LLNL-TR-409155

LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

Covariance Modifications to Subspace Bases

D. B. Harris

December 3, 2008

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

Covariance Modifications to Subspace Bases

D. B. Harris

Introduction

Adaptive signal processing algorithms that rely upon representations of signal and noise subspaces often require updates to those representations when new data become available. Subspace representations frequently are estimated from available data with singular value (SVD) decompositions. Subspace updates require modifications to these decompositions. Updates can be performed inexpensively provided they are low-rank.

A substantial literature on SVD updates exists, frequently focusing on rank-1 updates (see e.g. [Karasalo, 1986; Comon and Golub, 1990, Badeau, 2004]). In these methods, data matrices are modified by addition or deletion of a row or column, or data covariance matrices are modified by addition of the outer product of a new vector. A recent paper by Brand [2006] provides a general and efficient method for arbitrary rank updates to an SVD. The purpose of this note is to describe a closely-related method for applications where right singular vectors are not required. This note also describes the SVD updates to a particular scenario of interest in seismic array signal processing.

The particular application involve updating the wideband subspace representation used in seismic subspace detectors [Harris, 2006]. These subspace detectors generalize waveform correlation algorithms to detect signals that lie in a subspace of waveforms of dimension $d \geq 1$. They potentially are of interest because they extend the range of waveform variation over which these sensitive detectors apply. Subspace detectors operate by projecting waveform data from a detection window into a subspace specified by a collection of orthonormal waveform basis vectors (referred to as the template). Subspace templates are constructed from a suite of normalized, aligned master event waveforms that may be acquired by a single sensor, a three-component sensor, an array of such sensors or a sensor network. The template design process entails constructing a data matrix whose columns contain the master event waveform data, then performing a singular value decomposition on the data matrix to extract an orthonormal basis for the waveform suite. The template typically is comprised of a subset of the left singular vectors corresponding to the larger singular values.

The application involves updating a subspace template when new data become available, i.e. when new defining events are detected for a particular source. It often is the case that the waveforms corresponding to a particular source drift over time [Harris, 2001]. The Green's functions describing propagation can be altered because of changes in the source region. For example, if the source is a mine, signals from explosions may change gradually as a pit is extended (the source moves) or the scattering topography is altered by excavation. This motivates a tracking adjustment to a subspace representation.

This note also comments on SVD updates for a related problem. In realistic pipeline operations it often is the case that data from one or more channels of an array are unusable (dead channels, channels with prolific dropouts, etc.). In such cases it is desirable to modify an array subspace template to operate on data only from the remaining usable channels. Furthermore, it is desirable to modify the templates directly without recourse to the original data matrix. Usually the template design process is separate from the application of the template in a detector to a continuous data stream. Consequently, the original data matrix may not be available for template modification at detector run time.

Subspace Update for an Augmented Data Matrix

In this section we assume that a desired subspace is estimated from an $N \times M$ data matrix

$$\underline{X} = \lfloor \underline{x}_1 \quad \underline{x}_2 \quad \cdots \quad \underline{x}_M \rfloor \tag{1}$$

through a singular value decomposition:

$$\underline{X} = \underline{U}\underline{\Sigma}\underline{V}^T . \tag{2}$$

The data matrix is comprised of M column vectors; typically, each column holds signal snapshot data of some type. We assume further that only the matrices $\underline{U}$ of left singular vectors and $\underline{\Sigma}$ of singular values are of interest. In the "thin" SVD of $\underline{X}$, $\underline{U}$ is an $N \times d$ matrix $\underline{\Sigma}$ is a $d \times d$ matrix with $d \le M$ and $d = M$ only if $\underline{X}$ is full rank.

The question we pose is how to modify this representation when the data matrix is augmented by an $N \times r$ matrix $\underline{Y}$ of r new column vectors.

$$\underline{\widetilde{X}} = \left[\underline{X} \,|\, \underline{Y} \right] \tag{3}$$

We desire updates $\underline{\widetilde{U}}$ and $\underline{\widetilde{\Sigma}}$ to $\underline{U}$ and $\underline{\Sigma}$ that correspond to the new SVD

$$\underline{\widetilde{X}} = \underline{\widetilde{U}}\underline{\widetilde{\Sigma}}\underline{\widetilde{V}}^T . \tag{4}$$

However, we want to compute these updates without constructing the augmented data matrix. It may be the case that the original data matrix $\underline{X}$ is unavailable. As we shall see it is possible to update the subspace with knowledge only of $\underline{U}$, $\underline{\Sigma}$ and $\underline{Y}$.

Consider the related outer product ("covariance") matrix:

$$\underline{\Psi} = \underline{X}\,\underline{X}^T \tag{5}$$

As is well known, this symmetric positive (possibly semi-) definite matrix $\underline{\Psi}$ has an eigendecomposition in terms of the left singular vectors and singular values of $\underline{X}$:

$$\underline{\Psi} = \underline{U}\underline{\Sigma}^2\underline{U}^T \tag{6}$$

This identity is easy to derive by substituting the definition of equation (2) into (5) and noting that $\underline{V}$ is orthonormal.

One route to updating $\underline{U}$ and $\underline{\Sigma}$ is to update the eigendecomposition of $\underline{\Psi}$ instead of the SVD of $\underline{X}$. The augmented outer product matrix $\tilde{\underline{\Psi}}$ is:

$$\tilde{\underline{\Psi}} = \tilde{\underline{X}}\tilde{\underline{X}}^T = \underline{X}\,\underline{X}^T + \underline{Y}\,\underline{Y}^T = \underline{U}\underline{\Sigma}^2\underline{U}^T + \underline{Y}\,\underline{Y}^T \tag{7}$$

for which we seek the eigendecomposition:

$$\tilde{\underline{\Psi}} = \tilde{\underline{U}}\tilde{\underline{\Sigma}}^2\tilde{\underline{U}}^T \tag{8}$$

The updated $\tilde{\underline{U}}$ and $\tilde{\underline{\Sigma}}$ of equation (8) are the same quantities that appear in equation (4).

Following Brand [1], we define an orthonormal matrix $\underline{C}$ that spans the innovation that $\underline{Y}$ represents with respect to $\underline{X}$, i.e. $\underline{C}$ spans the projection of $\underline{Y}$ into the nullspace of $\underline{X}$:

$$\underline{C}\,\underline{C}^T\left(\underline{I} - \underline{U}\,\underline{U}^T\right)\underline{Y} = \left(\underline{I} - \underline{U}\,\underline{U}^T\right)\underline{Y} \tag{9}$$

$\underline{C}$ has the following properties:

$$\underline{C}^T\underline{U} = \underline{0}$$

$$\underline{C}^T\underline{C} = \underline{I} \tag{10}$$

$$\underline{Y} = \underline{U}\,\underline{U}^T\underline{Y} + \underline{C}\,\underline{C}^T\underline{Y}$$

$\underline{C}$ can be computed with a "thin" QR decomposition of the $N \times r$ matrix $\left(\underline{I} - \underline{U}\,\underline{U}^T\right)\underline{Y}$. These operations can be inexpensive compared to the original computation of the SVD of $\underline{X}$ if $r << M$.

With $\underline{C}$ defined, we can write the augmented $\tilde{\underline{\Psi}}$ as:

$$\tilde{\underline{\Psi}} = \underline{U}\underline{\Sigma}^2\underline{U}^T + \left(\underline{U}\,\underline{U}^T + \underline{C}\,\underline{C}^T\right)\underline{Y}\,\underline{Y}^T\left(\underline{U}\,\underline{U}^T + \underline{C}\,\underline{C}^T\right) \tag{11}$$

A series of manipulations produce a simple expression for $\tilde{\underline{\Psi}}$. First:

$$\tilde{\underline{\Psi}} = \underline{U}\underline{\Sigma}^2\underline{U}^T + \left[\underline{U} \quad \underline{C}\right]\begin{bmatrix}\underline{U}^T\underline{Y} \\ \underline{C}^T\underline{Y}\end{bmatrix}\left[\underline{Y}^T\underline{U} \quad \underline{Y}^T\underline{C}\right]\begin{bmatrix}\underline{U}^T \\ \underline{C}^T\end{bmatrix} \tag{12}$$

Defining $\underline{F} = \begin{bmatrix} \underline{U}^T \underline{Y} \\ \underline{C}^T \underline{Y} \end{bmatrix}$, we obtain:

$$\underline{\tilde{\Psi}} = \begin{bmatrix} \underline{U} & \underline{C} \end{bmatrix} \left(\begin{bmatrix} \underline{\Sigma}^2 & \underline{0} \\ \underline{0} & \underline{0} \end{bmatrix} + \underline{F}\,\underline{F}^T \right) \begin{bmatrix} \underline{U}^T \\ \underline{C}^T \end{bmatrix} \tag{13}$$

Note that $\begin{bmatrix} \underline{U} & \underline{C} \end{bmatrix}$ is an orthonormal matrix (because $\underline{U}$ and $\underline{C}$ are individually orthonormal and orthogonal to each other).

In a manner similar to Brand's SVD update, we now are in a position to compute the update to $\underline{\tilde{\Psi}}$ through a rotation of the $\begin{bmatrix} \underline{U} & \underline{C} \end{bmatrix}$ matrix. Defining the eigendecomposition

$$\begin{bmatrix} \underline{\Sigma}^2 & \underline{0} \\ \underline{0} & \underline{0} \end{bmatrix} + \underline{F}\,\underline{F}^T = \underline{G}\,\underline{\Gamma}\,\underline{G}^T \tag{14}$$

the updated decomposition (8) can be found by substituting (14) into (13):

$$\underline{\tilde{\Psi}} = \begin{bmatrix} \underline{U} & \underline{C} \end{bmatrix} \left(\underline{G}\,\underline{\Gamma}\,\underline{G}^T \right) \begin{bmatrix} \underline{U}^T \\ \underline{C}^T \end{bmatrix} \tag{15}$$

We obtain the updated matrices $\underline{\tilde{U}}$ and $\underline{\tilde{\Sigma}}$ by inspection:

$$\underline{\tilde{U}} = \begin{bmatrix} \underline{U} & \underline{C} \end{bmatrix} \underline{G} \qquad \underline{\tilde{\Sigma}} = \underline{\Gamma}^{1/2} \tag{16}$$

This algorithm may not have numerical properties (e.g. sensitivity to rounding error) as favorable as those of more expensive alternatives which do not square the singular values. However, this is not likely to be a major problem in the seismic application since updates are performed infrequently (only when a new event is detected and not every time a new datum is available, as in continuous subspace tracking algorithms). In addition, low-rank approximations to the data matrix are likely to be used, limiting the dynamic range of the singular values to that of only the largest singular values.

Exponential age weighting

It may be desirable to alter the algorithm to employ weights that cause the updated subspace progressively to "forget" the contributions defined by previous events. For this purpose, the augmented matrix $\underline{\tilde{X}}$ is modified by:

$$\underline{\tilde{X}} = \begin{bmatrix} \lambda \underline{X} \mid \underline{Y} \end{bmatrix} \tag{17}$$

where λ is some positive factor slightly less than one. In this case, $\underline{\Sigma}^2$ in equations (13) and (14) is replaced by $\lambda^2\underline{\Sigma}^2$. The method is otherwise the same.

Subspace Correction for Deleted Data

We now consider the problem of data deletion for a "bad" channel. The data matrix for design of templates in a subspace detector has the form:

$$\underline{X} = \begin{bmatrix} \underline{x}_{11} & \underline{x}_{12} & \cdots & \underline{x}_{1M} \\ \underline{x}_{21} & \underline{x}_{22} & \cdots & \underline{x}_{2M} \\ \vdots & \vdots & \cdots & \vdots \\ \underline{x}_{i1} & \underline{x}_{i2} & \cdots & \underline{x}_{iM} \\ \vdots & \vdots & \cdots & \vdots \\ \underline{x}_{N_c1} & \underline{x}_{N_c2} & \cdots & \underline{x}_{N_cM} \end{bmatrix} = \underline{U}\underline{\Sigma}\underline{V}^T \tag{18}$$

Typically the data are sampled waveforms from an array of N_c sensors (channels). The column vectors are comprised of trace-sequential compositions of N_c signals with N_T time samples each. Hence, the total number of rows in the matrix is $N = N_c \cdot N_T$. The individual constituent vectors in (17) are:

$$\underline{x}_{ij} = \begin{bmatrix} x_{ij}(0\cdot\Delta t) \\ x_{ij}(1\cdot\Delta t) \\ \vdots \\ x_{ij}((N_T-1)\cdot\Delta t) \end{bmatrix} \tag{19}$$

where the function $x_{ij}(t)$ is the signal observed by the i^{th} channel of the array for the j^{th} event. At issue is what to do when data from channel *i* have to be removed.

In this case, the data matrix is modified by zeroing the elements corresponding to the i^{th} channel:

$$\hat{\underline{X}} = \begin{bmatrix} \underline{x}_{11} & \underline{x}_{12} & \cdots & \underline{x}_{1M} \\ \underline{x}_{21} & \underline{x}_{22} & \cdots & \underline{x}_{1M} \\ \vdots & \vdots & \cdots & \vdots \\ \underline{0} & \underline{0} & \cdots & \underline{0} \\ \vdots & \vdots & \cdots & \vdots \\ \underline{x}_{N1} & \underline{x}_{N2} & \cdots & \underline{x}_{1M} \end{bmatrix} = \begin{bmatrix} \underline{x}_{11} & \underline{x}_{12} & \cdots & \underline{x}_{1M} \\ \underline{x}_{21} & \underline{x}_{22} & \cdots & \underline{x}_{1M} \\ \vdots & \vdots & \cdots & \vdots \\ \underline{x}_{i1} & \underline{x}_{i2} & \cdots & \underline{x}_{iM} \\ \vdots & \vdots & \cdots & \vdots \\ \underline{x}_{N1} & \underline{x}_{N2} & \cdots & \underline{x}_{1M} \end{bmatrix} + \begin{bmatrix} \underline{0} & \underline{0} & \cdots & \underline{0} \\ \underline{0} & \underline{0} & \cdots & \underline{0} \\ \vdots & \vdots & \cdots & \vdots \\ -\underline{x}_{i1} & -\underline{x}_{i2} & \cdots & -\underline{x}_{iM} \\ \vdots & \vdots & \cdots & \vdots \\ \underline{0} & \underline{0} & \cdots & \underline{0} \end{bmatrix} \tag{20a}$$

$$= \underline{X} + \underline{E}_i\left(-\underline{E}_i^T\underline{X}\right) = \left(\underline{I} - \underline{E}_i\underline{E}_i^T\right)\underline{X}$$

where:

$$\underline{E}_i = \begin{bmatrix} \underline{0} \\ \underline{0} \\ \vdots \\ \underline{I}_{N_T \times N_T} \\ \vdots \\ \underline{0} \end{bmatrix} \tag{20b}$$

Note the identity matrix in $\underline{E}_i$ is in the position corresponding to the i^{th} data channel. .

We seek left singular vectors and singular values of the SVD $\underline{\hat{X}} = \underline{\hat{U}}\underline{\hat{\Sigma}}\underline{\hat{V}}^T$ knowing only $\underline{U}$ and $\underline{\Sigma}$. We can do so by developing the eigendecomposition of:

$$\underline{\hat{\Psi}} = \underline{\hat{X}}\underline{\hat{X}}^T = \left(\underline{I} - \underline{E}_i\underline{E}_i^T\right)\underline{X}\underline{X}^T\left(\underline{I} - \underline{E}_i\underline{E}_i^T\right) = \left(\underline{I} - \underline{E}_i\underline{E}_i^T\right)\underline{U}\underline{\Sigma}^2\underline{U}^T\left(\underline{I} - \underline{E}_i\underline{E}_i^T\right) \tag{21}$$

From (20) it is apparent that we can compute the eigendecomposition of $\underline{\hat{\Psi}}$ by computing the SVD of $\left(\underline{I} - \underline{E}_i\underline{E}_i^T\right)\underline{U}\underline{\Sigma}$. The implied algorithm has three steps: (1) compute the product matrix $\underline{U}\underline{\Sigma}$, (2) modify the matrix by setting to zero all elements of the rows corresponding to the dropped channel(s), and (3) compute the SVD of the modified matrix. $\underline{\hat{U}}$ and $\underline{\hat{\Sigma}}$ are obtained as the factors of this SVD.

This approach has the same computational expense as the original SVD of the data matrix, but does not require knowledge of the original data matrix. A low-rank modification of the matrix (similar to the update of the previous section) is possible when a few rows are set to zero. However, the modification contemplated here results in the elimination of hundreds or thousands of rows, so that a recomputation of the SVD is most economical.

Approximate Matrix Representations

If the subspace being modified has lower rank than the original data matrix, the modification does not correct the SVD factors of $\underline{\hat{X}} = \left(\underline{I} - \underline{E}_i\underline{E}_i^T\right)\underline{X}$, but rather the factors of $\underline{\hat{X}}_d = \left(\underline{I} - \underline{E}_i\underline{E}_i^T\right)\underline{X}_d$, where

$$\underline{X}_d = \underline{U}_d\underline{\Sigma}_d\underline{V}_d^T \tag{22}$$

$\underline{X}_d \sim \underline{X}$ is the rank-d approximation of the original data matrix frequently used when multiple event waveform data are strongly linear dependent. The approximation is formed by retaining only the d largest singular values and associated left- and right-singular vectors in the approximating SVD factors. In subspace detectors, low-rank approximations are used to balance the probability of detection against the false alarm probability. There is significant incentive to reduce the rank in order to bound the false alarm probability.

References

Badeau, R., G. Richard and B. David (2004), Sliding window adaptive SVD Algorithms, IEEE Trans. on Sig. Proc., **52**(1), 1-10.

Brand, M. (2006) Fast low-rank modifications of the thin singular value decomposition, *Linear Algebra and its Applications* **415**(1), 20-30.

Comon, P. and G. H. Golub (1990), Tracking a few extreme singular values and vectors in signal processing, *Proc. of the IEEE*, **78**(8), 1327-1343.

Harris, D. B. (2001), Subspace techniques for detecting repeating events, abstract in *Seismological Research Letters*, Vol. **72**(2), 245.

Harris, D. B. (2006), Subspace detectors: theory, Lawrence Livermore National Laboratory Technical Report UCRL-TR-222758, 46 pages, Livermore, CA.

Karasalo, I. (1986), Estimating the covariance matrix by signal subspace averaging, IEEE Trans. on ASSP, **ASSP-34**(1), 8-12.

Auspices

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.