Proof of Theorem 1: It is clear that $P(w^{(t)}, \gamma^{(t)}) \leq \gamma^{(t)}\Omega(w^{(t-1)})$, since $(w = w^{(t-1)}, \zeta = 0)$ is a feasible solution for problem (10). So,

$$
\gamma^{(t)}\zeta^{(t)} + \frac{1}{2}\|w^{(t)} - w^{(t-1)}\|^2 \leq \gamma^{(t)}\Omega(w^{(t-1)}) - \Omega(w^{(t)})).
$$

The LHS is lower bounded by $\gamma^{(t)}g(w^{(t-1)})T(w^{(t)} - w^{(t-1)}) + \frac{1}{2}\|w^{(t)} - w^{(t-1)}\|^2 \geq \gamma^{(t)}g(w^{(t-1)})^2$. Then, we would have $P(w^{(t)}, \gamma^{(t)}) \geq \gamma^{(t)}g(w^{(t-1)})^2$. So, $-\frac{\gamma^{(t)}}{2}R^2 \leq -\gamma^{(t)}i_g(w^{(t-1)})^2 \leq P(w^{(t)}, \gamma^{(t)}) \leq \gamma_0\Omega(w^{(t)}) = \gamma_0D$. It is clear that in problem (5), $R = C\max \{\max_i \|B_{ij}\|, \max_j \|B_{ij}\|\}$. \hfill \Box

Proof of Theorem 2: The algorithm terminates under two conditions. We show under either of the two conditions, the algorithm terminates after finite steps.

1) Condition 1: As shown in step (8) of Table 1, if $\gamma^{(t)}$ is below a specific threshold $\epsilon_1$, the algorithm terminates. It is clear that the method cannot execute the step (7) to step (9) more than $\log \frac{\epsilon}{\epsilon_0}/\log(\eta)$ times.

2) Condition 2: Suppose the conditions stated in step (4) hold. By rearranging Eq.(1), we can get: $\zeta^{(t)} + \Omega(w^{(t)}) - \Omega(w^{(t-1)}) \leq \frac{1}{2\gamma_0^2}\|w^{(t)} - w^{(t-1)}\|^2$. So, $\zeta^{(t)} + \Omega(w^{(t)}) - \Omega(w^{(t-1)}) \geq \frac{1}{2\gamma_0^2}\|w^{(t)} - w^{(t-1)}\|^2 \geq \frac{1}{2\gamma_0^2}g^2$. It means that, each execution of step (4) to step (6) will decrease $F(w)$ by at least $\frac{1}{2\gamma_0^2}g^2$. Since $F(w)$ is upper bounded by $E$ and lower bounded by 0. Step (4) to step (6) cannot be executed for at most $\frac{2E}{\gamma_0^2m}\eta$ steps.

By summarizing these two conditions, we can conclude that the total number iterations should not exceed $\log \frac{\epsilon}{\epsilon_0}/\log(\eta) + \frac{2E}{\gamma_0^2m}\eta$.

\hfill \Box

1Here, to avoid confusion, $g(w^{(t-1)}) = g_{t-1}$.

The problem (5) is upper bounded by $nC$, since $(w = 0, \xi_1 = 1), i = 1, \ldots, n$ is a feasible solution. So, the maximum number of required iterations is $\log \frac{\epsilon}{\epsilon_0}/\log(\eta) + \frac{2nC}{\gamma_0^2m}\eta$.

Proof of Theorem 3: The Rademacher Complexity of $F_B$ can be calculated as:

$$
R_n(F_B) = E_{\omega}[\sup_{f_{FB}}\frac{2}{n}\sum_{i=1}^{n} \sigma_i f_{\omega}(B_{ij}, B_{ij})] \\
\leq \frac{2}{n}E_{\omega}[\sum_{i=1}^{n} \sigma_i C_{\max} \|B_{ij}\|, B_{ij} + (1-\lambda)w^T_{ij}B_{ij}] \\
\leq \frac{2B}{n}E_{\omega}[\sum_{i=1}^{n} \sigma_i \lambda_{\max} \|B_{ij}\|, B_{ij} + \frac{B}{n}\sum_{i=1}^{n} \sigma_i (1-\lambda)w^T_{ij}B_{ij}] \\
\leq \frac{2B}{n}E_{\omega}[\sum_{i=1}^{n} \lambda_{\max}^2 K(B_{ij}, B_{ij}) + \frac{B}{n}\sum_{i=1}^{n} (1-\lambda)^2 K(B_{ij}, B_{ij})] \\
\leq \frac{2B}{n} \max_{\phi_{ij} \geq 0, \phi_{ij}} \left( \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_{ij}^2 \phi_{ij}^2 K(B_{ij}, B_{ij}) + \sum_{i=1}^{n} (1-\lambda)^2 K(B_{ij}, B_{ij}) \right)
$$

where $j^*_i = \arg \max_j w^T_{ij}B_{ij}$, $w^*_i = \arg \max_{w_i} \sum_{i=1}^{n} \sigma_i \lambda_{\max} \|B_{ij}w^T_{ij}\|$ and $w^*_2 = \arg \max_{w_i} \sum_{i=1}^{n} (1-\sigma_i) \lambda_{\max} \|B_{ij}w^T_{ij}\|$.

It can be seen that the Rademacher Complexity is composed of two different parts. The first part is mainly derived from the multiple instance setting, while the second part is from the traditional setting.

Proof of Theorem 5: It is clear that

$$
\frac{2B}{n} \max_{\phi_{ij} \geq 0, \phi_{ij}} \left( \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_{ij}^2 \phi_{ij}^2 K(B_{ij}, B_{ij}) + \sum_{i=1}^{n} (1-\lambda)^2 K(B_{ij}, B_{ij}) \right) \\
\leq \max \left( \frac{C_2}{C_1} a, 1 - \frac{C_2}{C_1} (1-a) \right) C_1 + (1-a) C_2
$$
If $C_1 < C_2$, \(\max\{\frac{C_2}{C_1}a, 1 - \frac{C_2}{C_1}(1-a)\} = \frac{C_2}{C_1}a\), So,
\[
\max\{\frac{C_2}{C_1}a, 1 - \frac{C_2}{C_1}(1-a)\}C_1 + (1-a)C_2 = \frac{C_2}{C_1}aC_1 + (1-a)C_2 = C_2.
\]
If $C_2 \leq C_1$, \(\max\{\frac{C_2}{C_1}a, 1 - \frac{C_2}{C_1}(1-a)\} = 1 - \frac{C_2}{C_1}(1-a)\), So,
\[
\max\{\frac{C_2}{C_1}a, 1 - \frac{C_2}{C_1}(1-a)\}C_1 + (1-a)C_2 = (1 - \frac{C_2}{C_1}(1-a))C_1 + (1-a)C_2 = C_1.
\]
By summarizing these two conditions, we can get the conclusion. □