Simultaneous clustering of multiview biomedical data using manifold optimization: Supplementary materials

December 26, 2018

1 Detailed inference of Algorithm 1

The Stiefel manifold \( St(p,n) \) is defined as the set of all \( n \times p \) orthonormal matrices, i.e.,

\[
St(p,n) := \{ X \in \mathbb{R}^{n \times p} : X^T X = I_p \}
\]

Thus, our model

\[
\min_{U_m^T U_m = I_K, m=1,2,...,M} \text{trace}(U^T LU)
\]

can be written as

\[
\min_{\{U_m\}_{m=1}^M \in St(K,N)} \text{trace}(U^T LU) \tag{1}
\]

which is an optimization problem defined on manifold \( \mathcal{M} := \{ U \in \mathbb{R}^{MN \times K} : U_m \in St(K,N), \ m = 1, \cdots, M \} \).

To solve (1), we use an iterative approach. Let \( U^{(t)} = ([U_1^{(t)}]^T, \cdots, [U_M^{(t)}]^T)^T \) denote the result at iteration \( t \). We take three steps to get the result at iteration \( t + 1 \).

**Step 1**: Project the negative gradient direction of the objective function to the tangent vector space of the manifold \( \mathcal{M} \). For each stiefel manifold \( St(K,N) \) its tangent vector space at \( U_m^{(t)} \) has the form: \( T_{U_m^{(t)}} St(K,N) = \{ U_m^{(t)} B + (I - (U_m^{(t)})^T U_m) C : B = -B^T, C \in \mathbb{R}^{N \times K} \} \). Let \( Y_m \in \mathbb{R}^{N \times K} \), with orthogonal
projection of $Y_m$ to the tangent vector space of $St(K, N)$ at $U_m^{(t)}$, we get $(I - U_m^{(t)}(U_m^{(t)})^T)Y_m + U_m^{(t)}\text{skew}((U_m^{(t)})^TY_m)$, where $\text{skew}(A) = \frac{1}{2}(A - A^T)$ (P. A. Absil, chapter 3) [13].

Let $Y = (Y_1^T, Y_2^T, \ldots, Y_M^T)^T$ be any point in $\mathbb{R}^{MN \times K}$. Projecting $Y$ to the tangent vector space of the manifold $\mathcal{M}$ is equivalent to projecting each $Y_m$ to the manifold $St(K, N)$. Thus by projecting $Y$ to $T_{U(t)}\mathcal{M}$, we get:

$$P_{U(t)}Y = ((P_1^{(t)}Y_1)^T, \ldots, (P_M^{(t)}Y_M)^T)^T,$$

where

$$P_m^{(t)}Y_m = (I - U_m^{(t)}(U_m^{(t)})^T)Y_m + U_m^{(t)}\text{skew}((U_m^{(t)})^TY_m), \quad \forall m = 1, \ldots, M.$$

Let’s denote the negative gradient direction of the objective function at iteration $t$ as:

$$Z^{(t)} = -\nabla_{U(t)}\text{trace}(U^TLU) = -LU^{(t)} = ((Z_1^{(t)})^T, \ldots, (Z_M^{(t)})^T)^T.$$

thus, the descent direction $\eta^{(t)} = P_{U(t)}Z^{(t)} = ((\eta_1^{(t)})^T, \ldots, (\eta_M^{(t)})^T)^T$, where

$$\eta_m^{(t)} = P_m^{(t)}Z_m^{(t)} = (I - U_m^{(t)}(U_m^{(t)})^T)Z_m^{(t)} + \frac{1}{2}U_m^{(t)}((U_m^{(t)})^TZ_m^{(t)} - (Z_m^{(t)})^TU_m^{(t)}),$$

$$= Z_m^{(t)} - \frac{1}{2}U_m^{(t)}((U_m^{(t)})^TZ_m^{(t)} + (Z_m^{(t)})^TU_m^{(t)}).$$

**Step 2:** Do line-search on the tangent vector space $T_{U(t)}\mathcal{M}$ : $\tilde{U}^{(t+1)} = U^{(t)} + \alpha_t\eta^{(t)}$, where $\alpha$ is an Armijo step size we will explain below. Since $U^{(t)} \in T_{U(t)}\mathcal{M}$ and $\eta^{(t)} \in T_{U(t)}\mathcal{M}$, $\tilde{U}^{(t+1)}$ also belongs to $T_{U(t)}\mathcal{M}$.

**Step 3:** Retract the obtained point $\tilde{U}^{(t+1)}$ to the manifold $\mathcal{M}$. Retraction $R$ is a smooth mapping from the tangent bundle $\bigcup_{U \in \mathcal{M}} T_U\mathcal{M}$ to $\mathcal{M}$. Let $R_{U(t)}$ denotes the restriction of $R$ to $T_{U(t)}\mathcal{M}$ (P. A. Absil, chapter 4)[13] which satisfies:

1. $R_{U(t)}(0_{U(t)}) = U^{(t)}$, where $0_{U(t)}$ is the zero element of $T_{U(t)}\mathcal{M}$;
2. $\frac{d}{dt}R_{U(t)}(tZ)|_{t=0} = Z$, for all $Z \in T_{U(t)}\mathcal{M}$.

Then we define the corresponding retraction $R_{U(t)}$ as:

$$R_{U(t)}(\tilde{U}^{(t+1)}) = R_{U(t)}\begin{pmatrix} \tilde{U}_1^{(t+1)} \\ \vdots \\ \tilde{U}_M^{(t+1)} \end{pmatrix} = \begin{pmatrix} \text{svd}(\tilde{U}_1^{(t+1)}) \\ \vdots \\ \text{svd}(\tilde{U}_M^{(t+1)}) \end{pmatrix}.$$
where $\tilde{U}^{(t+1)}_m = W_m \Sigma_m V_T^m$ is the SVD decomposition of $\tilde{U}^{(t+1)}_m$, and $\text{svd}(\cdot)$ is defined as: $\text{svd}(\tilde{U}^{(t+1)}_m) = W_m V_T^m$, $\forall m = 1, \cdots, M$.

With these three steps, we finish one iteration and have $U^{(t+1)} = R_{U^{(t)}}(\tilde{U}^{(t+1)})$. We iteratively run these steps and get the final results. In the following we will analyze the convergence property of above method by introducing the convergence of Accelerated Line Search algorithm (ALS) from P. A. Absil (chapter 4, Algorithm 1) and prove that our algorithm is under the frame of ALS. That’s to say, we can guarantee our algorithm’s convergence under some conditions.

2 Convergence analysis

We first introduce the Accelerated Linear Search algorithm (ALS) in [13] and its convergence property. ALS is actually a line-search framework for Riemannian manifold. We then prove that our proposed algorithm falls in this framework and thus has the convergence results.

In ALS, assume $x_k$ is the estimate from the $k$-th iteration, then for convenience, $x_{k+1} = R_{x_k}(\alpha_k \eta_k) := R_{x_k}(x_k + \alpha_k \eta_k)$, where $\eta_k$ is the descent direction and $\alpha_k$ is the step size. To obtain the global convergence results, some restrictions must be imposed on $\eta_k$ and $\alpha_k$.

**Definition 1 (Gradient-related sequence)**[13] Given a cost function $f$ on a Riemannian manifold $\mathcal{M}$, the sequence $\{\eta_k\}$ is gradient-related, if for any subsequence $\{x_k\}_{k \in K_0}$ of $\{x_k\} \in \mathcal{M}$ that converges to a noncritical point of $f$, the corresponding subsequence $\{\eta_k\}_{k \in K_0}$ is bounded and satisfies:

$$\lim_{k \to \infty} \sup_{k \in K_0} \langle \nabla f(x_k), \eta_k \rangle < 0 \quad (2)$$

**Definition 2 (Armijo point)**[13] Given a cost function $f$ on a Riemannian manifold $\mathcal{M}$ with retraction $R$, a point $x \in \mathcal{M}$, a tangent vector $\eta \in T_x \mathcal{M}$, scalars $\gamma > 0$, $\beta, \sigma \in (0, 1)$, $R_x$ is the restriction of $R$ to $T_x \mathcal{M}$, then the Armijo point $\hat{\eta} = \alpha \eta = \beta^m \gamma \eta$, where $m$ is the smallest nonnegative inter such that

$$f(x) - f(R_x(\beta^m \gamma \eta)) \geq -\sigma \langle \nabla f(x), \beta^m \gamma \eta \rangle \quad (3)$$

and $\alpha$ is the Armijo step size.

Then, the Accelerated Line Search Algorithm is summarized as follows:

**Accelerated Line Search Algorithm (ALS)**[13]
Require: Riemannian manifold $\mathcal{M}$, continuously differentiable scalar field $f$ on $\mathcal{M}$, retraction $R$ from $\bigcup_{x \in \mathcal{M}} T_x \mathcal{M}$ to $\mathcal{M}$, scalars $\gamma > 0, c, \beta, \sigma \in (0, 1)$.

Input: Initial iterate $x_0 \in \mathcal{M}$.

Output: Sequence of iterates $\{x_k\}$.

1. for $k = 1, 2, \ldots$ do:
   2. pick $\eta_k$ in $T_{x_k} \mathcal{M}$ such that the sequence $\{\eta_i\}_{i=0,1,\ldots}$ is gradient-related;
   3. select $x_{k+1}$ such that $f(x_k) - f(x_{k+1}) \geq c(f(x_k) - f(R_{x_k}(\alpha_k \eta_k)))$;

where $\alpha_k$ is the Armijo step size for the given $\gamma, \beta, \sigma, \eta_k$

4. end for

For the above ALS algorithm, we have the following convergence theorem:

**Theorem 1 (P. A. Absil, Theorem 4.3.1) [13]** Let $\{x_k\}$ be an infinite sequence of iterates generated by ALS. Then every accumulation point of $\{x_k\}$ is a critical point of the cost function $f$.

Based on the results of ALS algorithm, we have the following convergence results for our proposed algorithm.

**Theorem 2** Our proposed algorithm 1 converges to the critical point of the objective function.

**Proof.** Firstly, we need to show that the manifold we considered is a Riemannian manifold. For our algorithm, the Riemannian metric is inherited from the embedding space $\mathbb{R}^{MN \times K}$: $\langle X, Y \rangle = \text{trace}(X^T Y)$. As a result, $\mathcal{M} = \{U \in \mathbb{R}^{MN \times K} : U_m \in \text{St}(K, N), \forall m\}$ is a Riemannian manifold. And the objective function $f = \text{trace}(U^T LU)$, where $U \in \mathcal{M}$, is apparently a continuously differentiable scalar field on $\mathcal{M}$.

Secondly, in our method, since $\mathcal{M}$ is the submanifold of Riemannian manifold $\mathbb{R}^{MN \times K}$, let $\bar{f} = \text{trace}(U^T LU)$, where $U \in \mathbb{R}^{MN \times K}$, then $f$ is the restriction of $\bar{f}$ to $\mathcal{M}$. And the gradient of $\bar{f}$ satisfies: $\text{grad} f(U) = P_U \text{grad} \bar{f}(U)$, which is exactly the way we compute descent direction $\eta^{(t)}$ in each iteration of our algorithm. As a result, $\langle \text{grad} f(U^{(t)}), \eta^{(t)} \rangle = \langle \text{grad} f(U^{(t)}), -\text{grad} f(U^{(t)}) \rangle < 0$, so the sequence $\{\eta^{(t)}\}$ is gradient-related.

Thirdly, we select the step length $\alpha_t$ as an Armijo step size with $\beta, \sigma \in (0, 1)$ and $\gamma$ be the BB stepsize $\frac{|\zeta_t \cdot \xi_t|}{\langle \zeta_t, \xi_t \rangle}$ where $\zeta_t = \eta^{(t)} - \eta^{(t-1)}$ and $\xi_t = U^{(t)} - U^{(t-1)}$. That’s to say, we have $f(U^{(t)}) - f(U^{(t+1)}) = f(U^{(t)}) - f(R_{U^{(t)}}(\alpha_t \eta^{(t)}))$,
Table 1: Sample size of each cancer subtype

<table>
<thead>
<tr>
<th>Disease</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>GBM</td>
<td>(101, 26, 34)</td>
</tr>
<tr>
<td>COAD</td>
<td>(17, 10, 42)</td>
</tr>
<tr>
<td>LSCC</td>
<td>(16, 10, 28, 26)</td>
</tr>
<tr>
<td>KRCCC</td>
<td>(19, 70, 9)</td>
</tr>
</tbody>
</table>

where \( \{ \eta^{(t)} \} \) is gradient-related and \( \alpha_t \) is the Armijo step size, which is just the equation (2.3) when \( c = 1 \). Therefore, our method is actually a special form of ALS, which means that every accumulation point of \( \{ U^{(t)} \} \) is a critical point of the objective function \( f \).

3 Sample size of each cancer subtype

We put it in the Table 1.

References


