[18] L. Zhou, G. Li, Z. Song, and S. J. Qin, "Autoregressive dynamic latent variable models for process monitoring," *IEEE Trans. Control Syst. Technol.*, vol. 25, pp. 366–373, 2016.

[19] R. T. Behrens and L. L. Scharf, "Signal processing applications of oblique projection operators," *IEEE Trans. Signal Process.*, vol. 42, pp. 1413–1424, 1994.

[20] Y. Dong, Y. Liu, and S. J. Qin, "Efficient dynamic latent variable analysis for high-dimensional time series data," *IEEE Trans. Ind. Informat.*, vol. 16, pp. 4068–4076, 2019.

[21] G. Pillonetto, F. Dinuzzo, T. Chen, G. De Nicolao, and L. Ljung, "Kernel methods in system identification, machine learning and function estimation: A survey," *Automatica*, vol. 50, pp. 657–682, 2014.

[22] M. Khosravi and R. S. Smith, "The existence and uniqueness of solutions for kernel-based system identification," *Automatica*, vol. 148, p. 110728, 2023.

[23] J. Bai and K. Li, "Statistical analysis of factor models of high dimension," *Ann. Stat.*, vol. 40, pp. 436–465, 2012.

[24] C. F. J. Wu, "On the convergence properties of the EM algorithm," *Ann. Stat.*, pp. 95–103, 1983.

[25] L. Qiu, Y. Zhang, and C.-K. Li, "Unitarily invariant metrics on the grassmann space," *SIAM J. Matrix Anal. Appl.*, vol. 27, pp. 507–531, 2005.