Simultaneous Maximization of the Trace of a Set of 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

Nov. 14, 2009

Problem description

Given $\mathbf{M}_1, \ldots, \mathbf{M}_K \in \mathbb{R}^{N \times N}$, and $R < N$, we wish to maximize

$$f(\mathbf{a}, \mathbf{Q}) = \sum_{k=1}^{K} \sum_{r=1}^{R} a_k (\tilde{\mathbf{M}}_k)_{rr}, \tag{1}$$

in which $\mathbf{Q} \in \mathbb{R}^{N \times N}$ is orthogonal, $\mathbf{a} \in \mathbb{R}^K$ is unit-norm and $\tilde{\mathbf{M}}_k = \mathbf{Q} \cdot \mathbf{M}_k \cdot \mathbf{Q}^T$, $k = 1, \ldots, K$. We will also consider the case where $\mathbf{a}$ is either unit-norm or zero, and nonnegative. The matrices $\mathbf{M}_k$ are not necessarily positive (semi)definite. Eq. (1) will be written as

$$f(\mathbf{a}, \mathbf{Q}) = \sum_{k=1}^{K} a_k \, \mathrm{trace}_R(\tilde{\mathbf{M}}_k). \tag{2}$$

Note that $\mathrm{trace}_R(\cdot)$ is a linear function, i.e.,

$$\mathrm{trace}_R(\alpha \mathbf{A} + \beta \mathbf{B}) = \alpha \, \mathrm{trace}_R(\mathbf{A}) + \beta \, \mathrm{trace}_R(\mathbf{B}). \tag{3}$$

Alternating algorithm

We can compute the solution by means of an alternating algorithm:

1. Update $\mathbf{a}$ for given estimate of $\mathbf{Q}$.
2. Update $\mathbf{Q}$ for given estimate of $\mathbf{a}$.
3. Go back to 1 until convergence.

Convergence to a local optimum is guaranteed, because in each step we will maximally increase the objective function f. It may be necessary to reinitialize a number of times in order to find the global optimum.

Updating a.

Given $\mathbf{Q}$, f in (1) can be written as

$$\begin{aligned} f(\mathbf{a}) &= \sum_{k=1}^{K} a_k \sum_{r=1}^{R} (\tilde{\mathbf{M}}_k)_{rr} \\ &= \mathbf{a}^T \cdot \left(\sum_{r=1}^{R} (\tilde{\mathbf{M}}_1)_{rr}, \ldots, \sum_{r=1}^{R} (\tilde{\mathbf{M}}_K)_{rr} \right)^T \\ &\stackrel{\text{def}}{=} \mathbf{a}^T \cdot \mathbf{x}. \end{aligned}$$

The optimal vector $\mathbf{a}$ is given by $\frac{\mathbf{x}}{\|\mathbf{x}\|}$ and the optimal value of f is given by $\|\mathbf{x}\|$.

Special case: $\mathbf{a}$ *is either unit-norm or zero, and nonnegative.* Denote by $\tilde{\mathbf{x}} \in \mathbb{R}^K$ the vector obtained from $\mathbf{x}$ by setting the negative components equal to zero. If $\tilde{\mathbf{x}} \neq \mathbf{0}$, then the optimal vector $\mathbf{a}$ is given by $\frac{\tilde{\mathbf{x}}}{\|\tilde{\mathbf{x}}\|}$ and the optimal value of f is given by $\|\tilde{\mathbf{x}}\|$. If $\tilde{\mathbf{x}} = \mathbf{0}$, then we restart the iteration from a new initial point. After a number of restarts, we decide that the overall optimal vector $\mathbf{a}$ and the overall optimal value of f are zero.

Updating Q.

Given $\mathbf{a}$, f can be written as

$$f(\mathbf{Q}) = \mathrm{trace}_R(\mathbf{Q} \cdot (\sum_{k=1}^{K} a_k \mathbf{M}_k) \cdot \mathbf{Q}^T). \tag{4}$$

Hence, we have to determine the optimal $\mathbf{Q}$ for only one matrix, namely $\sum_{k=1}^{K} a_k \mathbf{M}_k$, which we write as $\mathbf{Y} \in \mathbb{R}^{N \times N}$. Only the first R rows of $\mathbf{Q}$ are of importance. If $\mathbf{Y}$ has not less than R positive eigenvalues, then the first R rows of $\mathbf{Q}$ are taken equal to the R eigenvectors that correspond to the largest eigenvalues of $\mathbf{Y}$. If $\mathbf{Y}$ has only $\tilde{R} < R$ positive eigenvalues, then the first R rows of $\mathbf{Q}$ are taken equal to the $\tilde{R}$ eigenvectors that correspond to the positive eigenvalues of $\mathbf{Y}$, complemented with the $R - \tilde{R}$ eigenvectors that correspond to the least negative eigenvalues of $\mathbf{Y}$.

Initialization.

The (first R rows of the) matrix $\mathbf{Q}$ may be initialized as (part of) the identity matrix, or the matrix obtained from the truncated MLSVD, or the matrix obtained from HOOI, or a set of random mutually orthonormal vectors.

Preprocessing: dimensionality reduction.

Stack the matrices $\mathbf{M}_k$ in a tensor $\mathcal{M} \in \mathbb{R}^{N \times N \times K}$. Let the MLSVD of $\mathcal{M}$ be given by

$$\mathcal{M} = \mathcal{S} \bullet_1 \mathbf{U} \bullet_2 \mathbf{U} \bullet_3 \mathbf{V}. \tag{5}$$

Assume that for some $R' \geqslant R$, the mode-1 singular values $\sigma^{(1)}_{R'+1}$, $\ldots$, $\sigma^{(1)}_{N}$ are very small, in the sense that the corresponding part of $\mathcal{S}$ does not really contribute to $\mathcal{M}$. Then we may do the analysis for $\mathcal{S}(1:R', 1:R', 1)$, $\ldots$, $\mathcal{S}(1:R', 1:R', K)$ and afterwards multiply by $\mathbf{U}(:, 1:R')$. (This is a bit delicate when, for instance, all the $\mathbf{M}_k$ are negative semidefinite, but that is a detail; beyond R', we should actually have a margin of at least R, to avoid

problems with negativity; to be discussed.) One may wish to conclude with a few fine-tuning steps of the algorithm in the original space.

Acknowledgments

Research supported by: (1) Research Council K.U.Leuven: GOA-MaNet, CoE EF/05/006 Optimization in Engineering (OPTEC), CIF1, STRT1/08/023, (2) F.W.O.: (a) project G.0427.10N, (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.