Supplementary Material

Section A: Derivation of Tensor Factor Tracking

Under the assumption that the tensor factors are either static or slowly varying (i.e. $\mathbf{D}_t \approx \mathbf{D}_{t-1}$) at time t, the corrupted entries of $\boldsymbol{\mathcal{X}}_t$ can be recovered by using the following rule:

$$[\hat{\boldsymbol{\mathcal{X}}}_t]_{i_1 i_2 \dots i_N} = \begin{cases} \big[\boldsymbol{\mathcal{H}}_{t-1} \times_{N+1} \mathbf{u}_t^\top\big]_{i_1 i_2 \dots i_N}, & \text{if } [\boldsymbol{\mathcal{P}}_t]_{i_1 i_2 \dots i_N} = 0 \\ [\boldsymbol{\mathcal{X}}_t]_{i_1 i_2 \dots i_N}, & \text{if } [\boldsymbol{\mathcal{P}}_t]_{i_1 i_2 \dots i_N} = 1. \end{cases}$$

With a set of full estimated slices $\{\hat{\boldsymbol{\mathcal{X}}}_k\}_{k=1}^t$, we can consider an alternative of (15) in the main manuscript as follows:

$$\mathbf{U}_t^{(n)} = \underset{\mathbf{U}^{(n)}}{\operatorname{argmin}}\; g_t\big(\mathbf{U}^{(n)}, .\big), \quad \text{with} \tag{A1}$$
$$g_t\big(\mathbf{U}^{(n)}, .\big) = \frac{1}{L_t} \sum_{k=t-L_t+1}^{t} \lambda^{t-k} \Big\| \hat{\underline{\mathbf{X}}}_k^{(n)} - \mathbf{U}^{(n)}\big(\mathbf{W}_k^{(n)}\big)^\top \Big\|_F^2,$$

where $\hat{\underline{\mathbf{X}}}_k^{(n)}$ is the mode-n unfolding matrix of $\hat{\boldsymbol{\mathcal{X}}}_k$. The only difference from (15) is that we remove the binary mask $\boldsymbol{\mathcal{P}}_k$ out of the objective function, and replace it with $\hat{\boldsymbol{\mathcal{X}}}_k$.

Accordingly, the minimization (18) can be rewritten as

$$\mathbf{u}_{t,m}^{(n)} = \underset{\mathbf{u}_m^{(n)}}{\operatorname{argmin}}\, g_t\big(\mathbf{u}_m^{(n)}, .\big), \text{with} \tag{A2}$$
$$g_t\big(\mathbf{u}_m^{(n)}, .\big) = \frac{1}{L_t} \sum_{k=t-L_t+1}^{t} \lambda^{t-k} \Big\| \big(\hat{\underline{\mathbf{x}}}_{k,m}^{(n)}\big)^\top - \mathbf{W}_k^{(n)}\big(\mathbf{u}_m^{(n)}\big)^\top \Big\|_2^2,$$

The recursive rule for updating $\mathbf{S}_{t,m}^{(n)}$ in (22) becomes

$$\mathbf{S}_{t,m}^{(n)} = \lambda \mathbf{S}_{t-1,m}^{(n)} + \big(\widetilde{\mathbf{W}}_t^{(n)}\big)^\top \widetilde{\mathbf{W}}_t^{(n)}. \tag{A3}$$

Clearly, (A3) is the same for all m. This leads to a simplified updating rule for $\mathbf{S}_{t,m}^{(n)}$ and $\mathbf{V}_{t,m}^{(n)}$ as follows

$$\mathbf{S}_{t,m}^{(n)} \triangleq \mathbf{S}_t^{(n)} = \lambda \mathbf{S}_{t-1}^{(n)} + \big(\widetilde{\mathbf{W}}_t^{(n)}\big)^\top \widetilde{\mathbf{W}}_t^{(n)}, \tag{A4}$$
$$\mathbf{V}_{t,m}^{(n)} \triangleq \mathbf{V}_t^{(n)} = \big(\mathbf{S}_t^{(n)}\big)^{-1} \big(\widetilde{\mathbf{W}}_t^{(n)}\big)^\top. \tag{A5}$$

As a result, the updating rule of (23) can be modified as

$$\mathbf{U}_t^{(n)} = \mathbf{U}_{t-1}^{(n)} + \big(\widetilde{\hat{\mathbf{X}}}_t^{(n)} - \mathbf{U}_{t-1}^{(n)}\big(\widetilde{\mathbf{W}}_t^{(n)}\big)^\top\big)\big(\mathbf{V}_t^{(n)}\big)^\top. \tag{A6}$$

where $\widetilde{\hat{\mathbf{X}}}_t^{(n)} = [\hat{\mathbf{X}}_t^{(n)} \;\; \hat{\mathbf{X}}_{t-L_t,m}^{(n)}]$. It should be noted that when the (i,j)-th entry of $\mathbf{X}_t^{(n)}$ is missing or affected by outliers, $[\hat{\mathbf{X}}_t^{(n)} - \mathbf{U}_{t-1}^{(n)}(\mathbf{W}_t^{(n)})^\top]_{i,j} = 0$. To sum up, the tensor factor $\mathbf{U}_t^{(n)}$ can be updated via

$$\mathbf{U}_t^{(n)} = \mathbf{U}_{t-1}^{(n)} + \underline{\widetilde{\mathbf{P}}}_t^{(n)} \circledast \big(\widetilde{\mathbf{X}}_t^{(n)} - \mathbf{U}_{t-1}^{(n)}\big(\widetilde{\mathbf{W}}_t^{(n)}\big)^\top\big)\big(\mathbf{V}_t^{(n)}\big)^\top. \tag{A7}$$

We exploit an interesting fact from the alternative (A2) that if the column $\underline{\mathbf{x}}_{t,m}^{(n)}$ is completely corrupted by outliers or missing data, then $\mathbf{u}_{t,m}^{(n)} = \operatorname{argmin} g_t\big(\mathbf{u}_m^{(n)}, .\big) = \mathbf{u}_{t-1,m}^{(n)}$ when we use the exponential window, i.e. $L_t = t$. In such a case, the modified tracker seems to ignore the m-th row of $\mathbf{U}_t^{(n)}$ which is consistent with the original update rule (23). In fact, we can rewrite (A2) as follows:

$$t g_t\big(\mathbf{u}_m^{(n)}, .\big) = t\lambda g_{t-1}\big(\mathbf{u}_m^{(n)}, .\big) + \Big\| \mathbf{W}_t^{(n)}\big(\mathbf{u}_{t-1,m}^{(n)} - \mathbf{u}_m^{(n)}\big)^\top \Big\|_2^2. \tag{A8}$$

It is known that $\mathbf{u}_{t-1,m}^{(n)} = \operatorname{argmin} g_{t-1}(\mathbf{u}_m^{(n)}, .)$ and the second term of (A8) is equal to zero when $\mathbf{u}_m^{(n)} = \mathbf{u}_{t-1,m}^{(n)}$. Accordingly, (A8) is minimized at $\mathbf{u}_{t-1,m}^{(n)}$.

Section B: RACP as Second-Order Stochastic Gradient Descent

Without loss of generality, we can reshape $\mathbf{U}^{(n)}$ into a column vector $\mathbf{u}^{(n)} = \big[\mathbf{u}_1^{(n)}, \mathbf{u}_2^{(n)}, \dots, \mathbf{u}_{I_n}^{(n)}\big]^\top$ where $\mathbf{u}_m^{(n)}$ is the m-th row of $\mathbf{U}^{(n)}$. Accordingly, we can rewrite $\tilde{f}_t(\mathbf{U}^{(n)}, .)$ as follows

$$\tilde{f}_t\big(\mathbf{u}^{(n)}, .\big) = \frac{1}{L_t} \sum_{k=t-L_t+1}^{t} \lambda^{t-k} \tilde{\ell}_k\big(\mathbf{u}^{(n)}, .\big), \text{ where} \tag{B1}$$
$$\tilde{\ell}_k\big(\mathbf{u}^{(n)}, .\big) = \big\|\underline{\mathbf{O}}_k^{(n)}\big\|_1 + \frac{\rho}{2}\Big\| \mathbf{P}_k^{(n)}\big(\hat{\mathbf{x}}_k^{(n)} - \mathbf{W}_k^{(n)}\mathbf{u}^{(n)}\big)\Big\|_2^2, \tag{B2}$$

where $\hat{\mathbf{x}}_k^{(n)}$ is the vectorized form of $\widehat{\underline{\mathbf{X}}}_k^{(n)}$ arranged by rows and the mask $\mathbf{P}_k^{(n)} = \operatorname{diag}\big(\underline{\mathbf{P}}_{k,1}^{(n)}, \underline{\mathbf{P}}_{k,2}^{(n)}, \dots, \underline{\mathbf{P}}_{k,I_n}^{(n)}\big)$. Setting $\partial \tilde{f}/\partial \mathbf{u}^{(n)}$ to zero yields

$$\sum_{k=t-L_t+1}^{t} \lambda^{t-k}\big(\mathbf{W}_k^{(n)}\big)^\top \mathbf{P}_k^{(n)} \Big[\hat{\mathbf{x}}_{k,1}^{(n)}, \hat{\mathbf{x}}_{k,2}^{(n)}, \dots, \hat{\mathbf{x}}_{k,I_n}^{(n)}\Big]^\top \tag{B3}$$
$$= \sum_{k=t-L_t+1}^{t} \lambda^{t-k}\big(\mathbf{W}_k^{(n)}\big)^\top \mathbf{P}_k^{(n)} \mathbf{W}_k^{(n)} \Big[\mathbf{u}_1^{(n)}, \mathbf{u}_2^{(n)}, \dots, \mathbf{u}_{I_n}^{(n)}\Big]^\top.$$

Breaking (B3) into I_n equations w.r.t each row $\mathbf{u}_m^{(n)}$ results in (19). It explains why we can decompose the minimization (16) into sub-problems for each row $\mathbf{u}_m^{(n)}$ of $\mathbf{U}^{(n)}$ as presented in Section III.A. The Hessian matrix of $\tilde{f}_t(\mathbf{u}^{(n)}, .)$ is then given by

$$\mathbf{H}\big(\mathbf{u}_{t-1}^{(n)}\big) = \frac{\rho}{L_t} \sum_{k=t-L_t+1}^{t} \lambda^{t-k}\big(\mathbf{W}_k^{(n)}\big)^\top \mathbf{P}_k^{(n)} \mathbf{W}_k^{(n)}. \tag{B4}$$

Accordingly, the update rule (23) can be rewritten as

$$\mathbf{u}_{t,m}^{(n)} = \mathbf{u}_{t-1,m}^{(n)} - \mathbf{H}\big(\mathbf{u}_{t-1,m}^{(n)}\big)^{-1} \frac{\partial \tilde{f}}{\partial \mathbf{u}_m^{(n)}}\bigg|_{\mathbf{u}=\mathbf{u}_{t-1}}, \tag{B5}$$

which is indeed a second-order stochastic gradient descent.

Section C: Proof of Proposition 1

1./ Boundedness: $\{\mathbf{D}_t, \boldsymbol{\mathcal{O}}_t, \mathbf{u}_t\}_{t=1}^\infty$ are uniformly bounded.

At each time $t > 0$, the outlier $\boldsymbol{\mathcal{O}}_t$ and the coefficient vector $\mathbf{u}_t$ are derived from the minimization (7) in the main manuscript. Accordingly, we always have

$$\tilde{\ell}\big(\mathbf{D}_{t-1}, \boldsymbol{\mathcal{P}}_t, \boldsymbol{\mathcal{X}}_t, \boldsymbol{\mathcal{O}}_t, \mathbf{u}_t\big) \le \tilde{\ell}\big(\mathbf{D}_{t-1}, \boldsymbol{\mathcal{P}}_t, \boldsymbol{\mathcal{X}}_t, \mathbf{0}, \mathbf{0}\big). \tag{C1}$$

It is therefore that

$$\|\boldsymbol{\mathcal{O}}_t\|_1 + \frac{\rho}{2}\big\|\boldsymbol{\mathcal{P}}_t \circledast (\boldsymbol{\mathcal{X}}_t - \boldsymbol{\mathcal{O}}_t - \boldsymbol{\mathcal{H}}_{t-1} \times_{N+1} \mathbf{u}_t)\big\|_F^2 \le \frac{\rho}{2}\|\boldsymbol{\mathcal{P}}_t \circledast \boldsymbol{\mathcal{X}}_t\|_F^2.$$

Due to the two facts that $\|\mathbf{M}\|_F + \|\mathbf{N}\|_F \ge \|\mathbf{M} - \mathbf{N}\|_F \ge \|\mathbf{M}\|_F - \|\mathbf{N}\|_F$, and $\|\mathbf{M}\|_F \le \|\mathbf{M}\|_1$ [1], we then obtain

$$\big\|\boldsymbol{\mathcal{O}}_t\big\|_F \le \big\|\boldsymbol{\mathcal{O}}_t\big\|_1 \le \frac{\rho}{2}\big\|\boldsymbol{\mathcal{P}}_t \circledast \boldsymbol{\mathcal{X}}_t\big\|_F^2 \le \frac{\rho}{2}M_x^2 < \infty, \tag{C2}$$
$$\big\|\mathbf{P}_t \mathbf{H}_{t-1}\mathbf{u}_t\big\|_2 \le 2\big\|\boldsymbol{\mathcal{P}}_t \circledast \boldsymbol{\mathcal{X}}_t\big\|_F + \big\|\boldsymbol{\mathcal{P}}_t \circledast \boldsymbol{\mathcal{O}}_t\big\|_F < \infty, \tag{C3}$$

where M_x is the upper bound of $\|\boldsymbol{\mathcal{X}}_t\|_F$ (see Assumption A1). Thanks to (C2), $\boldsymbol{\mathcal{O}}_t$ is uniformly bound.

We indicate the bound of the solution $\mathbf{u}_t$ and $\mathbf{D}_t = [\mathbf{U}_t^{(1)}, \mathbf{U}_t^{(2)}, \dots, \mathbf{U}_t^{(N)}]$ by using the mathematical induction.

We first recall that the proposed RACP algorithm begins with N full-rank matrices $\big\{\mathbf{U}_0^{(n)}\big\}_{n=1}^N$ and a set of matrices $\mathbf{S}_{0,m}^{(n)} = \delta_n \mathbf{I}, m = 1, 2, \dots, I_n$.

The base case: At $t = 1$, the matrix $\mathbf{H}_0 = \odot_{n=1}^N \mathbf{U}_0^{(n)}$ is then full rank, i.e., the null space of $\mathbf{H}_0$ admits only $\mathbf{0}$ as a vector. Accordingly, $\mathbf{u}_1$ is bounded, thanks to (C3).

To indicate the bound of $\mathbf{U}_1^{(n)}$ for $n = 1, 2, \dots, N$, we show that each row $\mathbf{u}_{1,m}^{(n)}$ of $\mathbf{U}_1^{(n)}$ is bounded. We first obtain the inequality

$$\big\|\mathbf{u}_{1,m}^{(n)}\big\|_2 \le \Big\|\underline{\mathbf{P}}_{1,m}^{(n)}\big(\big(\underline{\mathbf{x}}_{1,m}^{(n)}\big)^\top - \mathbf{W}_1^{(n)}\big(\mathbf{u}_{0,m}^{(n)}\big)^\top\big)\Big\|_2 \big\|\mathbf{V}_{1,m}^{(n)}\big\|_2 + \big\|\mathbf{u}_{0,m}^{(n)}\big\|_2.$$

In fact, three matrices $\mathbf{W}_{1,m}^{(n)}$, $\mathbf{S}_{1,m}^{(n)}$ and $\mathbf{V}_{1,m}^{(n)}$ for updating $\mathbf{u}_{1,m}^{(n)}$ are bounded due to the bound of $\big\{\mathbf{U}_0^{(n)}\big\}_{n=1}^N$. Accordingly, its right hand side is finite, thus $\mathbf{u}_{1,m}^{(n)}$ is bounded for all m. It implies that $\mathbf{U}_1^{(n)}$ is bounded.

The induction step: We assume that $\{\mathbf{U}_i^{(n)}\}_{i=1}^k$ generated by RACP are bounded at time $t = k > 1$, we will prove that at $t = k+1$, $\mathbf{U}_{k+1}^{(n)}$ is also bounded.

Since $\{\mathbf{U}_k^{(n)}\}_{n=1}^N$ are assumed to be bounded, $\mathbf{u}_{k+1}$ and $\mathbf{W}_{k+1,m}^{(n)}$ are then bounded. In parallel, we exploit that $\mathbf{S}_{k+1,m}^{(n)}$ can be expressed by $\mathbf{S}_{k+1,m}^{(n)} = \lambda \mathbf{S}_{k,m}^{(n)} + \sum_i \underline{p}_{k+1,m}^{(n)}(i)\mathbf{w}_i^\top \mathbf{w}_i$, where $\mathbf{w}_i$ is the i-th row of $\mathbf{W}_{k+1,m}^{(n)}$. Thanks to Woodbury matrix identity [2] and $\mathbf{S}_{0,m}^{(n)} = \delta \mathbf{I}$ with $\delta > 0$, we obtain $\mathbf{S}_{k+1,m}^{(n)} \succ \mathbf{0}$, i.e., $\mathbf{S}_{k+1,m}^{(n)}$ is nonsingular with the smallest eigenvalue $\sigma_{\min}\big(\mathbf{S}_{k+1,m}^{(n)}\big) \geq \delta > 0$. Thus $\mathbf{V}_{k+1,m}^{(n)}$ is always existent. For given $\mathbf{M} \succ \mathbf{0}$, we always have $\|\mathbf{M}\|_F \leq \sqrt{r}\|\mathbf{M}\|_2 = \sqrt{r}\sigma_{\max}(\mathbf{M})$, and $\big\|\mathbf{M}^{-1}\big\|_2 = \sigma_{\min}^{-1}(\mathbf{M})$ where $\sigma_{\max}(\mathbf{M})$ and $\sigma_{\min}(\mathbf{M})$ are the largest and smallest eigenvalue of $\mathbf{M}$ [1]. Accordingly, we derive $\|\mathbf{V}_{k+1,m}^{(n)}\|_F \leq \sqrt{r}/\delta < \infty$, i.e., $\mathbf{V}_{k+1,m}^{(n)}$ is bounded. As a result, $\mathbf{u}_{k+1,m}^{(n)}$ is bounded for all $m = 1, 2, \dots, I_n$. Thanks to the mathematical induction, we can conclude that the solution $\mathbf{U}_t^{(n)}$ generated by RACP is bounded for $t \geq 1$.

2./ Forward Monotonicity: $\tilde{f}_t(\mathbf{D}_{t-1}) \geq \tilde{f}_t(\mathbf{D}_t)$.

We have

$$\tilde{f}_t(\mathbf{D}_{t-1}) - \tilde{f}_t(\mathbf{D}_t) \tag{C4}$$

$$= \begin{cases} \sum_{n=1}^N \tilde{f}_t\big(\mathbf{U}_{t-1}^{(1)}, \dots, \mathbf{U}_{t-1}^{(n-1)}, \mathbf{U}_{t-1}^{(n)}, \dots, \mathbf{U}_{t-1}^{(N)}\big) \\ \quad - \tilde{f}_t\big(\mathbf{U}_{t-1}^{(1)}, \dots, \mathbf{U}_{t-1}^{(n-1)}, \mathbf{U}_t^{(i)}, \dots, \mathbf{U}_{t-1}^{(N)}\big) & \text{[Jacobi]} \\ \sum_{n=1}^N \tilde{f}_t\big(\mathbf{U}_t^{(1)}, \dots, \mathbf{U}_t^{(n-1)}, \mathbf{U}_{t-1}^{(n)}, \dots, \mathbf{U}_{t-1}^{(N)}\big) \\ \quad - \tilde{f}_t\big(\mathbf{U}_t^{(1)}, \dots, \mathbf{U}_t^{(n-1)}, \mathbf{U}_t^{(n)}, \dots, \mathbf{U}_{t-1}^{(N)}\big) & \text{[Gauss-Seidel]} \end{cases}$$

Recall that $\mathbf{U}_t^{(n)}$ is the minimizer of $\tilde{f}_t\big(\mathbf{U}_{t-1}^{(1)}, \dots, \mathbf{U}_{t-1}^{(n-1)}, \mathbf{U}, \mathbf{U}_{t-1}^{(n+1)}, \dots, \mathbf{U}_{t-1}^{(N)}\big)$ if using Jacobi scheme or $\tilde{f}_t\big(\mathbf{U}_t^{(1)}, \dots, \mathbf{U}_t^{(n-1)}, \mathbf{U}, \mathbf{U}_{t-1}^{(n+1)}, \dots, \mathbf{U}_{t-1}^{(N)}\big)$ if using Gauss-Seidel scheme. Therefore, we always have

$$\tilde{f}_t\big(\mathbf{U}_{t-1}^{(1)}, \dots, \mathbf{U}_{t-1}^{(n-1)}, \mathbf{U}_{t-1}^{(n)}, \dots, \mathbf{U}_{t-1}^{(N)}\big) \geq \tilde{f}_t\big(\mathbf{U}_{t-1}^{(1)}, \dots, \mathbf{U}_{t-1}^{(n-1)}, \mathbf{U}_t^{(i)}, \dots, \mathbf{U}_{t-1}^{(N)}\big) \quad \text{[Jacobi]} \tag{C5}$$

$$\tilde{f}_t\big(\mathbf{U}_{t-1}^{(1)}, \dots, \mathbf{U}_{t-1}^{(n-1)}, \mathbf{U}_{t-1}^{(n)}, \dots, \mathbf{U}_{t-1}^{(N)}\big) \geq \tilde{f}_t\big(\mathbf{U}_{t-1}^{(1)}, \dots, \mathbf{U}_{t-1}^{(n-1)}, \mathbf{U}_t^{(i)}, \dots, \mathbf{U}_{t-1}^{(N)}\big) \quad \text{[Gauss-Seidel]} \tag{C6}$$

As a result, $\tilde{f}_t(\mathbf{D}_{t-1}) \geq \tilde{f}_t(\mathbf{D}_t)$.

3./ Backward Monotonicity: $\tilde{f}_t(\mathbf{D}_t) \leq \tilde{f}_t(\mathbf{D}_{t+1})$.

Applying the similar arguments above, we obtain $\tilde{f}_t(\mathbf{D}_t) \leq \tilde{f}_t(\mathbf{D}_{t+1})$.

4./ Stability of Estimates: $\|\mathbf{D}_t - \mathbf{D}_{t-1}\|_F = \mathcal{O}(1/t)$.

We first prove that the surrogate $\tilde{f}_t(.)$ w.r.t. each factor is Lipschitz continuous. Since $\mathbf{U}_t^{(n)} = \operatorname{argmin} \tilde{f}_t(\mathbf{U}^{(n)}, .)$, we have $\tilde{f}_t(\mathbf{U}_t^{(n)}, .) \leq \tilde{f}_t(\mathbf{U}_{t-1}^{(n)}, .) \forall t$ and hence

$$\tilde{f}_{t-1}\big(\mathbf{U}_t^{(n)}, .\big) - \tilde{f}_{t-1}\big(\mathbf{U}_{t-1}^{(n)}, .\big) \leq \Big\{\tilde{f}_{t-1}\big(\mathbf{U}_t^{(n)}, .\big) - \tilde{f}_t\big(\mathbf{U}_t^{(n)}, .\big)\Big\} - \Big\{\tilde{f}_{t-1}\big(\mathbf{U}_{t-1}^{(n)}, .\big) - \tilde{f}_t\big(\mathbf{U}_{t-1}^{(n)}, .\big)\Big\}. \tag{C7}$$

Let us denote the error function $d_t(\mathbf{U}^{(n)}, .) = \tilde{f}_{t-1}(\mathbf{U}^{(n)}, .) - \tilde{f}_t(\mathbf{U}^{(n)}, .)$. We have

$$\nabla d_t\big(\mathbf{U}^{(n)}, .\big) = \mathbf{U}^{(n)}\bigg(\frac{\mathbf{A}_{t-1}}{t-1} - \frac{\mathbf{A}_t}{t}\bigg) + \bigg(\frac{\mathbf{B}_{t-1}}{t-1} - \frac{\mathbf{B}_t}{t}\bigg), \tag{C8}$$

where $\mathbf{A}_t = \sum_{k=1}^t \lambda^{t-k}\big(\mathbf{W}_k^{(n)}\big)^\top \mathbf{W}_k^{(n)}$, $\mathbf{B}_t = \sum_{k=1}^t \lambda^{t-k}\big(\underline{\mathbf{P}}_k^{(n)} \circledast (\underline{\mathbf{X}}_k^{(n)} - \mathbf{O}_k^{(n)})\big)\mathbf{W}_k^{(n)}$. Thanks to the two facts that $\|\mathbf{M}\mathbf{N}\|_F \leq \|\mathbf{M}\|_F \|\mathbf{N}\|_F$ and $\|\mathbf{M} + \mathbf{N}\|_F \leq \|\mathbf{M}\|_F + \|\mathbf{N}\|_F$ [1], we obtain

$$\big\|\nabla d_t(\mathbf{U}^{(n)}, .)\big\|_F \leq \kappa_U \bigg\|\frac{\mathbf{A}_{t-1}}{t-1} - \frac{\mathbf{A}_t}{t}\bigg\|_F + \bigg\|\frac{\mathbf{B}_{t-1}}{t-1} - \frac{\mathbf{B}_t}{t}\bigg\|_F = c_n,$$

where κ_U is the upper bound for $\|\mathbf{U}^{(n)}\|_F$. As a result, the error function $d_t(\mathbf{U}^{(n)})$ is Lipschitz with parameter $c_n = \mathcal{O}(1/t)$, i.e.,

$$\begin{aligned} \tilde{f}_{t-1}\big(\mathbf{U}_t^{(n)}, .\big) - \tilde{f}_{t-1}\big(\mathbf{U}_{t-1}^{(n)}, .\big) &\leq d_t\big(\mathbf{U}_t^{(n)}, .\big) - d_t\big(\mathbf{U}_{t-1}^{(n)}, .\big) \\ &\leq c_n \big\|\mathbf{U}_t^{(n)} - \mathbf{U}_{t-1}^{(n)}\big\|_F. \end{aligned} \tag{C9}$$

Moreover, $\tilde{f}_t(\mathbf{U}^{(n)}, .)$ is a m-strongly convex function

$$\tilde{f}_{t-1}\big(\mathbf{U}_t^{(n)}, .\big) - \tilde{f}_{t-1}\big(\mathbf{U}_{t-1}^{(n)}, .\big) \geq m\big\|\mathbf{U}_t^{(n)} - \mathbf{U}_{t-1}^{(n)}\big\|_F^2. \tag{C10}$$

From (C9) and (C10), we obtain the asymptotic variation of $\mathbf{U}^{(n)}$ as follows $\big\|\mathbf{U}_t^{(n)} - \mathbf{U}_{t-1}^{(n)}\big\|_F \leq \frac{c_n}{m} = \mathcal{O}\big(1/t\big)$. Therefore, we can conclude that $\sum_{n=1}^N \big\|\mathbf{U}_t^{(n)} - \mathbf{U}_{t-1}^{(n)}\big\|_F^2 = \|\mathbf{D}_t - \mathbf{D}_{t-1}\|_F^2 = \mathcal{O}\big(1/t^2\big)$ or $\|\mathbf{D}_t - \mathbf{D}_{t-1}\|_F = \mathcal{O}\big(1/t\big)$.

5./ Stability of Errors: $\big|e_t(\mathbf{D}_t) - e_{t-1}(\mathbf{D}_{t-1})\big| = \mathcal{O}(1/t)$.

We begin with verifying the differentiable property of the loss function $\ell(\mathbf{D}, \boldsymbol{\mathcal{P}}_t, \boldsymbol{\mathcal{X}}_t)$ at each time t.

Proposition 1. *Given an observation $\boldsymbol{\mathcal{P}}_t \circledast \boldsymbol{\mathcal{X}}_t$ and the past estimation of $\mathbf{D}$, let $\boldsymbol{\mathcal{O}}_t, \mathbf{u}_t^*$ be the minimizer of $\tilde{\ell}(\mathbf{D}, \boldsymbol{\mathcal{P}}_t, \boldsymbol{\mathcal{X}}_t, \boldsymbol{\mathcal{O}}, \mathbf{u})$:*

$$\{\mathbf{u}_t^*, \boldsymbol{\mathcal{O}}_t^*\} = \underset{\mathbf{u}, \boldsymbol{\mathcal{O}}}{\operatorname{argmin}} \|\boldsymbol{\mathcal{O}}\|_1 + \frac{\rho}{2}\big\|\boldsymbol{\mathcal{P}}_t \circledast \big(\boldsymbol{\mathcal{X}}_t - \boldsymbol{\mathcal{O}} - \boldsymbol{\mathcal{H}} \times_{N+1} \mathbf{u}\big)\big\|_F^2.$$

where $\boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{I}} \prod_{n=1}^N \times_n \mathbf{U}^{(n)}$. We obtain that $\ell(\mathbf{D}, \boldsymbol{\mathcal{P}}_t, \boldsymbol{\mathcal{X}}_t) = \min_{\mathbf{u}, \boldsymbol{\mathcal{O}}} \tilde{\ell}(\mathbf{D}, \boldsymbol{\mathcal{P}}_t, \boldsymbol{\mathcal{X}}_t, \boldsymbol{\mathcal{O}}, \mathbf{u})$ is a continuously differentiable function and its partial derivative w.r.t. $\mathbf{U}^{(n)}$ is given by

$$\frac{\partial \ell(\mathbf{D}, \boldsymbol{\mathcal{P}}_t, \boldsymbol{\mathcal{X}}_t)}{\partial \mathbf{U}^{(n)}} = 2\underline{\mathbf{P}}_t^{(n)} \circledast \Big(\underline{\mathbf{X}}_t^{(n)} - \mathbf{O}_t^{(n)} - \mathbf{U}^{(n)}\big(\bar{\mathbf{W}}_t^{(n)}\big)^\top\Big)\bar{\mathbf{W}}_t^{(n)},$$

where $\bar{\mathbf{W}}_t^{(n)} = \Big(\bigodot_{i=1, i\neq n}^N \mathbf{U}_{t-1}^{(i)}\Big) \odot (\mathbf{u}_t^*)^\top$.

Proof. The result follows intermediately Theorem 4.1 in [3]. □

Accordingly, $f_t(\mathbf{D}) = L_t^{-1}\sum_{k=t-L_t+1}^t \lambda^{t-k}\ell(\mathbf{D}, \boldsymbol{\mathcal{P}}_k, \boldsymbol{\mathcal{X}}_k)$ is continuously differentiable. Now, let us denote $\bar{f}_t\big(\mathbf{U}^{(n)}, .\big) = f_{t-1}\big(\mathbf{U}^{(n)}, .\big) - f_t\big(\mathbf{U}^{(n)}, .\big)$. Applying the same arguments in Sec. C.4, we also obtain

$$\big\|\nabla \bar{f}_t(\mathbf{U}^{(n)}, .)\big\|_F \leq \kappa_U \bigg\|\frac{\bar{\mathbf{A}}_{t-1}^{(n)}}{t-1} - \frac{\bar{\mathbf{A}}_t^{(n)}}{t}\bigg\|_F + \bigg\|\frac{\bar{\mathbf{B}}_{t-1}^{(n)}}{t-1} - \frac{\bar{\mathbf{B}}_t^{(n)}}{t}\bigg\|_F = d_n,$$

where $\bar{\mathbf{A}}_t^{(n)} = \sum_{k=1}^t \lambda^{t-k}\big(\bar{\mathbf{W}}_k^{(n)}\big)^\top \bar{\mathbf{W}}_k^{(n)}$, and $\bar{\mathbf{B}}_t^{(n)} = \sum_{k=1}^t \lambda^{t-k}\big(\underline{\mathbf{P}}_k^{(n)} \circledast (\underline{\mathbf{X}}_k^{(n)} - \mathbf{O}_k^{(n)})\big)\bar{\mathbf{W}}_k^{(n)}$. Accordingly, $\nabla \bar{f}_t\big(\mathbf{U}^{(n)}, .\big)$ is bounded and hence $f_t\big(\mathbf{U}_{t-1}^{(n)}, .\big) - f_t\big(\mathbf{U}_t^{(n)}, .\big) \leq d_n\big\|\mathbf{U}_{t-1}^{(n)} - \mathbf{U}_t^{(n)}\big\|_F$. It implies that $f_t(.)$ is Lipschitz continuous. Since $\tilde{f}_t(\mathbf{D})$ and $f_t(\mathbf{D})$ are both Lipschitz continuous functions, we then have

$$\begin{aligned} &\big|e_t(\mathbf{D}_t) - e_{t-1}(\mathbf{D}_{t-1})\big| \\ &\quad = \big|\big(\tilde{f}_t(\mathbf{D}_t) - f_t(\mathbf{D}_t)\big) - \big(\tilde{f}_{t-1}(\mathbf{D}_{t-1}) - f_{t-1}(\mathbf{D}_{t-1})\big)\big| \\ &\quad \leq \big|\tilde{f}_t(\mathbf{D}_t) - \tilde{f}_t(\mathbf{D}_{t-1})\big| + \big|f_t(\mathbf{D}_t) - f_t(\mathbf{D}_{t-1})\big| \\ &\quad \leq \sum_{n=1}^N (c_n + d_n)\big\|\mathbf{U}_{t-1}^{(n)} - \mathbf{U}_t^{(n)}\big\|_F = \mathcal{O}(1/t). \end{aligned} \tag{C11}$$

It ends the proof.

Section D: Proof of Lemma 1

We apply the similar arguments of Proposition 7 in our companion work [4] to prove Lemma 1.

1./ Almost sure convergence of $\{\tilde{f}_t(\mathbf{D}_t)\}_{t=1}^\infty$.

Main approach: We prove the convergence of the sequence $\tilde{f}_t(\mathbf{D}_t)$ by showing that the stochastic positive process $u_t :=$

$\tilde{f}_t(\mathbf{D}_t)$ is a quasi-martingale. In particular, if the sum of the positive difference of u_t is bounded, u_t is a quasi-martingale, and the sum converges almost surely, thanks to the following quasi-martingale theorem:

Proposition 2 (Quasi-martingale Theorem [5, Theorem 9.4 & Proposition 9.5])**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\{u_t\}_{t>0}$ be a stochastic process on the probability space and $\{\mathcal{F}_t\}_{t>0}$ be a filtration by the past information at time instant t. Let us define the indicator function δ_t as follows*

$$\delta_t \triangleq \begin{cases} 1 & \textit{if} \quad \mathbb{E}[u_{t+1} - u_t | \mathcal{F}_t] > 0, \\ 0 & \textit{otherwise.} \end{cases}$$

For all t, if $u_t \geq 0$ and $\sum_{i=1}^{\infty} \mathbb{E}[\delta_i (u_{i+1} - u_i) | \mathcal{F}_i] < \infty$, then u_t is a quasi-martingale and converges almost surely, i.e.,

$$\sum_{t=1}^{\infty} \mathbb{E}[u_{t+1} - u_t | \mathcal{F}_t] < \infty.$$

Now, we begin with the following relation when $L_t = t$

$$\begin{aligned} \tilde{f}_{t+1}(\mathbf{D}_t) &= \frac{1}{t+1} \sum_{k=1}^{t+1} \lambda^{t+1-k} \tilde{\ell}(\mathbf{D}_t, \boldsymbol{\mathcal{P}}_k, \boldsymbol{\mathcal{X}}_k, \boldsymbol{\mathcal{O}}_k, \mathbf{u}_k) \\ &= \frac{\tilde{\ell}(\mathbf{D}_t, \boldsymbol{\mathcal{P}}_{t+1}, \boldsymbol{\mathcal{X}}_{t+1}, \boldsymbol{\mathcal{O}}_{t+1}, \mathbf{u}_{t+1})}{t+1} + \frac{t(\lambda - 1)}{t+1} \tilde{f}_t(\mathbf{D}_t) + \frac{t}{t+1} \tilde{f}_t(\mathbf{D}_t). \end{aligned} \tag{D1}$$

Thanks to Proposition 1 and $\lambda \leq 1$, we obtain $\tilde{f}_{t+1}(\mathbf{D}_{t+1}) \leq \tilde{f}_{t+1}(\mathbf{D}_t)$ and

$$\begin{aligned} \frac{\tilde{f}_t(\mathbf{D}_t) - f_t(\mathbf{D}_t)}{t+1} &\leq \tilde{f}_t(\mathbf{D}_t) - \tilde{f}_{t+1}(\mathbf{D}_{t+1}) \\ &+ \frac{\tilde{\ell}(\mathbf{D}_t, \boldsymbol{\mathcal{P}}_{t+1}, \boldsymbol{\mathcal{X}}_{t+1}, \boldsymbol{\mathcal{O}}_{t+1}, \mathbf{u}_{t+1}) - f_t(\mathbf{D}_t)}{t+1}. \end{aligned} \tag{D2}$$

Since $f_t(\mathbf{D}_t) \leq \tilde{f}_t(\mathbf{D}_t)\ \forall t$, we then have

$$\tilde{f}_{t+1}(\mathbf{D}_{t+1}) - \tilde{f}_t(\mathbf{D}_t) \leq \frac{\tilde{\ell}(\mathbf{D}_t, \boldsymbol{\mathcal{P}}_{t+1}, \boldsymbol{\mathcal{X}}_{t+1}, \boldsymbol{\mathcal{O}}_{t+1}, \mathbf{u}_{t+1}) - f_t(\mathbf{D}_t)}{t+1},$$

Define by $\{\mathcal{F}_t\}_{t>0}$ a filtration associated to $\{u_t\}_{t>0}$ where $\mathcal{F}_t = \{\mathbf{D}_k, \boldsymbol{\mathcal{O}}_k, \mathbf{u}_k\}_{1 \leq k \leq t}$ records all past estimates of RACP at time t. By definition, for every $i \leq t$, $\mathcal{F}_i \subseteq \mathcal{F}_t$, and thus, the filtration is interpreted as streams of all historical but not future information generated by RACP. Now, taking the expectation of the inequality above conditioned on $\mathcal{F}_t$ results in

$$\mathbb{E}\Big[\tilde{f}_{t+1}(\mathbf{D}_{t+1}) - \tilde{f}_t(\mathbf{D}_t) \big| \mathcal{F}_t\Big] \leq \frac{f(\mathbf{D}_t) - f_t(\mathbf{D}_t)}{t+1}, \tag{D3}$$

because the expected cost function $f(.)$ is given by $f(\mathbf{D}) = \lim_{k \to \infty} f_k(\mathbf{D})$, $\mathbb{E}\big[\ell(\mathbf{D}_t, \boldsymbol{\mathcal{P}}_{k+1} \boldsymbol{\mathcal{X}}_{k+1})\big] = f(\mathbf{D}_t), \forall \mathbf{D}_t$ and $\forall t$; and $\ell(\mathbf{D}_t, \boldsymbol{\mathcal{P}}_{t+1}, \boldsymbol{\mathcal{X}}_{t+1}) = \tilde{\ell}(\mathbf{D}_t, \boldsymbol{\mathcal{P}}_{t+1}, \boldsymbol{\mathcal{X}}_{t+1}, \boldsymbol{\mathcal{O}}_{t+1}, \mathbf{u}_{t+1})$ due to $\{\boldsymbol{\mathcal{O}}_{t+1}, \mathbf{u}_{t+1}\} = \arg\min_{\boldsymbol{\mathcal{O}}, \mathbf{u}} \tilde{\ell}(\mathbf{D}, \boldsymbol{\mathcal{P}}_{t+1}, \boldsymbol{\mathcal{X}}_{t+1}, \boldsymbol{\mathcal{O}}, \mathbf{u})$ at time t.

Next, let us denote the following indicator function

$$\delta_t \triangleq \begin{cases} 1 & \text{if} \quad \mathbb{E}\big[\tilde{f}_{t+1}(\mathbf{D}_{t+1}) - \tilde{f}_t(\mathbf{D}_t) \big| \mathcal{F}_t\big] > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{D4}$$

Here, the process $\{\delta_t\}_{t>0}$ is adapted to the filtration $\{\mathcal{F}_t\}_{t>0}$ as δ_t is measurable w.r.t. $\mathcal{F}_t$ for every t. From (D3), we obtain

$$\begin{aligned} &\mathbb{E}\Big[\delta_t \mathbb{E}\big[\tilde{f}_{t+1}(\mathbf{D}_{t+1}) - \tilde{f}_t(\mathbf{D}_t) \big| \mathcal{F}_t\big]\Big] \\ &\leq \mathbb{E}\left[\frac{f(\mathbf{D}_t) - f_t(\mathbf{D}_t)}{(t+1)}\right] = \mathbb{E}\Big[\sqrt{t}\big(f(\mathbf{D}_t) - f_t(\mathbf{D}_t)\big)\Big] \frac{1}{\sqrt{t}(t+1)}. \end{aligned} \tag{D5}$$

As the solutions $\{\mathbf{D}_t, \boldsymbol{\mathcal{O}}_t, \mathbf{u}_t\}_{t>0}$ are bounded thanks to Proposition 1, we exploit that the set of measurable functions $\{\ell(\mathbf{D}_t, \boldsymbol{\mathcal{P}}, \boldsymbol{\mathcal{X}})\}_{t>0}$, which is composed of a quadratic norm term and ℓ_1-norm term, is $\mathbb{P}$-Donsker. It is therefore that the centered and scaled version of $f_t(\mathbf{D}_t)$ satisfies $\mathbb{E}\big[\sqrt{t}\big(f(\mathbf{D}_t) - f_t(\mathbf{D}_t)\big)\big] = \mathcal{O}(1)$, thanks to the Donsker theorem [6, Section 19.2]. In addition, we have $\int_{t=1}^{+\infty} \frac{1}{\sqrt{t}(t+1)} dt = \frac{\pi}{4}$. Hence, $\sum_{t=1}^{+\infty} 1/\sqrt{t}(t+1) < \infty$ too. Accordingly, we obtain

$$\sum_{t=1}^{\infty} \mathbb{E}\Big[\delta_t \mathbb{E}\big[\tilde{f}_{t+1}(\mathbf{D}_{t+1}) - \tilde{f}_t(\mathbf{D}_t) \big| \mathcal{F}_t\big]\Big] < \infty. \tag{D6}$$

Thanks to Proposition 2, $\{\tilde{f}_t(\mathbf{D}_t)\}_{t=1}^{\infty}$ converges almost surely

$$\sum_{t=1}^{\infty} \mathbb{E}\Big[\tilde{f}_{t+1}(\mathbf{D}_{t+1}) - \tilde{f}_t(\mathbf{D}_t) \big| \mathcal{F}_t\Big] < \infty. \tag{D7}$$

2./ As $t \to \infty$, $\tilde{f}_t(\mathbf{D}_t) \to f_t(\mathbf{D}_t)$ almost surely.
We prove $\{f_t(\mathbf{D}_t)\}_{t=1}^{\infty}$ and $\{\tilde{f}_t(\mathbf{D}_t)\}_{t=1}^{\infty}$ converge to the same limit by showing $\sum_{t=1}^{\infty} \frac{\tilde{f}_t(\mathbf{D}_t) - f_t(\mathbf{D}_t)}{t+1} < \infty$.
According to (D2), we know that $\frac{e_t(\mathbf{D}_t)}{t+1}$ is bounded by $\tilde{f}_t(\mathbf{D}_t) - \tilde{f}_{t+1}(\mathbf{D}_{t+1})$ and $\frac{\ell(\mathbf{D}_t, \boldsymbol{\mathcal{P}}_{t+1}, \boldsymbol{\mathcal{X}}_{t+1}) - f_t(\mathbf{D}_t)}{(t+1)}$. Moreover, we have $\sum_{t=1}^{\infty} \tilde{f}_t(\mathbf{D}_t) - \tilde{f}_{t+1}(\mathbf{D}_{t+1}) < \infty$, and the sum of $\frac{\ell(\mathbf{D}_t, \boldsymbol{\mathcal{P}}_{t+1}, \boldsymbol{\mathcal{X}}_{t+1}) - f_t(\mathbf{D}_t)}{t+1}$ also converges due to the convergence of $\frac{\mathbb{E}[f(\mathbf{D}_t) - f_t(\mathbf{D}_t)]}{t+1}$ and $\mathbb{E}[\ell(\mathbf{D}_t, \boldsymbol{\mathcal{P}}, \boldsymbol{\mathcal{X}})] = f(\mathbf{D}_t) \forall t$. Since $\sum_{t=1}^{\infty} \frac{1}{t+1} = \infty$ and $\big|e_t(\mathbf{D}_t) - e_{t-1}(\mathbf{D}_{t-1})\big| = \mathcal{O}(1/t)$, we obtain $\sum_{t=1}^{\infty} \tilde{f}_t(\mathbf{D}_t) - f_t(\mathbf{D}_t) < \infty$, or

$$\tilde{f}_t(\mathbf{D}_t) \to f_t(\mathbf{D}_t)\ a.s., \tag{D8}$$

thanks to [7, Lemma 3].

Section E: Proof of Lemma 2

In what follows, we prove that when $t \to \infty$, $\nabla \tilde{f}_t\big(\mathbf{D}_t\big) \to \nabla f_t\big(\mathbf{D}_t\big)$ and $\nabla \tilde{f}_t(\mathbf{D}_t) \to 0$ almost surely.

1./ As $t \to \infty$, $\nabla \tilde{f}_t\big(\mathbf{D}_t\big) \to \nabla f_t\big(\mathbf{D}_t\big)$ almost surely.
Let us denote by $\mathbf{D}_\infty$ the dictionary $\mathbf{D}_t$ at $t \to \infty$. We know that $\tilde{f}_t(\mathbf{D})$ is a majorant function of $f_t(\mathbf{D})$, i.e.,

$$\tilde{f}_t(\mathbf{D} + a_t \mathbf{V}) \geq f_t(\mathbf{D} + a_t \mathbf{V})\ \forall \mathbf{D}, \mathbf{V} \in \mathcal{D}, a_t. \tag{E1}$$

Taking the Taylor expansion of (E1) at $t \to \infty$ results in

$$\begin{aligned} &f_\infty\big(\mathbf{D}_\infty\big) + \operatorname{tr}\left[a_t \mathbf{V}^\top \nabla f_\infty\big(\mathbf{D}_\infty\big)\right] + \boldsymbol{o}\big(a_t \mathbf{V}\big) \\ &\quad \leq \tilde{f}_\infty\big(\mathbf{D}_\infty\big) + \operatorname{tr}\left[a_t \mathbf{V}^\top \nabla f_\infty\big(\mathbf{D}_\infty\big)\right] + \boldsymbol{o}\big(a_t \mathbf{V}\big), \end{aligned} \tag{E2}$$

where $\tilde{f}_\infty = \lim_{t \to \infty} \tilde{f}_t(.)$. As indicated in Lemma 1, $\tilde{f}_\infty\big(\mathbf{D}_\infty\big) = f_\infty\big(\mathbf{D}_\infty\big)$ and hence $\operatorname{tr}\left[a_t \mathbf{V}^\top \nabla f_\infty\big(\mathbf{D}_\infty\big)\right] \leq \operatorname{tr}\left[a_t \mathbf{V}^\top \nabla \tilde{f}_\infty\big(\mathbf{D}_\infty\big)\right]$. Since the above inequality must hold for all $\mathbf{V}$ and a_t, we obtain $\operatorname{tr}\left[\nabla \tilde{f}_\infty\big(\mathbf{D}_\infty\big) - \nabla f_\infty\big(\mathbf{D}_\infty\big)\right] \to 0$ or $\nabla \tilde{f}_\infty\big(\mathbf{D}_\infty\big) = \nabla f_\infty\big(\mathbf{D}_\infty\big)$ a.s.

2./ As $t \to \infty$, $\nabla \tilde{f}_\infty\big(\mathbf{D}_\infty\big) = \mathbf{0}$.
This property is proved by applying immediately the following stages:

1) Stage 1: $\lim_{t \to \infty} \operatorname{tr}\left[(\mathbf{D}_t - \mathbf{D}_{t+1})^\top \nabla \tilde{f}_{t+1}\big(\mathbf{D}_{t+1}\big)\right] = 0$;
2) Stage 2: $\operatorname{tr}\Big[(\mathbf{D}_t - \mathbf{D}_{t+1})^\top \nabla \tilde{f}_{t+1}\big(\mathbf{D}_{t+1}\big)\Big] \leq c_1 \operatorname{tr}\Big[(\mathbf{D} - \mathbf{D}_t)^\top \nabla \tilde{f}_{t+1}\big(\mathbf{D}_t\big)\Big] + c_2 \big\|\mathbf{D}_{t+1} - \mathbf{D}_t\big\|_F^2\ \forall t, \mathbf{D} \in \mathcal{D}$;
3) Stage 3: $\big(\nabla \tilde{f}_\infty(\mathbf{D}_\infty)\big)^\top (\mathbf{D} - \mathbf{D}_\infty) \geq \mathbf{0}\ \forall \mathbf{D}$.

Stage 1: When $L_t = t$, we can recast the surrogate function $\tilde{f}_t(.)$ into the following form

$$\begin{aligned} \tilde{f}_t(\mathbf{D}) &= \frac{\rho}{t} \operatorname{tr}\Big[\mathbf{A}_t\Big(\big[(\mathbf{U}^{(N)})^\top \mathbf{U}^{(N)}\big] \circledast \cdots \circledast \big[(\mathbf{U}^{(1)})^\top \mathbf{U}^{(1)}\big]\Big)\Big] \\ &- \frac{2\rho}{t} \operatorname{tr}\Big[\mathbf{B}_t\big(\mathbf{U}^{(N)} \odot \mathbf{U}^{(N-1)} \odot \cdots \odot \mathbf{U}^{(1)}\big)^\top\Big] + \mathbf{R}_{\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{O}}}, \end{aligned} \tag{E3}$$

where $\mathbf{A}_t = \lambda \mathbf{A}_{t-1} + \mathbf{u}_t \mathbf{u}_t^\top$, and $\mathbf{B}_t$ is the $(N+1)$-unfolding matrix of the tensor $\boldsymbol{\mathcal{B}}_t = \lambda \boldsymbol{\mathcal{B}}_{t-1} + \boldsymbol{\mathcal{P}}_t \circledast (\boldsymbol{\mathcal{X}}_t - \boldsymbol{\mathcal{O}}_t) \times_{N+1} \mathbf{u}_t^\top$, and $\mathbf{R}_{\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{O}}} =$

$\frac{\rho}{t}\sum_{k=1}^{t}\|\boldsymbol{\mathcal{P}}_t \circledast \boldsymbol{\mathcal{X}}_t\|_F^2 + \frac{1}{t}\sum_{k=1}^{t}\lambda^{t-k}\|\boldsymbol{\mathcal{O}}_k\|_1$ independent of $\mathbf{D}$. With respect to each $\mathbf{U}^{(n)}$, we can further express $\tilde{f}_t(\mathbf{D})$ as

$$\tilde{f}_t(\mathbf{D}) = \frac{\rho}{t}\operatorname{tr}\Big[\big(\mathbf{U}^{(n)}\big)^\top \mathbf{U}^{(n)}\mathbf{A}_{t,n}\Big] - \frac{2\rho}{t}\operatorname{tr}\Big[\big(\mathbf{U}^{(n)}\big)^\top \mathbf{B}_{t,n}\Big] + \mathbf{R}_{\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{O}}}.$$

Here, the two matrices $\mathbf{A}_{t,n}$ and $\mathbf{B}_{t,n}$ are given by

$$\begin{aligned}\mathbf{A}_{t,n} = \mathbf{A}_t \circledast \big[(\mathbf{U}^{(1)})^\top\mathbf{U}^{(1)}\big] \circledast \cdots \circledast \big[(\mathbf{U}^{(n-1)})^\top\mathbf{U}^{(n-1)}\big] \circledast \\ \circledast \big[(\mathbf{U}^{(n+1)})^\top\mathbf{U}^{(n+1)}\big] \circledast \cdots \circledast \big[(\mathbf{U}^{(1)})^\top\mathbf{U}^{(1)}\big],\\ \mathbf{B}_{t,n} = \sum_{j=1}^{r}\mathbf{B}_t^{(j)} \times_1 \mathbf{U}^{(1)}(:,j) \times_2 \cdots \times_{n-1} \mathbf{U}^{(n-1)}(:,j) \times_{n+1} \\ \times_{n+1}\mathbf{U}^{(n+1)}(:,j)\cdots \times_N \mathbf{U}^{(N)}(:,j),\end{aligned}$$

where $\mathbf{B}_t^{(j)} \in \mathbb{R}^{I_1\times I_2 \cdots \times I_N}$ denote the j-th mode-$(N+1)$ slices of $\mathbf{B}_t$. It is easy to see that $\tilde{f}_t(\mathbf{D})$ is a multi-block convex and differentiable function and its partial derivative w.r.t. each block is Lipschitz continuous with constant $\tilde{L}_{t,n} = \|\mathbf{A}_{t,n}\|_F$. Accordingly, we have

$$\begin{aligned}\Big|\tilde{f}_{t+1}\big(\mathbf{D}_t\big) - \tilde{f}_{t+1}\big(\mathbf{D}_{t+1}\big) - \operatorname{tr}\big[(\mathbf{D}_t - \mathbf{D}_{t+1})^\top \nabla \tilde{f}_{t+1}(\mathbf{D}_{t+1})\big]\Big| \\ \leq \tilde{L}\big\|\mathbf{D}_t - \mathbf{D}_{t+1}\big\|_F^2,\end{aligned} \quad \text{(E4)}$$

with $\tilde{L} = \max_n(\tilde{L}_{t,n}/2)$. Thanks to the triangle inequality, we then obtain

$$\begin{aligned}\Big|\operatorname{tr}\big[(\mathbf{D}_t - \mathbf{D}_{t+1})^\top \nabla \tilde{f}_{t+1}(\mathbf{D}_{t+1})\big]\Big| \\ \leq \tilde{L}\big\|\mathbf{D}_t - \mathbf{D}_{t+1}\big\|_F^2 + \tilde{f}_{t+1}(\mathbf{D}_t) - \tilde{f}_{t+1}(\mathbf{D}_{t+1}).\end{aligned} \quad \text{(E5)}$$

Accordingly, we have

$$\begin{aligned}\sum_{t=1}^{\infty}\Big|\mathbb{E}\Big[\operatorname{tr}\big[(\mathbf{D}_t - \mathbf{D}_{t+1})^\top \nabla \tilde{f}_{t+1}(\mathbf{D}_{t+1})\big]\big|\mathcal{F}_t\Big]\Big| \quad \text{(E6)}\\ \leq \tilde{L}\sum_{t=1}^{\infty}\mathbb{E}\Big[\big\|\mathbf{D}_t - \mathbf{D}_{t+1}\big\|_F^2\Big] + \sum_{t=1}^{\infty}\Big|\mathbb{E}\Big[\tilde{f}_{t+1}(\mathbf{D}_{t+1}) - \tilde{f}_{t+1}(\mathbf{D}_t)\big|\mathcal{F}_t\Big]\Big|.\end{aligned}$$

Recall that $\big\|\mathbf{D}_t - \mathbf{D}_{t+1}\|_F = \mathcal{O}(1/t)$ as indicated in Proposition 1, hence $\sum_{t=1}^{\infty}\big\|\mathbf{D}_t - \mathbf{D}_{t+1}\big\|_F^2 \leq d\sum_{t=1}^{\infty}\frac{1}{t^2} = d\frac{\pi}{6} < \infty$ for some constant $d > 0$. Accordingly, we obtain that RHS of (E6) is finite.

Also, it is well-known that $\mathbb{E}[|x|] < \infty$ implies $|x| < \infty$ almost surely for any random variable x, thus we obtain

$$\sum_{t=1}^{\infty}\Big|\operatorname{tr}\big[(\mathbf{D}_t - \mathbf{D}_{t+1})^\top \nabla \tilde{f}_{t+1}(\mathbf{D}_{t+1})\big]\Big| < \infty. \quad \text{(E7)}$$

Moreover, we always have

$$\begin{aligned}\sum_{t=1}^{\infty}\operatorname{tr}\big[(\mathbf{D}_t - \mathbf{D}_{t+1})^\top \nabla \tilde{f}_{t+1}(\mathbf{D}_{t+1})\big] \\ < \sum_{t=1}^{\infty}\Big|\operatorname{tr}\big[(\mathbf{D}_t - \mathbf{D}_{t+1})^\top \nabla \tilde{f}_{t+1}(\mathbf{D}_{t+1})\big]\Big| < \infty.\end{aligned} \quad \text{(E8)}$$

Therefore the series $\big\{\operatorname{tr}[(\mathbf{D}_t - \mathbf{D}_{t+1})^\top \nabla \tilde{f}_{t+1}(\mathbf{D}_{t+1})]\big\}_{t\geq 1}$ converges and we suppose it converges to $C < \infty$. Then, we have

$$\begin{aligned}&\lim_{t\to\infty}\sum_{k=1}^{t}\operatorname{tr}\big[(\mathbf{D}_k - \mathbf{D}_{k+1})^\top \nabla \tilde{f}_{k+1}(\mathbf{D}_{k+1})\big]\\ &= \lim_{t\to\infty}\operatorname{tr}\big[(\mathbf{D}_t - \mathbf{D}_{t+1})^\top \nabla \tilde{f}_{t+1}\big(\mathbf{D}_{t+1}\big)\big]\\ &+ \lim_{t\to\infty}\sum_{k=1}^{t-1}\operatorname{tr}\big[(\mathbf{D}_k - \mathbf{D}_{k+1})^\top \nabla \tilde{f}_{k+1}(\mathbf{D}_{k+1})\big] = C < \infty.\end{aligned} \quad \text{(E9)}$$

When $t \to \infty$, the following partial sum also converges to C, i.e.,

$$\lim_{t\to\infty}\sum_{k=1}^{t-1}\operatorname{tr}\big[(\mathbf{D}_k - \mathbf{D}_{k+1})^\top \nabla \tilde{f}_{k+1}(\mathbf{D}_{k+1})\big] = C. \quad \text{(E10)}$$

It implies that

$$\lim_{t\to\infty}\operatorname{tr}\Big[(\mathbf{D}_t - \mathbf{D}_{t+1})^\top \nabla \tilde{f}_{t+1}\big(\mathbf{D}_{t+1}\big)\Big] = 0. \quad \text{(E11)}$$

Step 2: Because $\mathbf{U}_{t+1}^{(n)} = \operatorname{argmin}_{\mathbf{U}^{(n)}}\tilde{f}_{t+1}\big(\mathbf{U}^{(n)},.\big)$, we have

$$\tilde{f}_{t+1}\big(\mathbf{U}_{t+1}^{(n)},.\big) \leq \tilde{f}_{t+1}\Big(\mathbf{U}_t^{(n)} + \frac{d_1}{tN}\big(\mathbf{U}^{(n)} - \mathbf{U}_t^{(n)}\big),.\Big) \quad \forall \mathbf{D} \in \mathcal{D}. \quad \text{(E12)}$$

Without loss of generality, we suppose that $\mathbf{D}$ is arbitrarily chosen in $\mathcal{D}$ such that $\|\mathbf{D} - \mathbf{D}_t\|_F = d_1/tN$ for some positive constant $d_1 > 0$, hence $\|\mathbf{U}^{(n)} - \mathbf{U}_t^{(n)}\|_F \leq d_1/Nt \ \forall n$.

As mentioned in Stage 1, $\nabla \tilde{f} = \big[\nabla_1\tilde{f}, \nabla_2\tilde{f}, \ldots, \nabla_N\tilde{f}\ \big]$ is Lipschitz where $\nabla_n\tilde{f}$ denote the partial derivative of $\tilde{f}$ w.r.t. the n-th factor $\mathbf{U}^{(n)}$. Thanks to [8, Lemma 1.2.3], there always exists a constant $d_2 > 0$ such that

$$\begin{aligned}\operatorname{tr}\Big[\big(\mathbf{U}_t^{(n)} - \mathbf{U}_{t+1}^{(n)}\big)^\top \nabla_n\tilde{f}_{t+1}\big(\mathbf{U}_{t+1}^{(n)},.\big)\Big] \quad \text{(E13)}\\ \leq \frac{d_1}{tN}\operatorname{tr}\Big[\big(\mathbf{U}^{(n)} - \mathbf{U}_t^{(n)}\big)^\top \nabla_n\tilde{f}_{t+1}\big(\mathbf{U}_t^{(n)},.\big)\Big] + \frac{\tilde{L}d_2}{t^2N^2}.\end{aligned}$$

Collecting these inequalities with $n = 1, 2, \ldots, N$ together, we derive

$$\begin{aligned}&\operatorname{tr}\Big[(\mathbf{D}_t - \mathbf{D}_{t+1})^\top\big[\nabla_1\tilde{f}_{t+1}(\mathbf{U}_{t+1}^{(1)},.), \ldots, \nabla_N\tilde{f}_{t+1}(\mathbf{U}_{t+1}^{(N)},.)\big]\Big] \quad \text{(E14)}\\ &\leq \frac{d_1}{tN}\operatorname{tr}\Big[\big(\mathbf{D} - \mathbf{D}_t\big)^\top\Big[\nabla_1\tilde{f}_{t+1}\big(\mathbf{U}_t^{(n)},.\big), \ldots, \nabla_N\tilde{f}_{t+1}\big(\mathbf{U}_t^{(n)},.\big)\Big]\Big] + \frac{\tilde{L}d_2}{t^2N^2}.\end{aligned}$$

It then follows that

$$\begin{aligned}\operatorname{tr}\Big[(\mathbf{D}_t - \mathbf{D}_{t+1})^\top \nabla \tilde{f}_{t+1}\big(\mathbf{D}_{t+1}\big)\Big] \leq \frac{d_1}{tN}\operatorname{tr}\Big[(\mathbf{D} - \mathbf{D}_t)^\top \nabla \tilde{f}_{t+1}\big(\mathbf{D}_t\big)\Big] \\ + \tilde{L}d_2\big\|\mathbf{D}_t - \mathbf{D}_{t+1}\big\|_F^2,\end{aligned} \quad \text{(E15)}$$

because of $\|\mathbf{D}_t - \mathbf{D}_{t+1}\|_F = \mathcal{O}(1/t)$. The inequality (E15) still holds for all $\mathbf{D} \in \mathcal{D}$ such that $\|\mathbf{D} - \mathbf{D}_t\|_F > d_1/tN$.

Step 3: We use the proof by contradiction to indicate that $\mathbf{D}_\infty$ is a stationary point of $\tilde{f}_\infty(.)$ over $\mathcal{D}$.

Assume that $\mathbf{D}_\infty$ is not a stationary point of $\tilde{f}_t$ over $\mathcal{D}$ when $t \to \infty$. Then there exists $\mathbf{D}' \in \mathcal{D}$ and $\epsilon_1 > 0$ such that

$$\operatorname{tr}\big[(\mathbf{D}' - \mathbf{D}_\infty)^\top \nabla \tilde{f}_\infty\big(\mathbf{D}_\infty\big)\big] \leq -\epsilon_1 < 0. \quad \text{(E16)}$$

Thanks to the triangle inequality, we have

$$\begin{aligned}&\big\|(\mathbf{D}' - \mathbf{D}_k)^\top \nabla \tilde{f}_{k+1}(\mathbf{D}_k) - (\mathbf{D}' - \mathbf{D}_\infty)^\top \nabla \tilde{f}_\infty(\mathbf{D}_\infty)\big\|_F \leq \\ &\big\|\nabla \tilde{f}_{k+1}(\mathbf{D}_k) - \nabla \tilde{f}_\infty(\mathbf{D}_\infty)\big\|_F\big\|\mathbf{D}' - \mathbf{D}_k\big\|_F \\ &+ \|\tilde{f}_\infty(\mathbf{D}_\infty)\|_F\|\mathbf{D}_\infty - \mathbf{D}_k\|_F.\end{aligned} \quad \text{(E17)}$$

It is easy to see that the RHS of (E17) approaches to zero as $k \to \infty$ because of $\mathbf{D}_k \to \mathbf{D}_\infty$ and $\nabla \tilde{f}_{k+1}(\mathbf{D}_k) \to \nabla \tilde{f}_\infty(\mathbf{D}_\infty)$. In parallel, we know that $\operatorname{tr}[\mathbf{A}] - \operatorname{tr}[\mathbf{B}] = \operatorname{tr}[\mathbf{A} - \mathbf{B}] \leq \sqrt{n}\|\mathbf{A} - \mathbf{B}\|_F$ and hence

$$\operatorname{tr}\Big[\big(\mathbf{D}' - \mathbf{D}_k\big)^\top \nabla \tilde{f}_{k+1}\big(\mathbf{D}_k\big)\Big] \leq -\epsilon_1 < 0. \quad \text{(E19)}$$

According to (E15), we obtain

$$\lim_{k\to\infty}\operatorname{tr}\Big[(\mathbf{D}_k - \mathbf{D}_{k+1})^\top \nabla \tilde{f}_{k+1}(\mathbf{D}_{k+1})\Big] \leq \frac{-d_1\epsilon}{tN\|\mathbf{D}' - \mathbf{D}_k\|_F} < 0,$$

which is a contradiction in (E11) in Step 1. Therefore, $\mathbf{D}_\infty$ is a stationary point of $\tilde{f}_\infty$.

References

[1] G. H. Golub and C. F. Van Loan, *Matrix Computations*. JHU press, 2012.

[2] N. J. Higham, *Accuracy and stability of numerical algorithms*. SIAM, 2002.

[3] J. F. Bonnans and A. Shapiro, "Optimization problems with perturbations: A guided tour," *SIAM Rev.*, vol. 40, no. 2, pp. 228–264, 1998.

[4] L. T. Thanh, K. Abed-Meraim *et al.*, "Tracking online low-rank approximations of incomplete high-order streaming tensors," *TechRxiv*, 2022.

[5] M. Métivier, *Semimartingales: A course on stochastic processes*. Walter de Gruyter, 1984.

[6] A. W. Van der Vaart, *Asymptotic Statistics*. Cambridge University Press, 2000.

[7] J. Mairal, F. Bach *et al.*, "Online learning for matrix factorization and sparse coding," *J. Mach. Learn.Res.*, vol. 11, pp. 19–60, 2010.

[8] Y. Nesterov, "Introductory lectures on convex programming volume I: Basic course," *Lecture Notes*, vol. 3, no. 4, 1998.