Preprints of the 23rd IFAC World Congress
Busan, Republic of Korea, August 23-28, 2026

A Novel State-Space Model Identification Method from A Behavioral System-Theoretic Perspective

Qingyuan Liu * **Yibo Wang** * **Tao Liu** ** **Zhongmei Li** ***
Xiao He * **Chao Shang** *

* *Department of Automation, Tsinghua University, Beijing 100084, China (e-mail: c-shang@tsinghua.edu.cn).*
** *Key Laboratory of Intelligent Control and Optimization for Industrial Equipment of Ministry of Education, Dalian University of Technology, Dalian 116024, China*
*** *State Key Laboratory of Industrial Control Technology, Ministry of Education, East China University of Science and Technology, Shanghai 200237, China*

Abstract: As the mainstream methodology for identifying state-space models, subspace identification relies on the orthogonality assumption between data spaces, and could thus lead to unsatisfactory identification accuracy in finite-sample regime. To overcome this limitation, a novel state-space model identification method is proposed by leveraging the capability of behavioral systems theory in characterizing system dynamics with finite-length data trajectory. In virtue of the innovation-based data-driven output predictor (DDOP), the state-space model identification is converted into an innovation estimation problem followed by a model reduction step. To achieve better identification accuracy, an improved innovation estimation strategy incorporating low-rank prior is further proposed, formulated as a rank-constrained programming problem and solved via the alternating direction method of multipliers (ADMM). Numerical and industrial dataset experiments demonstrate the superior modeling accuracy of the proposed method over existing subspace identification methods in both open-loop and closed-loop cases.

Keywords: Linear system identification, State-space model, Subspace identification, Behavioral systems theory, Rank-constrained programming, Closed-loop identification

1. INTRODUCTION

State-space model has been the most prevalent choice for modeling linear time-invariant (LTI) multi-input multi-output (MIMO) processes, with subspace identification being a prevalent tool (Verhaegen, 2021). By exploiting the statistical orthogonality between process data spaces, subspace identification seeks to recover via projection the system's observable subspace. Thereafter, matrix factorization techniques such as singular value decomposition (SVD) and QR decomposition can be applied to derive a low-dimensional subspace, from which the state-space system matrices can be readily attained (Van Overschee and De Moor, 1995). Because of its numerical stability and computational efficiency, subspace identification has received continual attention since its invention (Van Overschee and De Moor, 1995; Huang and Kadali, 2008).

However, since the statistical orthogonality between data spaces only holds as the data length goes to infinity, existing subspace identification methods can only guarantee asymptotic properties. This limitation may lead to significant modeling errors, especially when available data are inadequate for identification or corrupted with heavy measurement noise.

In recent years, the behavioral systems theory, which seeks to characterize process dynamics directly with finite input-output data, has received increasing attention. Central to this theory is the Willems' fundamental lemma (Willems et al., 2005), which states that all trajectories from a deterministic LTI system can be expressed as linear combinations of finite accessible trajectories. It has inspired a series of direct data-driven output predictor and predictive control (DDOP/DDPC) methods as promising alternatives to traditional identification-for-prediction/control schemes (Coulson et al., 2019; Markovsky and Dörfler, 2021; Dörfler et al., 2022; Breschi et al., 2023).

In a more recent work (Wang et al., 2025), this direct data-driven paradigm have been further extended to tackle stochastic systems by regarding the innovation as an augmented system input, yielding the so-called innovation-based DDOP. It is found that with finite input-output-innovation data, the innovation-based DDOP may achieve equivalent prediction performance to that of the oracle Kalman filter under ideal conditions. This suggests a new approach to state-space model identification. Specifically,

⋆ This work is supported by National Natural Science Foundation of China (Nos. 62373211 and 62327807), and the Open Research Project of the State Key Laboratory of Industrial Control Technology, China (Grant No. ICT2025B10).

©2026 the authors. Accepted by IFAC for publication under a Creative Commons Licence CC-BY-NC-ND

since the innovation-based DDOP is still underpinned by the LTI system assumption, which is also inherent in state-space models, a new route towards state-space model identification from finite data is possible from a behavioral system-theoretic perspective.

Motivated by this, we propose a novel state-space model identification method based on the behavioral systems theory. By transforming the innovation-based DDOP into an equivalent state-space formulation, the state-space model identification problem can be recast as an offline innovation estimation problem, thereby yielding a novel identification scheme. Furthermore, To achieve higher accuracy of innovation estimation than the trivial least-squares (LS) method, we propose an improved strategy by incorporating the low-rank structure prior, which modeling the innovation estimation problem as a rank-constrained programming (RCP). A tailored solution based on alternating direction method of multipliers (ADMM), which has emerged as a popular choice for solving RCP in recent system identification research (Liu et al., 2021, 2023), is subsequently presented to efficiently solve the problem. Numerical examples and industrial case studies showcase the superior identification accuracy of the proposed method over existing subspace identification methods under both open-loop and closed-loop identification cases.

The remainder of the paper is structured as follows. After introducing the preliminaries in Section 2, the equivalent state-space realization of the innovation-based DDOP as well as an improved offline innovation estimation strategy is proposed in Section 3 to implement the novel state-space identification method. Numerical examples and industrial case studies are presented in Sections 4 and 5. Section 6 concludes this work.

Notations: $\mathbb{R}$, $\mathbb{Z}$, and $\mathbb{N}$ denote the sets of real numbers, integers, and natural numbers, respectively. $\mathbf{0}$ and $\mathbf{1}$ denote matrices or vectors of all zeros and all ones with appropriate dimensions, respectively. $\mathbf{I}$ denotes a unitary matrix with appropriate dimensions. For matrices or vectors $\mathbf{M}_1, \mathbf{M}_2, \dots, \mathbf{M}_n$ of same dimensions, we define $\mathbf{col}(\mathbf{M}_1, \dots, \mathbf{M}_n) \triangleq \begin{bmatrix}\mathbf{M}_1^{\mathrm{T}} & \cdots & \mathbf{M}_n^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}}$. The block-Hankel operator and the lower-triangular block-Toeplitz operator are defined by

$$\mathcal{H}_s(\mathbf{M}_1, \dots, \mathbf{M}_n) \triangleq \begin{bmatrix}\mathbf{M}_1 & \cdots & \mathbf{M}_{n-s+1} \\ \vdots & \ddots & \vdots \\ \mathbf{M}_s & \cdots & \mathbf{M}_n\end{bmatrix},$$

$$\mathcal{T}(\mathbf{M}_1, \dots, \mathbf{M}_n) \triangleq \begin{bmatrix}\mathbf{M}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{M}_2 & \mathbf{M}_1 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{M}_n & \mathbf{M}_{n-1} & \cdots & \mathbf{M}_1\end{bmatrix}.$$

For discrete-time series $\mathbf{x}(t) = [x_1(t)\ x_2(t)\ \cdots\ x_n(t)]^{\mathrm{T}}$, we define $\mathbf{x}_{i:j} \triangleq \mathbf{col}(\mathbf{x}(i), \mathbf{x}(i+1), \dots, \mathbf{x}(j))$, $\mathcal{H}_s(\mathbf{x}_{i:j}) \triangleq \mathcal{H}_s(\mathbf{x}(i), \mathbf{x}(i+1), \dots, \mathbf{x}(j))$.

2. PRELIMINARIES

2.1 State-space model and subspace identification

Consider the following stochastic discrete-time state-space model in innovation form:

$$\begin{cases}\hat{\mathbf{x}}(t+1) = \mathbf{A}\hat{\mathbf{x}}(t) + \mathbf{B}\mathbf{u}(t) + \mathbf{K}\mathbf{e}(t) \\ \mathbf{y}(t) = \mathbf{C}\hat{\mathbf{x}}(t) + \mathbf{D}\mathbf{u}(t) + \mathbf{e}(t)\end{cases} \tag{1}$$

where $\mathbf{u}(t) \in \mathbb{R}^m$, $\mathbf{y}(t) \in \mathbb{R}^n$, $\mathbf{x}(t) \in \mathbb{R}^r$ and $\mathbf{e}(t) \in \mathbb{R}^n$ denote system inputs, outputs and states, respectively. $\mathbf{e}(t)$ is the system innovation. $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ are state-space matrices. $\mathbf{K}$ is the steady-state Kalman gain rendering $\mathbf{A} - \mathbf{KC}$ Schur stable.

On this basis, state-space model identification aims to estimate the optimal state-space matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{K}$ to best fit the true process from a finite collection of input-output data $\{\mathbf{u}(t), \mathbf{y}(t)\}_{t=0}^{L-1}$. Its principle can be briefly summarized as follows. First, by iterating model (1) over time, one can obtain the following extended state-space model (Van Overschee and De Moor, 1995):

$$\mathbf{Y}_{\mathrm{f}} = \mathbf{O}\hat{\mathbf{X}}_{\mathrm{f}} + \mathbf{L}_{uf}\mathbf{U}_{\mathrm{f}} + \mathbf{L}_{ef}\mathbf{E}_{\mathrm{f}}, \tag{2}$$

where $\mathbf{O}$ is the extended observability matrix, $\hat{\mathbf{X}}_{\mathrm{f}}$, $\mathbf{U}_{\mathrm{f}}$, $\mathbf{Y}_{\mathrm{f}}$, $\mathbf{E}_{\mathrm{f}}$ are block-Hankel data matrices. $\mathbf{L}_{uf}$ and $\mathbf{L}_{ef}$ are coefficient matrices consisting of Markov parameters.

Second, the unmeasurable system states $\hat{\mathbf{X}}_{\mathrm{f}}$ can be approximated with past inputs and outputs. By substituting the output equation into the state equation, the innovation form model (1) can be further reformulated into the following observer form:

$$\begin{cases}\hat{\mathbf{x}}(t+1) = \bar{\mathbf{A}}\hat{\mathbf{x}}(t) + \bar{\mathbf{B}}\mathbf{u}(t) + \mathbf{K}\mathbf{y}(t) \\ \mathbf{y}(t) = \mathbf{C}\hat{\mathbf{x}}(t) + \mathbf{D}\mathbf{u}(t) + \mathbf{e}(t)\end{cases} \tag{3}$$

where $\bar{\mathbf{A}} \triangleq \mathbf{A} - \mathbf{KC}$, $\bar{\mathbf{B}} \triangleq \mathbf{B} - \mathbf{KD}$. Owing to the Schur stability of $\bar{\mathbf{A}}$, $\bar{\mathbf{A}}^{T_{\mathrm{p}}} \approx \mathbf{0}$ for sufficiently large T_{p}. By iterating the state equation in (3) over time, one can approximate the system states $\hat{\mathbf{X}}_{\mathrm{f}}$ as

$$\hat{\mathbf{X}}_{\mathrm{f}} = \bar{\mathbf{A}}^{T_{\mathrm{p}}}\hat{\mathbf{X}}_{\mathrm{p}} + \mathbf{\Gamma}_u\mathbf{U}_{\mathrm{p}} + \mathbf{\Gamma}_y\mathbf{Y}_p \approx \mathbf{\Gamma}_u\mathbf{U}_{\mathrm{p}} + \mathbf{\Gamma}_y\mathbf{Y}_p. \tag{4}$$

By substituting (4) into the extended state-space model (2), one obtains the following input-output data relation:

$$\mathbf{Y}_{\mathrm{f}} \approx \mathbf{L}_{up}\mathbf{U}_{\mathrm{p}} + \mathbf{L}_{yp}\mathbf{Y}_{\mathrm{p}} + \mathbf{L}_{uf}\mathbf{U}_{\mathrm{f}} + \mathbf{L}_{ef}\mathbf{E}_{\mathrm{f}}, \tag{5}$$

where $\mathbf{L}_{up} \triangleq \mathbf{O}\mathbf{\Gamma}_u$, $\mathbf{L}_{yp} \triangleq \mathbf{O}\mathbf{\Gamma}_y$.

Third, by regarding $\mathbf{L}_{ef}\mathbf{E}_{\mathrm{f}}$ in (5) as residuals, the unknown coefficient matrices $\mathbf{L}_{up}$, $\mathbf{L}_{yp}$, $\mathbf{L}_{uf}$ can be estimated by solving a least-square problem. Then, a low-rank extended observability matrix $\hat{\mathbf{O}}$ can be estimated through SVD and its column vectors span the low-dimensional observable subspace, and the state-space matrices can be recovered from this subspace (Van Overschee and De Moor, 1995; Verhaegen and Verdult, 2007; Huang and Kadali, 2008). Such a paradigm encapsulates the majority of known subspace identification methods, including open-loop SIMs such as CVA, N4SID, MOESP, N2SID, and closed-loop SIMs such as closed-loop N4SID, SIMPCA, PBSID.

2.2 Direct data-driven output prediction of stochastic systems

Recently, a direct data-driven paradigm for output prediction of LTI systems has emerged as an alternative to the conventional model-based scheme (Hou and Wang, 2013), which builds upon the Willems' fundamental lemma (Willems et al., 2005), which enables a direct DDOP of deterministic LTI systems, and has underpinned a wide spectrum of direct DDPC methods. However, since the fundamental lemma considers the deterministic case, the

derived DDOP is prone to significant prediction error when handling stochastic cases with $\mathbf{e}(t) \neq \mathbf{0}$. To overcome this limitation, a more recent work (Wang et al., 2025) has extended the fundamental lemma for stochastic systems as follows:

Theorem 1. (Fundamental lemma in innovation form, Wang et al. (2025)). Consider the system (1), from which an input-output-innovation trajectory $\{\mathbf{u}(t), \mathbf{y}(t), \mathbf{e}(t)\}_{t=1}^{N}$ is generated offline. We define the online past data $\{\mathbf{u}_\mathrm{p}^\mathrm{on}(t), \mathbf{y}_\mathrm{p}^\mathrm{on}(t), \mathbf{e}_\mathrm{p}^\mathrm{on}(t)\}$ with $\mathbf{e}_\mathrm{p}^\mathrm{on}(t) = \mathbf{e}_{t-L_\mathrm{p}:t-1}^\mathrm{on}$, and the future inputs $\mathbf{u}_\mathrm{f}^\mathrm{on}(t)$. If $L_\mathrm{p} \geq l(A, B, C, D)$, $N \geq (m+n+1)L + r - 1$ and the extended inputs $\tilde{\mathbf{u}}_{1:N}$ defined by $\tilde{\mathbf{u}}(t) \triangleq \mathbf{col}(\mathbf{u}(t), \mathbf{e}(t))$ is persistently exciting of order $(L+m)$, a DDOP is given by

$$\hat{\mathbf{y}}_\mathrm{f}^\mathrm{on}(t) = \mathbf{Y}_\mathrm{f}\mathbf{g}(t), \tag{6}$$

where $\mathbf{g}(t)$ solving

$$\begin{aligned}&\mathbf{col}(\mathbf{u}_\mathrm{p}^\mathrm{on}(t), \mathbf{y}_\mathrm{p}^\mathrm{on}(t), \mathbf{e}_\mathrm{p}^\mathrm{on}(t), \mathbf{u}_\mathrm{f}^\mathrm{on}(t), \mathbf{0})\\ &= \mathbf{col}(\mathbf{U}_\mathrm{p}, \mathbf{Y}_\mathrm{p}, \mathbf{E}_\mathrm{p}, \mathbf{U}_\mathrm{f}, \mathbf{E}_\mathrm{f})\mathbf{g}(t),\end{aligned} \tag{7}$$

with $\mathbf{U}_\mathrm{p}$ contains the first (mL_p) rows of $\mathbf{U}$, $\mathbf{U}_\mathrm{f}$ contains the remaining (mL_f) rows, and $\mathbf{Y}_\mathrm{p}$, $\mathbf{Y}_\mathrm{f}$, $\mathbf{E}_\mathrm{p}$, $\mathbf{E}_\mathrm{f}$ are constructed by splitting $\mathbf{Y}$ in a similar way. In practical use, the offline innovations $\mathbf{e}_{1:N}$ are unmeasurable and it is suggested to use their estimates instead.

Theorem 1 provides a promising approach to the DDOP of stochastic systems, namely the innovation-based DDOP (Wang et al., 2025). Thanks to the use of innovations, the innovation-based DDOP better resembles the Kalman filter in comparison with the deterministic DDOP and thus exhibits desirable performance.

3. THE PROPOSED METHOD

In this section we describe a route towards state-space model identification inspired by the innovation-based DDOP.

3.1 State-space realization of innovation-based DDOP

We first show that the innovation-based DDOP in (6) and (7) admits an equivalent state-space formulation. By substituting (7) into (6), one obtains:

$$\hat{\mathbf{y}}_\mathrm{f}^\mathrm{on}(t) = \mathbf{M} \cdot \mathbf{col}(\mathbf{u}_\mathrm{p}^\mathrm{on}(t), \mathbf{y}_\mathrm{p}^\mathrm{on}(t), \mathbf{e}_\mathrm{p}^\mathrm{on}(t), \mathbf{u}_\mathrm{f}^\mathrm{on}(t)), \tag{8}$$

where

$$\mathbf{M} = \mathbf{Y}_\mathrm{f} \cdot \mathbf{col}(\mathbf{U}_\mathrm{p}, \mathbf{Y}_\mathrm{p}, \hat{\mathbf{E}}_\mathrm{p}, \mathbf{U}_\mathrm{f}, \hat{\mathbf{E}}_\mathrm{f})^\dagger \left[\mathbf{I}\ \mathbf{0}\right]. \tag{9}$$

From (8) it can be clearly observed that in the innovation-based DDOP, the predicted outputs $\mathbf{y}_\mathrm{f}^\mathrm{on}(t)$ is linearly dependent on $\mathbf{u}_\mathrm{p}^\mathrm{on}(t)$, $\mathbf{y}_\mathrm{p}^\mathrm{on}(t)$, $\mathbf{e}_\mathrm{p}^\mathrm{on}(t)$ and $\mathbf{u}_\mathrm{f}^\mathrm{on}(t)$. Hence, by defining $\mathbf{x}_\mathrm{DD}(t) = \mathbf{col}(\mathbf{u}_\mathrm{p}^\mathrm{on}(t), \mathbf{y}_\mathrm{p}^\mathrm{on}(t), \mathbf{e}_\mathrm{p}^\mathrm{on}(t)) \in \mathbb{R}^{(m+2n)L_\mathrm{p}}$ as the surrogate state, $\mathbf{u}_\mathrm{f}^\mathrm{on}(t)$ as input, $\mathbf{u}_\mathrm{f}^\mathrm{on}(t)$ as output, $\mathbf{e}_\mathrm{f}^\mathrm{on}$ as innovation, it is straightforward to derive the following state-space formulation of innovation-based DDOP:

$$\begin{cases}\mathbf{x}_\mathrm{DD}(t+1) = \mathbf{A}_\mathrm{DD}\mathbf{x}_\mathrm{DD}(t) + \mathbf{B}_\mathrm{DD}\mathbf{u}_\mathrm{f}^\mathrm{on}(t) + \mathbf{K}_\mathrm{DD}\mathbf{e}_\mathrm{f}^\mathrm{on}(t)\\ \hat{\mathbf{y}}_\mathrm{f}^\mathrm{on}(t) = \mathbf{C}_\mathrm{DD}\mathbf{x}_\mathrm{DD}(t) + \mathbf{D}_\mathrm{DD}\mathbf{u}_\mathrm{f}^\mathrm{on}(t) + \mathbf{e}_\mathrm{f}^\mathrm{on}(t)\end{cases} \tag{10}$$

$$\begin{aligned}\mathbf{A}_\mathrm{DD} = &\ \mathrm{diag}\{\mathbf{Z}_{up}, \mathbf{Z}_{yp}, \mathbf{Z}_{ep}\}\\ &+ \mathbf{col}(\mathbf{0}, \mathbf{Z}_{yf}, \mathbf{0})\left[\mathbf{M}_{up}\ \mathbf{M}_{yp}\ \mathbf{M}_{ep}\right],\end{aligned} \tag{11}$$

$$\mathbf{B}_\mathrm{DD} = \mathbf{col}(\mathbf{Z}_{uf}, \mathbf{0}, \mathbf{0}) + \mathbf{col}(\mathbf{0}, \mathbf{Z}_{yf}, \mathbf{0})\mathbf{M}_{uf}, \tag{12}$$

$$\mathbf{C}_\mathrm{DD} = \left[\mathbf{M}_{up}\ \mathbf{M}_{yp}\ \mathbf{M}_{ep}\right],\ \mathbf{D}_\mathrm{DD} = \mathbf{M}_{uf}, \tag{13}$$

$$\mathbf{K}_\mathrm{DD} = \mathbf{col}(\mathbf{0}, \mathbf{Z}_{yf}, \mathbf{Z}_{ef}), \tag{14}$$

where $\mathbf{Z}_{up}$ and $\mathbf{Z}_{uf}$ are time shift matrices defined by

$$\mathbf{u}_\mathrm{p}^\mathrm{on}(t+1) = \mathbf{Z}_{up}\mathbf{u}_\mathrm{p}^\mathrm{on}(t) + \mathbf{Z}_{uf}\mathbf{u}_\mathrm{f}^\mathrm{on}(t), \tag{15}$$

and $\mathbf{Z}_{yp}$, $\mathbf{Z}_{yf}$, $\mathbf{Z}_{ep}$, $\mathbf{Z}_{ef}$ are defined in a similar way.

With this state-space formulation, we arrive at a new approach to state-space model identification. Letting $L_\mathrm{f} = 1$, the following state-space model can be obtained:

$$\begin{cases}\mathbf{x}_\mathrm{DD}(t+1) = \mathbf{A}_\mathrm{DD}\mathbf{x}_\mathrm{DD}(t) + \mathbf{B}_\mathrm{DD}\mathbf{u}(t) + \mathbf{K}_\mathrm{DD}\mathbf{e}(t)\\ \mathbf{y}(t) = \mathbf{C}_\mathrm{DD}\mathbf{x}_\mathrm{DD}(t) + \mathbf{D}_\mathrm{DD}\mathbf{u}(t) + \mathbf{e}(t)\end{cases} \tag{16}$$

This indicates that the state-space model identification amounts to an innovation estimation problem, which, however, comes with two pitfalls. On the one hand, model (16) could be non-minimal in certain circumstances, even if the system inputs are persistently exciting. On the other hand, since the dimension of $\mathbf{x}_\mathrm{DD}$ is typically much higher than the true system order, the identification result may be prone to overfitting. Both issues suggest an additional model reduction step, which can be implemented by invoking the balanced truncation method (Varga, 1991).

3.2 RCP strategy: Improving offline innovation estimation with low-rank structure prior

The crux of our identification method lies in how to effectively estimate offline innovations $\hat{\mathbf{E}}_\mathrm{p}$ and $\hat{\mathbf{E}}_\mathrm{f}$. In Wang et al. (2025), it is suggested to estimate the innovations by fitting a VARX model with LS regression. Revisiting the approximated data equation (5), by taking $T_\mathrm{f} = 1$, one can derive the following approximation:

$$\begin{aligned}\mathcal{H}_1(\mathbf{y}_{1:N}) \approx &\ \mathbf{L}_{up}\mathcal{H}_{T_\mathrm{p}}(\mathbf{u}_{1-T_\mathrm{p}:N-1}) + \mathbf{L}_{yp}\mathcal{H}_{T_\mathrm{p}}(\mathbf{y}_{1-T_\mathrm{p}:N-1})\\ &+ \mathbf{D}\mathcal{H}_1(\mathbf{u}_{1:N}) + \mathcal{H}_1(\mathbf{e}_{1:N}),\end{aligned} \tag{17}$$

based on which $\mathbf{e}_{1:N}$ can be conveniently estimated with LS regression.

However, as the LS strategy relies on data equation (5), it requires a sufficiently large past window size T_p rendering $\bar{\mathbf{A}}^{T_\mathrm{p}} \approx \mathbf{0}$, to reduce the approximation errors. Nevertheless, large T_p will inevitably lead to high model complexity and overfitting. This dilemma limits the performance of the LS strategy. To mitigate this, we propose an improved offline innovation estimation strategy that leverages the low-rank structure prior of dynamical systems, named the rank-constrained programming (RCP) strategy.

By iterating the state equation and output equation of the observer form model (3), we arrive at the following data equation:

$$\mathbf{Y} = \bar{\mathbf{O}}\hat{\mathbf{X}} + \bar{\mathbf{L}}_u\mathbf{U} + \bar{\mathbf{L}}_y\mathbf{Y} + \mathbf{E}, \tag{18}$$

where $\hat{\mathbf{X}} \triangleq \mathcal{H}_1(\hat{\mathbf{x}}_{1:N+1-L})$, $\bar{\mathbf{O}} \triangleq \mathbf{col}(\mathbf{C}, \mathbf{C}\bar{\mathbf{A}}, \ldots, \mathbf{C}\bar{\mathbf{A}}^{L-1})$. $\mathbf{L}_u$ and $\mathbf{L}_y$ are lower-triangler block-Toeplitz coefficient matrices:

$$\begin{aligned}\mathbf{L}_u &\triangleq \mathcal{T}(\mathbf{D}, \mathbf{C}\bar{\mathbf{B}}, \mathbf{C}\bar{\mathbf{A}}\bar{\mathbf{B}}, \ldots, \mathbf{C}\bar{\mathbf{A}}^{L-2}\bar{\mathbf{B}}),\\ \mathbf{L}_y &\triangleq \mathcal{T}(\mathbf{0}, \mathbf{C}\mathbf{K}, \mathbf{C}\bar{\mathbf{A}}\mathbf{K}, \ldots, \mathbf{C}\bar{\mathbf{A}}^{L-2}\mathbf{K}).\end{aligned}$$

In data equation (18), the low-rank structure prior of the process is injected by $\mathrm{rank}\{\bar{\mathbf{O}}\hat{\mathbf{X}}\} \leq r$, based on which the offline innovations can be estimated with the following RCP:

$$\begin{aligned}\min_{\mathbf{e}, \mathbf{L}_u, \mathbf{L}_y}\ &\|\mathbf{e}_{1:N}\|_2^2\\ \text{s.t.}\quad &\mathrm{rank}\{\mathbf{L}_u\mathbf{U} + (\mathbf{L}_y - \mathbf{I})\mathbf{Y} + \mathbf{E}\} \leq r\end{aligned} \tag{19}$$

Different from the LS strategy, where the unmeasurable states are approximated with past inputs and outputs, here the states are directly estimated with consideration of their low-rank structure, thus avoiding the approximation error. However, the rank constraint renders the problem (19) non-convex and NP-hard.

To tackle this challenge, we further present a tailored algorithm for solving problem (19), which is based on the alternating direction method of multipliers (ADMM). First, we relax the structural constraints on $\mathbf{L}_u$ and $\mathbf{L}_y$ for the ease of solution. By focusing exclusively on the block-lower-triangular and block-Toeplitz structure, $\mathbf{L}_u$ and $\mathbf{L}_y$ can be reparameterized as

$$\mathbf{L}_u = \mathcal{T}(\mathbf{L}_u^{(0)}, \mathbf{L}_u^{(1)}, \mathbf{L}_u^{(2)}, \dots, \mathbf{L}_u^{(L-1)}), \tag{20}$$

$$\mathbf{L}_y = \mathcal{T}(\mathbf{0}, \mathbf{L}_y^{(1)}, \mathbf{L}_y^{(2)}, \dots, \mathbf{L}_y^{(L-1)}). \tag{21}$$

Next, introducing an auxiliary variable

$$\mathbf{\Xi} \triangleq -(\mathbf{L}_u\mathbf{U} + (\mathbf{L}_y - \mathbf{I})\mathbf{Y} + \mathbf{E}),$$

and considering that $\mathbf{E} \triangleq \mathcal{H}_L(\mathbf{e}_{1:N})$, problem (19) can be transformed into

$$\begin{aligned} \min_{\mathbf{e},\mathbf{L}_u,\mathbf{L}_y,\mathbf{\Xi}} \ & \|\mathbf{e}_{1:N}\|_2^2 \\ \text{s.t.} \ & \mathbf{L}_u\mathbf{U} + \mathbf{L}_y\mathbf{Y} + \mathcal{H}_L(\mathbf{e}_{1:N}) + \mathbf{\Xi} - \mathbf{Y} = \mathbf{0} \\ & \operatorname{rank}\{\mathbf{\Xi}\} \le r \end{aligned} \tag{22}$$

Thereafter, this problem can be readily solved with the nonconvex ADMM algorithm, as detailed here. The augmented Lagrangian of problem (22) is given by

$$\mathcal{L} = \|\mathbf{e}_{1:N}\|_2^2 - \langle \mathbf{\Lambda}, \mathbf{F} \rangle + \frac{\beta}{2}\|\mathbf{F}\|_{\mathrm{F}}^2, \tag{23}$$

where $\mathbf{F} \triangleq \mathbf{L}_u\mathbf{U} + \mathbf{L}_y\mathbf{Y} + \mathcal{H}_L(\mathbf{e}_{1:N}) + \mathbf{\Xi} - \mathbf{Y}$. $\mathbf{\Lambda} \in \mathbb{R}^{nL\times(N-L+1)}$ is the dual variable, and $\beta > 0$ is the penalty. On this basis, the ADMM-based solution is summarized in Algorithm 1.

Algorithm 1 ADMM-based solution to problem (19)

Inputs: Initial penalty β^0 and maximum $\beta_{\max}$
Penalty increasing rate $\rho > 1$
Initial values of $\mathbf{e}$, $\mathbf{L}_u$ and $\mathbf{L}_y$

1: Set iteration count $k = 0$
2: **repeat**
3: $\quad \mathbf{\Xi}^{k+1} \in \arg\min_{\mathbf{\Xi}} \ \mathcal{L} \quad \text{s.t. } \operatorname{rank}\{\mathbf{\Xi}\} \le r$
4: $\quad \left\{\mathbf{e}^{k+1}, \mathbf{L}_u^{k+1}, \mathbf{L}_y^{k+1}\right\} \in \arg\min_{\mathbf{e},\mathbf{L}_u,\mathbf{L}_y} \mathcal{L}$
5: $\quad \mathbf{\Lambda}^{k+1} = \mathbf{\Lambda}^k - \beta^k\mathbf{F}$
6: $\quad \beta^{k+1} = \min\left\{\rho\beta^k, \beta_{\max}\right\}$
7: **until** convergence
8: **return** $\mathbf{e}^*$

In addition, the two subproblems for updating $\mathbf{\Xi}$ and $\{\mathbf{e}, \mathbf{L}_u, \mathbf{L}_y\}$ in Algorithm 1 both admit concise closed-form solutions, as detailed below.

For updating $\mathbf{\Xi}$, the optimization objective $\mathcal{L}$ is a quadratic form with respect to $\mathbf{\Xi}$:

$$\mathcal{L} = \frac{\beta}{2}\left\|\mathbf{\Xi} + \mathbf{L}_u\mathbf{U} + \mathbf{L}_y\mathbf{Y} + \mathcal{H}_L(\mathbf{e}_{1:N}) - \mathbf{Y} - \frac{\mathbf{\Lambda}^k}{\beta^k}\right\|_{\mathrm{F}}^2 + \mathcal{C}, \tag{24}$$

where $\mathcal{C}$ contains all terms independent of $\mathbf{\Xi}$. Hence, the optimal solution is given by

$$\mathbf{\Xi}^{k+1} = \mathcal{U}^k\widetilde{\mathbf{\Sigma}}^k(\mathcal{V}^k)^{\mathrm{T}}, \tag{25}$$

where $\mathcal{U}^k$, $\mathbf{\Sigma}^k$ and $\mathcal{V}^k$ stem from the following SVD:

$$\mathcal{U}^k\mathbf{\Sigma}^k(\mathcal{V}^k)^{\mathrm{T}} = -\mathbf{L}_u\mathbf{U} - \mathbf{L}_y\mathbf{Y} - \mathcal{H}_L(\mathbf{e}_{1:N}) + \mathbf{Y} + \frac{\mathbf{\Lambda}^k}{\beta^k}, \tag{26}$$

and $\widetilde{\mathbf{\Sigma}}^k$ is the truncated version of singular value matrix $\mathbf{\Sigma}^k$, with the top r singular values retained, but the remaining ones set to 0.

For updating $\{\mathbf{e}, \mathbf{L}_u, \mathbf{L}_y\}$, it can be easily observed that the optimization problem is an unconstrained convex quadratic program on $\{\mathbf{e}, \mathbf{L}_u, \mathbf{L}_y\}$. Thus, the derivation of its optimal solution is straightforward.

Remark 1. In industrial practice, many processes are operated in closed-loop. As a result, the collected data exhibits correlation between the process input and past disturbances, which is a key challenge in subspace identification. The proposed identification method is applicable to such closed-loop scenarios because both the innovation-based DDOP and the RCP formulation (19) do not hinge on the uncorrelation between the input and the past innovation.

4. NUMERICAL EXAMPLES

In this section, we perform numerical studies on the following process to investigate the performance of the proposed identification method:

$$\begin{cases} \mathbf{x}(t+1) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) + \mathbf{w}(t) \\ \mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t) + \mathbf{v}(t) \end{cases}$$

where

$$\mathbf{A} = \begin{bmatrix} 0.98 & 0 & 0 \\ 0 & 0.92 & 0 \\ 0 & 0 & 0.85 \end{bmatrix}, \ \mathbf{B} = \begin{bmatrix} 0.2 & 0.06 \\ 0.2 & 0.3 \\ 0.2 & -0.16 \end{bmatrix},$$

$$\mathbf{C} = \begin{bmatrix} 0.1 & 0.1 & 0.1 \\ 0.02 & 0.1 & -0.05 \end{bmatrix}, \ \mathbf{D} = \mathbf{0},$$

$$\mathbf{w}(t) \sim \mathcal{N}(\mathbf{0}, \mathcal{A}\cdot\mathbf{I}), \ \mathbf{v}(t) \sim \mathcal{N}(\mathbf{0}, \mathcal{A}\cdot\mathbf{I}).$$

Here, $\mathcal{A} > 0$ is an amplitude coefficient for regulating the noise level. We carry out both open-loop and closed-loop identification experiments on this process.

4.1 Open-loop experiments

Open-loop offline tests are carried on the process with a total duration of $N = 1000$ s, and the input-output data $\{\mathbf{u}(t), \mathbf{y}(t)\}_{t=1}^N$ are collected at a sampling interval of $\Delta t = 1$ s for identification. To ensure persistent excitation, the input signal $\{\mathbf{u}(t)\}_{t=1}^N$ is designed as an M-sequence consisting of 1 and -1. By regulating the noise amplitude coefficient $\mathcal{A}$, three different noise levels (NSR $= 0.13, 0.30, 0.65$) are created for a comprehensive comparison.

Under every noise level, four different methods are adopted to identify the process, including two classic subspace identification methods and the proposed identification scheme implemented with two innovation estimation strategies:

- N4SID: A popular open-loop subspace identification method (Van Overschee and De Moor, 1994)
- PBSID: A popular closed-loop subspace identification method (Houtzager et al., 2009)
- LS: The proposed method with innovations estimated using LS

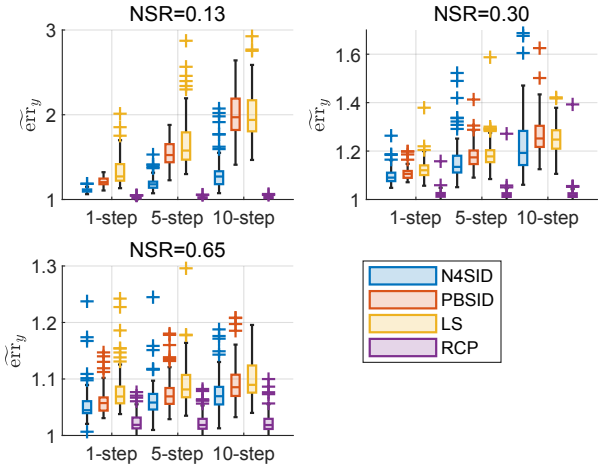


Fig. 1. Relative prediction error of different methods in the open-loop identification experiments

- RCP: The proposed method with innovations estimated using the RCP strategy

where N4SID is implemented with `n4sid` routine in MATLAB 2024b, PBSID is implemented with PBSID Toolbox (Houtzager et al., 2009), and the balanced truncation method is implemented with `balreal` and `xelim` routine. The hyperparameters are selected as $r = 3$, $L_\mathrm{p} = 4$, $T_\mathrm{p} = 7$, $L = 50$, via cross-validation to ensure the optimal performance of each method.

Upon the identification completed, another online test lasting $N^\mathrm{on} = 1000$ s is carried out on the process, where the input follows the standard norm distribution, i.e., $\mathbf{u}^\mathrm{on}(t) \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, and the noise follows the same distribution as in the offline test. The identification performance is evaluated with the online prediction error under three different prediction steps (1-step, 5-step, 10-step). The online prediction error err_y is defined by

$$\mathrm{err}_y = \frac{1}{N^\mathrm{on}} \sum_{t=1}^{N^\mathrm{on}} \|\mathbf{y}^\mathrm{on}(t) - \hat{\mathbf{y}}^\mathrm{on}(t)\|_2^2, \tag{27}$$

where $\mathbf{y}^\mathrm{on}(t)$ is the true output, and $\hat{\mathbf{y}}^\mathrm{on}(t)$ is the output predicted by the Kalman filter constructed with the identified model. For a more intuitive exhibition of identification performance, the prediction error of the true model err_y^0 is utilized as a reference, based on which the relative prediction error $\widetilde{\mathrm{err}}_y = \mathrm{err}_y/\mathrm{err}_y^0$ is calculated. This normalized metric is always lower-bounded by 1, with a lower value signifying better performance.

To evaluate the overall identification performance, 100 Monte Carlo tests are carried out under every noise level. The relative prediction error $\widetilde{\mathrm{err}}_y$ of different identification methods are depicted in Fig. 1. It can be clearly observed from the result that RCP outperforms two classic subspace identification methods, N4SID and PBSID, under all noise levels and prediction steps. This evidences the effectiveness of the proposed state-space model identification scheme based on behavioral systems theory. However, it can also be observed that LS exhibits even slightly worse performance than classic subspace identification methods, which is attributed to the poor offline innovation estimation accuracy of the LS strategy.

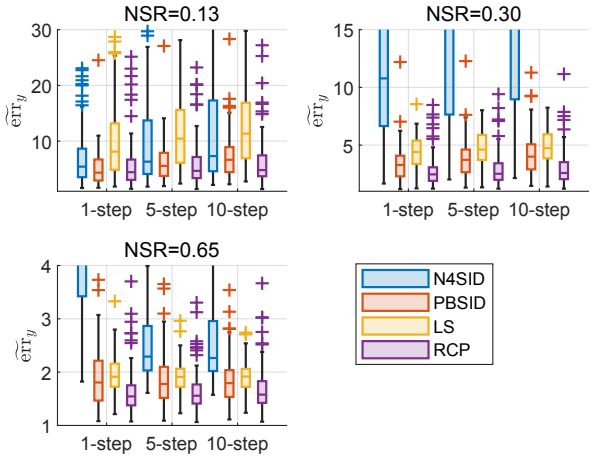


Fig. 2. Relative prediction error of different methods in the closed-loop identification experiments

4.2 Closed-loop experiments

To carry out closed-loop identification, we construct a closed-loop system by conducting a pole placement on the same process. By setting the desired closed-loop poles to $[0.1, 0.01, 0.01]$, the state-feedback matrix $\mathbf{K}_f$ is computed, and the system feedback is designed as $\mathbf{u}(t) = \mathbf{u}'(t) - \mathbf{K}_f\hat{\mathbf{x}}(t)$, where $\mathbf{u}(t)$ and $\mathbf{u}'(t)$ are the open-loop and closed-loop input of the process, respectively. On this basis, closed-loop offline tests lasting $N = 1000$ s are carried out on the process, and the open-loop input-output data $\{\mathbf{u}(t), \mathbf{y}(t)\}_{t=1}^N$ are collected at a sampling interval of $\Delta t = 1$ s for identification. The closed-loop input $\{\mathbf{u}'(t)\}_{t=1}^N$ is designed as an M-sequence consisting of 1 and -1 for persistent excitation. The remaining experimental setup follows that of the open-loop identification experiments, including the evaluated identification methods, identification performance metric, noise level setup, hyperparameter setup, and the number of Monte Carlo tests.

The relative prediction error $\widetilde{\mathrm{err}}_y$ of different identification methods are depicted in Fig. 2. Notice that different from the previous open-loop case where N4SID and PSBID achieve comparable performance, here the open-loop N4SID performs unsatisfactorily, especially under high noise level. By contrast, it is seen that RCP retains its performance advantage over N4SID, PBSID, and LS. This highlights that the proposed identification scheme and the improved RCP strategy remain effective in the closed-loop identification scenario.

5. EXPERIMENTS ON INDUSTRIAL DATASET

Next, we carry out case studies based on the following two industrial benchmark datasets in DaISy (Database for the Identification of Systems, De Moor et al. (1997)), including a glass furnace process (Van Overschee and De Moor, 1994) and an industrial evaporator process (Zhu et al., 1994). For both systems, a segment of historical data containing 500 samples is selected as the training dataset for identification, and another segment with 500 samples is utilized as the test dataset. The same four identification methods as in the numerical experiments are adopted to identify the processes, and the hyperparameter values are

Table 1. Hyperparameters in The Industrial Dataset Experiments

	r	L_{p}	T_{p}	L
Glass furnace example	6	3	50	4
Industrial evaporator example	7	7	50	4

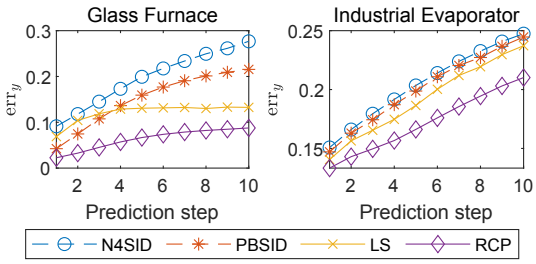


Fig. 3. Prediction error of different methods in the industrial dataset identification experiments

determined via cross-validation, as detailed in Table 1. The output prediction error err_y on the test dataset, as defined in (27), is utilized to evaluate the identification performance. For a more comprehensive comparison, the 1- to 10-step prediction errors of each method are recorded.

The test results are depicted in Fig. 3. It is seen that in both examples, the proposed RCP method achieves superior prediction accuracy as compared to the classic N4SID and PBSID methods under all settings, which highlights the effectiveness of the proposed identification scheme in real-world industrial scenarios. Also, RCP yields much better prediction accuracy than LS, which highlights the effectiveness and necessity of an improved offline innovation estimation strategy for handling complicated industrial processes.

6. CONCLUSION

In this work, a novel state-space model identification method was proposed from a behavioral systems theory perspective. By exploiting the equivalent state-space representation of the innovation-based DDOP, the state-space model identification problem was converted into an innovation estimation problem. An improved innovation estimation strategy incorporating the low-rank prior, as well as a tailored ADMM-based solution, was subsequently proposed to achieve better innovation estimation accuracy. Numerical and industrial dataset experiments demonstrated the superior identification accuracy of the proposed method over classic subspace identification methods under both open-loop and closed-loop cases. Also, the proposed innovation estimation strategy was evidenced to be able to achieve significantly better innovation estimation accuracy than existing LS strategy, showing the effectiveness and necessity of the proposed improvement strategy for innovation estimation.

REFERENCES

Breschi, V., Chiuso, A., and Formentin, S. (2023). Data-driven predictive control in a stochastic setting: a unified framework. *Automatica*, 152, 110961.

Coulson, J., Lygeros, J., and Dörfler, F. (2019). Data-enabled predictive control: In the shallows of the deepc. In *2019 18th European Control Conference (ECC)*, 307–312. IEEE.

De Moor, B., De Gersem, P., De Schutter, B., Favoreel, W., et al. (1997). DAISY: A database for identification of systems. *Journal A*, 38(3), 4–5.

Dörfler, F., Coulson, J., and Markovsky, I. (2022). Bridging direct and indirect data-driven control formulations via regularizations and relaxations. *IEEE Transactions on Automatic Control*, 68(2), 883–897.

Hou, Z.S. and Wang, Z. (2013). From model-based control to data-driven control: Survey, classification and perspective. *Information Sciences*, 235, 3–35.

Houtzager, I., van Wingerden, J.W., and Verhaegen, M. (2009). VARMAX-based closed-loop subspace model identification. In *Proceedings of the 48h IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference*, 3370–3375. IEEE.

Huang, B. and Kadali, R. (2008). *Dynamic Modeling, Predictive Control and Performance Monitoring: A Data-Driven Subspace Approach*. Springer.

Liu, Q., Shang, C., and Huang, D. (2021). Efficient low-order system identification from low-quality step response data with rank-constrained optimization. *Control Engineering Practice*, 107, 104671.

Liu, Q., Shang, C., Liu, T., and Huang, D. (2023). Efficient relay autotuner of industrial controllers via rank-constrained identification of low-order time-delay models. *IEEE Transactions on Control Systems Technology*, 31(4), 1787–1802.

Markovsky, I. and Dörfler, F. (2021). Behavioral systems theory in data-driven analysis, signal processing, and control. *Annual Reviews in Control*, 52, 42–64.

Van Overschee, P. and De Moor, B. (1994). N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica*, 30(1), 75–93.

Van Overschee, P. and De Moor, B. (1995). A unifying theorem for three subspace system identification algorithms. *Automatica*, 31(12), 1853–1864.

Varga, A. (1991). Balancing free square-root algorithm for computing singular perturbation approximations. In *[1991] Proceedings of the 30th IEEE Conference on Decision and Control*, 1062–1065. IEEE.

Verhaegen, M. (2021). Subspace techniques in system identification. In *Encyclopedia of Systems and Control*, 2235–2244. Springer.

Verhaegen, M. and Verdult, V. (2007). *Filtering and system identification: a least squares approach*. Cambridge university press.

Wang, Y., You, K., Huang, D., and Shang, C. (2025). Data-driven output prediction and control of stochastic systems: An innovation-based approach. *Automatica*, 171, 111897.

Willems, J.C., Rapisarda, P., Markovsky, I., and De Moor, B.L. (2005). A note on persistency of excitation. *Systems & Control Letters*, 54(4), 325–329.

Zhu, Y., van Overschee, P., de Moor, B., and Ljung, L. (1994). Comparison of three classes of identification methods. *IFAC Proceedings Volumes*, 27(8), 169–174.