Model Averaging and Dimension Selection for the Singular Value Decomposition

Seoyoung Shin

March 26, 2024

Table of Contents

Various kinds of social relation models

Social Relations Model(SRM)

$$y_{i,j} = \mu + a_i + b_j + \epsilon_{i,j}$$

Social Relations Regression Model(SRRM)

$$y_{i,j} = \boldsymbol{\beta}^T \boldsymbol{x}_{i,j} + \mu + a_i + b_j + \epsilon_{i,j}$$

Additive and Multiplicative Effects Model(AMEM)

$$y_{i,j} = \boldsymbol{\beta}^T \boldsymbol{x}_{i,j} + \boldsymbol{u}_i^T \boldsymbol{v}_j + \mu + a_i + b_j + \epsilon_{i,j}$$

- Third order dependence patterns such as transitivity and clustering can be represented with multiplicative random effects models.

- Multiplicative effects matrix $\mathbf{UV}^T$ provides a reduced-rank representation of $\mathbf{Y}$.

Construction of the Fixed-Rank Model

Model Averaging and Dimension Selection for the SVD

- Modeling data matrix $\mathbf{Y}$ as equal to a reduced-rank mean matrix $\mathbf{M}$ plus Gaussian noise $\mathbf{E}$, and estimating $\mathbf{M}$ after deciding on its rank.

- SVD allows interpretation of multiplicative model based on row, column factors.

Model Construction

$\mathbf{Y}_{m\times n} = \mathbf{M} + \mathbf{E} = \mathbf{UDV}' + \mathbf{E}, \quad y_{i,j} = \mathbf{u}_j'\mathbf{D}\mathbf{v}_j + e_{i,j},$ where

$\mathbf{M}$: Reduced rank(K) mean matrix,

$\mathbf{U}$: $m \times K$ orthonormal matrix, $\mathbf{V}$: $n \times K$ orthonormal matrix,

$\mathbf{E}$: Gaussian Noise $\sim \mathrm{N}(0, \sigma^2)$

Bayesian procedure would provide a mapping from a prior distribution $p(\mathbf{U}, \mathbf{D}, \mathbf{V}, \sigma^2)$ to a posterior distribution $p(\mathbf{U}, \mathbf{D}, \mathbf{V}, \sigma^2 | \mathbf{Y})$.

Recall Singular Value Decomposition(SVD)

Every $m \times n (m \geq n)$ matrix $\mathbf{M}$ has a representation of the form $\mathbf{M} = \mathbf{UDV}'$, where

$\mathbf{U}$: $m \times n$ matrix with orthonormal columns, $\mathbf{V}$: $n \times n$ matrix with orthonormal columns,

$\mathbf{D}$: $n \times n$ diagonal matrix, with diagonal elements $\{d_1, \cdots, d_n\}$.

Prior Distribution of U and V

$\mathbf{U}$: $m \times K$ orthonormal matrix, the set of such matrices is called the Stiefel manifold $\mathcal{V}_{K,m}$

* Stiefel manifold : The set of ordered orthonormal k-tuples of vectors in $\mathbb{R}^n$

ex) $n = 3$

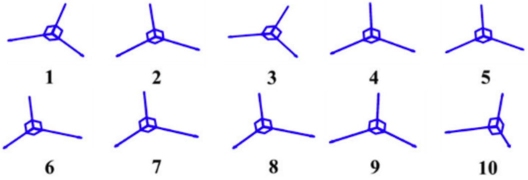


Figure: Samples from vMF with various κ values

Prior Distribution of U and V

We can generate random samples of $\mathbf{U}$ uniformly on $\mathcal{V}_{K,m}$ as follows:

1. Sample $\mathbf{u}_1$ uniformly from unit m sphere

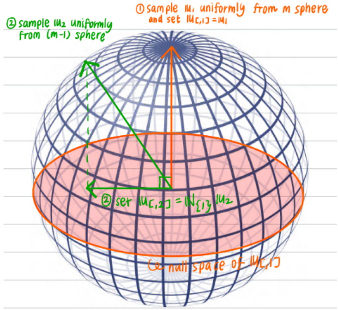
   and $\mathbf{U}_{[,1]} := \mathbf{u}_1$
2. Sample $\mathbf{u}_2$ uniformly from unit $(m-1)$ sphere

   and $\mathbf{U}_{[,2]} := \mathbf{N}_{\{1\}}\mathbf{u}_2$

$\vdots$

K. Sample $\mathbf{u}_K$ uniformly from unit $(m-K+1)$ sphere

   and $\mathbf{U}_{[,K]} := \mathbf{N}_{\{1,\dots,K-1\}}\mathbf{u}_K$



<ex. when K = 3>

* Notations

$\mathbf{U}_{[,A]}$: the columns of $\mathbf{U}$ that correspond to a subset of column labels $A \subset \{1,\dots,K\}$

$\mathbf{N}_A$: any $m \times (m - |A|)$ matrix whose columns form orthonormal basis for the null space of $U_{[,A]}$

Prior Distribution of D, E

- Prior of **D**

$$d_j \sim \mathrm{N}(\mu, \tfrac{1}{\psi}) \longleftarrow \left\{ \begin{array}{c} \mu \sim \mathrm{N}(\mu_0, v_0^2) \\ \psi \sim \mathrm{Gamma}\left(\frac{\eta_0}{2}, \frac{\eta_0 \tau_0^2}{2}\right) \end{array} \right\}$$

- Prior of **E**

$$E \overset{iid}{\sim} \mathrm{N}\left(0, \frac{1}{\phi}\right) \longleftarrow \phi \sim \mathrm{Gamma}\left(\frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2}\right)$$

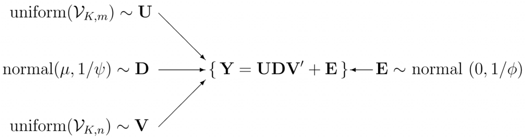


Figure: A Graphical Representation of the Model

Gibbs sampling for Fixed Rank Model

Likelihood function :

$$f(\mathbf{Y}|\mathbf{U},\mathbf{D},\mathbf{V},\phi) = \left(\frac{\phi}{2\pi}\right)^{\frac{mn}{2}} \exp\left(-\frac{1}{2}\phi\|\mathbf{Y}-\mathbf{UDV}'\|^2\right) \quad \because \mathbf{Y}-\mathbf{UDV}' \sim \mathrm{N}\left(0,\frac{1}{\phi}\right)$$

$$= \left(\frac{\phi}{2\pi}\right)^{\frac{mn}{2}} \exp\left(-\frac{1}{2}\phi\|\mathbf{E}_{-j}\|^2 + \phi d_j \mathbf{U}'_{[,j]}\mathbf{E}_{-j}\mathbf{V}_{[,j]} - \frac{1}{2}\phi d_j^2\right)$$

where $\mathbf{E}_{-j} = \mathbf{Y} - \mathbf{U}_{[,-j]}\mathbf{D}_{[-j,-j]}\mathbf{V}'_{[,-j]}$

Full conditional distribution :

$$P(\mathbf{u}_j|\mathbf{Y},\mathbf{U},\mathbf{D},\mathbf{V},\phi) \propto P(\mathbf{Y}|\mathbf{U},\mathbf{D},\mathbf{V},\phi) \times P(\mathbf{u}_j)$$

$$\propto \exp\left(\mathbf{u}'_j\phi d_j \mathbf{N}^{u'}_{-j}\mathbf{E}_{-j}\mathbf{V}_{[,j]}\right) \sim \mathrm{vMF}\left(\phi d_j \mathbf{N}^{u'}_{-j}\mathbf{E}_{-j}\mathbf{V}_{[,j]}\right)$$

$$P(\mathbf{v}_j|\mathbf{Y},\mathbf{U},\mathbf{D},\mathbf{V},\phi) \propto P(\mathbf{Y}|\mathbf{U},\mathbf{D},\mathbf{V},\phi) \times P(\mathbf{v}_j)$$

$$\propto \exp\left(\mathbf{v}'_j\phi d_j \mathbf{U}'_{[,j]}\mathbf{E}_{-j}\mathbf{N}^{v}_{\{-j\}}\right) \sim \mathrm{vMF}\left(\phi d_j \mathbf{U}'_{[,j]}\mathbf{E}_{-j}\mathbf{N}^{v}_{\{-j\}}\right)$$

Gibbs sampling for Fixed-Rank Model

Full conditional distribution :

$$P(d_j|\mathbf{Y},\mathbf{U},\mathbf{D}_{[-j,-j]},\mathbf{V},\phi,\mu,\psi) \propto P(\mathbf{Y}|\mathbf{U},\mathbf{D},\mathbf{V},\phi) \times P(d_j|\mu,\psi^{-1})$$
$$\sim \mathrm{N}\left(\frac{\mathbf{U}'_{[,j]}\mathbf{E}_{-j}\mathbf{V}_{[,j]}\phi + \mu\psi}{\phi+\psi}, \frac{1}{\phi+\psi}\right)$$

$$P(\phi|\mathbf{Y},\mathbf{U},\mathbf{D},\mathbf{V}) \propto P(\mathbf{Y}|\mathbf{U},\mathbf{D},\mathbf{V},\phi) \times P(\phi)$$
$$\sim \mathrm{Gamma}\left(\frac{\nu_0+mn}{2}, \frac{\nu_0\sigma_0^2 + \|\mathbf{Y}-\mathbf{UDV}'\|^2}{2}\right)$$

$$P(\mu|\mathbf{D},\psi) \propto P(\mathbf{D}|\mu,\psi^{-1}) \times P(\mu) \sim \mathrm{N}\left(\frac{\psi\sum_{j=1}^K d_j + \mu_0/v_0^2}{\psi K + 1/v_0^2}, \frac{1}{\psi K + 1/v_0^2}\right)$$

$$P(\psi|\mathbf{D},\mu) \propto P(\mathbf{D}|\mu,\psi^{-1}) \times P(\psi) \sim \mathrm{Gamma}\left(\frac{\eta_0+K}{2}, \frac{\eta_0\tau_0^2 + \sum_{j=1}^K (d_j-\mu)^2}{2}\right)$$

von Mises-Fisher distribution

von Mises-Fisher distribution

The probability density function of the von Mises-Fisher distribution for the random p-dimensional unit vector **x** is given by :

$$f_p(\mathbf{x}; \boldsymbol{\mu}, \kappa) = C_p(\kappa) \exp\left(\kappa \boldsymbol{\mu}^T \mathbf{x}\right)$$

where $\kappa \geq 0$, $\|\boldsymbol{\mu}\| = 1$ and $C_p(\kappa) = \frac{\kappa^{p/2-1}}{(2\pi)^{p/2} I_{p/2-1}(\kappa)}$,
where I_v denotes the modified Bessel function of the first kind at order v

i.e. $\mathbf{X} \in \mathbb{R}^p \sim \text{vMF}(\boldsymbol{\mu}, \kappa) \quad \Rightarrow \quad f_p(\mathbf{x}; \boldsymbol{\mu}, \kappa) = C_p(\kappa) \exp\left(\kappa \boldsymbol{\mu}^T \mathbf{x}\right)$

ex) $\boldsymbol{\mu} = (0,\ 0,\ 1) \in S^3,\ \kappa \in \{1,\ 5,\ 10,\ 20\}$

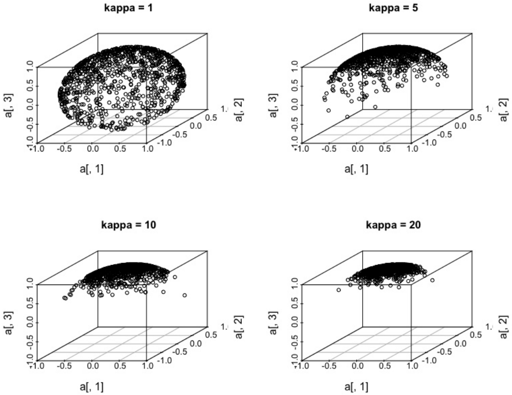


Figure: Samples from vMF with various κ values

Gibbs Sampling Procedure for Fixed-Rank Model

For $j \in \{1, \ldots, K\}$,

- Sample $(\mathbf{U}_{[,j]} \mid \mathbf{Y}, \mathbf{U}_{[,-j]}, \mathbf{D}, \mathbf{V}, \phi) = \mathbf{N}^{u}_{\{-j\}}\mathbf{u}_j$,

  where $\mathbf{u}_j \sim \text{vMF}(\phi d_j \mathbf{N}^{u'}_{-j} \mathbf{E}_{-j} \mathbf{V}_{[,j]})$

- Sample $(\mathbf{V}_{[,j]} \mid \mathbf{Y}, \mathbf{U}, \mathbf{D}, \mathbf{V}_{[,-j]}, \phi) = \mathbf{N}^{v}_{\{-j\}}\mathbf{v}_j$,

  where $\mathbf{v}_j \sim \text{vMF}(\phi d_j \mathbf{U}'_{[,j]} \mathbf{E}_{-j} \mathbf{N}^{v}_{\{-j\}})$

- Sample $(d_j \mid \mathbf{Y}, \mathbf{U}, \mathbf{D}_{[-j,-j]}, \mathbf{V}, \phi, \mu, \psi) \sim \text{N}\left(\frac{\mathbf{U}'_{[,j]}\mathbf{E}_{-j}\mathbf{V}_{[,j]}\phi + \mu\psi}{\phi + \psi}, \frac{1}{\phi+\psi}\right)$

Sample $(\phi \mid \mathbf{Y}, \mathbf{U}, \mathbf{D}, \mathbf{V}) \sim \text{Gamma}\left(\frac{\nu_0 + mn}{2}, \frac{\nu_0\sigma_0^2 + \|\mathbf{Y} - \mathbf{UDV}'\|^2}{2}\right)$

Sample $(\mu \mid \mathbf{D}, \psi) \sim \text{N}\left(\frac{\psi\sum_{j=1}^{K} d_j + \mu_0/v_0^2}{\psi K + 1/v_0^2}, \frac{1}{\psi K + 1/v_0^2}\right)$

Sample $(\psi \mid \mathbf{D}, \mu) \sim \text{Gamma}\left(\frac{\eta_0 + K}{2}, \frac{\eta_0\tau_0^2 + \sum_{j=1}^{K}(d_j - \mu)^2}{2}\right)$

Implementation scheme 1 : Orthonormal basis for the Null space

QR Decomposition

For a matrix $\mathbf{A}_{m\times n}(m \geq n)$, $r(\mathbf{A}) = n$, $\mathbf{A}$ can be decomposed as

$$\mathbf{A} = \mathbf{QR} = [\mathbf{Q}_1 \ \mathbf{Q}_2] \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{0} \end{bmatrix} = \mathbf{Q}_1\mathbf{R}_1,$$ where

$\mathbf{Q}_1$ is an $n \times n$ orthogonal matrix,

$\mathbf{Q}_2$ is an $(m-n) \times n$ orthogonal matrix,

$\mathbf{R}_1$ is an $n \times n$ upper triangle matrix

Proposition1 In the QR decomposition of the matrix $\mathbf{A}$, the vectors $\boldsymbol{q}_{n+1}, \ldots, \boldsymbol{q}_m$ are an orthonormal basis for the null space of $\mathbf{A}$.

Implementation scheme 2 : Sampling from vMF

1. $m \leftarrow$ length of $\boldsymbol{\mu}$
2. Sample the t marginal distribution
   (i) $b \leftarrow \frac{(m-1)}{2\kappa+\sqrt{4\kappa^2+(m-1)^2}}$
   (ii) $x_0 \leftarrow \frac{1-b}{1+b}$
   (iii) $c \leftarrow \kappa x_0 + (m-1)\log(1-x_0^2)$
   (iv) for $i = 1$ to N

   while $\kappa \mathrm{W} + (m-1)\log(1 - x_0 \mathrm{W}) - c < \log(\mathrm{U})$
   - Z $\leftarrow$ sample from Beta$\left(\frac{p-1}{2},\ \frac{p-1}{2}\right)$
   - U $\leftarrow$ sample from Unif$[0,\ 1]$
   - W $\leftarrow \frac{1-(1+b)Z}{1-(1-b)Z}$
3. $\xi \leftarrow$ sample uniform distribution on the sphere $\mathbb{S}^{m-2}$
4. Samples(:, 0)$\leftarrow t$, Samples(:, 1:)$\leftarrow \sqrt{1-t^2}$
5. Rotate each sample to the desired mean direction.

Derive Likelihood function

$$
\begin{aligned}
P(\mathbf{Y}|\mathbf{U},\mathbf{D},\mathbf{V},\phi) &= \left[\frac{1}{2\pi(1/\phi)}\right]\exp\left(-\frac{1}{2(1/\phi)}\|\mathbf{Y}-\mathbf{UDV}'\|^2\right) \qquad \because \mathbf{Y}-\mathbf{UDV}' \sim \mathrm{N}\left(0,\frac{1}{\phi}\right)\\
&= \left[\frac{1}{2\pi(1/\phi)}\right]\exp\left(-\frac{1}{2}\phi\left\{\|\mathbf{E}_{-j}\|^2 - 2d_j\mathbf{U}'_{[,j]}\mathbf{E}_{-j}\mathbf{V}_{[,j]} + d_j^2\right\}\right) \quad \cdots (*)\\
&= \left(\frac{\phi}{2\pi}\right)^{\frac{mn}{2}}\exp\left(-\frac{1}{2}\phi\|\mathbf{E}_{-j}\|^2 + \phi d_j\mathbf{U}'_{[,j]}\mathbf{E}_{-j}\mathbf{V}_{[,j]} - \frac{1}{2}\phi d_j^2\right)
\end{aligned}
$$

where $\mathbf{E}_{-j} = \mathbf{Y} - \mathbf{U}_{[,-j]}\mathbf{D}_{[-j,-j]}\mathbf{V}'_{[,-j]}$

$$
\begin{aligned}
(*)\ \|\mathbf{Y}-\mathbf{UDV}'\|^2 &= \|\mathbf{Y} - \mathbf{U}_{[,j]}\mathbf{D}_{[j,j]}\mathbf{V}'_{[,j]} - d_j\mathbf{U}_{[,-j]}\mathbf{V}'_{[,-j]}\|^2\\
&= \|\mathbf{E}_{-j} - d_j\mathbf{U}_{[,j]}\mathbf{V}'_{[,j]}\|^2\\
&= \|\mathbf{E}_{-j}\|^2 - 2d_j\mathbf{U}'_{[,j]}\mathbf{E}_{-j}\mathbf{V}_{[,j]} + \|d_j\mathbf{U}_{[,j]}\mathbf{V}'_{[,j]}\|^2\\
&= \|\mathbf{E}_{-j}\|^2 - 2d_j\mathbf{U}'_{[,j]}\mathbf{E}_{-j}\mathbf{V}_{[,j]} + d_j^2
\end{aligned}
$$

Derive conditional distribution of $\mathbf{u}_j$

$$
\begin{aligned}
P(\mathbf{u}_j|\mathbf{Y},\mathbf{U},\mathbf{D},\mathbf{V},\phi) &\propto P(\mathbf{Y}|\mathbf{U},\mathbf{D},\mathbf{V},\phi) \times P(\mathbf{u}_j) \\
&\propto P(\mathbf{Y}|\mathbf{U},\mathbf{D},\mathbf{V},\phi) \\
&\propto \exp\left(\phi d_j \mathbf{U}'_{[,j]} \mathbf{E}_{-j} \mathbf{V}_{[,j]}\right) \\
&= \exp\left(\phi d_j \left(\mathbf{N^u}_{-j} \mathbf{u}_j\right)' \mathbf{E}_{-j} \mathbf{V}_{[,j]}\right) \\
&= \exp\left(\phi d_j \mathbf{u}'_j \mathbf{N^u}_{-j}{}' \mathbf{E}_{-j} \mathbf{V}_{[,j]}\right) \\
&= \exp\left(\mathbf{u}'_j \phi d_j \mathbf{N^u}_{-j}{}' \mathbf{E}_{-j} \mathbf{V}_{[,j]}\right) \sim \mathbf{vMF}\left(\phi d_j \mathbf{N^u}_{-j}{}' \mathbf{E}_{-j} \mathbf{V}_{[,j]}\right)
\end{aligned}
$$

Derive conditional distribution of $\mathbf{d}_j$

$$
\begin{aligned}
P(d_j|\mathbf{Y},\mathbf{U},\mathbf{D},\mathbf{V},\phi,\mu,\psi) &\propto P(\mathbf{Y}|\mathbf{U},\mathbf{D},\mathbf{V},\phi) \times P(d_j|\mu,\psi^{-1}) \\
&\propto \exp\left(\phi d_j \mathbf{U}'_{[,j]}\mathbf{E}_{-j}\mathbf{V}_{[,j]} - \frac{1}{2}\phi d_j^2\right) \times \exp - \left\{\frac{\psi}{2}(d_j-\mu)^2\right\} \\
&= \exp\left\{-\frac{1}{2}(\phi+\psi)d_j^2 + (\phi\mathbf{U}'_{[,j]}\mathbf{E}_{-j}\mathbf{V}_{[,j]} + \mu\psi)d_j - \frac{\psi\mu^2}{2}\right\} \\
&\propto \exp\left[-\frac{1}{2}(\phi+\psi)\left\{d_j - \frac{\phi\mathbf{U}'_{[,j]}\mathbf{E}_{-j}\mathbf{V}_{[,j]} + \mu\psi}{\phi+\psi}\right\}^2\right] \\
&\sim \mathbf{N}\left(\frac{\mathbf{U}'_{[,j]}\mathbf{E}_{-j}\mathbf{V}_{[,j]}\phi + \mu\psi}{\phi+\psi}, \frac{1}{\phi+\psi}\right)
\end{aligned}
$$

Appendix B : Calculation for Full conditional distribution

Derive conditional distribution of ϕ

$$
\begin{aligned}
P(\phi|\mathbf{Y},\mathbf{U},\mathbf{D},\mathbf{V}) &\propto P(\phi|\mathbf{Y},\mathbf{U},\mathbf{D},\mathbf{V},\phi) \times P(\phi) \\
&\propto (\phi)^{\frac{mn}{2}} \exp\left[-\frac{1}{2(1/\phi)}\|\mathbf{Y}-\mathbf{UDV}'\|^2\right] \times \phi^{\frac{\nu_0}{2}-1} \exp\left(-\frac{\nu_0\sigma_0^2}{2}\phi\right) \\
&\propto (\phi)^{\frac{mn+\nu_0}{2}-1} exp\left[-\frac{\phi}{2}\left(\|\mathbf{Y}-\mathbf{UDV}'\|^2+\nu_0\sigma_0^2\right)\right] \\
&\sim \text{Gamma}\left(\frac{\nu_0+mn}{2}, \frac{\nu_0\sigma_0^2+\|\mathbf{Y}-\mathbf{UDV}'\|^2}{2}\right)
\end{aligned}
$$

Appendix B : Calculation for Full conditional distribution

Derive conditional distribution of μ

$$
\begin{aligned}
P(\mu|\mathbf{D},\psi) &\propto P(\mathbf{D}|\mu,\psi^{-1}) \times P(\mu) \\
&\propto \mathrm{N}(\mathbf{D}|\mu,\psi^{-1}) \times \mathrm{N}(\mu|\mu_0, v_0^2) \\
&\sim \mathrm{N}\left(\frac{\psi\sum_{j=1}^{K} d_j + \mu_0/v_0^2}{\psi K + 1/v_0^2},\ \frac{1}{\psi K + 1/v_0^2}\right)
\end{aligned}
$$

Derive conditional distribution of ψ

$$
\begin{aligned}
P(\psi|\mathbf{D},\mu) &\propto P(\mathbf{D}|\mu,\psi^{-1}) \times P(\psi) \\
&\propto \prod_{j=1}^{K} (\psi)^{\frac{1}{2}} \exp\left\{-\frac{1}{2\psi^{-1}}(d_j-\mu)^2\right\} \times (\psi)^{\frac{\eta_0}{2}-1} \exp\left(-\frac{\eta_0\tau_0^2}{2}\psi\right) \\
&= \psi^{\frac{K+\eta_0}{2}-1} \exp\left\{-\frac{\psi}{2}\left(\sum_{j=1}^{K}(d_j-\mu)^2 + \eta_0\tau_0^2\right)\right\} \\
&\sim \text{Gamma}\left(\frac{\eta_0+K}{2},\ \frac{\eta_0\tau_0^2+\sum_{j=1}^{K}(d_j-\mu)^2}{2}\right)
\end{aligned}
$$

Proposition 1 In the QR decomposition of the matrix $\mathbf{A}$, the vectors $\boldsymbol{q}_{n+1}, \ldots, \boldsymbol{q}_m$ are an orthonormal basis for the null space of $\mathbf{A}$.

pf) Note that $\mathbf{A} = \mathbf{QR}$ and the columns of $\mathbf{Q} = [\mathbf{Q}_1\ \mathbf{Q}_2]$ is an orthonormal vectors.

We want to show that $\mathbf{A}^T\mathbf{Q}_2 = \mathbf{0}$.

Let $\left\{\mathbf{q}_{n+1}, \cdots, \mathbf{q}_m\right\}$ be the columns of $\mathbf{Q}_2$.

Then $\mathbf{A}^T\mathbf{q}_j = (\mathbf{QR})^T\mathbf{q}_j = \mathbf{R}^T\mathbf{Q}^T\mathbf{q}_j = \mathbf{R}^T\mathbf{e}_j$.

Note that $\mathbf{R}^T\mathbf{e}_j$ is j^{th} column in $\mathbf{R}^T$ and so same as j^{th} row in $\mathbf{R}$.

Since $\mathbf{R}$ is an upper triangle matrix that has zero vectors in $(n+1)$ to m rows,

$\mathbf{A}^T\mathbf{q}_j = 0, \quad \forall j \in \{n+1, \cdots, m\}$.

Reference

- Wood, A. T. A., Simulation of the von Mises Fisher distribution, Communications in Statistics - Simulation and Computation, 23 , 157-164 (1994).

Thank you for your attention!