Matrix Eigen-decomposition via Doubly Stochastic Riemannian Optimization: Supplementary Material

Zhiqiang Xu XUZQ@I2R.A-STAR.EDU.SG
Peilin Zhao ZHAOP@I2R.A-STAR.EDU.SG
Jianneng Cao CAOJN@I2R.A-STAR.EDU.SG
Xiaoli Li XLLI@I2R.A-STAR.EDU.SG
Institute for Infocomm Research, A*STAR, Singapore

Preparation

First, based on the definitions of $\mathbf{A}_t$, $\mathbf{Y}_t$, $\tilde{\mathbf{Z}}_t$ and $\mathbf{Z}_t$, we can write

$$g_t = G(s_t, r_t, \mathbf{X}_t) = p_{s_t}^{-1} p_{r_t}^{-1} (\mathbf{I} - \mathbf{X}_t \mathbf{X}_t^\top)(\mathbf{E}_{s_t} \odot \mathbf{A})(\mathbf{E}_{\cdot r_t} \odot \mathbf{X}) = (\mathbf{I} - \mathbf{X}_t \mathbf{X}_t^\top)\mathbf{A}_t \mathbf{Y}_t.$$

Then from (6), we have

$$\mathbf{X}_{t+1} = \mathbf{X}_t + \alpha_t g_t \mathbf{W}_t - \frac{\alpha_t^2}{2} \mathbf{X}_t g_t^\top g_t \mathbf{W}_t.$$

Since

$$\mathbf{W}_t = (\mathbf{I} + \frac{\alpha_t^2}{4} g_t^\top g_t)^{-1} = \mathbf{I} - \frac{\alpha_t^2}{4} g_t^\top g_t + O(\alpha_t^4),$$

we get

$$\mathbf{X}_{t+1} = \mathbf{X}_t + \alpha_t \mathbf{A}_t \mathbf{Y}_t - \alpha_t \mathbf{X}_t \mathbf{X}_t^\top \mathbf{A}_t \mathbf{Y}_t - \frac{\alpha_t^2}{2} \mathbf{X}_t g_t^\top g_t - O(\alpha_t^3).$$

Let $\mathcal{F}_t$ be the set of all the random variables seen thus far[1] (i.e., from 0 to t).

Proof of Lemma 4.4

Proof. The proof technique follows (Balsubramani et al., 2013) and (Xie et al., 2015). Note that for two square matrices $\mathbf{Q}_i$, $i = 1, 2$, their products $\mathbf{Q}_1\mathbf{Q}_2$ and $\mathbf{Q}_2\mathbf{Q}_1$ have the same spectrum. The spectral norm (i.e., matrix 2-norm) is orthogonal invariant. Hence, we can write

$$\begin{aligned} \lambda_{\min}(\mathbf{Z}_t^\top \mathbf{V}\mathbf{V}^\top \mathbf{Z}_t) &= \lambda_{\min}(\mathbf{V}^\top \mathbf{Z}_t \mathbf{Z}_t^\top \mathbf{V}) = \min_{y \neq 0} \frac{\|\mathbf{V}^\top \mathbf{Z}_t y\|_2^2}{\|y\|_2^2} = \min_{y \neq 0} \frac{\|\mathbf{V}^\top \mathbf{Z}_t y\|_2^2}{\|\mathbf{Z}_t y\|_2^2} \\ &= \min_{y \neq 0} \frac{\|\mathbf{V}^\top (\mathbf{X}_t + \alpha_t \mathbf{A}_t \mathbf{Y}_t) y\|_2^2}{\|(\mathbf{X}_t + \alpha_t \mathbf{A}_t \mathbf{Y}_t) y\|_2^2}. \end{aligned}$$

[1]Mathematically, it's known as a filtration, i.e., sub-sigma algebras such that $\mathcal{F}_t \subset \mathcal{F}_{t+1}$.

First, we have the following two inequalities:

$$\|\mathbf{V}^\top(\mathbf{X}_t + \alpha_t \mathbf{A}_t \mathbf{Y}_t)y\|_2^2 \geq \|\mathbf{V}^\top \mathbf{X}_t y\|_2^2 + 2\alpha_t \langle \mathbf{V}^\top \mathbf{X}_t y, \mathbf{V}^\top \mathbf{A}_t \mathbf{Y}_t y\rangle$$

$$\begin{aligned}\|(\mathbf{X}_t + \alpha_t \mathbf{A}_t \mathbf{Y}_t)y\|_2^2 &= \|\mathbf{X}_t y\|_2^2 + 2\alpha_t \langle \mathbf{X}_t y, \mathbf{A}_t \mathbf{Y}_t y\rangle + \alpha_t^2 \|\mathbf{A}_t \mathbf{Y}_t y\|_2^2 \\ &= \|y\|_2^2 + 2\alpha_t \langle \mathbf{X}_t y, \mathbf{A}_t \mathbf{Y}_t y\rangle + \alpha_t^2 \|\mathbf{A}_t \mathbf{Y}_t y\|_2^2 \\ &\leq \|y\|_2^2 + 2\alpha_t \langle \mathbf{X}_t y, \mathbf{A}_t \mathbf{Y}_t y\rangle + \alpha_t^2 \|\mathbf{A}_t \mathbf{Y}_t\|_2^2 \|y\|_2^2.\end{aligned}$$

Letting $z = y/\|y\|_2$, then $\|z\|_2 = 1$ and we get

$$\begin{aligned}&\|\mathbf{V}^\top(\mathbf{X}_t + \alpha_t \mathbf{A}_t \mathbf{X}_t)y\|_2^2 / \|(\mathbf{X}_t + \alpha_t \mathbf{A}_t \mathbf{Y}_t)y\|_2^2 \\ \geq\ & \|\mathbf{V}^\top \mathbf{X}_t y\|_2^2 + 2\alpha_t \langle \mathbf{V}^\top \mathbf{X}_t y, \mathbf{V}^\top \mathbf{A}_t \mathbf{Y}_t y\rangle / (\|y\|_2^2 + 2\alpha_t \langle \mathbf{X}_t y, \mathbf{A}_t \mathbf{Y}_t y\rangle + \alpha_t^2 \|\mathbf{A}_t \mathbf{Y}_t\|_2^2 \|y\|_2^2) \\ =\ & \|\mathbf{V}^\top \mathbf{X}_t z\|_2^2 + 2\alpha_t \langle \mathbf{V}^\top \mathbf{X}_t z, \mathbf{V}^\top \mathbf{A}_t \mathbf{Y}_t z\rangle / (1 + 2\alpha_t \langle \mathbf{X}_t z, \mathbf{A}_t \mathbf{Y}_t z\rangle + \alpha_t^2 \|\mathbf{A}_t \mathbf{Y}_t\|_2^2) \\ \geq\ & (\|\mathbf{V}^\top \mathbf{X}_t z\|_2^2 + 2\alpha_t \langle \mathbf{V}^\top \mathbf{X}_t z, \mathbf{V}^\top \mathbf{A}_t \mathbf{Y}_t z\rangle)(1 - 2\alpha_t \langle \mathbf{X}_t z, \mathbf{A}_t \mathbf{Y}_t z\rangle - \alpha_t^2 \|\mathbf{A}_t \mathbf{Y}_t\|_2^2) \\ \geq\ & \|\mathbf{V}^\top \mathbf{X}_t z\|_2^2 + 2\alpha_t \langle \mathbf{V}^\top \mathbf{X}_t z, \mathbf{V}^\top \mathbf{A}_t \mathbf{Y}_t z\rangle - 2\alpha_t \|\mathbf{V}^\top \mathbf{X}_t z\|_2^2 \langle \mathbf{X}_t z, \mathbf{A}_t \mathbf{Y}_t z\rangle - 5\alpha_t^2 \|\mathbf{A}_t \mathbf{Y}_t\|_2^2 - 2\alpha_t^3 \|\mathbf{A}_t \mathbf{Y}_t\|_2^3 \\ =\ & \|\mathbf{V}^\top \mathbf{X}_t z\|_2^2 + 2\alpha_t z^\top \mathbf{X}_t^\top (\mathbf{V}\mathbf{V}^\top - \|\mathbf{V}^\top \mathbf{X}_t z\|_2^2 \mathbf{I}) \mathbf{A}_t \mathbf{Y}_t z - 5\alpha_t^2 \|\mathbf{A}_t \mathbf{Y}_t\|_2^2 - 2\alpha_t^3 \|\mathbf{A}_t \mathbf{Y}_t\|_2^3.\end{aligned}$$

Since $\|\mathbf{V}^\top \mathbf{X}_t z\|_2^2 + \|\mathbf{V}_\perp^\top \mathbf{X}_t z\|_2^2 = \|\mathbf{X}_t z\|_2^2 = \|z\|_2^2 = 1$, we have

$$\begin{aligned}\mathbf{V}\mathbf{V}^\top - \|\mathbf{V}^\top \mathbf{X}_t z\|_2^2 \mathbf{I} &= \mathbf{V}\mathbf{V}^\top - \|\mathbf{V}^\top \mathbf{X}_t z\|_2^2 (\mathbf{V}\mathbf{V}^\top + \mathbf{V}_\perp \mathbf{V}_\perp^\top) \\ &= \|\mathbf{V}_\perp^\top \mathbf{X}_t z\|_2^2 \mathbf{V}\mathbf{V}^\top - \|\mathbf{V}^\top \mathbf{X}_t z\|_2^2 \mathbf{V}_\perp \mathbf{V}_\perp^\top.\end{aligned}$$

Then

$$\begin{aligned}&\|\mathbf{V}^\top(\mathbf{X}_t + \alpha_t \mathbf{A}_t \mathbf{Y}_t)y\|_2^2 / \|(\mathbf{X}_t + \alpha_t \mathbf{A}_t \mathbf{Y}_t)y\|_2^2 \\ \geq\ & \|\mathbf{V}^\top \mathbf{X}_t z\|_2^2 + 2\alpha_t z^\top \mathbf{X}_t^\top (\|\mathbf{V}_\perp^\top \mathbf{X}_t z\|_2^2 \mathbf{V}\mathbf{V}^\top - \|\mathbf{V}^\top \mathbf{X}_t z\|_2^2 \mathbf{V}_\perp \mathbf{V}_\perp^\top) \mathbf{A}_t \mathbf{Y}_t z \\ & -5\alpha_t^2 \|\mathbf{A}_t \mathbf{Y}_t\|_2^2 - 2\alpha_t^3 \|\mathbf{A}_t \mathbf{Y}_t\|_2^3.\end{aligned}$$

Suppose $\tilde{y}$ makes

$$\lambda_{\min}(\mathbf{V}^\top \mathbf{Z}_t \mathbf{Z}_t^\top \mathbf{V}) = \|\mathbf{V}^\top(\mathbf{X}_t + \alpha_t \mathbf{A}_t \mathbf{Y}_t)\tilde{y}\|_2^2 / \|(\mathbf{X}_t + \alpha_t \mathbf{A}_t \mathbf{Y}_t)\tilde{y}\|_2^2$$

and accordingly let $\tilde{z} = \tilde{y}/\|\tilde{y}\|_2$. Further denote $b_t = 5\alpha_t^2 \|\mathbf{A}_t \mathbf{Y}_t\|_2^2 + 2\alpha_t^3 \|\mathbf{A}_t \mathbf{Y}_t\|_2^3$. Then

$$\begin{aligned}&\lambda_{\min}(\mathbf{V}^\top \mathbf{Z}_t \mathbf{Z}_t^\top \mathbf{V}) \\ \geq\ & \|\mathbf{V}^\top \mathbf{X}_t \tilde{z}\|_2^2 + 2\alpha_t \tilde{z}^\top \mathbf{X}_t^\top (\|\mathbf{V}_\perp^\top \mathbf{X}_t \tilde{z}\|_2^2 \mathbf{V}\mathbf{V}^\top - \|\mathbf{V}^\top \mathbf{X}_t \tilde{z}\|_2^2 \mathbf{V}_\perp \mathbf{V}_\perp^\top) \mathbf{A}_t \mathbf{Y}_t \tilde{z} - b_t.\end{aligned}$$

Let $a_t = \mathbb{E}[b_t | \mathcal{F}_{t-1}]$. Then we have

$$\begin{aligned}&\mathbb{E}[\lambda_{\min}(\mathbf{V}^\top \mathbf{Z}_t \mathbf{Z}_t^\top \mathbf{V}) | \mathcal{F}_{t-1}] \\ \geq\ & \|\mathbf{V}^\top \mathbf{X}_t \tilde{z}\|_2^2 + 2\alpha_t \tilde{z}^\top \mathbf{X}_t^\top [\|\mathbf{V}_\perp^\top \mathbf{X}_t \tilde{z}\|_2^2 \mathbf{V}\mathbf{V}^\top - \|\mathbf{V}^\top \mathbf{X}_t \tilde{z}\|_2^2 \mathbf{V}_\perp \mathbf{V}_\perp^\top] \mathbb{E}[\mathbf{A}_t \mathbf{Y}_t | \mathcal{F}_{t-1}] \tilde{z} - a_t \\ =\ & \|\mathbf{V}^\top \mathbf{X}_t \tilde{z}\|_2^2 + 2\alpha_t \tilde{z}^\top \mathbf{X}_t^\top [\|\mathbf{V}_\perp^\top \mathbf{X}_t \tilde{z}\|_2^2 \mathbf{V}\mathbf{V}^\top - \|\mathbf{V}^\top \mathbf{X}_t \tilde{z}\|_2^2 \mathbf{V}_\perp \mathbf{V}_\perp^\top] \mathbf{A} \mathbf{X}_t \tilde{z} - a_t \\ =\ & \|\mathbf{V}^\top \mathbf{X}_t \tilde{z}\|_2^2 + 2\alpha_t \|\mathbf{V}_\perp^\top \mathbf{X}_t \tilde{z}\|_2^2 \tilde{z}^\top \mathbf{X}_t^\top \mathbf{V} \Sigma \mathbf{V}^\top \mathbf{X}_t \tilde{z} - 2\alpha_t \|\mathbf{V}^\top \mathbf{X}_t \tilde{z}\|_2^2 \tilde{z}^\top \mathbf{X}_t^\top \mathbf{V}_\perp \Sigma_\perp \mathbf{V}_\perp^\top \mathbf{X}_t \tilde{z} - a_t \\ \geq\ & \|\mathbf{V}^\top \mathbf{X}_t \tilde{z}\|_2^2 + 2\alpha_t \lambda_q \|\mathbf{V}_\perp^\top \mathbf{X}_t \tilde{z}\|_2^2 \tilde{z}^\top \mathbf{X}_t^\top \mathbf{V}\mathbf{V}^\top \mathbf{X}_t \tilde{z} - 2\alpha_t \lambda_{q+1} \|\mathbf{V}^\top \mathbf{X}_t \tilde{z}\|_2^2 \tilde{z}^\top \mathbf{X}_t^\top \mathbf{V}_\perp \mathbf{V}_\perp^\top \mathbf{X}_t \tilde{z} - a_t \\ =\ & \|\mathbf{V}^\top \mathbf{X}_t \tilde{z}\|_2^2 + 2\alpha_t (\lambda_q - \lambda_{q+1}) \|\mathbf{V}^\top \mathbf{X}_t \tilde{z}\|_2^2 \|\mathbf{V}_\perp^\top \mathbf{X}_t \tilde{z}\|_2^2 - a_t \\ =\ & \|\mathbf{V}^\top \mathbf{X}_t \tilde{z}\|_2^2 (1 + 2\alpha_t \tau (1 - \|\mathbf{V}^\top \mathbf{X}_t \tilde{z}\|_2^2)) - a_t.\end{aligned}$$

When $\alpha_t < \frac{1}{4\tau}$, the function $f(x) = x[1 + a(1-x)] = -a(x - \frac{a+1}{2a})^2 + \frac{(a+1)^2}{4a}$ is monotonically increasing on $(-\infty, 1]$ where $a = 2\alpha_t \tau$. Since $1 \geq \|\mathbf{V}^\top \mathbf{X}_t \tilde{z}\|_2^2 \geq \lambda_{\min}(\mathbf{V}^\top \mathbf{X}_t \mathbf{X}_t^\top \mathbf{V})$, we get

$$\begin{aligned}&\|\mathbf{V}^\top \mathbf{X}_t \tilde{z}\|_2^2 (1 + 2\alpha_t \tau (1 - \|\mathbf{V}^\top \mathbf{X}_t \tilde{z}\|_2^2)) - a_t \\ \geq\ & \lambda_{\min}(\mathbf{V}^\top \mathbf{X}_t \mathbf{X}_t^\top \mathbf{V})(1 + 2\alpha_t \tau (1 - \lambda_{\min}(\mathbf{V}^\top \mathbf{X}_t \mathbf{X}_t^\top \mathbf{V}))) - a_t.\end{aligned}$$

Hence, we obtain

$$\mathbb{E}[\lambda_{\min}(\mathbf{V}^\top \mathbf{Z}_t \mathbf{Z}_t^\top \mathbf{V})|\mathcal{F}_{t-1}] \geq \lambda_{\min}(\mathbf{V}^\top \mathbf{X}_t \mathbf{X}_t^\top \mathbf{V})(1 + 2\alpha_t \tau(1 - \lambda_{\min}(\mathbf{V}^\top \mathbf{X}_t \mathbf{X}_t^\top \mathbf{V}))) - a_t.$$

By the assumption $\cos^2\langle \mathbf{X}_t, \mathbf{V}\rangle = \lambda_{\min}(\mathbf{V}^\top \mathbf{X}_t \mathbf{X}_t^\top \mathbf{V}) \geq 1/2$, we get

$$\begin{aligned}
1 - \mathbb{E}[\lambda_{\min}(\mathbf{V}^\top \mathbf{Z}_t \mathbf{Z}_t^\top \mathbf{V})|\mathcal{F}_{t-1}] &\leq 1 - \lambda_{\min}(\mathbf{V}^\top \mathbf{X}_t \mathbf{X}_t^\top \mathbf{V})(1 + 2\alpha_t \tau(1 - \lambda_{\min}(\mathbf{V}^\top \mathbf{X}_t \mathbf{X}_t^\top \mathbf{V}))) + a_t \\
&= (1 - \lambda_{\min}(\mathbf{V}^\top \mathbf{X}_t \mathbf{X}_t^\top \mathbf{V}))(1 - 2\alpha_t \tau \lambda_{\min}(\mathbf{V}^\top \mathbf{X}_t \mathbf{X}_t^\top \mathbf{V})) + a_t \\
&\leq (1 - \lambda_{\min}(\mathbf{V}^\top \mathbf{X}_t \mathbf{X}_t^\top \mathbf{V}))(1 - \alpha_t \tau) + a_t,
\end{aligned}$$

and thus

$$\mathbb{E}[1 - \mathbb{E}[\lambda_{\min}(\mathbf{V}^\top \mathbf{Z}_t \mathbf{Z}_t^\top \mathbf{V})|\mathcal{F}_{t-1}]] \leq (1 - \alpha_t \tau)\mathbb{E}[1 - \lambda_{\min}(\mathbf{V}^\top \mathbf{X}_t \mathbf{X}_t^\top \mathbf{V})] + \mathbb{E}[a_t],$$

where

$$\mathbb{E}[a_t] = 5\alpha_t^2 \mathbb{E}[\|\mathbf{A}_t \mathbf{Y}_t\|_2^2] + O(\alpha_t^3) \leq 5\alpha_t^2 \mathbb{E}[\|\mathbf{A}_t\|_2^2]\mathbb{E}[\|\mathbf{Y}_t\|_2^2] + O(\alpha_t^3).$$

Therefore, we arrive at

$$1 - \mathbb{E}[\lambda_{\min}(\mathbf{V}^\top \mathbf{Z}_t \mathbf{Z}_t^\top \mathbf{V})] \leq (1 - \alpha_t \tau)(1 - \mathbb{E}[\lambda_{\min}(\mathbf{V}^\top \mathbf{X}_t \mathbf{X}_t^\top \mathbf{V})]) + 5\alpha_t^2 \mathbb{E}[\|\mathbf{A}_t\|_2^2]\mathbb{E}[\|\mathbf{Y}_t\|_2^2] + O(\alpha_t^3),$$

i.e.,

$$1 - \mathbb{E}[\cos^2\langle \mathbf{Z}_t, \mathbf{V}\rangle] \leq (1 - \alpha_t \tau)(1 - \mathbb{E}[\cos^2\langle \mathbf{X}_t, \mathbf{V}\rangle]) + 5\beta_t \alpha_t^2 + O(\alpha_t^3).$$

□

Proof of Lemma 4.5

Proof. Let $\mathbf{\Phi}(\alpha_t) = (\tilde{\mathbf{Z}}_t^\top \tilde{\mathbf{Z}}_t)^{-1}$. Then $\mathbf{Z}_t \mathbf{Z}_t^\top = \tilde{\mathbf{Z}}_t \mathbf{\Phi}(\alpha_t) \tilde{\mathbf{Z}}_t^\top$ and

$$\begin{aligned}
\mathbf{\Phi}(\alpha_t) &= \mathbf{\Phi}(0) + \mathbf{\Phi}'(0)\alpha_t + \frac{1}{2}\mathbf{\Phi}''(0)\alpha_t^2 \\
\frac{d\mathbf{\Phi}(\alpha_t)}{d\alpha_t} &= -\mathbf{\Phi}(\alpha_t)\frac{d\mathbf{\Phi}^{-1}(\alpha_t)}{d\alpha_t}\mathbf{\Phi}(\alpha_t) \\
\frac{d^2\mathbf{\Phi}(\alpha_t)}{d\alpha_t^2} &= -\frac{d\mathbf{\Phi}(\alpha_t)}{d\alpha_t}\frac{d\mathbf{\Phi}^{-1}(\alpha_t)}{d\alpha_t}\mathbf{\Phi}(\alpha_t) - \mathbf{\Phi}(\alpha_t)\frac{d^2\mathbf{\Phi}^{-1}(\alpha_t)}{d\alpha_t^2}\mathbf{\Phi}(\alpha_t) - \mathbf{\Phi}(\alpha_t)\frac{d\mathbf{\Phi}^{-1}(\alpha_t)}{d\alpha_t}\frac{d\mathbf{\Phi}(\alpha_t)}{d\alpha_t} \\
\mathbf{\Phi}^{-1}(\alpha_t) &= \tilde{\mathbf{Z}}_t^\top \tilde{\mathbf{Z}}_t = \mathbf{I} + \alpha_t \mathbf{X}_t^\top \mathbf{A}_t \mathbf{Y}_t + \alpha_t \mathbf{Y}_t^\top \mathbf{A}_t^\top \mathbf{X}_t + \alpha_t^2 \mathbf{Y}_t^\top \mathbf{A}_t^\top \mathbf{A}_t \mathbf{Y}_t.
\end{aligned}$$

Hence, we can get

$$\begin{aligned}
\mathbf{\Phi}(0) &= \mathbf{I} \\
\mathbf{\Phi}'(0) &= -(\mathbf{X}_t^\top \mathbf{A}_t \mathbf{Y}_t + \mathbf{Y}_t^\top \mathbf{A}_t^\top \mathbf{X}_t) \\
\mathbf{\Phi}''(0) &= 2[\mathbf{\Phi}'(0)]^2 - 2\mathbf{Y}_t^\top \mathbf{A}_t^\top \mathbf{A}_t \mathbf{Y}_t \\
&= 2(\mathbf{X}_t^\top \mathbf{A}_t \mathbf{Y}_t \mathbf{X}_t^\top \mathbf{A}_t \mathbf{Y}_t + \mathbf{X}_t^\top \mathbf{A}_t \mathbf{Y}_t \mathbf{Y}_t^\top \mathbf{A}_t^\top \mathbf{X}_t + \mathbf{Y}_t^\top \mathbf{A}_t^\top \mathbf{X}_t \mathbf{X}_t^\top \mathbf{A}_t \mathbf{Y}_t + \\
&\quad \mathbf{Y}_t^\top \mathbf{A}_t^\top \mathbf{X}_t \mathbf{Y}_t^\top \mathbf{A}_t^\top \mathbf{X}_t - \mathbf{Y}_t^\top \mathbf{A}_t^\top \mathbf{A}_t \mathbf{Y}_t)
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{Z}_t \mathbf{Z}_t^\top &= (\mathbf{X}_t + \alpha_t \mathbf{A}_t \mathbf{Y}_t)(\mathbf{X}_t + \alpha_t \mathbf{A}_t \mathbf{Y}_t)^\top + \alpha_t(\mathbf{X}_t + \alpha_t \mathbf{A}_t \mathbf{Y}_t)\mathbf{\Phi}'(0)(\mathbf{X}_t + \alpha_t \mathbf{A}_t \mathbf{Y}_t)^\top + \\
&\quad (1/2)\alpha_t^2(\mathbf{X}_t + \alpha_t \mathbf{A}_t \mathbf{Y}_t)\mathbf{\Phi}''(0)(\mathbf{X}_t + \alpha_t \mathbf{A}_t \mathbf{Y}_t)^\top \\
&\triangleq \triangle_1 + \alpha_t \triangle_2 + \frac{1}{2}\alpha_t^2 \triangle_3 .
\end{aligned}$$

Expanding each item above, we have

$$
\begin{aligned}
\triangle_1 &= \mathbf{X}_t\mathbf{X}_t^\top + \alpha_t\mathbf{X}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top + \alpha_t\mathbf{A}_t\mathbf{Y}_t\mathbf{X}_t^\top + \alpha_t^2\mathbf{A}_t\mathbf{Y}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top \\
\triangle_2 &= -(\mathbf{X}_t + \alpha_t\mathbf{A}_t\mathbf{Y}_t)(\mathbf{X}_t^\top\mathbf{A}_t\mathbf{Y}_t + \mathbf{Y}_t^\top\mathbf{A}_t^\top\mathbf{X}_t)(\mathbf{X}_t + \alpha_t\mathbf{A}_t\mathbf{Y}_t)^\top \\
&= -\mathbf{X}_t\mathbf{X}_t^\top\mathbf{A}_t\mathbf{Y}_t\mathbf{X}_t^\top - \mathbf{X}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top\mathbf{X}_t\mathbf{X}_t^\top \\
&\quad -\alpha_t\mathbf{X}_t\mathbf{X}_t^\top\mathbf{A}_t\mathbf{Y}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top - \alpha_t\mathbf{X}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top\mathbf{X}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top - \alpha_t\mathbf{A}_t\mathbf{Y}_t\mathbf{X}_t^\top\mathbf{A}_t\mathbf{Y}_t\mathbf{X}_t^\top - \alpha_t\mathbf{A}_t\mathbf{Y}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top\mathbf{X}_t\mathbf{X}_t^\top \\
&\quad -\alpha_t^2\mathbf{A}_t\mathbf{Y}_t\mathbf{X}_t^\top\mathbf{A}_t\mathbf{Y}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top - \alpha_t^2\mathbf{A}_t\mathbf{Y}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top\mathbf{X}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top \\
\triangle_3 &= 2(\mathbf{X}_t\mathbf{X}_t^\top\mathbf{A}_t\mathbf{Y}_t\mathbf{X}_t^\top\mathbf{A}_t\mathbf{Y}_t\mathbf{X}_t^\top + \mathbf{X}_t\mathbf{X}_t^\top\mathbf{A}_t\mathbf{Y}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top\mathbf{X}_t\mathbf{X}_t^\top + \mathbf{X}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top\mathbf{X}_t\mathbf{X}_t^\top\mathbf{A}_t\mathbf{Y}_t\mathbf{X}_t^\top \\
&\quad +\mathbf{X}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top\mathbf{X}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top\mathbf{X}_t\mathbf{X}_t^\top - \mathbf{X}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top\mathbf{A}_t\mathbf{Y}_t\mathbf{X}_t^\top) + O(\alpha_t).
\end{aligned}
$$

Accordingly,

$$
\begin{aligned}
\mathbf{Z}_t\mathbf{Z}_t^\top &= \mathbf{X}_t\mathbf{X}_t^\top + \alpha_t(\mathbf{X}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top + \mathbf{A}_t\mathbf{Y}_t\mathbf{X}_t^\top - \mathbf{X}_t\mathbf{X}_t^\top\mathbf{A}_t\mathbf{Y}_t\mathbf{X}_t^\top - \mathbf{X}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top\mathbf{X}_t\mathbf{X}_t^\top) \\
&\quad +\alpha_t^2(\mathbf{A}_t\mathbf{Y}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top - \mathbf{X}_t\mathbf{X}_t^\top\mathbf{A}_t\mathbf{Y}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top - \mathbf{X}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top\mathbf{X}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top - \mathbf{A}_t\mathbf{Y}_t\mathbf{X}_t^\top\mathbf{A}_t\mathbf{Y}_t\mathbf{X}_t^\top - \mathbf{A}_t\mathbf{Y}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top\mathbf{X}_t\mathbf{X}_t^\top \\
&\quad +\mathbf{X}_t\mathbf{X}_t^\top\mathbf{A}_t\mathbf{Y}_t\mathbf{X}_t^\top\mathbf{A}_t\mathbf{Y}_t\mathbf{X}_t^\top + \mathbf{X}_t\mathbf{X}_t^\top\mathbf{A}_t\mathbf{Y}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top\mathbf{X}_t\mathbf{X}_t^\top + \mathbf{X}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top\mathbf{X}_t\mathbf{X}_t^\top\mathbf{A}_t\mathbf{Y}_t\mathbf{X}_t^\top \\
&\quad +\mathbf{X}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top\mathbf{X}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top\mathbf{X}_t\mathbf{X}_t^\top - \mathbf{X}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top\mathbf{A}_t\mathbf{Y}_t\mathbf{X}_t^\top) \pm O(\alpha_t^3).
\end{aligned}
$$

On the other hand for $\mathbf{X}_{t+1}$, we have

$$
\begin{aligned}
&\mathbf{X}_{t+1}\mathbf{X}_{t+1}^\top \\
= \;& (\mathbf{X}_t + \alpha_t\mathbf{A}_t\mathbf{Y}_t - \alpha_t\mathbf{X}_t\mathbf{X}_t^\top\mathbf{A}_t\mathbf{Y}_t - \frac{\alpha_t^2}{2}\mathbf{X}_t g_t^\top g_t - O(\alpha_t^3))(\mathbf{X}_t + \alpha_t\mathbf{A}_t\mathbf{Y}_t - \alpha_t\mathbf{X}_t\mathbf{X}_t^\top\mathbf{A}_t\mathbf{Y}_t - \frac{\alpha_t^2}{2}\mathbf{X}_t g_t^\top g_t - O(\alpha_t^3))^\top \\
= \;& \mathbf{X}_t\mathbf{X}_t^\top + \alpha_t\mathbf{X}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top - \alpha_t\mathbf{X}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top\mathbf{X}_t\mathbf{X}_t^\top - \frac{\alpha_t^2}{2}\mathbf{X}_t g_t^\top g_t\mathbf{X}_t^\top + \\
& \alpha_t\mathbf{A}_t\mathbf{Y}_t\mathbf{X}_t^\top + \alpha_t^2\mathbf{A}_t\mathbf{Y}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top - \alpha_t^2\mathbf{A}_t\mathbf{Y}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top\mathbf{X}_t\mathbf{X}_t^\top - \\
& \alpha_t\mathbf{X}_t\mathbf{X}_t^\top\mathbf{A}_t\mathbf{Y}_t\mathbf{X}_t^\top - \alpha_t^2\mathbf{X}_t\mathbf{X}_t^\top\mathbf{A}_t\mathbf{Y}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top + \\
& \alpha_t^2\mathbf{X}_t\mathbf{X}_t^\top\mathbf{A}_t\mathbf{Y}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top\mathbf{X}_t\mathbf{X}_t^\top - \frac{\alpha_t^2}{2}\mathbf{X}_t g_t^\top g_t\mathbf{X}_t^\top \pm O(\alpha_t^3).
\end{aligned}
$$

Note that $g_t = (\mathbf{I} - \mathbf{X}_t\mathbf{X}_t^\top)\mathbf{A}_t\mathbf{Y}_t$ and $\mathbf{X}_t g_t^\top g_t\mathbf{X}_t^\top = \mathbf{X}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top(\mathbf{I} - \mathbf{X}_t\mathbf{X}_t^\top)\mathbf{A}_t\mathbf{Y}_t\mathbf{X}_t^\top$. Thus,

$$
\begin{aligned}
\mathbf{X}_{t+1}\mathbf{X}_{t+1}^\top &= \mathbf{X}_t\mathbf{X}_t^\top + \alpha_t(\mathbf{X}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top - \mathbf{X}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top\mathbf{X}_t\mathbf{X}_t^\top + \mathbf{A}_t\mathbf{Y}_t\mathbf{X}_t^\top - \mathbf{X}_t\mathbf{X}_t^\top\mathbf{A}_t\mathbf{Y}_t\mathbf{X}_t^\top) + \\
&\quad \alpha_t^2(\mathbf{A}_t\mathbf{Y}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top - \mathbf{A}_t\mathbf{Y}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top\mathbf{X}_t\mathbf{X}_t^\top - \mathbf{X}_t\mathbf{X}_t^\top\mathbf{A}_t\mathbf{Y}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top + \mathbf{X}_t\mathbf{X}_t^\top\mathbf{A}_t\mathbf{Y}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top\mathbf{X}_t\mathbf{X}_t^\top - \\
&\quad \mathbf{X}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top\mathbf{A}_t\mathbf{Y}_t\mathbf{X}_t^\top + \mathbf{X}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top\mathbf{X}_t\mathbf{X}_t^\top\mathbf{A}_t\mathbf{Y}_t\mathbf{X}_t^\top) \pm O(\alpha_t^3).
\end{aligned}
$$

We find that

$$
\mathbf{X}_{t+1}\mathbf{X}_{t+1}^\top = \mathbf{Z}_t\mathbf{Z}_t^\top - \alpha_t^2\mathbf{M}_t \pm O(\alpha_t^3),
$$

where $\mathbf{M}_t = \mathbf{M}_t^\top$ and

$$
\mathbf{M}_t = -\mathbf{X}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top\mathbf{X}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top - \mathbf{A}_t\mathbf{Y}_t\mathbf{X}_t^\top\mathbf{A}_t\mathbf{Y}_t\mathbf{X}_t^\top + \mathbf{X}_t\mathbf{X}_t^\top\mathbf{A}_t\mathbf{Y}_t\mathbf{X}_t^\top\mathbf{A}_t\mathbf{Y}_t\mathbf{X}_t^\top + \mathbf{X}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top\mathbf{X}_t\mathbf{Y}_t^\top\mathbf{A}_t^\top\mathbf{X}_t\mathbf{X}_t^\top.
$$

Hence, we can write

$$
\begin{aligned}
\lambda_{\min}(\mathbf{V}^\top\mathbf{X}_{t+1}\mathbf{X}_{t+1}^\top\mathbf{V})] &= \lambda_{\min}(\mathbf{V}^\top\mathbf{Z}_t\mathbf{Z}_t^\top\mathbf{V} - \alpha_t^2\mathbf{V}^\top\mathbf{M}_t\mathbf{V}) \pm O(\alpha_t^3) \\
&\geq \lambda_{\min}(\mathbf{V}^\top\mathbf{Z}_t\mathbf{Z}_t^\top\mathbf{V}) - \alpha_t^2\lambda_{\max}(\mathbf{V}^\top\mathbf{M}_t\mathbf{V}) \pm O(\alpha_t^3).
\end{aligned}
$$

By the Poincare separation theorem, $\lambda_{\max}(\mathbf{V}^\top\mathbf{M}_t\mathbf{V}) \leq \lambda_{\max}(\mathbf{M}_t)$. Also note that $\lambda_{\max}(\mathbf{M}_t) \leq \|\mathbf{M}_t\|_2$ and $\|\mathbf{M}_t\|_2 \leq 4\|\mathbf{A}_t\mathbf{Y}_t\|_2^2$. Then we get

$$
\lambda_{\max}(\mathbf{V}^\top\mathbf{M}_t\mathbf{V}) \leq 4\|\mathbf{A}_t\mathbf{Y}_t\|_2^2 \leq 4\|\mathbf{A}_t\|_2^2\|\mathbf{Y}_t\|_2^2
$$

and finally arrive at

$$\lambda_{\min}(\mathbf{V}^\top \mathbf{X}_{t+1}\mathbf{X}_{t+1}^\top \mathbf{V})] \geq \lambda_{\min}(\mathbf{V}^\top \mathbf{Z}_t\mathbf{Z}_t^\top \mathbf{V}) - 4\alpha_t^2\|\mathbf{A}_t\|_2^2\|\mathbf{Y}_t\|_2^2 \pm O(\alpha_t^3),$$

i.e.,

$$\cos^2\langle \mathbf{X}_{t+1}, \mathbf{V}\rangle \geq \cos^2\langle \mathbf{Z}_t, \mathbf{V}\rangle - 4\alpha_t^2\|\mathbf{A}_t\|_2^2\|\mathbf{Y}_t\|_2^2 \pm O(\alpha_t^3).$$

□

Proof of Lemma 4.6

Proof. By Lemma 4.5 and 4.6, we have

$$\begin{aligned} 1 - \mathbb{E}[\cos^2\langle \mathbf{X}_{t+1}, \mathbf{V}\rangle] &\leq 1 - \mathbb{E}[\cos^2\langle \mathbf{Z}_t, \mathbf{V}\rangle] + 4\beta_t\alpha_t^2 \pm O(\alpha_t^3) \\ &\leq (1-\alpha_t\tau)(1 - \mathbb{E}[\cos^2\langle \mathbf{X}_t, \mathbf{V}\rangle]) + 9\beta_t\alpha_t^2 \pm O(\alpha_t^3). \end{aligned}$$

Thus, we can write

$$\Theta_{t+1} \leq (1-\alpha_t\tau)\Theta_t + \gamma\beta_t\alpha_t^2,$$

for some constant $\gamma > 9$.

□

Remark From the proof of Lemma 4.4, given $\cos^2\langle \mathbf{X}_t, \mathbf{V}\rangle \geq 1/2$, we have that

$$1 - \cos^2\langle \mathbf{Z}_t, \mathbf{V}\rangle \leq (1-\alpha_t\tau)(1-\cos^2\langle \mathbf{X}_t, \mathbf{V}\rangle) + 5b_t\alpha_t^2 + O(\alpha_t^3),$$

which combined with Lemma 4.5 yields

$$1 - \cos^2\langle \mathbf{X}_{t+1}, \mathbf{V}\rangle \leq (1-\alpha_t\tau)(1-\cos^2\langle \mathbf{X}_t, \mathbf{V}\rangle) + 5b_t\alpha_t^2 + O(\alpha_t^3).$$

Then with $\alpha_t = \frac{c}{t}$ where $c > 0$, we have

$$\begin{aligned} 1 - \cos^2\langle \mathbf{X}_{t+1}, \mathbf{V}\rangle &\leq \frac{1}{2}(1 - \frac{c\tau}{t}) + \frac{5b_tc^2}{t^2} + O(\frac{1}{t^3}) \\ &= \frac{1}{2} - \frac{c}{t}(\frac{\tau}{2} - \frac{5b_tc}{t} - O(\frac{1}{t^2})). \end{aligned}$$

Thus, when t is sufficiently large, we get $1 - \cos^2\langle \mathbf{X}_{t+1}, \mathbf{V}\rangle \leq \frac{1}{2}$, i.e., $\cos^2\langle \mathbf{X}_{t+1}, \mathbf{V}\rangle \geq \frac{1}{2}$. This says that for $t > s$ where s is sufficiently large, $\cos^2\langle \mathbf{X}_t, \mathbf{V}\rangle \geq 1/2$ implies $\cos^2\langle \mathbf{X}_{t+1}, \mathbf{V}\rangle \geq 1/2$.

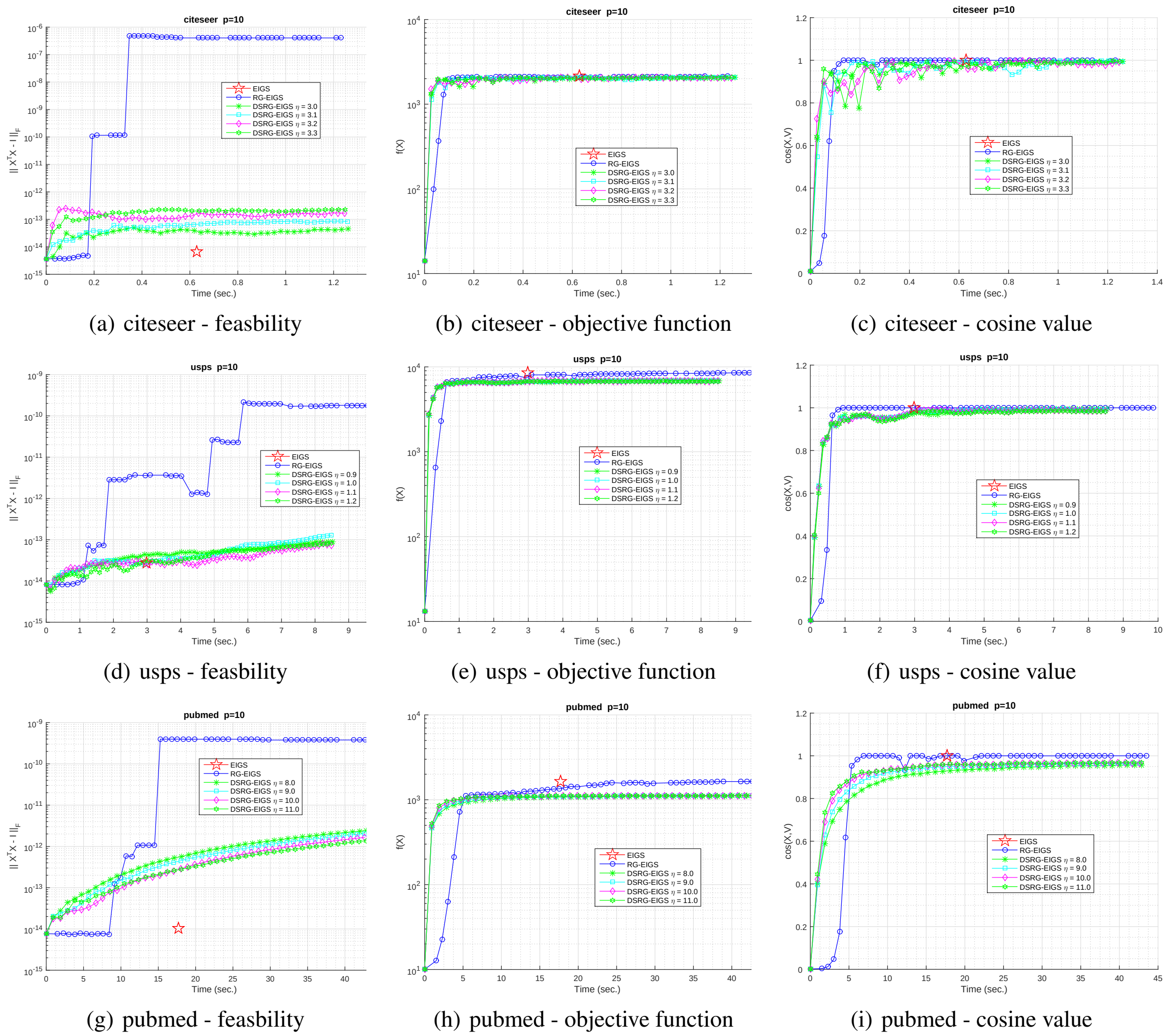
Proof of Theorem 4.2

Proof. We only need to prove that there exists a sufficiently large constant $\sigma > 0$ such that $\Theta_t \leq \frac{\sigma}{t}$. Induction is employed. It's clear that Θ_s satisfies the inequality. Suppose it holds for $t > s$. Then by Lemma 4.6 as well as assumptions that $c\tau > 2$ and σ is sufficiently large,

$$\begin{aligned} \Theta_{t+1} &\leq (1 - \frac{c}{t}\tau)\frac{\sigma}{t} + \gamma\beta_t\frac{c^2}{t^2} \\ &= \frac{1}{t^2}(\sigma(t - c\tau) + \gamma\beta_tc^2) \\ &= \frac{1}{t^2}(\sigma(t-1) + \sigma(1-c\tau) + \gamma\beta_tc^2) \\ &\leq \frac{1}{t^2}(\sigma(t-1) - (\sigma - \gamma\beta_tc^2)) \\ &\leq \frac{\sigma(t-1)}{t^2} \leq \frac{\sigma(t-1)}{t^2-1} = \frac{\sigma}{t+1}. \end{aligned}$$

□

Performance on Dense Matrices



(a) citeseer - feasbility (b) citeseer - objective function (c) citeseer - cosine value

(d) usps - feasbility (e) usps - objective function (f) usps - cosine value

(g) pubmed - feasbility (h) pubmed - objective function (i) pubmed - cosine value

Figure 1. Performance on dense matrices.

References

Balsubramani, Akshay, Dasgupta, Sanjoy, and Freund, Yoav. The fast convergence of incremental PCA. In *Advances in Neural Information Processing Systems 26, December 5-8, 2013, Lake Tahoe, Nevada, United States.*, pp. 3174–3182, 2013.

Xie, Bo, Liang, Yingyu, and Song, Le. Scale up nonlinear component analysis with doubly stochastic gradients. *CoRR*, abs/1504.03655, 2015.