APPENDIX

This section briefly explains the sentiment propagation mechanism in the IP learning for the CSNN.

A. Effect of IP learning for WOSL, SSL, WLCSL

Notation. Let us define $R(\cdot)$ and $PN(\cdot)$ as

$$R(w_t^{\mathbf{Q}}) := \begin{cases} -1 & (sentiment\ of\ w_t^{\mathbf{Q}}\ is\ shifted) \\ 1 & (otherwise) \end{cases}, \tag{8}$$

$$PN(w_t^{\mathbf{Q}}, \mathbf{Q}) := \begin{cases} 1 & (\mathrm{sign}(d^{\mathbf{Q}} - 0.5) \neq (R(w_t^{\mathbf{Q}})) \\ -1 & (d^{\mathbf{Q}} = (\mathrm{sign}(d^{\mathbf{Q}} - 0.5) = (R(w_t^{\mathbf{Q}})) \end{cases} \tag{9}$$

where $PN(w_t^{\mathbf{Q}}, \mathbf{Q}) = 1$ denotes the case where the sentiment of term $w_t^{\mathbf{Q}}$ is shifted in a negative review $\mathbf{Q}$ or the sentiment of term $w_t^{\mathbf{Q}}$ is not shifted in a positive review, and $PN(w_t^{\mathbf{Q}}, \mathbf{Q}) = -1$ denotes the opposite case.

Moreover, we define Condition A.1 as

Condition A.1: if word w_i in S^d then,

$$w_i^p \begin{cases} > 0 & (OS(w_i^p) > 0) \\ < 0 & (OS(w_i^p) < 0) \end{cases} \tag{10}$$

is satisfied, where

$$OS(w_j^p) := E[PN(w_t^{\mathbf{Q}}, \mathbf{Q}) | w_t^{\mathbf{Q}} = w_j^p, \mathbf{Q} \in \Omega^{tr}].$$

and Ω^{tr} is a set of reviews in a training dataset.

Here, the following Proposition A.2 is satisfied:

Proposition A.2: If Condition A.1 is satisfied, and $\min_{w_i \in S^d} \|\boldsymbol{w}_i^{em} - \boldsymbol{w}_j^{em}\|_2 < \delta$ where $\delta(> 0)$ is sufficiently small, then, the following equations are satisfied for word w_j after sufficient iterations through IP learning:

$$E[w_j^p] \begin{cases} > 0 & (OS(w_j^p) > 0) \\ < 0 & (OS(w_j^p) < 0) \end{cases}, \tag{11}$$

$$E[s_t^{\mathbf{Q}}] \begin{cases} > 0 & (R(w_t^{\mathbf{Q}}) > 0) \\ < 0 & (R(w_t^{\mathbf{Q}}) < 0) \end{cases} \tag{12}$$

Proposition A.2 denotes that if the meaning of a term w_j is sufficiently similar to any of words in S^d and S^d satisfies Condition A.1, then, its word-level original sentiments and sentiment shifts are expected to be accurately assigned by the CSNN. The quality of the word sentiment dictionary is important for the success of propagation, where $|S^d|$ should not be too small and each word in S^d must satisfy Condition A.1. This proposition can be explained using the following propositions A.3–A.6.

Proposition A.3: For every $c_t^{\mathbf{Q}} \in \{\{c_t^{\mathbf{Q}}\}_{t=1}^n | \mathbf{Q} \in \Omega^{tr}\}$,

$$\frac{\partial L^{\mathbf{Q}}}{\partial c_t^{\mathbf{Q}}} \begin{cases} < 0 & (d^{\mathbf{Q}} = 1) \\ > 0 & (d^{\mathbf{Q}} = 0) \end{cases} \tag{13}$$

where

$$c_t^{\mathbf{Q}} := p_t^{\mathbf{Q}} \cdot s_t^{\mathbf{Q}}.$$

Proposition A.4: If $d^{\mathbf{Q}} = 1$ and $w_i^p > 0$, or $d^{\mathbf{Q}} = 0$ and $w_i^p < 0$ word $w_i = w_t^{\mathbf{Q}}$, then, $\frac{\partial L^{\mathbf{Q}}}{\partial s_t^{\mathbf{Q}}} < 0$. In the opposite case, $\frac{\partial L^{\mathbf{Q}}}{\partial s_t^{\mathbf{Q}}} > 0$.

Proposition A.5: Let $D_i^{\mathbf{Q}}$ be a set of passages that include word w_i, and $t'(\mathbf{Q}, i)$ be $\{t' | w_{t'}^{\mathbf{Q}} = w_i, w_{t'}^{\mathbf{Q}} \in \mathbf{Q}\}$. If

$$\|\boldsymbol{w}_i^{em} - \boldsymbol{w}_j^{em}\|_2 < \delta, \tag{14}$$

$$\min_{t' \in t'(\mathbf{Q},i), \mathbf{Q} \in D_i^{\mathbf{Q}}} \|\overrightarrow{\boldsymbol{h}}_t^{\mathbf{Q}} - \overrightarrow{\boldsymbol{h}}_{t'}^{\mathbf{Q}}\|_2 < T'\delta, \tag{15}$$

and

$$\min_{t' \in t'(\mathbf{Q},i), \mathbf{Q} \in D_i^{\mathbf{Q}}} \|\overleftarrow{\boldsymbol{h}}_t^{\mathbf{Q}} - \overleftarrow{\boldsymbol{h}}_{t'}^{\mathbf{Q}}\|_2 < T''\delta, \tag{16}$$

where $\delta > 0$ is established, then,

$$\min_{t' \in t'(\mathbf{Q},i), \mathbf{Q} \in D_i^{\mathbf{Q}}} \|s_t^{\mathbf{Q}} - s_{t'}^{\mathbf{Q}}\|_2 < T'''\delta \tag{17}$$

where $T' > 0, T'' > 0$, $T''' > 0$, and $w_j = w_t^{\mathbf{Q}}$.

Proposition A.6: If w_j satisfies

$$\begin{cases} s_t^{\mathbf{Q}} < 0 & (R(w_t^{\mathbf{Q}}) = -1, w_t = w_j) \\ s_t^{\mathbf{Q}} > 0 & (R(w_t^{\mathbf{Q}}) = 1, w_t = w_j) \end{cases} \tag{18}$$

, then, w_j satisfies

$$\begin{cases} \frac{\partial L^{\mathbf{Q}}}{\partial w_j^p} < 0 (PN(w_t^{\mathbf{Q}}, \mathbf{Q}) = 1 \wedge w_t^{\mathbf{Q}} = w_j) \\ \frac{\partial L^{\mathbf{Q}}}{\partial w_j^p} > 0 (PN(w_t^{\mathbf{Q}}, \mathbf{Q}) = -1 \wedge w_t^{\mathbf{Q}} = w_j) \end{cases}. \tag{19}$$

Proposition A.7: Let the values of $\boldsymbol{W}^O$ before and after performing *Update* in Algorithm 1 in the *tth* iteration be $\boldsymbol{W}_t^{O,a}$ and $\boldsymbol{W}_t^{O,b}$, respectively. Then,

$$\frac{\|\boldsymbol{W}_t^{O,a} - \boldsymbol{W}_t^{O,b}\|_2}{\|\boldsymbol{W}_t^{O,b}\|_2} \xrightarrow[t \to \infty]{} 0. \tag{20}$$

B. Effect of IP learning for GIL

In IP learning, the values of GIL are assigned in the form that terms with strong sentiment are attentioned:

$$\frac{\partial L^{\mathbf{Q}}}{\partial \alpha_t^{\mathbf{Q}}} = \boldsymbol{M}_t^{\mathbf{Q}T} \Delta_o^{\mathbf{Q}} \cdot s_t^{\mathbf{Q}} \cdot p_t^{\mathbf{Q}} \tag{21}$$

where

$$\boldsymbol{M}_t^{\mathbf{Q}} := \boldsymbol{W}^o \boldsymbol{b}_t^{\mathbf{Q}} \mathrm{diag}(1 - (\tanh(\sum_{t=1}^n \boldsymbol{v}_t^{\mathbf{Q}}))^2), \tag{22}$$

$$\Delta_o^{\mathbf{Q}} := \begin{cases} (\boldsymbol{a}^{\mathbf{Q}} - (1,0)) & (d^{\mathbf{Q}} = 0) \\ (\boldsymbol{a}^{\mathbf{Q}} - (0,1)) & (d^{\mathbf{Q}} = 1) \end{cases}, \tag{23}$$

$$\boldsymbol{M}_t^{\mathbf{Q}T} \Delta_o^{\mathbf{Q}} \begin{cases} > 0 & (d^{\mathbf{Q}} = 0) \\ < 0 & (d^{\mathbf{Q}} = 1) \end{cases}. \tag{24}$$

This attention manner is known to be natural for humans [6].

Proposition A.8: When Init is used, then, if $\min_{w_j \in S^d} |e_t^{\mathbf{Q}} - \boldsymbol{w}_j^{em}| < \epsilon$ where $\epsilon > 0$ is sufficiently small, then,

$$\mathrm{sign}\left(\frac{\partial L^{\mathbf{Q}'}}{\partial p_{t'}^{\mathbf{Q}'}}\right) = \begin{cases} R(w_{t'}^{\mathbf{Q}'(w_{t'}^{\mathbf{Q}'}, w_j)}) & (d^{\mathbf{Q}} = 0) \\ -R(w_{t'}^{\mathbf{Q}'(w_{t'}^{\mathbf{Q}'}, w_j)}) & (d^{\mathbf{Q}} = 1) \end{cases}.$$

where

$$I(a,b) := \begin{cases} 1 & (a = b) \\ 0 & (a \neq b), \end{cases}, \Psi(a,b) := \begin{cases} 1 & (a = b) \\ -1 & (a \neq b) \end{cases}$$

is established.

Proposition A.8 explains the effect of Init for the word-level original sentiment assignment property of CSNN.

C. Proofs of Propositions A.3–A.7

We introduce the proofs of Propositions A.3–A.7

1) Proof of Proposition A.3:

$$\frac{\partial L^{\mathbf{Q}}}{\partial c_t^{\mathbf{Q}}}$$

$$= \frac{\partial L^{\mathbf{Q}}}{\partial \boldsymbol{a}^{\mathbf{Q}}} \frac{\partial \boldsymbol{a}^{\mathbf{Q}}}{\partial(\tanh(\sum_{t=1}^{n} \boldsymbol{v}_t^{\mathbf{Q}}))} \frac{\partial(\tanh(\sum_{t=1}^{n} \boldsymbol{v}_t^{\mathbf{Q}}))}{\partial \sum_{t=1}^{n} \boldsymbol{v}_t^{\mathbf{Q}}} \frac{\partial \sum_{t=1}^{n} \boldsymbol{v}_t^{\mathbf{Q}}}{\partial g_t^{\mathbf{Q}}}$$

$$= \boldsymbol{\Delta}_o^{\mathbf{Q}} \boldsymbol{W}^o \boldsymbol{b}_t^{\mathbf{Q}} \mathrm{diag}(1 - (\tanh(\sum_{t=1}^{n} \boldsymbol{v}_t^{\mathbf{Q}}))^2) \alpha_t^{\mathbf{Q}}$$

$$= \boldsymbol{M}_t^{\mathbf{Q}^T} \boldsymbol{\Delta}_o^{\mathbf{Q}} \alpha_t^{\mathbf{Q}}$$

where

$$\boldsymbol{M}_t^{\mathbf{Q}} = \boldsymbol{W}^o \mathrm{diag}(1 - (\tanh(\sum_{t=1}^{n} \boldsymbol{v}_t^{\mathbf{Q}}))^2) \boldsymbol{b}_t^{\mathbf{Q}}$$

$$\boldsymbol{\Delta}_o^{\mathbf{Q}} = \begin{cases} \boldsymbol{a}^{\mathbf{Q}} - (1,0)^T & (d^{\mathbf{Q}} = 0) \\ \boldsymbol{a}^{\mathbf{Q}} - (0,1)^T & (d^{\mathbf{Q}} = 1) \end{cases}$$

Here, $\frac{\partial L}{\partial c^{\mathbf{Q}}}$ is positive and negative when $d^{\mathbf{Q}} = 0$ and $d^{\mathbf{Q}} = 1$, respectively, $(t = 1, 2, \cdots, n,)$ because $m_t^{\mathbf{Q}}, 0 \leq 0$ and $m_t^{\mathbf{Q}}, 1 \geq 0$ by $Update$. Therefore, the proposition is established.

2) Proof of Proposition A.4:

$$\frac{\partial L^{\mathbf{Q}}}{\partial s_t^{\mathbf{Q}}} = \frac{\partial L^{\mathbf{Q}}}{\partial c_t^{\mathbf{Q}}} \frac{\partial c_t^{\mathbf{Q}}}{\partial s_t^{\mathbf{Q}}} = \boldsymbol{M}_t^{\mathbf{Q}^T} \boldsymbol{\Delta}_o^{\mathbf{Q}} \alpha_t^{\mathbf{Q}} p_t^{\mathbf{Q}}$$

$$\frac{\partial L^{\mathbf{Q}}}{\partial s_t^{\mathbf{Q}}} = \frac{\partial L^{\mathbf{Q}}}{\partial c_t^{\mathbf{Q}}} \frac{\partial c_t^{\mathbf{Q}}}{\partial s_t^{\mathbf{Q}}} = \frac{\partial L^{\mathbf{Q}}}{\partial c_t^{\mathbf{Q}}} p_t^{\mathbf{Q}} \tag{25}$$

where word $w_i = w_t^{\mathbf{Q}}$,and the $p_{t,i}$ is the ith element of $\boldsymbol{p}_t$. Therefore, from Proposition A.3 and the above Eq (25), this proposition is established.

3) Proof of Proposition A.5: Proposition A.5 can be explained as follows. Here, Eq (15) can be explained from the property of the skip-gram method: if $\|\boldsymbol{w}_i^{em} - \boldsymbol{w}_j^{em}\| < \delta$ and the value of δ is sufficiently small, then, the appearance patterns of the word w_i and w_j are similar.

Proof.

For every $t' \in t'(\mathbf{Q}, i), \mathbf{Q} \in D_i^Q$,

$$\|\boldsymbol{s}_t^{\mathbf{Q}} - \boldsymbol{s}_{t'}^{\mathbf{Q}'}\| = \|\tanh(\boldsymbol{v}^{left^T} \overleftarrow{\boldsymbol{h}}_t^{\mathbf{Q}}) \cdot \tanh(\boldsymbol{v}^{right^T} \overrightarrow{\boldsymbol{h}}_t^{\mathbf{Q}}) -$$

$$\tanh(\boldsymbol{v}^{left^T} \overleftarrow{\boldsymbol{h}}_{t'}^{\mathbf{Q}'}) \cdot \tanh(\boldsymbol{v}^{right^T} \overrightarrow{\boldsymbol{h}}_{t'}^{\mathbf{Q}'})\|$$

$$= \|\tanh(\boldsymbol{v}^{left^T} (\overleftarrow{\boldsymbol{h}}_t^{\mathbf{Q}} - \overleftarrow{\boldsymbol{h}}_{t'}^{\mathbf{Q}'}) \cdot \tanh(\boldsymbol{v}^{right^T} \overleftarrow{\boldsymbol{h}}_t^{\mathbf{Q}})$$

$$+ \tanh(\boldsymbol{v}^{left^T} (\overleftarrow{\boldsymbol{h}}_{t'}^{\mathbf{Q}'})) \cdot \tanh(\boldsymbol{v}^{right^T} (\overrightarrow{\boldsymbol{h}}_t^{\mathbf{Q}} - \overrightarrow{\boldsymbol{h}}_{t'}^{\mathbf{Q}'}))\|$$

$$< \|\tanh(\boldsymbol{v}^{left^T} (\overleftarrow{\boldsymbol{h}}_t^{\mathbf{Q}} - \overleftarrow{\boldsymbol{h}}_{t'}^{\mathbf{Q}'})) \cdot \tanh(\boldsymbol{v}^{right^T} \overrightarrow{\boldsymbol{h}}_t^{\mathbf{Q}})\|$$

$$+ \|\tanh(\boldsymbol{v}^{left^T} (\overleftarrow{\boldsymbol{h}}_{t'}^{\mathbf{Q}'})) \cdot \tanh(\boldsymbol{v}^{right^T} (\overrightarrow{\boldsymbol{h}}_t^{\mathbf{Q}} - \overrightarrow{\boldsymbol{h}}_{t'}^{\mathbf{Q}'}))\|$$

$$< \delta(\|\boldsymbol{v}^{right}\| + \|\boldsymbol{v}^{left}\|)$$

Thus,

$$\min_{t' \in t'(Q), Q \in D_i^{\mathbf{Q}}} \|r_t^{\mathbf{I}} - r_{t'}^{\mathbf{Q}}\| < T''' \delta$$

where $T''' = \|\boldsymbol{v}^{right}\| + \|\boldsymbol{v}^{left}\|$ is established. Therefore this proposition is established.,

4) Proof of Proposition A.6: In the update process using $\mathbf{Q} \in D^{tr}$,

$$\frac{\partial L^{\mathbf{Q}}}{\partial w_j^p} = \sum_{t=1}^{n} \frac{\partial L^{\mathbf{Q}}}{\partial c_t^{\mathbf{Q}}} \frac{\partial c_t^{\mathbf{Q}}}{\partial p_t^{\mathbf{Q}}} I(w_j, p_t^{\mathbf{Q}}) = \sum_{t=1}^{n} \frac{\partial L^{\mathbf{Q}}}{\partial c_t^{\mathbf{Q}}} s_t^{\mathbf{Q}} I(w_j, p_t^{\mathbf{Q}}) \tag{26}$$

is established. Here, when $w_j = p_t^{\mathbf{Q}}$, all the values of $\{s_t^{\mathbf{Q}}\}$ satisfy Eq (18). Thus, this proposition is established.

5) Proof of Proposition A.7: Proof After the sufficient time of update iterations, for every j,

$$\boldsymbol{u}^{3,\mathbf{Q}} := \tanh(\sum_{t=1}^{n} \boldsymbol{v}_t^{\mathbf{Q}})$$

$$E\left[\frac{\partial L^{\mathbf{Q}}}{\partial w_{1,j}^O}\right] = E\left[\sum_{\mathbf{Q} \in D^{mini}} \Delta_{o,1}^{\mathbf{Q}} (u_j^{3,\mathbf{Q}})^T\right] > 0$$

$$E\left[\frac{\partial L^{\mathbf{Q}}}{\partial w_{2,j}^O}\right] = E\left[\sum_{\mathbf{Q} \in D^{mini}} \Delta_{o,2}^{\mathbf{Q}} (u_j^{3,\mathbf{Q}})^T\right] < 0$$

where $w_{i,j}^O$ is the (i,j) element of $\boldsymbol{W}^O$ and D^{mini} is the mini-batch dataset in the learning. Therefore, considering that each value of $\boldsymbol{u}^{3,\mathbf{Q}}$ is negative and positive when $d^{\mathbf{Q}} = 0$ and $d^{\mathbf{Q}} = 1$, respectively, is established because Proposition A.3 is established. Therefore, Proposition A.7 is established.

6) Proof of Proposition A.8: First,

$$\frac{\partial L^{\mathbf{Q}}}{\partial p_t^{\mathbf{Q}}} = \frac{\partial L^{\mathbf{Q}}}{\partial c_t^{\mathbf{Q}}} \frac{\partial c_t^{\mathbf{Q}}}{\partial p_t^{\mathbf{Q}}} = \boldsymbol{M}_t^{\mathbf{Q}^T} \boldsymbol{\Delta}_o^{\mathbf{Q}} \alpha_t^{\mathbf{Q}} s_t^{\mathbf{Q}},$$

$$\frac{\partial L^{\mathbf{Q}}}{\partial s_t^{\mathbf{Q}}} = \frac{\partial L^{\mathbf{Q}}}{\partial c_t^{\mathbf{Q}}} \frac{\partial c_t^{\mathbf{Q}}}{\partial s_t^{\mathbf{Q}}} = \boldsymbol{M}_t^{\mathbf{Q}^T} \boldsymbol{\Delta}_o^{\mathbf{Q}} \alpha_t^{\mathbf{Q}} p_t^{\mathbf{Q}},$$

and

$$\frac{\partial L^{\mathbf{Q}}}{\partial w_{1,j}^O} = \Delta_{o,1}^{\mathbf{Q}} (u_j^{3,\mathbf{Q}})^T, \frac{\partial L^{\mathbf{Q}}}{\partial w_{2,j}^O} = \Delta_{o,2}^{\mathbf{Q}} (u_j^{3,\mathbf{Q}})^T \tag{27}$$

where

$$\Delta_{o,1}^{\mathbf{Q}} = -\Delta_{o,2}^{\mathbf{Q}}$$

are established. Therefore,

$$-w_{1,j}^O = w_{2,j}^O (= \omega_j) \tag{28}$$

is established. Moreover,

$$\frac{\partial L^{\mathbf{Q}}}{\partial \boldsymbol{u}^{3,\mathbf{Q}}} = \frac{\partial L^{\mathbf{Q}}}{\partial \boldsymbol{a}} \frac{\partial \boldsymbol{a}}{\partial w_j^o} = \boldsymbol{\Delta}_o^{\mathbf{Q}} \boldsymbol{W}^o = \boldsymbol{\Delta}_o^{\mathbf{Q}} [-\boldsymbol{\omega}; \boldsymbol{\omega}]$$

$$= \begin{cases} 2|\Delta_{o,1}^{\mathbf{Q}}| \boldsymbol{\omega} & (d^{\mathbf{Q}} = 0) \\ -2|\Delta_{o,1}^{\mathbf{Q}}| \boldsymbol{\omega} & (d^{\mathbf{Q}} = 1) \end{cases} \tag{29}$$

is established. Therefore, after the sufficient time of iterations,

$$E[\boldsymbol{u}^{3,\mathbf{Q}}] = \begin{cases} -k\boldsymbol{\omega} & (d^{\mathbf{Q}} = 0) \\ k\boldsymbol{\omega} & (d^{\mathbf{Q}} = 1) \end{cases}$$

where $k > 0$ is expected to be established.

Therefore,

$$\boldsymbol{M}_t^{\mathbf{Q}^T} \boldsymbol{\Delta}_o^{\mathbf{Q}} = \boldsymbol{b}_t^{\mathbf{Q}^T} \mathrm{diag}(1 - (\tanh(\sum_{t=1}^{n} \boldsymbol{v}_t^{\mathbf{Q}}))^2) \boldsymbol{W}^{oT} \boldsymbol{\Delta}_o^{\mathbf{Q}}$$

$$= \boldsymbol{b}_t^{\mathbf{Q}^T} \mathrm{diag}(\mathbf{1} - (\tanh(\sum_{t=1}^{n} \boldsymbol{v}_t^{\mathbf{Q}}))^2) \boldsymbol{W}^{oT} \boldsymbol{\Delta}_o^{\mathbf{Q}}$$

$$= \begin{cases} 2|\Delta_{o,2}^{\mathbf{Q}}| \boldsymbol{b}_t^{\mathbf{Q}^T} \mathrm{diag}(\mathbf{1} - (\tanh(\sum_{t=1}^{n} \boldsymbol{v}_t^{\mathbf{Q}}))^2) \boldsymbol{\omega} & (d^{\mathbf{Q}} = 0) \\ -2|\Delta_{o,2}^{\mathbf{Q}}| \boldsymbol{b}_t^{\mathbf{Q}^T} \mathrm{diag}(\mathbf{1} - (\tanh(\sum_{t=1}^{n} \boldsymbol{v}_t^{\mathbf{Q}}))^2) \boldsymbol{\omega} & (d^{\mathbf{Q}} = 1) \end{cases}$$

Moreover, after the sufficient times of iterations,

$$\mathrm{sign}(\omega_1) = \mathrm{sign}(\omega_2) = \cdots = \mathrm{sign}(\omega_k) \tag{30}$$

is established because Eq (27) and

$$\boldsymbol{u}^{3,\mathbf{Q}} = \sum_{t=1}^{n} \boldsymbol{v}_t^{\mathbf{Q}} = \sum_{t=1}^{n} g_t^{\mathbf{Q}} \boldsymbol{b}_t^{\mathbf{Q}}$$

where

$$\mathrm{sign}(v_{t,1}^{\mathbf{Q}}) = \mathrm{sign}(v_{t,2}^{\mathbf{Q}}) = \cdots = \mathrm{sign}(v_{t,k}^{\mathbf{Q}}).$$

are satisfied, and in sufficient times of cases,

$$\boldsymbol{u}^{3,\mathbf{Q}} \simeq g_{\hat{t}}^{\mathbf{Q}} \boldsymbol{b}_{\hat{t}}^{\mathbf{Q}} \tag{31}$$

where

$$\hat{t} = \mathrm{argmax}_t g_t^{\mathbf{Q}}.$$

Eq (31) occurs because

$$E[p_t^{\mathbf{Q}} | w_t^{\mathbf{Q}} \in S^d] >> E[p_t^{\mathbf{Q}} | w_t^{\mathbf{Q}} \notin S^d]$$

$$E[\alpha_t^{\mathbf{Q}} | \min_{w_j \in S^d} |w_t^{\mathbf{Q}} - \boldsymbol{w}_j^{em}| < \epsilon] >> E[\alpha_t^{\mathbf{Q}} | \min_{w_j \in S^d} |w_t^{\mathbf{Q}} - \boldsymbol{w}_j^{em}| \geq \epsilon],$$

where ϵ is sufficiently small, and $|S^d|$ is sufficiently small.

Therefore,

$$\mathrm{sign}(\boldsymbol{M}_t^{\mathbf{Q}^T} \boldsymbol{\Delta}_o^{\mathbf{Q}}) = \begin{cases} -\chi & (d^{\mathbf{Q}} = 0) \\ \chi & (d^{\mathbf{Q}} = 1) \end{cases}.$$

Thus,

$$s_t^{\mathbf{Q}} \simeq \epsilon - \sum_{\mathbf{Q}' \in \Omega^{tr}} \sum_{t'=1}^{|\mathbf{Q}'|} \boldsymbol{M}_{t'}^{\mathbf{Q}'^T} \boldsymbol{\Delta}_o^{\mathbf{Q}'} \alpha_{t'}^{\mathbf{Q}'} p_{t'}^{\mathbf{Q}'} I(|e_t^{\mathbf{Q}} - e_{t'}^{\mathbf{Q}}| < \epsilon, True).$$

$$\simeq - \sum_{\mathbf{Q}' \in \Omega^{tr}} \sum_{t'=1}^{|\mathbf{Q}'|} \boldsymbol{M}_{t'}^{\mathbf{Q}'^T} \boldsymbol{\Delta}_o^{\mathbf{Q}'} \alpha_{t'}^{\mathbf{Q}'} p_{t'}^{\mathbf{Q}'} I(|e_t^{\mathbf{Q}} - e_{t'}^{\mathbf{Q}}| < \epsilon, True) I(w_{t'}^{\mathbf{Q}} \in S^d, True).$$

Here,

$$\mathrm{sign}(\boldsymbol{M}_{t'}^{\mathbf{Q}'^T} \boldsymbol{\Delta}_o^{\mathbf{Q}'} \alpha_{t'}^{\mathbf{Q}'} p_{t'}^{\mathbf{Q}'}) = \chi R(w_{t'}^{\mathbf{Q}'})$$

Thus, if $w_t^{\mathbf{Q}} \in S^d$, then,

$$\mathrm{sign}(s_t^{\mathbf{Q}}) = -\chi R(w_t^{\mathbf{Q}})$$

is established because each $w_j \in S^d$ satisfies Condition A.1. Therefore, in such a situation,

$$\mathrm{sign}\left(\frac{\partial L^{\mathbf{Q}}}{\partial p_t^{\mathbf{Q}}}\right) = \mathrm{sign}(\boldsymbol{M}_t^{\mathbf{Q}^T} \boldsymbol{\Delta}_o^{\mathbf{Q}} \alpha_t^{\mathbf{Q}} s_t^{\mathbf{Q}})$$

$$\simeq \begin{cases} \chi^2 R(w_t^{\mathbf{Q}}) = R(w_t^{\mathbf{Q}}) & (d^{\mathbf{Q}} = 0) \\ -\chi^2 R(w_t^{\mathbf{Q}}) = -R(w_t^{\mathbf{Q}}) & (d^{\mathbf{Q}} = 1). \end{cases}$$

Therefore, if $|e_t^{\mathbf{Q}} - \boldsymbol{w}_j^{em}| < \epsilon$ where $\epsilon > 0$ is sufficiently small, then, the following equation is established.

$$\frac{\partial L^{\mathbf{Q}'}}{\partial p_{t'}^{\mathbf{Q}'}} = \boldsymbol{M}_{t'}^{\mathbf{Q}'^T} \boldsymbol{\Delta}_o^{\mathbf{Q}'} \alpha_{t'}^{\mathbf{Q}'} s_{t'}^{\mathbf{Q}'} \simeq \boldsymbol{M}_{t'}^{\mathbf{Q}'^T} \boldsymbol{\Delta}_o^{\mathbf{Q}'} \alpha_{t'}^{\mathbf{Q}'} s_{t'}^{\mathbf{Q}'(w_{t'}^{\mathbf{Q}'}, w_j)}$$

Therefore,

$$\mathrm{sign}\left(\frac{\partial L^{\mathbf{Q}'}}{\partial p_{t'}^{\mathbf{Q}'}}\right) = \begin{cases} R(w_{t'}^{\mathbf{Q}'(w_{t'}^{\mathbf{Q}'}, w_j)}) & (d^{\mathbf{Q}} = 0) \\ -R(w_{t'}^{\mathbf{Q}'(w_{t'}^{\mathbf{Q}'}, w_j)}) & (d^{\mathbf{Q}} = 1) \end{cases}.$$

because

$$s_{t'}^{\mathbf{Q}'} \simeq s_{t'}^{\mathbf{Q}'(w_{t'}^{\mathbf{Q}'}, w_t^{\mathbf{Q}})}$$

due to $|e_t^{\mathbf{Q}} - e_{t'}^{\mathbf{Q}'}| < \epsilon$ where $\epsilon > 0$ is sufficiently small, is established.

Thus, Proposition A.8 is established.