# Supplementary Material: Optimality of Belief Propagation for Crowdsourced Classification

## A. Proof of Lemma 1

We start with the conditional probability of error given $A$ in the following:

$$\Pr[s_\rho \neq \hat{s}^*_\rho(A) \mid A] \;=\; \min\left\{\Pr[s_\rho = +1 \mid A], \Pr[s_\rho = -1 \mid A]\right\}.$$

This directly implies that

$$\begin{aligned}\Delta(\hat{s}^*_\rho(A)) &= \mathbb{E}\left[\frac{1}{2} - \Pr[s_\rho \neq \hat{s}^*_\rho(A) \mid A]\right] \\ &= \frac{1}{2}\,\mathbb{E}\left[\big|\Pr[s_\rho = +1 \mid A] - \Pr[s_\rho = -1 \mid A]\big|\right]\end{aligned} \tag{18}$$

Then, by simple algebra, it follows that

$$\begin{aligned}&\Delta(\hat{s}^*_\rho(A)) \\ &= \frac{1}{2}\sum_A \Pr[A]\cdot\big|\Pr[s_\rho = +1 \mid A] - \Pr[s_\rho = -1 \mid A]\big| \\ &= \frac{1}{2}\sum_A |\Pr[A, s_\rho = +1] - \Pr[A, s_\rho = -1]| \\ &= \frac{1}{2}\sum_A \frac{1}{2}|\Pr[A \mid s_\rho = +1] - \Pr[A \mid s_\rho = -1]|\end{aligned}$$

where for the last equality we use $\Pr[s_\rho = +1] = \Pr[s_\rho = -1] = 1/2$.

Let $\phi^+_\rho$ denote the distribution of $A$ given $s_\rho = +1$, and let $\phi^-_\rho$ be the distribution of $A$ given $s_\rho = -1$, i.e.,

$$\phi^+_i(A) = \Pr[A \mid s_i = +1] \text{ and } \phi^-_i(A) = \Pr[A \mid s_i = -1]$$

Then we have a simple expression of $\Delta(\hat{s}^*_\rho(A))$ as follows:

$$\Delta(\hat{s}^*_\rho(A)) = d_{\mathtt{TV}}(\phi^+_\rho, \phi^-_\rho) \tag{19}$$

where we let $d_{\mathtt{TV}}$ denotes the total variation distance, i.e., for distributions $\phi$ and $\psi$ on the same space $\Omega$, we define

$$d_{\mathtt{TV}}(\phi, \psi) := \frac{1}{2}\sum_{\sigma\in\Omega} |\phi(\sigma) - \psi(\sigma)|.$$

Next we note that since $\partial V_{\rho,2k}$ blocks every path from the outside of $G_{\rho,2k}$ to $\rho$, the information on the outside of $G_{\rho,2k}$, $A \setminus A_{\rho,2k}$, is independent of $s_\rho$ given $s_{\partial V_{\rho,2k}}$, i.e.,

$$\Pr[s_\rho \mid A_{\rho,2k}, s_{\partial V_{\rho,2k}}] = \Pr[s_\rho \mid A, s_{\partial V_{\rho,2k}}]. \tag{20}$$

Hence if we set $\psi^+_{\rho,2k}$ to be the distribution of $A$ and $s_{\partial V_{\rho,2k}}$ given $s_\rho = +1$ and similarly for $\psi^-_{\rho,2k}$, we have

$$\Delta(\hat{z}^*_\rho(A_{\rho,2k})) = d_{\mathtt{TV}}(\psi^+_{\rho,2k}, \psi^-_{\rho,2k}).$$

Noting that $\phi_\rho^+$ and $\phi_\rho^-$ can be obtained by marginalizing out $s_{\partial V_{\rho,2k}}$ in $\psi_{\rho,2k}^+$ and $\psi_{\rho,2k}^-$, it follows that

$$\begin{aligned}
&d_{\mathtt{TV}}(\phi_\rho^+, \phi_\rho^-) \\
&= \frac{1}{2}\sum_A |\phi_\rho^+(A) - \phi_\rho^-(A)| \\
&= \frac{1}{2}\sum_A \left| \sum_{s_{\partial V_{\rho,2k}}} \left(\psi_i^+(A, s_{\partial V_{\rho,2k}}) - \psi_i^-(A, s_{\partial V_{\rho,2k}})\right) \right| \\
&\le \frac{1}{2}\sum_A \sum_{s_{\partial V_{\rho,2k}}} \left|\psi_i^+(A, s_{\partial V_{\rho,2k}}) - \psi_i^-(A, s_{\partial V_{\rho,2k}})\right| \\
&= d_{\mathtt{TV}}(\psi_{\rho,2k}^+, \psi_{\rho,2k}^-)
\end{aligned} \tag{21}$$

which implies $\Delta(\hat{z}^*(A_{\rho,2k})) \ge \Delta(\hat{s}^*(A))$.

We now study $\Delta(\hat{z}^*(A_{\rho,2k}))$ with different $k$. Observe that $\partial V_{\rho,2k}$ blocks every path from $\partial V_{\rho,2k+2}$ to $\rho$, i.e., $s_{\partial V_{\rho,2k+2}}$ is independent of $s_\rho$ given $s_{\partial V_{\rho,2k}}$. Thus from (20) it follows that

$$\Pr[s_\rho \mid A, s_{\partial V_{\rho,2k}}] = \Pr[s_\rho \mid A, s_{\partial V_{\rho,2k}}, s_{\partial V_{\rho,2k+2}}].$$

Therefore, $\psi_{\rho,2k+2}^+$ and $\psi_{\rho,2k+2}^-$ can be obtained from $\psi_{\rho,2k}^+$ and $\psi_{\rho,2k}^-$ by marginalizing out $s_{\partial V_{\rho,2k+2}}$. Similarly to (21), we have

$$d_{\mathtt{TV}}(\psi_{\rho,2k+2}^+, \psi_{\rho,2k+2}^-) \le d_{\mathtt{TV}}(\psi_{\rho,2k}^+, \psi_{\rho,2k}^-)$$

which completes the proof of Lemma 1.

## B. Proof of Lemma 2

The proof of Lemma 2 is analog to that of Lemma 1. Let $\varphi_\rho^+$ be the distribution of $A'$ given $s_\rho = +1$ and $\varphi_\rho^-$ be the distribution of $A'$ given $s_\rho = -1$, i.e.,

$$\Delta(\hat{s}_\rho^*(A')) = d_{\mathsf{TV}}(\varphi_\rho^+, \varphi_\rho^-).$$

Since $\varphi_\rho^+$ and $\varphi_\rho^-$ can be obtained by marginalizing out $A \setminus A'$ from $\phi_\rho^+$ and $\phi_\rho^-$ in (19), using the same logic for (21), we have

$$d_{\mathtt{TV}}(\varphi_\rho^+, \varphi_\rho^-) \le d_{\mathtt{TV}}(\phi_\rho^+, \phi_\rho^-)$$

which completes the proof of Lemma 2.

## C. Proof of Lemma 3

We start with several notations which we study for the proof. For $i \in V_{\rho,2k}$, let $T_i = (V_i, W_i, E_i)$ be the subtree rooted from $i$ including all the offsprings of $i$ in tree $G_{\rho,2k}$. We let $\partial V_i$ denote the leaves in $T_i$ and $A_i := \{A_{ju} : (j,u) \in E_i\}$. Define

$$X_i := \Pr[s_i = +1 \mid A_i] - \Pr[s_i = -1 \mid A_i]$$

Here $X_i$ is often called the *magnetization* of $s_i$ given $A_i$. Let $\partial V_i$ denote the leaves in $T_i$ and $A_i := \{A_{ju} : (j,u) \in E_i\}$. Similarly, given $A_i$ and $s_{\partial V_i}$, we define the *biased* magnetization $Y_i$:

$$Y_i := \Pr[s_i = +1 \mid A_i, s_{\partial V_i}] - \Pr[s_i = -1 \mid A_i, s_{\partial V_i}].$$

Using the alternative expression of $\Delta$ in (18), one can check that

$$0 \le \Delta(\hat{z}_i^*(A_i)) - \Delta(\hat{s}_i^*(A_i))$$

$$
\begin{aligned}
&= \frac{1}{2}\,\mathbb{E}\left[|Y_i| - |X_i|\right] \\
&\le \mathbb{E}[|Y_i - X_i|]
\end{aligned}
$$

where the expectation is taken with respect to $A_i$ and $s_{\partial V_i}$.

Next, for $0 \le t \le k$, we define $i(t) \in \partial V_{\rho,2k-2t}$ to be a random node chosen uniformly at random so that $i(0)$ is a leaf node in $G_{\rho,2k}$, i.e., $X_{i(0)} = 0$ thus $|X_{i(0)} - Y_{i(0)}| \le 1$, and $i(k)$ is the root $\rho$, i.e., $\Delta(\hat{z}^*_\rho(A_\rho)) - \Delta(\hat{s}^*_\rho(A_\rho)) = \frac{1}{2}\,\mathbb{E}\left[|Y_\rho| - |X_\rho|\right]$. Therefore it is enough to show that for each $0 \le t < k$

$$
\mathbb{E}\left[\sqrt{\left|X_{i(t+1)} - Y_{i(t+1)}\right|}\right] \le \frac{1}{2}\,\mathbb{E}\left[\sqrt{\left|X_{i(t)} - Y_{i(t)}\right|}\right] \tag{22}
$$

since this implies $\mathbb{E}\left[\sqrt{\left|Y_\rho - X_\rho\right|}\right] \le 2^{-k}$, i.e., $\mathbb{E}\left[\left|Y_\rho - X_\rho\right|\right] \to 0$ as $k \to \infty$. Here $\mathbb{E}\left[\sqrt{|X_{i(t)} - Y_{i(t)}|}\right]$ quantifies the correlation from the information at the leaves $\partial V_{i(t)}$ to $s_{i(t)}$. We study the correlation after taking the square root. The square root magnifies the correlation especially when it is small and the magnification provides an analytical consistency even after the correlation becomes small. In particular, we will show that the correlation exponentially decays with respect to $0 \le t < k$ in what follows.

To do so we study certain recursions describing relations among $X$ and $Y$. Let $\partial i$ be the set of all the second offsprings of $i$ and let $u(ij)$ denote the worker to whom assigned tasks $i$ and $j \in \partial i$ because $r = 2$. Then, using Bayes' rule with $\Pr[s_j \mid A_j] = \frac{1+s_j X_j}{2}$, we have the following recurrence for $X$:

$$
\begin{aligned}
X_i &= h_i(X_{\partial i}) \\
&:= \frac{\prod_{j\in\partial i} g^+_{ij}(X_j) - \prod_{j\in\partial i} g^-_{ij}(X_j)}{\prod_{j\in\partial i} g^+_{ij}(X_j) + \prod_{j\in\partial i} g^-_{ij}(X_j)}
\end{aligned} \tag{23}
$$

where for $i$ and $j \in \partial i$, the functions $g^+_{ij}$ and $g^-_{ij}$ are the marginal probability of $s_i = +1$ and $s_i = -1$ given $A_{iu(ij)}, A_{ju(ij)}$ and $X_j$, respectively, i.e.,

$$
\begin{aligned}
g^+_{ij}(X_j) &\propto \Pr\left[s_i = +1 \mid A_{iu(ij)}, A_{ju(ij)}, X_j\right] \\
&= \sum_{s_j} f_{u(ij)}(s_i = +1, s_j)\left(\frac{1 + s_j X_j}{2}\right) \\
g^-_{ij}(X_j) &\propto \Pr\left[s_i = -1 \mid A_{iu(ij)}, A_{ju(ij)}, X_j\right] \\
&= \sum_{s_j} f_{u(ij)}(s_i = -1, s_j)\left(\frac{1 + s_j X_j}{2}\right)
\end{aligned}
$$

In particular, for notational convenience, we use $g^+_{ij}$ and $g^-_{ij}$ in what follows:

$$
g^+_{ij}(x_j) := \begin{cases}
(1+\mu)(1+\delta x_j) & \text{if } \sigma_{ij} = ++ \\
(1+\mu)(1-\delta x_j) & \text{if } \sigma_{ij} = +- \\
(1-\mu)(1-\delta' x_j) & \text{if } \sigma_{ij} = -+ \\
(1-\mu)(1+\delta' x_j) & \text{if } \sigma_{ij} = --
\end{cases}
$$

$$
g^-_{ij}(x_j) := \begin{cases}
(1-\mu)(1-\delta' x_j) & \text{if } \sigma_{ij} = ++ \\
(1-\mu)(1+\delta' x_j) & \text{if } \sigma_{ij} = +- \\
(1+\mu)(1+\delta x_j) & \text{if } \sigma_{ij} = -+ \\
(1+\mu)(1-\delta x_j) & \text{if } \sigma_{ij} = --
\end{cases}
$$

where for $i$ and $j \in \partial i$, we let $\sigma_{ij}$ denote the *ordered* pair of signs of $A_{iu(ij)}$ and $A_{ju(ij)}$, and using

$$
\mu := \mathbb{E}\left[2p_u - 1\right] \quad \text{and} \quad \theta := \frac{\mathrm{Var}[p_u]}{\mathbb{E}[p_u]\,\mathbb{E}[1-p_u]},
$$

we define

$$\delta := 1 - (1-\theta)(1-\mu) \quad \text{and} \quad \delta' := 1 - (1-\theta)(1+\mu).$$

From the same reasoning for (23), we have the same recurrence when every instance of $X$ is replaced by $Y$, i.e.,

$$Y_i = h_i(Y_{\partial i}). \tag{24}$$

From the above initialization of $X_{i(0)}$ and $Y_{i(0)}$, the recurrence in (23) and (24) generates $X_{i(t)}$ and $Y_{i(t)}$ for all $1 \le t \le k$, where the recurrence is the function of $A_{\rho,2k}$ since the functions $g^+_{ij}$ and $g^+_{ij}$ depend on $\sigma_{ij}$.

Assuming the true label $s_i = +1$ for all $i \in G_{\rho,2k}$, for each $i$ and $j \in \partial i$, $\sigma_{ij}$ is a i.i.d. random variable with the following distribution:

$$\sigma_{ij} = \begin{cases} ++ & \text{w.p. } \mathbb{E}[p_u^2] = \left(\frac{1+\mu}{2}\right)\left(\frac{1+\delta}{2}\right) \\ +- & \text{w.p. } \mathbb{E}[p_u(1-p_u)] = (1-\theta)\left(\frac{1-\mu^2}{4}\right) \\ -+ & \text{w.p. } \mathbb{E}[(1-p_u)p_u] = (1-\theta)\left(\frac{1-\mu^2}{4}\right) \\ -- & \text{w.p. } \mathbb{E}[(1-p_u)^2] = \left(\frac{1-\mu}{2}\right)\left(\frac{1+\delta'}{2}\right). \end{cases} \tag{25}$$

We note that for any $i$, the true label $s_i$ is uniformly distributed and it is clear that $|\partial i| \ge l-1$. In addition, the choice of $i(t)$ in (22) is uniform. Therefore, without loss of generality, we focus on a non-leaf node $i \in V_{\rho,2k} \setminus \partial V_{\rho,2k}$ and show that

$$\mathbb{E}^+\left[\sqrt{|X_i - Y_i|}\right] \;\le\; \frac{1}{2(l-1)} \sum_{j \in \partial i} \mathbb{E}^+\left[\sqrt{|X_j - Y_j|}\right]. \tag{26}$$

where we let $\mathbb{E}^+$ denote the conditional expectation given $s_j = +1$ for all $j$.

To this end, we will use the mean value theorem. We first obtain a bound on gradient of $h_i(x)$ for $x \in [-1,1]^{\partial i}$. Define $g^+_i(x) := \prod_{j\in\partial i} g^+_{ij}(x_j)$ and $g^-_i(x) := \prod_{j\in\partial i} g^-_{ij}(x_j)$. Then, by basic calculus, we have

$$\begin{aligned} \frac{\partial h_i}{\partial x_j} &= \frac{\partial}{\partial x_j}\frac{g^+_i - g^-_i}{g^+_i + g^-_i} \\ &= 2\frac{g^-_i \cdot \frac{\partial g^+_i}{\partial x_j} - g^+_i \cdot \frac{\partial g^-_i}{\partial x_j}}{(g^+_i + g^-_i)^2} \\ &= \begin{cases} \frac{g^+_i g^-_i}{(g^+_i+g^-_i)^2}\frac{4\theta}{(1+\delta x_j)(1-\delta' x_j)} & \text{if } \sigma_{ij} = ++ \\ -\frac{g^+_i g^-_i}{(g^+_i+g^-_i)^2}\frac{4\theta}{(1-\delta x_j)(1+\delta' x_j)} & \text{if } \sigma_{ij} = +- \\ -\frac{g^+_i g^-_i}{(g^+_i+g^-_i)^2}\frac{4\theta}{(1+\delta x_j)(1-\delta' x_j)} & \text{if } \sigma_{ij} = -+ \\ \frac{g^+_i g^-_i}{(g^+_i+g^-_i)^2}\frac{4\theta}{(1-\delta x_j)(1+\delta' x_j)} & \text{if } \sigma_{ij} = --. \end{cases} \end{aligned}$$

Since $x \in [-1,1]^{\partial i}$, both $g^+_i$ and $g^-_i$ are positive. Thus we have

$$\left|\frac{g^+_i g^-_i}{(g^+_i + g^-_i)^2}\right| \le \frac{g^-_i}{g^+_i}. \tag{27}$$

We note here that one can replace the RHS with $g^+_i/g^-_i$. However, in our analysis, we focus on the case of $s_i = +1$ and (27) because plugging $X_{\partial i}$ or $Y_{\partial i}$ into $x$, $h_i(x)$, which is the magnetization $X_i$ or $Y_i$, will be large thus $g^-_i/g^+_i$ will be a tighter upper bound than $g^+_i/g^-_i$. Our analysis covers all the general cases because the same analysis with $g^+_i/g^-_i$ will work with $s_i = -1$ conversely.

From (27), it follows that for $x \in [-1,1]^{\partial i}$

$$\left|\frac{\partial h_i}{\partial x_j}(x)\right| \le g'_{ij}(x_j) \prod_{j' \ne j} \frac{g^-_{ij'}(x_{j'})}{g^+_{ij'}(x_{j'})}$$

where we define

$$g'_{ij}(x_j) = \begin{cases} \frac{4\theta}{(1+\delta x_j)^2}\left(\frac{1-\mu}{1+\mu}\right) & \text{if } \sigma_{ij} = ++ \\ \frac{4\theta}{(1-\delta x_j)^2}\left(\frac{1-\mu}{1+\mu}\right) & \text{if } \sigma_{ij} = +- \\ \frac{4\theta}{(1-\delta' x_j)^2}\left(\frac{1+\mu}{1-\mu}\right) & \text{if } \sigma_{ij} = -+ \\ \frac{4\theta}{(1+\delta' x_j)^2}\left(\frac{1+\mu}{1-\mu}\right) & \text{if } \sigma_{ij} = --\,. \end{cases}$$

By simple algebra with $\mu > 0$, one can check that for given $\sigma_{ij}$ and $x_j \in [-1,1]$, each of $g'_{ij}(x_j)$ and $g^-_{ij}(x_j)/g^+_{ij}(x_j)$ is strictly positive and monotonically decreasing or increasing. Thus, for $x, y \in [-1,1]^{\partial i}$ and $\lambda \in [0,1]$, we have

$$\left|\frac{\partial h_i}{\partial x_j}(\lambda x + (1-\lambda)y)\right| \;\le\; \max\left\{g'_{ij}(x_j), g'_{ij}(y_j)\right\} \prod_{j' \neq j} \max\left\{\frac{g^-_{ij'}(x_{j'})}{g^+_{ij'}(x_{j'})}, \frac{g^-_{ij'}(y_{j'})}{g^+_{ij'}(y_{j'})}\right\}.$$

Then, plugging $X_{\partial i}$ and $Y_{\partial i}$ into $x$ and $y$ and using the mean value theorem, it follows that

$$\begin{aligned} &|h_i(X_{\partial i}) - h_i(Y_{\partial i})| \\ &\le \sum_{j \in \partial i} |X_j - Y_j| \cdot \left|\frac{\partial h_i}{\partial x_j}(\lambda X_{\partial i} + (1-\lambda)Y_{\partial i})\right| \\ &\le \sum_{j \in \partial i} |X_j - Y_j| \cdot \max\left\{g'_{ij}(X_j), g'_{ij}(Y_j)\right\} \times \prod_{j' \neq j} \max\left\{\frac{g^-_{ij'}(X_{j'})}{g^+_{ij'}(X_{j'})}, \frac{g^-_{ij'}(Y_{j'})}{g^+_{ij'}(Y_{j'})}\right\}. \end{aligned} \tag{28}$$

Note that in the RHS, the first two terms depend on each other. To remove the dependence, we use a constant bound $\eta$ of $\max\left\{g'_{ij}(X_j), g'_{ij}(Y_j)\right\}$ in what follows:

$$\begin{aligned} \eta &:= \max_{x_j \in [-1,1]} g'_{ij}(x_j) \\ &= \max\left\{\tfrac{4\theta}{(1-\delta')^2}\left(\tfrac{1+\mu}{1-\mu}\right), \tfrac{4\theta}{(1+\delta')^2}\left(\tfrac{1+\mu}{1-\mu}\right)\right\} \end{aligned}$$

where the last equality is straightforward from simple algebra with the fact that $\mu \in [0,1]$ and $\theta \in [0,1]$. Then, by taking square root of (28) with $\eta$, it follows that

$$\begin{aligned} &\mathbb{E}^+\left[\sqrt{|X_i - Y_i|}\right] \\ &\le \sum_{j \in \partial i} \mathbb{E}^+\left[\sqrt{|X_j - Y_j|}\right] \times \sqrt{\eta} \prod_{j' \neq j} \mathbb{E}^+\left[\max\left\{\sqrt{\tfrac{g^-_{ij'}(X_{j'})}{g^+_{ij'}(X_{j'})}}, \sqrt{\tfrac{g^-_{ij'}(Y_{j'})}{g^+_{ij'}(Y_{j'})}}\right\}\right]. \end{aligned}$$

The square root is taken to .

We obtain a bound of the last term in the following lemma whose proof is presented in Appendix D.

**Lemma 4.** *For given $\mu := \mathbb{E}[2p_u - 1] > 0$, there exists a constant $C'_\mu$ such that for any $l \ge C'_\mu$,*

$$\mathbb{E}^+\left[\max\left\{\sqrt{\tfrac{g^-_{ij}(X_j)}{g^+_{ij}(X_j)}}, \sqrt{\tfrac{g^-_{ij}(Y_j)}{g^+_{ij}(Y_j)}}\right\}\right] \le \sqrt{1 - \tfrac{\mu^2}{2}}$$

Then we can find a constant $C_\mu \ge C'_\mu$ such that if $l - 1 \ge C_\mu$,

$$\eta\left(1 - \tfrac{\mu^2}{2}\right)^{C_\mu/2} \;\le\; \frac{1}{2C_\mu} \;\le\; \frac{1}{2l}$$

which implies (26) and completes the proof of Lemma 3.

# D. Proof of Lemma 4

For notational convenience, we let $\Gamma(x)$ denote the conditional expectation of $\sqrt{\frac{g^{-}_{ij}(X_j)}{g^{+}_{ij}(X_j)}}$ given $s_i = s_j = +1$ with $X_j = x$. Using the conditional distribution of $\sigma_{ij}$ given $s_i = s_j = +1$ in (25), it follows that for $-1 \le x \le 1$

$$\begin{aligned}\Gamma(x) &:= \mathbb{E}^{+}\left[\sqrt{\frac{g^{-}_{ij}(x)}{g^{+}_{ij}(x)}}\right] \\ &= \frac{\sqrt{1-\mu^2}}{4}\left[(1+\delta)\sqrt{\frac{1-\delta' x}{1+\delta x}} + (1-\delta)\sqrt{\frac{1+\delta' x}{1-\delta x}} + (1-\delta')\sqrt{\frac{1+\delta x}{1-\delta' x}} + (1+\delta')\sqrt{\frac{1-\delta x}{1+\delta' x}}\right].\end{aligned}$$

By simple algebra, it is not hard to check that for $-1 \le x \le 1$, $\frac{d\Gamma(x)}{dx} \le 0$, i.e., $\Gamma(x)$ is non-increasing. Hence we have

$$\begin{aligned}\mathbb{E}^{+}\left[\max\left\{\sqrt{\frac{g^{-}_{ij}(X_j)}{g^{+}_{ij}(X_j)}}, \sqrt{\frac{g^{-}_{ij}(Y_j)}{g^{+}_{ij}(Y_j)}}\right\}\right] &\le \Gamma(-1)\Pr{}^{+}[X_j \le 0, Y_j \le 0] \;+\; \Gamma(0)\left(1 - \Pr{}^{+}[X_j \le 0, Y_j \le 0]\right) \\ &\le \Gamma(-1)\left(\Pr{}^{+}[X_j \le 0] + \Pr{}^{+}[Y_j \le 0]\right) \;+\; \Gamma(0) \qquad (29)\end{aligned}$$

where it is straightforward to check that

$$\Gamma(0) = \sqrt{1-\mu^2}.$$

Then we will show that

$$\Pr{}^{+}[Y_j \le 0] \;\le\; \Pr{}^{+}[X_j \le 0] \;\le\; \exp\left(-\frac{(|\partial j| - 1)\mu^2}{2}\right). \qquad (30)$$

Combining (29) and (30) it follows that

$$\mathbb{E}^{+}\left[\max\left\{\sqrt{\frac{g^{-}_{ij}(X_j)}{g^{+}_{ij}(X_j)}}, \sqrt{\frac{g^{-}_{ij}(Y_j)}{g^{+}_{ij}(Y_j)}}\right\}\right] \le \Gamma(-1)\cdot 2\exp\left(-\frac{(|\partial j| - 1)\mu^2}{2}\right) \;+\; \sqrt{1-\mu^2}$$

where we can find constantly large $C'_\mu$ such that for all $|\partial j| > C'_\mu$ the first term in RHS is arbitrarily small, i.e., less than $\sqrt{1-\mu^2/2} - \sqrt{1-\mu^2}$ because $\Gamma(-1)$ is constant but $2\exp\left(-\frac{(|\partial j|-1)\mu^2}{2}\right)$ with $\mu > 0$ exponentially decreases with respect to $|\partial j|$.

To prove (30), we first note that the MAP estimator $\hat{s}^*_j(A_j)$ of $s_j$ given $A_j$ is identical to estimating $s_j = +1$ if $X_j$ is positive and $s_j = -1$ otherwise. Here the MAP estimator $\hat{s}^*_j(A_j)$ outperforms MV with $\{A_{ju(jj')} : j' \in \partial j\}$. Using Hoeffding's bound the error probability of MV is bounded as follows:

$$\Pr{}^{+}[X_j \le 0] \;\le\; \Pr{}^{+}[s_j \ne \hat{s}^{\mathsf{MV}}_j] \;\le\; \exp\left(-\frac{(|\partial j| - 1)\mu^2}{2}\right)$$

where Lemma 2 implies the first inequality. In addition, $\hat{z}^*_j(A_j)$ of $s_j$ given $A_j$ and $\partial V_i$ is identical to estimating $s_j = +1$ if $Y_j$ is positive and $s_j = -1$ otherwise. Hence, from Lemma 1, it directly follows the first inequality in (30). This completes the proof of Lemma 4.