A. Proof of the statement in Eq. (3)

In order to show the the result in Eq. (3), we breakdown the process in Eq. (2) into two steps: Let us denote $\widetilde{X} = \sum_{\phi_k \in \Phi} \alpha_k \phi_k$ and $X = d(\widetilde{X})$ where Φ is a set of s basis functions. Since the set of $\widetilde{X}$ functions create a linear subspace, every member can be written as a linear combination of at least s other functions:

$$\widetilde{X}_i = \sum_{j \neq i} \beta_j \widetilde{X}_j. \tag{7}$$

Given the fact that the set of deformations is a group, the inverse of deformation operators are also in the set and we can rewrite Eq. (7) as

$$d_i^{-1}(X_i) = \sum_{j \neq i} \beta_j d_j^{-1}(X_j), \tag{8}$$

$$X_i = d_i\bigg(\sum_{j \neq i} \beta_j d_j^{-1}(X_j)\bigg). \tag{9}$$

Since the operators are assumed to be linear maps, we can rewrite Eq. (9) as follows

$$X_i = \sum_{j \neq i} \beta_j d_i(d_j^{-1}(X_j)). \tag{10}$$

Group's closure property guarantees that for all i and j, there exists $\widetilde{d}_j$ in the group such that $\widetilde{d}_j = d_i \circ d_j^{-1}$. Thus we can rewrite Eq. (10) as

$$X_i = \sum_{j \neq i} \beta_j \widetilde{d}_j(X_j).$$

B. Proof of the Theorem

To prove the statement of the theorem, we need to show that by selection of the termination criterion as the theorem suggests, the Algorithm 1 will stop before adding any functions from other subspaces. In other words, Let us study the correctness of the theorem for neighbors of an arbitrary function Y_i; heretoafter we drop the i index for simplicity of notation whenever it is not ambiguous. If R_k denotes the residual at kth step, define the normalized residual as $\bar{R}_k = R_k / \|R\|_2$; we need to show that the following quality cannot be larger than ϵ:

$$\max_{V \notin \mathcal{Y}_i, d} \left\langle \bar{R}_k, \bar{d}(V) \right\rangle < \epsilon.$$

where $\bar{d}(Y) = d(Y)/\|d(Y)\|_2$ for any function Y. Furthermore, define

$$\mu_\ell = \max_{\ell' \neq \ell} \sup_{V \in S_\ell, Z \in S_{\ell'}, d, d'} \frac{|\langle d(V), d'(U)\rangle|}{\|d(V)\|_2 \|d'(U)\|_2}.$$

Algorithm 2: Spectral clustering for FSC.

Data: Affinity matrix $\boldsymbol{A}$

Result: Clustering assignments for Y_i, $i = 1, \ldots, n$.

1 $\boldsymbol{D} \leftarrow \operatorname{diag}(\boldsymbol{A}\boldsymbol{1})$

2 $\boldsymbol{L} \leftarrow \boldsymbol{D}^{-\frac{1}{2}} \boldsymbol{A} \boldsymbol{D}^{-\frac{1}{2}}$

3 $\boldsymbol{\lambda}, \boldsymbol{V} \leftarrow \operatorname{eig}(\boldsymbol{L})$

4 $m^\star \leftarrow \operatorname{argmax}_{i=1,\ldots,n-1}(\lambda_i - \lambda_{i+1})$

5 Apply k-means to the first $m^\star$ column of $\boldsymbol{V}$.

We note that we always have $\mu_\ell \leq \theta_\ell$, as $\mathcal{Y}_\ell \subset S_\ell$. Also, let us define the span of $d(\mathcal{Y}_\ell)$ as the span of the set of functions $\{d(Y) \big| Y \in \mathcal{Y}_\ell\}$.

To prove the main statement, we proceed with induction, as in (Dyer et al., 2013). Given the assumptions and the quantity of ϵ in the theorem, the first step holds, because $R_k = Y_i$. Now, assume that at kth iteration all of the previous functions have been selected from the correct subspace. Given the result in Eq. (3), the residual is still in the span of the $d(\mathcal{Y}_\ell)$. Thus, we can write $\bar{R}_k = \bar{d}_1(U) + E$ where U is the closes function in $\mathcal{Y}_\ell$ to $\bar{R}_k$ and $E \in S_i$, the latter is due to the assumption that d is a linear map. We can write:

$$\begin{aligned}
\max_{Y_j \notin \mathcal{Y}_1, d_1, d_2} & |\bar{R}_k, \bar{d}_2(Y_j)\rangle| \\
&= \max_{Y_j \notin \mathcal{Y}_1, d_1, d_2} |\langle \bar{d}_1(U) + E, \bar{d}_2(Y_j)\rangle| \\
&\leq \max_{Y_j \notin \mathcal{Y}_1, d_1, d_2} |\langle \bar{d}_1(U), \bar{d}_2(Y_j)\rangle| + |\langle E, \bar{d}_2(Y_j)\rangle| \\
&\leq \mu_\ell + \max_{Y_j \notin \mathcal{Y}_1, d_1, d_2} |\langle E, \bar{d}_2(Y_j)\rangle| \\
&\leq \mu_\ell + \cos\theta_0 \|E\|_2 \|\bar{d}_2(Y_j)\|_2,
\end{aligned} \tag{11}$$

where θ is the minimum principal angle between S_i and all other subspaces. We can bound the $\|E\|_2$ as follows:

$$\begin{aligned}
\|E\|_2 &= \|\bar{R}_k - \bar{d}_1(U)\|_2 \\
&= \sqrt{\|\bar{R}_k\|_2^2 + \|\bar{d}_1(U)\|_2^2 - 2\langle \bar{R}_k, \bar{d}_1(U)\rangle} \\
&\leq \sqrt{2 - 2\sqrt{1 - (r_\ell/2)^2}}.
\end{aligned} \tag{12}$$

Plugging the result in Eq. (12) in Eq. (11) yields:

$$\begin{aligned}
\max_{Y_j \notin \mathcal{Y}_1, d_1, d_2} |\bar{R}_k, \bar{d}_2(Y_j)\rangle| &\leq \mu_\ell + \cos\theta_0 \sqrt{2 - \sqrt{4 - r_\ell^2}} \\
&\leq \mu_\ell + \cos\theta_0 \frac{r_\ell}{\sqrt[4]{12}},
\end{aligned}$$

where the last step is due to (Dyer et al., 2013, Lemma 1). Given the fact that $\cos\theta$ is an upper bound for μ_ℓ, we can obtain the statement in the theorem.

C. Spectral Clustering

Note that we use the eigen-gap statistic (Line 4 in Algorithm 2 to determine the dimension of the embedding (Tib-

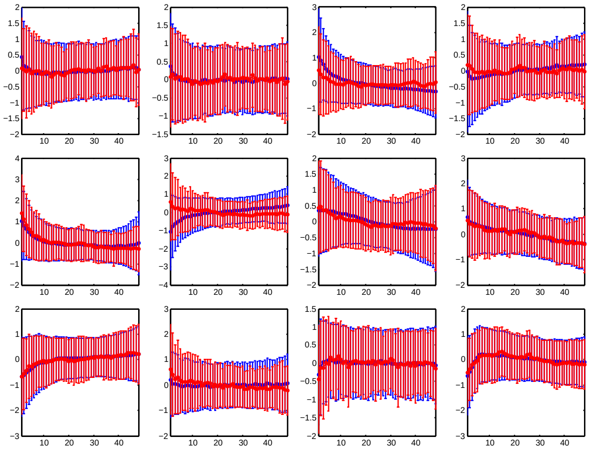

Figure 4. Mean and standard deviation trajectories for twelve variables in Physionet dataset, for patients who *survived* (blue) and *deceased* (red). Note the similarity of time series and the fact that they are almost indistinguishable by naked eye.

shirani et al., 2001; Von Luxburg, 2007).