1. Details of Coefficients Calculation

The objective function of DL-GSGC is:

\[
\min_{D, x} \left\{ \left| \sum_{i=1}^{N} y_i - DX_i \right|_F^2 + \sum_{i=1}^{N} \left| \sum_{j=1}^{N_i} x_{ij} \right|_1 + \lambda_2 \left\| X \right\|_F^2 \right\}
\]

subject to \( \left\| d_k \right\|_2^2 = 1, \quad \forall k = 1, 2, \ldots, K \)

The Eq. 1 can be rewritten as:

\[
\min_{D, x} \left\{ \left| \sum_{i=1}^{N} y_i - DX_i \right|_F^2 + \sum_{i=1}^{N} \left| \sum_{j=1}^{N_i} x_{ij} \right|_1 \right\}
\]

After fixing the dictionary \( D \), the coefficients for the \( j \)-th feature in class \( i \), \( x_{ij} \), can be calculated by solving the following convex problem:

\[
\min_{x_{ij}} \left\{ \left| y_j - DX_j \right|_F^2 + \left| y_j - D_{\not\in i}X_j \right|_F^2 + \lambda_1 \left| \sum_{j=1}^{N_j} x_{ij} \right|_1 \right\}
\]

where \( I \in \mathbb{R}^{K \times K} \) is an identity matrix. To remove the influence of shared features among classes, we use templates belonging to the same class as the input feature for similarity measure at this stage. Therefore, the objective function becomes:

\[
\min_{x_{ij}} \left\{ \left| s_j - D_{\not\in i}X_j \right|_F^2 + \lambda_1 \left| x_{ij} \right|_1 + \lambda_3 \sum_{m=1}^{A_i} \left\| \alpha_m - x_{ij} \right\|_2^2 \right\}
\]

where

\[
s_j = [y_j^f; y_j^f; 0; \ldots; 0] \quad \text{for } j = f + K
\]

\[
D_{\not\in i}^j = [D_{\not\in i}; D_{\not\in i}; \sqrt{\lambda_2}I]
\]

\[
L(x_j^j) = \sum_{m=1}^{A_i} \left\| \alpha_m - x_j^j \right\|_2^2 w_{mj}
\]

To use the feature-sign search for coefficients calculation, Eq. 4 can be rewritten as:

\[
\min_{x} \left\{ \left| s_j - D_{\not\in i}x^j \right|_2^2 + \lambda_1 \left| x^j \right|_1 + \lambda_3 \sum_{m=1}^{A_i} \left\| \alpha_m - x^j \right\|_2^2 w_{mj} \right\}
\]

where \( x^j = \sqrt{2 + \lambda_2}x^j \) and \( D_{\not\in i}^j \) is the \( l_2 \) column-wise normalized version of \( D_{\not\in i} \). To simplify notation, we present the following equivalent optimization problem:

\[
\min_{x} f(x) = h(x) + \lambda|x|_1
\]

\[
h(x) = \left| s - Bx \right|_2^2 + \gamma \sum_{m=1}^{A_i} \left\| \beta_m - x \right\|_2^2 w_m
\]

where \( B = D_{\not\in i}^T \), \( \lambda = \frac{\lambda_1}{\sqrt{2 + \lambda_2}} \), \( \gamma = \frac{\lambda_3}{2 + \lambda_2} \) and \( \beta_m = \sqrt{2 + \lambda_2}\alpha_m \). Then the first derivative of \( h(x) \) over \( x \) can be represented as:

\[
\nabla h(x) = -2B^Ts - 2\gamma \sum_{m=1}^{A_i} w_m \beta_m
\]

Details of coefficients calculation using the feature-sign search method [1] are provided in Algorithm 1.

According to the Algorithm 1, solutions for Eq. 8 can be obtained as \( x^j \). The coefficients should be \( x^j = \frac{1}{\sqrt{2 + \lambda_2}}x^j \).

References

Algorithm 1 Coefficients Calculation

Input: $\beta = [\beta_1, \ldots, \beta_A]$, $W = [w_1, \ldots, w_A]$, $\lambda$, $\gamma$, $B$, $s$, and $h(x)$
Output: $x^T = [x_1, \ldots, x_K]$

1: **Step 1: Initialization**
2: $x = 0$, $\theta = 0$, and active set $\mathcal{H} = \{\}$, where $\theta_i = \{1, 0, -1\}$ denotes $\text{sign}(x_i)$.
3: **Step 2: Active Set Update**
4: From zero coefficients of $x$, choose the one with the maximum absolute value of the first derivative: $i = \arg \max [\nabla h(x_i)]$
5: if $\nabla h(x_i) > \lambda$ then
6: set $\theta_i = -1$, $\mathcal{H} = \{i\} \cup \mathcal{H}$
7: end if
8: if $\nabla h(x_i) < -\lambda$ then
9: set $\theta_i = 1$, $\mathcal{H} = \{i\} \cup \mathcal{H}$
10: end if
11: **Step 3: Feature-sign step**
12: Let $\hat{B}$ be a submatrix of $B$ that contains only the columns corresponding to the active set $\mathcal{H}$. Let $\hat{x}$, $\hat{\beta}_m$, and $\hat{\theta}$ be subvectors of $x$, $\beta_m$, and $\theta$ corresponding to the active set $\mathcal{H}$.
13: Compute the analytical solution according to Eqs. 9 and 10: $\hat{x}^{\text{new}} = (\hat{B}^T \hat{B} + \gamma \sum_{m=1}^{A_i} w_m I)^{-1} (\hat{B}^T \hat{s} + \gamma \sum_{m=1}^{A_i} w_m \hat{\beta}_m - \frac{\lambda}{\gamma})$
14: Perform a discrete line search on the closed line segment from $\hat{x}$ to $\hat{x}^{\text{new}}$: check the objective value at $\hat{x}^{\text{new}}$ and all points where any coefficients changes sign; and then update $\hat{x}$ to the point with the lowest objective value.
15: Remove zero coefficients of $\hat{x}$ from the active set $\mathcal{H}$ and update $\theta = \text{sign}(x)$.
16: **Step 4: Check the Optimality Conditions**
17: if $\nabla h(x_j) + \lambda \text{sign}(x_j) = 0$, $\forall x_j \neq 0$ then
18: if $|\nabla h(x_j)| \leq \lambda$, $\forall x_j = 0$ then
19: Return $x$ as solution
20: else
21: Go to Step 2
22: end if
23: else
24: Go to Step 3
25: end if