Question: Suppose $X, Y \in \mathbb{R}^n$, $X \sim \mathcal{N}(0, I_n)$, $Y \sim \mathcal{N}(0, \Sigma)$, $A \in \mathbb{R}^{r \times n}$, then we have $X_{\mathrm{A}} = AX \sim \mathcal{N}(0, AA^T)$ and $Y_A \sim \mathcal{N}(0, A\Sigma A^T)$. Let $Z = A\Sigma A^T$.
Find A to maximize the following:

$$\min\left\{1, \sqrt{\sum_{i=1}^{r} \beta_i^2}\right\}$$

in which $\beta_1, \beta_2, \dots, \beta_r$ are eigenvalues of $(AA^T)^{-1}Z - I_r$.

Solution:

To maximize $\min\left\{1, \sqrt{\sum_{i=1}^{r} \beta_i^2}\right\}$, it is sufficient (but not necessary) to maximize $\sum_{i=1}^{r} \beta_i^2$. Assume Σ is positive definite and $\mathrm{rank}(A) = r$, then we can deduce that both AA^T and $T = A\Sigma A^T$ are positive definite.

In addition, $(AA^T)^{-1}Z$ have the same spectrum (eigenvalues) as $(AA^T)^{-\frac{1}{2}}Z(AA^T)^{-\frac{1}{2}}$, so $(AA^T)^{-1}Z - I_r$ have the same spectrum (eigenvalues) as $(AA^T)^{-\frac{1}{2}}Z(AA^T)^{-\frac{1}{2}} - I_r$.
By SVD decomposition, we have:

$$A^T = P\begin{pmatrix} M \\ 0 \end{pmatrix}Q = (P_1, P_2)\begin{pmatrix} M \\ 0 \end{pmatrix}Q$$

in which $P \in \mathbb{R}^{n \times n}$, $Q \in \mathbb{R}^{r \times r}$ are orthogonal matrices, $P_1 \in \mathbb{R}^{n \times r}$ and $P_2 \in \mathbb{R}^{n \times (n-r)}$ are submatrices of P, $M \in \mathbb{R}^{r \times r}$ is a diagonal matrix, and $0 \in \mathbb{R}^{(n-r) \times r}$.

So we have:

$$AA^{\mathrm{T}} = Q^T(\mathrm{M} \quad 0)\begin{pmatrix} P_1^T \\ P_2^T \end{pmatrix}(P_1, P_2)\begin{pmatrix} M \\ 0 \end{pmatrix}Q = Q^T\mathrm{M}^2Q$$

So $(AA^T)^{-\frac{1}{2}} = Q^T\mathrm{M}^{-1}Q$, $Z = A\Sigma A^T = Q^{\mathrm{T}}(\mathrm{M} \quad 0)P^T\Sigma P\begin{pmatrix} \mathrm{M} \\ 0 \end{pmatrix}Q$. Therefore, we have:

$$(AA^T)^{-\frac{1}{2}}Z(AA^T)^{-\frac{1}{2}} = Q^TP^T\Sigma PQ = Q^TP_1^T\Sigma P_1Q$$

So, to study $(AA^T)^{-1}Z - I_r$'s eigenvalues is to study $(AA^T)^{-\frac{1}{2}}Z(AA^T)^{-\frac{1}{2}} - I_r$'s eigenvalues, which is to study $Q^TP_1^T\Sigma P_1Q - I_r$'s eigenvalues. Since Q is orthogonal, $Q^TP_1^T\Sigma P_1Q - I_r$'s eigenvalues are the same as $P_1^T\Sigma P_1 - I_r$'s eigenvalues.

Suppose by eigenvalue decomposition we have $\Sigma = V\Lambda V^{\mathrm{T}}$ in which $V \in \mathbb{R}^{n \times n}$ is orthogonal and $\Lambda \in \mathbb{R}^{n \times n}$ is diagonal, then our aim is to study $P_1^TV\Lambda V^TP_1 - I_r$'s eigenvalues.

Suppose the diagonal matrix Λ has n eigenvalues, $\lambda_1 \geq \cdots \geq \lambda_n$, and $P_1^TV\Lambda V^TP_1$ has r eigenvalues, $\gamma_1 \geq \cdots \geq \gamma_r$. It is obvious that $\beta_i = \gamma_i - 1$. So, to maximize $\sum_{i=1}^{r} \beta_i^2$ is to maximize $\sum_{i=1}^{r}(\gamma_i - 1)^2$.

Because $P \in \mathbb{R}^{n \times n}$ is orthogonal, we have $P_1^TP_1 = I_r$. Since V is orthogonal, we have $P_1^{\mathrm{T}}VV^TP_1 = I_r$, which means that P_1^TV is semi-orthogonal. So, by Cauchy's interlace theorem, we have:

$$\lambda_1 \geq \gamma_1 \geq \lambda_{n-r+1}$$
$$\lambda_2 \geq \gamma_2 \geq \lambda_{n-r+2}$$

$$\cdots$$

$$\lambda_{\mathrm{r}} \geq \gamma_r \geq \lambda_n$$

So, similar to the discussion of K-L divergence and Hellinger distance, we have the following 3 cases.

Conclusion:

Case 1: If $\lambda_1 \geq \cdots \geq \lambda_n \geq 1$, then $\sum_{i=1}^{r}(\gamma_i - 1)^2$ yields its maximum by taking $\gamma_1 = \lambda_1$, $\gamma_2 = \lambda_2$,…,$\gamma_r = \lambda_r$. This is possible, since we can construct *A* so that the SVD decomposition of *A* is $A = Q^T(M \quad 0)P^T$, $P = (P_1 \quad P_2)$ and $P_1 = (v_1 \quad v_2 \quad \cdots \quad v_r)$.

Case 2: If $1 \geq \lambda_1 \geq \cdots \geq \lambda_n$, then $\sum_{i=1}^{r}(\gamma_i - 1)^2$ yields its maximum by taking $\gamma_1 = \lambda_{n-r+1}$, $\gamma_2 = \lambda_{n-r+2}$,…,$\gamma_r = \lambda_n$. This is possible as long as $P_1 = (v_{n-r+1} \quad v_{n-r+2} \quad \cdots \quad v_n)$.

Case 3: Suppose that there are both eigenvalues that are greater than 1 and eigenvalues that are less than 1. Notice that each term $(\gamma_i - 1)^2$ yields its maximum either at the upper bound of γ_i, λ_i, or at the lower bound of γ_i, λ_{n-r+i}. So we can decide each γ_i's value by comparison of $|\lambda_i - 1|$ and $|\lambda_{n-r+i} - 1|$. Then we can construct P_1 and *A* accordingly.