Author’s response to the referees’ reports on "Recent progress in structured low-rank approximation"

Ivan Markovsky

I thank the referees and the associate editor for their relevant and useful comments. In this document, I quote in **bold face** comments/questions from the reports. My replies follow in ordinary prints. In blue, I quote passages from the revised manuscript.

Reviewer #2

- **Section 7 should be moved to a separate Annex 1.**

  Done

- **the detailed description of the Matlab source code should be omitted.**

  Done

- **p.2, 2nd to last line: missing data estimation is in Section 6.3, not 7.1**

  Section 6.3 indeed shows an application of missing data estimation. The sentence under question, however, refers to *methods* for missing data estimation. The methods are discussed in Section 7.1.

- **p.2, (1): state that D is $m \times n$-dimensional (?)**

  The size of the matrix D depends on the application and is stated later in the text for each of the applications reviewed in the paper. The following explanation has been added:

  The matrix D is structured, *i.e.*, it is a function of the data. Its size and structure depend on the example and are specified later on.

- **p.2 to p.5: insert subtitles to mark start of description of examples**

  Done

- **p.6, p.25, p.36: bordered text boxes for important statements seem more appropriate for lectures, rather than for a journal paper.**

  The bordered text boxes have been replaced by text in italics.

- **Section 5, p.10: explain why the title is "variable projection-type."**

  The "-type" is a hint that there is a difference between the conventional variable projection method and the variable projection method used in the solution of the structured low-rank approximation problem. This is explained in the following paragraph:

  The approach, described above, for solving (13) is similar to the variable projection method [GP03] for solution of separable unconstrained non-linear least squares problems

  $$\text{minimize} \quad \text{over } x \text{ and } \theta \quad \|A(\theta)x - b\|_2^2, \tag{1}$$

  where matrix valued function A and the vector y are the given data and the vector x is the to be found parameter. The low-rank approximation problem (13), however, is different from the nonlinear least squares problem (1). Instead of the explicit function $\widehat{b} = A(\theta)x$, where x is unconstrained, an implicit relation $R\mathscr{S}(\widehat{p}) = 0$ is considered, where the variable R is constrained to have full row rank. This fact requires new type of algorithms where the nonlinear least squares problem is an optimization problem on a Grassmann manifold, see [AMS08, AMSD02].

- **p.10, (11): It is not clear why the SLRA can be written in the "equivalent (?)" form (11). Why minimize over R? Without further motivation, this criterion seems arbitrary.**

  The equivalence of the two optimization problems follows from the basic linear algebra fact that

  $$\operatorname{rank}\big(\mathscr{S}(\widehat{p})\big) \leq r \quad \iff \quad \begin{array}{l}\text{there is a full row rank matrix } R \in \mathbb{R}^{(m-r)\times m},\\ \text{such that } R\mathscr{S}(\widehat{p}) = 0.\end{array} \tag{rank$_R$}$$

  The constrain of the SLRA problem (which is the left-hand-side in the above equivalence) is replaced by the equation $R\mathscr{S}(\widehat{p}) = 0$ (which is the right-hand-side of the equivalence). The reformulation creates a new optimization variable R which parametrizes the left kernel of $\mathscr{S}(\widehat{p})$.

- **p.10, (14): How does (14) follow from (13)?**

  This step is explain in subsection "Analytical solution of the inner minimization problem" of section "Algorithmic details". As suggested, in the revised version of the paper, this material has been moved to an appendix and a reference has been added on page 10.

- **Please give the precise definition of vec(.)**

  Done:

  Here vec is the column-wise vectorization operator: for an $m \times n$ matrix $A = \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix}$, $\operatorname{vec}(A)$ is the mn-dimensional vector $\begin{bmatrix} a_1^\top & \cdots & a_n^\top \end{bmatrix}^\top$.

- **p.10, (15): what is $\Psi(\alpha)$?**

  The matrix valued function Ψ and the vector y are the given data for the nonlinear least squares problem (1). In the revised version of the paper, we have changed the notation to the more common $Ax = b$ one (see the quoted passage above) and have added the following explanation:

  where matrix valued function A and the vector y are the given data and the vector x is the to be found parameter.

- **p.11, line 33: delete "compulsory" (wrong English term).**

  Done

- **p.11, [Mu12a]: from the website, it is not clear whether and how the package is made available to interested readers.**

  We now give explicit reference to the software and instructions of how the examples can be reproduced:

  The reported numerical results are reproducible in the sense of [BD95] by downloading the `slra` package [MU12] from

  `https://github.com/slra/slra`

  and following the instructions in `/doc/overview-readme.txt`.

- **p.11, (17): explain meaning and selection criteria of matrices Φ and θ.**

  The Ψ matrix is part of the mosaic-Hankel-like structure specification:

  $$\mathscr{S}(p) := \Phi \mathscr{H}_{\mathbf{m},\mathbf{n}}(p), \tag{2}$$

  An example of the selection and usage of Ψ is "Approximate greatest common divisor of N polynomials" (Section 7.6 in the revised version of the paper).

  The θ parameter is not selected by the user; it is computed by the optimization method. In this case, the user choice is the Ψ matrix which defines structure of the kernel parameter~R. An example of selection and usage of Ψ is "Autonomous system identification with fixed poles" (Section 7.2 in the revised version of the paper). The following explanation was added:

  Applications leading to structured low-rank approximation with linearly structured kernel are identification with fixed poles (see Section 7.2) and approximate common divisor computation for three or more polynomials (see Section 7.6).

- **p.12, "in the software package, problem (12) is solved by": what numerical solver is used ?**

  (SLRA_{θ}) is a standard nonlinear least squares problem, for which existing methods, *e.g.*, the Levenberg-Marquardt algorithm [Mar63] implemented the GNU Scientific Library [G$^+$], are used.

- **p.12, (18): why minimization over R? (as in (11))**

  R is the only optimization variable left after the elimination of $\widehat{p}$.

- **p.15, line 40: y considered in the previous section -> y as defined in the previous section**

  Corrected

- **p.15, line 50: y_{val} = y_{v} of p.14 line 9 ? use same notation.**

  Corrected

- **p.16, line 45 - 49: add symbols: noisy trajectory *y**, optimal approx. **y hat**, true trajectory **y overbar**

  A caption has been added to the figure.

- **p.21, Section 7: move to Annex 1.**

  Done

- **p.25, line 8: delete "of the paper". line 57: in the paper -> in this paper**

  Done

Reviewer #3

- **The manuscript deals about an interesting topic. It is well organized and is clearly written.**

  Thank you for the encouraging remark.

- **However, these new results are not relevant enough to justify an article in this journal: the importance and interest of the new applications of Section 6 are not justified (they seem to be straightforward applications of the general method).**

  This comment raises two independent questions:

1. Are the applications considered in Section 6 of interest for the signal processing community?
2. Are the methods described in the paper appropriate for solving these applications?

The reviewer suggests that the answers to both questions is negative because 1) "importance and interest of the new applications are not justified" and 2) "they seem to be straightforward applications of the general method". I disagree on both counts:

1. The sum-of-damped-exponential modeling problem (Sections 7.1–3) is *the* prototypical signal processing application. System identification (Sections 7.2) is a generalization of the sum-of-damped-exponential modeling problem to linear time-invariant system with inputs. Data-driven simulation is a non orthodox approach for model-free signal processing. The approximate greatest common divisor (Section 7.6) appears in signal processing applications as shown in [ASÅ04].
2. There are robust and efficient methods for structured low-rank approximation. Therefore, it is useful to convert an application in any field, in particular signal processing, to this problem and thus take advantage of the existing methods. In addition, the paper shows connections among applications that may not be obvious for researchers working in the corresponding sub-fields.

- **missing data estimation has been reported in a previous article [MU12b], the fast algorithm of Section 7.2 has already been published in [UM12b], and the software package is interesting but it does not justify by itself this publication.**

  The results mentioned above are indeed not original and actually there is no claim for originality. The paper is an overview and not a research article. This is clearly stated in the abstract

  This paper gives an overview of recent progress in efficient local optimization algorithms for solving weighted mosaic-Hankel structured low-rank approximation problems.

  and the introduction

  This paper is an update of [Mar08] and [MV07] with new applications, methods for missing data estimation, fast approximation of mosaic-Hankel-like matrices, and availability of robust and efficient methods implementing the theory.

References

[AMS08] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, Princeton, NJ, 2008.

[AMSD02] P.-A. Absil, R. Mahony, R. Sepulchre, and P. Van Dooren. A Grassmann–Rayleigh quotient iteration for computing invariant subspaces. *SIAM Review*, 44(1):57–73, 2002.

[ASÅ04] M. Agarwal, P. Stoica, and T. P Åhgren. Common factor estimation and two applications in signal processing. *Signal Processing*, 84:421–429, 2004.

[BD95] J. Buckheit and D. Donoho. *Wavelets and statistics*, chapter "Wavelab and reproducible research". Springer-Verlag, Berlin, New York, 1995.

[G$^+$] M. Galassi et al. GNU scientific library reference manual. `http://www.gnu.org/software/gsl/`.

[GP03] G. Golub and V. Pereyra. Separable nonlinear least squares: the variable projection method and its applications. *Institute of Physics, Inverse Problems*, 19:1–26, 2003.

[Mar63] D. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *SIAM J. Appl. Math.*, 11:431–441, 1963.

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

[MU12] I. Markovsky and K. Usevich. Software for weighted structured low-rank approximation. Technical report, Vrije Univ. Brussel, 2012.

[MV07] I. Markovsky and S. Van Huffel. Overview of total least squares methods. *Signal Proc.*, 87:2283–2302, 2007.