Supplementary Material to "Spectral Learning of General Weighted Automata via Constrained Matrix Completion"

Borja Balle
Universitat Politècnica de Catalunya
bballe@lsi.upc.edu

Mehryar Mohri
Courant Institute and Google Research
mohri@cims.nyu.edu

For convenience we begin by recalling the statement of our main result and the key assumptions used in the proof.

Assumption 1 *There exists a constant $\nu > 0$ such that if $(x, y) \sim \mathcal{D}$, then $|y| \leq \nu$ almost surely.*

Assumption 2 *There exist constants $c, \eta > 0$ such that $\mathbb{P}_{x \sim \mathcal{D}_\Sigma}[|x| \geq t] \leq \exp(-ct^{1+\eta})$ holds for all $t \geq 0$.*

Theorem 1 *Let Z be a sample formed by m i.i.d. examples generated from some distribution $\mathcal{D}$ satisfying Assumptions 1 and 2. Let A_Z be the WFA returned by algorithm $\mathsf{HMC}_{\mathsf{p},\ell}+\mathsf{SM}$ with $p = 2$ and loss function $\ell(y, y') = |y - y'|$. Then, for any $\delta > 0$, the following holds with probability at least $1 - \delta$ for $f_Z = t_\nu \circ f_{A_Z}$:*

$$R(f_Z) \leq \widehat{R}_Z(f_Z) + O\left(\frac{\nu^4 |\mathcal{P}|^2 |\mathcal{S}|^{3/2}}{\tau \sigma^3 \rho \pi} \frac{\ln m}{m^{1/3}} \sqrt{\ln \frac{1}{\delta}}\right) \ .$$

1 Perturbation and stability tools

In this section, we list a series of known perturbation results for singular values, pseudo-inverses, and singular vectors, and other stability results needed for the proofs given in this appendix.

Lemma 2 ([4]) *Let $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{d_1 \times d_2}$. Then, for any $n \in [1, \min\{d_1, d_2\}]$, the following inequality holds: $|\sigma_n(\mathbf{A}) - \sigma_n(\mathbf{B})| \leq \|\mathbf{A} - \mathbf{B}\|$.*

Lemma 3 ([4]) *Let $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{d_1 \times d_2}$. Then the following upper bound holds for the norm of the difference of the pseudo-inverses of matrices $\mathbf{A}$ and $\mathbf{B}$:*

$$\|\mathbf{A}^+ - \mathbf{B}^+\| \leq \frac{1 + \sqrt{5}}{2} \max\left\{\|\mathbf{A}^+\|^2, \|\mathbf{B}^+\|^2\right\} \|\mathbf{A} - \mathbf{B}\|$$

Lemma 4 ([5]) *Let $\mathbf{A} \in \mathbb{R}^{d \times d}$ be symmetric positive semidefinite matrix and $\mathbf{E} \in \mathbb{R}^{d \times d}$ a symmetric matrix such that $\mathbf{B} = \mathbf{A} + \mathbf{E}$ is positive semidefinite. Fix $n \leq \operatorname{rank}(\mathbf{A})$ and suppose that $\|\mathbf{E}\|_F \leq (\lambda_n(\mathbf{A}) - \lambda_{n+1}(\mathbf{A}))/4$. Then, writing $\mathbf{V}_n$ for the top n eigenvectors of $\mathbf{A}$ and $\mathbf{W}_n$ for the top n eigenvectors of $\mathbf{B}$, we have*

$$\|\mathbf{V}_n - \mathbf{W}_n\|_F \leq \frac{4\|\mathbf{E}\|_F}{\lambda_n(\mathbf{A}) - \lambda_{n+1}(\mathbf{A})} \ . \tag{1}$$

This last lemma will be most useful to us in the form given in this next corollary.

Corollary 5 *Let* $\mathbf{A}, \mathbf{E} \in \mathbb{R}^{d_1 \times d_2}$ *and write* $\mathbf{B} = \mathbf{A} + \mathbf{E}$. *Suppose* $n \leq \operatorname{rank}(\mathbf{A})$ *and* $\|\mathbf{E}\|_F \leq \sqrt{\sigma_n(\mathbf{A})^2 - \sigma_{n+1}(\mathbf{A})^2}/4$. *If* $\mathbf{V}_n, \mathbf{W}_n$ *contain the first* n *right singular vectors of* $\mathbf{A}$ *and* $\mathbf{B}$ *respectively, then*

$$\|\mathbf{V}_n - \mathbf{W}_n\|_F \leq \frac{8\|\mathbf{A}\|_F\|\mathbf{E}\|_F + 4\|\mathbf{E}\|_F^2}{\sigma_n(\mathbf{A})^2 - \sigma_{n+1}(\mathbf{A})^2} \ .$$

Proof. Using that $\|\mathbf{A}^\top\mathbf{A} - \mathbf{B}^\top\mathbf{B}\|_F \leq 2\|\mathbf{A}\|_F\|\mathbf{E}\|_F + \|\mathbf{E}\|_F^2$ and $\lambda_n(\mathbf{A}^\top\mathbf{A}) = \sigma_n(\mathbf{A})^2$, we can apply Lemma 4 to get the bound on $\|\mathbf{V}_n - \mathbf{W}_n\|_F$ under the condition that $\|\mathbf{A}^\top\mathbf{A} - \mathbf{B}^\top\mathbf{B}\|_F \leq (\sigma_n(\mathbf{A})^2 - \sigma_{n+1}(\mathbf{A})^2)/4$. To see that this last condition is satisfied, observe that for all $x, y \geq 0$ one has $\sqrt{1+\sqrt{2}}\sqrt{x+y} \geq \sqrt{x} + \sqrt{y}$. Thus, we get

$$\begin{aligned}
\|\mathbf{E}\|_F &\leq \frac{\sqrt{\sigma_n(\mathbf{A})^2 - \sigma_{n+1}(\mathbf{A})^2}}{4} \\
&\leq \frac{\sqrt{\sigma_n(\mathbf{A})^2 - \sigma_{n+1}(\mathbf{A})^2} + \sqrt{4\|\mathbf{A}\|_F^2} - 2\|\mathbf{A}\|_F}{2\sqrt{1+\sqrt{2}}} \\
&\leq \frac{\sqrt{4\|\mathbf{A}\|_F^2 + \sigma_n(\mathbf{A})^2 - \sigma_{n+1}(\mathbf{A})^2} - 2\|\mathbf{A}\|_F}{2} \ ,
\end{aligned}$$

and this last inequality implies $2\|\mathbf{A}\|_F\|\mathbf{E}\|_F + \|\mathbf{E}\|_F^2 \leq (\sigma_n(\mathbf{A})^2 - \sigma_{n+1}(\mathbf{A})^2)/4$. □

The next two results give useful extensions of McDiarmid's inequality to deal with functions that do not satisfy the bounded difference assumption almost surely [2].

Definition 6 *Let* $X = (X_1, \ldots, X_m)$ *be a random variable on a probability space* Ω^m. *We say that a function* $\Phi \colon \Omega^m \to \mathbb{R}$ *is* strongly difference-bounded *by* (b, c, δ) *if the following holds: there exists a measurable subset* $E \subseteq \Omega^m$ *with* $\mathbb{P}[E] \leq \delta$, *such that*

- *if* X *and* X' *differ only by one coordinate and* $X \notin E$, *then* $|\Phi(X) - \Phi(X')| \leq c$;
- *for all* X, X' *that differ only by one coordinate* $|\Phi(X) - \Phi(X')| \leq b$.

Theorem 7 *Let* Φ *be a function over a probability space* Ω^m *that is strongly difference-bounded by* (b, c, δ) *with* $b \geq c > 0$. *Then, for any* $t > 0$,

$$\mathbb{P}\left[\Phi - \mathrm{E}[\Phi] \geq t\right] \leq \exp\left(\frac{-t^2}{8mc^2}\right) + \frac{mb\delta}{c} \ .$$

Furthermore, the same upper bound holds for $\mathbb{P}[\mathrm{E}[\Phi] - \Phi \geq t]$.

Corollary 8 *Let* Φ *be a function over a probability space* Ω^m *that is strongly difference-bounded by* $(b, \theta/m, \exp(-Km))$. *Then, for any* $0 < t \leq 2\theta\sqrt{K}$ *and* $m \geq \max\{b/\theta, (9+18/K)\ln(3+6/K)\}$,

$$\mathbb{P}\left[\Phi - \mathrm{E}[\Phi] \geq t\right] \leq 2\exp\left(\frac{-t^2 m}{8\theta^2}\right) \ .$$

Furthermore, the same upper bound holds for $\mathbb{P}[\mathrm{E}[\Phi] - \Phi \geq t]$.

The following is another useful form of the previous Corollary.

Corollary 9 *Let* Φ *be a function over a probability space* Ω^m *that is strongly difference-bounded by* $(b, \theta/m, \exp(-Km))$. *Then, for any* $\delta > 0$ *and any* $m \geq \max\{b/\theta, (9 + 18/K)\ln(3 + 6/K), (2/K)\ln(2/\delta)\}$, *each of the following holds with probability at least* $1 - \delta$:

$$\Phi \geq \mathrm{E}[\Phi] - \sqrt{\frac{8\theta^2}{m}\ln\left(\frac{2}{\delta}\right)} \ ,$$

$$\Phi \leq \mathrm{E}[\Phi] + \sqrt{\frac{8\theta^2}{m}\ln\left(\frac{2}{\delta}\right)} \ .$$

2 Proof of Theorem 1

To analyze the stability of our algorithm, we consider a sample $Z' = (z_1, \ldots, z_{m-1}, z'_m)$ that differs from Z only by the last point (z'_m instead of z_m). Example z'_m is an arbitrary point in the domain of $\mathcal{D}$. Throughout the analysis, $h = h_Z$ and $h' = h_{Z'}$ denote the functions in $\mathbb{H}$ obtained by solving (HMC-h) respectively with training samples Z and Z' respectively. We also denote by $\mathbf{H} = \mathbf{H}_Z$ and $\mathbf{H}' = \mathbf{H}_{Z'}$ their corresponding Hankel matrices.

The following technical lemma will be used to study the algorithmic stability of the optimization problem (HMC-h).

Lemma 10 *The following inequality holds for all samples Z and Z' differing by only one point:*

$$2\tau \|h - h'\|_2^2 \leq \widehat{R}_{\widetilde{Z}}(h') - \widehat{R}_{\widetilde{Z}}(h) + \widehat{R}_{\widetilde{Z}'}(h) - \widehat{R}_{\widetilde{Z}'}(h') \enspace .$$

Proof. The argument is the same as the one presented in [3] to bound the stability of kernel ridge regression. The following inequality is first shown using the expansion of $\|h - h'\|_2^2$ in terms of the corresponding inner product:

$$2\tau \|h - h'\|_2^2 \leq \tau (B_N(h' \| h) + B_N(h \| h')) \leq B_{F_Z}(h' \| h) + B_{F_{Z'}}(h \| h') \enspace ,$$

where B_F denotes the Bregman divergence associated to F. Next, using the optimality of h and h', which implies $\nabla F_Z(h) = 0$ and $\nabla F_{Z'}(h') = 0$, we can write $B_{F_Z}(h' \| h) + B_{F_{Z'}}(h \| h') = \widehat{R}_{\widetilde{Z}}(h') - \widehat{R}_{\widetilde{Z}}(h) + \widehat{R}_{\widetilde{Z}'}(h) - \widehat{R}_{\widetilde{Z}'}(h')$. □

Our next lemma bounds the stability of the first stage of the algorithm using Lemma 10.

Lemma 11 *Assume that $\mathcal{D}$ satisfies Assumption 1. Then, the following holds:*

$$\|\mathbf{H} - \mathbf{H}'\|_F \leq \min \left\{ 2\nu \sqrt{|\mathcal{P}||\mathcal{S}|}, \frac{1}{\tau \min\{\widetilde{m}, \widetilde{m}'\}} \right\} \enspace .$$

Proof. Note that by Assumption 1, for all (x, y) in $\widetilde{Z}$, or $\widetilde{Z}'$, we have $|y| \leq \nu$. Therefore, we must have $|\mathbf{H}(u, v)| \leq \nu$ for all $u \in \mathcal{P}$ and $v \in \mathcal{S}$, otherwise the value of $F_Z(\mathbf{H})$ is not minimal because decreasing the absolute value of an entry $|\mathbf{H}(u, v)| > \nu$ decreases the value of $F_Z(\mathbf{H})$. The same holds for $\mathbf{H}'$. Thus, the first bound follows from $\|\mathbf{H} - \mathbf{H}'\|_F \leq \|\mathbf{H}\|_F + \|\mathbf{H}'\|_F \leq 2\nu\sqrt{|\mathcal{P}||\mathcal{S}|}$.

Now we proceed to show the second bound. Since by definition $\|\mathbf{H} - \mathbf{H}'\|_F = \|h - h'\|_2$, it is sufficient to bound this second quantity. By Lemma 10, we have

$$2\tau \|h - h'\|_2^2 \leq \widehat{R}_{\widetilde{Z}}(h') - \widehat{R}_{\widetilde{Z}}(h) + \widehat{R}_{\widetilde{Z}'}(h) - \widehat{R}_{\widetilde{Z}'}(h') \enspace . \tag{2}$$

We can consider four different situations for the right-hand side of this expression, depending on the membership of x_m and x'_m in the set $\mathcal{PS}$.

If $x_m, x'_m \notin \mathcal{PS}$, then $\widetilde{Z} = \widetilde{Z}'$. Therefore, $\widehat{R}_{\widetilde{Z}}(h) = \widehat{R}_{\widetilde{Z}'}(h)$, $\widehat{R}_{\widetilde{Z}}(h') = \widehat{R}_{\widetilde{Z}'}(h')$, and $\|h - h'\|_2 = 0$.

If $x_m, x'_m \in \mathcal{PS}$, then $\widetilde{m} = \widetilde{m}'$, and the following equalities hold:

$$\widehat{R}_{\widetilde{Z}'}(h) - \widehat{R}_{\widetilde{Z}}(h) = \frac{|h(x'_m) - y'_m| - |h(x_m) - y_m|}{\widetilde{m}} \enspace ,$$
$$\widehat{R}_{\widetilde{Z}}(h') - \widehat{R}_{\widetilde{Z}'}(h') = \frac{|h'(x_m) - y_m| - |h'(x'_m) - y'_m|}{\widetilde{m}} \enspace .$$

Thus, in view of (2), we can write

$$2\tau \|h - h'\|_2^2 \leq \frac{|h(x_m) - h'(x_m)| + |h(x'_m) - h'(x'_m)|}{\widetilde{m}} \leq \frac{2}{\widetilde{m}} \|h - h'\|_2 \enspace ,$$

where the first inequality follows from $||h(x) - y| - |h'(x) - y|| \leq |h(x) - h'(x)|$, and the second from $|h(x) - h'(x)| \leq \|h - h'\|_2$.

If $x_m \in \mathcal{PS}$ and $x'_m \notin \mathcal{PS}$, the right-hand side of (2) equals

$$\sum_{z \in \widetilde{Z}'} \left(\frac{|h'(x)-y|}{\widetilde{m}} - \frac{|h'(x)-y|}{\widetilde{m}'} + \frac{|h(x)-y|}{\widetilde{m}'} - \frac{|h(x)-y|}{\widetilde{m}} \right) + \frac{|h'(x_m)-y_m|}{\widetilde{m}} - \frac{|h(x_m)-y_m|}{\widetilde{m}} \ .$$

Now, since $\widetilde{m} = \widetilde{m}' + 1$ we can write

$$2\tau \|h - h'\|_2^2 \leq \sum_{z \in \widetilde{Z}'} \frac{|h(x) - h'(x)|}{\widetilde{m}\,\widetilde{m}'} + \frac{|h(x_m) - h'(x_m)|}{\widetilde{m}} \leq \frac{2}{\widetilde{m}} \|h - h'\|_2 \ .$$

By symmetry, a similar bound holds in the case where $x_m \notin \mathcal{PS}$ and $x'_m \in \mathcal{PS}$. Combining these four bounds yields the desired inequality. □

The next three lemmas contain the main technical tools needed to bound the difference $|f_{A_Z}(x) - f_{A_{Z'}}(x)|$ in our agnostic setting.

Lemma 12 *Let $A = \langle \boldsymbol{\alpha}, \boldsymbol{\beta}, \{\mathbf{A}_a\} \rangle$ and $A' = \langle \boldsymbol{\alpha}', \boldsymbol{\beta}', \{\mathbf{A}'_a\} \rangle$ be two weighted automata with n states. Let γ be such that both A and A' are γ-bounded. Then, the following inequality holds for any string $x \in \Sigma^\star$:*

$$|f_A(x) - f_{A'}(x)| \leq \gamma^{|x|+1} \left(\|\boldsymbol{\alpha} - \boldsymbol{\alpha}'\| + \|\boldsymbol{\beta} - \boldsymbol{\beta}'\| + \sum_{i=1}^{|x|} \|\mathbf{A}_{x_i} - \mathbf{A}'_{x_i}\| \right) \ .$$

Proof. Follows by induction on $|x|$ using techniques similar to those used to prove Lemmas 11 and 12 in [1]. □

Lemma 13 *Let $\gamma = \nu \sqrt{|\mathcal{P}||\mathcal{S}|} / \sigma_n(\mathbf{H}_\epsilon)$. The weighted automaton A_Z is γ-bounded.*

Proof. Since $\|\mathbf{H}_a\| \leq \|\mathbf{H}_a\|_F \leq \nu\sqrt{|\mathcal{P}||\mathcal{S}|}$, simple calculations show that $\|\boldsymbol{\alpha}^\top\| \leq \nu\sqrt{|\mathcal{S}|}$, $\|\boldsymbol{\beta}\| \leq \nu\sqrt{|\mathcal{P}|}/\sigma_n(\mathbf{H}_\epsilon)$, and $\|\mathbf{A}_a\| \leq \nu\sqrt{|\mathcal{P}||\mathcal{S}|}/\sigma_n(\mathbf{H}_\epsilon)$. □

Let us define the following quantities in terms of the vectors and matrices that define A and A':

$$\begin{aligned}
\varepsilon_\epsilon &= \|\mathbf{H}_\epsilon - \mathbf{H}'_\epsilon\| \ , \\
\varepsilon_a &= \|\mathbf{H}_a - \mathbf{H}'_a\| \ , \\
\varepsilon_V &= \|\mathbf{V} - \mathbf{V}'\| \ , \\
\varepsilon_{\mathcal{S}} &= \|\mathbf{h}_{\lambda,\mathcal{S}} - \mathbf{h}'_{\lambda,\mathcal{S}}\| \ , \\
\varepsilon_{\mathcal{P}} &= \|\mathbf{h}_{\mathcal{P},\lambda} - \mathbf{h}'_{\mathcal{P},\lambda}\| \ .
\end{aligned}$$

Now we state a result that will be used in the proof of Lemma 15.

Lemma 14 *The following three bounds hold:*

$$\begin{aligned}
\|\mathbf{A}_a - \mathbf{A}'_a\| &\leq \frac{\varepsilon_a + \varepsilon_V \|\mathbf{H}'_a\|}{\sigma_n(\mathbf{H}_\epsilon \mathbf{V})} + \frac{1+\sqrt{5}}{2} \frac{\|\mathbf{H}'_a\|(\varepsilon_\epsilon + \varepsilon_V \|\mathbf{H}'_\epsilon\|)}{\min\{\sigma_n(\mathbf{H}_\epsilon \mathbf{V})^2, \sigma_n(\mathbf{H}'_\epsilon \mathbf{V}')^2\}} \ , \\
\|\boldsymbol{\alpha} - \boldsymbol{\alpha}'\| &\leq \varepsilon_{\mathcal{S}} + \varepsilon_V \|\mathbf{h}_{\lambda,\mathcal{S}}\| \ , \\
\|\boldsymbol{\beta} - \boldsymbol{\beta}'\| &\leq \frac{\varepsilon_{\mathcal{P}}}{\sigma_n(\mathbf{H}_\epsilon \mathbf{V})} + \frac{1+\sqrt{5}}{2} \frac{\|\mathbf{h}'_{\mathcal{P},\lambda}\|(\varepsilon_\epsilon + \varepsilon_V \|\mathbf{H}'_\epsilon\|)}{\min\{\sigma_n(\mathbf{H}_\epsilon \mathbf{V})^2, \sigma_n(\mathbf{H}'_\epsilon \mathbf{V}')^2\}} \ .
\end{aligned}$$

Proof. Using the triangle inequality, the submultiplicativity of the operator norm, and the properties of the pseudo-inverse, we can write

$$\begin{aligned}
\|\mathbf{A}_a - \mathbf{A}'_a\| &= \|(\mathbf{H}_\epsilon \mathbf{V})^+ (\mathbf{H}_a \mathbf{V} - \mathbf{H}'_a \mathbf{V}') + ((\mathbf{H}'_\epsilon \mathbf{V}')^+ - (\mathbf{H}_\epsilon \mathbf{V})^+)\mathbf{H}'_a \mathbf{V}'\| \\
&\leq \|(\mathbf{H}_\epsilon \mathbf{V})^+\| \|\mathbf{H}_a \mathbf{V} - \mathbf{H}'_a \mathbf{V}'\| + \|(\mathbf{H}_\epsilon \mathbf{V})^+ - (\mathbf{H}'_\epsilon \mathbf{V}')^+\| \|\mathbf{H}'_a \mathbf{V}'\| \\
&\leq \sigma_n(\mathbf{H}_\epsilon \mathbf{V})^{-1} \|\mathbf{H}_a \mathbf{V} - \mathbf{H}'_a \mathbf{V}'\| + \|\mathbf{H}'_a\| \|(\mathbf{H}_\epsilon \mathbf{V})^+ - (\mathbf{H}'_\epsilon \mathbf{V}')^+\| \ ,
\end{aligned}$$

where we used that $\|(\mathbf{H}_\epsilon \mathbf{V})^+\| = \sigma_n(\mathbf{H}_\epsilon \mathbf{V})$ by the properties of pseudo-inverse and operator norm, and $\|\mathbf{H}'_a \mathbf{V}'\| \leq \|\mathbf{H}'_a\|$ by sub-multiplactivity and $\|\mathbf{V}'\| = 1$. Now note that we also have

$$\|\mathbf{H}_a \mathbf{V} - \mathbf{H}'_a \mathbf{V}'\| \leq \|\mathbf{V}\| \|\mathbf{H}_a - \mathbf{H}'_a\| + \|\mathbf{H}'_a\| \|\mathbf{V} - \mathbf{V}'\| \leq \varepsilon_a + \varepsilon_V \|\mathbf{H}'_a\| \ .$$

Furthermore, using Lemma 3 we obtain

$$\begin{aligned}
\|(\mathbf{H}_\epsilon \mathbf{V})^+ - (\mathbf{H}'_\epsilon \mathbf{V}')^+\| &\leq \frac{1+\sqrt{5}}{2} \|\mathbf{H}_\epsilon \mathbf{V} - \mathbf{H}'_\epsilon \mathbf{V}'\| \max\{\|(\mathbf{H}_\epsilon \mathbf{V})^+\|^2, \|(\mathbf{H}'_\epsilon \mathbf{V}')^+\|^2\} \\
&\leq \frac{1+\sqrt{5}}{2} \frac{\|\mathbf{H}_\epsilon - \mathbf{H}_\epsilon\| \|\mathbf{V}\| + \|\mathbf{H}'_\epsilon\| \|\mathbf{V} - \mathbf{V}'\|}{\min\{\sigma_n(\mathbf{H}_\epsilon \mathbf{V})^2, \sigma_n(\mathbf{H}'_\epsilon \mathbf{V}')^2\}} \\
&= \frac{1+\sqrt{5}}{2} \frac{\varepsilon_\epsilon + \varepsilon_V \|\mathbf{H}'_\epsilon\|}{\min\{\sigma_n(\mathbf{H}_\epsilon \mathbf{V})^2, \sigma_n(\mathbf{H}'_\epsilon \mathbf{V}')^2\}} \ .
\end{aligned}$$

Thus we get the first of the bounds. The second bound follows straightforwardly from

$$\|\mathbf{V}^\top \mathbf{h}_{\lambda,\mathcal{S}} - \mathbf{V}'^\top \mathbf{h}'_{\lambda,\mathcal{S}}\| \leq \|\mathbf{V}^\top - \mathbf{V}'^\top\| \|\mathbf{h}_{\lambda,\mathcal{S}}\| + \|\mathbf{V}'^\top\| \|\mathbf{h}_{\lambda,\mathcal{S}} - \mathbf{h}'_{\lambda,\mathcal{S}}\| = \varepsilon_\mathcal{S} + \varepsilon_V \|\mathbf{h}_{\lambda,\mathcal{S}}\| \ ,$$

which uses that $\|\mathbf{M}^\top\| = \|\mathbf{M}\|$ holds for the operator norm.

Finally, the last bound follows from the following inequalities, where we use Lemma 3 again:

$$\begin{aligned}
\|\boldsymbol{\beta} - \boldsymbol{\beta}'\| &\leq \|(\mathbf{H}_\epsilon \mathbf{V})^+\| \|\mathbf{h}_{\mathcal{P},\lambda} - \mathbf{h}'_{\mathcal{P},\lambda}\| + \|\mathbf{h}'_{\mathcal{P},\lambda}\| \|(\mathbf{H}_\epsilon \mathbf{V})^+ - (\mathbf{H}'_\epsilon \mathbf{V}')^+\| \\
&\leq \frac{\|\mathbf{h}_{\mathcal{P},\lambda} - \mathbf{h}'_{\mathcal{P},\lambda}\|}{\sigma_n(\mathbf{H}_\epsilon \mathbf{V})} + \frac{1+\sqrt{5}}{2} \frac{\|\mathbf{h}'_{\mathcal{P},\lambda}\| \|\mathbf{H}_\epsilon \mathbf{V} - \mathbf{H}'_\epsilon \mathbf{V}'\|}{\min\{\sigma_n(\mathbf{H}_\epsilon \mathbf{V})^2, \sigma_n(\mathbf{H}'_\epsilon \mathbf{V}')^2\}} \\
&\leq \frac{\varepsilon_\mathcal{P}}{\sigma_n(\mathbf{H}_\epsilon \mathbf{V})} + \frac{1+\sqrt{5}}{2} \frac{\|\mathbf{h}'_{\mathcal{P},\lambda}\| (\varepsilon_\epsilon + \varepsilon_V \|\mathbf{H}'_\epsilon\|)}{\min\{\sigma_n(\mathbf{H}_\epsilon \mathbf{V})^2, \sigma_n(\mathbf{H}'_\epsilon \mathbf{V}')^2\}} \ .
\end{aligned}$$

□

Lemma 15 *Let* $\varepsilon = \|\mathbf{H} - \mathbf{H}'\|_F$*,* $\widehat{\sigma} = \min\{\sigma_n(\mathbf{H}_\epsilon), \sigma_n(\mathbf{H}'_\epsilon)\}$*, and* $\widehat{\rho} = \sigma_n(\mathbf{H}_\epsilon)^2 - \sigma_{n+1}(\mathbf{H}_\epsilon)^2$*. Suppose* $\varepsilon \leq \sqrt{\widehat{\rho}}/4$*. There exists a universal constant* $c_1 > 0$ *such that the following inequalities hold for all* $a \in \Sigma$*:*

$$\begin{aligned}
\|\mathbf{A}_a - \mathbf{A}'_a\| &\leq c_1 \frac{\varepsilon \nu^3 |\mathcal{P}|^{3/2} |\mathcal{S}|^{1/2}}{\widehat{\rho} \widehat{\sigma}^2} \ , \\
\|\boldsymbol{\alpha} - \boldsymbol{\alpha}'\| &\leq c_1 \frac{\varepsilon \nu^2 |\mathcal{P}|^{1/2} |\mathcal{S}|}{\widehat{\rho}} \ , \\
\|\boldsymbol{\beta} - \boldsymbol{\beta}'\| &\leq c_1 \frac{\varepsilon \nu^3 |\mathcal{P}|^{3/2} |\mathcal{S}|^{1/2}}{\widehat{\rho} \widehat{\sigma}^2} \ .
\end{aligned}$$

Proof. We begin with a few observations that will help us apply Lemma 14. First note that $\|\mathbf{H}_a - \mathbf{H}'_a\| \leq \|\mathbf{H}_a - \mathbf{H}'_a\|_F \leq \varepsilon$ for all $a \in \Sigma'$, as well as $\|\mathbf{h}_{\mathcal{P},\lambda} - \mathbf{h}'_{\mathcal{P},\lambda}\| \leq \varepsilon$ and $\|\mathbf{h}_{\lambda,\mathcal{S}} - \mathbf{h}'_{\lambda,\mathcal{S}}\| \leq \varepsilon$. Furthermore, $\|\mathbf{H}_a\| \leq \|\mathbf{H}_a\|_F \leq \nu \sqrt{|\mathcal{P}||\mathcal{S}|}$ and $\|\mathbf{H}'_a\| \leq \nu \sqrt{|\mathcal{P}||\mathcal{S}|}$ for all $a \in \Sigma'$. In addition, we have $\|\mathbf{h}_{\lambda,\mathcal{S}}\| \leq \nu \sqrt{|\mathcal{S}|}$ and $\|\mathbf{h}'_{\mathcal{P},\lambda}\| \leq \nu \sqrt{|\mathcal{P}|}$. Finally, by construction we also have $\sigma_n(\mathbf{H}_\epsilon \mathbf{V}) = \sigma_n(\mathbf{H}_\epsilon)$ and $\sigma_n(\mathbf{H}'_\epsilon \mathbf{V}') = \sigma_n(\mathbf{H}'_\epsilon)$. Therefore, it only remains to bound $\|\mathbf{V} - \mathbf{V}'\|$, which by Corollary 5 is

$$\|\mathbf{V} - \mathbf{V}'\| \leq \frac{4\varepsilon}{\widehat{\rho}} (2\nu \sqrt{|\mathcal{P}||\mathcal{S}|} + \varepsilon) \leq \frac{16 \varepsilon \nu \sqrt{|\mathcal{P}||\mathcal{S}|}}{\widehat{\rho}} \ ,$$

where the last inequality follows from Lemma 11.

Plugging all the bounds above in Lemma 14 yields the following inequalities:

$$\begin{aligned}
\|\mathbf{A}_a - \mathbf{A}'_a\| &\leq \frac{\varepsilon}{\widehat{\sigma}} \left(1 + \frac{16 \nu |\mathcal{P}|^{1/2} |\mathcal{S}|^{1/2}}{\widehat{\rho}}\right) + \frac{1+\sqrt{5}}{2} \frac{\varepsilon \nu |\mathcal{P}|^{1/2} |\mathcal{S}|^{1/2}}{\widehat{\sigma}^2} \left(1 + \frac{16 \nu^2 |\mathcal{P}||\mathcal{S}|}{\widehat{\rho}}\right) \ , \\
\|\boldsymbol{\alpha} - \boldsymbol{\alpha}'\| &\leq \varepsilon \left(1 + \frac{16 \nu^2 |\mathcal{P}|^{1/2} |\mathcal{S}|}{\widehat{\rho}}\right) \ , \\
\|\boldsymbol{\beta} - \boldsymbol{\beta}'\| &\leq \frac{\varepsilon}{\widehat{\sigma}} + \frac{1+\sqrt{5}}{2} \frac{\varepsilon \nu |\mathcal{P}|^{1/2}}{\widehat{\sigma}^2} \left(1 + \frac{16 \nu^2 |\mathcal{P}||\mathcal{S}|}{\widehat{\rho}}\right) \ .
\end{aligned}$$

The result now follows from an adequate choice of c_1. $\square$

We now define the properties that make Z a good sample and show that for large enough m they are satisfied with high probability.

Definition 16 *We say that a sample Z of m i.i.d. examples from $\mathcal{D}$ is* good *if the following conditions are satisfied for any $z'_m = (x'_m, y'_m) \in \mathrm{supp}(\mathcal{D})$:*

- $|x_i| \leq ((1/c)\ln(4m^4))^{1/(1+\eta)}$ *for all* $1 \leq i \leq m$*;*
- $\|\mathbf{H} - \mathbf{H}'\|_F \leq 4/(\tau\pi m)$*;*
- $\min\{\sigma_n(\mathbf{H}_\epsilon), \sigma_n(\mathbf{H}'_\epsilon)\} \geq \sigma/2$*;*
- $\sigma_n(\mathbf{H}_\epsilon)^2 - \sigma_{n+1}(\mathbf{H}_\epsilon)^2 \geq \rho/2$.

Lemma 17 *Suppose $\mathcal{D}$ satisfies Assumptions 1 and 2. There exists a quantity $M = \mathsf{poly}(\nu, \pi, \sigma, \rho, \tau, |\mathcal{P}|, |\mathcal{S}|)$ such that if $m \geq M$, then Z is good with probability at least $1 - 1/m^3$.*

Proof. First note that by Assumption 2, writing $L = ((1/c)\ln(4m^4))^{1/(1+\eta)}$ a union bound yields

$$\mathbb{P}\left[\bigvee_{i=1}^{m} |x_i| > L\right] \leq m\exp(-cL^{1+\eta}) = \frac{1}{4m^3} \quad .$$

Now let $\bar{m} = (x_1, \ldots, x_{m-1}) \cap (\mathcal{PS})$. Note that we have $\min\{\widetilde{m}, \widetilde{m}'\} \geq \bar{m}$ and $\mathrm{E}_Z[\bar{m}] = \pi(m-1)$. Thus, for any $\Delta \in (0,1)$ the Chernoff bound gives

$$\mathbb{P}[\bar{m} < \pi(m-1)(1-\Delta)] \leq \exp\left(-\frac{(m-1)\pi\Delta^2}{2}\right) \leq \exp\left(-\frac{m\pi\Delta^2}{4}\right) \quad ,$$

where we have used that $(m-1)/m \geq 1/2$ for $m \geq 2$.

Taking $\Delta = \sqrt{(4/m\pi)\ln(4m^3)}$ above we see that $\min\{\widetilde{m}, \widetilde{m}'\} \geq (m-1)\pi(1-\Delta) \geq m\pi(1-\Delta)/2$ holds with probability at least $1 - 1/(4m^3)$. Now note that $m \geq (16/\pi)\ln(4m^3)$ implies $\Delta \leq 1/2$. Therefore, by Lemma 11 we have that $m \geq \max\{2, (16/\pi)\ln(4m^3), 2/(\tau\pi\nu\sqrt{|\mathcal{P}||\mathcal{S}|})\}$ implies that $\|\mathbf{H} - \mathbf{H}'\|_F \leq 4/(\tau\pi m)$ holds with probability at least $1 - 1/(4m^3)$.

For the third claim note that by Lemma 2 we have $|\sigma_n(\mathbf{H}_\epsilon) - \sigma_n(\mathbf{H}'_\epsilon)| \leq \|\mathbf{H}_\epsilon - \mathbf{H}'_\epsilon\|_F \leq \|\mathbf{H} - \mathbf{H}'\|_F$. Thus, from the argument we just used in the previous bound we can see that when $m \geq 2$ the function $\Phi(Z) = \sigma_n(\mathbf{H}_\epsilon)$ is strongly difference-bounded by $(b_\sigma, \theta_\sigma/m, \exp(-K_\sigma m))$ with $b_\sigma = 2\nu\sqrt{|\mathcal{P}||\mathcal{S}|}$, $\theta_\sigma = 2/(\tau\pi(1-\Delta))$, and $K_\sigma = \pi\Delta^2/4$ for any $\Delta \in (0,1)$. Now note that by Lemma 2 and the previous goodness condition on $\|\mathbf{H} - \mathbf{H}'\|_F$ we have $\min\{\sigma_n(\mathbf{H}_\epsilon), \sigma_n(\mathbf{H}'_\epsilon)\} \geq \sigma_n(\mathbf{H}_\epsilon) - \|\mathbf{H} - \mathbf{H}'\|_F \geq \sigma_n(\mathbf{H}_\epsilon) - 4/(\nu\pi m)$. Furthermore, taking $\Delta = 1/2$ and assuming that

$$m \geq \max\left\{\frac{\nu\tau\pi\sqrt{|\mathcal{P}||\mathcal{S}|}}{2}, \left(9 + \frac{288}{\pi}\right)\ln\left(3 + \frac{96}{\pi}\right), \frac{32}{\pi}\ln(8m^3)\right\} \quad ,$$

we can apply Corollary 9 with $\delta = 1/(4m^3)$ to see that

$$\sigma_n(\mathbf{H}_\epsilon) - \frac{4}{\nu\pi m} \geq \sigma - \sqrt{\frac{128}{\tau^2\pi^2 m}\ln(8m^3)} - \frac{4}{\nu\pi m}$$

holds with probability at least $1 - 1/(4m^3)$. Hence, for any sample size such that $m \geq \max\{16/(\nu\pi\sigma), (2048/\tau^2\pi^2\sigma^2)\ln(8m^3)\}$, we get

$$\min\{\sigma_n(\mathbf{H}_\epsilon), \sigma_n(\mathbf{H}'_\epsilon)\} \geq \sigma - \sqrt{\frac{128}{\tau^2\pi^2 m}\ln(8m^3)} - \frac{4}{\nu\pi m} \geq \sigma - \frac{\sigma}{4} - \frac{\sigma}{4} = \frac{\sigma}{2} \quad .$$

To prove the fourth bound we shall study the stability of $\Phi(Z) = \sigma_n(\mathbf{H}_\epsilon)^2 - \sigma_{n+1}(\mathbf{H}_\epsilon)^2$. We begin with the following chain of inequalities, which follows from Lemma 2 and $\sigma_n(\mathbf{H}_\epsilon) \geq \sigma_{n+1}(\mathbf{H}_\epsilon)$:

$$\begin{aligned}
|\Phi(Z) - \Phi(Z')| &= \left|(\sigma_n(\mathbf{H}_\epsilon)^2 - \sigma_{n+1}(\mathbf{H}_\epsilon)^2) - (\sigma_n(\mathbf{H}'_\epsilon)^2 - \sigma_{n+1}(\mathbf{H}'_\epsilon)^2)\right| \\
&\leq |\sigma_n(\mathbf{H}_\epsilon)^2 - \sigma_n(\mathbf{H}'_\epsilon)^2| + |\sigma_{n+1}(\mathbf{H}_\epsilon)^2 - \sigma_{n+1}(\mathbf{H}'_\epsilon)^2| \\
&= |\sigma_n(\mathbf{H}_\epsilon) + \sigma_n(\mathbf{H}'_\epsilon)||\sigma_n(\mathbf{H}_\epsilon) - \sigma_n(\mathbf{H}'_\epsilon)| + |\sigma_{n+1}(\mathbf{H}_\epsilon) + \sigma_{n+1}(\mathbf{H}'_\epsilon)||\sigma_{n+1}(\mathbf{H}_\epsilon) - \sigma_{n+1}(\mathbf{H}'_\epsilon)| \\
&\leq (2\sigma_n(\mathbf{H}_\epsilon) + \|\mathbf{H}_\epsilon - \mathbf{H}'_\epsilon\|) \|\mathbf{H}_\epsilon - \mathbf{H}'_\epsilon\| + (2\sigma_{n+1}(\mathbf{H}_\epsilon) + \|\mathbf{H}_\epsilon - \mathbf{H}'_\epsilon\|) \|\mathbf{H}_\epsilon - \mathbf{H}'_\epsilon\| \\
&\leq 4\sigma_n(\mathbf{H}_\epsilon)\|\mathbf{H} - \mathbf{H}'\|_F + 2\|\mathbf{H} - \mathbf{H}'\|_F^2 \ .
\end{aligned}$$

Now we can use this last bound to show that $\Phi(Z)$ is strongly difference-bounded by $(b_\rho, \theta_\rho/m, \exp(-K_\rho m))$ with the definitions: $b_\rho = 16\nu^2|\mathcal{P}||\mathcal{S}|$, $\theta_\rho = 64\sigma/(\tau\pi)$ and $K_\rho = \min\{\sigma^2\tau^2\pi^2/256, \pi/64\}$. For b_ρ just observe that from Lemma 11 and $\sigma_n(\mathbf{H}_\sigma) \leq \|\mathbf{H}_\sigma\|_F \leq \nu\sqrt{|\mathcal{P}||\mathcal{S}|}$ we get

$$4\sigma_n(\mathbf{H}_\epsilon)\|\mathbf{H} - \mathbf{H}'\|_F + 2\|\mathbf{H} - \mathbf{H}'\|_F^2 \leq 16\nu^2|\mathcal{P}||\mathcal{S}| \ .$$

By the same arguments used above, if m is large enough we have $\|\mathbf{H} - \mathbf{H}'\|_F \leq 4/(\tau\pi m)$ with probability at least $1 - \exp(-m\pi/16)$. Furthermore, by taking $\Delta = 1/2$ in the stability argument given above for $\sigma_n(\mathbf{H}_\epsilon)$, and invoking Corollary 9 with $\delta = 2\exp(-Km)$ for some $0 < K \leq K_\sigma/2 = \pi/32$, we get

$$\sigma_n(\mathbf{H}_\epsilon) \leq \sigma + \sqrt{\frac{128K}{\tau^2\pi^2}} \ ,$$

with probability at least $1 - 2\exp(-Km)$. Thus, taking $K = \min\{\pi/32, \sigma^2\tau^2\pi^2/128\}$ we get $\sigma_n(\mathbf{H}_\epsilon) \leq 2\sigma$. If we now combine the bounds for $\|\mathbf{H} - \mathbf{H}'\|_F$ and $\sigma_n(\mathbf{H}_\epsilon)$, we get

$$4\sigma_n(\mathbf{H}_\epsilon)\|\mathbf{H} - \mathbf{H}'\|_F + 2\|\mathbf{H} - \mathbf{H}'\|_F^2 \leq \frac{32\sigma}{\tau\pi m} + \frac{32}{\tau^2\pi^2 m^2} \leq \frac{64\sigma}{\tau\pi m} = \frac{\theta_\rho}{m} \ ,$$

where have assumed that $m \geq 1/(\tau\pi\sigma)$. To get K_ρ note that the above bound holds with probability at least

$$1 - e^{-m\pi/16} - 2e^{-Km} \geq 1 - 3e^{-Km} \geq 1 - e^{-Km/2} = 1 - e^{-K_\rho m} \ ,$$

where we have used that $K \leq \pi/16$ and assumed that $m \geq 2\ln(3)/K$. Finally, applying Corollary 9 to $\Phi(Z)$ we see that with probability at least $1 - 1/(4m^3)$ one has

$$\sigma_n(\mathbf{H}_\epsilon)^2 - \sigma_{n+1}(\mathbf{H}_\epsilon)^2 \geq \rho - \sqrt{\frac{2^{15}\sigma^2}{\tau^2\pi^2 m}\ln(8m^3)} \geq \frac{\rho}{2} \ ,$$

whenever $m \geq \max\{(2^{17}\sigma^2/\tau^2\pi^2\rho^2)\ln(8m^3), \nu^2\tau\pi|\mathcal{P}||\mathcal{S}|/(4\sigma), (9 + 18/K_\rho)\ln(3 + 6/K_\rho), (2/K_\rho)\ln(8m^3)\}$. $\square$

We can now analyze how the change of one sample point in Z can affect the difference $R(f_Z) - \widehat{R}_Z(f_Z)$. Our main result will be obtained by applying Theorem 7 to this difference.

Lemma 18 *Let* $\gamma_1 = 64\nu^4|\mathcal{P}|^2|\mathcal{S}|^{3/2}/(\tau\sigma^3\rho\pi)$ *and* $\gamma_2 = 2\nu|\mathcal{P}|^{1/2}|\mathcal{S}|^{1/2}/\sigma$. *If* $m \geq \max\{M, 16\sqrt{2}/(\tau\pi\sqrt{\rho}), \exp(6\ln\gamma_2(1.2c\ln\gamma_2)^{1/\eta})\}$, *then the function* $\Phi(Z) = R(f_Z) - \widehat{R}_Z(f_Z)$ *is strongly difference-bounded by* $(4\nu + 2\nu/m, c_2\gamma_1 m^{-5/6}\ln m, 1/m^3)$ *for some constant* $c_2 > 0$.

Proof. We will write for short $f = f_Z$ and $f' = f_{Z'}$. Let $\beta_1 = \mathrm{E}_{x\sim\mathcal{D}_\Sigma}[|f(x) - f'(x)|]$ and $\beta_2 = \max_{1\leq i\leq m-1}|f(x_i) - f'(x_i)|$. We first show that $|\Phi(Z) - \Phi(Z')| \leq \beta_1 + \beta_2 + 2\nu/m$. By definition of Φ we can write

$$|\Phi(Z) - \Phi(Z')| \leq |R(f) - R(f')| + |\widehat{R}_Z(f) - \widehat{R}_{Z'}(f')| \ .$$

By Jensen's inequality, the first term can be upper bounded by $\mathrm{E}_{(x,y)\sim\mathcal{D}}[||f(x)-y|-|f'(x)-y||] \leq \beta_1$. Now, using the triangle inequality and $|f(x_m) - y_m|, |f'(x'_m) - y'_m| \leq 2\nu$, the second term can be bounded as follows:

$$|\widehat{R}_Z(f) - \widehat{R}_{Z'}(f')| \leq \frac{2\nu}{m} + \frac{1}{m}\sum_{i=1}^{m-1}|f(x_i) - f'(x_i)| \leq \frac{2\nu}{m} + \beta_2\frac{m-1}{m} \ .$$

Observe that for any samples Z and Z' we have $\beta_1, \beta_2 \leq 2\nu$. This provides an almost-sure upper bound needed in the definition of strongly difference-boundedness. We use this bound when the sample Z is not good. By Lemma 17, when m is large enough this event will occur with probability at most $1/m^3$.

It remains to bound β_1 and β_2 assuming that Z is good. Note that by Lemma 17, $m \geq \max\{M, 16\sqrt{2}/(\tau\pi\sqrt{\rho})\}$ implies $\|\mathbf{H} - \mathbf{H}'\|_F \leq \sqrt{\rho}/4$. Thus, by combining Lemmas 12, 13, 15, and 17, we see that the following holds for any $x \in \Sigma^\star$:

$$|f(x) - f'(x)| \leq \left(\frac{2\nu|\mathcal{P}|^{1/2}|\mathcal{S}|^{1/2}}{\sigma}\right)^{|x|+1} \frac{32c_1(|x|+2)\nu^3|\mathcal{P}|^{3/2}|\mathcal{S}|}{m\tau\pi\sigma^2\rho}$$
$$= \frac{c_1\gamma_1}{m}\exp(|x|\ln\gamma_2 + \ln(|x|+2)) \ .$$

In particular, for $|x| \leq L = ((1/c)\ln(4m^4))^{1/(1+\eta)}$ and $m \geq \exp(6\ln\gamma_2(1.2c\ln\gamma_2)^{1/\eta})$, a simple calculation shows that $|f(x) - f'(x)| \leq C\gamma_1 m^{-5/6}\ln m$ for some constant C. Thus, we can write

$$\beta_1 \leq \mathop{\mathrm{E}}_{x\sim\mathcal{D}_\Sigma}[|f(x) - f'(x)| \mid |x| \leq L] + 2\nu\mathbb{P}_{x\sim\mathcal{D}_\Sigma}[|x| \geq L] \leq C\gamma_1 m^{-5/6}\ln m + \nu/2m^3$$

and $\beta_2 \leq C\gamma_1 m^{-5/6}\ln m$, where the last bound follows from the goodness of Z. Combining these bounds yields the desired result. □

The following is the proof of our main result.

Proof.[of Theorem 1] The result follows from an application of Theorem 7 to $\Phi(Z)$, defined as in Lemma 18. In particular, for large enough m, the following holds with probability at least $1-\delta$:

$$R(f_Z) \leq \widehat{R}_Z(f_Z) + \mathop{\mathrm{E}}_{Z\sim\mathcal{D}^m}[\Phi(Z)] + \sqrt{C\gamma_1^2\frac{\ln^2 m}{m^{2/3}}\ln\left(\frac{1}{\delta - \frac{6\nu}{C'\gamma_1}\frac{1}{m^{7/6}\ln m}}\right)} \ ,$$

for some constants C, C' and $\gamma_1 = \nu^4|\mathcal{P}|^2|\mathcal{S}|^{3/2}/\tau\sigma^3\rho\pi$. Thus, it remains to bound $\mathrm{E}_{Z\sim\mathcal{D}^m}[\Phi(Z)]$.

First note that we have $\mathrm{E}_{Z\sim\mathcal{D}^m}[R(f_Z)] = \mathrm{E}_{Z,z\sim\mathcal{D}^{m+1}}[|f_Z(x) - y|]$. On the other hand, we can also write $\mathrm{E}_{Z\sim\mathcal{D}^m}[\widehat{R}_Z(f_Z)] = \mathrm{E}_{Z,z\sim\mathcal{D}^{m+1}}[|f_{Z'}(x) - y|]$, where Z' is a sample of size m containing z and $m-1$ other points in Z chosen at random. Thus, by Jensen's inequality we can write

$$|\mathop{\mathrm{E}}_{Z\sim\mathcal{D}^m}[\Phi(Z)]| \leq \mathop{\mathrm{E}}_{Z,z\sim\mathcal{D}^{m+1}}[|f_Z(x) - f_{Z'}(x)|] \ .$$

Now an argument similar to the one used in Lemma 18 for bounding β_1 can be used to show that, for large enough m, the following inequality holds:

$$\left|\mathop{\mathrm{E}}_{Z\sim\mathcal{D}^m}[\Phi(Z)]\right| \leq C\gamma_1\frac{\ln m}{m^{5/6}} + \frac{2\nu}{m^3} \ ,$$

which completes the proof. □

References

[1] D. Hsu, S. M. Kakade, and T. Zhang. A spectral algorithm for learning hidden Markov models. *COLT*, 2009.

[2] S. Kutin. Extensions to McDiarmid's inequality when differences are bounded with high probability. Technical report, TR-2002-04, University of Chicago, 2002.

[3] M. Mohri, A. Rostamizadeh, and A. Talwalkar. *Foundations of Machine Learning*. The MIT Press, 2012.

[4] G.W. Stewart and J. Sun. *Matrix perturbation theory*. Academic press New York, 1990.

[5] L. Zwald and G. Blanchard. On the convergence of eigenspaces in kernel principal component analysis. *NIPS*, 2006.