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

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

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

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

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

$$\left|(1 + x)^{-\frac{1}{2}} - 1 + \frac{x}{2}\right| = \frac{3}{8} \frac{1}{(1 + \theta x)^{\frac{1}{2}}} x^2 \leq \frac{3}{8} \frac{1}{(1 - \frac{5}{9})^{\frac{1}{2}}} x^2 = \frac{36}{25} x^2 \leq 3x^2.$$ 

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

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

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

$$\left\| x \right\| = \left\| 2\eta u_k^T X_k Y_k^T v_k + \eta^2 v_k^T Y_k X_k^T X_k Y_k^T v_k \right\|
\leq 2\eta \left\| u_k \right\| \left\| X_k \right\| \left\| Y_k \right\| \left\| v_k \right\| + \eta^2 \left\| v_k \right\| \left\| Y_k \right\|\left\| X_k \right\| \left\| Y_k \right\| \left\| v_k \right\|$$

$$= 2\eta Bd + \eta^2 B^2 d^2 < \frac{5}{9},$$

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

$$T_1 := (1 + 2\eta u_k^T X_k Y_k^T v_k + \eta^2 v_k^T Y_k X_k^T X_k Y_k^T v_k)^{-\frac{1}{2}} - 1 + \eta u_k^T X_k Y_k^T v_k + \frac{1}{2} \eta^2 v_k^T Y_k X_k^T X_k Y_k^T v_k.$$ 

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

$$u_{k+1} - u_k = \left\| u_k + \eta X_k Y_k^T v_k \right\|_2^{-1} \left( u_k + \eta X_k Y_k^T v_k \right) - u_k$$

$$= (1 - \eta u_k^T X_k Y_k^T v_k)(u_k + \eta X_k Y_k^T v_k) - u_k + T_1(u_k + \eta X_k Y_k^T v_k)$$

$$= (1 - \eta u_k^T X_k Y_k^T v_k)(u_k + \eta X_k Y_k^T v_k) - u_k + R = \eta (X_k Y_k^T v_k - u_k^T X_k Y_k^T v_k) + R_k,$$

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

$$\left\| R_k^{(i)} \right\| \leq \left( 3(\eta u_k^T X_k Y_k^T v_k + \eta^2 v_k^T Y_k X_k^T X_k Y_k^T v_k)^2 + \frac{1}{2} \eta^2 B^2 d^2 \right) \left( u_k^{(i)} + \eta X_k Y_k^T v_k \right)$$

$$\leq \left( 3\eta^2 (dB)^2 (2 + \eta dB)^2 + \frac{1}{2} \eta^2 (dB)^2 \right) (1 + \eta dB)$$

$$\leq \left( 3\eta^2 (dB)^2 \frac{9}{4} + \frac{1}{2} \eta^2 (dB)^2 \right) \frac{4}{3} \leq 20\eta^2 (dB)^2.$$ 

\[\square\]
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 $u$.

\[
\frac{d}{dt} \mathbb{E} \left( \eta^{(j)}(t) - U_{\eta}(0) \right)^2 \bigg|_{t=0} \\
\leq \eta^{-1} \text{tr} \left( \mathbb{E} \left( \eta(\eta) - U_{\eta}(0) \right) (U_{\eta}(\eta) - U_{\eta}(0))^\top \right) U_{\eta}(0) = u_k, \ V_{\eta}(0) = v_k \\
= \eta^{-1} \mathbb{E} \left[ \eta \left( X_k Y_k^\top u_k - u_k^\top X_k Y_k^\top v_k u_k \right) + f_k(u_k, v_k) \right]^\top \left( \eta \left( X_k Y_k^\top u_k - u_k^\top X_k Y_k^\top v_k u_k \right) + f_k(u_k, v_k) \right] \\
= \eta \mathbb{E} \left( u_k^\top Y_k X_k^\top X_k Y_k^\top u_k - 2u_k^\top Y X_k^\top u_k u_k^\top X_k Y_k^\top v_k + u_k^\top u_k (u_k^\top X_k Y_k^\top v_k)^2 \right) + O(\eta^2).
\]

Furthermore, by Cauchy-Schwarz inequality, we have

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

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. \qed

A.3. Proof of Theorem 3.6

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

\[
\frac{d}{dt} H^{(i)} = \lambda_i H^{(i)} - \sum_{j=1}^{2d} \lambda_j (H^{(j)})^2 H^{(i)} \\
= \lambda_i \sum_{j=1}^{2d} (H^{(j)})^2 H^{(i)} - \sum_{j=1}^{2d} \lambda_j (H^{(j)})^2 H^{(i)} \\
= H^{(i)} \sum_{j=1}^{2d} (\lambda_i - \lambda_j) (H^{(j)})^2.
\]

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

\[
S^{(j)}(t) = H^{(j)}(0) \exp(\lambda_j t).
\]

Then we have

\[
H^{(i)}(t) = \frac{S^{(i)}(t)}{\sqrt{\sum_{j=1}^{2d} (S^{(j)}(t))^2}}.
\]

Now we only need to verify

\[
\frac{d}{dt} H^{(i)}(t) = \frac{\left( \lambda_i S^{(i)}(t) \right) \sqrt{\sum_{j=1}^{2d} (S^{(j)}(t))^2} - \left( \sum_{j=1}^{2d} \lambda_j (S^{(j)}(t))^2 \right) S^{(i)}(t)}{\sum_{j=1}^{2d} (S^{(j)}(t))^2} \\
= \lambda_i \frac{S^{(i)}(t)}{\sqrt{\sum_{j=1}^{2d} (S^{(j)}(t))^2}} - \sum_{j=1}^{2d} \lambda_j \frac{(S^{(j)}(t))^2}{\sqrt{\sum_{j=1}^{2d} (S^{(j)}(t))^2}} S^{(i)}(t) \\
= \lambda_i H^{(i)}(t) - \sum_{j=1}^{2d} \lambda_j \left( H^{(j)}(t) \right)^2 H^{(i)}(t),
\]

which completes the proof. \qed
B. Proof Detailed Proofs in Section 4

B.1. Proof of Theorem 4.1

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

\[
\begin{align*}
\frac{d}{dt} \mathbb{E} Z^{(i)}_{\eta}(t) \big|_{t=0} &= \eta^{-1} \mathbb{E} \left[ Z^{(i)}_{\eta}(\eta) - Z^{(i)}_{\eta}(0) \big| H_{\eta}(0) = h \right] \\
&= \eta^{-1} \mathbb{E} \left[ \eta^{-1/2} \left( H^{(i)}_{\eta}(\eta) - H^{(i)}_{\eta}(0) \right) \big| H_{\eta}(0) = h \right] \\
&= \eta^{-1/2} h^{(i)} \sum_{l=1}^{2d} (\lambda_l - \lambda_0) (h^{(l)})^2 + O(\eta) = Z^{(i)}(\lambda_i - \lambda_j) + o(1),
\end{align*}
\]

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 $\hat{\Lambda}$,

\[
\hat{\Lambda} = P^T Q P = \frac{1}{2} \begin{pmatrix}
\bar{Y} \bar{X}^T + \bar{X} \bar{Y}^T & \bar{Y} \bar{X}^T - \bar{X} \bar{Y}^T \\
-\bar{Y} \bar{X}^T + \bar{X} \bar{Y}^T & -\bar{Y} \bar{X}^T - \bar{X} \bar{Y}^T
\end{pmatrix},
\]

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

\[
e_i^T \hat{\Lambda} e_j = \begin{cases}
\frac{1}{2} \left( \bar{X}^{(i)} (j) + \bar{X}^{(j)} (i) \right) & \text{if } \max(i, j) \leq d, \\
\frac{1}{2} \left( -\bar{X}^{(j)} (i-d) + \bar{X}^{(i-d)} (j) \right) & \text{if } j \leq d < i, \\
\frac{1}{2} \left( \bar{X}^{(j-d)} (i) + \bar{X}^{(i)} (j-d) \right) & \text{if } i \leq d < j, \\
\frac{1}{2} \left( -\bar{X}^{(i-d)} (j-d) - \bar{X}^{(j-d)} (i-d) \right) & \text{if } \min(i, j) > d,
\end{cases}
\]

which further implies

\[
\begin{align*}
\frac{d}{dt} \mathbb{E} (Z^{(i)}_{\eta}(t) - Z^{(i)}_{\eta}(0))^2 \big|_{t=0} &= \eta^{-1} \mathbb{E} \left[ (Z^{(i)}_{\eta}(\eta) - Z^{(i)}_{\eta}(0))^2 \big| H_{\eta}(0) = h \right] \\
&= \eta^{-2} \mathbb{E} [\eta^2 (\hat{\Lambda} h - h^T \hat{\Lambda} h)(\hat{\Lambda} h - h^T \hat{\Lambda} h)^T]_{i,i} + O(\eta) \\
&= \mathbb{E} (e_i^T \hat{\Lambda} e_j)_{i,i} \alpha_{ij} + o(1) \\
&= \frac{1}{4} \left( \gamma_i \omega_j + \gamma_j \omega_i + 2 \text{sign}(i-d-1/2) \cdot \text{sign}(j-1/2-d) \cdot \alpha_{ij} \right).
\end{align*}
\]

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 $z^{(1)}_{\eta,k}$ by its continuous approximation $Z^{(1)}_{\eta}(t)$, which is normal distributed at time $t$. By simple manipulation, we have

\[
\mathbb{P} \left( \left( h^{(2)}_{\eta,N_1} \right)^2 \leq 1 - \delta^2 \right) = \mathbb{P} \left( \left( z^{(2)}_{\eta,N_1} \right)^2 \leq \eta^{-1} (1 - \delta^2) \right) \geq \mathbb{P} (|z^{(1)}_{\eta,N_1}| \geq \eta^{-\frac{1}{2}} \delta).
\]

We then prove \( P \left( \left| z^{(1)}_{\eta,N_1} \right| \geq \eta^{-\frac{1}{2}} \delta \right) > 1 - \nu \). At time $t$, $z^{(1)}_{\eta,N_1}$ 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

\[
\mathbb{P} \left( \left| z^{(1)}_{\eta,N_1} \right| \geq \frac{\sqrt{\beta_{12}^2 \cdot \left[ \exp \left( 2(\lambda_1 - \lambda_2)\eta N_1 \right) - 1 \right]}}{2(\lambda_1 - \lambda_2)} \cdot \Phi^{-1} \left( \frac{1 + \nu}{2} \right) \right) \approx 1 - \nu.
\]
which requires
\[
\eta^{-\frac{1}{2}} \delta \leq \Phi^{-1}\left(\frac{1 + \nu}{2}\right) \cdot \sqrt{\frac{\beta_1^2}{2(\lambda_1 - \lambda_2)}} \cdot \left[\exp(2(\lambda_1 - \lambda_2)\eta N_1) - 1\right].
\]
Solving the above inequality, we get
\[
N_1 = \frac{\eta^{-1}}{2(\lambda_1 - \lambda_2)} \log \left(\frac{2\eta^{-1}\delta^2(\lambda_1 - \lambda_2)}{\Phi^{-1}\left(\frac{1 + \nu}{2}\right)^2 \beta_1^2} + 1\right).
\]

**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
\[
\left(h_{\eta,N_2}^{(1)}(t)\right)^2 = \left(H^{(1)}(\eta N_2)\right)^2 = \left(\sum_{j=1}^{2d} \left(\left(H^{(j)}(0)\right)^2 \exp(2\lambda_j \eta N_2)\right)\right)\left(H^{(1)}(0)\right)^2 \exp(2\lambda_1 \eta N_2)
\]
\[
\geq (\delta^2 \exp(2\lambda_1 \eta N_2) + (1 - \delta^2) \exp(2\lambda_2 \eta N_2))^{-1} \delta^2 \exp(2\lambda_1 \eta N_2),
\]
which requires
\[
(\delta^2 \exp(2\lambda_1 \eta N_2) + (1 - \delta^2) \exp(2\lambda_2 \eta N_2))^{-1} \delta^2 \exp(2\lambda_1 \eta N_2) \geq \eta^{-1}(1 - \delta^2).
\]
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,
\[
\frac{d}{dt} \mathbb{E}\left[Z^{(i)}_n(t)\right]_{t=t_0} = \eta^{-1} \mathbb{E} \left[Z^{(i)}_n(t_0 + \eta) - Z^{(i)}_n(t_0)\right] H^{(n)}(t_0) = h
\]
\[
= \eta^{-1/2} h \sum_{j=1}^{2d} (\lambda_i - \lambda_j) h_j^2 + O(\eta) = Z^{(i)}(\lambda_i - \lambda_1) + o(1),
\]
\[
\frac{d}{dt} \mathbb{E}\left(Z^{(i)}_n(t) - Z^{(i)}_n(t_0)\right)^2_{t=t_0} = \eta^{-1} \mathbb{E} \left[Z^{(i)}_n(t_0 + \eta) - Z^{(i)}_n(t_0)\right]^2 H^{(n)}(t_0) = h
\]
\[
= \eta^{-2} \mathbb{E} \left[\eta^2 (\Lambda h - h^T \Lambda h)(\Lambda h - h^T \Lambda h)^T\right]_{i,i} + O(\eta)
\]
\[
= \mathbb{E}(e_i^T \Lambda e_i e_i^T \Lambda^T e_i) + o(1) = \frac{1}{4} (\gamma_i \omega_i + \gamma_i \omega_i - 2 \text{sign}(i - d - 1/2)\alpha_{i1}) + o(1).
\]

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 \(\eta^{-1/2} h_{\eta,[\eta^{-1}]}^{(k)}\) 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 \(\sum_{i=2}^{2d} (Z^{(i)}_{n,0})^2 = \eta^{-1} \delta^2\). Since \(Z^{(i)}_{n,k}\) approximates to \(Z^{(i)}(\eta k)\) and its second moment:
\[
\mathbb{E}\left(Z^{(i)}(t)\right)^2 = \frac{\beta_1^2}{2(\lambda_1 - \lambda_i)} + \left(\left(Z^{(i)}(0)\right)^2 - \frac{\beta_1^2}{2(\lambda_1 - \lambda_i)}\right) \exp[-2(\lambda_1 - \lambda_i)t], \text{ for } i \neq 1,
\]
we use the Markov inequality:

$$
P\left( \sum_{i=2}^{2d} \left( \hat{h}_{\eta,N_3}^{(i)} \right)^2 > \epsilon \right) \leq \frac{\mathbb{E} \left( \sum_{i=2}^{2d} \left( \hat{h}_{\eta,N_3}^{(i)} \right)^2 \right)}{\epsilon} = \frac{\mathbb{E} \left( \sum_{i=2}^{2d} \left( z_{\eta,N_3}^{(i)} \right)^2 \right)}{\eta^{-1}\epsilon} \\
= \frac{1}{\eta^{-1}\epsilon} \sum_{i=2}^{2d} \frac{\beta_1^2}{2(\lambda_1 - \lambda_i)} \left( 1 - \exp \left( -2(\lambda_1 - \lambda_i)\eta N_3 \right) \right) + \left( z_{\eta,0}^{(i)} \right)^2 \exp \left[ -2(\lambda_1 - \lambda_i)\eta N_3 \right] \\
\leq \frac{1}{\eta^{-1}\epsilon} \max_{2 \leq i \leq 2d} \left( \frac{\beta_1^2}{2(\lambda_1 - \lambda_2)} \right) \left( 1 - \exp \left( -2(\lambda_1 - \lambda_d)\eta N_3 \right) \right) \\
+ \frac{d}{2(\lambda_1 + \lambda_d)} \left( 1 - \exp \left( -4\lambda_1\eta N_3 \right) \right) + \delta^2 \exp \left[ -2(\lambda_1 - \lambda_d)\eta N_3 \right] \\
\leq \frac{1}{\eta^{-1}\epsilon} \left( \frac{d}{\lambda_1 - \lambda_2} \right) + \delta^2 \exp \left[ -2(\lambda_1 - \lambda_d)\eta N_3 \right].
$$

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

$$
N_3 \geq \frac{\eta^{-1}}{2(\lambda_1 - \lambda_2)} \log \left( \frac{4(\lambda_1 - \lambda_2)\delta^2}{(\lambda_1 - \lambda_2)\epsilon\eta^{-1} - 4d \max_{1 \leq i \leq d} \beta_1^2} \right).
$$

\hfill \Box

**B.6. Proof of Corollary 4.6**

**Proof.** First, we prove that $\|u_{\eta,k} - \tilde{u}\|_2^2 + \|v_{\eta,k} - \tilde{v}\|_2^2$ can be bounded by $3 \sum_{i=2}^{2d} \left( \hat{h}_{\eta,k}^{(i)} \right)^2$, when it is near the optima. Recall that $h_{\eta,k} = \frac{1}{\sqrt{2}} P^T (u_{\eta,k}^T v_{\eta,k}^T)$ and $c_1 = \tilde{h} = \frac{1}{\sqrt{2}} P(\tilde{u}^T \tilde{v}^T)^T$. Our analysis has shown that when $k$ is large enough, the SGD iterates near the optima. Then we have

$$
\|u_{\eta,k} - \tilde{u}\|_2^2 + \|v_{\eta,k} - \tilde{v}\|_2^2 = 4 - 2\langle u_{\eta,k}, \tilde{u} \rangle - 2\langle v_{\eta,k}, \tilde{v} \rangle = 4 - 4h_{\eta,k}^1 \\
= 4 - 4\sqrt{1 - \sum_{i=2}^{2d} \left( \hat{h}_{\eta,k}^{(i)} \right)^2} = \frac{16 \sum_{i=2}^{2d} \left( \hat{h}_{\eta,k}^{(i)} \right)^2}{4 + 4\sqrt{1 - \sum_{i=2}^{2d} \left( \hat{h}_{\eta,k}^{(i)} \right)^2}} \leq 3 \sum_{i=2}^{2d} \left( \hat{h}_{\eta,k}^{(i)} \right)^2,
$$

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

$$
N = N_1 + N_2 + N_3.
$$

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

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

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

$$
N_3 = \frac{\eta^{-1}}{2(\lambda_1 - \lambda_2)} \log \left( \frac{4(\lambda_1 - \lambda_2)\delta^2}{(\lambda_1 - \lambda_2)\epsilon\eta^{-1} - 4d \max_{1 \leq i \leq d} \beta_1^2} \right).
$$
Given a small enough $\epsilon$, we choose $\eta$ as follow:

$$\eta \asymp \frac{\epsilon(\lambda_1 - \lambda_2)}{d \max_{1 \leq i \leq d} \beta_{1i}^2}.$$  

(B.8)

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

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

(B.9)

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

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

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