Journal of Scientific Computing (2022) 90:37
https://doi.org/10.1007/s10915-021-01729-z

CORRECTION

Correction to: An Approximate Augmented Lagrangian Method for Nonnegative Low-Rank Matrix Approximation

Hong Zhu[1] · Michael K. Ng[2] · Guang-Jing Song[3]

Received: 5 September 2021 / Accepted: 30 October 2021 / Published online: 7 December 2021

Correction to: Journal of Scientific Computing (2021) 88:45 https://doi.org/10.1007/s10915-021-01556-2

The original version of this article [4] unfortunately contained an error. The authors would like to correct the error with this corrigendum.

In [4], the optimization formulation is not presented correctly. We should use the following model:

$$\min_X \ \Phi(X) := \frac{1}{2}||A - X||_F^2 \quad \text{subject to} \quad \text{rank}(X) \le r, \ X \ge 0. \tag{1}$$

instead of (2) stated in [4] to avoid rank-deficient iteration points caused by the fact that the constraint set $\mathcal{M}_r = \{X \in \mathbb{R}^{m\times n} | \text{rank}(X) = r\}$ used in [4] is not closed.

In order to derive the Augmented Lagrangian (AL) method for (1), we need to add some geometric properties of $\mathcal{M}_{\le r} = \{X \in \mathbb{R}^{m\times n} | \text{rank}(X) \le r\}$ which is a real-algebraic variety [3]. In the rank-deficient point $X \in \mathcal{M}_{\le r}$ with $\text{rank}(X) = s < r$, the tangent cone $T_X\mathcal{M}_{\le r}$ is given by [2]

$$T_X\mathcal{M}_{\le r} = T_X\mathcal{M}_s \oplus \{\Xi_{r-s} \in \mathcal{U}^\perp \otimes \mathcal{V}^\perp \mid \text{rank}(\Xi_{r-s}) \le r - s\},$$

where $T_X\mathcal{M}_s$ is the tangent space of $\mathcal{M}_s$ at X, $\oplus$ denotes direct sum and $\otimes$ denotes Kronecker product. Then for any $Z \in \mathbb{R}^{m\times n}$, the orthogonal projection of Z onto $T_X\mathcal{M}_{\le r}$ follows

$$\mathcal{P}_{T_X\mathcal{M}_{\le r}}(Z) = \mathcal{P}_{T_X\mathcal{M}_s}(Z) + \Xi_{r-s},$$

The original article can be found online at https://doi.org/10.1007/s10915-021-01556-2.

✉ Michael K. Ng
mng@maths.hku.hk

Hong Zhu
zhuhongmath@126.com

Guang-Jing Song
sgjshu@163.com

1 School of Mathematical Sciences, Jiangsu University, Zhenjiang 212013, Jiangsu, People's Republic of China

2 Department of Mathematics, The University of Hong Kong, Pokfulam, Hong Kong

3 School of Mathematics and Information Sciences, Weifang University, Weifang 261061, People's Republic of China

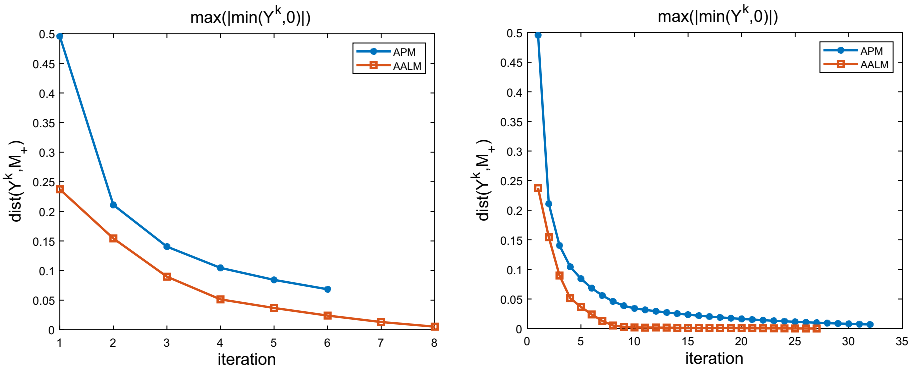


Fig. 1 Correction to Figure 1 in [4]. Iteration convergence on $\operatorname{dist}(Y^k, \mathcal{M}_+)$ for face dataset UMist with Gaussian noise $\mathcal{N}(0, (30/255)^2)$. Left: $(\varepsilon_p, \varepsilon_f) = (e^{-7}, e^{-9})$; right: $(\varepsilon_p, \varepsilon_f) = (e^{-10}, e^{-13})$

where Ξ_{r-s} is a best rank-$(r-s)$ approximation of $Z - \mathcal{P}_{T_X \mathcal{M}_r}(Z)$ in the Frobenius norm. For differentiable function Φ, the critical point X^* of $\min_{X \in \mathcal{M}_{\leq r}} \Phi(X)$ satisfies [3]

$$0 = \|\nabla \Phi(X^*)\|^2 - \operatorname{dist}(-\nabla \Phi(X^*), T_{X^*}\mathcal{M}_{\leq r})^2, \tag{2}$$

where

$$\operatorname{dist}(Z, T_{X^*}\mathcal{M}_{\leq r}) = \|Z - \mathcal{P}_{T_{X^*}\mathcal{M}_{\leq r}}(Z)\|_F.$$

Recall that $\mathcal{P}_{T_{X^*}\mathcal{M}_{\leq r}}(Z)$ is a orthogonal projection of Z onto $T_{X^*}\mathcal{M}_{\leq r}$, it holds that

$$\|\mathcal{P}_{T_{X^*}\mathcal{M}_{\leq r}}(-\nabla \Phi(X^*))\|^2 = \|-\nabla \Phi(X^*)\|^2 - \|-\nabla \Phi(X^*) - \mathcal{P}_{T_{X^*}\mathcal{M}_{\leq r}}(-\nabla \Phi(X^*))\|_F^2.$$

Therefore, (2) is equivalent to

$$\mathcal{P}_{T_{X^*}\mathcal{M}_{\leq r}}(-\nabla \Phi(X^*)) = 0.$$

If $\operatorname{rank}(X^*) = r$, according to the definition of $\mathcal{P}_{T_{X^*}\mathcal{M}_{\leq r}}(\cdot)$, it holds that $\mathcal{P}_{T_{X^*}\mathcal{M}_r}(-\nabla \Phi(X^*)) = 0$.

The AL subproblem can be rewritten as

$$\min_{X,Y} \mathcal{L}_{\rho_{k-1}}(X, Y, \Lambda^{k-1}) + \delta_{\mathcal{M}_+}(X) + \delta_{\mathcal{M}_{\leq r}}(Y). \tag{3}$$

According to the above discussion of geometric properties of $\mathcal{M}_+$ and $\mathcal{M}_{\leq r}$, the stationary point (X_*^k, Y_*^k) of (3) satisfies

$$\begin{cases} 0 \in \nabla_X \mathcal{L}_{\rho_{k-1}}(X_*^k, Y_*^k, \Lambda^{k-1}) + \partial \delta_{\mathcal{M}_+}(X_*^k); \\ 0 = \|\nabla_Y \mathcal{L}_{\rho_{k-1}}(X_*^k, Y_*^k, \Lambda^{k-1})\|^2 - \operatorname{dist}(-\nabla_Y \mathcal{L}_{\rho_{k-1}}(X_*^k, Y_*^k, \Lambda^{k-1}), T_{Y_*^k}\mathcal{M}_{\leq r})^2. \end{cases}$$

Then Algorithm 1 of [4] can be revised.

The main convergence result of Theorem 1 for Algorithm 1 stated in the original article [4] still holds. By substituting $\mathcal{M}_{\leq r}$ for $\mathcal{M}_r$, the proof is the same except formula (17) in [4] should be replaced by

$$\left| \|\frac{1}{2}(Y^k - A) + \Lambda^k\|^2 - \operatorname{dist}(-\frac{1}{2}(Y^k - A) - \Lambda^k, T_{Y^k}\mathcal{M}_{\leq r})^2 \right| \leq \epsilon_k/2, \ \forall k \in \mathbb{N}. \tag{8}$$

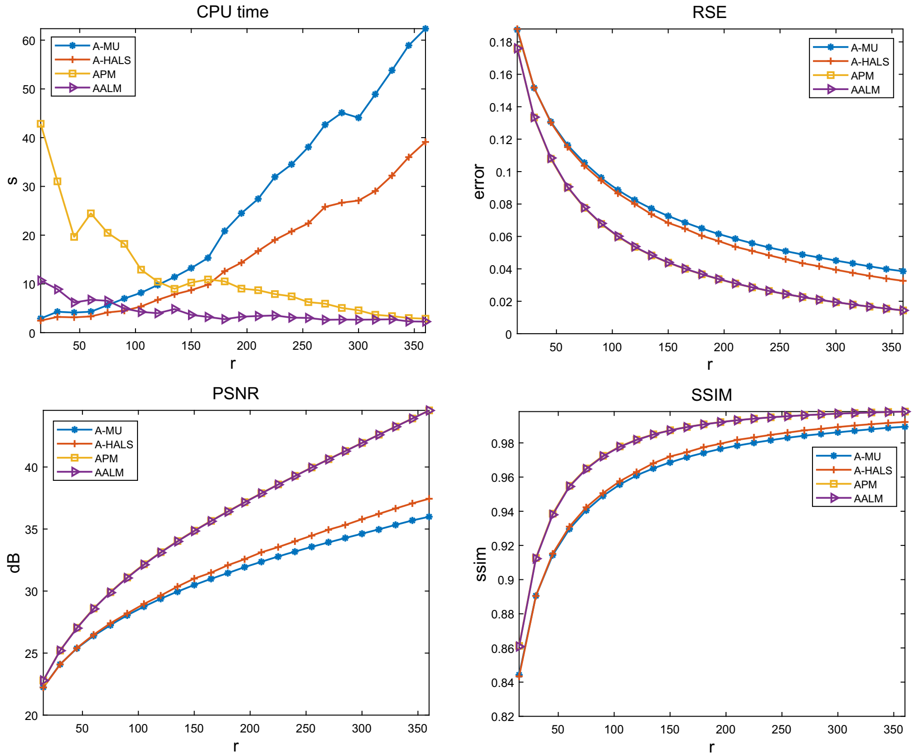
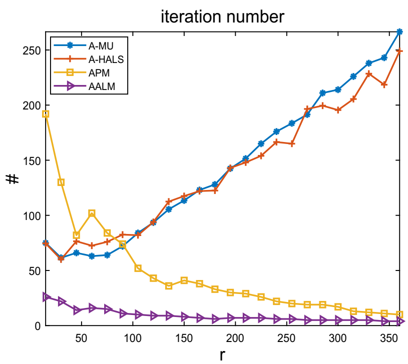
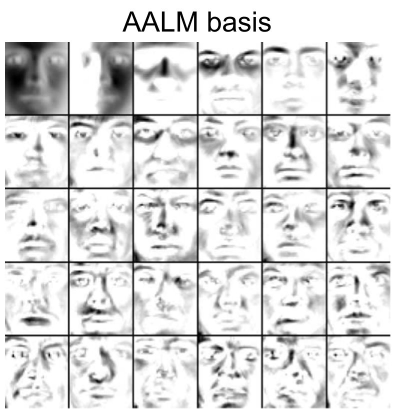


Fig. 2 Correction to Figure 2 in [4]. CPU time, recover qualities RSE, PSNR and SSIM as function of the number of testing rank r for dataset YaleB



(a) Number of iterations as function of the number of testing rank r for dataset YaleB.



(b) Basis elements obtained by AALM for dataset YaleB.

Fig. 3 Left: correction to Figure 3 in [4]. Right: correction to Figure 5 in [4]

Algorithm 1 (AALMNLR: Approximate Augmented Lagrangian Method for NLR model.

Require: $\{\epsilon_k\}_{k\in\mathbb{N}} \downarrow 0$, $\mu > 1$, $k = 1$, $\rho_0 > 0$, $\theta > 0$, $(X^{\text{feasi}}, Y^{\text{feasi}}) \in \mathcal{M}_+ \times \mathcal{M}_r$, $\Lambda^0 \in \mathbb{R}^{m\times n}$, $\widetilde{\Psi} \geq \max\{\Psi(X^{\text{feasi}}, Y^{\text{feasi}}), \mathcal{L}_{\rho_0}(X^{\text{feasi}}, Y^{\text{feasi}}; \Lambda^0)\}$.

Ensure: A sequence $\{(X^k, \Lambda^k)\}_{k\in\mathbb{N}}$.

1: **while** stopping criterion is not satisfied **do**

2: **Step 1.** For given ρ_{k-1}, Λ^{k-1}, compute (X^k, Y^k) for the AL subproblem (3) such that $(X^k, Y^k) \in \mathcal{M}_+ \times \mathcal{M}_{\leq r}$,

$$\text{dist}\left(-\nabla_X \mathcal{L}_{k-1}(X^k, Y^k; \Lambda^{k-1}), \mathcal{N}_{\mathcal{M}_+}(X^k)\right) \leq \epsilon_k/2; \tag{4}$$

$$\left| \left\| \nabla_Y \mathcal{L}_{k-1}(X^k, Y^k; \Lambda^{k-1}) \right\|_F^2 - \text{dist}\left(-\nabla_Y \mathcal{L}_{k-1}(X^k, Y^k; \Lambda^{k-1}), T_{Y^k}\mathcal{M}_{\leq r}\right)^2 \right| \leq \epsilon_k/2; \tag{5}$$

and

$$\mathcal{L}_{\rho_{k-1}}(X^k, Y^k, \Lambda^{k-1}) \leq \widetilde{\Psi}. \tag{6}$$

3: **Step 2.** Update the Lagrangian multipliers

$$\Lambda^k = \Lambda^{k-1} - \rho_{k-1}(X^k - Y^k).$$

4: **Step 3.** Update the penalty parameter

$$\rho_k = \max\{\mu\rho_{k-1}, \|\Lambda^k\|_\infty^{1+\theta}\}. \tag{7}$$

5: **end while**

6: **return** $\{(X^k, \Lambda^k)\}_{k\in\mathbb{N}}$.

Recall (15) and (16) in [4], taking limits as $\mathcal{K} \ni k \to \infty$ on both sides of (8), there exists $Z^* \in \partial\delta_{\mathcal{M}_+}(X^*)$, such that

$$\|\nabla\Phi(X^*) + Z^*\|^2 - \text{dist}(-\nabla\Phi(X^*) - Z^*, T_{X^*}\mathcal{M}_{\leq r})^2 = 0,$$

which implies that X^* is a stationary point of problem (1).

Note that the AL subproblem is a composite optimization problem. It can be checked that (a) $\delta_{\mathcal{M}_+}$ and $\delta_{\mathcal{M}_{\leq r}}$ are proper lower semicontinuous; (b) $\Theta(X, Y) \triangleq \Psi(X, Y) - \langle\Lambda^{k-1}, X - Y\rangle + \frac{\rho_{k-1}}{2}\|X - Y\|_F^2$ is a $\mathcal{C}^1$ function and $\nabla\Theta$ is Lipschitz continuous on bounded subset of $\mathbb{R}^{m\times n} \times \mathbb{R}^{m\times n}$ (for the case that penalty parameter ρ_k tends to infinity, the AL subproblem can be scaled by $1/\rho_k$ to ensure that $\nabla_X L_k(X, Y)$ and $\nabla_Y L_k(X, Y)$ are Lipschitz continuous.); (c) $\mathcal{M}_{\leq r}$ is a real-algebraic variety [3] and $L_k(X, Y)$ is a semi-algebraic function as a finite sum of semi-algebraic functions [1]. Thus the AL subproblem has the KL property. Therefore, the inner-loop solver is valid.

Since the optimality condition of the AL subproblem is changed when substituting $\mathcal{M}_{\leq r}$ for $\mathcal{M}_r$, Theorem 2(ii) in [4] should be replaced by (ii) $\left(X^{k,j}, Y^{k,j}\right)$ converges to a critical point of L_k. Let $\left(X^{k,*}, Y^{k,*}\right)$ be the limit point of $\{\left(X^{k,j}, Y^{k,j}\right)\}_{j\in\mathbb{N}}$. Then

$$0 \in \nabla_X \mathcal{L}_{\rho_{k-1}}(X^{k,*}, Y^{k,*}, \Lambda^{k-1}) + \partial\delta_{\mathcal{M}_+}(X^{k,*});$$
$$0 = \|\nabla_Y \mathcal{L}_{\rho_{k-1}}(X^{k,*}, Y^{k,*}, \Lambda^{k-1})\|^2 - \text{dist}(-\nabla_Y \mathcal{L}_{\rho_{k-1}}(X^{k,*}, Y^{k,*}, \Lambda^{k-1}), T_{Y^{k,*}}\mathcal{M}_{\leq r})^2.$$

Table 1 Correction to Table 1 in [4]. Average results over 10 tests on synthetic data with fixed m

m	n	r	σ^2	AALM			
				CPU	RSE	O-ITER	I-ITER
500	200	10	0	**0.00**	**0.000**	0	0.00
			0.1	**0.13**	0.553	8	1.34
			0.2	0.50	0.674	12	2.50
			0.3	2.50	0.737	17	10.57
	400	20	0	**0.01**	**0.000**	0	0.00
			0.1	**0.39**	**0.480**	7	1.14
			0.2	**0.63**	**0.601**	10	1.40
			0.3	**0.93**	**0.669**	14	1.41
	600	30	0	**0.02**	**0.000**	0	0.00
			0.1	**0.54**	**0.441**	7	1.00
			0.2	**0.87**	**0.560**	9	1.33
			0.3	**1.45**	**0.629**	15	1.35
	800	40	0	**0.02**	**0.000**	0	0.00
			0.1	**0.69**	0.424	7	1.00
			0.2	**1.10**	**0.541**	9	1.30
			0.3	**1.87**	**0.610**	15	1.26
1000	200	10	0	**0.01**	**0.000**	0	0.00
			0.1	**0.35**	**0.557**	12	1.18
			0.2	**0.41**	**0.679**	16	1.15
			0.3	**0.74**	**0.743**	24	1.13
	400	20	0	**0.02**	**0.000**	0	0.00
			0.1	**0.76**	**0.498**	9	1.23
			0.2	**1.17**	**0.621**	12	1.28
			0.3	**2.10**	**0.689**	24	1.22
	600	30	0	**0.04**	**0.000**	0	0.00
			0.1	**1.42**	**0.459**	8	1.26
			0.2	**2.28**	**0.581**	12	1.31
			0.3	**4.29**	**0.651**	23	1.28
	800	40	0	**0.08**	**0.000**	0	0.00
			0.1	**2.20**	**0.434**	7	1.29
			0.2	**3.61**	**0.555**	10	1.45
			0.3	7.09	**0.625**	23	1.33

By substituting $\mathcal{M}_{\leq r}$ for $\mathcal{M}_r$, the stopping criterion for AALM (i.e., formula (19) in the original article [4]) should be replaced by

$$\begin{cases} \|X^k - Y^k\| < \varepsilon, \\ \|\mathcal{P}_{T_{X^k}\mathcal{M}_+}\left(\frac{1}{2}(X^k - A) - \Lambda^k\right)\| < \varepsilon, \\ \left|\|\frac{1}{2}(Y^k - A) + \Lambda^k\|^2 - \mathrm{dist}(-\frac{1}{2}(Y^k - A) - \Lambda^k, T_{Y^k}\mathcal{M}_{\leq r})^2\right| < \varepsilon. \end{cases}$$

Notice that both projections onto $\mathcal{M}_r$ and $\mathcal{M}_{\leq r}$ can be obtained by SVD. The only difference between solving the AL subproblem over $\mathcal{M}_r$ and $\mathcal{M}_{\leq r}$ occurs when the rank of iteration point is less than r. We have retested numerical experiments listed in [4]. Rank-

Table 2 Correction to Table 2 in [4]. Average results over 10 tests on synthetic data with fixed n

m	n	r	σ^2	AALM			
				CPU	RSE	O-ITER	I-ITER
500	200	10	0	**0.00**	**0.000**	0	0.00
			0.1	**0.13**	0.553	10	1.11
			0.2	**0.18**	0.674	10	1.43
			0.3	2.29	0.737	15	10.41
	400	20	0	**0.01**	**0.000**	0	0.00
			0.1	0.41	**0.480**	8	1.04
			0.2	**0.63**	**0.601**	9	1.45
			0.3	1.97	**0.669**	14	2.51
	600	30	0	**0.02**	**0.000**	0	0.00
			0.1	**0.56**	**0.438**	7	1.03
			0.2	**0.91**	**0.556**	8	1.35
			0.3	2.90	**0.625**	16	2.33
	800	40	0	**0.02**	**0.000**	0	0.00
			0.1	**0.62**	**0.410**	6	1.00
			0.2	**1.05**	**0.526**	8	1.33
			0.3	**1.88**	**0.595**	15	1.28
1000	200	10	0	**0.01**	**0.000**	0	0.00
			0.1	**0.34**	**0.561**	8	1.38
			0.2	**0.50**	**0.683**	11	1.43
			0.3	11.76	**0.746**	18	23.89
	400	20	0	**0.02**	**0.000**	0	0.00
			0.1	**0.68**	**0.493**	8	1.08
			0.2	**1.17**	**0.617**	10	1.47
			0.3	7.15	**0.685**	14	5.56
	600	30	0	0.04	**0.000**	0	0.00
			0.1	**1.21**	0.465	8	1.03
			0.2	**2.04**	0.587	9	1.47
			0.3	**3.02**	0.657	13	1.53
	800	40	0	**0.07**	**0.000**	0	0.00
			0.1	**1.74**	**0.434**	7	1.06
			0.2	**3.04**	**0.555**	9	1.39
			0.3	9.55	**0.625**	13	2.54

deficient has not been observed when the rank of the initial point equals r. Hence, the performance of the revised algorithm is similar to that in [4]. The revised numerical results are given in Figs. 1, 2, 3, and Tables 1, 2, 3, 4 corresponding to the update in the figures and tables of [4].

Table 3 Correction to Table 4 in [4]. Average results for Face Datasets with different *r*

Datasets	YaleB							
r	75		90		105		120	
alg.	CPU	RSE	CPU	RSE	CPU	RSE	CPU	RSE
AALM	5.65	**0.078**	**4.30**	**0.068**	**3.97**	**0.060**	**3.60**	**0.054**
Datasets	ORL							
r	20		30		40		50	
alg.	CPU	RSE	CPU	RSE	CPU	RSE	CPU	RSE
AALM	6.55	**0.176**	**5.61**	**0.159**	**4.40**	**0.147**	**3.82**	**0.138**
Datasets	CBCL							
r	19		29		39		49	
alg.	CPU	RSE	CPU	RSE	CPU	RSE	CPU	RSE
AALM	1.48	**0.120**	**1.34**	**0.099**	**1.11**	**0.085**	**1.24**	**0.074**
Datasets	UMsit							
r	20		40		60		80	
alg.	CPU	RSE	CPU	RSE	CPU	RSE	CPU	RSE
AALM	23.48	**0.187**	15.82	**0.144**	**8.80**	**0.121**	**7.83**	**0.106**
Datasets	Olivetti							
r	10		20		30		40	
alg.	CPU	RSE	CPU	RSE	CPU	RSE	CPU	RSE
AALM	2.72	**0.143**	2.48	**0.118**	**2.00**	**0.104**	**1.83**	**0.094**
Datasets	Frey							
r	20		30		40		50	
alg.	CPU	RSE	CPU	RSE	CPU	RSE	CPU	RSE
AALM	1.45	**0.082**	**1.27**	**0.071**	**1.10**	**0.064**	**0.86**	**0.059**

Table 4 Correction to Table 5 in [4]. Denoising results for face Datasets

Datasets		YaleB ($r = 30$)			ORL ($r = 40$)			UMist($r = 60$)			Olivetti ($r = 30$)		
σ	Algs.	CPU	RSE	PSNR	CPU	RSE	PSNR	CPU	RSE	PSNR	CPU	RSE	PSNR
20	AALM	98.17	**0.14**	**24.80**	**5.26**	0.16	**22.44**	63.65	**0.14**	25.18	**1.94**	**0.11**	**23.92**
30	AALM	113.73	**0.15**	**24.31**	**6.18**	0.17	21.86	51.51	**0.16**	24.10	**2.18**	0.12	**23.25**
40	AALM	138.02	**0.16**	**23.67**	**6.77**	0.18	21.14	52.05	**0.18**	22.92	**2.73**	0.13	22.41

References

1. Attouch, H., Bolte, J., Redont, P., Soubeyran, A.: Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-łojasiewicz inequality. Math. Oper. Res. **35**(2), 438–457 (2010)
2. Cason, T., Absil, P., Van Dooren, P.: Iterative methods for low rank approximation of graph similarity matrices. Linear Algebra Appl. **438**(4), 1863–1882 (2013)
3. Schneider, R., Uschmajew, A.: Convergence results for projected line-search methods on varieties of low-rank matrices via lojasiewicz inequality. SIAM J. Optim. **25**(1), 622–646 (2015)
4. Zhu, H., Ng, M., Song, G.: Augmented Lagrangian methods for nonnegative low-rank matrix approximation. J. Sci. Comput. **88** (2021)