The Binomial likelihood equation of the HLSM is:

$$L = \prod_k \prod_{i,j} \sigma(\eta_{ijk})^{y_{ijk}} (1 - \sigma(\eta_{ijk}))^{1-y_{ijk}}$$

where,

$$\eta_{ijk} = -\alpha_k + \|z_{ik} - z_{jk}\|^2 = -\alpha_k + \|b_i + \epsilon_{ik} - b_j - \epsilon_{jk}\|^2$$

$$z_{ik} = b_i + \epsilon_{ik}$$

$$\sigma(\eta_{ijk}) = \frac{1}{1 + \exp(\eta_{ijk})}$$

Then, the log likelihood is

$$\begin{aligned}
\ell &= \sum_k \sum_{i,j} y_{ijk} \ln \sigma(\eta_{ijk}) + (1 - y_{ijk}) \ln(1 - \sigma(\eta_{ijk})) \\
&= \sum_k \sum_{i,j} -y_{ijk} \ln(1 + \exp(\eta_{ijk})) + (1 - y_{ijk}) \ln\left(\frac{\exp(\eta_{ijk})}{1 + \exp(\eta_{ijk})}\right) \\
&= \sum_k \sum_{i,j} -y_{ijk} \ln(1 + \exp(\eta_{ijk})) + (1 - y_{ijk})[\eta_{ijk} - \ln(1 + \exp(\eta_{ijk}))] \\
&= \sum_k \sum_{i,j} (1 - y_{ijk})\eta_{ijk} - \ln(1 + \exp(\eta_{ijk}))
\end{aligned}$$

We define our decision variables as: $\{b_i\} \; \forall \; i$ and $\{\epsilon_{ik}\} \; \forall \; i, k$.
Then, the gradients are:

$$\frac{\partial \eta_{ijk}}{\partial b_{ik}} = 2(b_i + \epsilon_{ik} - b_j - \epsilon_{jk})$$

$$\frac{\partial \eta_{ijk}}{\partial b_{jk}} = -2(b_i + \epsilon_{ik} - b_j - \epsilon_{jk})$$

$$\frac{\partial \eta_{ijk}}{\partial \epsilon_{ik}} = 2(b_i + \epsilon_{ik} - b_j - \epsilon_{jk})$$

$$\frac{\partial \eta_{ijk}}{\partial \epsilon_{jk}} = -2(b_i + \epsilon_{ik} - b_j - \epsilon_{jk})$$

$$
\begin{aligned}
\frac{\partial \ell}{\partial b_{ik}} &= \sum_j (1 - y_{ijk}) \frac{\partial \eta_{ijk}}{\partial b_{ik}} - \frac{\partial \ln(1 + \exp(\eta_{ijk}))}{\partial b_{ik}} \\
&= 2 \sum_j (1 - y_{ijk})(b_i + \epsilon_{ik} - b_j - \epsilon_{jk}) - \frac{1}{1 + \exp(\eta_{ijk}))} \exp(\eta_{ijk})(b_i + \epsilon_{ik} - b_j - \epsilon_{jk}) \\
&= 2 \sum_j (1 - y_{ijk})(b_i + \epsilon_{ik} - b_j - \epsilon_{jk}) - \frac{\exp(\eta_{ijk})}{1 + \exp(\eta_{ijk}))} (b_i + \epsilon_{ik} - b_j - \epsilon_{jk})
\end{aligned}
$$

$$
\frac{\partial \ell}{\partial b_{jk}} = -2 \sum_i (1 - y_{ijk})(b_i + \epsilon_{ik} - b_j - \epsilon_{jk}) - \frac{\exp(\eta_{ijk})}{1 + \exp(\eta_{ijk}))} (b_i + \epsilon_{ik} - b_j - \epsilon_{jk})
$$

$$
\frac{\partial \ell}{\partial \epsilon_{ik}} = 2 \sum_j (1 - y_{ijk})(b_i + \epsilon_{ik} - b_j - \epsilon_{jk}) - \frac{\exp(\eta_{ijk})}{1 + \exp(\eta_{ijk}))} (b_i + \epsilon_{ik} - b_j - \epsilon_{jk})
$$

$$
\frac{\partial \ell}{\partial \epsilon_{jk}} = -2 \sum_i (1 - y_{ijk})(b_i + \epsilon_{ik} - b_j - \epsilon_{jk}) - \frac{\exp(\eta_{ijk})}{1 + \exp(\eta_{ijk}))} (b_i + \epsilon_{ik} - b_j - \epsilon_{jk})
$$

Finally, our complete objective function is to minimize the sum between the negative log likelihood and a lasso penalty on the deviations ϵ_{ik}.

$$
\min_{\epsilon, b} \sum_k \sum_{i,j} [(y_{ijk} - 1)\eta_{ijk} + \ln(1 + \exp(\eta_{ijk}))] + \lambda \sum_k \sum_i \|\epsilon_{i,k}\|
$$