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

Ivan Markovsky

September 20, 2013

I thank the referees and the 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.

Editor's comments

Several reviewers comment on the limited scope of the review article focusing on the authors' own previous work. Please broaden the scope to include a review of other relevant recent work on "low rank approximation".

The following revision steps have been done.

1. In response to reviewer #3, new references about application of low-rank approximation methods to signal processing problems have been added.
2. A Subsection "Other methods" is added to the Section "Related paradigms, problems, and algorithms" in order to state the limitations of the overview.

   A subarea of low-rank approximation that is not covered in this overview is *tensors low-rank approximation* [WVB10, LV00]. Tensor methods are used in higher order statistical signal processing problems, such as independent component analysis, and multidimensional signal processing, such as spatio-temporal modeling and video processing, to name a few. Other areas of research on low-rank approximation that we do not cover are nonnegative low-rank approximation and matrix completion [Mar11]. Nonnegative constraints appear, for example, in chemometrics, image, and text processing [BBL+07]. These constraints are imposed in solution methods by a rank revealing factorization with nonnegative factors. In addition, upper bounds on the elements are imposed in the method of [KIP12].

3. Note that the material in the overview based on my own previous work (Section 6 and the Appendix) comprises 5 out of the 35 pages. The rest of the paper reviews problems of generic interest that are not related to a specific method. In order to make it explicitly clear that the recent results in the paper are about a particular approach (variables projection) for solving the generic problem, the title has been modified as follows:

   Recent progress on variable projection methods for structured low-rank approximation

The comment of reviewer #6 that the topic of the paper is **"a very specific subject where the only relevant results have been developed by the author and coworkers"** has been answered in a rebuttal. Indeed, the topic of low-rank is also closely related to (and can be viewed as a generalization of) established and widely used techniques, such as principal component analysis and total least squares.

Reviewer #1

- **It would have been helpful for the reviewers if authors had also delivered a report with detailed responses to reviewer's comments.**

  A document with detailed point-to-point answers to the reviewers comments and explanation of the revision steps was submitted together with the revised version of the paper. I am surprised that this letter had not been made available to the reviewers.

- **I couldn't find the overview-readme.txt in the doc folder.**

  Thank you for pointing out this omission. The file is now added in `git`.

- **p.2, line 40: Section 7.3, not 7.5**

  The reference is to "data-driven simulation", which is indeed presented in Section 7.5.

- **p. 18, line 38: State that SLRA succeeds in identifying the system, while classical methods are not applicable, since SLRA uses a different estimation criterion than the classical method.**

  The sentence was modified as follows:

  Although the classical methods are not applicable in this case, methods for solving (SLRA) can be used by assigning zero weights in the cost function to the missing data elements.

- **SLRA estimate is not "optimum" in the classical sense.**

  The SLRA cost function generalizes the classical one in order to allow for missing data. In the case when all data are specified, however, the classical and SLRA cost functions coincide. See, [Mar13] where the SLRA approach is compared with classical methods for model reduction and system identification problem. Equivalent results are obtained (modulo convergence to different local minima).

- **p.22, line 10: please explain why 0.1924 should be a a good approximate common root of the 3 given polynomials?**

  The philosophy behind the SLRA problem is:

  > modify the given data as little as possible in order to ensure that an exact model exists for the modified data.

  The root 0.1924 is a good *approximate* common root of the three given polynomials because it is the *exact* common root of the three perturbed polynomials, obtained with a minimal perturbation of the coefficients of the given polynomials. This notion of approximate common factor is standard in the literature, see, *e.g.*, [ASÅ04]. The following explanation has been added in the second revised version of the paper:

  The motivation for this problem formulation is to modify the given polynomial coefficients as little as possible in order ensure that the modified polynomials have a common factor of a desired order. Then, the exact common factor of the modified polynomials is by definition the approximate common factor for the original polynomials.

Reviewer #3

- **There is little discussion about alternative approaches for solving the SLRA with pros and cons of each strategy.**

  A complete (to the best of the author's knowledge) list of approaches for structured low-rank approximation is given on page 12 of the paper:

  The main types of methods, with a few representatives for each type, are:

  - global solution methods (see [UM12, Section 3]),
    * semi-definite programming relaxations of rational function minimization [JdK06],
    * methods based on solving systems of polynomial equations [Ste04];
  - local optimization methods,
    * variable projection [MVP05],
    * alternating projections [WAH+ 97]; and
  - heuristics based on convex relaxations,
    * multistage methods [VD96],
    * nuclear norm heuristic [Faz02].

  The rest of section 5 discusses pros and cons of the approaches. In our experience the most promising approach for batch model identification is local optimization using the variable projection approach, so that we present this approach in more details in Section 6.

- **The connection to the signal processing literature is also weak and needs to be strengthened.**

  Nine new references ([MMH03, Sch91, TS93, BBL$^+$07, AG07, KIP12, WVB10, LV00, BA11]) using low-rank approximation techniques in signal processing have been added in the revised version of the paper.

- **page 8 : unclear relation between solutions to the two problems, please clarify.**

  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 $\mathscr{S}(\widehat{p})$.

- **In the final expression of the page, there should be exact equality ($\widehat{A}X = \widehat{B}$).**

  Thank you. The typo is now corrected.

- **page 24: the references to Jackson and Joliffe are both included twice.**

  Corrected.

- **Please include legends in the plots to improve clarity.**

  Done.

Reviewer #6

1. **The keyword of the subject is "low-rank approximation". Checking the titles of the cited references, the only ones that include these words in the title are the ones coauthored by the author of this article. This may mean that: a) the title "low rank approximation" does not embrace all the subject and techniques described in this article and is inappropriate, or b) this is a very specific subject where the only relevant results have been developed by the author and coworkers (in which case I doubt about the general interest of the scientific community on this matter).**

   As discussed in section "Related paradigms, problems, and algorithms" there are a number of terms in the literature referring to problems that are equivalent or closely related to the concept of low-rank approximation:

   - principal component analysis (PCA),
   - latent semantic analysis (LSA),
   - total least squares (TLS),
   - errors-in-variables (EIV) modeling,
   - behavioral approach,
   - rank minimization.

   These terms are field specific

   - PCA — machine learning,
   - LSA — speech processing,
   - TLS — numerical linear algebra,
   - EIV — statistics,
   - behavioral approach — systems and control theory,
   - rank minimization — linear algebra,

   and have connotations (statistical properties, specific applications, *etc*). I prefer the term "low-rank approximation" because it states exactly what the problem is: "approximation of a matrix by another matrix which has reduced rank", and is not bound to a specific application specific domain. Low-rank approximation emphasizes the generality of the problem which is also the goal of this overview.

   The comment of the reviewer has been addressed in the revised version of the paper by:

   (a) Adding new references [MMH03, Sch91, TS93, BBL$^+$07, AG07, KIP12, WVB10, LV00, BA11] about application of low-rank approximation to signal processing.

   (b) Adding a Subsection "Other methods" to Section "Related paradigms, problems, and algorithms" that states the limitations of the overview:

   A subarea of low-rank approximation that is not covered in this overview is *tensors low-rank approximation* [WVB10, LV00]. Tensor methods are used in higher order statistical signal processing problems, such as independent component analysis, and multidimensional signal processing, such as spatio-temporal modeling and video processing, to name a few. Other areas of research on low-rank approximation that we do not cover are nonnegative low-rank approximation and matrix completion [Mar11]. Nonnegative constraints appear, for example, in chemometrics, image, and text processing [BBL$^+$07]. These constraints are imposed in solution methods by a rank revealing factorization with nonnegative factors. In addition, upper bounds on the elements are imposed in the method of [KIP12].

2. **The part of the title "Recent progress" is also unclear. The author refers to two previous surveys written by himself, published in 2007 and 2008. Then new results should be foreseen in the time interval 2008–2013. There are only 12 cites to works in this period (in the References Section) and 9 of them were written by the author and coworkers. The remaining three are: [AMS08] which does not seem to report any new result on low-rank approximation, [LV09] which is cited as an example of relation between this technique and rank minimization, and [Byd10] which is not cited in the text. Then the only new results have been provided by the author.**

   I agree that the more technical Section 6 and the Appendix are focused on results obtained by collaborators of mine and myself. Note, however, that this material is approximately 5 pages out the total 35 pages. The rest of the presentation is about general problems, which are not dependent on a specific approach, *i.e.*, the rest of the paper can be read skipping the details on the variable projections method presented in Section 6. In order to clarify the "Recent progress" part of the title, the title was modified as follows:

   Recent progress on variable projection methods for structured low-rank approximation

References

[AG07] J. Antoni and L. Garibaldi. A reduced-rank spectral approach to multi-sensor enhancement of large numbers of sinusoids in unknown noise fields. *Signal Processing*, 87(3):453–478, March 2007.

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

[BA11] N. Boumal and P.-A. Absil. RTRMC: A Riemannian trust-region method for low-rank matrix completion. In J. Shawe-Taylor, R.S. Zemel, P. Bartlett, F.C.N. Pereira, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 24 (NIPS)*, pages 406–414. 2011.

[BBL+07] M. Berry, M. Browne, A. Langville, P. Pauca, and R. Plemmons. Algorithms and applications for approximate nonnegative matrix factorization. *Comput. Stat. Data Anal.*, 52(1):155–173, September 2007.

[KIP12] R. Kannan, M. Ishteva, and H. Park. Bounded matrix low rank approximation. In *12th Int. Conf. on Data Mining*, pages 319–328, 2012.

[LV00] B. Lathauwer, L. De Moor and J. Vandewalle. On the best rank-1 and rank-(R_1,R_2,...,R_N) approximation of higher-order tensors. *SIAM J. Matrix Anal. Appl.*, 21(4):1324–1342, 2000.

[Mar11] I. Markovsky. *Algorithms and literate programs for weighted low-rank approximation with missing data*, volume 3 of *Springer Proc. Mathematics*, pages 255–273. Springer, 2011.

[Mar13] I. Markovsky. A software package for system identification in the behavioral setting. *Control Engineering Practice*, 21:1422–1436, 2013.

[MMH03] J. Manton, R. Mahony, and Y. Hua. The geometry of weighted low-rank approximations. *IEEE Trans. Signal Process.*, 51(2):500–514, 2003.

[Sch91] L. Scharf. The SVD and reduced rank signal processing. *Signal Processing*, 25(2):113–133, 1991.

[TS93] D. Tufts and A. Shah. Estimation of a signal waveform from noisy data using low-rank approximation to a data matrix. *IEEE Trans. Signal Proc.*, 41(4):1716–1721, 1993.

[WVB10] S. Weiland and F. Van Belzen. Singular value decompositions and low rank approximations of tensors. *IEEE Trans. Signal Proc.*, 58(3):1171–1182, 2010.