SCIENCE CHINA
Information Sciences

• LETTER •

August 2026, Vol. 69, Iss. 8, 189203:1–189203:2
https://doi.org/10.1007/s11432-025-4784-3

Robust identification of multi-model systems with a synergistic reweighted expectation algorithm based on Krylov subspace

Ronghuan LI[1], Yujie MA[3], Junxia MA[1,2*] & Weili XIONG[1,2]

[1]*School of Automation and Intelligent Science (School of IOT), Jiangnan University, Wuxi 214122, China*
[2]*Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Jiangnan University, Wuxi 214122, China*
[3]*Department of Electrical and Electronic Engineering, The University of Manchester, Manchester M13 9PL, UK*

Received 19 August 2025/Revised 5 November 2025/Accepted 24 November 2025/Published online 6 May 2026

Citation Li R H, Ma Y J, Ma J X, et al. Robust identification of multi-model systems with a synergistic reweighted expectation algorithm based on Krylov subspace. Sci China Inf Sci, 2026, 69(8): 189203, https://doi.org/10.1007/s11432-025-4784-3

Data-driven system identification plays a pivotal role in process control [1]. The linear parameter-varying (LPV) model, as a multi-model framework, alleviates the inability of a single model to capture nonlinear industrial dynamics across wide operating ranges. However, conventional expectation-based LPV identification suffers from two major issues: excessive computational cost due to matrix inversions and biased estimation under data outliers. To address these limitations, this study proposes a synergistic reweighted expectation (SRE) algorithm that integrates a Krylov subspace-based parameterization to avoid matrix inversion with an adaptive reweighting scheme for outlier suppression. This synergistic design jointly ensures efficiency and robustness. The proposed method is validated on two benchmark processes: a mass-spring-damper (MSD) system and a continuous stirred-tank reactor (CSTR). Key notations are summarized in Appendix A.

Problem formulation. Consider a multi-model LPV system with M local auto-regressive with exogenous input (ARX) sub-models. The global output y_k at time k is given by

$$y_k = \sum_{m=1}^{M} \zeta_{m,w_k}(\boldsymbol{\phi}_k^{\mathrm{T}}\boldsymbol{\theta}_m) + v_k,$$

where $\boldsymbol{\phi}_k$ is the information vector, $\boldsymbol{\theta}_m$ is the parameter vector of the m-th sub-model, and v_k is noise. The weighting function ζ_{m,w_k}, governed by the scheduling variable w_k, ensures a smooth transition between operating points. The complete definitions of $\boldsymbol{\phi}_k$, $\boldsymbol{\theta}_m$, and ζ_{m,w_k} are provided in Appendix A.

Let N denote the number of samples with input $U = \{u_k\}_{k=1}^N$ and output $Y = \{y_k\}_{k=1}^N$. $\mathcal{H} = \{\mathcal{H}_m\}_{m=1,\cdots,M}$ denotes the set of predesigned operating points. The sub-model identity (SMI), denoted by $\mathcal{I} = \{\mathcal{I}_k\}_{k=1}^N$, is a hidden variable and cannot be directly estimated. The observation set is $C_{obs} = \{U, Y, \mathcal{H}\}$, and the hidden set is $C_{hid} = \{\mathcal{I}\}$. Model parameters are denoted by $\boldsymbol{\theta} = \{\boldsymbol{\theta}_m\}_{m=1}^M$ with $\hat{\cdot}^s$ as their s-th iteration estimates.

To estimate both the SMI and model parameters, the expectation-maximization (EM) algorithm is employed, which iteratively maximizes a $\mathbb{Q}$-function [2]. The parameter update in the M-step as

$$\hat{\boldsymbol{\theta}}_m^{s+1} = \left[\sum_{k=1}^{N} \eta_{(k,m)}\boldsymbol{\phi}_k\boldsymbol{\phi}_k^{\mathrm{T}}\right]^{-1} \sum_{k=1}^{N} \eta_{(k,m)}\boldsymbol{\phi}_k y_k. \tag{1}$$

The posterior probability $\eta_{(k,m)}$ and the derivation of (1) from the $\mathbb{Q}$-function are detailed in Appendix A.

Remark 1. The parameter estimates are obtained using the EM algorithm. However, the matrix inversion introduces computational costs (see (1)), and outliers degrade modeling accuracy. Hence, the SRE is proposed.

Synergistic reweighted expectation algorithm. $\imath_k$ denotes the output outlier. Let $\hat{\imath}_k$ denote the estimate of $\imath_k$, and define $\hat{\boldsymbol{o}} := \left[\hat{\imath}_1, \hat{\imath}_2, \ldots, \hat{\imath}_N\right]$. To estimate parameters $\boldsymbol{\theta}_m$, we formulate a cost function that penalizes model residual and outlier sparsity via an ℓ_0 pseudo-norm:

$$L(\hat{\boldsymbol{\theta}}^s, \hat{\imath}_k) = \sum_{k=1}^{N} \left\{ \left(y_k - \sum_{m=1}^{M} \eta_{(k,m)} y_{m,k} - \hat{\imath}_k\right)^2 + \lambda \|\hat{\boldsymbol{o}}\|_0 \right\}, \tag{2}$$

where $\|\cdot\|_0$ denotes the ℓ_0 pseudo norm, λ is a design factor to control the sparsity of $\hat{\boldsymbol{o}}$. Since Eq. (2) is NP-hard [3], an approximate solution (3) is obtained through convex relaxation:

$$\mathcal{L}(\hat{\boldsymbol{\theta}}^s, \hat{\imath}_k) = \sum_{k=1}^{N} \left\{ \left(y_k - \sum_{m=1}^{M} \eta_{(k,m)} y_{m,k} - \hat{\imath}_k\right)^2 + \lambda |\hat{\imath}_k| \right\}. \tag{3}$$

By setting the derivative of (3) with respect to $\hat{\imath}_k$ to zero and applying the soft-thresholding principle, the outlier estimate can be obtained (see Appendix B for derivation).

Remark 2. The design factor λ balances outlier detection and noise suppression. λ is adaptively updated using output errors, with initialization following the 3σ rule.

An adaptive reweighting mechanism feedback loop is used between outlier detection and parameter update, avoiding matrix

* Corresponding author (email: jxma@jiangnan.edu.cn)

© Science China Press 2026

inversion. The cost function is optimized as

$$\begin{aligned}\mathbb{F}(\boldsymbol{\theta}_m) &= \frac{1}{2}\left\{\sum_{i=1}^{N_o}\gamma_i\eta_{(k,m)}\left[y_i-\boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{\theta}_m-\hat{\imath}_i\right]^2\right.\\ &\quad\left.+\sum_{j=1}^{N_n}\gamma_j\eta_{(k,m)}\left[y_j-\boldsymbol{\phi}_j^{\mathrm{T}}\boldsymbol{\theta}_m\right]^2\right\}\\ &= \frac{1}{2}\sum_{k=1}^{N}\gamma_k\eta_{(k,m)}\left[y_k-\boldsymbol{\phi}_k^{\mathrm{T}}\boldsymbol{\theta}_m-\hat{\imath}_k\right]^2, \end{aligned} \tag{4}$$

where N_o, N_n denote the sets of outliers and normal samples, γ is a positive weight used to vary the output errors over time given as $\gamma_k = \frac{1}{\tau+\beta|\hat{\imath}_k|}$, τ controls the maximum weight, and β is the weight factor.

The parameter can be updated by gradient descent calculated by the partial derivative of (4) (see Appendix B for full derivation):

$$\begin{aligned}\boldsymbol{\theta}_m^{s+1} &= \left[\boldsymbol{I}-\alpha\sum_{k=1}^{N}\gamma_k\eta_{(k,m)}\boldsymbol{\phi}_k\boldsymbol{\phi}_k^{\mathrm{T}}\right]\boldsymbol{\theta}_m^s\\ &\quad+\alpha\sum_{k=1}^{N}\gamma_k\eta_{(k,m)}\boldsymbol{\phi}_k\left(y_k-\hat{\imath}_k\right), \end{aligned} \tag{5}$$

where $\boldsymbol{I}$ denotes the identity matrix, and α is the step size.

Remark 3. Although the update (5) avoids matrix inversion, it requires computing the largest eigenvalue to ensure convergence (see Appendix B for detailed derivation), which remains challenging in high dimensions.

Based on the above analysis, the SRE algorithm is introduced with an Arnoldi-based optimization. Instead of computing the full gradient, the update is constrained to a low-dimensional Krylov subspace, requiring only $l < (n_{\mathrm{a}}+n_{\mathrm{b}})$ directions per iteration. Consider the Krylov subspace $\bar{\boldsymbol{\chi}}_{m,l}^s = \mathrm{span}\{\bar{\boldsymbol{\chi}}_{m,1}^s, Q_m\bar{\boldsymbol{\chi}}_{m,1}^s, \cdots, (Q_m)^{l-1}\bar{\boldsymbol{\chi}}_{m,1}^s\}$, $Q_m \in \mathbb{R}^{(n_a+n_b)\times(n_a+n_b)}$ is nonsingular, $\bar{\boldsymbol{\chi}}_{m,1}^s \in \mathbb{R}^{(n_a+n_b)}$ with $\|\bar{\boldsymbol{\chi}}_{m,1}^s\| = 1$. The orthonormal basis is given as $\{\bar{\boldsymbol{\chi}}_{m,1}^s, \bar{\boldsymbol{\chi}}_{m,2}^s, \cdots, \bar{\boldsymbol{\chi}}_{m,l}^s\}$ and the detailed steps for generating it are presented in Appendix C.

The process begins by projecting the optimization problem onto a Krylov subspace. Let $Q_m^s = \sum_{k=1}^{N}\gamma_k\eta_{(k,m)}\boldsymbol{\phi}_k\boldsymbol{\phi}_k^{\mathrm{T}}$ be the curvature matrix, and initialize with the normalized gradient direction:

$$\begin{aligned}\boldsymbol{\chi}_{m,1}^s &= \sum_{k=1}^{N}\gamma_k\eta_{(k,m)}\boldsymbol{\phi}_k\left[y_k-\boldsymbol{\phi}_k^{\mathrm{T}}\boldsymbol{\theta}_m^s-\hat{\imath}_k\right],\\ \bar{\boldsymbol{\chi}}_{m,1}^s &= \frac{\boldsymbol{\chi}_{m,1}^s}{\|\boldsymbol{\chi}_{m,1}^s\|} = \frac{\boldsymbol{\chi}_{m,1}^s}{\xi^s}. \end{aligned} \tag{6}$$

An orthonormal basis ${}^l\bar{\boldsymbol{\chi}}_m^s = [\bar{\boldsymbol{\chi}}_{m,1}^s, \bar{\boldsymbol{\chi}}_{m,2}^s, \cdots, \bar{\boldsymbol{\chi}}_{m,l}^s]$ is constructed through the Arnoldi process [4, 5] (detailed procedure in Appendix C), ensuring $[\bar{\boldsymbol{\chi}}_{m,i}^s]^{\mathrm{T}}\bar{\boldsymbol{\chi}}_{m,j}^s = 0$ for $i \neq j$.

The model parameters are updated within this low-dimensional subspace:

$$\boldsymbol{\theta}_m^{s+1} = \boldsymbol{\theta}_m^s + {}^l\bar{\boldsymbol{\chi}}_m^s\upsilon^s, \tag{7}$$

where υ^s is obtained by solving a low-dimensional optimization problem. The Arnoldi process yields the relationship:

$$Q_m^{s\ l}\bar{\boldsymbol{\chi}}_m^s = {}^{l+1}\bar{\boldsymbol{\chi}}_m^s\boldsymbol{P}^s,$$

where $\boldsymbol{P}^s \in \mathbb{R}^{(l+1)\times l}$ is an upper Hessenberg matrix. This allows us to transform the gradient norm minimization into a low-dimensional problem:

$$\begin{aligned}&\left\|\sum_{k=1}^{N}\gamma_k\eta_{(k,m)}\boldsymbol{\phi}_k\left[y_k-\boldsymbol{\phi}_k^{\mathrm{T}}\boldsymbol{\theta}_m^{s+1}-\hat{\imath}_k\right]\right\|\\ &\quad= \left\|\xi^{s+1}\boldsymbol{r}_{l+1,1}-\boldsymbol{P}^s\upsilon^s\right\|,\\ &\boldsymbol{r}_{l+1,1} = [1,0,\ldots,0]^{\mathrm{T}} \in \mathbb{R}^{l+1}. \end{aligned} \tag{8}$$

The optimal step size υ^s is obtained by solving the low-dimensional least-squares problem in (8). Through QR decomposition of $\boldsymbol{P}^s$ and appropriate orthogonal transformation (see Appendix C for detailed derivation), we obtain the closed-form solution:

$$\upsilon^s = (\tilde{\boldsymbol{P}}^s)^{-1}\boldsymbol{\mu}, \tag{9}$$

where $\tilde{\boldsymbol{P}}^s$ is an upper triangular matrix from the QR decomposition, and $\boldsymbol{\mu}$ is derived from the transformed residual.

Theorem 1. With the parameter update in (7) and the optimal step size υ^s from (9), the gradient norm does not increase.

See Appendix C for the complete proof with detailed mathematical derivations.

Experiments and conclusion. The proposed algorithm is evaluated on MSD and CSTR systems. The results demonstrate that the proposed algorithm achieves model identification and robust parameter estimation in both cases, with detailed experimental setups and comparative analyses provided in Appendix D.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 61803183, 61773182) and Natural Science Foundation of Jiangsu Province (Grant No. BK20180591).

Supporting information Appendixes A–D. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Yu Y, Xu C-Z, Pei H-L, et al. State and parameter identification of linearized water wave equation via adjoint method. Sci China Inf Sci, 2024, 67: 202205
2 Liu X, Hai Y, Dai W. Robust LPV system identification with skewed and asymmetric measurement noise. IEEE Trans Automat Sci Eng, 2025, 22: 4988–4999
3 Mao Y, Xu C, Chen J. Regularization based reweighted estimation algorithms for nonlinear systems in presence of outliers. Nonlinear Dyn, 2024, 112: 13131–13146
4 Chen J, Gao J, Liu Y, et al. A comprehensive expectation identification framework for multirate time-delayed systems. IEEE Trans Ind Inf, 2022, 19: 7011–7021
5 Saad Y, Schultz M H. GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems. SIAM J Sci Stat Comput, 1986, 7: 856–869