Applied Mathematics Letters 45 (2015) 12–17

Contents lists available at ScienceDirect

Applied Mathematics Letters

journal homepage: www.elsevier.com/locate/aml

An efficient monotone projected Barzilai–Borwein method for nonnegative matrix factorization

Yakui Huang *, Hongwei Liu, Sha Zhou

School of Mathematics and Statistics, Xidian University, Xi'an 710126, PR China

A R T I C L E I N F O

Article history:
Received 14 October 2014
Received in revised form 10 January 2015
Accepted 10 January 2015
Available online 20 January 2015

Keywords:
Nonnegative matrix factorization
Alternating nonnegative least squares
Projected Barzilai–Borwein method
Monotone

A B S T R A C T

In this paper, we present an efficient method for nonnegative matrix factorization based on the alternating nonnegative least squares framework. Our approach adopts a monotone projected Barzilai–Borwein (MPBB) method as an essential subroutine where the step length is determined without line search. The Lipschitz constant of the gradient is exploited to accelerate convergence. Global convergence of the proposed MPBB method is established. Numerical results are reported to demonstrate the efficiency of our algorithm.

1. Introduction

Nonnegative matrix factorization (NMF) [1,2] is to approximate a given matrix $V \in \mathbb{R}^{m\times n}$, $V \geq 0$ by the product of two nonnegative matrices $W \in \mathbb{R}^{m\times r}$ and $H \in \mathbb{R}^{r\times n}$ with $r \ll \min\{m, n\}$ being a specified positive integer. Using the Frobenius norm to measure the distance between V and WH, NMF can be formulated as:

$$\begin{aligned} &\min_{W,H}\ F(W,H) := \frac{1}{2}\|V - WH\|_F^2 \\ &s.t.\quad W \in \mathbb{R}_+^{m\times r},\ H \in \mathbb{R}_+^{r\times n}. \end{aligned} \tag{1}$$

NMF has received considerable attention in the past decade due to its usefulness in dimension reduction of image, text, and signal data, see [3–8] and references therein. Numerous algorithms have been proposed for solving (1). Lee and Seung [9] developed the multiplicative update (MU) algorithm which updates the two matrices by multiplying each entry with a positive factor in each iteration. Although MU is easy to implement, it has been observed to converge relatively slowly, especially when dealing with dense matrices [3,10,11]. Paatero and Tapper [2] suggested to use the alternating nonnegative least squares (ANLS) framework:

$$W^{k+1} = \arg\min_{W\geq 0} F(W, H^k), \tag{2}$$

$$H^{k+1} = \arg\min_{H\geq 0} F(W^{k+1}, H). \tag{3}$$

Notice that problem (1) is nonconvex and NP-hard [12]. The ANLS framework allows us to solve two convex subproblems (2) and (3) for which optimal solutions can be found. Recently, Grippo and Sciandrone [13] proved the convergence of the ANLS framework to a stationary point of (1).

Clearly, at each iteration, the main cost of the ANLS framework is in solving the subproblems (2) and (3). Many algorithms aim to solve the two subproblems efficiently have been developed, for example, the active set method [14], the projected

* Corresponding author.
E-mail address: huangyakui2006@gmail.com (Y. Huang).

http://dx.doi.org/10.1016/j.aml.2015.01.003
0893-9659/

gradient (PG) method [11], the projected Barzilai–Borwein (BB) method [15], the projected Newton method [16], and the projected quasi-Newton method [10,17]. Guan et al. [18] pointed out that these methods may be inefficient because of using time consuming line searches. They applied Nesterov's optimal gradient method (OGM) [19] to solve the subproblems without line search and developed the NeNMF solver. However, it was observed by Huang et al. [20] that OGM may take thousands of iterations to reach a given tolerance which will degrade the efficiency of NeNMF. In [20], the authors proposed a quadratic regularization projected Barzilai–Borwein (QRPBB) method. By making use of the Lipschitz constant of the gradient, the QRPBB method improves the performance of the projected BB method significantly and outperforms other three solvers including PG, APBB2 [15], and NeNMF. However, the QRPBB method has to calculate two projections and two gradients at each iteration which is expensive for a large matrix. Moreover, it needs to perform a nonmonotone line search to determine the step length at each iteration.

In this paper, follow the ANLS framework, we present an efficient monotone projected BB (MPBB) method to solve the subproblems. The MPBB method exploits the Lipschitz constant of the gradient and makes use of the two BB stepsizes [21] alternately to accelerate convergence. Unlike the QRPBB method, our MPBB method only computes two projections and two gradients at even steps. Moreover, the proposed MPBB method determines the step length without using any line search. Global convergence of the MPBB method is established. Numerical results are reported to demonstrate the efficiency of our algorithm.

As is well known that BB-like methods are often more efficient with nonmonotone schemes. We will show by experiments that our MPBB method outperforms the APBB2 method which employs the Grippo–Lampariello–Lucidi nonmonotone line search [22]. Our results provide the possibility that, by proper modification, monotone BB-like methods can win the nonmonotone ones in some cases.

The rest of this paper is organized as follows. Section 2 introduces the MPBB method for solving the subproblems and presents its global convergence result. Experimental comparisons among several NMF solvers are presented in Section 3.

2. Monotone projected Barzilai–Borwein method and its convergence

Since W and H is perfectly symmetric, we focus only on the update of the matrix W. At the kth iteration of the ANLS framework, we have to solve

$$\min_{W\geq 0} f^k(W) := F(W, H^k) = \frac{1}{2}\|V - WH^k\|_F^2. \tag{4}$$

To simplify the notation, we use $f(W)$ rather than $f^k(W)$ in the rest of the paper.

Let $P(\cdot)$ be the operator that projects all the negative entries of an input matrix to zero. It is well known that W is a stationary point of (4) if and only if, for any fixed $\alpha > 0$,

$$\|P[W - \alpha\nabla f(W)] - W\|_F = 0. \tag{5}$$

By Lemma 1 in [20], we know that $f(W)$ is convex and its gradient $\nabla f(W)$ is Lipschitz continuous with constant $L = \|H^k(H^k)^T\|_2$. Since $H^k(H^k)^T$ is an $r \times r$ matrix and $r \ll \min\{m, n\}$, the Lipschitz constant L is not expensive to obtain.

Our monotone projected Barzilai–Borwein (MPBB) method is presented in Algorithm 1, where W^k and H^k are obtained from the previous iteration in the ANLS framework.

Algorithm 1. Monotone projected BB method

Step 1. Choose constants $\sigma \in (0, 1)$, $\alpha_{\max} > \alpha_{\min} > 0$. Compute $L = \|H^k(H^k)^T\|_2$. Set $W_0 = W^k$, $\alpha_0 = 1$, and $t = 0$.
Step 2. If W_t is a stationary point of (4), stop.
Step 3. If t is odd, set $Z_t = W_t$; otherwise, compute Z_t by

$$Z_t = P\left[W_t - \frac{1}{L}\nabla f(W_t)\right]. \tag{6}$$

Step 4. Compute $D_t = P[Z_t - \alpha_t\nabla f(Z_t)] - Z_t$ and $\delta_t = \langle D_t, H^k(H^k)^T D_t\rangle$. If $\delta_t = 0$, set $\lambda_t = 1$; otherwise, set $\lambda_t = \min\{\tilde{\lambda}_t, 1\}$, where

$$\tilde{\lambda}_t = -\frac{(1-\sigma)\langle\nabla f(Z_t), D_t\rangle}{\delta_t}. \tag{7}$$

Set $W_{t+1} = Z_t + \lambda_t D_t$.
Step 5. Define $S_t = W_{t+1} - W_t$ and $Y_t = \nabla f(W_{t+1}) - \nabla f(W_t)$. If $\langle S_t, Y_t\rangle \leq 0$, set $\alpha_{t+1} = \alpha_{\max}$; otherwise, compute

$$\alpha_{t+1}^{BB} = \begin{cases} \dfrac{\langle S_t, S_t\rangle}{\langle S_t, Y_t\rangle}, & \text{for odd } t; \\ \dfrac{\langle S_t, Y_t\rangle}{\langle Y_t, Y_t\rangle}, & \text{for even } t. \end{cases} \tag{8}$$

Set $\alpha_{t+1} = \min\{\alpha_{\max}, \max\{\alpha_{\min}, \alpha_{t+1}^{BB}\}\}$.
Step 6. Set $t = t + 1$ and go to Step 2.