Appendix: Supplementary Material

A. Proof of Theorem 1

Proof. Denote the *unnormalized graph Laplacian* by $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{K}$ and the *normalized graph Laplacian* by

$$\boldsymbol{L}^* = \boldsymbol{D}^{-1/2}\boldsymbol{L}\boldsymbol{D}^{-1/2} = \boldsymbol{I}_n - \boldsymbol{D}^{-1/2}\boldsymbol{K}\boldsymbol{D}^{-1/2},$$

where $\boldsymbol{I}_n$ is the identity matrix of size n. Optimization (5) can be rewritten as

$$\min_{\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_c\in\mathbb{R}^n} \Delta(p,q) + \gamma' \sum\nolimits_{y\in\mathcal{Y}} \boldsymbol{\alpha}_y^\top \boldsymbol{L}^* \boldsymbol{\alpha}_y + \lambda' \sum\nolimits_{y\in\mathcal{Y}} \frac{1}{2}\|\boldsymbol{\alpha}_y\|_2^2, \tag{12}$$

where $\gamma' = \gamma c/2n > 0$ and $\lambda' = \lambda - \gamma c/n > 0$ are regularization parameters. Notice that $\forall y \in \mathcal{Y}$,

$$\boldsymbol{\alpha}_y^\top \boldsymbol{L}^* \boldsymbol{\alpha}_y = \frac{1}{2}\sum_{i,j=1}^{n} \left(\frac{\alpha_{y,i}}{d_i} - \frac{\alpha_{y,j}}{d_j}\right)^2 \boldsymbol{K}_{i,j},$$

and then the second term of (12) is convex since $\boldsymbol{K}_{i,j} \geq 0$.

The loss function $\Delta(p,q)$ is convex w.r.t. $q(y \mid \boldsymbol{x};\boldsymbol{\alpha})$, and $q(y \mid \boldsymbol{x};\boldsymbol{\alpha})$ is linear w.r.t. $\boldsymbol{\alpha}_y$, so $\Delta(p,q)$ is convex w.r.t. $\boldsymbol{\alpha}_y$. The ℓ_2-norm of $\boldsymbol{\alpha}_y$ is strictly convex w.r.t. $\boldsymbol{\alpha}_y$, i.e., it takes zero if and only if $\boldsymbol{\alpha}_y$ is identically zero. Therefore, optimization (12) is strictly convex and there exists a unique globally optimal solution. □

B. Derivation of the Error Bounds

B.1. Definitions

To begin with, we state the inductive definition of Rademacher complexity following El-Yaniv & Pechyony (2009).

Definition 1. *Suppose that $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n$ are independent observations according to $p(\boldsymbol{x})$. Let $\mathcal{F}$ be a class of functions mapping from $\mathcal{X}$ to $\mathbb{R}$, and $\sigma_1,\ldots,\sigma_n$ be independent uniformly $\{\pm 1\}$-valued random variables, i.e., Rademacher variables. Subsequently, the empirical Rademacher complexity conditioned on $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n$ is defined as*

$$\widehat{\mathcal{R}}_n(\mathcal{F}) := \mathbb{E}_{\sigma_1,\ldots,\sigma_n}\left\{\sup_{f\in\mathcal{F}} \frac{2}{n}\sum_{i=1}^{n}\sigma_i f(\boldsymbol{x}_i)\right\},$$

and the inductive Rademacher complexity is defined as

$$\mathcal{R}_n(\mathcal{F}) := \mathbb{E}_{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n}\left\{\widehat{\mathcal{R}}_n(\mathcal{F})\right\}.$$

There exist various definitions of $\widehat{\mathcal{R}}_n(\mathcal{F})$: The definition in Bartlett & Mendelson (2002) is

$$\widehat{\mathcal{R}}_n(\mathcal{F}) = \mathbb{E}_{\sigma_1,\ldots,\sigma_n}\left\{\sup_{f\in\mathcal{F}} \frac{2}{n}\left|\sum_{i=1}^{n}\sigma_i f(\boldsymbol{x}_i)\right|\right\},$$

the definition in Koltchinskii (2001) uses

$$\widehat{\mathcal{R}}_n(\mathcal{F}) = \mathbb{E}_{\sigma_1,\ldots,\sigma_n}\left\{\sup_{f\in\mathcal{F}} \frac{1}{n}\left|\sum_{i=1}^{n}\sigma_i f(\boldsymbol{x}_i)\right|\right\},$$

while the definition in Meir & Zhang (2003) adopt

$$\widehat{\mathcal{R}}_n(\mathcal{F}) = \mathbb{E}_{\sigma_1,\ldots,\sigma_n}\left\{\sup_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^{n}\sigma_i f(\boldsymbol{x}_i)\right\}.$$

The definition in El-Yaniv & Pechyony (2009) is consistent with Bartlett & Mendelson (2002) for function classes that are closed under negation, and is always equal to or less than the one in Bartlett & Mendelson (2002).

Nevertheless, a vital disagreement arises when considering comparison theorems and thus the famous *contraction principle* of Rademacher averages. If $\psi : \mathbb{R} \mapsto \mathbb{R}$ is Lipschitz continuous with a Lipschitz constant L_ψ and satisfies $\psi(0) = 0$, then

$$\widehat{\mathcal{R}}_n(\psi \circ \mathcal{F}) \leq L_\psi \widehat{\mathcal{R}}_n(\mathcal{F})$$

for El-Yaniv & Pechyony (2009) and

$$\widehat{\mathcal{R}}_n(\psi \circ \mathcal{F}) \leq 2L_\psi \widehat{\mathcal{R}}_n(\mathcal{F})$$

for Bartlett & Mendelson (2002). When all involved error bounds are single-sided concentration results, those definitions without the absolute value in the argument of the supremum (El-Yaniv & Pechyony, 2009; Meir & Zhang, 2003) are more natural and powerful.

B.2. Proof of Theorem 2

Let $\boldsymbol{\beta}_{\mathcal{F}} = \boldsymbol{K}^{-1/2}\boldsymbol{D}^{-1/2}\boldsymbol{\alpha}^*_{\mathcal{F}}$, then

$$B^2_{\mathcal{F}} = \|\boldsymbol{D}^{-1/2}\boldsymbol{\alpha}^*_{\mathcal{F}}\|_2^2 = \boldsymbol{\beta}_{\mathcal{F}}^\top \boldsymbol{K}\boldsymbol{\beta}_{\mathcal{F}},$$
$$B'_{\mathcal{F}} = \|\boldsymbol{K}^{-1/2}\boldsymbol{D}^{-1/2}\boldsymbol{\alpha}^*_{\mathcal{F}}\|_1 = \|\boldsymbol{\beta}_{\mathcal{F}}\|_1.$$

Define the class of functions $\mathcal{F}$ as

$$\mathcal{F} := \left\{ \boldsymbol{x} \mapsto \sum_{i=1}^n \beta_i k(\boldsymbol{x}, \boldsymbol{x}'_i) \;\middle|\; \boldsymbol{x}'_i \in \mathcal{X}, \beta_i \in \mathbb{R}, \sum_{i=1}^n |\beta_i| \leq B'_{\mathcal{F}}, \sum_{i,j=1}^n \beta_i\beta_j k(\boldsymbol{x}'_i, \boldsymbol{x}'_j) \leq B^2_{\mathcal{F}} \right\}.$$

It is easy to verify that $f(\boldsymbol{x}) = \langle \Phi_n(\boldsymbol{x}), \boldsymbol{\beta}_{\mathcal{F}} \rangle \in \mathcal{F}$, where $f(\boldsymbol{x})$ is the decision function defined in Eq. (9). By Lemma 22 of Bartlett & Mendelson (2002), we get

$$\widehat{\mathcal{R}}_n(\mathcal{F}) \leq \frac{2B_{\mathcal{F}}}{n}\left(\sum_{i=1}^n k(\boldsymbol{x}_i, \boldsymbol{x}_i)\right)^{1/2} \leq \frac{2B_k B_{\mathcal{F}}}{\sqrt{n}}. \tag{13}$$

Applying Lemma 22 of Bartlett & Mendelson (2002) again gives us

$$\widehat{\mathcal{R}}_l(\mathcal{F}) \leq \frac{2B_{\mathcal{F}}}{l}\left(\sum_{i=1}^l k(\boldsymbol{x}_i, \boldsymbol{x}_i)\right)^{1/2} \leq \frac{2B_k B_{\mathcal{F}}}{\sqrt{l}}. \tag{14}$$

where $\widehat{\mathcal{R}}_l(\mathcal{F})$ is the empirical Rademacher complexities of $\mathcal{F}$ conditioned only on $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_l$.

In the following, we only focus on the proof of inequality (11) based on inequality (13). Inequality (10) can be derived by the exactly same way based on inequality (14). Let

$$\ell_\eta \circ \mathcal{F} := \{(\boldsymbol{x}, y) \mapsto \ell_\eta(yf(\boldsymbol{x})) \mid f \in \mathcal{F}\},$$

which is a class of functions mapping from $\mathcal{X} \times \mathcal{Y}$ to the interval $[0, 1]$. The rest of the proof consists of two steps. The first step bounds $\mathcal{R}_n(\ell_\eta \circ \mathcal{F})$ from above, and the second step bounds $\mathbb{E}\ell(yf(\boldsymbol{x}))$ using $\mathcal{R}_n(\ell_\eta \circ \mathcal{F})$.

B.2.1. Step 1

The following lemma relates the inductive Rademacher complexity of a class of bounded functions to the corresponding empirical Rademacher complexity.

Lemma 3 (Concentration Lemma)**.** *Let $\mathcal{F}_C$ be a class of functions mapping to the interval $[-C, C]$. With probability at least $1 - \delta/2$, we have*

$$\mathcal{R}_n(\mathcal{F}_C) \leq \widehat{\mathcal{R}}_n(\mathcal{F}_C) + 4C\sqrt{\frac{\ln(2/\delta)}{2n}}.$$

Similarly, let $\mathcal{F}^+_C$ be a class of functions mapping to the interval $[0, C]$. With probability at least $1 - \delta/2$, we have

$$\mathcal{R}_n(\mathcal{F}^+_C) \leq \widehat{\mathcal{R}}_n(\mathcal{F}^+_C) + 2C\sqrt{\frac{\ln(2/\delta)}{2n}}.$$

Proof. Recall that $\widehat{\mathcal{R}}_n(\mathcal{F}_C)$ conditioned on $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ is a random variable defined as

$$\widehat{\mathcal{R}}_n(\mathcal{F}_C) = \mathbb{E}_{\sigma_1,\ldots,\sigma_n}\left\{\sup_{f\in\mathcal{F}_C} \frac{2}{n}\sum_{i=1}^{n}\sigma_i f(\boldsymbol{x}_i)\right\}.$$

When an observation $\boldsymbol{x}_i$ changes to $\boldsymbol{x}_i'$, the change of $\widehat{\mathcal{R}}_n(\mathcal{F}_C)$ is no more than $4C/n$, and thus *McDiarmid's inequality* (McDiarmid, 1989) implies that

$$\Pr\left\{\mathcal{R}_n(\mathcal{F}_C) - \widehat{\mathcal{R}}_n(\mathcal{F}_C) \geq \epsilon\right\} \leq \exp\left(-\frac{\epsilon^2 n}{8C^2}\right).$$

The first bound can be obtained by equating the right-hand side of the above inequality to $\delta/2$.

For $\mathcal{F}_C^+$, when an observation $\boldsymbol{x}_i$ changes to $\boldsymbol{x}_i'$, the change of $\widehat{\mathcal{R}}_n(\mathcal{F}_C^+)$ is no more than $2C/n$. The lemma follows by the same argument as above. □

The next lemma is a variation of the comparison lemma in Meir & Zhang (2003), where the comparison is done for two sets of functions under a Bayesian framework, and its validity follows Lemma 5 of El-Yaniv & Pechyony (2009) by setting $p = 1/2$.

Lemma 4 (Comparison Lemma)**.** *Let*

$$\mathcal{H} := \{\boldsymbol{h} = (h_1, \ldots, h_n)^\top \mid h_i = y_i f(\boldsymbol{x}_i), f \in \mathcal{F}\},$$

and $\psi, \psi' : \mathbb{R} \mapsto \mathbb{R}$ be real-valued functions. If for all $\boldsymbol{h}, \boldsymbol{h}' \in \mathcal{H}$ and $i = 1, \ldots, n$,

$$|\psi(h_i) - \psi(h_i')| \leq |\psi'(h_i) - \psi'(h_i')|,$$

then

$$\mathbb{E}_{\sigma_1,\ldots,\sigma_n}\left\{\sup_{h\in\mathcal{H}}\sum_{i=1}^{n}\sigma_i\psi(h_i)\right\} \leq \mathbb{E}_{\sigma_1,\ldots,\sigma_n}\left\{\sup_{h\in\mathcal{H}}\sum_{i=1}^{n}\sigma_i\psi'(h_i)\right\}.$$

Now $\widehat{\mathcal{R}}_n(\ell_\eta \circ \mathcal{F})$ and $\mathcal{R}_n(\ell_\eta \circ \mathcal{F})$ can be bounded from above by $\widehat{\mathcal{R}}_n(\mathcal{F})$ and $\mathcal{R}_n(\mathcal{F})$ based on the comparison lemma.

Lemma 5 (Contraction Lemma)**.** *For any $\eta > 0$, we have*

$$\widehat{\mathcal{R}}_n(\ell_\eta \circ \mathcal{F}) \leq \frac{1}{\eta}\widehat{\mathcal{R}}_n(\mathcal{F}),$$
$$\mathcal{R}_n(\ell_\eta \circ \mathcal{F}) \leq \frac{1}{\eta}\mathcal{R}_n(\mathcal{F}).$$

Proof. Note that $\ell_\eta(z)$ satisfies the Lipschitz condition

$$|\ell_\eta(z) - \ell_\eta(z')| \leq \frac{1}{\eta}|z - z'|, \quad \forall z, z' \in \mathbb{R}.$$

Let $\psi(h_i) = \ell_\eta(y_i f(\boldsymbol{x}_i))$ and $\psi'(h_i) = y_i f(\boldsymbol{x}_i)/\eta$, then

$$\begin{aligned}\mathbb{E}_{\sigma_1,\ldots,\sigma_n}\left\{\sup_{f\in\mathcal{F}}\sum_{i=1}^{n}\sigma_i\ell_\eta(y_i f(\boldsymbol{x}_i))\right\} &\leq \frac{1}{\eta}\mathbb{E}_{\sigma_1,\ldots,\sigma_n}\left\{\sup_{f\in\mathcal{F}}\sum_{i=1}^{n}\sigma_i y_i f(\boldsymbol{x}_i)\right\} \\ &= \frac{1}{\eta}\mathbb{E}_{\sigma_1,\ldots,\sigma_n}\left\{\sup_{f\in\mathcal{F}}\sum_{i=1}^{n}\sigma_i f(\boldsymbol{x}_i)\right\},\end{aligned}$$

where the first step is a corollary of the comparison lemma, and the second step is due to the same distribution of each $\sigma_i y_i$ and σ_i. This completes the proof. □

As a result, if we contract $\widehat{\mathcal{R}}_n(\mathcal{F})$ and then concentrate $\widehat{\mathcal{R}}_n(\ell_\eta \circ \mathcal{F})$, we could know

$$\begin{aligned}\mathcal{R}_n(\ell_\eta \circ \mathcal{F}) &\leq \widehat{\mathcal{R}}_n(\ell_\eta \circ \mathcal{F}) + 2\sqrt{\frac{\ln(2/\delta)}{2n}} \\ &\leq \frac{2B_k B_{\mathcal{F}}}{\eta\sqrt{n}} + 2\sqrt{\frac{\ln(2/\delta)}{2n}},\end{aligned} \tag{15}$$

since ℓ_η maps to the interval $[0, 1]$. On the other hand, for any $f \in \mathcal{F}$,

$$\|f\|_\infty = \sup_{\boldsymbol{x}\in\mathcal{X}} \left| \sum_{i=1}^{n} \beta_i k(\boldsymbol{x}, \boldsymbol{x}'_i) \right| \leq B_k^2 B'_{\mathcal{F}},$$

which says that $\mathcal{F}$ is a class of functions mapping to the interval $[-B_k^2 B'_{\mathcal{F}}, B_k^2 B'_{\mathcal{F}}]$. Thus, if we concentrate $\widehat{\mathcal{R}}_n(\mathcal{F})$ before contract $\mathcal{R}_n(\mathcal{F})$, we can obtain

$$\begin{aligned}\mathcal{R}_n(\ell_\eta \circ \mathcal{F}) &\leq \frac{1}{\eta}\mathcal{R}_n(\mathcal{F}) \\ &\leq \frac{1}{\eta}\left(\frac{2B_k B_{\mathcal{F}}}{\sqrt{n}} + 4B_k^2 B'_{\mathcal{F}}\sqrt{\frac{\ln(2/\delta)}{2n}}\right).\end{aligned} \tag{16}$$

Combining inequalities (15) and (16) finalizes the first step of the proof, that is,

$$\mathcal{R}_n(\ell_\eta \circ \mathcal{F}) \leq \frac{2B_k B_{\mathcal{F}}}{\eta\sqrt{n}} + \min\left(2, \frac{4B_k^2 B'_{\mathcal{F}}}{\eta}\right)\sqrt{\frac{\ln(2/\delta)}{2n}}.$$

B.2.2. Step 2

This step is composed of a single concentration inequality, that is, with probability at least $1 - \delta/2$,

$$\mathbb{E}\ell(yf(\boldsymbol{x})) \leq \hat{\mathbb{E}}_n \ell_\eta(yf(\boldsymbol{x})) + \mathcal{R}_n(\ell_\eta \circ \mathcal{F}) + \sqrt{\frac{\ln(2/\delta)}{2n}}. \tag{17}$$

Since $\forall z \in \mathbb{R}$, $\ell(z)$ is always equal to or less than $\ell_\eta(z)$, for any $f \in \mathcal{F}$ we can write

$$\begin{aligned}\mathbb{E}\ell(yf(\boldsymbol{x})) &\leq \mathbb{E}\ell_\eta(yf(\boldsymbol{x})) \\ &\leq \hat{\mathbb{E}}_n \ell_\eta(yf(\boldsymbol{x})) + \sup_{\psi\in\ell_\eta\circ\mathcal{F}} (\mathbb{E}\psi - \hat{\mathbb{E}}_n\psi).\end{aligned}$$

Any function $\psi(\boldsymbol{x}, y) = \ell_\eta(yf(x)) \in \ell_\eta \circ \mathcal{F}$ satisfies $0 \leq \psi(\boldsymbol{x}, y) \leq 1$, so when $(\boldsymbol{x}_i, y_i)$ changes to $(\boldsymbol{x}'_i, y'_i)$, the change of $\sup_{\psi\in\ell_\eta\circ\mathcal{F}}(\mathbb{E}\psi - \hat{\mathbb{E}}_n\psi)$ cannot be more than $1/n$. Hence, McDiarmid's inequality implies that

$$\Pr\left\{\sup_{\psi\in\ell_\eta\circ\mathcal{F}} (\mathbb{E}\psi - \hat{\mathbb{E}}_n\psi) - \mathbb{E}_{(\boldsymbol{x}_1,y_1),\ldots,(\boldsymbol{x}_n,y_n)} \sup_{\psi\in\ell_\eta\circ\mathcal{F}} (\mathbb{E}\psi - \hat{\mathbb{E}}_n\psi) \geq \epsilon\right\} \leq \exp(-2\epsilon^2 n),$$

or equivalently, with probability at least $1 - \delta/2$,

$$\sup_{\psi\in\ell_\eta\circ\mathcal{F}} (\mathbb{E}\psi - \hat{\mathbb{E}}_n\psi) \leq \mathbb{E}_{(\boldsymbol{x}_1,y_1),\ldots,(\boldsymbol{x}_n,y_n)} \sup_{\psi\in\ell_\eta\circ\mathcal{F}} (\mathbb{E}\psi - \hat{\mathbb{E}}_n\psi) + \sqrt{\frac{\ln(2/\delta)}{2n}}.$$

It remains to bound the expectation $\mathbb{E}_{(\boldsymbol{x}_1,y_1),\ldots,(\boldsymbol{x}_n,y_n)} \sup_{\psi\in\ell_\eta\circ\mathcal{F}}(\mathbb{E}\psi - \hat{\mathbb{E}}_n\psi)$ by the complexity $\mathcal{R}_n(\ell_\eta \circ \mathcal{F})$. Suppose that

$$\{(\boldsymbol{x}'_1, y'_1), \ldots, (\boldsymbol{x}'_n, y'_n) \mid (\boldsymbol{x}'_i, y'_i) \sim p(\boldsymbol{x}, y)\}$$

is a ghost sample for symmetrization, then

$$
\begin{aligned}
\mathbb{E}_{(\boldsymbol{x}_i,y_i)} \sup_{\psi\in\ell_\eta\circ\mathcal{F}} (\mathbb{E}\psi - \hat{\mathbb{E}}_n\psi) &= \mathbb{E}_{(\boldsymbol{x}_i,y_i)} \sup_{\psi\in\ell_\eta\circ\mathcal{F}} \left(\mathbb{E}_{(\boldsymbol{x}'_i,y'_i)}[\hat{\mathbb{E}}_n\psi(\boldsymbol{x}'_i,y'_i)] - \hat{\mathbb{E}}_n\psi(\boldsymbol{x}_i,y_i)\right) \\
&= \mathbb{E}_{(\boldsymbol{x}_i,y_i)} \sup_{\psi\in\ell_\eta\circ\mathcal{F}} \left(\mathbb{E}_{(\boldsymbol{x}'_i,y'_i)}[\hat{\mathbb{E}}_n\psi(\boldsymbol{x}'_i,y'_i) - \hat{\mathbb{E}}_n\psi(\boldsymbol{x}_i,y_i)]\right) \\
&\le \mathbb{E}_{(\boldsymbol{x}_i,y_i),(\boldsymbol{x}'_i,y'_i)} \sup_{\psi\in\ell_\eta\circ\mathcal{F}} \left(\hat{\mathbb{E}}_n\psi(\boldsymbol{x}'_i,y'_i) - \hat{\mathbb{E}}_n\psi(\boldsymbol{x}_i,y_i)\right) \qquad (18) \\
&= \mathbb{E}_{(\boldsymbol{x}_i,y_i),(\boldsymbol{x}'_i,y'_i)} \sup_{\psi\in\ell_\eta\circ\mathcal{F}} \frac{1}{n}\sum_{i=1}^{n}(\psi(\boldsymbol{x}'_i,y'_i) - \psi(\boldsymbol{x}_i,y_i)) \\
&= \mathbb{E}_{\sigma_i,(\boldsymbol{x}_i,y_i),(\boldsymbol{x}'_i,y'_i)} \sup_{\psi\in\ell_\eta\circ\mathcal{F}} \frac{1}{n}\sum_{i=1}^{n}\sigma_i(\psi(\boldsymbol{x}'_i,y'_i) - \psi(\boldsymbol{x}_i,y_i)) \qquad (19) \\
&\le \mathbb{E}_{(\boldsymbol{x}'_i,y'_i),\sigma_i} \sup_{\psi\in\ell_\eta\circ\mathcal{F}} \frac{1}{n}\sum_{i=1}^{n}\sigma_i\psi(\boldsymbol{x}'_i,y'_i) + \mathbb{E}_{(\boldsymbol{x}_i,y_i),\sigma_i} \sup_{\psi\in\ell_\eta\circ\mathcal{F}} \frac{1}{n}\sum_{i=1}^{n}(-\sigma_i)\psi(\boldsymbol{x}_i,y_i) \\
&= 2\mathbb{E}_{(\boldsymbol{x}_i,y_i),\sigma_i} \sup_{\psi\in\ell_\eta\circ\mathcal{F}} \frac{1}{n}\sum_{i=1}^{n}\sigma_i\psi(\boldsymbol{x}_i,y_i) \qquad (20) \\
&= \mathcal{R}_n(\ell_\eta\circ\mathcal{F}),
\end{aligned}
$$

where (18) uses the fact that the supremum is a convex function and then we apply *Jensen's inequality*, (19) is due to the symmetry of the ghost sample and the original sample and thus the same distribution of $\psi(\boldsymbol{x}'_i,y'_i)-\psi(\boldsymbol{x}_i,y_i)$ and $\sigma_i(\psi(\boldsymbol{x}'_i,y'_i)-\psi(\boldsymbol{x}_i,y_i))$, and (20) is valid since σ_i and $-\sigma_i$ have the same distribution while the original and ghost samples also have the same distribution. □