Authors' response to the referee's report on "Software for weighted structured low-rank approximation"

Ivan Markovsky and Konstantin Usevich

February 26, 2013

We thank the referee for their relevant and useful comments. In this document, we quote in **bold face** comments/questions from the report. Our replies follow in ordinary print. In blue, we quote passages from the revised manuscript.

- **"since the authors introduce a new interface to the statistical software R, I think that it would be interesting if the authors could work out a statistical example in detail"**

An example of errors-in-variables estimation, motivated by [Mar08, Sections 3.1 and 3.7], replaces the sum-of-damped exponentials example in the first version of the paper. Asymptotic statistical analysis (*e.g.*, consistency and asymptotic normality of the estimator) can not be done in the case of a sum-of-damped exponentials model because the data sequence converges exponentially to zero. For this reason, we selected a new example where a persistently exciting input signal is present and the asymptotic analysis is applicable.

As illustrated by the simulation results in the revised Section 5.1, the solution of the structured low-rank approximation problem yields a consistent estimator in the errors-in-variables setup. (This is an empirical confirmation for the theoretical results in [KMV05].) In the case of Gaussian noise, the estimator is moreover maximum likelihood and has error convergence rate of $1/\sqrt{T}$, where T is the sample size. Apart from the error convergence, in the new example, we show the confidence ellipsoid for the parameter estimates.

Errors-in-variables deconvolution

The convolution $h \star u$ of the sequences

$$u = \big(u(1), \dots, u(T)\big) \qquad \text{and} \qquad h = \big(h(0), h(1), \dots, h(n-1)\big)$$

is a sequence y, defined by the convolution sum

$$y(t) := \sum_{\tau=0}^{n} h(\tau)u(t-\tau), \qquad \text{for} \quad t = n, n+1, \dots, T. \tag{CONV}$$

With some abuse of notation, we denote by u, h, and y both the sequences and the corresponding vectors:

$$u := \begin{bmatrix} u(1) \\ \vdots \\ u(T) \end{bmatrix} \in \mathbb{R}^T, \qquad h := \begin{bmatrix} h(0) \\ \vdots \\ h(n-1) \end{bmatrix} \in \mathbb{R}^n, \qquad y := \begin{bmatrix} y(n) \\ \vdots \\ y(T) \end{bmatrix} \in \mathbb{R}^{T-n+1}.$$

Using the Toeplitz matrix

$$\mathscr{T}(u) = \begin{bmatrix} u(n) & u(n-1) & \cdots & u(2) & u(1) \\ u(n+1) & u(n) & \ddots & \ddots & u(2) \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ u(T) & u(T-1) & \cdots & u(T-n+2) & u(T-n+1) \end{bmatrix} \in \mathbb{R}^{(T-n+1)\times n},$$

the convolution sum (CONV) can be written as a matrix-vector product

$$y = \mathscr{T}(u)h. \tag{CONV'}$$

Consider the errors-in-variables data generating model [KMV05]:

$$u = \bar{u} + \widetilde{u} \qquad \text{and} \qquad y = \bar{y} + \widetilde{y}, \tag{EIV}$$

where (u, y) is the measured data, $(\bar{u}, \bar{y})$ is its "true value", and $(\widetilde{u}, \widetilde{y})$ is the measurement noise. The true data satisfy the relation $\bar{y} = \mathscr{T}(\bar{u})\bar{h}$, for some "true parameter" vector $\bar{h}$. The goal in the errors-in-variables problem is to estimate consistently and efficiently $\bar{h}$ from the noisy data.

The true parameter vector, however, is not identifiable (a solution is not unique) unless there is prior knowledge about the measurement noise. We assume that the measurement noise elements $\widetilde{u}(t)$ and $\widetilde{y}(t)$ are zero mean, independent, and identically distributed, but the noise variance is unknown. It is proven in [KMV05] that in this case the structured total least squares approximate solution of the overdetermined system of linear equations (CONV') yields a consistent estimator. If in addition, the noise distribution is normal, it is a maximum likelihood estimator and is asymptotically normal.

The structured total least squares problem is equivalent to low-rank approximation of the matrix $\begin{bmatrix} \mathscr{T}(u) & y \end{bmatrix}^\top$ (see Note **??**). The structure is mosaic-Hankel-like

$$\mathscr{S}\left(\begin{bmatrix} u \\ y \end{bmatrix}\right) = \begin{bmatrix} \mathscr{T}^\top(u) \\ y^\top \end{bmatrix} = \underbrace{\begin{bmatrix} \begin{bmatrix} & & 1 \\ & \iddots & \\ 1 & & \end{bmatrix} & \\ & 1 \end{bmatrix}}_{\Phi} \begin{bmatrix} \mathscr{H}_{n,T-n+1}(u) \\ \mathscr{H}_{1,T-n+1}(y) \end{bmatrix} = \Phi \mathscr{H}_{[n\ 1],T-n+1}\left(\begin{bmatrix} u \\ y \end{bmatrix}\right).$$

Therefore, the maximum likelihood estimator for $\bar{h}$ can be computed with the `slra` function

⟨*call* `slra`⟩≡
```
[uyh, info] = slra([u; y], s, n);
```
using the structure specification

⟨*structure specification for deconvolution*⟩≡
```
s.m = [n 1]; s.phi = blkdiag(fliplr(eye(n)), 1);
```
The estimate $\widehat{h}$ is obtained from the parameter vector $\widehat{R}$ (see (**??**)) by normalization:

$$\widehat{R}' := -\widehat{R}/\widehat{R}_{n+1}, \qquad \widehat{h}(\tau) = \widehat{R}'_{\tau+1}, \quad \text{for } \tau = 0, 1, \dots, n-1.$$

Indeed,

$$\widehat{R}' \begin{bmatrix} \mathscr{T}^\top(\widehat{u}) \\ \widehat{y}^\top \end{bmatrix} = 0 \qquad \Longleftrightarrow \qquad \mathscr{T}(\widehat{u})\widehat{h} = \widehat{y}.$$

⟨$\widehat{R} \mapsto \widehat{h}$⟩≡
```
hh = - info.Rh(1:n)' / info.Rh(n + 1);
```

Note 1 (Identification of a finite impulse response system). In system theory and signal processing, (CONV) defines a finite impulse response linear time-invariant dynamical system. The sequence h is a parameter, u is the input, and y is the output of the system. The deconvolution problem is therefore a system identification problem: estimate the true data generating system from noisy data.

Numerical example

We illustrate empirically the consistency of the structured total least squares estimator. A true parameter vector $\bar{h}$ and a true input sequence $\bar{u}$ are randomly generated. The corresponding true output $\bar{y}$ is computed by convolution of $\bar{h}$ and $\bar{u}$. Zero mean independent and normally distributed noise is added to the true data according to the errors-in-variables model (EIV) and the `slra` function is evoked for the computation of the estimate.

The experiment is repeated $K = 500$ times with independent noise realizations (but fixed true values). Let $\widehat{h}^{(i)}$ be the total least squares solution obtained in the ith repetition. Figure 1, left, shows the root-mean-square error

$$e := \sqrt{\frac{1}{K}\sum_{i=1}^{K} (\|h - \widehat{h}^{(i)}\|_2^2}$$

as a function of the sample size T. Theoretically, the maximum likelihood estimation error converges to zero at a rate that is proportional to the inverse square root of the sample size ($1/\sqrt{T}$ convergence). The simulation results confirm the theoretical convergence rate.

Figure 1, right, shows the true parameter value $\bar{h}$ (red cross), the 500 estimates $\widehat{h}^{(i)}$ (blue dots), and the 95% confidence ellipsoid, computed from the covariance matrix `info.Vh`, corresponding to $\widehat{h}^{(500)}$ and translated to $\bar{h}$. (Note that we do not plot the confidence ellipsoid around each estimate of the parameter in order to simplify the picture.) The fact that about 475 estimates have the true value of the parameter in the confidence region is an empirical confirmation that the confidence regions are correct.

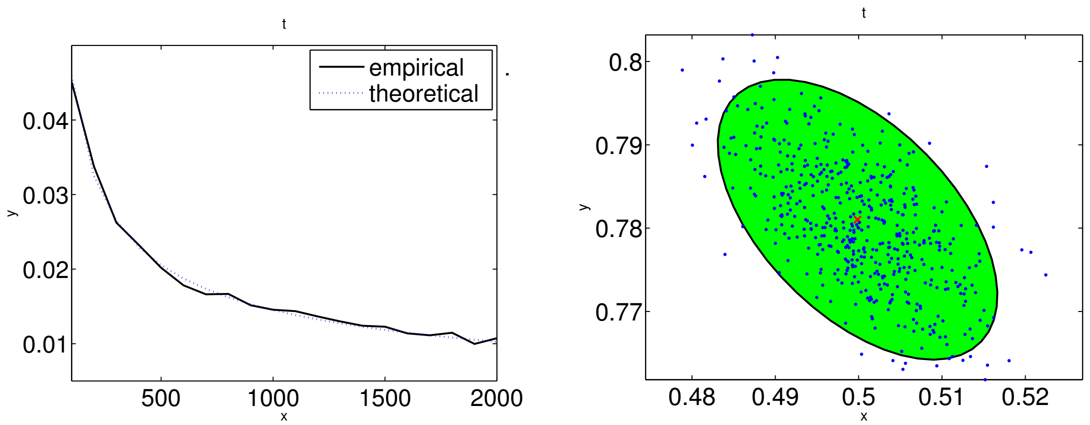

Figure 1: Left: empirical (solid line) and theoretical (dotted line) root-mean-square estimation error e as a function of the sample size T; Right: 95% confidence ellipsoid.

- **"The statistical context $n \gg m$ mentioned by the authors in the abstract is known in the statistical literature as the "large p small n" context, where p stands for the dimension of the data, and n is the sample size.**

In the structured low-rank approximation problem m is related to the dimension of the data and n is related to the sample size. Therefore, the $n \gg m$ context considered in the paper is the opposite to the "large p small n" context mentioned by the referee. Modeling high dimensional data by using a small number of samples is outside the scope of the paper.

- **"Could the authors clarify the definition of the sequence y, since it is defined as a sum of complex exponentials, but the trajectory is in $\mathbb{R}$."**

In the removed from the paper "sum of complex exponentials" example, the complex exponents were chosen to be in complex conjugate pairs, so that their sum is a real valued sequence.

- **"The paragraph "Approximation is needed … is the noise" is incorrect in my view: there exist many statistical problems that can be solved consistently in the presence of noisy data. I would suggest that the authors either develop and clarify this paragraph, or simply delete it."**

In the removed from the paper “sum of complex exponentials” example, the statement about the need of approximation refers to the situation when a finite sample size is used.

- **"typo: p. 2, before (S), “structures of the form” … p. 11 “white noise”."**

Thank you. The typos have been corrected.

- **suggested additional references**

Thank you for the reference. Indeed, it is relevant for our work and we have added it in the reference list.

References

[KMV05] A. Kukush, I. Markovsky, and S. Van Huffel. Consistency of the structured total least squares estimator in a multivariate errors-in-variables model. *J. Statist. Plann. Inference*, 133(2):315–358, 2005.

[Mar08] I. Markovsky. Structured low-rank approximation and its applications. *Automatica*, 44(4):891–909, 2008.