

Numerical Methods for Q-Weighted Nonnegative Matrix Tri-Factorization

Chun-Mei Li^{1,*}, Lin-Zhang Lu^{1,2} and Zhen Chen¹

¹*School of Mathematical Sciences, Guizhou Normal University, Guiyang 550001, P.R. China.*

²*School of Mathematical Sciences, Xiamen University, Xiamen 361005, P.R. China.*

Received 15 April 2019; Accepted (in revised version) 14 July 2019.

Abstract. The Q-weighted nonnegative matrix tri-factorisation problem arising in clustering analysis is considered and a necessary condition for its optimal solution is obtained. The problem is solved by the proximal alternating nonnegative least squares method supplemented by an acceleration scheme. The convergence of the methods is studied. Numerical examples show the feasibility and efficiency of the methods proposed.

AMS subject classifications: 15A48, 15A23, 65F30

Key words: Nonnegative matrix factorisation, necessary condition, PANLS method, ELS technique, clustering analysis.

1. Introduction

Let $R^{m \times n}$, $R_+^{m \times n}$ and $SR_{++}^{n \times n}$ be, respectively, the sets of real $m \times n$ matrices, nonnegative real $m \times n$ matrices and symmetric positive definite $n \times n$ matrices. Moreover, A^T refers to the transpose of the matrix A and $\text{tr}(A)$ to the trace of A . If $B \in R^{m \times n}$, the vector $\vec{B} \in R^{mn}$ consists of the elements B_{ij} , $i, j = 1, 2, \dots, n$ arranged in the reverse lexicographic order of the indices. It is easily seen that for any row vector $b \in R^{mn}$, there is an $m \times n$ matrix B such that $\vec{B}^T = b$. The symbols $A \circ B$ and $A \otimes B$ stand for Hadmard and Kronecker products of matrices A and B , respectively. For differentiable function f , we denote by $\nabla_X f(X_0)$ the gradient of f at X_0 with respect to the variable X , i.e.

$$[\nabla_X f(X_0)]_{ij} = \frac{\partial f}{\partial X_{ij}}(X_0).$$

Further, let $\|A\|_F$ be the Frobenius norm of the matrix A induced by the inner product $\langle B, C \rangle := \text{tr}(C^T B)$ for $B, C \in R^{m \times n}$. If $Q \in SR_{++}^{mn \times mn}$, the Q-weighted norm of $A \in R^{m \times n}$ is defined by

*Corresponding author. *Email addresses:* lengyue123@126.com (C.-M. Li), lzlu@xmu.edu.cn (L.-Z. Lu), zchen@gznu.edu.cn (Z. Chen)

$$\|A\|_Q = \sqrt{\vec{A}^T Q \vec{A}}.$$

It is induced by the inner product

$$\langle A, B \rangle_Q = \vec{A}^T Q \vec{B}, \quad A, B \in R^{m \times n}.$$

For $Q = I_{mn}$, the Q-weighted norm coincides with the Frobenius norm — i.e.

$$\|A\|_F^2 = \|A\|_Q^2, \quad Q = I_{mn}.$$

If $W \in R^{m \times n}$, the W-weighted norm is defined by

$$\|A\|_W = \|W \circ A\|_F, \quad A \in R^{m \times n}.$$

The H-weighted norm is defined by

$$\|A\|_H = \|H^{1/2} A H^{1/2}\|_F, \quad A \in R^{n \times n},$$

where $H \in SR_{++}^{n \times n}$ and $H^{1/2}$ is the square root of the matrix H . According to [4], W- and H-weighted norms are special cases of the Q-weighted norm — viz.

$$\begin{aligned} \|A\|_W^2 &= \|A\|_Q^2, \quad Q = \text{diag}(\vec{W} \circ W), \\ \|A\|_H^2 &= \|A\|_Q^2, \quad Q = H \otimes H. \end{aligned}$$

It is worth noting that Q-weighted norm is a convenient tool to measure distance between matrices and is widely used in signal processing [27] and system identification [29].

This work deals with the following Q-weighted nonnegative matrix tri-factorisation (QNMFTF) problem.

Problem 1. Given matrices $A \in R^{m \times n}$ and $Q \in SR_{++}^{mn \times mn}$, find matrices $\widehat{R} \in R_+^{m \times r}$, $\widehat{B} \in R_+^{r \times t}$ and $\widehat{C} \in R_+^{t \times n}$, such that

$$\|A - \widehat{R} \widehat{B} \widehat{C}\|_Q^2 = \min_{R \in R_+^{m \times r}, B \in R_+^{r \times t}, C \in R_+^{t \times n}} \|A - RBC\|_Q^2.$$

Problem 1 arises in clustering analysis, including document clustering, collaborative filtering and microarray analysis [6, 7, 12, 18]. The data can be formulated as a two-dimensional matrix of dyadic data. The term dyadic means a domain with two sets of objects $X = \{x_1, x_2, \dots, x_n\}$ and $Y = \{y_1, y_2, \dots, y_m\}$ and the dyad is a scalar $w(x, y)$, e.g. the frequency of co-occurrence. For a scalar dyad, the data can always be represented as an $n \times m$ matrix A by assigning the row indices to the elements in X and the column indices to the ones in Y , so that every $w(x, y)$ corresponds to an entry of A . We are interested in simultaneous splitting of X and Y into r and t disjoint clusters, respectively. This problem is equivalent to representing A as an $r \times t$ block matrix such that in each submatrix the entries are, in a sense, similar to each other and dissimilar from the ones belonging to the