Supplementary Material

A. Derivation of ΔU_t

We first recall that

$$\boldsymbol{U}_t = \boldsymbol{U}_{t-1}\boldsymbol{E}_t + \boldsymbol{\Delta U}_t, \quad \text{(A1)}$$

where $\boldsymbol{\Delta U}_t$ represents the distinctive new information in $\boldsymbol{U}_t$. Thanks to (10) in the main text, we have

$$\boldsymbol{x}_t = \boldsymbol{y}_t + \boldsymbol{U}_{t-1}\boldsymbol{z}_t. \quad \text{(A2)}$$

The matrix $\boldsymbol{S}_t$ in (7) can be expressed as follows

$$\begin{aligned}\boldsymbol{S}_t &= \boldsymbol{R}_t\boldsymbol{U}_{t-1} = \beta\boldsymbol{R}_{t-1}\boldsymbol{U}_{t-1} + \boldsymbol{x}_t\boldsymbol{z}^\top \\ &= \beta\boldsymbol{R}_{t-1}\underbrace{\begin{bmatrix}\boldsymbol{U}_{t-2} & \boldsymbol{U}_{t-2,\perp}\end{bmatrix}\begin{bmatrix}\boldsymbol{U}_{t-2} & \boldsymbol{U}_{t-2,\perp}\end{bmatrix}^\top}_{=\boldsymbol{I}_n}\boldsymbol{U}_{t-1} + \boldsymbol{x}_t\boldsymbol{z}_t^\top \\ &= \beta\boldsymbol{R}_{t-1}\boldsymbol{U}_{t-2}\boldsymbol{U}_{t-2}^\top\boldsymbol{U}_{t-1} + \beta\boldsymbol{R}_{t-1}\boldsymbol{U}_{t-2,\perp}\boldsymbol{U}_{t-2,\perp}^\top\boldsymbol{U}_{t-1} + \boldsymbol{x}_t\boldsymbol{z}_t^\top,\end{aligned} \quad \text{(A3)}$$

where $\boldsymbol{U}_{t-2,\perp}^\top\boldsymbol{U}_{t-2} = \boldsymbol{0}_{n-r\times r}$ and $\boldsymbol{U}_{t-2}^\top\boldsymbol{U}_{t-2,\perp} = \boldsymbol{0}_{r\times n-r}$ by definition. As the underlying subspace is assumed to be fixed or slowly varying with time, $\boldsymbol{U}_{t-1}$ is nearly orthogonal to the noise subspace of $\boldsymbol{U}_{t-2}$, i.e., $\boldsymbol{U}_{t-2,\perp}^\top\boldsymbol{U}_{t-1} \approxeq \boldsymbol{0}$. Therefore, the second term of (A3) is negligible and can be discarded. In what follows, we indicate that the QR decomposition of $\boldsymbol{S}_t$ in (A3) can be expressed in terms of the augmented and updated terms of the QR decomposition of $\boldsymbol{S}_{t-1}$. Denote by $\boldsymbol{U}_k\boldsymbol{R}_{U,k}$ the QR representation of $\boldsymbol{S}_k$ for $k = 1, 2, \dots, t$. Note that $\boldsymbol{S}_{t-1} = \boldsymbol{R}_{t-1}\boldsymbol{U}_{t-2}$, (A3) is further expressed as follows

$$\begin{aligned}\boldsymbol{S}_t &\approxeq \beta\boldsymbol{R}_{t-1}\boldsymbol{U}_{t-2}\boldsymbol{U}_{t-2}^\top\boldsymbol{U}_{t-1} + \boldsymbol{x}_t\boldsymbol{z}_t^\top \\ &= \beta\underbrace{\boldsymbol{U}_{t-1}\boldsymbol{R}_{U,t-1}}_{\text{QR}(\boldsymbol{S}_{t-1})}\underbrace{\boldsymbol{E}_{t-1}}_{\boldsymbol{U}_{t-2}^\top\boldsymbol{U}_{t-1}} + \boldsymbol{x}_t\boldsymbol{z}_t^\top \\ &= \boldsymbol{U}_{t-1}\big(\beta\boldsymbol{R}_{U,t-1}\boldsymbol{E}_{t-1} + \boldsymbol{z}_t\boldsymbol{z}_t^\top\big) + \boldsymbol{y}_t\boldsymbol{z}_t^\top \\ &= \underbrace{\begin{bmatrix}\boldsymbol{U}_{t-1} & \boldsymbol{y}_t\end{bmatrix}}_{\text{augmented term}}\underbrace{\begin{bmatrix}\beta\boldsymbol{R}_{U,t-1}\boldsymbol{E}_{t-1} + \boldsymbol{z}_t\boldsymbol{z}_t^\top \\ \boldsymbol{z}_t^\top\end{bmatrix}}_{\text{updated term}}.\end{aligned} \quad \text{(A4)}$$

Without loss of generality, we suppose that the Givens method is used to compute the QR decomposition of $\boldsymbol{S}_t$. By using a sequence of Givens rotations, (A4) is recast into the following form

$$\begin{aligned}\boldsymbol{S}_t &\approxeq \Big(\begin{bmatrix}\boldsymbol{U}_{t-1} & \boldsymbol{y}_t\end{bmatrix}\boldsymbol{G}_t^\top\Big)\left(\boldsymbol{G}_t\begin{bmatrix}\beta\boldsymbol{R}_{U,t-1}\boldsymbol{E}_{t-1} + \boldsymbol{z}_t\boldsymbol{z}_t^\top \\ \boldsymbol{z}_t^\top\end{bmatrix}\right) \\ &= \Big(\begin{bmatrix}\boldsymbol{U}_{t-1} & \bar{\boldsymbol{y}}_t\end{bmatrix}\boldsymbol{G}_t^\top\Big)\left(\boldsymbol{G}_t\begin{bmatrix}\beta\boldsymbol{R}_{U,t-1}\boldsymbol{E}_{t-1} + \boldsymbol{z}_t\boldsymbol{z}_t^\top \\ \|\boldsymbol{y}_t\|_2\boldsymbol{z}_t^\top\end{bmatrix}\right),\end{aligned} \quad \text{(A5)}$$

where $\bar{\boldsymbol{y}} = \boldsymbol{y}_t/\|\boldsymbol{y}_t\|_2$ is the normalized vector of $\boldsymbol{y}_t$, and $\boldsymbol{G}_t$ is a $(r+1)\times(r+1)$ orthogonal matrix representing the sequence of Givens rotations. The Givens rotations in $\boldsymbol{G}_t$ should be selected such that the second term of (A5) is transformed into an upper triangular matrix, i.e.,

$$\boldsymbol{G}_t\begin{bmatrix}\beta\boldsymbol{R}_{U,t-1}\boldsymbol{E}_{t-1} + \boldsymbol{z}_t\boldsymbol{z}_t^\top \\ \|\boldsymbol{y}_t\|_2\boldsymbol{z}_t^\top\end{bmatrix} = \begin{bmatrix}\boldsymbol{R}_{U,t} \\ \boldsymbol{0}_{1\times r}\end{bmatrix}, \quad \text{(A6)}$$

to obtain the R-factor $\boldsymbol{R}_{U,t}$ of $\boldsymbol{S}_t$. Now, let $\boldsymbol{u}_t = \begin{bmatrix}\boldsymbol{U}_{t-1} & \bar{\boldsymbol{y}}_t\end{bmatrix}\boldsymbol{g}_t^\top$ where $\boldsymbol{g}_t$ is the last row of the Givens matrix $\boldsymbol{G}_t$. To form the QR representation of $\boldsymbol{S}_t$, we have

$$\begin{aligned}\boldsymbol{S}_t &\overset{\text{QR}}{=} \boldsymbol{U}_t\boldsymbol{R}_{U,t} \\ &= \underbrace{\Big(\begin{bmatrix}\boldsymbol{U}_{t-1} & \bar{\boldsymbol{y}}_t\end{bmatrix}\boldsymbol{G}_t^\top\Big)}_{\begin{bmatrix}\boldsymbol{U}_t & \boldsymbol{u}_t\end{bmatrix}}\underbrace{\left(\boldsymbol{G}_t\begin{bmatrix}\beta\boldsymbol{R}_{U,t-1}\boldsymbol{E}_{t-1} + \boldsymbol{z}_t\boldsymbol{z}_t^\top \\ \|\boldsymbol{y}_t\|_2\boldsymbol{z}_t^\top\end{bmatrix}\right)}_{\begin{bmatrix}\boldsymbol{R}_{U,t} \\ \boldsymbol{0}_{1\times r}\end{bmatrix}}.\end{aligned}$$

This gives rise the following recursion for updating $\boldsymbol{U}_t$ at time t

$$\begin{bmatrix}\boldsymbol{U}_t & \boldsymbol{u}_t\end{bmatrix} = \begin{bmatrix}\boldsymbol{U}_{t-1} & \bar{\boldsymbol{y}}_t\end{bmatrix}\boldsymbol{G}_t^\top. \quad \text{(A7)}$$

As $\bar{\boldsymbol{y}}_t^\top\bar{\boldsymbol{y}}_t = 1$ and $\begin{bmatrix}\boldsymbol{U}_{t-1}^\top \\ \bar{\boldsymbol{y}}_t^\top\end{bmatrix}\begin{bmatrix}\boldsymbol{U}_{t-1} & \bar{\boldsymbol{y}}_t\end{bmatrix} = \boldsymbol{I}$, we can express the rotation matrix $\boldsymbol{G}_t$ as follows

$$\begin{aligned}\boldsymbol{G}_t^\top &= \begin{bmatrix}\boldsymbol{U}_{t-1}^\top \\ \bar{\boldsymbol{y}}_t^\top\end{bmatrix}\begin{bmatrix}\boldsymbol{U}_t & \boldsymbol{u}_t\end{bmatrix} = \begin{bmatrix}\boldsymbol{U}_{t-1}^\top\boldsymbol{U}_t & \boldsymbol{U}_{t-1}^\top\boldsymbol{u}_t \\ \bar{\boldsymbol{y}}_t^\top\boldsymbol{U}_t & \bar{\boldsymbol{y}}_t^\top\boldsymbol{u}_t\end{bmatrix} \\ &= \begin{bmatrix}\boldsymbol{E}_t & \boldsymbol{U}_{t-1}^\top\boldsymbol{u}_t \\ \boldsymbol{h}_t^\top & \bar{\boldsymbol{y}}_t^\top\boldsymbol{u}_t\end{bmatrix},\end{aligned} \quad \text{(A8)}$$

where $\boldsymbol{E}_t = \boldsymbol{U}_{t-1}^\top\boldsymbol{U}_t$ and $\boldsymbol{h}_t = \boldsymbol{U}_t^\top\bar{\boldsymbol{y}}_t$ are defined as in (8) and (11), respectively. By substituting (A8) into (A7), we obtain

$$\begin{bmatrix}\boldsymbol{U}_{t-1} & \bar{\boldsymbol{y}}_t\end{bmatrix}\begin{bmatrix}\boldsymbol{E}_t & \boldsymbol{U}_{t-1}^\top\boldsymbol{u}_t \\ \boldsymbol{h}_t^\top & \bar{\boldsymbol{y}}_t^\top\boldsymbol{u}_t\end{bmatrix} = \begin{bmatrix}\boldsymbol{U}_t & \boldsymbol{u}_t\end{bmatrix}, \quad \text{(A9)}$$

and hence,

$$\boldsymbol{U}_t = \boldsymbol{U}_{t-1}\boldsymbol{E}_t + \bar{\boldsymbol{y}}_t\boldsymbol{h}_t^\top. \quad \text{(A10)}$$

It implies that $\boldsymbol{\Delta U}_t = \bar{\boldsymbol{y}}_t\boldsymbol{h}_t^\top$, according to (A1).

B. Proof of Lemma 1

Because $\boldsymbol{U}_{t,\mathcal{F}}$ is the Q-factor of $\boldsymbol{S}_t$, we obtain $\theta(\boldsymbol{A}, \boldsymbol{U}_{t,\mathcal{F}}) = \theta(\boldsymbol{A}, \boldsymbol{S}_t)$ and hence

$$\tan\theta(\boldsymbol{A}, \boldsymbol{U}_{t,\mathcal{F}}) = \max_{\|\boldsymbol{v}\|_2=1}\left\{f(\boldsymbol{v}) = \frac{\|\boldsymbol{A}_\perp^\top\boldsymbol{S}_t\boldsymbol{v}\|_2}{\|\boldsymbol{A}^\top\boldsymbol{S}_t\boldsymbol{v}\|_2}\right\}. \quad \text{(B1)}$$

For any vector $\boldsymbol{v} \in \mathbb{R}^{r\times 1}$ and $\|\boldsymbol{v}\|_2 = 1$, we can rewrite $f(\boldsymbol{v})$ in (B1) as follows

$$\begin{aligned}f(\boldsymbol{v}) &= \frac{\|\boldsymbol{A}_\perp^\top\boldsymbol{R}_t\boldsymbol{U}_{t-1}\boldsymbol{v}\|_2}{\|\boldsymbol{A}^\top\boldsymbol{R}_t\boldsymbol{U}_{t-1}\boldsymbol{v}\|_2} = \frac{\|\boldsymbol{A}_\perp^\top\big(t(\boldsymbol{C} + \boldsymbol{\Delta C}_t)\big)\boldsymbol{U}_{t-1}\boldsymbol{v}\|_2}{\|\boldsymbol{A}^\top\big(t(\boldsymbol{C} + \boldsymbol{\Delta C}_t)\big)\boldsymbol{U}_{t-1}\boldsymbol{v}\|_2} \\ &= \frac{\|\boldsymbol{A}_\perp^\top\big(\sigma_x^2\boldsymbol{A}\boldsymbol{A}^\top + \sigma_n^2\boldsymbol{I}_N + \boldsymbol{\Delta C}_t\big)\boldsymbol{U}_{t-1}\boldsymbol{v}\|_2}{\|\boldsymbol{A}^\top\big(\sigma_x^2\boldsymbol{A}\boldsymbol{A}^\top + \sigma_n^2\boldsymbol{I}_N + \boldsymbol{\Delta C}_t\big)\boldsymbol{U}_{t-1}\boldsymbol{v}\|_2} \\ &\overset{(i)}{=} \frac{\|\sigma_n^2\boldsymbol{A}_\perp^\top\boldsymbol{U}_{t-1}\boldsymbol{v} + \boldsymbol{A}_\perp^\top\boldsymbol{\Delta C}_t\boldsymbol{U}_{t-1}\boldsymbol{v}\|_2}{\|(\sigma_x^2 + \sigma_n^2)\boldsymbol{A}^\top\boldsymbol{U}_{t-1}\boldsymbol{v} + \boldsymbol{A}^\top\boldsymbol{\Delta C}_t\boldsymbol{U}_{t-1}\boldsymbol{v}\|_2} \\ &\overset{(ii)}{\le} \frac{\sigma_n^2\|\boldsymbol{A}_\perp^\top\boldsymbol{U}_{t-1}\|_2 + \|\boldsymbol{A}_\perp^\top\boldsymbol{\Delta C}_t\boldsymbol{U}_{t-1}\|_2}{(\sigma_x^2 + \sigma_n^2)\|\boldsymbol{A}^\top\boldsymbol{U}_{t-1}\|_2 - \|\boldsymbol{A}^\top\boldsymbol{\Delta C}_t\boldsymbol{U}_{t-1}\|_2} \\ &\overset{(iii)}{\le} \frac{\sigma_n^2\|\boldsymbol{A}_\perp^\top\boldsymbol{U}_{t-1}\|_2 + \|\boldsymbol{\Delta C}_t\|_2}{(\sigma_x^2 + \sigma_n^2)\sqrt{1 - \|\boldsymbol{A}_\perp^\top\boldsymbol{U}_{t-1}\|_2^2} - \|\boldsymbol{\Delta C}_t\|_2}.\end{aligned} \quad \text{(B2)}$$

Here, (i) is due to $\boldsymbol{A}_\perp^\top \boldsymbol{A} = \boldsymbol{0}$ (orthogonal complement); (ii) uses the inequality $\|\boldsymbol{P}\|_2 - \|\boldsymbol{Q}\|_2 \le \|\boldsymbol{P}+\boldsymbol{Q}\|_2 \le \|\boldsymbol{P}\|_2 + \|\boldsymbol{Q}\|_2$, $\forall \boldsymbol{P},\boldsymbol{Q}$ of the same size; and (iii) is derived from the following facts: $\|\boldsymbol{P}\boldsymbol{\Delta C}_t\|_2 \le \|\boldsymbol{P}\|_2 \|\boldsymbol{\Delta C}_t\|_2$, $\|\boldsymbol{A}\|_2 = \|\boldsymbol{A}_\perp\|_2 = \|\boldsymbol{U}_{t-1}\|_2 = 1$, and

$$\lambda_{\min}^2(\boldsymbol{A}^\top \boldsymbol{U}_{t-1}) + \lambda_{\max}^2(\boldsymbol{A}_\perp^\top \boldsymbol{U}_{t-1}) = 1, \tag{B3}$$

where $\lambda_{\max}(\boldsymbol{P})$ and $\lambda_{\min}(\boldsymbol{P})$ represent the largest and smallest singular value of $\boldsymbol{P}$, respectively.

Indeed, the relation (B3) leads to

$$\begin{aligned}\left\|\boldsymbol{A}^\top \boldsymbol{U}_{t-1}\right\|_2 &= \lambda_{\max}(\boldsymbol{A}^\top \boldsymbol{U}_{t-1}) \ge \lambda_{\min}(\boldsymbol{A}^\top \boldsymbol{U}_{t-1}) \\ &= \sqrt{1 - \lambda_{\max}^2\big(\boldsymbol{A}_\perp^\top \boldsymbol{U}_{t-1}\big)} = \sqrt{1 - \|\boldsymbol{A}_\perp^\top \boldsymbol{U}_{t-1}\|_2^2}, \end{aligned} \tag{B4}$$

and thus, (iii) follows.

In parallel, it is well known that $\sin\psi = 1\big/\sqrt{1+\tan^{-2}\psi}$ $\forall \psi \in [0,\pi/2]$ and $h(x) = 1/\sqrt{1+x^{-2}}$ is an increasing function in the domain $(0,\infty)$, i.e., $x_1 \le x_2$ implies $h(x_1) \le h(x_2)$. Accordingly, we obtain

$$\begin{aligned}\left\|\boldsymbol{A}_\perp^\top \boldsymbol{U}_{t,\mathcal{F}}\right\|_2 &\le \frac{1}{\sqrt{1 + \big[\max_{\boldsymbol{v}} f(\boldsymbol{v})\big]^{-2}}} \\ &= \frac{\sigma_n^2 \left\|\boldsymbol{A}_\perp^\top \boldsymbol{U}_{t-1}\right\|_2 + \|\boldsymbol{\Delta C}_t\|_2}{\left(\Big[(\sigma_x^2+\sigma_n^2)\sqrt{1-\|\boldsymbol{A}_\perp^\top \boldsymbol{U}_{t-1}\|^2} - \|\boldsymbol{\Delta C}_t\|_2\Big]^2 + \Big[\sigma_n^2\left\|\boldsymbol{A}_\perp^\top \boldsymbol{U}_{t-1}\right\|_2 + \|\boldsymbol{\Delta C}_t\|_2\Big]^2\right)^{1/2}}. \end{aligned} \tag{B5}$$

It ends the proof.

C. Proof of Lemma 2

We first recast $\left\|\boldsymbol{U}_{t,\perp}^\top \boldsymbol{U}_{t,\mathcal{F}}\right\|_2$ into the following form

$$\begin{aligned}\left\|\boldsymbol{U}_{t,\perp}^\top \boldsymbol{U}_{t,\mathcal{F}}\right\|_2 &= \left\|\boldsymbol{U}_{t,\mathcal{F},\perp}^\top \boldsymbol{U}_t\right\|_2 \\ &= \left\|\boldsymbol{U}_{t,\mathcal{F},\perp}^\top \big(\boldsymbol{U}_t - \boldsymbol{U}_{t,\mathcal{F}}\big)\right\|_2 = \left\|\boldsymbol{U}_{t,\mathcal{F},\perp}^\top \boldsymbol{\Delta U}_t\right\|_2. \end{aligned} \tag{C1}$$

Under the following condition

$$(1+\sqrt{2})\kappa(\boldsymbol{S}_t)\left\|\boldsymbol{S}_t - \hat{\boldsymbol{S}}_t\right\|_F < \left\|\boldsymbol{S}_t\right\|_2, \tag{C2}$$

where $\boldsymbol{\Delta S}_t = \boldsymbol{S}_t - \hat{\boldsymbol{S}}_t$ and $\kappa(\boldsymbol{S}_t) = \|\boldsymbol{S}_t^{\#}\|_2 \|\boldsymbol{S}_t\|_2$, we can bound this distance as follows

$$\begin{aligned}\left\|\boldsymbol{U}_{t,\mathcal{F},\perp}^\top \boldsymbol{\Delta U}_t\right\|_2 &\le \left\|\boldsymbol{U}_{t,\mathcal{F},\perp}^\top \boldsymbol{\Delta U}_t\right\|_F \\ &\overset{(i)}{\le} \frac{\kappa(\boldsymbol{S}_t)\dfrac{\|\boldsymbol{U}_{t,\mathcal{F},\perp}^\top \boldsymbol{\Delta S}_t\|_F}{\|\boldsymbol{S}_t\|_2}}{1 - (1+\sqrt{2})\kappa(\boldsymbol{S}_t)\dfrac{\|\boldsymbol{\Delta S}_t\|_F}{\|\boldsymbol{S}_t\|_2}} \\ &\overset{(ii)}{\le} \frac{\|\boldsymbol{\Delta S}_t\|_F}{\lambda_{\min}(\boldsymbol{S}_t) - (1+\sqrt{2})\|\boldsymbol{\Delta S}_t\|_F}. \end{aligned} \tag{C3}$$

Here, (i) follows immediately the perturbation theory for QR decomposition [1, Theorem 3.1] and (ii) is obtained from the facts that $\|\boldsymbol{U}_{t,\mathcal{F},\perp}\|_2 = 1$, $\|\boldsymbol{P}\boldsymbol{Q}\|_F \le \|\boldsymbol{P}\|_2 \|\boldsymbol{Q}\|_F$, and $\|\boldsymbol{P}^{\#}\|_2 = \lambda_{\min}^{-1}(\boldsymbol{P})$ $\forall \boldsymbol{P},\boldsymbol{Q}$ of suitable sizes.

We also know that there always exists two coefficient matrices $\boldsymbol{H}_t \in \mathbb{R}^{r\times r}$ and $\boldsymbol{K}_t \in \mathbb{R}^{(n-r)\times r}$ satisfying $\boldsymbol{U}_{t-1} = \boldsymbol{A}\boldsymbol{H}_t + \boldsymbol{A}_\perp \boldsymbol{K}_t$ (i.e., projection of $\boldsymbol{U}_{t-1}$ onto the subspace $\boldsymbol{A}$) and

$$\lambda_{\max}(\boldsymbol{H}_t) = \|\boldsymbol{A}^\top \boldsymbol{U}_{t-1}\|_2, \lambda_{\min}(\boldsymbol{H}_t) = \sqrt{1 - \|\boldsymbol{A}_\perp^\top \boldsymbol{U}_{t-1}\|_2^2}, \tag{C4}$$

$$\lambda_{\max}(\boldsymbol{K}_t) = \|\boldsymbol{A}_\perp^\top \boldsymbol{U}_{t-1}\|_2, \lambda_{\min}(\boldsymbol{K}_t) = \sqrt{1 - \|\boldsymbol{A}^\top \boldsymbol{U}_{t-1}\|_2^2}. \tag{C5}$$

Accordingly, we can express $\boldsymbol{S}_t$ by

$$\begin{aligned}\boldsymbol{S}_t &= \boldsymbol{R}_t \boldsymbol{U}_{t-1} = t(\boldsymbol{C}\boldsymbol{U}_{t-1} + \boldsymbol{\Delta C}_t \boldsymbol{U}_{t-1}) \\ &= t\big(\boldsymbol{A}\boldsymbol{\Sigma}_x \boldsymbol{A}^\top + \sigma_n^2 \boldsymbol{I}_n(\boldsymbol{A}\boldsymbol{H}_t + \boldsymbol{A}_\perp \boldsymbol{K}_t) + \boldsymbol{\Delta C}_t \boldsymbol{U}_{t-1}\big) \\ &= t\big(\boldsymbol{A}(\sigma_x^2 \boldsymbol{I}_r + \sigma_n^2 \boldsymbol{I}_r)\boldsymbol{H}_t + \sigma_n^2 \boldsymbol{A}_\perp \boldsymbol{K}_t + \boldsymbol{\Delta C}_t \boldsymbol{U}_{t-1}\big). \end{aligned} \tag{C6}$$

Thanks to the fact that $\lambda_i(\boldsymbol{P}+\boldsymbol{Q}) \ge \lambda_i(\boldsymbol{P}) - \lambda_{\max}(\boldsymbol{Q})$ $\forall \boldsymbol{P},\boldsymbol{Q}$ of the same size, the lower bound on $\lambda_{\min}(\boldsymbol{S}_t)$ is given by

$$\begin{aligned}\lambda_{\min}(\boldsymbol{S}_t) &\ge t\Big(\lambda_{\min}\big((\sigma_x^2+\sigma_n^2)\boldsymbol{A}\boldsymbol{H}_t\big) - \lambda_{\max}\big(\sigma_n^2 \boldsymbol{A}_\perp \boldsymbol{K}_t\big) - \lambda_{\max}\big(\boldsymbol{\Delta C}_t \boldsymbol{U}_{t-1}\big)\Big) \\ &\ge t\big((\sigma_x^2+\sigma_n^2)\lambda_{\min}(\boldsymbol{H}_t) - \sigma_n^2 \lambda_{\max}(\boldsymbol{K}_t) - \|\boldsymbol{\Delta C}_t\|_2\big) \\ &= t\Big((\sigma_x^2+\sigma_n^2)\sqrt{1 - \|\boldsymbol{A}_\perp^\top \boldsymbol{U}_{t-1}\|_2^2} - \sigma_n^2 \|\boldsymbol{A}_\perp^\top \boldsymbol{U}_{t-1}\|_2 - \|\boldsymbol{\Delta C}_t\|_2\Big), \end{aligned} \tag{C7}$$

In what follows, we derive an upper bound on $\|\boldsymbol{\Delta S}_t\|_F$. For short, let us denote the support of $\boldsymbol{A}$, $\boldsymbol{U}_{t-1}$, and $\boldsymbol{U}_t$ by $\mathcal{T}_A$, $\mathcal{T}_{t-1}$, and $\mathcal{T}_t$, respectively, and $\mathcal{S}_t = \mathcal{T}_A \cup \mathcal{T}_{t-1} \cup \mathcal{T}_t$. Here, we also know that $\boldsymbol{S}_{t,\mathcal{S}_t} = \boldsymbol{R}_{t,\mathcal{S}_t\times\mathcal{S}_t} \boldsymbol{U}_{t-1}$ and $\hat{\boldsymbol{S}}_t = \boldsymbol{S}_{t,\mathcal{T}_t} = \tau(\boldsymbol{S}_{t,\mathcal{S}_t}, k)$. Accordingly, we can bound $\|\boldsymbol{\Delta S}_t\|_F$ as follows

$$\begin{aligned}\left\|\boldsymbol{\Delta S}_t\right\|_F &= \left\|\boldsymbol{S}_{t,\mathcal{S}_t} - \boldsymbol{S}_{t,\mathcal{T}_t}\right\|_F \overset{(i)}{\le} \left\|\boldsymbol{S}_{t,\mathcal{S}_t} - \boldsymbol{S}_{t,\mathcal{T}_A}\right\|_F \\ &= t\left\|\sigma_n^2 \boldsymbol{A}_\perp \boldsymbol{K}_t + \boldsymbol{\Delta C}_t \boldsymbol{U}_{t-1}\right\|_F \\ &\le t\sqrt{r}\left\|\sigma_n^2 \boldsymbol{A}_\perp \boldsymbol{K}_t + \boldsymbol{\Delta C}_t \boldsymbol{U}_{t-1}\right\|_2 \le t\sqrt{r}\big(\sigma_n^2 \|\boldsymbol{K}_t\|_2 + \|\boldsymbol{\Delta C}_t\|_2\big) \\ &= t\sqrt{r}\big(\sigma_n^2 \|\boldsymbol{A}_\perp^\top \boldsymbol{U}_{t-1}\|_2 + \|\boldsymbol{\Delta C}_t\|_2\big), \end{aligned} \tag{C8}$$

where (i) is due to $|\mathcal{T}_t| \ge |\mathcal{T}_A|$ $\forall t$ (i.e., $|\mathcal{S}_t \setminus \mathcal{T}_t| \le |\mathcal{S}_t \setminus \mathcal{T}_A|$), thanks the thresholding operator $\tau(\cdot)$ with $n\omega_{\text{sparse}} \le k \le \sqrt{n/\log n}$.

In parallel, we can rewrite the sufficient and necessary condition (C2) as

$$(1+\sqrt{2})\left\|\boldsymbol{S}_t^{\#}\right\|_2 \left\|\boldsymbol{\Delta S}_t\right\|_F \le 1. \tag{C9}$$

Since $\|\boldsymbol{S}_t^{\#}\|_2 = \lambda_{\min}^{-1}(\boldsymbol{S}_t)$, substituting the (C7) for $\|\boldsymbol{S}_t^{\#}\|_2$ and (C8) for $\|\boldsymbol{\Delta S}_t\|_F$ results in

$$\frac{\sigma_n^2 \left\|\boldsymbol{A}_\perp^\top \boldsymbol{U}_{t-1}\right\|_2 + \|\boldsymbol{\Delta C}_t\|_2}{(\sigma_x^2+\sigma_n^2)\sqrt{1 - \|\boldsymbol{A}_\perp^\top \boldsymbol{U}_{t-1}\|^2}} \le \frac{\sqrt{2}-1}{\sqrt{r}-1+\sqrt{2}}. \tag{C10}$$

Under the condition (C10), the upper bound on $\|\boldsymbol{U}_{t,\perp}^\top \boldsymbol{U}_{t,\mathcal{F}}\|_2$ is

$$\begin{aligned}&\left\|\boldsymbol{U}_{t,\perp}^\top \boldsymbol{U}_{t,\mathcal{F}}\right\|_2 \\ &\le \frac{\sqrt{r}\big(\sigma_n^2 \left\|\boldsymbol{A}_\perp^\top \boldsymbol{U}_{t-1}\right\|_2 + \|\boldsymbol{\Delta C}_t\|_2\big)}{\Big((\sigma_x^2+\sigma_n^2)\sqrt{1 - \|\boldsymbol{A}_\perp^\top \boldsymbol{U}_{t-1}\|_2^2} - \sigma_n^2\left\|\boldsymbol{A}_\perp^\top \boldsymbol{U}_{t-1}\right\|_2 - \|\boldsymbol{\Delta C}_t\|_2 - \sqrt{r}\big(1+\sqrt{2}\big)\big(\sigma_n^2\left\|\boldsymbol{A}_\perp^\top \boldsymbol{U}_{t-1}\right\|_2 + \|\boldsymbol{\Delta C}_t\|_2\big)\Big)} \\ &= \frac{\sqrt{r}\big(\sigma_n^2 \left\|\boldsymbol{A}_\perp^\top \boldsymbol{U}_{t-1}\right\|_2 + \|\boldsymbol{\Delta C}_t\|_2\big)}{\Big((\sigma_x^2+\sigma_n^2)\sqrt{1 - \|\boldsymbol{A}_\perp^\top \boldsymbol{U}_{t-1}\|_2^2} - \big(1+\sqrt{r}(1+\sqrt{2})\big) \times \big(\sigma_n^2\left\|\boldsymbol{A}_\perp^\top \boldsymbol{U}_{t-1}\right\|_2 + \|\boldsymbol{\Delta C}_t\|_2\big)\Big)}, \end{aligned} \tag{C11}$$

thanks to (C3). It ends the proof.

D. Proof of Lemma 3

We begin the proof with the following proposition:

Proposition 1. *Given two sets of random variable vectors* $\{\boldsymbol{a}_i\}_{i=1}^N$ *and* $\{\boldsymbol{b}_i\}_{i=1}^N$ *where* $\boldsymbol{a}_i \overset{i.i.d.}{\sim} \mathcal{N}(\boldsymbol{0}, \sigma_a^2 \boldsymbol{I}_n)$, $\boldsymbol{b}_i \overset{i.i.d.}{\sim} \mathcal{N}(\boldsymbol{0}, \sigma_b^2 \boldsymbol{I}_m)$, *and*

$\boldsymbol{a}_i$ is independent of $\boldsymbol{b}_j, \forall i,j$. The following inequality holds with a probability at least $1-\delta$:

$$\left\|\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{a}_i\boldsymbol{b}_i^\top\right\|_2 \le C\sigma_a\sigma_b\sqrt{\log(2/\delta)\frac{\max\{n,m\}}{N}}. \tag{D1}$$

where $0<\delta\ll 1$ and $C>0$ is a universal positive number.

Proof. Its proof follows immediately Lemma 15 in [2]. □

Since $\boldsymbol{x}_i = \boldsymbol{A}\boldsymbol{w}_i + \boldsymbol{n}_i$, we always have

$$\begin{aligned}\|\boldsymbol{\Delta C}_t\|_2 \le\ & \|\boldsymbol{A}\|_2^2\left\|\frac{1}{tW}\sum_{i=1}^{tW}\boldsymbol{w}_i\boldsymbol{w}_i^\top - \sigma_x^2\boldsymbol{I}_r\right\|_2 + \\ &+ 2\|\boldsymbol{A}\|_2\left\|\frac{1}{tW}\sum_{i=1}^{tW}\boldsymbol{w}_i\boldsymbol{n}_i^\top\right\|_2 + \left\|\frac{1}{tW}\sum_{i=1}^{tW}\boldsymbol{n}_i\boldsymbol{n}_i^\top - \sigma_n^2\boldsymbol{I}_n\right\|_2,\end{aligned} \tag{D2}$$

please see (D3) for a detailed derivation of (D2). Accordingly, with a probability at least $1-\delta$ ($0<\delta\ll 1$), three components in the right hand side of (D2) are respectively bounded by

$$\left\|\frac{1}{tW}\sum_{i=1}^{tW}\boldsymbol{w}_i\boldsymbol{w}_i^\top - \sigma_w^2\boldsymbol{I}_r\right\|_2 \le C_1\sqrt{\log(2/\delta)}\sigma_w^2\sqrt{\frac{r}{tW}}, \tag{D4}$$

$$\left\|\frac{1}{tW}\sum_{i=1}^{tW}\boldsymbol{n}_i\boldsymbol{n}_i^\top - \sigma_n^2\boldsymbol{I}_n\right\|_2 \le C_2\sqrt{\log(2/\delta)}\sigma_n^2\sqrt{\frac{n}{tW}}, \tag{D5}$$

$$\left\|\frac{1}{tW}\sum_{i=1}^{tW}\boldsymbol{w}_i\boldsymbol{n}_i^\top\right\|_2 \le C_3\sqrt{\log(2/\delta)}\sigma_w\sigma_n\sqrt{\frac{n}{tW}}, \tag{D6}$$

where C_1, C_2, C_3 are universal positive parameters, thanks to Proposition 1 and [3, Proposition 2.1]. As a result, we obtain

$$\|\boldsymbol{\Delta C}_t\|_2 \le c_\delta\left(\sigma_w^2\sqrt{\frac{r}{tW}} + \left(2\sigma_n\sigma_w + \sigma_n^2\right)\sqrt{\frac{n}{tW}}\right), \tag{D7}$$

where $c_\delta = \max\{C_1, C_2, C_3\}\sqrt{\log(2/\delta)}$. It ends the proof.

E. Proof of Lemma 4

We first use proof by induction to prove $d_t \le \omega_0 = \max\{d_0, \epsilon\}$. Particularly, we already have the base case of $d_0 \le \omega_0$. In the induction step, we suppose $d_{t-1} \le \omega_0$ and then prove $d_t \le \omega_0$ still holds. After that, we indicate that $d_t \le \epsilon$ is achievable when the two conditions (17) and (18) are met.

Thanks to Lemma 3, when t satisfies (17), i.e.,

$$t \ge \frac{C\log(2/\delta)r^2}{W\epsilon^2\rho^2}\left(\sqrt{r} + \left(\frac{\sigma_n^2}{\sigma_x^2} + 2\frac{\sigma_n}{\sigma_x}\right)\sqrt{n}\right)^2, \tag{E1}$$

we obtain $\|\boldsymbol{\Delta C}_t\|_2 \le r^{-1}\rho\sigma_x^2\epsilon$ with $0<\rho\le r$. In what follows, two case studies $d_{t-1}\ge\epsilon$ and $d_{t-1}\le\epsilon$ are investigated.

Case 1: When $d_{t-1}\ge\epsilon$, i.e., $\|\boldsymbol{\Delta C}_t\|_2 \le r^{-1}\rho\sigma_x^2 d_{t-1}$.

We can rewrite $\|\boldsymbol{A}_\perp^\top\boldsymbol{U}_{t,\mathcal{F}}\|_2$ as follows

$$\begin{aligned}\|\boldsymbol{A}_\perp^\top\boldsymbol{U}_{t,\mathcal{F}}\|_2 &\le \frac{(\sigma_n^2 + r^{-1}\rho\sigma_x^2)d_{t-1}}{\Big(\big[(\sigma_n^2+\sigma_x^2)\sqrt{1-d_{t-1}^2} - r^{-1}\rho\sigma_x^2 d_{t-1}\big]^2 + (\sigma_n^2+\sigma_x^2\rho/r)^2 d_{t-1}^2\Big)^{1/2}} \\ &\overset{(i)}{\le} \frac{(\sigma_n^2 + r^{-1}\rho\sigma_x^2)d_{t-1}}{\Big(\big[(\sigma_n^2+\sigma_x^2)\sqrt{1-\omega_0^2} - r^{-1}\rho\sigma_x^2\omega_0\big]^2 + (\sigma_n^2+r^{-1}\rho\sigma_x^2)^2\omega_0^2\Big)^{1/2}} \\ &\overset{(ii)}{\le} \frac{(\sigma_n^2 + r^{-1}\rho\sigma_x^2)d_{t-1}}{\Big((1+\gamma^2r^2)\sigma_n^4 + (1-\rho\gamma)^2\sigma_x^4 + 2(1-\rho\gamma+\gamma^2r^2)\sigma_x^2\sigma_n^2\Big)^{1/2}\sqrt{1-\omega_0^2}}.\end{aligned} \tag{E2}$$

Here, (i) is obtained from the fact that $g(x) = \big((a\sqrt{1-x^2}-bx)^2 + cx^2\big)^{-1/2}$ is an increasing function in the range $[0, \sqrt{2}/2]$ where a, b, and c are defined therein[1] and (ii) is simple due to the fact that there always exists a small parameter $\gamma > 0$ such that $\rho\gamma < 1$ and $\omega_0 \le \gamma r\sqrt{1-\omega_0^2}$.

In the similar way, we obtain the following upper bound on $\|\boldsymbol{U}_{t,\perp}^\top\boldsymbol{U}_{t,\mathcal{F}}\|_2$:

$$\begin{aligned}\|\boldsymbol{U}_{t,\perp}^\top\boldsymbol{U}_{t,\mathcal{F}}\|_2 &\le \frac{\sqrt{r}\big(\sigma_n^2 + r^{-1}\rho\sigma_x^2\big)d_{t-1}}{\big(\sigma_x^2+\sigma_n^2\big)\sqrt{1-d_t^2} - \big(1+\sqrt{r}(1+\sqrt{2})\big)\times\big(\sigma_n^2+r^{-1}\rho\sigma_x^2\big)d_{t-1}} \\ &\overset{(i)}{\le} \frac{\sqrt{r}(\sigma_n^2 + r^{-1}\rho\sigma_x^2)d_{t-1}}{\big(\sigma_x^2+\sigma_n^2\big)\sqrt{1-\omega_0^2} - \big(1+\sqrt{r}(1+\sqrt{2})\big)\big(\sigma_n^2+r^{-1}\rho\sigma_x^2\big)\omega_0} \\ &\overset{(ii)}{\le} \frac{\sqrt{r}(\sigma_n^2 + r^{-1}\rho\sigma_x^2)}{(\sigma_x^2+\sigma_n^2)(1-\varrho)\sqrt{1-\omega_0^2}}d_{t-1},\end{aligned} \tag{E4}$$

where $\varrho = \gamma\big(1+\sqrt{r}(1+\sqrt{2})(r\sigma_n^2+\rho\sigma_x^2)\big)(\sigma_x^2+\sigma_n^2)^{-1}$. Specifically, (i) is due to the increasing property of $z(x) = (a\sqrt{1-x^2}-bx)^{-1}$, and (ii) thanks to $\omega_0 \le \gamma r\sqrt{1-\omega_0^2}$.

Thanks to (E2) and (E4), we obtain

$$d_t \le \|\boldsymbol{A}_\perp^\top\boldsymbol{U}_{t,\mathcal{F}}\|_2 + \|\boldsymbol{U}_{t,\perp}^\top\boldsymbol{U}_{t,\mathcal{F}}\|_2 \le \frac{r\sigma_n^2+\rho\sigma_x^2}{r\xi\sqrt{1-\omega_0^2}}d_{t-1}, \tag{E5}$$

where

$$\begin{aligned}\xi = 0.5\max\Big\{&\Big((1+\gamma^2r^2)\sigma_n^4 + (1-\rho\gamma)^2\sigma_x^4 \\ &+ 2(1-\rho\gamma+\gamma^2r^2)\sigma_x^2\sigma_n^2\Big)^{1/2}, (\sigma_x^2+\sigma_n^2)(1-\varrho)/\sqrt{r}\Big\}.\end{aligned} \tag{E6}$$

Note that in order to utilize the two bounds (E2) and (E4), the condition (C10) must be satisfied which is equivalent to

$$\frac{(\sigma_n^2 + r^{-1}\rho\sigma_x^2)\omega_0}{(\sigma_x^2+\sigma_n^2)\sqrt{1-\omega_0^2}} \le \frac{\sqrt{2}-1}{\sqrt{r}-1+\sqrt{2}}. \tag{E7}$$

Accordingly, we obtain $\omega_0 \le \left(\frac{\alpha(r,\rho)}{1-\alpha(r,\rho)}\right)^{1/2}$ where

$$\alpha(r,\rho) = \frac{(3-2\sqrt{2})(\sigma_x^2+\sigma_n^2)^2}{\big(r+2\sqrt{r}(\sqrt{2}-1)+3-2\sqrt{2}\big)\big(\sigma_n^2+r^{-1}\rho\sigma_x^2\big)^2}. \tag{E8}$$

In parallel, $\alpha(r,\rho) \ge \frac{3-2\sqrt{2}}{r+2\sqrt{r}(\sqrt{2}-1)+3-2\sqrt{2}}$ for every $0<\rho\le r$. Thus, we obtain $\omega_0 \le \left(\frac{3-2\sqrt{2}}{r+2\sqrt{r}(\sqrt{2}-1)}\right)^{1/2}$ which is exactly the condition (18) in Theorem 1. Moreover, there are various options of $p \in (0, r]$ satisfying $\rho\sigma_x^2 < r\xi\sqrt{1-\omega_0^2} - r\sigma_n^2$, e.g., when the value of ρ is very close to zero. In such cases, d_t will decrease in each time t, i.e., $d_t \le d_{t-1} \le \omega_0$.

Case 2: When $d_{t-1}\le\epsilon$, applying the same arguments in Case 1, we also obtain $d_t \le \frac{r\sigma_n^2+\rho\sigma_x^2}{r\xi\sqrt{1-\omega_0^2}}\epsilon \le \epsilon \le \omega_0$.

[1]Writing $x = \sin y$, the domain of y is $[0, \pi/4]$. Here, we can recast $g(x)$ into $g(y) = \big((a\cos y - b\sin y)^2 + c\sin^2 y\big)^{-1/2}$. The derivative $g'(y)$ is given by

$$\begin{aligned}g'(y) = 0.5\big((a\cos y - b\sin y)^2 + c\sin^2 y\big)^{-3/2}\times \\ \times\big((a^2-b^2-c)\sin(2y) + ab\cos(2y)\big).\end{aligned} \tag{E3}$$

Since $a^2-b^2>c$ by their definition, $g'(y)>0$ $\forall y\in[0,\pi/4]$ and hence $g'(x) = g'(y)dy/dx = g'(y)/\sqrt{1-x^2} > 0$ $\forall x\in[0,\sqrt{2}/2]$. Accordingly, $d_{t-1}\le\omega_0\le\sqrt{2}/2$ implies $g(d_{t-1}) \le g(\omega_0)$ which (i) then follows.

$$\left\|\boldsymbol{\Delta C}_t\right\|_2 = \left\|\frac{1}{tW}\sum_{i=1}^{tW}\boldsymbol{x}_i\boldsymbol{x}_i^\top - \boldsymbol{C}\right\|_2 = \left\|\frac{1}{tW}\sum_{i=1}^{tW}\Big(\boldsymbol{A}\boldsymbol{w}_i\boldsymbol{w}_i^\top\boldsymbol{A}^\top + \boldsymbol{n}_i\boldsymbol{n}_i^\top + \boldsymbol{A}\boldsymbol{w}_i\boldsymbol{n}_i^\top + \boldsymbol{n}_i\boldsymbol{w}_i^\top\boldsymbol{A}^\top\Big) - \sigma_x^2\boldsymbol{A}\boldsymbol{A}^\top - \sigma_n^2\boldsymbol{I}_n\right\|_2$$
$$\leq \left\|\boldsymbol{A}\bigg(\frac{1}{tW}\sum_{i=1}^{tW}\boldsymbol{w}_i\boldsymbol{w}_i^\top - \sigma_x^2\boldsymbol{I}_r\bigg)\boldsymbol{A}^\top\right\|_2 + \left\|\frac{1}{tW}\sum_{i=1}^{tW}\boldsymbol{n}_i\boldsymbol{n}_i^\top - \sigma_n^2\boldsymbol{I}_N\right\|_2 + 2\left\|\boldsymbol{A}\bigg(\frac{1}{tW}\sum_{i=1}^{tW}\boldsymbol{w}_i\boldsymbol{n}_i^\top\bigg)\right\|_2$$
$$\leq \left\|\boldsymbol{A}\right\|_2^2\left\|\frac{1}{tW}\sum_{i=1}^{tW}\boldsymbol{w}_i\boldsymbol{w}_i^\top - \sigma_x^2\boldsymbol{I}_r\right\|_2 + \left\|\frac{1}{tW}\sum_{i=1}^{tW}\boldsymbol{n}_i\boldsymbol{n}_i^\top - \sigma_n^2\boldsymbol{I}_n\right\|_2 + 2\left\|\boldsymbol{A}\right\|_2\left\|\frac{1}{tW}\sum_{i=1}^{tW}\boldsymbol{w}_i\boldsymbol{n}_i^\top\right\|_2, \tag{D3}$$

thanks to the inequality $\|\boldsymbol{PQ}\|_2 \leq \|\boldsymbol{P}\|_2\|\boldsymbol{Q}\|_2$ for all $\boldsymbol{P}$ and $\boldsymbol{Q}$ of suitable sizes.

To sum up, if the two conditions (17) and (18) are satisfied, then $d_t \leq \max\{d_{t-1}, \epsilon\} = \omega_0$. As a result, the statement $d_t \leq \epsilon$ holds if and only if

$$\left(\frac{r\sigma_n^2 + \rho\sigma_x^2}{r\xi\sqrt{1-\omega_0^2}}\right)^{tW}\omega_0 \leq \epsilon. \tag{E9}$$

Specifically, (E9) is equivalent to

$$t \geq \frac{\log(\epsilon/\omega_0)}{W\Big(\log(r\sigma_n^2 + \rho\sigma_x^2) - \log(r\xi\sqrt{1-\omega_0^2})\Big)}. \tag{E10}$$

which is lower than the bound (17). Therefore, we can conclude that $d_t \leq \epsilon$ holds and it ends the proof.

F. Decomposition of U_t

Let $\boldsymbol{U}_{t,\mathcal{F}} = \boldsymbol{D}_t$ and $\boldsymbol{U}_{t,\mathcal{F},\perp} = \boldsymbol{D}_{t,\perp}$ for easy of representation. Now, our objective is to demonstrate the existence of two matrices $\boldsymbol{W}_1 \in \mathbb{R}^{r\times r}$ and $\boldsymbol{W}_2 \in \mathbb{R}^{(n-r)\times r}$ such that

$$\boldsymbol{U}_t = \boldsymbol{D}_t\boldsymbol{W}_1 + \boldsymbol{D}_{t,\perp}\boldsymbol{W}_2.$$

Proof. Given a full-rank matrix $\boldsymbol{P} \in \mathbb{R}^{n\times n}$, we always find a matrix $\boldsymbol{W} \in \mathbb{R}^{n\times r}$ such that

$$\boldsymbol{U}_t = \boldsymbol{PW} \quad \text{or} \quad \boldsymbol{u}_t^{(i)} = \boldsymbol{P}\boldsymbol{w}^{(i)}, i = 1, 2, \ldots, r,$$

where $\boldsymbol{u}_t^{(i)}$ and $\boldsymbol{w}^{(i)}$ are the i-th column of $\boldsymbol{U}_t$ and $\boldsymbol{W}$, respectively. It is because $\boldsymbol{w}^{(i)} = \boldsymbol{P}^{-1}\boldsymbol{u}_t^{(i)}$ always exists. Form $\boldsymbol{P} = [\boldsymbol{D}_t \;\; \boldsymbol{D}_{t,\perp}]$ (of size $n \times n$, full rank n), we then obtain

$$\boldsymbol{U}_t = \big[\boldsymbol{D}_t \quad \boldsymbol{D}_{t,\perp}\big]\boldsymbol{W} = \big[\boldsymbol{D}_t \quad \boldsymbol{D}_{t,\perp}\big]\begin{bmatrix}\boldsymbol{W}_1\\ \boldsymbol{W}_2\end{bmatrix} = \boldsymbol{D}_t\boldsymbol{W}_1 + \boldsymbol{D}_{t,\perp}\boldsymbol{W}_2,$$

where $\boldsymbol{W}_1 \in \mathbb{R}^{r\times r}$ and $\boldsymbol{W}_2 \in \mathbb{R}^{(n-r)\times r}$ are sub-matrices of $\boldsymbol{W}$. It implies that we always decompose $\boldsymbol{U}_t$ into two components as

$$\boldsymbol{U}_t = \boldsymbol{U}_{t,\mathcal{F}}\boldsymbol{W}_1 + \boldsymbol{U}_{t,\mathcal{F},\perp}, \boldsymbol{W}_2.$$

References

[1] X.-W. Chang, “On the perturbation of the q-factor of the qr factorization,” *Numerical Linear Algebra with Applications*, vol. 19, no. 3, pp. 607–619, 2012.

[2] I. Mitliagkas, C. Caramanis, and P. Jain, “Memory limited, streaming PCA,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 2886–2894.

[3] R. Vershynin, “How close is the sample covariance matrix to the actual covariance matrix?” *J. Theor. Probab.*, vol. 25, no. 3, pp. 655–686, 2012.