Geosci. Instrum. Method. Data Syst. Discuss., 4, C191–C195, 2014
www.geosci-instrum-method-data-syst-discuss.net/4/C191/2014/

Geoscientific Instrumentation Methods and Data Systems
Discussions

Interactive comment on "COSIMA data analysis using multivariate techniques" *by* J. Silén et al.

P. Paatero (Referee)

Pentti.Paatero@helsinki.fi

Received and published: 10 October 2014

This is an interesting report of experiments where random projections (RP) are applied in order to compress a large matrix without compromising its information content. In particular, such compression is intended as a preliminary step before a PCA or SVD decomposition is computed of the matrix. The RP method is applied on mineralogical data, to be collected by a very advanced space probe. In this review, I only consider the mathematical-statistical aspects of the work.

I believe that the experimental results reported in this work will contribute to the understanding of compression that is achievable with RP. Thus I recommend this manuscript for publication after enhancements discussed below are performed.

The manuscript contains discussion that may not be justified or that at least must be

carefully qualified so that unfounded claims are not made.

Before going into these details, I summarize some aspects of how data may be linearly transformed in preparation of a SVD computation (or PCA) whose purpose is to separate random noise and non-random "signal" contained in the matrix. The given matrix is denoted by X. Estimates of uncertainty in elements of X is assumed to be available, and different elements of X may possess different estimates of uncertainty.

In the following, operations on columns are mostly discussed. It appears that rows might be handled similarly as columns although columns are discussed in literature. Three transformations for enhancing the S/N separating capability of SVD/PCA are the following:

First transformation is omission of such columns that contain no or very little signal in comparison to the noise contained in those columns. Paatero and Hopke (2003) demonstrated that signal/noise separation power of a SVD is enhanced if a significant number of high-noise columns are omitted from the matrix. This simple transformation is of the form

(1)..... Y = X D

where a number of elements of the diagonal matrix D are zero.

Another transformation is of the form

(2)..... Y = X Q D

where Q is an orthogonal matrix and D is a diagonal matrix with a number of zero-valued diagonal elements. Q is formulated so that some columns of product (X Q) will contain mostly noise, and the zero elements of D will omit these high-noise columns from Y. This kind of transformation may significantly enhance the S/N separation capability of SVD, while simultaneously compressing X into Y. Note that a successful Q may not always exist or may not be known even if its existence is theoretically predicted. The formulation of a Q is efficient if correlations of columns of X contain a structure, this is the case with the original matrix of spectra. In contrast, a matrix produced by the RP approach does not possess such structure, thus a successful Q may not be easy to find for the RP - projected matrix.

A third transformation is

(3)..... $Y = D_1 X D_2$

where diagonal matrices D_1 and D_2 are defined so that they minimize the variation (row-to-row variation and column-to-column variation) of noise amplitude in Y. Paatero and Tapper (1993) discussed these transformations. The S/N separation power of SVD is enhanced by this transformation if noise amplitude variation in Y is less than noise amplitude variation in X. The ms should discuss the fact that optimality of SVD (and of PCA) depends on the variation of noise amplitude in X. Instead of scaling X so that all columns are normalized to same norm, matrix D_2 should be formulated so that noise amplitude is equal in all columns of the scaled matrix Y. This follows because SVD and PCA are bilinear (not linear) least squares processes and they are almost optimal (for S/N separation) if all elements of the matrix have the same uncertainty.

The manuscipt claims that the random projection method is "optimal". These claims have been taken from one or a few references listed in the ms. These optimality claims are not well specified, not in the references nor in this manuscript: it is not clear what classes of methods are considered and how is the "goodness" defined. Either, the optimality should be carefully defined and proved, or (recommended), optimality claims should be removed from the ms. In particular, if optimality is claimed, then also the transformations discussed in preceding paragraphs of this report should be considered: is RP method as good as or better than best possible transformations (1), (2), and (3) when attempting a best possible S/N separation by SVD/PCA of the compressed matrix.

This is a mathematical work. Equations are the language of mathematics. However, there are verbal explanations describing what was done in this work. These explana-

tions are hard to understand, different readers may understand them differently. The description of what was done should be rewritten so that all operations are described by equations, possibly accompanied by verbal support. Also, it must be stated what operations were performed many times, iteratively, in the Monte-Carlo sense, and what operations were performed only once. E.g. is the RP projection computed only once, with one coefficient matrix, or is it computed many times so that the best outcome is then selected for further processing.

Details that should be corrected:

The ms uses the word "robust" in a colloquial manner, as a synonym of "reliable", say. On the other hand, "robust" is a well-defined statistical concept. In order to avoid misunderstandings, "robust" may only be used in the statistical sense. The computation of SVD (or PCA) for S/N separation is essentially a least squares fit. Hence, SVD/PCA are extremely non-robust, as all methods based on a LS fit are known to be. Thus, in this ms, robust may only be used in a sentence saying that "SVD/PCA are non-robust methods". Such words as "reliable" or "stable" might perhaps describe the property that the authors have empirically determined.

SVD and PCA must not be called linear. They are bilinear, and their peculiar properties derive from their bilinear nature. In the SVD decomposition $X = U S V'$, columns of U are linear combinations of columns of X. However, this does not mean that SVD, as a concept, would be linear.

Statistically, optimality of an estimator often depends on the shapes of distributions of random variables (errors, say). Example: the arithemtic mean is optimal estimate of location only if errors are normally distributed with equal variance. If distributions contain more tail, then e.g. median may be better than arithmetic mean. If optimality is claimed, then I would assume that also the assumed distributions of errors should be specified.

References for this review:

P. Paatero and U. Tapper, Analysis of Different Modes of Factor Analysis as Least Squares Fit Problems, Chemometrics and Intelligent Laboratory Systems 18(2)183-194 (1993)

Pentti Paatero and Philip K. Hopke, Discarding or downweighting high-noise variables in factor analytic models. Analytica Chimica Acta 490 (2003) 277-289.

Interactive comment on Geosci. Instrum. Method. Data Syst. Discuss., 4, 455, 2014.