Proof for uniqueness of the choice of A

Proposition 1:

Suppose $\Sigma \in \mathbb{R}^{n\times n}$ has n different eigenvalues, $\lambda_1 > \cdots > \lambda_n$, and $v_1, \dots, v_n \in \mathbb{R}^{n\times 1}$ are the corresponding eigenvectors that form an orthonormal basis of $\mathbb{R}^n$.

Suppose $A \in \mathbb{R}^{r\times n}$ is a semi-orthogonal matrix that satisfies $AA^T = I_r, r \leq n$.

Then we have the following statements:

a) if $A\Sigma A^T = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_r \end{pmatrix}$, then we have $\mathrm{A} = \begin{pmatrix} \pm v_1^T \\ \pm v_2^T \\ \vdots \\ \pm v_r^T \end{pmatrix}$. The choice of A is unique up to a sign of $\pm$.

b) if $A\Sigma A^T$'s eigenvalues are $\lambda_1, \dots, \lambda_r$, then we have $\mathrm{A} = \mathrm{O}\begin{pmatrix} \pm v_1^T \\ \pm v_2^T \\ \vdots \\ \pm v_r^T \end{pmatrix}$, $\mathrm{O} \in \mathbb{R}^{r\times r}$ is an orthogonal matrix.

To conclude, if $A\Sigma A^T$'s eigenvalues, $\lambda_1, \dots, \lambda_r$, are exactly the *r largest eigenvalues* of Σ, and Σ's eigenvalues $\lambda_1, \dots, \lambda_n$ are different from each other, then the choice of semi-orthogonal matrix $A \in \mathbb{R}^{r\times n}$ is unique up to an orthogonal transformation $\mathrm{O} \in \mathbb{R}^{r\times r}$.

Proof: We use mathematical induction to prove this.

1) If $r = 1$, then $A \in \mathbb{R}^{1\times n}$. As $v_1, \dots, v_n \in \mathbb{R}^{n\times 1}$ form an orthonormal basis of $\mathbb{R}^n$, we assume $A = \alpha_1 v_1^T + \cdots + \alpha_n v_n^T$. Since *A* is semi-orthogonal, $AA^T = I_r = 1$, we have $\alpha_1^2 + \cdots + \alpha_n^2 = 1$.

Now suppose a) $A\Sigma A^T = \lambda_1, \lambda_1 > \cdots > \lambda_n$.

By eigenvalue decomposition we have:

$$\Sigma = \sum_{i=1}^{n} \lambda_i v_i^T v_i$$

So we have:

$$A\Sigma A^T = \sum_{i=1}^{n} \alpha_i^2 \lambda_i = \lambda_1$$

Since $\alpha_1^2 + \cdots + \alpha_n^2 = 1$ and $\lambda_1 > \cdots > \lambda_n$, we have $\alpha_1^2 = 1$ and $\alpha_2^2 = \cdots = \alpha_n^2 = 0$. So we have:

$$\mathrm{A} = \pm v_1^T$$

which proves that statement a) is correct when $r = 1$.

If $r = 1$, then statement a) is equivalent to statement b). Hence, we proved that statement a), b) are correct when $r = 1$.

2) Suppose statement a) b) are correct for all $r \leq k < n$, now we prove that statement a) b) are

correct when $r = k + 1 \leq n$.

For statement a), suppose $A = \begin{pmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_{k+1}^T \end{pmatrix}$ and $A\Sigma A^T = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_{k+1} \end{pmatrix}$. Then we have:

$$\begin{pmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_k^T \end{pmatrix} \Sigma \begin{pmatrix} u_1 & u_2 & \cdots & u_k \end{pmatrix} = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_k \end{pmatrix}$$

and $u_{k+1}^T \Sigma u_{k+1} = \lambda_{k+1}$.

Since statement a) is true when $r = k$, we deduce that the choice of $\begin{pmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_k^T \end{pmatrix}$ is unique up to a sign $\pm$:

$$\begin{pmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_k^T \end{pmatrix} = \begin{pmatrix} \pm v_1^T \\ \pm v_2^T \\ \vdots \\ \pm v_k^T \end{pmatrix}$$

Since A is semi-orthogonal, we have $u_{\mathrm{k}+1}^{\mathrm{T}} u_1 = \cdots = u_{k+1}^T u_k = 0$ and $u_{\mathrm{k}+1}^T u_{k+1} = 1$. So $u_{k+1} \in span(v_{k+1}, \dots, v_n)$. Assume $u_{k+1}^T = \alpha_{k+1} v_{k+1}^T + \cdots + \alpha_n v_n^T$, then we have $\alpha_{k+1}^2 + \cdots + \alpha_n^2 = 1$.

By eigenvalue decomposition we have $\Sigma = \sum_{i=1}^n \lambda_i v_i^T v_i$ and $u_{k+1}^T \Sigma u_{k+1} = \sum_{i=k+1}^n \alpha_i^2 \lambda_i = \lambda_{k+1}$, so we have $\alpha_{k+1}^2 = 1$ and $\alpha_{k+2}^2 = \cdots = \alpha_n^2 = 0$ and $u_{k+1}^T = \pm v_{k+1}^T$.

Hence, we proved that the choice of $\begin{pmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_{k+1}^T \end{pmatrix}$ is unique up to a sign $\pm$, which means that the statement a) is true when $r = k + 1$.

For statement b): if $A\Sigma A^T$'s eigenvalues are $\lambda_1, \dots, \lambda_{k+1}$, then we have:

$$A\Sigma A^T = O \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_{k+1} \end{pmatrix} O^T$$

where $O \in \mathbb{R}^{(k+1)\times(k+1)}$ is an orthogonal matrix. So we have:

$$O^T A\Sigma A^T O = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_{k+1} \end{pmatrix}$$

According to statement a), we deduce that $O^T A$ is unique up to a sign $\pm$:

$$O^T A = \begin{pmatrix} \pm v_1^T \\ \pm v_2^T \\ \vdots \\ \pm v_{k+1}^T \end{pmatrix}$$

So we proved that $A = O\begin{pmatrix} \pm v_1^T \\ \pm v_2^T \\ \vdots \\ \pm v_{k+1}^T \end{pmatrix}$, $O \in \mathbb{R}^{(k+1)\times(k+1)}$ is an orthogonal matrix. Hence, we completed the proof for statement b).

Putting the deductions in 1) and 2) together, we conclude that the choice of semi-orthogonal matrix $A \in \mathbb{R}^{r\times n}$ is unique up to an orthogonal transformation $\mathrm{O} \in \mathbb{R}^{r\times r}$, if $A\Sigma A^T$'s eigenvalues, $\lambda_1, \dots, \lambda_r$, are exactly the *r largest eigenvalues* of Σ.

Similar to proposition 1, we can prove that $A \in \mathbb{R}^{r\times n}$ is unique up to an orthogonal transformation $\mathrm{O} \in \mathbb{R}^{r\times r}$, if $A\Sigma A^T$'s eigenvalues, $\lambda_{n-r+1}, \dots, \lambda_n$, are exactly the *r smallest eigenvalues* of Σ and Σ's eigenvalues $\lambda_1, \dots, \lambda_n$ are different from each other.

Proposition 2:

Suppose $\Sigma \in \mathbb{R}^{n\times n}$ has *n* different eigenvalues, $\lambda_1 > \cdots > \lambda_n$, and $v_1, \dots, v_n \in \mathbb{R}^{n\times 1}$ are the corresponding eigenvectors that form an orthonormal basis of $\mathbb{R}^{\mathrm{n}}$.

Suppose $A \in \mathbb{R}^{r\times n}$ is a semi-orthogonal matrix that satisfies $AA^T = I_r, r \le n$.

Then we have the following statements:

a) if $A\Sigma A^T = \begin{pmatrix} \lambda_{n-r+1} & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$, then we have $\mathrm{A} = \begin{pmatrix} \pm v_{n-r+1}^T \\ \pm v_{n-r+2}^T \\ \vdots \\ \pm v_n^T \end{pmatrix}$. The choice of A is unique up to a sign of $\pm$.

b) if $A\Sigma A^T$'s eigenvalues are $\lambda_{n-r+1}, \dots, \lambda_n$, then we have $\mathrm{A} = \mathrm{O}\begin{pmatrix} \pm v_{n-r+1}^T \\ \pm v_{n-r+2}^T \\ \vdots \\ \pm v_n^T \end{pmatrix}$, $\mathrm{O} \in \mathbb{R}^{r\times r}$ is an orthogonal matrix.

Sketch of the Proof: The proof of proposition 2 is quite similar to the proof of proposition 1. First, we can verify that if $r = 1$, then the only choice of A is $A = \pm v_n^T$. Then we assume that for any $r \le k < n$, statement a) b) are correct. Following the very same deduction for proposition 1, we can confirm that statement a) b) are correct when $r = k + 1 \le n$. This suffice to prove that proposition 2 is correct.

Then, we can prove that, if $A\Sigma A^T$'s r eigenvalues are composed of r_1 *largest eigenvalues* of Σ and r_2 *smallest eigenvalues* of Σ, $\mathrm{r}_1 + \mathrm{r}_2 = r$ and Σ's eigenvalues $\lambda_1, \dots, \lambda_n$ are different from each other, then the choice of A is also unique up to an orthogonal transformation $\mathrm{O} \in \mathbb{R}^{r\times r}$.

Proposition 3:

Suppose $\Sigma \in \mathbb{R}^{n\times n}$ has *n* different eigenvalues, $\lambda_1 > \cdots > \lambda_n$, and $v_1, \dots, v_n \in \mathbb{R}^{n\times 1}$ are the corresponding eigenvectors that form an orthonormal basis of $\mathbb{R}^{\mathrm{n}}$.

Suppose $A \in \mathbb{R}^{r\times n}$ is a semi-orthogonal matrix that satisfies $AA^T = I_r, r \le n$.

Then we have the following statements:

a) if $A\Sigma A^T = \begin{pmatrix} \lambda_1 &&&&& \\ & \ddots &&&& \\ && \lambda_{r_1} &&& \\ &&& \lambda_{n-r+r_1+1} && \\ &&&& \ddots & \\ &&&&& \lambda_n \end{pmatrix}$, then we have $A = \begin{pmatrix} \pm v_1^T \\ \vdots \\ \pm v_{r_1}^T \\ \pm v_{n-r+r_1+1}^T \\ \vdots \\ \pm v_n^T \end{pmatrix}$. The choice of A is unique up to a sign of $\pm$.

b) if $A\Sigma A^T$'s eigenvalues are $\lambda_1, \dots, \lambda_{r_1}, \lambda_{n-r+r_1+1}, \dots, \lambda_n$, then we have $A = O \begin{pmatrix} \pm v_1^T \\ \vdots \\ \pm v_{r_1}^T \\ \pm v_{n-r+r_1+1}^T \\ \vdots \\ \pm v_n^T \end{pmatrix}$, $O \in \mathbb{R}^{r\times r}$ is an orthogonal matrix.

Proof: For statement a), suppose $A = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}, A_1 \in \mathbb{R}^{r_1\times n}, A_2 \in \mathbb{R}^{(r-r_1)\times n}$. Since $A\Sigma A^T = \begin{pmatrix} \lambda_1 &&&&& \\ & \ddots &&&& \\ && \lambda_{r_1} &&& \\ &&& \lambda_{n-r+r_1+1} && \\ &&&& \ddots & \\ &&&&& \lambda_n \end{pmatrix}$, we have $A_1\Sigma A_1^T = \begin{pmatrix} \lambda_1 && \\ & \ddots & \\ && \lambda_{r_1} \end{pmatrix}$ and $A_2\Sigma A_2^T = \begin{pmatrix} \lambda_{n-r+r_1+1} && \\ & \ddots & \\ && \lambda_n \end{pmatrix}$. From proposition 1 we have $A_1 = \begin{pmatrix} \pm v_1^T \\ \vdots \\ \pm v_{r_1}^T \end{pmatrix}$ is unique and from proposition 2 we have $A_2 = \begin{pmatrix} \pm v_{n-r+r_1+1}^T \\ \vdots \\ \pm v_n^T \end{pmatrix}$ is unique, so $A = \begin{pmatrix} \pm v_1^T \\ \vdots \\ \pm v_{r_1}^T \\ \pm v_{n-r+r_1+1}^T \\ \vdots \\ \pm v_n^T \end{pmatrix}$ is unique up to a sign of $\pm$.

For statement b): if $A\Sigma A^T$'s eigenvalues are $\lambda_1, \dots, \lambda_{r_1}, \lambda_{n-r+r_1+1}, \dots, \lambda_n$, then we have:

$$A\Sigma A^T = O \begin{pmatrix} \lambda_1 &&&&& \\ & \ddots &&&& \\ && \lambda_{r_1} &&& \\ &&& \lambda_{n-r+r_1+1} && \\ &&&& \ddots & \\ &&&&& \lambda_n \end{pmatrix} O^T$$

where $O \in \mathbb{R}^{r\times r}$ is an orthogonal matrix. So we can deduce that $O^T A$ is unique up to a sign $\pm$:

$$O^T A = \begin{pmatrix} \pm v_1^T \\ \vdots \\ \pm v_{r_1}^T \\ \pm v_{n-r+r_1+1}^T \\ \vdots \\ \pm v_n^T \end{pmatrix}$$

So we proved that $A = O \begin{pmatrix} \pm v_1^T \\ \vdots \\ \pm v_{r_1}^T \\ \pm v_{n-r+r_1+1}^T \\ \vdots \\ \pm v_n^T \end{pmatrix}$, $O \in \mathbb{R}^{r\times r}$ is an orthogonal matrix. Hence, we completed the proof for statement b).

We have proved that to maximize each *f*-divergence is to extract some of the largest eigenvalues and some of the smallest eigenvalues of Σ, so according to proposition 3, we have the following:

Proposition 4:
If $\Sigma \in \mathbb{R}^{n\times n}$ has *n* different eigenvalues, $\lambda_1 > \cdots > \lambda_n$, and the evaluation function $g(\lambda)$ attains *n* different values $g(\lambda_i)$ at $\lambda = \lambda_i$, then to maximize each *f*-divergence, the optimal choice of the semi-orthogonal matrix $A \in \mathbb{R}^{t\times n}$ is unique up to an orthogonal transformation $O \in \mathbb{R}^{r\times r}$.