Bayesian Inference of Total Least-Squares With Known Precision

FRIML, D.; VÁCLAVEK, P.

2022 IEEE 61st Conference on Decision and Control (CDC)

eISBN: 978-1-6654-6761-2

DOI: https://doi.org/10.1109/CDC51059.2022.9992409

Accepted manuscript

Bayesian Inference of Total Least-Squares With Known Precision

Dominik Friml[1] and Pavel Václavek[2]

***Abstract*— This paper provides a Bayesian analysis of the total least-squares problem with independent Gaussian noise of known variance. It introduces a derivation of the likelihood density function, conjugate prior probability-density function, and the posterior probability-density function. All in the shape of the Bingham distribution, introducing an unrecognized connection between orthogonal least-squares methods and directional analysis. The resulting Bayesian inference expands on available methods with statistical results. A recursive statistical identification algorithm of errors-in-variables models is laid-out. An application of the introduced inference is presented using a simulation example, emulating part of the identification process of linear permanent magnet synchronous motor drive parameters. The paper represents a crucial step towards enabling Bayesian statistical methods for problems with errors in variables.**

I. INTRODUCTION

One of the most widely used linear regression methods is the linear regression of the model $\widetilde{y} = \overline{H}\theta + e_y$, where θ is an unknown deterministic estimand or vector of parameters. H is the design matrix, and e_y is the Gaussian noise vector [1]. This model is still used in cases where the noise is not only present in the measurements of the observation vector y, but also in the design matrix. The method's popularity in such cases is explained by the number and convenience of the available statistical methods for this model.

Although there are some attempts to incorporate design matrix perturbations into the least-squares algorithms, mainly through robust least-squares [2], [3] or minimax mean squared error [4], the proper way of finding the solution is by rephrasing the model to the errors-in-variables (EIV) problem. In such a case, the design matrix H is known up to the additional Gaussian noise matrix E_H [5]. It has been shown [6] that for an independent, identically distributed Gaussian noise matrix E_H, the maximum likelihood solution of such model coincides with the total least-squares estimator introduced by [7].

While a detailed analysis of the maximum likelihood solution has shown progress in recent years [8], a Bayesian analysis of the EIV problem has proved to be difficult, achieving impractically complicated [9], [10] or quasi solutions [11], making statistical research of EIV problems to continue falling behind the rapidly evolving research of statistical methods based on ordinary least-squares.

This paper offers an alternative approach of analyzing the total least-squares problem, achieving a practical Bayesian statistical inference with the conjugate prior, allowing for recursive identification because the likelihood, prior, and posterior functions are shown to take the shape of the Bingham distribution. The result represents an important step towards Bayesian statistical methods in EIV problems with known precision, by discovering a connection between orthogonal regression and directional statistics.

II. BINGHAM DISTRIBUTION

The Bingham distribution naturally arises from a zero-mean multivariate normal random variable in $\mathbb{R}^q$ constrained to lie on the unit sphere $\mathbb{S}^{q-1}$. It was first introduced by [12] as an antipodally symmetric distribution on the sphere $\mathbb{S}^2$, later generalized to higher-dimensional spheres $\mathbb{S}^{q-1}$. The density function is

$$p(z|A) = c(A)^{-1} \exp\left(-z^T A z\right), \tag{1}$$

where $z \in \mathbb{R}^q$, $z^T z = 1$ and A is a $q \times q$ symmetric, positive definite matrix. The $c(A)$ is an intractable integration constant, the main complication of the Bingham distribution.

This distribution has particular applications in, for example, paleomagnetic studies [13], wind speed modelling [14], biomedical image analysis, [15], crystal orientations analysis [16] or orientation estimation [17].

The covariance matrix A can be spectrally decomposed to $A = \Gamma\Lambda\Gamma^T$, where Γ is an orthogonal matrix constructed from the eigenvectors of A, and $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_q)$ is the diagonal matrix of the eigenvalues of A. Without loss of generality, we can assume that concentration parameters Λ fulfill the identifiability constraint $\lambda_1 \geq \cdots \geq \lambda_{q-1} \geq \lambda_q = 0$. These constraints ensure identifiability, as the density does not change if a positive constant is added to the λ_is.

By decomposing covariance matrix A, concentration parameter matrix Λ and concentration matrix $\Gamma = [\gamma_1, \gamma_2, \dots, \gamma_q]$ are obtained, where $\gamma_1, \gamma_2, \dots, \gamma_{q-1}$ are concentration axes vectors with corresponding concentration parameters $\lambda_1, \lambda_2, \dots, \lambda_{q-1}$. The γ_q is the mean concentration vector or the modus vector.

The standard form of the distribution is

$$p(z|\Lambda) = c(\Lambda)^{-1} \exp\left\{-\sum_{i=1}^{q-1} \lambda_i z_i^2\right\} \tag{2}$$

with respect to a uniform measure on the sphere and

$$c(\Lambda) = \int_{z\in\mathbb{S}^{q-1}} \exp\left\{\sum_{i=1}^{q-1} \lambda_i z_i^2\right\} d\mathbb{S}^{q-1}(z). \tag{3}$$

[1]Dominik Friml is with the Faculty of Electrical Engineering and Communication, Brno University of Technology, and with the Central European Institute of Technology, Technická 12 Brno, Czech Republic dominik.friml@vutbr.cz

[2]Pavel Václavek is with the Faculty of Electrical Engineering and Communication, Brno University of Technology, and with the Central European Institute of Technology, Technická 12 Brno, Czech Republic pavel.vaclavek@vutbr.cz

Note, that the standard form does not include concentration vectors Γ, which conveniently follows from the rotation of the distribution such that the concentration vectors are aligned with the main axes. If Z follows a Bingham distribution with density $p(z|A)$, then $W = \Gamma X$ follows a Bingham distribution with density $p(z|\Lambda)$, see [18] and [12] for a more detailed explanation.

Suppose $Z^{(N)}$ to be a set of N samples of unit vectors in $\mathbb{S}^{q-1}$ sampled from a Bingham distribution with density $p(z|\Lambda)$. The likelihood function is obtained as

$$\mathcal{L}(\Lambda) = c(\Lambda)^{-N} \exp\left\{-\sum_{i=1}^{q-1} \lambda_i \sum_{j=1}^{N} \left(z_j^i\right)^2\right\} \tag{4}$$

$$= c(\Lambda)^{-N} \exp\left\{-\sum_{i=1}^{q-1} \lambda_i \tau_i\right\}, \tag{5}$$

where $\tau_i = \frac{1}{N}\sum_{j=1}^{N}\left(z_j^i\right)^2$. The sufficient statistics for Λ is the set of $\{N, \tau_1, \ldots, \tau_{q-1}\}$.

Moments of the Bingham distribution have, up to the authors' knowledge, not yet been derived. There are moments derived for von Misses and Fisher distributions using a method which is expected to apply also to the Bingham Distribution [19]. Numerically approximated moments can be obtained, for example, using a fast algorithm introduced by [20].

The establishment of confidence limits based on concentration parameters has been proved to be complicated. An approximate formula has been discovered by [12]. Another notable method of establishing confidence limits popular in paleomagnetism is presented in [21].

Although the cumulative distribution function of the Bingham distribution can not be expressed due to the doubly intractable normalizing constant [22], [23], sampling using the Metropolis-Hastings algorithm is possible using an approximation of the normalization constant [24]. Other methods bypassing the normalizing constant altogether are presented in [25] and [26]. In this paper, the reversible-jump MCMC [27] method introduced by [24] is utilized for convenience of the MATLAB implementation in libDirectional [28].

III. Formulation of the Errors-in-Variables Problem

An ordinary least-squares estimation is widely known for being the best linear unbiased estimator for problems, where the design matrix H is noiseless, and the noise e_y of the observation vector y is Gaussian with zero mean. This model can be expressed by

$$\widetilde{y} = \overline{y} + e_y = \overline{H}\theta + e_y,\ e_y \approx \mathcal{N}\left(e_y|0, \lambda_y^{-1}\right), \tag{6}$$

where $y \in \mathbb{R}^N$, $e_y \in \mathbb{R}^N$, $H \in \mathbb{R}^{N\times q-1}$, the sought after parameter vector $\theta \in \mathbb{R}^{q-1}$, and λ_y is the precision of the Gaussian noise present on the observation vector.

However, this model is insufficient when the error is also present in the design matrix. In such a case, it is required to extend the model with

$$\widetilde{H} = \overline{H} + E_H,\ \operatorname{vec}(E_H) \approx \mathcal{N}(\operatorname{vec}(E_H)|0, \Lambda_H^{-1}). \tag{7}$$

All observations and the design matrix contaminated with noise can be expressed in a new observation matrix $\widetilde{\Phi} = \left[\widetilde{H}, \widetilde{y}\right]$ with additional Gaussian noise E_Φ, with zero mean and precision matrix Λ_Φ^{-1}. In combination with the extended vector of unknown parameters $\vartheta = \left[\theta^T, -1\right]^T$, the problem can be reformulated as

$$\left(\widetilde{\Phi} - \Delta\Phi\right)\vartheta = 0, \tag{8}$$

where $\Delta\Phi \in \mathbb{R}^{N\times q}$ is the matrix of corrections, such that the equation holds.

The widely used solution is obtained using orthogonal regression, also known as total least-squares. The total least-squares define the corrections as orthogonal distance to the regression hyperplane. The total least-squares solution to the problem is discussed in the following section.

The problem is solved under the assumption that the observation noise variable e_y and the design matrix noise variable E_H are independent, and orthogonal to the model. Using the normalized vector of true model $\bar{z} = \frac{\bar{\vartheta}}{|\bar{\vartheta}|}$, the noise precision matrix can be provided as $\lambda_n \bar{z}\bar{z}^T = \Lambda_n$. This assumption can be summarized into

$$\widetilde{\varphi}_i = \overline{\varphi}_i + e_{\varphi i},\ e_{\varphi i} \approx \mathcal{N}(e_{\varphi i}|0, \Lambda_n^{-1}). \tag{9}$$

IV. Total Least-Squares

As stated in the previous section, total least-squares is a widely used solution to the EIV problem [7], [29]. It is also proved [30] to be an unbiased maximum likelihood estimator of the sought-after parameters θ within first-order error terms.

The total least-squares utilizes the minimizing of $\Delta\Phi(\widetilde{\Phi}, \vartheta)$ to obtain the optimal solution, leading to the following minimization problem

$$\begin{aligned} \hat{\theta} := \underset{\theta}{\text{minimize}} \left|\left|\Delta\Phi(\widetilde{\Phi}, \vartheta)\right|\right|_F \\ \text{subject to } 0 = \left(\widetilde{\Phi} - \Delta\Phi(\widetilde{\Phi}, \vartheta)\right)\vartheta. \end{aligned} \tag{10}$$

In the total least-squares, the measure of distance $\Delta\Phi$ is defined as a function of the system parameters ϑ and the observation matrix $\widetilde{\Phi}$.

$$\Delta\varphi_i(\widetilde{\varphi}_i, \vartheta) = \frac{\widetilde{\varphi}_i^T\vartheta}{\vartheta^T\vartheta}\vartheta,\ i \in \{1, 2, \ldots, N\} \tag{11}$$

As apparent from (11), the distance measure $\Delta\Phi(\widetilde{\Phi}, \vartheta)$ is an orthogonal euclidean distance of the observation vectors $\widetilde{\varphi}_i$ from the hyperplane orthogonal to ϑ, further denoted as the ϑ-hyperplane. Orthogonal projection of the observation vectors to the ϑ-hyperplane are called nuisance variables

$$\hat{\varphi}_i(\widetilde{\varphi}_i, \vartheta) = \widetilde{\varphi}_i - \frac{\widetilde{\varphi}_i^T\vartheta}{\vartheta^T\vartheta}\vartheta,\ i \in \{1, 2, \ldots, N\}. \tag{12}$$

The following theorem formulated by [29] gives the TLS solution and its uniqueness assumptions:

Theorem 1. *Solution of the total least-squares problem. Let*

$$\widetilde{\Phi} = V \Sigma V^T, \text{ where } \Sigma = diag\,(\sigma_1, \ldots, \sigma_q)$$

be a singular value decomposition of $\widetilde{\Phi}$ *and* σ_i, $i \in \{1,2,\ldots,q\}$ *be the singular values of* $\widetilde{\Phi}$. *After defining the partitioning*

$$V = \begin{array}{cc} \scriptstyle q-1 & \scriptstyle 1 \\ \multicolumn{2}{c}{\begin{bmatrix} V_{11} & v_{12} \\ v_{12}^T & v_{22} \end{bmatrix}} \end{array}, \Sigma = \begin{array}{cc} \scriptstyle q-1 & \scriptstyle 1 \\ \multicolumn{2}{c}{\begin{bmatrix} \Sigma_1 & 0 \\ 0 & \sigma_q \end{bmatrix}} \end{array},$$

a TLS solution exists if and only if v_{22} *is non-zero. In addition, it is unique if and only if* $\sigma_{q-1} \neq \sigma_q$. *In the case when the TLS solution exists and is unique, it is given by*

$$\hat{\theta} = \frac{-v_{12}}{v_{22}},$$

the corresponding TLS correction matrix is

$$\Delta\Phi = -V\,diag\,(0, \sigma_{n+m+1})\,V^T$$

and nuisance variables are

$$\hat{\Phi} = \widetilde{\Phi} + \Delta\Phi = V\,diag\,(\Sigma_1, 0)\,V^T.$$

As total least-squares methods are readily used in many applications [29], it has proved its value as an identification method for EIV problems. It, however, does not provide any information on the uncertainty of the model, which should be taken into account.

In practical applications, singular value decomposition is often avoided for its high computational complexity. The solution is usually computed as the minimum of the Raleigh quotient [31]

$$\hat{\vartheta}_{ML} := \underset{\vartheta}{\text{minimize}} \frac{\vartheta^T \left(\widetilde{\Phi}^T \widetilde{\Phi}\right) \vartheta}{\vartheta^T \vartheta}. \tag{13}$$

V. Sampling Probability-Density Function

We start by considering only one observation vector $\widetilde{\varphi}_i$ from the system of equations (8), which will in this section be abbreviated to just $\widetilde{\varphi}$. Defining the probability-density function $p(\widetilde{\varphi}|\theta)$, from which this observation was sampled, is crucial for finding the likelihood function, as the likelihood function is the product of the sampling densities for multiple observations.

This probability has to fulfill the following properties: I. the maximum of the sampling probability must lay on the ϑ-hyperplane; II. the likelihood function maximum derived from the sampling probability is required to coincide with the maximum likelihood solution stated in the previous section.

It is possible to utilize the standard form of the sampling probability function, known from the ordinary least-squares.

$$p(\widetilde{\varphi}|\hat{\varphi}, \lambda_n, \theta) = \mathcal{N}\left(\widetilde{\varphi}|\hat{\varphi} - \widetilde{\varphi}, \Lambda_n^{-1}\right). \tag{14}$$

Accounting for the stated properties, noise assumptions presented in section III and orthogonality of the nuisance variables (12), the sampling probability is formulated as

$$p(\widetilde{\varphi}|\hat{\varphi}, \Lambda_n, \theta) = p(\widetilde{\varphi}|\lambda_n, \theta), \tag{15}$$

which is easily shown to be a degenerate normal distribution

$$p(\widetilde{\varphi}|\lambda_n, \theta) \propto \lambda_n^{\frac{1}{2}} \exp\left(-\frac{\lambda_n}{2} \widetilde{\varphi}^T \left(\frac{\vartheta\vartheta^T}{\vartheta^T\vartheta}\right)^{\dagger} \widetilde{\varphi}\right), \tag{16}$$

where $(\cdot)^{\dagger}$ denotes pseudo-inverse.

VI. Posterior Density

In order to make statements about the probability of parameters ϑ given the set of N measurement vectors $\widetilde{\varphi}^{(N)} = \{\widetilde{\varphi}_1, \widetilde{\varphi}_2, \ldots, \widetilde{\varphi}_N\}$, a model providing the posterior probability-density function is required.

The unnormalized posterior density is a product of the prior distribution $p(z)$ and the product of sampling distribution densities for a given set of measurements, denoted as the likelihood function

$$\mathcal{L}(\theta|\lambda_n, \widetilde{\varphi}^{(N)}) \propto \prod_{i=1}^{N} p(\widetilde{\varphi}_i|\lambda_n, \theta), \tag{17}$$

$$\mathcal{L}(\theta|\lambda_n, \widetilde{\varphi}^{(N)}) \propto \exp\left(-\frac{\lambda_n}{2} \frac{\vartheta^T \left(\widetilde{\Phi}^T \widetilde{\Phi}\right) \vartheta}{\vartheta^T \vartheta}\right), \tag{18}$$

where the observation matrix $\widetilde{\Phi}$ is constructed by stacking individual transposed measurement vectors $\widetilde{\varphi}_j$ underneath each other. Compared to known likelihood functions for EIV, this likelihood function does not suffer from dimensionality growth due to nuisance variables [11].

Notice that the exponent is in the form of a scaled Rayleigh quotient. Maximizing the logarithmic likelihood function is identical to minimizing the Rayleigh quotient, which leads to the maximum likelihood solution presented in section IV.

Let us introduce the following substitutions:

$$\frac{\vartheta}{\sqrt{\vartheta^T \vartheta}} = \frac{\vartheta}{||\vartheta||_2} = z, \tag{19}$$

$$\frac{\lambda_n}{2}\left(\widetilde{\Phi}^T \widetilde{\Phi}\right) = A, \tag{20}$$

which changes the shape of the likelihood function to the Bingham distribution

$$\mathcal{L}(z|A^{(N)}) \propto c(A)^{-1} \exp\left(-z^T A z\right) = \mathcal{B}(z|A) \tag{21}$$

under constraint $||z||_2 = 1$, where $c(A)$ is the normalization function of A. The Bingham distribution is a known and studied probability density function, which brings noticeable advantages compared to existing TLS likelihood function formulations [9].

Note that by this substitution, the understanding of the sought-after extended parameter vector ϑ changed to z. Until now, the likelihood function expressed the likelihood of ϑ

has constrained to the last value of the vector equal to -1. The new understanding expresses the likelihood for vector z, lying on the unit circle. The vector θ can be obtained from z by normalization, using $\theta = -\frac{z_{1\dots q-1}}{z_q}$. The likelihood function $\mathcal{L}(z|A^{(N)})$ expresses the likelihood of a vector orthogonal to the sought-after hyperplane, defining the hyperplane unambiguously.

In a further derivation, reproducibility $\mathcal{B}(z|X)\mathcal{B}(z|Y) = \mathcal{B}(z|X+Y)$ of the Bingham distribution is used.

The reproducible likelihood distribution function allows the prior to be formulated in the form of the Bingham distribution $p(z) = \mathcal{B}(B)$, making the posterior distribution $p(z|C^{(N)})$ also a Bingham distribution

$$p(z|C^{(N)}) = \mathcal{B}(z|A^{(N)} + B) \propto \mathcal{L}(z|A^{(N)})p(z). \tag{22}$$

Reproducibility also allows for recursive inference

$$p(z|C^{(N+1)}) \propto \mathcal{L}(z|A^{(N)})p(z)p(z|A_{(N+1)}), \tag{23}$$

where

$$A_{(N+1)} = \frac{\lambda_n}{2} z^T \left(\widetilde{\varphi}_{(N+1)}\widetilde{\varphi}^T_{(N+1)}\right) z \tag{24}$$

.

VII. Application

Application of the presented inference can be demonstrated using simulations, as the practical application is now in the preparation phase. To indicate practical use, we will simulate identification of parameters in a linear permanent magnet synchronous motor drive with a locked mover, meaning that mechanical speed $v_m = 0$. The full discrete model can be closely studied in [32]. For simplicity, let us consider only the difference equation of the stator q current component

$$i_q(k+1) = \theta_1 i_q(k) + \theta_2 i_d(k) v_m(k) + \theta_3 v_m + \theta_4 u_q(k), \tag{25}$$

where i_d and i_q denote stator current components in the dq frame and u_q denotes the stator voltage q component. Since $v_m = 0$, parameters θ_2 and θ_3 are unidentifiable and will not be identified. This is convenient, because the low number of estimated parameters allows for graphical visualization of results.

The problem can be rearranged to (9) with a known noise precision $\lambda_n = 0.05$. The observation matrix $\widetilde{\Phi}$ is constructed by stacking transposed individual current and voltage measurement vectors $\widetilde{\varphi}_j = [u_q(j), i_q(j), i_q(j+1)]$ underneath each other. The sought after parameter vector is constructed as $\vartheta = [\theta_4, -\theta_1, -1]^T$.

While the maximum likelihood estimate does not provide any information regarding identification quality, the shape of the posterior density does. The identification quality is directly affected by the selected identification signal injected as u_q. The simulations are done with two identification signals, one providing a poor and the other providing a better identification quality, while the maximum likelihood solution is similar. The first signal is a pseudorandom binary sequence (PRBS), while the latter is a pseudorandom gaussian sequence (PRGS). The signals used can be studied in Fig. 1.

Simulation allows for a priori knowledge of the true parameters $\overline{\theta} = [0.3 \ -0.4]$. The maximum likelihood solution obtained from a singular value decomposition for $N = 1000$ samples is:

$$\hat{\theta}_{ML,\text{PRGS}} = [0.2988 \ -0.39589]^T ;$$
$$\hat{\theta}_{ML,\text{PRBS}} = [0.29921 \ -0.39926]^T ; \tag{26}$$

for PRGS and PRBS identification signals, respectively. Both maximum likelihood solutions seem to estimate the actual parameters accurately, leading to the conclusion that quality of identification for both signals is equivalent. This is misleading, as seen from the posterior density.

To avoid subjectivity, the prior distribution function is selected as $\mathcal{B}(z|A^{(50)})$, where $A^{(50)}$ denotes using only the first 50 samples of the simulation, similarly to (24). The likelihood function takes the shape of $\mathcal{B}(z|A^{(N)}_{(51)})$.

The posterior density function is expressed as

$$p(z|\Gamma\Lambda\Gamma^T) \propto \mathcal{B}(z|A^{(N)}_{(51)})\mathcal{B}(z|A^{(50)}), \tag{27}$$

with concentration parameters for signal PRBS and PRGS respectively

$$\Lambda_{\text{PRGS}} = \text{diag}\,(1.0920, 0.0100, 0)\,,$$
$$\Lambda_{\text{PRBS}} = \text{diag}\,(4.3305, 0.1686, 0)\,. \tag{28}$$

And mean directions for PRBS and PRGS respectively

$$\gamma_{31} = [-0.2678, 0.3557, 0.8954]^T\,,$$
$$\gamma_{32} = [-0.2690, 0.3664, 0.8907]^T\,. \tag{29}$$

As follows from the construction of prior probability distribution, mean directions of the posterior density function

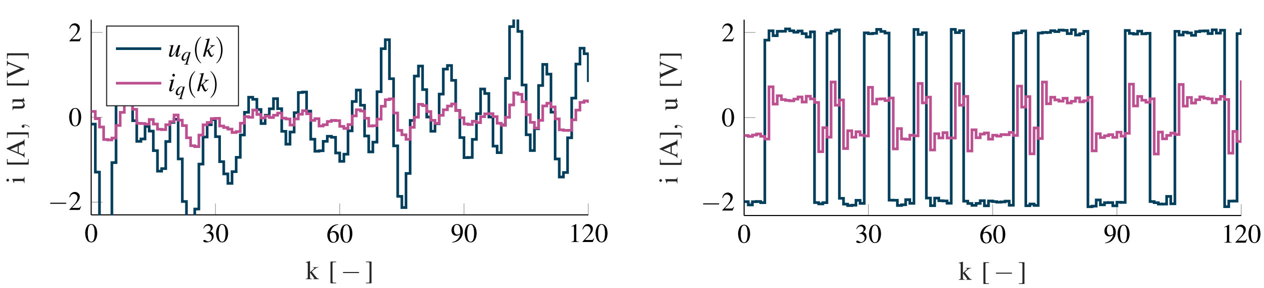

Fig. 1. Comparison of simulated measurements for PRGS and PRBS on the left and right respectively.

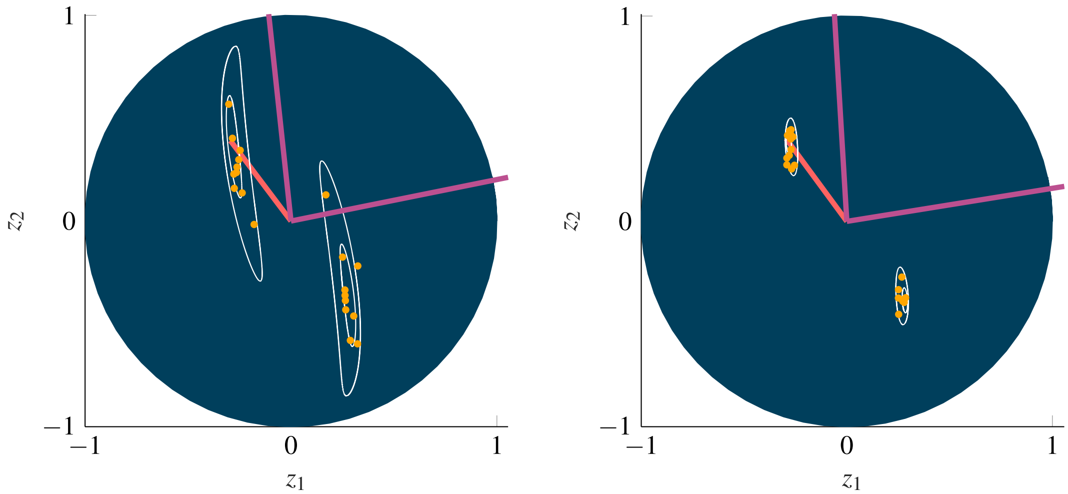

Fig. 2. Orthographic projection of the posterior density function contour plot for identification signals PRGS and PRBS on the left and right, respectively. Mean direction is depicted in red, and concentration axes in violet. Sampled parameters are depicted in yellow. Superior identification quality of PRBS is apparent from the lower variability of sampled parameters and tighter contours.

denormalized using $\hat{\theta} = -\frac{\gamma_{1\ldots q-1}}{\gamma_q}$ from Theorem 1 coincide with the maximum likelihood solutions presented in (26).

Low values of the concentration parameters Λ_{PRGS} compared to Λ_{PRBS} show a higher uncertainty. PRGS, therefore, achieves a lower identification quality than the PRBS. This fact is also apparent from Fig. 2, where the wider contour plot ellipses present the higher uncertainty in the case of the PRGS compared to PRBS. Poor quality of identification in the case of PRGS is also apparent from the wide variability of sampled parameters.

This statement can be further supported by sampling from the posterior density and plotting step responses using sampled parameters. The resulting step responses of 20 samples can be inspected in Fig. 3.

Bayesian inference of model parameters exposed the low identification quality of the PRGS based signal used and resulted in applicable estimation of parameters density.

VIII. Conclusions

This paper introduced a novel approach to statistical identification of problems with errors in variables.

We have shown that the favoured solution to the errors-in-variables problem, called total least-squares, can be reformulated as a maximum likelihood solution. The formulation of the likelihood function is different from available literature, as orthogonality of the TLS method is exploited to achieve likelihood in a distribution known from directional statistics, the Bingham distribution.

The conjugate prior for the Bingham likelihood function is derived, achieving Bayesian inference allowing for a recursive statistical identification for errors in variables, which up to the authors' knowledge, has not been achieved yet.

Formulation of the posterior density function allows for the derivation of Bayesian methods for EIV, allowing for defining the maximum likelihood estimators, maximum a posteriori estimators, establishing confidence limits, sampling from the posterior density, and obtaining statistical identification results. The form of the posterior density function also allows for recursive identification, making this readily usable in practice. We presented an example of such possibilities on the identification of a linear permanent magnet synchronous motor, simplified to allow the graphical representation of results.

Further research will expand on this idea of using Bayesian total least-squares methods, mainly in deriving the inference

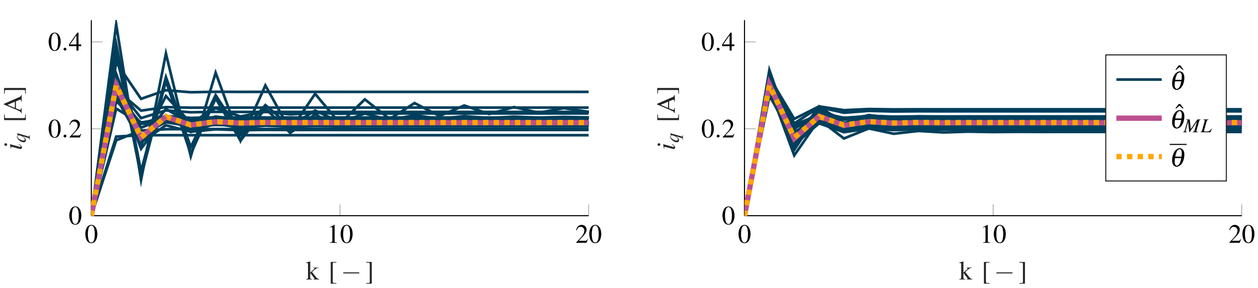

Fig. 3. Step responses of the simulated current system using true parameters $\overline{\theta}$, maximum likelihood parameters $\hat{\theta}_{ML}$ and parameters $\hat{\theta}$ sampled from the posterior density function for identification signals PRGS and PRBS on the left and right respectively. Better identification quality of the PRBS is apparent from the lower variability of the responses.

for cases with an unknown precision λ_n. Other possibilities are available, such as deriving smoothing, filtering, or estimating algorithms for EIV problems, or using total least-squares Bayesian inference in the decision-making for EIV problems.

ACKNOWLEDGMENT

This research has been financially supported by the Technology Agency of the Czech Republic, under the project NCK KUI TN01000024.

The completion of this paper was made possible by the grant No. FEKT-S-20-6205 - "Research in Automation, Cybernetics and Artificial Intelligence within Industry 4.0", financially supported by the Internal Science Fund of Brno University of Technology.

The work was supported by the RICAIP that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 857306, and from the Ministry of Education, Youth and Sports under the OP RDE grant agreement No. CZ.02.1.01/0.0/0.0/17_043/0010085.

REFERENCES

[1] U. Forssell and L. Ljung, "Closed-loop identification revisited," *Automatica*, vol. 35, no. 7, pp. 1215–1241, jul 1999.

[2] L. El Ghaoui and H. Lebret, "Robust solutions to least-squares problems with uncertain data," *SIAM Journal on Matrix Analysis and Applications*, vol. 18, no. 4, pp. 1035–1064, jul 1997.

[3] S. Chandrasekaran, G. H. Golub, M. Gu, and A. H. Sayed, "Parameter Estimation in the Presence of Bounded Data Uncertainties," *SIAM Journal on Matrix Analysis and Applications*, vol. 19, no. 1, pp. 235–252, jul 1998.

[4] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, "Robust mean-squared error estimation in the presence of model uncertainties," *IEEE Transactions on Signal Processing*, vol. 53, no. 1, pp. 168–181, jan 2005.

[5] T. Söderström, "Errors-in-variables methods in system identification," *Automatica*, vol. 43, no. 6, pp. 939–958, jun 2007.

[6] L. A. Stefanski, "Measurement error models," in *Statistics in the 21st Century*. Wiley, 2001, pp. 461–470.

[7] S. Van Huffel, "The Generalized Total Least Squares Problem : Formulation, Algorithm and Properties," in *Numerical Linear Algebra, Digital Signal Processing and Parallel Algorithms*. Springer, Berlin, Heidelberg, 1991, pp. 651–660.

[8] J. L. Crassidis and Y. Cheng, "Maximum likelihood analysis of the total least squares problem with correlated errors," *Journal of Guidance, Control, and Dynamics*, vol. 42, no. 6, pp. 1204–1217, apr 2019.

[9] O. Nestares, D. J. Fleet, and D. J. Heeger, "Likelihood functions and confidence bounds for total-least-squares problems," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1. IEEE Computer Society, 2000, pp. 523–530.

[10] N. A. Rozliman, A. I. N. Ibrahim, and R. M. Yunus, "Bayesian approach to errors-in-variables in regression models," in *AIP Conference Proceedings*, vol. 1842, no. 1. AIP Publishing LLC AIP Publishing, may 2017, p. 030018.

[11] X. Fang, B. Li, H. Alkhatib, W. Zeng, and Y. Yao, "Bayesian inference for the Errors-In-Variables model," *Studia Geophysica et Geodaetica*, vol. 61, no. 1, pp. 35–52, aug 2017.

[12] C. Bingham, "An Antipodally Symmetric Distribution on the Sphere," *The Annals of Statistics*, vol. 2, no. 6, pp. 1201–1225, 1974.

[13] T. C. Onstott, "Application of the Bingham distribution function in palaeomagnetic studies." *Journal of Geophysical Research*, vol. 85, no. B3, pp. 1500–1510, mar 1980.

[14] S. L. Goh, M. Chen, D. H. Popović, K. Aihara, D. Obradovic, and D. P. Mandic, "Complex-valued forecasting of wind profile," *Renewable Energy*, vol. 31, no. 11, pp. 1733–1750, sep 2006.

[15] F. Zhang, E. R. Hancock, C. Goodlett, and G. Gerig, "Probabilistic white matter fiber tracking using particle filtering and von Mises-Fisher sampling," *Medical Image Analysis*, vol. 13, no. 1, pp. 5–18, feb 2009.

[16] M. A. Bingham, B. K. Lograsso, and F. C. Laabs, "A statistical analysis of the variation in measured crystal orientations obtained through electron backscatter diffraction," *Ultramicroscopy*, vol. 110, no. 10, pp. 1312–1319, sep 2010.

[17] I. Gilitschenski, G. Kurz, S. J. Julier, and U. D. Hanebeck, "Unscented Orientation Estimation Based on the Bingham Distribution," *IEEE Transactions on Automatic Control*, vol. 61, no. 1, pp. 172–177, jan 2016.

[18] A. Kume and S. G. Walker, "Sampling from compositional and directional distributions," *Statistics and Computing*, vol. 16, no. 3, pp. 261–265, sep 2006.

[19] T. Hillen, K. J. Painter, A. C. Swan, and A. D. Murtha, "Moments of von Mises and Fisher distributions and applications," *Mathematical Biosciences and Engineering*, vol. 14, no. 3, pp. 673–694, dec 2017.

[20] Y. Luo, J. Xu, and P. Zhang, "A Fast Algorithm for the Moments of Bingham Distribution," *Journal of Scientific Computing*, vol. 75, no. 3, pp. 1337–1350, jun 2018.

[21] J. J. Love, "Paleomagnetic Principles and Practice," *Eos, Transactions American Geophysical Union*, vol. 81, no. 16, pp. 172–172, 2000.

[22] J. Møller, A. N. Pettitt, R. Reeves, and K. K. Berthelsen, "Miscellanea an efficient Markov chain Monte Carlo method for distributions with intractable normalising constants," *Biometrika*, vol. 93, no. 2, pp. 451–458, jun 2006.

[23] I. Murray, Z. Ghahramani, and D. J. MacKay, "MCMC for doubly-intractable distributions," in *Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence, UAI 2006*, jun 2006, pp. 359–366.

[24] A. Kume and S. G. Walker, "On the Bingham distribution with large dimension," *Journal of Multivariate Analysis*, vol. 124, pp. 345–352, feb 2014.

[25] C. J. Fallaize and T. Kypraios, "Exact Bayesian inference for the Bingham distribution," *Statistics and Computing*, vol. 26, no. 1-2, pp. 349–360, jan 2016.

[26] M. G. Tsionas, "Note on posterior inference for the Bingham distribution," *Communications in Statistics - Theory and Methods*, vol. 47, no. 12, pp. 3022–3028, jun 2018.

[27] P. J. Green, "Reversible jump Markov chain monte carlo computation and Bayesian model determination," *Biometrika*, vol. 82, no. 4, pp. 711–732, dec 1995.

[28] G. Kurz, I. Gilitschenski, F. Pfaff, L. Drude, U. D. Hanebeck, R. Haeb-Umbach, and R. Y. Siegwart, "Directional Statistics and Filtering Using libDirectional," *Journal of Statistical Software*, vol. 89, no. 4, pp. 1–31, 2019.

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

[30] G. H. Golub and C. F. van Loan, "An Analysis of the Total Least Squares Problem," *SIAM Journal on Numerical Analysis*, vol. 17, no. 6, pp. 883–893, jul 1980.

[31] S. Rhode, F. Bleimund, and F. Gauterin, "Recursive generalized total least squares with noise covariance estimation," in *IFAC Proceedings Volumes (IFAC-PapersOnline)*, vol. 19, no. 3. Elsevier, jan 2014, pp. 4637–4643.

[32] M. Kozubik and P. Vaclavek, "Speed Control of PMSM with Finite Control Set Model Predictive Control Using General-purpose Computing on GPU," in *IECON Proceedings (Industrial Electronics Conference)*, vol. 2020-Octob. IEEE Computer Society, oct 2020, pp. 379–383.