# A. Proof Detailed Proofs in Section 3

## A.1. Proof of Proposition 3.4

*Proof.* Before we proceed, we first introduce the following lemma.

**Lemma A.1.** For $|x| \leq \frac{5}{9}$, we have

$$\left|(1+x)^{\frac{1}{2}} - 1 + \frac{x}{2}\right| \leq 3x^2.$$

*Proof of Lemma A.1.* By the Taylor Expansion with Lagrange remainder, we have

$$(1+x_0)^{-\frac{1}{2}} = 1 - \frac{1}{2}x_0 + \frac{1}{2}\frac{\partial^2(1+x_0)^{-\frac{1}{2}}}{\partial x^2}\Big|_{x=\theta x_0} x_0^2$$

for some $\theta \in (0,1)$. Then for $|x| \leq \frac{5}{9}$, we have

$$\left|(1+x)^{-\frac{1}{2}} - 1 + \frac{1}{2}x\right| = \frac{3}{8}\frac{1}{(1+\theta x)^{\frac{5}{2}}}x^2 \leq \frac{3}{8}\frac{1}{(1-\frac{5}{9})^{\frac{5}{2}}}x^2 = \frac{3^6}{2^8}x^2 \leq 3x^2.$$

□

We then proceed with the main proof. Since the optimization problem is symmetric about $u$ and $v$, we only prove the claim for $u$. Specifically, we first compute $u_{k+1} - u_k$. By (2.2) and (2.3), we have

$$u_{k+1} = \frac{u_k + \eta X_k Y_k^\top v_k}{\left\|u_k + \eta X_k Y_k^\top v_k\right\|_2}.$$

Since $\eta B d \leq \frac{1}{4}$, by Cauchy-Schwarz inequality, we have

$$\begin{aligned} |x| &= \left|2\eta {u_k}^\top X_k Y_k^\top v_k + \eta^2 v_k^\top Y_k X_k^\top X_k Y_k^\top v_k\right| \\ &\leq 2\eta\|u_k\|_2 \, \|X_k\|_2 \, \|Y_k\|_2 \, \|v_k\|_2 + \eta^2\|v_k\|_2 \, \|Y_k\|_2 \, \|X_k\|_2^2 \, \|Y_k\|_2 \, \|v_k\|_2 \\ &= 2\eta B d + \eta^2 B^2 d^2 < \frac{5}{9}, \end{aligned}$$

which satisfies the condition of Lemma A.1. We denote

$$\begin{aligned} T_1 :=& (1 + 2\eta {u_k}^\top X_k Y_k^\top v_k + \eta^2 v_k^\top Y_k X_k^\top X_k Y_k^\top v_k)^{-\frac{1}{2}} - 1 + \eta {u_k}^\top X_k Y_k^\top v_k \\ &+ \frac{1}{2}\eta^2 v_k^\top Y_k X_k^\top X_k Y_k^\top v_k - \frac{1}{2}\eta^2 v_k^\top Y_k X_k^\top X_k Y_k^\top v_k. \end{aligned}$$

Then by Lemma A.1, we have $|T_1| \leq 3\left(\left|2\eta {u_k}^\top X_k Y_k^\top v_k + \eta^2 v_k^\top Y_k X_k^\top X_k Y_k^\top v_k\right|\right)^2 + \frac{1}{2}\eta^2 B^2 d^2$. Therefore, we have

$$\begin{aligned} u_{k+1} - u_k &= \|u_k + \eta X_k Y_k^\top v_k\|_2^{-1}(u_k + \eta X_k Y_k^\top v_k) - u_k \\ &= (1 - \eta u_k^\top X_k Y_k^\top v_k)(u_k + \eta X_k Y_k^\top v_k) - u_k + T_1(u_k + \eta X_k Y_k^\top v_k) \\ &= (1 - \eta u_k^\top X_k Y_k^\top v_k)(u_k + \eta X_k Y_k^\top v_k) - u_k + R_k = \eta(X_k Y_k^\top v_k - u_k^\top X_k Y_k^\top v_k u_k) + R_k, \end{aligned}$$

where $R_k = \left(R_k^{(1)},\ R_k^{(2)}, \dots, R_k^{(d)}\right)^\top$ with

$$\begin{aligned} \left|R_k^{(i)}\right| &\leq \left|\left(3(2\eta {u_k}^\top X_k Y_k^\top v_k + \eta^2 v_k^\top Y_k X_k^\top X_k Y_k^\top v_k)^2 + \frac{1}{2}\eta^2 B^2 d^2\right)(u_k^{(i)} + \eta X_k Y_k^\top v_k)\right| \\ &\leq \left(3\eta^2 (dB)^2 (2 + \eta(dB))^2 + \frac{1}{2}\eta^2 (dB)^2\right)(1 + \eta d B) \\ &\leq \left(3\eta^2 (dB)^2 \frac{9}{4}^2 + \frac{1}{2}\eta^2 (dB)^2\right)\frac{4}{5} \leq 20\eta^2 (dB)^2. \end{aligned}$$

□

### A.2. Proof of Theorem 3.5

*Proof.* We first bound the infinitesimal conditional variance. Since the optimization problem is symmetric about $u$ and $v$, we only prove the claim for $u$.

$$\begin{aligned}
&\frac{d}{dt}\mathbb{E}\left(U_\eta^{(j)}(t)-U_\eta^{(j)}(0)\right)^2\Big|_{t=0}\\
&\leq \eta^{-1}tr\left(\mathbb{E}\left[(U_\eta(\eta)-U_\eta(0))\,(U_\eta(\eta)-U_\eta(0))^\top\right)\Big|U_\eta(0)=u_k,\ V_\eta(0)=v_k\right]\\
&=\eta^{-1}\mathbb{E}\left[\left(\eta\left(X_kY_k^\top u_k-u_k^\top X_kY_k^\top v_ku_k\right)+f_k(u_k,v_k)\right)^\top\left(\eta\left(X_kY_k^\top u_k-u_k^\top X_kY_k^\top v_ku_k\right)+f_k(u_k,v_k)\right)\right]\\
&=\eta\mathbb{E}\left(u_k^\top Y_kX_k^\top X_kY_k^\top u_k-2u_k^\top YX_k^\top u_ku_k^\top X_kY_k^\top v_k+u_k^\top u_k(u_k^\top X_kY_k^\top v_k)^2\right)+O(\eta^2).
\end{aligned}$$

Furthermore, by Cauchy-Schwarz inequality, we have

$$\frac{d}{dt}\mathbb{E}\left(U_\eta^{(j)}(t)-U_\eta^{(j)}(0)\right)^2\Big|_{t=0}\leq\eta\mathbb{E}\left((dB)^2+2(dB)^2+(dB)^2\right)+O(\eta^2)\leq\eta 4(dB)^2+O(\eta^2)=O(\eta).$$

By Section 4 of Chapter 7 in (Ethier and Kurtz, 2009), we know that, as $\eta\to 0^+$, $U_\eta(t)$ and $V_\eta(t)$ weakly converge to the solution of (3.1) and (3.2) with the same initial. By definition of $U_\eta(t)$ and $V_\eta(t)$, we complete the proof. □

### A.3. Proof of Theorem 3.6

*Proof.* Since $P$ is an orthonormal matrix, $\|H_j\|_2=\|W_j\|_2=1$ for all $j=1,...,d$. Thus, we have

$$\begin{aligned}
\frac{d}{dt}H^{(i)}&=\lambda_iH^{(i)}-\sum_{j=1}^{2d}\lambda_j(H^{(j)})^2H^{(i)}\\
&=\lambda_i\sum_{j=1}^{2d}(H^{(j)})^2H^{(i)}-\sum_{j=1}^{2d}\lambda_j(H^{(j)})^2H^{(i)}\\
&=H^{(i)}\sum_{j=1}^{2d}\left(\lambda_i-\lambda_j\right)(H^{(j)})^2.
\end{aligned}$$

We then verify (3.8) satisfies (3.7). By (Evans, 1988), we know that since $H_j(t)$ is continuously differentiable in $t$, the solution to the ODE is unique. For notational simplicity, we denote

$$S^{(j)}(t)=H^{(j)}(0)\exp(\lambda_jt).$$

Then we have

$$H^{(i)}(t)=\frac{S^{(i)}(t)}{\sqrt{\sum_{j=1}^{2d}\left(S^{(j)}(t)\right)^2}}.$$

Now we only need to verify

$$\begin{aligned}
\frac{d}{dt}H^{(i)}(t)&=\frac{\left(\lambda_iS^{(i)}(t)\right)\sqrt{\sum_{j=1}^{2d}\left(S^{(j)}(t)\right)^2}-\frac{\left(2\sum_{j=1}^{2d}\lambda_j\left(S^{(j)}(t)\right)^2\right)S^{(i)}(t)}{2\sqrt{\sum_{j=1}^{2d}\left(S^{(j)}(t)\right)^2}}}{\sum_{j=1}^{2d}\left(S^{(j)}(t)\right)^2}\\
&=\lambda_i\frac{S^{(i)}(t)}{\sqrt{\sum_{j=1}^{2d}\left(S^{(j)}(t)\right)^2}}-\sum_{j=1}^{2d}\lambda_j\frac{\left(S^{(j)}(t)\right)^2}{\sum_{j=1}^{2d}\left(S^{(j)}(t)\right)^2}\frac{S^{(i)}(t)}{\sqrt{\sum_{j=1}^{2d}\left(S^{(j)}(t)\right)^2}}\\
&=\lambda_iH^{(i)}(t)-\sum_{j=1}^{2d}\lambda_j\left(H^{(j)}(t)\right)^2H^{(i)}(t),
\end{aligned}$$

which completes the proof. □

## B. Proof Detailed Proofs in Section 4

### B.1. Proof of Theorem 4.1

*Proof.* We calculate the infinitesimal conditional expectation and variance for $Z_\eta^{(i)},\ i \neq j$.

$$\begin{aligned}\frac{d}{dt}\mathbb{E}Z_\eta^{(i)}(t)\big|_{t=0} &= \eta^{-1}\mathbb{E}\left[Z_\eta^{(i)}(\eta) - Z_\eta^{(i)}(0)\big|H_\eta(0)=h\right]\\ &= \eta^{-1}\mathbb{E}\left[\eta^{-1/2}\left(H_\eta^{(i)}(\eta) - H_\eta^{(i)}(0)\right)\big|H_\eta(0)=h\right]\\ &= \eta^{-1/2}h^{(i)}\sum_{l=1}^{2d}(\lambda_i-\lambda_l)\,(h^{(l)})^2 + O(\eta) = Z_\eta^{(i)}\,(\lambda_i-\lambda_j) + o(1),\end{aligned} \tag{B.1}$$

where the last equality comes from the assumption that the algorithm starts near $j^{th}$ column of $P,\ j \neq 1$, i.e., $h \approx e_j$. To compute variance, we first compute $\widehat{\Lambda}$,

$$\widehat{\Lambda} = P^\top Q P = \frac{1}{2}\begin{pmatrix} \overline{Y}\,\overline{X}^\top + \overline{X}\,\overline{Y}^\top & \overline{Y}\,\overline{X}^\top - \overline{X}\,\overline{Y}^\top \\ -\overline{Y}\,\overline{X}^\top + \overline{X}\,\overline{Y}^\top & -\overline{Y}\,\overline{X}^\top - \overline{X}\,\overline{Y}^\top\end{pmatrix},$$

where $Q$ is defined in (3.3). Then we analyze $e_i^\top\widehat{\Lambda}e_j$ by cases:

$$e_i^\top\widehat{\Lambda}e_j = \begin{cases}\frac{1}{2}\left(\overline{X}^{(i)}\overline{Y}^{(j)} + \overline{X}^{(j)}\overline{Y}^{(i)}\right) & \text{if } \max(i,j)\le d,\\ \frac{1}{2}\left(-\overline{X}^{(j)}\overline{Y}^{(i-d)} + \overline{X}^{(i-d)}\overline{Y}^{(j)}\right) & \text{if } j\le d<i,\\ \frac{1}{2}\left(\overline{X}^{(j-d)}\overline{Y}^{(i)} - \overline{X}^{(i)}\overline{Y}^{(j-d)}\right) & \text{if } i\le d<j,\\ \frac{1}{2}\left(-\overline{X}^{(i-d)}\overline{Y}^{(j-d)} - \overline{X}^{(j-d)}\overline{Y}^{(i-d)}\right) & \text{if } \min(i,j)>d,\end{cases}$$

which further implies

$$\begin{aligned}\frac{d}{dt}\mathbb{E}(Z_\eta^{(i)}(t) - Z_\eta^{(i)}(0))^2\big|_{t=0} &= \eta^{-1}\mathbb{E}\left[\left(Z_\eta^{(i)}(\eta) - Z_\eta^{(i)}(0)\right)^2\big|H_\eta(0)=h\right]\\ &= \eta^{-2}\mathbb{E}[\eta^2(\widehat{\Lambda}h - h^\top\widehat{\Lambda}hh)(\widehat{\Lambda}h - h^\top\widehat{\Lambda}hh)^\top]_{i,i} + O(\eta)\\ &= \mathbb{E}(e_i^\top\widehat{\Lambda}e_je_j^\top\widehat{\Lambda}^\top e_i) + o(1)\\ &= \frac{1}{4}\left(\gamma_i\omega_j + \gamma_j\omega_i + 2\,\text{sign}(i-d-1/2)\cdot\text{sign}(j-1/2-d)\cdot\alpha_{ij}\right).\end{aligned} \tag{B.2}$$

By (B.1) and (B.2), we get the limit stochastic differential equation,

$$dZ^{(i)}(t) = -(\lambda_j - \lambda_i)Z^{(i)}(t)dt + \beta_{ij}dB(t).$$

□

### B.2. Proof of Proposition 4.2

*Proof.* Our analysis is based on approximating $z_{\eta,k}^{(1)}$ by its continuous approximation $Z_\eta^{(1)}(t)$, which is normal distributed at time $t$. By simple manipulation, we have

$$\mathbb{P}\left((h_{\eta,N_1}^{(2)})^2 \le 1-\delta^2\right) = \mathbb{P}\left((z_{\eta,N_1}^{(2)})^2 \le \eta^{-1}(1-\delta^2)\right) \ge \mathbb{P}(|z_{\eta,N_1}^{(1)}| \ge \eta^{-\frac{1}{2}}\delta).$$

We then prove $P\left(\left|z_{\eta,N_1}^{(1)}\right| \ge \eta^{-\frac{1}{2}}\delta\right) \ge 1-\nu$. At time t, $z_{\eta,k}^{(1)}$ approximates to a normal distribution with mean 0 and variance $\frac{\beta_{12}^2}{2(\lambda_1-\lambda_2)}\left[\exp\left(2(\lambda_1-\lambda_2)\eta N_1\right) - 1\right]$. Therefore, let $\Phi(x)$ be the CDF of $N(0,1)$, we have

$$\mathbb{P}\left(\frac{\left|z_{\eta,N_1}^{(1)}\right|}{\sqrt{\frac{\beta_{12}^2}{2(\lambda_1-\lambda_2)}\cdot\left[\exp\left(2(\lambda_1-\lambda_2)\eta N_1\right)-1\right]}} \ge \Phi^{-1}\left(\frac{1+\nu}{2}\right)\right) \approx 1-\nu,$$

which requires

$$\eta^{-\frac{1}{2}}\delta \leq \Phi^{-1}\left(\frac{1+\nu}{2}\right)\cdot\sqrt{\frac{\beta_{12}^2}{2(\lambda_1-\lambda_2)}\cdot[\exp\left(2(\lambda_1-\lambda_2)\eta N_1\right)-1]}.$$

Solving the above inequality, we get

$$N_1=\frac{\eta^{-1}}{2(\lambda_1-\lambda_2)}\log\left(\frac{2\eta^{-1}\delta^2(\lambda_1-\lambda_2)}{\Phi^{-1}\left(\frac{1+\nu}{2}\right)^2\beta_{12}^2}+1\right).$$

□

### B.3. Proof of Proposition 4.3

*Proof.* After Phase I, we restart our counter, i.e., $h_{\eta,0}^{(1)}=\delta$. By (3.8) and $h_{\eta,N_2}^{(1)}$ approximating to the process $H^{(1)}(\eta N_2)$, we obtain

$$\begin{aligned}\left(h_{\eta,N_2}^{(1)}(t)\right)^2=\left(H^{(1)}(\eta N_2)\right)^2&=\left(\sum_{j=1}^{2d}\left(\left(H^{(j)}(0)\right)^2\exp\left(2\lambda_j\eta N_2\right)\right)\right)^{-1}\left(H^{(1)}(0)\right)^2\exp\left(2\lambda_1\eta N_2\right)\\&\geq\left(\delta^2\exp(2\lambda_1\eta N_2)+(1-\delta^2)\exp(2\lambda_2\eta N_2)\right)^{-1}\delta^2\exp(2\lambda_1\eta N_2),\end{aligned}$$

which requires

$$\left(\delta^2\exp(2\lambda_1\eta N_2)+(1-\delta^2)\exp(2\lambda_2\eta N_2)\right)^{-1}\delta^2\exp(2\lambda_1\eta N_2)\geq\eta^{-1}(1-\delta^2).$$

Solving the above inequality, we get

$$N_2=\frac{\eta^{-1}}{2(\lambda_1-\lambda_2)}\log\frac{1-\delta^2}{\delta^2}.$$

□

### B.4. Proof of Theorem 4.4

*Proof.* For $i=2,...,2d$, we compute the infinitesimal conditional expectation and variance,

$$\begin{aligned}\frac{d}{dt}\mathbb{E}Z_\eta^{(i)}(t)\big|_{t=t_0}&=\eta^{-1}\mathbb{E}\left[Z_\eta^{(i)}(t_0+\eta)-Z_\eta^{(i)}(t_0)\big|H^\eta(t_0)=h\right]\\&=\eta^{-1/2}h_i\sum_{j=1}^{2d}(\lambda_i-\lambda_j)\,h_j^2+O(\eta)=Z^{(i)}\left(\lambda_i-\lambda_1\right)+o(1),\\\frac{d}{dt}\mathbb{E}\left(Z_\eta^{(i)}(t)-Z_\eta^{(i)}(t_0)\right)^2\big|_{t=t_0}&=\eta^{-1}\mathbb{E}\left[\left(Z_\eta^{(i)}(t_0+\eta)-Z_\eta^{(i)}(t_0)\right)^2\big|H^\eta(t_0)=h\right]\\&=\eta^{-2}\mathbb{E}\left[\eta^2(\widehat{\Lambda}h-h^\top\widehat{\Lambda}hh)(\widehat{\Lambda}h-h^\top\widehat{\Lambda}hh)^\top\right]_{i,i}+O(\eta)\\&=\mathbb{E}(e_i^\top\widehat{\Lambda}e_1e_1^\top\widehat{\Lambda}^\top e_i)+o(1)=\frac{1}{4}\left(\gamma_i\omega_1+\gamma_1\omega_i-2\,\mathrm{sign}(i-d-1/2)\alpha_{i1}\right)+o(1).\end{aligned}$$

Following similar lines to the proof of Theorem 4.1, by Section 4 of Chapter 7 in (Ethier and Kurtz, 2009), we have for each $k=2,...,2d$, if $Z^{(i)}(0)=\eta^{-1/2}h_{\eta,0}^{(i)}$ as $\eta\to 0^+$, then the stochastic process $\eta^{-1/2}h_{\eta,\lfloor t\eta^{-1}\rfloor}^{(k)}$ weakly converges to the solution of the stochastic differential equation (4.3). □

### B.5. Proof of Proposition 4.5

*Proof.* Since we restart our counter, we have $\sum_{i=2}^{2d}(z_{\eta,0}^{(i)})^2=\eta^{-1}\delta^2$. Since $z_{\eta,k}^{(i)}$ approximates to $Z^{(i)}(\eta k)$ and its second moment:

$$\mathbb{E}\left(Z^{(i)}(t)\right)^2=\frac{\beta_{i1}^2}{2(\lambda_1-\lambda_i)}+\left(\left(Z^{(i)}(0)\right)^2-\frac{\beta_{i1}^2}{2(\lambda_1-\lambda_i)}\right)\exp\left[-2(\lambda_1-\lambda_i)t\right],\quad\text{for } i\neq 1,$$

we use the Markov inequality:

$$\begin{aligned}\mathbb{P}\left(\sum_{i=2}^{2d}\left(h_{\eta,N_3}^{(i)}\right)^2>\epsilon\right)&\leq\frac{\mathbb{E}\left(\sum_{i=2}^{2d}\left(h_{\eta,N_3}^{(i)}\right)^2\right)}{\epsilon}=\frac{\mathbb{E}\left(\sum_{i=2}^{2d}\left(z_{\eta,N_3}^{(i)}\right)^2\right)}{\eta^{-1}\epsilon}\\&=\frac{1}{\eta^{-1}\epsilon}\sum_{i=2}^{2d}\frac{\beta_{i1}^2}{2(\lambda_1-\lambda_i)}\Big(1-\exp\big(-2(\lambda_1-\lambda_i)\eta N_3\big)\Big)+\left(z_{\eta,0}^{(i)}\right)^2\exp\left[-2(\lambda_1-\lambda_i)\eta N_3\right]\\&\leq\frac{1}{\eta^{-1}\epsilon}\Big(\frac{d\max\limits_{2\leq i\leq d}(\beta_{i1}^2)}{2(\lambda_1-\lambda_2)}\Big(1-\exp\big(-2(\lambda_1-\lambda_d)\eta N_3\big)\Big)\\&\qquad+\frac{d\max\limits_{d+1\leq i\leq 2d}(\beta_{i1}^2)}{2(\lambda_1+\lambda_d)}\Big(1-\exp\big(-4\lambda_1\eta N_3\big)\Big)+\delta^2\exp\left[-2(\lambda_1-\lambda_2)\eta N_3\right]\Big)\\&\leq\frac{1}{\eta^{-1}\epsilon}\left(\frac{d\max\limits_{1\leq i\leq d}(\beta_{i1}^2)}{(\lambda_1-\lambda_2)}+\delta^2\exp\left[-2(\lambda_1-\lambda_2)\eta N_3\right]\right).\end{aligned}$$

To guarantee $\frac{1}{\eta^{-1}\epsilon}\left(\frac{d\max_{1\leq i\leq d}(\beta_{i1}^2)}{(\lambda_1-\lambda_2)}+\delta^2\exp\left[-2(\lambda_1-\lambda_2)\eta N_3\right]\right)\leq\frac{1}{4}$, we get:

$$N_3\geq\frac{\eta^{-1}}{2(\lambda_1-\lambda_2)}\log\left(\frac{4(\lambda_1-\lambda_2)\delta^2}{(\lambda_1-\lambda_2)\epsilon\eta^{-1}-4d\max\limits_{1\leq i\leq d}\beta_{i1}^2}\right).$$

□

### B.6. Proof of Corollary 4.6

*Proof.* First, we prove that $\|u_{\eta,k}-\widehat{u}\|_2^2+\|v_{\eta,k}-\widehat{v}\|_2^2$ can be bounded by $3\sum_{i=2}^{2d}\left(h_{\eta,k}^{(i)}\right)^2$, when it is near the optima. Recall that $h_{\eta,k}=\frac{1}{\sqrt{2}}P^\top(u_{\eta,k}^\top\ v_{\eta,k}^\top)^\top$ and $e_1=\widehat{h}=\frac{1}{\sqrt{2}}P(\widehat{u}^\top\ \widehat{v}^\top)^\top$. Our analysis has shown that when $k$ is large enough, the SGD iterates near the optima. Then we have

$$\begin{aligned}\|u_{\eta,k}-\widehat{u}\|_2^2+\|v_{\eta,k}-\widehat{v}\|_2^2&=4-2\langle u_{\eta,k},\widehat{u}\rangle-2\langle v_{\eta,k},\widehat{v}\rangle=4-4h_{\eta,k}^1\\&=4-4\sqrt{1-\sum\nolimits_{i=2}^{2d}\left(h_{\eta,k}^{(i)}\right)^2}=\frac{16\sum_{i=2}^{2d}\left(h_{\eta,k}^{(i)}\right)^2}{4+4\sqrt{1-\sum_{i=2}^{2d}\left(h_{\eta,k}^{(i)}\right)^2}}\leq 3\sum\nolimits_{i=2}^{2d}\left(h_{\eta,k}^{(i)}\right)^2,\end{aligned}\tag{B.3}$$

where the last inequality holds since $k$ is large enough such that $\sum_{i=2}^{2d}\left(h_{\eta,k}^{(i)}\right)^2$ is sufficiently small. By Propositions 4.2, 4.3, and 4.5, the total iteration number is

$$N=N_1+N_2+N_3.\tag{B.4}$$

To explicitily bound $N$ in (B.4) in terms of sample size n, we consider

$$N_1=\frac{\eta^{-1}}{2(\lambda_1-\lambda_2)}\log\left(\frac{2\eta^{-1}\delta^2(\lambda_1-\lambda_2)}{\Phi^{-1}\left(\frac{1+\nu}{2}\right)^2\beta_{12}^2}+1\right),\tag{B.5}$$

$$N_2=\frac{\eta^{-1}}{2(\lambda_1-\lambda_2)}\log\frac{1-\delta^2}{\delta^2},\tag{B.6}$$

$$N_3=\frac{\eta^{-1}}{2(\lambda_1-\lambda_2)}\log\left(\frac{4(\lambda_1-\lambda_2)\delta^2}{(\lambda_1-\lambda_2)\epsilon\eta^{-1}-4d\max\limits_{1\leq i\leq d}\beta_{i1}^2}\right).\tag{B.7}$$

Given a small enough $\epsilon$, we choose $\eta$ as follow:

$$\eta \asymp \frac{\epsilon(\lambda_1 - \lambda_2)}{d \max_{1 \le i \le d} \beta_{i1}^2}. \tag{B.8}$$

Combining the above sample complexities (B.5), (B.6), (B.7), and (B.8), we get

$$N = O\left[\frac{d}{\epsilon(\lambda_1 - \lambda_2)^2} \log\left(\frac{d}{\epsilon}\right)\right]. \tag{B.9}$$

By Proposition 4.5 with (B.3), given $\nu < 1/9$, after at most $N$ iterations, we have

$$\|u_{\eta,n} - \widehat{u}\|_2^2 + \|v_{\eta,n} - \widehat{v}\|_2^2 \le 3\|h_{\eta,n} - \widehat{h}\|_2^2 \le 3\epsilon,$$

with probability at least $\frac{2}{3}$. □