An Extension of the Generalized SVD for More than Two Matrices

Lieven De Lathauwer

Katholieke Universiteit Leuven
Group Science, Engineering and Technology, 8500 Kortrijk, Belgium
and E.E. Dept. (ESAT) - SCD - SISTA, B-3001 Leuven, Belgium
Email: Lieven.DeLathauwer@kuleuven-kortrijk.be and
Lieven.DeLathauwer@esat.kuleuven.be

Problem description

Given $\mathbf{A}_1 \in \mathbb{R}^{I_1 \times J}$, ..., $\mathbf{A}_K \in \mathbb{R}^{I_K \times J}$, with $I_k \gg J$ for all k, find an approximate decomposition of the type

$$\begin{aligned} \mathbf{A}_1 &= \mathbf{U}_1 \cdot \mathbf{D}_1 \cdot \mathbf{M}^T \\ &\vdots \\ \mathbf{A}_K &= \mathbf{U}_K \cdot \mathbf{D}_K \cdot \mathbf{M}^T, \end{aligned}$$

in which $\mathbf{U}_k \in \mathbb{R}^{I_k \times J}$ is column-wise orthonormal, $\mathbf{D}_k \in \mathbb{R}^{J \times J}$ diagonal and $\mathbf{M} \in \mathbb{R}^{J \times J}$ not necessarily orthogonal ($1 \leq k \leq K$). For $K = 2$, this corresponds to the Generalized SVD. For $K > 2$, the problem is overdetermined (more equations than unknowns). We will solve it in least-squares sense, i.e., we will minimize

$$f = \sum_{k=1}^{K} \|\mathbf{A}_k - \mathbf{U}_k \cdot \mathbf{D}_k \cdot \mathbf{M}^T\|^2, \tag{1}$$

in which the norm is the Frobenius-norm.

Alternating Least Squares Algorithm

We can compute the solution by means of an Alternating Least Squares algorithm (ALS):

1. Update $\{\mathbf{U}_k\}$ for given estimates of $\{\mathbf{D}_k\}$ and $\mathbf{M}$.
2. Update $\{\mathbf{D}_k\}$ and $\mathbf{M}$ for given estimates of $\{\mathbf{U}_k\}$.
3. Go back to 1 until convergence.

Convergence to a local optimum is guaranteed, because in each step we will further decrease the cost function f. It may be necessary to reinitialize a number of times in order to find the global optimum. ALS is a very basic approach in comparison with the advanced techniques in current numerical linear algebra (for instance for the computation of the GSVD). However, it is quite common in multilinear algebra, where numerical aspects are still largely unexplored.

Updating $\{\mathbf{D}_k\}$ and $\mathbf{M}$.

Since $\mathbf{U}_k$ are column-wise orthonormal, we have

$$f = \sum_{k=1}^{K} \|\mathbf{U}_k^T \mathbf{A}_k - \mathbf{D}_k \cdot \mathbf{M}^T\|^2.$$

Call $\mathbf{B}_j \in \mathbb{R}^{K \times J}$ the matrix in which the jth rows of $\mathbf{U}_k^T \mathbf{A}_k$, $1 \leq k \leq K$, are stacked. Define $E_j = (\mathbf{D}_1(j,j) \ \ldots \ \mathbf{D}_K(j,j))^T$ and let M_j denote the jth column of $\mathbf{M}$, $1 \leq j \leq J$. Then we have

$$f = \sum_{j=1}^{J} \|\mathbf{B}_j - E_j M_j^T\|^2.$$

This corresponds to a set of decoupled best rank-1 approximation problems. E_j is the dominant left singular vector of $\mathbf{B}_j$, and M_j is the dominant right singular vector times the dominant singular value of $\mathbf{B}_j$, for all j. Actually it does not matter which of the two vectors is multiplied by the singular value, since a scalar can be exchanged between E_j and M_j.

Updating $\{\mathbf{U}_k\}$.

If $\langle \cdot, \cdot \rangle$ denotes the inner product, then we have

$$\begin{aligned} f &= \sum_{k=1}^{K} \langle \mathbf{A}_k - \mathbf{U}_k \cdot \mathbf{D}_k \cdot \mathbf{M}^T, \mathbf{A}_k - \mathbf{U}_k \cdot \mathbf{D}_k \cdot \mathbf{M}^T \rangle \\ &= \sum_{k=1}^{K} \big(\langle \mathbf{A}_k, \mathbf{A}_k \rangle - \langle \mathbf{A}_k, \mathbf{U}_k \cdot \mathbf{D}_k \cdot \mathbf{M}^T \rangle - \langle \mathbf{U}_k \cdot \mathbf{D}_k \cdot \mathbf{M}^T, \mathbf{A}_k \rangle \\ &\quad + \langle \mathbf{U}_k \cdot \mathbf{D}_k \cdot \mathbf{M}^T, \mathbf{U}_k \cdot \mathbf{D}_k \cdot \mathbf{M}^T \rangle \big) \\ &= \sum_{k=1}^{K} \big(\|\mathbf{A}_k\|^2 + \|\mathbf{U}_k \cdot \mathbf{D}_k \cdot \mathbf{M}^T\|^2 - 2\langle \mathbf{U}_k \cdot \mathbf{D}_k \cdot \mathbf{M}^T, \mathbf{A}_k \rangle \big) \end{aligned}$$

Since $\mathbf{U}_k$ is column-wise orthonormal, $\|\mathbf{U}_k \cdot \mathbf{D}_k \cdot \mathbf{M}^T\|^2 = \|\mathbf{D}_k \cdot \mathbf{M}^T\|^2$ does not depend on $\mathbf{U}_k$. Hence, we have

$$\begin{aligned} f &= c - 2 \sum_{k=1}^{K} \langle \mathbf{U}_k \cdot \mathbf{D}_k \cdot \mathbf{M}^T, \mathbf{A}_k \rangle \\ &= c - 2 \sum_{k=1}^{K} \langle \mathbf{U}_k, \mathbf{A}_k \mathbf{M} \mathbf{D}_k \rangle \end{aligned}$$

Up to a constant, the same expression would have been obtained if we had started from

$$g = \sum_{k=1}^{K} \|\mathbf{A}_k \mathbf{M} \mathbf{D}_k - \mathbf{U}_k\|^2 \tag{2}$$

The latter equation corresponds to a set of decoupled Procrustes problems (find the best column-wise orthonormal approximation of a given matrix) [1]. If the SVD of $\mathbf{A}_k \mathbf{M} \mathbf{D}_k$ is given by $\mathbf{P}_k \mathbf{\Sigma}_k \mathbf{Q}_k^T$, then the optimal $\mathbf{U}_k$ is equal to $\mathbf{P}_k \mathbf{Q}_k^T$, for all k. ($\mathbf{\Sigma}_k$ contains only the nonzero singular values of $\mathbf{A}_k \mathbf{M} \mathbf{D}_k$; the dimensions of $\mathbf{P}_k$ and $\mathbf{Q}_k$ are defined accordingly.)

Note that the columns of the optimal $\mathbf{U}_k$ always lie in the column space of $\mathbf{A}_k \mathbf{M} \mathbf{D}_k$, hence in the column space of $\mathbf{A}_k$. This means that prior to the ALS algorithm, we may reduce the dimensions by projecting on the column space of $\mathbf{A}_k$:

1. Compute the K SVD's

$$\mathbf{A}_k = \tilde{\mathbf{U}}_k \cdot \tilde{\mathbf{S}}_k \cdot \tilde{\mathbf{V}}_k^T$$

2. Substitute

$$\mathbf{A}_k \leftarrow \tilde{\mathbf{U}}_k^T \mathbf{A}_k \qquad \mathbf{U}_k \leftarrow \tilde{\mathbf{U}}_k^T \mathbf{U}_k$$

3. ALS for square matrices $\{\mathbf{A}_k\}$, $\{\mathbf{U}_k\}$, $\{\mathbf{D}_k\}$, $\mathbf{M}$.

4. Substitute $\mathbf{U}_k \leftarrow \tilde{\mathbf{U}}_k \mathbf{U}_k$.

Acknowledgments

Research supported by: (1) Research Council K.U.Leuven: GOA-Ambiorics, GOA-MaNet, CoE EF/05/006 Optimization in Engineering (OPTEC), CIF1, STRT1/08/023, (2) F.W.O.: (a) project G.0321.06, (b) Research Communities ICCoS, ANMMM and MLDM, (3) the Belgian Federal Science Policy Office: IUAP P6/04 (DYSCO, "Dynamical systems, control and optimization", 2007–2011), (4) EU: ERNSI.

References

[1] Golub, G.H. and Van Loan, C.F. 1996. *Matrix Computations*, Baltimore, Maryland: Johns Hopkins University Press.