Low-rank tensor completion: a Riemannian manifold preconditioning approach Supplementary material

Hiroyuki Kasai KASAI@IS.UEC.AC.JP

The University of Electro-Communications,1-5-1, Chofu-gaoka, Chofu-shi, Tokyo, 182-8585, Japan

Bamdev Mishra BAMDEVM@AMAZON.COM

Amazon Development Centre India, Bengaluru 560055, Karnataka, India

A Proof and derivation of manifold-related ingredients

The concrete computations of the optimization-related ingredients presented in the paper are discussed below.

The total space is $\mathcal{M} := \mathrm{St}(r_1, n_1) \times \mathrm{St}(r_2, n_2) \times \mathrm{St}(r_3, n_3) \times \mathbb{R}^{r_1 \times r_2 \times r_3}$. Each element $x \in \mathcal{M}$ has the matrix representation $(\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3, \boldsymbol{\mathcal{G}})$. Invariance of Tucker decomposition under the transformation $(\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3, \boldsymbol{\mathcal{G}}) \mapsto (\mathbf{U}_1\mathbf{O}_1, \mathbf{U}_2\mathbf{O}_2, \mathbf{U}_3\mathbf{O}_3, \boldsymbol{\mathcal{G}} \times_1 \mathbf{O}_1^T \times_2 \mathbf{O}_2^T \times_3 \mathbf{O}_3^T)$ for all $\mathbf{O}_d \in \mathcal{O}(r_d)$, the set of orthogonal matrices of size of $r_d \times r_d$ results in equivalence classes of the form $[x] = [(\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3, \boldsymbol{\mathcal{G}})] := \{(\mathbf{U}_1\mathbf{O}_1, \mathbf{U}_2\mathbf{O}_2, \mathbf{U}_3\mathbf{O}_3, \boldsymbol{\mathcal{G}} \times_1 \mathbf{O}_1^T \times_2 \mathbf{O}_2^T \times_3 \mathbf{O}_3^T) : \mathbf{O}_d \in \mathcal{O}(r_d)\}$.

A.1 Tangent space characterization and the Riemannian metric

The tangent space, $T_x\mathcal{M}$, at x given by $(\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3, \boldsymbol{\mathcal{G}})$ in the total space $\mathcal{M}$ is the product space of the tangent spaces of the individual manifolds. From (Absil et al., 2008), the tangent space has the matrix characterization

$$
\begin{array}{rcl}
T_x\mathcal{M} & = & \{(\mathbf{Z}_{\mathbf{U}_1}, \mathbf{Z}_{\mathbf{U}_2}, \mathbf{Z}_{\mathbf{U}_3}, \mathbf{Z}_{\boldsymbol{\mathcal{G}}}) \in \mathbb{R}^{n_1 \times r_1} \times \mathbb{R}^{n_2 \times r_2} \times \mathbb{R}^{n_3 \times r_3} \times \mathbb{R}^{r_1 \times r_2 \times r_3} \\
 & & : \mathbf{U}_d^T \mathbf{Z}_{\mathbf{U}_d} + \mathbf{Z}_{\mathbf{U}_d}^T \mathbf{U}_d = 0, \text{ for } d \in \{1, 2, 3\}\}.
\end{array}
\tag{A.1}
$$

The proposed metric $g_x : T_x\mathcal{M} \times T_x\mathcal{M} \rightarrow \mathbb{R}$ is

$$
g_x(\xi_x, \eta_x) \quad = \quad \langle \xi_{\mathbf{U}_1}, \eta_{\mathbf{U}_1}(\mathbf{G}_1\mathbf{G}_1^T)\rangle + \langle \xi_{\mathbf{U}_2}, \eta_{\mathbf{U}_2}(\mathbf{G}_2\mathbf{G}_2^T)\rangle + \langle \xi_{\mathbf{U}_3}, \eta_{\mathbf{U}_3}(\mathbf{G}_3\mathbf{G}_3^T)\rangle + \langle \xi_{\boldsymbol{\mathcal{G}}}, \eta_{\boldsymbol{\mathcal{G}}}\rangle,
\tag{A.2}
$$

where $\xi_x, \eta_x \in T_x\mathcal{M}$ are tangent vectors with matrix characterizations $(\xi_{\mathbf{U}_1}, \xi_{\mathbf{U}_2}, \xi_{\mathbf{U}_3}, \xi_{\boldsymbol{\mathcal{G}}})$ and $(\eta_{\mathbf{U}_1}, \eta_{\mathbf{U}_2}, \eta_{\mathbf{U}_3}, \eta_{\boldsymbol{\mathcal{G}}})$, respectively and $\langle \cdot, \cdot \rangle$ is the Euclidean inner product.

A.2 Characterization of the normal space

Given a vector in $\mathbb{R}^{n_1 \times r_1} \times \mathbb{R}^{n_2 \times r_2} \times \mathbb{R}^{n_3 \times r_3} \times \mathbb{R}^{r_1 \times r_2 \times r_3}$, its projection onto the tangent space $T_x\mathcal{M}$ is obtained by extracting the component *normal*, in the metric sense, to the tangent space. This section describes the characterization of the *normal space*, $N_x\mathcal{M}$.

Let $\zeta_x = (\zeta_{\mathbf{U}_1}, \zeta_{\mathbf{U}_2}, \zeta_{\mathbf{U}_3}, \zeta_{\boldsymbol{\mathcal{G}}}) \in N_x\mathcal{M}$, and $\eta_x = (\eta_{\mathbf{U}_1}, \eta_{\mathbf{U}_2}, \eta_{\mathbf{U}_3}, \eta_{\boldsymbol{\mathcal{G}}}) \in T_x\mathcal{M}$. Since ζ_x is orthogonal to η_x, i.e., $g_x(\zeta_x, \eta_x) = 0$, the conditions

$$
\mathrm{Trace}(\mathbf{G}_d\mathbf{G}_d^T \zeta_{\mathbf{U}_d}^T \eta_{\mathbf{U}_d}) = 0, \text{ for } d \in \{1, 2, 3\}
\tag{A.3}
$$

must hold for all η_x in the tangent space. Additionally from (Absil et al., 2008), $\eta_{\mathbf{U}_d}$ has the characterization

$$
\eta_{\mathbf{U}_d} \quad = \quad \mathbf{U}_d\mathbf{\Omega} + \mathbf{U}_{d\perp}\mathbf{K},
\tag{A.4}
$$

where $\mathbf{\Omega}$ is any skew-symmetric matrix, $\mathbf{K}$ is a any matrix of size $(n_d - r_d) \times r_d$, and $\mathbf{U}_{d\perp}$ is any $n_d \times (n_d - r_d)$ that is orthogonal complement of $\mathbf{U}_d$. Let $\tilde{\zeta}_{\mathbf{U}_d} = \zeta_{\mathbf{U}_d}\mathbf{G}_d\mathbf{G}_d^T$ and let $\tilde{\zeta}_{\mathbf{U}_d}$ is defined as

$$\tilde{\zeta}_{\mathbf{U}_d} = \mathbf{U}_d\mathbf{A} + \mathbf{U}_{d\perp}\mathbf{B} \tag{A.5}$$

without loss of generality, where $\mathbf{A} \in \mathbb{R}^{r_d \times r_d}$ and $\mathbf{B} \in \mathbb{R}^{(n_d - r_d)\times r_d}$ are to be characterized from (A.3) and (A.4). A few standard computations show that $\mathbf{A}$ has to be symmetric and $\mathbf{B} = \mathbf{0}$. Consequently, $\tilde{\zeta}_{\mathbf{U}_d} = \mathbf{U}_d\mathbf{S}_{\mathbf{U}_d}$, where $\mathbf{S}_{\mathbf{U}_d} = \mathbf{S}_{\mathbf{U}_d}^T$. Equivalently, $\zeta_{\mathbf{U}_d} = \mathbf{U}_d\mathbf{S}_{\mathbf{U}_d}(\mathbf{G}_d\mathbf{G}_d^T)^{-1}$ for a symmetric matrix $\mathbf{S}_{\mathbf{U}_d}$. Finally, the normal space $N_x\mathcal{M}$ has the characterization

$$\begin{array}{rcl} N_x\mathcal{M} & = & \{(\mathbf{U}_1\mathbf{S}_{\mathbf{U}_1}(\mathbf{G}_1\mathbf{G}_1^T)^{-1}, \mathbf{U}_2\mathbf{S}_{\mathbf{U}_2}(\mathbf{G}_2\mathbf{G}_2^T)^{-1}, \mathbf{U}_3\mathbf{S}_{\mathbf{U}_3}(\mathbf{G}_3\mathbf{G}_3^T)^{-1}, 0) \\ & & : \mathbf{S}_{\mathbf{U}_d} \in \mathbb{R}^{r_d\times r_d}, \mathbf{S}_{\mathbf{U}_d}^T = \mathbf{S}_{\mathbf{U}_d}, \text{ for } d \in \{1,2,3\}\}. \end{array} \tag{A.6}$$

A.3 Characterization of the vertical space

The horizontal space projector of a tangent vector is obtained by removing the component along the vertical direction. This section shows the matrix characterization of the vertical space $\mathcal{V}_x$.

$\mathcal{V}_x$ is the defined as the linearization of the equivalence class $[(\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3, \boldsymbol{\mathcal{G}})]$ at $x = [(\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3, \boldsymbol{\mathcal{G}})]$. Equivalently, $\mathcal{V}_x$ is the linearization of $(\mathbf{U}_1\mathbf{O}_1, \mathbf{U}_2\mathbf{O}_2, \mathbf{U}_3\mathbf{O}_3, \boldsymbol{\mathcal{G}}\times_1\mathbf{O}_1^T\times_2\mathbf{O}_2^T\times_3\mathbf{O}_3^T)$ along $\mathbf{O}_d \in \mathcal{O}(r_d)$ at the *identity element* for $d \in \{1,2,3\}$. From the characterization of linearization of an orthogonal matrix (Absil et al., 2008), we have the characterization for the vertical space as

$$\begin{array}{rcl} \mathcal{V}_x & = & \{(\mathbf{U}_1\mathbf{\Omega}_1, \mathbf{U}_2\mathbf{\Omega}_2, \mathbf{U}_3\mathbf{\Omega}_3, -(\boldsymbol{\mathcal{G}}\times_1\mathbf{\Omega}_1 + \boldsymbol{\mathcal{G}}\times_2\mathbf{\Omega}_2 + \boldsymbol{\mathcal{G}}\times_3\mathbf{\Omega}_3)) : \\ & & \mathbf{\Omega}_d \in \mathbb{R}^{r_d\times r_d}, \mathbf{\Omega}_d^T = -\mathbf{\Omega}_d \text{ for } d \in \{1,2,3\}\}. \end{array} \tag{A.7}$$

A.4 Characterization of the horizontal space

The characterization of the horizontal space $\mathcal{H}_x$ is derived from its orthogonal relationship with the vertical space $\mathcal{V}_x$.

Let $\xi_x = (\xi_{\mathbf{U}_1}, \xi_{\mathbf{U}_2}, \xi_{\mathbf{U}_3}, \xi_{\boldsymbol{\mathcal{G}}}) \in \mathcal{H}_x$, and $\zeta_x = (\zeta_{\mathbf{U}_1}, \zeta_{\mathbf{U}_2}, \zeta_{\mathbf{U}_3}, \zeta_{\boldsymbol{\mathcal{G}}}) \in \mathcal{V}_x$. Since ξ_x must be orthogonal to ζ_x, which is equivalent to $g_x(\xi_x, \zeta_x) = 0$ in (A.2), the characterization for ξ_x is derived from (A.2) and (A.7).

$$\begin{array}{rcl} g_x(\xi_x,\zeta_x) & = & \langle \xi_{\mathbf{U}_1}, \zeta_{\mathbf{U}_1}(\mathbf{G}_1\mathbf{G}_1^T)\rangle + \langle \xi_{\mathbf{U}_2}, \zeta_{\mathbf{U}_2}(\mathbf{G}_2\mathbf{G}_2^T)\rangle + \langle \xi_{\mathbf{U}_3}, \zeta_{\mathbf{U}_3}(\mathbf{G}_3\mathbf{G}_3^T)\rangle + \langle \xi_{\boldsymbol{\mathcal{G}}}, \zeta_{\boldsymbol{\mathcal{G}}}\rangle \\ & = & \langle \xi_{\mathbf{U}}, (\mathbf{U}_1\mathbf{\Omega}_1)(\mathbf{G}_1\mathbf{G}_1^T)\rangle + \langle \xi_{\mathbf{U}_2}, (\mathbf{U}_2\mathbf{\Omega}_2)(\mathbf{G}_2\mathbf{G}_2^T)\rangle + \langle \xi_{\mathbf{U}_3}, (\mathbf{U}_3\mathbf{\Omega}_3)(\mathbf{G}_3\mathbf{G}_3^T)\rangle \\ & & +\langle \xi_{\boldsymbol{\mathcal{G}}}, -(\boldsymbol{\mathcal{G}}\times_1\mathbf{\Omega}_1 + \boldsymbol{\mathcal{G}}\times_2\mathbf{\Omega}_2 + \boldsymbol{\mathcal{G}}\times_3\mathbf{\Omega}_3)\rangle \\ & & \text{(Switch to unfoldings of } \boldsymbol{\mathcal{G}}.) \\ & = & \mathrm{Trace}((\mathbf{G}_1\mathbf{G}_1^T)\xi_{\mathbf{U}_1}^T(\mathbf{U}_1\mathbf{\Omega}_1)) + \mathrm{Trace}((\mathbf{G}_2\mathbf{G}_2^T)\xi_{\mathbf{U}_2}^T(\mathbf{U}_2\mathbf{\Omega}_2)) + \mathrm{Trace}((\mathbf{G}_3\mathbf{G}_3^T)\xi_{\mathbf{U}_3}^T(\mathbf{U}_3\mathbf{\Omega}_3)) \\ & & +\mathrm{Trace}(\xi_{\mathbf{G}_1}(-\mathbf{\Omega}_1\mathbf{G}_1)^T) + \mathrm{Trace}(\xi_{\mathbf{G}_2}(-\mathbf{\Omega}_2\mathbf{G}_2)^T) + \mathrm{Trace}(\xi_{\mathbf{G}_3}(-\mathbf{\Omega}_3\mathbf{G}_3)^T) \\ & = & \mathrm{Trace}\left[\left\{(\mathbf{G}_1\mathbf{G}_1^T)\xi_{\mathbf{U}_1}^T\mathbf{U}_1 + \xi_{\mathbf{G}_1}\mathbf{G}_1^T\right\}\mathbf{\Omega}_1\right] + \mathrm{Trace}\left[\left\{(\mathbf{G}_2\mathbf{G}_2^T)\xi_{\mathbf{U}_2}^T\mathbf{U}_2 + \xi_{\mathbf{G}_2}\mathbf{G}_2^T\right\}\mathbf{\Omega}_2\right] \\ & & +\mathrm{Trace}\left[\left\{(\mathbf{G}_3\mathbf{G}_3^T)\xi_{\mathbf{U}_3}^T\mathbf{U}_3 + \xi_{\mathbf{G}_3}\mathbf{G}_3^T\right\}\mathbf{\Omega}_3\right], \end{array}$$

where $\xi_{\mathbf{G}_d}$ is the mode-d unfolding of $\xi_{\boldsymbol{\mathcal{G}}}$. Since $g_x(\xi_x, \zeta_x)$ above should be zero for all skew-matrices $\mathbf{\Omega}_d$, $\xi_x = (\xi_{\mathbf{U}_1}, \xi_{\mathbf{U}_2}, \xi_{\mathbf{U}_3}, \xi_{\boldsymbol{\mathcal{G}}}) \in \mathcal{H}_x$ must satisfy

$$(\mathbf{G}_d\mathbf{G}_d^T)\xi_{\mathbf{U}_d}^T\mathbf{U}_d + \xi_{\mathbf{G}_d}\mathbf{G}_d^T \quad \text{is symmetric for } d \in \{1,2,3\}. \tag{A.8}$$

A.5 Proof of Proposition 1

We first introduce the following lemma:

Lemma 1. *Let* $(\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3, \boldsymbol{\mathcal{G}}) \in \mathrm{St}(r_1, n_1) \times \mathrm{St}(r_2, n_2) \times \mathrm{St}(r_3, n_3) \times \mathbb{R}^{r_1\times r_2\times r_3}$ *and* $\xi_{[(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3,\boldsymbol{\mathcal{G}})]}$ *be a tangent vector to the quotient manifold at* $[(\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3, \boldsymbol{\mathcal{G}})]$. *The horizontal lifts of* $\xi_{[(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3,\boldsymbol{\mathcal{G}})]}$ *at* $(\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3, \boldsymbol{\mathcal{G}})$ *and* $(\mathbf{U}_1\mathbf{O}_1, \mathbf{U}_2\mathbf{O}_2, \mathbf{U}_3\mathbf{O}_3, \boldsymbol{\mathcal{G}}\times_1\mathbf{O}_1^T\times_2\mathbf{O}_2^T\times_3\mathbf{O}_3^T)$ *are related for* $\mathbf{O}_d \in \mathcal{O}(r_d)$ *as follows,*

$$(\xi_{\mathbf{U}_1\mathbf{O}_1}, \xi_{\mathbf{U}_2\mathbf{O}_2}, \xi_{\mathbf{U}_3\mathbf{O}_3}, \xi_{\boldsymbol{\mathcal{G}}\times_1\mathbf{O}_1^T\times_2\mathbf{O}_2^T\times_3\mathbf{O}_3^T}) = (\xi_{\mathbf{U}_1}\mathbf{O}_1, \xi_{\mathbf{U}_2}\mathbf{O}_2, \xi_{\mathbf{U}_3}\mathbf{O}_3, \xi_{\boldsymbol{\mathcal{G}}}\times_1\mathbf{O}_1^T\times_2\mathbf{O}_2^T\times_3\mathbf{O}_3^T). \tag{A.9}$$

Proof. Let $f : (\mathrm{St}(r_1,n_1)\times \mathrm{St}(r_2,n_2)\times \mathrm{St}(r_3,n_3)\times \mathbb{R}^{r_1\times r_2\times r_3}/(\mathcal{O}(r_1)\times\mathcal{O}(r_2)\times\mathcal{O}(r_3))) \rightarrow \mathbb{R}$ be an arbitrary smooth function, and define

$$\bar{f} \quad := f\circ\pi : (\mathrm{St}(r_1,n_1)\times \mathrm{St}(r_2,n_2)\times \mathrm{St}(r_3,n_3)\times \mathbb{R}^{r_1\times r_2\times r_3}/(\mathcal{O}(r_1)\times\mathcal{O}(r_2)\times\mathcal{O}(r_3))) \rightarrow \mathbb{R},$$

where π is the mapping $\pi : \mathcal{M}\rightarrow\mathcal{M}/\sim$ defined by $x\mapsto[x]$.

Consider the mapping

$$h : (\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3,\boldsymbol{\mathcal{G}}) \mapsto (\mathbf{U}_1\mathbf{O}_1,\mathbf{U}_2\mathbf{O}_2,\mathbf{U}_3\mathbf{O}_3,\boldsymbol{\mathcal{G}}\times_1\mathbf{O}_1^T\times_2\mathbf{O}_2^T\times_3\mathbf{O}_3^T),$$

where $\mathbf{O}_d\in\mathcal{O}(r_d)$. Since $\pi(h(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3,\boldsymbol{\mathcal{G}}))=\pi(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3,\boldsymbol{\mathcal{G}})$ for all $(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3,\boldsymbol{\mathcal{G}})$, we have

$$\bar{f}(h(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3,\boldsymbol{\mathcal{G}}))=\bar{f}(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3,\boldsymbol{\mathcal{G}}).$$

By taking the differential of both sides,

$$\mathrm{D}\bar{f}(h(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3,\boldsymbol{\mathcal{G}}))[\mathrm{D}h(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3,\boldsymbol{\mathcal{G}})[(\xi_{\mathbf{U}_1},\xi_{\mathbf{U}_2},\xi_{\mathbf{U}_3},\xi_{\boldsymbol{\mathcal{G}}})]]=\mathrm{D}\bar{f}(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3,\boldsymbol{\mathcal{G}})[(\xi_{\mathbf{U}_1},\xi_{\mathbf{U}_2},\xi_{\mathbf{U}_3},\xi_{\boldsymbol{\mathcal{G}}})]. \tag{A.10}$$

By noting the definition of $(\xi_{\mathbf{U}_1},\xi_{\mathbf{U}_2},\xi_{\mathbf{U}_3},\xi_{\boldsymbol{\mathcal{G}}})$, i.e., $\mathrm{D}\pi(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3,\boldsymbol{\mathcal{G}})[\xi_{\mathbf{U}_1},\xi_{\mathbf{U}_2},\xi_{\mathbf{U}_3},\xi_{\boldsymbol{\mathcal{G}}}]=\xi_{[(\xi_{\mathbf{U}_1},\xi_{\mathbf{U}_2},\xi_{\mathbf{U}_3},\xi_{\boldsymbol{\mathcal{G}}})]}$, the right side of (A.10) is

$$\begin{aligned}\mathrm{D}\bar{f}(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3,\boldsymbol{\mathcal{G}})[(\xi_{\mathbf{U}_1},\xi_{\mathbf{U}_2},\xi_{\mathbf{U}_3},\xi_{\boldsymbol{\mathcal{G}}})] &= \mathrm{D}f(\pi(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3,\boldsymbol{\mathcal{G}}))[\mathrm{D}\pi(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3,\boldsymbol{\mathcal{G}})[(\xi_{\mathbf{U}_1},\xi_{\mathbf{U}_2},\xi_{\mathbf{U}_3},\xi_{\boldsymbol{\mathcal{G}}})]\\ &= \mathrm{D}f(\pi(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3,\boldsymbol{\mathcal{G}}))[\xi_{[(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3,\boldsymbol{\mathcal{G}})]}],\end{aligned}$$

where the chain rule is applied to the first equality.

Moreover, from the directional derivatives of the mapping h, the bracket of the left side of (A.10) is obtained as

$$\mathrm{D}h(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3,\boldsymbol{\mathcal{G}})[(\xi_{\mathbf{U}_1},\xi_{\mathbf{U}_2},\xi_{\mathbf{U}_3},\xi_{\boldsymbol{\mathcal{G}}})] \quad = \quad (\xi_{\mathbf{U}_1}\mathbf{O}_1,\xi_{\mathbf{U}_2}\mathbf{O}_2,\xi_{\mathbf{U}_3}\mathbf{O}_3,\xi_{\boldsymbol{\mathcal{G}}}\times_1\mathbf{O}_1^T\times_2\mathbf{O}_2^T\times_3\mathbf{O}_3^T).$$

Therefore, (A.10) yields

$$\begin{aligned}&\mathrm{D}\bar{f}(\mathbf{U}_1\mathbf{O}_1,\mathbf{U}_2\mathbf{O}_2,\mathbf{U}_3\mathbf{O}_3,\boldsymbol{\mathcal{G}}\times_1\mathbf{O}_1^T\times_2\mathbf{O}_2^T\times_3\mathbf{O}_3^T)[(\xi_{\mathbf{U}_1}\mathbf{O}_1,\xi_{\mathbf{U}_2}\mathbf{O}_2,\xi_{\mathbf{U}_3}\mathbf{O}_3,\xi_{\boldsymbol{\mathcal{G}}}\times_1\mathbf{O}_1^T\times_2\mathbf{O}_2^T\times_3\mathbf{O}_3^T)]\\ &\quad=\mathrm{D}f(\pi(\mathbf{U}_1\mathbf{O}_1,\mathbf{U}_2\mathbf{O}_2,\mathbf{U}_3\mathbf{O}_3,\boldsymbol{\mathcal{G}}\times_1\mathbf{O}_1^T\times_2\mathbf{O}_2^T\times_3\mathbf{O}_3^T))[\xi_{[(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3,\boldsymbol{\mathcal{G}})]}],\end{aligned} \tag{A.11}$$

where we address the equivalence class $\pi(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3,\boldsymbol{\mathcal{G}})=\pi(\mathbf{U}_1\mathbf{O}_1,\mathbf{U}_2\mathbf{O}_2,\mathbf{U}_3\mathbf{O}_3,\boldsymbol{\mathcal{G}}\times_1\mathbf{O}_1^T\times_2\mathbf{O}_2^T\times_3\mathbf{O}_3^T)$. The left side of (A.11) is further transformed by the chain rule as

$$\begin{aligned}&\mathrm{D}\bar{f}(\mathbf{U}_1\mathbf{O}_1,\mathbf{U}_2\mathbf{O}_2,\mathbf{U}_3\mathbf{O}_3,\boldsymbol{\mathcal{G}}\times_1\mathbf{O}_1^T\times_2\mathbf{O}_2^T\times_3\mathbf{O}_3^T)[(\xi_{\mathbf{U}_1}\mathbf{O}_1,\xi_{\mathbf{U}_2}\mathbf{O}_2,\xi_{\mathbf{U}_3}\mathbf{O}_3,\xi_{\boldsymbol{\mathcal{G}}}\times_1\mathbf{O}_1^T\times_2\mathbf{O}_2^T\times_3\mathbf{O}_3^T)]\\ &\quad=\mathrm{D}f(\pi(\mathbf{U}_1\mathbf{O}_1,\mathbf{U}_2\mathbf{O}_2,\mathbf{U}_3\mathbf{O}_3,\boldsymbol{\mathcal{G}}\times_1\mathbf{O}_1^T\times_2\mathbf{O}_2^T\times_3\mathbf{O}_3^T))[\mathrm{D}\pi(\mathbf{U}_1\mathbf{O}_1,\mathbf{U}_2\mathbf{O}_2,\mathbf{U}_3\mathbf{O}_3,\boldsymbol{\mathcal{G}}\times_1\mathbf{O}_1^T\times_2\mathbf{O}_2^T\times_3\mathbf{O}_3^T)]\\ &\qquad\quad[(\xi_{\mathbf{U}_1}\mathbf{O}_1,\xi_{\mathbf{U}_2}\mathbf{O}_2,\xi_{\mathbf{U}_3}\mathbf{O}_3,\xi_{\boldsymbol{\mathcal{G}}}\times_1\mathbf{O}_1^T\times_2\mathbf{O}_2^T\times_3\mathbf{O}_3^T)].\end{aligned} \tag{A.12}$$

By comparing the right sides of (A.11) and (A.12), since this equality holds for any smooth function f, it implies that

$$\begin{aligned}&\mathrm{D}\pi(\mathbf{U}_1\mathbf{O}_1,\mathbf{U}_2\mathbf{O}_2,\mathbf{U}_3\mathbf{O}_3,\boldsymbol{\mathcal{G}}\times_1\mathbf{O}_1^T\times_2\mathbf{O}_2^T\times_3\mathbf{O}_3^T)[(\xi_{\mathbf{U}_1}\mathbf{O}_1,\xi_{\mathbf{U}_2}\mathbf{O}_2,\xi_{\mathbf{U}_3}\mathbf{O}_3,\xi_{\boldsymbol{\mathcal{G}}}\times_1\mathbf{O}_1^T\times_2\mathbf{O}_2^T\times_3\mathbf{O}_3^T)]\\ &\quad=\xi_{[(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3,\boldsymbol{\mathcal{G}})]}.\end{aligned} \tag{A.13}$$

Finally, we check whether $(\xi_{\mathbf{U}_1}\mathbf{O}_1,\xi_{\mathbf{U}_2}\mathbf{O}_2,\xi_{\mathbf{U}_3}\mathbf{O}_3,\xi_{\boldsymbol{\mathcal{G}}}\times_1\mathbf{O}_1^T\times_2\mathbf{O}_2^T\times_3\mathbf{O}_3^T)$ is an element of $\mathcal{H}_{(\mathbf{U}_1\mathbf{O}_1,\mathbf{U}_2\mathbf{O}_2,\mathbf{U}_3\mathbf{O}_3,\boldsymbol{\mathcal{G}}\times_1\mathbf{O}_1^T\times_2\mathbf{O}_2^T\times_3\mathbf{O}_3^T)}$. Addressing that the mode-1 unfolding of $\boldsymbol{\mathcal{G}}\times_1\mathbf{O}_1^T\times_2\mathbf{O}_2^T\times_3\mathbf{O}_3^T$ is $\mathbf{O}_1^T\mathbf{G}_1(\mathbf{O}_3^T\otimes\mathbf{O}_2^T)^T$, plugging $(\xi_{\mathbf{U}_1}\mathbf{O}_1,\xi_{\mathbf{U}_2}\mathbf{O}_2,\xi_{\mathbf{U}_3}\mathbf{O}_3,\xi_{\boldsymbol{\mathcal{G}}}\times_1\mathbf{O}_1^T\times_2\mathbf{O}_2^T\times_3\mathbf{O}_3^T)$ into $(\mathbf{G}_d\mathbf{G}_d^T)\xi_{\mathbf{U}_d}^T\mathbf{U}_d+\xi_{\mathbf{G}_d}\mathbf{G}_d^T$ in (A.8) yields

$$\begin{aligned}&(\mathbf{O}_1^T\mathbf{G}_1(\mathbf{O}_3^T\otimes\mathbf{O}_2^T)^T)\left(\mathbf{O}_1^T\mathbf{G}_1(\mathbf{O}_3^T\otimes\mathbf{O}_2^T)^T\right)^T(\xi_{\mathbf{U}_1}\mathbf{O}_1)^T(\mathbf{U}_1\mathbf{O}_1)+(\mathbf{O}_1^T)^T\xi_{\mathbf{G}_1}(\mathbf{O}_3^T\otimes\mathbf{O}_2^T)(\mathbf{O}_1^T\mathbf{G}_1(\mathbf{O}_3^T\otimes\mathbf{O}_2^T)^T)^T\\ &\quad=\mathbf{O}_1^T\mathbf{G}_1\mathbf{G}_1^T\mathbf{O}_1+\mathbf{O}_1^T\xi_{\mathbf{G}_1}\mathbf{G}_1^T\mathbf{O}_1\\ &\quad=\mathbf{O}_1^T((\mathbf{G}_1\mathbf{G}_1^T)\xi_{\mathbf{U}_1}^T\mathbf{U}_1+\xi_{\mathbf{G}_1}\mathbf{G}_1^T)\mathbf{O}_1.\end{aligned} \tag{A.14}$$

Since $(\xi_{\mathbf{U}_1},\xi_{\mathbf{U}_2},\xi_{\mathbf{U}_3},\xi_{\boldsymbol{\mathcal{G}}})$ is a symmetric matrix, the obtained result is also symmetric. Therefore, $(\xi_{\mathbf{U}_1}\mathbf{O}_1,\xi_{\mathbf{U}_2}\mathbf{O}_2,\xi_{\mathbf{U}_3}\mathbf{O}_3,\xi_{\boldsymbol{\mathcal{G}}}\times_1\mathbf{O}_1^T\times_2\mathbf{O}_2^T\times_3\mathbf{O}_3^T)$ is a horizontal vector at $(\mathbf{U}_1\mathbf{O}_1,\mathbf{U}_2\mathbf{O}_2,\mathbf{U}_3\mathbf{O}_3,\boldsymbol{\mathcal{G}}\times_1\mathbf{O}_1^T\times_2\mathbf{O}_2^T\times_3\mathbf{O}_3^T)$. This implies that $(\xi_{\mathbf{U}_1}\mathbf{O}_1,\xi_{\mathbf{U}_2}\mathbf{O}_2,\xi_{\mathbf{U}_3}\mathbf{O}_3,\xi_{\boldsymbol{\mathcal{G}}}\times_1\mathbf{O}_1^T\times_2\mathbf{O}_2^T\times_3\mathbf{O}_3^T)$ is the horizontal lift of ξ at $(\mathbf{U}_1\mathbf{O}_1,\mathbf{U}_2\mathbf{O}_2,\mathbf{U}_3\mathbf{O}_3,\boldsymbol{\mathcal{G}}\times_1\mathbf{O}_1^T\times_2\mathbf{O}_2^T\times_3\mathbf{O}_3^T)$, and the proof is completed. $\square$

Now, the proof of **Proposition 1** is given below using the result (A.9) in **Lemma 1**.

Proof. Plugging $\xi'_{\mathbf{U}_1} = \xi_{\mathbf{U}_1\mathbf{O}_1}$, $\eta'_{\mathbf{U}_1} = \eta_{\mathbf{U}_1\mathbf{O}_1}$, and $\mathbf{G}'_1 = \mathbf{O}_1^T\mathbf{G}_1(\mathbf{O}_3^T \otimes \mathbf{O}_2^T)$ into the first term of (A.2) yields

$$
\begin{array}{rcl}
\langle \xi_{\mathbf{U}_1\mathbf{O}_1}, \eta_{\mathbf{U}_1\mathbf{O}_1}(\mathbf{G}'_1\mathbf{G}'^T_1)\rangle & = & \mathrm{Trace}(\xi^T_{\mathbf{U}_1\mathbf{O}_1}\eta_{\mathbf{U}_1\mathbf{O}_1}(\mathbf{G}'_1\mathbf{G}'^T_1)) \\
& \overset{(A.9)}{=} & \mathrm{Trace}((\xi_{\mathbf{U}_1}\mathbf{O}_1)^T\eta_{\mathbf{U}_1}\mathbf{O}_1(\mathbf{G}'_1\mathbf{G}'^T_1)) \\
& = & \mathrm{Trace}\left[(\xi_{\mathbf{U}_1}\mathbf{O}_1)^T(\eta_{\mathbf{U}_1}\mathbf{O}_1)(\mathbf{O}_1^T\mathbf{G}_1(\mathbf{O}_3^T\otimes\mathbf{O}_2^T)^T)\left(\mathbf{O}_1^T\mathbf{G}_1(\mathbf{O}_3^T\otimes\mathbf{O}_2^T)^T\right)^T\right] \\
& = & \mathrm{Trace}\left[(\xi_{\mathbf{U}_1}\mathbf{O}_1)^T(\eta_{\mathbf{U}_1}\mathbf{O}_1)\mathbf{O}_1^T\mathbf{G}_1(\mathbf{O}_3^T\otimes\mathbf{O}_2^T)^T(\mathbf{O}_3^T\otimes\mathbf{O}_2^T)\mathbf{G}_1^T\mathbf{O}_1\right] \\
& = & \mathrm{Trace}\left[\mathbf{O}_1^T\xi^T_{\mathbf{U}_1}\eta_{\mathbf{U}_1}\mathbf{O}_1\mathbf{O}_1^T\mathbf{G}_1\mathbf{G}_1^T\mathbf{O}_1\right] \\
& = & \mathrm{Trace}\left[\xi^T_{\mathbf{U}_1}\eta_{\mathbf{U}_1}\mathbf{G}_1\mathbf{G}_1^T\right] \\
& = & \langle \xi_{\mathbf{U}_1}, \eta_{\mathbf{U}_1}(\mathbf{G}_1\mathbf{G}_1^T)\rangle.
\end{array}
$$

Since the same equalities against the each term in the metric (A.2) corresponding to $\mathbf{U}_2$, $\mathbf{U}_3$ and $\boldsymbol{\mathcal{G}}$ hold, we finally obtain the invariant property that the proposition claims;

$$
\begin{array}{l}
g_{(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3,\boldsymbol{\mathcal{G}})}((\xi_{\mathbf{U}_1},\xi_{\mathbf{U}_2},\xi_{\mathbf{U}_3},\xi_{\boldsymbol{\mathcal{G}}}),(\eta_{\mathbf{U}_1},\eta_{\mathbf{U}_2},\eta_{\mathbf{U}_3},\eta_{\boldsymbol{\mathcal{G}}})) \\
\quad = g_{(\mathbf{U}_1\mathbf{O}_1,\mathbf{U}_2\mathbf{O}_2,\mathbf{U}_3\mathbf{O}_3,\boldsymbol{\mathcal{G}}\times_1\mathbf{O}_1^T\times_2\mathbf{O}_2^T\times_3\mathbf{O}_3^T)}((\xi_{\mathbf{U}_1\mathbf{O}_1},\xi_{\mathbf{U}_2\mathbf{O}_2},\xi_{\mathbf{U}_3\mathbf{O}_3},\xi_{\boldsymbol{\mathcal{G}}\times_1\mathbf{O}_1^T\times_2\mathbf{O}_2^T\times_3\mathbf{O}_3^T}), \\
\qquad (\eta_{\mathbf{U}_1\mathbf{O}_1},\eta_{\mathbf{U}_2\mathbf{O}_2},\eta_{\mathbf{U}_3\mathbf{O}_3},\eta_{\boldsymbol{\mathcal{G}}\times_1\mathbf{O}_1^T\times_2\mathbf{O}_2^T\times_3\mathbf{O}_3^T})).
\end{array}
$$

□

A.6 Proof of Proposition 2 (derivation of the tangent space projector)

Proof. The tangent space $T_x\mathcal{M}$ projector is obtained by extracting the component normal to $T_x\mathcal{M}$ in the ambient space. The normal space $N_x\mathcal{M}$ has the matrix characterization shown in (A.6). The operator $\Psi_x : \mathbb{R}^{n_1\times r_1}\times\mathbb{R}^{n_2\times r_2}\times\mathbb{R}^{n_3\times r_3}\times\mathbb{R}^{r_1\times r_2\times r_3}\to T_x\mathcal{M} : (\mathbf{Y}_{\mathbf{U}_1},\mathbf{Y}_{\mathbf{U}_2},\mathbf{Y}_{\mathbf{U}_3},\mathbf{Y}_{\boldsymbol{\mathcal{G}}})\mapsto\Psi_x(\mathbf{Y}_{\mathbf{U}_1},\mathbf{Y}_{\mathbf{U}_2},\mathbf{Y}_{\mathbf{U}_3},\mathbf{Y}_{\boldsymbol{\mathcal{G}}})$ has the expression

$$
\Psi_x(\mathbf{Y}_{\mathbf{U}_1},\mathbf{Y}_{\mathbf{U}_2},\mathbf{Y}_{\mathbf{U}_3},\mathbf{Y}_{\boldsymbol{\mathcal{G}}}) = (\mathbf{Y}_{\mathbf{U}_1}-\mathbf{U}_1\mathbf{S}_{\mathbf{U}_1}(\mathbf{G}_1\mathbf{G}_1^T)^{-1},\mathbf{Y}_{\mathbf{U}_2}-\mathbf{U}_2\mathbf{S}_{\mathbf{U}_2}(\mathbf{G}_2\mathbf{G}_2^T)^{-1},\mathbf{Y}_{\mathbf{U}_3}-\mathbf{U}_3\mathbf{S}_{\mathbf{U}_3}(\mathbf{G}_3\mathbf{G}_3^T)^{-1},\mathbf{Y}_{\boldsymbol{\mathcal{G}}}). \tag{A.15}
$$

From the definition of the tangent space in (A.1), $\mathbf{U}_d$ should satisfy

$$
\begin{array}{rcl}
\eta^T_{\mathbf{U}_d}\mathbf{U}_d + \mathbf{U}_d^T\eta_{\mathbf{U}_d} & = & (\mathbf{Y}_{\mathbf{U}_d}-\mathbf{U}_d\mathbf{S}_{\mathbf{U}_d}(\mathbf{G}_d\mathbf{G}_d^T)^{-1})^T\mathbf{U}_d + \mathbf{U}_d^T(\mathbf{Y}_{\mathbf{U}_d}-\mathbf{U}_d\mathbf{S}_{\mathbf{U}_d}(\mathbf{G}_d\mathbf{G}_d^T)^{-1}) \\
& = & \mathbf{Y}^T_{\mathbf{U}_d}\mathbf{U}_d - (\mathbf{G}_d\mathbf{G}_d^T)^{-1}\mathbf{S}^T_{\mathbf{U}_d}\mathbf{U}_d^T\mathbf{U}_d + \mathbf{U}_d^T\mathbf{Y}_{\mathbf{U}_d} - \mathbf{U}_d^T\mathbf{U}_d\mathbf{S}_{\mathbf{U}_d}(\mathbf{G}_d\mathbf{G}_d^T)^{-1} \\
& = & \mathbf{Y}^T_{\mathbf{U}_d}\mathbf{U}_d - (\mathbf{G}_d\mathbf{G}_d^T)^{-1}\mathbf{S}_{\mathbf{U}_d} + \mathbf{U}_d^T\mathbf{Y}_{\mathbf{U}_d} - \mathbf{S}_{\mathbf{U}_d}(\mathbf{G}_d\mathbf{G}_d^T)^{-1} \\
& = & 0.
\end{array}
$$

Multiplying $(\mathbf{G}_d\mathbf{G}_d^T)$ from the right and left sides results in

$$
\begin{array}{rcl}
(\mathbf{G}_d\mathbf{G}_d^T)^{-1}\mathbf{S}_{\mathbf{U}_d} + \mathbf{S}_{\mathbf{U}_d}(\mathbf{G}_d\mathbf{G}_d^T)^{-1} & = & \mathbf{Y}^T_{\mathbf{U}_d}\mathbf{U}_d + \mathbf{U}_d^T\mathbf{Y}_{\mathbf{U}_d} \\
\mathbf{S}_{\mathbf{U}_d}\mathbf{G}_d\mathbf{G}_d^T + \mathbf{G}_d\mathbf{G}_d^T\mathbf{S}_{\mathbf{U}_d} & = & \mathbf{G}_d\mathbf{G}_d^T(\mathbf{Y}^T_{\mathbf{U}_d}\mathbf{U}_d + \mathbf{U}_d^T\mathbf{Y}_{\mathbf{U}_d})\mathbf{G}_d\mathbf{G}_d^T.
\end{array}
$$

Finally, we obtain the *Lyapunov* equation as

$$
\mathbf{S}_{\mathbf{U}_d}\mathbf{G}_d\mathbf{G}_d^T + \mathbf{G}_d\mathbf{G}_d^T\mathbf{S}_{\mathbf{U}_d} = \mathbf{G}_d\mathbf{G}_d^T(\mathbf{Y}^T_{\mathbf{U}_d}\mathbf{U}_d + \mathbf{U}_d^T\mathbf{Y}_{\mathbf{U}_d})\mathbf{G}_d\mathbf{G}_d^T \ \text{ for } d\in\{1,2,3\}, \tag{A.16}
$$

that are solved efficiently with the Matlab's `lyap` routine.

□

A.7 Proof of Proposition 3 (derivation of the horizontal space projector)

Proof. We consider the projection of a tangent vector $\eta_x = (\eta_{\mathbf{U}_1}, \eta_{\mathbf{U}_2}, \eta_{\mathbf{U}_3}, \eta_{\mathcal{G}}) \in T_x\mathcal{M}$ into a vector $\xi_x = (\xi_{\mathbf{U}_1}, \xi_{\mathbf{U}_2}, \xi_{\mathbf{U}_3}, \xi_{\mathcal{G}}) \in H_x$. This is achieved by subtracting the component in the vertical space $\mathcal{V}_x$ in (A.7) as

$$\left\{\begin{array}{lcl} \eta_{\mathbf{U}_1} & = & \underbrace{\eta_{\mathbf{U}_1} - \mathbf{U}_1\mathbf{\Omega}_1}_{=\xi_{\mathbf{U}_1}\in\mathcal{H}_x} + \underbrace{\mathbf{U}_1\mathbf{\Omega}_1}_{\in\mathcal{V}_x}, \\ \eta_{\mathbf{U}_2} & = & \eta_{\mathbf{U}_2} - \mathbf{U}_2\mathbf{\Omega}_2 + \mathbf{U}_2\mathbf{\Omega}_2, \\ \eta_{\mathbf{U}_3} & = & \eta_{\mathbf{U}_3} - \mathbf{U}_3\mathbf{\Omega}_3 + \mathbf{U}_3\mathbf{\Omega}_3, \\ \eta_{\mathcal{G}} & = & \eta_{\mathcal{G}} - (-(\mathcal{G}\times_1\mathbf{\Omega}_1 + \mathcal{G}\times_2\mathbf{\Omega}_2 + \mathcal{G}\times_3\mathbf{\Omega}_3)) + (-(\mathcal{G}\times_1\mathbf{\Omega}_1 + \mathcal{G}\times_2\mathbf{\Omega}_2 + \mathcal{G}\times_3\mathbf{\Omega}_3)). \end{array}\right.$$

As a result, the horizontal operator $\Pi_x : T_x\mathcal{M} \to \mathcal{H}_x : \eta_x \mapsto \Pi_x(\eta_x)$ has the expression

$$\Pi_x(\eta_x) \quad = \quad (\eta_{\mathbf{U}_1} - \mathbf{U}_1\mathbf{\Omega}_1, \eta_{\mathbf{U}_2} - \mathbf{U}_2\mathbf{\Omega}_2, \eta_{\mathbf{U}_3} - \mathbf{U}_3\mathbf{\Omega}_3, \eta_{\mathcal{G}} - (-(\mathcal{G}\times_1\mathbf{\Omega}_1 + \mathcal{G}\times_2\mathbf{\Omega}_2 + \mathcal{G}\times_3\mathbf{\Omega}_3))), \tag{A.17}$$

where $\eta_x = (\eta_{\mathbf{U}_1}, \eta_{\mathbf{U}_2}, \eta_{\mathbf{U}_3}, \eta_{\mathcal{G}}) \in T_x\mathcal{M}$ and $\mathbf{\Omega}_d$ is a skew-symmetric matrix of size $r_d \times r_d$. The skew-matrices $\mathbf{\Omega}_d$ for $d = \{1, 2, 3\}$ that are identified based on the conditions (A.8).

It should be noted that the tensor $\mathcal{G}\times_1\mathbf{\Omega}_1 + \mathcal{G}\times_2\mathbf{\Omega}_2 + \mathcal{G}\times_3\mathbf{\Omega}_3$ in (A.7) has the following equivalent unfoldings.

$$\begin{array}{lcl} \mathcal{G}\times_1\mathbf{\Omega}_1 + \mathcal{G}\times_2\mathbf{\Omega}_2 + \mathcal{G}\times_3\mathbf{\Omega}_3 & \overset{\text{mode }-1}{\Longleftrightarrow} & \mathbf{\Omega}_1\mathbf{G}_1 + \mathbf{G}_1(\mathbf{I}_{r_3}\otimes\mathbf{\Omega}_2)^T + \mathbf{G}_1(\mathbf{\Omega}_3\otimes\mathbf{I}_{r_2})^T \\ & \overset{\text{mode }-2}{\Longleftrightarrow} & \mathbf{G}_2(\mathbf{I}_{r_3}\otimes\mathbf{\Omega}_1)^T + \mathbf{\Omega}_2\mathbf{G}_2 + \mathbf{G}_2(\mathbf{\Omega}_3\otimes\mathbf{I}_{r_1})^T \\ & \overset{\text{mode }-3}{\Longleftrightarrow} & \mathbf{G}_3(\mathbf{I}_{r_2}\otimes\mathbf{\Omega}_1)^T + \mathbf{G}_3(\mathbf{\Omega}_2\otimes\mathbf{I}_{r_1})^T + \mathbf{\Omega}_3\mathbf{G}_3. \end{array}$$

Plugging $\xi_{\mathbf{U}_1} = \eta_{\mathbf{U}_1} - \mathbf{U}_1\mathbf{\Omega}_1$ and $\xi_{\mathbf{G}_1} = \eta_{\mathbf{G}_1} + \mathbf{\Omega}_1\mathbf{G}_1 + \mathbf{G}_1(\mathbf{I}_{r_3}\otimes\mathbf{\Omega}_2)^T + \mathbf{G}_1(\mathbf{\Omega}_3\otimes\mathbf{I}_{r_2})^T$ into (A.8) and using the relation $(\mathbf{A}\otimes\mathbf{B})^T = \mathbf{A}^T\otimes\mathbf{B}^T$ results in

$$\begin{array}{lcl} (\mathbf{G}_1\mathbf{G}_1^T)\xi_{\mathbf{U}_1}^T\mathbf{U}+\xi_{\mathbf{G}_1}\mathbf{G}_1^T & = & (\mathbf{G}_1\mathbf{G}_1^T)(\eta_{\mathbf{U}_1} - \mathbf{U}_1\mathbf{\Omega}_1)^T\mathbf{U}_1 + \left\{\eta_{\mathbf{G}_1} + (\mathbf{\Omega}_1\mathbf{G}_1 + \mathbf{G}_1(\mathbf{I}_{r_3}\otimes\mathbf{\Omega}_2)^T + \mathbf{G}_1(\mathbf{\Omega}_3\otimes\mathbf{I}_{r_2})^T)\right\}\mathbf{G}_1^T \\ & = & (\mathbf{G}_1\mathbf{G}_1^T)\eta_{\mathbf{U}_1}^T\mathbf{U}_1 - (\mathbf{G}_1\mathbf{G}_1^T)(\mathbf{U}_1\mathbf{\Omega}_1)^T\mathbf{U}_1 + \eta_{\mathbf{G}_1}\mathbf{G}_1^T + \mathbf{\Omega}_1\mathbf{G}_1\mathbf{G}_1^T + \mathbf{G}_1(\mathbf{I}_{r_3}\otimes\mathbf{\Omega}_2)^T\mathbf{G}_1^T \\ & & +\mathbf{G}_1(\mathbf{\Omega}_3\otimes\mathbf{I}_{r_2})^T\mathbf{G}_1^T \\ & = & (\mathbf{G}_1\mathbf{G}_1^T)\eta_{\mathbf{U}_1}^T\mathbf{U}_1 + (\mathbf{G}_1\mathbf{G}_1^T)\mathbf{\Omega}_1 + \eta_{\mathbf{G}_1}\mathbf{G}_1^T + \mathbf{\Omega}_1\mathbf{G}_1\mathbf{G}_1^T - \mathbf{G}_1(\mathbf{I}_{r_3}\otimes\mathbf{\Omega}_2)\mathbf{G}_1^T - \mathbf{G}_1(\mathbf{\Omega}_3\otimes\mathbf{I}_{r_2})\mathbf{G}_1^T, \end{array}$$

which should be a symmetric matrix due to (A.8), i.e., $(\mathbf{G}_1\mathbf{G}_1^T)\xi_{\mathbf{U}_1}^T\mathbf{U}+\xi_{\mathbf{G}_1}\mathbf{G}_1^T = ((\mathbf{G}_1\mathbf{G}_1^T)\xi_{\mathbf{U}_1}^T\mathbf{U}+\xi_{\mathbf{G}_1}\mathbf{G}_1^T)^T$.

Subsequently,

$$\begin{array}{l} (\mathbf{G}_1\mathbf{G}_1^T)\eta_{\mathbf{U}_1}^T\mathbf{U}_1 + (\mathbf{G}_1\mathbf{G}_1^T)\mathbf{\Omega}_1 + \eta_{\mathbf{G}_1}\mathbf{G}_1^T + \mathbf{\Omega}_1\mathbf{G}_1\mathbf{G}_1^T - \mathbf{G}_1(\mathbf{I}_{r_3}\otimes\mathbf{\Omega}_2)\mathbf{G}_1^T - \mathbf{G}_1(\mathbf{\Omega}_3\otimes\mathbf{I}_{r_2})\mathbf{G}_1^T \\ \quad = \mathbf{U}_1^T\eta_{\mathbf{U}_1}(\mathbf{G}_1\mathbf{G}_1^T) - \mathbf{\Omega}_1\mathbf{G}_1\mathbf{G}_1^T + \mathbf{G}_1\eta_{\mathbf{G}_1}^T - \mathbf{G}_1\mathbf{G}_1^T\mathbf{\Omega}_1 + \mathbf{G}_1(\mathbf{I}_{r_3}\otimes\mathbf{\Omega}_2)\mathbf{G}_1^T + \mathbf{G}_1(\mathbf{\Omega}_3\otimes\mathbf{I}_{r_2})\mathbf{G}_1^T, \end{array}$$

which is equivalent to

$$\mathbf{G}_1\mathbf{G}_1^T\mathbf{\Omega}_1 + \mathbf{\Omega}_1\mathbf{G}_1\mathbf{G}_1^T - \mathbf{G}_1(\mathbf{I}_{r_3}\otimes\mathbf{\Omega}_2)\mathbf{G}_1^T - \mathbf{G}_1(\mathbf{\Omega}_3\otimes\mathbf{I}_{r_2})\mathbf{G}_1^T \quad = \quad \mathrm{Skew}(\mathbf{U}_1^T\eta_{\mathbf{U}_1}\mathbf{G}_1\mathbf{G}_1^T) + \mathrm{Skew}(\mathbf{G}_1\eta_{\mathbf{G}_1}^T).$$

Here $\mathrm{Skew}(\cdot)$ extracts the skew-symmetric part of a square matrix, i.e., $\mathrm{Skew}(\mathbf{D}) = (\mathbf{D} - \mathbf{D}^T)/2$.

Finally, we obtain the *coupled* Lyapunov equations

$$\left\{\begin{array}{l} \mathbf{G}_1\mathbf{G}_1^T\mathbf{\Omega}_1 + \mathbf{\Omega}_1\mathbf{G}_1\mathbf{G}_1^T - \mathbf{G}_1(\mathbf{I}_{r_3}\otimes\mathbf{\Omega}_2)\mathbf{G}_1^T - \mathbf{G}_1(\mathbf{\Omega}_3\otimes\mathbf{I}_{r_2})\mathbf{G}_1^T \\ \qquad\qquad = \mathrm{Skew}(\mathbf{U}_1^T\eta_{\mathbf{U}_1}\mathbf{G}_1\mathbf{G}_1^T) + \mathrm{Skew}(\mathbf{G}_1\eta_{\mathbf{G}_1}^T), \\ \mathbf{G}_2\mathbf{G}_2^T\mathbf{\Omega}_2 + \mathbf{\Omega}_2\mathbf{G}_2\mathbf{G}_2^T - \mathbf{G}_2(\mathbf{I}_{r_3}\otimes\mathbf{\Omega}_1)\mathbf{G}_2^T - \mathbf{G}_2(\mathbf{\Omega}_3\otimes\mathbf{I}_{r_1})\mathbf{G}_2^T \\ \qquad\qquad = \mathrm{Skew}(\mathbf{U}_2^T\eta_{\mathbf{U}_2}\mathbf{G}_2\mathbf{G}_2^T) + \mathrm{Skew}(\mathbf{G}_2\eta_{\mathbf{G}_2}^T), \\ \mathbf{G}_3\mathbf{G}_3^T\mathbf{\Omega}_3 + \mathbf{\Omega}_3\mathbf{G}_3\mathbf{G}_3^T - \mathbf{G}_3(\mathbf{I}_{r_2}\otimes\mathbf{\Omega}_1)\mathbf{G}_3^T - \mathbf{G}_3(\mathbf{\Omega}_2\otimes\mathbf{I}_{r_1})\mathbf{G}_3^T \\ \qquad\qquad = \mathrm{Skew}(\mathbf{U}_3^T\eta_{\mathbf{U}_3}\mathbf{G}_3\mathbf{G}_3^T) + \mathrm{Skew}(\mathbf{G}_3\eta_{\mathbf{G}_3}^T), \end{array}\right. \tag{A.18}$$

that are solved efficiently with the Matlab's `pcg` routine that is combined with a specific preconditioner resulting from the Gauss-Seidel approximation of (A.18).

□

A.8 Proof of Proposition 4 (derivation of the Riemannian gradient formula)

Proof. Let $f(\mathcal{X}) = \|\mathcal{P}_\Omega(\mathcal{X}) - \mathcal{P}_\Omega(\mathcal{X}^\star)\|_F^2/|\Omega|$ and $\mathcal{S} = 2(\mathcal{P}_\Omega(\mathcal{G}\times_1\mathbf{U}_1\times_2\mathbf{U}_2\times_3\mathbf{U}_3) - \mathcal{P}_\Omega(\mathcal{X}^\star))/|\Omega|$ be an auxiliary sparse tensor variable that is interpreted as the Euclidean gradient of f in $\mathbb{R}^{n_1\times n_2\times n_3}$.

The partial derivatives of $f(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3,\mathcal{G})$ are

$$\left\{\begin{array}{lcl}
\dfrac{\partial f_1(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3,\mathbf{G}_1)}{\partial \mathbf{U}_1} &=& \dfrac{2}{|\Omega|}(\mathcal{P}_\Omega(\mathbf{U}_1\mathbf{G}_1(\mathbf{U}_3\otimes\mathbf{U}_2)^T) - \mathcal{P}_\Omega(\mathbf{X}_1^\star))(\mathbf{U}_3\otimes\mathbf{U}_2)\mathbf{G}_1^T\\
&=& \mathbf{S}_1(\mathbf{U}_3\otimes\mathbf{U}_2)\mathbf{G}_1^T,\\
\dfrac{\partial f_2(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3,\mathbf{G}_2)}{\partial \mathbf{U}_2} &=& \dfrac{2}{|\Omega|}(\mathcal{P}_\Omega(\mathbf{U}_2\mathbf{G}_2(\mathbf{U}_3\otimes\mathbf{U}_1)^T) - \mathcal{P}_\Omega(\mathbf{X}_2^\star))(\mathbf{U}_3\otimes\mathbf{U}_1)\mathbf{G}_2^T\\
&=& \mathbf{S}_2(\mathbf{U}_2\otimes\mathbf{U}_1)\mathbf{G}_2^T,\\
\dfrac{\partial f_3(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3,\mathbf{G}_3)}{\partial \mathbf{U}_3} &=& \dfrac{2}{|\Omega|}(\mathcal{P}_\Omega(\mathbf{U}_3\mathbf{G}_3(\mathbf{U}_2\otimes\mathbf{U}_1)^T) - \mathcal{P}_\Omega(\mathbf{X}_3^\star))(\mathbf{U}_2\otimes\mathbf{U}_1)\mathbf{G}_3^T\\
&=& \mathbf{S}_3(\mathbf{U}_2\otimes\mathbf{U}_1)\mathbf{G}_3^T,\\
\dfrac{\partial f(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3,\mathcal{G})}{\partial \mathcal{G}} &=& \dfrac{2}{|\Omega|}(\mathcal{P}_\Omega(\mathcal{G}_{\times 1}\mathbf{U}_1\times_2\mathbf{U}_2\times_3\mathbf{U}_3) - \mathcal{P}_\Omega(\mathcal{X}^\star))\times_1\mathbf{U}_1^T\times_2\mathbf{U}_2^T\times_3\mathbf{U}_3^T\\
&=& \mathcal{S}\times_1\mathbf{U}_1^T\times_2\mathbf{U}_2^T\times_3\mathbf{U}_3^T,
\end{array}\right.$$

where $\mathbf{X}_d^\star$ is mode-d unfolding of $\mathcal{X}^\star$ and

$$\left\{\begin{array}{lcl}
\mathbf{S}_1 &=& \dfrac{2}{|\Omega|}(\mathcal{P}_\Omega(\mathbf{U}_1\mathbf{G}_1(\mathbf{U}_3\otimes\mathbf{U}_2)^T) - \mathcal{P}_\Omega(\mathbf{X}_1^\star))\\
\mathbf{S}_2 &=& \dfrac{2}{|\Omega|}(\mathcal{P}_\Omega(\mathbf{U}_2\mathbf{G}_2(\mathbf{U}_3\otimes\mathbf{U}_1)^T) - \mathcal{P}_\Omega(\mathbf{X}_2^\star))\\
\mathbf{S}_3 &=& \dfrac{2}{|\Omega|}(\mathcal{P}_\Omega(\mathbf{U}_3\mathbf{G}_3(\mathbf{U}_2\otimes\mathbf{U}_1)^T) - \mathcal{P}_\Omega(\mathbf{X}_3^\star))\\
\mathcal{S} &=& \dfrac{2}{|\Omega|}(\mathcal{P}_\Omega(\mathcal{G}\times_1\mathbf{U}_1\times_2\mathbf{U}_2\times_3\mathbf{U}_3) - \mathcal{P}_\Omega(\mathcal{X}^\star)).
\end{array}\right.$$

Due to the specific scaled metric (A.2), the partial derivatives of f are further scaled by $((\mathbf{G}_1\mathbf{G}_1^T)^{-1},(\mathbf{G}_2\mathbf{G}_2^T)^{-1},(\mathbf{G}_3\mathbf{G}_3^T)^{-1},\mathcal{I})$, denoted as $\mathrm{egrad}_x f$ (after scaling), i.e.,

$$\begin{array}{lcl}
\mathrm{egrad}_x f &=& (\mathbf{S}_1(\mathbf{U}_3\otimes\mathbf{U}_2)\mathbf{G}_1^T(\mathbf{G}_1\mathbf{G}_1^T)^{-1}, \mathbf{S}_2(\mathbf{U}_3\otimes\mathbf{U}_1)\mathbf{G}_2^T(\mathbf{G}_2\mathbf{G}_2^T)^{-1}, \mathbf{S}_3(\mathbf{U}_2\otimes\mathbf{U}_1)\mathbf{G}_3^T(\mathbf{G}_3\mathbf{G}_3^T)^{-1},\\
&& \mathcal{S}\times_1\mathbf{U}_1^T\times_2\mathbf{U}_2^T\times_3\mathbf{U}_3^T).
\end{array}$$

Consequently, from the relationship that horizontal lift of $\mathrm{grad}_{[x]}f$ is equal to $\mathrm{grad}_x f = \Psi(\mathrm{egrad}_x f)$, we obtain that, using (A.15),

$$\begin{array}{lcl}
\text{the horizontal lift of } \mathrm{grad}_{[x]}f &=& (\mathbf{S}_1(\mathbf{U}_3\otimes\mathbf{U}_2)\mathbf{G}_1^T(\mathbf{G}_1\mathbf{G}_1^T)^{-1} - \mathbf{U}_1\mathbf{B}_{\mathbf{U}_1}(\mathbf{G}_1\mathbf{G}_1^T)^{-1},\\
&& \mathbf{S}_2(\mathbf{U}_3\otimes\mathbf{U}_1)\mathbf{G}_2^T(\mathbf{G}_2\mathbf{G}_2^T)^{-1} - \mathbf{U}_2\mathbf{B}_{\mathbf{U}_2}(\mathbf{G}_2\mathbf{G}_2^T)^{-1},\\
&& \mathbf{S}_3(\mathbf{U}_2\otimes\mathbf{U}_1)\mathbf{G}_3^T(\mathbf{G}_3\mathbf{G}_3^T)^{-1} - \mathbf{U}_3\mathbf{B}_{\mathbf{U}_3}(\mathbf{G}_3\mathbf{G}_3^T)^{-1},\\
&& \mathcal{S}\times_1\mathbf{U}_1^T\times_2\mathbf{U}_2^T\times_3\mathbf{U}_3^T),
\end{array}$$

From the requirements in (A.16) for a vector to be in the tangent space, we have the following relationship for mode-1.

$$\mathbf{B}_{\mathbf{U}_1}\mathbf{G}_1\mathbf{G}_1^T + \mathbf{G}_1\mathbf{G}_1^T\mathbf{B}_{\mathbf{U}_1} = \mathbf{G}_1\mathbf{G}_1^T(\mathbf{Y}_{\mathbf{U}_1}^T\mathbf{U}_1 + \mathbf{U}_1^T\mathbf{Y}_{\mathbf{U}_1})\mathbf{G}_1\mathbf{G}_1^T,$$

where $\mathbf{Y}_{\mathbf{U}_1} = (\mathbf{S}_1(\mathbf{U}_3\otimes\mathbf{U}_2)\mathbf{G}_1^T(\mathbf{G}_1\mathbf{G}_1^T)^{-1}$.

Subsequently,

$$\begin{array}{lcl}
\mathbf{G}_1\mathbf{G}_1^T(\mathbf{Y}_{\mathbf{U}_1}^T\mathbf{U}_1 + \mathbf{U}_1^T\mathbf{Y}_{\mathbf{U}_1})\mathbf{G}_1\mathbf{G}_1^T &=& \mathbf{G}_1\mathbf{G}_1^T\Big\{((\mathbf{S}_1(\mathbf{U}_3\otimes\mathbf{U}_2)\mathbf{G}_1^T(\mathbf{G}_1\mathbf{G}_1^T)^{-1})^T\mathbf{U}_1\\
&& + \mathbf{U}_1^T(\mathbf{S}_1(\mathbf{U}_3\otimes\mathbf{U}_2)\mathbf{G}_1^T(\mathbf{G}_1\mathbf{G}_1^T)^{-1}\Big\}\mathbf{G}_1\mathbf{G}_1^T\\
&=& ((\mathbf{S}_1(\mathbf{U}_3\otimes\mathbf{U}_2)\mathbf{G}_1^T)^T\mathbf{U}_1\mathbf{G}_1\mathbf{G}_1^T + \mathbf{G}_1\mathbf{G}_1^T\mathbf{U}_1^T(\mathbf{S}_1(\mathbf{U}_3\otimes\mathbf{U}_2)\mathbf{G}_1^T\\
&=& (\mathbf{G}_1\mathbf{G}_1^T\mathbf{U}_1^T(\mathbf{S}_1(\mathbf{U}_3\otimes\mathbf{U}_2)\mathbf{G}_1^T)^T + \mathbf{G}_1\mathbf{G}_1^T\mathbf{U}_1^T(\mathbf{S}_1(\mathbf{U}_3\otimes\mathbf{U}_2)\mathbf{G}_1^T\\
&=& 2\mathrm{Sym}(\mathbf{G}_1\mathbf{G}_1^T\mathbf{U}_1^T(\mathbf{S}_1(\mathbf{U}_3\otimes\mathbf{U}_2)\mathbf{G}_1^T).
\end{array}$$

Finally, $\mathbf{B}_{\mathbf{U}_d}$ for $d \in \{1,2,3\}$ are obtained by solving the Lyapunov equations

$$\left\{ \begin{array}{lcl} \mathbf{B}_{\mathbf{U}_1}\mathbf{G}_1\mathbf{G}_1^T + \mathbf{G}_1\mathbf{G}_1^T\mathbf{B}_{\mathbf{U}_1} & = & 2\mathrm{Sym}(\mathbf{G}_1\mathbf{G}_1^T\mathbf{U}_1^T(\mathbf{S}_1(\mathbf{U}_3 \otimes \mathbf{U}_2)\mathbf{G}_2^T), \\ \mathbf{B}_{\mathbf{U}_2}\mathbf{G}_2\mathbf{G}_2^T + \mathbf{G}_2\mathbf{G}_2^T\mathbf{B}_{\mathbf{U}_2} & = & 2\mathrm{Sym}(\mathbf{G}_2\mathbf{G}_2^T\mathbf{U}_2^T(\mathbf{S}_2(\mathbf{U}_3 \otimes \mathbf{U}_1)\mathbf{G}_2^T), \\ \mathbf{B}_{\mathbf{U}_3}\mathbf{G}_3\mathbf{G}_3^T + \mathbf{G}_3\mathbf{G}_3^T\mathbf{B}_{\mathbf{U}_3} & = & 2\mathrm{Sym}(\mathbf{G}_3\mathbf{G}_3^T\mathbf{U}_3^T(\mathbf{S}_3(\mathbf{U}_2 \otimes \mathbf{U}_1)\mathbf{G}_3^T), \end{array} \right.$$

where $\mathrm{Sym}(\cdot)$ extracts the symmetric part of a square matrix, i.e., $\mathrm{Sym}(\mathbf{D}) = (\mathbf{D}+\mathbf{D}^T)/2$. The above Lyapunov equations are solved efficiently with the Matlab's `lyap` routine. □

B Additional numerical comparisons

In addition to the representative numerical comparisons in the paper, we show additional numerical experiments spanning synthetic and real-world datasets.

Experiments on synthetic datasets:

Case S1: comparison with the Euclidean metric. We first show the benefit of the proposed metric (A.2) over the conventional choice of the Euclidean metric that exploits the product structure of $\mathcal{M}$ and symmetry. We compare *steepest descent* algorithms with Armijo backtracking linesearch for both the metric choices. Figure A.1 shows that the algorithm with the metric (A.2) gives a superior performance in *test error* than that of the conventional metric choice.

Case S2: small-scale instances. We consider tensors of size $100 \times 100 \times 100$, $150 \times 150 \times 150$, and $200 \times 200 \times 200$ and ranks $(5,5,5)$, $(10,10,10)$, and $(15,15,15)$. OS is $\{10,20,30\}$. Figures A.2(a)-(c) and Figures A.3(a)-(c) show the convergence behavior of different algorithms on a train set Ω and on a test set Γ, where Figures A.3(b) is identical to the figure in the manuscript paper. Figures A.2(d)-(f) and A.3(d)-(f) show the mean square error on Ω and Γ on each algorithm. Furthermore, Figure A.2(g)-(i) and Figure A.3(g)-(i) show the mean square error on Ω and Γ when OS is 10 in all the five runs. From Figures A.2 and Figures A.3, our proposed algorithm is consistently competitive or faster than geomCG, HalRTC, and TOpt. In addition, the mean square errors on a train set Ω and a test set Γ are consistently competitive or lower than those of geomCG and HalRTC, especially for lower sampling ratios, e.g, for OS 10.

Case S3: large-scale instances. We consider large-scale tensors of size $3000 \times 3000 \times 3000$, $5000 \times 5000 \times 5000$, and $10000 \times 10000 \times 10000$ and ranks **r**=$(5,5,5)$ and $(10,10,10)$. OS is 10. We compare our proposed algorithm to geomCG. Figure A.4 and Figure A.5 show the convergence behavior of the algorithms. The proposed algorithm outperforms geomCG in all the cases.

Case S4: influence of low sampling. We look into problem instances which result from scarcely sampled data. The test requires completing a tensor of size $10000 \times 10000 \times 10000$ and rank **r**=$(5,5,5)$. Figure A.6 and Figure A.7 show the convergence behavior when OS is $\{8,6,5\}$. The case of OS $= 5$ is particularly interesting. In this case, while the mean square errors on Ω and Γ increase for geomCG, the proposed algorithm stably decreases the error in all the five runs.

Case S5: influence of ill-conditioning and low sampling. We consider the problem instance of **Case S4** with OS $= 5$. Additionally, for generating the instance, we impose a diagonal core $\boldsymbol{\mathcal{G}}$ with exponentially decaying *positive* values of condition numbers (CN) 5, 50, and 100. Figure A.8 shows that the proposed algorithm outperforms geomCG for all the considered CN values on a train set Ω.

Case S6: influence of noise. We evaluate the convergence properties of algorithms under the presence of noise The tensor size and rank are same as in **Case S4** and OS is 10. Figure A.9 shows that the train error on a train set Ω for each ϵ is almost identical to the $\epsilon^2\|\mathcal{P}_\Omega(\boldsymbol{\mathcal{X}}^\star)\|_F^2$, but our proposed algorithm converges faster than geomCG.

Case S7: rectangular instances. We consider instances where dimensions and ranks along certain modes are different than others. Two cases are considered. Case (7.a) considers tensors size $20000 \times 7000 \times 7000$, $30000 \times 6000 \times 6000$, and $40000 \times 5000 \times 5000$ and rank $\mathbf{r} = (5,5,5)$. Case (7.b) considers a tensor of size $10000 \times 10000 \times 10000$ with ranks $\mathbf{r} = (7,6,6)$, $(10,5,5)$, and $(15,4,4)$. Figures A.10(a)-(c) and Figures A.11(a)-(c) show that the convergence behavior of our proposed algorithm is superior to that of geomCG on Ω and Γ, respectively. Our proposed algorithm also outperforms geomCG for the asymmetric rank cases as shown in Figure A.10(d)-(f) and Figure A.11(d)-(f).

Case S8: medium-scale instances. We additionally consider medium-scale tensors of size $500 \times 500 \times 500$, $1000 \times 1000 \times 1000$, and $1500 \times 1500 \times 1500$ and ranks $\mathbf{r} = (5,5,5)$, $(10,10,10)$, and $(15,15,15)$. OS is $\{10,20,30,40\}$. Our proposed algorithm and geomCG are only compared as the other algorithms cannot handle these scales efficiently. Figures

A.12(a)-(c) and A.13(a)-(c) show the convergence behavior on Ω and Γ, respectively. Figures A.12(d)-(f) and Figures A.13(d)-(f) also show the mean square error on Ω and Γ of rank $\mathbf{r} = (15, 15, 15)$ in all the five runs. The proposed algorithm performs better than geomCG in all the cases.

Experiments on real-world datasets:

Case R1: hyperspectral image. We also show the performance of our algorithm on the hyperspectral image "Ribeira". We show the mean square error on Ω and Γ when OS is $\{11, 22\}$ in Figure A.14 and Figure A.15, where Figure A.15(a) is identical to the figure in the manuscript paper. Our proposed algorithm gives lower test errors than those obtained by the other algorithms. We also show the image recovery results. Figures A.16 and A.17 show the reconstructed images when OS is $\{11, 22\}$, respectively. From these figures, we find that the proposed algorithm shows a good performance, especially for the lower sampling ratio.

Case R2: MovieLens-10M. Figure A.18 and Figure A.19 show the convergence plots for all the five runs of ranks $\mathbf{r} = (4, 4, 4)$, $(6, 6, 6)$, $(8, 8, 8)$ and $(10, 10, 10)$ on Ω and Γ, respectively. These figures show the superior performance of our proposed algorithm.

Experiments for online algorithms:

Case O: online instances. Figure A.20 and A.21 show the convergence plots for all the five runs on tensors of ranks $100 \times 100 \times 5000$, and $100 \times 100 \times 10000$ with rank $\mathbf{r} = (5, 5, 5)$ on Ω and Γ, respectively. These figures show that the proposed stochastic gradient descent algorithm gives similar or faster convergence than the proposed batch gradient descent algorithm.

Figure A.22 and A.23 show the convergence speed comparisons in the train error and the test error of the proposed online and batch algorithms with TeCPSGD and OLSTEC with rank $\mathbf{r} = (5, 5, 5)$ on the real-world video sequence Airport Hall dataset. These figures show that the proposed stochastic gradient descent algorithm gives similar or faster convergence than the proposed batch algorithm. In addition, Table B shows that the final train and test MSEs show the superior performance of the proposed algorithms.

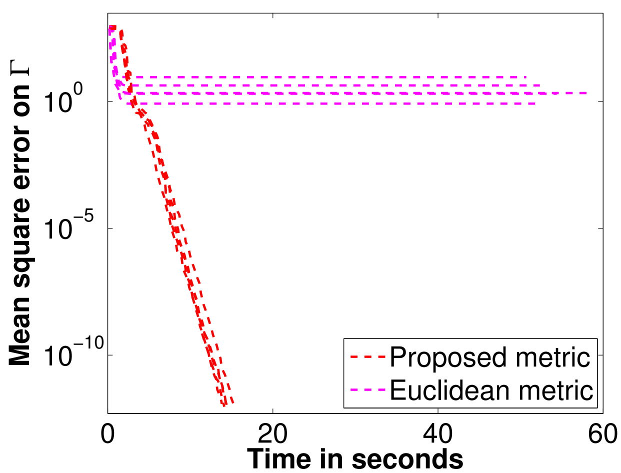
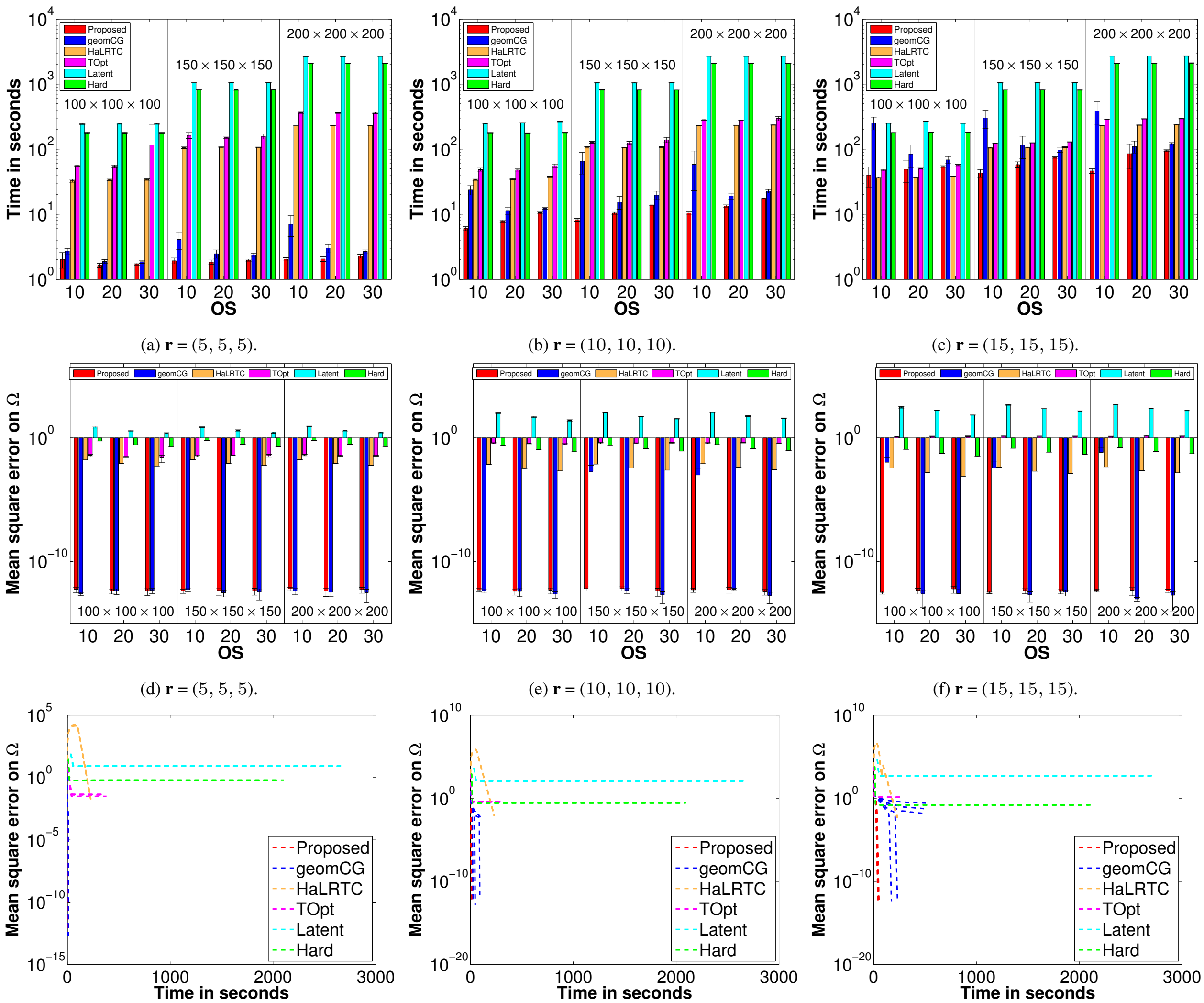


Figure A.1. **Case S1:** comparison between metrics (test error)

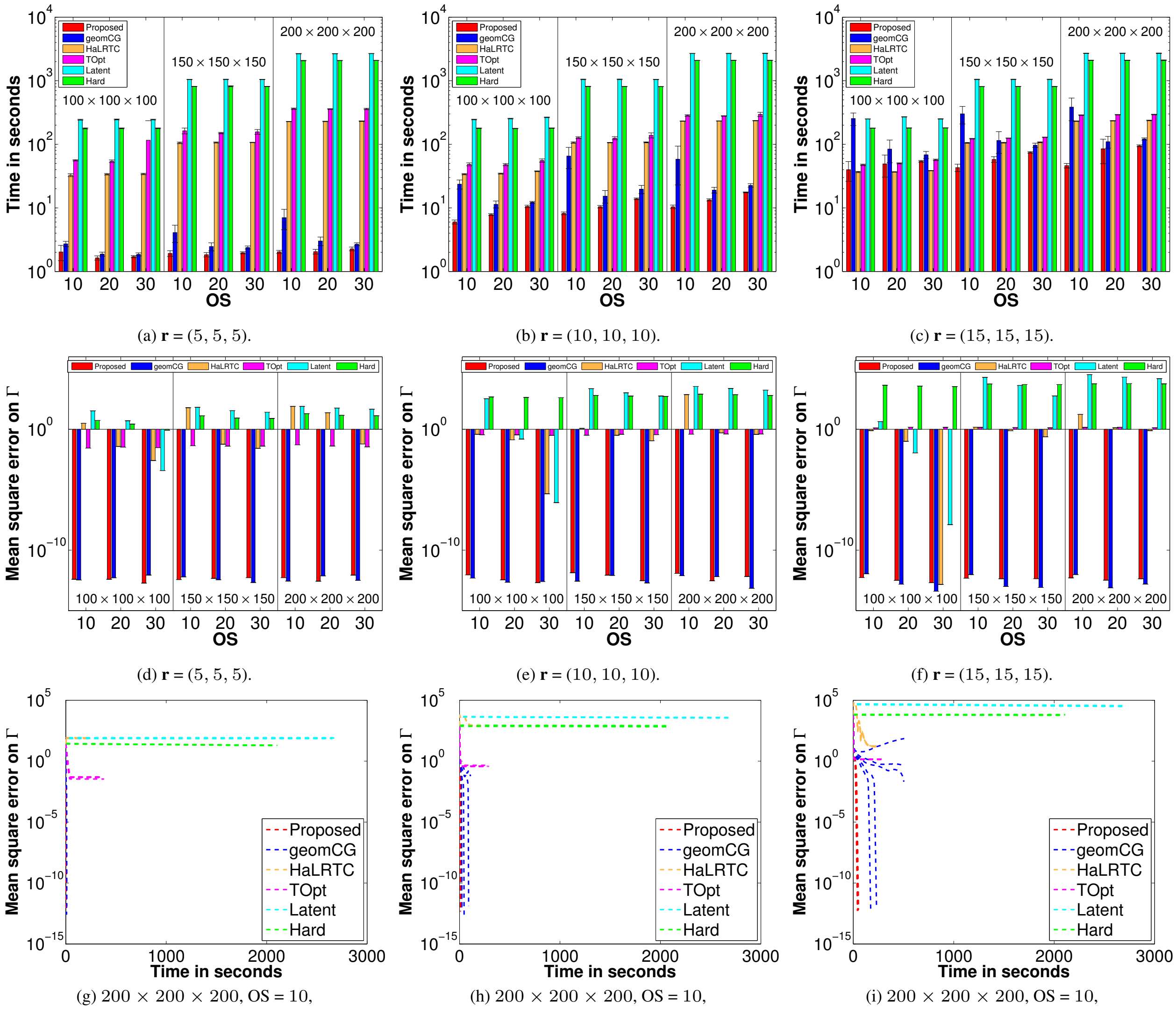


(a) $\mathbf{r} = (5, 5, 5)$. (b) $\mathbf{r} = (10, 10, 10)$. (c) $\mathbf{r} = (15, 15, 15)$.

(d) $\mathbf{r} = (5, 5, 5)$. (e) $\mathbf{r} = (10, 10, 10)$. (f) $\mathbf{r} = (15, 15, 15)$.

(g) $200 \times 200 \times 200$, OS = 10, $\mathbf{r} = (5, 5, 5)$. (h) $200 \times 200 \times 200$, OS = 10, $\mathbf{r} = (10, 10, 10)$. (i) $200 \times 200 \times 200$, OS = 10, $\mathbf{r} = (15, 15, 15)$.

Figure A.2. **Case S2:** small-scale comparisons on Ω (train error).



(a) **r** = (5, 5, 5).

(b) **r** = (10, 10, 10).

(c) **r** = (15, 15, 15).

(d) **r** = (5, 5, 5).

(e) **r** = (10, 10, 10).

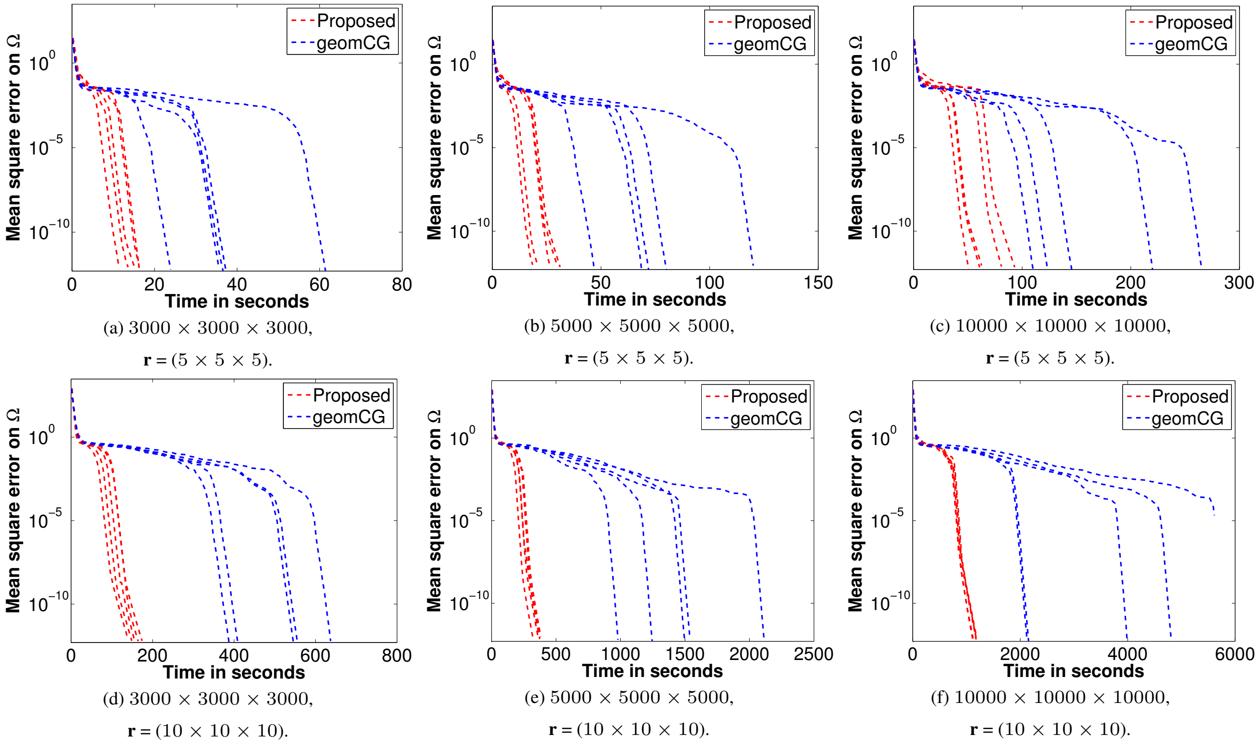
(f) **r** = (15, 15, 15).

(g) 200 × 200 × 200, OS = 10, **r** = (5, 5, 5).

(h) 200 × 200 × 200, OS = 10, **r** = (10, 10, 10).

(i) 200 × 200 × 200, OS = 10, **r** = (15, 15, 15).

Figure A.3. **Case S2:** small-scale comparisons on Γ (test error).



(a) $3000 \times 3000 \times 3000$, $\mathbf{r} = (5 \times 5 \times 5)$.

(b) $5000 \times 5000 \times 5000$, $\mathbf{r} = (5 \times 5 \times 5)$.

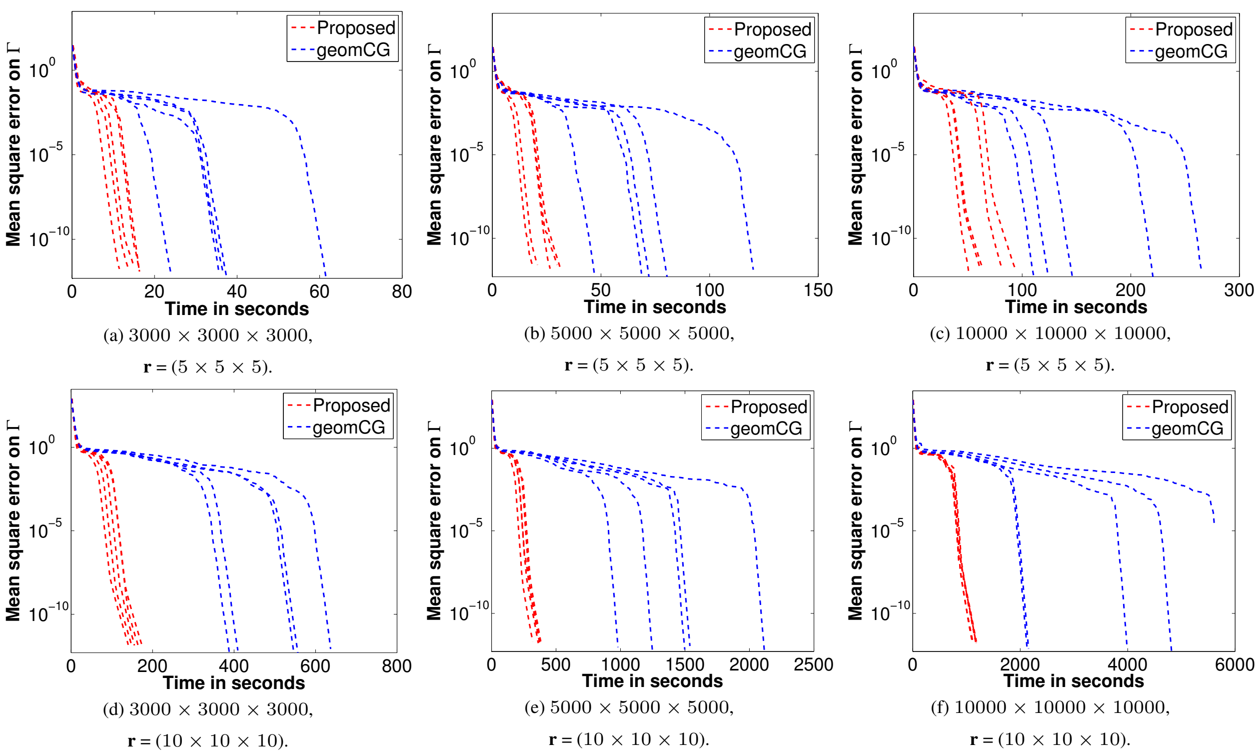
(c) $10000 \times 10000 \times 10000$, $\mathbf{r} = (5 \times 5 \times 5)$.

(d) $3000 \times 3000 \times 3000$, $\mathbf{r} = (10 \times 10 \times 10)$.

(e) $5000 \times 5000 \times 5000$, $\mathbf{r} = (10 \times 10 \times 10)$.

(f) $10000 \times 10000 \times 10000$, $\mathbf{r} = (10 \times 10 \times 10)$.

Figure A.4. **Case S3:** large-scale comparisons on Ω (train error).



(a) $3000 \times 3000 \times 3000$, $\mathbf{r} = (5 \times 5 \times 5)$.

(b) $5000 \times 5000 \times 5000$, $\mathbf{r} = (5 \times 5 \times 5)$.

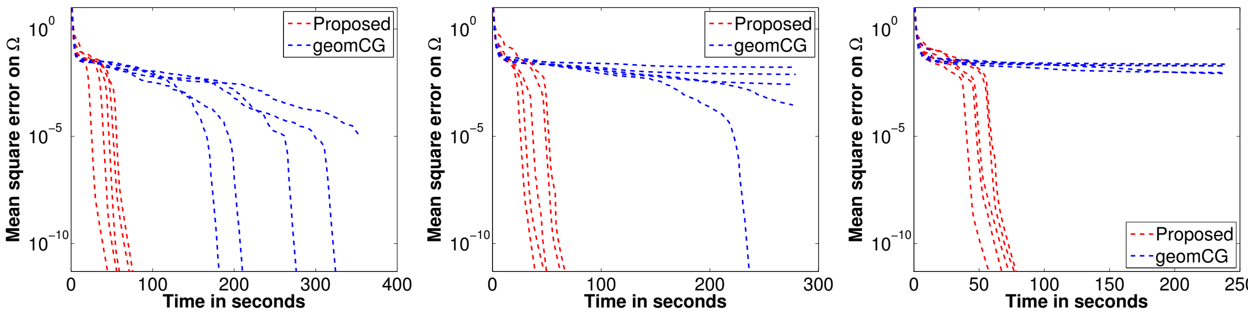
(c) $10000 \times 10000 \times 10000$, $\mathbf{r} = (5 \times 5 \times 5)$.

(d) $3000 \times 3000 \times 3000$, $\mathbf{r} = (10 \times 10 \times 10)$.

(e) $5000 \times 5000 \times 5000$, $\mathbf{r} = (10 \times 10 \times 10)$.

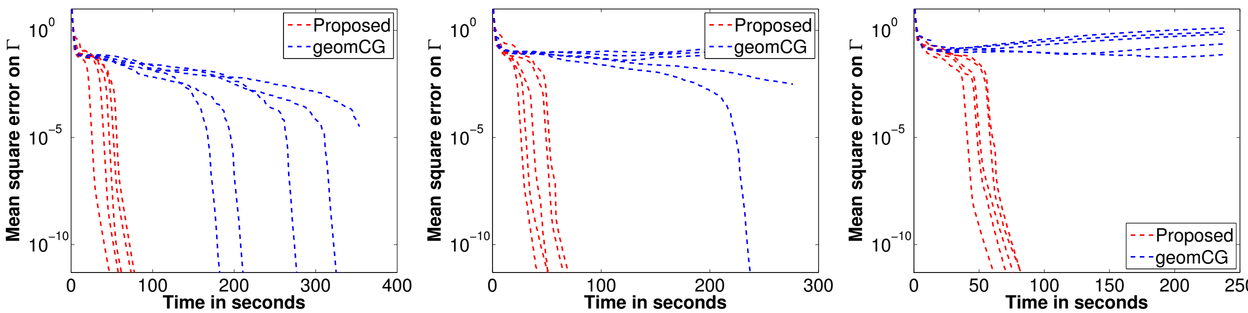
(f) $10000 \times 10000 \times 10000$, $\mathbf{r} = (10 \times 10 \times 10)$.

Figure A.5. **Case S3:** large-scale comparisons on Γ (test error).



(a) OS = 8. (b) OS = 6. (c) OS = 5.

Figure A.6. **Case S4:** low-sampling comparisons on Ω (train error).



(a) OS = 8. (b) OS = 6. (c) OS = 5.

Figure A.7. **Case S4:** low-sampling comparisons on Γ (test error).

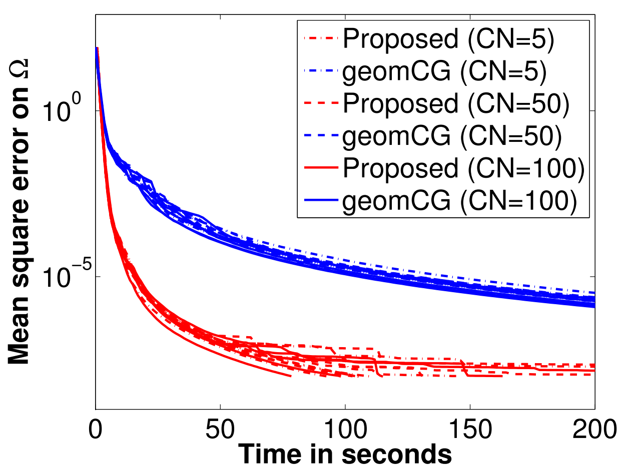


Figure A.8. **Case S5:** CN = $\{5, 50, 100\}$ on Ω (train error).

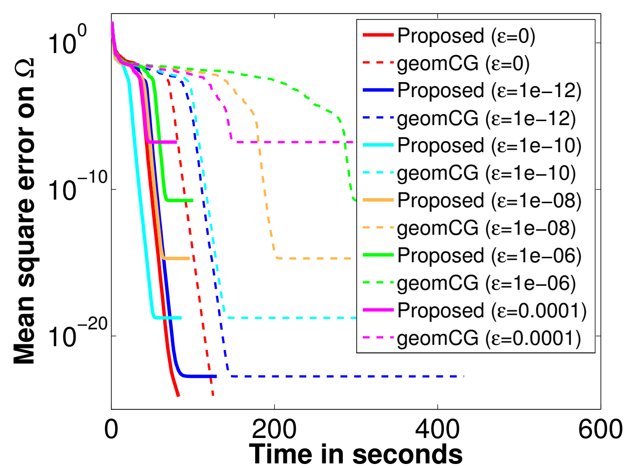
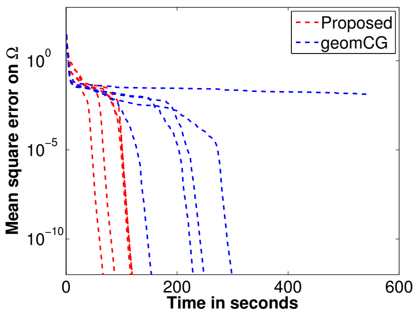
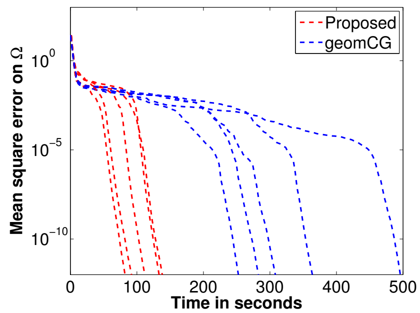


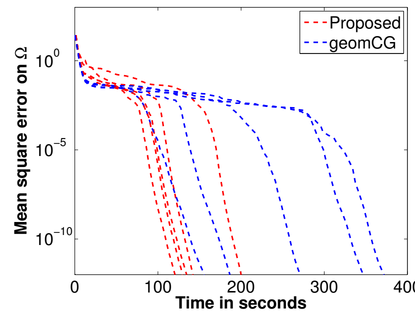
Figure A.9. **Case S6:** noisy data on Ω (train error).



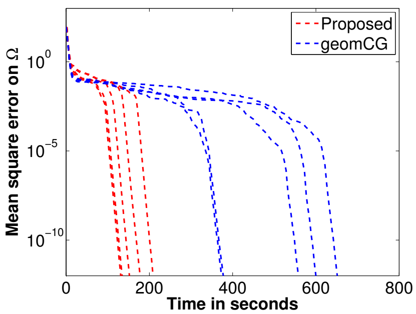
(a) $20000 \times 7000 \times 7000$, $\mathbf{r} = (5 \times 5 \times 5)$.



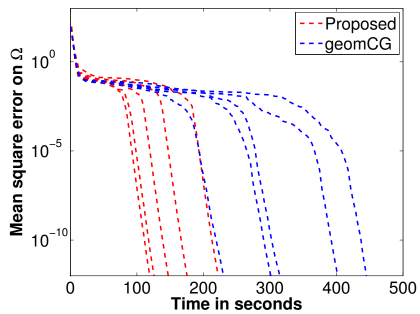
(b) $30000 \times 60000 \times 60000$, $\mathbf{r} = (5 \times 5 \times 5)$.



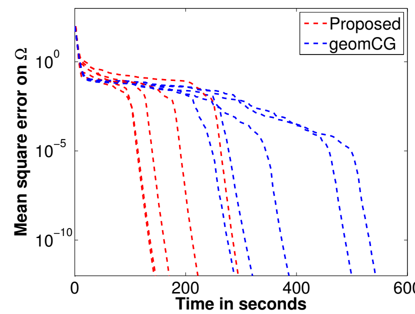
(c) $40000 \times 5000 \times 5000$, $\mathbf{r} = (5 \times 5 \times 5)$.



(d) $\mathbf{r} = (7 \times 6 \times 6)$, $10000 \times 10000 \times 10000$.

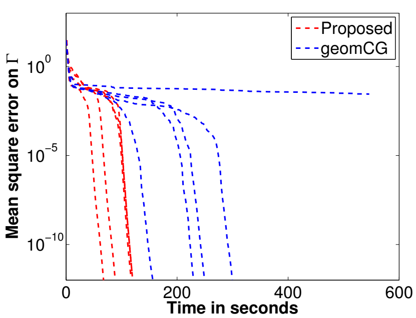


(e) $\mathbf{r} = (10 \times 5 \times 5)$, $10000 \times 10000 \times 10000$.

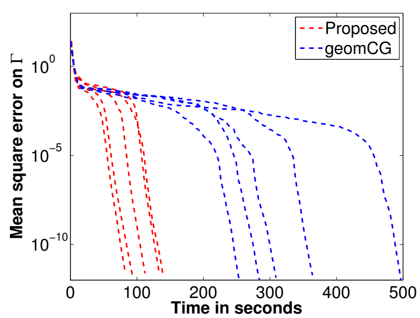


(f) $\mathbf{r} = (15 \times 4 \times 4)$, $10000 \times 10000 \times 10000$.

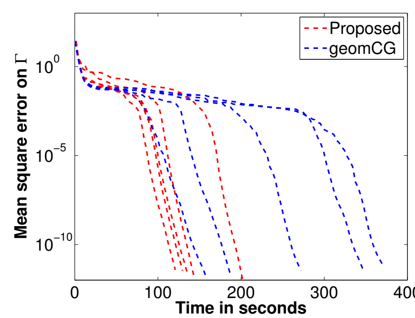
Figure A.10. **Case S7:** rectangular comparisons on Ω (train error).



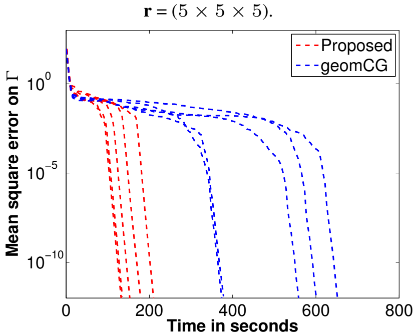
(a) $20000 \times 7000 \times 7000$, $\mathbf{r} = (5 \times 5 \times 5)$.



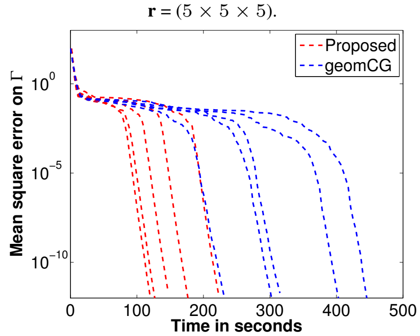
(b) $30000 \times 60000 \times 60000$, $\mathbf{r} = (5 \times 5 \times 5)$.



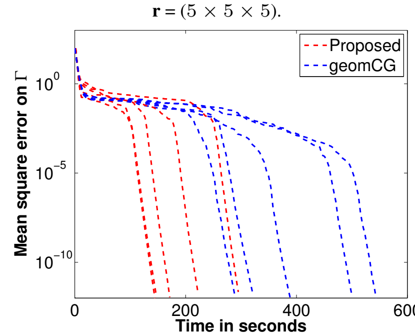
(c) $40000 \times 5000 \times 5000$, $\mathbf{r} = (5 \times 5 \times 5)$.



(d) $\mathbf{r} = (7 \times 6 \times 6)$, $10000 \times 10000 \times 10000$.

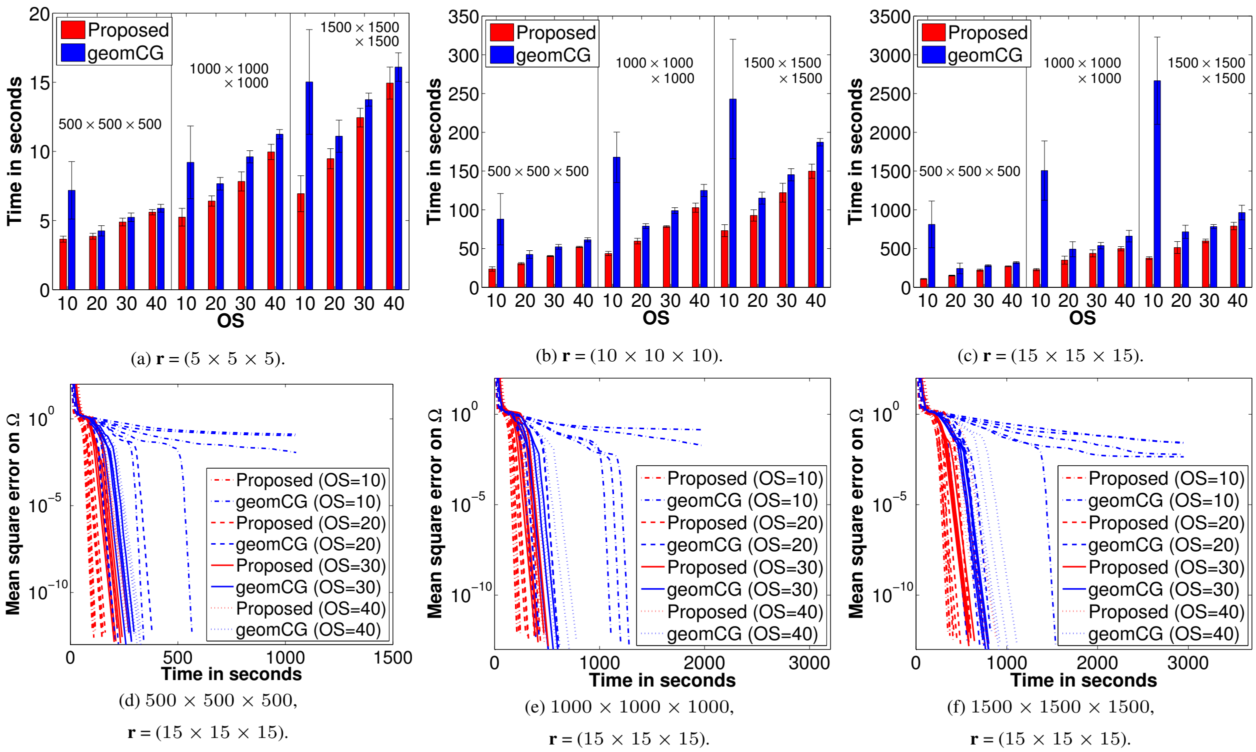


(e) $\mathbf{r} = (10 \times 5 \times 5)$, $10000 \times 10000 \times 10000$.



(f) $\mathbf{r} = (15 \times 4 \times 4)$, $10000 \times 10000 \times 10000$.

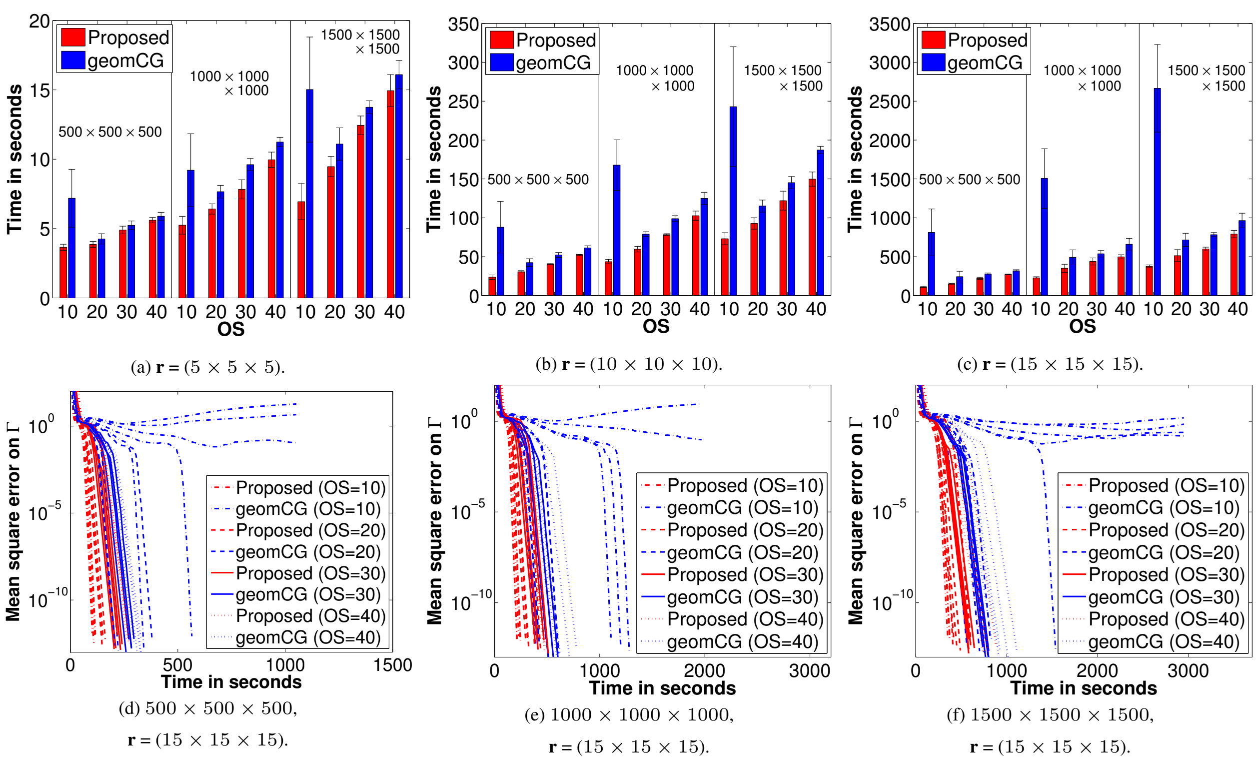
Figure A.11. **Case S7:** rectangular comparisons on Γ (test error).



(a) $\mathbf{r} = (5 \times 5 \times 5)$. (b) $\mathbf{r} = (10 \times 10 \times 10)$. (c) $\mathbf{r} = (15 \times 15 \times 15)$.

(d) $500 \times 500 \times 500$, $\mathbf{r} = (15 \times 15 \times 15)$. (e) $1000 \times 1000 \times 1000$, $\mathbf{r} = (15 \times 15 \times 15)$. (f) $1500 \times 1500 \times 1500$, $\mathbf{r} = (15 \times 15 \times 15)$.

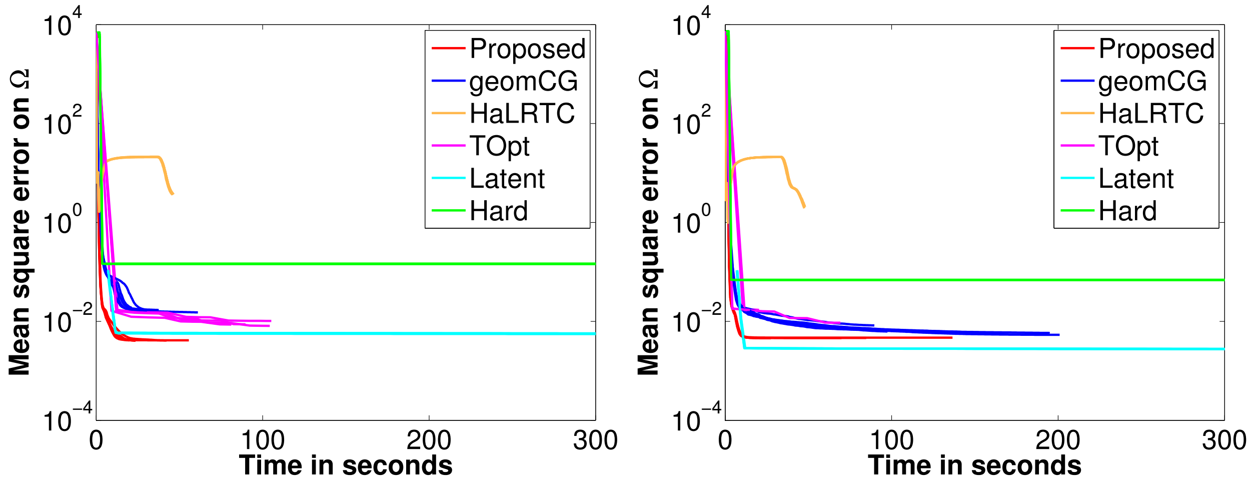
Figure A.12. **Case S8:** medium-scale comparisons on Ω (train error).



(a) $\mathbf{r} = (5 \times 5 \times 5)$. (b) $\mathbf{r} = (10 \times 10 \times 10)$. (c) $\mathbf{r} = (15 \times 15 \times 15)$.

(d) $500 \times 500 \times 500$, $\mathbf{r} = (15 \times 15 \times 15)$. (e) $1000 \times 1000 \times 1000$, $\mathbf{r} = (15 \times 15 \times 15)$. (f) $1500 \times 1500 \times 1500$, $\mathbf{r} = (15 \times 15 \times 15)$.

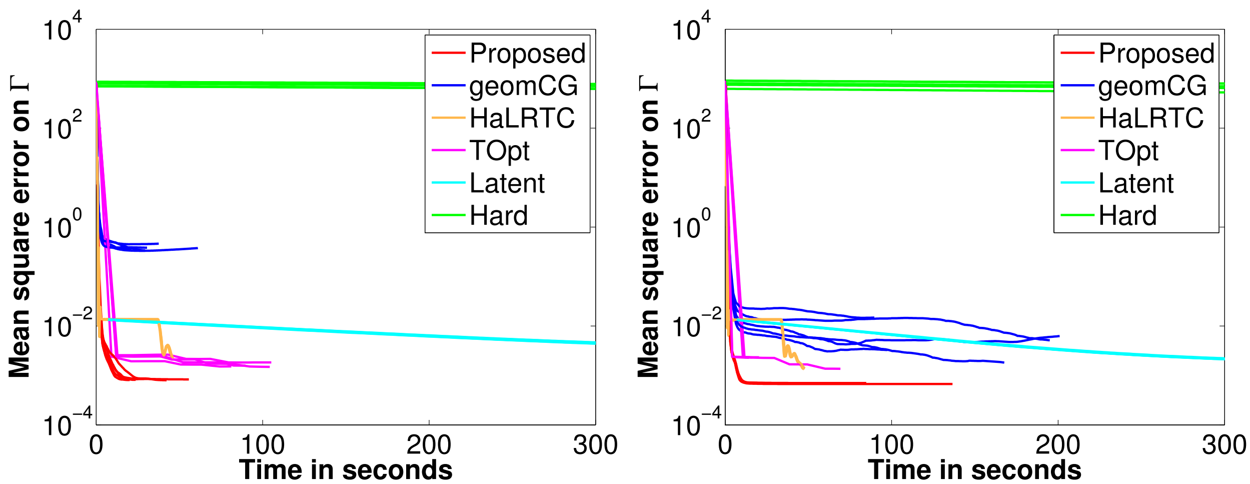
Figure A.13. **Case S8:** medium-scale comparisons on Γ (test error).



(a) OS = 11.

(b) OS = 22.

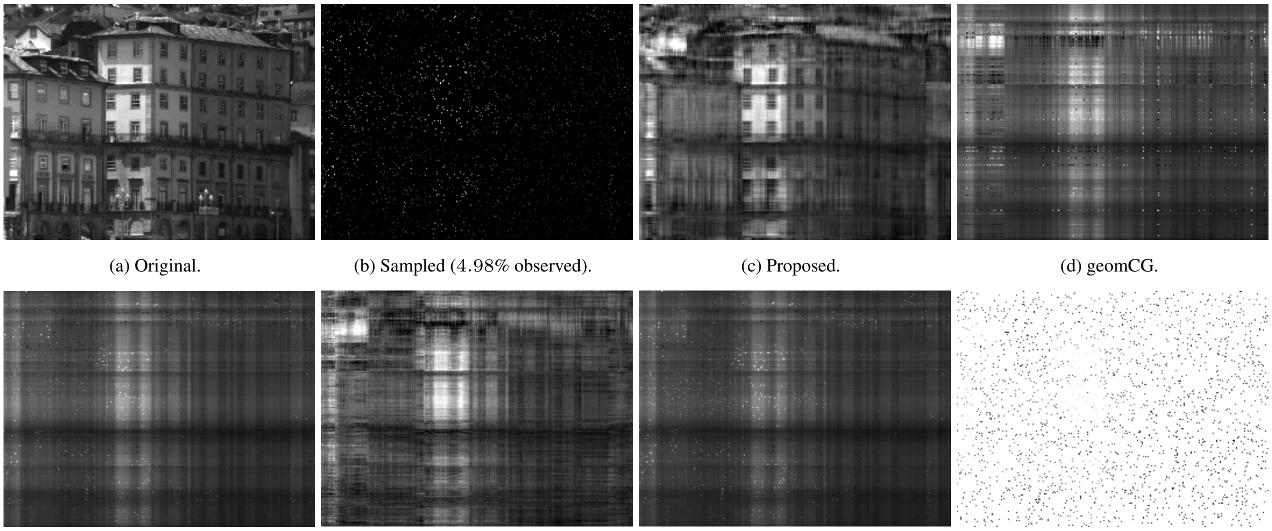
Figure A.14. **Case R1:** mean square error on Ω (train error).



(a) OS = 11.

(b) OS = 22.

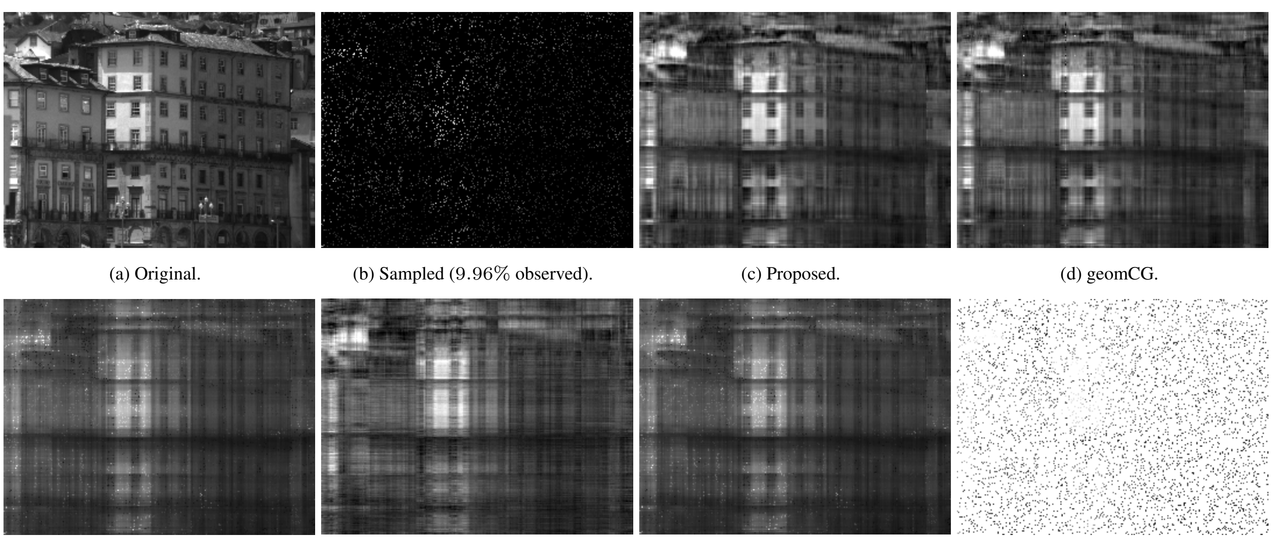
Figure A.15. **Case R1:** mean square error on Γ (test error).



(a) Original. (b) Sampled (4.98% observed). (c) Proposed. (d) geomCG.

(e) HaLRTC. (f) TOpt. (g) Latent. (h) Hard.

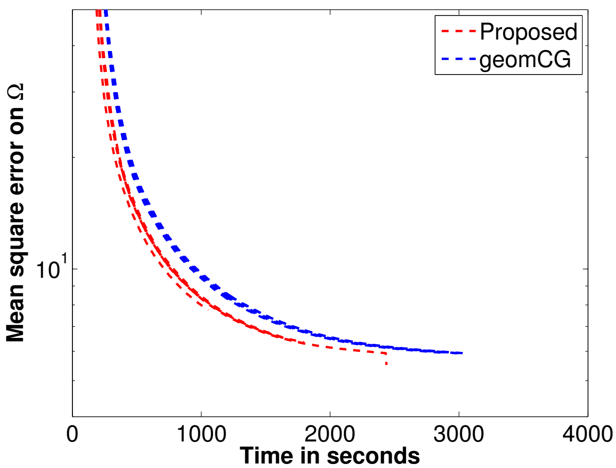
Figure A.16. **Case R1:** recovery results on the hyperspectral image “Ribeira” (frame = 16, OS = 11).



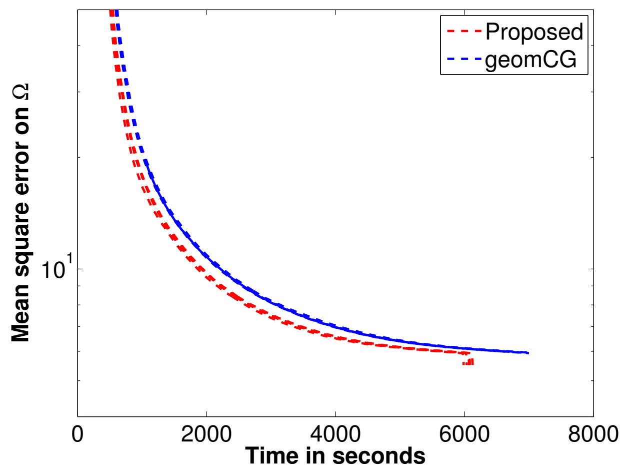
(a) Original. (b) Sampled (9.96% observed). (c) Proposed. (d) geomCG.

(e) HaLRTC. (f) TOpt. (g) Latent. (h) Hard.

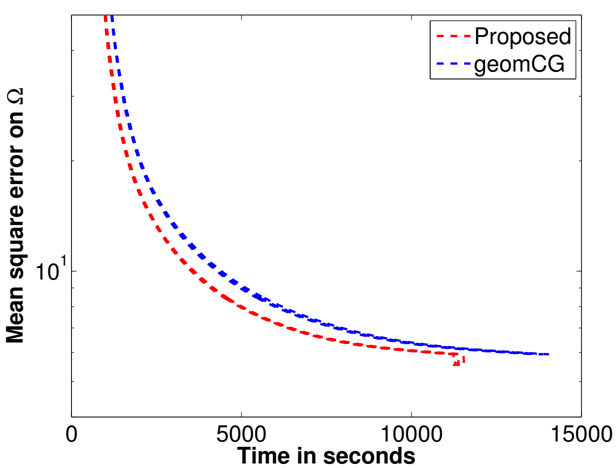
Figure A.17. **Case R1:** recovery results on the hyperspectral image “Ribeira” (frame = 16, OS = 22).



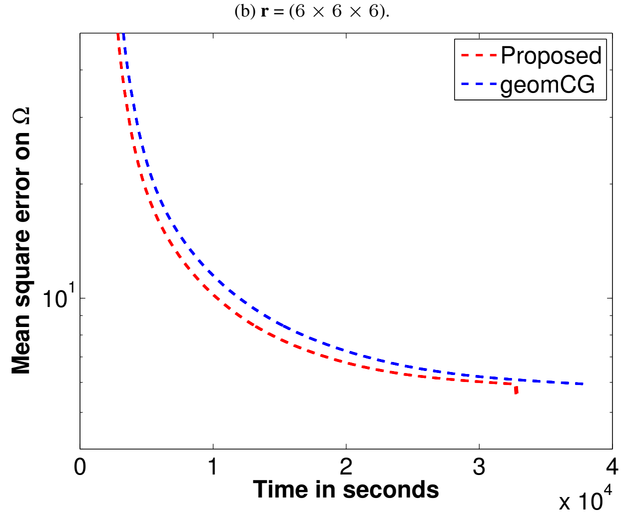
(a) $\mathbf{r} = (4 \times 4 \times 4)$.



(b) $\mathbf{r} = (6 \times 6 \times 6)$.

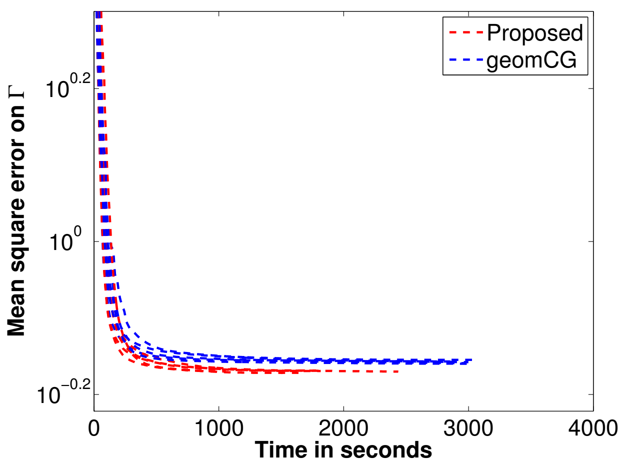


(c) $\mathbf{r} = (8 \times 8 \times 8)$.

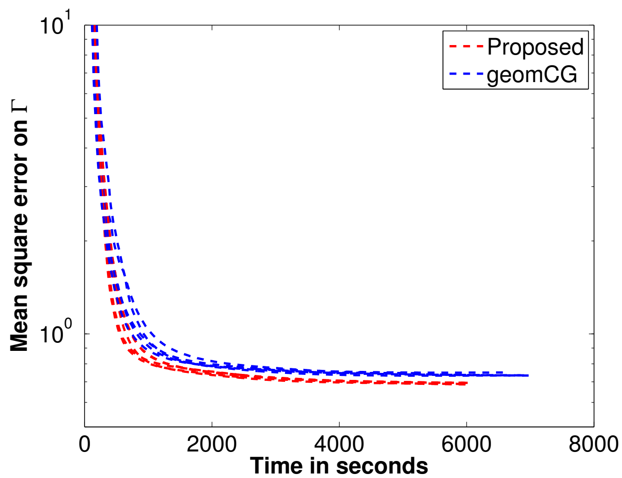


(d) $\mathbf{r} = (10 \times 10 \times 10)$.

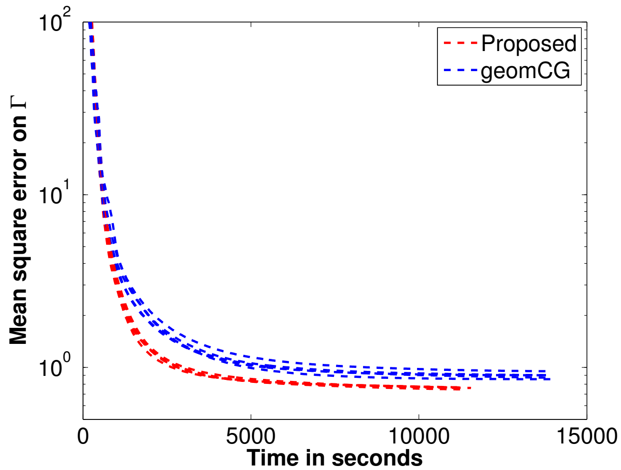
Figure A.18. **Case R2:** mean square error on Ω (train error).



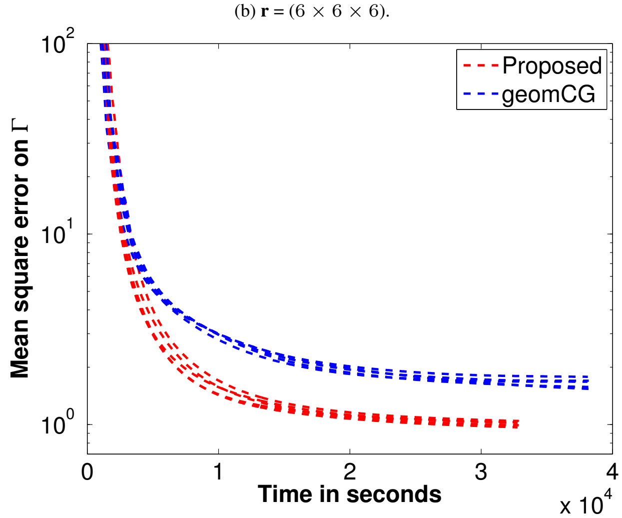
(a) $\mathbf{r} = (4 \times 4 \times 4)$.



(b) $\mathbf{r} = (6 \times 6 \times 6)$.

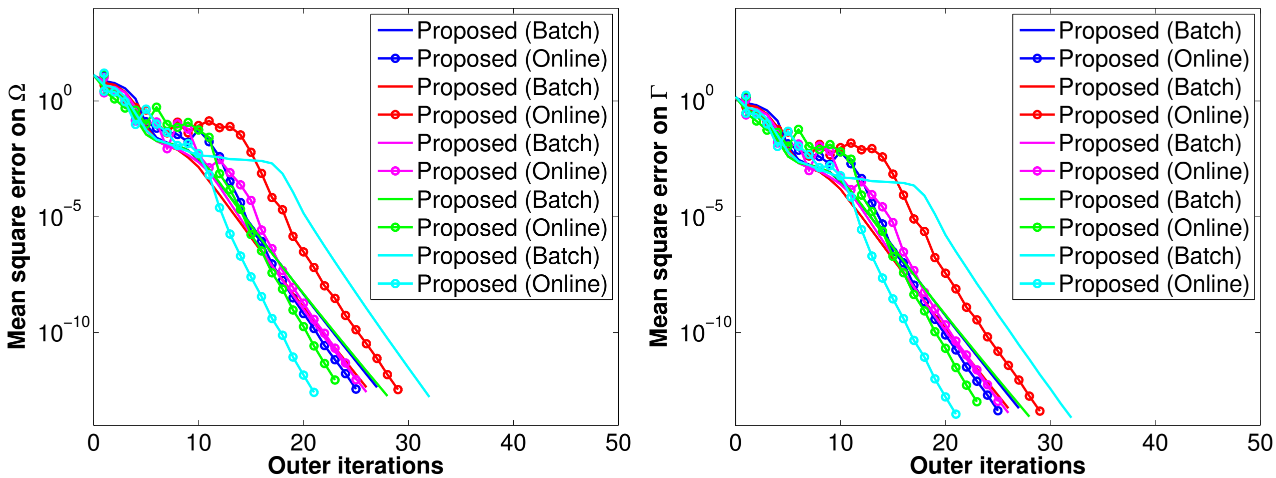


(c) $\mathbf{r} = (8 \times 8 \times 8)$.



(d) $\mathbf{r} = (10 \times 10 \times 10)$.

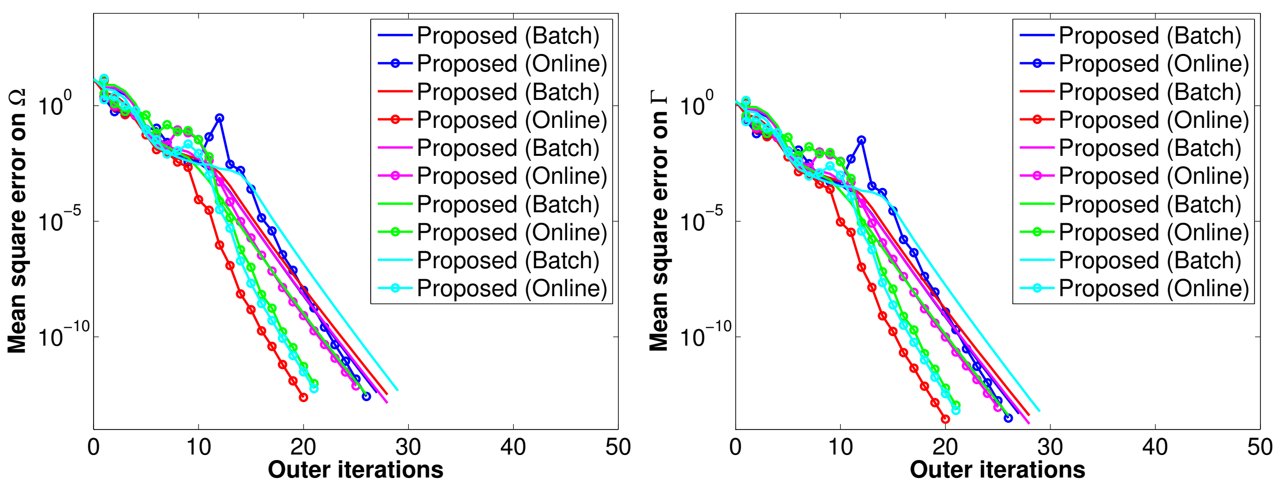
Figure A.19. **Case R2:** mean square error on Γ (test error).



(a) Mean square error on Ω (train error).

(b) Mean square error on Γ (test error).

Figure A.20. **Case O**: mean square error on synthetic instance of size $100 \times 100 \times 5000$.



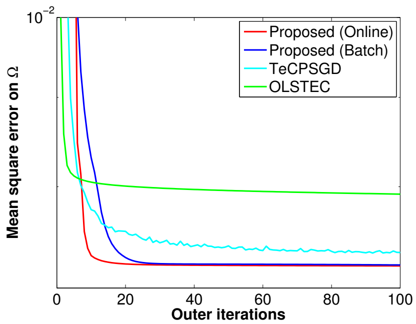
(a) Mean square error on Ω (train error).

(b) Mean square error on Γ (test error).

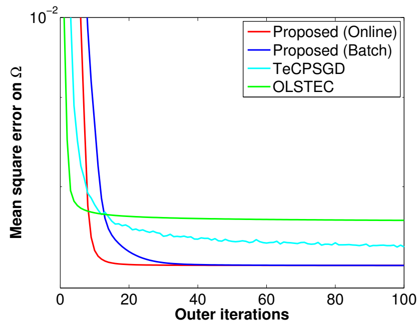
Figure A.21. **Case O**: mean square error on synthetic instance of size $100 \times 100 \times 10000$.

Table A.1. **Case O**: mean square error (5 runs) on Airport Hall dataset.

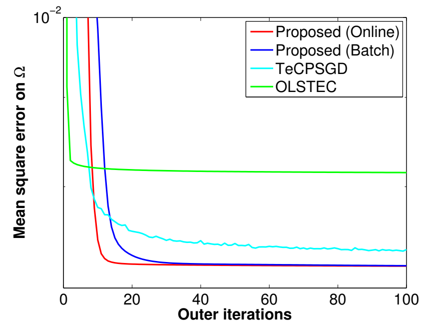
Error type	Algorithm	run 1	run 2	run 3	run 4	run 5
Training error on Ω	Proposed (Online)	7.210000	7.211718	7.205027	7.255203	7.230000
	Proposed (Batch)	7.215763	7.211496	7.208463	7.282901	7.218042
	TeCPSGD	7.335320	7.389269	7.364065	7.393318	7.390530
	OLSTEC	7.922385	7.653096	8.150799	8.248936	7.753596
Test error on Γ	Proposed (Online)	7.462097	7.440332	7.452799	7.443505	7.450065
	Proposed (Batch)	7.471942	7.440508	7.446072	7.492786	7.218042
	TeCPSGD	7.592109	7.601955	7.600740	7.579759	7.600621
	OLSTEC	8.205765	7.840107	8.599819	8.625715	7.965405



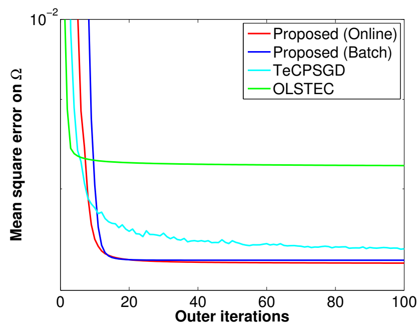
(a) Run 1.



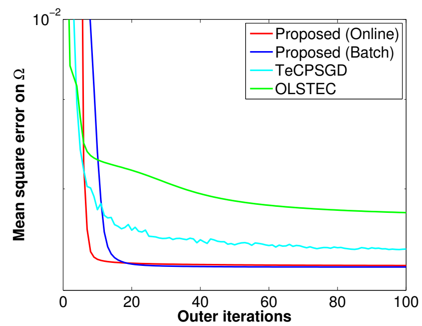
(b) Run 2.



(c) Run 3.

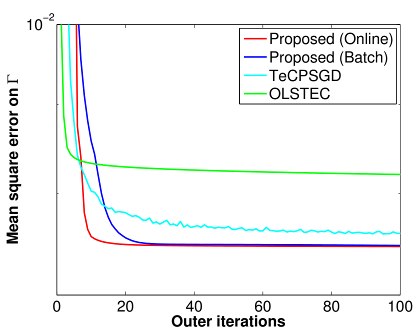


(d) Run 4.

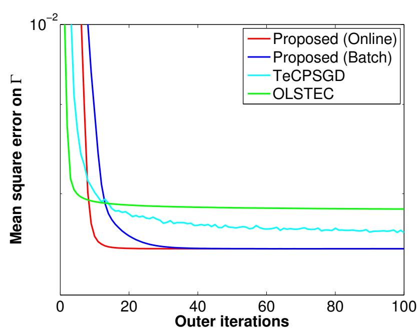


(e) Run 5.

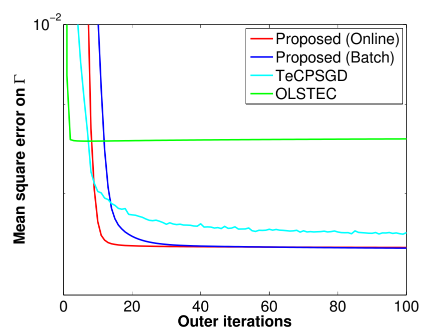
Figure A.22. **Case O**: mean square error on the training set Ω of the Airport Hall dataset.



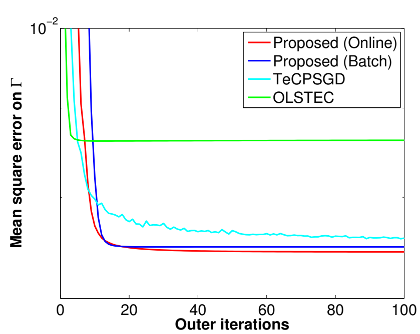
(a) Run 1.



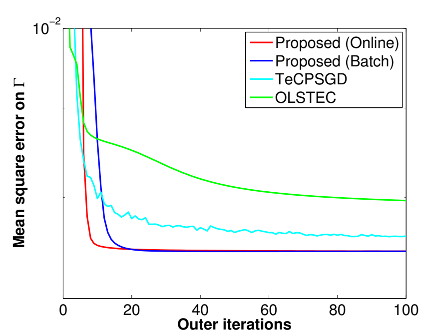
(b) Run 2.



(c) Run 3.



(d) Run 4.



(e) Run 5.

Figure A.23. **Case O**: mean square error on Γ (test error) for the Aiport Hall dataset.

References

Absil, P.-A., Mahony, R., and Sepulchre, R. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008.