SUPPLEMENTARY MATERIAL

Appendix: Convergence analysis

In this section, we prove all the theorems in Section 5.

1. Proof of Theorem 1

Proof (i) If $J(A_{(N)}^{k+1}) \le J(A_{(N)}^k) - \rho_t \|A_{(N)}^{k+1} - Y_{(N)}^k\|_F^2$, then the Theorem 1(i) is true. If not, since Eq. (10) and Proposition 4, we know that

$$H(A_1^{k+1}, \{A_i^k\}_{i=2}^N) \le \langle \nabla_{A_1} H(A_{(N)}^k), A_1^{k+1} - A_1^k \rangle + \frac{L_{\nabla_{A_1^k} H}}{2} \|A_1^{k+1} - A_1^k\|_F^2 + H(A_{(N)}^k). \tag{15}$$

From Eq. (12), we can also obtain

$$F_1(A_1^k) \ge F_1(A_1^{k+1}) + \frac{1}{2\sigma_1^k} \|A_1^{k+1} - A_1^k\|_F^2 + \langle \nabla_{A_1} H(A_{(N)}^k), A_1^{k+1} - A_1^k \rangle. \tag{16}$$

Then sum of the Eq. (15) and Eq. (16), we have

$$H(A_1^{k+1}, \{A_i^k\}_{i=2}^N) + F_1(A_1^{k+1}) \le H(A_{(N)}^k) + F_1(A_1^k) - \rho \|A_1^{k+1} - A_1^k\|_F^2, \tag{17}$$

where $\rho = \min(\frac{1}{2\sigma_1^k} - \frac{L_{\nabla_{A_1^k} H}}{2})$. Assuming Theorem 1(i) holds when $n = N-1$, i.e.,

$$H(A_{(N-1)}^{k+1}, A_N^k) + \sum_{i=1}^{N-1} F_i(A_i^{k+1}) \le H(A_{(N)}^k) + \sum_{i=1}^{N-1} F_i(A_i^k) - \rho \|A_{(N-1)}^{k+1} - A_{(N-1)}^k\|_F^2, \tag{18}$$

where $\rho = \min(\{\frac{1}{2\sigma_i^k} - \frac{L_{\nabla_{A_i^k} H}}{2}\}_{i=1}^{N-1})$. Similarly, from Eq. (12), we infer

$$F_N(A_N^k) \ge F_N(A_N^{k+1}) + \frac{1}{2\sigma_N^k} \|A_N^{k+1} - A_N^k\|_F^2 + \langle \nabla_{A_N} H(A_{(N-1)}^{k+1}, A_N^k), A_N^{k+1} - A_N^k \rangle. \tag{19}$$

From Eq. (10) and Proposition 4, we also have

$$H(A_{(N)}^{k+1}) \le H(A_{(N-1)}^{k+1}, A_N^k) + \langle \nabla_{A_N} H(A_{(N-1)}^{k+1}, A_N^k), A_N^{k+1} - A_N^k \rangle + \frac{L_{\nabla_{A_N^k} H}}{2} \|A_N^{k+1} - A_N^k\|_F^2. \tag{20}$$

Then sum of the Eq. (18), Eq. (19) and Eq. (20), we have

$$J(A_{(N)}^{k+1}) \le J(A_{(N)}^k) - \rho \|z^{k+1} - c^k\|_F^2, \tag{21}$$

where $\rho = \min(\{\frac{1}{2\sigma_i^k} - \frac{L_{\nabla_{A_i^k} H}}{2}\}_{i=1}^N)$. This shows the Theorem 1(i) is true.

(ii) We know that $\frac{\alpha}{2}\operatorname{Tr}((A_N)^T L A_N)$ in Eq. (6) is nonnegative. Hence, Eq. (6) is nonnegative. From Eq. (21), we have

$$\rho \|z^{k+1} - c^k\|_F^2 \le J(A_{(N)}^k) - J(A_{(N)}^{k+1}).$$

Sum of both side, since Eq. (6) is nonnegative, thus we have

$$\begin{aligned} \rho \sum_{k=1}^{\infty} \|z^{k+1} - c^k\|_F^2 &\le \sum_{k=1}^{\infty} (J(A_{(N)}^k) - J(A_{(N)}^{k+1})) \\ &= J(A_{(N)}^1) - \inf J \\ &< \infty. \end{aligned}$$

It follows that

$$\lim_{k\to\infty} \|z^{k+1} - c^k\|_F^2 = 0.$$

2. Proof of Lemma 1

Proof By Proposition 2 and Proposition 3, Eq. (12) follows that

$$0 \in \partial_{A_i} F_i(A_i^{k+1}) - \frac{1}{\sigma_i^k}(Y_i^k - A_i^{k+1}) + \nabla_{A_i} H(\{A_j^{k+1}\}_{j=1}^{i-1}, Y_i^k, \{A_j^k\}_{j=i+1}^N). \tag{22}$$

Thus we also have

$$\begin{aligned} f_i^{k+1} &= \nabla_{A_i} H(\{A_j^{k+1}\}_{j=1}^{i}, \{A_j^k\}_{j=i+1}^N) - \frac{1}{\sigma_i^k}(A_i^{k+1} - Y_i^k) \\ &- \nabla_{A_i} H(\{A_j^{k+1}\}_{j=1}^{i-1}, Y_i^k, \{A_j^k\}_{j=i+1}^N) \\ &\in \nabla_{A_i} H(\{A_j^{k+1}\}_{j=1}^{i-1}, A_i^{k+1}, \{A_j^k\}_{j=i+1}^N) \\ &+ \partial_{A_i}(\sum_{j=1}^{i} F_j(A_j^{k+1}) + \sum_{j=i+1}^{N} F_j(A_j^k)) \\ &= \partial_{A_i} J(\{A_j^{k+1}\}_{j=1}^{i-1}, A_i^{k+1}, \{A_j^k\}_{j=i+1}^N). \end{aligned} \tag{23}$$

Since Eq. (10), we have

$$\begin{aligned} \|f_i^{k+1}\|_F &= \|\nabla_{A_i} H(\{A_j^{k+1}\}_{j=1}^{i}, \{A_j^k\}_{j=i+1}^N) + \frac{1}{\sigma_i^k}(Y_i^k - A_i^{k+1}) \\ &- \nabla_{A_i} H(\{A_j^{k+1}\}_{j=1}^{i-1}, Y_i^k, \{A_j^k\}_{j=i+1}^N)\|_F, \\ &\le \|\nabla_{A_i} H(\{A_j^{k+1}\}_{j=1}^{i}, \{A_j^k\}_{j=i+1}^N) \\ &- \nabla_{A_i} H(\{A_j^{k+1}\}_{j=1}^{i-1}, Y_i^k, \{A_j^k\}_{j=i+1}^N)\|_F \\ &+ \|\frac{1}{\sigma_i^k}(A_i^{k+1} - Y_i^k)\|_F \\ &\le (L_{\nabla_{A_i^k} H} + \frac{1}{\sigma_i^k}) \|A_i^{k+1} - Y_i^k\|_F. \end{aligned} \tag{24}$$

Thus we infer

$$\|\{f_i^{k+1}\}_{i=1}^N\|_F \le \rho_b \|z^{k+1} - c^k\|_F, \tag{25}$$

where $\rho_b = \max(\{\frac{1}{\sigma_i^k} + L_{\nabla_{A_i^k} H}\}_{i=1}^N)$.

3. Proof of Theorem 2

Proof (i) Observe that z' can be viewed as an intersection of compact sets

$$z' = \cap_{q\in\mathbb{N}}\overline{\cup_{k\geq q}\{z^k\}}.$$

From proposition 1, z' is compact.

(ii) $\forall \overline{z} \in z'$, there exists a subsequence z^{k_j} such that

$$\lim_{j\to\infty} z^{k_j} = \overline{z}.$$

Let

$$F(z^{k_j}) = \sum_{i=1}^{N} F_i(A_i^{k_j}). \tag{26}$$

Since F_i is lower semicontinuous [1], from Definition 3 and Eq. (26), we obtain that

$$\lim_{j\to\infty} \inf F(z^{k_j}) \geq F(\overline{z}). \tag{27}$$

Choosing $k = k_j - 1$, from Eq. (12), we infer

$$\begin{aligned}\lim_{j\to\infty} \sup F_i(A_i^{k_j}) \leq{}& \lim_{j\to\infty} \sup (F_i(\overline{A_i}) + \frac{1}{2\sigma_i^k}\|\overline{A_i} - Y_i^k\|_F^2 \\ &+ \langle \nabla_{A_i} H(\{A_n^{k+1}\}_{n=1}^{i-1}, Y_i^k, \{A_n^k\}_{n=i+1}^{N}), \overline{A_i} - Y_i^k\rangle),\end{aligned} \tag{28}$$

where $\lim_{j\to\infty} A_i^{k_j} = \overline{A_i}$. Since $\lim_{j\to\infty} z^{k_j} = \overline{z}$ and $\lim_{j\to\infty} \|z^{k_j} - c^{k_j-1}\|_F = 0$ (Theorem 1(ii)), we can get that

$$\begin{aligned}\lim_{j\to\infty} \|\overline{A_i} - Y_i^k\|_F &\leq \lim_{j\to\infty} \|\overline{A_i} - A_i^{k_j}\|_F + \lim_{j\to\infty} \|A_i^{k_j} - Y_i^{k_j-1}\|_F \\ &= 0,\end{aligned} \tag{29}$$

From Eq. (28) and Eq. (29), we infer

$$\lim_{j\to\infty} \sup F_i(A_i^{k_j}) \leq \lim_{j\to\infty} \sup F_i(\overline{A_i}), \tag{30}$$

where $\lim_{j\to\infty} A_i^{k_j} = \overline{A_i}$. Therefore, from Eq. (26) and Eq. (30), we have

$$\lim_{j\to\infty} \sup F(z^{k_j}) \leq \lim_{j\to\infty} \sup F(\overline{z}). \tag{31}$$

From Eq. (26), Eq. (31) and Eq. (27), we infer

$$\lim_{j\to\infty} F(z^{k_j}) = F(\overline{z}). \tag{32}$$

Since Theorem 1, we have

$$\lim_{j\to\infty} H(z^{k_j}) = H(\overline{z}). \tag{33}$$

Thus from Eq. (32) and Eq. (33), we infer

$$\begin{aligned}\lim_{j\to\infty} H(z^{k_j}) + \lim_{j\to\infty} F(z^{k_j}) &= \lim_{j\to\infty} (H(z^{k_j}) + F(z^{k_j})) \\ &= \lim_{j\to\infty} J(z^{k_j}) \\ &= J(\overline{z}).\end{aligned}$$

This means J is constant on z'.

(iii) From Eq. (23), we know that

$$f_i^{k+1} \in \partial_{A_i} J(\{A_j^{k+1}\}_{j=1}^{i}, \{A_j^k\}_{j=i+1}^{N}), \tag{34}$$

where the definitions of f_i^{k+1} is the same as in Eq. (14). From Theorem 1(ii) and Lemma 1, we infer

$$\begin{aligned}\lim_{k\to\infty} (\|\{f_i^{k+1}\}_{i=1}^{N}\|_F &\leq \lim_{k\to\infty} \rho_b \|z^{k+1} - c^k\|_F \\ &= 0.\end{aligned} \tag{35}$$

Let

$$z_i^k = A_i^k. \tag{36}$$

From Theorem 1(i), we know that

$$\begin{aligned}\lim_{k\to\infty} J(\{z_i^{k+1}\}_{i=1}^{j-1}, z_j^{k+1}, \{z_i^k\}_{i=j+1}^{N}) &= \lim_{k\to\infty} J(z^k) \\ &= J(z^*).\end{aligned} \tag{37}$$

Therefore, from Eq. (34), Eq. (35), Eq. (36) and Eq. (37), we have

$$\begin{aligned}0 &\in \partial_{z_j} \lim_{k\to\infty} J(\{z_i^{k+1}\}_{i=1}^{j-1}, z_j^{k+1}, \{z_i^k\}_{i=j+1}^{N}) \\ &= \partial_{z_j} J(z^*).\end{aligned}$$

This means $0 \in \partial J(z^*)$.

4. Proof of Theorem 3

Proof From Eq. (10), we know that $\frac{\alpha}{2}\,\mathrm{Tr}((A_N)^T L A_N)$ is an infinitely differentiable function, and the norm of its derivatives of any order is also continuous, so $\frac{\alpha}{2}\,\mathrm{Tr}((A_N)^T L A_N)$ is a real analytic function (real analytic functions are all KŁ functions [2]). Since the Frobenius norm, ℓ_p-norm and Eq. (7) are also all KŁ functions [1], it follows that Eq. (6) is a KŁ function. Therefore, from Definition 6, there exists a concave function ϕ so that

$$\phi'(J(z^k) - J(\overline{z}))dist(0, \partial J(z^k)) \geq 1. \tag{38}$$

From ϕ is the convex function, we have

$$\begin{aligned}\phi(J(z^{k+1}) - J(\overline{z})) \leq{}& \phi(J(z^k) - J(\overline{z})) \\ &+ \phi'(J(z^k) - J(\overline{z}))(J(z^{k+1}) - J(z^k)).\end{aligned} \tag{39}$$

From Lemma 1 and Theorem 2, we infer

$$dist(0, \partial J(z^k)) \leq \rho_b \|z^k - c^{k-1}\|_F. \tag{40}$$

Since the Eq. (38) and Eq. (40), we infer

$$\begin{aligned}\phi^{'}(J(z^k) - J(\overline{z})) &\geq \frac{1}{dist(0, \partial J(z^k))} \\ &\geq \frac{1}{\rho_b \|z^k - c^{k-1}\|_F}.\end{aligned} \tag{41}$$

Let $J(k) = J(z^k) - J(\overline{z})$, from Eq. (39), Eq. (40), Eq. (41) and Theorem 1(i), we have

$$\begin{aligned}\phi(J(k)) - \phi(J(k+1)) &\geq \phi^{'}(J(k))(J(k) - J(k+1)) \\ &\geq \frac{J(k) - J(k+1)}{\rho_b \|z^k - c^{k-1}\|_F} \\ &\geq \frac{\rho \|z^{k+1} - c^k\|_F^2}{\rho_b \|z^k - c^{k-1}\|_F}.\end{aligned}$$

Define C=$\frac{\rho}{\rho_b}$, C is a constant, so we infer

$\|z^{k+1} - c^k\|_F^2 \leq C(\phi(J(k)) - \phi(J(k+1)))\|z^k - c^{k-1}\|_F$.

Using the fact that $2ab \leq a^2 + b^2$

$2\|z^{k+1} - c^k\|_F \leq C(\phi(J(k)) - \phi(J(k+1))) + \|z^k - c^{k-1}\|_F$.

Sum both sides

$$\begin{aligned}2\sum_{k=l+1}^{K} \|z^{k+1} - c^k\|_F \leq& \sum_{k=l+1}^{K} \|z^k - c^{k-1}\|_F \\ &+ C(\phi(J(l+1)) - \phi(J(K+1))) \\ &= C(\phi(J(l+1)) - \phi(J(K+1))) \\ &+ \|z^{l+1} - c^l\|_F + \sum_{k=l+1}^{K} \|z^{k+1} - c^k\|_F.\end{aligned} \tag{42}$$

From Eq. (42), we can get that

$$\begin{aligned}\lim_{K\to\infty} \sum_{k=l+1}^{K} \|z^{k+1} - c^k\|_F \leq& \|z^{l+1} - c^l\|_F + C\phi(J(l+1)) \\ &- \lim_{K\to\infty} C\phi(J(K+1)) \\ &< \infty.\end{aligned}$$

Thus we have

$$\sum_{k=l+1}^{\infty} \|z^{k+1} - c^k\|_F < \infty. \tag{43}$$

From Algorithm 1, no matter $c^k = z^k + \beta_k(z^k - z^{k-1})$ or $c^k = z^k$, we always have

$$\|z^{k+1} - z^k\|_F - \beta_k \|z^k - z^{k-1}\|_F \leq \|z^{k+1} - c^k\|_F. \tag{44}$$

From Eq. (43) and Eq. (44), we know that

$$\begin{aligned}\sum_{k=l+1}^{\infty} (\|z^{k+1} - z^k\|_F - \beta_k \|z^k - z^{k-1}\|_F) \leq& \sum_{k=l+1}^{\infty} \|z^{k+1} - c^k\|_F \\ &< \infty.\end{aligned}$$

Since

$$\begin{aligned}&\|z^{k+1} - z^k\|_F - \beta_{max}\|z^k - z^{k-1}\|_F \leq \\ &\|z^{k+1} - z^k\|_F - \beta_k\|z^k - z^{k-1}\|_F,\end{aligned}$$

we also have

$$\begin{aligned}&\sum_{k=l+1}^{\infty} (\|z^{k+1} - z^k\|_F - \beta_{max}\|z^k - z^{k-1}\|_F) \\ &\leq \sum_{k=l+1}^{\infty} (\|z^{k+1} - z^k\|_F - \beta_k\|z^k - z^{k-1}\|_F) \\ &< \infty,\end{aligned} \tag{45}$$

and

$$\begin{aligned}&\sum_{k=l+1}^{\infty} \|z^{k+1} - z^k\|_F - \sum_{k=l+1}^{\infty} \beta_{max}\|z^k - z^{k-1}\|_F \\ &= \sum_{k=l+1}^{\infty} (\|z^{k+1} - z^k\|_F - \beta_{max}\|z^{k+1} - z^k\|_F) + \beta_{max}\|z^{l+1} - z^l\|_F \\ &= \sum_{k=l+1}^{\infty} (1 - \beta_{max})\|z^{k+1} - z^k\|_F + \beta_{max}\|z^{l+1} - z^l\|_F.\end{aligned} \tag{46}$$

Let $s^k = \|z^{k+1} - z^k\|_F$, from Eq. (45) and Eq. (46), we have

$$\begin{aligned}\sum_{k=l+1}^{\infty} (1 - \beta_{max})s^k - \beta_{max}s^l &= \sum_{k=l}^{\infty} (s^{k+1} - \beta_{max}s^k) \\ &\leq \sum_{k=l+1}^{\infty} \|z^{k+1} - c^k\|_F. \\ &< \infty.\end{aligned}$$

Thus we infer

$$\sum_{k=l+1}^{\infty} (1 - \beta_{max})s^k < \infty. \tag{47}$$

From $(1 - \beta_{max})$ is a positive constant and Eq. (47), we infer

$$\sum_{k=l+1}^{\infty} \|z^{k+1} - z^k\|_F = \sum_{k=l+1}^{\infty} s^k < \infty.$$

Thus we have

$$\sum_{k=0}^{\infty} \|z^{k+1} - z^k\|_F = \sum_{k=0}^{\infty} s^k < \infty.$$

This shows that

$$\begin{aligned}\lim_{K\to\infty}\|z^{K+p}-z^{K}\|_F &= \lim_{K\to\infty}\|\sum_{k=K+1}^{K+p}(z^{k+1}-z^{k})\|_F \\ &\leq \lim_{K\to\infty}\sum_{k=K+1}^{K+p}\|z^{k+1}-z^{k}\|_F \\ &= \lim_{K\to\infty}\sum_{k=K+1}^{\infty}s^{k} \\ &= 0. \end{aligned} \tag{48}$$

This means the Theorem 3 is true.

5. Proof of Theorem 4

Proof We use the MTTKRP technology [3] to calculate $X^{(n)}B_n$. Therefore, the computational complexity of $X^{(n)}B_n$ reaches

$$R(\prod_{i=1}^{N} I_i). \tag{49}$$

Similarly, $(B_n)^T B_n$ can be calculated efficiently by $((A_N)^T A_N) * ...((A_{n+1})^T A_{n+1}) * ((A_{n-1})^T A_{n-1}).... * ((A_1)^T A_1)$, where $*$ represents the Hadamard product that is the elementwise matrix product. Therefore, the computational complexity of $(B_n)^T B_n$ is

$$R^2 \sum_{j=1, j\neq n}^{N} I_i. \tag{50}$$

Therefore, from Eq. (49) and Eq. (50), we know that the computational cost of $\nabla_{A_i} H(A_{(N)})$ is

$$\mathcal{O}(R(\prod_{i=1}^{N} I_i) + R^2 \sum_{j=1, j\neq i}^{N} I_i). \tag{51}$$

The cost of computing the Lipschitz constant, projection to nonnegative, and tensor unfolding is negligible compared to the cost of computing partial gradient $\nabla_{A_i} H(A_{(N)})$. Similarly, the cost of the sparsity projection is also negligible compared to the cost of computing the $\nabla_{A_i} H(A_{(N)})$, as the time complexity of finding the largest s elements in an array of p elements is $\mathcal{O}(p + s log_2(s))$. Therefore, the time complexity of the Algorithm 1 in each iteration is approximately estimated as

$$\mathcal{O}(NR(\prod_{i=1}^{N} I_i) + R^2 \sum_{i=1}^{N}(\sum_{j=1, j\neq i}^{N} I_i)).$$

References

[1] J. Bolte, S. Sabach, and M. Teboulle, Proximal alternating linearized minimization for nonconvex and nonsmooth problems, *Math. Program.*, vol.146, no.1, pp.459–494, 2014.

[2] Y. Xu and W. Yin, A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion, *Siam. J. Imaging. Sci.*, vol.6, no.3, pp.1758–1789, 2013.

[3] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, *et al.*, Tensor decomposition for signal processing and machine learning, *IEEE Trans. Signal Process.*, vol.65, no.13, pp.3551–3582, 2017.