9. Supplementary Material

9.1. Proof of Lemma 6.1

Note that the update rule (24) can be written as

$$\mathbf{Z}^{k+1} := \mathbf{Z}^k + \mathbf{W}\mathbf{Z}^k - \tilde{\mathbf{W}}\mathbf{Z}^{k-1} - \alpha(\hat{\mathcal{B}}^k(\mathbf{Z}^{k+1}) - \hat{\mathcal{B}}^{k-1}(\mathbf{Z}^k)), \tag{47}$$

from the definition of $\tilde{\mathbf{W}}$. To prove the first part of the lemma, by summing (47) from $k = 1$ to t and (25), one has

$$\mathbf{Z}^{t+1} = (\mathbf{W} - \tilde{\mathbf{W}}) \sum_{k=0}^{t} \mathbf{Z}^k + \tilde{\mathbf{W}}\mathbf{Z}^t - \alpha\hat{\mathcal{B}}^t(\mathbf{Z}^{t+1}). \tag{48}$$

From the definition of $\mathbf{U}$ and $\mathbf{Q}^t$ and the identity $\mathbf{I} = 2\tilde{\mathbf{W}} - \mathbf{W}$, we have

$$\alpha\hat{\mathcal{B}}^t(\mathbf{Z}^{t+1}) = \tilde{\mathbf{W}}(\mathbf{Z}^t - \mathbf{Z}^{t+1}) - \mathbf{U}\mathbf{Q}^{t+1}. \tag{49}$$

By subtracting the optimality condition (15), we have the result.

From first part, we have

$$\begin{aligned}
&\langle \mathbf{Z}^{t+1} - \mathbf{Z}^*, \alpha[\mathcal{B}(\mathbf{Z}^*) - \hat{\mathcal{B}}^t(\mathbf{Z}^{t+1})] \rangle \\
=&\langle \mathbf{Z}^{t+1} - \mathbf{Z}^*, -\tilde{\mathbf{W}}(\mathbf{Z}^t - \mathbf{Z}^{t+1}) + \mathbf{U}(\mathbf{Q}^{t+1} - \mathbf{Q}^*) \rangle \\
=&\langle \mathbf{Z}^{t+1} - \mathbf{Z}^*, \mathbf{Z}^{t+1} - \mathbf{Z}^t \rangle_{\tilde{\mathbf{W}}} + \langle \mathbf{Z}^{t+1} - \mathbf{Z}^*, \mathbf{U}(\mathbf{Q}^{t+1} - \mathbf{Q}^*) \rangle \\
=&\langle \mathbf{Z}^{t+1} - \mathbf{Z}^*, \mathbf{Z}^{t+1} - \mathbf{Z}^t \rangle_{\tilde{\mathbf{W}}} + \langle \mathbf{Q}^{t+1} - \mathbf{Q}^t, \mathbf{Q}^{t+1} - \mathbf{Q}^* \rangle,
\end{aligned} \tag{50}$$

where the last equality uses the definition of $\mathbf{Q}^t$ and that $\mathbf{U}\mathbf{Z}^* = \mathbf{0}$. By applying the generalized Law of cosines $2\langle a, b\rangle = \|a\|^2 + \|b\|^2 - \|a - b\|^2$ with $a = \mathbf{X}^{t+1} - \mathbf{X}^*$ and $b = \mathbf{X}^{t+1} - \mathbf{X}^t$, we have the second part.

9.2. Proof of Lemma 6.2

We have $T^{t+1} \geq \frac{1}{L} S^{t+1}$ from the definition of cocoerciveness. Expanding the definition of $\hat{\mathcal{B}}^t(\mathbf{Z}^{t+1})$, we have

$$\begin{aligned}
&\mathbb{E}\langle \mathbf{Z}^{t+1} - \mathbf{Z}^*, \mathcal{B}(\mathbf{Z}^*) - \hat{\mathcal{B}}^t(\mathbf{Z}^{t+1}) \rangle \\
=& \sum_{n=1}^{N} -\mathbb{E}_{i_n^t} \langle \mathbf{z}_{n,i_n^t}^{t+1} - \mathbf{z}^*, \mathcal{B}_{n,i_n^t}(\mathbf{z}_{n,i_n^t}^{t+1}) - \mathcal{B}_{n,i_n^t}(\mathbf{z}^*) \rangle \\
&+ \mathbb{E}_{i_n^t} \langle \mathbf{z}_{n,i_n^t}^{t+1} - \mathbf{z}^*, [\mathcal{B}_{n,i_n^t}(\mathbf{y}_{n,i_n^t}^t) - \mathcal{B}_{n,i_n^t}(\mathbf{z}^*)] - [\frac{1}{q}\sum_{i=1}^{q} \mathcal{B}_{n,i}(\mathbf{y}_{n,i}^t) - \mathcal{B}_n(\mathbf{z}^*)] \rangle.
\end{aligned} \tag{51}$$

The first term is exactly $-\frac{1}{2}T^{t+1}$, and is bounded by $-\frac{1}{2}T^{t+1} \leq -\frac{\theta}{2L}S^{t+1} - \frac{1-\theta}{2}T^{t+1}$ for $0 \leq \theta \leq 1$. Since

$$\mathbb{E}_{i_n^t}\{[\mathcal{B}_{n,i_n^t}(\mathbf{y}_{n,i_n^t}^t) - \mathcal{B}_{n,i_n^t}(\mathbf{z}^*)] - [\frac{1}{q}\sum_{i=1}^{q} \mathcal{B}_{n,i}(\mathbf{y}_{n,i}^t) - \mathcal{B}_n(\mathbf{z}^*)]\} = \mathbf{0}, \tag{52}$$

and $\mathbf{z}_n^t$ is independent of i_n^t, we have

$$\mathbb{E}_{i_n^t} \langle \mathbf{z}_n^t - \mathbf{z}^*, [\mathcal{B}_{n,i_n^t}(\mathbf{y}_{n,i_n^t}^t) - \mathcal{B}_{n,i_n^t}(\mathbf{z}^*)] - [\frac{1}{q}\sum_{i=1}^{q} \mathcal{B}_{n,i}(\mathbf{y}_{n,i}^t) - \mathcal{B}_n(\mathbf{z}^*)] \rangle = 0. \tag{53}$$

We bound the second term by

$$\begin{aligned}
&\sum_{n=1}^{N}\mathbb{E}_{i_n^t}\langle \mathbf{z}_{n,i_n^t}^{t+1}-\mathbf{z}^*,[\mathcal{B}_{n,i_n^t}(\mathbf{y}_{n,i_n^t}^t)-\mathcal{B}_{n,i_n^t}(\mathbf{z}^*)]-[\frac{1}{q}\sum_{i=1}^{q}\mathcal{B}_{n,i}(\mathbf{y}_{n,i}^t)-\mathcal{B}_n(\mathbf{z}^*)]\rangle\\
=&\sum_{n=1}^{N}\mathbb{E}_{i_n^t}\langle \mathbf{z}_{n,i_n^t}^{t+1}-\mathbf{z}_n^t,[\mathcal{B}_{n,i_n^t}(\mathbf{y}_{n,i_n^t}^t)-\mathcal{B}_{n,i_n^t}(\mathbf{z}^*)]-[\frac{1}{q}\sum_{i=1}^{q}\mathcal{B}_{n,i}(\mathbf{y}_{n,i}^t)-\mathcal{B}_n(\mathbf{z}^*)]\rangle\\
\leq&\sum_{n=1}^{N}\frac{\eta}{2}\mathbb{E}_{i_n^t}\|[\mathcal{B}_{n,i_n^t}(\mathbf{y}_{n,i_n^t}^t)-\mathcal{B}_{n,i_n^t}(\mathbf{z}^*)]-[\frac{1}{q}\sum_{i=1}^{q}\mathcal{B}_{n,i}(\mathbf{y}_{n,i}^t)-\mathcal{B}_n(\mathbf{z}^*)]\|^2+\frac{1}{2\eta}\mathbb{E}_{i_n^t}\|\mathbf{z}_{n,i_n^t}^{t+1}-\mathbf{z}_n^t\|^2\\
\leq&\sum_{n=1}^{N}\frac{\eta}{2}\mathbb{E}_{i_n^t}\|\mathcal{B}_{n,i_n^t}(\mathbf{y}_{n,i_n^t}^t)-\mathcal{B}_{n,i_n^t}(\mathbf{z}^*)\|^2+\frac{1}{2\eta}\mathbb{E}_{i_n^t}\|\mathbf{z}_{n,i_n^t}^{t+1}-\mathbf{z}_n^t\|^2\\
=&\frac{1}{2\eta}\mathbb{E}\|\mathbf{Z}^{t+1}-\mathbf{Z}^t\|^2+\frac{\eta}{4}D^t,
\end{aligned}\tag{54}$$

where we use $\langle a,b\rangle\leq\frac{1}{2\eta}\|a\|^2+\frac{\eta}{2}\|b\|^2$ in first inequality and $\|a-\mathbb{E}a\|^2\leq\|a\|^2$ in the second one.

9.3. Proof of Lemma 6.3

From the definition of $\hat{\mathcal{B}}^t(\mathbf{Z}^{t+1})$, on node n, we have

$$\hat{\mathcal{B}}_n^t(\mathbf{z}_n^{t+1})-\mathcal{B}_n(\mathbf{z}^*)=[\mathcal{B}_{n,i_n^t}(\mathbf{z}_{n,i_n^t}^{t+1})-\mathcal{B}_{n,i_n^t}(\mathbf{z}^*)]-[\mathcal{B}_{n,i_n^t}(\mathbf{y}_{n,i_n^t}^t)-\mathcal{B}_{n,i_n^t}(\mathbf{z}^*)]+[\frac{1}{q}\sum_{i=1}^{q}\mathcal{B}_{n,i}(\mathbf{y}_{n,i}^t)-\mathcal{B}_n(\mathbf{z}^*)].\tag{55}$$

Using $\|a+b\|^2\leq 2\|a\|^2+2\|b\|^2$, we have

$$\begin{aligned}
&\mathbb{E}\|\hat{\mathcal{B}}^t(\mathbf{Z}^{t+1})-\mathcal{B}(\mathbf{Z}^*)\|^2\\
\leq&\sum_{n=1}^{N}2\mathbb{E}_{i_n^t}\|\mathcal{B}_{n,i_n^t}(\mathbf{z}_{n,i_n^t}^{t+1})-\mathcal{B}_{n,i_n^t}(\mathbf{z}^*)\|^2+2\mathbb{E}_{i_n^t}\|[\mathcal{B}_{n,i_n^t}(\mathbf{y}_{n,i_n^t}^t)-\mathcal{B}_{n,i_n^t}(\mathbf{z}^*)]-[\frac{1}{q}\sum_{i=1}^{q}\mathcal{B}_{n,i}(\mathbf{y}_{n,i}^t)-\mathcal{B}_n(\mathbf{z}^*)]\|^2\\
\leq& S^{t+1}+D^t,
\end{aligned}\tag{56}$$

where the last inequality uses the definition of D^t and S^{t+1} and $\|a-\mathbb{E}a\|^2\leq\|a\|^2$.

9.4. Proof of Lemma 6.4

Expand $\|\mathbf{X}^t-\mathbf{X}^*\|_M^2$ by the definition of $\mathbf{X}^t$ and $\|\cdot\|_M$ and suppose $\mathbf{Z}^{t+1}$ and $\mathbf{Q}^{t+1}$ are generated from some fixed $i_n^t, n\in[N]$. Using $\|a+b\|^2\leq 2\|a\|^2+2\|b\|^2$, we have

$$\begin{aligned}
\|\mathbf{X}^t-\mathbf{X}^*\|_M^2&=\|\mathbf{Z}^t-\mathbf{Z}^*\|_{\tilde{\mathbf{W}}}^2+\|\mathbf{Q}^t-\mathbf{Q}^*\|^2\\
&\leq 2\|\mathbf{Z}^{t+1}-\mathbf{Z}^t\|_{\tilde{\mathbf{W}}}^2+2\|\mathbf{Z}^{t+1}-\mathbf{Z}^*\|_{\tilde{\mathbf{W}}}^2+2\|\mathbf{Q}^{t+1}-\mathbf{Q}^t\|^2+2\|\mathbf{Q}^{t+1}-\mathbf{Q}^*\|^2.
\end{aligned}\tag{57}$$

We now bound the second term and last term. Using

$$\|\mathbf{Z}^{t+1}-\mathbf{Z}^*\|_{\tilde{\mathbf{W}}}^2\leq\|\mathbf{Z}^{t+1}-\mathbf{Z}^*\|^2\tag{58}$$

since $\tilde{\mathbf{W}}\preccurlyeq I$, and the μ-strongly monotonicity of $\mathcal{B}_{n,i_n^t}$, we have

$$\|\mathbf{Z}^{t+1}-\mathbf{Z}^*\|_{\tilde{\mathbf{W}}}^2\leq\frac{1}{\mu}\sum_{n=1}^{N}\langle\mathbf{z}_{n,i_n^t}^{t+1}-\mathbf{z}^*,\mathcal{B}_{n,i_n^t}(\mathbf{z}_{n,i_n^t}^{t+1})-\mathcal{B}_{n,i_n^t}(\mathbf{z}^*)\rangle.\tag{59}$$

From the construction of $\mathbf{Q}^{t+1}$ and $\mathbf{Q}^*$, every column of $\mathbf{Q}^{t+1}-\mathbf{Q}^*$ is in $\mathrm{span}(U)$, thus we have

$$\gamma\|\mathbf{Q}^{t+1}-\mathbf{Q}^*\|^2\leq\|U(\mathbf{Q}^{t+1}-\mathbf{Q}^*)\|^2,\tag{60}$$

where γ is the smallest nonzero singular value of $U^2 = \tilde{\mathbf{W}} - W$. From Lemma 6.1, we write

$$\begin{aligned}\|U(\mathbf{Q}^{t+1} - \mathbf{Q}^*)\|^2 &= \|\alpha[\hat{\mathcal{B}}^t(\mathbf{Z}^{t+1}) - \mathcal{B}(\mathbf{Z}^*)] + \tilde{\mathbf{W}}(\mathbf{Z}^{t+1} - \mathbf{Z}^t)\|^2 \\ &\leq 2\alpha^2\|\hat{\mathcal{B}}^t(\mathbf{Z}^{t+1}) - \mathcal{B}(\mathbf{Z}^*)\|^2 + 2\|\mathbf{Z}^{t+1} - \mathbf{Z}^t\|^2_{\tilde{\mathbf{W}}}. \end{aligned} \tag{61}$$

Substituting these two upper bounds into (57), we have

$$\begin{aligned}\|\mathbf{X}^t - \mathbf{X}^*\|^2_M \leq (2 + \frac{4}{\gamma})\|\mathbf{Z}^{t+1} - \mathbf{Z}^t\|^2_{\tilde{\mathbf{W}}} + 2\|\mathbf{Q}^{t+1} - \mathbf{Q}^t\|^2 + \frac{2}{\mu}\sum_{n=1}^{N}\langle \mathbf{z}^{t+1}_{n,i^t_n} - \mathbf{z}^*, \mathcal{B}_{n,i^t_n}(\mathbf{z}^{t+1}_{n,i^t_n}) - \mathcal{B}_{n,i^t_n}(\mathbf{z}^*)\rangle \\ + \frac{4\alpha^2}{\gamma}\|\hat{\mathcal{B}}^t(\mathbf{Z}^{t+1}) - \mathcal{B}(\mathbf{Z}^*)\|^2. \end{aligned} \tag{62}$$

Taking expectation and using Lemma 6.3, we have the result.

9.5. Proof of Theorem 6.1

From Lemma 6.1 and 6.2, we have

$$\begin{aligned}&\mathbb{E}\|\mathbf{X}^{t+1} - \mathbf{X}^*\|^2_{\mathbf{M}} - \|\mathbf{X}^t - \mathbf{X}^*\|^2_{\mathbf{M}} + \mathbb{E}\|\mathbf{X}^{t+1} - \mathbf{X}^t\|^2_{\mathbf{M}} \\ &= 2\alpha\mathbb{E}\langle \mathbf{Z}^{t+1} - \mathbf{Z}^*, \mathcal{B}(\mathbf{Z}^*) - \hat{\mathcal{B}}^t(\mathbf{Z}^{t+1})\rangle \\ &\leq \frac{\alpha}{\eta}\mathbb{E}\|\mathbf{Z}^{t+1} - \mathbf{Z}^t\|^2 + \frac{\eta\alpha}{2}D^t - \frac{\theta\alpha}{L}S^{t+1} - (1-\theta)\alpha T^{t+1}. \end{aligned} \tag{63}$$

Also for D^{t+1}, we have

$$\begin{aligned}\mathbb{E}D^{t+1} &= \sum_{n=1}^{N}\frac{2}{q}\sum_{i=1}^{q}\mathbb{E}_{i^t_n}\|\mathcal{B}_{n,i}(\mathbf{y}^{t+1}_{n,i}) - \mathcal{B}_{n,i}(\mathbf{z}^*)\|^2 \\ &= \sum_{n=1}^{N}\frac{2}{q}\sum_{i=1}^{q}\{\frac{1}{q}\|\mathcal{B}_{n,i}(\mathbf{z}^{t+1}_{n,i}) - \mathcal{B}_{n,i}(\mathbf{z}^*)\|^2 + (1-\frac{1}{q})\|\mathcal{B}_{n,i}(\mathbf{y}^t_{n,i}) - \mathcal{B}_{n,i}(\mathbf{z}^*)\|^2\} \\ &= (1-\frac{1}{q})D^t + \frac{1}{q}S^{t+1}. \end{aligned} \tag{64}$$

By adding cD^{t+1} and rearranging terms, we have

$$\begin{aligned}\mathbb{E}[\|\mathbf{X}^{t+1} - \mathbf{X}^*\|^2_M + cD^{t+1}] \leq& \|\mathbf{X}^t - \mathbf{X}^*\|^2_{\mathbf{M}} - \mathbb{E}\|\mathbf{X}^{t+1} - \mathbf{X}^t\|^2_{\mathbf{M}} + (1-\frac{1}{q})cD^t + \frac{c}{q}S^{t+1} \\ &+ \frac{\alpha}{\eta}\mathbb{E}\|\mathbf{Z}^{t+1} - \mathbf{Z}^t\|^2 + \frac{\eta\alpha}{2}D^t - \frac{\theta\alpha}{L}S^{t+1} - (1-\theta)\alpha T^{t+1}. \end{aligned} \tag{65}$$

If we further have

$$\begin{aligned}(1-\delta)[\|\mathbf{X}^t - \mathbf{X}^*\|^2_M + cD^t] \geq& \|\mathbf{X}^t - \mathbf{X}^*\|^2_{\mathbf{M}} - \mathbb{E}\|\mathbf{X}^{t+1} - \mathbf{X}^t\|^2_{\mathbf{M}} + (1-\frac{1}{q})cD^t + \frac{c}{q}S^{t+1} \\ &+ \frac{\alpha}{\eta}\mathbb{E}\|\mathbf{Z}^{t+1} - \mathbf{Z}^t\|^2 + \frac{\eta\alpha}{2}D^t - \frac{\theta\alpha}{L}S^{t+1} - (1-\theta)\alpha T^{t+1}, \end{aligned} \tag{66}$$

then we have the result. The above inequality is equivalent to

$$\begin{aligned}&(\frac{c}{q} - c\delta - \frac{\alpha\eta}{2})D^t + (\frac{\alpha\theta}{L} - \frac{c}{q})S^{t+1} + \alpha(1-\theta)T^{t+1} \\ &\geq \underbrace{\delta\|\mathbf{X}^t - \mathbf{X}^*\|^2_{\mathbf{M}} - \|\mathbf{X}^{t+1} - \mathbf{X}^t\|^2_{\mathbf{M}} + \frac{\alpha}{\eta}\mathbb{E}\|\mathbf{Z}^{t+1} - \mathbf{Z}^t\|^2,}_{\Lambda} \end{aligned} \tag{67}$$

and hence a sufficient condition is that an upper bound of the right hand side is less than the left hand side.

To bound Λ, using Lemma 6.4 for the first term, the definition of $\|\mathbf{X}^{t+1}-\mathbf{X}^t\|_{\mathbf{M}}^2$ for the second term, and

$$\frac{1}{2}\|\mathbf{Z}^{t+1}-\mathbf{Z}^t\|^2 \le \|\mathbf{Z}^{t+1}-\mathbf{Z}^t\|_{\tilde{\mathbf{W}}}^2 \tag{68}$$

for the third term since $\frac{1}{2}I \preccurlyeq \tilde{\mathbf{W}}$, we have

$$\begin{aligned}\Lambda \le& \delta[(2+\frac{4}{\gamma})\mathbb{E}\|\mathbf{Z}^{t+1}-\mathbf{Z}^t\|_{\tilde{\mathbf{W}}}^2 + \frac{1}{\mu}T^{t+1} + 2\mathbb{E}\|\mathbf{Q}^{t+1}-\mathbf{Q}^t\|^2 + \frac{4\alpha^2}{\gamma}(S^{t+1}+D^t)]\\ &- \mathbb{E}\|\mathbf{Z}^{t+1}-\mathbf{Z}^t\|_{\tilde{\mathbf{W}}}^2 - \mathbb{E}\|\mathbf{Q}^{t+1}-\mathbf{Q}^t\|^2 + \frac{2\alpha}{\eta}\mathbb{E}\|\mathbf{Z}^{t+1}-\mathbf{Z}^t\|_{\tilde{\mathbf{W}}}^2.\end{aligned} \tag{69}$$

Uniting like terms gives us the following sufficient condition for Theorem 6.1 to stand:

$$\begin{aligned}&(\frac{c}{q} - c\delta - \frac{\alpha\eta}{2} - \frac{4\delta\alpha^2}{\gamma})D^t + (\frac{\alpha\theta}{L} - \frac{c}{q} - \frac{4\delta\alpha^2}{\gamma})S^{t+1} + (\alpha(1-\theta) - \frac{\delta}{\mu})T^{t+1}\\ &+ (1-2\delta)\mathbb{E}\|\mathbf{Q}^{t+1}-\mathbf{Q}^t\|^2 + (1-(2+\frac{4}{\gamma})\delta - \frac{2\alpha}{\eta})\mathbb{E}\|\mathbf{Z}^{t+1}-\mathbf{Z}^t\|_{\tilde{\mathbf{W}}}^2 \ge 0.\end{aligned} \tag{70}$$

Since every term in the above inequality is nonnegative, this inequality holds when every bracket is nonnegative. Let

$$\alpha = \frac{\tau}{L}, \eta = 4\alpha, \theta = \frac{1}{2}, c = \frac{mq}{L^2}, \tag{71}$$

where τ and m are constant to be set. The non-negativity of of the first two brackets equivalents to

$$\begin{cases} c(\frac{1}{3q}-\delta) + \frac{2m}{3L^2} - \frac{2\tau^2}{L^2} - \frac{\delta}{\gamma}\frac{4\tau^2}{L^2} \ge 0 \\ \frac{\tau}{2L^2} - \frac{m}{L^2} - \frac{\delta}{\gamma}\frac{4\tau^2}{L^2} \ge 0 \end{cases} \tag{72}$$

Taking $\tau = \frac{1}{24}, m = \frac{1}{96}, \delta \le \min\{\frac{\gamma}{12}, \frac{\mu}{48L}, \frac{1}{3q}, \frac{1}{4}\}$, we have the result.

9.6. Resolvent of Logistic Regression

In Logistic Regression, each component operator $\mathcal{B}_{n,i}$ is defined as $\mathcal{B}_{n,i}(\mathbf{z}) = \frac{-y_{n,i}}{1+\exp(y_{n,i}\cdot\mathbf{a}_{n,i}^\top\mathbf{z})}\mathbf{a}_{n,i}$, where $\mathbf{a}_{n,i}\in\mathbb{R}^d$ is the feature vector of a sample and $y_{n,i}\in\{-1,+1\}$ is its class label. The resolvent, $\mathcal{J}_{\alpha\mathcal{B}_{n,i}}(\mathbf{z})$, does not admit a closed form solution, but can be computed efficiently by the following newton iteration: let $a_0 = 0$, $b = \mathbf{a}_{n,i}^\top\mathbf{z}$

$$e_k = \frac{-y_{n,i}}{1+\exp(y_{n,i}a_k)} \text{ and } a_{k+1} = a_k - \frac{\alpha e_k + a_k - b}{1-\alpha y_{n,i}e_k - \alpha e_k^2}. \tag{73}$$

When the iterate converges, the resolvent is obtain by

$$\mathcal{J}_{\alpha\mathcal{B}_{n,i}}(\mathbf{z}) = \mathbf{z} - (b-a_k)\mathbf{a}_{n,i}. \tag{74}$$

In our experiments, 20 newton iteration is sufficient for DSBA.

9.7. Resolvent of AUC maximization

In the ℓ_2-relaxed AUC maximization, the variable $\mathbf{z}\in\mathbb{R}^{d+3}$ is a $d+3$-dimensional augmented vector, where d is the dimension of the dataset. For simplicity, we decompose $\mathbf{z}$ as $\mathbf{z} = [\mathbf{w}^\top; a; b; \theta]$ with $\mathbf{w}\in\mathbb{R}^d$, $a\in\mathbb{R}$, $b\in\mathbb{R}$, $\theta\in\mathbb{R}$. For a positive sample, i.e. $y_{n,i} = +1$, the component operator $\mathcal{B}_{n,i}$ is then defined as

$$\mathcal{B}_{n,i}(\mathbf{z}) = \begin{bmatrix} 2(1-p)((\mathbf{a}_{n,i}^\top\mathbf{w}-a) - (1+\theta))\mathbf{a}_{n,i} \\ -2(1-p)(\mathbf{a}_{n,i}^\top\mathbf{w}-a) \\ 0 \\ 2p(1-p)\theta + 2(1-p)\mathbf{a}_{n,i}^\top\mathbf{w} \end{bmatrix} \tag{75}$$

and for a negative sample, i.e. $y_{n,i} = -1$

$$\mathcal{B}_{n,i}(\mathbf{z}) = \begin{bmatrix} 2p((\mathbf{a}_{n,i}^\top \mathbf{w} - b) + (1+\theta))\mathbf{a}_{n,i} \\ 0 \\ -2p(\mathbf{a}_{n,i}^\top \mathbf{w} - b) \\ 2p(1-p)\theta - 2p\mathbf{a}_{n,i}^\top \mathbf{w} \end{bmatrix} \tag{76}$$

where $p = \frac{\#\text{positive samples}}{\#\text{samples}}$ is the positive ratio of the dataset. Similar to RR, the resolvent of $\mathcal{B}_{n,i}$ also admits a closed form solution, which we now derive. For a positive sample, define

$$\mathbf{A}^+ = \begin{bmatrix} 1+2(1-p)\alpha & -2(1-p)\alpha & 0 & -2(1-p)\alpha \\ -2(1-p)\alpha & 1+2(1-p)\alpha & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 2(1-p)\alpha & 0 & 0 & 1+2p(1-p)\alpha \end{bmatrix} \tag{77}$$

and

$$\mathbf{b}^+ = \begin{bmatrix} \mathbf{a}_{n,i}^\top \mathbf{w} + 2(1-p)\alpha \\ a \\ b \\ \theta \end{bmatrix}. \tag{78}$$

Let $\mathbf{b}_r^+ = (\mathbf{A}^+)^{-1}\mathbf{b}^+ \in \mathbb{R}^4$ and decompose it as $\mathbf{b}_r^+ = [z_r^+; a_r^+; b_r^+; \theta_r^+]$. The resolvent is obtain as

$$\mathcal{J}_{\alpha\mathcal{B}_{n,i}}(\mathbf{z}) = \mathbf{z}_r^+ = \begin{bmatrix} [\mathbf{w} - 2(1-p)\alpha[(z_r^+ - a) - (1+\theta)]\mathbf{a}_{n,i} \\ a_r^+ \\ b_r^+ \\ \theta_r^+ \end{bmatrix} \tag{79}$$

We can do the similar derivation for a negative sample. Define

$$\mathbf{A}^- = \begin{bmatrix} 1+2p\alpha & 0 & -2p\alpha & 2p\alpha \\ 0 & 1 & 0 & 0 \\ -2p\alpha & 0 & 1+2p\alpha & 0 \\ -2p\alpha & 0 & 0 & 1+2p(1-p)\alpha \end{bmatrix} \tag{80}$$

and

$$\mathbf{b}^+ = \begin{bmatrix} \mathbf{a}_{n,i}^\top \mathbf{w} - 2p\alpha \\ a \\ b \\ \theta \end{bmatrix} \tag{81}$$

Let $\mathbf{b}_r^- = (\mathbf{A}^-)^{-1}\mathbf{b}^- \in \mathbb{R}^4$ and decompose it as $\mathbf{b}_r^- = [z_r^-; a_r^-; b_r^-; \theta_r^-]$ The resolvent is obtain as

$$\mathcal{J}_{\alpha\mathcal{B}_{n,i}}(\mathbf{z}) = \mathbf{z}_r^+ = \begin{bmatrix} [\mathbf{w} - 2p\alpha[(z_r^- - b) - (1+\theta)]\mathbf{a}_{n,i} \\ a_r^- \\ b_r^- \\ \theta_r^- \end{bmatrix}. \tag{82}$$