Without loss of generality, let us instantiate the objective function with the regularizer \( \lambda \) of empirical error on the labeled data for a family of hash codes.

\[
f(W, Q) = \min \sum_{i} \| W_{k} P_{k}^i - H_{k}^i \|_F^2 + \lambda \sum_{i,k} \sum_{u,v} -s_{i,u,v} r_{i,u,v}^2\]  
\tag{A-1}
\]

One can express the second term of the above objective function in a compact matrix form by defining matrices \( S_{k}^i \in \mathbb{R}^{N_i \times N_i} \) incorporating the pairwise similarity of \( P_{k}^i \) and the pairwise relationship of \( P^j \) for labeled information respectively. Then, \( S((W_{k})_{k=1}^K) \) can be represented as

\[
S((W_{k})_{k=1}^K) = \lambda \sum_{i,k} \text{tr}(-S_{k}^i R^i) \]  
\tag{A-2}

and using (1) and absorbing the constant term into \( \lambda \) we can also represent it as

\[
S((W_{k})_{k=1}^K) = \lambda \sum_{i,k} \text{tr}(-H_{k}^i R^i H_{k}^i) \]  
\tag{A-3}

The objective function becomes the follows.

\[
f(W, Q) = \min \sum_{i} \| W_{k}^T P_{k}^i - \text{sign}(Q_{k}^i) \|_F^2 + \lambda \sum_{i,k} \sum_{u,v} \text{tr}(-\text{sign}(Q_{k}^i)^R \text{sign}(Q_{k}^i)^T) \]  
\tag{A-4}

Here, we concern two possible problems due to the sign function for \( f \). First, \( Q \) may not be a unique solution and thus the objective function is difficult to converge without considering any regularizer about \( Q \). We add a Frobenius norm regularizer \( \eta \). In addition, the objective function \( f(W, Q) \) is nondifferentiable in terms of \( Q \). We can approximate the sign function with the surrogate function (A-5).

\[
S_{\xi}(Q_{k}^i) = (Q_{k}^i + \xi)^{\frac{1}{2}} \circ Q_{k}^i \]  
\tag{A-5}

Here, \( \xi \) is a positive constant close to zero and \( \circ \) is the hadamard (elementwise) product. Finally, we have the following objective function.

\[
f(W, Q) = \min \sum_{i} \| W_{k}^T P_{k}^i - S_{\xi}(Q_{k}^i) \|_F^2 + \lambda \sum_{i,k} \sum_{u,v} \text{tr}(-S_{\xi}(Q_{k}^i)^R S_{\xi}(Q_{k}^i)^T + \eta \sum_{i,k} \| Q_{k}^i \|_F^2 \]  
\tag{A-6}

To minimize the objective function applying (A-6), we use a Newton-Raphson algorithm \[55\] and iteratively solve the objective function (A-5).

\[
\frac{\delta f}{\delta W_{k}} = \sum_{i,j} 2P_{k}^i (W_{k}^T P_{k}^j - S_{\xi}(Q_{k}^j) - (Q_{k}^i \circ Q_{k}^j + \xi)^{\frac{1}{2}} \circ Q_{k}^i)^{-1} \frac{\delta f}{\delta Q_{k}^j}\]  
\tag{A-7}

This approach allows us to update \( W_{k} \) for all \( k \) \((1 \leq k \leq K)\) simultaneously \( W_{k}^{\text{new}} = W_{k} - \frac{\delta f}{\delta W_{k}} \). Similarly with \( W \) fixed we can update \( Q \). To be specific, we update \( Q_k^i \) for all combinations of \((i,k)\), \(1 \leq i \leq M, 1 \leq k \leq K\) at the same time \( Q_k^{i,\text{new}} = Q_k^i - \frac{\delta^2 f}{\delta Q_{k}^i}^{-1} \frac{\delta f}{\delta Q_{k}^i} \). The first derivative of \( Q_{k}^i \) is

\[
\frac{\delta f}{\delta Q_{k}^i} = -2(W_{k}^T P_{k}^i - S_{\xi}(Q_{k}^i)) \circ S_{\xi}'(Q_{k}^i) - \lambda \nabla tr + 2\eta Q_{k}^i \]  
\tag{A-8}

where \( \nabla tr \) is defined in an elementwise manner

\[
(\nabla tr)_{p,q} = S_{\xi}'(Q_{k}^i; p,q) \{ (\sum_{t=1}^{N_i} S_{\xi}(Q_{k}^i; t,q) R_{i,p}^{t}) + S_{\xi}(Q_{k}^i; p,q) \} \]  
\tag{A-9}

and \( S_{\xi}'(Q_{k}^i) = (Q_{k}^i \circ \xi)^{\frac{1}{2}} \). Moreover, we calculate second derivatives for the Hessian matrix. We first consider the \( W_{k} \)

\[
\frac{\delta^2 f}{\delta W_{k}^2} = \sum_{i,k} \sum_{p} 2P_{k}^i P_{k}^j \]  
\tag{A-10}

In terms of \( Q_{k}^i \), \( \frac{\delta^2 f}{\delta Q_{k}^i} \) is derived as

\[
\text{vec} \{ -2(W_{k}^T P_{k}^i) S_{\xi}'(Q_{k}^i) + 2 \{S_{\xi}'(Q_{k}^i) \circ S_{\xi}'(Q_{k}^i) + S_{\xi}(Q_{k}^i) \circ S_{\xi}'(Q_{k}^i) \} \} \text{vec}(L_{(N_i,B_k)})^T - \lambda \nabla^2 tr + 2\eta I_{(N_i,B_k)} \]  
\tag{A-11}

Where

\[
\nabla^2 tr = \{I_{(B_k)} \otimes \{(I_{(N_i)} + I_{(N_i)}) \otimes R^i\}\} \ast \{\text{vec}(S_{\xi}'(Q_{k}^i)) \circ (S_{\xi}'(Q_{k}^i)) \} \} + \text{diag} (\text{vec}(A)). \]  
\tag{A-12}

Here, \( \otimes \) is the Kronecker product and \( \text{vec} \) and \( \text{diag} \) are to transform a matrix to a vector and a vector to a diagonal matrix respectively. In addition, \( J \) is a matrix of ones and \( A \) is defined in an elementwise manner as follows

\[
(J)_{p,q} = S_{\xi}(Q_{k}^i; p,q) \{ (\sum_{t=1}^{N_i} S_{\xi}(Q_{k}^i; t,q) R_{i,p}^{t}) + S_{\xi}(Q_{k}^i; p,q) \} \]  
\tag{A-13}

where \( S_{\xi}'(Q_{k}^i; p,q) = (Q_{k}^i \circ \xi)^{\frac{1}{2}}\).