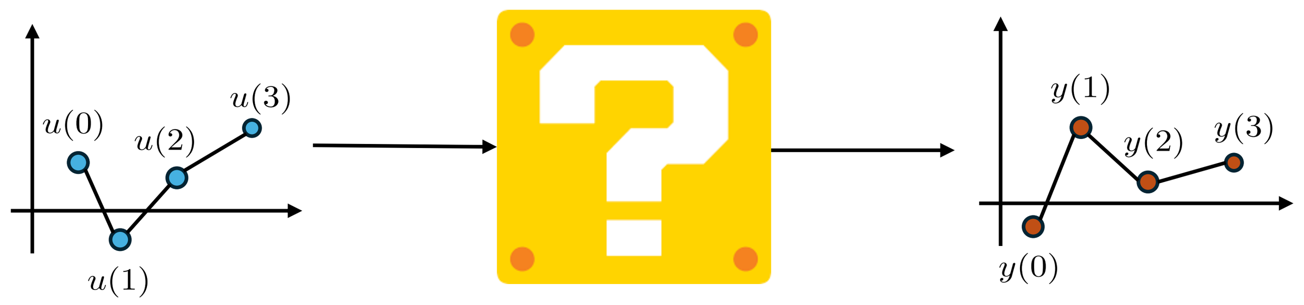
Online Subspace Learning on Flag Manifolds for System Identification

Department of Electrical and Computer Engineering
UNIVERSITY OF WISCONSIN–MADISON

Dian Jin, Jeremy Coulson
University of Wisconsin-Madison

System Identification Via Subspaces



$$\mathrm{im}\begin{bmatrix} u(0) & u(1) & \cdots & u(T-L) \\ \vdots & \vdots & & \vdots \\ u(L-1) & u(L) & \cdots & u(T-1) \\ y(0) & y(1) & \cdots & y(T-L) \\ \cdots & \vdots & & \vdots \\ y(L-1) & y(L) & \cdots & y(T-1) \end{bmatrix} = \underbrace{\text{Set of all L-length trajectories}}_{\mathcal{B}_L}$$

For LTI systems, finite-length input-output trajectories form a linear behavior subspace. Willems' Fundamental Lemma [1] gives a data-driven basis for this subspace.

Problem:
What if the system changes over time?
- Dynamics may drift
- System dimension / complexity may change

From LTI to LTV Behavior

LTI (Linear Time-Invariant) behavior: one fixed subspace
LTV (Linear Time-Varying) behavior: subspaces changing with time
We model such subspaces as:

$$\mathbf{U}^t \in \bigcup_{j=1}^{d} \mathrm{Gr}(p, q_j), \quad 0 < q_1 < \cdots < q_d < p.$$

Here $\mathrm{Gr}(p, q_j)$ is the set of all q_j -dimensional linear subspaces of $\mathbb{R}^p$, called the **Grassmannian.**

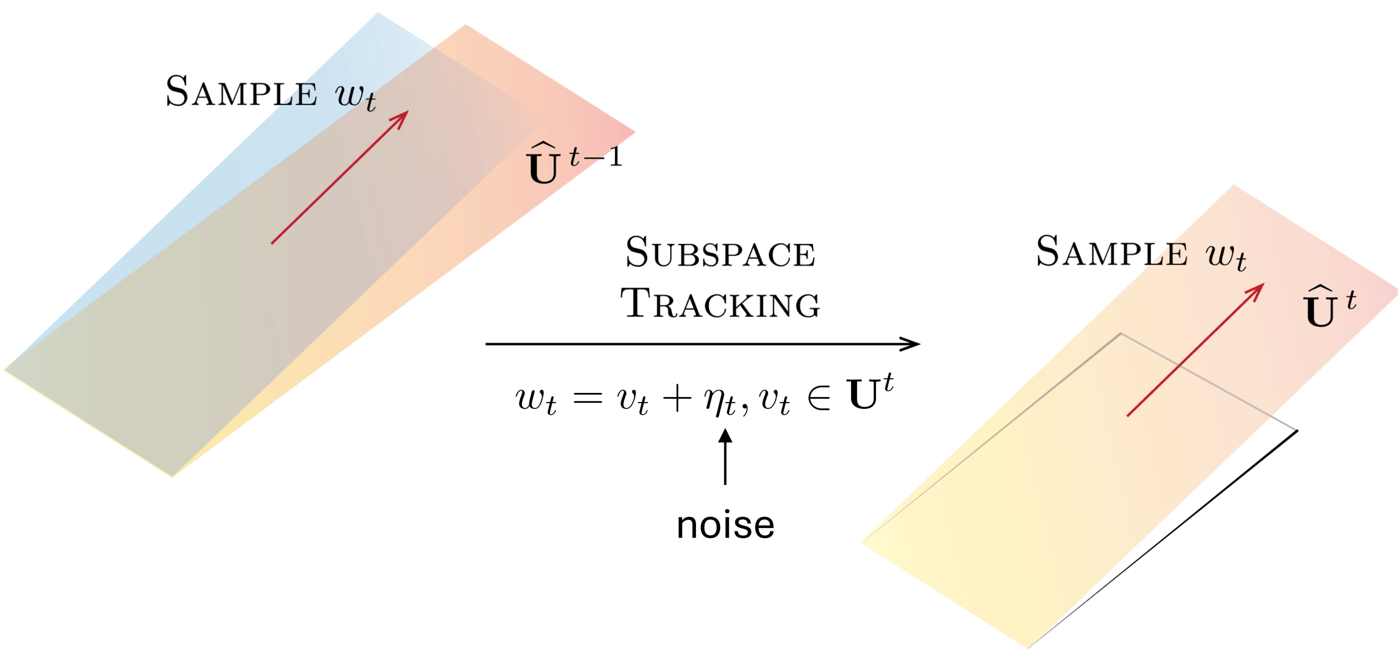


Fig 1. Subspace tracking: identify the subspace from streaming data

Key Challenge:
The subspace changes, and its dimension may also change (only an upper bound q_d of the subspace dimension is known)

Geometric Tool: Flag Manifolds

A **flag** (denoted by $\mathbf{U}_{1:d}$) is a hierarchy of nested subspaces $\{\mathbf{U}_j\}_{j=1}^d$ with increasing dimensions:

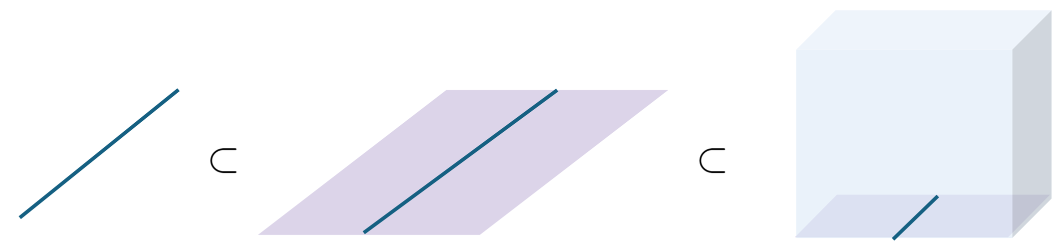
$$\{0\} \subset \mathbf{U}_1 \subset \mathbf{U}_2 \subset \cdots \subset \mathbf{U}_d \subset \mathbb{R}^p, \dim(\mathbf{U}_i) = q_i.$$

We call $(q_1, \ldots, q_d)$ the **signature.** The set of all flags in $\mathbb{R}^p$ with signature $(q_1, \ldots, q_d)$ is called the **flag manifold**, denoted by $\mathrm{Flag}(p, (q_1, \ldots, q_d))$.

For computation, a flag is represented by an orthonormal matrix whose column prefixes span the nested subspaces:

$$U = [u_1 \ \ldots \ u_{q_d}], \qquad \mathbf{U}_i = \mathrm{span}\{u_1, \ldots, u_{q_i}\}.$$

Example $[e_1 \ e_2 \ e_3] \Rightarrow \mathrm{span}\{e_1\} \subset \mathrm{span}\{e_1, e_2\} \subset \mathrm{span}\{e_1, e_2, e_3\}.$



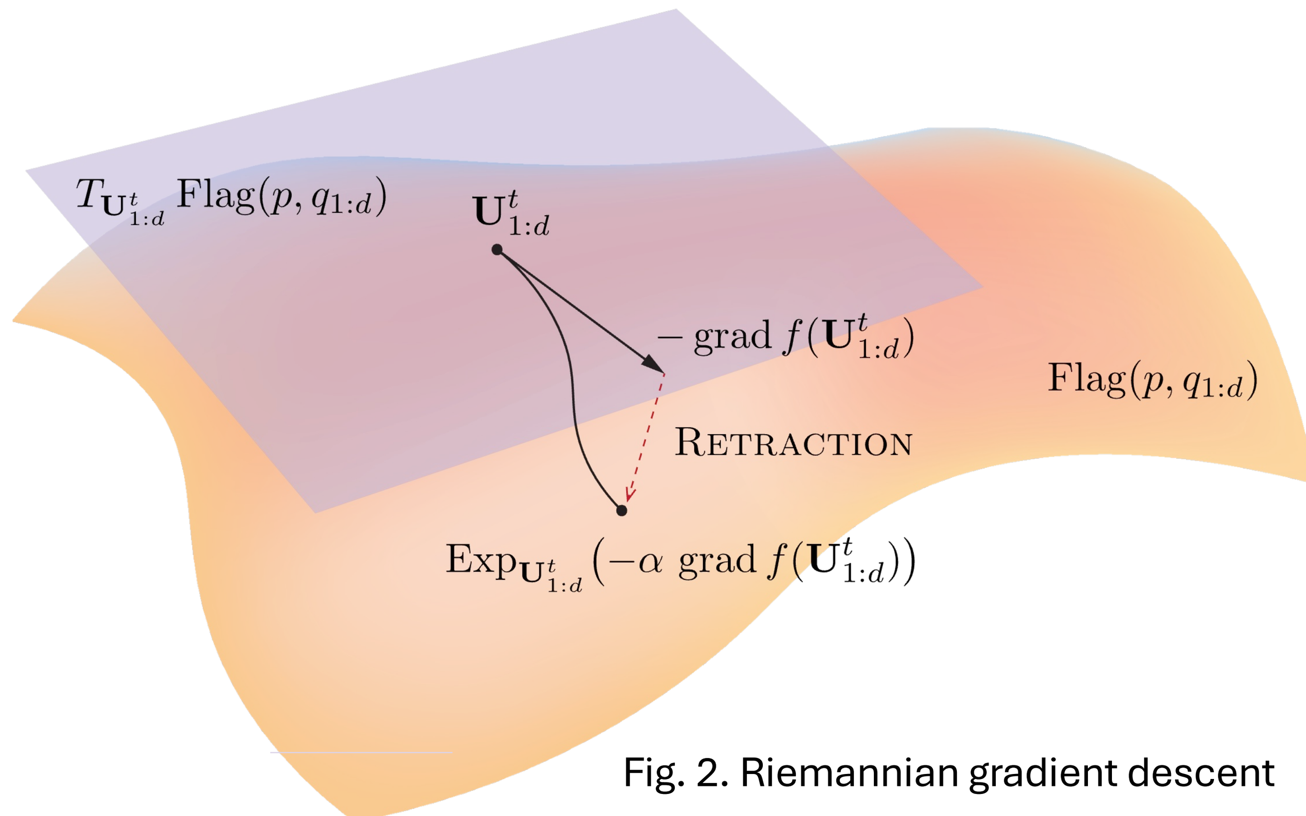


Fig. 2. Riemannian gradient descent

Algorithm: Flag Recursive ONline Tracking (FRONT)

1. Form recent data window

$$W_t = [w_{t-T+1} \ \cdots \ w_t].$$

2. Minimize "flag trick" [2] projection error

$$f_{W_t}(\mathbf{U}^t_{1:d}) = \left\| W_t - \frac{1}{d}\sum_{i=1}^{d} \Pi_{\mathbf{U}^t_i} W_t \right\|_F^2.$$

3. Update the flag by Riemannian gradient descent [4]

$$\hat{\mathbf{U}}^{t,k+1}_{1:d} = \mathrm{Exp}_{\hat{\mathbf{U}}^{t,k}_{1:d}}\left(-\alpha_t \operatorname{grad} f_{W_t}(\hat{\mathbf{U}}^{t,k}_{1:d})\right).$$

Repeat K gradient steps, then set

$$\hat{\mathbf{U}}^{t+1}_{1:d} = \hat{\mathbf{U}}^{t,K}_{1:d}.$$

Theoretical Analysis
When d=1, $\mathrm{Flag}(p,(q)) = \mathrm{Gr}(p,q)$, so FRONT reduces to GREAT [3]. Therefore, existing GREAT convergence guarantees [Theorem 1, 3] cover the fixed-rank special case of FRONT.

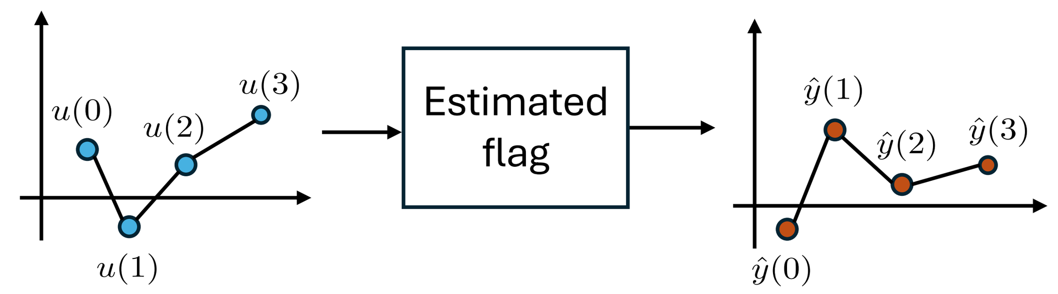
Numerical Study

Setup: a switched system

$$\begin{cases} y_t = 0.3y_{t-1} - 0.02y_{t-2} + 0.6u_{t-1} + 0.2u_{t-2}, & t < T_{\text{switch}}, \\ y_t = 1.5y_{t-1} - 0.74y_{t-2} + 0.12y_{t-3} + 0.6u_{t-1} + 0.2u_{t-2} + 0.05u_{t-3}, & t \geq T_{\text{switch}}, \end{cases}$$

Given:
- Past input-output trajectory
- Future input
- Learned subspace / flag

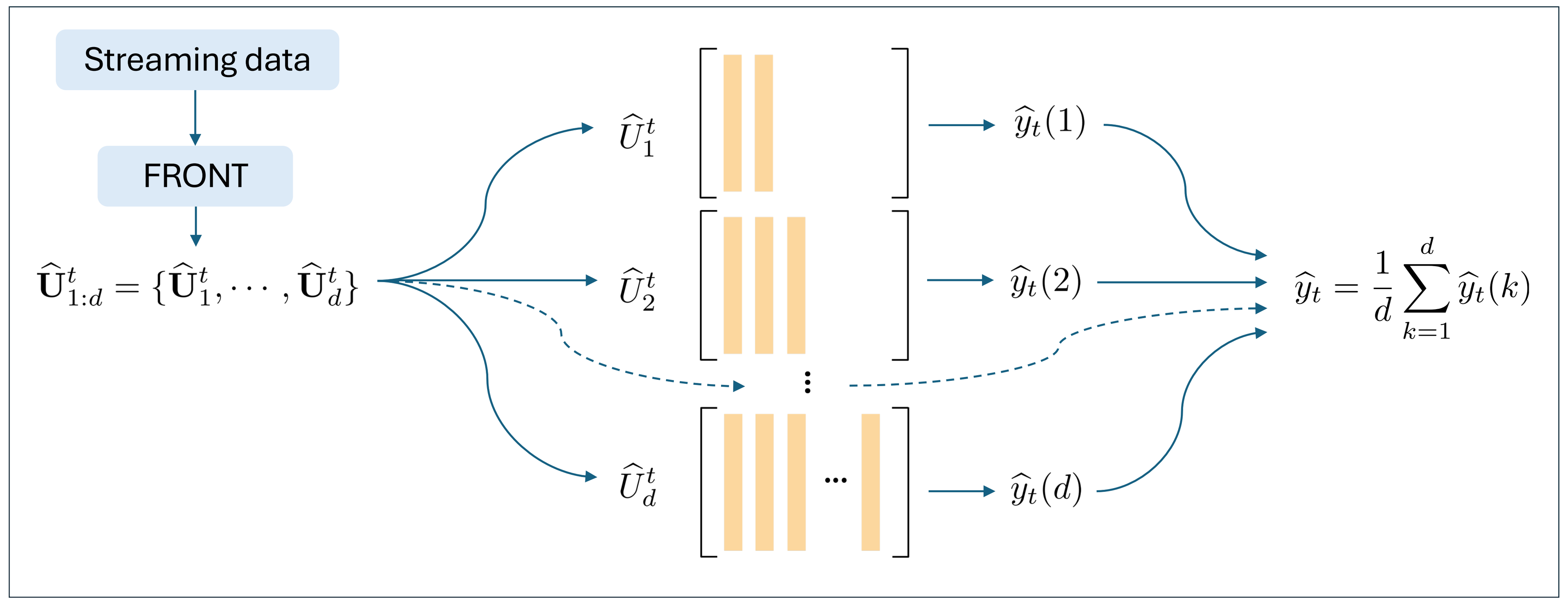
Predict:
- Future output

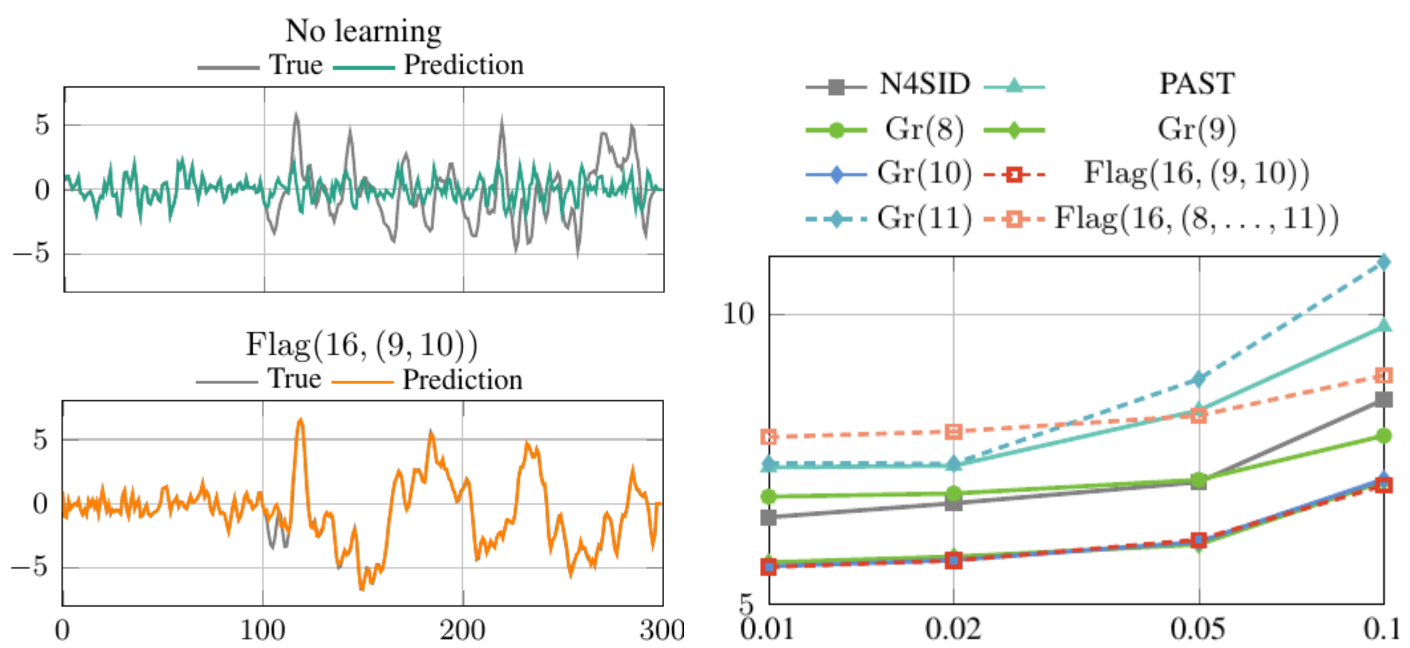


Each nested subspace gives a prediction using the subspace predictor:

$$\hat{y}_t(k) = Y_{f,k}\begin{bmatrix} U_{p,k} \\ U_{f,k} \\ Y_{p,k} \end{bmatrix}^\dagger \begin{bmatrix} u_{[t-T_{\text{ini}},t-1]} \\ u_{[t,t+T_f-1]} \\ y_{[t-T_{\text{ini}},t-1]} \end{bmatrix}, \text{ where } \hat{U}^t_{1:k} = \begin{bmatrix} U_{p,k} \\ U_{f,k} \\ Y_{p,k} \\ Y_{f,k} \end{bmatrix} \text{ is the flag estimate at time } t.$$

Since the true system dimension is unknown, average across ranks:





Left: x-axis is time step; y-axis is output value. Gray curves show the true output, and colored curves show predicted output.
Right: x-axis is noise level; y-axis is prediction error. Lower values indicate better prediction performance.

- Online adaptation matters. Without learning, predictions fail after the system switch; FRONT tracks the changing behavior and restores accurate prediction.
- Robust to unknown dimension. By averaging across nested ranks, the flag ensemble avoids committing to one fixed model order.
- Strong performance under noise. The flag-based predictors match the best single-rank methods and consistently outperform N4SID/PAST across noise levels.

Future work
- convergence guarantees for full flag-manifold learning
- adaptive weighting across ranks instead of uniform averaging

References

[1] Willems, Jan C., et al. "A note on persistency of excitation." *Systems & Control Letters* 54.4 (2005): 325-329.
[2] Szwagier, Tom, and Xavier Pennec. "Nested subspace learning with flags." *arXiv preprint arXiv:2502.06022* (2025).
[3] Sasfi, András, et al. "GREAT: Grassmannian REcursive Algorithm for Tracking & Online System Identification." *IEEE Transactions on Automatic Control* (2025).
[4] Ye, Ke, Ken Sze-Wai Wong, and Lek-Heng Lim. "Optimization on flag manifolds." *Mathematical Programming* 194.1 (2022): 621-660.