Authors' response to the referee's report

Paper title: Block-Toeplitz/Hankel Structured Total Least Squares
Authors: I. Markovsky, S. Van Huffel, and R. Pintelon

26 April 2004

1 Response to the main criticism

First we answer to the main criticism of the referee. We quote in **bold face** a statement from the report and give our replay next.

. . . the main problem is that it is difficult to understand from the presentation what the significant structural properties of the problem are.

The significant structural property is (quoting from page 2 of the paper):

> The data matrix C can be partitioned into blocks $C = \begin{bmatrix} C^{(1)} & \cdots & C^{(q)} \end{bmatrix}$, where each of the blocks $C^{(l)}$, for $l = 1, \dots, q$, is Hankel, Toeplitz, unstructured, or noise free.

This is our main assumption. It is emphasized in the abstract (see the first sentence) and in the introduction (see the 3rd and 4th paragraphs on page 2). Also it is stated formally in assumption (2.6). In our view the presentation do make it clear for the reader.

. . . (2.3) is a statement that involves the function $\mathcal{S}$ (via (2.2)), . . .

Assumption (2.3) does not involve the *function* $\mathcal{S}$! ($\mathcal{S}$ is defined by the matrices $\{S_i\}_{i=1}^{n_p}$ and is completely independent from assumption (2.3).) In other words, (2.3) does not have anything to do with the structure of the data matrix C. We agree that it is a confusing assumption for those who are not familiar with the stochastic setting of the problem. In the revised version of the paper, we removed it and gave a completely deterministic proof of the main result. See Section 2 of this document for more explanations.

. . . but it is said only very loosely how (2.6) is related to that structure.

We do not agree that this is a main problem of the paper; it could be a misunderstanding of the notation. Obviously (2.6) is a special case of (2.2), so that whenever one assumes (2.6), one implicitly assumes also (2.2). Maybe the link is better displayed by writing

$$\mathcal{S}(p) = \begin{bmatrix} C^{(1)} & \cdots & C^{(q)} \end{bmatrix}, \text{ where } C^{(l)},\ l = 1, \dots, q, \text{ is } \begin{cases} \texttt{T} & \text{block-Toeplitz,} \\ \texttt{H} & \text{block-Hankel,} \\ \texttt{U} & \text{unstructured, or} \\ \texttt{F} & \text{noise free} \end{cases}, \qquad \text{for all } p \in \mathbb{R}^{n_p}?$$

In the revised version of the manuscript, we added an example (see Example 1) with a block-Toeplitz data matrix that shows the matrices S_i for this structure. It is not difficult (but laborious) to write down explicitly the S_i's for the general case of a structured matrix satisfying assumption (2.6). Such an explicit specification is not needed for our purposes. It is enough to note that under assumption (2.6), there exist such S_i's, so they can be used in the derivations.

. . . it is said that (2.2) is replaced by (2.6). It does not make sense.

"replaced" wrongly suggests that (2.2) is no longer active as an assumption. This is wrong because (2.6) *implies* (2.2). We changed "replace" with "*in addition to*", which should avoid the confusion.

. . . the most prominent structure of Γ is the Toeplitz structure.

This statement depends on what "prominent" means. From the point of view of the paper, *i.e.*, computational complexity, more important is the banded structure because it is responsible for the $O(m)$ computational complexity. The Toeplitz structure alone implies computational complexity $O(m^2)$.

Why does that structure appear?

Formally the answer is in the proof of the main result. Informally, using probabilistic terminology, the Toeplitz structure is a consequence of the *stationarity* of the problem, and the bandedness is a consequence of the *independence* of the elements in the data matrix that are apart from each other at least as many rows as is the bandwidth. Depending on the structure of C, we can have both properties, one of the properties but not the other, or none of them. Under the given assumptions both properties are present. In the conclusion of the paper, we make a comment that relaxing the stationarity assumption but keeping the independence, the computational complexity is still $O(m)$.

In the revised version of the manuscript, we modified the proof and better clarified how the structure appears in the weight matrix.

Does it have anything to do with the Toeplitz/Hankel structure? I guess not . . .

The Toeplitz/Hankel structure ensures the independence property mentioned above. If in addition, the cost function is $\|\Delta p\|_2^2$, as we assume, then the problem is stationary. The referee is not right: assumption (2.6) is essential for the result of the paper.

A completely different question is whether other structures result in the same properties of Γ. The answer is positive, but the Toeplitz and Hankel structures are the most important ones for the applications that we have in mind (dynamic systems), so in our view it is a fairly minor restriction to stick to the Toeplitz/Hankel case.

Where does the Toeplitz/Hankel structure show up?

Assumption (2.6) is referred to in several places in the proof of the main result. These are the places where "the Toeplitz/Hankel structure show up".

. . . it must be possible to state it [the result] and prove it [the result] in such a way that the structure(s) in the problem are not hidden but rather explained.

We believe that the revised version of the paper better meets this goal.

It may happen that this paper would be considerably easier to read, if one had read [14], but the paper should be essentially understandable on its own.

We followed the suggestion and extended Section 2 with the missing details (referred to from [14]). Also an appendix is added that shows the derivation of f_0'. Now the paper does not rely on results from [14].

2 Explanation on how the paper is revised

In revising the manuscript according to the criticism of the referee (see Section 1), we concluded that assumption (2.3) is not natural in the proof of the main result (the structure of the weight matrix Γ). It rather artificially brings stochasticity in the problem and makes the arguments probabilistic: independence of random variables, covariance matrices, *etc.* The real assumption that in a sense corresponds to (2.3) is that the cost function of the optimization problem is $\|\Delta p\|^2$. The link is fixed by the maximum likelihood estimation principle that relates the cost function of the optimization problem to the covariance matrix of the measurement error $\tilde{p}$.

Our work has its origin in the field of stochastic estimation, where (2.3) is a central assumption and the tool is statistics, so we adopted them in the present context. The question about the structure of Γ, however, is completely deterministic, so now we find that it is more appropriate to prove it without involving stochastic reasoning. This is what we did in the revised version of the manuscript. Assumption (2.3) is avoided. The new proof is longer but our hope is that overall the presentation is improved.

Although the stochastic framework and tools are not essential in proving the main result, they do give interesting interpretation and further insight. We find it useful to explain this interpretation. A new section is added in the revised version of the manuscript that gives the statistical meaning of the weight matrix Γ and its structural properties.

3 Response to the "suggestions and remarks"

Next we answer to the suggestions and remarks of the referee.

1. corrected
2. corrected
3. the outline is deleted
4. corrected
5. We do not understand the question. The sentence that needs correction is deleted.
6. We clarified what is meant by "repeated block". The presentation on page 3 is extended in order to be self contained and avoid the reference to [14].
7. We agree and deleted Section 3.
8. This comment is related to the main criticism that we have answered above. Now the text of Sections 2 and 4 is rewritten taking into account the comments.

9. We agree and changed as suggested.

10. The variance σ^2 of the measurement errors $\tilde{p}_i$ scales the cost function $f_0 = r^\top \Gamma^{-1} r$ because, as defined, Γ does not depend on σ^2 and r is proportional to σ. Scaling of f_0 does not change the minimum point. Therefore the result, *i.e.*, the minimum point $\hat{X}$, does not depend on our explicit knowledge of σ^2. Moreover, the value of the cost function $f_0(\hat{X})$ at the minimum point is an estimate of σ^2. The W matrices are the corresponding covariance matrices V for $\sigma = 1$. They are just convenient notation for V/σ^2.

11. We agree and deleted the comparison.

12. According to the philosophy of the journal, each author is encouraged to link matrix analysis with applications, therefore we think that this application section can not be removed. The examples are real-life application of the STLS problem. They point where the theory and the derived algorithms are useful. The structures are stated quite explicitly and a reference is given. We quote from Section 6.1

    > ...can be formulated as an STLS problem (with a data matrix composed of two Hankel or Toeplitz structured blocks next to each other, see [1, Sec. 4.6]).

    and from Section 6.2

    > ...can be formulated as an STLS problem (with a block Hankel structured data matrix).

    Giving details on how these applications are related to the STLS problem and therefore how the structures occur, however, is outside the scope of the present paper. We gave the additional reference [2] where the reader can find the details.

References

[1] B. De Moor. Structured total least squares and L_2 approximation problems. *Linear Algebra Appl.*, 188–189:163–207, 1993.

[2] I. Markovsky, J. C. Willems, S. Van Huffel, B. De Moor, and R. Pintelon. Application of structured total least squares for system identification and model reduction. *IEEE Trans. Automat. Control*, 50(10):1490–1500, 2005.