HOGSVD and orthogonolization of $\boldsymbol{U}$

Saket Choudhary
saketkc@gmail.com

November 2019

GSVD

Consider two matrices $\boldsymbol{X}_{n_1 \times p}$ and $\boldsymbol{Y}_{n_2 \times p}$ represented as $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3 \dots \boldsymbol{x}_n]^T$ where $\boldsymbol{x}_{i \in [1 \dots n]} \in \mathbb{R}^{1 \times p}$ represents a row vector and similarly $\boldsymbol{Y}_{n_2 \times p} = [\boldsymbol{y}_1, \boldsymbol{y}_2 \dots \boldsymbol{y}_n]^T$ where $\boldsymbol{y}_{i \in [1 \dots n]} \in \mathbb{R}^{1 \times p}$. We consider $n_1 > p$ and $n_2 > p$.

GSVD looks for the following decomposition:

$$\boldsymbol{X}_{n_1 \times p} = \boldsymbol{U}_{n_1 \times p} \boldsymbol{A}_{p \times p} \boldsymbol{T}^T_{p \times p}$$
$$\boldsymbol{Y}_{n_2 \times p} = \boldsymbol{V}_{n_2 \times p} \boldsymbol{B}_{p \times p} \boldsymbol{T}^T_{p \times p}$$

where $\boldsymbol{A}, \boldsymbol{B}$ are both diagonal elements with diagoal elements $(a_1, a_2 \dots a_p)$ and $(b_1, b_2 \dots b_p)$ respectively satisfying $a_j^2 + b_j^2 = 1 \; \forall j \in [1, \dots p]$

$\boldsymbol{A}$ and $\boldsymbol{B}$ are column orthonormal, *i.e.* $\boldsymbol{a}_i \boldsymbol{a}_i^T = 1$. $\boldsymbol{T}^T$ relates the two matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$.

The rows of matrix $\boldsymbol{T}^T$, *i.e.* the columns of $\boldsymbol{T} := [\boldsymbol{t}_1, \boldsymbol{t}_2 \dots \boldsymbol{t}_p]$ can be thought of as an expression of p latent factors or "eigengenes", representative of both the datasets simultaneously. The relative contribution of these factors is captured by the elements in $\boldsymbol{A}$ and $\boldsymbol{B}$ matrix. It is measure as the relative contribution of each eigen gene to each dataset given the ratio of the square of corresponding entry in the matrix $\boldsymbol{A}$ or $\boldsymbol{B}$ with the sum scaled with the norm of the corresponding eigenvector.

$$R_j^X = \frac{a_j^2 ||\boldsymbol{t}_j||}{\sum_{l=1}^p a_l^2 ||\boldsymbol{t}_l||} \text{ and } R_j^Y = \frac{b_j^2 ||\boldsymbol{t}_j||}{\sum_{l=1}^p b_l^2 ||\boldsymbol{t}_l||}, j = 1, 2 \dots p$$

Once we are able to find matrices $\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{A}, \boldsymbol{B}$ and $\boldsymbol{T}$, we can find a projection of the n_1 and n_2 genes of $\boldsymbol{X}, \boldsymbol{Y}$ respectively onto the p eigengenes:

$$\boldsymbol{P}^{\mathrm{X}}_{n_1 \times p} = \boldsymbol{U}_{n_1 \times p} \boldsymbol{T}_{p \times p} \text{ and } \boldsymbol{P}^{\mathrm{Y}}_{n_1 \times p} = \boldsymbol{V}_{n_2 \times p} \boldsymbol{T}_{p \times p}$$

HOGSVD

Consider N matrices $\boldsymbol{X}_1, \boldsymbol{X}_2, \dots \boldsymbol{X}_N$ such that they have same number of columns, but possibly different number of rows. Higher order GSVD performs the following decomposition:

$$
\begin{aligned}
\boldsymbol{X}_1^{(n_1\times m)} &= \boldsymbol{U}_1\boldsymbol{\Sigma}_1\boldsymbol{V}^T\\
\boldsymbol{X}_2^{(n_2\times m)} &= \boldsymbol{U}_2\boldsymbol{\Sigma}_2\boldsymbol{V}^T\\
&\vdots\\
\boldsymbol{X}_N^{(n_N\times m)} &= \boldsymbol{U}_N\boldsymbol{\Sigma}_N\boldsymbol{V}^T
\end{aligned}
$$

In order to solve for $\boldsymbol{V}$, we make use of the following relations:

$$
\begin{aligned}
\boldsymbol{A}_i &= \boldsymbol{X}_i^T\boldsymbol{X}_i,\\
S_{ij} &= \frac{1}{2}\left(A_iA_j^{-1} + A_jA_i^{-1}\right),\\
\boldsymbol{S} &\equiv \frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{j>i}\left(A_iA_j^{-1} + A_jA_i^{-1}\right)\\
\boldsymbol{SV} &= \boldsymbol{V}\Lambda,\\
V &\equiv (v_1,\dots v_m), \Lambda = \text{diag}(\lambda_i).
\end{aligned}
$$

Thus, $\boldsymbol{V}$ can be obtained by eigen decomposition of matrix $\boldsymbol{S}$. Having obtained $\boldsymbol{S}$, we now solve for matrices $\boldsymbol{Z}_i$ to obtain $\boldsymbol{U}_i$:

$$
\begin{aligned}
\boldsymbol{V}\boldsymbol{Z}_i^{\mathrm{T}} &= \boldsymbol{X}_i^T\\
\boldsymbol{Z}_i &\equiv (z_{i1},\dots,z_{im}), i\in[1,N]\\
\boldsymbol{\Sigma}_{ik} &= ||z_{ik}||,\\
\boldsymbol{\Sigma}_i &= \text{diag}(\boldsymbol{\Sigma}_{ik}),\\
\boldsymbol{Z}_i &= \boldsymbol{\Sigma}_i\boldsymbol{U}_i
\end{aligned}
$$

Orthogonalizing U

The HOGSVD algorithm, in general, will not preserve the orthogonality of the $\boldsymbol{U}_i$ matrix. However, orthogonality might be a desired property for some of the applications. In order to make $\boldsymbol{U}$ columnwise orthonormal, we use the decompisition: $\boldsymbol{U}_{\text{ortho}} = \boldsymbol{U}(\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}})^{-1/2}$. The following theorem proves that such decomposition gives the "nearest" orthogonal matrix to $\boldsymbol{X}$:

Theorem 1. *Given an $n\times m$ ($n\geq m$) matrix $\boldsymbol{A}_{n\times m}$ the nearest possible orthogonal matrix $\boldsymbol{Q}$ to $\boldsymbol{A}$ is given by $\boldsymbol{Q} = \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-\frac{1}{2}}$. $\boldsymbol{Q}$ minimises both the frobenius $||\boldsymbol{A}-\boldsymbol{Q}||_{\mathrm{F}}$ and spectral norm $||\boldsymbol{A}-\boldsymbol{Q}||_2$*

Proof:

Minimizing Frobenius Norm:

In order to find an orthonormal $\boldsymbol{Q}$ such that the frobenius norm $||\boldsymbol{A}-\boldsymbol{Q}||_{\mathrm{F}}$ is minimized, we solve the following optimization problem:

$$
\min ||\boldsymbol{A}-\boldsymbol{Q}||_{\mathrm{F}} \text{ s.t. } \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} = \boldsymbol{I}
$$

$$
\begin{aligned}
||\boldsymbol{A}-\boldsymbol{Q}||_{\mathrm{F}}^2 &= \sum_{i=1}^{n}\sum_{j=1}^{m}|A_{ij}-Q_{ij}|^2\\
&= \mathrm{tr}\left((\boldsymbol{A}-\boldsymbol{Q})^{\mathrm{T}}(\boldsymbol{A}-\boldsymbol{Q})\right)
\end{aligned}
$$

Now,

$$\begin{aligned}\min ||\boldsymbol{A}-\boldsymbol{Q}||_{\mathrm{F}}^2 &= \min\ \mathrm{tr}\left((\boldsymbol{A}-\boldsymbol{Q})^{\mathrm{T}}(\boldsymbol{A}-\boldsymbol{Q})\right)\\ &= \min\ \mathrm{tr}\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}+\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}-2\boldsymbol{A}^{\mathrm{T}}Q\right)\\ &= \max\ \mathrm{tr}\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Q}\right)\end{aligned}$$

Now consider the singular value decomposition of $\boldsymbol{A}$ as $\boldsymbol{A}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$.

$$\begin{aligned}\max\ \mathrm{tr}\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Q}\right) &= \max\ \mathrm{tr}\left(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{A}\right)\\ &= \max\ \mathrm{tr}\left(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}\right)\\ &= \max\ \mathrm{tr}\left((\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{U}\boldsymbol{\Sigma})\boldsymbol{V}^{\mathrm{T}}\right)\\ &= \max\ \mathrm{tr}\left((\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{U}\boldsymbol{\Sigma})\boldsymbol{V}^{\mathrm{T}}\right)\\ &= \max\ \mathrm{tr}\left(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{U}\boldsymbol{\Sigma}\right) \quad (\because \mathrm{tr}(\boldsymbol{A}\boldsymbol{B})=\mathrm{tr}(\boldsymbol{B}\boldsymbol{A}))\end{aligned}$$

Let $\boldsymbol{Z}=\boldsymbol{V}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{U}$ where $\boldsymbol{V},\boldsymbol{Q},\boldsymbol{U}$ are all orthonormal matrices and so is $\boldsymbol{Z}$, *i.e* $\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}=\boldsymbol{I}$.

$$\begin{aligned}\mathrm{tr}\left(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{U}\boldsymbol{\Sigma}\right) &= \mathrm{tr}(\boldsymbol{Z}\boldsymbol{\Sigma})\\ &= \sum_{i=1}^{n} Z_{ii}\boldsymbol{\Sigma}_{ii}\\ &\leq \sum_{i=1}^{n}\boldsymbol{\Sigma}_{ii} \qquad (\boldsymbol{\Sigma}_{ii}\geq 0 \text{ and } |Z_{ii}|\leq 1)\end{aligned}$$

Thus, $\boldsymbol{Z}=\boldsymbol{I}$ and $\boldsymbol{Q}=\boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}$.

Minimizing spectral Norm:

In order to minimize the spectral norm $||\boldsymbol{A}-\boldsymbol{Q}||_2=\max_{||x||=1}||\boldsymbol{A}-\boldsymbol{Q}x||$, we again rely on the the singular value decomposition of $\boldsymbol{A}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$. Also, $||\boldsymbol{A}||_2\leq||\boldsymbol{A}||_{\mathrm{F}}$.

Then, $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{V}=\boldsymbol{\Sigma},\boldsymbol{U}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{V}=\boldsymbol{R},\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}=\boldsymbol{I}$.

$$\begin{aligned}\min ||\boldsymbol{A}-\boldsymbol{Q}||_2^2 &= \min ||\boldsymbol{U}^T(\boldsymbol{A}-\boldsymbol{Q})\boldsymbol{V}^2\\ &= \min ||\boldsymbol{\Sigma}-\boldsymbol{R}||_2^2\\ &= \min ||\boldsymbol{\Sigma}-\boldsymbol{I}+\boldsymbol{I}-\boldsymbol{R}||_2^2\\ &= \min ||\boldsymbol{\Sigma}-\boldsymbol{I}-(\boldsymbol{R}-\boldsymbol{I})\,||_2^2\\ &= \min \max_{||\boldsymbol{x}||=1} ||\left(\boldsymbol{\Sigma}-\boldsymbol{I}-(\boldsymbol{R}-\boldsymbol{I})\right)\boldsymbol{x}^{\mathrm{T}}||\end{aligned}$$

Consider $||\boldsymbol{\Sigma}-\boldsymbol{R}||$:

$$\begin{aligned}(\boldsymbol{\Sigma}-\boldsymbol{R})^{\mathrm{T}}(\boldsymbol{\Sigma}-\boldsymbol{R}) &= \left(\boldsymbol{\Sigma}-\boldsymbol{I}-(\boldsymbol{I}-\boldsymbol{R})\right)^{\mathrm{T}}\left(\boldsymbol{\Sigma}-\boldsymbol{I}-(\boldsymbol{I}-\boldsymbol{R})\right)\\ &= (\boldsymbol{\Sigma}-\boldsymbol{I})^{\mathrm{T}}(\boldsymbol{\Sigma}-\boldsymbol{I})-(\boldsymbol{\Sigma}-\boldsymbol{I})^{\mathrm{T}}(\boldsymbol{R}-\boldsymbol{I})-(\boldsymbol{R}-\boldsymbol{I})^{\mathrm{T}}(\boldsymbol{\Sigma}-\boldsymbol{I})+(\boldsymbol{R}-\boldsymbol{I})^{\mathrm{T}}(\boldsymbol{R}-\boldsymbol{I})\\ &= (\boldsymbol{\Sigma}-\boldsymbol{I})^{\mathrm{T}}(\boldsymbol{\Sigma}-\boldsymbol{I})+\boldsymbol{\Sigma}(\boldsymbol{I}-\boldsymbol{R})+(\boldsymbol{I}-\boldsymbol{R}^{\mathrm{T}})\boldsymbol{\Sigma}+(\boldsymbol{R}-\boldsymbol{I})^{\mathrm{T}}(\boldsymbol{R}-\boldsymbol{I})\end{aligned}$$

Without loss of generality we consider $\boldsymbol{x} = \boldsymbol{e}$ such that $\boldsymbol{e} = (1, 0, 0, \ldots, 0)$.

$$\min \max_{||\boldsymbol{x}||=1} || \left(\boldsymbol{\Sigma} - \boldsymbol{I} - (\boldsymbol{R} - \boldsymbol{I})\right) \boldsymbol{x}^{\mathrm{T}} ||^2 = \boldsymbol{e}(\boldsymbol{\Sigma} - \boldsymbol{R})^{\mathrm{T}}(\boldsymbol{\Sigma} - \boldsymbol{R})\boldsymbol{e}^{\mathrm{T}}$$

$$\geq \boldsymbol{e}(\boldsymbol{\Sigma} - \boldsymbol{I})^T(\boldsymbol{\Sigma} - \boldsymbol{I})\boldsymbol{e}^{\mathrm{T}} \qquad (\because \text{diag}(\boldsymbol{I} - \boldsymbol{R}) \geq 0 \text{ and } \text{diag}(\boldsymbol{I} - \boldsymbol{R}^{\mathrm{T}}) \geq 0)$$

The minima for the ultimate inequality is obtained when $\boldsymbol{R} = \boldsymbol{I}$ implying $\boldsymbol{Q} = \boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}$.

$Q = \boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}$ is equivalent to taking the SVD of $\boldsymbol{A}$ and setting all its singular values to be 1. Alternatively, $\boldsymbol{Q} = \boldsymbol{A}\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\right)^{-\frac{1}{2}}$