Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

Recursive Generalized Total Least Squares with Noise Covariance Estimation ⋆

Stephan Rhode * **Felix Bleimund** * **Frank Gauterin** *

* *Institute of Vehicle System Technology, Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany {stephan.rhode, felix.bleimund, frank.gauterin}@kit.edu*

Abstract: We propose a recursive generalized total least-squares (RGTLS) estimator that is used in parallel with a noise covariance estimator (NCE) to solve the errors-in-variables problem for multi-input-single-output linear systems with unknown noise covariance matrix. Simulation experiments show that the suggested RGTLS with NCE procedure outperforms the common recursive least squares (RLS) and recursive total instrumental variables (RTIV) estimators when all measured inputs and the measured output are noisy. Moreover, when all measured inputs are noise-free, RGTLS with NCE performs similarly to RLS, which in this special case is the optimal estimator, and again RTIV was inferior compared with the RGTLS and NCE procedure.

Keywords: System identification; Errors-in-variables models; Linear systems; Recursive algorithms; Recursive filters; Kalman filters; Signal processing

1. INTRODUCTION

Unbiased estimates of unknown parameters (X) are required in many multi-input-single-output (MISO) problems. The common recursive filters, such as recursive least squares (RLS) or the Kalman filter (KF), solve the constrained perturbation problem. This problem is known as output error model and corrections were solely applied to the measured output (B). However, RLS and KF are known to produce biased estimates if the measured inputs (A) are noisy. In this case, the unconstrained perturbation problem, or precisely the errors-in-variables (EIV) model, is preferred. EIV methods can be divided into two classes. The first class requires knowledge of the noise covariance matrix ($\widetilde{P}$). The bias-compensating RLS algorithm by Ding et al. [2006] adds a correction term, that is built from the noise variance and a cross-correlation matrix and adjusts the estimates. Furthermore, numerous recursive total least-squares algorithms based on the minimization of the Rayleigh quotient (RQ) were proposed by Davila [1994]; Feng et al. [2004]; Lim et al. [2005]; Feng and Zheng [2007]; Arablouei and Dogancay [2012]. All of these methods try to solve the EIV problem with a cost function that considers data corrections in all elements of the augmented data (Z). In the second class, the bias in the parameter estimates ($\widehat{X}$) is reduced by properly chosen instruments ($\mathfrak{A}$). A recursive total instrumental variables (RTIV) estimator was introduced by Feng and Zheng [2007]. The main advantage of instrumental variables (IV) methods is that no knowledge of $\widetilde{P}$ is required. An extensive overview of EIV methods is given by Söderström [2007].

After a brief review of the output error model and the RLS filter in Sec. 2, the EIV model is presented in Sec. 3. The proposed recursive generalized total least-squares (RGTLS) estimator in Sec. 3.1 falls into the first class, where knowledge of $\widetilde{P}$ is required. As $\widetilde{P}$ is commonly unknown in practice, we present a noise covariance estimator (NCE) in Sec. 3.3 that is based on a novel polynomial Kalman smoother (PKS) described in Sec. 3.2. We use simulation experiments with two noise settings in Sec. 4 to compare the results of RLS, RTIV, and RGTLS with NCE in Sec. 5 and finish with conclusions.

2. OUTPUT ERROR MODEL

The constrained perturbation problem is

$$\overline{A}X \approx B, \quad B = \overline{B} + \widetilde{B}, \tag{1}$$

with the true input ($\overline{A}$), $\overline{A} \in \mathbb{R}^{m\times n}$ unknown X, $X \in \mathbb{R}^{n\times 1}$ and B, $B \in \mathbb{R}^{m\times 1}$, that consists of the true output ($\overline{B}$) and output noise ($\widetilde{B}$). The noise covariance matrix of dimension $\widetilde{P} \in \mathbb{R}^{q\times q}$ with $q = n + 1$ is assumed as diagonal matrix where only the last element $\widetilde{P}_{q,q}$ is greater than zero and σ^2 is an unknown multiplier.

$$\operatorname{diag}(\widetilde{P}) = \sigma^2[0, \ldots, 1]^\top \tag{2}$$

The assumption in (2) is rather restrictive and will result in biased estimates if there is any kind of noise correlation or if input noise ($\widetilde{A}$) is present. The well-known RLS estimator [Ljung, 1999, 365] is shown in Alg. 1. RLS is the exponentially weighted recursive version of batch least squares (LS) with the cost function and closed-form solution [Ljung, 1999, 206]

$$\min_{X,\widehat{B}\in\mathbb{R}^{m\times 1}} \|B - \widehat{B}\|_2, \quad \text{s.t. } AX = \widehat{B} \tag{3a}$$

$$\widehat{X} = (A^\top A)^{-1} A^\top B. \tag{3b}$$

⋆ The associated Matlab code can be downloaded from: http://digbib.ubka.uni-karlsruhe.de/volltexte/1000038517

Alg. 1: Recursive least squares (RLS)

1 for $t \leftarrow 1$ to m do
  input: $\widehat{X}_{t-1}, P_{t-1}, A_t, B_t, \lambda$
2 $L_t = \left(P_{t-1}A_t^\top\right)\left(\lambda + A_t P_{t-1} A_t^\top\right)^{-1}$
3 $\widehat{X}_t = \widehat{X}_{t-1} + L_t\left(B_t - A_t\widehat{X}_{t-1}\right)$
4 $P_t = (I - L_t A_t)P_{t-1}\frac{1}{\lambda}$
  output: $\widehat{X}_t, P_t$

3. ERRORS-IN-VARIABLES MODEL

The unconstrained perturbation problem

$$AX \approx B, \quad A = \overline{A} + \widetilde{A}, \quad B = \overline{B} + \widetilde{B} \tag{4}$$

is known as EIV model and considers input noise and output noise. Generally speaking, EIV is a more realistic perturbation model, but requires knowledge of $\widetilde{P}$. Total least squares (TLS) is the optimal estimator if

$$\widetilde{P} = \sigma^2 I. \tag{5}$$

That means the noise is independently identically distributed (i.i.d.), Markovsky and Van Huffel [2007]. This condition is as restrictive as (2) in the output error model. However, there are weighted versions of TLS that can handle any $\widetilde{P}$ that is created from noise with zero mean and normal distribution, Markovsky and Van Huffel [2007]. And some of these estimators, specifically generalized total least squares (GTLS), yield the maximum likelihood estimates for EIV problems with a closed-form solution if the problem can be formulated as

$$\min_{X, \widehat{Z} \in \mathbb{R}^{m \times q}} \|W_{\mathrm{l}}(Z - \widehat{Z})W_{\mathrm{r}}\|_{\mathrm{F}} \quad \text{s.t. } \widehat{Z}\begin{bmatrix} X \\ -I \end{bmatrix} = 0, \tag{6}$$

with $Z = [A, B]$ and $\widehat{Z} = [\widehat{A}, \widehat{B}]$, respectively. One can use the left weighting matrix (W_{l}) as diagonal matrix that considers the row-wise exponential forgetting $\mathrm{diag}(W_{\mathrm{l}}) = [\ldots, \lambda^2, \lambda^1, \lambda^0]^\top$ and compute the right weighting matrix (W_{r}) from the Cholesky factor (C) of $\widetilde{P}$ in the sense that W_{r} transforms Z and $\widehat{Z}$ into Z' and $\widehat{Z}'$ where any $\widetilde{P} \neq \sigma^2 I$ becomes $\widetilde{P}' = \sigma^2 I$ (i.i.d.) and TLS is the optimal estimator. Alg. 2 shows the GTLS solution by data scaling through Cholesky factorization ($\mathrm{chol}(\cdot)$) of $\widetilde{P}$, Schuermans et al. [2005]. Recursive versions of GTLS with data scaling were shown in Kubus et al. [2008]; Rhode and Gauterin [2013]. These algorithms replace the batch $\mathrm{svd}(\cdot)$ in Alg. 2 line 4 with efficient $\mathrm{svd}(\cdot)$ update schemes, Brand [2002, 2006]; Gu and Eisenstat [1993]. A drawback of data scaling is that one cannot assume one or more measured inputs as noise-free, because a scaling with zero would neglect these measured inputs. However, these algorithms provide a closed-form solution.

Alg. 2: Generalized total least squares (GTLS)

  input: $A, B, \widetilde{P}$
1 $C = \mathrm{chol}(\widetilde{P})$
2 $W_{\mathrm{r}} = C^{-1} := \begin{matrix} \\ n \\ 1 \end{matrix}\begin{matrix} \begin{matrix} n & 1 \end{matrix} \\ \begin{bmatrix} W_{\mathrm{r}11} & W_{\mathrm{r}12} \\ 0 & W_{\mathrm{r}22} \end{bmatrix} \end{matrix}$
3 $Z' = [A', B'] = [A, B]W_{\mathrm{r}}$
4 $\sim S'V'^\top = \mathrm{svd}(Z')$
5 $V' := \begin{matrix} \\ n \\ 1 \end{matrix}\begin{matrix} \begin{matrix} n & 1 \end{matrix} \\ \begin{bmatrix} V'_{11} & V'_{12} \\ V'_{21} & V'_{22} \end{bmatrix} \end{matrix}$
6 $\widehat{X} = \left(W_{\mathrm{r}11}(-V'_{12}{V'_{22}}^{-1}) - W_{\mathrm{r}12}\right)W_{\mathrm{r}22}^{-1}$
  output: $\widehat{X}$

The majority of recursive total least-squares algorithms use power methods, such as inverse iteration (II) or Rayleigh quotient iteration (RQI). Davila [1994] showed that the minimization of the generalized Rayleigh quotient (GRQ)

$$\min_{V_{:,q}} \frac{V_{:,q}^\top (Z^\top Z) V_{:,q}}{V_{:,q}^\top \widetilde{P} V_{:,q}} \tag{7}$$

provides the eigenvector $V_{:,q}$ that corresponds to the smallest eigenvalue $S_{q,q}$. And this eigenvector is involved in the GTLS solution. The notation $V_{:,q}$ means all (:) rows in the rightmost (q-th) column of V and $S_{q,q}$ the last element of S.

Lemma 1. The minimization of (7) results in asymptotically unbiased and consistent GTLS solution X in case of additive, zero mean, Gaussian noise and known noise covariance matrix up to a multiplier by the substitution of $V_{:,q}$ with $[X^\top, -1]^\top$ and the substitution of $Z^\top Z$ with the expectation ($\mathbb{E}$) of the sample cross-correlation matrix (R) for large enough t, where

$$R_t = \frac{1}{t}\sum_i^t Z_i^\top Z_i. \tag{8}$$

Proof. See Proof of Theorem 1 in Davila [1994] for the substitution of $Z^\top Z$ with $\mathbb{E}(R_t)$ and Sec. B in Feng et al. [2004] for the substitution of $V_{:,q}$ with $[X^\top, -1]^\top$.

Hence, as shown in Feng et al. [2004], the eigenvector $V_{:,q}$ can be replaced with $[X^\top, -1]^\top$ in (7) and the minimization simplifies to the constrained generalized Rayleigh quotient (CGRQ)

$$\min_X \frac{[X^\top, -1](Z^\top Z)[X^\top, -1]^\top}{[X^\top, -1]\widetilde{P}[X^\top, -1]^\top}. \tag{9}$$

Following [Golub and Van Loan, 1996, 465], one can solve (7) with generalized inverse iteration (GII) as shown in Alg. 3. It is obvious that the while loop in Alg. 3 line 3 does not allow a closed-form solution, but GII converges in the most cases very fast within a few iterations. Because of that, GII is suitable for online algorithms assuming that for each time step one iteration is sufficient to follow the smallest eigenvector.

Alg. 3: Generalized inverse iteration (GII)

  input: $Z, \widetilde{P}$
1 $V_{:,q;t} = [1, 1, \ldots, 1]^\top$
2 $V_{:,q;t-1} = [0, 0, \ldots, 0]^\top$
3 while $\|V_{:,q;t-1} - V_{:,q;t}\|_2 >$ threshold do
4 $V_{:,q;t-1} = V_{:,q;t}$
5 $V_{:,q;t} = (Z^\top Z)^{-1}(\widetilde{P}V_{:,q;t})$
6 $V_{:,q;t} = V_{:,q;t}/\|V_{:,q;t}\|_2$
7 $\widehat{X} = -V_{1:n,q;t}/V_{q,q;t}$
  output: $\widehat{X}$

3.1 Recursive generalized total least squares (RGTLS)

The herein proposed RGTLS algorithm that is shown in Alg. 4, is based on the optimization procedure (9) and the recursive update of the augmented data covariance matrix. Apart from using Z_t instead of A_t, the update in Alg. 4 line 3 conforms with Alg. 1 line 4. The constrained generalized inverse iteration (CGII) is performed in Alg. 4 from line 4–line 5 and $\widetilde{P}$ is replaced with an estimated noise covariance matrix ($\widehat{P}$). $\widehat{P}$ can also be used as fixed user input in the form of (2) for a RLS solution or (5) for a RTLS solution. A method for the estimation of $\widehat{P}$ is shown in Sec. 3.2 and Sec. 3.3.

Alg. 4: Recursive generalized total least squares (RGTLS)

1 for $t \leftarrow 1$ to m do
  input: $\widehat{X}_{t-1}, P_{t-1}, Z_t, \widehat{P}_t, \lambda$
2 $L_t = \left(P_{t-1}Z_t^\top\right)\left(\lambda + Z_t P_{t-1} Z_t^\top\right)^{-1}$
3 $P_t = (I - L_t Z_t) P_{t-1} \frac{1}{\lambda}$
4 $V_{:,q;t-1} = [\widehat{X}_{t-1}^\top, -1]^\top$
5 $V'_{:,q;t} = P_t(\widehat{P}_t V_{:,q;t-1})$
6 $\widehat{X}_t = -V'_{1:n,q;t} / V'_{q,q;t}$
  output: $\widehat{X}_t, P_t$

3.2 Polynomial Kalman smoother (PKS)

Polynomial functions have been shown suitable for extracting signals from noisy time series, Savitzky and Golay [1964]. They are able to preserve the original signal level and have a well-defined and tunable delay.

The Savitzky Golay filter (SGF) uses convolution arrays to obtain the smoothed signal and the smoothed derivatives at one preselected point. Like proposed in Savitzky and Golay [1964], we use a smoothing window with equal left window (w_l) and right window (w_r) and model the signal with a time-varying polynomial function of order $n-1$.

SGF performs a weighted sum of the measured signal within the smoothing window. Hence, SGF requires one buffer of size $w_\mathrm{l} + 1 + w_\mathrm{r}$ for the measured data within the smoothing window and at least one buffer of same size for the weights, that are derived in Madden [1978]. Additional buffers with specific weights are required as the number of desired derivatives increases.

We use a specific form of the Kalman filter, which we call polynomial Kalman smoother (PKS), to extract noise from noisy measurements. PKS is based on the principles of SGF. The herein proposed PKS approach can be seen as the recursive version of SGF with exponentially weighted data. Conversely to SGF, PKS yields estimates of the polynomial function parameters. Hence, one can evaluate the polynomial function as well as its derivatives at multiple points. Due to the recursive approach, PKS outperforms SGF significantly in matters of memory.

The requirement for the polynomial function approach is that the time series is built from auto-correlated signals. The following procedure would fail if the signal is a random process, because the time-varying polynomial function would not properly model the underlying signal from the measured signal in this case.

However, the proposed RGTLS estimator could also be used in this case but then requires a user-defined $\widetilde{P}$ instead of the estimated $\widehat{P}$. Note that any kind of IV estimator would fail in this case, because the instruments are not properly correlated with the true input, which is a requirement in IV estimation, see Sec. 3.5.

We use the state-space representation in (10) with the state transition matrix ($\mathcal{A}$), the polynomial parameters X and the measurement vector ($\mathcal{C}$) as polynomial control input vector and perform a random walk model [Ljung and Gunnarsson, 1990] of the time-varying polynomial function, while $\mathcal{B}$ and $\mathcal{D}$ are zero.

$$X_t = \mathcal{A}X_{t-1} + \mathcal{B}A_t \tag{10a}$$
$$B_t = \mathcal{C}X_t + \mathcal{D}A_t \tag{10b}$$

For a fixed unit shift of the polynomial function, the state transition matrix $\mathcal{A} \in \mathbb{Z}^{n\times n}$ becomes a time-invariant upper triangular square matrix and contains binomial coefficients $\binom{n}{k}$

$$\mathcal{A}_{i,j} = \begin{cases} \binom{j-1}{j-i} & \forall j \geq i \\ 0 & \forall j < i, \end{cases} \tag{11}$$

see Zima [1997] for general shifts. For instance, $\mathcal{A}$ yields for a polynomial function with four parameters (third order) to

$$\mathcal{A} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{12}$$

If PKS is used a smoother with $w_\mathrm{l} = w_\mathrm{r}$, $\mathcal{C}$ is given with

$$\mathcal{C} = [(w_\mathrm{l} + 1 + w_\mathrm{r})^0, (w_\mathrm{l} + 1 + w_\mathrm{r})^1, \ldots, (w_\mathrm{l} + 1 + w_\mathrm{r})^{n-1}]. \tag{13}$$

The PKS algorithm is shown in Alg. 5, with the noisy measurement B_t and the forgetting factor (λ).

Alg. 5: Polynomial Kalman smoother (PKS)

1 for $t \leftarrow 1$ to m do
  input: $\widehat{X}_{t-1}, P_{t-1}, B_t, \mathcal{A}, \mathcal{C}, \lambda$
2 $\widehat{X}'_{t-1} = \mathcal{A}\widehat{X}_{t-1}$
3 $P'_{t-1} = \mathcal{A}P_{t-1}\mathcal{A}^\top$
4 $K_t = \left(P'_{t-1}\mathcal{C}^\top\right)\left(\lambda + \mathcal{C}P'_{t-1}\mathcal{C}^\top\right)^{-1}$
5 $P_t = \frac{1}{\lambda}\left(P'_{t-1} - K_t\mathcal{C}P'_{t-1}\right)$
6 $\widehat{X}_t = \widehat{X}'_{t-1} + K_t\left(B_t - \mathcal{C}\widehat{X}'_{t-1}\right)$
  output: $\widehat{X}_t, P_t$

The smoothed signal at the center of the window ($w_\mathrm{l} + 1$) is gained by

$$\widehat{B}_{t-w_\mathrm{r}} = [(w_\mathrm{l} + 1)^0, (w_\mathrm{l} + 1)^1, \ldots, (w_\mathrm{l} + 1)^{n-1}]\widehat{X}_t. \tag{14}$$

The estimated noise can be extracted from the measurement with

$$\widehat{B}_{t-w_\mathrm{r}} = B_{t-w_\mathrm{r}} - \widehat{B}_{t-w_\mathrm{r}}. \tag{15}$$

Note that we need a delay of w_r samples in the measured signal B to synchronize the smoothed signal $\widehat{B}$ from PKS. This requires additional memory for each measurement. However, in many cases, the accuracy of smoothing with

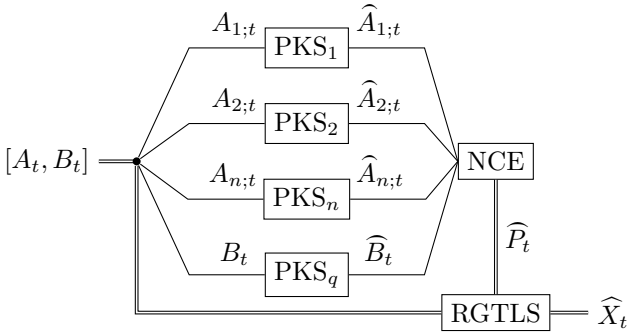


Fig. 1. Block diagram of RGTLS with NCE.

equal left window and right window outperforms a filter without right window. Nevertheless, we can modify PKS into a filter by setting $w_\mathrm{r} = 0$.

3.3 Noise covariance estimator (NCE)

In MISO system identification, $\widehat{P}$ is a square matrix with $\widehat{P} \in \mathbb{R}^{q\times q}$. A simple noise covariance update formula with forgetting is

$$\widehat{Z}_t = [\widehat{A}_{1;t}, \ldots, \widehat{A}_{n;t}, \widehat{B}_t], \tag{16a}$$

$$\widehat{P}_t = \lambda \widehat{P}_{t-1} + (1-\lambda)(\widehat{Z}_t^\top \widehat{Z}_t), \tag{16b}$$

where (16b) is the multidimensional version of the noise variance estimator in Zou et al. [2000].

Fig. 1 shows that we need q-independent PKS to compute $\widehat{Z}_t$ in (16a). The noisy measurement B_t in Alg. 5 is $A_{1;t}$ for PKS$_1$, …, $A_{n;t}$ for PKS$_n$ and B_t for PKS$_q$.

3.4 RGTLS with NCE

Finally, Fig. 1 gives the block diagram of RGTLS with NCE. Note that the smoothing of the q-independent PKS is only used in NCE to compute $\widehat{P}$, while RGTLS uses raw measured data.

3.5 Recursive total instrumental variables (RTIV)

The IV method can yield bias-free estimates if the instruments are chosen in such a way that they are highly correlated with $\overline{A}$ and uncorrelated with the noise [Ljung, 1999, 224]. IV estimators are easy to apply, because knowledge of $\widetilde{P}$ is not needed. The closed-form solution becomes [Ljung, 1999, 224]

$$\widehat{X} = (\mathfrak{A}^\top A)^{-1}\mathfrak{A}^\top B, \tag{17}$$

and is very similar to (3b).

The RTIV estimator in Alg. 6 was introduced by Feng and Zheng [2007] and is comparable with the proposed RGTLS estimator in Alg. 4 in terms of the used II. RTIV serves as benchmarking method for the proposed RGTLS with NCE estimator.

4. SIMULATION EXPERIMENTS

4.1 Simulation data setup

Fig. 2 shows $\overline{A}$ and $\overline{B}$ for $t = 400\,\mathrm{s}$ to $1000\,\mathrm{s}$. In particular, $\overline{A}$, X, and $\overline{B}$ were generated with

Alg. 6: Recursive total instrumental variables (RTIV)

for $t \leftarrow 1$ to m do
  input: $R_{t-1}, V_{:,1;t-1}, V_{:,q;t-1}, Z_t, \mathfrak{Z}_t, \lambda$
  $R_t = \lambda R_{t-1} + (\mathfrak{Z}_t^\top Z_t)$
  $V'_{:,1;t} = R_t^\top R_t V_{:,1;t-1}$
  $S_{1,1;t} = \|V'_{:,1;t}\|_2$
  $V_{:,1;t} = V'_{:,1;t}/S_{1,1;t}$
  $V'_{:,q;t} = R_t^\top R_t V_{:,q;t-1}$
  $S_{q,q;t} = \|V'_{:,q;t}\|_2$
  $V''_{:,q;t} = (0.5S_{1,1;t} + S_{q,q;t})V_{:,q;t-1} - V'_{:,q;t}$
  $V_{:,q;t} = V''_{:,q;t}/\|V''_{:,q;t}\|_2$
  $\widehat{X}_t = -V_{1:n,q;t}/V_{q,q;t}$
  output: $\widehat{X}_t, R_t, V_{:,1;t}, V_{:,q;t},$

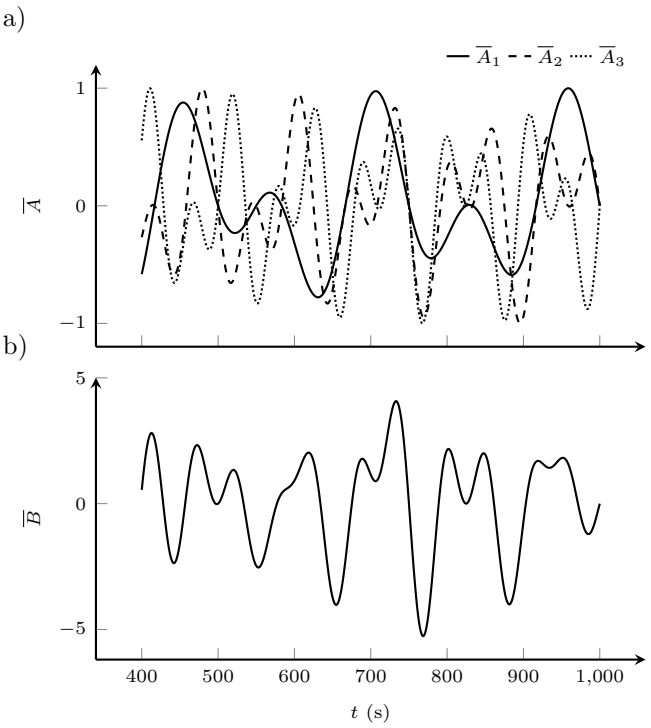


Fig. 2. True inputs (Fig. 2a) and true output (Fig. 2b) modeled as sine waves.

$$t = [1, 2, 3, \ldots, m]^\top,\ m = 10\,000\,\mathrm{s} \tag{18a}$$

$$\overline{A}_1 = \sin(2\pi\, t\, 0.006)\sin(2\pi\, t\, 0.006/3.3) \tag{18b}$$

$$\overline{A}_2 = \sin(2\pi\, t\, 0.012)\sin(2\pi\, t\, 0.012/3.3) \tag{18c}$$

$$\overline{A}_3 = \sin(2\pi\, t\, 0.014)\sin(2\pi\, t\, 0.014/3.3) \tag{18d}$$

$$X = \begin{cases} [1,2,3]^\top & 1\,\mathrm{s} \le t < 5000\,\mathrm{s} \\ [2,2,3]^\top & 5000\,\mathrm{s} < t \le 10\,000\,\mathrm{s} \end{cases} \tag{18e}$$

$$\overline{B} = \overline{A} \odot X. \tag{18f}$$

Note that there is a parameter step in (18e) for X_1 to test the tracking performance of RLS, RTIV and RGTLS.

Two different noise settings were applied. For each noise setting, 1000 independent experiments with white Gaussian noise were generated. The noise variances were selected as follows.

Noise setting №1: The first noise setting conforms to weighted total least squares (WTLS), where each measured

input and measured output is noisy. The noise covariance matrix is uncorrelated and unequally sized, which means that all off-diagonal elements of $\widetilde{P}$ are zero.

$$\sigma^2(\widetilde{A}_1) = 0.1,\ \sigma^2(\widetilde{A}_2) = 0.2,\ \sigma^2(\widetilde{A}_3) = 0.4 \tag{19a}$$

$$\sigma^2(\widetilde{B}) = 1 \tag{19b}$$

Noise setting №2: The noise variances in the second noise setting are

$$\sigma^2(\widetilde{A}_1) = \sigma^2(\widetilde{A}_2) = \sigma^2(\widetilde{A}_3) = 0 \tag{20a}$$

$$\sigma^2(\widetilde{B}) = 1. \tag{20b}$$

This setting corresponds to the assumptions of RLS filtering, where input noise is not considered. Hence, $\widetilde{P}_{4,4} = 1$ and all other elements are zero.

4.2 Estimator setup

All estimators were initialized with parameter estimates $(\widehat{X}_{t-1})$ of batch LS at $t = 30\,\mathrm{s}$, while $\widehat{X}_{t<30\,\mathrm{s}} = 0$. The initial covariance matrix (P_{t-1}) for RLS and RGTLS was also gained from batch LS, while the initial cross-correlation matrix (R_{t-1}) for RTIV was computed with $R_{t=30\,\mathrm{s}} = Z^{\top}_{t=1\,\mathrm{s}\,\mathrm{to}\,30\,\mathrm{s}} Z_{t=1\,\mathrm{s}\,\mathrm{to}\,30\,\mathrm{s}}$.

The forgetting factor was fixed to $\lambda = 0.998$ for the RLS, RTIV, NCE and RGTLS estimator. However, the PKS estimator was used with $\lambda = 0.9$.

RTIV setup: The augmented instruments were built with a delay of four from the augmented data. Hence, $\mathfrak{Z}_t = Z_{t-4}$.

PKS setup: We used $w_\mathrm{l} = w_\mathrm{r} = 30$ and a polynomial function with five parameters (fourth order). This setup leads with (11), (13) and (14) to

$$\mathcal{A} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \\ 0 & 0 & 1 & 3 & 6 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{21}$$

$$\mathcal{C} = [1, 61, 3721, 226981, 13845841] \tag{22}$$

$$\widehat{B}_{t-w_\mathrm{r}} = [1, 31, 961, 29791, 923521]\widehat{X}_t. \tag{23}$$

NCE setup: The estimated noise covariance matrix was initialized with $\widehat{P}_{t<30\,\mathrm{s}} = 0.3I$. This setting corresponds to a TLS-like noise assumption, with i.i.d. noise.

5. RESULTS AND DISCUSSION

In accordance with Davila [1994]; Feng and Zheng [2007], the squared error vector norm $(\mathcal{E})$

$$\mathcal{E}_t = \left\| \widehat{X}_t - X_t \right\|_2^2, \tag{24}$$

was computed as performance index for each estimator, noise setting, and experiment. Afterwards, the arithmetic mean $(\mu(\cdot))$ was used to compute the expectation $(\mathbb{E})$ of $\mathcal{E}$ over the 1000 experiments for each noise setting and estimator.

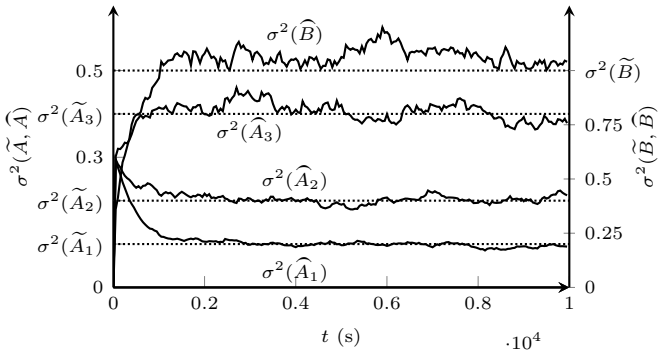


Fig. 3. Input and output noise variance estimation in noise setting №1 of one experiment. Note that $\sigma^2(\widetilde{B})$ and $\sigma^2(\widehat{B})$ belong to the right ordinate. True values are shown with ····, while estimated values are given with —.

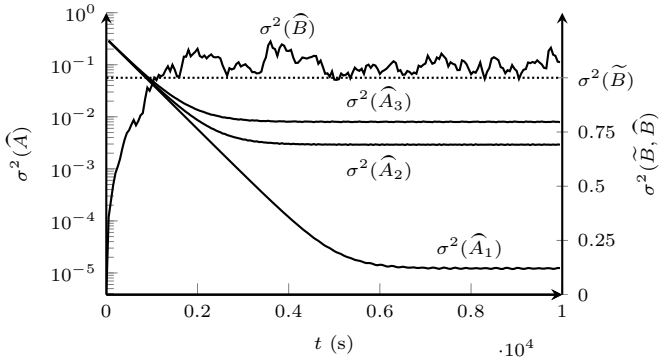


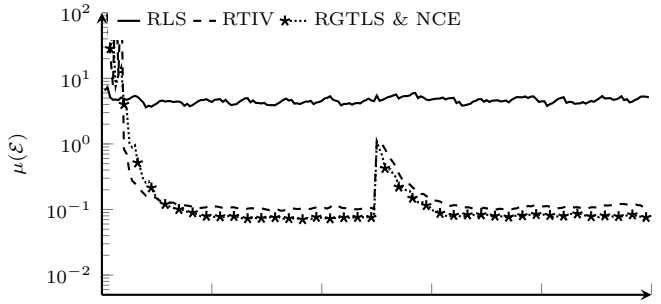
Fig. 4. Input and output noise variance estimation in noise setting №2 of one experiment with noise free inputs. Note that $\sigma^2(\widetilde{B})$ and $\sigma^2(\widehat{B})$ belong to the right ordinate. The true input noise variance is $\sigma^2(\widetilde{A}_{1:3}) = 0$.

5.1 Results of NCE

Fig. 3 provides the diagonal elements of $\widehat{P}_t$ for one experiment in noise setting №1, where $\widetilde{P}$ was adjusted in accordance with (19). NCE gives accurate estimates from $t > 1000\,\mathrm{s}$ on, while $\sigma^2(\widehat{A}_1)$ and $\sigma^2(\widehat{A}_2)$ are more accurate than $\sigma^2(\widehat{A}_3)$ and $\sigma^2(\widehat{B})$. The reason for this is that we used the same polynomial order and forgetting factor for all PKS filters. Hence, the higher frequent signals $\overline{A}_3$ and $\overline{B}$ were less accurate modeled by the polynomial function as the lower frequent $\overline{A}_1$ and $\overline{A}_2$. However, with a specific adjustment for each individual PKS, more accurate results are to be expected.

The NCE result for noise setting №2 of one experiment is given in Fig. 4. As expected, the accuracy of $\sigma^2(\widehat{B})$ is comparable with Fig. 3, while the estimates of $\sigma^2(\widehat{A}_{1:3})$ depend once again on the frequency of $\overline{A}_{1:3}$, compare (18). The lower frequent $\overline{A}_1$ has the most accurate noise variance estimate $\sigma^2(\widehat{A}_1)$ followed by $\sigma^2(\widehat{A}_2)$ and finally $\sigma^2(\widehat{A}_3)$.

a) Noise setting №1



b) Noise setting №2

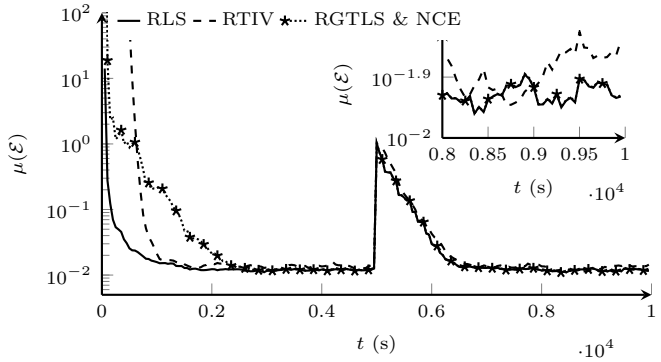


Fig. 5. Squared error vector norm ensemble averaged over 1000 independent experiments. The RGTLS estimator outperforms RLS and RTIV in noise setting №1 (Fig. 5a). Additionally, practically no difference can be observed between RGTLS and RLS in noise setting №2 (Fig. 5b), where RLS is the optimal estimator. Note the inset in Fig. 5b. RGTLS outperforms RTIV in both noise settings.

5.2 Results of RLS, RTIV, and RGTLS with NCE

Finally, we compare the performance of RLS, RTIV, and RGTLS with NCE in Fig. 5. The poor performance of RLS in Fig. 5a for noise setting №1 motivates to use EIV methods. Note that RLS fails entirely to indicate the parameter step change of X_1 at $t = 5000\,\mathrm{s}$. As this observation is quite obvious because RLS is not designed for EIV problems such as noise setting №1, the superior performance of RGTLS with NCE compared with RTIV is remarkable. However, RTIV converges faster than RGTLS for $t < 1000\,\mathrm{s}$. The reason is that NCE requires some time to converge.

Noise setting №2 in Fig. 5b is designed for RLS. Moreover, RLS provides the optimal solution in this case and is suitable for benchmarking RTIV and RGTLS with NCE. RGTLS with NCE is the slowest converging algorithm. Once again, this is due to the time that NCE requires to provide an accurate $\widehat{P}$ result. However, the inset in Fig. 5b shows that practically no difference is observed between the optimal RLS and the proposed RGTLS with NCE, while RTIV here again is inferior. This considerable performance of RGTLS was not expected here if we take into account the biased estimates of $\sigma^2(\widehat{A}_{1:3})$ in Fig. 4. This indicates that an approximate estimate of $\widehat{P}$ is sufficient to use RGTLS for various noise settings.

6. CONCLUSION

The simulation experiments show that the proposed recursive generalized total least-squares (RGTLS) estimator together with the introduced noise covariance estimator (NCE) yield highly accurate parameter estimates ($\widehat{X}$) in unknown noise environments. RGTLS with NCE outperforms the recursive total instrumental variables (RTIV) estimator in noise setting №1, where all measured inputs (A) and the measured output (B) were noisy as well as in noise setting №2, where only B is noisy. In addition, although the estimated noise covariance matrix ($\widehat{P}$) of NCE was biased in noise setting №2, RGTLS with NCE produced similar results as the, in this case optimal, recursive least squares (RLS) estimator. Once again the RTIV estimator performed poorer than RGTLS with NCE. The only drawback of RGTLS with NCE is the slower convergence during the first iterations compared with RTIV in both noise settings.

ACKNOWLEDGEMENTS

This work was accomplished in cooperation with the Energy Management Complete Vehicle Department at Dr. Ing. h.c. F. Porsche AG, Weissach, Germany.
The authors are grateful to valuable comments of the reviewers that helped to improve this paper.

REFERENCES

R. Arablouei and K. Dogancay. Linearly-constrained recursive total least-squares algorithm. *Signal Processing Letters, IEEE*, 19(12):821–824, 12 2012.

M. Brand. Incremental singular value decomposition of uncertain data with missing values. In *Computer Vision-ECCV 2002*, pages 707–720. Springer, 2002.

M. Brand. Fast low-rank modifications of the thin singular value decomposition. *Linear algebra and its applications*, 415(1):20–30, 2006.

C.E. Davila. An efficient recursive total least squares algorithm for fir adaptive filtering. *Signal Processing, IEEE Transactions on*, 42(2):268–280, 2 1994.

F. Ding, T. Chen, and L. Qiu. Bias compensation based recursive least-squares identification algorithm for miso systems. *Circuits and Systems II: Express Briefs, IEEE Transactions on*, 53(5):349–353, 5 2006.

D.-Z. Feng and W.X. Zheng. Recursive total instrumental-variable algorithm for solving over-determined normal equations and its applications. *Signal Processing*, 87(5): 918–936, 2007.

D.-Z. Feng, X.-D. Zhang, D.-X. Chang, and W.X. Zheng. A fast recursive total least squares algorithm for adaptive fir filtering. *Signal Processing, IEEE Transactions on*, 52(10):2729–2737, 10 2004.

G.H. Golub and C.F. Van Loan. *Matrix computations*, volume 3. Johns Hopkins Univ Pr, 1996.

M. Gu and S.C. Eisenstat. A stable and fast algorithm for updating the singular value decomposition. *New Haven: Yale University Department of Computer Science. RR-939*, 1993.

D. Kubus, T. Kroger, and F.M. Wahl. On-line estimation of inertial parameters using a recursive total least-squares approach. In *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*, pages 3845–3852, 9 2008.

J. Lim, N. Choi, and K. Sung. Robust recursive tls (total least square) method using regularized udu decomposed for fnn (feedforward neural network) training. In *Advances in Neural Networks*, pages 133–137. Springer, 2005.

L. Ljung. *System identification : theory for the user*. Prentice-Hall information and system sciences series. Prentice Hall, Upper Saddle River, NJ, 1999.

L. Ljung and S. Gunnarsson. Adaptation and tracking in system identification : A survey. *Automatica*, 26(1):7–21, 1990.

H.H. Madden. Comments on the savitzky-golay convolution method for least-squares-fit smoothing and differentiation of digital data. *Analytical Chemistry*, 50(9):1383–1386, 1978.

I. Markovsky and S. Van Huffel. Overview of total least-squares methods. *Signal Processing*, 87(10):2283–2302, 2007.

S. Rhode and F. Gauterin. Online estimation of vehicle driving resistance parameters with recursive least squares and recursive total least squares. In *Intelligent Vehicles Symposium (IV), 2013 IEEE*, pages 269–276, 6 2013.

A. Savitzky and M.J.E. Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8):1627–1639, 1964.

M. Schuermans, I. Markovsky, P. Wentzell, and S. Van Huffel. On the equivalence between total least squares and maximum likelihood pca. *Analytica Chimica Acta*, 544: 254–267, 2005.

T. Söderström. Errors-in-variables methods in system identification. *Automatica*, 43(6):939–958, 2007.

E.V. Zima. Fast parallel computation of the polynomial shift. In *Parallel Processing Symposium, 1997. Proceedings., 11th International*, pages 402–406, 4 1997.

Y. Zou, S.C. Chan, and T.S. Ng. A recursive least m-estimate (rlm) adaptive filter for robust filtering in impulse noise. *Signal Processing Letters, IEEE*, 7(11): 324–326, 11 2000.

Appendix A. ACRONYMS

CGII. . . . . constrained generalized inverse iteration
CGRQ.constrained generalized Rayleigh quotient
EIV. . . . . errors-in-variables
GII. . . . . generalized inverse iteration
GRQ. generalized Rayleigh quotient
GTLS. . . . generalized total least squares
i.i.d. . . . . independently identically distributed
II. . . . . inverse iteration
IV. . instrumental variables
KF. . . . . Kalman filter
LS. . . . . least squares
MISO. . . multi-input-single-output
NCE. . . . . noise covariance estimator
PKS. . . polynomial Kalman smoother
RGTLS. . recursive generalized total least squares
RLS.recursive least squares
RQ. . . . . Rayleigh quotient
RQI. . . . . Rayleigh quotient iteration
RTIV. . . . . recursive total instrumental variables
RTLS. . recursive total least squares
SGF. . Savitzky Golay filter
s.t. . . . . subject to
TLS. . . . . total least squares
WTLS. weighted total least squares

Appendix B. LIST OF SYMBOLS

$\widehat{A}$. . . . . estimated input
$\mathfrak{A}$. . . . . instruments
$\widehat{\Delta A}$. . . . estimated input noise
$\widetilde{A}$. . . . . input noise
$\overline{A}$. . . . . true input
A. . . . . measured input
$\mathcal{A}$. . . state transition matrix
$\widehat{B}$. . . . . estimated output
$\widehat{\Delta B}$. . estimated output noise
$\widetilde{B}$. . . . . output noise
$\overline{B}$. . . . . true output
B. . . . . measured output
$\mathcal{B}$. . . . . state input matrix
C. . . . . Cholesky factor
$\mathcal{C}$. . . . . measurement vector
$\mathrm{chol}(\cdot)$. . . . . Cholesky factorization
$\mathcal{D}$. . . . . measurement input matrix
$\mathrm{diag}(\cdot)$. . . diagonal elements
$\mathbb{E}$. . . . . expectation
$\odot$. . . . element-wise product
$\mathcal{E}$. . . . . squared error vector norm
$\|\cdot\|_2$. . . . . Euclidean norm
$\|\cdot\|_\mathrm{F}$. . . . . Frobenius norm
I. . . . . identity matrix
K. . . . . Kalman gain
L. . . . . correction vector
λ. . . . . forgetting factor
m. . . . . samples
$\mu(\cdot)$. . . . . arithmetic mean
n. . . . number of parameters
$\widehat{P}$. . . . . estimated noise covariance matrix
$\widehat{\Delta P}$. . noise covariance matrix
P. . . . . covariance matrix
q. . . . . $n+1$
R. . cross-correlation matrix
$\mathbb{R}$. . . . . rational numbers
S. . . . matrix of eigenvalues
$\sigma^2(\cdot)$. . . . . variance
$\mathrm{svd}(\cdot)$. . . . . singular value decomposition
t (s). . . . . time
V. . . matrix of eigenvectors
W_l. . . left weighting matrix
W_r. . right weighting matrix
w_l. . . . . left window
w_r. . . . . right window
$\widehat{X}$. . . . . parameter estimates
X. . . . . parameter
$\widehat{Z}$. . . . estimated augmented data
$\mathfrak{Z}$. . augmented instruments
$\widehat{\Delta Z}$. . . . augmented estimated noise
Z. . . . . augmented data
$\mathbb{Z}$. . . . . integer numbers