Supplementary Material for
"On Truly Block Eigensolvers via Riemannian Optimization"

Zhiqiang Xu
zhiqiangxu2001@gmail.com

For ease of exposition, we use ρ, η, ξ to represent positive numerical constants with possibly varying values at different places or cases even in the same line.

Part A: Getting Started

We start from defining a unified update as

$$\mathbf{X}_{t+1} = R(\mathbf{X}_t, \alpha_{t+1}\tilde{\nabla} f(\mathbf{X}_t) + \alpha_{t+1}\mathbf{W}_t),$$

where $\mathbf{W}_t$ represents the stochastic zero-mean term. Specifically,

$$\mathbf{W}_t = \begin{cases} \mathbf{0}, & \text{Solver 1} \\ (\mathbf{I} - \mathbf{X}_t\mathbf{X}_t^\top)(\mathbf{A}_{t+1} - \mathbf{A})\mathbf{X}_t, & \text{Solver 2} \\ (\mathbf{I} - \mathbf{X}_t\mathbf{X}_t^\top)(\mathbf{A}_{t+1} - \mathbf{A})(\mathbf{X}_t - \tilde{\mathbf{X}}\mathbf{B}_t) + & \\ (\mathbf{I} - \mathbf{X}_t\mathbf{X}_t^\top)\tilde{\mathbf{X}}\tilde{\mathbf{X}}^\top(\mathbf{A}_{t+1} - \mathbf{A})\tilde{\mathbf{X}}\mathbf{B}_t - & \\ \mathbf{X}_t \mathrm{skew}(\mathbf{X}_t^\top(\mathbf{I} - \tilde{\mathbf{X}}\tilde{\mathbf{X}}^\top)(\mathbf{A}_{t+1} - \mathbf{A})\tilde{\mathbf{X}}\mathbf{B}_t), & \text{Solver 3} \end{cases}.$$

Without loss of generality, assume that $l \geq k$. We then can write

$$\det(\mathbf{X}_t^\top \mathbf{V}_l \mathbf{V}_l^\top \mathbf{X}_t) = \frac{\det(a_1(\mathbf{X}_t) + b_1(\mathbf{W}_t))}{\det(a_2(\mathbf{X}_t) + b_2(\mathbf{W}_t))},$$

where

$$\begin{aligned} \mathbf{Y}_t &= \mathbf{X}_t + \alpha_{t+1}\tilde{\nabla} f(\mathbf{X}_t), \\ a_1(\mathbf{X}_t) &= \mathbf{Y}_t^\top \mathbf{V}_l \mathbf{V}_l^\top \mathbf{Y}_t, \\ b_1(\mathbf{W}_t) &= 2\alpha_{t+1}\mathrm{sym}(\mathbf{Y}_t^\top \mathbf{V}_l \mathbf{V}_l^\top \mathbf{W}_t) + \alpha_{t+1}^2 \mathbf{W}_t^\top \mathbf{V}_l \mathbf{V}_l^\top \mathbf{W}_t, \\ a_2(\mathbf{X}_t) &= \mathbf{Y}_t^\top \mathbf{Y}_t + \alpha_{t+1}^2 \mathbf{W}_t^\top \mathbf{W}_t, \\ b_2(\mathbf{W}_t) &= 2\alpha_{t+1}\mathrm{sym}(\mathbf{Y}_t^\top \mathbf{W}_t), \end{aligned}$$

and

$$\mathbb{E}[b_1(\mathbf{W}_t)|\mathbf{X}_t] \succeq \mathbf{0}, \quad \mathbb{E}[b_2(\mathbf{W}_t)|\mathbf{X}_t] = \mathbf{0}.$$

Due to $\mathbf{W}_t^\top \mathbf{W}_t \preceq \beta_t \mathbf{I}$, we have

$$\det(\mathbf{X}^\top \mathbf{V}_l \mathbf{V}_l^\top \mathbf{X}) \geq \frac{\det(a_1(\mathbf{X}_t) + b_1(\mathbf{W}_t))}{\det(a_2(\mathbf{X}_t) + b_2(\mathbf{W}_t))},$$

where now

$$a_2(\mathbf{X}_t) = \mathbf{Y}_t^\top \mathbf{Y}_t + \alpha_{t+1}^2 \beta_t \mathbf{I}$$

with a bit abuse of notation, and conditioned on $\mathbf{X}_t$ for $i = 1, 2$, $a_i(\mathbf{X}_t)$ are deterministic functions of $\mathbf{X}_t$ while $b_i(\mathbf{W}_t)$ are stochastic functions of $\mathbf{W}_t$.

Part B. Auxiliary Lemmas

Lemma B.1. For any $\mathbf{X} \in \mathrm{St}(n,k)$ and $\mathbf{Y} \in \mathrm{St}(n,l)$,

$$\Psi(\mathbf{X},\mathbf{Y}) \leq \Theta(\mathbf{X},\mathbf{Y}) = \min\{k,l\} - \|\mathbf{X}^\top \mathbf{Y}\|_F^2 \leq \min\{k,l\}\Psi(\mathbf{X},\mathbf{Y}).$$

Proof. Let $p = \min\{k,l\}$. Note that

$$\Psi(\mathbf{X},\mathbf{Y}) = 1 - \prod_{i=1}^{p} \cos^2\theta_i \quad \text{and} \quad \Theta(\mathbf{X},\mathbf{Y}) = p - \sum_{i=1}^{p} \cos^2\theta_i.$$

We prove the left inequality by induction. When $p = 1$,

$$\Psi(\mathbf{X},\mathbf{Y}) = 1 - \cos^2\theta_1 = \Theta(\mathbf{X},\mathbf{Y}).$$

Given $\Psi(\mathbf{X},\mathbf{Y}) \leq \Theta(\mathbf{X},\mathbf{Y})$ for p, then for $p+1$,

$$\begin{aligned}
\Theta(\mathbf{X},\mathbf{Y}) &= p + 1 - \sum_{i=1}^{p+1} \cos^2\theta_i \\
&= p - \sum_{i=1}^{p} \cos^2\theta_i + 1 - \cos^2\theta_{p+1} \\
&\geq 1 - \prod_{i=1}^{p} \cos^2\theta_i + 1 - \cos^2\theta_{p+1} - \left(1 - \prod_{i=1}^{p+1} \cos^2\theta_i\right) + 1 - \prod_{i=1}^{p+1} \cos^2\theta_i \\
&= \left(1 - \cos^2\theta_{p+1}\right)\left(1 - \prod_{i=1}^{p} \cos^2\theta_i\right) + 1 - \prod_{i=1}^{p+1} \cos^2\theta_i \\
&\geq 1 - \prod_{i=1}^{p+1} \cos^2\theta_i = \Psi(\mathbf{X},\mathbf{Y}).
\end{aligned}$$

For the right inequality, by the generalized mean inequality, we have

$$\sum_{i=1}^{p} \cos^2\theta_i \quad = \quad p\left(\frac{\sum_{i=1}^{p} \cos^2\theta_i}{p}\right)^{\frac{1}{2}\cdot 2} \geq p\left(\prod_{i=1}^{p} \cos\theta_i\right)^{\frac{2}{p}} \geq p\left(\prod_{i=1}^{p} \cos\theta_i\right)^{2}.$$

Thus, we get

$$\Theta(\mathbf{X},\mathbf{Y}) \leq p - p\left(\prod_{i=1}^{p} \cos\theta_i\right)^2 = p\left(1 - \prod_{i=1}^{p} \cos^2\theta_i\right) = p\Psi(\mathbf{X},\mathbf{Y}).$$

□

Lemma B.2. For any $\mathbf{X},\mathbf{Y},\mathbf{Z} \in \mathrm{Grass}(n,k)$,

$$\Psi^{\frac{1}{2}}(\mathbf{X},\mathbf{Y}) \leq \Psi^{\frac{1}{2}}(\mathbf{X},\mathbf{Z}) + \Psi^{\frac{1}{2}}(\mathbf{Z},\mathbf{Y}).$$

Proof. Let $\mathbf{x} \in \mathbb{R}^{\binom{n}{k}}$ be a column vector of all the $k \times k$ minors[1] of $\mathbf{X}$ in certain order. Similarly, let $\mathbf{y}$ and $\mathbf{z}$ be the counterparts for $\mathbf{Y}$ and $\mathbf{Z}$, respectively, with minors placed in the same order as $\mathbf{x}$. According to the Binet-Cauchy formula, we have $\det(\mathbf{X}^\top\mathbf{Y}) = \mathbf{x}^\top\mathbf{y}$. Then for any $\mathbf{X},\mathbf{Y},\mathbf{Z} \in \mathrm{Grass}(n,k)$, we have

$$\left\|\mathbf{x}\mathbf{x}^\top - \mathbf{y}\mathbf{y}^\top\right\|_F \leq \left\|\mathbf{x}\mathbf{x}^\top - \mathbf{z}\mathbf{z}^\top\right\|_F + \left\|\mathbf{z}\mathbf{z}^\top - \mathbf{y}\mathbf{y}^\top\right\|_F,$$

[1] A $k \times k$ minor of a matrix $\mathbf{A}$ is the determinant of a $k \times k$ sub-matrix in $\mathbf{A}$.

where

$$\left\|\mathbf{x}\mathbf{x}^\top - \mathbf{y}\mathbf{y}^\top\right\|_F^2 = \mathrm{tr}(\mathbf{x}\mathbf{x}^\top\mathbf{x}\mathbf{x}^\top) - 2\mathrm{tr}(\mathbf{x}\mathbf{x}^\top\mathbf{y}\mathbf{y}^\top) + \mathrm{tr}(\mathbf{y}\mathbf{y}^\top\mathbf{y}\mathbf{y}^\top)$$
$$= \det{}^2(\mathbf{X}^\top\mathbf{X}) - 2\det{}^2(\mathbf{X}^\top\mathbf{Y}) + \det{}^2(\mathbf{Y}^\top\mathbf{Y})$$
$$= 2\left(1 - \det{}^2(\mathbf{X}^\top\mathbf{Y})\right).$$

Thus, we get

$$\left(1 - \det{}^2(\mathbf{X}^\top\mathbf{Y})\right)^{1/2} \le \left(1 - \det{}^2(\mathbf{X}^\top\mathbf{Z})\right)^{1/2} + \left(1 - \det{}^2(\mathbf{Z}^\top\mathbf{Y})\right)^{1/2}.$$

□

Remark The lemma holds on Stiefel manifolds as well.

Lemmma B.3. Let $\beta = \max_i \|\tilde{\mathbf{A}}_i - \mathbf{A}\|_F^2$. Then

$$\|\mathbf{W}_t\|_F^2 \le \beta_t = \begin{cases} 0, & \text{Solver 1} \\ \beta, & \text{Solver 2} \\ 24k\beta\left(\Psi(\mathbf{X}_t, \mathbf{Y}) + \Psi(\tilde{\mathbf{X}}_{s-1}, \mathbf{Y})\right), & \text{Solver 3} \end{cases}$$

for any $\mathbf{Y} \in \mathrm{St}(n, k)$.

Proof. For brevity, we omit subscript t here. Note that

$$\|\mathbf{AB}\|_F \le \|\mathbf{A}\|_2 \|\mathbf{B}\|_F \quad \text{and} \quad \|\mathbf{X}\|_2 = \|\mathbf{X}_\perp\|_2 = 1,$$

where $\mathbf{X}_\perp$ represents the orthogonal complement of $\mathbf{X}$ in $\mathbb{R}^{n\times n}$, i.e.,

$$[\mathbf{X}\ \mathbf{X}_\perp]\,[\mathbf{X}\ \mathbf{X}_\perp]^\top = [\mathbf{X}\ \mathbf{X}_\perp]^\top\,[\mathbf{X}\ \mathbf{X}_\perp] = \mathbf{I}.$$

For Solver 1, $\beta_t = 0$ by definition. For Solver 2, we have

$$\|\mathbf{W}\|_F^2 = \left\|\left(\mathbf{I} - \mathbf{X}\mathbf{X}^\top\right)(\mathbf{A}_{t+1} - \mathbf{A})\,\mathbf{X}\right\|_F^2$$
$$= \left\|\mathbf{X}_\perp\mathbf{X}_\perp^\top(\mathbf{A}_{t+1} - \mathbf{A})\,\mathbf{X}\right\|_F^2$$
$$\le \|\mathbf{X}_\perp\|_2^4\,\|\mathbf{A}_{t+1} - \mathbf{A}\|_F^2\,\|\mathbf{X}\|_2^2$$
$$= \|\mathbf{A}_{t+1} - \mathbf{A}\|_F^2$$
$$\le \max_i \|\tilde{\mathbf{A}}_i - \mathbf{A}\|_F^2 \triangleq \beta_t.$$

For Solver 3, we get

$$\|\mathbf{W}\|_F^2 = \Big\|\left(\mathbf{I} - \mathbf{X}\mathbf{X}^\top\right)(\mathbf{A}_{t+1} - \mathbf{A})\left(\mathbf{X} - \tilde{\mathbf{X}}\mathbf{Q}\right) + \left(\mathbf{I} - \mathbf{X}\mathbf{X}^\top\right)\tilde{\mathbf{X}}\tilde{\mathbf{X}}^\top(\mathbf{A}_{t+1} - \mathbf{A})\,\tilde{\mathbf{X}}\mathbf{Q} - \mathbf{X}\mathrm{skew}\left(\mathbf{X}^\top\left(\mathbf{I} - \tilde{\mathbf{X}}\tilde{\mathbf{X}}^\top\right)(\mathbf{A}_{t+1} - \mathbf{A})\,\tilde{\mathbf{X}}\mathbf{Q}\right)\Big\|_F^2$$
$$\le \beta\left(\left\|\mathbf{X} - \tilde{\mathbf{X}}\mathbf{Q}\right\|_F + \left\|\mathbf{X}_\perp^\top\tilde{\mathbf{X}}\right\|_F + \left\|\mathbf{X}^\top\tilde{\mathbf{X}}_\perp\right\|_F\right)^2$$
$$\le 3\beta\left(\left\|\mathbf{X} - \tilde{\mathbf{X}}\mathbf{Q}\right\|_F^2 + \left\|\mathbf{X}_\perp^\top\tilde{\mathbf{X}}\right\|_F^2 + \left\|\mathbf{X}^\top\tilde{\mathbf{X}}_\perp\right\|_F^2\right).$$

For the first term above, we have

$$\left\|\mathbf{X} - \tilde{\mathbf{X}}\mathbf{B}\right\|_F^2 = 2\left(k - \mathrm{tr}(\mathbf{X}^\top\tilde{\mathbf{X}}\mathbf{B})\right) = 2\left(k - \mathrm{tr}(\hat{\mathbf{P}}\mathbf{\Lambda}\check{\mathbf{P}}^\top\check{\mathbf{P}}\hat{\mathbf{P}}^\top)\right)$$
$$\le 2\left(k - \mathrm{tr}\left(\mathbf{\Lambda}^2\right)\right) = 2\left(k - \left\|\mathbf{X}^\top\tilde{\mathbf{X}}\right\|_F^2\right)$$
$$= 2\Theta(\mathbf{X}, \mathbf{Y}) \le 2k\Psi(\mathbf{X}, \mathbf{Y})$$
$$\le 4k\left(\Psi(\mathbf{X}, \mathbf{Y}) + \Psi(\mathbf{Y}, \tilde{\mathbf{X}})\right). \quad \text{(Lemmas A.1-A.2)}$$

For the second term, it could be derived as follows,

$$\begin{aligned}\left\|\mathbf{X}_{\perp}^{\top}\tilde{\mathbf{X}}\right\|_F^2 &= \left\|\mathbf{X}_{\perp}^{\top}\left(\mathbf{Y}\mathbf{Y}^{\top}+\mathbf{Y}_{\perp}\mathbf{Y}_{\perp}^{\top}\right)\tilde{\mathbf{X}}\right\|_F^2\\ &\leq \left(\left\|\mathbf{X}_{\perp}^{\top}\mathbf{Y}\right\|_F\|\mathbf{Y}\|_2\left\|\tilde{\mathbf{X}}\right\|_2+\|\mathbf{X}_{\perp}\|_2\|\mathbf{Y}_{\perp}\|_2\left\|\mathbf{Y}_{\perp}^{\top}\tilde{\mathbf{X}}\right\|_F\right)^2\\ &= \left(\left(k-\left\|\mathbf{X}^{\top}\mathbf{Y}\right\|_F^2\right)^{1/2}+\left(k-\left\|\mathbf{Y}^{\top}\tilde{\mathbf{X}}\right\|_F^2\right)^{1/2}\right)^2\\ &\leq 2\left(k-\left\|\mathbf{X}^{\top}\mathbf{Y}\right\|_F^2+k-\left\|\tilde{\mathbf{X}}^{\top}\mathbf{Y}\right\|_F^2\right)\\ &\leq 2k\left(\Psi(\mathbf{X},\mathbf{Y})+\Psi(\tilde{\mathbf{X}},\mathbf{Y})\right). \quad \text{(Lemmas A.1)}\end{aligned}$$

Similarly, we have $\left\|\mathbf{X}^{\top}\tilde{\mathbf{X}}_{\perp}\right\|_F^2 \leq 2k\left(\Psi(\mathbf{X},\mathbf{Y})+\Psi(\tilde{\mathbf{X}},\mathbf{Y})\right)$ for the last term. Therefore, we can write

$$\|\mathbf{W}\|_F^2 \leq 24k\beta\left(\Psi(\mathbf{X},\mathbf{Y})+\Psi(\tilde{\mathbf{X}}_{s-1},\mathbf{Y})\right).$$

□

Remark We have $0 \leq \beta_t \leq 48k\beta$ as $\Psi(\mathbf{X},\mathbf{Y})$,$\Psi(\tilde{\mathbf{X}}_{s-1},\mathbf{Y}) \in [0,1]$.

Lemma B.4. Let $a = \|\mathbf{A}\|_2$ and $l \geq k$. If $\det(\mathbf{X}_t^{\top}\mathbf{V}_l\mathbf{V}_l^{\top}\mathbf{X}_t) > \gamma$ and $2\alpha_{t+1}\delta_t + 2\alpha_{t+1}\beta_t^{1/2}(1+\alpha_{t+1}\delta_t) < \gamma$, $0 < \gamma < 1$, then

$$\mathbb{E}[\det(\mathbf{X}_{t+1}^{\top}\mathbf{V}_l\mathbf{V}_l^{\top}\mathbf{X}_{t+1})|\mathbf{X}_t] \geq \frac{\det(a_1(\mathbf{X}_t))}{\det(a_2(\mathbf{X}_t))} - \alpha_{t+1}^2\xi_t\beta_t,$$

where $\delta_t^2 = 4ka^2\Psi(\mathbf{X}_t,\mathbf{V}_k)$ and

$$\begin{aligned}\xi_t &= \frac{2k+1}{2}\left(\frac{(1+\alpha_{t+1}\delta_t+\alpha_{t+1}\beta_t^{1/2})^2}{1-2\alpha_{t+1}\beta_t^{1/2}(1+\alpha_{t+1}\delta_t)}\right)^k\\ &\quad\left(\left(\frac{2+2\alpha_{t+1}\delta_t+\alpha_{t+1}\beta_t^{1/2}}{\gamma-2\alpha_{t+1}\delta_t-2\alpha_{t+1}\beta_t^{1/2}(1+\alpha_{t+1}\delta_t)}\right)^2+4\left(\frac{1+\alpha_{t+1}\delta_t}{1-2\alpha_{t+1}\beta_t^{1/2}(1+\alpha_{t+1}\delta_t)}\right)^2\right).\end{aligned}$$

Proof. Note that

$$\begin{aligned}\det(\mathbf{X}_{t+1}^{\top}\mathbf{V}_l\mathbf{V}_l^{\top}\mathbf{X}_{t+1}) &= \frac{\det\left(a_1(\mathbf{X}_t)+b_1(\mathbf{W}_t)\right)}{\det\left(a_2(\mathbf{X}_t)+b_2(\mathbf{W}_t)\right)},\\ \mathbf{X}_t^{\top}\tilde{\nabla}f(\mathbf{X}_t) &= \mathbf{X}_t^{\top}(\mathbf{I}-\mathbf{X}_t\mathbf{X}_t^{\top})\mathbf{A}\mathbf{X}_t = \mathbf{0},\\ \mathbf{A} &= \mathbf{V}_k\Sigma_k\mathbf{V}_k^{\top}+\mathbf{V}_k^{\perp}\Sigma_k^{\perp}\left(\mathbf{V}_k^{\perp}\right)^{\top}.\end{aligned}$$

Then

$$\begin{aligned}\left\|\tilde{\nabla}f(\mathbf{X}_t)\right\|_F^2 &= \left\|\mathbf{X}_t^{\perp}\left(\mathbf{X}_t^{\perp}\right)^{\top}\mathbf{A}\mathbf{X}_t\right\|_F^2\\ &= \left\|\mathbf{X}_t^{\perp}\left(\mathbf{X}_t^{\perp}\right)^{\top}\mathbf{V}_k\Sigma_k\mathbf{V}_k^{\top}\mathbf{X}_t+\mathbf{X}_t^{\perp}\left(\mathbf{X}_t^{\perp}\right)^{\top}\mathbf{V}_k^{\perp}\Sigma_k^{\perp}\left(\mathbf{V}_k^{\perp}\right)^{\top}\mathbf{X}_t\right\|_F^2\\ &\leq 2\|\mathbf{A}\|_2^2\left(\left\|\left(\mathbf{X}_t^{\perp}\right)^{\top}\mathbf{V}_k\right\|_F^2+\left\|\left(\mathbf{V}_k^{\perp}\right)^{\top}\mathbf{X}_t\right\|_F^2\right)\\ &\leq 4k\|\mathbf{A}\|_2^2\Psi(\mathbf{X}_t,\mathbf{V}_k)\\ &\triangleq \delta_t^2.\end{aligned}$$

Next, we have

$$\begin{aligned} a_2(\mathbf{X}_t) &= \left(\mathbf{X}_t + \alpha_{t+1}\tilde{\nabla} f(\mathbf{X}_t)\right)^\top \left(\mathbf{X}_t + \alpha_{t+1}\tilde{\nabla} f(\mathbf{X}_t)\right) + \alpha_{t+1}^2\beta_t\mathbf{I} \\ &= (1+\alpha_{t+1}^2\beta_t)\mathbf{I} + \alpha_{t+1}\left(\mathbf{X}_t^\top\tilde{\nabla} f(\mathbf{X}_t) + \left(\mathbf{X}_t^\top\tilde{\nabla} f(\mathbf{X}_t)\right)^\top\right) + \alpha_{t+1}^2\tilde{\nabla} f(\mathbf{X}_t)^\top\tilde{\nabla} f(\mathbf{X}_t) \\ &= (1+\alpha_{t+1}^2\beta_t)\mathbf{I} + \alpha_{t+1}^2\left(\mathbf{X}_t^\top\mathbf{A}\mathbf{X}_t^\perp\right)\left(\mathbf{X}_t^\top\mathbf{A}\mathbf{X}_t^\perp\right)^\top \\ &\succcurlyeq (1+\alpha_{t+1}^2\beta_t)\mathbf{I} \\ &\succ \mathbf{0}. \end{aligned}$$

On the other hand, $b_2(\mathbf{W}_t)$ is symmetric and

$$\begin{aligned} \|b_2(\mathbf{W}_t)\|_2 &= \left\|\alpha_{t+1}\left(\mathbf{X}_t + \alpha_{t+1}\tilde{\nabla} f(\mathbf{X}_t)\right)^\top \mathbf{W}_t + \alpha_{t+1}\mathbf{W}_t^\top\left(\mathbf{X}_t + \alpha_{t+1}\tilde{\nabla} f(\mathbf{X}_t)\right)\right\|_2 \\ &\leq 2\alpha_{t+1}\left\|\mathbf{X}_t + \alpha_{t+1}\tilde{\nabla} f(\mathbf{X}_t)\right\|_2 \|\mathbf{W}_t\|_2 \\ &\leq 2\alpha_{t+1}\left(1+\alpha_{t+1}\left\|\tilde{\nabla} f(\mathbf{X}_t)\right\|_2\right)\|\mathbf{W}_t\|_2. \end{aligned}$$

Thus, for $\varsigma \in [0,1]$, we get

$$\begin{aligned} a_2(\mathbf{X}_t) + \varsigma b_2(\mathbf{W}_t) &\succcurlyeq (1+\alpha_{t+1}^2\beta_t)\mathbf{I} - \|b_2(\mathbf{W}_t)\|_2\mathbf{I} \\ &= \left(1+\alpha_{t+1}^2\beta_t - 2\alpha_{t+1}\left(1+\alpha_{t+1}\left\|\tilde{\nabla} f(\mathbf{X}_t)\right\|_2\right)\|\mathbf{W}_t\|_2\right)\mathbf{I} \\ &\succ \mathbf{0}, \end{aligned}$$

and now can define the function

$$f(\varsigma) = \frac{\det\left(a_1(\mathbf{X}_t) + \varsigma b_1(\mathbf{W}_t)\right)}{\det\left(a_2(\mathbf{X}_t) + \varsigma b_2(\mathbf{W}_t)\right)}, \quad \varsigma \in [0,1].$$

In a similar vein, we have

$$\begin{aligned} & a_1(\mathbf{X}_t) + \varsigma b_1(\mathbf{W}_t) \\ = \ & \mathbf{X}_t^\top\mathbf{V}_l\mathbf{V}_l^\top\mathbf{X}_t + \alpha_{t+1}\left(\mathbf{X}_t^\top\mathbf{V}_l\mathbf{V}_l^\top\tilde{\nabla} f(\mathbf{X}_t) + \left(\mathbf{X}_t^\top\mathbf{V}_l\mathbf{V}_l^\top\tilde{\nabla} f(\mathbf{X}_t)\right)^\top\right) + \\ & \varsigma\alpha_{t+1}\left(\left(\mathbf{X}_t + \alpha_{t+1}\tilde{\nabla} f(\mathbf{X}_t)\right)^\top\mathbf{V}_l\mathbf{V}_l^\top\mathbf{W}_t + \mathbf{W}_t^\top\mathbf{V}_l\mathbf{V}_l^\top\left(\mathbf{X}_t + \alpha_{t+1}\tilde{\nabla} f(\mathbf{X}_t)\right)\right) + \\ & \alpha_{t+1}^2\left(\tilde{\nabla} f(\mathbf{X}_t)\right)^\top\mathbf{V}_l\mathbf{V}_l^\top\tilde{\nabla} f(\mathbf{X}_t) + \varsigma\alpha_{t+1}^2\mathbf{W}_t^\top\mathbf{V}_l\mathbf{V}_l^\top\mathbf{W}_t \\ \preccurlyeq \ & \mathbf{I} + \alpha_{t+1}\left\|\mathbf{X}_t^\top\mathbf{V}_l\mathbf{V}_l^\top\tilde{\nabla} f(\mathbf{X}_t) + \left(\mathbf{X}_t^\top\mathbf{V}_l\mathbf{V}_l^\top\tilde{\nabla} f(\mathbf{X}_t)\right)^\top\right\|_2\mathbf{I} + \\ & \alpha_{t+1}\left\|\left(\mathbf{X}_t + \alpha_{t+1}\tilde{\nabla} f(\mathbf{X}_t)\right)^\top\mathbf{V}_l\mathbf{V}_l^\top\mathbf{W}_t + \mathbf{W}_t^\top\mathbf{V}_l\mathbf{V}_l^\top\left(\mathbf{X}_t + \alpha_{t+1}\tilde{\nabla} f(\mathbf{X}_t)\right)\right\|_2\mathbf{I} + \\ & \alpha_{t+1}^2\left\|\left(\tilde{\nabla} f(\mathbf{X}_t)\right)^\top\mathbf{V}_l\mathbf{V}_l^\top\tilde{\nabla} f(\mathbf{X}_t)\right\|_2 + \alpha_{t+1}^2\left\|\mathbf{W}_t^\top\mathbf{V}_l\mathbf{V}_l^\top\mathbf{W}_t\right\|_2 \\ \preccurlyeq \ & \left(1+2\alpha_{t+1}\left\|\tilde{\nabla} f(\mathbf{X}_t)\right\|_2 + 2\alpha_{t+1}(1+\alpha_{t+1}\left\|\tilde{\nabla} f(\mathbf{X}_t)\right\|_2)\|\mathbf{W}_t\|_2 + \alpha_{t+1}^2\left\|\tilde{\nabla} f(\mathbf{X}_t)\right\|_2^2 + \alpha_{t+1}^2\|\mathbf{W}_t\|_2^2\right)\mathbf{I} \\ = \ & \left(1+\alpha_{t+1}\left\|\tilde{\nabla} f(\mathbf{X}_t)\right\|_2 + \alpha_{t+1}\|\mathbf{W}_t\|_2\right)^2\mathbf{I}. \end{aligned}$$

Moreover, since

$$\det(\mathbf{X}_t^\top\mathbf{V}_l\mathbf{V}_l^\top\mathbf{X}_t) = \prod_{i=1}^k \cos^2\theta_i \leq \min\cos^2\theta_i,$$

we have

$$\mathbf{X}_t^\top \mathbf{V}_l \mathbf{V}_l^\top \mathbf{X}_t \succcurlyeq \gamma \mathbf{I}.$$

Then

$$\begin{aligned}
& a_1(\mathbf{X}_t) + \varsigma b_1(\mathbf{W}_t) \\
\succcurlyeq\ & \mathbf{X}_t^\top \mathbf{V}_l \mathbf{V}_l^\top \mathbf{X}_t + \alpha_{t+1} \left(\mathbf{X}_t^\top \mathbf{V}_l \mathbf{V}_l^\top \tilde{\nabla} f(\mathbf{X}_t) + \left(\mathbf{X}_t^\top \mathbf{V}_l \mathbf{V}_l^\top \tilde{\nabla} f(\mathbf{X}_t) \right)^\top \right) + \\
& \varsigma \alpha_{t+1} \left(\left(\mathbf{X}_t + \alpha_{t+1} \tilde{\nabla} f(\mathbf{X}_t) \right)^\top \mathbf{V}_l \mathbf{V}_l^\top \mathbf{W}_t + \mathbf{W}_t^\top \mathbf{V}_l \mathbf{V}_l^\top \left(\mathbf{X}_t + \alpha_{t+1} \tilde{\nabla} f(\mathbf{X}_t) \right) \right) \\
\succcurlyeq\ & \gamma \mathbf{I} - \alpha_{t+1} \left\| \mathbf{X}_t^\top \mathbf{V}_l \mathbf{V}_l^\top \tilde{\nabla} f(\mathbf{X}_t) + \left(\mathbf{X}_t^\top \mathbf{V}_l \mathbf{V}_l^\top \tilde{\nabla} f(\mathbf{X}_t) \right)^\top \right\|_2 \mathbf{I} - \\
& \alpha_{t+1} \left\| \left(\mathbf{X}_t + \alpha_{t+1} \tilde{\nabla} f(\mathbf{X}_t) \right)^\top \mathbf{V}_l \mathbf{V}_l^\top \mathbf{W}_t + \mathbf{W}_t^\top \mathbf{V}_l \mathbf{V}_l^\top \left(\mathbf{X}_t + \alpha_{t+1} \tilde{\nabla} f(\mathbf{X}_t) \right) \right\|_2 \mathbf{I} \\
\succcurlyeq\ & \left(\gamma - 2\alpha_{t+1} \left\| \tilde{\nabla} f(\mathbf{X}_t) \right\|_2 - 2\alpha_{t+1} \left(1 + \alpha_{t+1} \left\| \tilde{\nabla} f(\mathbf{X}_t) \right\|_2 \right) \|\mathbf{W}_t\|_2 \right) \mathbf{I} \\
\succ\ & \mathbf{0},
\end{aligned}$$

which shows that

$$a_1(\mathbf{X}_t) \succcurlyeq \left(\gamma - 2\alpha_{t+1} \left\| \tilde{\nabla} f(\mathbf{X}_t) \right\|_2 \right) \mathbf{I} \succ \mathbf{0}$$

and both $a_1(\mathbf{X}_t)$ and $a_1(\mathbf{X}_t) + \varsigma b_1(\mathbf{W}_t)$ are invertible as well. For brevity, let

$$\mathbf{H}_i = a_i(\mathbf{X}_t) + \varsigma b_i(\mathbf{W}_t), \quad i = 1, 2.$$

Then the first-order and second order derivatives of $f(\varsigma) = \frac{\det(\mathbf{H}_1)}{\det(\mathbf{H}_2)}$ can be derived as follows,

$$f'(\varsigma) = \left(\frac{d}{d\varsigma} \det(\mathbf{H}_1) \right) \det^{-1}(\mathbf{H}_2) - \det(\mathbf{H}_1) \det^{-2}(\mathbf{H}_2) \left(\frac{d}{d\varsigma} \det(\mathbf{H}_2) \right),$$

and

$$\begin{aligned}
f''(\varsigma) \ =\ & \left(\frac{d^2}{d\varsigma^2} \det(\mathbf{H}_1) \right) \det^{-1}(\mathbf{H}_2) - 2 \left(\frac{d}{d\varsigma} \det(\mathbf{H}_1) \right) \det^{-2}(\mathbf{H}_2) \left(\frac{d}{d\varsigma} \det(\mathbf{H}_2) \right) \\
& + 2 \det(\mathbf{H}_1) \det^{-3}(\mathbf{H}_2) \left(\frac{d}{d\varsigma} \det(\mathbf{H}_2) \right)^2 - \det(\mathbf{H}_1) \det^{-2}(\mathbf{H}_2) \left(\frac{d^2}{d\varsigma^2} \det(\mathbf{H}_2) \right).
\end{aligned}$$

Note that for an invertible matrix function $\mathbf{F}(x)$ of scalar variable x,

$$\frac{d}{dx} \det(\mathbf{F}(x)) = \det(\mathbf{F}(x)) \mathrm{tr} \left(\mathbf{F}^{-1}(x) \frac{d}{dx} \mathbf{F}(x) \right) \quad \text{and} \quad \frac{d}{dx} \mathbf{F}^{-1}(x) = -\mathbf{F}^{-1}(x) \left(\frac{d}{dx} \mathbf{F}(x) \right) \mathbf{F}^{-1}(x).$$

Hence, we can write

$$\begin{aligned}
\frac{d}{d\varsigma} \det(\mathbf{H}_i) &= \det(\mathbf{H}_i) \,\mathrm{tr}\left(\mathbf{H}_i^{-1} b_i(\mathbf{W}_t) \right), \\
\frac{d^2}{d\varsigma^2} \det(\mathbf{H}_i) &= \det(\mathbf{H}_i) \,\mathrm{tr}^2\left(\mathbf{H}_i^{-1} b_i(\mathbf{W}_t) \right) - \det(\mathbf{H}_i) \,\mathrm{tr}\left(\left(\mathbf{H}_i^{-1} b_i(\mathbf{W}_t) \right)^2 \right),
\end{aligned}$$

and then

$$f'(\varsigma) = \det(\mathbf{H}_1) \det^{-1}(\mathbf{H}_2) \left(\mathrm{tr}\left(\mathbf{H}_1^{-1} b_1(\mathbf{W}_t) \right) - \mathrm{tr}\left(\mathbf{H}_2^{-1} b_2(\mathbf{W}_t) \right) \right),$$

and

$$\begin{aligned}
& f''(\varsigma) \\
= \;& \det\left(\mathbf{H}_1\right)\det{}^{-1}\left(\mathbf{H}_2\right)\Bigg(\mathrm{tr}^2\left(\mathbf{H}_1^{-1}b_1(\mathbf{W}_t)\right) - \mathrm{tr}\left(\left(\mathbf{H}_1^{-1}b_1(\mathbf{W}_t)\right)^2\right) - \\
& 2\mathrm{tr}\left(\mathbf{H}_1^{-1}b_1(\mathbf{W}_t)\right)\mathrm{tr}\left(\mathbf{H}_2^{-1}b_2(\mathbf{W}_t)\right) + \mathrm{tr}^2\left(\mathbf{H}_2^{-1}b_2(\mathbf{W}_t)\right) + \mathrm{tr}\left(\left(\mathbf{H}_2^{-1}b_2(\mathbf{W}_t)\right)^2\right)\Bigg) \\
= \;& \frac{\det\left(\mathbf{H}_1\right)}{\det\left(\mathbf{H}_2\right)}\left(\left(\mathrm{tr}\left(\mathbf{H}_1^{-1}b_1(\mathbf{W}_t)\right) - \mathrm{tr}\left(\mathbf{H}_2^{-1}b_2(\mathbf{W}_t)\right)\right)^2 - \mathrm{tr}\left(\left(\mathbf{H}_1^{-1}b_1(\mathbf{W}_t)\right)^2\right) + \mathrm{tr}\left(\left(\mathbf{H}_2^{-1}b_2(\mathbf{W}_t)\right)^2\right)\right).
\end{aligned}$$

Thus, we get

$$f'(0) = \frac{\det\left(a_1(\mathbf{X}_t)\right)}{\det\left(a_2(\mathbf{X}_t)\right)}\left(\mathrm{tr}\left(a_1^{-1}(\mathbf{X}_t)b_1(\mathbf{W}_t)\right) - \mathrm{tr}\left(a_2^{-1}(\mathbf{X}_t)b_2(\mathbf{W}_t)\right)\right),$$

and

$$\begin{aligned}
& |f''(\varsigma)| \\
\leq \;& \frac{\det\left(\mathbf{H}_1\right)}{\det\left(\mathbf{H}_2\right)}\left(2\mathrm{tr}^2\left(\mathbf{H}_1^{-1}b_1(\mathbf{W}_t)\right) + 2\mathrm{tr}^2\left(\mathbf{H}_2^{-1}b_2(\mathbf{W}_t)\right) + \mathrm{tr}\left(\left(\mathbf{H}_1^{-1}b_1(\mathbf{W}_t)\right)^2\right) + \mathrm{tr}\left(\left(\mathbf{H}_2^{-1}b_2(\mathbf{W}_t)\right)^2\right)\right) \\
\leq \;& (2k+1)\frac{\det\left(\mathbf{H}_1\right)}{\det\left(\mathbf{H}_2\right)}\left(\mathrm{tr}\left(\left(\mathbf{H}_1^{-1}b_1(\mathbf{W}_t)\right)^2\right) + \mathrm{tr}\left(\left(\mathbf{H}_2^{-1}b_2(\mathbf{W}_t)\right)^2\right)\right),
\end{aligned}$$

by the Cauchy-Schwartz inequality. In particular, since $a_1(\mathbf{X}_t) \succ \mathbf{0}$ and $\mathbb{E}\left[b_1(\mathbf{W}_t)|\mathbf{X}_t\right] \succcurlyeq \mathbf{0}$, we have

$$\mathbb{E}\left[\mathrm{tr}\left(a_1^{-1}(\mathbf{X}_t)b_1(\mathbf{W}_t)\right)|\mathbf{X}_t\right] = \mathrm{tr}\left(a_1^{-1}(\mathbf{X}_t)\mathbb{E}\left[b_1(\mathbf{W}_t)|\mathbf{X}_t\right]\right) \geq 0,$$

and thus

$$\mathbb{E}\left[f'(0)|\mathbf{X}_t\right] = \frac{\det\left(a_1(\mathbf{X}_t)\right)}{\det\left(a_2(\mathbf{X}_t)\right)}\left(\mathrm{tr}\left(a_1^{-1}(\mathbf{X}_t)\mathbb{E}\left[b_1(\mathbf{W}_t)|\mathbf{X}_t\right]\right) - \mathrm{tr}\left(a_2^{-1}(\mathbf{X}_t)\mathbb{E}\left[b_2(\mathbf{W}_t)|\mathbf{X}_t\right]\right)\right) \geq 0.$$

In addition, note that

$$\begin{aligned}
\frac{\det\left(\mathbf{H}_1\right)}{\det\left(\mathbf{H}_2\right)} \;&\leq\; \left(\frac{\left(1+\alpha_{t+1}\left\|\tilde{\nabla}f(\mathbf{X}_t)\right\|_2 + \alpha_{t+1}\left\|\mathbf{W}_t\right\|_2\right)^2}{1+\alpha_{t+1}^2\beta_t - 2\alpha_{t+1}\left(1+\alpha_{t+1}\left\|\tilde{\nabla}f(\mathbf{X}_t)\right\|_2\right)\left\|\mathbf{W}_t\right\|_2}\right)^k \\
&\leq\; \left(\frac{\left(1+\alpha_{t+1}\left\|\tilde{\nabla}f(\mathbf{X}_t)\right\|_2 + \alpha_{t+1}\beta_t^{1/2}\right)^2}{1+\alpha_{t+1}^2\beta_t - 2\alpha_{t+1}\beta_t^{1/2}\left(1+\alpha_{t+1}\left\|\tilde{\nabla}f(\mathbf{X}_t)\right\|_2\right)}\right)^k,
\end{aligned}$$

and

$$\begin{aligned}
& \mathrm{tr}\left(\left(\mathbf{H}_i^{-1}b_i(\mathbf{W}_t)\right)^2\right) \\
\leq \;& \left\|\mathbf{H}_i^{-1}b_i(\mathbf{W}_t)\right\|_F^2 \leq \left\|\mathbf{H}_i^{-1}\right\|_2^2\left\|b_i(\mathbf{W}_t)\right\|_F^2 = \frac{\left\|b_i(\mathbf{W}_t)\right\|_F^2}{\lambda_{min}^2(\mathbf{H}_i)} \\
\leq \;& \left(\frac{2\alpha_{t+1}\left(1+\alpha_{t+1}\left\|\tilde{\nabla}f(\mathbf{X}_t)\right\|_2\right)\left\|\mathbf{W}_t\right\|_F + \frac{1-(-1)^i}{2}\alpha_{t+1}^2\left\|\mathbf{W}_t\right\|_F^2}{\frac{1-(-1)^i}{2}\left(\gamma - 2\alpha_{t+1}\left\|\tilde{\nabla}f(\mathbf{X}_t)\right\|_2\right) + \frac{1+(-1)^i}{2}(1+\alpha_{t+1}^2\beta_t) - 2\alpha_{t+1}\left(1+\alpha_{t+1}\left\|\tilde{\nabla}f(\mathbf{X}_t)\right\|_2\right)\left\|\mathbf{W}_t\right\|_2}\right)^2 \\
\leq \;& \left(\frac{2\alpha_{t+1}\beta_t^{1/2}\left(1+\alpha_{t+1}\left\|\tilde{\nabla}f(\mathbf{X}_t)\right\|_2\right) + \frac{1-(-1)^i}{2}\alpha_{t+1}^2\beta_t}{\frac{1-(-1)^i}{2}\left(\gamma - 2\alpha_{t+1}\left\|\tilde{\nabla}f(\mathbf{X}_t)\right\|_2\right) + \frac{1+(-1)^i}{2}(1+\alpha_{t+1}^2\beta_t) - 2\alpha_{t+1}\beta_t^{1/2}\left(1+\alpha_{t+1}\left\|\tilde{\nabla}f(\mathbf{X}_t)\right\|_2\right)}\right)^2.
\end{aligned}$$

Thus $|f''(\varsigma)|$ can be bounded as follows:

$$\begin{aligned}
& |f''(\varsigma)| \\
\leq\ & (2k+1)\alpha_{t+1}^2\beta_t \left(\frac{\left(1+\alpha_{t+1}\delta_t + \alpha_{t+1}\beta_t^{1/2}\right)^2}{1-2\alpha_{t+1}\beta_t^{1/2}\left(1+\alpha_{t+1}\delta_t\right)} \right)^k \\
& \left(\left(\frac{2+2\alpha_{t+1}\delta_t + \alpha_{t+1}\beta_t^{1/2}}{\gamma - 2\alpha_{t+1}\delta_t - 2\alpha_{t+1}\beta_t^{1/2}\left(1+\alpha_{t+1}\delta_t\right)} \right)^2 + 4\left(\frac{1+\alpha_{t+1}\delta_t}{1-2\alpha_{t+1}\beta_t^{1/2}\left(1+\alpha_{t+1}\delta_t\right)} \right)^2 \right) \\
\triangleq\ & 2\alpha_{t+1}^2\beta_t\xi_t,
\end{aligned}$$

where we have used that $\left\|\tilde{\nabla} f(\mathbf{X}_t)\right\|_2 \leq \left\|\tilde{\nabla} f(\mathbf{X}_t)\right\|_F \leq \delta_t$. It is easy to see that ξ_t is a monotonically increasing function with respect to each input variable in their given ranges.

Finally, we could write

$$\begin{aligned}
& \mathbb{E}\left[\frac{\det\left(a_1(\mathbf{X}_t) + b_1(\mathbf{W}_t)\right)}{\det\left(a_2(\mathbf{X}_t) + b_2(\mathbf{W}_t)\right)} \Big| \mathbf{X}_t \right] \\
=\ & \mathbb{E}\left[f(1)|\mathbf{X}_t\right] = f(0) + \mathbb{E}\left[f'(0)|\mathbf{X}_t\right] + \frac{1}{2}\mathbb{E}\left[f''(\varsigma) \Big| \mathbf{X}_t \right], \quad \varsigma \in [0,1] \\
\geq\ & f(0) + \mathbb{E}\left[f'(0)|\mathbf{X}_t\right] - \frac{1}{2}\mathbb{E}\left[\max_{\varsigma\in[0,1]} |f''(\varsigma)| \Big| \mathbf{X}_t \right] \\
\geq\ & f(0) - \alpha_{t+1}^2\beta_t\xi_t = \frac{\det\left(a_1(\mathbf{X}_t)\right)}{\det\left(a_2(\mathbf{X}_t)\right)} - \alpha_{t+1}^2\beta_t\xi_t.
\end{aligned}$$

□

Lemma B.5. If $\det(\mathbf{X}_t^\top \mathbf{V}_l \mathbf{V}_l^\top \mathbf{X}_t) > \gamma$ and $2\alpha_{t+1}\delta_t < \gamma$, $0 < \gamma < 1$, then

$$\det(a_1(\mathbf{X}_t)) \geq \det(\mathbf{X}_t^\top \mathbf{V}_l \mathbf{V}_l^\top \mathbf{X}_t) + 2\alpha_{t+1}\det(\mathbf{X}_t^\top \mathbf{V}_l \mathbf{V}_l^\top \mathbf{X}_t)E(\mathbf{X}_t) - \eta_t\alpha_{t+1}^2\delta_t^2 \triangleq \varrho_t$$

and

$$\det(a_2(\mathbf{X}_t)) \leq 1 + \zeta_t\alpha_{t+1}^2(k^{1/2}\beta_t + \delta_t^2),$$

where

$$\begin{aligned}
\eta_t &= 2(k+1)(1+\alpha_{t+1}\delta_t)^{2k}\left(\frac{1+\alpha_{t+1}\delta_t}{\gamma - 2\alpha_{t+1}\delta_t}\right)^2, \\
\zeta_t &= k(1+\alpha_{t+1}^2\beta_t + \alpha_{t+1}^2\delta_t^2)^k, \\
E(\mathbf{X}_t) &= \mathrm{tr}(\mathbf{Q}_{l,t}^\top \mathbf{\Sigma}_l \mathbf{Q}_{l,t}) - \mathrm{tr}(\mathbf{X}_t^\top \mathbf{A} \mathbf{X}_t),
\end{aligned}$$

with $\mathbf{\Sigma}_l = \mathrm{diag}(\lambda_1, \cdots, \lambda_l)$ and $\mathbf{Q}_{l,t}$ from the thin SVD:

$$\mathbf{X}_t^\top \mathbf{V}_l = \mathbf{P}_{l,t}\mathbf{\Lambda}_{l,t}\mathbf{Q}_{l,t}^\top,$$

i.e., $\mathbf{Q}_{l,t} \in \mathrm{St}(l,k)$.

Proof. Define functions

$$\begin{aligned}
h_1(\varsigma) &= \det\left(a_1(\mathbf{X}_t;\varsigma)\right) \triangleq \det\left(\left(\mathbf{X}_t + \varsigma\alpha_{t+1}\tilde{\nabla} f(\mathbf{X}_t)\right)^\top \mathbf{V}_l\mathbf{V}_l^\top \left(\mathbf{X}_t + \varsigma\alpha_{t+1}\tilde{\nabla} f(\mathbf{X}_t)\right) \right), \\
h_2(\varsigma) &= \det\left(a_2(\mathbf{X}_t;\varsigma)\right) \triangleq \det\left(\left(\mathbf{X}_t + \varsigma^{1/2}\alpha_{t+1}\tilde{\nabla} f(\mathbf{X}_t)\right)^\top \left(\mathbf{X}_t + \varsigma^{1/2}\alpha_{t+1}\tilde{\nabla} f(\mathbf{X}_t)\right) + \varsigma\alpha_{t+1}^2\beta_t\mathbf{I} \right) \\
&= \det\left(\mathbf{I} + \varsigma\alpha_{t+1}^2\left(\beta_t\mathbf{I} + \tilde{\nabla} f(\mathbf{X}_t)^\top\tilde{\nabla} f(\mathbf{X}_t)\right) \right),
\end{aligned}$$

where $\varsigma \in [0,1]$. From the proof of the preceding lemma, we have that for $i = 1, 2$,

$$\begin{aligned} h_i'(\varsigma) &= \det\left(a_i(\mathbf{X}_t;\varsigma)\right) \operatorname{tr}\left(a_i^{-1}(\mathbf{X}_t;\varsigma)\frac{d\, a_i(\mathbf{X}_t;\varsigma)}{d\varsigma}\right) \\ h_i''(\varsigma) &= \det\left(a_i(\mathbf{X}_t;\varsigma)\right)\left(\operatorname{tr}^2\left(a_i^{-1}(\mathbf{X}_t;\varsigma)\frac{d\, a_i(\mathbf{X}_t;\varsigma)}{d\varsigma}\right) - \operatorname{tr}\left(\left(a_i^{-1}(\mathbf{X}_t;\varsigma)\frac{d\, a_i(\mathbf{X}_t;\varsigma)}{d\varsigma}\right)^2\right)\right) \\ &\quad + \det\left(a_i(\mathbf{X}_t;\varsigma)\right) \operatorname{tr}\left(a_i^{-1}(\mathbf{X}_t;\varsigma)\frac{d^2\, a_i(\mathbf{X}_t;\varsigma)}{d\varsigma^2}\right) \end{aligned}$$

and

$$\begin{aligned} |h_i''(\varsigma)| &\le (k+1)\det\left(a_i(\mathbf{X}_t;\varsigma)\right) \operatorname{tr}\left(\left(a_i^{-1}(\mathbf{X}_t;\varsigma)\frac{d\, a_i(\mathbf{X}_t;\varsigma)}{d\varsigma}\right)^2\right) \\ &\le (k+1)\det\left(a_i(\mathbf{X}_t;\varsigma)\right)\left(\frac{\left\|\frac{d\, a_i(\mathbf{X}_t;\varsigma)}{d\varsigma}\right\|_F^2 + \left\|\frac{d^2\, a_i(\mathbf{X}_t;\varsigma)}{d\varsigma^2}\right\|_F \lambda_{min}\left(a_i(\mathbf{X}_t;\varsigma)\right)}{\lambda_{min}^2\left(a_i(\mathbf{X}_t;\varsigma)\right)}\right). \end{aligned}$$

Note that

$$\begin{aligned} \frac{d\, a_1(\mathbf{X}_t;\varsigma)}{d\varsigma} &= \alpha_{t+1}\left(\tilde\nabla f(\mathbf{X}_t)\right)^\top \mathbf{V}_l\mathbf{V}_l^\top\left(\mathbf{X}_t + \varsigma\alpha_{t+1}\tilde\nabla f(\mathbf{X}_t)\right) + \\ &\quad \alpha_{t+1}\left(\mathbf{X}_t + \varsigma\alpha_{t+1}\tilde\nabla f(\mathbf{X}_t)\right)^\top \mathbf{V}_l\mathbf{V}_l^\top\tilde\nabla f(\mathbf{X}_t), \\ \frac{d^2\, a_1(\mathbf{X}_t;\varsigma)}{d\varsigma^2} &= 2\alpha_{t+1}\left(\tilde\nabla f(\mathbf{X}_t)\right)^\top \mathbf{V}_l\mathbf{V}_l^\top\left(\tilde\nabla f(\mathbf{X}_t)\right), \\ \frac{d\, a_2(\mathbf{X}_t;\varsigma)}{d\varsigma} &= \alpha_{t+1}^2\left(\beta_t\mathbf{I} + \tilde\nabla f(\mathbf{X}_t)^\top\tilde\nabla f(\mathbf{X}_t)\right), \end{aligned}$$

and

$$\begin{aligned} \mathbf{0} \prec (\gamma - 2\alpha_{t+1}\delta_t)\,\mathbf{I} \preccurlyeq \quad & a_1(\mathbf{X}_t;\varsigma) \preccurlyeq (1+\alpha_{t+1}\delta_t)^2\,\mathbf{I}, \\ \mathbf{0} \prec \mathbf{I} \preccurlyeq \quad & a_2(\mathbf{X}_t;\varsigma) \preccurlyeq \left(1 + \alpha_{t+1}^2\beta_t + \alpha_{t+1}^2\delta_t^2\right)\mathbf{I}. \end{aligned}$$

Thus, we get

$$\begin{aligned} h_1'(0) &= 2\alpha_{t+1}\det\left(\mathbf{X}_t^\top\mathbf{V}_l\mathbf{V}_l^\top\mathbf{X}_t\right)\operatorname{tr}\left(\left(\mathbf{X}_t^\top\mathbf{V}_l\mathbf{V}_l^\top\mathbf{X}_t\right)^{-1}\mathbf{X}_t^\top\mathbf{V}_l\mathbf{V}_l^\top\tilde\nabla f(\mathbf{X}_t)\right) \\ &= 2\alpha_{t+1}\det\left(\mathbf{X}_t^\top\mathbf{V}_l\mathbf{V}_l^\top\mathbf{X}_t\right)\left(\operatorname{tr}\left(\left(\mathbf{X}_t^\top\mathbf{V}_l\mathbf{V}_l^\top\mathbf{X}_t\right)^{-1}\mathbf{X}_t^\top\mathbf{V}_l\mathbf{V}_l^\top\mathbf{A}\mathbf{X}_t\right) - \operatorname{tr}\left(\mathbf{X}_t^\top\mathbf{A}\mathbf{X}_t\right)\right) \\ &= 2\alpha_{t+1}\det\left(\mathbf{X}_t^\top\mathbf{V}_l\mathbf{V}_l^\top\mathbf{X}_t\right)\left(\operatorname{tr}\left(\mathbf{\Sigma}_l\mathbf{V}_l^\top\mathbf{X}_t\left(\mathbf{X}_t^\top\mathbf{V}_l\mathbf{V}_l^\top\mathbf{X}_t\right)^{-1}\mathbf{X}_t^\top\mathbf{V}_l\right) - \operatorname{tr}\left(\mathbf{X}_t^\top\mathbf{A}\mathbf{X}_t\right)\right). \end{aligned}$$

Let $\mathbf{X}_t^\top\mathbf{V}_l = \mathbf{P}_{l,t}\mathbf{\Lambda}_{l,t}\mathbf{Q}_{l,t}^\top \in \mathbf{R}^{k\times l}$ be its thin SVD. Then

$$\mathbf{V}_l^\top\mathbf{X}_t\left(\mathbf{X}_t^\top\mathbf{V}_l\mathbf{V}_l^\top\mathbf{X}_t\right)^{-1}\mathbf{X}_t^\top\mathbf{V}_l = \mathbf{Q}_{l,t}\mathbf{Q}_{l,t}^\top.$$

Thus,

$$h_1'(0) = 2\alpha_{t+1}\det\left(\mathbf{X}_t^\top\mathbf{V}_l\mathbf{V}_l^\top\mathbf{X}_t\right)\left(\operatorname{tr}\left(\mathbf{Q}_{l,t}^\top\Sigma_l\mathbf{Q}_{l,t}\right) - \operatorname{tr}\left(\mathbf{X}_t^\top\mathbf{A}\mathbf{X}_t\right)\right).$$

In addition, we have

$$\begin{aligned} 0 \le h_2'(\varsigma) &= \det\left(a_2(\mathbf{X}_t;\varsigma)\right)\operatorname{tr}\left(a_2^{-1}(\mathbf{X}_t;\varsigma)\frac{d\, a_2(\mathbf{X}_t;\varsigma)}{d\varsigma}\right) \\ &\le k\det\left(a_2(\mathbf{X}_t;\varsigma)\right)\frac{\left\|\frac{d\, a_2(\mathbf{X}_t;\varsigma)}{d\varsigma}\right\|_F}{\lambda_{min}\left(a_2(\mathbf{X}_t;\varsigma)\right)} \\ &\le k\alpha_{t+1}^2\left(1+\alpha_{t+1}^2\beta_t + \alpha_{t+1}^2\delta_t^2\right)^k\left(k^{1/2}\beta_t + \delta_t^2\right). \end{aligned}$$

and

$$|h_1''(\varsigma)| \leq 8(k+1)\alpha_{t+1}^2\delta_t^2\left(1+\alpha_{t+1}\delta_t\right)^{2k}\left(\frac{1+\alpha_{t+1}\delta_t}{\gamma-2\alpha_{t+1}\delta_t}\right)^2.$$

We now can arrive at

$$\begin{aligned}
\det\left(a_1(\mathbf{X}_t)\right) &= h_1(1) \\
&= h_1(0)+h_1'(0)+\frac{1}{2}h_1''(\varsigma), \quad \varsigma\in[0,1] \\
&\geq h_1(0)+h_1'(0)-\frac{1}{2}\max_{\varsigma\in[0,1]}|h_1''(\varsigma)| \\
&\geq \det\left(\mathbf{X}_t^\top\mathbf{V}_l\mathbf{V}_l^\top\mathbf{X}_t\right)+2\alpha_{t+1}\det\left(\mathbf{X}_t^\top\mathbf{V}_l\mathbf{V}_l^\top\mathbf{X}_t\right)E(\mathbf{X}_t) \\
&\quad -4(k+1)\alpha_{t+1}^2\delta_t^2\left(1+\alpha_{t+1}\delta_t\right)^{2k}\left(\frac{1+\alpha_{t+1}\delta_t}{\gamma-2\alpha_{t+1}\delta_t}\right)^2,
\end{aligned}$$

and

$$\begin{aligned}
\det\left(a_2(\mathbf{X}_t)\right) &= h_2(1) \\
&\leq h_1(0)+\max_{\varsigma\in[0,1]}|h_1'(\varsigma)| \\
&\leq 1+k\alpha_{t+1}^2\left(1+\alpha_{t+1}^2\beta_t+\alpha_{t+1}^2\delta_t^2\right)^k\left(k^{1/2}\beta_t+\delta_t^2\right).
\end{aligned}$$

Let

$$\begin{aligned}
\eta_t &= 4(k+1)\left(1+\alpha_{t+1}\delta_t\right)^{2k}\left(\frac{1+\alpha_{t+1}\delta_t}{\gamma-2\alpha_{t+1}\delta_t}\right)^2, \\
\zeta_t &= k\left(1+\alpha_{t+1}^2\beta_t+\alpha_{t+1}^2\delta_t^2\right)^k.
\end{aligned}$$

Then we can write

$$\begin{aligned}
\det\left(a_1(\mathbf{X}_t)\right) &\geq \det\left(\mathbf{X}_t^\top\mathbf{V}_l\mathbf{V}_l^\top\mathbf{X}_t\right)+2\alpha_{t+1}\det\left(\mathbf{X}_t^\top\mathbf{V}_l\mathbf{V}_l^\top\mathbf{X}_t\right)E(\mathbf{X}_t)-\eta_t\alpha_{t+1}^2\delta_t^2 \\
&\triangleq \varrho_t, \\
\det\left(a_2(\mathbf{X}_t)\right) &\leq 1+\zeta_t\alpha_{t+1}^2\left(k^{1/2}\beta_t+\delta_t^2\right).
\end{aligned}$$

□

Lemma B.6. Bounded difference of potential functions:

$$|\Psi(\mathbf{X}_{t+1},\mathbf{V}_l)-\Psi(\mathbf{X}_t,\mathbf{V}_l)| \leq \omega_t\alpha_{t+1},$$

where

$$\omega_t = \frac{2k\beta_t^{1/2}(1+\alpha_{t+1}\delta_t)+\alpha_{t+1}\beta_t}{\gamma-2\alpha_{t+1}\delta_t}+2k\beta_t^{1/2}(1+\alpha_{t+1}\delta_t)+\alpha_{t+1}\beta_t\xi_t.$$

Proof. From the proof of Lemma 3.8, we have that

$$\Psi(\mathbf{X}_{t+1},\mathbf{V}_l) = f(0)+f'(0)+\frac{1}{2}f''(\varsigma), \quad \varsigma\in[0,1]$$

where

$$\begin{aligned}
f(0) &= \Psi(\mathbf{X}_t,\mathbf{V}_l), \\
f'(0) &= \Psi(\mathbf{X}_t,\mathbf{V}_l)\left(\mathrm{tr}\left(a_1^{-1}(\mathbf{X}_t)b_1(\mathbf{W}_t)\right)-\mathrm{tr}\left(a_2^{-1}(\mathbf{X}_t)b_2(\mathbf{W}_t)\right)\right).
\end{aligned}$$

and $|f''(\varsigma)| \le 2\alpha_{t+1}^2\beta_t\xi_t$. To bound the difference, we need the following

$$\begin{aligned}
\mathbf{0} \prec (\gamma - 2\alpha_{t+1}\delta_t)\,\mathbf{I} \preccurlyeq \quad & a_1(\mathbf{X}_t) & \preccurlyeq (1+\alpha_{t+1}\delta_t)^2\,\mathbf{I},\\
\mathbf{0} \prec \mathbf{I} \preccurlyeq \quad & a_2(\mathbf{X}_t) = \mathbf{Y}_t^\top\mathbf{Y}_t + \alpha_{t+1}^2\mathbf{W}_t^\top\mathbf{W}_t & \preccurlyeq \left(1+\alpha_{t+1}^2\beta_t + \alpha_{t+1}^2\delta_t^2\right)\mathbf{I},
\end{aligned}$$

and

$$\begin{aligned}
\|b_1(\mathbf{W}_t)\|_F &\le 2\alpha_{t+1}\|\mathrm{sym}(\mathbf{Y}_t^\top\mathbf{V}_l\mathbf{V}_l^\top\mathbf{W}_t)\|_F + \alpha_{t+1}^2\|\mathbf{W}_t^\top\mathbf{V}_l\mathbf{V}_l^\top\mathbf{W}_t\|_F\\
&\le 2\alpha_{t+1}\beta_t^{1/2}(1+\alpha_{t+1}\delta_t) + \alpha_{t+1}^2\beta_t,\\
\|b_2(\mathbf{W}_t)\|_F &\le 2\alpha_{t+1}\|\mathrm{sym}(\mathbf{Y}_t^\top\mathbf{W}_t)\|_F\\
&\le 2\alpha_{t+1}\beta_t^{1/2}(1+\alpha_{t+1}\delta_t).
\end{aligned}$$

Thus, we get

$$\begin{aligned}
& |\Psi(\mathbf{X}_{t+1},\mathbf{V}_l) - \Psi(\mathbf{X}_t,\mathbf{V}_l)|\\
\le\ & |f'(0)| + \frac{1}{2}|f''(\varsigma)|\\
\le\ & \frac{k\|b_1(\mathbf{W}_t)\|_F}{\lambda_{min}(a_1(\mathbf{X}_t))} + \frac{k\|b_2(\mathbf{W}_t)\|_F}{\lambda_{min}(a_2(\mathbf{X}_t))} + \alpha_{t+1}^2\beta_t\xi_t\\
\le\ & \frac{2k\alpha_{t+1}\beta_t^{1/2}(1+\alpha_{t+1}\delta_t) + \alpha_{t+1}^2\beta_t}{\gamma - 2\alpha_{t+1}\delta_t} + 2k\alpha_{t+1}\beta_t^{1/2}(1+\alpha_{t+1}\delta_t) + \alpha_{t+1}^2\beta_t\xi_t\\
=\ & \alpha_{t+1}\left(\frac{2k\beta_t^{1/2}(1+\alpha_{t+1}\delta_t) + \alpha_{t+1}\beta_t}{\gamma - 2\alpha_{t+1}\delta_t} + 2k\beta_t^{1/2}(1+\alpha_{t+1}\delta_t) + \alpha_{t+1}\beta_t\xi_t\right).
\end{aligned}$$

□

Lemma B.7. If $\det(\mathbf{X}_t^\top\mathbf{V}_l\mathbf{V}_l^\top\mathbf{X}_t) > \gamma$, $0 < \gamma < 1$ and

$$0 < \alpha_{t+1} \le \min\left\{\frac{\gamma}{8k^{1/2}\left(\|\mathbf{A}\|_2 + (4+\gamma)\beta^{1/2}\right)}, \frac{\gamma^{1/2}}{2k^{1/2}\eta^{1/2}\|\mathbf{A}\|_2}\right\},$$

then

$$\begin{aligned}
&2\alpha_{t+1}\delta_t + 2\alpha_{t+1}\beta_t^{1/2}(1+\alpha_{t+1}\delta_t) < \gamma/2, && 0 < \eta_t \le \eta \triangleq 2(k+1)(1+\gamma/4)^{2(k+1)}(\gamma/2)^{-2},\\
&0 < \zeta_t \le \zeta \triangleq k(1+(\gamma/4)^2)^k, && 0 < \xi_t \le \xi \triangleq (2k+1)(2+\gamma/2)^{2(k+1)}(\gamma/2)^{-2},\\
&0 \le \varrho_t \le \varrho \triangleq 1+\gamma^{3/2}, && 0 < \omega_t \le \omega \triangleq 7k^{3/2}\beta^{1/2}(1+\frac{\gamma}{4})(\xi + (\frac{\gamma}{4})^{-1}).
\end{aligned}$$

Proof. First let $\alpha_{t+1} \le \frac{\gamma}{8k^{1/2}\|\mathbf{A}\|_2}$. Then

$$\begin{aligned}
2\alpha_{t+1}\delta_t &= 4\alpha_{t+1}k^{1/2}\|\mathbf{A}\|_2\Psi(\mathbf{X}_t,\mathbf{V}_k)\\
&\le 4\alpha_{t+1}k^{1/2}\|\mathbf{A}\|_2\\
&\le \frac{\gamma}{2},
\end{aligned}$$

and

$$\begin{aligned}
& 2\alpha_{t+1}\delta_t + 2\alpha_{t+1}\beta_t^{1/2}(1+\alpha_{t+1}\delta_t)\\
\le\ & 4\alpha_{t+1}k^{1/2}\|\mathbf{A}\|_2 + 2\alpha_{t+1}\sqrt{48k\beta}(1+\frac{\gamma}{4})\\
\le\ & 4k^{1/2}\left(\|\mathbf{A}\|_2 + (4+\gamma)\beta^{1/2}\right)\alpha_{t+1} \le \frac{\gamma}{2}.
\end{aligned}$$

Thus, it is easy to see that

$$\begin{aligned}\eta_t &= 2(k+1)\left(1+\alpha_{t+1}\delta_t\right)^{2k}\left(\frac{1+\alpha_{t+1}\delta_t}{\gamma-2\alpha_{t+1}\delta_t}\right)^2 \\ &\leq 2(k+1)\left(1+\frac{\gamma}{4}\right)^{2(k+1)}\left(\frac{\gamma}{2}\right)^{-2}.\end{aligned}$$

Since $\alpha_{t+1}\delta_t+\alpha_{t+1}\beta_t^{1/2}\leq\frac{\gamma}{4}$, we have

$$\alpha_{t+1}^2\beta_t+\alpha_{t+1}^2\delta_t^2\leq\left(\alpha_{t+1}\delta_t+\alpha_{t+1}\beta_t^{1/2}\right)^2\leq\left(\frac{\gamma}{4}\right)^2.$$

Hence, we can write

$$\zeta_t=k\left(1+\alpha_{t+1}^2\beta_t+\alpha_{t+1}^2\delta_t^2\right)^k\leq k\left(1+\left(\frac{\gamma}{4}\right)^2\right)^k.$$

For ξ_t we have

$$\begin{aligned}\xi_t &= \frac{2k+1}{2}\left(\frac{\left(1+\alpha_{t+1}\delta_t+\alpha_{t+1}\beta_t^{1/2}\right)^2}{1-2\alpha_{t+1}\beta_t^{1/2}\left(1+\alpha_{t+1}\delta_t\right)}\right)^k \\ &\quad \left(\left(\frac{2+2\alpha_{t+1}\delta_t+\alpha_{t+1}\beta_t^{1/2}}{\gamma-2\alpha_{t+1}\delta_t-2\alpha_{t+1}\beta_t^{1/2}\left(1+\alpha_{t+1}\delta_t\right)}\right)^2+4\left(\frac{1+\alpha_{t+1}\delta_t}{1-2\alpha_{t+1}\beta_t^{1/2}\left(1+\alpha_{t+1}\delta_t\right)}\right)^2\right) \\ &\leq \frac{2k+1}{2}\left(\frac{\left(1+\frac{\gamma}{4}\right)^2}{1-\frac{\gamma}{2}}\right)^k\left(\left(\frac{2+\frac{\gamma}{2}}{\gamma-\frac{\gamma}{2}}\right)^2+4\left(\frac{1+\frac{\gamma}{4}}{1-\frac{\gamma}{2}}\right)^2\right) \\ &\leq \frac{2k+1}{2}\left(\frac{\left(1+\frac{\gamma}{4}\right)^2}{1-\frac{1}{2}}\right)^k\left(\left(\frac{2+\frac{\gamma}{2}}{\gamma-\frac{\gamma}{2}}\right)^2+4\left(\frac{1+\frac{\gamma}{4}}{\gamma-\frac{\gamma}{2}}\right)^2\right) \\ &= \frac{2k+1}{2}\left(2+\frac{\gamma}{2}\right)^{2(k+1)}\left(\frac{\gamma}{2}\right)^{-2}.\end{aligned}$$

For ϱ_t, on one hand,

$$\begin{aligned}\varrho_t &= \det\left(\mathbf{X}_t^\top\mathbf{V}_l\mathbf{V}_l^\top\mathbf{X}_t\right)+2\alpha_{t+1}\det\left(\mathbf{X}_t^\top\mathbf{V}_l\mathbf{V}_l^\top\mathbf{X}_t\right)E(\mathbf{X}_t)-\eta_t\alpha_{t+1}^2\delta_t^2 \\ &\geq \gamma-4k\eta\alpha_{t+1}^2\|\mathbf{A}\|_2^2 \\ &\geq \gamma-4k\eta\|\mathbf{A}\|_2^2\left(\frac{\gamma^{1/2}}{2k^{1/2}\eta^{1/2}\|\mathbf{A}\|_2}\right)^2 \\ &= 0.\end{aligned}$$

On the other hand, since $\alpha_{t+1}^2\|\mathbf{A}\|_2^2\leq\frac{\gamma}{4k\eta}$,

$$\begin{aligned}\varrho_t &\leq \det\left(\mathbf{X}_t^\top\mathbf{V}_l\mathbf{V}_l^\top\mathbf{X}_t\right)+2\alpha_{t+1}\det\left(\mathbf{X}_t^\top\mathbf{V}_l\mathbf{V}_l^\top\mathbf{X}_t\right)E(\mathbf{X}_t) \\ &\leq 1+4k\alpha_{t+1}\|\mathbf{A}\|_2\Psi(\mathbf{X}_t,\mathbf{V}_l)\leq 1+4k\alpha_{t+1}\|\mathbf{A}\|_2 \\ &\leq 1+4k\sqrt{\frac{\gamma}{4k\eta}} \\ &= 1+4k\frac{\frac{\gamma}{2}}{\sqrt{2(k+1)}\left(1+\frac{\gamma}{4}\right)^{k+1}}\sqrt{\frac{\gamma}{4k}} \\ &\leq 1+\gamma^{3/2}.\end{aligned}$$

Last, since $1+\alpha_{t+1}\delta_t+\alpha_{t+1}\beta_t^{1/2}\leq 1+\frac{\gamma}{4}$,

$$\begin{aligned}0<\omega_t &\leq \frac{2k\beta_t^{1/2}(1+\alpha_{t+1}\delta_t)+\alpha_{t+1}\beta_t}{\gamma-2\alpha_{t+1}\delta_t}+2k\beta_t^{1/2}(1+\alpha_{t+1}\delta_t)+\alpha_{t+1}\beta_t\xi_t\\ &\leq k\frac{2\beta_t^{1/2}(1+\alpha_{t+1}\delta_t+\alpha_{t+1}\beta_t^{1/2})}{\gamma-2\alpha_{t+1}\delta_t}+k\beta_t^{1/2}\xi_t(1+\alpha_{t+1}\delta_t+\alpha_{t+1}\beta_t^{1/2})\\ &\leq 7k^{3/2}\beta^{1/2}(1+\frac{\gamma}{4})(\xi+(\frac{\gamma}{4})^{-1}).\end{aligned}$$

□

Lemma B.8. For any $\iota\in(0,1)$, we have $\det(\mathbf{X}_t^\top\mathbf{V}_l\mathbf{V}_l^\top\mathbf{X}_t)>\gamma$ all $t=1,2,\cdots,t_0$ with probability at least $1-\iota$, provided that $\alpha_{t+1}\in(0,\rho)$ and ρ, t_0 satisfy

$$\det(\mathbf{X}_t^\top\mathbf{V}_l\mathbf{V}_l^\top\mathbf{X}_t)>\gamma+t_0\rho^2+\rho^2\sqrt{2t_0\log(1/\iota)}.$$

Proof. Consider the stochastic process $\{\Psi(\mathbf{X}_0,\mathbf{V}_l),\Psi(\mathbf{X}_1,\mathbf{V}_l),\ldots,\Psi(\mathbf{X}_{t_0},\mathbf{V}_l)\}$ and the filtration $\mathcal{F}=\{\mathcal{F}_t\}$ induced by random variables y_t. By Lemmas 4.6 and B.4, we have

$$\begin{aligned}\mathbb{E}[\Psi(\mathbf{X}_{t+1},\mathbf{V}_l)|\mathbf{X}_t] &\leq \Psi(\mathbf{X}_t,\mathbf{V}_l)-2\alpha_{t+1}\det(\mathbf{X}_t^\top\mathbf{V}_l\mathbf{V}_l^\top\mathbf{X}_t)E(\mathbf{X}_t)+\\ &\quad (\eta_t+\varrho_t\zeta_t)\alpha_{t+1}^2\delta_t^2+(k^{1/2}\varrho_t\zeta_t+\xi_t)\alpha_{t+1}^2\beta_t\\ &\leq \Psi(\mathbf{X}_t,\mathbf{V}_l)+\rho^2.\end{aligned}$$

Then we can define

$$\Phi_t=\Psi(\mathbf{X}_t,\mathbf{V}_l)-\rho^2 t$$

for $t=0,1,2,\ldots,t_0$, which clearly has a natural continuation such that

$$|\Phi_t|\leq\Psi(\mathbf{X}_t,\mathbf{V}_l)+\rho^2t_0\leq 1+\rho^2t_0,$$

for any t including $t>t_0$. And

$$\begin{aligned}\mathbb{E}[\Phi_{t+1}|\mathbf{X}_t] &= \mathbb{E}[\Psi(\mathbf{X}_{t+1},\mathbf{V}_l)|\mathbf{X}_t]-\rho^2(t+1)\\ &\leq \Psi(\mathbf{X}_t,\mathbf{V}_l)+\rho^2-\rho^2(t+1)\\ &= \Psi(\mathbf{X}_t,\mathbf{V}_l)-\rho^2t=\Phi_t,\end{aligned}$$

for $t=0,1,\ldots,t_0$, while it is clear that $\mathbb{E}[\Phi_{t+1}|\mathbf{X}_t]=\Phi_t$ for $t>t_0$. Thus, $\{\Phi_t\}$ constitutes a super-martingale. On the other hand, by Lemma B.3., we have

$$\begin{aligned}|\Phi_{t+1}-\Phi_t| &\leq |\Psi(\mathbf{X}_{t+1},\mathbf{V}_l)-\Psi(\mathbf{X}_t,\mathbf{V}_l)|+\rho^2\\ &\leq \omega\rho+\rho^2\\ &= \rho(\omega+\rho).\end{aligned}$$

Thus, we are now able to apply the Azuma-Hoeffding inequality to the super-martingale $\{\Phi_t\}$ with bounded difference at $\rho(\omega+\rho)$, and then have that for any $t>0$ and $r>0$,

$$\begin{aligned}P(\Phi_t-\Phi_0\geq r) &\leq \exp\{-\frac{r^2}{2\sum_{i=1}^t\rho^2(\omega+\rho)^2}\}\\ &\leq \exp\{-\frac{r^2}{2t_0\rho^2(\omega+\rho)^2}\}\\ &= \iota.\end{aligned}$$

Hence, $r=\rho(\omega+\rho)\sqrt{2t_0\log(1/\iota)}$ for any $\iota\in(0,1)$. That is, with probability at least $1-\iota$, we have $\Phi_t-\Phi_0<r$ and then

$$\begin{aligned}\Psi(\mathbf{X}_t,\mathbf{V}_l) &< r+\rho^2t+\Phi_0\\ &\leq \rho(\omega+\rho)\sqrt{2t_0\log(1/\iota)}+\rho^2t_0+\Phi_0,\end{aligned}$$

for all $t = 1, \dots, t_0$. Therefore, for any $\iota \in (0,1)$ and all $t = 1, \dots, t_0$, if ρ and t_0 are chosen such that

$$\rho^2\sqrt{2t_0 \log(1/\iota)} + \rho^2 t_0 + \Phi_0 < 1 - \gamma,$$

then we have $\Psi(\mathbf{X}_t, \mathbf{V}_l) < 1 - \gamma$, namely $\det(\mathbf{X}_t^\top \mathbf{V}_l \mathbf{V}_l^\top \mathbf{X}_t) > \gamma$, with probability at least $1 - \iota$.

□

Remark When $\Psi(\mathbf{X}_0, \mathbf{V}_l) < 1 - \gamma$, there exists ρ such that $\rho^2\sqrt{2t_0 \log(1/\iota)} + \rho^2 t_0 + \Phi_0 < 1 - \gamma$ holds.

Part C. Proofs of Main Lemmas

Lemma 4.5. If $\Psi(\mathbf{X}_t, \mathbf{V}_l) < 1 - \gamma$, $0 < \alpha_{t+1} < \rho$, and $0 < \gamma < 1$, we have

$$\mathbb{E}[\Psi(\mathbf{X}_{t+1}, \mathbf{V}_l)|\mathbf{X}_t] \le \Psi(\mathbf{X}_t, \mathbf{V}_l) - 2\alpha_{t+1}(1 - \Psi(\mathbf{X}_t, \mathbf{V}_l))E(\mathbf{X}_t) + \alpha_{t+1}^2 \eta \Psi(\mathbf{X}_t, \mathbf{V}_k) + \alpha_{t+1}^2 \xi \beta_t,$$

where

$$E(\mathbf{X}_t) = \begin{cases} \operatorname{tr}(\mathbf{Q}_{l,t}^\top \mathbf{\Sigma}_l \mathbf{Q}_{l,t}) - \operatorname{tr}(\mathbf{X}_t^\top \mathbf{A} \mathbf{X}_t), & l \ge k \\ \operatorname{tr}(\mathbf{\Sigma}_l) - \operatorname{tr}(\mathbf{Q}_{l,t}^\top \mathbf{X}_t^\top \mathbf{A} \mathbf{X}_t \mathbf{Q}_{l,t}), & l < k \end{cases},$$

$$\mathbf{X}_t^\top \mathbf{V}_l = \begin{cases} \mathbf{P}_{l,t} \mathbf{\Lambda}_{l,t} \mathbf{Q}_{l,t}^\top, & l \ge k \\ \mathbf{Q}_{l,t} \mathbf{\Lambda}_{l,t} \mathbf{P}_{l,t}^\top, & l < k \end{cases}$$

represents the thin SVD of $\mathbf{X}_t^\top \mathbf{V}_l$, and

$$\beta_t = \begin{cases} 0, & \text{Solver 1} \\ 1, & \text{Solver 2} \\ \Psi(\mathbf{X}_t, \mathbf{V}_k) + \Psi(\tilde{\mathbf{X}}_{s-1}, \mathbf{V}_k), & \text{Solver 3} \end{cases}.$$

Proof. We only consider the case that $l \ge k$, as the case of $l < k$ is similar. When $\varrho_t \ge 0$, by Lemma B.5 we have

$$\begin{aligned}
& \frac{\det(a_1(\mathbf{X}_t))}{\det(a_2(\mathbf{X}_t))} \\
\ge\ & \frac{\varrho_t}{1 + \zeta_t \alpha_{t+1}^2 \left(k^{1/2}\beta_t + \delta_t^2\right)} \\
=\ & \frac{\left(1 - \zeta_t \alpha_{t+1}^2 \left(k^{1/2}\beta_t + \delta_t^2\right)\right) \varrho_t}{1 - \left(\zeta_t \alpha_{t+1}^2 \left(k^{1/2}\beta_t + \delta_t^2\right)\right)^2} \\
\ge\ & \left(1 - \zeta_t \alpha_{t+1}^2 \left(k^{1/2}\beta_t + \delta_t^2\right)\right) \varrho_t \\
=\ & \varrho_t - \zeta_t \alpha_{t+1}^2 \left(k^{1/2}\beta_t + \delta_t^2\right) \varrho_t \\
\ge\ & \det(\mathbf{X}_t^\top \mathbf{V}_l \mathbf{V}_l^\top \mathbf{X}_t) + 2\alpha_{t+1} \det(\mathbf{X}_t^\top \mathbf{V}_l \mathbf{V}_l^\top \mathbf{X}_t) E(\mathbf{X}_t) - \eta_t \alpha_{t+1}^2 \delta_t^2 - \zeta_t \alpha_{t+1}^2 \left(k^{1/2}\beta_t + \delta_t^2\right) \varrho_t \\
=\ & \det(\mathbf{X}_t^\top \mathbf{V}_l \mathbf{V}_l^\top \mathbf{X}_t) + 2\alpha_{t+1} \det(\mathbf{X}_t^\top \mathbf{V}_l \mathbf{V}_l^\top \mathbf{X}_t) E(\mathbf{X}_t) - (\eta_t + \zeta_t \varrho_t)\, \alpha_{t+1}^2 \delta_t^2 - k^{1/2} \zeta_t \varrho_t \alpha_{t+1}^2 \beta_t.
\end{aligned}$$

We then can get

$$\begin{aligned}
\mathbb{E}\left[\Psi(\mathbf{X}_{t+1}, \mathbf{V}_l) \middle| \mathbf{X}_t\right] \le\ & \Psi(\mathbf{X}_t, \mathbf{V}_l) - 2\alpha_{t+1} \det(\mathbf{X}_t^\top \mathbf{V}_l \mathbf{V}_l^\top \mathbf{X}_t) E(\mathbf{X}_t) + \\
& (\eta_t + \zeta_t \varrho_t)\, \alpha_{t+1}^2 \delta_t^2 + \left(k^{1/2} \zeta_t \varrho_t + \xi_t\right) \alpha_{t+1}^2 \beta_t \\
\le\ & \Psi(\mathbf{X}_t, \mathbf{V}_l) - 2\alpha_{t+1}(1 - \Psi(\mathbf{X}_t, \mathbf{V}_l)) E(\mathbf{X}_t) + \\
& \alpha_{t+1}^2 \eta \Psi(\mathbf{X}_t, \mathbf{V}_k) + \alpha_{t+1}^2 \xi \beta_t,
\end{aligned}$$

where some missing constants are absorbed into the constants η and ξ.

□

Lemma 4.6.

1) $E(\mathbf{X}) \le (\lambda_1 - \lambda_n)\Theta(\mathbf{X}, \mathbf{V}_l) \le k(\lambda_1 - \lambda_n)\Psi(\mathbf{X}, \mathbf{V}_l)$,
2) $E(\mathbf{X}) \ge \Delta_l \Theta(\mathbf{X}, \mathbf{V}_l) \ge \Delta_l \Psi(\mathbf{X}, \mathbf{V}_l)$,

Proof. We consider $l \ge k$ and omit subscript t here. Note that $\mathbf{A} = \mathbf{V}_l \Sigma_l \mathbf{V}_l^\top + \mathbf{V}_l^\perp \Sigma_l^\perp \left(\mathbf{V}_l^\perp\right)^\top$. Then

$$\begin{aligned}
E(\mathbf{X}) &= \operatorname{tr}(\mathbf{Q}_{l,t}^\top \mathbf{\Sigma}_l \mathbf{Q}_{l,t}) - \operatorname{tr}(\mathbf{X}^\top \mathbf{A}\mathbf{X}) \\
&= \operatorname{tr}(\mathbf{Q}_{l,t}^\top \mathbf{\Sigma}_l \mathbf{Q}_{l,t}) - \operatorname{tr}(\mathbf{X}^\top \mathbf{V}_l \Sigma_l \mathbf{V}_l^\top \mathbf{X}) - \operatorname{tr}(\mathbf{X}^\top \mathbf{V}_l^\perp \Sigma_l^\perp \left(\mathbf{V}_l^\perp\right)^\top \mathbf{X}) \\
&= \operatorname{tr}(\mathbf{\Sigma}_l \mathbf{Q}_{l,t} \mathbf{Q}_{l,t}^\top) - \operatorname{tr}(\Sigma_l \mathbf{Q}_{l,t} \mathbf{\Lambda}_{l,t}^2 \mathbf{Q}_{l,t}^\top) - \operatorname{tr}(\mathbf{X}^\top \mathbf{V}_l^\perp \Sigma_l^\perp \left(\mathbf{V}_l^\perp\right)^\top \mathbf{X}) \\
&= \operatorname{tr}(\Sigma_l \mathbf{Q}_{l,t} (\mathbf{I} - \mathbf{\Lambda}_{l,t}^2) \mathbf{Q}_{l,t}^\top) - \operatorname{tr}(\mathbf{X}^\top \mathbf{V}_l^\perp \Sigma_l^\perp \left(\mathbf{V}_l^\perp\right)^\top \mathbf{X}) \\
&\le \lambda_1 \operatorname{tr}(\mathbf{Q}_{l,t} (\mathbf{I} - \mathbf{\Lambda}_{l,t}^2) \mathbf{Q}_{l,t}^\top) - \lambda_n \operatorname{tr}(\mathbf{X}^\top \mathbf{V}_l^\perp \left(\mathbf{V}_l^\perp\right)^\top \mathbf{X}) \\
&= \lambda_1 \operatorname{tr}(\mathbf{I} - \mathbf{\Lambda}_{l,t}^2) - \lambda_n \operatorname{tr}(\mathbf{X}^\top (\mathbf{I} - \mathbf{V}_l \mathbf{V}_l^\top) \mathbf{X}) \\
&= (\lambda_1 - \lambda_n)(k - \|\mathbf{X}^\top \mathbf{V}_l\|_F^2) = (\lambda_1 - \lambda_n)\Theta(\mathbf{X}, \mathbf{V}_l) \\
&\le k(\lambda_1 - \lambda_n)\Psi(\mathbf{X}, \mathbf{V}_l),
\end{aligned}$$

and

$$\begin{aligned}
E(\mathbf{X}) &= \operatorname{tr}(\Sigma_l \mathbf{Q}_{l,t} (\mathbf{I} - \mathbf{\Lambda}_{l,t}^2) \mathbf{Q}_{l,t}^\top) - \operatorname{tr}(\mathbf{X}^\top \mathbf{V}_l^\perp \Sigma_l^\perp \left(\mathbf{V}_l^\perp\right)^\top \mathbf{X}) \\
&\ge \lambda_l \operatorname{tr}(\mathbf{Q}_{l,t} (\mathbf{I} - \mathbf{\Lambda}_{l,t}^2) \mathbf{Q}_{l,t}^\top) - \lambda_{l+1} \operatorname{tr}(\mathbf{X}^\top \mathbf{V}_l^\perp \left(\mathbf{V}_l^\perp\right)^\top \mathbf{X}) \\
&= \Delta_l \Theta(\mathbf{X}, \mathbf{V}_l) \ge \Delta_l \Psi(\mathbf{X}, \mathbf{V}_l).
\end{aligned}$$

□

Lemma 4.7. For a uniformly sampled point $\mathbf{Y} \in \operatorname{Grass}(n, l)$ with $l < \frac{n+1}{2}$ and $0 < \gamma < 1$, we have that

$$\det{}^2(\mathbf{Y}^\top \mathbf{V}_l) > \gamma$$

with probability at least

$$1 - p_l(\gamma) = \frac{\Gamma(\frac{l+1}{2})\Gamma(\frac{n-l+1}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{n+1}{2})} (\sin(\cos^{-1}(\gamma^{\frac{1}{2l}})))^{l(n-l)} {}_2F_1(\frac{n-l}{2}, \frac{1}{2}, \frac{n+1}{2}; \mathbf{I}_{l\times l} \sin^2(\cos^{-1}(\gamma^{\frac{1}{2l}}))),$$

where ${}_2F_1$ is the Gaussian hypergeometric function of matrix argument.

Proof. Let $\theta_{\max} = \max_i \theta_i(\mathbf{Y}, \mathbf{V}_l)$. First, we have

$$\begin{aligned}
P\left(\det{}^2(\mathbf{Y}^\top \mathbf{V}_l) > \gamma\right) &= P\left(\prod_{i=1}^{l} \cos^2 \theta_i \ge \gamma\right) \\
&\ge P\left(\left(\min_i \cos \theta_i\right)^{2l} \ge \gamma\right) \\
&= P\left(\cos^{2l} \theta_{\max} \ge \gamma\right) \\
&= P\left(\theta_{\max} \le \cos^{-1} \gamma^{\frac{1}{2l}}\right).
\end{aligned}$$

According to Absil et al. (2006),

$$\begin{aligned}
& P\left(\theta_{\max} \le \cos^{-1} \gamma^{\frac{1}{2l}}\right) \\
= \; & \frac{\Gamma(\frac{l+1}{2})\Gamma(\frac{n-l+1}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{n+1}{2})} \left(\sin\left(\cos^{-1}\left(\gamma^{\frac{1}{2l}}\right)\right)\right)^{l(n-l)} {}_2F_1\left(\frac{n-l}{2}, \frac{1}{2}, \frac{n+1}{2}; \mathbf{I}_{l\times l} \sin^2\left(\cos^{-1}\left(\gamma^{\frac{1}{2l}}\right)\right)\right) \\
\triangleq \; & 1 - p_l(\gamma).
\end{aligned}$$

□

Remark 1 There is no doubt that $p_l(\gamma) \to 0$ as $\gamma \to 0$, because $\det^2(\mathbf{Y}^\top \mathbf{V}_l) \in [0,1]$. In fact, $p_l(\gamma)$ is the integral of the probability density function of the largest principal angle $\max_i \theta_i$ between two random points in $\text{Grass}(n,l)$, on the interval $[\cos^{-1}(\gamma^{\frac{1}{2l}}), \frac{\pi}{2}] \subset [0, \frac{\pi}{2}]$. It says that although in the high yet finite dimensional regime two random points in $\text{Grass}(n,l)$ are nearly orthogonal to each other with high probability , the probability of attaining the orthogonality of high precision is quite small, especially that the probability $P(\max_i \theta_i(\mathbf{Y}, \mathbf{V}_l) = \frac{\pi}{2})$ is zero because $\{\mathbf{X} \in \text{Grass}(n,l) : \max_i \theta_i(\mathbf{Y}, \mathbf{V}_l) = \frac{\pi}{2}\}$ is a zero-measure set. See Absil et al. (2006) for more details as well as a pictorial view of the density function in a low-dimensional setting.

Remark 2 What we need from this lemma is to get an initial $\mathbf{X}_0 \in \text{Grass}(n,k)$ such that $\Psi(\mathbf{X}_0, \mathbf{V}_l) < 1 - \gamma$ with high probability. There is no problem when $l = k$ as it can be directly given by the lemma. However, if $l \leq k$, i.e., $l = k_{\min}$ or $l = k_{\max}$ in our case, this would be different. Theoretically, there are four cases:

1. When $\Delta_k > 0$, it is a direct application of the lemma by letting $l = k$ in the lemma.
2. When $\Delta_k = 0$ and $k_{\max} = n$, we only need to show that $\Psi(\mathbf{X}_t, \mathbf{V}_{k_{\min}}) < \epsilon$. Thus, we need $\Psi(\mathbf{X}_0, \mathbf{V}_{k_{\min}}) < 1-\gamma$ with high probability. Fortunately, we can make it as in the above case. The reason is that $\Psi(\mathbf{X}_0, \mathbf{V}_k) < 1 - \gamma$ implies that $\Psi(\mathbf{X}_0, \mathbf{V}_{k_{\min}}) < 1 - \gamma$, and thus the latter's probability will be no less than the former's.
3. When $\Delta_k = 0$ and $k_{\min} = 0$, we only need to show that $\Psi(\mathbf{X}_t, \mathbf{V}_{k_{\max}}) < \epsilon$. We now need $\Psi(\mathbf{X}_0, \mathbf{V}_{k_{\max}}) < 1 - \gamma$ with high probability. To this end, we set $\mathbf{X}_0 = \mathbf{Y}_1$ with $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2)$ sampled uniformly from $\text{Grass}(n, k_{\max})$. This is only the theoretical soundness of the convergence proof. In practice, we may need to choose a l large enough to cover $k_{\max}$, i.e., $k_{\max} < l$, then sample $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2)$ uniformly from $\text{Grass}(n,l)$, and set $\mathbf{X}_0 = \mathbf{Y}_1$.
4. When $\Delta_k = 0$, $0 < k_{\min}$ and $k_{\max} < n$, we need to show that both $\Psi(\mathbf{X}_t, \mathbf{V}_{k_{\min}}) < \epsilon$ and $\Psi(\mathbf{X}_t, \mathbf{V}_{k_{\max}}) < \epsilon$. For two values of l, i.e., $k_{\min}$ and $k_{\max}$, we only need to sample one $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2)$ sampled uniformly from $\text{Grass}(n, k_{\max})$ and then set $\mathbf{X}_0 = \mathbf{Y}_1$, as $\Psi(\mathbf{Y}, \mathbf{V}_{k_{\max}}) < 1 - \gamma$ ensures $\Psi(\mathbf{X}_0, \mathbf{V}_{k_{\min}}) < 1 - \gamma$ and $\Psi(\mathbf{X}_0, \mathbf{V}_{k_{\max}}) < 1 - \gamma$ simultaneously.

References

P.-A. Absil, A. Edelman, and P. Koev. On the largest principal angle between random subspaces. *Linear Algebra Appl.*, 414(1):288–294, 2006.