Supplementary Material
A Simple Algorithm for Semi-supervised Learning with Improved Generalization Error Bound

Your Name EMAIL@YOURDOMAIN.EDU

Your Fantastic Institute, 314159 Pi St., Palo Alto, CA 94306 USA

Your CoAuthor's Name EMAIL@COAUTHORDOMAIN.EDU

Their Fantastic Institute, 27182 Exp St., Toronto, ON M6H 2T1 CANADA

Proof of Proposition 1

First we note that

$$\begin{aligned}
\mathrm{E}_{\mathbf{x}}\left[(\widehat{g}(\mathbf{x}) - g_s(\mathbf{x}))^2\right] &\leq \mathrm{E}_{\mathbf{x}}\left[\left(\sum_{i=1}^{s}(\gamma_i^* - \alpha_i)\phi_i(\mathbf{x})\right)^2\right] \\
= \quad & \mathrm{E}_{\mathbf{x}}\left[\sum_{i=1}^{s}(\gamma_i^* - \alpha_i)^2\phi_i^2(\mathbf{x})\right] + \sum_{i\neq j}(\gamma_i^* - \alpha_i)(\gamma_j^* - \alpha_j)\mathrm{E}_{\mathbf{x}}\left[\phi_i(\mathbf{x})\phi_j(\mathbf{x})\right] \\
= \quad & \|\gamma^* - \alpha^s\|_2^2
\end{aligned}$$

Second, since γ^* is the optimal solution to $\mathcal{L}(\gamma)$, we have

$$\mathcal{L}(\alpha^s) \geq \mathcal{L}(\gamma^*) + (\alpha^s - \gamma^*)ZZ^\top(\alpha^s - \gamma^*)$$

Since $0 \leq \mathcal{L}(\gamma^*) \leq \mathcal{L}(\alpha^s)$, we have

$$(\alpha^s - \gamma^*)ZZ^\top(\alpha^s - \gamma^*) \leq \mathcal{L}(\alpha^s).$$

Then $\|\gamma^* - \alpha^s\|_2^2 \leq \mathcal{L}(\alpha^s)/\lambda_{\min}(ZZ^\top)$. Third, since $\mathcal{L}(\alpha^s)/n$ is the empirical regression error of $g_s(\mathbf{x})$, by the Talagrand inequality (Koltchinskii, 2011) and Lemma 1, we have , with a probability at least $1 - N^{-3}$, $\mathcal{L}(\alpha^s)/n \leq \eta^2$. We complete the proof by combining the above results.

Proof of Proposition 2

To bound $\lambda_{\min}(ZZ^\top)/n$, we need the following proposition.

Proposition 1 (Concentration Inequaltiy). *(Proposition 1 (Smale & Zhou, 2009)) Let ξ be a random variable on $(\mathcal{X}, P_{\mathcal{X}})$ with values in a Hilbert space $(\mathcal{H}, \|\cdot\|)$. Assume $\|\xi\| \leq M < \infty$ almost sure. Then with a probability at least $1-\delta$, we have*

$$\left\|\frac{1}{n}\sum_{i=1}^{m}\xi(\mathbf{x}_i) - \mathrm{E}[\xi]\right\| \leq \frac{4M\ln(2/\delta)}{\sqrt{n}}.$$

We rewrite $ZZ^\top/n$ as

$$\frac{1}{n}ZZ^\top = \frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i^\top$$

Let $\|Z\|_2$, $\|Z\|_F$ be the spectral norm and Frobenius norm of Z, repectively. Since

$$\mathrm{E}_i[\mathbf{z}_i\mathbf{z}_i^\top] = I, \|\mathbf{z}_i\mathbf{z}_i^\top\|_F = \mathbf{z}_i^\top\mathbf{z}_i = \sum_{j=1}^s \phi_j^2(\mathbf{x}_i) \leq M(s),$$

using the concentration inequality in above proposition 1 (Smale & Zhou, 2009, Proposition 1), we have, with a probability at least $1-\delta$,

$$\left\|\frac{1}{n}ZZ^\top - I\right\|_2 \leq \left\|\frac{1}{n}ZZ^\top - I\right\|_F \leq \frac{4M(s)\ln(2/\delta)}{\sqrt{n}}$$

We complete the proof by setting $\delta = N^{-3}$ and using the fact that $\lambda_{\min}(ZZ^\top/n) \geq 1 - \left\|\frac{1}{n}ZZ^\top - I\right\|_2$.

Proof of Lemma 3

We bound $\mathrm{E}_\mathbf{x}\left[(h_s(\mathbf{x}) - f(\mathbf{x}))^2\right]$ by

$$\mathrm{E}_\mathbf{x}\left[(h_s(\mathbf{x}) - f(\mathbf{x}))^2\right] \leq 2\mathrm{E}_\mathbf{x}\left[(g_s(\mathbf{x}) - f(\mathbf{x}))^2\right] + 2\mathrm{E}_\mathbf{x}\left[(h_s(\mathbf{x}) - g_s(\mathbf{x}))^2\right] \tag{1}$$

Below we will bound the two terms on R.H.S of the above inequality.

To bound the first term in (1), we use Proposition 1, and with a probability $1-2N^{-3}$, we have

$$\|L - \widehat{L}_N\|_{HS} \leq \frac{12\ln N}{\sqrt{N}} = \tau_N$$

According to Lidskii's inequality, we have

$$\sum_i |\lambda_i - \widehat{\lambda}_i| \leq \frac{12\ln N}{\sqrt{N}} = \tau_N$$

Following the same analysis as Lemma 1, we have

$$\begin{aligned}\sum_{i=s+1}^\infty \alpha_i^2 &\leq R^2 \sum_{i=s+1}^\infty \lambda_i \leq R^2 \sum_{i=s+1}^N \widehat{\lambda}_i + R^2 \sum_i |\lambda_i - \widehat{\lambda}_i| \\ &\leq R^2\left(\frac{a^2}{s^{p-1}} + \frac{12\ln N}{\sqrt{N}}\right) \leq \frac{2R^2a^2}{s^{p-1}}\end{aligned}$$

where the last step we use the condition

$$\frac{12\ln N}{\sqrt{N}} \leq \frac{a^2}{s^{p-1}}$$

Hence, by the same analysis in the proof of Lemma 1, with a probability $1-2N^{-3}$, we have

$$\mathrm{E}_\mathbf{x}[(f(\mathbf{x}) - g_s(\mathbf{x}))^2] \leq \frac{4a^2R^2}{s^{p-1}} + 2\varepsilon^2 \leq 2\varepsilon_s^2$$

To bound the second term on (1), we use the following corrolary.

Corollary 1. *Let N ba sufficiently large number such that $\widehat{\phi}_i \in span(\phi_1, \ldots, \phi_N)$. Define*

$$\begin{aligned}\Theta &= \left(\widehat{\phi}_1, \ldots, \widehat{\phi}_s\right), \\ \Phi &= \left(\sqrt{\lambda_1}\phi_s, \ldots, \sqrt{\lambda_s}\phi_s\right), \\ \overline{\Phi} &= \left(\sqrt{\lambda_{s+1}}\phi_{s+1}, \ldots, \sqrt{\lambda_N}\phi_N\right)\end{aligned}$$

Assume

$$r_s = (\lambda_s - \lambda_{s+1}) > 3\|L - \widehat{L}_N\|_{HS}.$$

Then, there exists a matrix $P \in \mathbb{R}^{(N-s)\times s}$ satisfying

$$\|P\|_F \leq \frac{3\|L - \widehat{L}_N\|_{HS}}{r_s}$$

such that

$$\Theta = (\Phi + \overline{\Phi}P)(I + P^\top P)^{-1/2}$$

The above lemma follows from the following perturbation result.

Corollary 2. *(Theorem 2.7 of Chapter 6 (Stewart & guang Sun, 1990)) Let $(\lambda_i, \mathbf{v}_i), i \in [n]$ be the eigenvalues and eigenvectors of a symmetric matrix $A \in \mathbb{R}^{n\times n}$ ranked in the descending order of eigenvalues. Set $X = (\mathbf{v}_1, \ldots, \mathbf{v}_r)$ and $Y = (\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n)$. Given a symmetric perturbation matrix E, let*

$$\widehat{E} = (X, Y)^\top E(X, Y) = \begin{pmatrix} \widehat{E}_{11} & \widehat{E}_{12} \\ \widehat{E}_{21} & \widehat{E}_{22} \end{pmatrix}$$

Let $\|\cdot\|$ represent a consistent family of norms and set

$$\gamma = \|\widehat{E}_{21}\|, \delta = \lambda_r - \lambda_{r+1} - \|\widehat{E}_{11}\| - \|\widehat{E}_{22}\|$$

If $\delta > 0$ and $2\gamma < \delta$, then there exists a unique matrix $P \in \mathbb{R}^{(n-r)\times r}$ satisfying $\|P\| < \frac{2\gamma}{\delta}$ such that

$$X' = (X + YP)(I + P^\top P)^{-1/2}$$
$$Y' = (Y - XP^\top)(I + PP^\top)^{-1/2}$$

are the eigenvectors of $A + E$.

Proof of Corollary 1. Let $\varphi_i = \sqrt{\lambda_i}\phi_i$, it can be shown that $\langle \varphi_i, \varphi_j\rangle_{\mathcal{H}_\kappa} = \delta_{ij}$. Define matrix B as

$$B_{i,j} = \frac{1}{N}\sum_{k=1}^{N} \widehat{\lambda}_k \langle \widehat{\phi}_k, \varphi_i\rangle\langle \widehat{\phi}_k, \varphi_j\rangle.$$

Let $\mathbf{z}_i$ be the eigenvector of B corresponding to eigenvalue $\widehat{\lambda}_i/N$. It is straightforward to show that

$$\mathbf{z}_i = (\langle \varphi_1, \widehat{\phi}_i\rangle_{\mathcal{H}_\kappa}, \ldots, \langle \varphi_N, \widehat{\phi}_i\rangle_{\mathcal{H}_\kappa})^\top, i \in [N]$$

and therefore we have

$$\widehat{\phi}_i = \sum_{k=1}^{N} z_{i,k}\varphi_k, i \in [N], \text{ or } \Theta = (\Phi, \overline{\Phi})Z$$

where $Z = (\mathbf{z}_1, \cdots, \mathbf{z}_s)$. To decide the relationship between $\{\widehat{\phi}_i\}_{i=1}^s$ and $\{\varphi_i\}_{i=1}^N$, we need to determine matrix Z. We define matrix $D = \text{diag}(\lambda_1, \ldots, \lambda_N)$ and matrix $E = B - D$, i.e.

$$E_{i,j} = B_{i,j} - \lambda_i\delta_{i,j} = \langle \varphi_i, (\widehat{L}_N - L)\varphi_j\rangle_{\mathcal{H}_\kappa}.$$

Following the notation of Theorem 2, we define $X = (e_1, \ldots, e_s)$ and $Y = (e_{s+1}, \ldots, e_N)$, where $e_1, \ldots, e_N$ are the canonical bases of $\mathbb{R}^N$, which are also eigenvectors of D. Define δ and γ as follows

$$\gamma = \sqrt{\sum_{i=1}^{s}\sum_{j=s+1}^{N} \langle \varphi_i, (L - \widehat{L}_N)\varphi_j\rangle^2_{\mathcal{H}_\kappa}}$$

$$\delta = r_s - \sqrt{\sum_{i,j=1}^{s} \langle \varphi_i, (L - \widehat{L}_N)\varphi_j\rangle^2_{\mathcal{H}_\kappa}} - \sqrt{\sum_{i,j=s+1}^{N} \langle \varphi_i, (L - \widehat{L}_N)\varphi_j\rangle^2_{\mathcal{H}_\kappa}}$$

where $r_s = \lambda_s - \lambda_{s+1}$. It is easy to verify that γ, δ are defined with respect to the Frobenius norm of $\widehat{E}$ in Theorem 2. In order to apply the result in Theorem 2, we need to show $\delta > 0$ and $\gamma < \delta/2$. To this end, we need to provide the lower and upper bounds for γ and δ, respectively. We first bound δ as

$$\delta - r_s \quad \geq \quad -\sqrt{\sum_{i,j=1}^{N} \langle \varphi_i, (L - \widehat{L}_N)\varphi_j \rangle_{\mathcal{H}_\kappa}^2} = -\|L - \widehat{L}_N\|_{HS}$$

We then bound γ as

$$\gamma = \sqrt{\sum_{i=1}^{r}\sum_{j=r+1}^{N} \langle \varphi_i, (L - \widehat{L}_N)\varphi_j \rangle_{\mathcal{H}_\kappa}^2} \leq \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N} \langle \varphi_i, (L - \widehat{L}_N)\varphi_j \rangle_{\mathcal{H}_\kappa}^2} = \|L - \widehat{L}_N\|_{HS}$$

Hence, when $r_s > 3\|L - \widehat{L}_N\|_{HS}$, we have $\delta > 2\gamma > 0$, which satisfies the condition specified in Theorem 2. Thus, according to Theorem 2, there exists a $P \in \mathbb{R}^{(N-s)\times s}$ satisfying $\|P\| < 2\gamma/\delta$, such that

$$Z = (\mathbf{z}_1, \ldots, \mathbf{z}_s) = (X + YP)(I + P^\top P)^{-1/2}$$

implying

$$\Theta = (\Phi, \overline{\Phi})Z = (\Phi + \overline{\Phi}P)(I + P^\top P)^{-1/2}$$

□

By Corollary 1, since $r_s \geq 3\tau_N^{2/3} > 3\tau_N \geq 3\|L - \widehat{L}_N\|_{HS}$, by the above theorem, we have

$$\sum_{i=1}^{s} \|\widehat{\phi}_i - \sqrt{\lambda_i}\phi_i\|_{\mathcal{H}_\kappa}^2 = \|\Theta - \Phi\|_F^2 = \|\Phi(I - [I + P^\top P]^{-1/2})\|_F^2 + \|\overline{\Phi}P(I + P^\top P)^{-1/2}\|_F^2$$
$$\leq 2\|P^\top P\|_F^2 \leq \frac{18\|L - \widehat{L}_N\|_{HS}^2}{r_s^2} \leq \frac{18\tau_N^2}{r_s^2} \text{ (w.p. } 1 - 2N^{-3})$$

Then, we have

$$\mathrm{E}_{\mathbf{x}}\left[(h_s(\mathbf{x}) - g_s(\mathbf{x}))^2\right] = \mathrm{E}_{\mathbf{x}}\left[\left(\sum_{i=1}^{s} \frac{\alpha_i}{\sqrt{\lambda_i}}\left(\widehat{\phi}_i(\mathbf{x}) - \sqrt{\lambda_i}\phi_i(\mathbf{x})\right)\right)^2\right]$$
$$= \sum_{i=1}^{s} \frac{\alpha_i^2}{\lambda_i}\mathrm{E}_{\mathbf{x}}\left[\sum_{i=1}^{s}\left(\widehat{\phi}_i(\mathbf{x}) - \sqrt{\lambda_i}\phi_i(\mathbf{x})\right)^2\right] \leq \sum_{i=1}^{s} \frac{\alpha_i^2}{\lambda_i}\sum_{i=1}^{s}\left\|\widehat{\phi}_i(\cdot) - \sqrt{\lambda_i}\phi_i(\cdot)\right\|_{\mathcal{H}_\kappa}^2 \leq \frac{18\tau_N^2 R^2}{r_s^2} \text{ (w.p. } 1 - 2N^{-3})$$

Combining the above results, with a probability $1 - 2N^{-3}$, we have

$$\mathrm{E}_{\mathbf{x}}\left[(h_s(\mathbf{x}) - f(\mathbf{x}))^2\right] \leq 4\varepsilon_s^2 + \frac{36\tau_N^2 R^2}{N r_s^2}$$

1. Proof of Lemma 4

We bound as follows:

$$\mathrm{E}_{\mathbf{x}}\left[(\widehat{g}(\mathbf{x}) - h_s(\mathbf{x}))^2\right] \leq \mathrm{E}_{\mathbf{x}}\left[\left(\sum_{i=1}^{s}(\widehat{\gamma}_i^* - \alpha_i)\frac{\widehat{\phi}_i(\mathbf{x})}{\sqrt{\lambda_i}}\right)^2\right]$$
$$= \quad \mathrm{E}_{\mathbf{x}}\left[\left(\sum_{i=1}^{s}(\widehat{\gamma}_i^* - \alpha_i)\phi_i(\mathbf{x}) + (\widehat{\gamma}_i^* - \alpha_i)\left(\frac{\widehat{\phi}_i(\mathbf{x})}{\sqrt{\lambda_i}} - \phi_i(\mathbf{x})\right)\right)^2\right]$$
$$\leq \quad 2\mathrm{E}_{\mathbf{x}}\left[\left(\sum_{i=1}^{s}(\widehat{\gamma}_i^* - \alpha_i)\phi_i(\mathbf{x})\right)^2\right] + 2\mathrm{E}_{\mathbf{x}}\left[\left(\sum_{i=1}^{s}(\widehat{\gamma}_i^* - \alpha_i)\left(\frac{\widehat{\phi}_i(\mathbf{x})}{\sqrt{\lambda_i}} - \phi_i(\mathbf{x})\right)\right)^2\right]$$

For the first term in the above inequality, using the fact $\mathrm{E}_{\mathbf{x}}[\phi_i(\mathbf{x})\phi_j(\mathbf{x})] = \delta_{ij}$, we have

$$\mathrm{E}_{\mathbf{x}}\left[\left(\sum_{i=1}^{s}(\widehat{\gamma}_i^* - \alpha_i)\phi_i(\mathbf{x})\right)^2\right] = \|\gamma^* - \alpha^s\|_2^2$$

For the second term, we bound it as

$$\begin{aligned}
\mathrm{E}_{\mathbf{x}}\left[\left(\sum_{i=1}^{s}(\widehat{\gamma}_i^* - \alpha_i)\left(\frac{\widehat{\phi}_i(\mathbf{x})}{\sqrt{\lambda_i}} - \phi_i(\mathbf{x})\right)\right)^2\right] &\leq \|\widehat{\gamma}^* - \alpha^s\|_2^2 \mathrm{E}_{\mathbf{x}}\left[\sum_{i=1}^{s}\left(\frac{\widehat{\phi}_i(\mathbf{x})}{\sqrt{\lambda_i}} - \phi_i(\mathbf{x})\right)^2\right] \\
&\leq \frac{\|\widehat{\gamma}^* - \alpha^s\|_2^2}{\lambda_s}\mathrm{E}_{\mathbf{x}}\left[\sum_{i=1}^{s}\left(\widehat{\phi}_i(\mathbf{x}) - \sqrt{\lambda_i}\phi_i(\mathbf{x})\right)^2\right] \leq \frac{18\tau_N^2\|\widehat{\gamma}^* - \alpha^s\|_2^2}{\lambda_s r_s^2} \leq \frac{18\tau_N^2\|\widehat{\gamma}^* - \alpha^s\|_2^2}{r_s^3} \text{ (w.p. } 1 - 2N^{-3})
\end{aligned}$$

Similar to the infinite case, we introduce $\mathbf{z}_i = (\widehat{\phi}_1(\mathbf{x}_i)/\sqrt{\lambda_1}, \ldots, \widehat{\phi}_s(\mathbf{x}_i)/\sqrt{\lambda_s})^\top$ and $Z = (\mathbf{z}_1, \ldots, \mathbf{z}_n)$. Then by the similar analysis to Proposition 1 and Proposition 2, with a probability $1 - 2N^{-3}$, we have $\|\alpha^s - \widehat{\gamma}^*\|_2 \leq n\widehat{\eta}^2/\lambda_{min}(ZZ^\top) \leq 2\widehat{\eta}^2$. We then complete the proof by using the assumption **B3** that $r_s^3 \geq 27\tau_N^2$.

References

Koltchinskii, Vladimir. *Oracle Inequalities in Empirical Risk Minimization and Sparse Recovery Problems.* Springer, 2011.

Smale, Steve and Zhou, Ding-Xuan. Geometry on probability spaces. *Constructive Approximation*, 30:311–323, 2009.

Stewart, G. W. and guang Sun, Ji. *Matrix Perturbation Theory.* Academic Press, 1990.