A Recursive Restricted Total Least-Squares Algorithm

Stephan Rhode, Konstantin Usevich, Ivan Markovsky, and Frank Gauterin

Abstract—We show that the generalized total least squares (GTLS) problem with a singular noise covariance matrix is equivalent to the restricted total least squares (RTLS) problem and propose a recursive method for its numerical solution. The method is based on the generalized inverse iteration. The estimation error covariance matrix and the estimated augmented correction are also characterized and computed recursively. The algorithm is cheap to compute and is suitable for online implementation. Simulation results in least squares (LS), data least squares (DLS), total least squares (TLS), and restricted total least squares (RTLS) noise scenarios show fast convergence of the parameter estimates to their optimal values obtained by corresponding batch algorithms.

Index Terms—Generalized total least squares (GTLS), recursive estimation, restricted total least squares (RTLS), subspace tracking, system identification, total least squares (TLS).

I. INTRODUCTION

MANY problems in mathematics and engineering lead to an approximate solution of an overdetermined system $AX \approx B$ of m linear equations with $n < m$ unknowns. In general, $A \in \mathbb{R}^{m \times n}$ is a matrix of measured inputs, $X \in \mathbb{R}^{n \times d}$ is a parameter matrix, and $B \in \mathbb{R}^{m \times d}$ is a matrix that contains the measured outputs. It is convenient to concatenate A and B in the augmented data matrix $Z = [A\ B] \in \mathbb{R}^{m \times q}$, where $q = n + d$. The elements of Z contain all measurements.

The stochastic model, corresponding to this problem, is the errors-in-variables (EIV) model [1], [2]:

$$A = \overline{A} + \tilde{A}, \qquad B = \overline{B} + \tilde{B}, \qquad \overline{A}\overline{X} = \overline{B}, \qquad (1)$$

where $\overline{A} \in \mathbb{R}^{m \times n}$ and $\overline{B} \in \mathbb{R}^{m \times d}$ are the true input matrix and true output matrix, respectively, and $\overline{X} \in \mathbb{R}^{n \times d}$ is the true parameter vector. In (1), the vectorized perturbation matrix $\mathrm{vec}([\tilde{A}\ \tilde{B}])$ is a random, normally distributed vector, with zero mean and covariance $\sigma^2 \widetilde{P}$, where $\widetilde{P} \in \mathbb{R}^{qm \times qm}$ is known but the scaling factor σ^2 is unknown. We use the term noise covariance matrix for $\widetilde{P}$ (although the actual noise covariance matrix is $\sigma^2 \widetilde{P}$).

In the following sections, the multi-input single-output case with $d := 1$ is considered. Hence, X becomes an $n \times 1$ parameter vector and B reduces to an $m \times 1$ vector. In the case $\widetilde{P} = I_{qm}$, the maximum-likelihood estimator for $\overline{X}$ is given by the solution of the total least-squares (TLS) problem [3]

$$\widehat{X} = \arg \min_{X, \widehat{A}, \widehat{B}} \quad \left\| [A\quad B] - [\widehat{A}\quad \widehat{B}] \right\|_{\mathrm{F}}$$
$$\text{subject to} \quad \widehat{A}X = \widehat{B}. \qquad (2)$$

(Here $\| \cdot \|_{\mathrm{F}}$ is the Frobenius norm.) The covariance of the estimation error $\overline{X} - \widehat{X}$ can also be computed from the data, see [4, Section 8.5]. The assumption $\widetilde{P} = I$, however, implies that all elements of the data matrix are measured with equal precision and the errors are uncorrelated, which is often unrealistic in practice.

For nonsingular $\widetilde{P}$, the maximum-likelihood estimator for $\overline{X}$ is the solution of the weighted total least-squares (WTLS) problem

$$\min_{X, \widehat{A}, \widehat{B}} \quad \left\| [A\quad B] - [\widehat{A}\quad \widehat{B}] \right\|_{\widetilde{P}^{-1}}$$
$$\text{subject to} \quad \widehat{A}X = \widehat{B}, \qquad (3)$$

where $\| \cdot \|_W$ is the weighted matrix norm

$$\|\Delta Z\|_W := \sqrt{\mathrm{vec}(\Delta Z)^\top W \mathrm{vec}(\Delta Z)},$$

defined for a weight matrix $W \in \mathbb{R}^{qm \times qm}$. Although problem (3) differs from (2) only by replacing the Frobenius norm $\| \cdot \|_{\mathrm{F}}$ by the weighted matrix norm, (3) does not have an analytic solution, in general, while problem (2) does.

For a general positive semi-definite covariance matrix $\widetilde{P}$, model (1) is a special case of semi-linear model in [5, Ch. 17]. Hence, the maximum likelihood estimator for $\overline{X}$ in (1) is given by the solution of

$$\min_{X, \Delta A, \Delta B} \quad \left\| [\Delta A\quad \Delta B] \right\|_{\widetilde{P}^\dagger}$$
$$\text{subject to} \quad (A + \Delta A)X = (B + \Delta B),$$
$$\mathrm{vec}\left([\Delta A\quad \Delta B]\right) \in \mathrm{image}(\widetilde{P}), \qquad (4)$$

where $\widetilde{P}^\dagger$ denotes the pseudoinverse of $\widetilde{P}$. Due to the linear constraint, problem (4) is a structured total least-squares (STLS) problem [6, Ch. 4]. Similar to (3), the STLS problem is non-convex and has no analytic solution [7].

In this paper, we consider a special case of the STLS problem (4) with $\widetilde{P}$ of the form

$$\widetilde{P} = P_r \otimes P_l, \tag{5}$$

where the right noise covariance matrix (P_r) and the left noise covariance matrix (P_l) are positive semi-definite matrices of dimension $P_r \in \mathbb{R}^{q \times q}$ and $P_l \in \mathbb{R}^{m \times m}$, respectively, and $\otimes$ is the Kronecker product. In this case, i.e., with noise covariance matrix (5), the maximum likelihood estimation problem (4) is called a generalized total least-squares (GTLS) problem.

In the special case of $P_r = I_q$ and $P_l = I_m$, the GTLS problem reduces to the TLS problem (2). The case of non-singular noise covariance matrices P_r and P_l also admits an analytic solution, see Section II. In Section III-A, we show that, in general for singular covariance matrices P_r and P_l, the GTLS problem is equivalent to the restricted total least-squares (RTLS) problem [8]

$$\min_{X,E} \quad \|E\|_F \text{ subject to } \begin{array}{l} (A + \Delta A)X = (B + \Delta B) \\ [\Delta A \quad \Delta B] = DEF^\top, \end{array} \tag{6}$$

where $E \in \mathbb{R}^{r_m \times r_n}$ is an arbitrary matrix, and $D \in \mathbb{R}^{m \times r_m}$ and $F \in \mathbb{R}^{q \times r_n}$ are full column rank matrices such that

$$P_l = DD^\top \quad \text{and} \quad P_r = FF^\top.$$

The RTLS problem can be solved off-line by the restricted singular value decomposition (RSVD) [9]. Our main contribution is the development of iterative and recursive methods for solution of RTLS in Section III-B and Section IV. The method is based on the generalized inverse iteration and uses a result from [10] that relates the RTLS problem to the minimization of a Rayleigh quotient. A similar approach for solution of nonsingular generalized singular value decomposition (GSVD) problems is used in [11]–[15], while [16], [17] use singular value decomposition update algorithms with data weighting to solve nonsingular GSVD problems recursively. In addition, we show how the uncertainty bounds of the parameter estimates and the estimated augmented correction can be computed as a byproduct of the recursive algorithm. This result is a generalization of formula (8.47) in [4, p. 242] ((7) in Section II-A of this paper). Our derivation provides an interpretation of the asymptotic covariance matrix from the point of view of the Cramér-Rao lower bound for the Markov estimator [5, Section 17.3].

In [18] the more general WTLS problem is solved with an iterative procedure based on a Newton-Gauss approach, and a solution for the computation of the uncertainty bounds is presented. However, an online implementation for this approach seems challenging.

Outline of the Paper: In Section II, we review some relevant results from the literature: computation of the TLS estimate and its covariance as well as the solution of RTLS with nonsingular noise covariance matrix by reduction to TLS. The new results are reported in Section III. In Section III-A, we show the equivalence of the GTLS with singular right covariance matrix to the RTLS problem. Then, in Section III-B, based on a lemma

from [10] we propose a recursive method for solving the RTLS problem. The method is detailed in Section IV where the computational algorithm is outlined. In Section V, simulation results illustrating the performance of the proposed algorithm are shown. Finally, conclusions are given in Section VI.

II. REVIEW OF EXISTING RESULTS

A. Computation of the TLS Estimate and Its Covariance

Consider the singular value decomposition $(\text{svd}(\cdot))$ of Z

$$Z = USV^\top, \quad \text{where } V \in \mathbb{R}^{q \times q} \text{ and } S = (S_{1,1}, \ldots, S_{q,q}).$$

Then the solution $\widehat{X}$ of (2) and the corresponding optimal augmented correction $\Delta Z := [\Delta A \ \Delta B]$ is obtained from the smallest right singular vector $(v := V_{:,q})$, where $V_{:,q}$ means last column of the matrix of eigenvectors (V):

$$\widehat{X} = -\frac{v_{1:n}}{v_q}, \qquad \Delta Z = Z(vv^\top).$$

In [4, p. 242], the following approximation of covariance matrix $\text{cov}\widehat{X}$ of the TLS estimate $\widehat{X}$ was proposed:

$$\text{cov}(\widehat{X}) \approx \left(1 + \left\|\widehat{X}\right\|_2^2\right)\widehat{\sigma}^2(A^\top A - m\widehat{\sigma}^2 I)^{-1},$$
$$\text{where } \widehat{\sigma}^2 := S_{q,q}^2/m, \quad (7)$$

B. RTLS With Nonsingular Noise Covariance Matrix

In the case of positive definite P_r and P_l, the solution of the RTLS problem can be obtained from the solution of the TLS problem by data scaling and this is then called GTLS [19, Alg. 1]. Consider the Cholesky factor of P_r (C_r), the Cholesky factor of P_l (C_l) and their inverses:

$$C_r = \text{chol}(P_r), \qquad C_l = \text{chol}(P_l),$$
$$W_r = C_r^{-1}, \qquad W_l = C_l^{-1}.$$

Then the solution of the GTLS problem can be obtained from the solution of the TLS problem for the scaled matrix

$$Z' = W_l \begin{bmatrix} A & B \end{bmatrix} \quad W_r = W_l Z W_r$$
$$\widehat{X}' = \text{total least} - \text{squares solution for } Z'$$
$$\widehat{X} = \frac{W_{r11}\widehat{X}' - W_{r12}}{W_{r22}}. \tag{8}$$

The augmented correction can be obtained as

$$\Delta Z = \Delta Z' C_r = Z'v'v'^\top C_r$$

where v' is the smallest right singular vector of Z' corresponding to the smallest singular value and

$$W_r = C_r^{-1} := \begin{array}{c} n \\ 1 \end{array} \begin{bmatrix} \overset{n}{W_{r11}} & \overset{1}{W_{r12}} \\ 0 & W_{r22} \end{bmatrix}$$

is the partition of W_r. Also, from (8), it is easy to derive the parameter covariance matrix:

$$\text{cov}(\widehat{X}) = \text{cov}\left(\frac{W_{r11}^\top \widehat{X}'}{W_{r22}}\right) = \frac{W_{r11}^\top \text{cov}(\widehat{X}')W_{r11}}{W_{r22}^2}, \tag{9}$$

where an estimate of $\text{cov}(\widehat{X}')$ can be obtained from (7).

III. NEW RESULTS

A. Restricted Total Least Squares

For the special case of the noise covariance matrix (5), problem (4) can be simplified as follows. In this instance, the constraint in (4)

$$\mathrm{vec}\left([\Delta A \quad \Delta B]\right) \in \mathrm{image}(\widetilde{P})$$

is equivalent to

$$[\Delta A \quad \Delta B] = DEF^\top.$$

Moreover, the following lemma holds true.

Lemma 1: For the noise covariance matrix of the form (5), the problem (4) is equivalent to the RTLS problem (6).
Proof: Indeed for $\Delta Z = DEF^\top$,

$$\|\Delta Z\|_{P^\dagger}^2 = \mathrm{vec}(\Delta Z)^\top (FF^\top \otimes DD^\top)^\dagger \mathrm{vec}(\Delta Z)$$
$$= \mathrm{vec}(E)^\top \mathrm{vec}(E) = \|E\|_{\mathrm F}^2,$$

which completes the proof. ∎

The solution of (6) is given by RSVD [9]. Moreover, for $P_1 = I_m$, several special cases of RTLS are well known: least squares (LS) where A is noise-free, i.e.,

$$\mathrm{diag}(P_{\mathrm r}) = \sigma^2 [0 \; \cdots \; 0 \; 1]^\top,$$

and data least squares (DLS) [20] where B is noise-free, i.e.,

$$\mathrm{diag}(P_{\mathrm r}) = \sigma^2 [1 \; \cdots \; 1 \; 0]^\top.$$

Mixed LS-TLS problems can also be solved as special cases of RTLS problem. Another special case, considered in [21], is when the data is partitioned into columns that are exactly known, and for which the covariance matrix is known up to a scalar value. However, the rigid partitioning of data in [21] seems complicated for developing a general online algorithm.

In what follows, we use a reformulation of the problem (6) as a minimization of the generalized Rayleigh quotient $(\mathrm{rq}(\cdot))$.

Lemma 2 ([10, Lemma 2.2]): Let P_1 be nonsingular. Then the problem (6) is equivalent to

$$\min_X \mathrm{rq}(X), \quad \text{where } \mathrm{rq}(X) := \frac{X_{\mathrm{ext}}^\top Z^\top P_1^{-1} Z X_{\mathrm{ext}}}{X_{\mathrm{ext}}^\top P_{\mathrm r} X_{\mathrm{ext}}},$$
$$\text{and } X_{\mathrm{ext}} := \begin{bmatrix} X \\ -1 \end{bmatrix}. \quad (10)$$

We will use a modification of generalized inverse iteration for (10) in order to develop a recursive restricted total least-squares algorithm.

B. RTLS Solved by Generalized Inverse Iteration

The generalized inverse iteration

$$\|v_t\|_2 = 1, \quad v_{t-1} = [0 \; 0 \; \cdots \; 0]^\top$$
$$\text{while } \|v_{t-1} - v_t\|_2 > \text{threshold}$$
$$v_{t-1} = v_t \quad (11\mathrm a)$$
$$v_t = (Z^\top P_1 Z)^{-1} (P_{\mathrm r} v_t) \quad (11\mathrm b)$$
$$v_t = \frac{v_t}{\|v_t\|_2} \quad (11\mathrm c)$$

end
$$\widehat{X} = -v_{1:n;t}/v_{q;t} \quad (11\mathrm d)$$

is an iterative procedure that yields the smallest right singular vector, and is strongly related to the generalized Rayleigh quotient iteration [22, p. 465]. Upon convergence, the smallest right singular vector is used in the RTLS solution (11d).

C. Computation of the Augmented Correction in Generalized Inverse Iteration

Proposition 1: For nonsingular, P_1 the augmented correction for the problem (6) is

$$\Delta Z = Z(vv^\top) \frac{P_{\mathrm r}}{v^\top P_{\mathrm r} v}. \quad (12)$$

Proof: Let $\widehat{X}$ be the solution of (10) and define $\widehat{X}_{\mathrm{ext}}$ as in (10). From (6), the ΔZ is equal to $\Delta Z = DE^* F^\top$, where E^* is the solution of

$$\min_E \quad \|E\|_{\mathrm F}$$
$$\text{subject to} \quad (Z + DEF^\top)\widehat{X}_{\mathrm{ext}} = 0,$$

which can be rewritten as

$$\min_E \quad \|E\|_{\mathrm F}$$
$$\text{subject to} \quad E(F^\top \widehat{X}_{\mathrm{ext}}) = D^{-1} Z \widehat{X}_{\mathrm{ext}}.$$

Then, the optimal E^* is given by

$$E^* = D^{-1} Z \widehat{X}_{\mathrm{ext}} \frac{(F^\top \widehat{X}_{\mathrm{ext}})^\top}{\|F^\top \widehat{X}_{\mathrm{ext}}\|_2^2} = D^{-1} Z \frac{\widehat{X}_{\mathrm{ext}} \widehat{X}_{\mathrm{ext}}^\top F}{\widehat{X}_{\mathrm{ext}}^\top P_{\mathrm r} \widehat{X}_{\mathrm{ext}}},$$

and hence

$$\Delta Z = Z \widehat{X}_{\mathrm{ext}} \widehat{X}_{\mathrm{ext}}^\top \frac{P_{\mathrm r}}{\widehat{X}_{\mathrm{ext}}^\top P_{\mathrm r} \widehat{X}_{\mathrm{ext}}}.$$

Since $\widehat{X}_{\mathrm{ext}} = \alpha v$ and $\alpha \neq 0$, the proposition is proved. ∎

D. Hessian of the Generalized Rayleigh Quotient at the Minimum Point

In this section, we will derive the Hessian of the generalized Rayleigh quotient $(\mathcal{H}(\cdot))$ at its minimum point, which will be helpful for constructing an approximation of the parameter covariance matrix.

Lemma 3: The Hessian of the generalized Rayleigh quotient at the minimum point $\widehat{X}$ is equal to

$$\mathcal{H}(\widehat{X}) = 2 \frac{A^\top P_1^{-1} A - \mathrm{rq}(\widehat{X}) P_{\mathrm r 1:n,1:n}}{\widehat{X}_{\mathrm{ext}}^\top P_{\mathrm r} \widehat{X}_{\mathrm{ext}}}$$

Proof: We introduce a notation for denominator of the generalized Rayleigh quotient $(\mathrm{rqd}(\cdot))$, i.e.

$$\mathrm{rqd}(\widehat{X}) := X_{\mathrm{ext}}^\top P_{\mathrm r} X_{\mathrm{ext}}.$$

By straightforward differentiation,

$$\nabla \mathrm{rq}(X) = \frac{2}{\mathrm{rqd}(X)} \left(-\mathrm{rq}(X) P_{\mathrm{r}:,1:n} X + A^\top P_1^{-1} Z X_{\mathrm{ext}} \right).$$

Then the Hessian $\mathcal{H}_f(\widehat{X})$ is expressed as

$$\mathcal{H}(X) = -2 \frac{\nabla^\top \mathrm{rq}(X) P_{\mathrm{r}:,1:n} X + X^\top P_{\mathrm{r}1:n,:} \nabla \mathrm{rq}(X)}{\mathrm{rqd}(X)}$$
$$+ 2 \frac{A^\top P_1^{-1} A - \mathrm{rq}(X) P_{\mathrm{r}1:n,1:n}}{\mathrm{rqd}(X)}.$$

Since $\widehat{X}$ is a stationary point of $\mathrm{rq}(\cdot)$, we have that $\nabla \mathrm{rq}(\widehat{X}) = 0$ and the lemma is proved. ∎

E. Cramer-Rao Lower Bound and Its Approximation in Generalized Inverse Iteration

The model (1) is a special case of semi-linear model of [5, Ch. 17]. Therefore, the Cramer-Rao lower bound $(\mathrm{CRB}(\cdot))$ for $\widehat{X}$ can be derived from the Hessian of the Markov estimator cost function, which is $\frac{1}{2\sigma^2} \mathrm{rq}(X)$. Then, the Cramer-Rao lower bound is given by [5, Section 17.3]

$$\mathrm{CRB}(\overline{X}) = \left(\frac{1}{2\sigma^2} \mathcal{H}(\overline{X}, \overline{Z}) \right)^{-1}. \tag{13}$$

The Markov estimator is in many cases asymptotically efficient [5] (however, it is not proven that it is asymptotically efficient in the case of RTLS); we use an approximation of $\mathrm{CRB}(\overline{X})$ as an approximation of the covariance matrix of the parameter.

Since the true parameter vector $(\overline{X})$ and the true augmented data matrix $(\overline{Z})$ are not known, similarly to [5] we replace $\overline{X}$ with X, $\overline{Z}$ with Z, and σ^2 with its approximation $\widehat{\sigma}^2$.

The resulting expression is given by

$$\mathrm{cov}(\widehat{X})$$
$$\approx \widehat{\sigma}^2 \left(\widehat{X}_{\mathrm{ext}}^\top P_{\mathrm{r}} \widehat{X}_{\mathrm{ext}} \right) (A^\top P_1^{-1} A - m\widehat{\sigma}^2 P_{\mathrm{r}1:n,1:n})^{-1}$$
$$= \widehat{\sigma}^2 \left(1 + \|\widehat{X}\|_2^2 \right) (v^\top P_{\mathrm{r}} v)$$
$$\times (A^\top P_1^{-1} A - m\widehat{\sigma}^2 P_{\mathrm{r}1:n,1:n})^{-1} \tag{14a}$$
$$:= \frac{\mathrm{rq}(\widehat{X})}{m}. \tag{14b}$$

In the case $P_1 = I$ and $P_{\mathrm{r}} = I$, the expression (14) coincides with (9). Thus, we have obtained an interpretation of (9) as an approximation of Cramer-Rao lower bound, obtained by taking the inverse of the Hessian of the generalized Rayleigh quotient at the minimum point of the generalized Rayleigh quotient. Note that in the original [4, p. 242], neither the connection to Cramer-Rao lower bound $(\mathrm{CRB}(\cdot))$ or $\mathcal{H}(\cdot)$ was mentioned, nor any other interpretation was provided.

F. Exponential Forgetting and Covariance Matrix

In recursive algorithms, it is customary to use exponential scaling of the data. For a forgetting factor (λ), $0 < \lambda \leq 1$, the following left scaling matrix is constructed:

$$\mathrm{diag}(W_l) := [\sqrt{\lambda}^{m-1} \cdots \sqrt{\lambda}^2 \sqrt{\lambda}^1 \sqrt{\lambda}^0]^\top. \tag{15}$$

Now consider the GTLS problem with $P_1^{-1} = W_l^\top W_l$, and $P_{\mathrm{r}} \in \mathbb{R}^{q \times q}$. In this case, since P_1 is nonsingular, so that generalized inverse iteration (GII) can be used to compute estimated parameter vector and (12) can be used to compute augmented correction.

However, if the data is generated according to the model

$$Z = \overline{Z} + \widetilde{Z},$$

where $\widetilde{Z} \sim N(0, \sigma^2 P_{\mathrm{r}})$, $\widetilde{Z}$ are i.i.d., the left scaling matrix defined in (15) is not consistent with the noise model. Thus the formula (14) is no longer valid for computing the parameter covariance matrix. In order to address this problem, we first consider the case of the least squares estimator with exponential forgetting.

Lemma 4: Let P_{r} be a diagonal matrix such that $\mathrm{diag}(P_{\mathrm{r}}) = [0, \ldots, 0, 1]^\top$, and W_l is given by (15). Then

$$\mathrm{cov}(\widehat{X}) = \sigma^2 (A^\top P_1^{-1} A)^{-1} (A^\top P_1^{-2} A) (A^\top P_1^{-1} A)^{-1}. \tag{16}$$

Proof: This is a special case of the weighted least squares estimator, and (16) follows from [23, (2.16)]. ∎

Remark 1: If W_l was consistent with the noise model (i.e. if $\mathrm{vec}(\widetilde{B}) \sim N(0, \sigma^2 P_{\mathrm{r}} \otimes P_1)$), then the covariance of $\widehat{X}$ would be equal to $\sigma^2 (A^\top P_1^{-1} A)^{-1}$, which is a special case of (14). Based on Lemma 4 and Remark 1, we propose the following adjustment of (14) to the case of exponential forgetting:

$$\mathrm{cov}(\widehat{X}) \approx \widehat{\sigma}^2 \left(\widehat{X}_{\mathrm{ext}}^\top P_{\mathrm{r}} \widehat{X}_{\mathrm{ext}} \right)$$
$$\times \left(R_{1:n,1:n} - m_1 \widehat{\sigma}^2 P_{\mathrm{r}1:n,1:n} \right)^{-1}$$
$$\times \left(Q_{1:n,1:n} - m_2 \widehat{\sigma}^2 P_{\mathrm{r}1:n,1:n} \right)$$
$$\times \left(R_{1:n,1:n} - m_1 \widehat{\sigma}^2 P_{\mathrm{r}1:n,1:n} \right)^{-1}$$
$$\widehat{\sigma}^2 := \frac{\mathrm{rq}(\widehat{X})}{m_1}, \quad m_1 := \sum_{j=0}^{m-1} \lambda^j, \quad m_2 := \sum_{j=0}^{m-1} \lambda^{2j}. \tag{17}$$

where

$$R := Z^\top P_{\mathrm{r}}^{-1} Z, \quad Q := Z^\top P_{\mathrm{r}}^{-2} Z, \tag{18}$$

and therefore

$$R_{1:n,1:n} = A^\top P_{\mathrm{r}}^{-1} A, \quad Q_{1:n,1:n} := A^\top P_{\mathrm{r}}^{-2} A.$$

For $\lambda = 1$, $m = m_1 = m_2$. For $\lambda < 1$, they are

$$m_1 = \frac{1 - \lambda^m}{1 - \lambda}, \quad m_2 = \frac{1 - \lambda^{2m}}{1 - \lambda^2}.$$

The number m_1 is an equivalent window length (the multiplier for the noise variance). For $m \to \infty$, m_1 converges to $\frac{1}{1-\lambda}$ (or to ∞ if $\lambda = 1$). In recursive algorithms, it is common to replace m_1 with $\min\left(m, \frac{1}{1-\lambda}\right)$, however it may be a rough approximation for small m. m_2 is the equivalent window length for (Q).

IV. RECURSIVE RESTRICTED TOTAL LEAST SQUARES

The recursive restricted TLS (recursive RTLS) algorithm described in this section is designed to solve the RTLS problem online. At each step of the recursion (for each t), a new row-vector $Z_t \in \mathbb{R}^{1 \times q}$ is supplied. The algorithm computes an estimate $\widehat{X}_t$ for the RTLS problem with the augmented data matrix $Z_{1:t} := [Z_1^\top \cdots Z_t^\top]^\top$.

In order to allow the algorithm to follow time-varying parameters, an forgetting factor is used such that the (W_1) is defined as in (15). The right noise covariance matrix is also allowed to be time-varying and is represented by $P_{r_t} \in \mathbb{R}^{q \times q}$.

The algorithm does not store the whole matrix $Z_{1:t}$. Instead, it has access to the previous estimate $\widehat{X}_{t-1}$ and to the inverse of the data covariance matrix P_{t-1}, where $P_j = R_j^{-1}$ and R_j is the data covariance matrix defined in (18). For computation of the covariance of $\widehat{X}$, the algorithm has access to R_{t-1} and Q_{t-1} (data covariance matrix and doubly scaled data covariance matrix defined in (18)). In addition to $\widehat{X}_t$, the algorithm computes the updated matrix P_t, the parameter covariance matrix and the last row $\Delta Z_t \in \mathbb{R}^{1 \times q}$ of the augmented correction. The algorithm computes updated matrices R_t and Q_t, and also the estimated variance.

The recursion starts from the time instant $t = q+1$, where the algorithm is initialized using the batch solution of RTLS. However, other typical initializations are possible, such as $\widehat{X}_{t-1} = [0 \cdots 0]^\top$, $P_{t-1} = 100I$, $R_{t-1} = P_{t-1}^{-1}$ and $Q_{t-1} = R_{t-1}$, see the minimal working example of recursive RTLS in the supplementary material of this article.

$$
\text{input}: Z_t, P_{t-1}, \widehat{X}_{t-1}, P_{r_t}, \lambda, t, R_{t-1}, Q_{t-1}
$$

$$
L_t = \left(P_{t-1} Z_t^\top \right) \left(\lambda + Z_t P_{t-1} Z_t^\top \right)^{-1} \tag{19a}
$$

$$
P_t = (I - L_t Z_t) P_{t-1} \frac{1}{\lambda} \tag{19b}
$$

$$
v'_{t-1} = \left[\widehat{X}_{t-1}^\top \quad -1 \right]^\top \tag{19c}
$$

$$
v'_t = P_t \left(P_{r_t} v'_{t-1} \right) \tag{19d}
$$

$$
\widehat{X}_t = -v'_{1:n;t}/v'_{q;t} \tag{19e}
$$

$$
v_t = \left[\widehat{X}_t^\top, -1 \right]^\top \left(1 + \widehat{X}_t^\top \widehat{X}_t \right)^{-\frac{1}{2}} \tag{19f}
$$

$$
\Delta Z_t = Z_t \left(v_t v_t^\top \right) \frac{P_{r_t}}{v_t^\top P_{r_t} v_t} \tag{19g}
$$

$$
R_t = \lambda R_{t-1} + Z_t^\top Z_t, \quad Q_t = \lambda^2 Q_{t-1} + Z_t^\top Z_t, \tag{19h}
$$

$$
m_1 = \frac{1 - \lambda^t}{1 - \lambda}, \quad m_2 = \frac{1 - \lambda^{2t}}{1 - \lambda^2}, \tag{19i}
$$

$$
\widehat{\sigma}_t^2 = \frac{1}{m_1} \frac{v_t^\top P_t^{-1} v_t}{v_t^\top P_{rt} v_t} \tag{19j}
$$

$$
\begin{aligned}
\mathrm{cov}(\widehat{X}_t) = {} & \left(1 + \|\widehat{X}_t\|_2^2 \right) \widehat{\sigma}_t^2 \left(v_t^\top P_{r_t} v_t \right) \\
& \times \left(R_{1:n,1:n;t} - m_1 \widehat{\sigma}_t^2 P_{r1:n,1:n;t} \right)^{-1} \\
& \times \left(Q_{1:n,1:n;t} - m_2 \widehat{\sigma}_t^2 P_{r1:n,1:n;t} \right) \\
& \times \left(R_{1:n,1:n;t} - m_1 \widehat{\sigma}_t^2 P_{r1:n,1:n;t} \right)^{-1}
\end{aligned}
$$

$$
\text{output}: \widehat{X}_t, P_t, \mathrm{cov}(\widehat{X}_t), \Delta Z_t, R_t, Q_t, \widehat{\sigma}_t^2 \tag{19k}
$$

The recursive RTLS algorithm in (19a)–(19k) consists of three parts. First, the inverse of the data covariance matrix P_{t-1}

TABLE I
SUMMARY OF TLS PROBLEMS

problem (acronym)	definition	analytic solution
total least squares (TLS)	(2)	yes
weighted total least squares (WTLS)	(3)	no
structured total least squares (STLS)	—	no
generalized total least squares (GTLS)	(3) + (5)	yes
restricted total least squares (RTLS)	(6)	yes

is recursively updated with the matrix inversion lemma in (19a), (19b), [24, p. 364]. Second, in (19c)–(19d), a one-step generalized inverse iteration is performed (a modification of (11a)–(11c) with the previous estimate $\widehat{X}_{t-1}$ used to build the smallest right singular vector v'_{t-1}). The generalized inverse iteration is completed by computing $\widehat{X}$ with the non-normalized smallest right singular vector v'_t in (19e).

Using one-step generalized inverse iteration ((19c), (19d)) instead of running the whole GII recursion was initially proposed in [14]. Although it saves memory because there is no need to track the time-varying smallest right singular vector separately, the algorithm loses optimality compared with the batch procedure in (11a)–(11c). Notably, this fact is often not emphasized in literature about similar algorithms.

The steps (19a)–(19e) of the recursive RTLS algorithm were already introduced in [25]. The final part (19f)–(19k) of the algorithm extends the algorithm of [25] to the computation of the augmented correction in (19g), and the parameter covariance matrix in (19k). These computations are performed using (12) and (17).

The recursive RTLS algorithm in (19a)–(19k) has a computational complexity of $O(q^3)$ due to matrix inversion in (19k). However, the computational complexity may be reduced to $O(q^2)$ using the matrix inversion lemma also for (19k). This would require additional input and output parameters.

The GTLS estimator of Section II-B is of much higher computational complexity. The involved singular value decomposition of the m by q matrix Z requires $O(mq^2 + q^3)$ multiplications, see [22, p. 254] for different complex svd implementations. In addition to this, the computational complexity of GTLS is unbounded because m grows for each t to ∞, which is typical for batch estimators.

V. SIMULATION RESULTS

A. Simulation Data and Performance Index

The proposed recursive RTLS algorithm was extensively compared with appropriate batch estimators in four different right noise covariance matrix scenarios. Table II links the used optimal batch estimator, the diagonal elements of P_r (the noise variance) and the name of the Matlab function from the supplementary material of this article for each P_r scenario. We used the GTLS estimator of Section II-B for P_r scenario #1 and #2. The LS estimator was used from [26, pp. 279–281] in P_r scenario #3 and the DLS estimator from [20] in P_r scenario #4. For more details, we refer the reader to the provided Matlab implementation that contains all required steps (simulation data, estimator implementation, result figures) to rerun the experiments.

TABLE II
OVERVIEW OF SIMULATION EXPERIMENTS

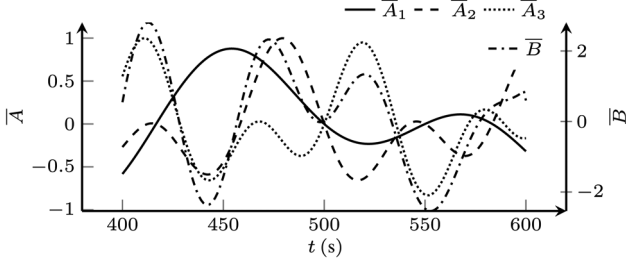| #	optimal estimator	$\text{diag}(P_r)$	function
1	RTLS Section II-B	$\begin{bmatrix} 0.1 & 0.2 & 0.4 & 1 \end{bmatrix}^\top$	fcn_gtls
2	TLS Section II-B	$\begin{bmatrix} 0.3 & 0.3 & 0.3 & 0.3 \end{bmatrix}^\top$	fcn_gtls
3	LS [26, 279–281]	$\begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}^\top$	fcn_wls
4	DLS [20]	$\begin{bmatrix} 0.3 & 0.3 & 0.3 & 0 \end{bmatrix}^\top$	fcn_dls

Fig. 1. True inputs ($\overline{A}$) and true output ($\overline{B}$) modeled as sine waves, [25].

Fig. 2. Estimated parameter vector ($\widehat{X}$) of one experiment for right noise co-variance matrix (P_r) scenario #1 of Table II. The true parameters ($\overline{X}$) are shown by ———.

Fig. 3. Estimated parameter covariance ($\text{cov}(\widehat{X})$) of one experiment for right noise covariance matrix (P_r) scenario #1 of Table II.

Analogically to [25] Fig. 1 shows $\overline{A}$ and $\overline{B}$ that were generated with

$$t = [1, 2, 3, \ldots, m]^\top, \quad m = 1000 \text{ s},$$

$$\overline{A}_{:,1} = \sin(2\pi t 0.006) \sin\left(\frac{2\pi t 0.006}{3.3}\right),$$

$$\overline{A}_{:,2} = \sin(2\pi t 0.012) \sin\left(\frac{2\pi t 0.012}{3.3}\right),$$

$$\overline{A}_{:,3} = \sin(2\pi t 0.014) \sin\left(\frac{2\pi t 0.014}{3.3}\right),$$

$$\overline{X} = \begin{cases} \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}^\top & 1 \text{ s} \leq t < 600 \text{ s} \\ \begin{bmatrix} 2 & 2 & 3 \end{bmatrix}^\top & 600 \text{ s} < t \leq 1000 \text{ s}, \end{cases}$$

$$\overline{B} = \overline{A} \odot \overline{X},$$

where $\overline{A}_{:,1}, \overline{A}_{:,2}$ and $\overline{A}_{:,3}$ are the first, second and third columns of $\overline{A}$ respectively, and $\odot$ is the element-wise product, also known as Hadamard product. The step change in $\overline{X}_1$ at $t = 600$ second was used to check the recursive RTLS algorithm tracking performance for time-varying systems.

For each P_r scenario of Table II, 500 independent experiments were conducted with additive zero mean white Gaussian noise, with covariance $\sigma^2 P_r$, where $\sigma^2 = 1$. Such a choice of σ^2 leads to signal-to-noise ratio (SNR) of 6.84, 8.26, 9.36, and 9.45 dB respectively for each noise scenario. Therefore we consider the case of poor SNR. Note that the SNR is computed as $20 \text{ dB} \cdot \log_{10} \frac{\text{rms}(\text{vec}(\overline{Z}))}{\text{rms}(\text{vec}(\widetilde{Z}))}$. Thus, the presented SNR for P_r scenario #3 and #4 are artificially increased, since we do not remove the exactly known columns of the augmented data matrix from the calculation.

The forgetting factor was set to $\lambda = 0.997$ for all estimators. While λ is directly involved in the recursive restricted TLS (recursive RTLS) algorithm (19a)–(19k), the optimal batch estimators w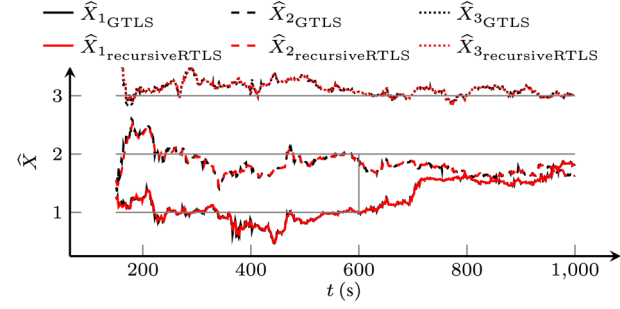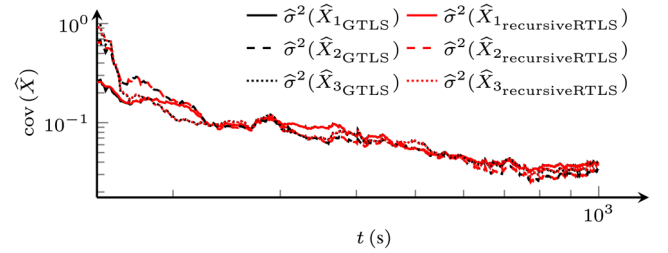ere applied for each t with row-wise scaled data $W_l Z$, where W_l was adjusted as shown in (15). The recursive RTLS algorithm was initialized with the corresponding batch estimator from Table II at $t = 4$ second.

Similarly to [11], [14], the performance index squared error vector norm ($\mathcal{E}$) was computed for each estimator, right noise covariance matrix scenario and experiment, and averaged with the arithmetic mean ($\mu(\cdot)$) to compute the expectation of $\mathcal{E}$ over 500 independent experiments

$$\mathcal{E}_t = \|\widehat{X}_t - \overline{X}_t\|_2^2.$$

B. Nonsingular Right Noise Covariance Matrix Scenario #1

Fig. 2 shows the parameter vector estimates of P_r scenario #1 for the optimal GTLS estimator and the recursive RTLS algorithm. For $t \geq 200$ second recursive RTLS follows GTLS quite accurately, despite the fact that recursive RTLS is subop-timal (only one step of GII is performed). In addition recursive RTLS, algorithm shows smoother results than the optimal GTLS estimator, which can be explained by the fact that the solution is updated recursively. This implicit regularization of recursive RTLS can be beneficial in practical applications where a smooth solution is desirable. Similar observations hold for the diagonal elements of the estimated parameter covariance in Fig. 3.

Fig. 4, shows the squared error vector norm averaged over 500 experiments in dB. Between 150 s to 200 s we can observe large deviations between recursive RTLS and GTLS which decay for $t \geq 200$ s, which indicates that the initialization of recursive RTLS is finished after approximately 200 s. More-over, after the initialization ($t \geq 200$ s) the difference between the optimal GTLS estimator and the proposed recursive RTLS algorithm is negligible. The step change in $\overline{X}_1$ at $t = 600$ s can be clearly seen as drastic increase in $\mathcal{E}$. However, the latter
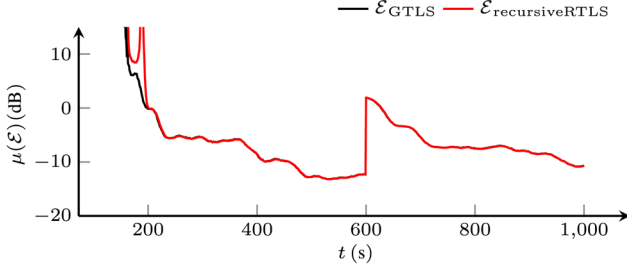
Fig. 4. Squared error vector norm averaged over 500 independent noise realizations for right noise covariance matrix (P_r) scenario #1 of Table II.
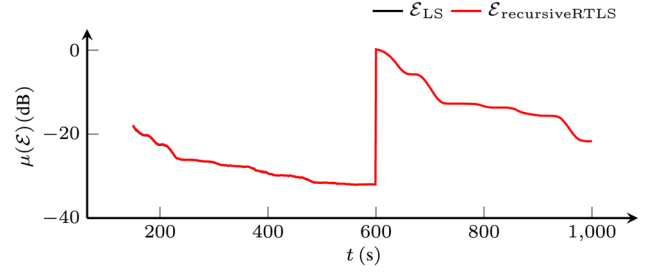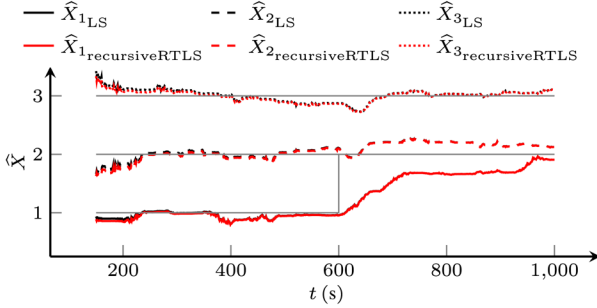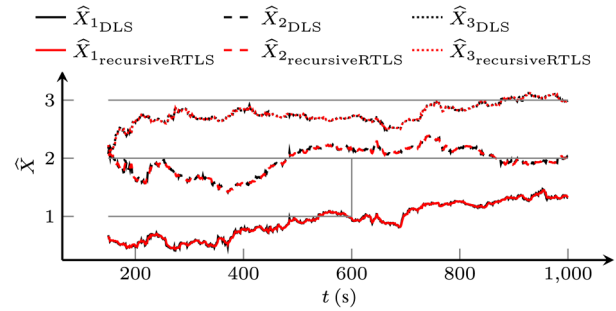

Fig. 5. Estimated parameter vector ($\widehat{X}$) of one experiment for right noise covariance matrix (P_r) scenario #3 of Table II. The true parameters ($\overline{X}$) are shown by ———.
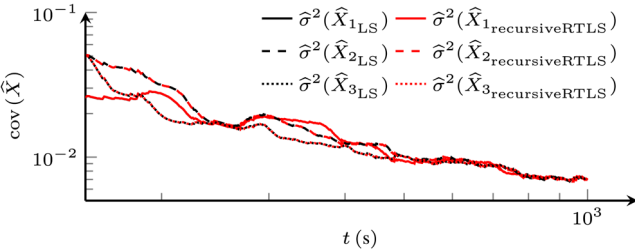

Fig. 6. Estimated parameter covariance ($\mathrm{cov}(\widehat{X})$) of one experiment for right noise covariance matrix (P_r) scenario #3 of Table II.

decrease of $\mathcal{E}$ shows the adaptability properties of the recursive RTLS algorithm with forgetting factor.

C. Nonsingular Right Noise Covariance Matrix Scenario #2

As discussed in Section II-B, TLS is included in GTLS as a special case. Hence, the performance of recursive RTLS compared with the optimal TLS estimator does not differ significantly from the statements made in Section V-B, and we omit the results for this P_r scenario. However, these results can be easily observed by using the mentioned Matlab implementation.

D. Singular Right Noise Covariance Matrix Scenario #3

Let us compare the results of recursive RTLS with the optimal LS estimator for P_r scenario #3. Fig. 5 shows the improved accuracy of recursive RTLS in the parameter vector estimates for this scenario compared to the results in Fig. 2. There is practically no initialization of recursive RTLS observable in Fig. 5. The performance of recursive RTLS with regards to the estimated parameter covariance in Fig. 6 is also improved compared

Fig. 7. Squared error vector norm averaged over 500 independent noise realizations for right noise covariance matrix (P_r) scenario #3 of Table II.

Fig. 8. Estimated parameter vector ($\widehat{X}$) of one experiment for right noise covariance matrix (P_r) scenario #4 of Table II. The true parameters ($\overline{X}$) are shown by ———.

with Fig. 3. Fig. 7 shows squared error vector norm $\mathcal{E}$ averaged over 500 experiments where recursive RTLS performs with high accuracy compared to the optimal LS estimator.

E. Singular Right Noise Covariance Matrix Scenario #4

Fig. 8 shows the parameter vector estimates of P_r scenario #4 for the optimal DLS estimator and the recursive RTLS algorithm. The accuracy of recursive RTLS for this P_r scenario seems appropriate for the most practical applications. Also, the implicit regularization of recursive RTLS is satisfactory observable in Fig. 8. We do not yet have an implementation for the estimated parameter covariance of DLS, therefore the input correction (ΔA) is shown in Fig. 9 for 150 s to 200 s. Recursive RTLS shows similar results as the optimal DLS solution also in terms of the input correction. The initialization between 150 s to 200 s, as well as the implicit regularization of recursive RTLS (250 s to 400 s) is satisfactory, observable by the averaged squared error vector norm in Fig. 10.

Finally, Fig. 11 gives the averaged squared error vector norm where the sine waves of Fig. 1 were substituted with white Gaussian noise with zero mean and unit variance for $\overline{A}$. The SNR is approximately 13 dB in this case. Compared with Fig. 10 the squared error vector norm is drastically reduced and there is no initialization or deviation between recursive RTLS and DLS observable. A possible explanation is that $\overline{A}$ built from white Gaussian noise gives better system excitation than $\overline{A}$ built from sine waves in Fig. 1.

For the purpose of brevity, only an excerpt of the simulation results was provided, and we encourage the reader to rerun the experiments with the provided Matlab implementation.
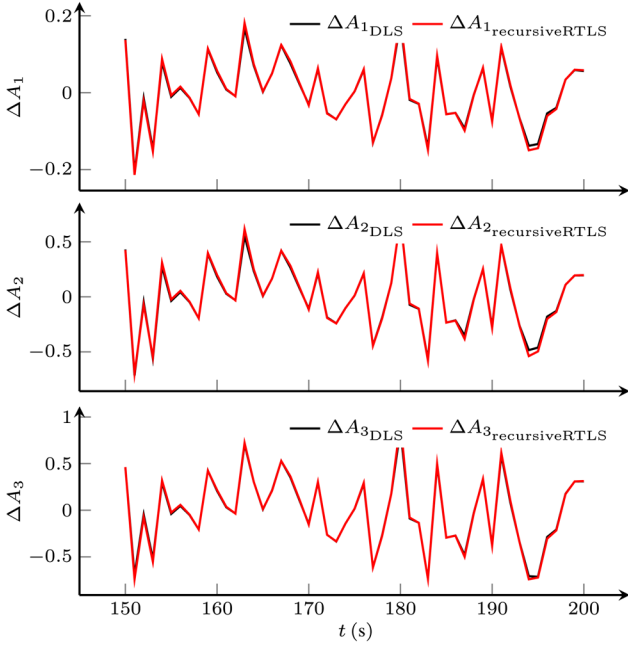
Fig. 9. Input correction (ΔA) of one experiment for right noise covariance matrix (P_r) scenario #4 of Table II.

Fig. 10. Squared error vector norm averaged over 500 independent noise realizations for right noise covariance matrix (P_r) scenario #4 of Table II.
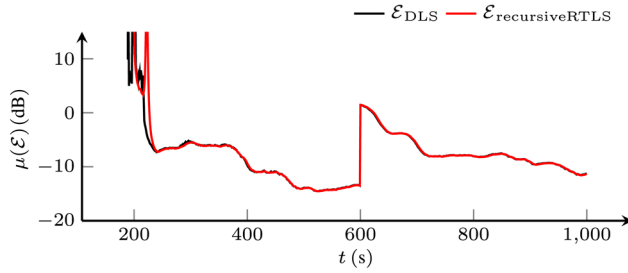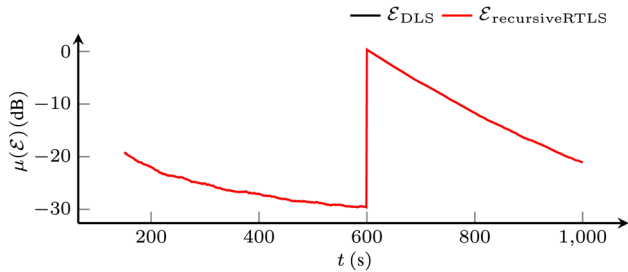

Fig. 11. Squared error vector norm averaged over 500 independent noise realizations for right noise covariance matrix (P_r) scenario #4 of Table II. Contrary to all previous results, the true inputs ($\overline{A}$) were built from white Gaussian noise with zero mean and unit variance.

F. Sample Covariance and Cramer-Rao Lower Bound

Finally, we compare the estimated covariance matrix (17) with sample covariance matrix and the Cramer-Rao lower bound. We consider the P_r scenario #3, but use the time-invariant case $\overline{X} = [1\ 2\ 3]^\top$ (without the step in the parameters). In this case, from (13) the $\mathrm{CRB}(\cdot)$ can be computed as

$$\mathrm{CRB}(\overline{X})_t = \sigma^2 \left(X_\mathrm{ext}^\top P_\mathrm{r} X_\mathrm{ext} \right) \left(\overline{A}_{1:t,:}^\top \overline{A}_{1:t,:} \right)^{-1},$$
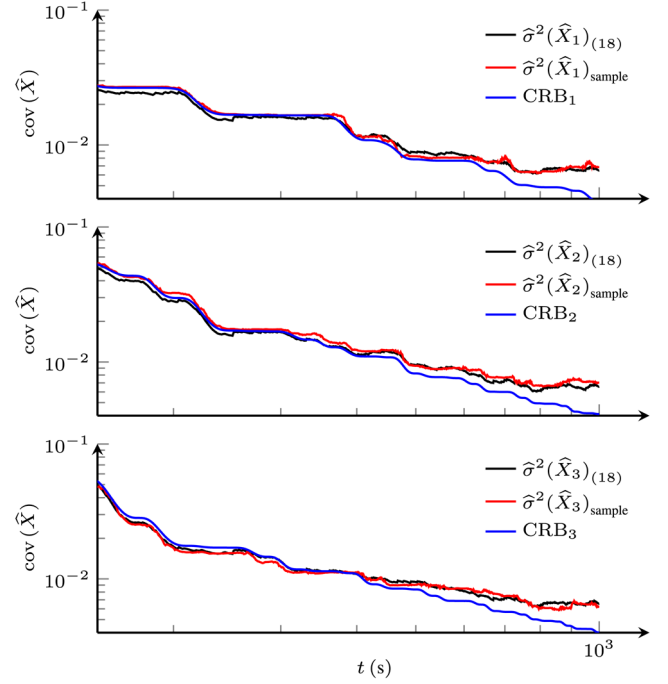

Fig. 12. Estimated parameter covariance ($\mathrm{cov}(\widehat{X})$) for right noise covariance matrix (P_r) scenario #3 of Table II. Note that the step change in $\overline{X}_1$ is omitted here for a better comparison with the Cramer-Rao lower bound. Hence, $\overline{X}$ is time-invariant ($\overline{X} = [1\ 2\ 3]^\top$). Black lines show the recursive RTLS result from (17) of one experiment, red lines the sample covariance obtained from 500 independent recursive RTLS parameter estimates and blue lines the Cramer-Rao lower bound.

which corresponds to the covariance matrix of estimated parameter vector ($\widehat{X}$) for the batch estimator without exponential forgetting.

In Fig. 12, we compare the diagonal elements of the theoretical parameter covariance matrix of (17) for one experiment with the diagonal elements of the sample covariance matrix of 500 independent parameter estimates for recursive RTLS, and the diagonal elements of the $\mathrm{CRB}(\overline{X})_t$. It can be seen that the sample covariance matches the estimated covariance quite well, thus the approximation (17) of the covariance matrix is quite accurate. (This was also verified for other noise scenarios.)

Compared to $\mathrm{CRB}(\cdot)$, we see that covariance of the parameter is bounded from below, whereas the $\mathrm{CRB}(\cdot)$ converges to zero. Due to the exponential forgetting, only a part of data (corresponding to the equivalent window length) is used, and thus the recursive RTLS is not consistent with $t \to \infty$. This is a typical feature of the recursive estimators that consistency is sacrificed for the ability to adapt to changes in the parameters.

VI. CONCLUSION

Starting with a short review of known errors-in-variables (EIV) estimators of the total least squares (TLS) type we have shown that generalized total least squares (GTLS) is equivalent to the restricted total least-squares (RTLS) problem with singular ($\widetilde{P}$). The proposed recursive restricted total least-squares (recursive RTLS) algorithm is based upon the generalized inverse iteration (GII), and provides the estimated parameter vector ($\widehat{X}$), along with a solution for the estimated parameter covariance matrix and the estimated augmented correction.

Compared with the corresponding optimal batch estimators, the suboptimal recursive RTLS algorithm has significantly lower computational complexity ($O(q^3)$ that can be reduced to $O(q^2)$ multiplications per iteration against $O(q^3 + mq^2)$, with $m \rightarrow \infty$). Moreover, numerous simulation experiments suggest that the recursive RTLS algorithm provides good estimates of the parameters, which are smoother than the batch estimates. We also proposed a formula for covariance of the parameters, which was shown to be accurate by the simulation experiments. Finally, due to forgetting factor, the algorithm is able to follow time-varying parameters. As it is common for recursive estimators, this is achieved at a price of reduced efficiency of the estimate compared to Cramer-Rao lower bound. The recursive RTLS algorithm and all presented results are available as a ready to run Matlab implementation in the supplementary material.

APPENDIX A

LIST OF SYMBOLS

A	measured input
$\widehat{A}$	estimated input
$\tilde{A}$	input noise
$\overline{A}$	true input matrix
B	measured output
$\widehat{B}$	estimated output
$\widetilde{B}$	output noise
$\overline{B}$	true output matrix
$\mathrm{chol}(\cdot)$	Cholesky factorization
C_1	Cholesky factor of P_1
$\mathrm{cov}(\cdot)$	covariance operator
C_r	Cholesky factor of P_r
$\mathrm{CRB}(\cdot)$	Cramer-Rao lower bound
d	number of outputs
ΔA	input correction
ΔB	output correction
ΔZ	augmented correction
$\mathrm{diag}(\cdot)$	diagonal elements
$\mathbb{E}$	expectation operator
$\mathcal{E}$	squared error vector norm
$\| \cdot \|_\mathrm{F}$	Frobenius norm
$\mathcal{H}(\cdot)$	Hessian of the generalized Rayleigh quotient
I	identity matrix
$\mathrm{image}(\cdot)$	image operator
$\otimes$	Kronecker product

L	correction vector
λ	forgetting factor
m	samples
$\mu(\cdot)$	arithmetic mean
r_m	rank of P_1
n	number of parameters
r_n	rank of P_1
$\| \cdot \|_2$	Euclidean norm
$O(\cdot)$	computational complexity
P	inverse of the data covariance matrix
P_1	left noise covariance matrix
$\widetilde{P}$	noise covariance matrix
P_r	right noise covariance matrix
$\odot$	element-wise product
Q	doubly scaled data covariance matrix
q	$n + d$
R	data covariance matrix
$\mathbb{R}$	rational numbers
$\mathrm{rq}(\cdot)$	generalized Rayleigh quotient
$\mathrm{rqd}(\cdot)$	denominator of the generalized Rayleigh quotient
E	middle factor in restricted TLS
D	left factor in restricted TLS
F	right factor in restricted TLS
S	matrix of eigenvalues
$\mathrm{svd}(\cdot)$	singular value decomposition
$t(\mathrm{s})$	time
U	matrix of left-singular vectors
V	matrix of eigenvectors
v	smallest right singular vector
$\sigma^2(\cdot)$	variance
$\widehat{\sigma}^2(\cdot)$	estimated variance
$\mathrm{vec}(\cdot)$	vectorization operator
W	scaling matrix
W_1	left scaling matrix
W_r	right scaling matrix
X	parameter vector
X_ext	extended parameter vector
$\widehat{X}$	estimated parameter vector
X_ext	extended estimated parametervector
$\overline{X}$	true parameter vector

Z	augmented data matrix
$\widetilde{Z}$	noise on augmented correction
$\overline{Z}$	true augmented data matrix

APPENDIX B
ACRONYMS

DLS	data least squares
EIV	errors-in-variables
GII	generalized inverse iteration
GRQI	generalized Rayleigh quotientiteration
GSVD	generalized singular valuedecomposition
GTLS	generalized TLS
LS	least squares
recursive RTLS	recursive restricted TLS
RSVD	restricted singular value decomposition
RTLS	restricted TLS
SNR	signal-to-noise ratio
STLS	structured TLS
TLS	total least squares
WTLS	weighted TLS

REFERENCES

[1] W. Fuller, *Measurement Error Models*. New York, NY, USA: Wiley, 1987.

[2] T. Söderström, "Errors-in-variables methods in system identification," *Automatica*, vol. 43, pp. 939–958, 2007.

[3] I. Markovsky and S. Van Huffel, "Overview of total least-squares methods," *Signal Process.*, vol. 87, no. 10, pp. 2283–2302, 2007.

[4] S. Van Huffel and J. Vandewalle, *The Total Least Squares Problem: Computational Aspects and Analysis*, ser. Frontiers in applied mathematics. Philadelphia, PA, USA: SIAM, 1991, vol. 9.

[5] R. Pintelon and J. Schoukens, *System Identification: A Frequency Domain Approach*, 2nd ed. New York, NY, USA: Wiley, 2012.

[6] I. Markovsky, J. C. Willems, B. De Moor, and S. Van Huffel, *Exact and Approximate Modeling of Linear Systems: A Behavioral Approach*, ser. Monographs on Mathematical Modeling and Computation. Philadelphia, PA, USA: SIAM, 2006, no. 11.

[7] I. Markovsky, "Structured low-rank approximation and its applications," *Automatica*, vol. 44, no. 4, pp. 891–909, 2008.

[8] S. Van Huffel and H. Zha, "The restricted total least squares problem: Formulation, algorithm and properties," *SIAM J. Matrix Anal. Appl.*, vol. 12, no. 2, pp. 292–309, 1991.

[9] H. Zha, "The restricted singular value decomposition of matrix triplets," *SIAM J. Matrix Anal. Appl.*, vol. 12, no. 1, pp. 172–194, 1991.

[10] A. Beck, "The matrix-restricted total least squares problem," *Signal Process.*, vol. 87, no. 10, pp. 2303–2312, 2007.

[11] C. Davila, "An efficient recursive total least squares algorithm for fir adaptive filtering," *IEEE Trans. Signal Process.*, vol. 42, no. 2, pp. 268–280, 2 1994.

[12] D.-Z. Feng, X.-D. Zhang, D.-X. Chang, and W. Zheng, "A fast recursive total least squares algorithm for adaptive fir filtering," *IEEE Trans. Signal Process.*, vol. 52, no. 10, pp. 2729–2737, Oct. 2004.

[13] 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*. New York, NY, USA: Springer, 2005, pp. 133–137.

[14] D.-Z. Feng and W. Zheng, "Fast approximate inverse power iteration algorithm for adaptive total least-squares fir filtering," *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 4032–4039, 10 2006.

[15] R. Arablouei and K. Dogancay, "Linearly-constrained recursive total least-squares algorithm," *IEEE Signal Process. Lett.*, vol. 19, no. 12, pp. 821–824, Dec. 2012.

[16] D. Kubus, T. Kroger, and F. Wahl, "On-line estimation of inertial parameters using a recursive total least-squares approach," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Systems (IROS)*, Sep. 09 2008, pp. 3845–3852.

[17] S. Rhode and F. Gauterin, "Online estimation of vehicle driving resistance parameters with recursive least squares and recursive total least squares," in *Proc. IEEE Intell. Veh. Symp. (IV)*, Jun. 2013, pp. 269–276.

[18] Y. Shen, B. Li, and Y. Chen, "An iterative solution of weighted total least-squares adjustment," *J. Geodesy*, vol. 85, pp. 229–238, 2011.

[19] M. Schuermans, I. Markovsky, P. Wentzell, and S. Van Huffel, "On the equivalence between total least squares and maximum likelihood PCA," *Analytica Chimica Acta*, vol. 544, pp. 254–267, 2005.

[20] R. DeGroat and E. Dowling, "The data least squares problem and channel equalization," *IEEE Trans. Signal Process.*, vol. 41, no. 1, pp. 407–411, 1993.

[21] S. Van Huffel and J. Vandewalle, "Analysis and properties of the generalized total least squares problem $ax \approx b$ when some or all columns in a are subject to error," *SIAM J. Matrix Anal. Appl.*, vol. 10, no. 3, pp. 294–315, 1989.

[22] G. Golub and C. Van Loan, *Matrix Computations*. Baltimore, MD, USA: Johns Hopkins Univ. Press, 1996, vol. 3.

[23] H. W. Sorenson, *Parameter Estimation: Principles and Problems*. New York, NY, USA: Marcel Dekker, 1980.

[24] L. Ljung, *System Identification: Theory for the User*, ser. Prentice-Hall information and system sciences series. Upper Saddle River, NJ, USA: Prentice-Hall, 1999.

[25] S. Rhode, F. Bleimund, and F. Gauterin, "Recursive generalized total least squares with noise covariance estimation," in *IFAC World Congr.*, 2014, accepted for publication.

[26] R. Isermann and M. Münchhof, *Identification of Dynamic Systems: An Introduction With Applications*. Heidelberg, Germany: Springer, 2011.

Stephan Rhode received a Dipl.-Ing. (DH) degree (B.Eng.) in Mechanical Engineering in 2004 from Baden-Württemberg Cooperative State University, Karlsruhe, Germany. During 2004–2009 he worked as a development engineer in mechanical design of electric servo motors at SEW Eurodrive, Bruchsal, Germany. He received an M.Sc. degree in Mechanical Engineering and Mechatronics from the University of Applied Science Karlsruhe, Germany. In September 2009 he joined the Institute of Vehicle System Technology at the Karlsruhe Institute of Technology as a Ph.D. student. His research interests are system identification, errors-in-variables, and robust recursive identification methods with applications to vehicle dynamics.

Konstantin Usevich obtained a specialist degree in 2007 and Ph.D. degree in 2011, both from the Department of Statistical Modeling, St. Petersburg State University, Russia. He held postdoctoral research positions at the University of Southampton, U.K., and at the Department ELEC, Vrije Universiteit Brussel, Belgium. Since July 2014, he is a postdoctoral researcher at GIPSA-lab, Centre National de la Recherche Scientifique, France. His research interests are in low-rank approximation of structured matrices, system identification, signal and image processing, and approximate polynomial computations.

Ivan Markovsky obtained M.S. degree in Control and Systems Engineering from the Technical University of Sofia in July 1998 and Ph.D. degree in Electrical Engineering from the Katholieke Universiteit Leuven in February 2005. From January 2007 to September 2012 he was a lecturer at the School of Electronics and Computer Science of the University of Southampton, U.K. Since October 2012 he has been with the Department ELEC of the Vrije Universiteit Brussel. His current research interests are structured low-rank approximation, system identification, and data-driven control.

Frank Gauterin received a Diploma degree in Physics (M.Sc.) in 1998 from the Westfälische Wilhelms-Universtät in Münster, Germany, and the Dr. rer. nat. (Ph.D) degree in 1994 from the University of Oldenburg, Germany. During 1993–2000 he worked as acoustics engineer at Continental AG, Hannover, Germany, and during 2000–2006 he worked as director of noise vibration harshness (NVH) engineering at Continental AG with locations in Germany, USA, and Malaysia. Since 2006 he has been a full professor in Vehicle Science at the Karlsruhe Institute of Technology (KIT), head of the Institute of Vehicle System Technology, and scientific spokesman of KIT-Center Mobility Systems. Dr. Gauterin is member of the Scientific Advisory Boards of e-mobil BW (Stuttgart), Automotive Engineering Network (Karlsruhe), Commercial Vehicle Cluster (Kaiserslautern), and the Automobiltechnische Zeitschrift (ATZ). He is member of the Scientific Society of Vehicle and Motor Technology WKM (Berlin), Heinrich Hertz Society (Karlsruhe), Association of German Engineers VDI (Düsseldorf), and the German Society of Acoustics DEGA (Berlin). He has served on numerous technical program committees for ATZ and VDI conferences. His research interests are vehicle NVH, tire acoustics, driver assistance systems, vehicle dynamics, as well as vehicle modeling and identification methods.