Weighted Null Space Fitting (WNSF): A Link between The Prediction Error Method and Subspace Identification

Jiabao He and Håkan Hjalmarsson

I. Introduction

Consider the following discrete-time LTI system in the predictor form:

$$x_{k+1} = A_K x_k + K y_k, \tag{1a}$$
$$y_k = C x_k + e_k. \tag{1b}$$

The main focus of this work is to estimate system's dynamical matrix A_K in a statistically optimal way using input and output data $\{u_k, y_k\}_{k=1}^{\bar{N}}$ from a single trajectory.

Subspace identification methods (SIMs) [1] have proven to be very useful and numerically robust for estimating state-space models. However, they are in general not believed to be as accurate as the prediction error method (PEM) [2]. Conversely, PEM, although more accurate, comes with non-convex optimization problems and therefore requires local non-linear optimization algorithms and good initialization points. A more recent weighted null space fitting (WNSF) [3] approach is able to combine some advantages of the two aforementioned mainstream approaches, being numerically robust and achieving asymptotic efficiency for many structured models. In this work, we extend the WNSF approach for estimating state-space models for multivariate time-series. For simplicity, we use SISO systems for illustration. For a full version of this work, see [4].

II. The Weighted Null-Space Fitting Method

There are multi-step least-squares in the WNSF method.

Step 1 (High Order AR Modeling): Based on the predictor form (1), the output is given by

$$y_k = C(qI - A_K)^{-1} K y_k + e_k = \sum_{i=1}^{\infty} g_i y_{k-i} + e_k, \tag{2}$$

where predictor Markov parameters $g_i = C A_K^{i-1} K$. After selecting a sufficient large order n, equation (2) is truncated to a high order AR (HOAR) model

$$y_k \approx \sum_{i=1}^{n} g_i y_{k-i} + e_k = \boldsymbol{g}_n \boldsymbol{y}_n(k) + e_k, \tag{3}$$

where $\boldsymbol{y}_n(k) = \begin{bmatrix} y_{k-1}^\top & y_{k-2}^\top & \cdots & y_{k-n}^\top \end{bmatrix}^\top$ and $\boldsymbol{g}_n = \begin{bmatrix} g_1 & g_2 & \cdots & g_n \end{bmatrix}$. An estimate of the first n Markov parameters is

$$\hat{\boldsymbol{g}}_n = r_n R_n^{-1}, \tag{4}$$

where $r_n := \frac{1}{N}\sum_{t=1}^{N} y_k \boldsymbol{y}_n^\top(k)$y, $R_n := \frac{1}{N}\sum_{k=1}^{N} \boldsymbol{y}_n(k)\boldsymbol{y}_n^\top(k)$ and $N = \bar{N} - n + 1$. Moreover, assuming the truncation bias of the HOAR model is negligible, which should be close to zero for sufficient large $\bar{N}$, the asymptotic distribution of the estimation error can be approximated as

$$\sqrt{N}(\hat{\boldsymbol{g}}_n - \boldsymbol{g}_n) \sim \mathrm{As}\mathcal{N}\left(0, \sigma_e^2 \bar{R}_n^{-1}\right), \tag{5}$$

where $\bar{R}_n := \bar{\mathbb{E}}\left[\boldsymbol{y}_n(k)\boldsymbol{y}_n^\top(k)\right]$.

Step 2 (Ordinary Least-Squares): With the non-parametric HOAR model in Step 1, we proceed to show how to get a parametric state-space model in the following Steps 2 and 3. According to the Cayley-Hamilton theorem, for matrix A_K, we have

$$A_K^{n_x} + a_1 A_K^{n_x-1} + \cdots + a_{n_x-1} A_K + a_{n_x} I = 0, \tag{6}$$

where $\{a_i\}_{i=1}^{n_x}$ are coefficients of the characteristic polynomial of matrix A_K. The extended observability matrix is then given by

$$\Gamma_{n_x} = \begin{bmatrix} C^\top & (CA_K)^\top & \cdots & (CA_K^{n_x})^\top \end{bmatrix}^\top, \tag{7}$$

where $\mathrm{rank}\left(\Gamma_{n_x}\right) = n_x$, and $\dim\left(\mathrm{Null}(\Gamma_{n_x}^\top)\right) = 1$. Using equation (6), we have

$$\begin{bmatrix} a_{n_x} & a_{n_x-1} & \cdots & a_1 & 1 \end{bmatrix} \Gamma_{n_x} = 0, \tag{8}$$

i.e., the left null space of Γ_{n_x} is fully parameterized by the coefficients $\{a_i\}_{i=1}^{n_x}$. moreover, a canonical realization of matrix A_K is obtained using the these coefficients. For simplicity of illustration, we define

$$\boldsymbol{a} := \begin{bmatrix} a_{n_x} & a_{n_x-1} & \cdots & a_1 \end{bmatrix}. \tag{9}$$

Similar to the Ho-Kalman algorithm, we construct a Hankle matrix from the first n Markov parameters:

$$H_{n_x n} := \left[\begin{array}{cccc} g_1 & g_2 & \cdots & g_p \\ g_2 & g_3 & \cdots & g_{p+1} \\ \vdots & \vdots & \ddots & \vdots \\ \hline g_{n_x+1} & g_{n_x+2} & \cdots & g_n \end{array}\right] := \left[\begin{array}{c} H_{n_x n}^+ \\ \hline H_{n_x n}^- \end{array}\right], \tag{10}$$

where the column number $p = n - n_x$. It is well known that the above Hankel matrix is the product of the extended observability matrix and controllability matrix, i.e.,

$$H_{n_x n} = \Gamma_{n_x} L_p. \tag{11}$$

where L_p is the extended controllability matrix. The above Hankel matrix satisfies $\mathrm{rank}(H_{n_x n}) = n_x$. A key observation is that the null space of the extended observability matrix

Jiabao He and Håkan Hjalmarsson are with the Division of Decision and Control Systems, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, 100 44 Stockholm, Sweden. (Emails: jiabaoh, hjalmars@kth.se)

Γ_{n_x} is also the null space of the Hankle matrix $H_{n_x n}$, i.e., $\begin{bmatrix} \boldsymbol{a} & 1 \end{bmatrix} H_{n_x n} = 0$, which implies

$$\boldsymbol{a} H^+_{n_x n} + H^-_{n_x n} = 0. \tag{12}$$

Since we have estimates of Markov parameters $\{g_i\}_{i=1}^n$ from Step 1, after constructing $H_{n_x n}$ from these Markov parameters, an initial estimate of $\boldsymbol{a}$ is given by OLS as:

$$\hat{\boldsymbol{a}}_{ols} = -\hat{H}^-_{n_x n}(\hat{H}^+_{n_x n})^\top \left(\hat{H}^+_{n_x n}(\hat{H}^+_{n_x n})^\top\right)^{-1}. \tag{13}$$

Step 3 (Weighted Least-Squares): Now we refine our initial estimate $\hat{\boldsymbol{a}}_{ols}$ by using the distribution of $(\hat{\boldsymbol{g}}_n - \boldsymbol{g}_n)$ obtained in Step 1. The residual of $\boldsymbol{a}\hat{H}^+_{n_x n} + \hat{H}^-_{n_x n}$ is

$$\boldsymbol{a}\hat{H}^+_{n_x n} + \hat{H}^-_{n_x n} - \left(\boldsymbol{a}H^+_{n_x n} + H^-_{n_x n}\right) = (\hat{\boldsymbol{g}}_n - \boldsymbol{g}_n)\mathcal{T}_n(\boldsymbol{a}), \tag{14}$$

where $\mathcal{T}_n(\boldsymbol{a})$ is a Toeplitz matrix of $\boldsymbol{a}$. According to (5), we conclude that the distribution of the residual is

$$\sqrt{N}(\hat{\boldsymbol{g}}_n - \boldsymbol{g}_n)\mathcal{T}_n(\boldsymbol{a}) \sim \mathrm{As}\mathcal{N}\left(0, \bar{\Lambda}_n(\boldsymbol{a})\right), \tag{15}$$

where $\bar{\Lambda}_n(\boldsymbol{a}) = \sigma_e^2 \mathcal{T}_n^\top(\boldsymbol{a})\bar{R}_n^{-1}\mathcal{T}_n(\boldsymbol{a})$. Taking $\bar{\Lambda}_n^{-1}(\boldsymbol{a})$ as the optimal weighting, and replacing $\boldsymbol{a}$ and $\bar{R}_n$ with their consistent estimates $\hat{\boldsymbol{a}}_{ols}$ and R_n, we refine the estimate with WLS

$$\begin{aligned}\hat{\boldsymbol{a}}_{wls} = &- \hat{H}^-_{n_x n}\bar{\Lambda}_n^{-1}(\hat{\boldsymbol{a}}_{ols})(\hat{H}^+_{n_x n})^\top \times \\ &\left(\hat{H}^+_{n_x n}\bar{\Lambda}_n^{-1}(\hat{\boldsymbol{a}}_{ols})(\hat{H}^+_{n_x n})^\top\right)^{-1}.\end{aligned} \tag{16}$$

Same as the WNSF for ARMAX model, replacing $\boldsymbol{a}$ with its consistent estimate $\hat{\boldsymbol{a}}_{\text{ols}}$ will not affect the asymptotic optimality of $\hat{\boldsymbol{a}}_{\text{wls}}$, we therefore conjecture that $\hat{\boldsymbol{a}}_{\text{wls}}$ is asymptotically efficient.

III. Simulations

In this section, we provide one SISO system and one MIMO system to demonstrate the performance of WNSF with respect to state-of-the-art methods. For the implementation of SIMs and PEM involved in the comparison, we use the corresponding versions in System Identification Toolbox in MATLAB R2021a. Moreover, the PEM is initialized with the true parameters.

A. A Single-Output System

Consider the following ARMA model:

$$y_k + a y_{k-1} = e_k + c e_{k-1},$$

where $a = -0.8$ and $c = 0.9$, and the innovation $e_k \sim \mathcal{N}(0, 1)$. We show that WNSF is asymptotically efficient for estimating the matrix A_K, which in our case is the parameter c. We perform 500 Monte Carlo trails, with sample size $N \in \{300, 600, 1000, 3000\}$, order of the high order AR model $n \in \{30, 40, 50, 60\}$, and past and future horizons of SIMs $f = p \in \{5, 6, 7, 8\}$, respectively. The performance shown in Figure 1 is evaluated by the mean-squared error (MSE) of the parameter c. As we can see, with the increase of sample size N, the MSE of WNSF approaches the CRLB closely and is competitive with PEM. Moreover, two representative SIMs, SSARX and N4SID are not comparable with WNSF.

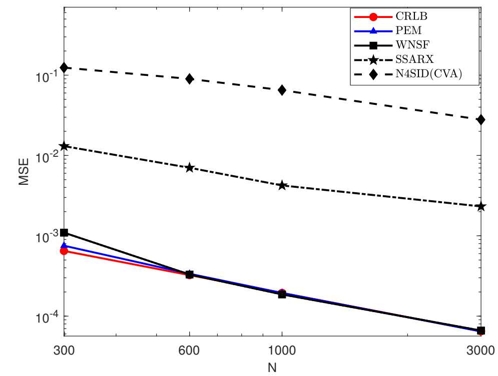

Fig. 1. MSE of parameter c from 500 Monte Carlo trials.

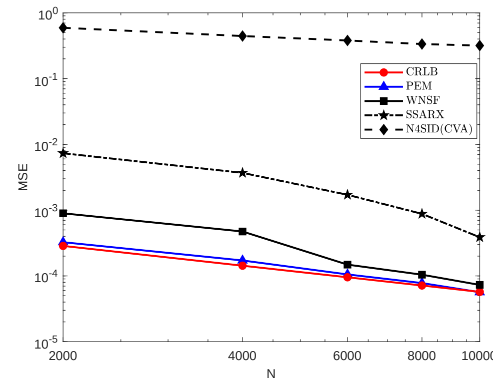

Fig. 2. MSE of coefficients $\boldsymbol{a}$ from 500 Monte Carlo trials.

B. A Multiple-Output System

Consider a MIMO state-space model (1) with the following system matrices:

$$A = \begin{bmatrix} 1 & -0.4 \\ 1 & 0 \end{bmatrix}, K = \begin{bmatrix} 2.45 & -1.05 \\ 2.95 & -0.65 \end{bmatrix}, C = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}.$$

The innovation $e_k \sim \mathcal{N}(0, I)$. We show the performance of different methods for estimating the coefficients of characteristic polynomial for A_K, i.e., the vector $\boldsymbol{a} = \begin{bmatrix} -1.3 & 0.845 \end{bmatrix}$. We perform 500 Monte Carlo trails, with sample size $N \in \{2000 : 2000 : 10000\}$, order of the high order AR model $n \in \{40 : 10 : 80\}$, and past and future horizons of SIMs $f = p \in \{10 : 2 : 18\}$, respectively. The performance shown in Figure 2 is evaluated by the MSE of the vector $\boldsymbol{a}$. As shown, with the increase of sample size N, the MSE of WNSF approaches the MSE of PEM and the CRLB, and performs better than SSARX and N4SID.

References

[1] P. Van Overschee and B. De Moor, "N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems," *Automatica*, vol. 30, no. 1, pp. 75–93, 1994.

[2] L. Ljung, *System identification: Theory for the user*. Prentice Hall information and system sciences series, Prentice Hall PTR, 1999.

[3] M. Galrinho, C. R. Rojas, and H. Hjalmarsson, "Parametric identification using weighted null-space fitting," *IEEE Trans. Autom. Control*, vol. 64, no. 7, pp. 2798–2813, 2018.

[4] J. He and H. Hjalmarsson, "Weighted null space fitting (wnsf): A link between the prediction error method and subspace identification," *arXiv preprint arXiv:2411.00506*, 2024.