

Bias and covariance of the least squares estimate in a structured errors-in-variables problem

Gustavo Quintana Carapia^{a,*}, Ivan Markovsky^a, Rik Pintelon^a, Péter Zoltán Csürcsia^{b,c}, Dieter Verbeke^b

^a*Vrije Universiteit Brussel, Department of Fundamental Electricity and Instrumentation, 1050, Brussels, Belgium*

^b*Vrije Universiteit Brussel, Department of Engineering Technology, 1050, Brussels, Belgium*

^c*Siemens Industry Software NV, RTD Test Division, Interleuvenlaan 68, 3001, Leuven, Belgium*

Abstract

A structured errors-in-variables (EIV) problem arising in metrology is studied. The observations of a sensor response are subject to perturbation. The input estimation from the transient response leads to a structured EIV problem. Total least squares (TLS) is a typical estimation method to solve EIV problems. The TLS estimator of an EIV problem is consistent, and can be computed efficiently when the perturbations have zero mean, and are independently and identically distributed (i.i.d). If the perturbation is additionally Gaussian, the TLS solution coincides with maximum-likelihood (ML). However, the computational complexity of structured TLS and total ML prevents their real-time implementation. The least-squares (LS) estimator offers a suboptimal but simple recursive solution to structured EIV problems with correlation, but the statistical properties of the LS estimator are unknown. To know the LS estimate uncertainty in EIV problems, either structured or not, to provide confidence bounds for the estimation uncertainty, and to find the difference from the optimal solutions, the bias and variance of the LS estimates should be quantified. Expressions to predict the bias and variance of LS estimators applied to unstructured and structured EIV problems are derived. The predicted bias and variance quantify the statistical properties of the LS estimate and give an approximation of the uncertainty and the mean squared error for comparison to the Cramér-Rao lower bound of the structured EIV problem.

Keywords: structured errors-in-variables problems, least-squares estimation, Cramér-Rao lower bound, statistical analysis, uncertainty assessment, Monte Carlo simulation

1. Introduction

Errors-in-variables (EIV) are linear estimation problems in which the regression matrix and the regressor are perturbed Van Huffel and Vandewalle (1991), Markovsky and Van Huffel (2007). In structured EIV problems, the regression matrix has a given structure that depends on the problem formulation. Hankel,
5 Toeplitz, or other application-specific matrices appear in problems of metrology Markovsky (2015), system identification Söderström (2007), image restoration Feiz and Rezghi (2017), nuclear magnetic resonance

*Corresponding author

Email address: gquintan@vub.be (Gustavo Quintana Carapia)

spectroscopy Cai et al. (2016), direction-of-arrival estimation Pan et al. (2018), and time-of-arrival estimation Jia et al. (2018).

In metrology, the direct estimation of the input from the sensor transient response is formulated as a structured EIV problem. The only observed signal is the sensor output. To estimate a step input, the regression matrix, and the regressor are built from the step response observations Markovsky (2015), and the structure in the regression matrix is block-Hankel. This data-driven estimation methodology reduces the estimation time of the classical two-stage approach where a sensor model is first identified, and later the input is estimated using the sensor model Azam et al. (2015); Niedźwiecki et al. (2016).

Total least-squares (TLS) is the typical estimator for unstructured EIV problems and is consistent when the perturbations have zero mean, and the covariance is a given positive definite matrix. When the perturbations are i.i.d. normally distributed, the solution of the TLS is equivalent to that of the maximum likelihood estimator (ML) Markovsky and Van Huffel (2007). For structured EIV problems, the TLS estimator does not give general results since each specific structure requires a particular treatment Van Huffel et al. (2007), and the ML estimator leads to non-convex optimization problems where finding the global optimum is not guaranteed Rhode et al. (2014). Moreover, the computational complexity of TLS and ML inhibits real-time implementation to solve structured EIV problems.

The least-squares (LS) estimator is a suboptimal but simple solution to structured EIV problems that admits a recursive form for easy real-time implementations. Some of the reported works that propose LS estimators for structured EIV problems include the design of a fast algorithm for matrices with small displacement rank Mastronardi and O’Leary (2007), the study of the estimator consistency Palanthandalam-Madapusi et al. (2010), the determination of the bias, and the mean squared error of the parameter estimates in the identification of AR models Kiviet and Phillips (2012) Kiviet and Phillips (2014), and a discussion of the causes of bias and inconsistency in homogeneous estimators Yeredor and De Moor (2004).

In measurement applications, it is highly relevant to assess the uncertainty of the input estimate. The uncertainty of the reported LS estimators for structured EIV problems has not been addressed, and then, remains unknown. The estimator uncertainty is expressed using the estimation bias and covariance Pintelon and Schoukens (2012). To know the LS estimator uncertainty, we quantified the bias and the covariance of the LS solution of EIV problems in the unstructured and structured cases. This extends the perturbation analysis of the LS estimator of unstructured and uncorrelated problems that was investigated in Stewart (1990) and in Vaccaro (1994). This paper presents a study of the statistics of the LS estimate of unstructured and structured EIV problems. We provide a discussion of the unstructured case as a reference, to get an insight of the impact that the structure of the regression matrix and the correlation between the regression matrix and the regressor perturbations have on the uncertainty of the LS estimate. The structured case is motivated by the metrology estimation problem Markovsky (2015), where the EIV problem has a Hankel structure, and the perturbations are correlated. The study of the metrology input estimate illustrates a

methodology to conduct statistical analysis for any structured EIV problem. The mathematical expectation of the second-order Taylor series expansion of the LS estimate boils down to expressions that quantify the first and second-order moments of the LS estimate. Via Monte Carlo simulations, we validated the accuracy of the bias, and the covariance approximations. The derived approximations predict the LS estimate bias, and the covariance, for given sample size and perturbation level. The predicted variance gives the uncertainty of the LS estimate. We observed that, for the step input estimation problem, the mean squared error of the LS estimate is near to the minimum variance limit given by the Cramér-Rao lower bound of the structured EIV problem. By following this methodology, the bias and variance of the solutions of EIV problems with other structures is determined, and therefore, the uncertainty of the estimate.

The organization of the paper is as follows. Section 2 presents a brief description of the structured EIV problem that illustrates the methodology. The problem is the estimation of the step input unknown level by processing of the sensor transient response. Section 3 describes the statistical analysis of the LS solution for unstructured and structured EIV problems. Section 4 discusses the results obtained from Monte Carlo simulations that validate the bias and variance expressions. Section 5 presents a summary of the findings of this work, and concludes the paper.

2. Step input estimation method

The step input estimation method estimates the unknown value $u \in \mathbb{R}$ of the input $\mathbf{u} = u\mathbf{s}$, where \mathbf{s} is the unit step function ($s(t) = 0$ if $t < 0$ and 1 elsewhere), applied to a bounded-input bounded-output stable linear time-invariant system of order n and given dc-gain $g \in \mathbb{R}$. The method processes the sequence of step response observations $(\tilde{y}(0), \dots, \tilde{y}(T))^\top$, where $\tilde{y}(t) \in \mathbb{R}$ for $t = 0, \dots, T$, where T is the number of samples, and

$$\tilde{y}(t) = y(t) + \epsilon(t). \quad (1)$$

The exact system response \mathbf{y} is affected by additive i.i.d. normally distributed perturbation $\boldsymbol{\epsilon}$ with zero mean and given variance σ_ϵ^2 . The observed response of the system is a step-invariant discretization of the continuous-time response.

The step input level estimation is formulated as the minimization problem

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \left\| \tilde{\mathbf{y}} - \tilde{\mathbf{K}}\mathbf{x} \right\|_2^2, \quad (2)$$

where $\tilde{\mathbf{y}} = (\tilde{y}(n+1), \dots, \tilde{y}(T))^\top$, the first element of the to-be-estimated vector $\mathbf{x} = (u, \boldsymbol{\ell}^\top)^\top$ is the step input level, the vector $\boldsymbol{\ell}$ is linked to the system initial conditions, and the matrix

$$\tilde{\mathbf{K}} = \begin{bmatrix} g & \Delta\tilde{y}(1) & \Delta\tilde{y}(2) & \cdots & \Delta\tilde{y}(n) \\ g & \Delta\tilde{y}(2) & \Delta\tilde{y}(3) & \cdots & \Delta\tilde{y}(n+1) \\ \vdots & \vdots & \vdots & & \vdots \\ g & \Delta\tilde{y}(T-n) & \Delta\tilde{y}(T-n+1) & \cdots & \Delta\tilde{y}(T-1) \end{bmatrix}, \quad (3)$$

is a Hankel matrix of $(T - n) -$ block rows, constructed from consecutive differences $\Delta\tilde{y}(t) = \tilde{y}(t) - \tilde{y}(t - 1)$ of the observed transient response, augmented in the left side with a $(T - n)$ -vector of elements equal to the known dc-gain g , (see Markovsky (2015)). The perturbations ϵ enter in matrix $\tilde{\mathbf{K}}$ and we can express

$$\tilde{\mathbf{K}} = \mathbf{K} + \mathbf{E}, \quad (4)$$

where \mathbf{K} is exact data information and \mathbf{E} is the additive perturbation noise given as

$$\mathbf{E} = \begin{bmatrix} 0 & \Delta\epsilon(1) & \Delta\epsilon(2) & \cdots & \Delta\epsilon(n) \\ 0 & \Delta\epsilon(2) & \Delta\epsilon(3) & \cdots & \Delta\epsilon(n+1) \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & \Delta\epsilon(T-n) & \Delta\epsilon(T-n+1) & \cdots & \Delta\epsilon(T-1) \end{bmatrix}. \quad (5)$$

The underlying system of equations $\tilde{\mathbf{y}} = \tilde{\mathbf{K}}\mathbf{x}$ in the minimization problem (2) is an errors-in-variables (EIV) problem with Hankel structure. For metrology applications, the least-squares (LS) approximate solution of this system of equations offers a simple alternative, in its recursive form, to implement the estimation method in real-time. The LS solution is examined even when it may have some bias because the perturbation errors in $\tilde{\mathbf{K}}$ are correlated to the perturbations in $\tilde{\mathbf{y}}$. The classical LS results for the bias and the covariance cannot be invoked because LS assumes that the additive perturbation only affects the regressor, and that there is no correlation between the regressor and the regression matrix. The recursive LS method allows for a real-time implementation of the step input estimation method. Details of the step input estimation method are described in Markovsky (2015).

3. Statistical analysis

For an overdetermined system of linear equations, the least-squares (LS) solution is given by

$$\hat{\mathbf{x}} = \tilde{\mathbf{K}}^\dagger \tilde{\mathbf{y}} = (\tilde{\mathbf{K}}^\top \tilde{\mathbf{K}})^{-1} \tilde{\mathbf{K}}^\top \tilde{\mathbf{y}}, \quad (6)$$

where $\tilde{\mathbf{K}}^\dagger$ is the pseudo-inverse of $\tilde{\mathbf{K}}$. The objective of the statistical analysis is to obtain the bias and the covariance of the solution $\hat{\mathbf{x}}$. The bias and the covariance are computed using the mathematical expectation operator $\mathbb{E}\{\cdot\}$. First we substitute (1) and (4) in (6) obtaining

$$\hat{\mathbf{x}} = \left((\mathbf{K} + \mathbf{E})^\top (\mathbf{K} + \mathbf{E}) \right)^{-1} (\mathbf{K} + \mathbf{E})^\top (\mathbf{y} + \epsilon),$$

which is equivalent to

$$\hat{\mathbf{x}} = (\mathbf{I} + \mathbf{M})^{-1} \mathbf{Q}^{-1} (\mathbf{K} + \mathbf{E})^\top (\mathbf{y} + \epsilon), \quad (7)$$

where

$$\mathbf{Q} = \mathbf{K}^\top \mathbf{K}, \quad \text{and} \quad \mathbf{M} = \mathbf{Q}^{-1} (\mathbf{K}^\top \mathbf{E} + \mathbf{E}^\top \mathbf{K} + \mathbf{E}^\top \mathbf{E}). \quad (8)$$

Applying a second-order Taylor expansion of the inverse matrix

$$(\mathbf{I} + \mathbf{M})^{-1} \approx \mathbf{I} - \mathbf{M} + \mathbf{M}^2, \quad (9)$$

that is valid when the SNR is sufficiently high, and therefore \mathbf{E} and \mathbf{M} are small, satisfying the constraint on the spectral radius $\|\mathbf{M}\| < 1$. The neglected term in the Taylor series expansion is of the order $O(\|\mathbf{M}\|^3)$.

We can express the estimate as

$$\hat{\mathbf{x}} \approx (\mathbf{I} - \mathbf{M} + \mathbf{M}^2) \mathbf{Q}^{-1}(\mathbf{K} + \mathbf{E})^\top (\mathbf{y} + \boldsymbol{\epsilon}). \quad (10)$$

Now that the perturbation variables $\boldsymbol{\epsilon}$ and \mathbf{E} are no longer inside an inverse matrix, we can compute the mathematical expectation of expressions derived from the Taylor series approximation (10) of $\hat{\mathbf{x}}$. The bias of the estimate $\hat{\mathbf{x}}$ is obtained from

$$\mathbf{b}(\hat{\mathbf{x}}) = \boldsymbol{\mu}(\hat{\mathbf{x}}) - \mathbf{x}, \quad (11)$$

where $\boldsymbol{\mu}(\hat{\mathbf{x}}) = \mathbb{E}\{\hat{\mathbf{x}}\}$, and $\mathbf{x} = \mathbf{K}^\dagger \mathbf{y} = \mathbf{Q}^{-1} \mathbf{K}^\top \mathbf{y}$ is the true value. The covariance of the estimate $\hat{\mathbf{x}}$ is obtained from

$$\mathbf{C}(\hat{\mathbf{x}}) = \mathbb{E}\left\{(\hat{\mathbf{x}} - \boldsymbol{\mu})(\hat{\mathbf{x}} - \boldsymbol{\mu})^\top\right\}. \quad (12)$$

The terms derived from (10) that do not contribute to the bias and to the covariance are filtered out by the mathematical expectation operator considering the following general rules that are valid regardless of the existence of structure in the regression problem:

- the expected values $\mathbb{E}\{\mathbf{E}\} = \mathbf{0}$, and $\mathbb{E}\{\boldsymbol{\epsilon}\} = \mathbf{0}$, since \mathbf{E} and $\boldsymbol{\epsilon}$ are zero-mean random variables, and
- the expected values of odd order moments, such as $\mathbb{E}\{\mathbf{E}^\top \mathbf{E} \mathbf{E}^\top\}$, are zero.

75

Moreover, the second-order approximation disregards moments of order four and higher.

After removing the terms with negligible expected value, we have expressions that are approximations of the LS estimation bias and covariance. These expressions are different depending on the type of EIV problem considered. Subsections 3.1 and 3.2 describe the resulting expressions for the unstructured and structured EIV problems, respectively. The perturbations in the considered unstructured EIV problem are independent. Comparing the expressions that result from the statistical analysis, we get an insight of what is the impact that the structure and the correlation have on the LS solution of the structured EIV problem.

80

3.1. Bias and covariance of the LS estimator for an unstructured EIV problem with uncorrelated noise

First, the statistics of the least squares estimator (6) of an unstructured EIV problem is discussed, provided that the perturbations of the regression matrix and the regressor are i.i.d. normally distributed with zero mean, and variances $\sigma_{\mathbf{E}}^2$ and $\sigma_{\boldsymbol{\epsilon}}^2$, respectively. Therefore, the terms in the Taylor series expansion (10) that contain products of \mathbf{E} and $\boldsymbol{\epsilon}$ have zero expected value. After removing the terms without contribution

to the bias, and to the covariance, with the mathematical expectation operator, the analytic approximation of the bias (11) is given by

$$\mathbf{b}_p(\hat{\mathbf{x}}) \approx \sigma_{\mathbf{E}}^2 (2 + 2n - T) \mathbf{Q}^{-1} \mathbf{x}, \quad (13)$$

and the covariance (12) is approximated by

$$\mathbf{C}_p(\hat{\mathbf{x}}) \approx \sigma_{\epsilon}^2 \mathbf{Q}^{-1} + \sigma_{\mathbf{E}}^2 \text{trace}(\mathbf{x}\mathbf{x}^\top) \mathbf{Q}^{-1} - \sigma_{\mathbf{E}}^4 (2 + 2n - T)^2 \mathbf{Q}^{-1} \mathbf{x}\mathbf{x}^\top \mathbf{Q}^{-1}, \quad (14)$$

where the subscript p stands for prediction of the bias and covariance. The derivation of equations (13) and (14) is described in Appendix 1. We use the results described in Vaccaro (1994) §3 and Stewart (1990) §2.1 for the expected values of products of unstructured matrices with perturbations. Equations (13) and (14) depend on the unobservable true values \mathbf{x} , \mathbf{K} , and on the variance of the perturbations. The observed variables are $\tilde{\mathbf{y}}$, $\tilde{\mathbf{K}}$, and from them we compute $\hat{\mathbf{x}}$. It is proposed to directly substitute the observed variables in the analytic expressions. The substitution gives an approximation of the estimate bias and covariance using the observed data. We have then

$$\tilde{\mathbf{b}}_p(\hat{\mathbf{x}}) \approx \sigma_{\mathbf{E}}^2 (2 + 2n - T) \tilde{\mathbf{Q}}^{-1} \hat{\mathbf{x}}, \quad (15)$$

$$\tilde{\mathbf{C}}_p(\hat{\mathbf{x}}) \approx \sigma_{\epsilon}^2 \tilde{\mathbf{Q}}^{-1} + \sigma_{\mathbf{E}}^2 \text{trace}(\hat{\mathbf{x}}\hat{\mathbf{x}}^\top) \tilde{\mathbf{Q}}^{-1} - \sigma_{\mathbf{E}}^4 (2 + 2n - T)^2 \tilde{\mathbf{Q}}^{-1} \hat{\mathbf{x}}\hat{\mathbf{x}}^\top \tilde{\mathbf{Q}}^{-1}. \quad (16)$$

In order to have a prediction of the estimate bias and covariance, the variances $\sigma_{\mathbf{E}}^2$ and σ_{ϵ}^2 and the observed variables $\tilde{\mathbf{y}}$, $\tilde{\mathbf{K}}$, and $\hat{\mathbf{x}}$ are needed.

3.2. Bias and covariance of the LS estimator for a structured EIV problem with noise correlation

This subsection describes the statistics of a structured EIV problem with correlation between the perturbations of the regression matrix and the regressor. The structured EIV problem is given by the step input estimation method (2). The correlation is a consequence of the construction of the block-Hankel matrix in the regression matrix $\tilde{\mathbf{K}}$ with the first difference of the elements in the regressor $\tilde{\mathbf{y}}$.

The mathematical expectation operator is applied to the Taylor series expansion of the LS estimate (10). After removing the terms with negligible expected value, and considering the structure of matrix \mathbf{K} , the estimation bias (11) is approximated by

$$\mathbf{b}_p(\hat{\mathbf{x}}) \approx \mathbf{Q}^{-1} \left((\mathbf{K}^\top \mathbf{B}_1 - \mathbf{B}_2) x - (\mathbf{K}^\top \mathbf{B}_3 - \mathbf{B}_4) \right), \quad (17)$$

whereas, the estimation covariance (12) is approximated by

$$\mathbf{C}_p(\hat{\mathbf{x}}) \approx \mathbf{K}^\dagger \left(\sigma_{\epsilon}^2 \mathbf{I}_{T-n} + \mathbf{C}_1 - \mathbf{C}_2 - \mathbf{C}_2^\top \right) \mathbf{K}^{\dagger\top} - \mathbf{b}_p(\hat{\mathbf{x}}) \mathbf{b}_p^\top(\hat{\mathbf{x}}), \quad (18)$$

where $\mathbf{B}_1 = \mathbb{E}\{\mathbf{E}\mathbf{K}^\dagger\mathbf{E}\}$, $\mathbf{B}_2 = \mathbb{E}\{\mathbf{E}^\top\mathbf{P}_\perp\mathbf{E}\}$, $\mathbf{B}_3 = \mathbb{E}\{\mathbf{E}\mathbf{K}^\dagger\epsilon\}$, $\mathbf{B}_4 = \mathbb{E}\{\mathbf{E}^\top\mathbf{P}_\perp\epsilon\}$, $\mathbf{C}_1 = \mathbb{E}\{\mathbf{E}\mathbf{x}\mathbf{x}^\top\mathbf{E}^\top\}$, $\mathbf{C}_2 = \mathbb{E}\{\mathbf{E}\mathbf{x}\epsilon^\top\}$, $\mathbf{P} = \mathbf{K}\mathbf{K}^\dagger$, and $\mathbf{P}_\perp = \mathbf{I} - \mathbf{P}$. The derivation of equations (17) and (18) is described in Appendix 1. The expected values \mathbf{B}_1 , \mathbf{B}_2 , \mathbf{B}_3 , \mathbf{B}_4 , \mathbf{C}_1 and \mathbf{C}_2 are obtained using the results of Lemma 1.

Lemma 1. Let $\mathbf{E} \in \mathbb{R}^{(T-n) \times (n+1)}$ be the partitioned matrix

$$\mathbf{E} = \begin{bmatrix} \mathbf{0}_{T-n \times 1} & \mathcal{H}(\boldsymbol{\epsilon}) \mathbf{D}_{n+1 \times n}^{1,0} \end{bmatrix},$$

where $\mathcal{H}(\boldsymbol{\epsilon}) \in \mathbb{R}^{(T-n) \times (n+1)}$ is the block-Hankel matrix of $T - n$ rows and n columns

$$\mathcal{H}(\boldsymbol{\epsilon}) = \begin{bmatrix} \epsilon(0) & \epsilon(1) & \cdots & \epsilon(n) \\ \epsilon(1) & \epsilon(2) & \cdots & \epsilon(n+1) \\ \vdots & \vdots & & \vdots \\ \epsilon(T-n-1) & \epsilon(T-n) & \cdots & \epsilon(T-1) \end{bmatrix},$$

constructed from samples of the i.i.d. normally distributed random variable $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$. Let $\mathbf{D}_{r \times c}^{1,k}$ and $\mathbf{D}_{r \times c}^{2,k}$ be the first and second-order finite differences matricial operators of dimensions $r \times c$ starting from the subdiagonal k , for example,

$$\mathbf{D}_{4 \times 3}^{1,-1} = \begin{bmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}, \quad \mathbf{D}_{4 \times 3}^{2,0} = \begin{bmatrix} -1 & 0 & 0 \\ 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}.$$

For a compatible deterministic matrix \mathbf{A} , or vector \mathbf{a} , the following expected values hold.

$$\mathbb{E} \{ \mathbf{E} \mathbf{A} \mathbf{E} \} = \mathbf{Z}, \quad \text{where } \mathbf{z}_{i1} = \mathbf{0}, \text{ and } z_{ij} = \sigma_\epsilon^2 \text{Tr} \left(\mathbf{A} \begin{bmatrix} \mathbf{0}_{T-n} & \mathbf{D}_{T-n \times n}^{2,j-i} \end{bmatrix} \right), \text{ for } i = 1, \dots, T-n, \text{ and } j = 2, \dots, n+1.$$

$$\mathbb{E} \{ \mathbf{E}^\top \mathbf{A} \mathbf{E} \} = \mathbf{Z}, \quad \text{where } \mathbf{z}_{1j} = \mathbf{0}, \mathbf{z}_{i1} = \mathbf{0}, \text{ and } z_{ij} = \sigma_\epsilon^2 \text{Tr} \left(\mathbf{A} \mathbf{D}_{T-n \times T-n}^{2,j-i+1} \right), \text{ for } i = 2, \dots, n+1, \text{ and } j = 2, \dots, n+1.$$

$$\mathbb{E} \{ \mathbf{E} \mathbf{A} \mathbf{E}^\top \} = \mathbf{Z}, \quad \text{where } z_{ij} = \sigma_\epsilon^2 \text{Tr} \left(\mathbf{A} \begin{bmatrix} 0 & \mathbf{0}_n^\top \\ \mathbf{0}_n & \mathbf{D}_{n \times n}^{2,j-i+1} \end{bmatrix} \right), \text{ for } i = 1, \dots, T-n, \text{ and } j = 1, \dots, T-n.$$

$$\mathbb{E} \{ \mathbf{E} \mathbf{A} \boldsymbol{\epsilon} \} = \mathbf{z}, \quad \text{where } z_i = \sigma_\epsilon^2 \text{Tr} \left(\mathbf{A} \begin{bmatrix} \mathbf{0}_{T-n} & \mathbf{D}_{T-n \times n}^{1,n+1-i} \end{bmatrix} \right), \text{ for } i = 1, \dots, T-n.$$

$$\mathbb{E} \{ \mathbf{E}^\top \mathbf{A} \boldsymbol{\epsilon} \} = \mathbf{z}, \quad \text{where } z_1 = 0, \text{ and } z_i = \sigma_\epsilon^2 \text{Tr} \left(\mathbf{A} \mathbf{D}_{T-n \times T-n}^{1,n+2-i} \right), \text{ for } i = 2, \dots, n+1.$$

$$\mathbb{E} \{ \mathbf{E} \mathbf{a} \boldsymbol{\epsilon}^\top \} = \mathbf{Z}, \quad \text{where each column } \mathbf{Z}_j = -\sigma_\epsilon^2 \mathbf{D}_{T-n \times n+1}^{1,-j} \mathbf{R}_{n+1} \mathbf{a}, \text{ for } j = 1, \dots, T-n, \text{ with } \mathbf{R}_{n+1} = \begin{bmatrix} \mathbf{R}_n \\ 0 \end{bmatrix},$$

where \mathbf{R}_n is a reversal matrix.

The proof of the lemma is given in Appendix 2.

95 The matrices $\mathbf{B}_1, \mathbf{B}_2, \mathbf{B}_3, \mathbf{B}_4, \mathbf{C}_1$ and \mathbf{C}_2 are considered in the different cases of Lemma 1. Each expected value in the lemma is a matrix, or a vector, whose elements are found following the indicated operations. These operations mainly compute the trace of a product of the corresponding deterministic matrix \mathbf{A} , and a matrix constructed from the finite differences matricial operator, which can be of first order \mathbf{D}^1 or of second order \mathbf{D}^2 . The total number of operations in the computation of the bias and the covariance is $O(T^3 + n^2)$,
100 but this order can be reduced since the \mathbf{D} matrices are sparse.

The formulas for the bias and covariance (17) and (18) depend on the perturbation variance and on the unobservable variables \mathbf{x} and \mathbf{K} . The substitution of the observed variables in the expression gives an approximation of the estimate bias and covariance based on the observed system response. We have then

$$\mathbf{b}_{\hat{\mathbf{x}}} \approx \tilde{\mathbf{Q}}^{-1} \left(\left(\tilde{\mathbf{K}}^\top \tilde{\mathbf{B}}_1 - \tilde{\mathbf{B}}_2 \right) \hat{\mathbf{x}} - \left(\tilde{\mathbf{K}}^\top \tilde{\mathbf{B}}_3 - \tilde{\mathbf{B}}_4 \right) \right), \quad (19)$$

and

$$\mathbf{C}_{\tilde{\mathbf{p}}}(\hat{\mathbf{x}}) \approx \tilde{\mathbf{K}}^\dagger \left(\sigma_\epsilon^2 \mathbf{I}_{T-n} + \tilde{\mathbf{C}}_1 - \tilde{\mathbf{C}}_2 - \tilde{\mathbf{C}}_2^\top \right) \tilde{\mathbf{K}}^{\dagger\top} - \mathbf{b}_{\tilde{\mathbf{p}}}(\hat{\mathbf{x}}) \mathbf{b}_{\tilde{\mathbf{p}}}^\top(\hat{\mathbf{x}}), \quad (20)$$

where $\tilde{\mathbf{B}}_1 = \mathbb{E}\{\mathbf{E}\tilde{\mathbf{K}}^\dagger\mathbf{E}\}$, $\tilde{\mathbf{B}}_2 = \mathbb{E}\{\mathbf{E}^\top\tilde{\mathbf{P}}_\perp\mathbf{E}\}$, $\tilde{\mathbf{B}}_3 = \mathbb{E}\{\mathbf{E}\tilde{\mathbf{K}}^\dagger\epsilon\}$, $\tilde{\mathbf{B}}_4 = \mathbb{E}\{\mathbf{E}^\top\tilde{\mathbf{P}}_\perp\epsilon\}$, $\tilde{\mathbf{C}}_1 = \mathbb{E}\{\mathbf{E}\hat{\mathbf{x}}\hat{\mathbf{x}}^\top\mathbf{E}^\top\}$, $\tilde{\mathbf{C}}_2 = \mathbb{E}\{\mathbf{E}\hat{\mathbf{x}}\epsilon^\top\}$, $\tilde{\mathbf{C}} = \tilde{\mathbf{K}}^\top\tilde{\mathbf{K}}$, $\tilde{\mathbf{P}} = \tilde{\mathbf{K}}\tilde{\mathbf{K}}^\dagger$, and $\tilde{\mathbf{P}}_\perp = \mathbf{I} - \tilde{\mathbf{P}}$.

3.3. Cramér-Rao lower bound of the structured errors-in-variables problem

The Cramér-Rao lower bound (CRLB) provides the lower limit on the variance of the estimate

$$\text{CRLB}(\mathbf{x}) = \left(\mathbf{I} + \frac{\partial \mathbf{b}(\hat{\mathbf{x}})}{\partial \hat{\mathbf{x}}} \right)^\top \mathbf{F}\mathbf{i}^{-1}(\mathbf{x}) \left(\mathbf{I} + \frac{\partial \mathbf{b}(\hat{\mathbf{x}})}{\partial \hat{\mathbf{x}}} \right), \quad (21)$$

where $\mathbf{b}(\hat{\mathbf{x}})$ is the bias of the estimate and $\mathbf{F}\mathbf{i}(\hat{\mathbf{x}})$ is the Fisher information matrix Pintelon and Schoukens (2012). The Fisher information matrix is defined as the expected value of the Hessian of the negative likelihood function

$$\mathbf{F}\mathbf{i}(x) = -\mathbb{E} \left\{ \frac{\partial^2 l(\hat{\mathbf{x}})}{\partial \hat{\mathbf{x}}^2} \right\}, \quad (22)$$

where the partial derivatives are evaluated in $\hat{\mathbf{x}} = \mathbf{x}$.

The minimization problem (2) is equivalent to a structured EIV problem that can be expressed as a linear in the measurements problem Pintelon and Schoukens (2012)

$$e(\hat{\mathbf{x}}, \tilde{\mathbf{z}}) = \mathbf{M}_1(\hat{\mathbf{x}}) \tilde{\mathbf{z}} = \begin{bmatrix} \mathbf{I}_{T-n} & -\hat{\mathbf{x}}^T \otimes \mathbf{I}_{T-n} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{y}} \\ \text{vec}(\tilde{\mathbf{K}}) \end{bmatrix} = 0. \quad (23)$$

where $\tilde{\mathbf{z}} = \mathbf{z} + \epsilon_{\mathbf{z}}$. The CRB requires that the true model $\mathbf{M}_1(\mathbf{x}) \mathbf{z} = 0$ exists. Under the assumption of the measurement perturbation $\epsilon_{\mathbf{z}}$ being normally distributed with covariance matrix $\mathbf{C}_{\mathbf{z}}$, the loglikelihood function of the structured EIV problem is

$$\ln l(\tilde{\mathbf{z}}, \hat{\mathbf{z}}, \hat{\mathbf{x}}) = -\frac{1}{2} (\tilde{\mathbf{z}} - \hat{\mathbf{z}})^\top \mathbf{C}_{\mathbf{z}}^{-1} (\tilde{\mathbf{z}} - \hat{\mathbf{z}}) + \text{constant}, \quad (24)$$

where $\hat{\mathbf{z}}$ are parameters of the measurements $\tilde{\mathbf{z}}$ that have to be estimated and satisfy $\mathbf{M}_1(\hat{\mathbf{x}}) \hat{\mathbf{z}} = 0$. The size of the Fisher information matrix $\mathbf{F}\mathbf{i}(\mathbf{x}, \mathbf{z})$ depends on the number of unknowns in $\hat{\mathbf{z}}$ and grows with the sample size. Moreover, in Chapter 19 of Pintelon and Schoukens (2012) it is shown that the Fisher information matrix $\mathbf{F}\mathbf{i}(\mathbf{x})$ can be obtained from $\mathbf{F}\mathbf{i}(\mathbf{x}, \mathbf{z})$ after doing inversion by parts, giving

$$\mathbf{F}\mathbf{i}(\mathbf{x}) = \left(\frac{\partial e(\hat{\mathbf{x}}, \mathbf{z})}{\partial \mathbf{x}} \right)^\top \left(\mathbf{M}_1(\mathbf{x}) \mathbf{C}_{\mathbf{z}} \mathbf{M}_1^\top(\mathbf{x}) \right)^{-1} \left(\frac{\partial e(\hat{\mathbf{x}}, \mathbf{z})}{\partial \mathbf{x}} \right), \quad (25)$$

where the partial derivatives are evaluated at the true values \mathbf{x} , and the covariance matrix of the measure-

ments is

$$\mathbf{C}_{\mathbf{z}} = \sigma_{\epsilon}^2 \begin{bmatrix} \mathbf{I}_{T-n} & \mathbf{0}_{T-n} & \mathbf{D}_{T-n \times T-n}^{1,n} & \mathbf{D}_{T-n \times T-n}^{1,n-1} & \cdots & \mathbf{D}_{T-n \times T-n}^{1,1} \\ \mathbf{0}_{T-n} & \mathbf{0}_{T-n} & \mathbf{0}_{T-n} & \mathbf{0}_{T-n} & \cdots & \mathbf{0}_{T-n} \\ \left(\mathbf{D}_{T-n \times T-n}^{1,n}\right)^{\top} & \mathbf{0}_{T-n} & \mathbf{D}_{T-n \times T-n}^{2,1} & \mathbf{D}_{T-n \times T-n}^{2,0} & \cdots & \mathbf{D}_{T-n \times T-n}^{2,2-n} \\ \left(\mathbf{D}_{T-n \times T-n}^{1,n-1}\right)^{\top} & \mathbf{0}_{T-n} & \mathbf{D}_{T-n \times T-n}^{2,2} & \mathbf{D}_{T-n \times T-n}^{2,1} & \cdots & \mathbf{D}_{T-n \times T-n}^{2,3-n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \left(\mathbf{D}_{T-n \times T-n}^{1,1}\right)^{\top} & \mathbf{0}_{T-n} & \mathbf{D}_{T-n \times T-n}^{2,n} & \mathbf{D}_{T-n \times T-n}^{2,n-1} & \cdots & \mathbf{D}_{T-n \times T-n}^{2,1} \end{bmatrix}. \quad (26)$$

The Cramér-Rao lower bound for an biased estimator of the minimization problem (2) is given by

$$\text{CRB}_{\mathbf{b}}(\mathbf{x}) = \left(\mathbf{I}_{n+1} + \frac{\partial \mathbf{b}(\hat{\mathbf{x}})}{\partial \mathbf{x}} \right)^{\top} \mathbf{F}\mathbf{i}^{-1}(\mathbf{x}) \left(\mathbf{I}_{n+1} + \frac{\partial \mathbf{b}(\hat{\mathbf{x}})}{\partial \mathbf{x}} \right), \quad (27)$$

105 and for an unbiased estimator it is $\text{CRB}_{\text{ub}}(\mathbf{x}) = \mathbf{F}\mathbf{i}^{-1}(\mathbf{x})$.

4. Simulation results

Two Monte Carlo (MC) simulation studies were conducted to test the obtained bias and variance formulas. One MC simulation was devoted to the unstructured EIV problem and the other to the structured EIV problem that corresponds to the step input estimation method (2). The MC simulations performed $N_{MC} = 10^6$ runs of the LS estimator with different realizations of the perturbation $[\mathbf{E} \ \epsilon]$. In the structured EIV problem the perturbation \mathbf{E} is correlated with ϵ . The perturbations variance was selected to have a signal-to-noise ratio (SNR) in the interval [30 dB, 80 dB], according to

$$\text{SNR} = 20 \log_{10} \frac{\sqrt{\frac{1}{T} \sum_{t=1}^T y(t)^2}}{\sigma_{\epsilon}}. \quad (28)$$

For high enough SNR, the constraint $\|\mathbf{M}\| < 1$ is valid and the Taylor series expansion (9) holds. In this case, the derived expressions predict the LS estimate bias and variance. Thus, it is relevant to monitor the evolution of the largest \mathbf{M} eigenvalue to detect the lower limit of the SNR that allows the validity of the

110 predictions.

The MC simulations provide empirical values of the bias and variance, through the sample mean and sample variance of the LS estimate $\hat{\mathbf{x}}$. The results presented in this Section focus the interest in the first element of $\hat{\mathbf{x}}$, because in the structured EIV problem, the step input estimate is $\hat{u} = \hat{\mathbf{x}}_{[1]}$.

4.1. Monte Carlo simulation results for an unstructured EIV problem with uncorrelated perturbations

115 The MC simulation of the unstructured EIV problem solution was conducted using the following settings. The exact data elements in the matrix $\mathbf{K} \in \mathbb{R}^{(T-n) \times (n+1)}$ are normally distributed with zero mean and variance one, for $T = 200$ and $n = 2$. The to-be-estimated exact data is the vector $x = [1 \ 2 \ 3]^{\top}$. The

perturbations \mathbf{E} and $\boldsymbol{\epsilon}$ are normally distributed with zero mean and variances subject to $\sigma_{\mathbf{E}}^2 = 2\sigma_{\boldsymbol{\epsilon}}^2$. These settings are similar to those of the structured case, aiming to have comparable situations.

The difference between the sample mean of the estimate $\widehat{\mathbf{x}}_{[1]}$ and its true value $\mathbf{x}_{[1]}$ is the empirical bias b_e .

$$b_e = \frac{1}{N_{\text{MC}}} \sum_{i=1}^{N_{\text{MC}}} (\widehat{\mathbf{x}}_{[1]})_i - \mathbf{x}_{[1]} \approx \mu(\widehat{\mathbf{x}}_{[1]}) - \mathbf{x}_{[1]}. \quad (29)$$

The standard deviation of the estimate $\widehat{\mathbf{x}}_{[1]}$ is used to obtain the standard error σ_e of the MC simulation, which decreases with respect to the square root of the number of MC runs N_{MC} Hammersley and Handscomb (1975):

$$\sigma_e = \frac{\sigma(\widehat{\mathbf{x}}_{[1]})}{\sqrt{N_{\text{MC}}}}, \quad \text{where} \quad \sigma^2(\widehat{\mathbf{x}}_{[1]}) = \frac{1}{N_{\text{MC}} - 1} \sum_{i=1}^{N_{\text{MC}}} \left((\widehat{\mathbf{x}}_{[1]})_i - \mu(\widehat{\mathbf{x}}_{[1]}) \right)^2. \quad (30)$$

In each of the N_{MC} runs we also obtain the approximations of the estimation bias $\mathbf{b}_{\widehat{\mathbf{p}}[1]}$ and variance $\mathbf{C}_{\widehat{\mathbf{p}}[1,1]}$ from observed data, using Equations (15), and (16). Similarly as before, we get the sample mean of the approximations to have bias and variance predictions from observed data

$$b_{\widehat{\mathbf{p}}} = \frac{1}{N_{\text{MC}}} \sum_{i=1}^{N_{\text{MC}}} (\mathbf{b}_{\widehat{\mathbf{p}}[1]})_i, \quad \text{and} \quad v_{\widehat{\mathbf{p}}} = \frac{1}{N_{\text{MC}}} \sum_{i=1}^{N_{\text{MC}}} (\mathbf{C}_{\widehat{\mathbf{p}}[1,1]})_i. \quad (31)$$

The standard error of the bias prediction $b_{\widehat{\mathbf{p}}}$ is given by

$$\sigma_{\widehat{\mathbf{p}}} = \sqrt{\frac{1}{N_{\text{MC}}(N_{\text{MC}} - 1)} \sum_{i=1}^{N_{\text{MC}}} \left((\mathbf{b}_{\widehat{\mathbf{p}}[1]})_i - b_{\widehat{\mathbf{p}}}(\widehat{\mathbf{x}}_{[1]}) \right)^2}. \quad (32)$$

120 The predicted bias $b_{\mathbf{p}} = \mathbf{b}_{\mathbf{p}[1]}$ and variance $v_{\mathbf{p}} = \mathbf{C}_{\mathbf{p}[1,1]}$ from exact data are obtained with one evaluation of the expressions (13) and (14).

Figure 1 shows the empirical bias and the bias predictions, with their corresponding standard errors, for the unstructured (plots in the left) and structured (plots in the right) EIV problems. The simulation settings for the structured and correlated EIV problem are described in the following subsection. In the figure, it can be observed that the empirical bias b_e is proportional to the perturbation noise variance while the standard error σ_e is proportional to the perturbation noise standard deviation. The bias predictions $b_{\mathbf{p}}$ and $b_{\widehat{\mathbf{p}}}$ are accurate since they coincide with the empirical bias b_e in all the SNR interval for unstructured EIV problems, In both unstructured and structured EIV cases, the standard errors of the MC estimates σ_e and $\sigma_{\widehat{\mathbf{p}}}$ are smaller than the bias estimates b_e and $b_{\widehat{\mathbf{p}}}$. The bias estimates are spread near their sample means. The uncertainty is smaller than the bias, therefore, the MC simulation is meaningful.

The absolute and relative errors between the predicted and the empirical bias are shown in Figure 2. The absolute errors decrease with respect to the perturbation variance. The relative errors are lower than 5% for SNR between 30 dB and 70 dB. There is an increment in the relative errors for SNR above 55 dB. As the SNR increases, the empirical and the predicted bias decrease, as well as the bias error between them. In order to reveal the bias error, more Monte Carlo runs are needed to reduce the uncertainty of the Monte

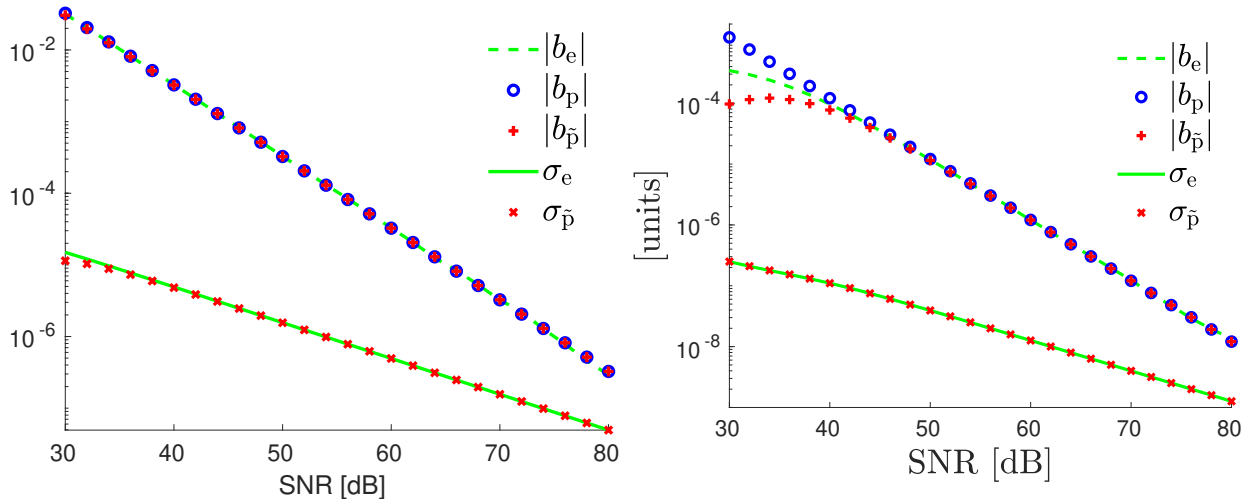


Figure 1: We observe the results of the LS solutions for the unstructured (left) and the structured (right) EIV problems. These results are the empirical bias b_e , the predicted bias from exact data b_p , the predicted bias from observed data $b_{\bar{p}}$, the empirical standard error σ_e , and the standard error from the estimations using observed data $\sigma_{\bar{p}}$. The estimation biases are proportional to the perturbation variance and the estimation standard errors are proportional to the perturbation standard deviation. Since the standard errors are smaller than the biases, the MC simulation is meaningful.

Carlo simulation that depends on the square root of N_{MC} , see Equation (30). If N_{MC} is insufficient, the uncertainty of the Monte Carlo simulation hides the bias error and we see this increasing effect of the relative errors.

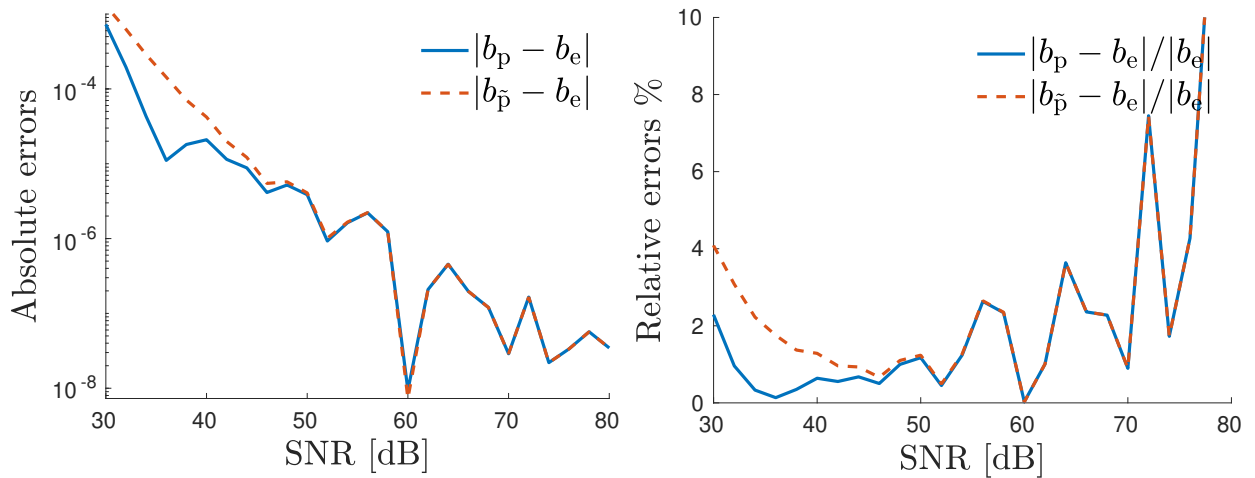


Figure 2: The MC simulation shows that when we solve an unstructured EIV problem with LS, the absolute errors (left) between the predicted bias and the empirical bias are proportional to the perturbation noise variance as it is expected, and the relative errors (right) are smaller than 5% for SNR below 70 dB. The bias prediction computed from exact data b_p is very similar to that computed using observed data $b_{\bar{p}}$.

The errors between the predicted and the empirical variance are shown in Figure 3. The absolute and relative errors decrease with respect to the perturbation variance. The relative errors are lower than 5% for SNR above 40 dB.

Figure 4 shows the absolute and the relative errors between the predictions computed from observed

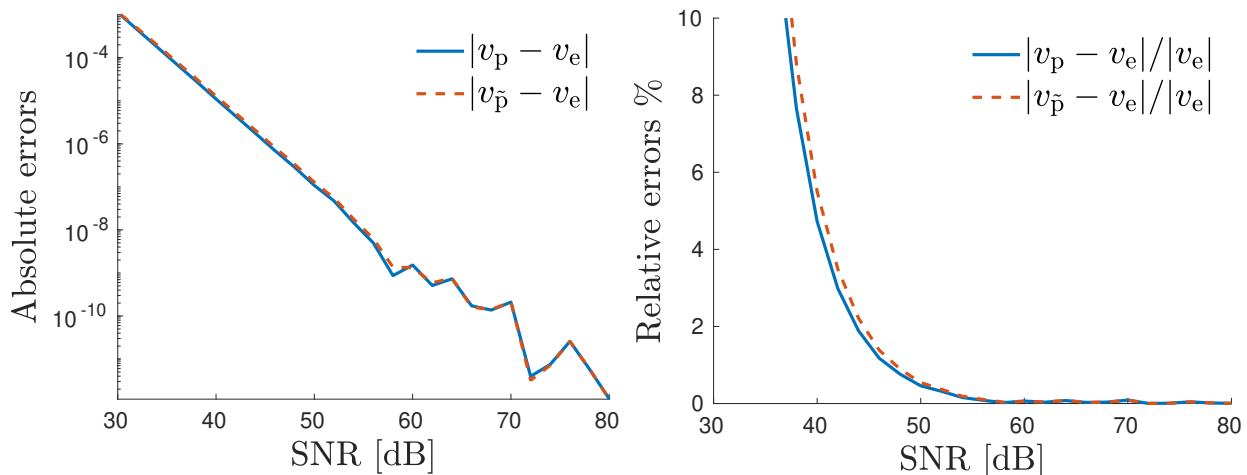


Figure 3: The MC simulation shows that when we solve an unstructured EIV problem with LS, the absolute errors (left) between the predicted variances and the empirical variance are proportional to the perturbation noise variance, and the relative errors (right) are smaller than 5% for SNR higher than 40 dB. The variance prediction computed using exact data v_p is very similar to that computed from observed data $v_{\bar{p}}$.

data and those computed from exact data. The absolute errors between both predictions are proportional to the perturbation noise variance. The bias and variance predictions, from either exact data or observed data, are equivalent for SNR above 35 dB since the relative errors are lower than 5%. The substitution of observed data on the prediction formulas is a valid procedure that allows the prediction of the estimate statistics.

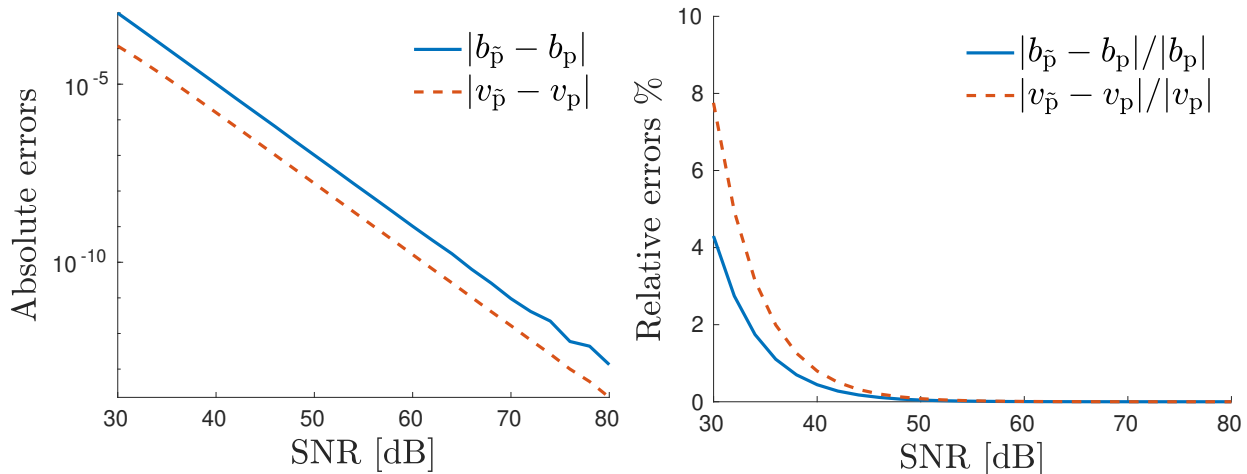


Figure 4: The MC simulation shows that when we solve an unstructured EIV problem with LS, the absolute errors (left) between the prediction with observed data and the prediction with exact data are proportional to the perturbation noise variance. The use of observed data instead of exact data in the prediction formulas is valid when the SNR is above 35 dB since the relative errors (right) are smaller than 5%.

4.2. Monte Carlo simulation results for a structured EIV problem with correlated perturbations

The MC simulation of the structured EIV problem (2) solution was conducted processing $T = 200$ samples of the transient response $\tilde{\mathbf{y}}$ generated by a linear time invariant system of order $n = 2$, after a

step input excitation with $u = 1$ units, where the units represent any physical quantity. The processed step response is shown in Figure 5. The steady state response of the system is reached after 400 samples because from there on the relative error between the instantaneous values of the transient response and the steady-state response value is smaller than 2%. Processing 200 samples ensures that the step input estimation is computed from transient data only.

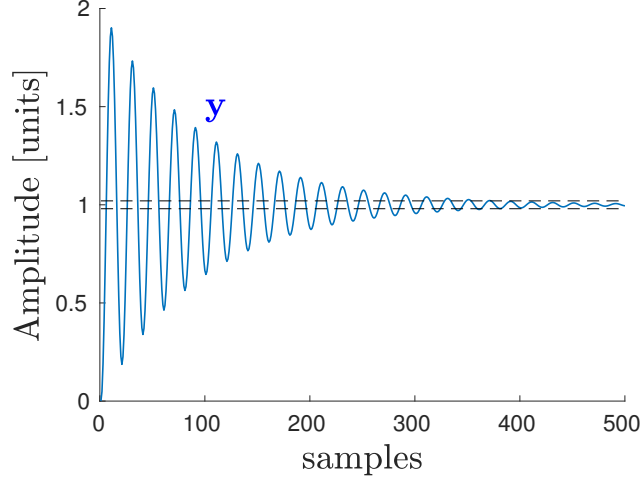


Figure 5: The structured EIV problem is constructed with 200 samples of the step response to ensure that only transient data is used. The relative errors between the values of \mathbf{y} and the steady state value are smaller than 2% after 400 samples, as it is indicated with dashed lines.

The empirical bias is the sample mean of the N_{MC} estimates minus the true value

$$b_e = \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} \hat{u}_i - u \approx \mu(\hat{u}) - u. \quad (33)$$

The standard error associated to this empirical bias estimation Hammersley and Handscomb (1975) is defined as

$$\sigma_e = \frac{\sigma(\hat{u})}{\sqrt{N_{MC}}}, \quad \text{where} \quad \sigma^2(\hat{u}) = \frac{1}{N_{MC} - 1} \sum_{i=1}^{N_{MC}} (\hat{u}_i - \mu(\hat{u}))^2. \quad (34)$$

The plots on the right side of Figure 1 show the empirical bias, the bias predictions (17) and (19), and their corresponding standard errors, for the structured EIV problem. The empirical bias b_e is proportional to the perturbation noise variance, and the bias predictions b_p and $b_{\bar{p}}$ coincide with the empirical bias b_e only for SNR above 40 dB. This indicates that the SNR drops to a point where the constraint $\|\mathbf{M}\| < 1$ is no longer satisfied. At an SNR of 30 dB the perturbation noise affects the bias prediction from observed data and it is three times smaller than the empirical bias.

The absolute and relative errors between the predicted and the empirical bias are shown in Figure 6. It can be seen that the absolute errors are proportional to the perturbation variance, and the relative errors are lower than 5% for SNR higher than 40 dB.

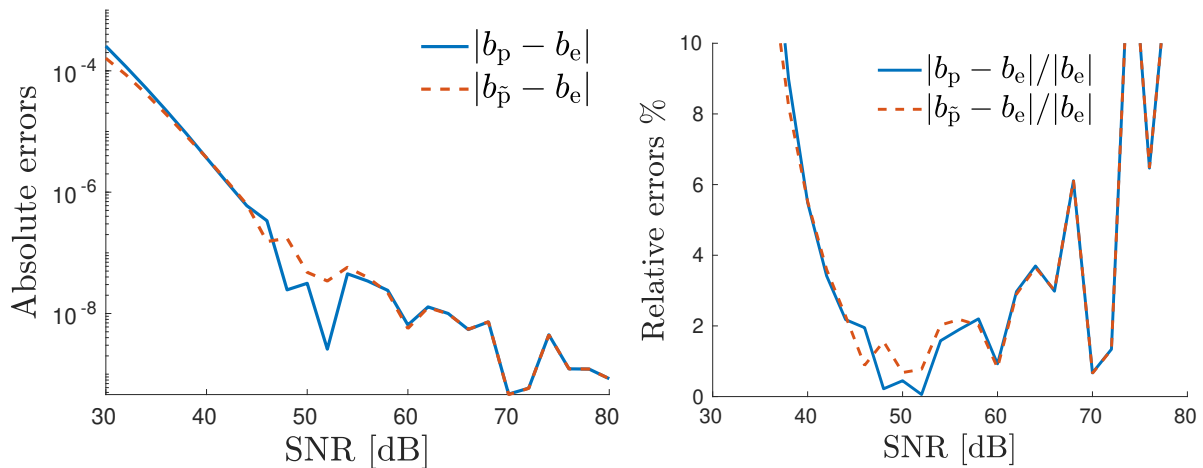


Figure 6: The MC simulation shows that when we solve a structured EIV problem with LS, the absolute errors (left) between the predicted bias and the empirical bias b_e are proportional to the perturbation noise variance, and the relative errors (right) are smaller than 5% only for SNR above 40 dB.

165 The absolute and relative errors between the empirical and the predicted variance, equations (18) and (20), are shown in Figure 7. These absolute errors are proportional to the perturbation noise variance, whereas the relative errors are lower than 5% for SNR higher than 45 dB.

The absolute and relative errors between the two predictions from observed data and from exact data, are shown in Figure 8. These absolute errors are also proportional to the perturbation noise variance, and the 170 relative errors show that the bias and variance predictions from either of the two alternatives are equivalent for SNR higher than 45 dB.

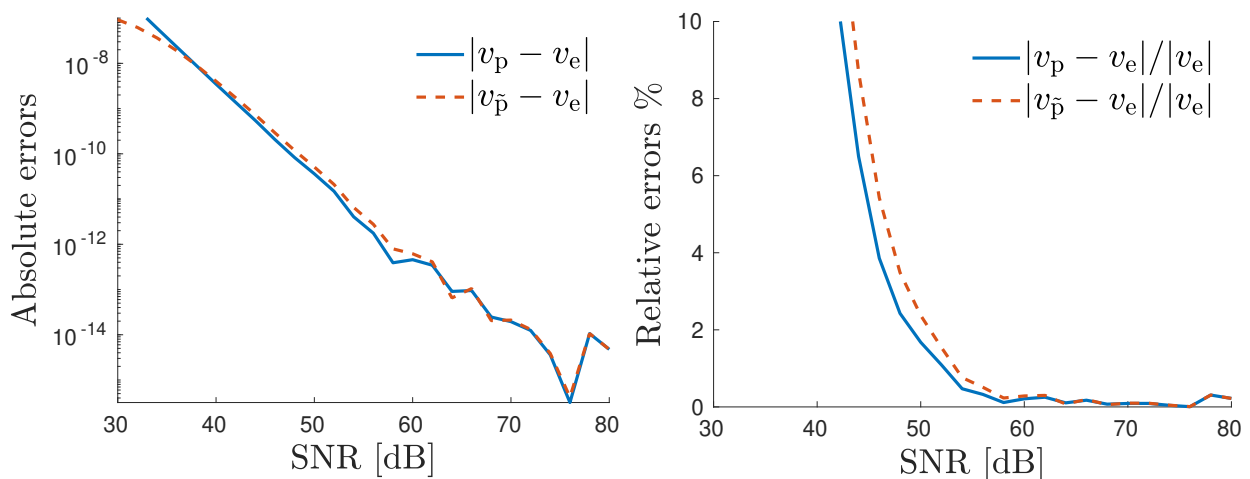


Figure 7: The MC simulation shows that when we solve a structured EIV problem with LS, the absolute errors (left) between the predicted and the empirical variance are proportional to the perturbation variance, and the relative errors (right) between the predicted and the empirical variance are smaller than 5% for SNR above 45 dB.

The simulation results show that the LS solution of the structured and correlated EIV problem is more sensitive to the perturbation. This represents a low limit in the SNR interval imposed by the noise level. In practical applications it is common to have SNRs of 40 dB and the user needs to be aware of the prediction error that the method has. We measure this prediction error with the mean squared error (MSE), defined

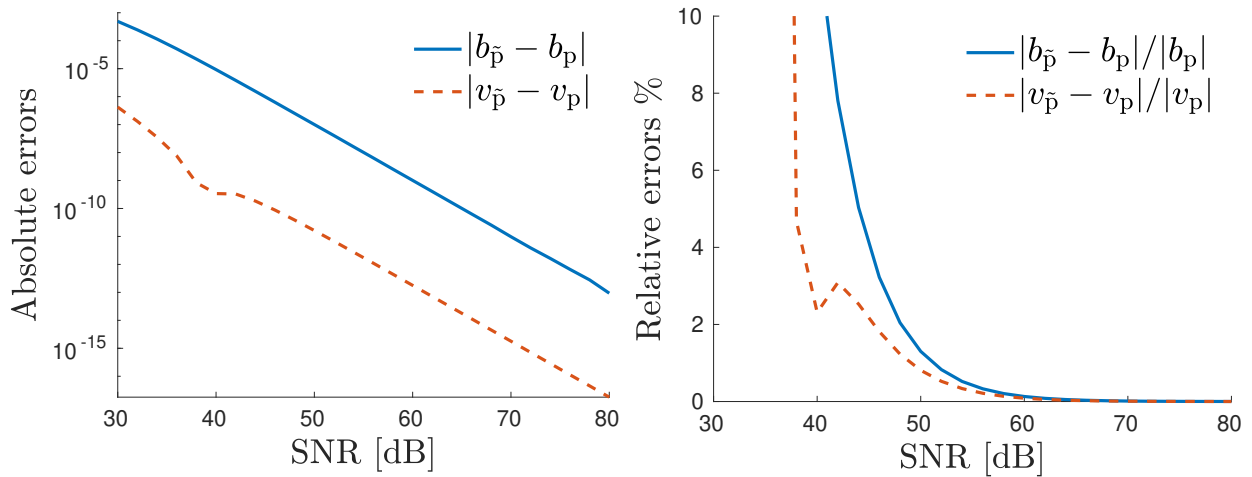


Figure 8: The MC simulation shows that when we solve a structured EIV problem with LS, the absolute errors (left) between the predictions computed from observed data and those from exact data are proportional to the perturbation noise variance. According to the relative errors (right), the substitution is valid for SNR higher than 45 dB.

as

$$\text{MSE} = \sigma^2 + b^2. \quad (35)$$

By comparing the different MSEs to the CRLB of the structured EIV problem, Figure 9 shows that the MSEs has the same proportionality as the CRLB with respect to the disturbing noise variance. The MSEs are three times larger than the CRLB. Since the difference between $\text{MSE}_{\hat{p}}$ and the CRLB is lower than one order of magnitude, the LS estimation of the structured EIV problem produces results that are comparable to the ML estimation. The $\text{MSE}_{\hat{p}}$ computed from observed data approaches the CRLB for SNR below 40 dB. This is due to the constraint violation of the Taylor series expansion.

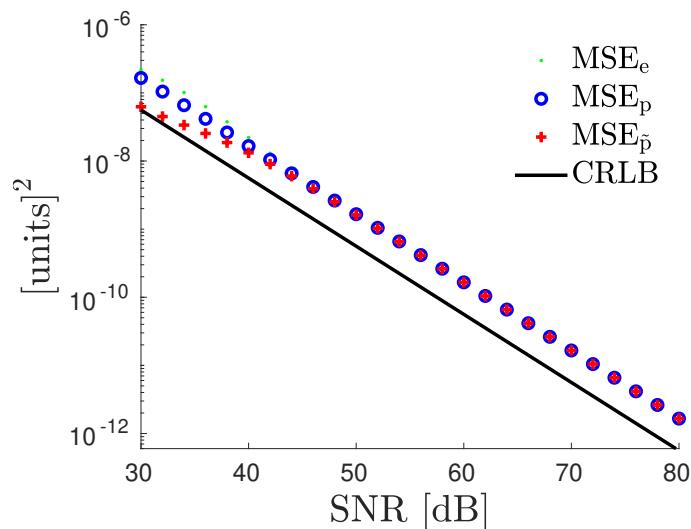


Figure 9: The mean squared errors of the LS estimate are close to the Cramér-Rao lower bound. This is a positive indication of the goodness of the LS estimator for the structured EIV problem. The mean squared error of the empirical estimates is represented by MSE_e , and those of the predictions are MSE_p and $\text{MSE}_{\hat{p}}$. The $\text{MSE}_{\hat{p}}$ is smaller than CRLB below 40 dB of SNR because of the introduced bias error, but the difference between $\text{MSE}_{\hat{p}}$ and the CRLB is lower than one order of magnitude.

5. Conclusions

We conducted a statistical analysis of a structured errors-in-variables (EIV) estimation problem with correlation to find the first and second moments of its least-squares solution. This estimation problem occurs in metrology when we estimate the value of a measurand directly from the sensor transient response. The data-driven estimation of the physical quantity is formulated as a structured EIV problem with correlation that uses the observed transient response to construct both the regression matrix and the regressor. The real-time implementation of the method uses a recursive least squares algorithm that is simple and has low computational complexity. The assessment of the uncertainty is done using the estimate bias and variance.

The conducted statistical analysis produced expressions that predict the estimate bias and variance for given sample size and perturbation level of the observed response. The Monte Carlo simulation validated the predictions. We compared the results of solving an unstructured and uncorrelated EIV problem with a structured and correlated EIV problem to understand how the structure and the correlation impacts in the estimation. We found that the predictions in the structured case are more susceptible to perturbations. This is due to the two approximations involved, a second-order Taylor series expansion of the estimate, and the substitution of perturbed data on the prediction expressions. The relative error results indicate that the estimate bias, and variance are predicted using the derived expressions, and the observed data. The mean squared error of the estimate is close to the results of the maximum likelihood estimate given by the Cramér-Rao lower bound.

The bias and variance can be accurately predicted, provided that the Taylor series expansion is valid. This constraint has to be taken into account to ensure the effectiveness of the method in practical applications. In the example, it was observed that when the SNR lies outside the validity region, the bias and variance estimation was at most three times larger than the empirical values.

The methodology presented in this paper can be applied to estimate the uncertainty of the solutions to other structured EIV problems. The bias and variance expressions obtained after the statistical analysis depend on each specific structure.

Acknowledgements

The research leading to these results has received funding from the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP7/2007–2013) / ERC Grant agreement number 258581 "Structured low-rank approximation: Theory, algorithms, and applications" and Fund for Scientific Research Vlaanderen (FWO) projects G028015N "Decoupling multivariate polynomials in non-linear system identification" and G090117N "Block-oriented nonlinear identification using Volterra series"; and Fonds de la Recherche Scientifique (FNRS) – FWO Vlaanderen under Excellence of Science (EOS) Project no 30468160 "Structured low-rank matrix / tensor approximation: numerical optimization-based

algorithms and applications”; and the Flemish Government (Methusalem award METH1), and VLAIO Innovation Mandate project number HBC.2016.0235. D. Verbeke is funded by a PhD grant from the Research Foundation Flanders (FWO).

Appendix 1. Derivation of bias and covariance expressions.

The bias and covariance of the least-squares (LS) estimate (10) are obtained using the mathematical expectation in the definitions (11) and (12). For an unstructured and uncorrelated EIV problem, the expected value, and the covariance of $\hat{\mathbf{x}}$ are approximated by

$$\begin{aligned}\mathbb{E}\{\hat{\mathbf{x}}\} &\approx \mathbf{x} + \mathbf{Q}^{-1}\mathbb{E}\left\{\mathbf{K}^{\top}\mathbf{E}\mathbf{Q}^{-1}\mathbf{K}^{\top}\mathbf{E} + \mathbf{E}^{\top}\mathbf{K}\mathbf{Q}^{-1}\mathbf{K}^{\top}\mathbf{E} - \mathbf{E}^{\top}\mathbf{E}\right\}\mathbf{x}, \quad \text{and} \\ \mathbf{C}(\hat{\mathbf{x}}) &\approx \mathbf{xx}^{\top} + \mathbf{Q}^{-1}\mathbb{E}\left\{\mathbf{K}^{\top}\boldsymbol{\epsilon}\boldsymbol{\epsilon}^{\top}\mathbf{K} + \mathbf{K}^{\top}\mathbf{E}\mathbf{xx}^{\top}\mathbf{E}^{\top}\mathbf{K}\right\}\mathbf{Q}^{-1} \\ &\quad + \mathbf{Q}^{-1}\mathbb{E}\left\{\mathbf{K}^{\top}\mathbf{E}\mathbf{Q}^{-1}\mathbf{K}^{\top}\mathbf{E} + \mathbf{E}^{\top}\mathbf{K}\mathbf{Q}^{-1}\mathbf{K}^{\top}\mathbf{E} - \mathbf{E}^{\top}\mathbf{E}\right\}\mathbf{xx}^{\top} \\ &\quad + \mathbf{xx}^{\top}\mathbb{E}\left\{\mathbf{E}^{\top}\mathbf{K}\mathbf{Q}^{-1}\mathbf{E}^{\top}\mathbf{K} + \mathbf{E}^{\top}\mathbf{K}\mathbf{Q}^{-1}\mathbf{K}^{\top}\mathbf{E} - \mathbf{E}^{\top}\mathbf{E}\right\}\mathbf{Q}^{-1} - \mathbb{E}\{\hat{\mathbf{x}}\}\mathbb{E}\{\hat{\mathbf{x}}\}^{\top},\end{aligned}\tag{36}$$

where we have considered the second order Taylor series approximation (9), and we have removed the terms of zero expected value, and the terms of order higher than 2. After an elementwise evaluation of the corresponding expected values in (36), the expressions result in

$$\begin{aligned}\mathbb{E}\{\hat{\mathbf{x}}\} &\approx \mathbf{x} + \mathbf{b}_p(\hat{\mathbf{x}}) = \mathbf{x} + \sigma_{\mathbf{E}}^2\mathbf{Q}^{-1}(2\mathbf{I} + 2n\mathbf{I} - T\mathbf{I})\mathbf{x}, \quad \text{and} \\ \mathbf{C}_p(\hat{\mathbf{x}}) &\approx \sigma_{\boldsymbol{\epsilon}}^2\mathbf{Q}^{-1} + \sigma_{\mathbf{E}}^2\text{trace}(\mathbf{xx}^{\top})\mathbf{Q}^{-1} - \sigma_{\mathbf{E}}^4(2 + 2n - T)^2\mathbf{Q}^{-1}\mathbf{xx}^{\top}\mathbf{Q}^{-1},\end{aligned}\tag{37}$$

215 from where equations (13) and (14) are obtained.

On the other hand, due to the correlation, the expressions that approximate the expected value of the LS estimate of the structured and EIV problem (2) have a different form:

$$\begin{aligned}\mathbb{E}\{\hat{\mathbf{x}}\} &\approx \mathbf{x} + \mathbf{Q}^{-1}\mathbb{E}\left\{\mathbf{K}^{\top}\mathbf{E}\mathbf{Q}^{-1}\mathbf{K}^{\top}\mathbf{E} + \mathbf{E}^{\top}\mathbf{K}\mathbf{Q}^{-1}\mathbf{K}^{\top}\mathbf{E} - \mathbf{E}^{\top}\mathbf{E}\right\}\mathbf{x} \\ &\quad + \mathbf{Q}^{-1}\mathbb{E}\left\{\mathbf{E}^{\top}\boldsymbol{\epsilon} - \mathbf{K}^{\top}\mathbf{E}\mathbf{Q}^{-1}\mathbf{K}^{\top}\boldsymbol{\epsilon} - \mathbf{E}^{\top}\mathbf{K}\mathbf{Q}^{-1}\mathbf{K}^{\top}\boldsymbol{\epsilon}\right\}, \quad \text{and} \\ \mathbf{C}(\hat{\mathbf{x}}) &\approx \mathbf{xx}^{\top} + \mathbf{Q}^{-1}\mathbb{E}\left\{\mathbf{K}^{\top}\boldsymbol{\epsilon}\boldsymbol{\epsilon}^{\top}\mathbf{K} + \mathbf{K}^{\top}\mathbf{E}\mathbf{xx}^{\top}\mathbf{E}^{\top}\mathbf{K} - \mathbf{K}^{\top}\mathbf{E}\mathbf{x}\boldsymbol{\epsilon}^{\top}\mathbf{K} - \mathbf{K}^{\top}\boldsymbol{\epsilon}\mathbf{x}^{\top}\mathbf{E}^{\top}\mathbf{K}\right\}\mathbf{Q}^{-1} \\ &\quad + \mathbf{Q}^{-1}\mathbb{E}\left\{\mathbf{K}^{\top}\mathbf{E}\mathbf{Q}^{-1}\mathbf{K}^{\top}\mathbf{E} + \mathbf{E}^{\top}\mathbf{K}\mathbf{Q}^{-1}\mathbf{K}^{\top}\mathbf{E} - \mathbf{E}^{\top}\mathbf{E}\right\}\mathbf{xx}^{\top} \\ &\quad + \mathbf{xx}^{\top}\mathbb{E}\left\{\mathbf{E}^{\top}\mathbf{K}\mathbf{Q}^{-1}\mathbf{E}^{\top}\mathbf{K} + \mathbf{E}^{\top}\mathbf{K}\mathbf{Q}^{-1}\mathbf{K}^{\top}\mathbf{E} - \mathbf{E}^{\top}\mathbf{E}\right\}\mathbf{Q}^{-1} - \mathbb{E}\{\hat{\mathbf{x}}\}\mathbb{E}\{\hat{\mathbf{x}}\}^{\top} \\ &\quad + \mathbf{Q}^{-1}\mathbb{E}\left\{\mathbf{E}^{\top}\boldsymbol{\epsilon} - \mathbf{K}^{\top}\mathbf{E}\mathbf{Q}^{-1}\mathbf{K}^{\top}\boldsymbol{\epsilon} - \mathbf{E}^{\top}\mathbf{K}\mathbf{Q}^{-1}\mathbf{K}^{\top}\boldsymbol{\epsilon}\right\}\mathbf{x}^{\top} \\ &\quad + \mathbf{x}\mathbb{E}\left\{\boldsymbol{\epsilon}^{\top}\mathbf{E} - \boldsymbol{\epsilon}^{\top}\mathbf{K}\mathbf{Q}^{-1}\mathbf{E}^{\top}\mathbf{K} - \boldsymbol{\epsilon}^{\top}\mathbf{K}\mathbf{Q}^{-1}\mathbf{K}^{\top}\mathbf{E}\right\}\mathbf{Q}^{-1}.\end{aligned}\tag{38}$$

These expressions have the non zero expected value terms, up to the second order. We have then

$$\begin{aligned}
\mathbb{E}\{\hat{\mathbf{x}}\} &= \mathbf{x} + \mathbf{b}_p(\hat{\mathbf{x}}) \approx \mathbf{x} + \underbrace{\mathbf{Q}^{-1}\mathbf{K}^\top \mathbb{E}\{\mathbf{E}\mathbf{Q}^{-1}\mathbf{K}^\top \mathbf{E}\}}_{\mathbf{B}_1} - \underbrace{\mathbf{Q}^{-1} \mathbb{E}\{\mathbf{E}^\top (\mathbf{I} - \mathbf{K}\mathbf{Q}^{-1}\mathbf{K}^\top) \mathbf{E}\}}_{\mathbf{B}_2} \mathbf{x} \\
&\quad - \underbrace{\mathbf{Q}^{-1}\mathbf{K}^\top \mathbb{E}\{\mathbf{E}\mathbf{Q}^{-1}\mathbf{K}^\top \boldsymbol{\epsilon}\}}_{\mathbf{B}_3} + \underbrace{\mathbf{Q}^{-1} \mathbb{E}\{\mathbf{E}^\top (\mathbf{I} - \mathbf{K}\mathbf{Q}^{-1}\mathbf{K}^\top) \boldsymbol{\epsilon}\}}_{\mathbf{B}_4}, \quad \text{and} \\
\mathbf{C}(\hat{\mathbf{x}}) &\approx \mathbf{Q}^{-1}\mathbf{K}^\top \left(\underbrace{\mathbb{E}\{\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top\}}_{\sigma_\epsilon^2 \mathbf{I}_{T-n}} + \underbrace{\mathbb{E}\{\mathbf{E}\mathbf{x}\mathbf{x}^\top \mathbf{E}^\top\}}_{\mathbf{C}_1} - \underbrace{\mathbb{E}\{\mathbf{E}\mathbf{x}\boldsymbol{\epsilon}^\top\}}_{\mathbf{C}_2} - \underbrace{\mathbb{E}\{\boldsymbol{\epsilon}\mathbf{x}^\top \mathbf{E}^\top\}}_{\mathbf{C}_2^\top} \right) \mathbf{K}\mathbf{Q}^{-1} - \mathbf{b}_p(\hat{\mathbf{x}}) \mathbf{b}_p^\top(\hat{\mathbf{x}}).
\end{aligned} \tag{39}$$

from where the expressions (17) and (18) are obtained.

Appendix 2. Proof of Lemma 1

PROOF. In the first case considered in the lemma, the elements of the expected value $\mathbf{Z} = \mathbb{E}\{\mathbf{E}\mathbf{A}\mathbf{E}\}$ are

$$z_{ij} = \mathbb{E}\{\mathbf{E}\mathbf{A}\mathbf{E}\}_{ij} = \mathbb{E}\{\mathbf{E}_i^\top \mathbf{A}\mathbf{E}_j\} = \text{tr}\left(\mathbf{A} \mathbb{E}\{\mathbf{E}_j \mathbf{E}_i^\top\}\right), \tag{40}$$

where \mathbf{E}_i^\top is the i -th row and \mathbf{E}_j is the j -th column of \mathbf{E} , for $i = 1, \dots, T-n$, and $j = 2, \dots, n+1$. The matrix $\mathbb{E}\{\mathbf{E}_j \mathbf{E}_i^\top\}$ is the product of σ_ϵ^2 times a matrix whose elements are 0 in the first column, 2 in the $(j-i-1)$ -th diagonal, and -1 in the $(j-i-2)$ -th and $(j-i)$ -th diagonals, and zeros elsewhere. Using the definition of the second differential operator we express

$$\mathbb{E}\{\mathbf{E}_j \mathbf{E}_i^\top\} = \sigma_\epsilon^2 \begin{bmatrix} \mathbf{0}_{T-n} & \mathbf{D}_{T-n \times n}^{2,j-i} \end{bmatrix}. \tag{41}$$

The proof of the other cases in the Lemma is similar. For the second case, the elements of the expected value $\mathbf{Z} = \mathbb{E}\{\mathbf{E}^\top \mathbf{A}\mathbf{E}\}$ are

$$z_{ij} = \mathbb{E}\{\mathbf{E}^\top \mathbf{A}\mathbf{E}\}_{ij} = \mathbb{E}\{\mathbf{E}_i^\top \mathbf{A}\mathbf{E}_j\} = \text{tr}\left(\mathbf{A} \mathbb{E}\{\mathbf{E}_j \mathbf{E}_i^\top\}\right), \tag{42}$$

where now \mathbf{E}_i^\top is the transposed i -th column and \mathbf{E}_j is the j -th column of \mathbf{E} , for $i = 2, \dots, n+1$, and $j = 2, \dots, n+1$. The matrix $\mathbb{E}\{\mathbf{E}_j \mathbf{E}_i^\top\}$ is σ_ϵ^2 times a matrix whose elements are 2 in the $(j-i)$ -th diagonal, and -1 in the $(j-i-1)$ -th and $(j-i+1)$ -th diagonals, and zeros elsewhere. Therefore we have

$$\mathbb{E}\{\mathbf{E}_j \mathbf{E}_i^\top\} = \sigma_\epsilon^2 \mathbf{D}_{T-n \times T-n}^{2,j-i+1}. \tag{43}$$

The expected values that involve the vector $\boldsymbol{\epsilon}$ are especial cases of the previous cases. The vector $\boldsymbol{\epsilon}$ in the expected values $\mathbb{E}\{\mathbf{E}\mathbf{A}\boldsymbol{\epsilon}\}$, $\mathbb{E}\{\mathbf{E}^\top \mathbf{A}\boldsymbol{\epsilon}\}$, and $\mathbb{E}\{\mathbf{E}\mathbf{A}\boldsymbol{\epsilon}^\top\}$ is

$$\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon(n+1) & \epsilon(n+2) & \cdots & \epsilon(T) \end{bmatrix}^\top, \tag{44}$$

as it is imposed by the input estimation method formulation.

References

220 References

- Azam, S., Chatzi, E., Papadimitriou, C., 2015. A dual Kalman filter approach for state estimation via output-only acceleration measurements. *Mechanical Systems and Signal Processing* 60-61, 866–886. doi:10.1016/j.ymsp.2015.02.001.
- Cai, J., Qu, X., Xu, W., Ye, G., 2016. Robust recovery of complex exponential signals from random Gaussian
225 projections via low rank Hankel matrix reconstruction. *Applied and Computational Harmonic Analysis* 41, 470–490. doi:10.1016/j.acha.2016.02.003.
- Feiz, R., Rezghi, M., 2017. A splitting method for total least squares color image restoration problem. *Journal of Visual Communication and Image Representation* 46, 48–57. doi:10.1016/j.jvcir.2017.03.001.
- Hammersley, J., Handscomb, D., 1975. *Monte Carlo Methods*. Fletcher & Son.
- 230 Jia, T., Wang, H., Shen, X., Jiang, Z., He, K., 2018. Target localization based on structured total least squares with hybrid TDOA-AOA measurements. *Signal Processing* 143, 211–221. doi:10.1016/j.sigpro.2017.09.011.
- Kiviet, J., Phillips, G., 2012. Higher-order asymptotic expansions of the least-squares estimation bias in first-order dynamic regression models. *Computational Statistics & Data Analysis* 56, 3705–3729. doi:10.1016/j.csda.2010.07.013.
235
- Kiviet, J., Phillips, G., 2014. Improved variance estimation of maximum likelihood estimators in stable first-order dynamic regression models. *Computational Statistics & Data Analysis* 76, 424–448. doi:10.1016/j.csda.2013.09.021.
- Markovsky, I., 2015. An application of system identification in metrology. *Control Engineering Practice* 43,
240 85–93. doi:10.1016/j.conengprac.2015.07.001.
- Markovsky, I., Van Huffel, S., 2007. Overview of total least-squares methods. *Signal Processing* 87, 2283–2302. doi:10.1016/j.sigpro.2007.04.004.
- Mastronardi, N., O’Leary, D., 2007. Fast robust regression algorithms for problems with Toeplitz structure. *Computational Statistics & Data Analysis* 52, 1119–1131. doi:10.1016/j.csda.2007.05.008.
- 245 Niedźwiecki, M., Meller, M., Pietrzak, P., 2016. System identification based approach to dynamic weighing revisited. *Mechanical Systems and Signal Processing* 80, 582–599. doi:10.1016/j.ymsp.2016.04.007.

- Palanthandalam-Madapusi, H., Van Pelt, T., Bernstein, D., 2010. Parameter consistency and quadratically constrained errors-in-variables least-squares identification. *International Journal of Control* 83, 862–877. doi:10.1080/00207170903470666.
- 250 Pan, Y., Luo, G.Q., Jin, H., Cao, W., 2018. Direction-of-arrival estimation with ULA: a spatial annihilating filter reconstruction perspective. *IEEE Access* 6, 23172–23179. doi:10.1109/ACCESS.2018.2828799.
- Pintelon, R., Schoukens, J., 2012. *System identification: a frequency domain approach*. 2 ed., IEEE Press, Piscataway, NJ. doi:10.1002/9781118287422.
- Rhode, S., Usevich, K., Markovsky, I., Gauterin, F., 2014. A recursive restricted total least-squares algorithm. *IEEE Transactions on Signal Processing* 62, 5652–5662. doi:10.1109/TSP.2014.2350959.
- 255 Söderström, T., 2007. Errors-in-variables methods in system identification. *Automatica* 43, 939–958. doi:10.1016/j.automatica.2006.11.025.
- Stewart, G., 1990. Stochastic perturbation theory. *SIAM Review* 32, 579–610. doi:10.1137/1032121.
- Vaccaro, R., 1994. A second-order perturbation expansion for the SVD. *SIAM J. Matrix Anal. Appl.* 15, 260 661–671. doi:10.1137/S0895479891224245.
- Van Huffel, S., Cheng, C., Mastronardi, N., Paige, C., Kukush, A., 2007. Total least squares and errors-in-variables modeling. *Computational Statistics & Data Analysis* 52, 1076–1079. doi:10.1016/j.csda.2007.07.001.
- Van Huffel, S., Vandewalle, J., 1991. *The total least squares problem: computational aspects and analysis*. 265 SIAM. doi:10.1137/1.9781611971002.
- Yeredor, A., De Moor, B., 2004. On homogeneous least-squares problems and the inconsistency introduced by mis-constraining. *Computational Statistics & Data Analysis* 47, 455–465. doi:10.1016/j.csda.2003.12.001.